Genome-wide transcriptome analysis of cassava challenged with Ugandan cassava brown streak virus (UCBSV)

by

Teddy Amuge

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

In the Faculty of Natural & Agricultural Sciences
Department of Plant and Soil Sciences
University of Pretoria
Pretoria

October 2019

Under the supervision of Prof Dave K Berger
and co-supervision of Dr Morag E Ferguson
and Prof Alexander A Myburg

Declaration

The research presented in this thesis is my own work and does not contain results that have been generated by other persons, except where due reference and acknowledgement have been made. The work presented in this thesis has not been previously submitted for any degree at any institution.

Teddy Amuge

February 2019

Thesis abstract

Cassava brown streak disease (CBSD) remains the second most destructive disease of cassava in Africa, causing up to 100% yield loss in farmers' fields. Two Potyviruses, *cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV) cause CBSD. The aim of this study was to determine the mechanisms underlying resistance to UCBSV in cassava variety, Namikonga. Thus, time-series transcriptome analysis (RNAseq) was conducted, including UCBSV susceptible variety, Albert. Simultaneously, accumulation of UCBSV in both varieties were assessed. Up to 55 differentially expressed defence genes selected from the transcriptome data were measured in three additional varieties (Kiroba, Mkombozi and NDL06/132), plus Namikonga and Albert, using QuantStudio. Leaf tissues were sampled at six and 12 hours after grafting (hag), 1, 2, 3, 4, 5, 6, 7, 8, 45, 46, 47, 48, 49, 50, 51, 52, 53 and 54 days after grafting (dag). From RNAseq, highest numbers of differentially expressed genes (DEGs) where identified at 2 (3887 DEGs) and 5 dag (4911 DEGs) in Namikonga. These DEGs included defence genes; gene ontology terms for 'translational elongation', 'translation factor activity', 'ribosomal subunit' and 'phosphorelay signal transduction', which were overrepresented only in Namikonga. From QuantStudio, defence genes were up regulated at 1 and 51 dag in Kiroba. Based on UCBSV load and root necrosis, Namikonga and Kiroba were confirmed as resistant to CBSD, Mkombozi as tolerant while Albert and NDL06/132 are susceptible. Therefore, time-period from inoculation, level of virus load and root necrosis are key markers of resistance/susceptibility in cassava infected with UCBSV.

Preface

Cassava brown streak disease (CBSD) is still the most economically significant disease of cassava in Africa, after cassava mosaic disease (CMD). No known cassava variety is immune to the disease; rather, resistance has been identified in some locally grown varieties. These resistant varieties are being adopted in CBSD-prone cassava growing areas, and breeders are also using them to generate new breeding lines that have disease resistance together with other culinary traits. Cassava is staple to >200 million people in Africa. It's a major source of dietary energy in West and East Africa, but CBSD has greatly affected production in East Africa (and by extension, West Africa where the disease is threatening to reach). CBSD is caused by two virus species: *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV), genus *Ipomovirus* within the family *Potyviridae*. The whitefly vectors (*Bemisia tabaci* and *Aleurodicus disperses*) vector naturally spreads both viruses in cassava fields where the disease is causing widespread losses. Control of CBSD, like other virus diseases, requires the deployment of disease resistant varieties. These varieties are not readily available, except through breeding. The work presented in this thesis is aimed at identifying CBSD-resistant varieties, and establishing the mechanism(s) underlying resistance in these varieties. The hypothesis is that the resistance to CBSD in these varieties is heritable. Future work can harness this trait to improve other cassava varieties for resistance to CBSD.

This study is based on collaborative research conducted at the Biosciences eastern and central Africa (BecA), International Institute of Tropical Agriculture (IITA)-Nairobi, University of California, Berkeley, CA, USA, and the University of Pretoria, Faculty of Natural & Agricultural Sciences, where the student is registered. The study is presented in five stand-alone chapters as shown below.

Chapter 1 is a literature review of the latest knowledge on cassava and cassava production in the face of CBSD. The mechanism of resistance to plant pathogens (pattern- and effector-triggered immunity), types and case studies of plant resistance proteins (R-proteins) are presented. Latest literature on other plant defence genes is presented. The application of these R-protein genes for defence against major plant pathogens is covered. Differential expression of R-proteins against Potyviruses is included. Transcriptome profiles of R-proteins in susceptible, tolerant and resistant hosts are discussed. Finally, the significance of cassava and CBSD in food security is discussed.

Chapter 2 covers fold change of UCBSV titre over a time course. The methods for identification of CBSD-causing viruses namely: CBSV and UCBSV were validated by samples from FERA. Sanger sequences of PCR products from UCBSV-inoculated samples matched sequences of NCBI's isolate LT560277-1, a UCBSV isolate sampled from Kampala, Uganda. In both the CBSD susceptible variety, Albert and CBSD resistant variety, Namikonga, UCBSV was first detected at six days after grafting (6 dag) in upper leaves. As expected, UCBSV detection was higher at later time points (45 dag through 54 dag) in UCBSV-inoculated samples of both varieties. At most time points, UCBSV was detected in upper leaves (UL) compared to lower leaves (LL). Virus titre was higher in Albert (highest titre was 2.1 fold at 5 dag) compared

to Namikonga (titre was 0.2 fold at 5 dag), which supports the hypothesis that Namikonga's defence functions by limiting virus multiplication and distribution.

Chapter 3 is based on the analysis of the transcriptome of Albert and Namikonga at eight time points including 6 hag, 1, 2, 5, 8, 45 and 54 dag. Over these seven time points, Namikonga had more differentially expressed genes (DEGs) (10,028 DEGs) compared to Albert (688 DEGs). The highest difference was at 2 dag and 5 dag where Namikonga had the highest number of DEGs compared to Albert. The DEGs identified in Namikonga at 2 dag and 5 dag include genes for growth, development and those for plant-pathogen defence. The defence genes encode LRR (Leucine-Rich Repeats), NB-ARC (Nucleotide binding Domain shared between human APAF-1 (Apoptotic Protease-Activating Factor 1), some plant R proteins, and CED-4 (*Caenorhabditus elegans* Dead protein 4)) in plants and NACHT (NAIP, CIITA, HETE and TP1), LEA (Late Embryogenesis Abundant protein), PR (Pathogenesis-related proteins), HSP (Heat Shock Proteins), chaperons and transcription factors (NAC, WRKY, NmrA, GATA and GRAS). A gene ontology (GO)-term enrichment of Namikonga's DEGs identified at 2 dag and 5 dag showed that GO-terms associated with elongation factors, ribosomal subunit and phosphorelay signal transduction were enriched. These enriched GO-terms contain notable plant defence proteins (including elongation factors and histidine kinases, which function in defence signalling).

Chapter 4 covers validation of 55 defence genes, which were differentially expressed in Namikonga's transcriptome. A time series analysis of two experiments: Experiment 1 and Experiment 2 were conducted with two (Albert and Namikonga) and five (Kiroba, Mkombozi and NDL06/132 in addition to Albert and Namikonga) cassava varieties from which leaves were sampled at UL and LL positions relative to the point of graft-inoculation on each plant. Based on manifestation of CBSD symptoms on leaves and storage roots, Namikonga and Kiroba are confirmed as resistant to UCBSV, Mkombozi as tolerant while NDL06/132 and Albert are susceptible. As earlier discovered in Chapter 3, defence in Namikonga is time bound and, is more active in tender leaves. Both Namikonga and Kiroba encode similar defence genes but in Kiroba, defence is expressed at early (1 dag) and late (51 dag) time points in older leaves yet Namikonga's defence genes are only expressed at early time points (3 dag, 5 dag and 7 dag) in younger leaves.

In **Chapter 5**, summaries of major results from all three research chapters are presented. The results are discussed in line with other studies where similar (and sometimes contradicting) results were obtained. Limitations of this study are acknowledged and recommendations are made for future studies.

As a result of the work in this thesis, this paper (Amuge et al., 2017)[1] below has been published in a peer-reviewed journal.

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

The following conference proceedings were based on this PhD work:

TEDDY AMUGE, M. E. Ferguson, A. A. Myburg, M. S. Katari, J. Harvey, S. Prochnik, J. Bredeson, S. Rounsley, D. K. Berger. Candidate genes conferring tolerance to cassava brown streak disease in cassava (*Manihot esculenta* Crantz). Plant and Animal Genome (PAG) conference, January 11th – 15th, 2014, San Diego, California, USA. (Poster presentation)

TEDDY AMUGE, Goldman S., M. S. Katari, A. A. Myburg, D. K. Berger and M. E. Ferguson. Time-series network of differentially expressed genes for cassava resistant to cassava brown streak disease (CBSD) World Congress on Root and Tuber Crops January 18th -22nd, 2016, Nanning, China. (Oral presentation)

Acknowledgement

Thank you, ALMIGHTY GOD, for this exceptional opportunity. Without YOU, I am nothing.

I sincerely thank my family for their faith in me, and for standing with me even when I had nothing left to hope for. From the beginning of this research to date, I have spent very limited time with my mother and siblings. I even missed it when Eric became unwell; I hadn't known that his time was running out. He left us, we all lost, a great loss. Rest in peace Atoto.

Unto Maa, the strongest, greatest mother in the world, thank you. Thank you for bringing us up almost entirely alone. I can't still understand, yet I know that as a widow, you suffered much for us. Thank you. May you continue living and enjoying the beauty in life.

I thank my siblings so much. Denis, Andrew, Paul, Beatrice and George. In you, I have the best family ever!

I give special thanks to my husband, Wilberforce Oduor. You have been a pillar, especially during the writing of these publications. Thank you for loving me when I was broken and almost hopeless. Your moral, spiritual and financial support is priceless. May you enjoy the coming days in our home, days without thesis or manuscript deadlines. Am deeply grateful for Berakah Joy. A peaceful baby she is, and so I got the time I needed to write this thesis.

To all my friends, thank you so much for your love and encouragement.

I have had the best academic supervisors: Prof. Dave Berger, Prof. Zander Myburg and Dr. Morag Ferguson. Thank you so much for the mentorship, guidance and support.

This study is successful because of the wonderful collaborators: Prof. Manpreet S. Katari, Dr. Steve Rounsley, Dr. Kathryn Denby, Dr. Simon Prochnik and their respective research teams. Thank you so much. Prof. Manpreet patiently trained me through the bioinformatics analysis process, and the output of that work is the core of this thesis. Thank you, Manny.

To all whose names are not mentioned here, your contribution is much appreciated. Thank you so much.

Dedication

UNTO JEHOVAH YAHWEH, THE SOVEREIGN LORD

Table of Contents

List of Figures

List of Tables

List of Supplementary Data

List of Electronic Supplementary Files

Supplementary Data S1. Number of UCBSV reads in each sample

Supplementary Data S2. Number of raw reads per sample

Supplementary Data S3. Number of mapped and unmapped RNAseq reads

Supplementary Data S4. HTseq counts of genes from mapped reads

Supplementary Data S5. DESeq_output_all_RNAseq_timepoints

Chapter 1

Literature review:

Plant defence genes with potential roles in resistance to cassava brown streak disease (CBSD)

T. Amuge, M. E. Ferguson, A. A. Myburg and D. K. Berger

Author Contributions

T.A.: Reviewed literature and drafted the chapter based on topics suggested by supervisors.

D.K.B., A.A.M. and **M.E.F** suggested topics to be reviewed and supervised the work.

1.1 Abstract

This review is divided into four parts: part 1 introduces cassava, its significance in food security and the threat of cassava brown streak disease (CBSD) to cassava production. Characteristics of CBSD and the causal agents are also discussed. Part 2 covers selected defence and resistance genes with NBS-LRR domain, their structural and functional features, which determine resistance to several pathogens. Part 3 is based on other defence genes lacking NBS-LRR domains, like the late embryogenesis abundant (LEA) proteins, pathogenesis-related (PR) proteins and transcription factors involved in defence (WRKY, NAC, GATA, GRAS and NmrA). The role of heat shock proteins and chaperones in defence is also reviewed. The last part covers application of these defence genes, with case studies of plant-potyvirus interactions. These genes have a huge potential for use to develop CBSD resistant cassava varieties with known, reliable and durable mechanisms of resistance. The possibility for their use in cassava against CBSD is proposed. This review is motivated by the discovery of quantitative resistance to CBSD in some cassava genotypes following the need for CBSD-resistant cassava varieties in Africa.

1.2 Introduction

1.2.1 Cassava: origin, evolution and importance

Cassava originates from a single wild *Manihot* species called *M. esculenta* ssp. *flabellifolia* (Pohl) Ciferri[1,2]. Botanical, geographical, agricultural and molecular-based phylogenetic studies altogether show that not only is *M. esculenta* ssp. *flabellifolia* (Pohl) Ciferri the single ancestor of cassava, but that the crop originates from the Amazonian basin of present day Brazil[2-4], contrary to earlier notions that cassava originated from multiple ancestors through introgression[5].

Cassava hybridizes naturally, even with wild progenitors in different tropical and sub-tropical environments[6,7]. Interspecific hybridisation (for speciation), polyploidisation (for fertility restoration) and apomixes (for maintenance/perpetuation of sterile hybrids) are key factors in the evolution of *Manihot* species[1]. Cassava seeds germinate well under high temperatures, including seeds that have been buried for years[6]. This characteristic behaviour supports the idea that cassava is naturally adapted to tropical environments, often characterised by wildfires.

Under relatively poor growth conditions, cassavas' yield is higher compared to other root crops[8]. Cassava is particularly drought tolerant, giving good yields in conditions where most cereal crops and potatoes would not survive[9]. These abilities explain the adoption of the crop in mainland Africa and Asia after its introduction by explorers and traders[10].

Most farmers use vegetative propagation, but seed germination is also used in research systems. Some farmers allow cassava plants derived from seed to continue growing in their fields (if these plantlets have desirable traits)[6]. This practice is common among Asian and African farmers. The starchy roots of cassava are used for food, feed, and industrial raw materials. Its leaves are consumed as a vegetable (for human consumption) and are used to make silage for animal feed. The stems are primarily used for crop propagation and for construction materials in rural settings.

Cassava is one of the 20 most important crops globally (FAOSTAT, 2012). Worldwide, >700 million people consume cassava on a regular basis. It is a major staple in Latin America, the Caribbean and sub-Saharan Africa, notably Nigeria, Democratic Republic of Congo, Ghana, Angola and East African countries (Uganda, Kenya and Tanzania)[8]. In East Africa alone, >280 million people rely on cassava as a staple (FAOSTAT, 2009). In 2012, 269 million metric tonnes were produced, more than half of this was in Africa (FAOSTAT, 2012).

Pests and diseases are the major challenge across all cassava growing areas. Cassava mosaic disease (CMD) and cassava brown streak disease (CBSD) are the two major diseases that cause loss in cassava production[11-13]. Although CBSD occurs in only in eastern Africa, it poses a serious threat to West Africa because its rate of spread.

So far, resistance to CMD is conferred by two loci mapped as polygenic recessive (*CMD1*) and monogenic dominant loci (*CMD2* and *CMD3*)[14-16]. Recently, a genome-wide association (GWAS) analysis of 6128 African cassava breeding lines identified a major resistance loci, *CMD2* on Chromosome 8[17]. An electronic PCR located using the same population showed that *CMD3*[15] was only 5 Mb away from *CMD2* loci, still on chromosome 8. Other loci, including those associated with susceptibility alleles were identified on other chromosomes, a sign that resistance to CMD is more multifaceted than known so far. Thus, these loci remain the only known and primary source of resistance to CMD in all CMD-resistant varieties of cassava. They are the basis for breeding CMD resistant cassava varieties in Africa[18].

Two whitefly species: *Bemisia tabaci*[19-22] and *Aleurodicus disperses*[23,24] transmit cassava mosaic geminiviruses (CMGs) that cause CMD and cassava brown streak viruses (CBSVs) that cause cassava brown streak disease (CBSD). Apart from transmitting these major viruses (CMGs and CBSVs), super-abundant populations of whiteflies themselves are shown to inflict significant losses in cassava[25,26]. A draft genome of *B. tabaci* and population genetics analysis shows six whitefly populations in Africa that spread both CMD and CBSD[27,28].

CBSD, caused by CBSVs was first reported in the coastal areas of Tanzania in the 1930s[22], and later in Lake Malawi in the 1940s[29]. Initially, the disease was thought to be limited in its occurrence and spread along the coastal Indian ocean covering parts of Kenya [19,30], Malawi[31], Mozambique [32] and Tanzania[33,34]. To date, CBSD has since been reported in many highland areas including Uganda[35], Burundi[36], Democratic Republic of Congo (DRC)[37], Kenya[38] and Mozambique[39]. Rwanda, which borders Uganda, Burundi, Tanzania and DRC has CBSD, but no official reports have been published yet. CBSD-like symptoms have also been observed in Madagascar although no diagnostic tests were carried out to confirm presence of the virus. But, CBSD has been identified in Mayotte Island, which lies between Mozambique and Madagascar[40]. In all the areas affected, use of CBSD resistant cassava varieties remains the most viable method of control. CMD and CBSD caused widespread losses in cassava production, but until recently, there was limited focus on CBSD because of the misconception that it was localised to the lowland areas[19,29,41-43]. The disease gained interest after the explosion of its spread into both lowland and highland areas[35-37,39,40]. Given the rate of spread in East Africa[44] and the explosion in CBSD vector populations[12], there is likelihood that CBSD will reach West Africa, the largest producer and consumer of cassava.

1.2.2 CBSD and causal viruses

CBSD is caused by two distinct species: *Cassava Brown Streak Virus* (CBSV) and *Ugandan Cassava Brown Streak Virus* (UCBSV)[45,46]. Both virus species have filamentous virus-like particles, 650 nm long with pinwheel inclusions in the virus cells[45]. They are members of genus *Ipomovirus*, family *Potyviridae*[45,46]. Both genomes are 9070 nucleotides (nt) and 9008 nt long, each yielding polypeptides of 2902 aa and 2916 aa for UCBSV and CBSV respectively[46-48].

The polyprotein translates to P1, P3, 6K1, CI, 6K2, VPg, Nla-Pro, Nlb and CP respectively. The PIPO protein is produced by a frame shift of the viral open reading frame (ORF). Protein HC-Pro is not produced

by both genomes but a new protein; Ham1-like was identified in both genome[46-48]. The new protein (Ham1) is inserted between Nlb and CP proteins. Among Ipomoviruses, Ham1 is novel to CBSVs and has only been identified in Euphobia ring spot virus (EuRSV), which infects members of the *Euphobiaceae* family[49]. The CBSVs are the first members of Potyviridae that contain a single P1 protein, lack HCPro and have Ham1 protein[46,48]. The introduction of Ham1, single instead of double P1 and deletion of HC-Pro in genomes of CBSVs is evidence that both genomes are undergoing rapid evolution[50]. Indeed, sequences of 12 new genomes of CBSV and UCBSV showed recombination events at 3' end, 5' end and mid-sections of CBSV and UCBSV genomes[51]. Within-species evolution is marked by homologous recombination in the Ham1 coding region, purifying- and positive-selection in the coat protein coding regions[50]. The single P1 silences suppression, a role normally played by HC-Pro in other potyviruses[47].

In sweet potato mild mottle virus (SPMMV), another member of the genus *Ipomovirus*, the role of HC-Pro is partly played by P1 protein. Considering (1) the genetic distance between SPMMV and CBSVs, (2) the roles of P1 and, (3) the complete absence of HC-Pro in CBSVs, it is expected that the CBSVs evolved recently by re-organizing its' genome to eliminate HC-Pro, transferring the roles of HC-Pro protein to P1 and introducing a new protein, HAM1h to balance the negative effects of evolution in both genomes[47,48]. Like other RNA viruses, there is no evidence of heterologous recombination in CBSVs[50].

Dual infection with both CBSV and UCBSV is common in field samples[45,46,52], possibly because of random movement of the vector of CBSVs. The viruses have been identified in wild species of cassava including *Manihot glaziovii* Müll.Arg. growing in Tanzania[52] and *Manihot carthaginensis* subsp. *glaziovii* (Müell-Arg.) Allem. growing in Mozambique[39]. In fact, the wild species growing in Mozambique reportedly showed typical symptoms of CBSD (score 4 on the 1-5 scale described by Hillocks et al.[33]), and tested positive for both CBSV and UCBSV. Two non-cassava perennial wild plants: velvet-fruited zanha (*Zanha africana* (Radlk.) Exell.) and camel bush (*Trichodesma zeylanicum* (Burm.f.) R.Br.) which were growing near cassava fields showed CBSD-like symptoms and tested positive for CBSV. Mechanical transmission of CBSV from cassava to the non-cassava host *T. zeylanicum* was successful, further confirming that *T. zeylanicum* is another host of CBSV[39].

In cassava, both virus species are transmitted by two known whitefly species: *B. tabaci*[19-22] and *A. disperses*[23,24]. Based on disease symptoms, *B. tabaci* transmits viruses more effectively than *A. disperses*[23]. Mechanical transmission in cassava is feasible with CBSV[53], but grafting is more preferred for study-purposes using single or both species. Cleft, bud[54] and top-wedge graft methods are equally effective and preferred[55,56].

Upon transmission into its host, immunohistochemical analysis of the species CBSV showed that it localises to the adaxial and abaxial epidermal cells positioned along the leaf blades and central veins of cassava[57]. Three tissue types including leaves sampled at six different positions along the length of the cassava plants, stems and storage roots, were taken from ten cassava varieties and analysed for presence and count of virus particles under field conditions. Results showed that virus titre varied in all three types of tissues and

varieties tested[58]. In some varieties, highest titre was observed is leaves, others in the stem and others in storage roots. The lowest virus titre was observed in the young leaves closest to the plant tip, followed by the oldest senescing leaves near the soil level while leaves positioned midway of the plant had the highest virus titre. Considering tissues types, all ten varieties showed highest titre in their storage roots compared to stems and leaves irrespective of single (CBSV or UCBSV) and mixed (CBSV and UCBSV) infections.

Typically, infected plants develop symptoms 3-10 weeks after inoculation irrespective of the method of transmission or virus species[20,23]. Symptoms include chlorosis on veins of leaves, brown streaks on stems, brown necrosis on starchy roots, rot of storage roots caused by bacterial infection, and plant dieback[41]. Symptoms vary between varieties, environments and among plants of different ages[19,41,59].

Breeding for resistance to CBSD is done at country level, having farmers' preference for respective country agro-ecologies[29,60,61]. Most 'improved' varieties with farmer-preferred traits do not have 100% resistance, as these 'improved' varieties were products of germplasm introduced from Tanzania (with various clones, including Namikonga, from the Amani breeding program, as the source of resistance) and wild cassava varieties from Brazil (with unknown response to CBSD, at the time), introduced in Uganda in 2013[62], which germplasm are not immune, but highly resistant to CBSD. In Uganda, multi-locational field trials of advanced breeding lines for their response to the two major virus diseases CBSD and CMD has confirmed that three varieties are adapted to the main agro-ecologies and are resistant to CMD, but only tolerant to CBSD[63]. All these breeding profiles could change because of the recent identification of absolute resistance to infection by CBSVs in seven varieties from South America[64]. The super varieties from South America have shown immunity to both CBSV (strain CBSV-Tan70) and UCBSV (strain UCBSV-Ke125). Considering that signatures of defence genes have been identified in the cassava genome[65-67], and sources discovered in South American varieties, there is hope for cassava farmers.

1.2.3 Identification of CBSVs

Visual identification based on shoot and root symptoms

Visual features of cassava shoot and root is used to determine CBSD-free and infected plants. Above ground, CBSD-affected cassava plants develop chlorotic feathery patterns on either sides of leaf veins[41]. Vein chlorosis is more visible on old mature leaves, with younger actively growing leaves remaining symptomless[19,35]. Characteristic brown streaks on stems and petioles of lower, and mid parts of the plant is typical of CBSD[41]. Under severe infection, CBSD causes plant die back. Highest damage of affected plants occurs in the starchy roots, the part that serves as food and feed. Roots of CBSD infected cassava develop brown, cork-like necrotic patterns covering up to 100% of the root. Constricted roots with rough coarse root- peel is a classic symptom of CBSD. Affected roots often rot due to opportunistic bacterial infection and are unfit for consumption, even to animals[68]. Severity of leaf and root symptoms are separately classified following a 1-5 scale where 1=no visible symptoms on either leaf or root and 5=severe symptoms[69]. Detection of visible symptoms is not robust for virus identification as symptoms vary within and between isolates, virus species and host[70]. In some cases, cassava plants infected with CBSVs do not show leaf symptoms at all[41,71]. This calls for a defined isolate-based method of identification.

Enzyme linked immuno-sorbent assay

An enzyme linked immuno-sorbent assay (ELISA) that detects both virus species has been developed by Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) (www.dsmz.de). The ELISA kits and antibodies were developed and marketed by DSMZ. The robustness of ELISA kits has not been demonstrated, and poor adoption has marred the benefits of this system for identifying CBSVs. Rather, for isolate detection, the more specific, sensitive approach of sequence-based PCR identification was developed and is widely applied in many research laboratories and national programs.

PCR- based methods of identification

For PCR methods, reverse-transcriptase polymerase chain reaction (RT-PCR) primers are used to transcribe viral RNA to cDNA, which is then amplified by PCR. The first RT-PCR primers were designed from partial coat protein sequence of CBSV[72]. Although specific to CBSV, these initial primers are still in use to-date[71,73-75]. Emergence of a new species (UCBSV) prompted the design of new, species-specific primers from partial and full-length sequences of viral genomes, because the first primers were less sensitive to other virus isolates and species[45,73]. However, very low virus titre in single or mixed infections[76] limits use of newly developed species-specific RT-PCR primers, hence the adoption of a quantitative method which allows one to identify and measure titre levels of the infecting virus.

Quantitative real time PCR (qPCR) is very sensitive to low count-isolates, measuring in real-time, the amplication during PCR, achieved using the standard curve[77] . This method has since been applied in the detection and measuring of low-titre counts of CBSVs[55,72,78]. Both Taqman and SYBR green chemistries have since been used to detect and quantify species-specific isolates[78]. The first set of housekeeping genes were optimized for use in qPCR identification of CBSVs[56]. These reference genes have since been applied in detection and quantification of CBSVs[55]. Besides qPCR, another assay of loop-mediated isothermal amplification (LAMP) assay was developed for use on-farm and in low-cost facilities[79]. This method is cheap, colour-based and convenient, but its robustness and sensitivity is yet to be proven.

1.2.4 Mechanisms of resistance to infection

Different mechanisms control immunity in plants and animals. Animals have specialized cells that control immunity to biotic and abiotic perturbations. Plants use two methods for innate immune response; pattern triggered immunity (PTI) and effector triggered immunity (ETI). In PTI, conserved pathogen-recognition receptors (PRRs) detect pathogen associated molecular patterns (PAMP). Most PRRs are receptor kinases with extracellular leucine rich repeats (LRR), similar to the mammalian Toll-like receptors[80]. ETI detects intracellular pathogen signals using kinase receptors, most of which have NBS-LRR (nucleotide-binding site leucine-rich repeat) in their structure[81]. These ETI receptors are highly polymorphic between plant species, detecting strain-specific pathogen effectors[82]. Plant NBS-LRR (in the class of R-proteins) detect structural conformational changes (indirect detection) or actual pathogen effectors (direct detection) within host cells[82].

Plant R-proteins are structurally and functionally similar to mammalian NOD (nucleotide-binding oligomerisation domain)-like receptors (NLRs), but NLR and R-proteins evolved independently, and from

different origins[83,84]. The functional similarity between R-proteins and NLRs, and their role in defence against pathogens is the basis for this review. The functional characteristics of plant R-proteins and their role in relation to other defence proteins is also reviewed. The comparison is based on cassava's resistance to a potyvirus (UCBSV) in which specific R-proteins and other defence proteins are expressed at specific time points, in defined magnitudes in specific varieties.

Basic module/structure of plant R-proteins versus animal NLR proteins

Plant R-genes and animal NLRs belong to the AAA+ (ATPases Associated with various cellular Activities) ATPase superfamily within the Signal Transduction ATPases with Numerous Domains (STAND) superfamily classification of AAA+ proteins[85-87]. Proteins of the STAND superfamily are functionally complex consisting of single and combined multi-domain proteins that function as regulatory switches, scaffold-forming units, sensors and signalling agents[86].

Structurally, STAND superfamily proteins have a tri-modular shape[88,89] consisting of the N-terminal region, a centrally located Nucleotide-Binding (NB) region and the C-terminus. At their N-termini, mammalian NLRs have Caspase-Activation and Recruitment Domains (CARD), Pyrin Domain (PD) or Baculovirus-*I*nhibitor-of-apoptosis Repeats (BIR)[88]. In plants, the N-terminus of an R protein is mainly of two types, the Toll-Interleukin 1 Receptor (TIR) type of N-termini, which forms the TNL class, and the coiled-coil (CC) type of N-termini, which forms the CNL class[90-92]. The N-terminus of plant R-proteins are involved in signalling and effector recognition[93,94].

The centrally located NB domain is termed NB-ARC (Nucleotide binding Domain shared between human APAF-1 (Apoptotic Protease-Activating Factor 1), some plant R proteins, and CED-4 (*Caenorhabditus elegans* Dead protein 4)) in plants and NACHT (NAIP, CIITA, HETE and TP1) in mammals. As named, these central domains coordinate binding of nucleotides[88].

Mammalian NLR proteins have a C-terminal region consisting of one or combinations of LRR, tetracopeptide, ankyrin and, tryptophan- and aspartic acid containing repeats. Plant R-proteins however, have leucine-rich repeat (LRR) domains at their C-termini. Some R-proteins may have armadillo or tetracopeptide repeats instead of LRR at their C-termini[84]. The C-terminal LRR domain determines specificity of effector recognition[95].

Structure and function of N-terminus (TIR/CC domain)

Amino acids spanning the TIR domains are conserved across Drosophila, humans and plants[96]. However, plants have a conserved unique subdomain within their TIR[97]. For R-proteins with the CC in their N-termini, two Arabidopsis R-proteins (Resistance to Pseudomonas Syringae5 (RPS5) and RPS2) showed 45% similarity in the sequences of their CC domains[98]. Beyond sequence similarities of the TIR-TIR and CC-CC domains of plant R-genes, functional roles of these proteins are highly dependent on specific protein conformations that mark "on" or "off" positions.

Fold

The TIR domain assumes a globular fold, which consists of alpha and beta sheets. In RPS4TIR, globular fold consists five-stranded parallel β-sheet (β-A and β-E) surrounded by five α-helical regions (a-A and a-E). The RPS4 protein associates with RRS1 for expression of resistance. Unlike RPS4, RRS1 contains only one helix in the α-D helical region[99]. A similar fold appearance is reported for the TIR domains of AtTIR[97] and L6[100]. Bernoux et al.,[100] observed similar fold arrangements as with AtTIR with bacterial and mammalian TIR domains.

For CC domain, the monomeric state shows an alpha helical form with an EDVID motif[94]. The EDVID motif is conserved in the CC domains of MLA10, Sr33 and Rx proteins[101]. In monomeric state, the MLA10 protein of barley has two long anti-parallel alpha helices linked by a short loop[102] while the monomeric state of Sr33 protein of wheat shows a bundle of four alpha helices[101], similar to the structure of the Rx protein of potato[103]. Taken together, the N-terminal region (CC or TIR) of both CNL and TNL proteins is important for formation of dimers (homo- /hetero-dimers)[99,101,103]. Dimerisation is a critical step for downstream expression of resistance in all known R-proteins and NLRs[104].

Dimerisation (homo and hetero) (also referred to as self-association)

When a protein interacts selectively and non-covalently with a domain within the same polypeptide, the phenomenon if termed self-association or homodimerisation. Where the interaction occurs between two different proteins, heterodimerisation is said to have occurred. In their regulation of resistance to pathogens, plant R-proteins undergo some form of self-association or heterodimerisation.

For *N. tabacum*, homodimerisation of the TIR domain of RPS4 was confirmed as critical for signalling cell death[99]. As seen with the NACHT L6 protein of flax, the TIR domain self-associates, a phenomenon governed by electrostatic interactions within the polymer[100]. In *N* gene of tobacco, the TIR domain provides the interaction site for oligomerisation[105]. The serine and histidine residues of RPS4 (Resistance to Pseudomonas Syringae 4) and TIR/TIR domain is required to form a functional complex that recognizes both AvrRps4 and PopP2 effectors in *N. tabacum* and Arabidopsis[99]. The TIR domain alone in the presence of EDS1 (enhanced disease susceptibility 1), SGT1 *(*suppressor of the G2 allele of skp1*)*, and HSP90 (heat shock protein 90) was sufficient to cause cell death[106].

In barley mildew A (MLA), a CNL protein, dimerisation is suggested to regulate interaction between MLA and WRKY transcription factors[93]. The CC domain harbours the dimerisation surface, which is essential and sufficient to cause cell death in *N. benthamiana*[102]. Indeed, the CC dimer functions as the minimum functional requirement for causing cell death in *N. benthamiana*[93]. A similar observation was made with recombinant MLA10, where, the CC domain alone was sufficient to cause cell death in *N. benthamiana*. Altogether, these observations demonstrate the critical role of dimerisation in the formation of a functional R-protein complex. How these interfaces operate to form dimers opens another avenue to appreciate these functional units and their roles in dimerisation.

Dimerisation interfaces

In MLA, homodimerisation of the CC domain involves large proportions of the alpha helix, which highly depend on residues located on α or β positions. This kind of interaction is common to the CC domains of other R-genes[93]. Dimerisation is strengthened by hydrophobic contacts between the α2a and the C-terminus of α1a and N-terminus of α2a of the other CC monomer[93]. Mutations in 17 amino acids lining the interior of the CC dimer of MLA10 did not abolish self-association except for residue F99E. Residue F99E is confirmed to harbour cell death signalling activation site, located in the CC domain of MLA10 protein[102].

When mapped to relative locations in the *N* protein of *N. tabacum*, *Arabidopsis thaliana's* TIR domain (AtTIR) mapped to two positions; the βC-loop within the N-terminus and αD3 adjacent to αC-helix[97]. In AtTIR, the α-D-helix (subdivided into αD1, αD2 and αD3) has a unique insertion site consisting 20 additional residues. Through mutagenesis, residues located at the αD3 were identified as crucial for the functioning of the TIR with regards to protein-protein interactions[97] A set of conserved hydrophobic residues from exon 2 of AtTIR aligned to similar regions within αE of mammalian Toll-like receptors (TLRs)[100]. Also, the flax L6 TIR/TIR structure revealed two interfaces; one made of αD1, αD3, αE, βE strands and the DE and EE loops (interface 1) and another formed by bonds between fewer amino acids (interface 2)[100]. In six out of eight amino acid residues of interphase 1, mutations inhibited self-association. This observation shows the significance of this interface 1 in self-association of L6 TIR domain. None of the mutations in interface 2 affected self-association, raising concerns whether or not it's a true interaction interface[100]. Conserved regions within L6 TIR and TIRs of tobacco *N* and Arabidopsis *RPS4* indicated the role of this region in self-association and signalling of L6 protein[100]. Deletion of the TIR domain from L6 had no effect on its recognition of AvrL567, showing that the TIR has no critical role in recognition[100].

For human NLRs, the CARD domains of NOD1/RIP2 (receptor-interacting protein 2) interact via the CARD interfaces 1, 2, 3 and 4 respectively[107,108]. Swapping of helices 1 while engaging helices 2 and 5 causes formation of CARD dimers. However, helix 2 is significant for both dimerisation and signalling. Alterations in residues E53K, D54K and E56K of helix 3 abolished NOD1 signalling. In the same way, mutation at helix 2 led to loss-of-function, suggesting significant role(s) of helix 2 in NOD1 signalling, through formation of functional oligomers[108].

Among conserved residues of mammalian NOD (namely Cys-39, Asn-43, Leu-44, Glu-53, Asp-54 and Glu-56), Cys-39 alone showed significance in dimerisation of NOD1 CARD. However, Glu-53, Asp-54 and Glu-56 are critical for interaction and signal transduction of RIP2. Structurally, Leu-44 forms part of the hydrophobic core as Asp-54 and Lys-78 are involved in formation of an internal salt bridge. Asn-43 was found conserved across species, bringing to light the role played by this residue in functioning of NOD1. Upon activation, both NOD1 and NOD2 partially localise to the plasma membrane, where they interact with RIP2 through CARD-CARD interactions[109]. Mutations in helices 2 and 3 do not affect translocation, but single alterations in helix 2 impair signalling, even though NOD1 and RIP2 remain unaffected. Undeniably, signalling impairment in helix 2 occurs due to alterations in the dimer-forming residues as shown by Asn-43. Type Ia interface in RIP2 controls RIP2-RIP2 type I interaction. Complex comparison demonstrated that

NOD1 residues form type IIIa interface as type IIIb are formed by RIP2. Therefore, NOD1 and RIP2 CARD interactions involve interface I and III, both of which locate to external surface of the dimer[108].

These observations demonstrate that major dimerisation interfaces within plant N-temini (CC and TIR) and animal NLR proteins have similar roles. So, dimerisation occurs upstream of the centrally located NB-ARC (or NATCH in mammalian NLR) domain and is critical for signalling.

Structure and function of plant NB-ARC (or mammalian NACHT)

Between the N-terminus and the C-terminus lies the centrally located plant NB-ARC (or mammalian NACHT) domain. Plant NB-ARC (Nucleotide-Binding adaptor shared by APAF-1, Resistance proteins and CED-4) is comparable to mammalian NACHT (NAIP (neuronal apoptosis inhibitor protein), C2TA (MHC class 2 transcription activator), HET-E (incompatibility locus protein from *Podospora anserina*) and TP1 (telomerase-associated protein))[110]. These two (NB-ARC and NACHT) domains are similar in architecture and show distribution of roles corresponding to subdivisions in their structure. For both mammalian NACHT and plant NB-ARC, the NB domain functions in nucleotide binding. Other subdomains including the four-helix bundle (ARC1 in plant), winged-helix fold (ARC2 in plants) and the helical bundle (ARC3) are part of the central domain structure. The NB, ARC1 and ARC2 subdomains are conserved in plants, but plants lack the ARC3 subdomain[89,104]. Plant NB, ARC1 and ARC2 subdomains form a nucleotide-binding pocket[89]. An *in-silico* structure of the NB-ARC domain was constructed based on the models of *RRS2* of Arabidopsis, *I2C1* of tomato (*Lycopersicon esculentum*), L6 of flax (*Linum usitatissimum*), *VMYR1* of urd-bean (*Vigna mungo*), Ced4 of nematode (*Caenorhabditis elegans*) and human Apaf-1 proteins. The proposed structure shows that ADP sits on a pocket in the P-loop of the NB-ARC domain using non-covalent bonds[111].

In a different study of two highly homologous plant immune receptors *Gpa2* (plant immune receptor which recognizes potato cyst nematode) and *Rx1* (plant immune receptor which recognizes potato virus X), *in-silico* models of *Gpa2*'s NB-ARC and *Rx1*'s LRR revealed a distinct pattern. In that pattern, NB, ARC1 and ARC2 form an ADP binding pocket, which appears as a compact globular structure. The N-terminus of the ARC2 region of *Gpa2* is significant for cooperation of NB-ARC with the LRR domain[112]. The N-terminal region of *Gpa2* and the NB subdomain interact with the C-terminal ARC2 subdomain via an acidic loop protruding from ARC2 subdomains of both *Gpa2* and *Rx1*. This loop is absent from the corresponding Apaf-1 structure[112]. Parts of the ARC2 subdomain and the first two N-terminal repeats of LRR have a regulatory role in *Gpa2* and *Rx1* respectively[112].

Transient expression of a construct lacking the NB-ARC domain in flax caused complete inhibition of self-association, further emphasizing the role of this region in negative auto-regulation of inactive *L6* [100]. In *N* protein, modifications at conserved positions 301 and 305 within the NBS region led to non-functionality of *N*, telling the importance of this domain in N-mediated resistance[96]. Also, fusion of the CC domain of *MLA* with tags did not affect its cell death activity; however, the CC-NB construct caused the strongest cell death response. These observations also suggest positive regulation of cell death by the NB domain[102].

In an earlier study, two CNLs; *Rx*, a potato (*Solanum tuberosum*) resistance protein and, *Bs2*, a pepper (*Capsicum annuum*) protein which confers resistance to PVX, the nematode *Globodera pallida*, and bacterium *Xanthomonas campestris*, respectively[113,114] differed in their interaction of ARC and LRR. Disruption of this interaction was not significant for initiation of signalling activity[94]. When tested on *N. benthamiana*, a construct lacking amino acids 293-382, which corresponds to the ARC1 subdomain failed to bind to LRR. Mutation studies of this segment (293-382) suggests that this subdomain physically interacts with LRR to cause binding[94]. The *Rx* LRR domain did not bind to *Rx* CC-NB or to any of the helper proteins *RAR1*, *SGT1* (suppressor of the G2 allele of skp1 (S-phase kinase-associated protein 1)) or coat protein, which means that LLR binding only occurs in ARC containing molecules[94].

In its structure, NB-ARC has a connecting surface by which it binds to LRR domain to form a complete NLR protein. Across all predicted linseed NLR proteins, the NBS domain had a highly conserved length, but LRR, the C-terminal region was highly variable in both in size and sequence[90]. The structural determinants of interaction between LRR and NB-ARC domains were identified[112].

Structure and function of C-terminal LRR

A typical LRR exhibits a curved horseshoe-like structure made of 2-45 motifs each containing 20-30 amino acids[115,116]. Lined with parallel β sheets on the concave side and helical elements to the convex side, the structure bears an extensive patch of conserved residues on its surface, suggesting a possible role in signal sensing or activation[87]. The concave surface is suggested to be a ligand-binding pocket, and its feature varies between proteins. The role of the residues on convex side remains unknown, but they are suspected to maintain the NLR proteins in an "inactive" state[87]. Two types of LRRs exist, type A in which cysteine occupies position nine and type B where asparagine takes up position nine.

The whole LRR domain is critical for expression of resistance phenotype in plants[96]. In *N. tabacum*, deletion mutations or alterations at the LRR region of *N* protein abolished *N*-mediated resistance to tobacco mosaic virus, suggesting that the whole LRR domain might be the binding site for other essential proteins[96]. As observed in potato *Rx* protein, the whole of LRR domain, not a section, is required for binding to ARC domain[94].

For mammals, similar structural features are noted, with minimal variations, unique to each NLR protein. In NOD2 protein, position nine in type A is occupied by threonine, giving this protein's LRR domain a distinct fold type[116]. By comparison, the human NOD9 (NLRX1) assumes a structure typical of LRRs with each motif consisting 28-29 amino acid residues. Motif LRR2, the shortest of all eight motifs has only 25 residues, and so introduces a groove that connects its α-helix to LRR3 β-helix[117]. To each of eight LRR motifs are single β-sheets (on the concave side), each corresponding to an α-helix (on the convex side) attached to two connecting loops[117]. NALP10 (or NLRP10) possessed the shortest LRR repeat, with typical structural conformity common to LRRs[87]. In NLRC4 however, the LRR domain adopts a typical solenoid fold consisting of six consecutive LRRs capped with an α-helix at the terminal end[116]. When compared to NLRP1 (NOD2), its LRR domain does not bind to muramyl dipeptide (MDP), a polymer that activates PAMPs in all known NLRs[116]. As with NOD9, each LRR consists one β-strand on its concave surface and an anti-parallel

α-helix on the convex side[116]. Altogether, the C-terminal LRR domain are intricately similar between plant R-proteins and mammalian NLRs. Beyond the three major domains: N-terminal region, centrally located nucleotide binding site and the C-termini, R-proteins often require presence of other non-R-proteins (including heat shock proteins, transcription factors) in order to express the resistance phenotype.

Proteins required by R-proteins for expression of resistance phenotype

Expression of resistance requires recruitment of other proteins downstream of the resistance. Some R-proteins even require presence of these 'other' proteins in order to trigger the resistance reaction. Several heat shock proteins including HSP2, HSP70, HSP80 and HSP90 are among the 'other' proteins. In rice, RAR1 (required for Mla12 resistance) and HSP90 (heat shock protein 90 kD) are required by RAC defence genes to express resistance against *Magnoporthe grisea* and *Xanthomonas oryzae*. In vivo, the RAC defence protein formed a complex with HSP90, HSP70 and RAR1. The I-2 and Mi-1 defence proteins need RSI2 (required for stability of I-2) to defend tomato against *Fusarium* spp[118]. However, I-2 also needs RSI2 and HSP20, SGT1 and HSP90 to develop HR in *N. benthamiana* [118]. The HSP80, a member of HSP90 family, is required to bind *N* protein to SGT1 and RAR1 during immune responses[118]. In *N. benthamiana* expressing *N* protein, EDS1 and HSP90 are required by *N* to express resistance phenotype against TMV[105,119]. Both HSP90 and SGT1 (Suppressor of G-2 Allele of Skip1) are required to enhance expression of PR proteins (PR-1a and PR 4), which lead to reduced disease development in *N. benthamiana* infected with *Ralstonia solanacearum* [120]. In barley, a minimal expression of RAR1 protein is required for MLA-mediated defence against *Blumeria graminis f. sp. hordei* (*Bgh*)[121]. In Arabidopsis infected with TEV, the defence gene *RTM1* (Restricted Tobacco Etch Virus Movement 1) requires both the C-terminal and jacalin domains in order to develop resistance phenotype[122]. On the other hand, *RTM2* protein needs small-HSPs, alpha helix domain and the C-terminus for expression of resistance[123].

Even in mammals, these helper-proteins are required. For example, the human Nod1 protein requires SGT1 for activation and signalling, and HSP90 for stability[124]. Altogether, the above are examples to demonstrate that the presence of a particular R-protein or defence proteins alone is not sufficient to cause resistance. Presence of other proteins (often functioning in different physiological pathways) is equally critical for defence against pathogens.

1.2.5 Application of R-proteins in plants, practical lessons

A common network governing plant defence in bacteria and fungi

An *in-silico* analysis of all immune-related proteins of *Arabidopsis thaliana* (a model plant) against effectors of *Pseudomonas syringae* (Psy) and *Hyaloperenospora arabidopsidis* (Hpa) identified a common network of pathogen- (i.e fungi and bacteria) defence proteins[125]. This network connects NBS-LRR to phytohormone pathways and conforms to the guard hypothesis of resistance. Out of 673 GO terms classified, 66 are related to defence while 165 effector targets regulate transcription, metabolism and nuclear location. The immune receptors were shown to have highly interconnected clusters of interacting genes[125].

Elsewhere, in-silico analysis using degenerate PCR and database mining were used to identify defence proteins in pepper[92]. At least 57 NBS-LRR proteins were identified. Identified proteins presented typical R-protein features, specifically showing an N-terminal TIR domain, a centrally located NB-ARC and the C-terminal LRR. Some defence proteins had the NBS-LRR domains but the N-termini were not of TIR type. Identified defence proteins also had conserved P-loop, kinase-2, and GLPL motifs. Their sequences mapped to other known R-proteins including *RPM1, Gpa2* and *Prf* [92]**.**

A gene mining of linseed (flax) genome identified 147 non-redundant NBS-LRR encoding genes[90]. Each had at least 24 amino acids spanning the NBS domain. Both CNL and TNL subfamilies showed distinct patterns of distribution with diverse domain arrangements in TNLs. The CNLs were similar in structure to At3g14470, an Arabidopsis CNL-C protein. The TNLs had complex structures often with short exonic LRR segments. Most of these short exons were caused by interruptions from phase-0 intron. Major domains (TIR, NB-ARC and LRR) were often separated by introns[90]. By phylogeny, TNL and CNL major clusters consisted of five and three subsets respectively. Truncated NLR genes (TN, CN, NL, and N) were dispersed throughout the tree, which means they are a result of domain loss. Most of the R-proteins originate from poplar. Type A was common among members of CNL-1, each showing similarities to Arabidopsis R-proteins. Among TNLs, clades TNL-2 and TNL-5 had the majority of R-proteins. The TNL-2 proteins were of type B. They clustered with linseed and poplar R-proteins. The TNL-5 clade was equally distributed between types C and D. They characterized together with linseed R-proteins. Clades TNL-1 and TNL-3 were clustered with WRKY transcription factor 52 of Arabidopsis and R-proteins of the Solanaceae family[90].

Using barley and Arabidopsis as host plants, MLA-mediated resistance was demonstrated to be conserved in both plants[95]. The network suggests that there is one conserved phytohormone-dependent mechanism of signalling controlled by all CNLs and that MLA1 connects to that pathway[95]. Jasmonic acid (JA), ethylene (ET) and salicylic acid (SA) activated an inter-connected network of genes in response to perturbation. Isolate specific resistance and activation of cell death to *Blumeria graminis f. sp. hordei* (*Bgh*) was retained in the presence JA, ET and SA encoding genes[95]. Therefore, defence caused by these phyto-hormones is applicable for MLA-1-dependent re-programming of transcription[95].

In Arabidopsis, SA and flagelin triggered differential expression of NBS-encoding genes, suggesting interaction between basal defence and SA-response pathways[91]. In another study on pepper, both SA and abscisic acid (ABA) phyto-hormones were transcriptionally induced following expression of R-proteins[92].

Cross-species application of NLR in plants

Two R-proteins (RPS4 and RRS1) were tested in cabbage, tomato, tobacco and cucumbers[126]. Results of these studies confirmed that interspecies transfer of known plant R-proteins is practical. In Brassica species, plants transformed with RPS4 and RRS1 genes were resistant to *Colletotrichum higginsianum* but wild type plants and those transformed to express either of RPS4 or RRS1 were highly susceptible. Similar observations were made with *N. benthamiana*. Hence, both R-proteins are required to control *C. higginsianum* in Brassica and non-Brassica species[126].

For tomato (*Solanum lycopersicum*), presence of both R-proteins caused resistance to *Ralstonia solanacearum* in the presence of PopP2 effector and to *P. syringae* pv tomato in the presence of AvrRps4 effector[126]. A cucumber (*Cucumis sativas*) carrying both RPS4 and RRS1 were also resistant to *C. orbitare*, further confirming the presence of an active defence conferred by this pair of defence proteins[126]. Therefore, RPS4 and RRS1 confer resistance to *Colletotrichum* sp in several crops including Brassica, Nicotiana, solanaceae and cucumber respectively[126]. The ability to control the same pathogen in different plant species confirms that there exists a robust, conserved defence pathway that targets *Colletotrichum* sp. in any plant[126]. With this conclusion in mind, gene expression studies of R-genes were tested in several plant transcriptomes as discussed below.

1.2.6 Transcriptional studies and differentially expressed genes in plants infected with Potyviruses

Susceptible responses

A plant variety is considered susceptible when upon infection; it develops classical symptoms, which lead to significant economic loss or death of the plant. The interaction between this plant and the pathogen is therefore compatible.

In sweet potato infected with PVY, 175 genes were differentially expressed (between PVY inoculated and mock inoculated plants). Most of these genes were identified at 14 days post inoculation (dpi) but very few at 7dpi[127]. The differentially expressed genes (DEGs) identified at 7 dpi (the time when necrotic spots are observable on infected leaves) were down regulated, yet many of the same genes were over-expressed at 14 dpi (the time when systemic symptoms become visible on infected plants)[128]. Those DEGs identified at 14 dpi are related to defence. Genes encoding for sporamin, cell expansion, and heat-shock proteins were down regulated at local and systemic infection phases[128].

In another study, sweet potato infected with sweet potato virus disease (SPVD) 157 differentially expressed genes were identified between virus infected and control plants[129]. When SPVD-susceptible plants were examined with individual SPVD-causing viruses (SPFMV (sweet potato feathery mottle virus) and SPCSV (sweet potato chlorotic stunt virus)), less than 15 genes were differentially expressed between infected and control plants[130]. However, 216 were identified in plants infected with both viruses (SPFMC and SPCSV). Majority of these genes are for protein synthesis. Among them, 123 were up regulated and 93 were down regulated[130].

In another study of potato infected with PVY on two susceptible varieties of potato, photosynthesis and signalling genes were down regulated at 12 hours post inoculation (hpi), as those for transport were down regulated at 48 hpi in infected potato plants[131].

In *N. benthamiana* infected with PVX (potato virus X), the number of DEGs increased with disease progression[132]. A maximum 1,816 DEGs that belong to 68 functional categories were identified[132]. Specifically, genes for photosynthesis, cell wall remodelling, glucan metabolism, gibberellic acid signalling,

heat shock proteins, DNA-binding proteins, abiotic stress (cold, drought and UV light), biosynthetic processes and histones were down regulated[132]. When *N. benthamiana* was infected with plum pox virus (PPV) and TuMV, ribosomal proteins were induced in susceptible host[133,134].

In rice and *A. thaliana* infected with rice dwarf virus, cell wall related genes were suppressed[135]. For rice, xyloglycugan endotransglucosylase/hydrolases (XTH) and expansins required for cell wall expansion were suppressed in susceptible plants.

A similar trend was seen in Arabidopsis infected with tobacco etch virus (TEV)[123]. Genes that encode PR proteins were up regulated at 14 dpi, while the defence protein HSP70 was down regulated at both 7 and 14 dpi[136]. In susceptible Arabidopsis infected with Turnip mosaic virus (TuMV), metabolism-related genes were repressed[134]. In *A. thaliana* infected with three positive-stranded RNA viruses (tobacco etch virus (TEV), potato virus Y (PVY) and tobacco vein mottling virus (TVMV)), b-1,3-glucanase genes were up regulated at 2, 4 and 5 dpi[137].

Altogether, these studies indicate that specific genes that code for developmental, maintenance and metabolic processes are differentially expressed in compatible host-virus interactions. Some defence-related genes are also differentially expressed. It is important to show profiles of gene expression when the host plant is tolerant to the infecting virus.

Tolerance responses

Tolerance here is defined as the ability of a crop variety to develop mild or no symptoms of the disease and result in negligible economic loss compared to the severe symptoms and significant economic loss registered in susceptible varieties.

In sweet potato varieties which are tolerant to SPVD, 145 genes were differentially expressed at 5 dpi[129]. Among these DEGs were ribosomal proteins, proteins involved in biosynthesis and signalling. Chloroplast-related proteins, including metallothionein and ubiquitin-conjugating enzymes were down regulated in SPVD-tolerant potato[129]. Defence related genes including ribosomal proteins and elongation factors, transcription factors (WRKY, homeodomain proteins, and NAC-like proteins) were also up regulated in tolerant sweet potato[129].

In melon, more genes were differentially expressed in the tolerant cultivar (1886 genes) compared to the susceptible (121 genes)[138]. Pathogen defence genes (WRKY, Myb-TF and PR) were up regulated. The heat shock proteins were strongly up regulated in tolerant genotype contrary to the susceptible variety[138]. The case studies show that more defence-related genes are differentially expressed in tolerant varieties compared to susceptible cultivar. An understanding of gene expression profiles disease resistant varieties will provide a better understanding of the mechanism of crop plant-pathogen interaction.

Resistance responses

Varieties that show no recognizable symptom of disease, and remain so throughout their life cycle, are considered resistant to the disease. Several studies have shown patterns of gene expression in these resistant varieties.

In non-heading Chinese cabbage, 400 transcript-derived fragments (TDFs) were identified in response to TuMV. Of these 314 were down regulated while 86 were up regulated in the resistant variety[139]. The differentially expressed genes include: glutathione reductase, leucine-rich repeat receptor-like protein kinase (LRR-RLK), RPP5 locus, HSP70, a drought induced protein, multi-drug resistance protein and virus resistance proteins[139]. Reactive oxygen species and glutathione reductase encoding genes, which are general stress responders accumulated in all TuMV-infected plants[139,140]. Between 8 and 48 hpi, a glutathione reductase gene was up regulated in resistant varieties. This gene did not change in mock-inoculated plants but its expression increased in the susceptible variety at 144 hpi. Up to 90% of validated DEGs were expressed early in the resistant variety compared to the susceptible variety. Of these, 70% displayed varied peaks across time-points, suggesting that resistance to TuMV is quantitative over time course.

In rice infected with rice dwarf virus, levels of defence hormones were significantly lower in mock-inoculated plants compared to levels in plants infected with the virus. Comparable magnitudes of gene expression for defence- and hormone-related genes were seen in susceptible and resistant treatments. Specific genes were down regulated in the resistant variety. The GTP-binding proteins, potassium ion trans-membrane transporters, and cell wall related genes were down regulated in the resistant variety[135].

In the resistant variety of cabbage infected with TuMV, LRR-RLK and HSP70 genes increased at 3 dpi and 4 dpi[139]. An RPP5 locus was strongly induced during the infection time-course, with a qPCR expression similar to that of LRR-RLK. Given the role of RPP5 in resistance to downy mildew of Arabidopsis[141], this gene might trigger resistance to TuMV in cabbage. Genes that code for calcium-dependent carriers and signal transduction were also up regulated at 72 hpi in the resistant variety, but was not differentially expressed again throughout the study[139]. All metabolism-related transcripts including insulinase family protein, proline dehydrogenase, enoyl-CoA hydratase, non-photochemical quenching, diaminopimelate decarboxylase, UDP-glucose:steril, glucosyltransferasen, cycloartenol synthase and phosphoinositide-binding mRNA were also up regulated in the resistant variety[139].

In *N. tabacum* inoculated with PVX, zinc-finger proteins were induced at 72 hpi and 96 hpi, matching with the onset of HR on PVX-inoculated *N. tabacum* plants[142]. Zinc finger proteins are required for transcription in eukaryotes[143], salt stress[144-146], cold stress[145-147], drought stress[148], oxidative stress[149] and light stress[150]. Apart from zinc-finger proteins, RNA polymerase II transcription factor, processing protein-related mRNA and rRNA processing protein were also up regulated.

In *N. benthamiana* infected with PVX, photosynthesis, cell wall remodelling, glucan metabolism and

gibberellic acid signalling genes were up regulated in all resistant variety[132]. Pathogenesis-related proteins, basic chitinases, mitogen-activated protein kinase, WRKY transcription factors and more were also up regulated[132]. At all time-points tested (1, 2, 4, 7, 8 and 11 dpi), JA, SA and ABA were over-represented in the resistant variety[132]. Similar observations were made in Arabidopsis during resistant and susceptible variety infected with *Pseudomonas syringae*[151].

Altogether, the above observations made on resistant varieties show that both defence-related genes, and genes involved in growth and development, are all up regulated. Defence genes are not constitutively expressed, but are only up regulated at specific time bounds. Also, defence is not only marked by up regulation of R-proteins (with NBS-LRR domains), but other defence-associated proteins including pathogenesis-related proteins, heat shock proteins and chaperons, late embryogenesis abundant protein (LEA), and several transcription factors including WRKY, NAC, GATA, GRAS and NmrA, which are up regulated in pathogen- and abiotic-stresses. In a cassava variety expressing resistance to CBSD, these defence proteins (other than typical NBS-LRR proteins) were up regulated, alongside R-proteins, as shown in Chapter 3. Because these defence-associated proteins are up regulated in resistant cassava variety (this study) and in other disease resistant crops, a summary of their expression in other crops is presented here.

1.2.7 Other types of defence-associated proteins

Heat shock proteins (HSP)

Heat shock proteins, are a group of structurally disordered proteins that function in diverse processes including regulatory, signalling and defence pathways. Also referred to as molecular chaperones, HSPs function in protein folding, translocation and metabolism of defence-associated and developmental proteins[152]. The ability of HSP to perform unique roles in different processes is governed by their intrinsic structural discord[153]. Chaperons are mainly involved in RNA binding and protein folding[153,154]. Some R-proteins require HSPs for their proper functioning[120,155]. Studies show that some HSPs function as defence-related proteins against biotic and abiotic stresses[156,157].

Pathogenesis-related (PR) genes

Pathogenesis-related (PR) genes are inducible defence-related proteins with varied molecular sizes (5 to 75 kDA)[158]. First discovered in tobacco resisting infection by tobacco mosaic virus (TMV), PR-1 has shown resistance beyond TMV[159]. So far, there are 17 recognized families of PRs[158,160]. Generally, PRs have RNase, DNAse activity coupled with antiviral properties[160-162]. A typical PR-mediated resistance causes hypersensitive response (HR) and programmed cell death (PCD) at the infection site[160,163]. Their expression marks prime systemic acquired resistance (SAR) against pathogens[164]. They also function in secondary metabolite biosynthesis, storage, phyto-hormone and ligand binding[162].

NAC transcription factors

Among transcription factors (TFs), the NAC (NAM, ATAF, and CUC) superfamily is one of the largest of families found only in plants. In a study of cassava infected with CBSV, five NAC proteins were validated as defence genes, prompting the analysis of the same in this study[55].

Structurally, NAC transcription factors have specific DNA-binding motifs by which they form homo- or heterodimer complexes, a critical step in the processes of expressing defence[165,166]. During compatible interaction with susceptible hosts, NACs are undetectable in the nuclei or plasma membrane[167,168]. For incompatible interaction (in resistant varieties), NAC proteins translocate to the nucleus[166,169] and plasma membrane[168] where they trigger a cascade of reactions leading to resistance.

Once in the nucleus, DNA fragmentation, hypersensitive response (HR) and cell death occur in incompatible interaction, leading to resistance in host plants[166]. The HR induced during incompatible interaction is associated with phosphorylation, which occurs downstream of the hydrogen peroxide (H_2O_2) pathway. The NAC transcription activator is located at the C-terminus[166,169,170]. This region is divergent, suggesting a specific role of the C-terminus in downstream perturbations. However, a functional SA pathway is required for NAC-directed defence against viral pathogens (like Turnip crinkle virus) but JA and ET pathways are not essential[171].

A NAC induced HR caused resistance to rice blast disease[172], *C. graminicola* in maize[173], *Botrytis cinera* and *P. syringae* pv. *tomato* (Pst) on Arabidopsis[168], turnip crinkle virus[171] and *Acidovorax avenae*. In rice, a NAC promoter region had a dehydration-response element (DRE) and another NAC promoter had two DREs. For Arabidopsis, abiotic stress factors like drought, salt, cold and ABA induced expression of NAC genes, implying the role of NAC in resistance to stress[168]. Similar results were recorded in maize[173], rice[166,170] and tomato[165,174].

WRKY transcription factors

The role of WRKY transcription factors in plant defence[175-177], their structural features and functional network has been reviewed[178]. An RNAseq analysis of cassava expressing resistance to UCBSV (Chapter 3) identified WRKY proteins among up regulated genes. Up regulation of these WRKY proteins occurred at the same time points when notable R-proteins were up regulated, hence, the suspected involvement of WRKY in cassava's resistance to UCBSV[65]. A WRKY protein has also been implied in the resistance to tobacco mosaic virus in *Medicago truncatula*[179], and they positively regulate expression of resistance to TMV in *Capsicum anuum*[180].

Elsewhere in rice, two alleles of one WRKY gene (OsWRKY45-1 and OsWRKY45-2) play opposite roles for resistance to bacteria[181]. The two alleles are 10-amino acids separate, each with a unique promoter region, supporting their specific, opposing function. Alone, OsWRKY45-1 reduced resistance to Xoo (*Xanthomonas oryzae* pv *oryzae*) and Xoc (*Xanthomonas oryzae* pv *oryzicola*) while OsWRKY45-2 enhanced resistance to Xoo and Xoc respectively[181]. With enhanced SA and JA production, OsWRKY45-1 provided resistance to Xoo and Xoc, but OsWRKY45-2 mediated resistance to Xoo and Xoc requires JA only[181].

In *Brassica campestris* ssp *chinesis* (Bcc), a WRKY (BcWRKY) was identified[182]. It localized to the nucleus and caused resistance to cold, salinity, heat and osmotic stress[182]. At 48 hours post-challenging, the WRKY gene was over-expressed in leaves of Bcc[182]. However, in Populus, 61 WRKY genes were over-expressed

in against *Marssonina brunnea*[183]. These WRKY genes grouped into three clusters, each having a defence-related cis-acting element at their promoter region[183]. Treatment with *M. brunnea*, wounding, SA, methyl jasmonate (MeJA), low temperature and salinity also caused differential expression of the WRKY genes in Populus[183]. The roles of WRKY in defence to plant viral (UCSBV and TMV), fungal (*Xanthomonas spp, M. brunnea*) and hormonal signals (SA and MeJA) are strong indicators of their role in defence.

Late embryogenesis abundant protein

The late embryogenesis abundant (LEA) type of proteins is associated with adaptation to water stress[184]. The LEA is abundant in plant[184,185], fungi[186] and mammalian genomes[187-191]. As the name suggests, LEAs are produced during late stages of embryogenesis, especially at seed development. They protect cells from desiccation[184], salinity[192] and extreme cold or hot temperatures[187]. Cell protection from abiotic stress occurs in a dose-dependent manner[190].

Amongst diverse species, a conserved high hydrophilicity region was identified[193]. Their main role is to protect cells from dehydration, salt and cold stress. ABA is the most significant factor that influences expression of LEA proteins. It directly reduces drought-caused harm[194]. LEA proteins function as multi-stress regulating proteins. They regulate metabolism of higher plants under stressful conditions through use of RNA-regulatory-proteins[194].

In soybean, expression of a LEA protein (MsLEA3) increased tolerance to salt stress. Salt stressed plants had increased levels of MsLEA3 starting at one hour after salt treatment reaching a substantial level at 12 hours post inoculation (12 hpi). By 24 hpi, its expression declined, an indicator that this gene acts within a very short time span. There was an increase of LEA transcripts in ABA-treated soybeans by 3 hpi. In seedling leaves, expression of LEA increased substantially by 6 hpi and reached a maxima at 48 hpi[192].

Arabidopsis genome has 51 LEA genes that are responsive to ABA, low temperature (cold), salt and drought[184]. This correlates with expression of senescence-associated gene (SAG) 21, a gene that also functions in senescence caused by natural or stress-induced conditions. *A. thaliana* LEA proteins are required for normal seed development. Mutant seeds lacking this protein showed premature seeds dehydration and impaired development, which is traceable to extreme desiccation[195]. A soybean LEA protein was tested in *E.coli* and found to be heat stable and hydrophilic. This protein called PM2, increased salt tolerance and not osmotic tolerance[185].

In cassava expressing resistance to UCBSV, several LEA proteins were up regulated in tandem with R-proteins and other defence genes (Chapter 3). This is an indicator of the possible role of LEAs in regulating defence by improving resilience in the resistant variety, so that these resistant varieties better compete for physiological resources when faced with competition, as is the case with viruses[196,197].

Beyond plants, fungi and mammals have expressed LEA genes in response to single and multiple stresses. For example, in bdelloid rotifers, LEA protein (ProLEA1), prevented desiccation of dehydrated rotifers[191]. Tested in yeast (*Saccharomyces cerevisiae*), a wheat LEA protein (Em) gave osmotic protection from

external osmotic challenge[198]. Larvae of African chironomid (*Polypedilum vanderplanki*) expressing LEA proteins (PvLEA1, PvLEA2, and PvLEA3) were tolerant to salt and water-loss[190]. All three LEA proteins are predominantly hydrophilic, a typical characteristic of LEA proteins. These proteins were induced by treatment of cells with NaCl, confirming their role in reduced salt stress. There was high expression of LEA in desiccating larvae at 8 and 24 hpi, but expression reduced to zero by 48 hpi. Instead, PvLEA2 was processed into smaller proteins at 48 hpi[190]. This might be a strategy for LEA to specify roles for salt and desiccation stress respectively. These proteins (PvLEA1, PvLEA2, and PvLEA3) are similar to those identified in desiccation-tolerant crustacean (*Artemia franciscana*), a member of anthropod class[189]. The *A. franciscana* LEA (AfrLEA1, AfrLEA2) were highly expressed at anhydrobiotic stages (diapause and post-diapause embryos). On the contrary, desiccation-intolerant larval stages (control) showed very low expression of these genes. This confirmed that LEA proteins are expressed as desiccation shield, whenever required.

In *Caenorhabditis elegans*, three LEA genes functionally reduced desiccation stress[187]. There was maximum resistance to desiccation at 8 hpi and 24 hpi respectively. Worms with silenced LEA genes had a significantly reduced survival rate compared to those expressing the LEA protein. Indeed, transcription in LEA-silenced worms was reduced 11-fold, due to desiccation, compared to worms with active LEA genes. The reduction was detected at 8 hpi and 24 hpi, correlating with the observed death rate when LEA silenced worms were exposed to desiccation. Silencing the same LEA genes reduced survival of worms during osmotic and heat stress.

LEA proteins act in a synergistic manner with certain sugars, sucrose for desiccating seeds of plants and trehalose for arthropods. It stabilizes in vitro proteins and other biological structures (especially of enzymes) from damage caused by diverse abiotic stresses. They preserve activity of desiccation-exposed citrate synthase (CS) and enzyme activity in a dose-dependent manner[188]. Altogether, LEA proteins act as a molecular shield against formation of destructive protein aggregates during water loss, whether by cold or heat.

1.3 Conclusion

This review covered examples and case studies of plants, plant R-proteins and other defence associated proteins, their responses in compatible and incompatible interactions with pathogens and abiotic stress factors. Based on up regulation of some of these R-proteins, and other defence genes in this study, it is likely that similar mechanisms are operational in UCBSV-resistant cassava. Cassava consumption (as food) is highest in developing countries of Africa where food security is a real problem. So far, no cassava variety is immune to CBSD, but varied levels of resistance are observed in some varieties. Planting of these resistant varieties lead to economically insignificant yields losses and is being promoted (locally and in research institutions) to sustain production in major cassava growing areas. There is a critical need to identify and exploit the mechanism(s) causing resistance in these varieties. But first the mechanism must be identified and appropriate functional studies done prior to their adoption in breeding programs. The work presented in this thesis contributes information on the mechanisms underlying resistance to UCBSV (and by extension, CBSD), and in doing so; re-directs the process of breeding for resistance to CBSD. To this end, the research questions asked, hypotheses tested and objectives for this study are presented below.

1.4 Research questions

1. To what extent is variety Namikonga resistant to UCBSV in relation to the susceptible variety Albert?
2. Which genes are differentially expressed between variety Albert infected with UCBSV and an uninfected Albert control?
3. Which genes are differentially expressed between variety Namikonga infected with UCBSV and an uninfected Namikonga control?
4. What expression patterns of pathogen defence genes are recognized in Namikonga and Albert infected with UCBSV over a time course?

1.5 Hypotheses of the study

The study was based on the hypotheses that:

1. UCBSV accumulates at significantly lower levels in Namikonga compared to accumulation rates in the susceptible Albert
2. Transcriptional changes associated with resistance to plant pathogens are significantly activated in Namikonga and not in Albert
3. Pathogen defence genes are expressed earlier in Namikonga compared to Albert over the time course
4. The pathogen defence genes that are significantly expressed in Namikonga differ from those observed in other CBSD-resistant cassava varieties: NDL06/132, Mkombozi and Kiroba

1.6 Objectives of the study

The main aim of this thesis was to analyse the transcriptome of cassava challenged with UCBSV.

The specific objectives were to:

1. Determine the trend of UCBSV accumulation in cassava plants of Namikonga and Albert infected with UCBSV
2. Determine gene families and pathways associated with susceptibility to CBSD observed in Albert at several time points
3. Determine gene families and pathways associated with tolerance to CBSD observed in Namikonga at several time points
4. Compare gene families and pathways associated with response to UCBSV infection in Namikonga and Albert
5. Determine and compare the patterns of gene expression of selected pathogen defense genes in Namikonga and Albert at several time-points after infection and before symptom emergence
6. Determine and compare the patterns of gene expression of selected pathogen defense genes associated with resistance to CBSD in other CBSD resistant cassava cultivars namely NDL06/132, Mkombozi and Kiroba

1.7 References

1 Allem, A. C. The origin of Manihot esculenta Crantz (Euphorbiaceae). Genet Resources Crop Evol (**1994**), **41**, 133-150.

2 Isendahl, C. The domestication and early spread of Manioc (Manihot esculenta Crantz): a brief synthesis. LAA (**2011**), **22**, 452-468.

3 Olsen, K. M. SNPs, SSRs and inferences on cassava's origin. Plant Mol Biol (**2004**), **56**, 517-526.

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

5 Chacon, J., Madrinan, S., Debouck, D., Rodriguez, F. & Tohme, J. Phylogenetic patterns in the genus Manihot (Euphorbiaceae) inferred from analyses of nuclear and chloroplast DNA regions. Mol Phylogenet Evol (**2008**), **49**, 260-267.

6 Duputie, A., David, P., Debain, C. & McKey, D. Natural hybridization between a clonally propagated crop, cassava (Manihot esculenta Crantz) and a wild relative in French Guiana. Mol Ecol (**2007**), 3025-3038.

7 Nassar, N. M. Cassava, Manihot esculenta Crantz, genetic resources: origin of the crop, its evolution and relationships with wild relatives. Genet Mol Res (**2002**), **1**, 298-305.

8 El-Sharkawy, M. A. Cassava biology and physiology. Plant Mol Biol (**2004**), **56**, 481-501.

9 Halsey, M. E., Olsen, K. M., Taylor, N. J. & Chavarriaga-Aguirre, P. Reproductive Biology of Cassava (Crantz) and Isolation of Experimental Field Trials. Crop Sci (**2008**), **48**, 49.

10 Meireles da Silva, R., Bandel, G. & Sodero Martins, P. Mating system in an experimental garden composed of cassava (Manihot esculenta Crantz) ethnovarieties. Euphytica (**2003**), **134**, 127-135.

11 Bisimwa, E., Walangululu, J. & Bragard, C. Cassava Mosaic Disease yield loss assessment under various altitude agroecosystems in the SudKivu Region, Democratic Republic of Congo. Tropicultura (**2015**), **33**, 101-110.

12 Legg, J. P., Sseruwagi, P., Boniface, S., Okao-Okuja, G., Shirima, R. & Bigirimana, S. Spatio-temporal patterns of genetic change amongst cassava Bemisia tabaci whiteflies driving virus pandemics in East and Central Africa. Virus Res. (**2014**), **186**, 96144.

13 Patil, B. L., Legg, J. P., Kanju, E. & Fauquet, C. M. Cassava brown streak disease: A threat to food security in Africa. J Gen Virol (**2014**), **96**, 956-968.

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

15 Okogbenin, E., Egesi, C. N., Olasanmi, B., Ogundapo, O., Kahya, S., Hurtado, P., Marin, J., Akinbo, O., Mba, C., Gomez, H., de Vicente, C., Baiyeri, S., Uguru, M., Ewa, F. & Fregene, M. Molecular marker analysis and validation of resistance to cassava mosaic disease in elite cassava genotypes in Nigeria. Crop Sci (**2012**), **52**, 2576-2586.

16 Rabbi, I. Y., Hamblin, M. T., Kumar, P. L., Gedil, M. A., Ikpan, A. S., Jannink, J. L. & Kulakow, P. A. High-resolution mapping of resistance to cassava mosaic geminiviruses in cassava using genotyping-by-sequencing and its implications for breeding. Virus Res (**2014**), **24**, 186-187.

17 Wolfe, M. D., Rabbi, I. Y., Egesi, C., Hamblin, M., Kawuki, R., Kulakow, P., Lozano, R., Carpio, D. P. D., Ramu, P. & Jannink, J. L. Genome-Wide Association and prediction reveals genetic

architecture of cassava mosaic disease Resistance and prospects for rapid genetic improvement. Plant Genome (**2016**), **9**, 2.

18 Houngue, J. A., Zandjanakou-Tachin, M., Ngalle, H. B., Pita, J. S., Cacaï, G. H. T., Ngatat, S. E., Bell, J. M. & Ahanhanzo, C. Evaluation of resistance to cassava mosaic disease in selected African cassava cultivars using combined molecular and greenhouse grafting tools. Physiol Mol Plant Pathol (**2019**), **105**, 47-53.

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

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

21 Mware, B. O., Ateka, E. M., Songa, J. M., Narla, R. D., Olubayo, F. & Amata, R. Transmission and distribution of cassava brown streak virus disease in cassava growing areas of Kenya. J Appl Biosci (**2009**), **16**, 864 - 870.

22 Storey, H. H. Virus diseases of East African plants. VI-A progress report on studies of the disease of cassava. E Afri Agric J (**1936**), **2**, 34-39.

23 Mware, B., Narla, R., Amata, R., Olubayo, F., Songa, J., Kyamanyua, S. & Ateka, E. M. Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. J Gen Mol Virol (**2009**), **1**, 40-45.

24 Mware, B., Olubayo, F., Narla, R., Songa, J., Amata, R., Kyamanywa, S. & Ateka, E. M. First record of spiraling whitefly in coastal Kenya: emergence, host range, distribution and association with cassava brown streak virus disease. Int J Agric & Biol (**2010**), **12**, 411–415.

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

26 Maruthi, M. N., Jeremiah, S. C., Mohammed, I. U. & Legg, J. P. The role of the whitefly, Bemisia tabaci (Gennadius), and farmer practices in the spread of cassava brown streak ipomoviruses. J Phytopathol (**2017**), **165**, 707-717.

27 Chen, W., Wosula, E. N., Hasegawa, D. K., Casinga, C., Shirima, R. R., Fiaboe, K. K. M., Hanna, R., Fosto, A., Goergen, G., Tamò, M., Mahuku, G., Murithi, H. M., Tripathi, L., Mware, B., Kumar, L. P., Ntawuruhunga, P., Moyo, C., Yomeni, M., Boahen, S., Edet, M., Awoyale, W., Wintermantel, W. M., Ling, K. S., Legg, J. P. & Fei, Z. Genome of the African cassava whitefly Bemisia tabaci and distribution and genetic diversity of cassava-colonizing whiteflies in Africa. Insect Biochem Mol Biol (**2019**), **110**, 112-120.

28 Mugerwa, H., Rey, M. E. C., Tairo, F., Ndunguru, J. & Sseruwagi, P. Two sub-Saharan Africa 1 populations of Bemisia tabaci exhibit distinct biological differences in fecundity and survivorship on cassava. Crop Protect (**2019**), **117**, 7-14.

29 Nichols, R. F. W. Breeding cassava for virus resistance. E Afri Agric J (**1947**), **12**, 184-194.

30 Were, H. K., Winter, S. & Maiss, E. Viruses Infecting Cassava in Kenya. Plant Dis (**2004**), **88**, 17-22.

31 Nichols, R. F. W. The Brown Streak Disease of Cassava. E Afri Agric J (**1950**), **15**, 154-160.

32 Hillocks, R. J., Thresh, J. M., Tomas, J., Botao, M., Macia, R. & Zavier, R. Cassava brown streak disease in northern Mozambique. Int J Pest Manage (**2002**), **48**, 178-181.

33 Hillocks, R. J., Raya, M. & Thresh, J. M. The association between root necrosis and above‐ground symptoms of brown streak virus infection of cassava in southern Tanzania. Int J Pest Manage (**1996**), **42**, 285-289.

34 Legg, J. P. & Raya, M. D. Survey of cassava virus diseases in Tanzania. Int J Pest Manage (**1998**), **44**, 17-23.

35 Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. & Colvin, J. Re-emergence of cassava brown streak disease in Uganda. Plant Dis (**2007**), **91**, 24-29.

36 Bigirimana, S., Barumbanze, P., Ndayihanzamaso, P., Shirima, R. & Legg, J. First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Burundi. New Dis Rep (**2011**), **24**, 2044-0588.

37 Mulimbi, W., Phemba, X., Assumani, B., Kasereka, P., Muyisa, S., Ugentho, H., Reeder, R., Legg, J., Laurenson, L. & Weekes, R. First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. New Dis Rep (**2012**), **26**, 2044-0588.

38 Were, M. N., Mukoye, B., Osogo, A. K., Mangeni, B. C., Nyamwamu, P. A., Ogemah, V. K., Muoma, J. V., Winter, S. & Were, H. K. Occurrence and Distribution of Begomoviruses infecting cassava in Western Kenya. Plant (**2016**), **4**, 108-113.

39 Amisse, J. J. G., Ndunguru, J., Tairo, F., Boykin, L. M., Kehoe, M. A., Cossa, N., Ateka, E., Rey, C. & Sseruwagi, P. First report of Cassava brown streak viruses on wild plant species in Mozambique. Physiol Mol Plant Pathol (**2019**), **105**, 88-95.

40 Roux-Cuvelier, M., Teyssedre, D., Chesneau, T., Jeffray, C., Massé, D., Jade, K., Karime, A. L. A., Hostachy, B., Reynaud, B. & Legg, J. P. First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Mayotte Island. New Dis Rep (**2014**), **30**, 28.

41 Hillocks, R., Raya, M. & Thresh, J. The association between root necrosis and above‐ground symptoms of brown streak virus infection of cassava in southern Tanzania. Int J Pest Manage (**1996**), **42**, 285-289.

42 Hillocks, R., Raya, M. & Thresh, J. Factors affecting the distribution, spread and symptom expression of cassava brown streak disease in Tanzania. African Journal of Root and Tuber Crops (**1999**), **3**, 57-61.

43 Jennings, D. L. Observations on virus diseases of Cassava in resistant and susceptible varieties. II. Brown streak disease. Empire J Exp Agri (**1960**), **28**, 261-270.

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

45 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Genetically distinct strains of cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch Virol (**2009**), **154**, 353-359.

46 Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M. & Butgereitt, A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen Virol (**2010**), **91**, 1365-1372.

47 Mbanzibwa, D. R., Tian, Y., Mukasa, S. B. & Valkonen, J. P. 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. J Virol (**2009**), **83**, 6934-6940.

48 Monger, W. A., Alicai, T., Ndunguru, J., Kinyua, Z. M., Potts, M., Reeder, R. H., Miano, D. W., Adams, I. P., Boonham, N., Glover, R. H. & Smith, J. The complete genome sequence of the Tanzanian strain of cassava brown streak virus and comparison with the Ugandan strain sequence. Arch Virol (**2010**), **155**, 429-433.

49 Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. & Ball, L. A. *Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses*. (Academic Press, 2005).

50 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. Evolution of cassava brown streak disease-associated viruses. J Gen Virol (**2011**), **92**, 974-987.

51 Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djikeng, A., Kehoe, M. & Boykin, L. M. Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation. PLoS One (**2015**), **10**, e0141939.

52 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. J Virol Methods (**2011**), **171**, 394-400.

53 Lister, R. M. Mechanical transmision of cassava brown streak virus. Nature (**1959**), **183**, 1588-1589

54 Wagaba, H., Beyene, G., Trembley, C., Alicai, T., Fauquet, C. M. & Taylor, N. J. Efficient transmission of cassava brown streak disease viral pathogens by chip bud grafting. BMC Res Not (**2013**), **6**, 516.

55 Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One (**2014**), **9**, e96642.

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

57 Saggaf, M. H., Ndunguru, J., Tairo, F., Sseruwagi, P., Ascencio-Ibáñez, J. T., Kilalo, D. & Miano, D. W. Immunohistochemical localization of Cassava brown streak virus and its morphological effect on cassava leaves. Physiol Mol Plant Pathol (**2019**), **105**, 67-76.

58 Ogwok, E., Alicai, T., Rey, M. E. C., Beyene, G. & Taylor, N. J. Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta*) plants. Plant Pathol (**2015**), **64**, 1235-1246.

59 Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. Adv Virol. (**2011**), **2012**, 795697.

60 Kanju, E. E., Masumba, E., Masawe, M., Tollano, S., Muli, B., Zacarias, A., Mahungu, N., Khizzah, B., Whyte, J. & Dixon, A. Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmer and consumer preferences. Proceedings 13th ISTRC Symp (**2007**), 95-101.

61 Kawuki, R. S., Pariyo, A., Amuge, T., Nuwamanya, E., Ssemakula, G., Tumwesigye, S., Bua, A., Baguma, Y., Omongo, C. A., Alicai, T. & Orone, J. A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). J Plant Breed Crop Sci (**2011**), **3**.

62 Kawuki, R. S., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I. S., Ozimati, A., Kyaligonza, V., Abaca, A., Orone, J. & Tumuhimbise, R. Eleven years of breeding efforts to combat cassava brown streak disease. Breed Sci. (**2016**), **66**, 560-571.

63 Mukiibi, D. R., Alicai, T., Kawuki, R., Okao-Okuja, G., Tairo, F., Sseruwagi, P., Ndunguru, J. & Ateka, E. M. Resistance of advanced cassava breeding clones to infection by major viruses in Uganda. Crop Protect. (**2019**), **115**, 104-112.

64 Sheat, S., Fuerholzner, B., Stein, B. & Winter, S. Resistance Against Cassava Brown Streak Viruses From Africa in Cassava Germplasm From South America. Frontiers In Plant Science (**2019**), **10**.

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

66 Anjanappa, R. B., Mehta, D., Okoniewski, M. J., Szabelska-Beresewicz, A., Gruissem, W. & Vanderschuren, H. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. Mol Plant Pathol (**2018**), **19**, 476-489.

67 Lozano, R., Hamblin, M. T., Prochnik, S. & Jannink, J. L. Identification and distribution of the NBS-LRR gene family in the Cassava genome. BMC Genomics (**2015**), **16**, 360.

68 Pennisi, E. Armed and Dangerous. Science (**2010**), **327**, 804-805.

69 Gondwe, F. M. T., Mahungu, N. M., Hillocks, R. J., Raya, M. D., Moyo, C. C., Soko, M. M., Chipungu, F. & Benesi, I. R. M. Economic losses experienced by small-scale famers in Malawi due to cassava brown streak virus disease. In: Cassava Brown Streak Disease: Past, Present and Future. J. P. Legg and R. J. Hillocks, eds. Proc. Int. Workshop, Mombasa, Kenya. 27-30 2002. Natural Resources International Limited, Aylesford, UK. (**2003**), 28-38.

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

71 Rwegasira, G. M. & Rey, C. M. E. Relationship between symptoms expression and virus detection in cassava brown virus streak-infected plants. J Agric Sci (**2012**), **4**, 246-253.

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

73 Abarshi, M. M., Mohammed, I. U., Jeremiah, S. C., Legg, J. P., Kumar, P. L., Hillocks, R. J. & Maruthi, M. N. 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. J Virol Methods (**2012**), **179**, 176-184.

74 Ogwok, E., Patil, B. L., Alicai, T. & Fauquet, C. M. Transmission studies with Cassava brown streak Uganda virus (Potyviridae: Ipomovirus) and its interaction with abiotic and biotic factors in Nicotiana benthamiana. J Virol Meth (**2010**), **169**, 296-304.

75 Rwegasira, G. M., Rey, M. E. C. & Nawabu, H. Approaches to diagnosis and detection of cassava brown streak virus (Potiviridae: Ipomovirus) in field-grown cassava crop. Afr J F Agric Nutri Develop (**2011**), **11**, 4739-4756.

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

77 Livak, K. & Schmittgen, T. Analysis of relative gene expression data using real-time quantitative PCR and the 2DeltaDeltaCT method. Methods (**2001**), **25**, 402 - 408.

78 Adams, I. P., Abidrabo, P., Miano, D. W., Alicai, T., Kinyua, Z. M., Clarke, J., Macarthur, R., Weekes, R., Laurenson, L., Hany, U., Peters, D., Potts, M., Glover, R., Boonham, N. & Smith, J. 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 Pathol (**2013**), **62**, 233-242.

79 Tomlinson, J. A., Ostoja-Starzewska, S., Adams, I. P., Miano, D. W., Abidrabo, P., Kinyua, Z., Alicai, T., Dickinson, M. J., Peters, D., Boonham, N. & Smith, J. Loop-mediated isothermal amplification for rapid detection of the causal agents of cassava brown streak disease. J Virol Methods (**2013**), **191**, 148-154.

80 Kumar, H., Kawai, T. & Akira, S. Toll-like receptors and innate immunity. Biochem Biophys Res Comm (**2009**), **388**, 621-625.

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

82 Caplan, J., Padmanabhan, M. & Dinesh-Kumar, S. P. Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming. Cell Host & Microbe (**2008**), **3**, 126-135.

83 Jacob, F., Vernaldi, S. & Maekawa, T. Evolution and Conservation of Plant NLR Functions. Front Immunol. (**2013**), **4**, 297.

84 Urbach, J. M. & Ausubel, F. M. The NBS-LRR architectures of plant R-proteins and metazoan NLRs evolved in independent events. PNAS (**2017**), **114**, 1063-1068.

85 Ammelburg, M., Frickey, T. & Lupas, A. N. Classification of AAA+ proteins. J Struct Biol (**2006**), **156**, 2-11.

86 Leipe, D. D., Koonin, E. V. & Aravind, L. STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. J Mol Biol (**2004**), **343**, 1-28.

87 Proell, M., Riedl, S. J., Fritz, J. H., Rojas, A. M. & Schwarzenbacher, R. The NOD-like receptor (NLR) family: a tale of similarities and differences. PLoS One (**2008**), **3**, e2119.

88 Maekawa, T., Kufer, T. A. & Schulze-Lefert, P. NLR functions in plant and animal immune systems: so far and yet so close. Nature Immunol (**2011**), **12** 818-826.

89 van Ooijen, G., van den Burg, H. A., Cornelissen, B. J. & Takken, F. L. Structure and function of resistance proteins in solanaceous plants. Annu Rev Phytopathol (**2007**), **45**, 43-72.

90 Kale, S. M., Pardeshi, V. C., Barvkar, V. T., Gupta, V. S. & Y., K. N. Genome-wide identification and characterization of nucleotide binding site leucine-rich repeat genes in linseed reveal distinct patterns of gene structure. Genome (**2013**), **56**, 91–99.

91 Tan, X., Meyers, B. C., Kozik, A., West, M. A., Morgante, M., St Clair, D. A., Bent, A. F. & Michelmore, R. W. Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in Arabidopsis. BMC Plant Biol (**2007**), **7**, 56.

92 Wan, H., Yuan, W., Ye, Q., Wang, R., Ruan, M., Li, Z., Zhou, G., Yao, Z., Zhao, J., Liu, S. & Yang, Y. Analysis of TIR- and non-TIR-NBS-LRR disease resistance gene analogous in pepper: characterization, genetic variation, functional divergence and expression patterns. BMC Genomics (**2012**), **13**, 3-15.

93 Maekawa, T., Cheng, W., Spiridon, L. N., Toller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q. H., Micluta, M. A., Somssich, I. E., Takken, F. L., Petrescu, A. J., Chai, J. & Schulze-Lefert, P. Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. Cell Host Microbe. (**2011**), **9**, 187-199.

94 Rairdan, G. J. & Moffett, P. Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation. Plant Cell (**2006**), **18**, 2082-2093.

95 Maekawa, T., Kracher, B., Vernaldi, S., van Themaat, E. V. L. & Schulze-Lefert, P. Conservation of NLR-triggered immunity across plant lineages. PNAS (**2012**), **109**, 20119–20123.

96 Dinesh-Kumar, S. P., Wai-HongTham & Baker, B. J. Structure–function analysis of the tobacco mosaic virus resistance gene N. PNAS (**2000**), **97**, 14789–14794.

97 Chan, S. L., Mukasa, T., Santelli, E., Low, L. Y. & Pascual, J. The crystal structure of a TIR domain from Arabidopsis thaliana reveals a conserved helical region unique to plants. Protein Sci (**2010**), **19**, 155-161.

98 Qi, D., DeYoung, B. J. & Innes, R. W. Structure-function analysis of the coiled-coil and leucine-rich repeat domains of the RPS5 disease resistance protein. Plant Physiol (**2012**), **158**, 1819-1832.

99 Williams, S. J., Sohn, K. H., Wan, L., Bernoux, M., Sarris, P. F., Segonzac, C., Ve, T., Ma, Y., Saucet, S. B., Ericsson, D. J., Casey, L. W., Lonhienne, T., Winzor, D. J., Zhang, X., Coerdt, A., Parker, J. E., Dodds, P. N., Kobe, B. & Jones, J. D. Structural basis for assembly and function of a heterodimeric plant immune receptor. Science (**2014**), **344**, 299-303.

100 Bernoux, M., Ve, T., Williams, S., Warren, C., Hatters, D., Valkov, E., Zhang, X., Ellis, J. G., Kobe, B. & Dodds, P. N. Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signaling, and autoregulation. Cell Host & Microbe (**2011**), **9**, 200-211.

101 Casey, L. W., Lavrencic, P., Bentham, A. R., Cesari, S., Ericsson, D. J., Croll, T., Turk, D., Anderson, P. A., Mark, A. E., Dodds, P. N., Mobli, M., Kobe, B. & Williams, S. J. The CC domain structure from the wheat stem rust resistance protein Sr33 challenges paradigms for dimerization in plant NLR proteins. PNAS (**2016**), **113**, 12856-12861.

102 Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., Wang, Q., Xiao, W., Liu, Y., Chai, J., Takken, F. L., Schulze-Lefert, P. & Shen, Q. H. Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance. PLoS Pathog (**2012**), **8**, e1002752.

103 Hao, W., Collier, S. M., Moffett, P. & Chai, J. Structural basis for the interaction between the potato virus X resistance protein (Rx) and its cofactor Ran GTPase-activating protein 2 (RanGAP2). J Biol Chem (**2013**), **288**, 35868-35876.

104 Qi, D. & Innes, R. W. Recent Advances in Plant NLR Structure, Function, Localization, and Signaling. Frontiers in Immunology (**2013**), **4**, 348.

105 Mestre, P. & Baulcombe, D. C. Elicitor-mediated oligomerization of the tobacco N disease resistance protein. Plant Cell (**2006**), **18**, 491-501.

106 Zhang, Y., Dorey, S., Swiderski, M. & Jones, J. D. Expression of RPS4 in tobacco induces an AvrRps4-independent HR that requires EDS1, SGT1 and HSP90. Plant J (**2004**), **40**, 213-224.

107 Manon, F., Favier, A., Núñez, G., Simorre, J. P. & Cusack, S. Solution structure of NOD1 CARD and mutational analysis of its interaction with the CARD of downstream kinase RICK. J Mol Biol (**2007**), **365**, 160-174.

108 Mayle, S., Boyle, J. P., Sekine, E., Zurek, B., Kufer, T. A. & Monie, T. P. Engagement of nucleotide-binding oligomerization domain-containing protein 1 (NOD1) by receptor-interacting protein 2 (RIP2) is insufficient for signal transduction. J Biol Chem (**2014**), **289**, 22900-22914.

109 Kufer, T. A., Kremmer, E., Adam, A. C., Philpott, D. J. & Sansonetti, P. J. The pattern‐recognition molecule Nod1 is localized at the plasma membrane at sites of bacterial interaction. Cell Microbiol (**2008**), **10**, 477-486.

110 van Der Biezen, E. A. & Jones, J. D. G. Plant disease-resistance proteins and the gene-for-gene concept. Trends Biochem Sci (**1998**), **23**, 454-456.

111 Chattopadhyaya, R. & Pal, A. Three-dimensional models of NB-ARC domains of disease resistance proteins in tomato, Arabidopsis, and flax. J Biomol Struct Dyn, **25**, 357-371.

112 Slootweg, E. J., Spiridon, L. N., Roosien, J., Butterbach, P., Pomp, R., Westerhof, L., Wilbers, R., Bakker, E., Bakker, J., Petrescu, A. J., Smant, G. & Goverse, A. Structural determinants at the interface of the ARC2 and leucine-rich repeat domains control the activation of the plant immune receptors Rx1 and Gpa2. Plant Physiol (**2013**), **162**, 1510-1528.

113 Tai, T. H., Dahlbeck, D., Clark, E. T., Gajiwala, P., Pasion, R., Whalen, M. C., Stall, R. E. & Staskawicz, B. J. Expression of the Bs2 pepper gene confers resistance to bacterial spot disease in tomato. PNAS (**1999**), **96**, 14153-14158.

114 van der Vossen, E. A., van der Voort, J. N., Kanyuka, K., Bendahmane, A., Sandbrink, H., Baulcombe, D. C., Bakker, J., Stiekema, W. J. & Klein-Lankhorst, R. M. Homologues of a single resistance-gene cluster in potato confer resistance to distinct pathogens: a virus and a nematode. Plant J (**2000**), **23**, 567-576.

115 Kajava, A. V. Structural diversity of leucine-rich repeat proteins1. J Mol Biol (**1998**), **277**, 519-527.

116 Reubold, T. F., Hahne, G., Wohlgemuth, S. & Eschenburg, S. Crystal structure of the leucine-rich repeat domain of the NOD-like receptor NLRP1: implications for binding of muramyl dipeptide. FEBS Lett (**2014**), **588**, 3327-3332.

117 Hong, M., Yoon, S. I. & Wilson, I. A. Structure and functional characterization of the RNA-binding element of the NLRX1 innate immune modulator. Immunity (**2012**), **36**, 337-347.

118 van Ooijen, G., Lukasik, E., Van Den Burg, H. A., Vossen, J. H., Cornelissen, B. J. & Takken, F. L. The small heat shock protein 20 RSI2 interacts with and is required for stability and function of tomato resistance protein I-2. Plant J (**2010**), **63**, 563-572.

119 Peart, J. R., Cook, G., Feys, B. J., Parker, J. E. & Baulcombe, D. C. An EDS1 orthologue is required for N-mediated resistance against tobacco mosaic virus. Plant J (**2002**), **29**, 569-579.

120 Ito, M., Ohnishi, K., Hikichi, Y. & Kiba, A. Molecular chaperons and co-chaperons, HSP90, RAR1, and SGT1 negatively regulate bacterial wilt disease caused by *Ralstonia solanacearum* in *Nicotiana benthamiana*. Plant Signal Behav (**2014**), **10**, e970410.

121 Bieri, S., Mauch, S., Shen, Q. H., Peart, J., Devoto, A., Casais, C., Ceron, F., Schulze, S., Steinbiß, H. H. & Shirasu, K. RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. The Plant Cell Online (**2004**), **16**, 3480-3495.

122 Chisholm, S. T., Mahajan, S. K., Whitham, S. A., Yamamoto, M. L. & Carrington, J. C. Cloning of the Arabidopsis RTM1 gene, which controls restriction of long-distance movement of tobacco etch virus. PNAS (**1999**), **97**, 489–494.

123 Whitham, S. A., Anderberg, R. J., Chisholm, S. T. & Carrington, J. C. Arabidopsis RTM2 gene is necessary for specific restriction of tobacco etch virus and encodes an unusual small heat shock-like protein. Plant Cell (**2000**), **12**, 569-582.

124 da Silva Correia, J., Miranda, Y., Leonard, N. & Ulevitch, R. SGT1 is essential for Nod1 activation. PNAS (**2007**), **104**, 6764-6769.

125 Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., Pevzner, S. J., Donovan, S. E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M. M., Gebreab, F., Gutierrez, B. J., Tam, S., Monachello, D., Boxem, M., Harbort, C. J., McDonald, N., Gai, L., Chen, H., He, Y., European Union Effectoromics, C., Vandenhaute, J., Roth, F. P., Hill, D. E., Ecker, J. R., Vidal, M., Beynon, J., Braun, P. & Dangl, J. L. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science (**2011**), **333**, 596-601.

126 Narusaka, M., Kubo, Y., Hatakeyama, K., Imamura, J., Ezura, H., Nanasato, Y., Tabei, Y., Takano, Y., Shirasu, K. & Narusaka, Y. Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. PLoS One (**2013**), **8**, e55954.

127 Pompe-Novak, M., Gruden, K., Baebler, Š., Krečič-Stres, H., Kovač, M., Jongsma, M. & Ravnikar, M. Potato virus Y induced changes in the gene expression of potato (Solanum tuberosum L.). Physiol Mol Plant Pathol (**2005**), **67**, 237-247.

128 Pompe-Novak, M., Gruden, K., Baebler, S., Krecˇicˇ-Stres, H., Kovacˇ, M., Maarten Jongsma, M. & Ravnikar, M. Potato virus Y induced changes in the gene expression of potato (Solanum tuberosum L.). Physiol Mol Plant Pathol (**2005**), **67**, 237-247.

129 McGregor, C. E., Miano, D. W., LaBonte, D. R., Hoy, M., Clark, C. A. & Rosa, G. J. M. Differential gene expression of resistant and susceptible sweetpotato plants after infection with the causal agents of sweet potato virus disease. J Am Soc Hortic Sci (**2009**), **134**, 658-666.

130 Kokkinos, C. D., Clark, C. A., MvGregor, C. E. & LaBonte, D. R. The effect of sweet potato virus disease and its viral components on gene expression levels in sweetpotato J. Amer. Soc. Hort (**2006**), **131**, 657-666.

131 Kogovsˇek, P., Pompe-Novak, M., Baebler, S., Rotter, A., Gow, L., Gruden, K., Foster, G. D., Boonham, N. & Ravnikar, M. Aggressive and mild Potato virus Y isolates trigger different specific responses in susceptible potato plants. Plant Pathol (**2010**), **59**, 1121-1132.

132 Pacheco, R., García-Marcos, A., Manzano, A., García de Lacoba, M., Camañes, G., García-Agustín, P., Díaz-Ruíz, J. R. & Tenllado, F. Comparative analysis of transcriptomic and hormonal responses to compatible and incompatible plant-virus interactions that lead to cell death. MPMI (**2012**), **25**, 709-723.

133 Dardick, C. Comparative expression profiling of Nicotiana benthamiana leaves systemically infected with three fruit tree viruses. MPMI (**2007**), **20**, 1004-1017.

134 Yang, C., Guo, R., Jie, F., Nettleton, D., Peng, J., Carr, T., Yeakley, J. M., Fan, J. & Whitham, S. A. Spatial Analysis of Arabidopsis thaliana Gene Expression in Response to Turnip mosaic virus Infection. MPMI (**2007**), **20**, 358 - 370.

135 Shimizu, T., Satoh, K., Kikuchi, S. & Omura, T. The Repression of Cell Wall- and Plastid-Related Genes and the Induction of Defense-Related Genes in Rice Plants Infected with Rice dwarf virus. MPMI (**2007**), **20**, 247-254.

136 Marathe, R., Guan, Z., Anandalakshmi, R., Zhao, H. & Dinesh-Kumar, S. P. Study of Arabidopsis thaliana resistome in response to cucumber mosaic virus infection using whole genome microarray. Plant Mol Biol (**2004**), **55**, 501-520.

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

138 Gonzalez-Ibeas, D., Cañizares, J. & Aranda, M. A. Microarray analysis shows that recessive resistance to watermelon mosaic virus in melon is associated with the induction of defense response genes. MPMI (**2012**), **25**, 107-118.

139 Peng, H., Wang, L., Li, Y., Li, Y., Guan, W., Yang, Y., Xu, X. & Hou, X. Differential gene expression in incompatible interaction between turnip mosaic virus and non-heading Chinese cabbage. Euro J Plant Pathol (**2011**), **132**, 393-406.

140 Kim, B. M., Suehiro, N., Natsuaki, T., Inukai, T. & Masuta, C. The P3 protein of turnip mosaic virus can alone induce hypersensitive response-like cell death in Arabidopsis thaliana carrying TuNI. MPMI (**2010**), **23**, 144-152.

141 Parker, J. E., Coleman, M. J., Szabò, V., Frost, L. N., Schmidt, R., van der Biezen, E. A., Moores, T., Dean, C., Daniels, M. J. & Jones, J. D. The Arabidopsis downy mildew resistance gene RPP5 shares similarity to the toll and interleukin-1 receptors with N and L6. The Plant Cell (**1997**), **9**, 879-894.

142 Lukhovitskaya, N. I., Ignatovich, I. V., Savenkov, E. I., Schiemann, J., Morozov, S. Y. & Solovyev, A. G. Role of the zinc-finger and basic motifs of chrysanthemum virus B p12 protein in nucleic acid binding, protein localization and induction of a hypersensitive response upon expression from a viral vector. J Gen Virol (**2009**), **90**, 723-733.

143 Takatsuji, H. Zinc-finger transcription factors in plants. Cellular and Molecular Life Sciences : CMLS (**1998**), **54**, 582-596.

144 Lippuner, V., Cyert, M. S. & Gasser, C. S. Two Classes of Plant cDNA Clones Differentially Complement Yeast Calcineurin Mutants and Increase Salt Tolerance of Wild-type Yeast. J Biol Chem (**1996**), **271**, 12859-12866.

145 Mukhopadhyay, A., Vij, S. & Tyagi, A. K. Overexpression of a zinc-finger protein gene from rice confers tolerance to cold, dehydration, and salt stress in transgenic tobacco. PNAS (**2004**), **16**, 6309-6314.

146 Sakamoto, H., Maruyama, K., Sakuma, Y., Meshi, T., Iwabuchi, M., Shinozaki, K. & Yamaguchi-Shinozaki, K. Arabidopsis Cys2/His2-Type Zinc-Finger Proteins Function as Transcription Repressors under Drought, Cold, and High-Salinity Stress Conditions. Plant Physiol (**2004**), **136**, 2734-2746.

147 Kim, Y. O., Kim, J. S. & Kang, H. Cold-inducible zinc finger-containing glycine-rich RNA-binding protein contributes to the enhancement of freezing tolerance in Arabidopsis thaliana. Plant Journal (**2005**), **42**, 890-900.

148 Sugano, S., Kaminaka, H., Rybka, Z., Catala, R., Salinas, J., Matsui, K., Ohme-Takagi, M. & Takatsuji, H. Stress-responsive zinc finger gene ZPT2-3 plays a role in drought tolerance in petunia. Plant J (**2003**), **36**, 830-841.

149 Rizhsky, L., Davletova, S., Liang, H. & Mittler, R. The zinc finger protein Zat12 is required for cytosolic ascorbate peroxidase 1 expression during oxidative stress in Arabidopsis. J Biol Chem (**2004**), **279**, 11736-11743.

150 Iida, A., Kazuoka, T., Torikai, S., Kikuchi, H. & Oeda, K. A zinc finger protein RHL41 mediates the light acclimatization response in Arabidopsis. Plant J (**2000**), **24**, 191-203.

151 Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.-S., Han, B., Zhu, T., Zou, G. & Katagiri, F. Quantitative Nature of Arabidopsis Responses during Compatible and Incompatible Interactions with the Bacterial Pathogen Pseudomonas syringae. The Plant Cell (**2003**), **15**, 317-330.

152 Park, C. J. & Seo, Y. S. Heat Shock Proteins: a review of the molecular chaperones for plant immunity. Plant Pathol (**2015**), **31**, 323-333.

153 Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. FASEB J (**2004**), **18**, 1169 - 1175.

154 Lashley, T., L., H. J., Verbeek, M. M., Rostagno, A., Bojsen-Moller, M., David, G., van Horssen, J., Braendgaard, H., Plant, G., Frangione, B., Ghiso, J. & Revesz, T. Molecular chaperons, amyloid and preamyloid lesions in the BRI2 gene-related dementias: a morphological study. Neuropathol Appl Neurobiol. (**2006**), **32**, 492-504.

155 Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., Peng, L., Zhang, L. & Lu, C. Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. Plant Cell (**2013**), **25**, 2925-2943.

156 Mayer, F. L., Wilson, D., Jacobsen, I. D., Miramón, P., Slesiona, S., Bohovych, I. M., Brown, A. J. P. & Hube, B. Small but Crucial: the novel small heat shock protein HSP21 mediates stress adaptation and virulence in *Candida albicans*. PLoS ONE (**2012**), **7**, e38584.

157 Neta-Sharir, I., Isaacson, T., Lurie, S. & Weiss, D. Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. Plant Cell (**2005**), **17**, 1829-1838.

158 Sels, J., Mathys, J., De Coninck, B., Cammue, B. & De Bolle, M. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. Plant Physiol Biochem (**2008**), **46**, 941 - 950.

159 van Loon, L. C., Gerritsen, Y. A. & Ritter, C. E. Identification, purification, and characterization of pathogenesis-related proteins from virus-infected Samsun NN tobacco leaves. Plant Mol Biol (**1987**), **9**, 593-609.

160 Agarwal, P., Bhatt, V., Singh, R., Das, M., Sopory, S. & Chikara, J. Pathogenesis-related gene, JcPR-10a from Jatropha curcas exhibit RNase and antifungal activity. Mol Biotechnol. (**2013**), **54**, 412 - 425.

161 Li, X., Xia, B., Jiang, Y., Wu, Q., Wang, C., He, L., Peng, F. & Wang, R. A new pathogenesis-related protein, LrPR4, from *Lycoris radiata*, and its antifungal activity against *Magnaporthe grisea*. Mol Biol Rep (**2010**), **37**, 995 - 1001.

162 Park, C. J., Kim, K. J., Shin, R., Park, J. M., Shin, Y. C. & Paek, K. H. Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. Plant J: Cell Mol Biol (**2004**), **37**, 186-198.

163 Guevara-Morato, M. A., Garcia de Lacoba, M., Garcia-Luque, I. & Serra, M. T. Characterization of a pathogenesis-related protein 4 (PR-4) induced in *Capsicum chinense* L3 plants with dual RNase and DNase activities. J Exp Bot (**2010**), **61**, 3259 - 3271.

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

165 Christiansen, M. W. & Gregersen, P. L. Members of the barley NAC transcription factor gene family show differential co-regulation with senescence-associated genes during senescence of flag leaves. J Exp Bot (**2014**), **65**, 4009-4022.

166 Takasaki, H., Maruyama, K., Kidokoro, S., Ito, Y., Fujita, Y., Shinozaki, K., Yamaguchi-Shinozaki, K. & Nakashima, K. The abiotic stress-responsive NAC-type transcription factor OsNAC5 regulates stress-inducible genes and stress tolerance in rice. Mol Genet Genomics (**2010**), **284**, 173-183.

167 Kaneda, T., Taga, Y., Takai, R., Iwano, M., Matsui, H., Takayama, S., Isogai, A. & Che, F. S. The transcription factor OsNAC4 is a key positive regulator of plant hypersensitive cell death. The EMBO Journal (**2009**), **28**, 926-936.

168 Liu, B., Ouyang, Z., Zhang, Y., Li, X., Hong, Y., Huang, L., Liu, S., Zhang, H., Li, D. & Song, F. Tomato NAC transcription factor SlSRN1 positively regulates defense response against biotic stress but negatively regulates abiotic stress response. PLoS One (**2014**), **9**, e102067.

169 Mao, X., Zhang, H., Qian, X., Li, A., Zhao, G. & Jing, R. TaNAC2, a NAC-type wheat transcription factor conferring enhanced multiple abiotic stress tolerances in Arabidopsis. J Exp Bot (**2012**).

170 Liang, C., Wang, Y., Zhu, Y., Tang, J., Hu, B., Liu, L., Ou, S., Wu, H., Sun, X., Chu, J. & Chu, C. OsNAP connects abscisic acid and leaf senescence by fine-tuning abscisic acid biosynthesis and directly targeting senescence-associated genes in rice. PNAS (**2014**), **111**, 10013-10018.

171 Donze, T., Qu, F., Twigg, P. & Morris, T. J. *Turnip crinkle virus* coat protein inhibits the basal immune response to virus invasion in Arabidopsis by binding to the NAC transcription factor TIP. Virology (**2014**), **449**, 207-214.

172 Nakashima, K., Tran, L. S., Van Nguyen, D., Fujita, M., Maruyama, K., Todaka, D., Ito, Y., Hayashi, N., Shinozaki, K. & Yamaguchi-Shinozaki, K. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. Plant J (**2007**), **51**, 617-630.

173 Voitsik, A. M., Muench, S., Deising, H. B. & Voll, L. M. Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. BMC Plant Biol (**2013**), **13**, 85.

174 Du, M., Zhai, Q., Deng, L., Li, S., Li, H., Yan, L., Huang, Z., Wang, B., Jiang, H., Huang, T., Li, C. B., Wei, J., Kang, L., Li, J. & Li, C. Closely related NAC transcription factors of tomato differentially regulate stomatal closure and reopening during pathogen attack. Plant Cell (**2014**), **26**, 3167-3184.

175 Banerjee, A. & Roychoudhury, A. WRKY proteins: signaling and regulation of expression during abiotic stress responses. Scientific World J (**2015**), **2015**, 17.

176 Chen, Y., Bates, D. L., Dey, R., Chen, P. H., Machado, A. C., Laird-Offringa, I. A., Rohs, R. & Chen, L. DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. Cell Reports (**2012**), **2**, 1197-1206.

177 Pandey, S. P. & Somssich, I. E. The role of WRKY transcription factors in plant immunity. Plant Physiol (**2009**), **150**, 1648-1655.

178 Eulgem, T. & Somssich, I. E. Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol (**2007**), **10**, 366-371.

179 Naoumkina, M. A., He, X. & Dixon, R. A. Elicitor-induced transcription factors for metabolic reprogramming of secondary metabolism in Medicago truncatula. BMC Plant Biol (**2008**), **8**, 132-132.

180 Huh, S. U., Lee, G. J., Jung, J. H., Kim, Y., Kim, Y. J. & Paek, K. H. Capsicum annuum transcription factor WRKYa positively regulates defense response upon TMV infection and is a substrate of CaMK1 and CaMK2. Sci Rep (**2015**), **5**, 7981.

181 Tao, Z., Liu, H., Qiu, D., Zhou, Y., Li, X., Xu, C. & Wang, S. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. Plant Physiol (**2009**), **151**, 936-948.

182 Tang, J., Wang, F., Wang, Z., Huang, Z., Xiong, A. & Hou, X. Characterization and co-expression analysis of WRKY orthologs involved in responses to multiple abiotic stresses in Pak-choi (Brassica campestris ssp. chinensis). BMC Plant Biol (**2013**), **13**, 188.

183 Jiang, Y., Duan, Y., Yin, J., Ye, S., Zhu, J., Zhang, F., Lu, W., Fan, D. & Luo, K. Genome-wide identification and characterization of the Populus WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses. J Exp Bot (**2014**), **65**, 6629-6644.

184 Hundertmark, M. & Hincha, D. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics (**2008**), **9**, 118.

185 Liu, Y. & Zheng, Y. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. Biochem Biophys Res Commun (**2005**), **311**, 325 - 332.

186 Shih, M., Lin, S., Hsieh, J., Tsou, C., Chow, T., Lin, T. & Hsing, Y. Gene cloning and characterization of a soybean (*Glycine max* L.) LEA protein, GmPM16. Plant Mol Biol (**2004**), **56**, 689 - 703.

187 Gal, T., Glazer, I. & Koltai, H. An LEA group 3 family member is involved in survival of C. elegans during exposure to stress. FEBS Lett (**2004**), **577**, 21 - 26.

188 Goyal, K., Walton, L. & Tunnacliffe, A. LEA proteins prevent protein aggregation due to water stress. Biochem J (**2005**), **388**, 151 - 157.

189 Hand, S. C., Jones, D., Menze, M. A. & Witt, T. L. Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. J Exp Zool (**2007**), **307A**, 62 - 66.

190 Kikawada, T., Nakahara, Y., Kanamori, Y., Iwata, K., Watanabe, M., McGee, B., Tunnacliffe, A. & Okuda, T. Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. Biochem Biophys Res Comm (**2006**), **348**, 56 - 61.

191 Tunnacliffe, A., Lapinski, J. & McGee, B. A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. Hydrobiologia (**2005**), **546**, 315 - 321.

192 Bai, Y., Yang, Q., Kang, J., Sun, Y., Gruber, M. & Chao, Y. Isolation and functional characterization of a *Medicago sativa* L. gene, MsLEA3-1. Mol Biol Rep (**2012**), **39**, 2883-2892.

193 Dure, L., Crouch, M., Harada, J., Ho, T.-H. D., Mundy, J., Quatrano, R., Thomas, T. & Sung, Z. R. Common amino acid sequence domains among the LEA proteins of higher plants. Plant Mol Biol (**1989**), **12**, 475 - 486.

194 Hong-Bo, S., Zong-Suo, L. & Ming-An, S. LEA proteins in higher plants: structure, function, gene expression and regulation. Colloids and Surfaces. B, Biointerfaces (**2005**), **45**, 131-135.

195 Manfre, A. J., Lanni, L. M. & Marcotte, W. R. J. The Arabidopsis group 1 LATE EMBRYOGENESIS ABUNDANT protein ATEM6 is required for normal seed development. Plant Physiol (**2006**), **140**, 140 - 149.

196 Kang, B. C., Yeam, I. & Jahn, M. M. Genetics of plant virus resistance. Annu Rev Phytopathol (**2005**), **43**, 581-621.

197 Xu, P., Chen, F., Mannas, J. P., Feldman, T., Sumner, L. W. & Roossinck, M. J. Virus infection improves drought tolerance. New Phytol (**2008**), **180**, 911-921.

198 Swire-Clark, G. A. & Marcotte, W. R. The wheat LEA protein Em functions as an osmoprotective molecule in Saccharomyces cerevisiae. Plant Mol Biol (**1999**), **39**, 117 - 128.

Chapter 2

Rate of multiplication and spread determines resistance/susceptibility to UCBSV

T. Amuge, M. E. Ferguson, A. A. Myburg and D. K. Berger

Author Contributions

T.A.: Conducted all field, greenhouse (planted experimental plants, graft-inoculated and collected samples) and laboratory procedures (identified inoculum, isolated RNA and prepared cDNA libraries). Performed data analysis, wrote the initial draft of the chapter and addressed comments raised by supervisors.

M.E.F., A.A.M. and D.K.B.: Supervised the study and reviewed the chapter

2.1 Abstract

Accurate CBSV/UCBSV diagnostics, both RT-PCR and quantitative PCR, are essential for meaningful results in this thesis. Results presented here include a RT-PCR protocol for CBSV/UCBSV identification, which was locally validated using cDNA and crushed leaf samples of isolates from the Food and Environment Research Agency (FERA), UK. The locally validated protocol was then used to study UCBSV distribution in cassava plants and virus accumulation over time. Two cassava varieties; Namikonga, which is resistant to CBSD and Albert, a CBSD susceptible variety were assayed at 20 time points between six hours after graft inoculation (6 hag) to 54 days after graft inoculation (dag). Time points were divided into phases of presumed pre-systemic virus infection (6 hag, 12 hag, 1, 2, 3, 4, 5, 6, 7 and 8 dag) and presumed systemic virus infection (45, 46, 47, 48, 49, 50, 51, 52, 53 and 54 dag). At each time point, leaves were sampled from upper position (beginning with upper most open leaf, then going downwards) and lower position (beginning with the lower most leaf near the soil level, and moving upwards). Using RT-PCR, UCBSV was detected at all the ten time points within the presumed systemic phase for both Albert and Namikonga. However, UCBSV was only detectable at 6 dag, out of the ten time points within the pre-systemic phase. Throughout the time course, titre was higher in Albert compared to Namikonga. Highest titre, measured using qPCR was recorded in Albert at 5 dag (2.1 fold relative to transcripts of the UBQ10 cassava reference gene), the same time point where Namikonga had the lowest virus titre (0.2 fold). In both RT-PCR and real-time methods, detection was greater in the upper (young, actively growing leaves, near the graft inoculation point) leaves compared to lower (older, senescing leaves far from graft inoculation point). No single method of diagnosis was, by itself, effective in producing 100% accurate results. Therefore, biological and technical repetitions are recommended, together with the use of a combination of methods including phenotypic assessment of symptom expression.

2.2 Introduction

To elucidate plant resistance and virus dynamics in plants, sensitive and accurate methods of virus identification are essential. In this chapter, we validated an RT-PCR system for identifying CBSVs[1,2] on positive and negative plants using local and imported samples. Other methods, including phenotypic scoring of plants based on CBSD-associated symptoms and qPCR were also employed on some local test plants.

Unlike CMD, CBSD cannot be readily identified through visual symptoms alone, so molecular diagnostics were required. Many protocols have been published for both RT-PCR[3-5] and qPCR[3,5,6] for CBSVs. In 2011/12 the UK Government's Food and Environment Research Agency (FERA), which operates accredited diagnostics facilities, was tasked with overseeing the virus status of cassava planting material being distributed to >1 million farmers in the Great Lakes region (Uganda, Kenya, Tanzania, Rwanda, Burundi and DR Congo). There was need to standardize procedures across laboratories in the region and verify the accuracy of their diagnostics through proficiency testing. To achieve this, 12 laboratories, ours included, were selected to participate in an independent test of unknown samples from a central, independent laboratory contracted by PhytoPAS (Plant health performance assessment scheme), FERA. The results of the proficiency test in our laboratory are reported here.

Few studies have examined within-host distribution and accumulation of CBSVs. Ogwok et al. (2014)[7] studied correlation of CBSD symptoms, and virus titre in leaf, stem and storage root samples of ten cassava varieties naturally infected with CBSVs. The study showed that in tolerant varieties, titre was higher in roots compared to leaves and stems while susceptible varieties had high titre in their leaves, stems and roots. In another study of naturally infected samples, Kaweesi et al. (2014)[8] found that both viruses accumulated over time when tested at 3, 5, 7, 9 and 11 months after planting. Throughout the experiment, UCBSV had lower titre than CBSV in both resistant and susceptible varieties. However, resistant varieties maintained very low titre of viruses throughout the experiment, an indicator that resistance to CBSD is partly caused by reducing rate of virus multiplication. However, no known studies have focused on measuring virus titre with single infection of cassava with either virus species (CBSV or UCBSV).

In this chapter, we report, for the first time, a study designed to measure titre of UCBSV in two cassava varieties graft inoculated with UCBSV. Varieties include: Namikonga, a CBSD-resistant variety and Albert, which is susceptible to CBSD. The objective of the study was to determine the trend of UCBSV accumulation in cassava plants of Namikonga and Albert graft inoculated with UCBSV. The assumption is that UCBSV accumulates at significantly lower levels in Namikonga compared with the accumulation rate in the susceptible Albert.

2.3 Materials and methods

2.3.1 Validation of virus testing protocol through the Food Analysis Performance Assessment Scheme (FAPAS), UK: an overview of FERA and procedures taken

FAPAS (the Food Analysis Performance Assessment Scheme), which is part of Food and Environmental Research Agency (FERA), is the leading global provider of proficiency testing schemes, and implements schemes in disease diagnostics under PhytoPAS (Plant health performance assessment scheme)) (https://fapas.com/). PhytoPAS established a set of cassava test samples with known CBSV and/or UCBSV content including negative controls, which were distributed to laboratories for proficiency testing of their PCR diagnostics assay. This FERA panel was used to validate the RT-PCR diagnostic assay used throughout this thesis at BecA (Biosciences east and central Africa), Nairobi, Kenya.

Through PhytoPAS, at FERA, samples for proficiency testing were extracted from sap inoculated fresh leaves of cassava using a methodology that applied magnetic bead separation and purification. As desired, RNA extracts from infected CBSV or UCBSV samples were diluted with comparable solutes from virus-negative plants of the same variety, same age, same growth conditions, except for virus inoculation. FERA panel samples were stored in 'RNAstable' tubes, having been dried using a rotorvac and stored at room temperature.

As part of setting up the FERA panel, sets of 10 samples from each batch were analysed using in-house primers at FERA using qPCR to confirm that virus positive samples (CBSV and/or UCBSV) were indeed positive and virus negative samples were negative. Data was scrutinized for consistency at every level and confirmed fit for proficiency testing by FERA. To verify that in-transit degradation had not occurred on samples, a subset of samples were taken by hand to and from Uganda and then analysed using the same procedures with a set that had been in storage at FERA. FERA indicated that the results were similar (data not shown), confirming that each shipment would remain stable in transit to respective laboratories. Subsequently, respective sample subsets were shipped to the 27 candidate laboratories including our laboratory for proficiency testing.

At BecA, panels of samples from FAPAS (for PhytoPAS) were received for proficiency test PLO153 on November 2011 and January 2012. As outlined in their protocol (available on request from info@fapas.com or testmaterials@fapas.com), received samples (A-G) were reconstituted in 30 µl of PCR reaction water and left to set for 20 mins. Dissolved samples were mixed by gently pipetting up and down, followed by brief centrifugation. From sample H which was received as dried leaves, ~0.1g was used to extract RNA using the CTAB method[9]. About 200-250 ng of RNA was used for cDNA synthesis using Superscript III (Invitrogen, Inc). One microlitre of cDNA was used in a 25 µl PCR reaction with CBSDDR and CBSDDF2 (5'-GGATATGGAGAAAGRKCTCC-3', 5'-GCTMGAAATGCYGGRTAYACAA-3') primers[10]. In the presence of UCBSV or CBSV, the primers yield PCR products of 438–440 nt or 344 nt respectively. The PCR reaction involved initial denaturing at 95°C for 1 min, 35 cycles of 94°C, 52°C, 72°C and final extension at 72°C for

10 mins. The PCR reaction was performed using DreamTaq DNA polymerase (Thermo Scientific), following manufacturer's protocol. Products from PCR were resolved on 2% agarose gel with gel red (Biotium Inc). Except for sterile disposable PCR reaction tubes and pipette tips, all reagent bottles and tubes were sterilized by treatment with DEPC then autoclaved for 1 hour at 160°C. Baking at 180°C for two hours was performed for mortars and pestles only.

2.3.2 Assessment of within-host distribution and accumulation of CBSVs over time

Variety selection, sampling and phenotypic assessment for disease (CBSD) symptoms

Virus negative stems of Namikonga and Albert were obtained from cassava stems at Sugarcane Research Institute (SRI) in Kibaha, Dar es salaam, Tanzania in June 2010. Stems of inoculum source (variety NDL06/132) were obtained from natural, whitefly infected field samples at the same institute. Both virus negative and virus-positive stems were planted for multiplication in separate glass house compartments at BecA, Nairobi, Kenya. Albert was chosen as the susceptible check because it develops clear CBSD-characteristic symptoms on its leaves, stems (under field conditions) and storage roots upon infection by either species of viruses (CBSV or UCBSV),[4,8] Namikonga however is known for its resistance to CBSD across East Africa. Infected plants of Namikonga may develop leaf symptoms but their storage roots remain non-necrotic or very minimal necrosis may develop in a few plants[4,8,10-12]. The two varieties were chosen because of their contrasting characteristics and the ability to further elucidate the underlying factors governing Namikonga's resistance to CBSD. Plants were sampled at two positions below the graft point, based on the assumption that virus accumulation varies between young, actively growing leaves at uppermost leaf positions (UL) and older, senescing leaves sampled from lowermost leaf positions (LL) Fig. 2.1.

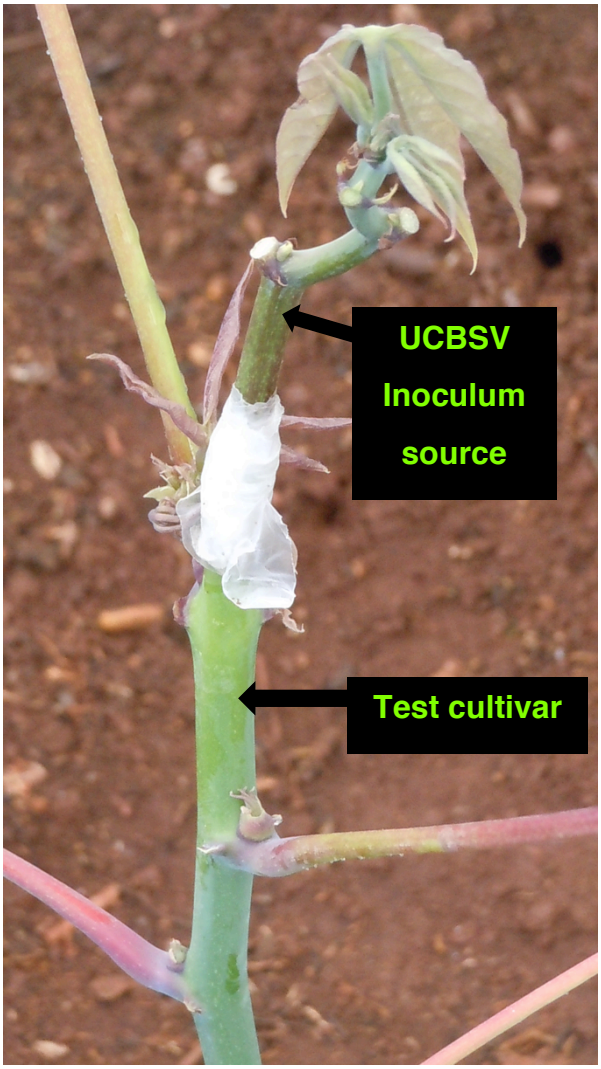


Fig. 2.1: Positions of leaf sampling from each biological replicate per variety.

2.3.3 Establishment of the time-course experiment

Establishment of the time-course experiment is fully described in Chapter 3. Initially, 45 cassava cuttings of each; Namikonga and Albert were planted in large pots in a glasshouse, and after two months were top grafted with scions either infected with CBSV, UCBSV or a combination of CBSV and UCBSV. For each variety, 15 plants were grafted with CBSV, ten plants with UCBSV, 10 plants with CBSV+UCBSV and 10 plants with virus-free scion. At 3 months after grafting (mag), plants were assessed for CBSD-related leaf chlorosis and presence of CBSVs using RT-PCR. In each variety, only three plants representing three biological replicates from each batch of CBSV-, UCBSV-, CBSV+UCBSV- and mock-inoculated plants were needed for further analysis. UCBSV-inoculated plants were selected for further analysis as is had the required number of biological replicates and represented a single virus species, hence, each of leaves were assessed for characteristic CBSD-related chlorosis using the 1-5 score[10,13-15]. Photographs of chlorotic leaves were taken for documentation.

Presence/absence of UCBSV was tested using RT-PCR at all 20 times points consisting presumed pre-systemic period of local infection (6 hag – 8 dag) and systemic infection (45 – 54 dag) while virus titre was

measured using real-time PCR at eight selected time points (6 hag, 1, 2, 5, 8, 45 and 54 dag). All plants were sampled at upper and lower leaf positions (Fig. 2.1).

2.3.4 RNA extraction, cDNA synthesis for RT-PCR and qPCR tests

RNA was extracted from fresh leaf samples using a CTAB protocol, with the amount of starting extraction buffer material modified from 600 µl to 800 µl[9]. Respective second-strand cDNAs were synthesized using Superscript III (Invitrogen, CA) according to manufacturer's instructions. RT-PCR was performed using primers that distinguish both viruses simultaneously; namely, CBSDDR and CBSDDF2[5] following the program and conditions described above in "*Validation of virus testing protocol through Food and Environmental Research Agency (FERA): an overview of FERA and procedures taken*".

qPCR

Virus titre was measured using Maxima SYBR Green/ROX qPCR master mix (2X) (Thermo Scientific) following the manufacturer's instructions. qPCR reactions were carried out on 7900HT Fast PCR system, using CBSDDR and CBSDDF2[5] primers. The qPCR program included initial denaturation at 94°C for 10 mins, 40 cycles of 94°C for 30s, 52°C for 1 min and 72°C for 2 mins. Reactions were held at 72°C or terminated. Data was collected at 72°C, corresponding to the exponential phase. All data were exported to MS Excel for filtering and statistical analyses. Each treatment had three technical and three biological replicates, from which mean Ct was obtained.

Virus titre was determined using the delta-delta Ct method[16]. This method determines Ct using the formula $(Ct, _{Target} - Ct, _{UBQ10})_{Time\ x} - (Ct, _{Target} - Ct, _{UBQ10})_{Time\ 0}$. For the cassava reference gene UBQ10 (5'-TGCATCTCGTTCTCCGATTG -3', 5'- GCGAAGATCAGTCGTTGTTGG -3')[17], Ct at time zero was not included in the qPCR runs but the Ct of UBQ10 at 6 hag was used as a baseline[16].

The standard curve method, combined with physical observation of amplification plots were used to determine the limit of detection (LOD). To generate the standard curved, six-point serial dilutions of "pooled-UCBSV-positive" samples were amplified using the above program. The undiluted "pooled-UCBSV-positive" sample was obtained by mixing 1 µl each of undiluted cDNA generated from all UCBSV positive samples of Albert, Namikonga and the inoculum source (NDL06/132). From the standard curve, samples within Ct 17.2 and 35.9 (supplementary fig. S2.1) were considered positive, and anything outside this range taken as background noise.

2.3.5 Sanger sequencing and alignment to UCBSV genomes

Samples amplified using CBSDDR and CBSDDF2 primers[5] were cleaned using AccuPrep PCR Purification *Kit (*Bioneer Inc, Smith's Parish, Bermuda). Cleaned PCR products were sequenced using ABI 3730 sequencer (Applied Biosystems corporation, Foster City, California, United States) with CBSDDR and CBSDDF2 reverse and forward primers. Sanger sequences were assessed for quality, trimmed and filtered using 4peaks software (Nucleobytes B.V. RA Aalsmeer, The Netherlands). Curated sequences were aligned to online available genomes of CBSV and UCBSV using NCBI's nucleotide BLAST. Sequences

were manually loaded one at a time and respective best hits recorded. Results of sanger sequencing are presented in Chapter 3[18].

2.4 Results

2.4.1 Validation of RT-PCR diagnostics using local and independently extracted cassava RNA samples from FERA

Results for samples labelled 1-3 (Fig. 2.2) are from the FERA leaf sample H (extracted in triplicate) alongside local samples (4-9), each extracted in duplicate (a and b) as shown in Table 2.1. Local samples were selected to demonstrate mixed infections, and single infections of CBSV and UCBSV. The results were as expected for samples 1-7, however samples 8 and 9 had unusually faint bands for UCBSV, contrary to the visibly clear band expected. Using FERA's samples (10-17), results for CBSV were 100% similar to the expected output by FERA, except for dual detection of both CBSV and UCBSV in samples 14 and 15 where only CBSV was expected. As expected by FERA (Table 2.1), samples 10, 11, 12, 13 and 17 were negative for CBSV as 14, 15 and 16 were positive for the same. For UCSBV, samples 10 and 12 were positive for UCBSV as expected by FERA. Sample 13 was negative for UCBSV on agarose gel, but according to FERA, it's expected to be positive with low titre of UCBSV. Samples 14 and 15 tested positive for UCBSV, contrary to the negative output expected by FERA. However, samples 11, 16 and 17 were negative for UCBSV, agreeing with FERA's expectation (Table 2.1). The blank sample (lane 18) had a feint CBSV band, contrary to the expected negative result.

Altogether, 90% of results of FERA's test samples were as expected by FERA (Table 2.1). A negative instead of positive result from sample 13, and positive instead of negative results for samples 14 and 15 caused the 10% difference. From local samples, 80% of the results were as expected (NB: Plants tested were artificially grafted inoculated with CBSV, UCBSV, CBSV+UCBSV). Faint bands of 440 bp (which represent UCBSV) observed in samples 8 and 9, which were expected to be as visibly clear as that observed in samples 6 and 7 caused the 20% uncertainty in results because faint bands were wrongly considered to be negative. Observation of false positive and false negative results highlight the need to further optimize conditions for virus testing within the laboratory, so that it can always meet the expected standards.

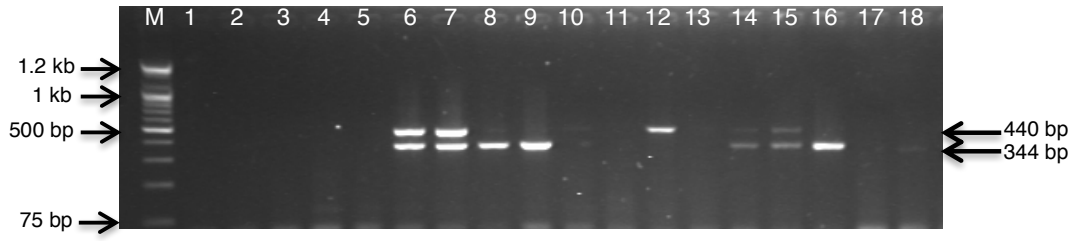


Fig. 2.2: Photograph taken from 2% agarose gel resolution of PCR products from RT-PCR of dual infected samples (CBSV and UCBSV) and singly infected samples (CBSV (344bp) and UCBSV (440bp)). Sampled were amplified using CBSDDR and CBSDDF2 primers[5]. Lane M: 1Kb+ ladder, lanes 1-3 are from FERA leaf samples extracted in triplicate at BecA; lanes 4-9 are local samples selected to demonstrate single and dual infections, lanes 10-17 are reconstituted FERA RNA samples while lane 18 is a non-template control.

Table 2.1: Results of PCR run with CBSDDR and CBSDDF2 primers on local samples (this study) and commercially extracted (FERA) cassava RNA samples.

Test material	Expected test result		Observed test result	
	UCBSV	CBSV	UCBSV	CBSV
1 (H1)	-	-	-	-
2 (H2)	-	-	-	-
3 (H3)	-	-	-	-
4 (N43 a)	-	-	-	-
5 (N43 b)	-	-	-	-
6 (N76 a)	+	+	+	+
7 (N76 b)	+	+	+	+
8 (N89 a)	+	+	+ (Very faint)	+
9 (N89 b)	+	+	+ (Very faint)	+
10 (A)	+	-	+	-
11 (B)	-	-	-	-
12 (C)	+	-	+	-
13 (D)	+ (Very low titre)	-	-	-
14 (E)	-	+	+	+
15 (F)	-	+	+	+
16 (G)	-	+	-	+
17 (H)	-	-	-	-
18 (blank)	-	-	-	+ (Very faint)

Results highlighted red are contrary to what was expected.

Selection of isolate

Graft inoculation was effective with dual (CBSV + UCBSV) infected scions in both Albert and Namikonga. In Namikonga, 60% of dual-virus inoculated plants were positive while 47% were positive in Albert (Table 2.2). With UCBSV, both Namikonga and Albert registered 70% infection. The lowest graft infection rate was observed with CBSV positive scions, where Albert had 47% infection and Namikonga only 7% infection. For sufficient statistical replication, at least three plants were required for further tests. Isolates UCBSV and CBSV+UCBSV had this required number of plants. However, only isolate UCBSV was selected for further

analysis while dual-virus inoculated plants (CBSV+UCBSV) were not considered because, in the absence of sufficient replicates of CBSV, it would be technically challenging to measure the contribution of each virus species. Sequences of the RT-PCR product from the inoculum source (Fig. 3.2, in Chapter 3) with the chosen UCBSV species, aligned to GenBank's accession KF878103.1 (UCBSV) with 98% identity, 100% query cover, E-value=3e179.

Table 2.2: Summary of results of RT-PCR and phenotyping done on, CBSV-, UCBSV-, CBSV+UCBSV-inoculated plants of Albert and Namikonga at 3 mag.

Variety and isolate used for grafting inoculation	No. of plants graft-inoculated[#]	No. of plants tested positive at 3 mag*	% Positive samples
Namikonga-CBSV	15	1	7
Namikonga-UCBSV	10	7	70
Namikonga-CBSV+UCBSV	10	9	90
Albert-CBSV	15	7	47
Albert-UCBSV	10	7	70
Albert-CBSV+UCBSV	10	7	70

[#]Scions of Cassava NDL06/132 were used to inoculate the CBSV and/or UCBSV onto rootstocks of Namikonga or Albert

* Gel photo of UCBSV- verses mock-inoculated plants are presented in Fig. 3.2

2.4.2 Virus identification on UCBSV-inoculated plants using PCR and leaf symptoms (chlorosis) at three months after graft inoculation (3 mag)

Alongside their respective controls, the three plants (representing three biological replicates) graft inoculated with UCBSV were tested at 3 mag for presence of UCBSV (gel photo not shown). According to leaf symptoms, two plants among Albert's UCBSV-inoculated plants were positive (Fig. 2.3A, 2.3B) while a third plant had unclear chlorosis (Fig. 2.3C, appearing like that caused by ageing. This third plant was classified as CBSD negative (-), based on leaf symptoms. Using RT-PCR, two out of three UCBSV-inoculated plants tested positive whereas a third plant, (Fig. 2.3B) was undetermined (na) as the sample spilled during RNA extraction, but that spilled out plant was positive (+) for CBSD leaf symptoms (Fig. 2.3B; Table 2.2). All mock inoculated plants of Albert were negative for UCBSV at both PCR and phenotype tests.

In Namikonga, all three UCBSV-inoculated plants tested positive for UCBSV either by PCR and/or symptoms. Based CBSD-related chlorosis on their leaves, two plants of UCBSV-inoculated Namikonga scored positive (Fig. 2.3E, 2.3G). One, out of three UCBSV-inoculated plants showed no CBSD-related chlorosis (Fig. 2.3F), and so was scored negative. At PCR, two plants were positive for UCBSV while one was negative. As in Albert, all mock-inoculated plants of Namikonga were negative for UCBSV as shown by PCR and phenotype test results.

Table 2.3: Results of RT-PCR and phenotyping done on UCBSV- and mock-inoculated plants of Albert and Namikonga at 3 mag.

Treatment	Biological replicate	Infecting virus	RT-PCR with CBDDR/CBSDDF2	Biological replicates as shown in (Fig. 2.3)	Photo diagnosis based on leaf chlorosis (Fig. 2.3)
Albert, UCBSV-inoculated	1	UCBSV	+	A	+
	2	UCBSV	(na)	B	+
	3	UCBSV	+	C	-
Albert, mock-inoculated	1	N	-	D	-
	2	N	-	ns	-
	3	N	-	ns	-
Namikonga, UCBSV-inoculated	1	UCBSV	-	E	+
	2	UCBSV	+	F	-
	3	UCBSV	+	G	+
Namikonga, mock-inoculated	1	N	-	H	-
	2	N	-	ns	-
	3	N	-	ns	-

RT-PCR tests were performed using CBSDDR and CBSDDF2 primers[5]. Sample (na) poured out during RNA extraction. **-** => Negative for UCBSV, **+** => positive for UCBSV. ns= photograph not shown.

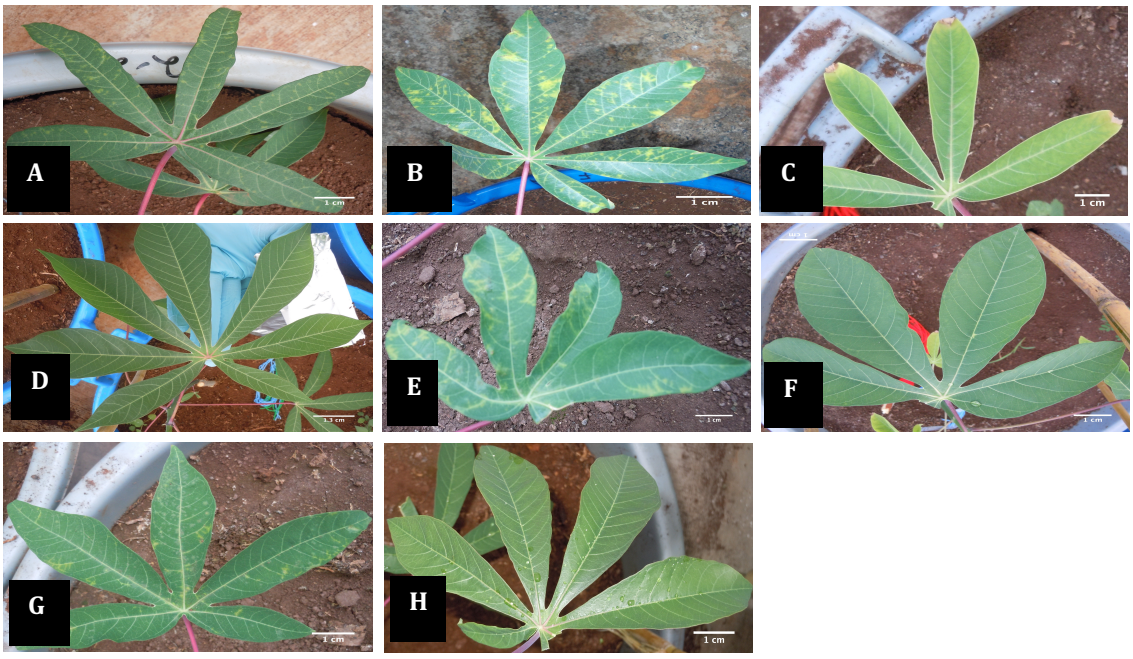


Fig. 2.3: UCBSV-inoculated (A-C), mock-inoculated (D) plants of Albert, and UCBSV-inoculated (E-G), mock-inoculated (H) plants of Namikonga at 3 mag. Photographs D and H are representative of mock-inoculated plants, all of which showed no observable CBSD-related chlorosis.

2.4.3 Identification of infecting virus

These results are based on leaves sampled directly below the graft point (UL) and those from lower senescing leaves near the soil level (LL) over 20 time points.

During the early sampling phase (6 hag through 8 dag), all three plants of both varieties were negative for UCBSV except at 6 dag when upper leaves of two of the three biological replicates of Albert tested positive (Fig. 2.4). At late sampling phase (45 through 54 dag), most upper leaves tested positive. Tests were performed using RT-PCR with CBSDDR and CBSDDF2 primers, as described in the methodology. In Albert, all three biological replicates tested positive for UCBSV in upper leaves at 47, 48, 49, 51 and 54 dag. At 45 and 50 dag, two of three biological replicates were positive in their upper leaves. On the lower leaves, Albert was positive at 45 and 49 dag, each with one of three biological replicates testing positive. The lower leaves of Albert were negative at all the other time points.

For Namikonga, only two biological replicates were positive for UCBSV in upper leaves at 45, 46, 51, 52, 53 and 54 dag. At 47 and 48 dag, only one biological replicate from upper leaves of Namikonga tested positive for UCBSV. However, all three biological replicates on upper leaves at 49 and 50 dag tested positive for UCBSV. From lower leaves of Namikonga, one of three biological replicates was UCBSV positive at 51 and 54 dag respectively.

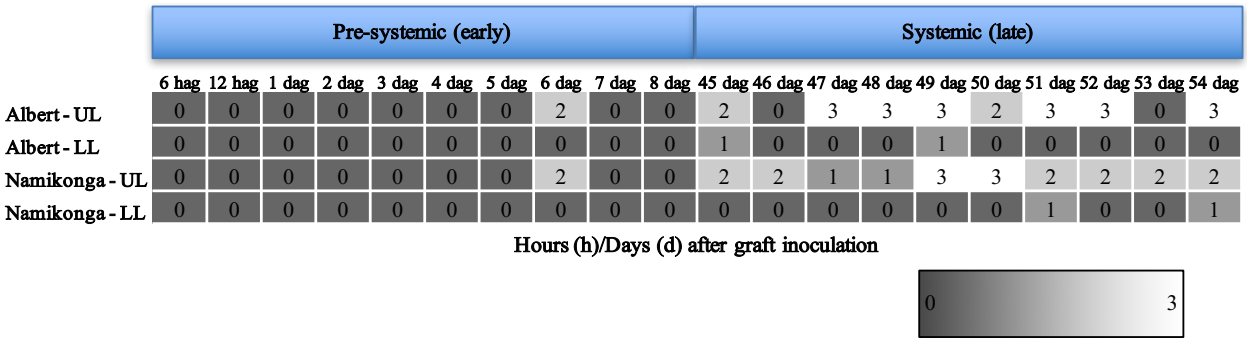


Fig. 2.4: Virus identification using RT-PCR of UCBSV-inoculated plants of Albert and Namikonga for 20 time points sampled at upper and lower leaf positions and amplified with CBSDDR and CBSDDF2 primers[5]. Numbers (0-3) on heat map represents the number of biological replicates that tested positive for UCBSV using these primers.

2.4.4 Relative virus titre in leaves of UCBSV-inoculated plants at UL and LL positions

Using the delta-delta Ct method, relative virus load relative to transcripts of a cassava housekeeping gene UBQ10 was determined in UCSBV-inoculated samples at seven selected time points (6 hag, 1, 2, 5, 8, 45 and 54 dag), the same time points used in Chapter 3[18]. In Albert, there was one unit of virus titre at UL and LL positions of UCBSV-inoculated plants relative to transcripts of the UBQ10 cassava reference gene in the same UCBSV-inoculated plants sampled at 6 hag from respective sampling positions. The units dropped at 1 dag, then took an upward trend from 2 dag, reaching the maximum of 2 units at 5 dag. From

8 dag, there was a gradual decline to 0.7 at 45 dag, before rising to 1 unit at 54 dag. The trend was similar for both UL and LL positions of Albert.

In Namikonga, UL and LL had different trends. For UL, one unit of virus titre was recorded at 6 hag, dropping considerably to 0.3 units at 5 dag. At 8 dag, 1.7 units were recorded, and it dropped again to <0.5 units by 54 dag. In the LL, there was an increasing trend beginning with 0.3 units at 6 hag and rising to 2.1 units at 45 dag, and later reducing to 0.5 units at 54 dag.

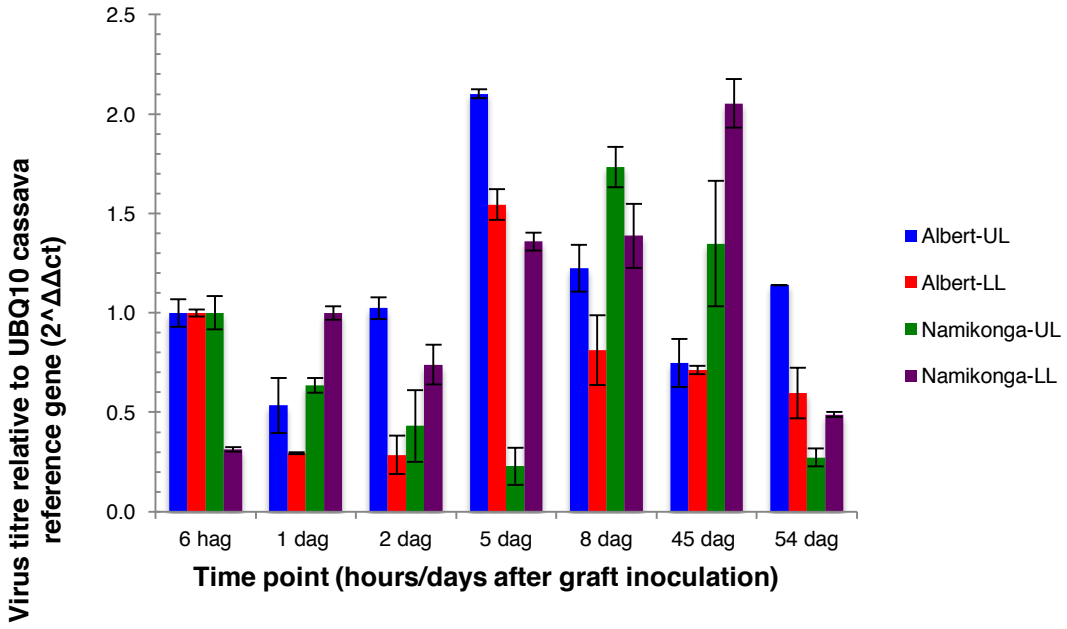


Fig. 2.5: Virus titre relative to transcripts of the UBQ10 cassava reference gene in UCBSV-inoculated plants of Albert and Namikonga sampled from upper (UL) and lower (LL) leaves at 6 hag, 1, 2, 5, 8, 45 and 54 dag, and measured using real-time qPCR.

2.5 Discussion

To date, the most robust, widely used method of identifying cassava brown streak disease (CBSD) causing viruses involves phenotypic scores of foliar and root symptoms. The other method involves use of PCR-based techniques. Some of the PCR-based techniques can distinguish between the two species; cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV). In this study, both phenotypic and PCR-based methods have been used to identify CBSV, UCBSV and CBSV+UCBSV after graft inoculation of two cassava varieties: Namikonga and Albert. Also, a qPCR-based method was used to determine virus titre in both varieties after inoculation with UCBSV.

2.5.1 UCBSV multiplies faster and more efficiently in Albert compared to Namikonga

The virus was detected in upper leaf samples of all three biological replicates of Albert at 47, 48, 51, 52 and 54 dag but Namikonga which had only one or two UCBSV-positive samples at the same time points (Fig. 2.3). A separate study showed that Namikonga had delayed accumulation of UCBSV even in 8-week-old plants[4].

Easier detection in all leaves of all three biological replicates of Albert at the systemic phase might coincide with a higher virus load in Albert. Detection in fewer biological replicates of Namikonga compared to biological replicates of Albert might be caused by low virus load caused by delayed, reduced or impaired multiplication of UCBSV in and around Namikonga's infection site. Delayed, reduced or impaired multiplication of virus particles is a resistance strategy used by many host plants against viruses[19-22]. In potato, resistance to phloem movement controls transport of potato leaf roll virus[23]. Some host plants use restricted virus movement as a strategy of defence. In Arabidopsis, **R**estricted **T**obacco Etch Virus **M**ovement (RTM) was shown to cause resistance to a wide range of viruses including Lettuce mosaic virus[24], Plum pox virus[25] and Tobacco etch virus (TEV)[20,21,26,27].

At the pre-systemic phase (Fig. 2.4), the virus was detected at 6 dag in the UL of both Albert and Namikonga, and not detected for the next two days (7 and 8 dag) in the same plants. This could indicate a non-uniform distribution in the plant[7]. It is also possible, as others have reported[28] in phloem transported viruses, that UCBSV inoculum multiplies within and around the infection site and the virus progressively moves to distant organs, and so, may remain temporarily low at the infection site. In the late time phase (systemic phase), the virus is quite well detected. Detection of UCBSV in more plants of Albert suggests that the virus multiplies faster in Albert compared to Namikonga.

2.5.2 Virus multiplies faster in upper, actively growing leaves compared to multiplication rates in lower senescing leaves

Detection of UCBSV was mostly in upper leaves of both varieties and phases. In Albert, the virus was detected in UL at 6, 45, 47, 48, 49, 50, 51, 52 and 54 dag. In LL however, UCBSV was only detectable at

45 and 49 dag. Curiously, Albert's UL and LL had undetectable levels of UCBSV at 46 and 53 dag. For Namikonga, UCBSV was detectable in UL at 6 dag and 45 through 54 dag. In lower leaves, it was only detectable at 51 and 54 dag.

These observations show that it's easier to detect UCBSV in upper leaves than in lower leaves of both Namikonga and Albert. Lower leaves are essentially older on the plant and are at more advanced stages of senescence as the plant grows. Therefore, undetectable levels of UCBSV in LL might have been caused by reduced virus multiplication as the leaf is at advanced stages of senescence, making it unable to sustain rapid virus proliferation compared to the younger, actively growing leaves[29]. In another study, Anjanappa[30] showed that Namikonga had higher virus load in the lower leaves as opposed to the upper leaves. It is possible that Namikonga's defence is more active in the upper than lower leaves. Moreover, under field conditions, Ogwok et al[7] recorded higher load in lower leaves of cassava infected with UCBSV.

2.5.3 Namikonga's resistance mechanism involves reduced virus load

Virus titre was generally higher in Albert (UL and LL) at all time points, except at 1 dag, when Namikonga was higher. At six out of seven time points measured, UL of Albert had higher virus titre than LL. However, Albert had its highest virus titre (UL and LL) at 5 dag, when Namikonga had the lowest titre in its UL (Fig. 2.5). According to Kaweesi et al.[8], virus titre is expected to be lower in the resistant variety, Namikonga compared to susceptible Albert, which is the case in this study. However, titre appears to fluctuate between time points, with some time points having very high titre while others have low titre. Fluctuation of titre between time points demonstrates that resistance in Namikonga, if it does involve reduced virus titre, occurs at specific time points, in this case 5 dag, where average virus titre in upper and lower leaves of UCBSV-inoculated Albert was four times higher than in Namikonga, and again at 54 dag (where average virus load in upper and lower leaves of UCBSV-inoculated Albert was two-times higher than in Namikonga). In another study, resistance to CBSD in variety KBH2006/8 was reported to result of inhibited virus replication[30].

Despite these interesting observations, these results are not conclusive as some samples showed very large standard deviation on qPCR results. Also, there was high virus titre in plants where the same virus was undetectably low with RT-PCR, revealing a possible flaw with qPCR over RT-PCR in detecting UCBSV.

The observably large standard deviation with qPCR could be caused by non-uniformity in of titre in samples as only one lobe was sampled at any time point. It is not known if virus titre is uniformly distributed within all leaves and leaf lobes of infected plants. Hence, sampling only fractions of a leaf, as done in this study may be restrictive because sampled fractions of a leaf may not have the true measure of virus titre. Therefore, future studies should consider sampling whole leaves for determination of virus titre or bulking leaves from several parts of the plant for accurate determination of load[31].

2.5.4 Comparing methods used for detecting UCBSV in this study

The RT-PCR protocol used in this study was validated using samples from FERA (Fig. 2.2 and Table 2.1), an internationally recognized proficiency testing organization. It confirms that indeed, the procedure is usable, albeit a few cases with false positive and false negative results. False negatives were reported in 10 out of 11 laboratories that participated in the round_12 proficiency tests organised by FERA, hence 91% of participating laboratories got at least one false negative result from low-titre UCBSV samples. With low-titre CBSV samples, there was a 64% rate of false negative detection in all participating laboratories. The observation of false negative results from low-titre CBSV and low-titre UCBSV prompted FERA to repeat their tests with the same samples, and they confirmed that indeed, titre was very low in the sample, causing a challenge in visualizing the band on gel analysis, which was the case with our study. This clearly shows that the false negative issue was not only observed in this study, but also with other laboratories. In future cases, laboratories should spread out the number of times the same plants are tested, so that false negatives caused by low titre can be captured later in the plants' life cycle when titre is high enough to be detected. The other option is to improve the quantitative detection methods (i.e. qPCR), which has the potential of detecting very low titre of viruses.

Within the same round_12 proficiency tests, there was 7% detection of false positives, and no further effort was taken to confirm otherwise, but laboratories were advised to improve accuracy when loading RT-PCR products during gel analysis. In this study, careful adjustments were made to ensure that no false positives were identified in subsequent tests. This adjustment included, loading fewer samples on a gel in order to avoid long-stay of initially loaded samples on gel, which is the case when many samples (>20) are all being loaded on the same gel. For record purposes, a print copy of the round_12_proficiency test report will be attached at the appendix section of this thesis.

For the other samples tested locally, it was difficult to detect UCBSV using RT-PCR in the first 8 dag, and the method did not provide for measuring virus titre, which we achieved using qPCR in seven time points (6 hag, 1 dag, 2 dag, 5 dag, 8 dag, 45 dag and 54 dag). The qPCR approach, which is to be the gold standard for virus identification was also not 100% effective.

The phenotypic scale used is the same as those applied by other cassava, CBSD research groups[1,10,11,13,15,32-35], and so was the qPCR protocol used[4,30]. All tests performed using phenotypic observation of CBSD-associated symptoms (summarized in Table 2.2), RT-PCR (Fig. 2.2, Fig. 2.3 and Table 2.2) or qPCR (Fig. 2.4) confirm, that none of the methods can singly identify both viruses with 100% accuracy. This observation emphasizes the need for a combined system of diagnosis involving rigorous optimization of these methods so that fewer (if any) false negatives/positives are identified. So far, several cassava research groups have reported independent, 'accurate' methods of identifying both viruses[5,6,36]. Despite these reports, there is need for a deliberate study to validate all such reports, identifying and minimizing overall experimental errors that come with reliance on any single approach.

2.5.5 Sanger sequencing confirmed isolate as UCBSV

Alignment of the RT-PCR product sequence from this study's isolate to the corresponding part of the already sequenced genome of UCBSV (NCBI isolate KF878103.1) confirms the presence of UCBSV in inoculated samples[18]. The isolate used in this study was sampled from the Sugarcane Research Institute (SRI) in Kibaha, Dar es Salaam, Tanzania, however isolate KF878103.1 (GenBank accession), to which our UCBSV Sanger sequenced isolate matched, was reportedly collected from NaCRRI Research Station, Namulonge in Kampala, Uganda. Alignment of isolates sampled from distant regions (Dar es Salaam in coastal Tanzania and Kampala, a high-altitude region) agrees with earlier findings where UCBSV was proliferating in both low and high altitude areas[37]. UCBSV is known to be widely distributed in the region, and has been identified in cassava fields along coastal Kenya and Tanzania[38].

2.6 Conclusion

These results suggest that resistant in Namikonga involves low virus load. This mechanism of resistance is most expressed at 5 dag, when both UL and LL of Namikonga had the lowest titre compared to Albert. These findings are confirmed in Chapter 3[18], which was based on RNAseq experimentation. The higher detection rate in UL suggests that rate of virus multiplication is higher within these actively growing leaf tissues compared to LL which are older and near senescence. Because of low titre and varied patterns of leaf chlorosis across variety and environments, no one-test system is efficient to identifying either virus species, which have various isolates across eastern and southern Africa where CBSD has been detected[38]. Results from this study suggest that RT-PCR, using the primers used in this study, is not 100% accurate, particularly in the early time phase, and must be complimented by phenotypic assessments for UCBSV identification. Indeed, samples interchangeably tested positive using both test methods or were positive by both methods, occurring without a distinct pattern specific to either PCR-based or phenotypic methods. Virus titre appears to fluctuate over time in and upper and lower leaves, However, titre consistently increased in upper leaves of UCBSV-inoculated Albert between 6 hag and 5 dag, while virus load consistently reduced in upper leaves of UCBSV-inoculated Namikonga at those time points. Between 8 dag and 54 dag, no defined pattern was observed in upper leaves of both UCBSV-inoculated Albert and Namikonga. Lower leaves had no consistent pattern, both in UCBSV-inoculated Albert and Namikonga throughout the time course (6 hag - 54 dag). For rapidly evolving species of viruses[39,40], continual validation of existing diagnostic primers and measurement of virus titre in shoot (leaf and stem) and roots is recommended for accurate identification and quantification of viruses isolated from farmers' fields.

2.7 References

1 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Genetically distinct strains of cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch Virol (**2009**), **154**, 353-359.

2 Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M. & Butgereitt, A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen Virol (**2010**), **91**, 1365-1372.

3 Abarshi, M. M., Mohammed, I. U., Jeremiah, S. C., Legg, J. P., Kumar, P. L., Hillocks, R. J. & Maruthi, M. N. 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. J Virol Methods (**2012**), **179**, 176-184.

4 Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One (**2014**), **9**, e96642.

5 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. J Virol Methods (**2011**), **171**, 394-400.

6 Adams, I. P., Abidrabo, P., Miano, D. W., Alicai, T., Kinyua, Z. M., Clarke, J., Macarthur, R., Weekes, R., Laurenson, L., Hany, U., Peters, D., Potts, M., Glover, R., Boonham, N. & Smith, J. 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 Pathol (**2013**), **62**, 233-242.

7 Ogwok, E., Alicai, T., Rey, M. E. C., Beyene, G. & Taylor, N. J. Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta*) plants. Plant Pathol (**2015**), **64**, 1235-1246.

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

9 Xu, J., Aileni, M., Abbagani, S. & Zhang, P. A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). Phytochem Anal. (**2010**), **21**, 395-398.

10 Kanju, E. E., Masumba, E., Masawe, M., Tollano, S., Muli, B., Zacarias, A., Mahungu, N., Khizzah, B., Whyte, J. & Dixon, A. Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmer and consumer preferences. Proceedings 13th ISTRC Symp (**2007**), 95-101.

11 Kulembeka, H. P., Ferguson, M., Herselman, L., Kanju, E., Mkamilo, G., Masumba, E., Fregene, M. & Labuschagne, M. T. Diallel analysis of field resistance to brown streak disease in cassava (Manihot esculenta Crantz) landraces from Tanzania. Euphytica (**2012**), **187**, 277-288.

12 Nichols, R. F. W. Breeding cassava for virus resistance. E Afri Agric J (**1947**), **12**, 184-194.

13 Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. Adv Virol. (**2011**), **2012**, 795697.

14 Rwegasira, G. M. & Rey, C. M. E. Relationship between symptoms expression and virus detection in cassava brown virus streak-infected plants. J Agric Sci (**2012**), **4**, 246-253.

15 Rwegasira, G. M., Rey, M. E. C. & Nawabu, H. Approaches to diagnosis and detection of cassava brown streak virus (Potiviridae: Ipomovirus) in field-grown cassava crop. Afr J F Agric Nutri Develop (**2011**), **11**, 4739-4756.

16 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2– ΔΔCT method. Methods (**2001**), **25**, 402-408.

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

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

19 Boevink, P. & Oparka, K. J. Virus-Host Interactions during Movement Processes. Plant Physiol (**2005**), **Vol. 138**, 1815–1821.

20 Chisholm, S. T., Parra, M. A., Anderberg, R. J. & Carrington, J. C. Arabidopsis RTM1 and RTM2 genes function in phloem to restrict long-distance movement of tobacco etch virus. Plant Physiol (**2001**), **127**, 1667-1675.

21 Cosson, P., Schurdi-Levraud, V., Le, Q. H., Sicard, O., Caballero, M., Roux, F., Gall, O. L., Candresse, T. & Revers, F. The RTM resistance to potyviruses in *Arabidopsis thaliana*: natural variation of the RTM genes and evidence for the implication of additional genes. PLoS One (**2012**), **7**.

22 Mahajan, S. K., Chisholm, S. T., Whitham, S. A. & Carrington, J. C. Identification and characterization of a locus (RTM1) that restricts long-distance movement of tobacco etch virus in Arabidopsis thaliana. Plant J (**1998**), **14**, 177-186.

23 Wilson, C. R. & Jones, R. A. C. Resistance to phloem transport of potato leafroll virus in potato plants. J. Gen. Virol (**1992**), **73**, 3219-3224.

24 Revers, F., Guiraud, T., Houvenaghel, M. C., Mauduit, T., Le Gall, O. & Candresse, T. Multiple resistance phenotypes to Lettuce mosaic virus among Arabidopsis thaliana accessions. Mol Plant Microbe Interact (**2003**), **16**, 608-616.

25 Decroocq, V., Sicard, O., Alamillo, J. M., Lansac, M., Eyquard, J. P., Garcia, J. A., Candresse, T., Le Gall, O. & Revers, F. Multiple resistance traits control Plum pox virus infection in Arabidopsis thaliana. Mol Plant Microbe Interact (**2006**), **19**, 541-549.

26 Chisholm, S. T., Mahajan, S. K., Whitham, S. A., Yamamoto, M. L. & Carrington, J. C. Cloning of the Arabidopsis RTM1 gene, which controls restriction of long-distance movement of tobacco etch virus. PNAS (**2000**), **97**, 489–494.

27 Cosson, P., Sofer, L., Le, Q. H., Leger, V., Schurdi-Levraud, V., Whitham, S. A., Yamamoto, M. L., Gopalan, S., Le Gall, O., Candresse, T., Carrington, J. C. & Revers, F. RTM3, which controls long-distance movement of potyviruses, is a member of a new plant gene family encoding a meprin and TRAF homology domain-containing protein. Plant Physiol (**2010**), **154**, 222-232.

28 Hipper, C., Brault, V., Ziegler-Graff, V. & Revers, F. Viral and Cellular Factors Involved in Phloem Transport of Plant Viruses. Front Plant Sci. (**2013**), **4**, 154.

29 Pallas, V. & García, J. A. How do plant viruses induce disease? Interactions and interference with host components. J. Gen. Virol (**2011**), **92**, 2691-2705.

30 Anjanappa, R. B., Mehta, D., Maruthi, M. N., Kanju, E., Gruissem, W. & Vanderschuren, H. Characterization of Brown Streak Virus–Resistant Cassava. Mol Plant Microbe Interact (**2016**), **29**, 527-534.

31 Otti, G., Bouvaine, S., Kimata, B., Mkamillo, G., Kumar, P. L., Tomlins, K. & Maruthi, M. N. High-throughput multiplex real-time PCR assay for the simultaneous quantification of DNA and RNA viruses infecting cassava plants. J Appl Microbiol (**2016**), **120**, 1346-1356.

32 Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. & Colvin, J. Re-emergence of cassava brown streak disease in Uganda. Plant Dis (**2007**), **91**, 24-29.

33 Hillocks, R. J., Raya, M. D., Mtunda, K. & Kiozia, H. Effects of brown streak virus disease on yield and quality of cassava in Tanzania. J Phytopathol (**2001**), **149**, 389-394.

34 Kawuki, R. S., Kaweesi, T., Esuma, W., Pariyo, A., Kayondo, I. S., Ozimati, A., Kyaligonza, V., Abaca, A., Orone, J. & Tumuhimbise, R. Eleven years of breeding efforts to combat cassava brown streak disease. Breed Sci. (**2016**), **66**, 560-571.

35 Kawuki, R. S., Pariyo, A., Amuge, T., Nuwamanya, E., Ssemakula, G., Tumwesigye, S., Bua, A., Baguma, Y., Omongo, C. A., Alicai, T. & Orone, J. A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). J Plant Breed Crop Sci (**2011**), **3**.

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

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

38 Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djikeng, A., Kehoe, M. & Boykin, L. M. Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation. PLoS One (**2015**), **10**, e0141939.

39 Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, M. A. & Boykin, L. M. Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. Sci Rep (**2016**), **6**, 36164.

40 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. Evolution of cassava brown streak disease-associated viruses. J Gen Virol (**2011**), **92**, 974-987.

Chapter 3

A time series transcriptome analysis of cassava (*Manihot esculenta* Crantz) varieties challenged with *Ugandan cassava brown streak virus*

T. Amuge, D. K. Berger, M. S. Katari, A. A. Myburg, S. L. Goldman and M. E. Ferguson

Author Contributions

T.A. Conducted all field, screen-house (planted experimental plants, graft-inoculated samples, collected samples) and laboratory procedures (identified inoculum, isolated RNA and prepared cDNA libraries). Performed data analysis and wrote the initial draft of the manuscript.

M.S.K. Led the data analysis component, trained and supervised T.A. in all aspects of RNAseq data analysis.

D.K.B. Suggested appropriate data analysis and led discussions on data interpretation. Critically reviewed and edited the manuscript. Provided University supervision.

A.A.M. Provided critical comments, guidance, and University supervision.

S.L.G. Compiled a cassava genome database at VirtualPlant.

M.E.F. Principle Investigator of the funding project. Identified cassava varieties, provided day-to-day supervision, and edited the manuscript

3.1 Abstract

A time-course transcriptome analysis of two cassava varieties that are resistant and susceptible to cassava brown streak disease (CBSD) was conducted using RNASeq, after graft inoculation with *Ugandan cassava brown streak virus* (UCBSV). From approximately 1.92 billion short reads, the largest number of differentially expressed genes (DEGs) was obtained in the resistant (Namikonga) variety at 2 days after grafting (dag) (3887 DEGs) and 5 dag (4911 DEGs). At the same time points, several defence response genes (encoding LRR-containing, NB-ARC-containing, pathogenesis-related, late embryogenesis abundant, selected transcription factors, chaperones, and heat shock proteins) were highly expressed in Namikonga. Also, defence-related GO terms of 'translational elongation', 'translation factor activity', 'ribosomal subunit' and 'phosphorelay signal transduction', were overrepresented in Namikonga at these time points. More reads corresponding to UCBSV sequences were recovered from the susceptible variety (Albert) (733 and 1660 read counts per million (cpm)) at 45 and 54 dag compared to Namikonga (10 and 117 cpm respectively). These findings suggest that Namikonga's resistance involves restriction of multiplication of UCBSV within the host. These findings can be used with other sources of evidence to identify candidate genes and biomarkers that would contribute substantially to knowledge-based resistance breeding.

3.2 Introduction

Cassava is among the six major crops of Africa, representing a staple food for >250 million people (FAOSTAT, 2010). Currently, production in parts of southern (Mozambique, Malawi and Angola), eastern (Uganda, Kenya, Tanzania, Rwanda and Burundi) and central Africa (D.R. Congo) is seriously affected by cassava brown streak disease (CBSD)[1,2]. At least two virus species cause CBSD: *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV)[3,4]. Here CBSVs is used to refer to both of these viruses. First reported in Tanzania, CBSD previously occurred at low levels primarily in coastal East Africa, Mozambique and around Lake Malawi and was thought to be restricted by altitude[5]. However, in the early 2000s, CBSD had begun to spread around Lake Victoria, and by 2004, typical CBSD symptoms were widespread in farmers' fields in central Uganda. The disease has spread steadily since then as far as DR Congo and South Sudan and now, together with cassava mosaic disease, causes over US$1 billion losses in production annually in Africa[1,6,7].

Both CBSV and UCBSV are members of genus *Ipomovirus*, family *Potyviridae*[3,4,8]. Both genomes have particles measuring 650 nm with pinwheel inclusions in their cells, distinctive of Potyviridae[3]. The CBSV and UCBSV genomes are 9008 nucleotides (nt) and 9070 nt long, respectively[4,8]. Each genome contains a single open reading frame that is translated into 2912 amino acids (aa) for CBSV and 2012 aa for UCBSV. Like other members of *Potyviridae*, both genomes encode the proteins P1, P3, 6K1, CI, 6K, VPg, Nla-Pro, Nlb and CP, with a new Ham1-like (Ham1) protein[3,4,8]. Neither genome encodes the helper-component protein (HC-Pro) that is encoded by other members of the family Potyviridae[3,4,8].

CBSVs are transmitted naturally by *Bemisia tabaci*[5,9,10] whiteflies in a semi-persistent manner, with the spiralling whitefly (*Aleurodicus dispersus*) being a possible alternative vector[11,12]. The main form of disease spread in cassava fields in Africa is thought to be through virus-positive stem cuttings in this clonally propagated crop[13]. Mechanical transmission[14,15] and graft inoculation[9,16] are used in research studies.

Infected susceptible plants develop chlorosis along leaf veins, brown streaks on stems, and root necrosis (Fig. 3.1), with severe infection causing shoot dieback. Dual infection with both virus species is common in farmers' fields, although there are no reports of synergistic virus interaction[3,4,8,17], and both viruses cause similar symptoms, although those of CBSV tend to be more severe than UCBSV[18,19].

Currently, no known cassava variety has been reported to be immune to CBSVs, but varied levels of resistance or tolerance have been identified. Here, resistance is defined as the ability of the cassava variety to maintain a low virus load and show minimal shoot symptoms coupled with little or no root necrosis at harvest. Using this criteria, Namikonga (also known as Kaleso) was identified as resistant to CBSV from greenhouse experiments[20] and classified as tolerant to CBSVs based on field symptoms and virus load in Uganda[21]. Cassava breeding is a lengthy process, and disease response is influenced by genotype-by-environment interactions[22]. To shorten the breeding cycle and improve the accuracy of variety selection, breeders are implementing genomics-based approaches. An understanding of the resistance mechanisms

involved, including biochemical pathways, and the identification of candidate genes and biomarkers would contribute substantially to knowledge-based genomics breeding, including marker-assisted selection (MAS).

The use of RNAseq (RNA sequencing) has enabled the high-throughput identification of new genes, exons and exon junctions, splice variants and promoter regions in sequenced transcriptomes[23-25]. Previously, in cassava, an RNAseq-based transcriptome analysis of CBSD-resistant and -susceptible cassava varieties infected with CBSV was conducted to identify genes putatively involved in disease resistance; however, the results were inconclusive[20].

In potyviruses, recessive resistance is known to involve >200 defence genes[26-28]. The majority of the cloned genes involve eukaryotic initiation factor (eIF) deployment[28-33]; however, other mechanisms involving defence response genes have been reported[34]. The eIF-mediated mechanism of resistance, also called 'passive'[28] or 'loss-of-susceptibility'[35] resistance, occurs when the host plant has a modified ribosomal protein sub-structure (usually eIF4E or eIF(iso)4E) so that the viral genome-linked (VPg) protein cap is unable to bind to the ribosome, thus preventing viral replication. An alternative mechanism of resistance to potyviruses was first identified in Arabidopsis plants infected with *Tobacco etch virus* (TEV) when a set of genes restricting the long-distance movement of TEV were cloned. These genes were named restricted TEV movement (RTM) genes. The RTM genes have since been found to cause resistance against other potyviruses including *Lettuce mosaic virus* (LMV) and *Plum pox virus* (PPV) by restricting the long-distance movement of virus particles[36]. Dominant resistance against potyviruses has been observed in pepper against Potato virus Y (PVY)[37] and in *Solanum lycopersicum* against TEV and *Pepper mottle virus* (PMV)[29]. This dominant resistance may be characterized by a hypersensitive response (HR), as in tobacco resistant to PMV[38] and *Tobacco mosaic virus* (TMV)[39] and in potato resistant to both PVY and TEV[40]. The identification of potyvirus defence genes, such as those described above, motivated this study, as both CBSVs are potyviruses affecting cassava, a major food crop in sub-Saharan Africa.

The greater goal of this study was to define biomarkers for genomics-based breeding of CBSD-resistant cassava varieties to help address food insecurity in Africa. To improve our understanding of the mechanism of resistance and potentially identify candidate genes involved in resistance to UCBSV, DEGs were determined from a time-course experiment involving UCBSV-inoculated and mock-inoculated plants of two cassava varieties with contrasting responses to UCBSV infection, Albert and Namikonga. In addition, we tested the hypothesis that UCBSV accumulates at significantly lower rates in Namikonga compared to Albert because the induction of defence genes in Namikonga restricts virus replication[20,21].

3.3 Materials and methods

3.3.1 Planting material sources

The study was conducted on Namikonga, a CBSD-resistant variety, and the susceptible variety Albert[21]. Namikonga is an interspecific hybrid between wild (*Manihot glaziovii* Müll. Arg.) and domesticated (*Manihot esculenta* Crantz) cassava, with an estimated 14% interspecific hybrid genome[41]. Albert is a farmer variety from southern Tanzania; it is susceptible to both virus species, showing characteristic CBSD-associated leaf chlorosis and severe root necrosis[21]. Virus-negative stem cuttings of Namikonga and Albert were obtained from plants maintained in a greenhouse at the Sugarcane Research Institute (SRI), Kibaha, Dar es Salaam, Tanzania. The source of inoculum, variety NDL06/132, was from a field trial at the SRI, Tanzania and was multiplied for scions in a greenhouse at BioSciences eastern and central Africa (BecA), Nairobi, Kenya. Experimental plants were tested for the presence or absence of both CBSV and UCBSV prior to grafting. For this, RNA was extracted from fresh leaf samples using a CTAB protocol, with the amount of starting extraction buffer material modified from 600 μl to 800 μl[42]. Respective second-strand cDNAs were synthesized using Superscript III (Invitrogen, CA). RT-PCR was performed using primers that distinguish both viruses simultaneously; namely, CBSDDR and CBSDDF2[43] and CBSVF2 and CBSVR8[44].

3.3.2 Establishment of the experiment, graft inoculation and sampling

The aim of this transcriptome time-series experiment was to determine the transcriptional signatures immediately after virus infection by grafting (early sampling) and several days before aerial symptom emergence (late sampling). Grafting is commonly used to transmit viruses in cassava, and was used here[4,16]. Artificial infectious clones were not available for inoculation, transmission using whiteflies was unreliable and mechanical transmission of CBSVs is only effective with model plants[15]. Initially 45 cassava cuttings of each variety were planted in large pots containing 3.85 kg gravel (construction grade) topped to 33.23 kg with forest soil and immediately watered with 3 liters of tap water. Thereafter, the plants were given 2 litres of tap water per week until harvest. At two months after planting, plants were ready for top grafting. Just prior to grafting a baseline leaf sample (time zero) was collected from the top leaves of all plants (see methodology below). Fifteen plants of Namikonga and Albert were grafted with scions positive for CBSV only, 10 plants of each variety were grafted with scions positive for UCBSV only and another 10 plants of each variety with scions positive for CBSV+UCBSV. Ten plants of each variety were grafted as controls with virus negative NDL06/132 scions. Samples were taken as described below from all grafted plants, and frozen until selection of biological replicates was made based on RT-PCR diagnostic results. An RT-PCR test at 3 mag (data not shown) confirmed that only one out of 15 grafted Namikonga plants was positive for CBSV, while seven of 10 plants each grafted with UCBSV or CBSV+UCBSV infected scions were positive for the respective virus. In Albert seven of the 15 plants grafted with CBSV infected scions were positive, and seven of 10 plants each infected with UCBSV or CBSV+UCBSV were positive. It is on the basis of successful graft transmission of UCBSV that this species was chosen for downstream studies. Three biological replicates for each of the two treatments (UCBSV-inoculated and mock-inoculated plants) of Albert and Namikonga were selected for RNAseq.

In this time-series experiment, samples were collected from the main experiment at 6 hag, 1, 2, 5, 8, 45 and 54 dag. The time points up to and including 8 dag constituted the “early sampling” and the remaining samples the “late sampling”. The “late sampling” time points were selected to coincide with the period just before foliar symptom emergence. To determine this a number of indicator plants were grafted three weeks before the main experiment providing ample time to plan for pre-symptom emergence sampling. At each time point leaves were sampled from below the graft point to avoid sampling the leaves of the sprouting NDL06/132 scion. Typically, both Albert and Namikonga plants have at least five lobes per leaf, each lobe weighing >400 g. Therefore, we sampled a different lobe of the same leaf at every different time point. Where lobes were fewer than needed to cover all successive time points, samples were taken from lobes on the leaf below in the same order. The samples were placed in pre-labelled aluminium foil envelopes and immediately frozen in liquid nitrogen, transferred to -80°C and stored prior to RNA extraction. At harvest (12 mag), roots from all plants were scored for CBSD-related necrosis.

CBSD symptom scoring

UCBSV-inoculated and mock-inoculated plants were scored for above-ground symptoms during the late sampling phase and at harvest (49 dag) for symptoms in storage roots, according to standard 1-5 scales[45]. According to the leaf severity scale, 1 implies no symptoms, and 5 implies clear leaf chlorosis covering >75% of leaves and streaking on the stem with shoot die back. On the storage root severity scale, a score of 5 implies more than 50% root necrosis[7].

3.3.3 RNA extraction, cDNA synthesis and RNAseq

Frozen leaf samples were ground to a fine powder in liquid nitrogen and divided into three aliquots for (a) cDNA library synthesis, (b) virus quantification with qPCR and (c) a back-up. RNA was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich) following the manufacturer’s protocol. RNA integrity was confirmed using a Qubit fluorometer (Life Technologies). The RNA extraction, cDNA library synthesis, shipment and Illumina sequencing were performed 16 months apart for two batches of samples that were sequenced at two facilities, the University of California, Berkeley, CA, USA, and Dow AgroSciences, Indianapolis, IN, USA, due to funding availability. The same person in the same laboratory at BecA conducted the RNA extraction and cDNA library synthesis of both batches. Each batch contained both the mock-inoculated and UCBSV-inoculated samples from a genotype at a specific time point, as listed in Table S3.1. The same shipping company was contracted to ship each batch of cDNA libraries to the sequencing facilities. The cDNA libraries were synthesized using the Illumina Truseq cDNA library synthesis kit (Illumina, San Diego, CA, USA) using set A indices for 50-bp sequencing at UC Berkeley and set B indices for 101-bp sequencing at Dow AgroSciences. Illumina Hi-Seq instruments were employed at both facilities, using in-house pre-sequencing preparations of cDNA library material and instrument run settings with the aim of generating 20 million reads of 50 bp (for UC Berkeley) or 101 bp (for Dow AgroSciences) in length for each biological replicate sample (Table S3.1). Control and UCBSV-infected samples from both varieties obtained at 6 hag and 8 dag were sequenced at UC Berkeley. Control and infected Albert samples collected at 45 and 54 dag were also sequenced at UC Berkeley. For both varieties, control and UCBSV-infected samples collected at 1, 2 and 5 dag, were sequenced at Dow. Dow also sequenced the Namikonga samples (mock- and UCBSV- inoculated) collected at 45 and 54 dag.

3.3.4 RNA extraction and cDNA synthesis for diagnostics

For RT-PCR and quantitative RT-PCR virus diagnostics, aliquots of RNA samples from the Spectrum extractions described above were used, and cDNA was synthesized using Superscript III (Invitrogen). The presence of UCBSV was positively verified with a single band at 440 bp, and the absence of CBSV was verified using the primer sets described above with appropriate CBSV and UCBSV positive and negative controls[43] (Fig. 3.2).

3.3.5 Detection of UCBSV in RNAseq reads

The availability of RNAseq data from this experiment was exploited as an additional method to detect UCBSV RNA virus molecules in the cassava samples. To confirm successful graft inoculation with UCBSV and that control samples were UCBSV-negative, RNAseq reads from all 96 samples were mined for UCBSV sequences. To retrieve virus sequences from RNAseq reads, all five fully sequenced UCBSV genomes from NCBI were used as a reference to map the RNAseq reads of UCBSV-infected and mock-inoculated samples. The five UCBSV genomes were isolates HG965222.1, FN434109.1, FJ185044.1, HM181930.1 and NC_014791.1 and were combined to form one UCBSV reference file. RNAseq reads that failed to map to the cassava reference genome (unmapped reads) were each aligned to the UCBSV reference genome using Bowtie2[46]. The results were obtained from the alignment summary and transferred to Excel for plotting.

3.3.6 Data analysis

Read quality analysis and mapping to the cassava reference genome

Individual files received as batches of ≤ four million reads (later combined to make one large file of 20 million reads) were analysed for sequence quality using FastQC (FastQC)[47]. All reads had a Phred score above 20, meaning that 99% of the bases were accurately called. The first low-quality 10-13 bp of each read, which is typical for Illumina sequencing, were trimmed to 37 bp (for 50-bp reads sequenced at UC Berkeley) and 88 bp (for 101-bp reads sequenced at Dow AgroSciences) using FASTX_toolkit (fastx_trimmer)[48]. The trimmed files were mapped to the cassava reference genome v4.1 (plant accession number AM560-2)[49] using Tophat2[46], which runs Bowtie2 in the background. The default settings were used, allowing up to two mismatches per read. Allowing two mismatches for both 37 bp and 88 bp reads did not affect mapping, as reads that map to the same gene model are only counted once, irrespective of their length, taking into account the two allowed mismatches. This was tested and confirmed using randomly selected samples of 37 bp or 88 bp (data not shown). The mapped reads were de-duplicated using Picard's dedup function, correcting for any amplification bias caused during the PCR used in constructing the sequencing libraries. Mapped reads with gene models on the cassava genome were counted with HTseq_count[50], a python script widely used to count RNAseq reads[50]. Genes without gene models (counted altogether as no_feature) or matching more than one gene model (ambiguous) were removed before clustering and identification of DEGs. The HTseq_count command no_feature counts reads that are not aligned to any gene model on the reference genome, while the command ambiguous counts reads spanning an intersection of two gene models or aligned to more than one gene model (Supplementary Data S4).

Filtering RNAseq reads

To reduce artifacts and increase the statistical power for identifying DEGs, the counted reads were filtered in four steps. First, genes with zero count values were deleted, followed, second, by genes with very few reads (rowsum≤log25) on the distribution curve (data not shown). Third, genes that passed stages (i) and (ii) but were below a set per time point minimum cut-off were deleted. The minimum cut-off per time point was calculated from row (per gene) median and standard deviation values (minimum cut-off = median - 2*standard deviation). The fourth and last filtering was based on how close the biological replicates clustered by respective treatments, variety and time point. Outlier biological replicates were removed from the downstream analysis, although at least two samples per treatment were retained. As was done for differential gene expression analysis, data for clustering were normalized using the estimateSizeFactors function of the DESeq package[51]. Data normalization was performed with combined gene counts from both Albert and Namikonga. To do this, only good-quality genes (genes that passed the four filtering stages described above) were used for the normalization analysis in both varieties. Clustering was performed using Pearson's algorithm. The algorithm clusters samples based on their covariance, providing a robust method of grouping samples as opposed to the distance-based Euclidean method. The filtering, clustering, selection of DEGs and plotting of data were performed using the R statistical package[52].

Identification of DEGs and cluster analysis

Compared to distribution-based methods and cuffdiff[53], negative binomial algorithms implemented by DESeq (edgeR and bayseq) provide a higher statistical power for identifying DEGs[50,54]. Using filtered reads, DESeq was applied to identify DEGs between control and treated samples of each variety and at each time point. In DESeq, each variety had a dataset for respective sampling time points (example: Albert_6hag). Each dataset contained filtered biological replicates for one time point and two treatments (UCBSV inoculated and mock inoculated). The function newCountDataSet, which imports gene counts with pre-defined conditions, was used to import samples under "infected" or "control", where "infected" identifies gene counts of UCBSV-inoculated samples and "control" identifies gene counts from mock-inoculated samples. Once imported, the samples were normalized by their respective library sizes. To perform 'normalization', the effective library size was determined using the estimateSizeFactors command. The per sample dispersion was then estimated using the estimateDispersions command, followed by differential expression analysis. Differential expression analyses were performed using the negative binomial algorithm with the command nbinomTest. After differential expression analysis, p-values were adjusted for multiple testing with the Benjamini-Hochberg procedure, which controls for the false discovery rate (FDR). For most RNAseq studies, a 10% FDR is the recommended limit for identifying significant DEGs[54,55]. We chose only to apply a FDR of 0.1 as the main filter to identify DEGs, and did not apply a further filter of Log2 fold change. Therefore, genes for treatment comparison with log2FC>0 and FDR <0.1 were classified as up regulated, whereas those with log2FC<0 and FDR <0.1 were down regulated. This was done to capture the maximum number of genes with differential expression across the time points.

Functional annotation, GO enrichment and network analysis of DEGs

The lists of DEGs were queried against the cassava genome database at VirtualPlant (http://www.virtualplant.org)[56] to determine functional annotations and GO enrichment. VirtualPlant is

annotated using cassavaCyc data (functional categories) from Phytozome's version 4.1 of the cassava reference genome (plant accession number AM560-2) (26). VirtualPlant contains 30,666 cassava genes, 28,610 of which have predicted Arabidopsis orthologues that have protein-coding annotations. Thus, the functional annotations are primarily metabolic, as available in cassavaCyc (http://www.plantcyc.org/). The GO terms for cassava genes in VirtualPlant are allocated based on the GO terms of their corresponding predicted Arabidopsis orthologues in TAIR (www.Arabidopsis.org). VirtualPlant therefore harbors GO terms for 10,902, 15,368, and 4,546 cassava genes in the biological process, molecular function and cellular compartment categories, respectively. The manual analysis of defence genes focused on the time points with the highest number of DEGs.

3.4 Results

3.4.1 Symptoms of CBSD in leaves and roots

Characteristic CBSD symptoms were observed in UCBSV-inoculated plants and varied in magnitude by variety and time of observation. In the early sampling phase, no visible symptoms were observed on the leaves of either UCBSV-inoculated or mock-inoculated plants of both varieties (Fig. 3.1). In Albert, at the late sampling phase, young leaves of UCBSV-inoculated plants showed chlorotic patterns along veins, expanding to form very large yellow areas (Fig. 3.1). At 12 months after grafting (mag), >50% of storage roots from UCBSV-inoculated Albert plants were necrotic, with a severity score of 4 (Fig. 3.1). Control Albert plants showed no disease symptoms on leaves and roots at 12 mag.

For Namikonga, leaf chlorotic spots (Fig. 3.1) were observed in the late sampling phase. At 3 mag chlorosis covering 2–3 leaves per plant (score 2) was observed. At 12 mag all storage roots were non-necrotic. The storage roots from mock-inoculated plants were also non-necrotic at 12 mag.

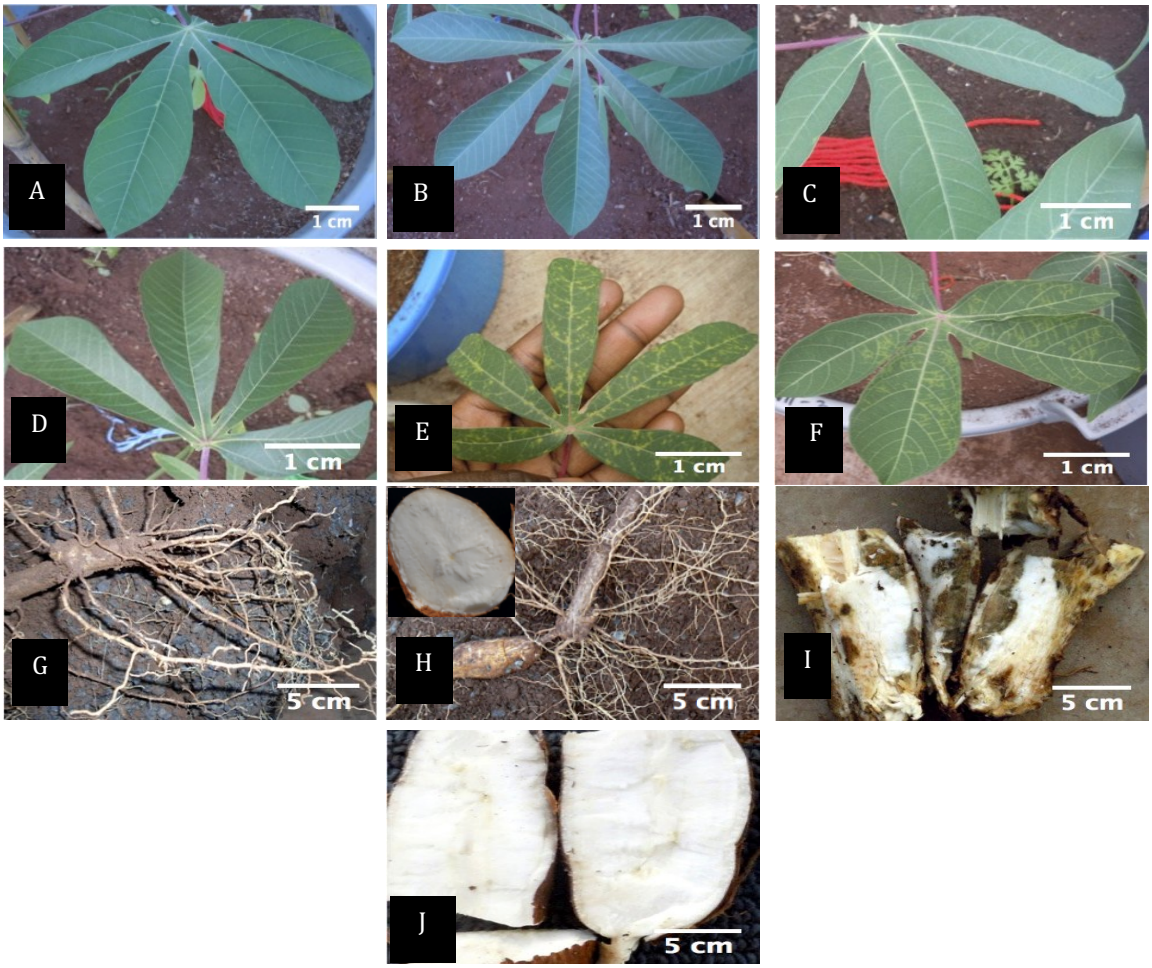


Fig. 3.1: Leaf and root symptoms of UCBSV-inoculated versus mock-inoculated Albert and Namikonga plants during the experiment. (A) Albert, UCBSV inoculated, 8 dag; (B) Albert, mock inoculated, 8 dag; (C) Namikonga, UCBSV inoculated, 8 dag; (D) Namikonga, mock inoculated, 8 dag; (E) Albert, UCBSV inoculated, 54 dag; (F) Namikonga, UCBSV inoculated, 54 dag; (G) Albert root, UCBSV inoculated, 3 mag; (H) Namikonga root, UCBSV inoculated, 3 mag; (I) Albert root, UCBSV inoculated, 12 mag; (J) Namikonga root, UCBSV inoculated, 12 mag.

3.4.2 Detection of UCBSV in UCBSV-inoculated cassava using RT-PCR and RNAseq

In the early phase, all samples (UCBSV- and mock-inoculated) were negative (non-detectable) for UCBSV except at 6 dag, when one of the ten UCBSV-inoculated plants of both the Albert and Namikonga varieties were positive for UCBSV using the RT-PCR assay (data not shown). Seven out of ten UCBSV-inoculated Albert plants were consistently positive at the late time phase (data not shown). For the Namikonga variety, seven out of ten UCBSV-inoculated plants tested positive at least one sampling point during the late time phase (data not shown). However, this was not as consistent as the results from the Albert variety.

To confirm the presence of UCBSV in UCBSV-inoculated plants, leaves were sampled at 3 mag, after sampling for early and late time phases. RT-PCR with the UCBSV-specific primers CBSDDR and CBSDDF2 was performed on these 3 mag samples of Albert and Namikonga and 7 of the 10 plants of each variety tested were positive for UCBSV, having a 440-bp RT-PCR fragment when resolved on a 2% agarose gel (Fig. 3.2). Mock-inoculated plants had no visible RT-PCR fragment on the same gel, confirming the absence of UCBSV. An RT-PCR fragment putatively diagnostic for UCBSV (approximately 440 nt) was gel purified and Sanger sequenced before being aligned to the full UCBSV genome sequences from GenBank. The best alignment (98% identity, 100% query cover, E value=3e179) was obtained with accession KF878103.1 (annotated as UCBSV). The same RT-PCR fragment size was amplified from the UCBSV inoculum source (variety NDL06/132), which indicates that the same virus species was present in the inoculum and UCBSV-inoculated plants (Fig. 3.2). Ten plants of each variety and each treatment were confirmed to be negative for UCBSV by RT-PCR prior to grafting. At 3 mag, after samples for RNAseq had been taken and frozen, RT-PCR was conducted on all 40 plants to confirm whether they were UCBSV positive. Sanger sequencing of the RT-PCR products was performed to confirm the identity of UCBSV. All ten Albert plants and seven out of ten Namikonga plants were UCBSV-positive. Three of the positive plants of each variety were randomly selected along with three mock-inoculated plants of each variety and used as biological replicates for the RNAseq study.

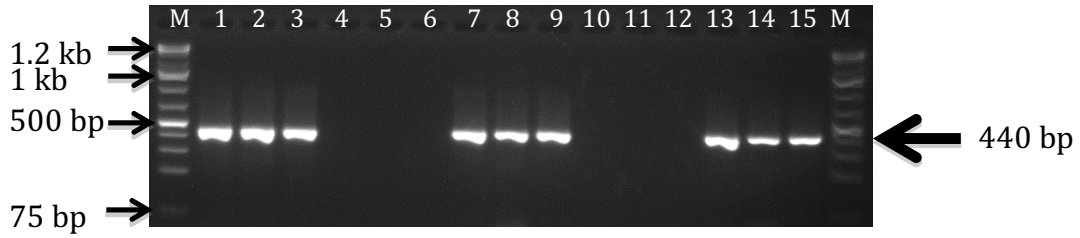


Fig. 3.2: Diagnostic RT-PCR for UCBSV in the inoculum source NDL06/132, Albert and Namikonga plants at 3 mag. RNA extracted from cassava plants from the UCBSV-inoculation experiment were amplified using the UCBSV diagnostic primers CBSDDR and CBSDDF2[43], and the RT-PCR products were separated by agarose gel electrophoresis (2%). Lanes 1-3: UCBSV inoculum source (variety NDL06/132); lanes 4-6: Albert, mock grafted with virus-negative scions of NLD06/132; lanes 7-9: Albert, UCBSV inoculated; lanes 10-12: Namikonga, mock grafted with virus-negative scions of NLD06/132; lanes 13-15: Namikonga, UCBSV inoculated and M: 1 Kb+ ladder.

The number of UCBSV reads detected in each RNAseq sample at each time point is given in Supplementary Data S1. UCBSV sequences were detected in the inoculated susceptible variety Albert at 45 and 54 dag (late sampling phase). At 45 dag, 733, 507 and 37 UCBSV read counts per million (cpm) were detected from each of the three UCBSV-inoculated biological replicates of Albert (Fig. 3.3). At 54 dag, the same samples had 1660, 940 and 80 UCBSV read cpm. Surprisingly, one of the three mock-inoculated biological replicates of Albert had six UCBSV reads at 45 dag, and at 54 dag, two replicates had two UCBSV reads each, and one replicate had a single read. The surprising occurrence of UCBSV reads in mock-inoculated plants could be explained by contamination or the fact that the virus-negative plants host minute levels of UCBSV, which are not detectable by routine RT-PCR but are detectable with deep sequencing, such as RNAseq.

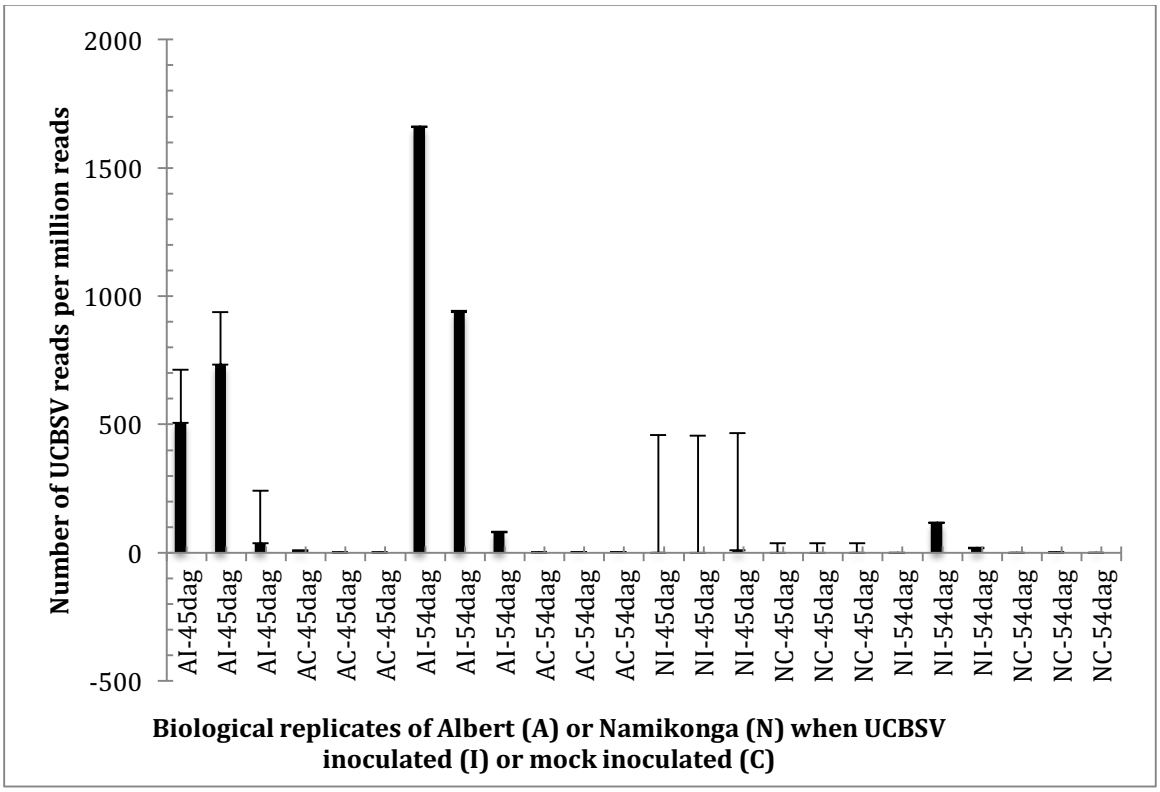


Fig. 3.3: Number of DEGs between UCBSV-inoculated and mock-inoculated cassava varieties Albert and Namikonga at 6 hag, 1, 2, 5, 8, 45 and 54 dag. The DEGs were identified using the DESeq method[51].

In Namikonga, very few UCBSV reads were detected in UCBSV-inoculated samples. The highest count (117 UCBSV sequences) was recorded in one biological replicate at 54 dag (Supplementary Data S1). This is 14 times lower than the highest count recorded in Albert at the same time point. The rest of the UCBSV-inoculated Namikonga samples had ≤10 UCBSV reads detected at any time point. For Namikonga, at 45 dag, one replicate from a mock-inoculated plant had one UCBSV sequence; the other two samples from mock-inoculated Namikonga plants had no detectable UCBSV sequence. The same was true for Namikonga at 54 dag.

3.4.3 Read depth and mapping to the cassava reference genome (v4.1)

A total of approximately 1.92 billion raw reads were sequenced from 96 cDNA Namikonga and Albert libraries at eight time points with a coverage of approximately 20 million reads per library (Table S3.1;

Supplementary Data S3.2). All sequences have been deposited in NCBI's Sequence Read Archive (SRA) SRP with BioProject ID PRJNA360340.

After sequence quality control (QC), including filtering based on Phred score and trimming, reads from each library were mapped to the cassava reference genome using Tophat2[46]. The cassava reference genome (version 4.1) is derived from a partial inbred line, AM560-2, a third generation self of the Latin American variety MCOL-1505[49]. Albert had 29,755,234,095 mapped reads and 8,231,859,633 unmapped reads, while Namikonga had 32,490,332,800 mapped and 8,477,326,725 unmapped reads (Supplementary Data S3.3). Therefore, Albert had 78.0% of its reads mapped and 22.0% unmapped to the cassava reference genome, while Namikonga had 79.0% mapped and 21.0% unmapped reads (Table 3.1). The cassava reference genome is derived from a partial inbred line, AM560-2, a third generation self of the Latin American variety MCOL-1505[49]. Table S3.2 contains a summary of the per variety statistics after QC. Altogether, reads from 40 out of the 48 libraries sequenced at Dow (101 bp) (Table S3.2b) mapped at least 84.0% to the cassava reference genome. Four samples mapped at 62.3%-78.7%, while the remaining four mapped at 24.0%-36.3%. Among the 50-bp libraries sequenced at UC Berkeley (Table S3.2a), one sample mapped poorly (3.9%), another was 59.4% mapped, ten samples mapped between 74.1%-79.7% and the remaining 36 libraries were >80.0% mapped to the reference genome. Out of the 96 sequenced cDNA libraries, all libraries that mapped poorly (<40.0%) also failed to cluster with respective biological replicates (see section 'Clustering of samples using data from RNAseq reads mapped to cassava genome'). These outlier libraries were removed from subsequent analyses.

Table 3.1: Total statistics for RNAseq reads from cassava varieties Albert and Namikonga sampled at eight time points after mock or graft inoculation with UCBSV. The time points were time zero (before graft inoculation), 6 hag, 1, 2, 5, 8, 45 and 54 dag. The figures reflect the number of RNAseq reads that mapped or did not map (unmapped) to the cassava reference genome v4.1[49].

Time point	Mapping to cassava genome	Albert UCBSV inoculated	Albert Mock inoculated	Namikonga UCBSV inoculated	Namikonga Mock inoculated
Time	Mapped	-	4,218,021,147	-	3,996,224,040
zero	Unmapped	-	835,908,591	-	929,195,862
6 hag	Mapped	1,974,374,444	2,311,923,325	1,739,750,996	2,034,621,580
	Unmapped	465,562,984	558,691,687	373,228,177	416,264,064
1 dag	Mapped	1,782,476,393	1,606,541,112	1,751,611,492	1,415,764,268
	Unmapped	469,842,785	281,839,135	380,522,348	892,832,192
2 dag	Mapped	2,311,174,776	2,258,285,184	2,278,828,029	2,195,695,847
	Unmapped	425,576,288	411,150,949	494,433,874	648,333,549
5 dag	Mapped	1,152,362,044	900,243,552	2,223,072,359	1,809,373,734
	Unmapped	900,243,552	988,890,194	402,279,032	713,771,063
8 dag	Mapped	1,845,319,525	1,948,946,548	2,265,995,556	1,756,458,519

Time point	Mapping to cassava genome	Albert UCBSV inoculated	Albert Mock inoculated	Namikonga UCBSV inoculated	Namikonga Mock inoculated
	Unmapped	352,441,534	373,483,677	493,741,019	547,890,775
45 dag	Mapped	1,701,360,044	1,258,543,240	2,269,455,275	2,155,716,341
	Unmapped	365,911,422	831,147,304	505,056,145	476,811,865
54 dag	Mapped	2,089,595,204	2,396,067,557	2,288,557,815	2,309,206,949
	Unmapped	473,446,221	497,723,310	545,859,978	657,106,782
Total	Mapped	12,856,662,430	16,898,571,665	14,817,271,522	17,673,061,278
	Unmapped	3,453,024,786	4,778,834,847	3,195,120,573	5,282,206,152

HTseq-counts[50] was used to count the number of reads that aligned to gene models of the cassava reference genome (v4.1). More than 22,000 out of 33,000 genes in the cassava reference genome v4.1 were represented by sequence information from either variety. Supplementary Data S4 summarizes the reads that could not be counted by HT-seq (ie reads that did not map to any gene models in the reference genome or reads that mapped to more than one gene model).

Clustering of samples using data from RNAseq reads mapped to the cassava genome

Using Pearson's correlation, the normalized reads of samples from both Albert and Namikonga clustered by variety (Supplementary Fig. S.3.1). In most treatments, all three biological replicates clustered together, irrespective of where the samples were sequenced. Respective biological replicates clustered by time point, variety and treatment (UCBSV inoculated and mock inoculated), irrespective of the laboratory where samples had been sequenced. This was expected, as RNAseq samples sequenced from different laboratories are comparable provided that recommended laboratory procedures are followed and sequence reads are filtered appropriately[55]. However, ten of 96 samples were outliers, and these were removed from further analyses (Table S3.3). Seven of these removed outlier samples were from Albert (1dag_Alb_Inf_2, 5dag_Alb_Inf_1, 1dag_Alb_Ctl_3, 2dag_Alb_Ctl_3, 5dag_Alb_Ctl_2, 8dag_Alb_Ctl_1 and 45_dag_Alb_Ctl_1) (Supplementary Fig. S.3.1a), and three were from Namikonga (2dag_Nam_Ctl_2, 54dag_Nam_Ctl_3 and 6hag_Nam_Ctl_2) (Supplementary Fig. S.3.1b).

In Albert, once the outlier samples were removed, and reads re-normalized, the median gene expression values were comparable across all time points (Supplementary Fig. S.3.2a). The treatments had a distinct median range of filtered, normalized reads cutting across UCBSV-inoculated and mock-inoculated treatments (approximately 5.8-6.2). The median of the gene expression values in the 1 and 45 dag samples was slightly lower (5.8-6.0) compared to other time points (6.0-6.2).

For Namikonga, the median of the re-normalized gene expression values was within the same range (approx. 6.5) (Supplementary Fig. S.3.2b), as was the number of outlier genes above the upper quartile range. A slightly higher number of outlier genes were recorded at 45 dag, and the reverse at 54 dag.

Samples with medians below the lower quartile were only observed at 6 hag (one UCBSV-inoculated and one mock-inoculated sample), 8 dag (three UCBSV-inoculated and one mock-inoculated sample) and 54 dag (two UCBSV-inoculated and mock-inoculated samples). After checking the distribution of the data, DESeq [51] was used to identify DEGs (Supplementary Data S5) using at least two biological replicates per treatment at each time point (Table S3.3).

3.4.4 Identification of DEGs between mock- and UCBSV-inoculated samples of susceptible (Albert) and resistant (Namikonga) varieties at different time points

To identify DEGs, UCBSV-inoculated samples were compared with mock-inoculated samples at each time point, per variety. We defined DEGs as those that were computed by DESeq to have a false discovery rate (FDR) for differential expression of less than 10% for a particular treatment comparison. We did not apply an additional filter of log2 fold change (log2FC) values, but DEGs with positive or negative log2FC values were classified as up regulated (log2FC≥0), or down regulated (log2FC≤0) for a treatment, respectively. In total, more genes were differentially expressed at particular time points in Namikonga (10,028) compared to Albert (688) (Fig. 3.4). It is unlikely that this is due to sequence differences between these varieties and the reference genome, since both had a similar proportion of unmapped reads (approximately 21%) (Table 3.1). For Namikonga, there were approximately equal numbers of up- and down regulated genes at all time points, except 1 dag, when only 26% of the DEGs were up regulated. For Albert, more than 75% of the DEGs were up regulated at 1 and 2 dag, whereas there were more down regulated genes at the other time points.

In Namikonga, the greatest differential gene expression (either down regulated or up regulated) occurred at the early time points, particularly 2 dag (3887 DEGs), 5 dag (4911 DEGs) and 8 dag (1438 DEGs) (Fig. 3.4). The number of DEGs in Albert in the early time phase was much lower than in Namikonga, with the highest being at 6 hag with only 92 DEGs. At the late time points, few genes were differentially expressed in either variety. Albert had 543 DEGs at 45 dag (which was the maximum for Albert) and none at 54 dag. Namikonga had no DEGs at 45 dag but 738 at 54 dag (Fig. 3.4). All statistically significant DEGs were further characterized by functional annotation and GO term enrichment.

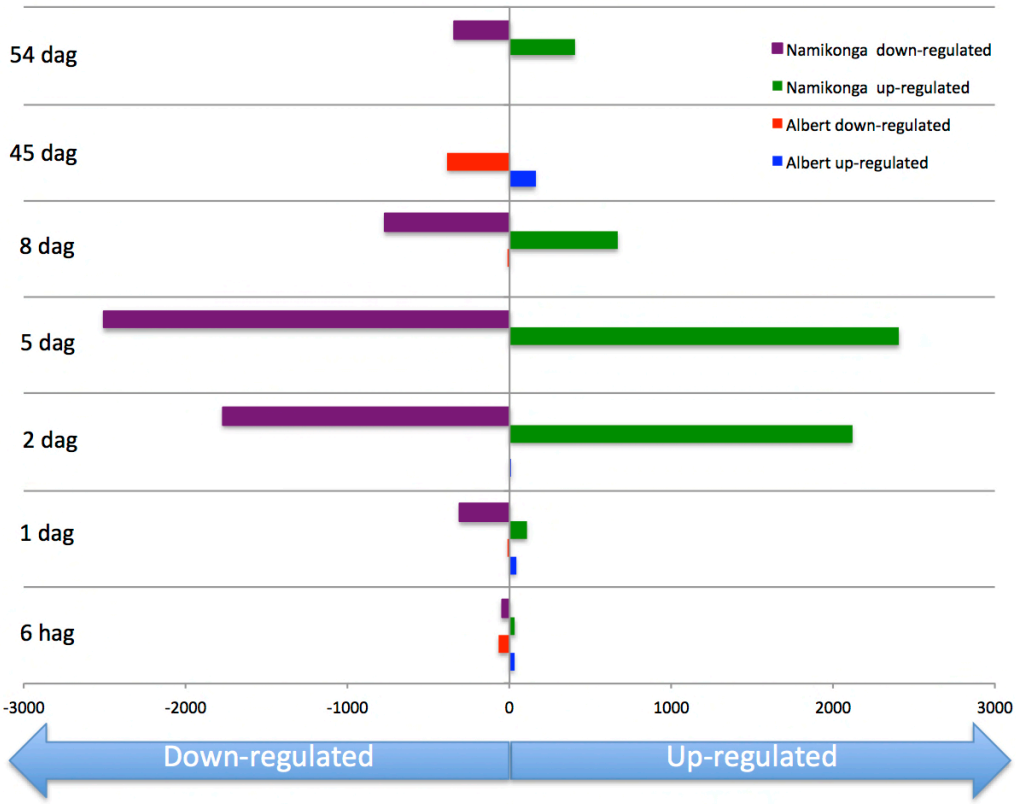


Fig. 3.4: Number of DEGs between UCBSV-inoculated and mock-inoculated cassava varieties Albert and Namikonga at 6 hag, 1, 2, 5, 8, 45 and 54 dag. The DEGs were identified using the DESeq method[51].

3.4.5 GO enrichment of DEGs

Over-represented GO terms were identified among the up regulated and down regulated genes separately at each time point and in each variety. The numbers of enriched GO terms for each treatment are presented in Table S3.4. Namikonga had the largest number of DEGS at 2 dag (3887 DEGs) and 5 dag (4911 DEGs), which also corresponded to the largest number of over-represented GO terms (Table S3.4). For this reason, these time points were the focus for further analysis, particularly the identification of enriched defence-related GO terms.

Enriched GO terms in Namikonga that are related to defence responses

Among the over-represented GO terms of Namikonga identified at 2 and 5 dag (Fig. 3.5), the following terms that are likely to be related to pathogen defence[33,57,58] were identified: translational elongation (GO:0006414), translation factor activity, nucleic acid binding (GO:0008135), ribosomal subunit (GO:0044391) and phosphorelay signal transduction (GO:0000160 and GO:0000156) were each represented by 12, 19, 13 and 18 (17 in GO:0000156) DEGs, respectively (Table 3.2). None of these defence-related GO terms were over-represented in Albert. Individual genes with these GO terms were significantly differentially expressed in Namikonga, but not in Albert. The expression of genes with these defence-related GO terms was further examined.

Table 3.2: GO term over-representation analysis of the up regulated genes of Namikonga at 2 and 5 dag.

GO ID	Enriched GO terms found within the up regulated genes of Namikonga at 2 dag	No. of DEGs	Adj. p-value
	Biological process	No. of DEGs	Adj. p-value
GO:0043623	Cellular protein complex assembly	10	2.00E-02
GO:0006152	Purine nucleoside catabolic process	10	5.E-03
GO:1901136	Carbohydrate derivative catabolic process	12	2.E-02
GO:1901565	Organonitrogen compound catabolic process	14	1.E-02
GO:0000160	Phosphorelay signal transduction system	15	2.E-02
GO:0006457	Protein folding	26	5.E-03
GO:0044710	Single-organism metabolic process	250	1.E-02
GO:0008152	Metabolic process	622	1.E-02
	Molecular function	No. of DEGs	Adj. p-value
GO:0016491	Oxidoreductase activity	161	3.E-02
GO:0003824	Catalytic activity	667	9.E-04
	Enriched GO terms found within the up regulated genes of Namikonga at 5 dag		
	Biological process	No. of DEGs	Adj. p-value
GO:0006414	Translational elongation	12	3.E-03
GO:0000160	Phosphorelay signal transduction system	18	1.E-02
GO:0006457	Protein folding	27	3.E-02
GO:0006412	Translation	205	3.E-68
GO:0034645	Cellular macromolecule biosynthetic process	238	1.E-59
GO:0009059	Macromolecule biosynthetic process	241	5.E-56
GO:0044249	Cellular biosynthetic process	288	3.E-43
GO:1901576	Organic substance biosynthetic process	292	1.E-42
GO:0009058	Biosynthetic process	309	5.E-40
GO:0044267	Cellular protein metabolic process	329	2.E-15
GO:0019538	Protein metabolic process	370	1.E-10
GO:0044260	Cellular macromolecule metabolic process	414	4.E-15
GO:0043170	Macromolecule metabolic process	447	3.E-11
GO:0044237	Cellular metabolic process	541	4.E-18
GO:0044238	Primary metabolic process	582	2.E-10
GO:0071704	Organic substance metabolic process	609	3.E-11
GO:0009987	Cellular process	688	6.E-13
GO:0008152	Metabolic process	786	1.E-04

GO ID	Enriched GO terms found within the up regulated genes of Namikonga at 2 dag	No. of DEGs	Adj. p-value
	Molecular function		
GO:0000156	Phosphorelay response regulator activity	17	5.E-03
GO:0008135	Translation factor activity, nucleic acid binding	19	1.E-02
GO:0003924	Gtpase activity	23	5.E-03
GO:0003723	RNA binding	86	4.E-12
GO:0003735	Structural constituent of ribosome	202	3.E-78
GO:0005198	Structural molecule activity	206	3.E-73
	Cellular component		
GO:0042719	Mitochondrial intermembrane space protein transporter complex	5	3.E-02
GO:0015934	Large ribosomal subunit	9	4.E-03
GO:0044391	Ribosomal subunit	13	3.E-04
GO:0005840	Ribosome	195	4.E-63
GO:0043228	Non-membrane-bounded organelle	198	9.E-56
GO:0030529	Ribonucleoprotein complex	206	3.E-64
GO:0005622	Intracellular	219	3.E-19
GO:0044444	Cytoplasmic part	249	1.E-47
GO:0032991	Macromolecular complex	251	9.E-28
GO:0043226	Organelle	255	2.E-21
GO:0044424	Intracellular part	340	2.E-14
GO:0044464	Cell part	416	9.E-09

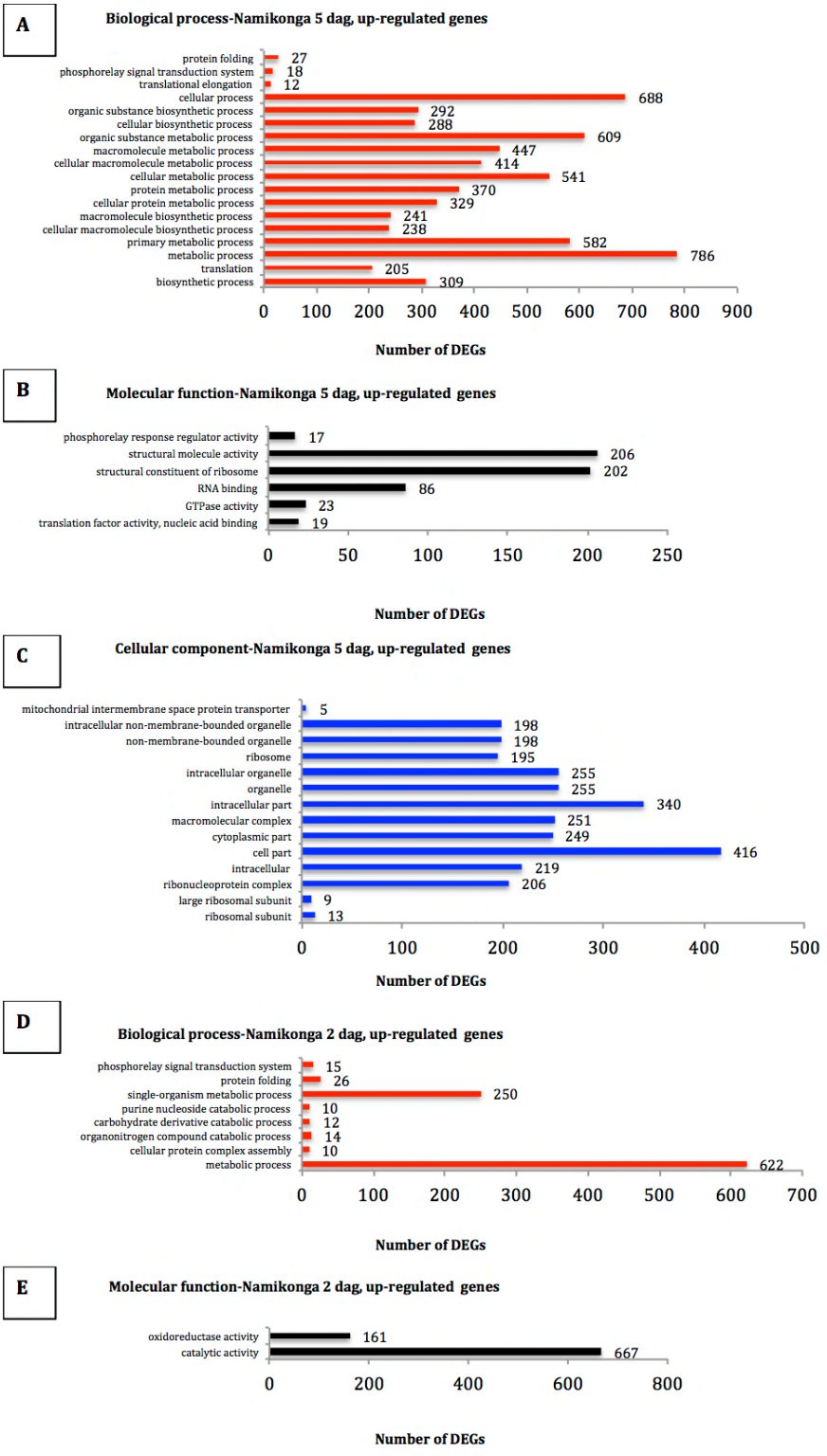


Fig. 3.5: Over-represented GO terms for up regulated genes in Namikonga at 2 and 5 dag. A=Over-represented GO terms of the category “biological process” in the Namikonga up regulated genes at 5 dag; B=Over-represented GO terms of the category “molecular function” in the Namikonga up regulated genes at 5 dag; C=Over-represented GO terms of the category “cellular component” in the Namikonga up regulated genes at 5 dag; D=Over-represented GO terms of the category “biological process” in the Namikonga up regulated genes at 2 dag; E=Over-represented GO terms of the category “molecular function” in the Namikonga up regulated genes at 2 dag. Results are presented for over-represented GO terms with adjusted p-values in the 95% confidence interval.

Translation initiation in the ribosomal subunit

The role of eIF4E genes in Potyvirus resistance has been demonstrated in many plants[28]. We identified two eIF4E genes (Cassava4.1_016620m.g and Cassava4.1_013223m.g) for which reads were mapped in our dataset. Neither gene was significantly differentially expressed in our experiment, except for Cassava4.1_016620m.g in Namikonga at a single time point (5 dag) (log2FC=0.72; adj. p-value=0.02) (Table S3.5). The eIF4E gene Cassava4.1_013223m.g [20] had a log2FC of -0.17 and 0.63 at 2 and 5 dag in Namikonga. This gene was up regulated by a log2FC of 0.1 and 0.4 in Albert at 2 and 5 dag, which was not statistically significant.

Although the data did not show transcriptional differences in eIF4E, there were three main enriched GO terms (GO:0006414, GO:0008135 and GO:0044391) associated with translation in Namikonga at the early time points. This indicates, in general, the potential role(s) of translation in CBSD resistance in Namikonga, and therefore individual genes are reported here. Within these GO terms, Namikonga had several families of translation factors, including eIF4E, EF1B, EF-Ts, 5A-1, eIF3 subunit 7, eIF2 subunit 1, 3B1, and IF2/1F5, each having one or two genes identified at 2 dag and/or 5 dag (Table 3.3). The two families of EF-T genes (5A-1 and GTP binding EF-Tu) had a log2FC≤1 in both varieties, yet were statistically significant and up regulated in Namikonga, in contrast to Albert. The EF1B gene was up regulated in Namikonga, but the results were not significant in either variety. Within the same three GO terms, genes encoding ribosomal protein L10 and 60S acidic ribosomal proteins were statistically significantly up regulated in Namikonga, but non-significantly down regulated in Albert (Table 3).

Table 3.3: Comparison of UCBSV-induced genes corresponding to the over-represented GO terms 'translational elongation', 'translation factor activity, nucleic acid binding', and 'ribosomal subunit' at 2 and 5 dag in Namikonga and Albert.

		2 dag		5 dag	
Gene ID	Gene annotation	Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
GO term: Translational elongation (GO:0006414)					
Cassava4.1_015319m.g	Translation elongation factor EF1B/ribosomal protein S6 family protein	0.0(1)	-0.7(1.3E-01)	0.2(1)	0.6(6.8E-02)
Cassava4.1_010349m.g	Translation elongation factor Ts (EF-Ts), putative	-0.4(1)	-0.5(5.7E-01)	0.5(1)	1.0(1.3E-02)
Cassava4.1_018059m.g	Eukaryotic elongation factor 5A-1	0.0(1)	0.5(3.1E-01)	-0.9(1)	0.9(6.0E-04)
Cassava4.1_018052m.g	Eukaryotic elongation factor 5A-1	-0.2(1)	0.2(8.1E-01)	-0.1(1)	0.9(5.5E-03)
Cassava4.1_007378m.g	GTP binding Elongation factor Tu family protein	-0.2(1)	0.6(3.1E-01)	-0.3(1)	1.7(4.5E-05)
Cassava4.1_011934m.g	Ribosomal protein L10 family protein	-0.1(1)	-0.2(8.7E-01)	-0.4(1)	0.8(7.5E-03)
Cassava4.1_019599m.g	60S acidic ribosomal protein family	0.0(1)	-0.5(4.4E-01)	-0.6(1)	1.9(6.7E-10)
Cassava4.1_019305m.g	60S acidic ribosomal protein family	0.0(1)	-0.1(1.0E+00)	-0.7(1)	0.9(3.7E-03)
Cassava4.1_020132m.g	60S acidic ribosomal protein family	0.1(1)	-0.2(8.2E-01)	-0.9(1)	0.7(1.5E-02)

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
Cassava4.1_019568m.g	60S acidic ribosomal protein family	-0.1(1)	-0.2(8.9E-01)	0(1)	0.6(4.4E-02)
Cassava4.1_019588m.g	60S acidic ribosomal protein family	0.0(1)	-0.1(9.4E-01)	-0.5(1)	0.8(4.2E-03)
Cassava4.1_018344m.g	60S acidic ribosomal protein family	0.0(1)	0.2(7.2E-01)	-0.7(1)	1.0(5.3E-04)
GO term: Translation factor activity, nucleic acid binding (GO:0008135)					
Cassava4.1_016620m.g	Eukaryotic initiation factor 4E protein	0.2(1)	-0.2(8.6E-01)	0.4(1)	0.7(2.0E-02)
Cassava4.1_013223m.g	Eukaryotic translation initiation factor 4E	0.1(1)	-0.2(9.1E-01)	0.4(1)	0.6(6.2E-02)
Cassava4.1_004498m.g	Eukaryotic translation initiation factor 3 subunit 7 (eif-3)	-0.1(1)	-0.4(5.0E-01)	0.1(1)	0.8(1.0E-02)
Cassava4.1_033528m.g	Eukaryotic initiation factor 3 gamma subunit family protein	-0.1(1)	0.9(5.8E-01)	1.7(1)	1.6(9.2E-02)
Cassava4.1_011050m.g	Eukaryotic translation initiation factor 2 subunit 1	0.0(1)	-0.5(3.2E-01)	-0.6(1)	0.7(2.4E-02)
Cassava4.1_017333m.g	Eukaryotic initiation factor 3 gamma subunit family protein	-0.2(1)	0.7(3.1E-01)	0.8(1)	1.5(1.0E-03)
Cassava4.1_018052m.g	Eukaryotic elongation factor 5A-1	-0.2(1)	0.2(8.1E-01)	-0.1(1)	0.9(5.5E-03)

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
Cassava4.1_018059m.g	Eukaryotic elongation factor 5A-1	0.0(1)	0.5(3.1E-01)	-0.9(1)	0.9(6.0E-04)
Cassava4.1_001203m.g	Eukaryotic translation initiation factor 3C	-0.2(1)	-0.5(3.2E-01)	0.3(1)	1.5(2.0E-07)
Cassava4.1_020990m.g	Eukaryotic release factor 1-3	-0.1(1)	-0.4(5.7E-01)	-0.2(1)	0.6(7.8E-02)
Cassava4.1_002530m.g	Translation initiation factor 3B1	0.1(1)	-0.1(9.7E-01)	-0.3(1)	0.8(1.3E-02)
Cassava4.1_002467m.g	Translation initiation factor 2, small GTP-binding protein	0.0(1)	-0.3(7.4E-01)	0.6(1)	0.6(1.1E-01)
Cassava4.1_015319m.g	Translation elongation factor EF1B/ribosomal protein S6 family protein	0.0(1)	-0.7(1.3E-01)	0.2(1)	0.6(6.8E-02)
Cassava4.1_007596m.g	Translation initiation factor IF2/IF5	-0.1(1)	0.1(9.9E-01)	-0.4(1)	0.7(1.5E-02)
Cassava4.1_007700m.g	Translation initiation factor IF2/IF5	0.1(1)	-0.6(1.7E-01)	-0.1(1)	0.6(4.3E-02)
Cassava4.1_010349m.g	Translation elongation factor Ts (EF-Ts), putative	-0.4(1)	-0.5(5.7E-01)	0.5(1)	1.0(1.3E-02)
Cassava4.1_007378m.g	GTP binding Elongation factor Tu family protein	-0.2(1)	0.6(3.1E-01)	-0.3(1)	1.7(4.5E-05)
Cassava4.1_018593m.g	Nucleic acid-binding, OB-fold-like protein	0.2(1)	-0.6(3.0E-01)	0(1)	0.9(5.1E-02)
Cassava4.1_005116m.g	Peptide chain release factor 2	-0.1(1)	1.2(4.5E-01)	1.6(1)	2.0(8.2E-04)

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
GO term: Ribosomal subunit (GO:0006414)					
Cassava4.1_018516m.g	Translation protein SH3-like family protein	-0.1(1)	-0.2(8.7E-01)	-0.5(1)	0.5(9.9E-02)
Cassava4.1_016258m.g	Ribosomal protein 5B	-0.2(1)	-0.4(4.7E-01)	0.2(1)	1.0(2.2E-04)
Cassava4.1_013486m.g	Ribosomal protein l18e/L15 superfamily protein	-0.1(1)	0.5(4.8E-01)	-0.2(1)	0.8(8.6E-02)
Cassava4.1_015972m.g	Ribosomal protein l1p/L10e family	-0.2(1)	0.3(6.2E-01)	-0.4(1)	1.1(7.9E-05)
Cassava4.1_015937m.g	Ribosomal protein l1p/L10e family	-0.1(1)	-0.1(9.8E-01)	0.1(1)	0.8(3.6E-03)
Cassava4.1_009639m.g	Ribosomal protein l1p/L10e family	0.3(1)	0.4(7.4E-01)	0.4(1)	2.0(9.5E-08)
Cassava4.1_009636m.g	Ribosomal protein l1p/L10e family	0.4(1)	0.0(1.0E+00)	0.3(1)	1.5(9.8E-05)
Cassava4.1_008473m.g	Ribosomal protein l1p/L10e family	-0.1(1)	-0.8(8.2E-02)	0.4(1)	0.7(3.6E-02)
Cassava4.1_017425m.g	Ribosomal protein l22p/L17e family protein	-0.2(1)	-0.2(8.3E-01)	-0.5(1)	1.1(8.8E-05)
Cassava4.1_019192m.g	Ribosomal protein l22p/L17e family protein	0.0(1)	-0.4(5.2E-01)	-0.4(1)	1.2(2.0E-05)
Cassava4.1_012305m.g	40s ribosomal protein SA	-0.3(1)	-0.1(1.0E+00)	1.1(1)	1.8(7.2E-08)

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
Cassava4.1_012175m.g	40s ribosomal protein SA	-0.3(1)	-0.2(8.8E-01)	0.6(1)	1.0(2.4E-03)
Cassava4.1_012280m.g	40s ribosomal protein SA	-0.3(1)	0.0(1.0E+00)	1.7(1)	1.3(6.4E-04)

Phosphorelay signal transduction and response regulation

There were two GO terms associated with phosphorelay signal transduction: phosphorelay signal transduction system (GO:0000160) and phosphorelay response regulator activity (GO:0000156). The first GO term contained 18 genes that encoded histidine-related proteins and response regulators. All 18 genes were up regulated in Namikonga (log2FC of 0.8-3.5), but in Albert, 12 genes were down regulated, while the other six were expressed at very low levels (log2FC of 0.0 - 0.8). The other GO term, phosphorelay response regulator activity, had 17 genes that were expressed in the same pattern as the first GO term (GO:0000160) (Table 3.4).

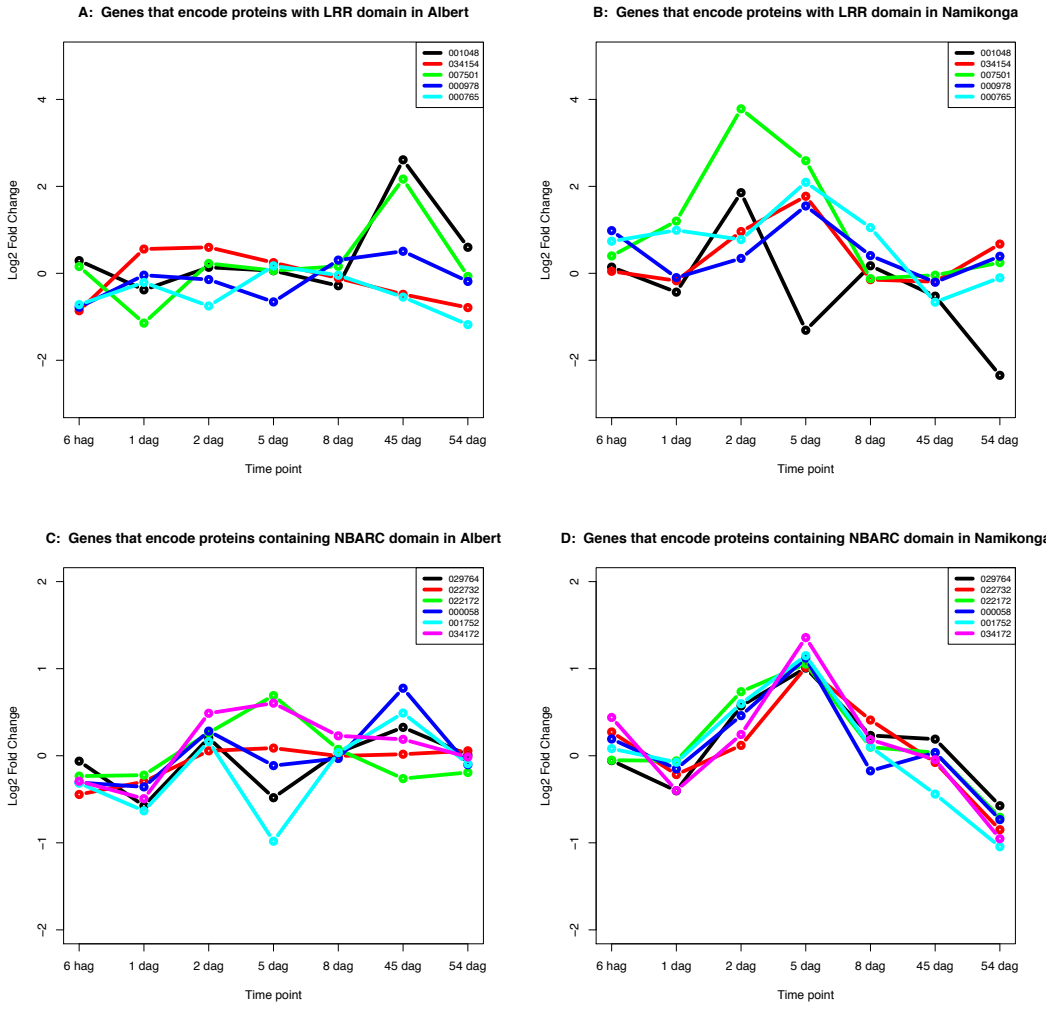
Table 3.4: Comparison of UCBSV-induced genes corresponding to the over-represented GO terms 'phosphorelay signal transduction system' and 'response regulator activity' at 2 and 5 dag in Namikonga and Albert.

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
GO term: Phosphorelay signal transduction system (GO:0000160) and response regulator activity (GO:0000156)					
Cassava4.1_000395m.g	Histidine kinase 2	0.1(1)	1.3(6.2E-04)	-0.9(1)	2.1(8.6E-14)
Cassava4.1_000859m.g	CHASE domain containing histidine kinase protein	0.1(1)	2.6(3.3E-06)	-0.3(1)	2.1(1.6E-11)
Cassava4.1_001670m.g	Histidine kinase 2	0.8(1)	4.7(1.6E-36)	0.1(1)	2.3(3.7E-18)
Cassava4.1_002119m.g	Pseudo-response regulator 7	0.3(1)	1.8(1.0E-07)	-0.1(1)	0.8(3.8E-03)
Cassava4.1_002915m.g	Response regulator 12	0.6(1)	1.5(1.7E-05)	0.1(1)	2.4(1.1E-10)
Cassava4.1_015446m.g	Response regulator 9	0.8(1)	4.6(1.3E-04)	-1.5(1)	3.5(1.4E-14)
Cassava4.1_015815m.g	Response regulator 9	-0.2(1)	4.0(6.9E-24)	-1.5(1)	1.8(2.3E-01)
Cassava4.1_018229m.g	Histidine-containing phosphotransmitter 1	0.4(1)	1.0(3.0E-01)	-0.5(1)	1.4(4.1E-07)
Cassava4.1_018306m.g	Histidine-containing phosphotransmitter 1	-0.4(1)	3.2(7.0E-07)	0.8(1)	-1.1(5.0E-01)
Cassava4.1_022288m.g	Response regulator 9	0.9(1)	4.1(1.6E-01)	0.3(1)	2.5(8.7E-08)
Cassava4.1_022850m.g	Pseudo-response regulator 7	0.2(1)	4.4(1.0E-14)	-0.5(1)	2.5(4.5E-08)
Cassava4.1_024410m.g	Histidine kinase 5	-0.2(1)	3.8(1.9E-02)	2.6(1)	0.1(1.0E+00)
Cassava4.1_025220m.g	Pseudo-response regulator 3	0.7(1)	4.3(2.5E-16)	-0.5(1)	1.2(1.7E-03)
Cassava4.1_026975m.g	Response regulator 3	0.4(1)	4.1(3.0E-03)	0.0(1)	3.5(1.3E-12)
Cassava4.1_033332m.g	Pseudo-response regulator 9	0.4(1)	1.3(5.6E-04)	-0.7(1)	1.0(2.9E-04)
Cassava4.1_002375m.g	Signal transduction histidine kinase, hybrid-type, ethylene sensor	-0.6(1)	-2.6(4.8E-06)	-0.2(1)	0.8(1.2E-02)

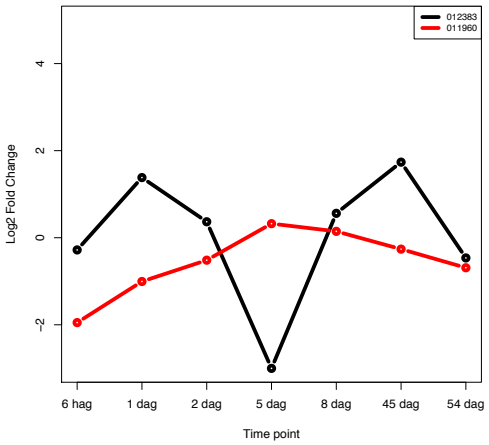
Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
Cassava4.1_015797m.g	Response regulator 5	-0.4(1)	1.5(5.0E-01)	-1.3(1)	1.3(1.6E-02)
Cassava4.1_023472m.g	Response regulator 3	-0.7(1)	0.2(9.4E-01)	0.6(1)	1.4(3.8E-03)
Cassava4.1_027609m.g	Response regulator 9	0.7(1)	0.2(9.6E-01)	-0.4(1)	1.8(2.3E-07)
Cassava4.1_027924m.g	Signal transduction histidine kinase, hybrid-type, ethylene sensor	0.1(1)	0.2(8.3E-01)	0.8(1)	0.8(5.2E-02)
Cassava4.1_028820m.g	Response regulator 2	0.2(1)	1.4(8.4E-02)	-0.2(1)	1.4(4.8E-02)

3.4.6 Specific defence response genes are highly expressed in Namikonga at specific time points

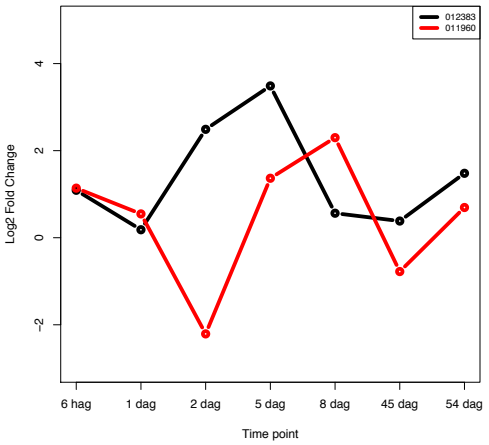
Expression patterns (shown by log2FC) of 55 manually selected defence response genes were examined over the time course (6 hag, 1, 2, 5, 8, 45 and 54 dag) in Albert and Namikonga (Fig. 3.6). The genes were manually selected based on earlier studies in which these genes were shown to function in pathogen defence. The genes included those with leucine-rich repeats (LRRs) (Fig. 3.6A and Fig. 3.6B), those that have a nucleotide binding domain (NB-ARC) (Fig. 3.6C and Fig. 3.6D), pathogenesis-related (PR) proteins (Fig. 3.6E and Fig. 3.6F)[59], late embryogenesis abundant (LEA) proteins (Fig. 3.6G and Fig. 3.6H), transcription factors WRKY; (NAM, ATAF and CUC (NAC)); NmrA; eIF; GATA; and GRAS) (Fig. 3.6I, Fig. 3.6J, Fig. 3.6O Fig. 3.6P), chaperones (Fig. 3.6K and Fig. 3.6L) and heat shock proteins (HSP) (Fig. 3.6M and Fig. 3.6N). Both LRR and NB-ARC domains are structural units of the NOD-like receptors (NLRs), which are defence proteins whose structure contains a TIR or CC domain at the N-terminus, LRRs at the C-terminus and a centrally located NB-ARC domain[60]. Of these, seven gene families (LRR, NB-ARC, PR, LEA, Chaperone, HSP and TF) were significantly different between Albert and Namikonga, and the details are presented below.



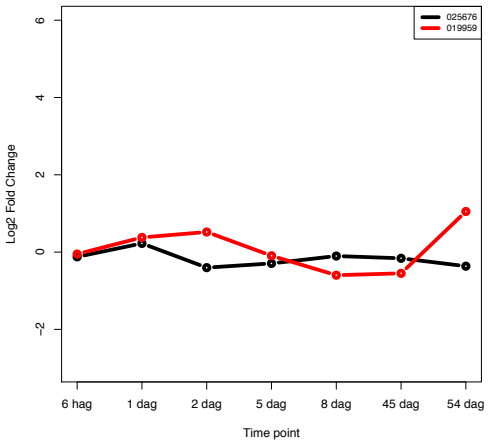
E: Genes that encode PR proteins in Albert



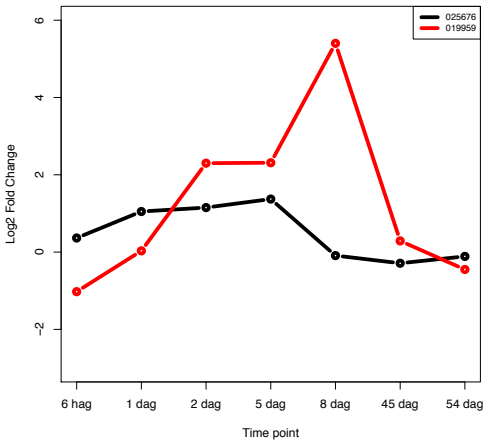
F: Genes that encode PR proteins in Namikonga



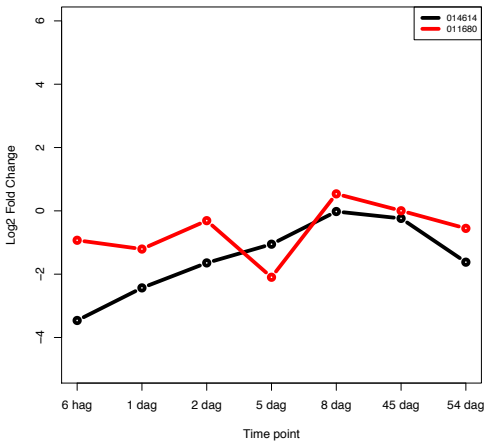
G: Genes that encode LEA proteins in Albert



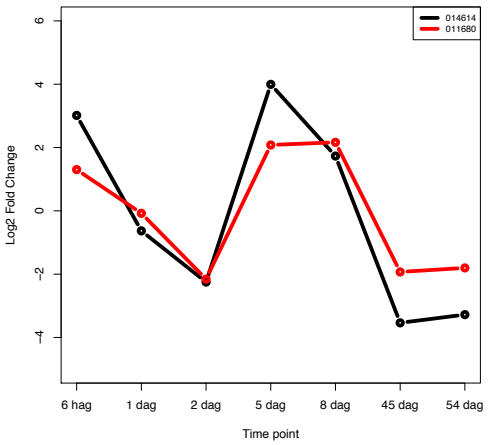
H: Genes that encode LEA proteins in Namikonga



I: Genes that encode WRKY proteins in Albert



J: Genes that encode WRKY proteins in Namikonga



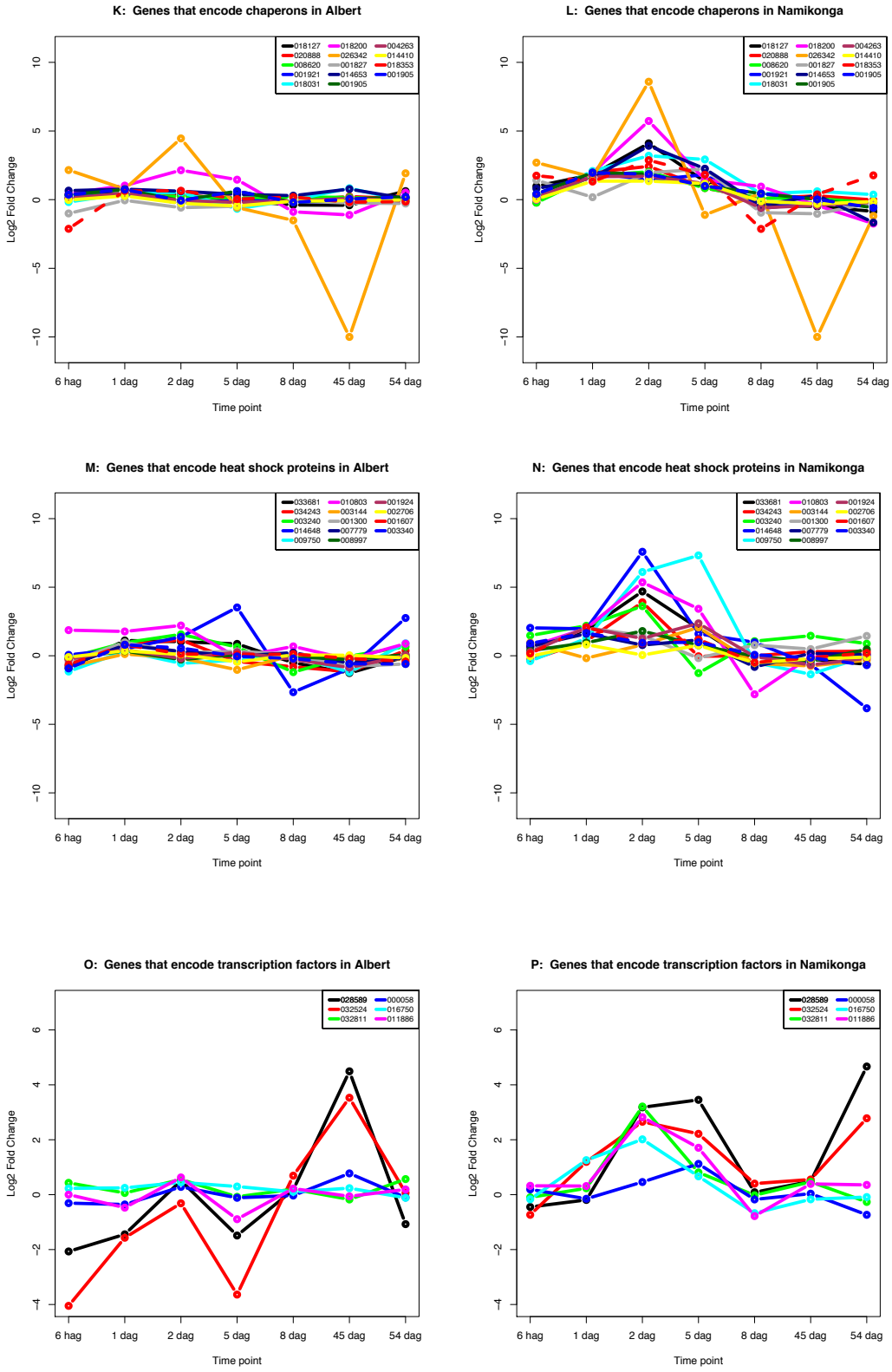


Fig. 3.6: Selected DEGs that were expressed at significantly different levels in susceptible (Albert) and resistant (Namikonga) varieties at different time points (RNAseq data). Gene families include A and B=Leucine-rich repeat (LRR)-containing proteins, C and D=NB-ARC-containing proteins, E and F=Pathogenesis-related (PR) proteins, G and H=Late embryogenesis abundant (LEA) proteins, I and J=WRKY DNA, K and L=Heat shock proteins, M and N=Chaperones and O and P=Transcription factors.

LRR domain

In Namikonga, two (007501 and 001048) of the five selected LRR genes were up- regulated at 2 dag and four (007501, 000978, 034154 and 000765) were up regulated at 5 dag. For Albert, only two genes (007501 and 001048) were up regulated at a later time point, 54 dag.

NB-ARC domain

For Namikonga, all six selected NB-ARC genes were strongly up regulated at 5 dag with similar profiles, suggesting their co-expression and co-regulation. For Albert, only a few genes were induced (034172 and 022172) at that same time point.

PR proteins (thaumatin-like superfamily, PR-5 family)

Two PR genes (012383 and 011960) that belong to the PR-5 family[61-63], thaumatin-like superfamily, were analysed in Namikonga and Albert. In Namikonga, gene 012383 was up regulated at 2 and 5 dag, while gene 011960 had delayed induction. In Albert, however, gene 012383 was up regulated at 1 dag and later at 45 dag.

LEA proteins

In Namikonga, gene 019959 was highly induced in Namikonga at 2, 5 and 8 dag, while gene 025676 was induced at 1, 2 and 5 dag but at a lower level than that of gene 019959. In Albert, however, both genes remained unchanged across all time points.

Chaperones

Of the 14 selected chaperones, gene 026342 was induced in both Namikonga and Albert at 2 dag but to a greater extent in Namikonga (log2FC=18) compared to Albert (log2FC=5). The same gene (026342) was repressed in both varieties at 45 dag. In Namikonga, several other chaperones were induced at 2 and 5 dag; however, gene 026342 showed the only interesting pattern in Albert, with the other genes remaining unchanged.

HSPs

In Namikonga, six (014648, 009750, 010803, 033681, 034243 and 003240) of the 14 selected HSPs were highly induced at 2 dag (log2FC=3.6 – 7.5). The other eight genes were also induced at other early time points (6 hag, 1, 2 and 5 dag) in Namikonga. For Albert, only two genes, 010803 (induced at 6 hag, 1 and 2 dag) and 003340 (induced at 45 dag), were interesting. The others remained unchanged.

TFs (WRKY, NmrA, GATA, GRAS and NAC)

Of the five families of TFs examined, only two showed interesting patterns. The first, WRKY, had two genes (011680 and 014614) that were both repressed at 2 dag and induced at 5 and 8 dag in Namikonga, whereas only 011680 was induced in Albert at 8 dag. The other TF families (NmrA, GATA, GRAS and NAC) were induced at 2 and 5 dag in Namikonga only. In Albert, two genes (032524 and 028589) that encode NmrA proteins were induced at 45 dag, while the others remained unchanged at all other time points.

3.5.7 Other defence-related genes identified in an earlier study by Maruthi et al., 2014

A transcriptome analysis of Kaleso (identical to Namikonga[64]) and Albert infected with CBSV reported that three NAC transcription factors (011029, 015961 and 023870) and one elongation factor, eIF(iso)4E (016601) were over-expressed. The NAC transcription factors were over-expressed in CBSV-infected Kaleso by 27-139 times compared to CBSV-infected Albert[20], while the eIF was two-fold over-expressed in CBSV-infected Kaleso compared to Albert, in RPKM values. We examined the fold change of the count values of these genes in the present study (Table S3.6 and Table S3.7). Preliminary filtering dropped one NAC gene (016601) together with other genes whose expression was below the expression threshold (see the details of filtering low-quality genes in the methodology section “Filtering RNAseq reads”). The remaining two NAC genes were expressed as follows: In Namikonga, both NAC genes (011029 and 015961) were up regulated at 2 dag (log2FC of 1.0 and 2.9) and 5 dag (log2FC of 5.1 and 1.0). In Albert, both genes were slightly up regulated (log2FC of 0.4) at 2 dag. At 5 dag, one NAC gene was substantially down regulated in Albert (log2FC=-2.4), and the other remained slightly up regulated (log2FC=0.4).

3.5 Discussion

CBSD is a major constraint to the production of the African staple food cassava[1]. Two causative agents have been identified, the Potyviruses CBSV and UCBSV. There is an urgent need to identify molecular markers or biomarkers associated with resistance or tolerance to support the efficient breeding of new cassava varieties. Here, we used an RNAseq approach to identify DEGs between UCBSV-inoculated and mock-inoculated plants in a time-course experiment with two Tanzanian cassava varieties, Namikonga (CBSD resistant) and Albert (CBSD susceptible). The results indicated that Namikonga plants restricted disease progression, limited symptoms to the leaves and maintained a low virus load, while allowing normal root expansion (root bulking) without necrosis or constriction. In Albert, where viral loads were higher, infection with UCBSV caused substantial leaf chlorosis and root necrosis. The results also indicated that a strong resistance response is invoked in Namikonga from 2 dag to 5 dag and reduces from 8 dag onwards to 54 dag. Comparatively little differential gene expression was observed at the early time points (6 hag to 1 dag). The over-represented GO terms indicate the involvement of 'translation initiation' in the 'ribosomal subunit' and 'phosphorelay signal transduction and response regulation'. In addition, many genes with defence-related roles, including those with LRR domains, NLRs, PR proteins, LEA proteins and TFs are highly differentially expressed at these time points, indicating a complex response. Although some of these genes are up regulated in Albert, significant up-regulation tends to occur much later, at 54 dag.

3.5.1 Low virus titre in Namikonga suggests that the defence mechanism controls virus replication

Typical CBSD symptoms were observed on both Namikonga and Albert test plants at harvest (49 dag), demonstrating that Namikonga is not immune to UCBSV and can be graft inoculated (Fig. 3.1). UCBSV-inoculated plants of both varieties tested positive for UCBSV (Fig. 3.1) using CBSDDR and CBSDDF2 primers[43]. No UCBSV sequences were identified in the Namikonga RNAseq reads, except at 45 and 54 dag. The high number of UCBSV sequences recovered from Albert at the early time points confirms that the virus multiplied at a higher rate in Albert than in Namikonga (Fig. 3.3). Susceptible cassava varieties have been shown to harbour a higher virus load than resistant varieties[21,65]. Low UCBSV titres have been reported in field samples of Namikonga that were naturally infected via the CBSVs vector, the whitefly[21,65]. We observed similar trends, as the UCBSV titre remained very low throughout the study and only increased at 54 dag, whereas that of Albert had doubled by 6 hag compared to virus levels in Namikonga at the same time. The slightly increased titre (117 UCBSV reads) at 54 dag in Namikonga might be a result of localized virus multiplication within the leaves of Namikonga, which was previously observed in another study[20], or may indicate the occurrence of a more general fluctuation in viral load, rather than loads remaining constantly low. There was an exponential increase in virus load in Albert at 1 dag, which reduced by 5 dag. Altogether, virus levels remained higher in Albert than in Namikonga up to 54 dag (1660 UCBSV reads), except at 5 dag, when both varieties had very low loads. This observation suggests that defence mechanism in Namikonga controls virus multiplication without necessarily eliminating the virus. Our data support earlier studies in which Namikonga was declared 'resistant' to CBSD[20].

3.5.2 Read mapping to the cassava reference genome

Over 75% of the raw reads mapped to genic regions of the cassava reference genome. Of the unmapped reads, 2% mapped to multiple gene models, irrespective of length. Other possible explanations for the unmapped reads could include sequence variation from SNPs between the cassava reference genome version and Namikonga and Albert. The cassava reference genome is from a partial inbred line (accession number AM560-2) derived by three generations of selfing from variety MCOL-1505 of Latin American origin[49], while Namikonga is a local variety derived from early cassava breeding work at Amani, Tanzania[66], and is known to contain approximately 14% *M. glaziovii – M. esculenta* hybrid genome[41]. Albert is a local, pure *M. esculenta* variety from Tanzania[41]. These differences could explain the unmapped reads, as population structure does exist between cassava germplasms from South America and Africa[67].

3.5.3 GO term enrichment of DEGs

The most over-represented GO terms were identified at 2 and 5 dag in Namikonga, which corresponded to the largest number of DEGs (3887 and 4911, respectively, in Namikonga, compared to two and zero, respectively, in Albert). Of the enriched GO terms at these time points, those involved in plant defence included 'translation elongation factors' (containing eIF genes), 'ribosomal subunit' and 'phosphorelay signal transduction'. None of these GO terms were over-represented in Albert. This observation suggests that translation elongation; the ribosomal subunit and phosphorelay signal transduction have play key roles in the resistance of Namikonga to CBSD.

Translation elongation is among the defence responses in Namikonga

Some viruses utilize the host plant's translation factors to replicate within the host. The virus-encoded protein cap structure (VPg) covalently linked to the 5' end of some viral genomes, including CBSVs, recognizes and binds to translation eIFs on ribosomal subunits. This delivers an RNA helicase to the 5' region, bridges the mRNA to the ribosome and circularizes the mRNA, enabling viral replication[33,68]. Mutations in translation initiation factors can result in structural changes in the protein, likely preventing the interaction of viral RNAs or proteins with host factors and therefore restricting the multiplication or movement of the virus, indirectly rendering the host resistant to that virus. The involvement of eIF genes in passive host plant resistance through an altered virus-host interaction surface has been recognized for many years. Two of the eIFs most commonly implicated in passive host plant resistance are eIF4E and eIF4G and their isoforms eIFiso4E and eIFiso4G. This passive resistance has been widely demonstrated in Potyviruses (reviewed in[28,33,69]), the family to which UCBSV belongs[70]. eIF4E has been repeatedly identified as a naturally occurring, recessively inherited resistance locus in species such as *Pisum sativum* (*sbm1*, *wlv/cyv2*[71]) and *Solanum habrochaites* (*pot-1*[31]). Alleles *pvr1* and *pvr12* from *Capsicum* spp. [32] confer resistance to the potyvirids TEV, PVY and PMV based on eIF4E. Similarly, over-expression of pvr1 from *Capsicum* spp. confers resistance to TEV and PMV in *Solanum lycopersicum* based on eIF4E, and over-expression of pvr12 confers resistance to PVY in *Solanum tuberosum*. The resistance gene pvr6, which functions with pvr1 or pvr12, confers resistance to PVMV and ChiVMV in *Capsicum* spp. based on eIF(iso)4E. In Arabidopsis, an eIF4E gene confers resistance to melon necrotic spot virus (MNSV)[30]. Over-representation of the 'translation initiation factor' and 'ribosomal subunit' GO terms in Namikonga at 2 and 5 dag (Table 3.2) indicates the presence of an eIF-mediated resistance mechanism in the Namikonga

defence response. In addition to these, the term 'phosphorelay signal transduction' was also enriched at 2 and 5 dag in Namikonga.

Phosphorelay signal transduction activates the defence response in Namikonga

The phosphorelay system is a complex interaction of multicomponent regulatory systems in which pathogen-triggered signals are transferred from host extracellular histidine- or histidine-containing proteins to intracellular regulatory proteins, which in turn trigger reactions that alter the plant's physiological patterns of defence[72-76].

More specifically, a phosphoryl group is transferred from a membrane-bound, conserved, sensor histidine kinase to an intracellular, receiver response regulator containing a conserved aspartate residue. The histidine kinases receive signals from extracellular sensors and transmit these signals via multi-step processes to cytoplasmic response regulators. The response regulators localize in the nucleus alone or in both the nucleus and the cytoplasm. The transfer of stimuli between the extracellular histidine kinases and intracellular response regulators occurs in a multi-step manner involving several plant proteins[72,73]. The response regulators in turn trigger a cascade of reactions that lead to pathogen defence and other physiological pathways.

Cytokinins, which are plant growth hormones, have been shown to receive and transmit signals in a manner similar to that of bacterial two-component phosphorelay signaling. In Arabidopsis, cytokinins have receptors (AHK2, AHK3 and AHK4)[77] whose role is similar to that of eukaryotic histidine kinases in the bacterial two-component system. These three proteins have extracellular cytokinin binding, cytoplasmic histidine transmitter and receiver domains, respectively[74,78]. Arabidopsis plants treated with cytokinins were resistant to P. syringae pv. tomato DC3000 (Pst)[79]. Defence against Pst occurs via the salicylic acid (SA) pathway, which is triggered by cytokinins[80]. A receptor-like kinase (P. syringae effector B) was shown to activate the expression of RPM1 (an NBS-LRR protein) against Pst in Arabidopsis[58]. The SA pathway confers resistance to the Potyviruses PVY in potato[81] and peanut mottle virus (PeMV)[82]. In tobacco, plants primed with cytokinins became tolerant to the strain PVYNTN of potato virus Y (PVY)[75]. In this study involving cassava and UCBSV, both histidine kinases and regulatory proteins were among the over-expressed genes in Namikonga at 2 and 5 dag (Table 3.4). In addition to these GO terms and the constituent genes, other defence-related genes were significantly expressed in Namikonga compared to Albert.

3.5.4 A coordinated set of defence genes collectively confer resistance to CBSD in Namikonga

In addition to eIFs and possibly linked to phosphorelay signal transduction, genes that encode LRR- and NB-ARC-containing proteins, PR proteins, LEA proteins, TFs (WRKY, GRAS, GATA and NmrA), chaperones and HSPs were all highly up regulated in Namikonga at 2 and 5 dag. These proteins have all been implicated in defence responses against viral, fungal and bacterial pathogens. Defence genes encoding proteins containing LRRs (Fig. 3.6A and Fig. 3.6B) and NB-ARCs (Fig. 3.6C and Fig. 3.6D) were substantially over-expressed at 2 and 5 dag in Namikonga, the latter time point being the peak of expression for most genes (Fig. 3.6). Other defence genes, including those encoding PR proteins (Fig. 3.6E and Fig.

3.6F), LEA proteins (Fig. 3.6G and Fig. 3.6H), TFs involved in pathogen defence (WRKY, NmrA, GATA, NAC and GRAS) (Fig. 3.6E-F, Fig. 3.6O-P), HSPs (Fig. 3.6M-N) and chaperones (Fig. 3.6K-L), were over-expressed at 2 and 5 dag. This suggests that Namikonga's resistance to UCBSV infection may be complex, involving several possibly interrelated strategies including NLR proteins with LRR and NB-ARC domains.

An NBS-LRR network contributes to Namikonga's resistance to UCBSV

In general, the NB-ARC genes were over-expressed at 5 dag in Namikonga and steadily decreased thereafter. In Albert, some genes were down regulated and others over-expressed (but with lower log2FCs than in Namikonga) at 5 dag. However, three (000058, 001752 and 029764) of these genes were over-expressed at 45 dag in Albert. Maximum LRR gene expression was observed in Namikonga at 2 and 5 dag. When the LRR genes were maximally expressed in Namikonga, their expression was either normal or down regulated in Albert. Two of these genes (001048 and 007501) were over-expressed at a much later time point (45 dag) in Albert, a time point when the expression of these genes in Namikonga had reduced to near normal.

Typically, LRRs, NB-ARCs and CCs or TIRs are subdomains of the tri-modular NLR protein, the most studied type of defence gene[60,83,84]. Alone, LRR and NB-ARC domains can function in pathogen defence. Some NLR proteins require the presence of other 'helper' proteins to develop a resistance phenotype. The 'helper' is often an HSP[85,86]. In the case of Namikonga, synchronized timing for the up-regulation of these HSPs, chaperones and NLRs implies a synergistic interaction of the three defence proteins to orchestrate Namikonga's resistance to CBSD.

HSPs are structurally disordered proteins with diverse functions in regulatory, signaling and defence pathways. Alone, HSPs may function as defence proteins against biotic and abiotic stresses[87,88]. Some HSPs are also chaperones. Chaperones are mainly involved in RNA binding and protein folding[89,90], a role that is key in the functioning of elongation factors[57]. Their potential ability to independently confer resistance phenotypes in Namikonga cannot be ruled out.

During the peak over-expression of LRR, NB-ARC and HSP genes, LEA proteins were also over-expressed, reaching a maximum at 8 dag in Namikonga (Fig. 3.6H). LEA proteins are involved in adaptation to water stress[91] and are abundant in plant[91,92], fungi[93] and mammalian genomes[94-98]. As the name suggests, LEA proteins are produced during the late stages of embryogenesis, especially at seed development. They protect cells from desiccation[91], salinity[99] and extreme cold or hot temperatures[94] in a dose-dependent manner[97]. The expression of LEA genes in Namikonga seems to ensure that the already infected plants maintain balanced cell solutes for normal functioning under the stressful conditions imposed by UCBSV.

Another type of defence proteins that were up regulated in Namikonga at 2 and 5 dag are the PR proteins. The PR proteins are inducible defence-related proteins of varied molecular sizes (5 to 75 kDa)[100]. They were first discovered in tobacco resistant to infection by TMV[59,62]. To date, PR proteins have been associated with resistance to several pathogens[59,62]. The proteins are RNase and DNase active, implying

that they function in defence by abolishing foreign RNA and DNA molecules in host plants. The PR proteins have antiviral properties[101-103], and resistance triggered by PR proteins causes an HR and programmed cell death at the infection site[101,102]. They are critical for secondary metabolite biosynthesis, storage, phytohormones and ligand binding[103]. Their expression marks the peak of systemic acquired resistance (SAR) against pathogens[62]. In Namikonga, PR proteins were up regulated at the peak of LRR, NB-ARC and HSP expression, but the role and extent of involvement of PR proteins in resistance to UCBSV is not clear.

At the same peak time points, five major groups of TFs (NAC, WRKY, GRAS, GATA and NmrA) were over-expressed in Namikonga. Although there is very little knowledge of the transcriptional networks that are activated in cassava plants in response to pathogens, the co-expression of these transcription factors in Namikonga with other defence genes in response to UCBSV inoculation suggests a role in defence.

The NAC (NAM, ATAF and CUC) superfamily of TFs is one of the largest TF families. It is found only in plants and was thus investigated here. The role of NAC proteins in defence is known in several plants. In Arabidopsis, a functional SA pathway is required for NAC-directed defence against viral pathogens such as TCV, but the jasmonic acid (JA) and ethylene (ET) pathways are not essential[104]. A NAC-mediated HR causes resistance in rice to rice blast disease[105], in maize to *C. graminicola*[106] and in Arabidopsis to *Botrytis cinerea*, *P. syringae* pv. *tomato* (Pst)[107], *Turnip crinkle virus*[104] and *Acidovorax avenae*[108]. In this study, NAC genes were differentially expressed at specific time points with other defence genes, implying that NAC has a role in defence. However, to date, no studies have focused on the role of NAC genes in cassava's defence against viral pathogens. Our data indicate that the induction of NAC genes in Namikonga at 2 and 5 dag contributes to NAC-mediated resistance to CBSD in Namikonga.

Another class of transcription factors, WRKY, was also up regulated in Namikonga at 2 and 5 dag. The role of WRKY transcription factors in plant defence[109-111] and their structural features and functional network have been reviewed[112]. In rice, two alleles of a WRKY gene (OsWRKY45-1 and OsWRKY45-2) conferred resistance to *Magnaporthe grisea*, yet each allele encoded a different set of defence genes regulated by distinct promoter regions[113].

In this study, Namikonga had two differentially expressed WRKY genes at 2 and 5 dag. However, one WRKY was down regulated at 2 dag, possibly to establish a network feedback loop to regulate their over-expression at 5 and 8 dag, respectively[113]. In some cases, SA and JA enhance the functionality of WRKY genes in defence. In rice, enhancing the levels of SA and JA production enabled OsWRKY45-1 to resist Xoo and Xoc, but OsWRKY45-2-mediated resistance to Xoo and Xoc required JA only[113]. Whether the SA levels enhanced the activity of the Namikonga WRKY genes against UCBSV was not examined in this study. There is a need to examine the roles of SA and JA in the Namikonga defence mechanism and if and how these phytohormones enhance the action of WRKY genes.

3.6 Conclusion

Based on these findings, a model for resistance to CBSD in Namikonga is proposed (Fig. 3.7). In this model, the challenge of Namikonga with UCBSV triggers an NBS-LRR signalling cascade, which subsequently induces a phosphorelay signal transduction pathway leading to defence. This is marked by induction of defence-related PR-proteins and stress response LEA, heat-shock and chaperone proteins, and a range of proteins associated with translation, including translation elongation factors. This may represent a mechanism in Namikonga of preventing efficient interaction of UCBSV with the host translational machinery, a known mechanism of resistance to potyviruses in other plants[31,32,70,71,114]. The outcome is that Namikonga shows reduced virus titre, limited leaf chlorosis and minimal (if any) root necrosis. In Albert, UCBSV's presence does not induce any recognizable defence pathway and virus titre increases unchecked (Fig. 3.7). The result is major leaf chlorosis and root necrosis. Our study confirms that resistance in Namikonga is associated with significant transcriptional re-programming soon after UCBSV challenge, in contrast to Albert in which the subterfuge of the virus appears to suppress defence gene expression (Fig. 3.7). Furthermore, our data provides a rich source of candidate alleles in the Namikonga genotype for testing hypothesis of resistance mechanisms by correlating gene expression with QTL mapping data, with the ultimate aim of developing robust biomarkers for cassava breeders to develop durable resistance to CBSD.

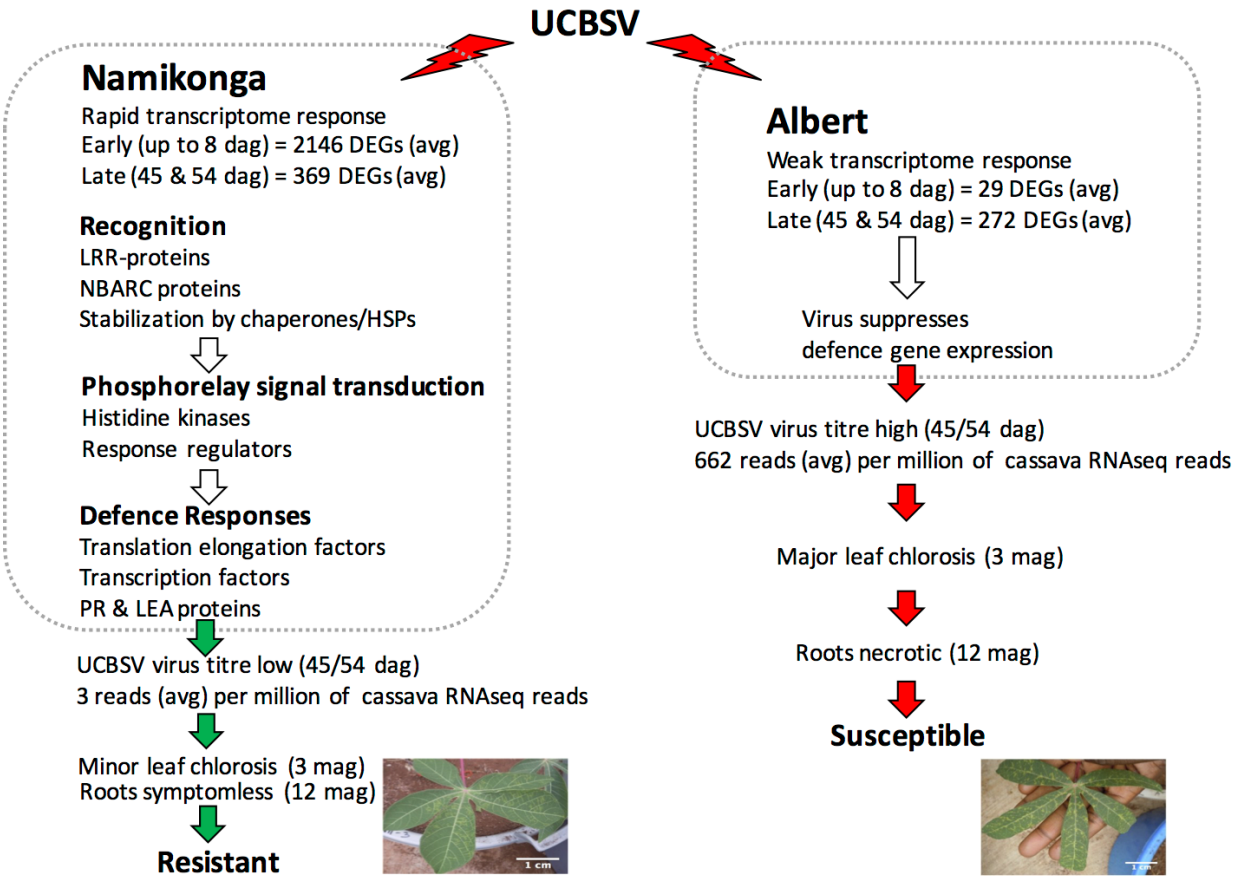


Fig. 3.7: Model of cassava responses to UCBSV based on transcriptomics. DEGs=differentially expressed genes; avg=average; dag=days after grafting and mag=months after grafting.

3.7 References

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

2 Patil, B. L., Ogwok, E., Wagaba, H., Mohammed, I. U., Yadav, J. S., Bagewadi, B., Taylor, N. J., Kreuze, J. F., Maruthi, M. N., Alicai, T. & Fauquet, C. M. RNAi-mediated resistance to diverse isolates belonging to two virus species involved in cassava brown streak disease. Mol Plant Pathol (**2011**), **12**, 31-41.

3 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Genetically distinct strains of cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch Virol (**2009**), **154**, 353-359.

4 Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M. & Butgereitt, A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen Virol (**2010**), **91**, 1365-1372.

5 Storey, H. H. Virus diseases of East African plants. VI-A progress report on studies of the disease of cassava. E Afri Agric J (**1936**), **2**, 34-39.

6 Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, M. A. & Boykin, L. M. Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. Sci Rep (**2016**), **6**, 36164.

7 Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. & Colvin, J. Re-emergence of cassava brown streak disease in Uganda. Plant Dis (**2007**), **91**, 24-29.

8 Monger, W. A., Alicai, T., Ndunguru, J., Kinyua, Z. M., Potts, M., Reeder, R. H., Miano, D. W., Adams, I. P., Boonham, N., Glover, R. H. & Smith, J. The complete genome sequence of the Tanzanian strain of cassava brown streak virus and comparison with the Ugandan strain sequence. Arch Virol (**2010**), **155**, 429-433.

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

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

11 Mware, B., Narla, R., Amata, R., Olubayo, F., Songa, J., Kyamanyua, S. & Ateka, E. M. Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. J Gen Mol Virol (**2009**), **1**, 40-45.

12 Mware, B., Olubayo, F., Narla, R., Songa, J., Amata, R., Kyamanywa, S. & Ateka, E. M. First record of spiraling whitefly in coastal Kenya: emergence, host range, distribution and association with cassava brown streak virus disease. Int J Agric & Biol (**2010**), **12**, 411–415.

13 Hillocks, R. J., Raya, M. D., Mtunda, K. & Kiozia, H. Effects of brown streak virus disease on yield and quality of cassava in Tanzania. J Phytopathol (**2001**), **149**, 389-394.

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

15 Ogwok, E., Patil, B. L., Alicai, T. & Fauquet, C. M. Transmission studies with cassava brown streak Uganda virus (Potyviridae: Ipomovirus) and its interaction with abiotic and biotic factors in *Nicotiana benthamiana*. J Virol Methods (**2010**), **169**, 296-304.

16 Wagaba, H., Beyene, G., Trembley, C., Alicai, T., Fauquet, C. M. & Taylor, N. J. Efficient transmission of cassava brown streak disease viral pathogens by chip bud grafting. BMC Res Not (**2013**), **6**, 516.

17 Adams, I. P., Abidrabo, P., Miano, D. W., Alicai, T., Kinyua, Z. M., Clarke, J., Macarthur, R., Weekes, R., Laurenson, L., Hany, U., Peters, D., Potts, M., Glover, R., Boonham, N. & Smith, J. 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 Pathol (**2013**), **62**, 233-242.

18 Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. Adv Virol. (**2011**), **2012**, 795697.

19 Rwegasira, G. M. & Rey, C. M. E. Relationship between symptoms expression and virus detection in cassava brown virus streak-infected plants. J Agric Sci (**2012**), **4**, 246-253.

20 Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One (**2014**), **9**, e96642.

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

22 Kawuki, R. S., Pariyo, A., Amuge, T., Nuwamanya, E., Ssemakula, G., Tumwesigye, S., Bua, A., Baguma, Y., Omongo, C. A., Alicai, T. & Orone, J. A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). J Plant Breed Crop Sci (**2011**), **3**.

23 Morozova, O., Hirst, M. & Marra, M. A. Applications of new sequencing technologies for transcriptome analysis. Annu Rev Genomics Hum Genet. (**2009**), **10**, 135-151.

24 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods (**2008**), **5**, 621-628.

25 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet (**2009**), **10**, 57-63.

26 de Ronde, D., Butterbach, P. & Kormelink, R. Dominant resistance against plant viruses. Front Plant Sci. (**2014**), **5**, 307.

27 Robaglia, C. & Caranta, C. Translation initiation factors: a weak link in plant RNA virus infection. Trends Plant Sci (**2006**), **11**, 40-45.

28 Truniger, V. & Aranda, M. A. Recessive resistance to plant viruses. Adv Virus Res (**2009**), **75**, 119-231.

29 Kang, B. C., Yeam, I., Li, H., Perez, K. W. & Jahn, M. M. Ectopic expression of a recessive resistance gene generates dominant potyvirus resistance in plants. Plant Biotech J (**2007**), **5**, 526-536.

30 Nieto, C., Morales, M., Orjeda, G., Clepet, C., Monfort, A., Sturbois, B., Puigdomenech, P., Pitrat, M., Caboche, M. & Dogimont, C. An eIF4E allele confers resistance to an uncapped and non‐polyadenylated RNA virus in melon. The Plant J (**2006**), **48**, 452-462.

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

32 Ruffel, S., Gallois, J. L., Lesage, M. L. & Caranta, C. The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. Mol Genet Genomics (**2005**), **274**, 346-353.

33 Duan, C. G., Wang, C. H. & Guo, H. S. Application of RNA silencing to plant disease resistance. Silence (**2012**), **3**, 5.

34 Gonzalez-Ibeas, D., Cañizares, J. & Aranda, M. A. Microarray analysis shows that recessive resistance to watermelon mosaic virus in melon is associated with the induction of defense response genes. MPMI (**2012**), **25**, 107-118.

35 Lellis, A. D., Kasschau, K. D., Whitham, S. A. & Carrington, J. C. Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF (iso) 4E during potyvirus infection. Curr Biol. (**2002**), **12**, 1046-1051.

36 Cosson, P., Schurdi-Levraud, V., Le, Q. H., Sicard, O., Caballero, M., Roux, F., Gall, O. L., Candresse, T. & Revers, F. The RTM resistance to potyviruses in *Arabidopsis thaliana*: natural variation of the RTM genes and evidence for the implication of additional genes. PLoS One (**2012**), **7**.

37 Palloix, A., Ayme, V. & Moury, B. Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. New Phytol (**2009**), **183**, 190 - 199.

38 Huh, S. U., Lee, G. J., Jung, J. H., Kim, Y., Kim, Y. J. & Paek, K. H. Capsicum annuum transcription factor WRKYa positively regulates defense response upon TMV infection and is a substrate of CaMK1 and CaMK2. Sci Rep (**2015**), **5**, 7981.

39 Wright, K., M., Duncan, G. H., Pradel, K. S., Carr, F., Wood, S., Oparka, K. J. & Cruz, S. S. Analysis of the N gene hypersensitive response induced by a fluorescently tagged *Tobacco mosaic virus*. Plant Physiol (**2000**), **123**, 1375 - 1385.

40 Hinrichs, J., Berger, S. & Shaw, J. G. A hypersensitive response-like mechanism is involved in resistance of potato plants bearing the Ry(sto) gene to the potyviruses *Potato virus Y* and *Tobacco etch virus*. J Gen Virol (**1998**), **79**, 167-176.

41 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. & Rokhsar, D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nature Biotechnol (**2016**), **34**, 562-570.

42 Xu, J., Aileni, M., Abbagani, S. & Zhang, P. A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). Phytochem Anal. (**2010**), **21**, 395-398.

43 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. J Virol Methods (**2011**), **171**, 394-400.

44 Abarshi, M. M., Mohammed, I. U., Jeremiah, S. C., Legg, J. P., Kumar, P. L., Hillocks, R. J. & Maruthi, M. N. 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. J Virol Methods (**2012**), **179**, 176-184.

45 Hillocks, R. J., Raya, M. & Thresh, J. M. The association between root necrosis and above‐ground symptoms of brown streak virus infection of cassava in southern Tanzania. Int J Pest Manage (**1996**), **42**, 285-289.

46 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. & Salzberg, S. L. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol (**2013**), **14**, R36.

47 Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (**2010**).

48 Hannon, G. FASTX-Toolkit v.0.0.14. Cold Spring Harbor Laboratory, Long Island. Available online at: http://hannonlab.cshl.edu/fastx_toolkit/ (**2009**).

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

50 Anders, S., Pyl, P. T. & Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. Biorxiv Preprint (**2014**).

51 Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol (**2010**), **11**, R106.

52 R: A language and environment for statistical computing. (R foundation for statistical computing, Vienna, Austria. Retrieved from: http://www.r-project.org/. **2013**).

53 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics (**2009**), **25**, 1105-1111.

54 Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C. E., Socci, N. D. & Betel, D. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. Genome Biol (**2013**), **14**, R95.

55 AC't Hoen, P., Friedländer, M. R., Almlöf, J., Sammeth, M., Pulyakhina, I., Anvar, S. Y., Laros, J. F. J., Buermans, H. P. J., Karlberg, O. & Brännvall, M. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. Nature Biotechnol (**2013**), **31**, 1015-1022.

56 Katari, M. S., Nowicki, S. D., Aceituno, F. F., Nero, D., Kelfer, J., Thompson, L. P., Cabello, J. M., Davidson, R. S., Goldberg, A. P., Shasha, D. E., Coruzzi, G. M. & Gutierrez, R. A. VirtualPlant: a software platform to support systems biology research. Plant Physiol (**2010**), **152**, 500-515.

57 Gingras, A., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. Annu Rev Biochem (**1999**), **68**, 913-963.

58 Liu, J., Elmore, J. M., Lin, Z.-J. D. & Coaker, G. A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. Cell Host Microbe. (**2011**), **9**, 137-146.

59 van Loon, L. C. & van Kammen, A. Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana tabacum* var. ‘Samsun’ and ‘Samsun NN’: II. changes in protein constitution after infection with *Tobacco mosaic virus*. Virology (**1970**), **40**, 199-211.

60 Proell, M., Riedl, S. J., Fritz, J. H., Rojas, A. M. & Schwarzenbacher, R. The NOD-like receptor (NLR) family: a tale of similarities and differences. PLoS One (**2008**), **3**, e2119.

61 Liu, J. J., Sturrock, R. & Ekramoddoullah, A. K. M. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. Plant Cell Rep (**2010**), **29**, 419-436.

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

63 van Loon, L. C. & van Strien, E. A. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Path (**1999**), **55**, 85-97.

64 Kanju, E., Mahungu, N., Dixon, A. G. O. & Whyte, J. Is resistance/tolerance to cassava brown streak disease associated with the zigzag stem trait. Roots (**2003**), **8**, 15-19.

65 Ogwok, E., Alicai, T., Rey, M. E. C., Beyene, G. & Taylor, N. J. Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta*) plants. Plant Pathol (**2015**), **64**, 1235-1246.

66 Kanju, E. E., Masumba, E., Masawe, M., Tollano, S., Muli, B., Zacarias, A., Mahungu, N., Khizzah, B., Whyte, J. & Dixon, A. Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmer and consumer preferences. Proceedings 13th ISTRC Symp (**2007**), 95-101.

67 Ferguson, M. E., Hearne, S. J., Close, T. J., Wanamaker, S., Moskal, W. A., Town, C. D., de Young, J., Marri, P. R., Rabbi, I. Y. & de Villiers, E. P. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. Theor Appl Genet (**2012**), **124**, 685-695.

68 Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. Curr Opin Struct Biol (**2003**), **13**, 56-63.

69 Kang, B. C., Yeam, I. & Jahn, M. M. Genetics of plant virus resistance. Annu Rev Phytopathol (**2005**), **43**, 581-621.

70 Maule, A. J., Caranta, C. & Boulton, M. I. Sources of natural resistance to plant viruses: status and prospects. Mol Plant Pathol (**2007**), **8**, 223-231.

71 Ashby, J. A., Stevenson, C. E., Jarvis, G. E., Lawson, D. M. & Maule, A. J. Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to *Pea seed-borne mosaic virus* in pea. PLoS One (**2011**), **6**, e15873.

72 Bauer, J., Reiss, K., Veerabagu, M., Heunemann, M., Harter, K. & Stehle, T. Structure–function analysis of *Arabidopsis thaliana* histidine kinase AHK5 bound to Its cognate phosphotransfer protein AHP1. Mol Plant (**2013**), **6**, 959-970.

73 Romir, J., Harter, K. & Stehle, T. Two-component systems in *Arabidopsis thaliana* – a structural view. Eur J Cell Biol (**2010**), **89**, 270-272.

74 Sheen, J. Phosphorelay and transcription control in cytokinin signal transduction. Science (**2002**), **296**, 1650-1652.

75 Spoustová, P., Hýsková, V., Müller, K., Schnablová, R., Ryšlavá, H., Čeřovská, N., Malbeck, J., Cvikrová, M. & Synková, H. Tobacco susceptibility to *Potato virus Y*NTN infection is affected by grafting and endogenous cytokinin content. Plant Sci (**2015**), **235**, 25-36.

76 West, A. H. & Stock, A. M. Histidine kinases and response regulator proteins in two-component signaling systems. Trends Biochem Sci (**2001**), **26**, 369-376.

77 To, J. P. C. & Kieber, J. J. Cytokinin signaling: two-components and more. Trends Plant Sci (**2008**), **13**, 85-92.

78 Mähönen, A. P., Higuchi, M., Törmäkangas, K., Miyawaki, K., Pischke, M. S., Sussman, M. R., Helariutta, Y. & Kakimoto, T. Cytokinins regulate a bidirectional phosphorelay network in Arabidopsis. Curr Biol. (**2006**), **16**, 1116-1122.

79 Choi, J., Huh, S. U., Kojima, M., Sakakibara, H., Paek, K.-H. & Hwang, I. The cytokinin-activated transcription factor ARR2 promotes plant immunity via TGA3/NPR1-dependent salicylic acid signaling in Arabidopsis. Dev Cell. (**2010**), **19**, 284-295.

80 Naseem, M. & Dandekar, T. The role of auxin-cytokinin antagonism in plant-pathogen interactions. PLoS Pathog (**2012**), **8**, e1003026.

81 Baebler, Š., Witek, K., Petek, M., Stare, K., Tušek-Žnidarič, M., Pompe-Novak, M., Renaut, J., Szajko, K., Strzelczyk-Żyta, D., Marczewski, W., Morgiewicz, K., Gruden, K. & Hennig, J. Salicylic acid is an indispensable component of the Ny-1 resistance-gene-mediated response against *Potato virus Y* infection in potato. J Exp Bot (**2014**), **65**, 1095-1109.

82 Kobeasy, M. I., El-Beltagi, H. S., El-Shazly, M. A. & Khattab, E. A. H. Induction of resistance in *Arachis hypogaea* L. against *Peanut mottle virus* by nitric oxide and salicylic acid. Physiol Mol Plant Path (**2011**), **76**, 112-118.

83 Maekawa, T., Cheng, W., Spiridon, L. N., Toller, A., Lukasik, E., Saijo, Y., Liu, P., Shen, Q. H., Micluta, M. A., Somssich, I. E., Takken, F. L., Petrescu, A. J., Chai, J. & Schulze-Lefert, P. Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. Cell Host Microbe. (**2011**), **9**, 187-199.

84 Marone, D., Russo, M. A., Laido, G., De Leonardis, A. M. & Mastrangelo, A. M. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. Int J Mol Sci (**2013**), **14**, 7302-7326.

85 Ito, M., Ohnishi, K., Hikichi, Y. & Kiba, A. Molecular chaperons and co-chaperons, HSP90, RAR1, and SGT1 negatively regulate bacterial wilt disease caused by *Ralstonia solanacearum* in *Nicotiana benthamiana*. Plant Signal Behav (**2014**), **10**, e970410.

86 Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., Peng, L., Zhang, L. & Lu, C. Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. Plant Cell (**2013**), **25**, 2925-2943.

87 Mayer, F. L., Wilson, D., Jacobsen, I. D., Miramón, P., Slesiona, S., Bohovych, I. M., Brown, A. J. P. & Hube, B. Small but crucial: the novel small heat shock protein HSP21 mediates stress adaptation and virulence in *Candida albicans*. PLoS One (**2012**), **7**, e38584.

88 Neta-Sharir, I., Isaacson, T., Lurie, S. & Weiss, D. Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. Plant Cell (**2005**), **17**, 1829-1838.

89 Lashley, T., L., H. J., Verbeek, M. M., Rostagno, A., Bojsen-Moller, M., David, G., van Horssen, J., Braendgaard, H., Plant, G., Frangione, B., Ghiso, J. & Revesz, T. Molecular chaperons, amyloid and preamyloid lesions in the BRI2 gene-related dementias: a morphological study. Neuropathol Appl Neurobiol. (**2006**), **32**, 492-504.

90 Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. FASEB J (**2004**), **18**, 1169 - 1175.

91 Hundertmark, M. & Hincha, D. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics (**2008**), **9**, 118.

92 Liu, Y. & Zheng, Y. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. Biochem Biophys Res Commun (**2005**), **311**, 325 - 332.

93 Shih, M., Lin, S., Hsieh, J., Tsou, C., Chow, T., Lin, T. & Hsing, Y. Gene cloning and characterization of a soybean (*Glycine max* L.) LEA protein, GmPM16. Plant Mol Biol (**2004**), **56**, 689 - 703.

94 Gal, T., Glazer, I. & Koltai, H. An LEA group 3 family member is involved in survival of *Caenorhabditis elegans* during exposure to stressh. FEBS Lett (**2004**), **577**, 21 - 26.

95 Goyal, K., Walton, L. & Tunnacliffe, A. LEA proteins prevent protein aggregation due to water stress. Biochem J (**2005**), **388**, 151 - 157.

96 Hand, S. C., Jones, D., Menze, M. A. & Witt, T. L. Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. J Exp Zool (**2007**), **307A**, 62 - 66.

97 Kikawada, T., Nakahara, Y., Kanamori, Y., Iwata, K., Watanabe, M., McGee, B., Tunnacliffe, A. & Okuda, T. Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. Biochem Biophys Res Comm (**2006**), **348**, 56 - 61.

98 Tunnacliffe, A., Lapinski, J. & McGee, B. A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. Hydrobiologia (**2005**), **546**, 315 - 321.

99 Bai, Y., Yang, Q., Kang, J., Sun, Y., Gruber, M. & Chao, Y. Isolation and functional characterization of a *Medicago sativa* L. gene, MsLEA3-1. Mol Biol Rep (**2012**), **39**, 2883-2892.

100 Sels, J., Mathys, J., De Coninck, B., Cammue, B. & De Bolle, M. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. Plant Physiol Biochem (**2008**), **46**, 941 - 950.

101 Guevara-Morato, M. A., Garcia de Lacoba, M., Garcia-Luque, I. & Serra, M. T. Characterization of a pathogenesis-related protein 4 (PR-4) induced in *Capsicum chinense* L3 plants with dual RNase and DNase activities. J Exp Bot (**2010**), **61**, 3259 - 3271.

102 Li, X., Xia, B., Jiang, Y., Wu, Q., Wang, C., He, L., Peng, F. & Wang, R. A new pathogenesis-related protein, LrPR4, from *Lycoris radiata*, and its antifungal activity against *Magnaporthe grisea*. Mol Biol Rep (**2010**), **37**, 995 - 1001.

103 Park, C. J., Kim, K. J., Shin, R., Park, J. M., Shin, Y. C. & Paek, K. H. Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. Plant J: Cell Mol Biol (**2004**), **37**, 186-198.

104 Donze, T., Qu, F., Twigg, P. & Morris, T. J. *Turnip crinkle virus* coat protein inhibits the basal immune response to virus invasion in Arabidopsis by binding to the NAC transcription factor TIP. Virology (**2014**), **449**, 207-214.

105 Nakashima, K., Tran, L. S. P., Van Nguyen, D., Fujita, M., Maruyama, K., Todaka, D., Ito, Y., Hayashi, N., Shinozaki, K. & Yamaguchi‐Shinozaki, K. Functional analysis of a NAC‐type transcription factor OsNAC6 involved in abiotic and biotic stress‐responsive gene expression in rice. The Plant J (**2007**), **51**, 617-630.

106 Voitsik, A. M., Muench, S., Deising, H. B. & Voll, L. M. Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. BMC Plant Biol (**2013**), **13**, 85.

107 Liu, B., Ouyang, Z., Zhang, Y., Li, X., Hong, Y., Huang, L., Liu, S., Zhang, H., Li, D. & Song, F. Tomato NAC transcription factor SlSRN1 positively regulates defense response against biotic stress but negatively regulates abiotic stress response. PLoS One (**2014**), **9**, e102067.

108 Kaneda, T., Taga, Y., Takai, R., Iwano, M., Matsui, H., Takayama, S., Isogai, A. & Che, F. S. The transcription factor OsNAC4 is a key positive regulator of plant hypersensitive cell death. The EMBO Journal (**2009**), **28**, 926-936.

109 Banerjee, A. & Roychoudhury, A. WRKY proteins: signaling and regulation of expression during abiotic stress responses. Scientific World J (**2015**), **2015**, 17.

110 Chen, L., Song, Y., Li, S., Zhang, L., Zou, C. & Yu, D. The role of WRKY transcription factors in plant abiotic stresses. Biochem Biophys Acta (**2012**), **1819**, 120-128.

111 Pandey, S. P. & Somssich, I. E. The role of WRKY transcription factors in plant immunity. Plant Physiol (**2009**), **150**, 1648-1655.

112 Eulgem, T. & Somssich, I. E. Networks of WRKY transcription factors in defense signaling. Curr Opin Plant Biol (**2007**), **10**, 366-371.

113 Tao, Z., Liu, H., Qiu, D., Zhou, Y., Li, X., Xu, C. & Wang, S. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. Plant Physiol (**2009**), **151**, 936-948.

114 Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C. & LeGall, O. The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus *Lettuce mosaic virus*. Plant Physiol (**2003**), **132**, 1272-1282.

Chapter 4

Quantitative expression of defence related genes in five cassava varieties challenged with UCBSV, using QuantStudio 12K Flex Real-Time PCR

T. Amuge, R. P. Eyles, M. E. Ferguson, A. A. Myburg and D. K. Berger

Author Contributions

T.A.: Conducted all field, screen-house (planted experimental plants, graft-inoculated and collected samples), laboratory procedures (identified inoculum, isolated RNA and prepared cDNA libraries), performed data analysis and wrote the initial draft of the manuscript.

Life Technologies (now Thermo Fisher Scientific): Designed primers and manufactured custom arrays for QuantStudio machine.

J. B.: Performed all wet-laboratory QuanstStudio analysis procedures from cDNA handling to data exportation (using protocols, and under the guidance and supervision of personnel from Life Technologies).

R.P.E.: Identified the method and guided identification of reference genes.

D.K.B.: Suggested appropriate data analysis procedures and extracted data from ExpressionSuite software (version 1.3) (Thermo Fisher Scientific (www.thermofisher.com)) into MSExcel. He led discussions on data interpretation, critically reviewed and extensively edited the chapter. He also provided University supervision.

M.E.F.: Principle Investigator of the funding project, acquired funding, identified varieties, provided day-to-day host institutional supervision, and edited the manuscript.

A.A.M.: Suggested the use of QuanstStudio technology for qPCR analysis, hosts The University of Pretoria's QuanstStudio wet laboratory platform (where all QuantStudio procedures were carried out). Coordinated collaboration between staff at Life Technologies, J. B., T. A., D.K.B and M.E.F. A.A.M. also provided critical comments, guidance, and University supervision.

D.K.B., A.A.M., **M.E.F** and **T.A.** designed the study.

NOTE: The contributions of J. B. (Jane Bredenkamp) and Life Technologies were a paid-service.

4.1 Abstract

The study was conducted as two experiments involving two cassava varieties (Experiment 1) and five cassava varieties (Experiment 2) sampled over 20 time points. For every variety, three plants were inoculated with UCBSV and three were mock-inoculated. At each time point, leaf tissues sampled from upper (UL) and lower (LL) positions (relative to the graft point) were assayed on QuantStudio 12K Flex Real-Time PCR system. However, poor quality of output data, possibly caused by low quality of cDNA or cDNA degradation during the qPCR workflow, resulted in selection of data from only few time points for differential gene expression analysis. For Experiment 1, three time points: three days after grafting (3 dag), 5 and 7 dag were selected while 1 and 51 dag were analysed in Experiment 2. At 12 months after grafting (mag) with UCBSV-inoculum, roots harvested from each sample were scored for CBSD-related root necrosis.

The following results were obtained: At 12 mag, roots of UCBSV-infected plants of Namikonga (from Experiments 1 and 2) and Kiroba (Experiment 2) were non-necrotic, one of the UCBSV-infected plants of variety Mkombozi had one necrotic spot in one of its storage roots. In NDL06/132 and Albert, UCBSV-infected plants of were clearly necrotic at harvest. Based on their roots scores, Namikonga and Kiroba were characterised as resistant to CBSD, Mkombozi as tolerant while NDL06/132 and Albert were characterized as susceptible.

QuantStudio results from Experiment 1 showed that in Namikonga, genes that encode LRR, PR, chaperone, HSP, LEA and TF (GATA, WRKY, NmrA, NAC) were all expressed at substantially high levels (between UCBSV-inoculated and mock-inoculated plants) at 3, 5 and 7 dag. In Albert, expression of the same genes was low at 3 and 5 dag, and higher at 7 dag.

In Experiment 2, one Leucine-rich repeat transmembrane protein kinase (QuantStudio code AIPAER8), among others, was expressed at substantially high levels (FC 670 at UL) in Kiroba at 1 dag, the same time point when two chaperones (QuantStudio codes AIX01ZS and AI0IYB, both of which encode Chaperone protein htpG family proteins) were also expressed at substantially high levels (FC 686 and FC 180) in Kiroba. However, the same LRR and chaperone genes were not differentially expressed other varieties (Namikonga, Mkombozi and Albert). Instead, that LRR gene was differentially expressed in LL of NDL06/132, albeit at a low level (FC 2.0).

Drawing from Experiments 1 and 2, its visible that Namikonga's and Kiroba's defence involve similar downstream defence genes (mainly chaperones and heat shock proteins (HSP), and transcription factors (TF)), which also function as 'helper' genes in NBS-LRR triggered defence. And, Namikonga's defence is expressed between 3 dag and 5 dag in UL while Kiroba's is expressed at 1 dag (UL and LL) and 51 dag (LL). Kiroba's ability to express defence genes at 1 and 51 dag suggests that its defence mechanism is triggered and sustained all through its interactive time with UCBSV. Between Namikonga and Kiroba, the

modes of resistance are similar with respect to ‘helper’ genes (chaperones and HSP) but different terms of specific NBS-LRR trigger (AI89LJS and AIBJYVU in Namikonga, AIPAER8 in Kiroba), and time points at which resistance is triggered (5 dag in Namikonga and 1 dag in Kiroba). Because both Experiments 1 and 2 were conducted on different time points, results drawn from this study remain inconclusive requiring confirmation analysis under similar time points.

4.2 Introduction

To date no cassava variety has been found to be immune to cassava brown streak disease (CBSD). Lack of immunity in cassava has prompted research and breeding work to rely on CBSD resistant and tolerant varieties. In cassava, resistance to CBSD is defined as the ability of a variety to maintain low virus titre after infection, and to have minimal root necrosis at maturity[1-3].

CBSD is the most critical threat to cassava production in East[4-6], Central[7] and southern Africa[8,9], where the disease has now spread[10]. In West Africa: the world's largest producer and consumer of cassava, the possible arrival of CBSD threatens food security. Cassava provides food and income to millions of people in Africa (FAOSTAT, 2014). CBSD is caused by two species of potyvirus: Cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV)[11,12]. Both virus species are naturally transmitted by whiteflies in a non-persistent manner[13-15]. In farmers' fields, CBSD is often caused by both species of the virus[11,12,16]. Leaf and root symptoms caused by single virus species are indistinguishable from symptoms caused by dual infection[17,18].

To date, studies on cassava's resistance to CBSD have focused mainly on the variety Namikonga, the first resistant cultivar to be identified[19]. Resistance to CBSD being the ability of a cassava variety to maintain a low virus load and show minimal shoot symptoms coupled with little or no root necrosis at harvest (Chapter 3, this thesis)[3,20]. Consequently, Namikonga has been used in breeding programs and research studies[19,21]. Sets of defence-associated genes and pathways that have direct bearing on Namikonga's resistance have been identified[1,3,22]. In addition, several other varieties, considered tolerant to CBSD have been identified, including Kiroba[2,19], Mkombozi[23] and NDL06/132[23] (Table 1). For breeding purposes, it is preferable to combine different mechanisms of resistance or tolerance, derived from different varieties. In fact Jennings, 1960[24] recommends that resistant or tolerant varieties should be inter-crossed to combine different sources in order to enhance both levels and durability of the resistance.

To determine whether or not different varieties showing resistance or tolerance to CBSD use different mechanisms, we ask the research question: "to what extent do patterns of pathogen defence genes up regulated in 'Namikonga' challenged with UCBSV correlate with responses in other CBSD tolerant cultivars namely Kiroba, Mkombozi and NDL06/132?" The hypothesis for this study is that the expression of these pathways (associated with resistance to CBSD in Namikonga), differ between Namikonga and other CBSD-tolerant cassava varieties (Kiroba, Mkombozi and NDL06/132). The objective of the study is to determine and compare the patterns of expression of selected pathogen-defence-genes associated with resistance to CBSD in Namikonga, in other CBSD-tolerant cassava cultivars (NDL06/132, Mkombozi and Kiroba). The pathogen defence genes include: leucine rich repeats (LRR), NB-ARC, late embryogenesis abundant proteins (LEA), pathogenesis related proteins (PR), transcription factors (WRKY, NAC, NmrA and GRAS), heat shock proteins (HSP) and chaperons. The study was established as two experiments. In Experiment 1, Namikonga's resistance is characterized alongside susceptible Albert first using RNAseq (Chapter 3)[1],

and then using QuantStudio 12K Flex Real-Time PCR System (this chapter). In Experiment 2, response to UCBSV infection in five varieties; Albert, Kiroba, Mkombozi, Namikonga and NDL06/132, was investigated using gene expression on a QuantStudio 12K Flex Real-Time PCR System and compared to the resistance profile displayed by Namikonga.

4.3 Materials and Methods

Two experiments were set to examine the expression of 55 defence-associated genes in cassava sampled at upper (UL) and lower (LL) leaf positions over 20 time points. In Experiment 1, two cassava varieties: Albert (CBSD susceptible) and Namikonga (CBSD resistant) were assayed using RNASeq data generated as described in Chapter 3. The RNAseq study was performed on upper leaves (UL) sampled at six hours after grafting (6 hag), one day after grafting (1 dag), 2, 5, 8, 45 and 54 dag. Plants used for RNAseq study were sampled in 2011 and their RNA extracted in May 2012. Afterwards, QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific, Massachusetts, USA), hereafter referred to as QuantStudio was used to analysed UL and LL sampled over 20 time points (described under "*Establishment of planting materials for Experiments 1 and 2*" below. The QuantStudio, technology performs real-time PCR in chip format, allowing for analysis of more samples per run and minimising variations caused by running fewer samples on many plates as done in traditional qPCR machines.

Experiment 2 involved analysis of five cassava varieties including Albert, Kiroba, Mkombozi, Namikonga and NDL06/132 sampled from UL and LL positions at 20 time points. This experiment (Experiment 2) was also established in a glasshouse, with the same design as for experiment 1, except that experiment 2 had five varieties. Time points are described in the section "*Establishment of planting materials for Experiments 1 and 2*" below. Kiroba, Mkombozi and NDL06/132 (also the source of UCBSV inoculum) were obtained from field trials at Sugarcane Research Institute (SRI), Tanzania while the sources of Namikonga and Albert are described in Chapter 3. Plants of Namikonga and Albert grown in Experiment 2 were different from those used in for RNAseq analysis (Chapter 3). Experiment 2 samples were obtained in May 2013 and RNA extracted in May 2014. All five cassava varieties have different responses to CBSD (Table 1).

Table 4.1: Origin of CBSD-resistant and CBSD susceptible cassava varieties used in Experiments 1 and 2

Variety	Response to CBSD	Origin	Reference
Namikonga	Resistant	Tanzania	Kanju et al. 2007[19]; Kaweesi et al., 2014[2]; Ferguson et al., 2015[23]
Kiroba	Tolerant	Tanzania	Kanju et al. 2007[19]; Kaweesi et al., 2014[2]
NDL06/132	Tolerant	Tanzania	Ferguson et al., 2015[23]
Mkombozi	Tolerant	Tanzania/ IITA-Nigeria	Ferguson et al., 2015[23]
Albert	Susceptible	Tanzania	Kulembeka et al., 2012[21]; Ferguson et al., 2015[23]

4.3.1 Establishment of planting materials for Experiments 1 and 2

Planting materials for both experiments 1 and 2 were established using the same protocol. A minimum of nine stem cuttings (approx. size) was taken from 6-month-old plants each of the five cassava varieties. Cuttings were planted in 33.23 kg of forest soil including 3.85 kg of gravel to overlay the perforated surfaces

of buckets, perforations serving as solute outlets, and watered with three litres of tap water. Planted cuttings were watered with two litres of tap water twice per-week. This watering regime was adhered to throughout the experimental period.

At two months after planting, a set of three plants representing biological replicates, were grafted with UCBSV-inoculated and another set of three plants mock-inoculated scions from pre-established plants of UCBSV-positive and UCBSV-negative NDL06/132. The top-wedge method of grafting was used [25]. Newly grafted plants were fastened with microfilm. To maintain humidity around the graft, plants were covered with transparent polythene bags for four days, with daily observation of humidity build-up. At each sampling, tissue for RNA extraction was obtained from a single leaf lobe of each of the three plants per treatment, as described in Chapter 3[1].

Sampling was done at 20 time points including: one hour before grafting (0 hours), six hours after grafting (6 hag), 1, 2, 3, 4, 5, 6, 7, 8, 45, 46, 47, 48, 49, 50, 51, 52, 53 and 54 dag (days after grafting). The first samples (0 hours) were taken from three biological replicates to function as a baseline (representing the state of plants per variety before plants are injured through grafting). At each sampling for both mock- and UCBSV-inoculated plants, leaf tissues were obtained from two positions of each plant: near graft point, located just below the plant tip (UL), and the leaf positioned nearest to the soil level (LL). The purpose for sampling at two positions is to know whether gene expression varies between young, tender leaves near the actively growing tip (UL) and senescing leaves (LL) at the base of the plant. All sampled leaf lobes were placed in pre-labelled aluminium foil sample bags and immediately frozen in liquid nitrogen and stored at -80°C prior to RNA extraction. At maturity (12 months after grafting), roots mock- and UCBSV-inoculated were evaluated for CBSD-related necrosis. Wet-laboratory analysis of sampled leaves and roots were performed differently as described in the following section.

4.3.2 RNA extraction, synthesis and validation of cDNA

In Experiment 1, frozen leaf tissues sampled from UL at 0 hours, 6 hag, 1, 2, 5, 8, 45 and 54 dag were ground to a fine powder in liquid nitrogen. Designated for analysis using RNAseq technology, the RNA extraction and cDNA synthesis from these samples were performed as described in chapter 3. Samples designated for analysis on QuantStudio platform were assayed following the same protocol used for Experiment 2.

In Experiment 2, frozen leaf tissues sampled from UL and LL designated for QuantStudio study were ground to a fine powder in liquid nitrogen. RNA was extracted using a modified CTAB method [26] in which, the amount of starting extraction buffer was increased from 600ul to 800ul. Complementary DNA (cDNA) was synthesized from 4 µl of total RNA using superscript III (Invitrogen, CA, USA). To test the quality of cDNA, a random set of six cDNA samples (1 µl) from each time point was used to amplify the housekeeping gene, Rubisco (Forward: 5'-CTTTCCAAGGCCCGCCTCA -3', Reverse: 5'-CATCATCTTTGGTAAAATCAAGTCCA-3')[27,28] in a 25 µl PCR reaction volume. The PCR parameters were as follows: initial denaturation at 94°C for 5 mins, 35 cycles of 94°C for 30 sec, annealing at 52°C and

extension at 72°C for one minute. A final extension step was performed at 72°C for 10 mins, after which the run was held at 10°C or stopped. PCR products were resolved on 2% agarose gel with an expected product of 171 bp[3]. Samples with poor quality cDNA (identified by faint or no PCR product from Rubisco primers) were re-synthesized from the respective RNA (data not shown).

4.3.3 Selection of 55 defence and one candidate reference gene used in the study

Fifty-five defence related genes were selected for this study, based on a combination of data from differential expression, from RNASeq data (Chapter 3), and evidence from the literature of genes involved in plant potyvirus defence networks. Plants resist Potyvirus by expression of known defence genes including leucine rich repeats (LRR)[29], NB-ARC[30], pathogenesis related (PR) proteins[31,32], transcription factors[33-36], late embryogenesis abundant proteins (LEA)[37], chaperones[38,39] and heat shock proteins[40,41] among others. From these listed categories of plant defence genes, genes that were up regulated in Namikonga and either marginally up regulated or down regulated in Albert at 2 and 5 dag were selected[1] for evaluation with QuantStudio qPCR (Table 4.2).

Gene AI20US1 (the 56th gene evaluated on QuantStudio), which encodes a CLP protease regulatory subunit X, had been pre-selected as a reference gene from among five reference genes reported to have stable expression in cassava infected with CBSV[42]. Based on RNAseq data, gene AI20US1 had stable Log2FC values in both Albert and Namikonga, while the other four genes (identified by Moreno et al., 2011[42]) showed varied expression patterns in Albert and Namikonga in all the seven RNAseq-assayed time points. Unfortunately, this RNAseq stable gene (Gene AI20US1), generated very low or undetermined Ct values from QuantStudio analyses, and so, was not used as a reference gene. A different approach, described in the section "*Identification of reference genes for the QuantStudio data analysis*" was used to identify reference genes from within the experiments.

Table 4.2: RNAseq expression ratios (from Chapter 3) for fifty-five defence and one candidate reference gene, which formed the basis for QuantStudio qPCR study

				Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam
Quantstudio code for gene	**Gene ID (Prefix "cassava4.1_", suffix "m.g")**	**Gene annotation based on VirtualPlant[43]**	**Time points where genes are substantially up regulated in Namikonga (why genes where chosen for validation) [N]**	**6 hag**		**1 dag**		**2 dag**		**5 dag**		**8 dag**		**45 dag**		**54 dag**	
				Log2 FC	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**
AIRSA4O	001048	Leucine-rich repeat transmembrane protein kinase	2 dag	0.3	0.1	-0.4	-0.4	0.1	1.9	0.1	-1.3	-0.3	0.2	2.6	-0.5	0.6	-2.3
AIBJYVU	034154	Leucine-rich repeat (LRR) family protein	2, 5 and 54 dag	-0.9	0	0.6	-0.2	0.6	1	0.2	1.8	-0.1	-0.1	-0.5	-0.2	-0.8	0.7
AI89LJS	007501	Leucine-rich repeat (LRR) family protein	1, 2 and 5 dag	0.2	0.4	-1.1	1.2	0.2	3.8	0.1	2.6	0.2	-0.1	2.2	0	-0.1	0.2
AIQJCYG	000978	Leucine-rich receptor-like protein kinase family protein	6 hag and 5 dag	-0.8	1	0	-0.1	-0.1	0.3	-0.7	1.5	0.3	0.4	0.5	-0.2	-0.2	0.4
AIPAER8	000765	Leucine-rich repeat transmembrane protein kinase	6 hag, 1, 2, 5 and 8 dag	-0.7	0.7	-0.2	1	-0.8	0.8	0.2	2.1	0	1.1	-0.5	-0.7	-1.2	-0.1
AI5IQ05	029764	NB-ARC domain-containing disease resistance protein	5 dag	-0.1	-0.1	-0.6	-0.4	0.2	0.6	-0.5	1	0	0.2	0.3	0.2	0	-0.6
AI0IYB9	022732	NB-ARC domain-containing disease resistance protein	5 dag	-0.4	0.3	-0.3	-0.2	0.1	0.1	0.1	1	0	0.4	0	-0.1	0.1	-0.8
AIY9Z51	022172	NB-ARC domain-containing disease resistance protein	5 dag	-0.2	-0.1	-0.2	-0.1	0.3	0.7	0.7	1.1	0.1	0.1	-0.3	0	-0.2	-0.7
AIVI5NC	001752	NB-ARC domain-containing disease resistance protein	5 dag	-0.3	0.1	-0.6	-0.1	0.2	0.6	-1	1.1	0	0.1	0.5	-0.4	-0.1	-1

				Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam
Quantstudio code for gene	**Gene ID (Prefix "cassava4.1_", suffix "m.g")**	**Gene annotation based on VirtualPlant[43]**	**Time points where genes are substantially up regulated in Namikonga (why genes where chosen for validation)** [N]	**6 hag**		**1 dag**		**2 dag**		**5 dag**		**8 dag**		**45 dag**		**54 dag**	
				Log2 FC	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**
AICSW12	034172	NB-ARC domain-containing disease resistance protein	5 dag	-0.3	0.4	-0.5	-0.4	0.5	0.2	0.6	1.4	0.2	0.2	0.2	0	0	-1
AII1N3E	000058	NB-ARC domain-containing disease resistance protein	5 dag	-0.3	0.2	-0.4	-0.2	0.3	0.5	-0.1	1.1	0	-0.2	0.8	0	-0.1	-0.7
AIKAL3D	012383	Pathogenesis-related thaumatin superfamily protein	6 hag, 2, 5 and 54 dag	-0.3	1.1	1.4	0.2	0.4	2.5	-3	3.5	0.6	0.6	1.7	0.4	-0.5	1.5
AII1NW5	011960	Pathogenesis-related thaumatin superfamily protein	6 hag, 5, 8 and 54 dag	-1.9	1.1	-1	0.5	-0.5	-2.2	0.3	1.4	0.1	2.3	-0.3	-0.8	-0.7	0.7
AI1RWIH	025676	Late embryogenesis abundant protein, group 2	1, 2 and 5 dag	-0.1	0.4	0.2	1	-0.4	1.2	-0.3	1.4	-0.1	-0.1	-0.2	-0.3	-0.4	-0.1
AIWR3TL	019959	Late Embryogenesis Abundant 4-5	2, 5 and 8 dag	-0.1	-1	0.4	0	0.5	2.3	-0.1	2.3	-0.6	5.4	-0.5	0.3	1.1	-0.4
AIMSIFT	014614	WRKY DNA-binding protein 40	6 hag, 5 and 8 dag	-3.5	3	-2.4	-0.6	-1.6	-2.2	-1.1	4	0	1.7	-0.2	-3.5	-1.6	-3.3
AIGJRKP	011680	WRKY DNA-binding protein 70	6 hag, 5 and 8 dag	-0.9	1.3	-1.2	-0.1	-0.3	-2.2	-2.1	2.1	0.5	2.2	0	-1.9	-0.6	-1.8
AIAA0PM	033681	Heat shock protein 90.1	1, 2 and 5 dag	-0.9	0.6	1.1	1.9	1	4.7	0.9	2	-0.5	-0.6	-1.3	-0.2	-0.1	-0.6
AID1U8A	034243	Heat-shock protein 70T-2	2 dag	-0.3	-0.3	0.8	1.1	1.1	3.9	-0.4	0	-0.8	0	-1.2	0.3	0.4	0.3

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based on VirtualPlant[43]	Time points where genes are substantially up regulated in Namikonga (why genes where chosen for validation) [N]	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam
				6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
				Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC
AI5IQ04	003240	Heat shock protein 70B	6 hag, 1, 2, 8 and 45 dag	-0.1	1.5	0.9	2.2	1.6	3.6	0.6	-1.3	-1.2	1.1	0	1.5	0.7	0.9
AIN1GL1	014648	Heat shock protein 21	6 hag, 1, 2 and 8 dag	0.1	2	0.5	2	1.4	7.6	3.5	1.6	-2.7	1	-0.9	-0.6	2.8	-3.8
AID1U79	009750	Heat shock transcription factor A2	1, 2 and 5 dag	-1.1	-0.4	0.3	1.3	-0.6	6.1	-0.3	7.3	0	-0.5	-1.2	-1.4	0.9	0
AIFATEH	010803	Heat shock transcription factor A6B	2 and 5 dag	1.9	0.7	1.8	2	2.2	5.4	-0.1	3.4	0.7	-2.8	-0.3	-0.1	0.9	0
AI39SUW	003144	Heat shock protein 70 (HSP 70) family protein	6 hag and 5 dag	-0.8	1	0.1	-0.2	-0.2	0.8	-1	2.1	0	-0.6	-0.1	-0.7	-0.3	-0.3
AIS09AW	001300	Heat shock protein 101	1, 2 and 54 dag	-0.5	0.4	1	2	0.2	1.5	0.3	-0.2	0.1	0.8	-0.8	0.5	-0.6	1.4
AIAA0PL	007779	DNAJ heat shock N-terminal domain-containing protein	6 hag, 1 and 5 dag	-0.7	0.7	0.8	1.6	0.3	0.8	0	1.2	0.2	-0.8	-0.5	0.2	-0.2	0.1
AICSW11	008997	HSP70-interacting protein 1	2 dag	-0.2	0.4	0.3	0.9	-0.3	1.8	0.1	0.8	-0.1	0	-0.3	-0.4	0.1	0.5
AI1RWIG	001924	Heat shock protein 89.1	1, 2 and 5 dag	-0.3	0.3	0.4	2.1	-0.2	1.3	0.3	2.4	-0.3	-0.1	-0.9	-0.7	0	-0.2
AI20UOO	002706	Chloroplast heat shock protein 70-2	5 dag	-0.1	0	0.4	0.8	0.1	0.1	-0.4	0.8	0.1	-0.5	0	-0.2	-0.2	-0.1
AIT97G4	001607	Heat shock protein 70 (HSP 70) family protein	1, 2 and 5 dag	-0.6	0.2	0.8	2.1	0.1	0.9	0.1	1.2	0.1	-0.5	-0.2	-0.2	-0.4	0.2
AI6RO7C	003340	Heat shock protein 70	6 hag, 1 and 5 dag	-0.9	0.9	0.8	1.7	0.6	0.9	0	1	-0.2	0.1	-0.6	-0.1	-0.6	-0.7

				Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam
Quantstudio code for gene	**Gene ID (Prefix "cassava4.1_", suffix "m.g")**	**Gene annotation based on VirtualPlant[43]**	**Time points where genes are substantially up regulated in Namikonga (why genes where chosen for validation)** [N]	**6 hag**		**1 dag**		**2 dag**		**5 dag**		**8 dag**		**45 dag**		**54 dag**	
				Log2 FC	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**	**Log2 FC**
AIS09AX	018127	HSP20-like chaperones superfamily protein	6 hag, 1, 2 and 5 dag	0.3	1	0.4	1.8	0.4	4.1	0	1.3	-0.4	-0.4	-0.4	-0.5	0.6	-0.9
AIX01ZT	020888	Trigger factor type chaperone family protein	1 and 2 dag	0.1	-0.2	0.3	2	0.3	2.4	0.1	0.8	-0.1	0.1	0.2	0.3	0	0
AIBJYVT	008620	Trigger factor type chaperone family protein	1 and 2 dag	0.3	-0.2	0.3	1.7	0.2	2	0.1	0.8	0.2	0.1	0.2	0.1	0	-0.2
AI0IYB8	001921	Chaperone protein htpG family protein	1, 2 and 5 dag	0.2	0.1	0.4	2	-0.3	1.4	-0.1	1.8	-0.1	-0.6	-0.2	-0.2	0.1	-0.5
AIRSA4P	018031	Chaperone DNAJ-domain superfamily protein	1, 2 and 5 dag	-0.2	0.2	0.4	2.1	0.5	3.2	-0.7	2.9	0	0.4	0.8	0.6	-0.1	0.4
AIT97G5	018200	HSP20-like chaperones superfamily protein	1, 2 ad 8 dag	0.4	0	1	2	2.1	5.7	1.5	1.5	-0.9	0.9	-1.1	-0.4	0.5	-1.7
AI20UOP	026342	HSP20-like chaperones superfamily protein	6 hag, 1 and 2 dag	2.2	2.7	0.8	1.6	4.5	8.6	-0.6	-1.1	-1.5	0.5	-10	-10	1.9	-1.2
AIWR3TK	001827	Chaperone protein htpG family protein	6 hag, 2 and 5 dag	-1	1.4	0	0.2	-0.6	2	-0.5	2.2	-0.1	-0.9	-0.2	-1	-0.3	-0.2
AIPAER9	014653	Chaperone DNAJ-domain superfamily protein	1, 2 and 5 dag	0.7	0.8	0.8	1.6	0.6	3.9	0.4	2.3	0.3	-0.4	0.8	0.3	0.1	-1.7

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based on VirtualPlant[43]	Time points where genes are substantially up regulated in Namikonga (why genes where chosen for validation) [N]	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam
				6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
				Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC
AIX01ZS	001905	Chaperone protein htpG family protein	1, 2 and 5 dag	0.4	0.4	0.7	2	-0.1	1.9	0.6	1	-0.2	0.4	0	0	0.2	-0.6
AI70NDK	004263	Chaperonin-60alpha	1, 2 and 5 dag	0.1	0.2	0.5	1.7	-0.1	1.7	-0.2	1.2	-0.1	-0.6	0.1	-0.4	-0.1	-0.2
AILJJ9L	014410	Chaperonin 20	1, 2 and 5 dag	0	0	0.3	1.3	-0.3	1.3	-0.5	1.2	-0.1	-0.1	-0.1	-0.3	0	-0.1
AIVI5ND	018353	Chaperone DNAJ-domain superfamily protein	6 hag, 1, 2, 5 and 45 dag	-2.1	1.7	0.6	1.3	0.6	2.9	0	1.8	0.2	-2.1	1	0.4	-0.1	1.8
AI89LJT	001921	Chaperone protein htpG family protein	1, 2 and 5 dag	0.2	0.1	0.4	2.0	-0.3	1.4	-0.1	1.8	-0.1	-0.6	-0.2	-0.2	0.1	-0.5
AI39SUX	028589	NmrA-like negative transcriptional regulator family protein	2, 5 and 54 dag	-2.1	-0.4	-1.4	-0.2	0.5	3.2	-1.5	3.5	0.2	0.1	4.5	0.5	-1.1	4.7
AI6RO7D	032524	NmrA-like negative transcriptional regulator family protein	1, 2, 5 and 54 dag	-4	-0.7	-1.6	1.2	-0.3	2.7	-3.6	2.2	0.7	0.4	3.5	0.6	0.1	2.8
AI70NDL	032811	GRAS family transcription factor	2 dag	0.4	-0.1	0.1	0.2	0.6	3.2	-0.1	0.8	0.2	0	-0.2	0.5	0.6	-0.3
AIQJCYH	016750	GATA type zinc finger transcription factor family protein	1 and 2 dag	0.2	-0.2	0.2	1.3	0.4	2	0.3	0.7	0.1	-0.7	0.2	-0.2	-0.1	-0.1
AIHSPQX	011886	GATA transcription factor 5	1, 2 and 5 dag	0	0.3	-0.5	0.3	0.6	2.8	-0.9	1.7	0.2	-0.8	-0.1	0.4	0.2	0.4
AIY9Z50	032811	GRAS family transcription factor	2 and 5 dag	0.4	-0.1	0.1	0.2	0.6	3.2	-0.1	0.8	0.2	0.0	-0.2	0.5	0.6	-0.3

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based on VirtualPlant[43]	Time points where genes are substantially up regulated in Namikonga (why genes where chosen for validation) [N]	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam	Alb	Nam
				6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
				Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC	Log2 FC
AI39SY9	011029	NAC domain containing protein 96	2 and 5 dag	-0.7	-1.5	-0.1	0.2	0.4	1	-2.4	5.1	-0.5	0.1	-0.4	-2	-0.7	0.6
AI5IQ5H	015961	NAC transcription factor-like 9	1, 2, 5 and 54 dag	0.6	0.8	-1.4	1.3	0.4	2.9	0.4	1	-3	-0.6	-2	-5.8	-3.2	0.8
AI70NHX	023870	NAC transcription factor-like 9	1, 2, 5 and 54 dag	0.3	0.4	0.2	1	0.6	2.2	-1.9	1.2	-1	-1.1	-0.5	-2.7	-0.4	1.3
AI6RPBP	016601	Eukaryotic initiation factor 4E protein	6 hag, 1 and 5 dag	-0.5	0.5	0	-0.3	0.1	0.4	-0.5	0.4	-0.4	-0.8	-0.7	-0.2	-0.1	0.3
AI20US1	004423	CLP protease regulatory subunit X	NB: pre-selected candidate reference gene	0	0.6	0.5	0.4	0.1	0	0	0	0.2	0.2	0.1	0	-0.1	0

[N] = Time points where genes were substantially up regulated in Namikonga but had low expression (or were down regulated) in Albert. Yellow cells = Log2FC (of RNAseq gene expression ratio) is up regulated in Namikonga and down regulated in Albert. Hag: hours after grafting. Dag: days after grafting. Nam: Namikonga. Alb: Albert. Log2FC: Log2FoldChange of count data (for respective genes) from the RNAseq study (Chapter 3) mapped to cassava reference genome version4.1[44].

4.3.4 Primer design and QuantStudio-based gene expression

All cDNA samples were shipped in 96-well PCR plates on dry ice via express courier to University of Pretoria, Department of Genetics laboratory for QuantStudio 12K Flex Real-Time PCR analysis. All primers were custom designed by Thermo Fisher Scientific and used to synthesize CustomTaqMan® Assays by Life technologies (now Thermo Fisher Scientific) (www.thermofisher.com). Target sequences of all genes for primer design were obtained from protein coding gene loci of the cassava reference genome v4.1[44] available in Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html). The cassava reference genome (version 4.1) is derived from a partial inbred line, AM560-2, a third generation self of the Latin American variety MCOL-1505[44]. All genes from version 4.1 are named with cassava4.1_" and suffix "m.g", and numbers specifying each gene, for example "cassava4.1_011886m.g"). Primers designed from respective genes were assigned code names, as shown in Table 4.2. The QuantStudio qPCR was performed as a laboratory service at the Forest Molecular Genetics (FMG) Laboratory, Department of Genetics, University of Pretoria, and included three technical replicates as well as the three biological replicates.

4.3.5 Data processing from QuantStudio

Raw data (Ct values) from QuantStudio were received as *.eds* files and were processed using ExpressionSuite software (version 1.3) (Thermo Fisher Scientific (www.thermofisher.com)); a freely available program designed for analysis of QuantStudio qPCR data. All data were exported to Microsoft Excel (2011) and R software[45] for statistical analysis.

Filtering QuantStudio data to identify time points and genes with the least "undetermined" Ct values

Using data for both Experiments 1 and 2, all qPCR data was first filtered by 'determined' and 'undetermined' Ct values. Samples with "undetermined' Ct values were marked NA (not available), and automatically dropped from further analysis. Next, all 'determined' Ct values were tested for normal distribution. Outlier samples (within 1st percentile (Ct value <16.7) and 99th percentile (Ct value >33.8)) were discarded, retaining only 'determined' Ct values that fall within 2nd and 98th percentile. After this initial filtering, data was now analysed by Experiment.

For Experiment 1, pre-filtered data (pre-filtered by determined/undetermined and 2nd – 98th percentile above) from the two assayed varieties, Albert and Namikonga were assembled by time point. For each time point, 2,688 data points were expected (56 genes X 48 samples), where 48 samples represent two varieties under two treatments represented by three biological replicates sampled from two plant positions (UL and LL). Data for each biological replicate is a mean of three technical replicates (data for technical replicates are not described). Total number of 'determined' Ct values after pre-filtering was counted per time point, and percentage calculated relative to 2,688. Five time points (3, 5, 7, 53 and 54 dag) with the highest numbers (and %) of 'determined' Ct values (Table S4.1) were selected. Within these five time points, 27 genes with determined Ct values at all the five time points in both varieties, Albert and Namikonga, were selected, as they fulfilled the requirement for normalization analysis. The other 29 genes had undetermined

Ct values in one or more of these time points, in either or both varieties, making them unusable for normalization analyses.

For Experiment 2, pre-filtered data (pre-filtered by determined/undetermined and 2^{nd} – 98^{th} percentile above) from the five assayed varieties (Albert, Kiroba, Mkombozi, Namikonga and NDL06/132) were assembled by time point. For each time point, 3,360 data points were expected (56 genes X 60 samples), where 60 samples represent five varieties under two treatments represented by three biological replicates sampled from two plant positions (UL and LL). Data for each biological replicate is a mean of three technical replicates (data for technical replicates are not described). Total number of 'determined' Ct values after pre-filtering were counted per time point, and percentage calculated relative to 3,360. Three time points (1, 46 and 51 dag) with the highest numbers (and %) of 'determined' Ct values (Table S4.2) were selected. Within these five time points, 25 genes with determined Ct values at all the three time points and five varieties (Albert, Kiroba, Mkombozi, Namikonga and NDL06/132), a requirement for normalization analysis, were selected. The other 31 genes had undetermined Ct values in one or more of these time points, within one or more varieties, hence, these 31 genes could not be used for normalization.

4.3.6 Identification of reference genes for the QuantStudio data analysis

NormFinder[46] was used to identify reference genes from 27 genes (Experiment 1) and 25 genes (Experiment 2). This software uses a model-based approach to determine intra- and inter-group variation in expression of candidate reference genes. The groups, in this case, are different treatments under which expression of the said genes have been measured. The NormFinder method is widely used to determine suitable reference genes from qPCR producing results consistent with other software designed for the same function such as BestKeeper and geNorm [47-50].

4.3.7 Data analysis

Each biological replicate had three technical replicates. The Ct value for each biological replicate was obtained from the geometric mean of its technical replicates. Standard deviations of technical replicates were calculated using MS Excel. Thereafter, the Ct value of each treatment was calculated from a geometric mean of three biological replicates. Treatments with only one biological replicate were dropped, as standard deviation cannot be calculated using one sample. The mean Ct values from biological replicates was used to determine fold change using $2^{-\Delta\Delta Ct}$ method [51]. Using this method, Ct1 ($Ct_{\text{gene X (inoculated)}} - Ct_{\text{reference gene (inoculated)}})_{\text{Time T}}$) – Ct 2($Ct_{\text{gene X (mock)}} - Ct_{\text{reference gene (mock)}})_{\text{Time T}}$) is generated as delta delta Ct ($^{\Delta\Delta}Ct$). The delta delta Ct is then used to compute fold change, where fold change = $2^{-\Delta\Delta Ct}$. Gene X refers to the respective gene within the list while time T refers to the time point for which fold change is being generated. From fold change values, log2foldchnge was calculated using the LOG formula in MS Excel. To determine significance of difference in means of mock and inoculated treatments, a students' t-test was performed in MS Excel. The paired sample t-test based on two-tailed hypothesis with 95% confidence level, was used to test for differences in mean Ct of mock and inoculated treatments per gene. As for standard deviation and fold change calculations, tests were performed using at least two replicates of each mock or inoculated treatment. Therefore, genes with only one no biological replicate in either mock

or inoculated treatments automatically returned nonsensical outputs (#DIV/0!) and are labelled NA in the result tables (see results section). Based on the above criteria, the resultant numbers of genes available for determining fold change between UCBSV-inoculated and mock-inoculated samples, and log2foldchange varied by experiment, variety, and sampling position.

4.4 Results

4.4.1 Resistance/susceptibility to CBSD based on presence/absence and extent of root necrosis

Experiment 1: Results based on two varieties (Chapter 3)[1]

Phenotypic responses from leaves and roots of Albert and Namikonga in Experiment 1 have been described in Chapter 3[1]. In summary, storage roots from UCBSV-inoculated Albert had typical CBSD-related necrosis but storage roots from mock-inoculated plants of Albert were non-necrotic. In Namikonga, storages roots harvested from both UCBSV- and mock-inoculated plants were non-necrotic.

Experiment 2: Results based on five varieties

In Experiment 2, there was no observable necrosis in UCBSV-inoculated and mock-inoculated plants of either Namikonga or Kiroba (Fig. 4.1). Mkombozi had one necrotic spot on roots from UCBSV-inoculated plants but the mock-inoculated plants had no observable necrosis (Fig. 4.1).

Roots harvested from UCBSV-inoculated plants of NDL06/132 were small, with several necrotic spots but the mock-inoculated plant was non-necrotic (Fig. 4.1). A similar pattern was observed in Albert except that it had the highest number of necrotic spots in roots of UCBSV-inoculated plants. Besides necrosis, root constriction was also observed on roots from UCBSV-inoculated plants of Albert. Roots harvested from UCBSV-inoculated plants of Albert and NDL06/132 were generally smaller in size, and those of NDL06/132 had a characteristic yellowish coloration.

Fig. 4.1: Tolerance and susceptibility to UCBSV-related root necrosis in five cassava varieties. Black arrows indicate necrotic spots. (A): Namikonga-UCBSV-inoculated; (B) Namikonga-Mock-inoculated; (C) Kiroba-UCBSV-inoculated; (D) Kiroba-Mock-inoculated; (E) Mkombozi-UCBSV-inoculated; (F) Mkombozi-Mock-inoculated; (G) NDL06/132-UCBSV-inoculated; (H) NDL06/132-Mock-inoculated; (J) Albert-UCBSV-inoculated and (K) Albert-Mock-inoculated

4.4.2 Experiments 1 and 2: Filtering of data sets to identify representative time points and differentially expressed genes

Identification of time points for data analysis

In both experiments, samples were taken during an early time phase (within hours and days after challenging plants with UCBSV) and a late time phase (several days prior to symptom emergence). The early time phase included 0 hours through 8 dag and the late time phase included 45 through 54 dag. However, due to possible degradation of cDNA in transit to the University of Pretoria (where all QuantStudio procedures were performed), much of the data was undetermined (had no Ct values) at qPCR level. A report from Life technologies (now Thermo Fisher Scientific), who designed primers, manufactured arrays and analysed samples on the QuantStudio machine) showed that all assays (except AII1N3E) performed well at manufacturing and were properly loaded on to the machine (Fig. S4.1A), without leakage (Fig. S4.1B). However, the post-qPCR florescence image showed that several wells had no amplification (Fig. S4.1C), suggesting a possible flaw in the quality of cDNA. This major failure of amplification at many data points led to the decision to only use specific genes from within specific time points which had sufficient data to represent respective time phases in both experiments. Further within these representative time points, genes with outlier Ct values (at 1st and 99th percentile of the normally distributed data) were removed from downstream analyses.

Experiment 1

For Experiment 1, which included varieties Albert and Namikonga, five time points were chosen (3, 5, 7, 53 and 54 dag) (Table 4.3) based on data for 56 (i.e 55 defence and one presumed reference) genes (Table S4.1). These time points had up to 81% determined Ct values spanning the 56 genes (Table S4.1). Out of the 56 genes (55 defence-related and one pre-selected house-keeping gene, which was later dropped at filtering), 27 had determined Ct values spanning both mock and inoculated treatments across all five-time points in Albert and Namikonga (Table S4.3). Therefore, these 27 genes at five time points were used for normalization analysis (i.e identification of reference genes). However, only three of the time points, (3, 5 and 7 dag), which had at least 70% determined Ct values (Table 4.3), were chosen for analysis of differentially expressed genes (DEG).

Table 4.3: Experiment 1 analysis of 55 defence-related and one ordinary gene in Namikonga and Albert at five time points with the highest numbers of determined Ct values after filtering.

	Time point				
	3 dag	**5 dag**	**7 dag**	**53 dag**	**54 dag**
Determined [T]	2186	2150	1885	1257	1373
%	81.3	80	70.1	46.8	51.1

[T] = Number of QuantStudio reactions that were scored as “determined” (row 1) and their percentage (row 2) at respective time points shown. Determined Ct values are between Ct 16.7 and 33.8. dag: Days after grafting

Experiment 2

For Experiment 2, the varieties Albert, Kiroba, Mkombozi, Namikonga and NDL06/132 were analysed (Table S.4.2). In the experiment, three time points (1, 46 and 51 dag) (Table 4.4) were chosen for NormFinder analysis to identify the most stable reference genes based on Ct values for 56 genes. These three time points had up to 96% determined Ct values spanning the 56 genes (Table 4.5, S4.2). Twenty-five of these genes had more determined Ct values across the three time points, than other genes, and were chosen for identification of reference genes (Table 4.5, S4.4). For identification of differentially expressed genes, the best two time points (1 and 51 dag) were chosen to identify DEG representing 'early' (represented by 1 dag, which had 96.3% determined Ct values) and 'late' (represented by 51 dag, which had 91.2% determined Ct values) time phases (Table 4.4).

Table 4.4: Experiment 2 analysis of 55 defence-related and one ordinary gene in Namikonga, Albert, Kiroba, NDL06/132 and Mkombozi at three time points with the highest numbers of determined Ct values after filtering.

	Time point		
	1 dag	**46 dag**	**51 dag**
Determined [T]	3235	2903	3063
%	96.3	86.4	91.2

[T] = Number of QuantStudio reactions that were scored as "determined" (row 1) and their percentage (row 2) at respective time points shown. Determined Ct values are between Ct 16.7 and 33.8. Dag: days after grafting.

Table 4.5: Selected candidate reference genes for Experiment 1 (27 genes) and Experiment 2 (25 genes) used for NormFinder[46] analysis.

Quantstudio code for gene	**Gene ID** [*]	**Gene annotation based on VirtualPlant**	**Experiment 1**[#]	**Experiment 2**[#]
AIRSA4O	001048	Leucine-rich repeat transmembrane protein kinase	NS	Y
AIBJYVU	034154	Leucine-rich repeat (LRR) family protein	Y	Y
AI89LJS	007501	Leucine-rich repeat (LRR) family protein	Y	NS
AIQJCYG	000978	Leucine-rich receptor-like protein kinase family protein	Y	Y
AIPAER8	000765	Leucine-rich repeat transmembrane protein kinase	NS	Y

Quantstudio code for gene	Gene ID *	Gene annotation based on VirtualPlant	Experiment 1#	Experiment 2#
AI5IQ05	029764	NB-ARC domain-containing disease resistance protein	Y	NS
AICSW12	034172	NB-ARC domain-containing disease resistance protein	Y	NS
AIWR3TL	019959	Late Embryogenesis Abundant 4-5	Y	NS
AIMSIFT	014614	WRKY DNA-binding protein 40	NS	Y
AIGJRKP	011680	WRKY DNA-binding protein 70	Y	Y
AIAA0PM	033681	Heat shock protein 90.1	NS	Y
AID1U8A	034243	Heat-shock protein 70T-2	NS	Y
AIN1GL1	014648	Heat shock protein 21	NS	Y
AI39SUW	003144	Heat shock protein 70 (HSP 70) family protein	NS	Y
AICSW11	008997	HSP70-interacting protein 1	Y	Y
AI1RWIG	001924	Heat shock protein 89.1	NS	Y
AI20UOO	002706	Chloroplast heat shock protein 70-2	Y	NS
AIT97G4	001607	Heat shock protein 70 (HSP 70) family protein	Y	Y
AI6RO7C	003340	Heat shock protein 70	Y	Y
AIS09AX	018127	HSP20-like chaperones superfamily protein	Y	NS
AIX01ZT	020888	Trigger factor type chaperone family protein	Y	Y
AIBJYVT	008620	Trigger factor type chaperone family protein	Y	NS
AI0IYB8	001921	Chaperone protein htpG family protein	Y	NS
AIRSA4P	018031	Chaperone DNAJ-domain superfamily protein	Y	Y
AIT97G5	018200	HSP20-like chaperones superfamily protein	NS	Y
AIWR3TK	001827	Chaperone protein htpG family protein	Y	NS

Quantstudio code for gene	Gene ID *	Gene annotation based on VirtualPlant	Experiment 1#	Experiment 2#
AIPAER9	014653	Chaperone DNAJ-domain superfamily protein	Y	Y
AIX01ZS	001905	Chaperone protein htpG family protein	Y	NS
AI70NDK	004263	Chaperonin-60alpha	Y	Y
AILJJ9L	014410	Chaperonin 20	Y	Y
AIVI5ND	018353	Chaperone DNAJ-domain superfamily protein	Y	NS
AI89LJT	001921	Chaperone protein htpG family protein	NS	Y
AI70NDL	032811	GRAS family transcription factor	Y	Y
AIQJCYH	016750	GATA type zinc finger transcription factor family protein	Y	NS
AIHSPQX	011886	GATA transcription factor 5	Y	Y
AIY9Z50	032811	GRAS family transcription factor	Y	NS
AI39SY9	011029	NAC domain containing protein 96	NS	Y
AI70NHX	023870	NAC transcription factor-like 9	NS	Y
AI6RPBP	016601	Eukaryotic initiation factor 4E protein	Y	NS

* (Prefix “cassava4.1_”, suffix “m.g”), # Y = candidate reference gene, NS = Data not used. Each gene marked Y was selected as a candidate reference because it had determined Ct values across all time points used for NormFinder analysis (i.e 3, 5, 7, 53 and 54 dag for Experiment 1; and 1, 46 and 51 dag for Experiment 2). Genes marked NS were not used for NormFinder analysis as they did not have determined Ct values at all time points chosen for reference gene identification.

4.4.3 Identification of reference genes for the QuantStudio data analysis

The 27 (for Experiment 1) and 25 (Experiment 2) genes selected for identification of reference genes using NormFinder are listed in Table 4.5 (above). By design, NormFinder stability values range between 0 and 1[46], and on this basis, appropriate reference genes were chosen for use on Experiments 1 and 2. Lower the stability values indicates that that , gene is more stable[46]. Besides providing the most stable gene, the NormFinder algorithm also provides the best combination of two genes, for studies where more than one reference gene is needed. In such a case, the user decides whether to use the most stable gene or the best combination of two genes as reference(s) for their study. In our case with Experiment 1, gene AIT97G4

(Heat shock protein 70 (HSP 70) family protein) was the most stable (NormFinder stability value of 0.107) out of 27 candidate reference genes (Table S4.5). The best combination of two genes included gene AICSW11 (HSP70-interacting protein 1) and gene AIY9Z50 (Chaperone protein htpG family protein), with a combined NormFinder stability of 0.080. To confirm the suitability of the most stable or best combination of genes for Experiment 1, a regression analysis was performed. The regression analysis showed that one of two best combination genes, gene AICSW11, had a positive slope in Albert and a negative slope in Namikonga, while the second gene (AIY9Z50) had negative slopes in both varieties (Table 4.6). It was not possible to establish the probable error(s) that could have been introduced, if the two genes with positive-negative and negative-negative slopes were combined to normalize both varieties. Therefore, the single most stable gene, AIT97G4, was used as a reference for Experiment 1. This gene, AIT97G4, identified as the best among 27 genes, had comparable Ct values in both Albert and Namikonga. At 5 dag, Ct value for this gene was similar in both varieties. A similar trend was observed at 7, 53 and 54 dag where the difference in Ct values was <1 in Albert and Namikonga. A regression analysis showed that this gene had a slope of 0.2323 in Albert and -0.1855 in Namikonga with R^2 values of 0.199 and 0.096 in Albert and Namikonga respectfully (Fig. 4.2, Table 4.6).

Similar observations have been made in other studies and genes with stability[52] values of 0.07-0.65 represented the most stable genes[47-50]. In two varieties of rice subjected to low temperature stress, the most stable reference genes had 0.199 – 0.3 under different conditions[47]. With human cell lines, the most stable gene had stability values of 0.10-0.23[48]. In another study of cell lines and tissues, the most stable genes had stability values of 0.13-0.65[50]. A study on apple shoots identified the most stable gene as having 0.077 stability value on NormFinder[49]. All these studies show a range of values that compare favourably with our study.

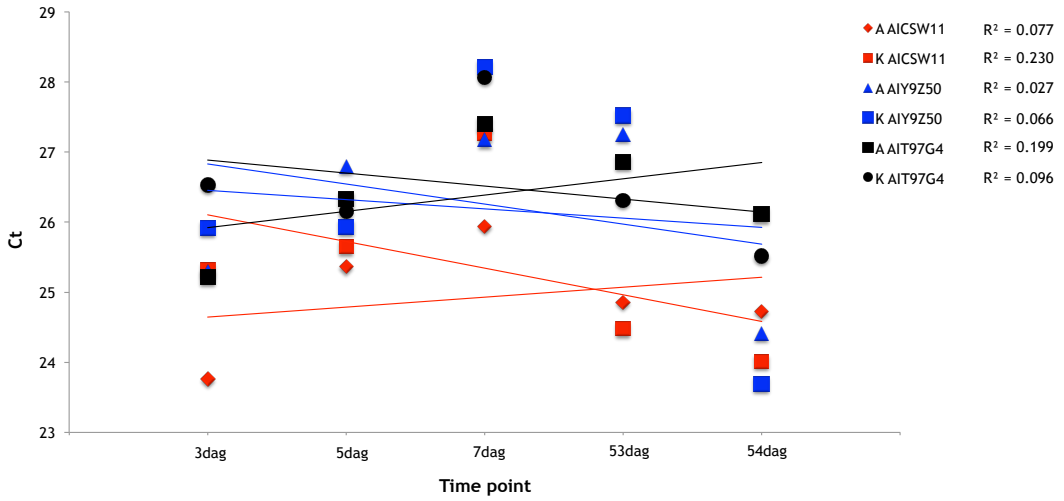


Fig. 4.2: Regression analysis of the Ct values of AIT97G4 (Heat shock protein 70 (HSP 70) family protein) reference gene over a time course in Albert (A) and Namikonga (K)

Table 4.6: Equation of the straight line showing slope and y-intercept for the most stable gene and the best combination of two genes in Albert and Namikonga in Experiment 1.

	AIT97G4	AICSW11	AIY9Z50
Albert	y = 0.2323x + 25.689	y = 0.142x + 24.503	y = -0.1321x + 26.585
Namikonga	y = -0.1855x + 27.071	y = -0.3803x + 26.484	y = -0.286x + 27.115

Gene **AIT97G4** was chosen for normalization.

For Experiment 2, the gene with the lowest NormFinder stability value (0.014) was AICSW11 (Late embryogenesis abundant protein, group 2) out of 25 candidate reference genes (Table S4.3). Genes AI70NDK (Chaperonin-60alpha) and AIBJYVU (Leucine-rich repeat (LRR) family protein) were identified as the best combination of two genes, with a combined stability value of 0.009. All three genes: the most stable (AICSW11) and the best combination of two genes, (AI70NDK and AIBJYVU) had varied slopes in all five varieties, varying from negative to positive in different varieties. However, the best combination of two genes was chosen as references because their Ct values followed a similar pattern at all four representative time points (1, 8, 46 and 51 dag) in all five varieties yet the most stable gene (AICSW11) had skewed Ct values at 8 dag and 46 dag (Fig. 4.3, Table 4.7) and so, was dropped. The geometrical mean of Ct values of AI70NDK and AIBJYVU was used as a reference value for normalizing data for Experiment 2.

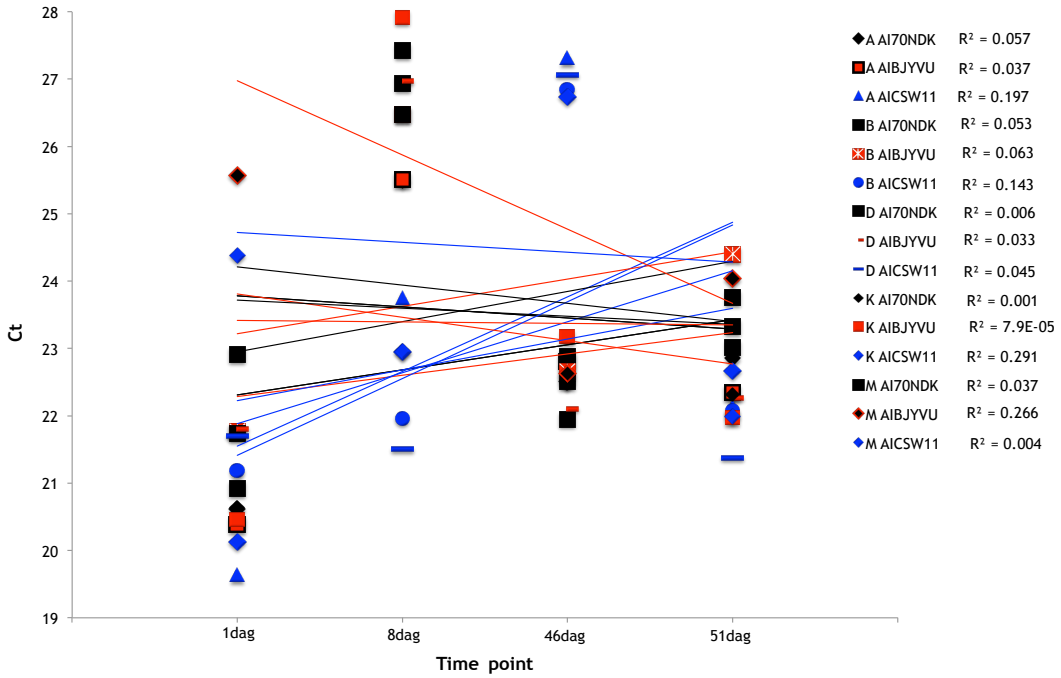


Fig. 4.3: Regression analysis of the Ct of AIBJYVU (Leucine-rich repeat (LRR) family protein) and AI70NDK (Chaperonin-60alpha) reference genes over a time course in Albert (A), Kiroba (B), NDL06/132 (D), Mkombozi (M) and Namikonga (K).

Table 4.7: Equation of the straight-line showing slope and y-intercept for the most stable gene and the best combination of two genes in Albert and Namikonga in Experiment 2

	AIBJYVU	**AI70NDK**	AICSW11
Albert	y = 0.3167x + 21.969	y = 0.3725x + 21.937	y = 1.1081x + 20.442
Namikonga	y = -0.0222x + 23.439	y = -0.1198x + 23.838	y = 1.1402x + 20.273
Mkombozi	y = -1.1021x + 28.078	y = -0.2706x + 24.482	y = -0.1468x + 24.869
Kiroba	y = 0.4058x + 22.813	y = 0.4482x + 22.502	y = 0.7568x + 21.126
NDL02/136	y = -0.3466x + 24.156	y = -0.1644x + 23.946	y = 0.4572x + 21.767

Genes **AIBJYVU** and **AI70NDK** were chosen for normalization.

4.4.4 Experiments 1 and 2: Identification of genes and treatments with fold change data from representative time points

Results in Tables S4.4 (Experiment 1) S4.5 (Experiment 2) shows details of genes per treatment that were retained (✔) and those that were dropped (✖) because they had insufficient Ct data necessary for calculating fold change between UCBSV- and mock-inoculated treatments respectively. To be able to calculate fold change between UCBSV- and mock-inoculated treatments, each gene needed to have Ct values from both UCBSV- and mock-inoculated treatments. This criterion was not met in all genes as some genes had Ct values for only UCBSV-inoculated or only mock-inoculated left after filtering. Therefore, only genes with fold change data were presented in the subsequent results sections (including selection of reference genes).

Experiment 1: Expression of genes at representative time points sampled from upper (UL) and lower (LL) leaf positions

Here, only three (3, 5 and 7 dag) out of five time points (Table 4.3) were used for analysis of differential gene expression. In Namikonga, eleven genes were differentially expressed (FC 2.1 to 8.22) in upper leaves (UL) at 3 and 5 dag (Table 4.8). At 3 dag, all 11 DEG (FC 2.15 to 8.22) had significant difference between mock- and UCBSV-inoculated samples of Namikonga. At 5 dag, seven of the 11 DEG (FC 3.7 to 11.43) had significant difference between mock- and UCBSV-inoculated samples while at 7 dag; only one of the eleven genes was differentially expressed. However, none of the 11 genes were significantly different between mock- and UCBSV-inoculated samples at 7 dag.

Table 4.8: Up regulated genes in Namikonga's upper leaves (UL) at 3, 5 and 7 dag of Experiment 1

Gene ID	Gene Annotation	3 dag			5 dag			7 dag		
		Fold change	Log2FC	P-value	Fold change	Log2FC	P-value	Fold change	Log2FC	P-value
AI39SUW	HSP	7.3	2.9	**0.001**	2.8	1.5	0.175	0.4	-1.3	0.241
AI70NDK	Chaperone	6.9	2.8	**0.001**	3.9	2.0	**0.045**	1.8	0.9	0.232
AIX01ZS	Chaperone	4.7	2.2	**0.001**	4.9	2.3	**0.012**	**2.1**	1.1	0.264
AIWR3TK	Chaperone	8.2	3.0	**0.004**	11.4	3.5	**0.004**	1.1	0.1	0.781
AI0IYB8	Chaperone	2.2	1.2	**0.004**	7.0	2.8	**0.003**	2.0	1.0	0.199
AIX01ZT	Chaperone	2.6	1.4	**0.011**	6.3	2.7	**0.005**	0.6	-0.8	0.452
AIRSA4P	Chaperone	2.1	1.1	**0.013**	5.0	2.3	**0.009**	0.5	-1.0	0.336
AI1RWIH	LEA	3.6	1.8	**0.022**	5.0	2.3	0.159	1.0	0.0	NA
AILJJ9L	Chaperone	3.0	1.6	**0.026**	3.7	1.9	**0.016**	1.6	0.7	0.299
AI6RPBP	TF, NAC	4.1	2.0	**0.027**	3.3	1.7	0.201	1.7	0.7	0.37
AI70NHX	TF, NAC	2.2	1.1	**0.048**	3.3	1.7	0.413	0.7	-0.6	0.651

Significant p-values (≤ 0.05) shown in **bold text** are based on a two-tailed paired-sampled t-test. NA represents genes that had less than two biological replicates for either mock or inoculated, hence t-test resulted in error value. Genes with FC ≥ 2 are considered differentially expressed.

When the 11 genes up regulated in Namikonga at 3 and 5 dag were observed in Albert, the following results were obtained. In Albert's UL sampled at 3, 5 and 7 dag (Table 4.9), six of the same 11 genes were differentially expressed between mock- and UCBSV-inoculated treatments (FC 2.04 to 4.72) at 3 dag. However, only eight of the 11 DEG were significantly different between mock- and UCBSV-inoculated samples at 3 dag. For 5 dag, seven of the 11 genes were differentially expressed in Albert's UL. Eight of the 11 genes had significant difference in expression between mock- and UCBSV-inoculated samples at 5 dag. At 7 dag, nine of the 11 genes were differentially expressed with the highest recorded fold change in Albert (FC 2.83 to 9.02), while seven of the 11 genes were significantly different between mock- and UCBSV-inoculated samples.

Table 4.9: Comparison of Namikonga's up regulated genes (identified at 3, 5 and 7 dag) in Albert's upper leaves (UL) of Experiment 1

Gene ID	Gene Annotation	3dag			5dag			7dag		
		Fold change	Log2FC	P-value	Fold change	Log2FC	P-value	Fold change	Log2FC	P-value
AI39SUW	HSP	1.0	0.0	**0.052**	2.6	1.4	**0.042**	2.8	1.5	0.088
AI70NDK	Chaperone	3.9	2.0	**0.012**	2.5	1.3	**0.004**	6.5	2.7	**0.002**
AIX01ZS	Chaperone	3.2	1.7	**0.023**	2.1	1.0	**0.021**	1.1	0.1	NA
AIWR3TK	Chaperone	2.7	1.4	**0.020**	2.7	1.4	0.094	9.0	3.2	NA
AI0IYB8	Chaperone	4.7	2.2	**0.010**	1.7	0.8	**0.011**	3.9	2.0	**0.014**
AIX01ZT	Chaperone	2.2	1.1	**0.028**	1.7	0.7	**0.034**	3.4	1.8	**0.057**
AIRSA4P	Chaperone	0.8	-0.4	0.212	2.4	1.3	**0.009**	3.8	1.9	**0.030**
AI1RWIH	LEA	0.9	-0.2	0.082	2.4	1.3	**0.047**	3.5	1.8	**0.014**
AILJJ9L	Chaperone	2.0	1.0	**0.035**	2.0	1.0	**0.007**	4.6	2.2	**0.002**
AI6RPBP	TF, NAC	1.1	0.1	0.077	1.2	0.3	0.358	5.7	2.5	**0.025**
AI70NHX	TF, NAC	1.5	0.5	**0.024**	1.6	0.7	0.299	1.2	0.3	NA

Significant p-values (≤ 0.05) shown in **bold text** are based on a two-tailed paired-sampled t-test. NA represents genes that had less than two biological replicates for either mock or treated treatments, hence t-test resulted in error value. Genes with FC ≥ 2 are considered differentially expressed.

From lower leaves (LL), Namikonga had only one DEG, AIRSA4O (encodes LRR) present and differentially expressed at 3 and 5 dag. At 7 dag, there was no DEG. In Albert, two genes: AIWR3TL (encodes LEA) and AIBJYVU (encodes LRR) were differentially expressed at 3 and 7 dag. No gene was differentially expressed in Albert at 5 dag.

Experiment 1: Genes unique to Albert-alone and Namikonga-alone at 3 dag, 5 dag and 7 dag

Experiment 1: Upper Leaves

At each of the three representative time points, (3, 5 and 7 dag) several genes were differentially expressed in Namikonga-alone and Albert-alone (Table 4.10). At 3 dag, seven genes (four HSP, one NBD, one chaperone and one TF, NmrA) were differentially expressed in Namikonga. Three of the seven DEG of Namikonga were significantly different between mock and inoculated samples (p-values = 0.002-0.009). However, Albert-alone had six DEG unique at 3 dag. These six genes include: three chaperones, one HSP, one WRKY and one TF, NmrA. Four of Albert's six DEG had significantly different expression between mock- and UCBSV-inoculated samples (p-value = 0.004-0.038).

At 5 dag, Namikonga-alone had 16 DEG (three HSP, three chaperones, four TF (two GATA, two NmrA and one NAC), two LRR, two WRKY and one PR) genes. Nine of these genes had significant p-value (p-value = 4.9E-04-0.052) between mock and inoculated treatments. In Albert however, only six genes were differentially expressed and unique at 5 dag. These include two chaperones, two HSP and two TF, GRAS. Five of the six genes had significant p-value between mock- and UCBSV-inoculated treatments (p-value = 0.002-0.009).

At 7 dag, Namikonga-alone registered six DEG including four TF (two GATA and two NmrA), one HSP and one chaperone. Only two of these genes were significantly different between mock and inoculated treatments (p-value = 0.023 and 0.046). In Albert-alone, nine genes were differentially expressed at 7 dag. The said genes include three HSP, three chaperones, and two TF (GATA and GRAS). Eight of the nine genes had significantly different expression between mock- and UCBSV-inoculated treatments (p-value = 0.001-0.031).

Table 4.10: DEG identified in Namikonga-alone and Albert-alone at 3, 5 and 7 dag in upper leaves of Experiment 1

3 dag

Namikonga					Albert				
Gene ID	Gene Annotation	Fold change	Log2FC	P-value	Gene ID	Gene Annotation	Fold change	Log2FC	P-value
AI6RO7C	HSP	2.3	1.2	**0.002**	AIGJRKP	WRKY	4.2	2.1	**0.004**
AI20UOO	HSP	3.7	1.9	**0.009**	AIBJYVT	Chaperone	2.0	1.0	**0.023**
AIS09AX	Chaperone	3.6	1.8	**0.009**	AIY9Z50	Chaperone	4.4	2.1	**0.037**
AI39SUX	TF, NmrA	15.5	4.0	0.068	AIVI5ND	Chaperone	2.6	1.4	**0.038**
AICSW12	NBD	4.8	2.3	NA	AI39SUX	TF, NmrA	9.9	3.3	0.077
AID1U79	HSP	3.7	1.9	NA	AIN1GL1	HSP	5.8	2.5	0.300
AIN1GL1	HSP	3.5	1.8	NA					

5 dag

Namikonga					Albert				
Gene ID	Gene Annotation	Fold change	Log2FC	P-value	Gene ID	Gene Annotation	Fold change	Log2FC	P-value
AI1RWIG	HSP	12.8	3.7	**4.9E-04**	AIY9Z50	Chaperone	3.4	1.8	**0.002**
AIQJCYH	TF, GATA	6.4	2.7	**0.002**	AI20UOO	HSP	2.1	1.1	**0.002**
AIGJRKP	WRKY	9.9	3.3	**0.002**	AI89LJT	TF, GRAS	3.6	1.8	**0.002**
AIMSIFT	WRKY	33.2	5.1	**0.004**	AICSW11	HSP	2.7	1.4	**0.005**
AIFATEH	HSP	15.7	4.0	**0.019**	AIBJYVT	Chaperone	3.6	1.8	**0.009**
AI89LJS	LRR	8.6	3.1	**0.023**	AI70NDL	TF, GRAS	2.3	1.2	0.080
AIKAL3D	PR	8.3	3.1	**0.031**					
AIBJYVU	LRR	4.4	2.1	**0.050**					
AIPAER9	Chaperone	3.9	2.0	**0.052**					

AI6RO7D	TF, NmrA	14.3	3.8	0.061
AIY9Z50	Chaperone	3.6	1.8	0.081
AIBJYVT	Chaperone	4.3	2.1	0.095
AID1U79	HSP	4.6	2.2	0.144
AIHSPQX	TF	2.2	1.1	0.403
AI39SUX	TF, NmrA	3.2	1.7	0.523
AI5IQ5H	TF, NAC	5.7	2.5	NA

7 dag

Namikonga					**Albert**				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
AIQJCYH	TF, GATA	6.0	2.6	**0.023**	AI6RO7C	HSP	2.3	1.2	**0.001**
AI6RO7D	TF, NmrA	25.0	4.6	**0.046**	AIS09AX	Chaperone	14.3	3.8	**0.003**
AI39SUX	TF, NmrA	5.6	2.5	0.133	AI20UOO	HSP	3.7	1.9	**0.006**
AIAA0PL	HSP	2.2	1.1	0.200	AI70NDL	TF, GRAS	3.2	1.7	**0.008**
AIHSPQX	TF, GATA	2.1	1.1	0.409	AIVI5ND	Chaperone	17.4	4.1	**0.009**
AIVI5ND	Chaperone	4.0	2.0	NA	AI5IQ05	NBD	2.3	1.2	**0.012**
					AICSW11	HSP	3.3	1.7	**0.015**
					AIQJCYH	TF, GATA	40.4	5.3	**0.031**
					AIPAER9	Chaperone	4.0	2.0	0.070
					AI39SUX	TF, NmrA	35.0	5.1	0.084
					AIKAL3D	PR	5.0	2.3	0.183
					AI89LJS	LRR	7.7	2.9	NA
					AIAA0PL	HSP	2.5	1.3	NA
					AID1U79	HSP	2.1	1.1	NA

Significant p-values (≤ 0.05) shown in **bold text** are based on a two-tailed paired-sampled t-test. NA represents genes that had less than two biological replicates for either mock and inoculated treatments; hence, their t-test resulted in error value. Genes with FC ≥ 2 are considered differentially expressed.

Experiment 1: Lower Leaves

From lower leaves of Experiment 1, several genes were differentially expressed and unique to both Namikonga and Albert at 3, 5 and 7 dag (Table 4.11). At 3 dag, Namikonga-alone had three genes that were differentially expressed, but none had p-value information (meaning, either mock- or UCBSV-inoculated treatments had only one biological replicate, which is insufficient to perform a t-test). On the hand, Albert-alone had five DEG at 3 dag. Two of these genes encode LRR proteins, and one each encoded LEA, TF (NAC) and chaperone proteins. None of the five genes had significantly different expression between mock and inoculated treatments.

At 5 dag, Namikonga-alone had seven DEG including one gene each of PR, chaperone, LEA, HSP, WRKY, LRR and TF (NAC) proteins. None of the seven genes were significantly different between mock- and UCBSV-inoculated samples. Albert-alone had no DEG at 5 dag.

At 7 dag, Namikonga-alone had no DEG. However, Albert-alone had seven DEG. Of the seven genes, two encode LRR proteins, the others, each encode LEA, HSP, PR, NBD and WRKY proteins.

Table 4.11: DEG identified in Namikonga-alone and Albert-alone at 3, 5 and 7 dag in lower leaves of Experiment 1

3 dag

Namikonga					Albert				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
AIBJYVU	LRR	2.3	1.2	NA	AIWR3TL	LEA	15.6	4.0	0.048
AIRSA4O	LRR	2.2	1.1	NA	AI20UOP	Chaperone	4.1	2.0	0.797
AI20UOP	Chaperone	2.0	1.0	NA	AI5IQ5H	TF, NAC	3.1	1.7	NA
					AIRSA4O	LRR	2.7	1.4	0.733
					AIBJYVU	LRR	2.7	1.4	0.595

5 dag

Namikonga					Albert				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
AII1NW5	PR	8.0	3.0	NA					
AIWR3TL	LEA	4.3	2.1	0.984					
AIT97G5	Chaperone	3.3	1.7	0.797					
AI6RO7C	HSP	3.1	1.6	0.578					
AIMSIFT	WRKY	3.0	1.6	0.732					
AIRSA4O	LRR	2.6	1.4	0.391					
AI70NHX	TF, NAC	2.3	1.2	NA					

7 dag

Namikonga					Albert				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
					AIWR3TL	LEA	8.1	3.0	0.659
					AIAA0PM	HSP	4.2	2.1	NA
					AII1NW5	PR	4.2	2.1	0.481
					AIBJYVU	LRR	3.3	1.7	0.090
					AI0IYB9	NBD	3.2	1.7	NA
					AIMSIFT	WRKY	2.6	1.4	NA
					AI89LJS	LRR	2.1	1.0	NA

Significant p-values (≤ 0.05) shown in **bold text** are based on a two-tailed paired-sampled t-test. NA represents genes that had less than two biological replicates for either mock and inoculated treatments; hence t-test resulted in error value. Genes with FC ≥ 2 are considered differentially expressed.

Expression of genes at representative time points of Experiment 2 sampled from upper (UL) and lower (LL) leaf positions

Experiment 2: Upper Leaves

In this case, only two (1 and 51 dag) out of three time points (Table 4.4) were used for analysis of differential gene expression. Namikonga had three DEG at 1 dag; none at 51 dag while Kiroba had seven and eight DEG at 1 dag (FC of 2-670) and 51 dag (FC 2 in all eight genes) (Table 4.12). None of the genes was common to either time points. Despite the very high fold change (up to 670) registered in upper leaves of Kiroba at 1 dag, none was statistically significant. Mkombozi had five and seven DEG at 1 and 51 dag respectively. Gene AIT97G4 (HSP) and AI6RO7D (TF, NmrA) were the only common genes at both time points in Mkombozi. None of the genes was statistically significant. Variety NDL06/132 had three DEG at 51 dag and zero DEG at 1 dag. Albert had no DEG at both 1 and 51 dag.

Table 4.12: DEG unique to one of 1 or 51 dag, and DEG found at both 1 and 51 dag in upper leaves of five varieties (Namikonga, Albert, Kiroba, Mkombozi and NDL06/132) in Experiment 2

1 dag

Kiroba					**Namikonga**					**Mkombozi**				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
AI5IQ04	HSP	10.0	3.3	0.061	AI20UOO	HSP	4.3	2.1	0.095	**AIT97G4**	**HSP**	**6.6**	**2.7**	**0.219**
AIPAER8	LRR	670.4	9.4	0.089	AIT97G4	HSP	5.7	2.5	0.165	AIGJRKP	WRKY	10.2	3.4	0.412
AICSW11	HSP	3.6	1.8	0.192	AI20US1	CLP	3.8	1.9	0.218	**AI6RO7D**	**TF, NmrA**	**3.2**	**1.7**	**0.559**
AIGJRKP	WRKY	12.4	3.6	0.424						AIQJCYH	TF, GATA	2.6	1.4	0.619
AIVI5ND	Chaperone	5.4	2.4	0.429						AICSW11	HSP	2.0	1.0	0.770
AID1U8A	HSP	2.3	1.2	0.510										
AIT97G4	HSP	2.7	1.4	NA										

51 dag

Kiroba					**NDL06/132**					**Mkombozi**				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**

AID1U79	HSP	2.4	1.2	0.652	AIY9Z50	Chaperone	6.3	2.6	0.369	AI6RPBP	TF, NAC	4.1	2.0	0.134
AI89LJT	TF, GRAS	2.4	1.2	0.661	AIBJYVT	Chaperone	4.1	2.0	0.439	AI20UOO	HSP	3.4	1.8	0.229
AI20UOP	Chaperone	2.2	1.1	0.722	AIAA0PL	HSP	3.0	1.6	NA	AI20US1	CLP	3.4	1.8	0.333
AIN1GL1	HSP	2.3	1.2	0.727						**AIT97G4**	**HSP**	**3.4**	**1.8**	**0.347**
AI6RO7C	HSP	2.2	1.2	0.747						AIX01ZT	Chaperone	2.2	1.2	0.918
AIWR3TL	LEA	2.1	1.1	0.776						**AI6RO7D**	**TF, NmrA**	**2.2**	**1.1**	**0.961**
AI39SY9	TF, NAC	2.3	1.2	0.780						AIWR3TK	Chaperone	2.1	1.1	0.994
AICSW12	NBD	2.1	1.1	0.796										

Significant p-values (≤ 0.05) shown in **bold text** are based on a two-tailed paired-sampled t-test. Genes found at both 1 and 51 dag of the same variety are in **bold text**. NA represents genes that had less than two biological replicates for either mock and inoculated treatments; hence t-test resulted in error value. Genes with FC ≥ 2 are considered differentially expressed.

Experiment 2: Lower Leaves

Here, Kiroba had 13 DEG at 1 dag and none at 51 dag. It was in lower leaves of Kiroba that the highest fold change was recorded (FC 686 in gene AIX01ZS, which encodes chaperone) (Table 4.13). Namikonga had three DEG at 1 dag and zero at 51 dag while Mkombozi had 16 DEG at 1 dag and 5 DEG at 51 dag. Gene AIX01ZT (Chaperone) was the only one common to both time points in Mkombozi. Variety NDL06/132 had six DEG at 51 dag and none at 1 dag. Albert had no DEG at both 1 and 51 dag.

Table 4.13: DEGs that were unique to any one time point and those which are found at both 1 and 51 dag (**bold text**) in Namikonga, Albert, Kiroba, Mkombozi and NDL06/132 at 1 and 51 dag in lower leaves of Experiment 2

1 dag

Kiroba					**Namikonga**					**Mkombozi**				
Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
AI5IQ05	NBD	2.0	1.0	**0.007**	AIVI5ND	Chaperone	4.4	2.1	0.302	AID1U8A	HSP	9.3	3.2	**0.001**
AIX01ZS	Chaperone	685.6	9.4	**0.013**	AIS09AW	HSP	2.2	1.1	0.425	AI5IQ04	HSP	5.6	2.5	**0.029**
AI0IYB8	Chaperone	179.8	7.5	**0.038**	AIT97G4	HSP	2.4	1.3	0.634	AIMSIFT	WRKY	2.3	1.2	**0.052**
AIX01ZT	Chaperone	4.6	2.2	0.077						AIS09AX	Chaperone	7.4	2.9	0.107
AI6RO7D	TF, NmrA	10.5	3.4	0.111						**AIX01ZT**	**Chaperone**	**3.4**	**1.8**	**0.138**
AIS09AX	Chaperone	3.8	1.9	0.247						AIBJYVU	LRR	3.3	1.7	0.185
AIRSA4O	LRR	5.4	2.4	0.252						AI5IQ05	NBD	3.8	1.9	0.205
AILJJ9L	Chaperone	2.6	1.4	0.275						AIHSPQX	TF, GATA	6.9	2.8	0.226
AIT97G4	HSP	13.8	3.8	0.317						AIGJRKP	WRKY	2.2	1.2	0.256
AIKAL3D	PR	18.3	4.2	0.321						AICSW11	HSP	3.5	1.8	0.289

AIWR3TL	LEA	66.6	6.1	0.322	AI89LJS	LRR	11.3	3.5	0.351
AIVI5ND	Chaperone	12.5	3.6	0.429	AIQJCYH	TF, GATA	2.2	1.1	0.356
AIS09AW	HSP	2.1	1.1	NA	AI1RWIG	HSP	8.9	3.2	NA
					AILJJ9L	Chaperone	8.2	3.0	NA
					AIAA0PL	HSP	6.5	2.7	NA
					AIT97G4	HSP	4.7	2.2	NA

51 dag

Kiroba	**NDL06/132**					**Mkombozi**				
	Gene ID	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**	**Gene ID**	**Gene Annotation**	**Fold change**	**Log2FC**	**P-value**
	AID1U79	HSP	3.7	1.9	0.128	AI6RO7D	TF, NmrA	2.6	1.4	0.242
	AI39SUW	HSP	3.3	1.7	0.468	AI70NDK	Chaperone	2.1	1.0	0.372
	AIN1GL1	HSP	2.7	1.4	0.532	AIY9Z51	NBD	4.7	2.2	NA
	AI0IYB9	NBD	3.2	1.7	0.672	AIBJYVT	Chaperone	2.7	1.4	NA
	AI70NDL	TF, GRAS	2.1	1.1	NA	**AIX01ZT**	**Chaperone**	**2.2**	**1.1**	**NA**
	AIPAER8	LRR	2.0	1.0	NA					

Significant p-values (≤ 0.05) shown in **bold text** are based on a two-tailed paired-sampled t-test. Genes found at both 1 and 51 dag of the same variety are in **bold text**. NA represents genes that had less than two biological replicates for either mock and inoculated treatments; hence t-test resulted in error value. Genes with FC ≥ 2 are considered differentially expressed.

4.5 Discussion

4.5.1 Namikonga's defence is time bound and more active in upper than lower leaves

Substantial differential expression was recorded in 11 genes in Namikonga's upper leaves at 3 dag (FC 2.2 to 8.2) and 5 dag (FC 2.8 to 11.4), while only subtle change was registered at 7 dag (FC 0.4 to 2.1) (Table 4.7). The eleven genes encode seven chaperons, one HSP, one LEA and two TF (NAC), A t-test showed that all eleven DEG were statistically significant at 3 dag, with very high significance levels in some chaperone and HSP (p-value = 0.001-0.004). At 5 dag, higher fold change was observed in the same genes, with significant p-values in some and others, non-significant. By 7 dag, all eleven genes were non-significant, with below average expression (FC <2), except for gene AIX01ZS, a chaperon, which had fold change of 2.1 at 7 dag. This observation underscores first, the significance of timing in Namikonga's resistance to UCBSV. A critical level of expression is set by 3 dag, reaches a peak at 5 dag and is almost shutdown by 7 dag. Similar observations were made in the RNAseq study of Albert and Namikonga[1]. In Albert, the same genes were differentially expressed, albeit with much higher p-values (0.012-0.052) and lower, often below average (i.e. <2) fold change (FC = 1.0-4.7) at the same time point (Table 8).

Although chaperones and HSP are capable of launching defence mechanisms against pathogens[41,53,54], they often play a 'helper' role when other defence genes, like NBS-LRR, have been triggered in hosts[41,54]. In this study, we observed that substantial expression of chaperons and HSP occur at the same time when transcription factors (NAC and NmrA) are also up regulated (Tables 7 and 9). This shows that chaperons and HSP could also act as 'helper' components in resistance triggered by transcription factors. The roles of transcription factors (TF) (WRKY, GATA, NmrA, NAC), chaperones, heat shock proteins (HSP), late embryogenesis abundant (LEA) proteins and leucine rich repeats (LRR) in Namikoga's resistance to UCBSV have been extensively discussed in chapter 3[1].

A similar trend was observed in lower leaves (Table 4.10). However, more genes were differentially expressed in upper leaves (total of 74 (40 in Namikonga and 34 in Albert)) compared to lower leaves (total of 22 (10 in Namikonga and 12 in Albert)), suggesting that defence is more rigorously employed within young, actively growing leaves contrary to older, senescing leaves of the same plant.

4.5.2 Both Kiroba and Namikonga invoke similar defence genes at different times and leaf positions

Out of the 29 DEGs up regulated in Namikonga's UL at 3 dag (7 DEG), 5 dag (16 DEG) and 7 dag (6 DEG) (Table 4.9), seven were also up regulated in Kiroba. In Kiroba, three of the seven genes were up regulated in upper leaves (AI6RO7C, AID1U79 and AIGJRKP, which encode HSP, HSP and WRKY respectively) at 1 dag and 51 dag, while the remaining four (AIS09AX, AIKAL3D, AI6RO7D and AIVI5ND) which encode

chaperone, pathogenesis-related protein and TF(NmrA) were up regulated in Kiroba's lower leaves at 1 dag only.

From Namikonga's lower leaves (Table 4.10), two genes; AI20UOP (up regulated at 3 dag) and AIWR3TL (up regulated at 5 dag) that encode chaperone and LEA proteins were also up regulated in Kiroba's upper leaves at 51 dag. Despite the similarities between genes up regulated in Kiroba and Namikonga, Kiroba distinctly expressed one LRR gene (AIPAER8) at an exceedingly high FC (670) at 1 dag in its upper leaves (Table 4.11). At its lower leaves, it also expressed two chaperones (AIX01ZS and AI0IYB8) at exceedingly high FC levels (686 and 180), compared to other genes within Kiroba and other varieties.

Besides these, comparative analysis of Namikonga's response at Experiments 1 and 2 showed that two chaperone encoding genes (AIX01ZS and AIX01ZT), which were among the 11 genes stated in Table 4.7 (i.e. up regulated in Namikonga's upper leaves at 3 dag, 5 dag and 7 dag) were also up regulated in Kiroba's lower leaves at 1 dag. In Kiroba, the said genes had much higher fold change (FC 686 for AIX01ZS and FC 4.64 for AIX01ZT) (Table 4.12), compared to their expression in Namikonga's upper leaves at 3 dag, 5 dag and 7 dag (Table 4.7), where their FC was between 2.10 and 6.31.

Altogether, these observations underscore the fact that Kiroba's defence involves some similar 'helper' genes as in Namikonga, and that Kiroba's defence is ignited earlier (i.e 1 dag) than in Namikonga where defence expression reached a peak at 3 dag and 5 dag. In addition, Kiroba expressed one LRR gene (AIPAER8) at an exceedingly high FC (670) at 1 dag in its upper leaves, and two chaperones (AIX01ZS and AI0IYB8) at exceedingly high FC levels (686 and 180) in lower leaves, this indicates major differences in defence mechanisms. A better, more reliable observation can only be drawn if both varieties are observed under similar conditions and timing.

This distinct type of response observed in Kiroba (Tables 4.11 and 4.12) agrees with earlier findings where Kiroba's QTLs were different from that identified in Namikonga[55]. Kiroba had QTL for CBSD foliar symptoms on chromosomes 4, 6, 11, 15, 17 and 18, and for root necrosis on chromosomes 5, 11, 12 and 15, whereas Namikonga had QTL for foliar symptoms on chromosomes 2, 14 and 17 and root necrosis on 2, 11 and 18. QTL were detected on the same arm of chromosome 17 for CBSD foliar symptoms in both Kiroba and Namikonga, however different arms of chromosome 11 for CBSD root necrosis. This supports the observation that Kiroba and Namikonga have somewhat different defence responses.

4.5.3 Mkombozi's defence occurs at 1 dag in lower leaf position although roots develop observable necrosis at harvest

The highest number of Mkombozi's up regulated defence genes (16 genes) was observed in lower leaves at 1 dag (Table 4.12). These genes included six HSP, three chaperones, two each of WRKY, LRR, TF (GATA) and one NBD. By 51 dag, only five genes were up regulated in Mkombozi's lower leaves, similar to the trend observed in its upper leaves at both 1 dag and 51 dag. However, the observation of one

necrotic spot characteristic of CBSD in one storage root from the same Mkombozi plant (Fig. 3.1) might imply that Mkombozi is only tolerant to UCBSV.

4.5.4 Summary of gene expression results in Experiments 1 and 2

Altogether, the findings of Experiments 1 and 2 confirm that indeed Namikonga and Kiroba are resistant or tolerant to UCBSV. Namikonga's resistance involves expression of specific defence genes at specific time bounds, in this case 3 dag and 5 dag. For Kiroba, defence was similar at 1 dag and 51 dag (Tables 4.11 and 4.12), suggesting that a steady state defence mechanism is set up in Kiroba and sustained throughout the plants' life interaction with UCBSV. This trend is also observed in Mkombozi, except that Mkombozi developed a characteristic UCBSV-root-necrosis at harvest, indicating that Mkombozi is only tolerant to UCBSV. Very limited expression of defence genes was observed with NDL06/132 and its roots developed nearly as much necrosis as that observed in Albert. This indicates that NDL06/132, like Albert, is susceptible to UCBSV.

4.5.5 Limitations of the study

The low expression of genes at 51 dag (late phase) in all five cassava varieties agrees with findings of chapter 3, where Namikonga, showed very limited expression in the late phase (chapter 3 late phase findings were drawn from observations at 45 dag and 54 dag). However, in that study (chapter 3), Albert had high expression at 45 dag, which was not the same time point used in this study. Therefore, its not possible to draw conclusions on the same time point, as in this study, Albert had insufficient data at 45 dag. It must also be noted that experiments 1 and 2 were both established in different years and seasons, hence, possible difference in gene expression and virus titre. For Experiment 1, plants were sampled in 2011 and their RNA extracted in May 2012 (RNAseq was performed on cDNA from this batch, which included Albert and Namikonga) while for Experiment 2, plants were sampled in May 2013 and their RNA extracted in May 2014.

At the beginning, both experiments were set to analyse 20 time points of each variety across 55-defence, and one reference gene in both experiments. However, majority (> 60%) of the data was undetermined at PCR, probably because the cDNA template was degraded during transit to the sequencing facility. Therefore, only a few time points with determined reads were presented here. The rest of the time points and genes which are not presented here had undetermined or low quality Ct values.

Among the undetermined reactions was the candidate reference gene. Therefore, adjustments were made to pick another set of genes (from the remaining defence-genes with determined Ct values) to be used as reference genes. There is no guarantee that these *ad hoc* reference genes fully represent, without bias, the role of a stable reference gene. Besides, each experiment ended up with a different reference gene based on the available Ct for that gene and its stability in respective experiments. This further compounds the problem of referencing, as possible errors introduced by each reference gene makes it less reliable to compare outputs from experiments 1 and 2. Because of this, Experiment 1 included time points representative of the early time phase alone and not the early and late time phase. This made it difficult to

compare the study's outcome with results from the RNAseq study (Chapter 3) where observations and conclusions were drawn from specific time points of the late phase. Altogether, the findings presented in this chapter are indicative, but inconclusive, and calls for re-analysis involving all designated time points analysed and extensively compared with findings from other studies, including the RNAseq study in chapter 3.

4.6 References

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

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

3 Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One (**2014**), **9**, e96642.

4 Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. & Colvin, J. Re-emergence of cassava brown streak disease in Uganda. Plant Dis (**2007**), **91**, 24-29.

5 Bigirimana, S., Barumbanze, P., Ndayihanzamaso, P., Shirima, R. & Legg, J. First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Burundi. New Dis Rep (**2011**), **24**, 2044-0588.

6 Storey, H. H. Virus diseases of East African plants. VI-A progress report on studies of the disease of cassava. E Afri Agric J (**1936**), **2**, 34-39.

7 Mulimbi, W., Phemba, X., Assumani, B., Kasereka, P., Muyisa, S., Ugentho, H., Reeder, R., Legg, J., Laurenson, L. & Weekes, R. First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. New Dis Rep (**2012**), **26**, 2044-0588.

8 Gondwe, F. M. T., Mahungu, N. M., Hillocks, R. J., Raya, M. D., Moyo, C. C., Soko, M. M., Chipungu, F. & Benesi, I. R. M. Economic losses experienced by small-scale famers in Malawi due to cassava brown streak virus disease. In: Cassava Brown Streak Disease: Past, Present and Future. J. P. Legg and R. J. Hillocks, eds. Proc. Int. Workshop, Mombasa, Kenya. 27-30 2002. Natural Resources International Limited, Aylesford, UK. (**2003**), 28-38.

9 Nichols, R. F. W. Breeding cassava for virus resistance. E Afri Agric J (**1947**), **12**, 184-194.

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

11 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Genetically distinct strains of cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. Arch Virol (**2009**), **154**, 353-359.

12 Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M. & Butgereitt, A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen Virol (**2010**), **91**, 1365-1372.

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

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

15 Mware, B. O., Ateka, E. M., Songa, J. M., Narla, R. D., Olubayo, F. & Amata, R. Transmission and distribution of cassava brown streak virus disease in cassava growing areas of Kenya. J Appl Biosci (**2009**), **16**, 864 - 870.

16 Mbanzibwa, D. R., Tian, Y. P., Tugume, A. K., Mukasa, S. B., Tairo, F., Kyamanywa, S., Kullaya, A. & Valkonen, J. P. Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. J Virol Methods (**2011**), **171**, 394-400.

17 Hillocks, R. J., Raya, M. & Thresh, J. M. The association between root necrosis and above‐ground symptoms of brown streak virus infection of cassava in southern Tanzania. Int J Pest Manage (**1996**), **42**, 285-289.

18 Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. Adv Virol. (**2011**), **2012**, 795697.

19 Kanju, E. E., Masumba, E., Masawe, M., Tollano, S., Muli, B., Zacarias, A., Mahungu, N., Khizzah, B., Whyte, J. & Dixon, A. Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmer and consumer preferences. Proceedings 13th ISTRC Symp (**2007**), 95-101.

20 Hillocks, R. J. & Jennings, D. L. Cassava brown streak disease: A review of present knowledge and research needs. Int J Pest Manage (**2003**), **49**, 225-234.

21 Kulembeka, H. P., Ferguson, M., Herselman, L., Kanju, E., Mkamilo, G., Masumba, E., Fregene, M. & Labuschagne, M. T. Diallel analysis of field resistance to brown streak disease in cassava (Manihot esculenta Crantz) landraces from Tanzania. Euphytica (**2012**), **187**, 277-288.

22 Shi, S., Zhang, X., Mandel, M. A., Zhang, P., Zhang, Y., Ferguson, M., Amuge, T., Rounsley, S., Liu, Z. & Xiong, Z. Variations of five eIF4E genes across cassava accessions exhibiting tolerant and susceptible responses to cassava brown streak disease. PLoS One (**2017**), **12**, e0181998.

23 Ferguson, M., Koga, T. M., Johnson, D. A., Koga, K. A., Hirsch, G. N., Lopez-Lavalle, L. A. B. & Messier, W. Identification of genes that have undergone adaptive evolution in cassava (Manihot esculenta) and that may confer resistance to cassava brown streak disease. AJB (**2015**), **14**, 96-107.

24 Jennings, D. L. Observations on virus diseases of Cassava in resistant and susceptible varieties. II. Brown streak disease. Empire J Exp Agri (**1960**), **28**, 261-270.

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

26 Xu, J., Aileni, M., Abbagani, S. & Zhang, P. A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). Phytochem Anal. (**2010**), **21**, 395-398.

27 Alabi, O. J., Kumar, P. L. & Naidu, R. A. Multiplex PCR for the detection of African cassava mosaic virus and East African cassava mosaic Cameroon virus in cassava. J Virol Methods (**2008**), **154**, 111-120.

28 Nassuth, A., Pollari, E., Helmeczy, K., Stewart, S. & Kofalvi, S. A. Improved RNA extraction and one-tube RT-PCR assay for simultaneous detection of control plant RNA plus several viruses in plant extracts. J Virol Methods (**2000**), **90**, 37-49.

29 Proell, M., Riedl, S. J., Fritz, J. H., Rojas, A. M. & Schwarzenbacher, R. The NOD-like receptor (NLR) family: a tale of similarities and differences. PLoS One (**2008**), **3**, e2119.

30 Marone, D., Russo, M. A., Laido, G., De Leonardis, A. M. & Mastrangelo, A. M. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. Int J Mol Sci (**2013**), **14**, 7302-7326.

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

32 van Loon, L. C. & van Kammen, A. Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana tabacum* var. 'Samsun' and 'Samsun NN': II. changes in protein constitution after infection with *Tobacco mosaic virus*. Virology (**1970**), **40**, 199-211.

33 Banerjee, A. & Roychoudhury, A. WRKY proteins: signaling and regulation of expression during abiotic stress responses. Scientific World J (**2015**), **2015**, 17.

34 Donze, T., Qu, F., Twigg, P. & Morris, T. J. *Turnip crinkle virus* coat protein inhibits the basal immune response to virus invasion in Arabidopsis by binding to the NAC transcription factor TIP. Virology (**2014**), **449**, 207-214.

35 Nakashima, K., Tran, L. S. P., Van Nguyen, D., Fujita, M., Maruyama, K., Todaka, D., Ito, Y., Hayashi, N., Shinozaki, K. & Yamaguchi‐Shinozaki, K. Functional analysis of a NAC‐type transcription factor OsNAC6 involved in abiotic and biotic stress‐responsive gene expression in rice. The Plant J (**2007**), **51**, 617-630.

36 Voitsik, A. M., Muench, S., Deising, H. B. & Voll, L. M. Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. BMC Plant Biol (**2013**), **13**, 85.

37 Hundertmark, M. & Hincha, D. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics (**2008**), **9**, 118.

38 Lashley, T., L., H. J., Verbeek, M. M., Rostagno, A., Bojsen-Moller, M., David, G., van Horssen, J., Braendgaard, H., Plant, G., Frangione, B., Ghiso, J. & Revesz, T. Molecular chaperons, amyloid and preamyloid lesions in the BRI2 gene-related dementias: a morphological study. Neuropathol Appl Neurobiol. (**2006**), **32**, 492-504.

39 Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. FASEB J (**2004**), **18**, 1169 - 1175.

40 Mayer, F. L., Wilson, D., Jacobsen, I. D., Miramón, P., Slesiona, S., Bohovych, I. M., Brown, A. J. P. & Hube, B. Small but crucial: the novel small heat shock protein HSP21 mediates stress adaptation and virulence in *Candida albicans*. PLoS One (**2012**), **7**, e38584.

41 Zhong, L., Zhou, W., Wang, H., Ding, S., Lu, Q., Wen, X., Peng, L., Zhang, L. & Lu, C. Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. Plant Cell (**2013**), **25**, 2925-2943.

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

43 Katari, M. S., Nowicki, S. D., Aceituno, F. F., Nero, D., Kelfer, J., Thompson, L. P., Cabello, J. M., Davidson, R. S., Goldberg, A. P., Shasha, D. E., Coruzzi, G. M. & Gutierrez, R. A. VirtualPlant: a software platform to support systems biology research. Plant Physiol (**2010**), **152**, 500-515.

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

45 Team, R. C. (Available from CRAN sites, 2000).

46 Andersen, C. L., Jensen, J. L. & Orntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Research (**2004**), **64**, 5245 - 5250.

47 Auler, P. A., Benitez, L. C., do Amaral, M. N., Vighi, I. L., Rodrigues, G. S., da Maia, L. C. & Braga, E. J. B. Selection of candidate reference genes and validation for real-time PCR studies in rice plants exposed to low temperatures. LID - 10.4238/gmr16029695 [doi]. Genet Mol Res (**2017**), **16**.

48 De Spiegelaere, W., Dern-Wieloch, J., Weigel, R., Schumacher, V., Schorle, H., Nettersheim, D., Bergmann, M., Brehm, R., Kliesch, S., Vandekerckhove, L. & Fink, C. Reference Gene Validation for RT-qPCR, a Note on Different Available Software Packages. PLoS ONE (**2015**), **10**, e0122515.

49 Kałużna, M., Kuras, A. & Puławska, J. Validation of reference genes for the normalization of the RT-qPCR gene expression of virulence genes of Erwinia amylovora in apple shoots. Sci Rep (**2017**), **7**, 2034.

50 Perez, L. J., Rios, L., Trivedi, P., D'Souza, K., Cowie, A., Nzirorera, C., Webster, D., Brunt, K., Legare, J.-F., Hassan, A., Kienesberger, P. C. & Pulinilkunnil, T. Validation of optimal reference genes for quantitative real time PCR in muscle and adipose tissue for obesity and diabetes research. Sci Rep (**2017**), **7**, 3612.

51 Livak, K. & Schmittgen, T. Analysis of relative gene expression data using real-time quantitative PCR and the 2DeltaDeltaCT method. Methods (**2001**), **25**, 402 - 408.

52 Brown, J. K., Fauquet, C. M., Briddon, R. W., Zerbini, M., Moriones, E. & Navas-Castillo, J. Family Geminiviridae. In: King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, editors. Virus Taxonomy: Classification and Nomenclature of Viruses—Ninth Report of the International Committee on Taxonomy of Viruses. . USA: Elsevier Academic Press, San Diego, CA. (**2012**), 351–373.

53 Bieri, S., Mauch, S., Shen, Q. H., Peart, J., Devoto, A., Casais, C., Ceron, F., Schulze, S., Steinbiß, H. H. & Shirasu, K. RAR1 positively controls steady state levels of barley MLA resistance proteins and enables sufficient MLA6 accumulation for effective resistance. The Plant Cell Online (**2004**), **16**, 3480-3495.

54 Ito, M., Ohnishi, K., Hikichi, Y. & Kiba, A. Molecular chaperons and co-chaperons, HSP90, RAR1, and SGT1 negatively regulate bacterial wilt disease caused by *Ralstonia solanacearum* in *Nicotiana benthamiana*. Plant Signal Behav (**2014**), **10**, e970410.

55 Nzuki, I., Katari, M. S., Bredeson, J. V., Masumba, E., Kapinga, F., Salum, K., Mkamilo, G. S., Shah, T., Lyons, J. B., Rokhsar, D. S., Rounsley, S., Myburg, A. A. & Ferguson, M. E. QTL Mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*. Fron Plant Sci (**2017**), **8**, 1168.

Chapter 5

Summary of Major Research Findings

T. Amuge, M.E. Ferguson A. A. Myburg and D. K. Berger

Author Contributions

T.A.: Drafted the chapter

M.E.F, A.A.M and D.K.B.: Suggested changes in the layout and format, and proof-read the chapter

5.1 Abstract

Both Namikonga and Albert are host plants of UCBSV, as shown by positive results from UCBSV identification by PCR methods (Chapter 2). Levels of virus load, which influence detection, suggest that Namikonga's resistance to UCBSV involves restricting replication and movement of the virus. Low detection levels and titre of UCBSV in Namikonga correspond to high number of DEGs, which include major plant defence genes. The reverse is observed in susceptible Albert where virus levels were very high at some time points with very low numbers of DEGs. Analysis of some defence genes using two separate assays (RNAseq and QuantStudio) performed in different RNA samples from both Albert and Namikonga showed limited correlation. Using an additional set of cassava varieties (Kiroba, Mkombozi and NDL02/132), Kiroba was identified as resistant based on expression of defence-associated genes and having non-necrotic roots even at 12 months after grafting (mag) with UCBSV. Namikonga also had no necrotic roots at 12 mag.

5.2 Time-period from inoculation to detection, level of virus load, and degree of root necrosis, are key markers of resistance, tolerance or susceptibility

5.2.1 Time before virus is detectable

In this thesis, a locally verified RT-PCR method was used to establish presence or absence of UCBSV in both varieties, while qPCR was applied to determine virus load in a section of samples. Higher load of UCBSV was detectable in Albert compared to Namikonga. Both varieties showed higher detection levels in upper rather than lower leaves, relative to the graft point as described in Chapter 2. At 3 mag, Albert had detectable levels of UCBSV in all three biological replicates. Also, a study of 20 time points showed higher detection levels in upper leaves, implying that virus load was higher in these upper leaves. For Namikonga, UCBSV was detected in two out of three UCBSV-inoculated plants at 3 mag. The 20 time series study showed that UCBSV was detected at fewer time points of Namikonga compared to Albert, confirming the theory that Namikonga is resistant to UCBSV[1,2]. A similar observation was reported with Namikonga (also called Kaleso) in a 12-month study with CBSV[3].

Based on leaf symptoms, both Albert and Namikonga had observable leaf symptoms at 6 dag. This observation is characteristic for most varieties as CBSD-related symptoms develop on shoots of virus-infected cassava by eight weeks after inoculation[4,5]. Resistant varieties, however, may have delayed symptom development, or remain asymptomatic by this time. In another study with Albert (susceptible), Kiroba (resistant) and Kaleso (another name for Namikonga, a CBSD resistant variety), CBSV was first detected in Albert's roots at 4 dag but roots of both Kiroba and Kaleso tested negative at 4 dag[3]. In leaves, both CBSV and UCBSV were detected one week after grafting (wag) in Albert, 2 wag in Kiroba, and 8 wag in Kaleso[3].

Using qPCR to quantify virus, UCBSV load was measured in UCBSV-inoculated and control plants relative to a cassava reference gene (UBQ10). Therefore, fold accumulation of virus per time point is calculated as the change of Ct in infected sample relative to transcripts of cassava housekeeping gene UBQ10 in that sample[6]. Albert had the highest measure of UCBSV titre at 5 dag (2.1 fold) in upper leaves. This is contrary to Namikonga in which titre was lowest at 5 dag (0.2 fold) in upper leaves. This difference further supports the hypothesis that Namikonga's resistance is partly caused by its ability to restrict virus multiplication at specific time points, in this case 2 dag (Albert = 1.0 fold and Namikonga = 0.4 fold) and 5 dag (Albert = 2.1 fold and Namikonga = 0.2 fold).

From the above observations (i.e. time when first leaf symptoms become observable, number of infected samples from among inoculated treatments and virus load), a pattern of resistance/susceptibility is

observed. This pattern of resistance/susceptibility is similar to that observed in other studies of cassava and cassava viruses. For example, other CBSD-resistant varieties (KBH 2006/18 and KBH 2006/26)[7,8], alongside a another susceptible variety (60444)[9], no CBSD-related shoot symptoms were observed in graft inoculated plants of resistant varieties until 16 wag. On the contrary, plants from a susceptible variety showed typical CBSD-related shoot symptoms as early as 4 wag.

When stem cuttings for these infected, resistant varieties (KBH 2006/18, KBH 2006/26) were replanted, scions of resistant lines remained asymptomatic while all cuttings obtained from susceptible variety (60444) were symptomatic. Overall, 92% of cuttings replanted from UCBSV-infected plants of resistant varieties regenerated well and remained asymptomatic. On the other hand, only 60% of the cuttings from UCBSV-infected plants of susceptible variety (60444) regenerated, all having highly symptomatic shoots. The rest of the replanted 60444 cuttings succumbed to CBSD[9].

The observations reported above show that the time period before symptom emergence is one parameter by which cassava varieties may be classified as resistant or susceptible, prior to measuring the level of virus load in these infected symptomatic, and asymptomatic plants. Tolerance to CBSD is defined relative to resistant and susceptible responses. Tolerant varieties essentially accumulate moderate virus load relative to the resistant varieties and their roots become necrotic, but the level of root necrosis is not economically significant. This was the case with variety Mkombozi (Chapter 4).

5.2.2 Low virus titre indicates resistance by restriction of viral replication

As seen in other studies, measuring virus relative titre/load is one key strategy being used to identify susceptible from tolerant and/or resistant varieties in plants[9,10]. Accordingly, having observed higher detection levels (by RT-PCR) and higher virus titre (by qPCR) in upper leaves, it is likely that UCBSV replicates more in young, tender leaves, tapping into cassava's photo-assimilates from actively photosynthesising leaves. On the other hand, detection of more virus particles in upper leaves might be a factor of more efficient extraction of virus particles during RNA isolation from upper, tender and easy-to-crush leaves[11]. Tender leaves are known to yield higher amounts and better quality of nucleic acids than older, often hardened leaves[11].

Using field samples, virus load for both CBSD- and CMD-causing viruses was associated with higher expression of disease phenotype in cassava[12,13]. Studies conducted on cassava varieties widely grown in Uganda and Tanzania determined virus titre and scored shoot symptoms at 3, 5, 7, 9 and 11 map (months after planting), and at harvest scored root necrosis on plants grown under open field conditions. Starting with virus negative planting materials, test plants were exposed to whiteflies carrying on or both of the viruses. Some plants tested positive for CBSV alone, others UCBSV alone and some positive for both CBSV and UCBSV[2]. In four out of seven varieties that were UCBSV-positive, a strong positive correlation ($R^2 > 0.92$) between virus titre and shoot (leaf and stem) symptoms of CBSD was observed. Plants that tested positive for UCBSV, and measured high virus titre for the same had severe CBSD-related foliar symptoms, but plants with low virus titre had low symptoms[2]. A similar correlation of virus titre and disease phenotype was observed in transgenic cassava plants[14].

In another study on cassava, rate of accumulation of virus particles (fold) was determined relative to the quantity of virus particles in the susceptible variety (Albert) at the first week of the experiment[3]. In the study graft inoculated cassava plants were tested weekly over a period of 36 weeks. Titre of UCBSV was very low (≤10 fold) in both Kiroba and Kaleso (another name for Namikonga), but very high (≥800 fold) in Albert, the susceptible variety. For CBSV, titre was low (<10 fold) in Kaleso, medium (<100) in Kiroba and very high (≥200 fold) in Albert. In both resistant and susceptible varieties, titre of CBSV was higher than UCBSV and disease phenotype was more severe with CBSV compared to UCBSV infection[3]. At harvest, Kaleso's roots were non-necrotic while Kiroba's had a score of 2 (using the 1-5 scale). Roots of Albert, the susceptible variety, were highly susceptible (score 4)[3].

Elsewhere, CBSD-resistant cultivars had undetectable levels of both viruses, up to 8 weeks after grafting (wag), but virus titre for both CBSV and UCBSV transmitted from rootstock infected with both virus species, was higher in the susceptible plants, which were also clearly symptomatic by 8 wag. However, titre of UCBSV was higher than that of CBSV in the susceptible, an observation contrary to Maruthi et al.[3] with the same species[9]. A protoplast transfection with CBSV and UCBSV showed that the resistance in KBH2006/18 interferes with viral replication[9]. Both KBH2006/18 and Namikonga have introgressed sections of *M. glaziovii*, the wild cassava from which CBSD resistance is traced in their genomes[7]. The above observations provide evidence that UCBSV replicates in and around the infection site before spreading to distant organs and tissues[2-4,9,14].

5.3 Determination of DEGs in cassava-virus interaction is important, but time of sampling is most essential

The susceptible variety (Albert) had the lowest number (688) of differentially expressed genes (DEGs) compared to the resistant variety (Namikonga) with 10,028 DEGs. The RNAseq study showed that the highest number of DEGs were recorded at 2 dag (3887 DEGs) and 5 dag (4911 DEGs) in Namikonga, the same time points at which Namikonga expressed the lowest measures of virus titre. Albert either had very few (2 DEGs at 2 dag) or no DEGs (as seen at 5 dag) (Fig 5.1). Further analyses showed that the majority of Albert's DEGs were down regulated. In Namikonga however, the number of up regulated and down regulated genes were similar among the DEGs.

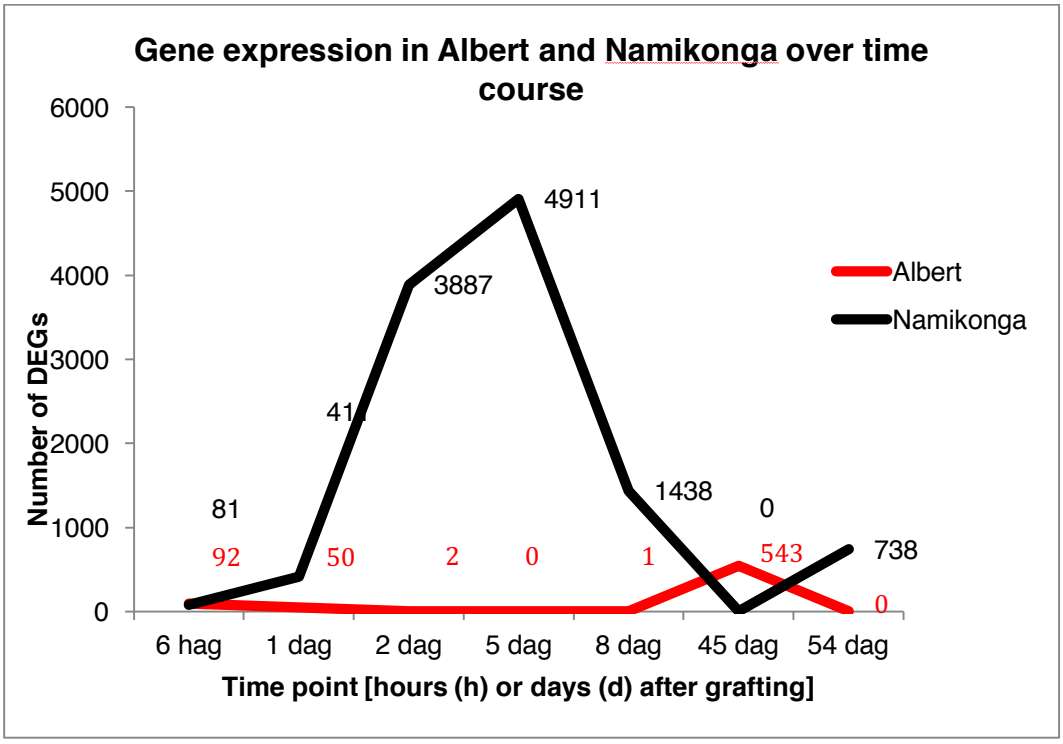


Fig. 5.1. Numbers of DEGs in Namikonga and Albert over time course.

The DEGs identified in Albert at 6 hag and 45 dag were enriched with GO terms related to sugar and organic acid metabolism. Similar GO terms were enriched in Namikonga at 1, 2, 5, 8 and 54 dag. However, other GO terms related to plant defence were enriched in Namikonga at 2 dag and 5 dag. These defence associate genes included phosphorelay signal transduction, elongation factors and the ribosomal subunit. Phosphorelay signal transduction governs transmission of signals about the infecting pathogens to host's response unit, so that resistance responses are triggered against the pathogen[15]. Their transmission normally involves kinases and response regulators[16,17]. Among enriched defence GO of Namikonga were elongation factors, which confer passive resistance to host plants by altering the virus-host interaction surface to a form unrecognizable by the virus[18,19]. The form of passive resistance largely occurs in the ribosome[20], a possible reason why ribosome proteins were substantially enriched in Namikonga.

Enrichment of these defence-related GO terms in Namikonga at 2 and 5 dag confirmed their significance in Namikonga's resistance to CBSD. It appears that defence in Namikonga becomes effective at 2 dag and the downstream events of the following days and weeks follow a cascade, which continuously render Namikonga resistant to CBSD. This kind of resistance may be a result of host genes within Namikonga, which upon detecting the virus, become expressed as a normal physiological process occurring all over the plant. None of the said gene ontologies were enriched in Albert at the same time points, but instead, Albert registered high UCBSV detection levels and high UCBSV virus titre, demonstrating absence of an appropriate means to control UCBSV replication.

Similar results were observed with other cassava-CBSD transcriptome analyses. Two studies; (Maruthi et al.[3] and Anjanappa et al.[21]) determined the numbers and types of genes that are differentially expressed between virus-infected (CBSVs) and control (or mock-infected) samples of cassava. In Maruthi et al.[3], test

plants were sampled 12 months after infection while Anjanappa et al.[21], DEGs were determined at 28 dag. Both studies showed interesting results.

For Maruthi et al.[3], 16 genes of resistant variety (Kaleso) had the highest RPKM (reads per kilobase per million reads) values (106 to 310). These 16 genes encode proteins involved in lipid transfer, cation exchange, phosphate response, chlorophyll binding and glucose-phosphate translocation. Some members of the NAC transcription family were among these top differentially expressed genes.

In the susceptible variety, the top most differentially expressed genes (RPKM 38 to 127) were those encoding proteins for myb domain, cytochrome P450, acyl-transferase, S-locus lectine protein kinase, NAD(P) binding protein (Rossman superfamily), DNA-binding protein and a senescence associated protein.

They further analysed genes amplifying the NAC domain. Their findings showed that some NAC encoding genes are differentially expressed in the resistant variety (24-140 fold) while the same genes were also differentially expressed in the susceptible variety (Albert), but at low scale (2-21 fold)[3]. When analysed in this thesis, expressions of these NAC genes followed a similar pattern with resistant variety having higher fold change compared to fold change observed in susceptible Albert.

More recently, Anjanappa et al.[21] determined the number of DEGs at 28 dag in resistant (KBH 2006/18) and susceptible (60444) varieties of cassava that were graft inoculated using root stock that tested positive for both CBSV and UCBSV. Results showed that the resistant variety had 585 DEGs while the susceptible variety had 1292 DEGs. Of all the DEGs, 158 were differentially expressed in both resistant and susceptible varieties, although 99 of the 158 genes were down regulated in the resistant variety.

One group of chloroplast proteins, the nuclear genes encoding chloroplast proteins (NGCPs) were more abundant in the resistant (19%) than in the susceptible (7%) variety. The NGCPs are induced during perception of PAMP signals[22].

In the susceptible variety, several genes were up regulated, including those that encode proteins involved in callose deposition at the plasmodesmata, enhanced viral movement through the plasmodesmata, and proteins that positively regulate cell-to-cell movement of virus particles. Also, the susceptible variety had up regulated levels of genes involved in lignin and salicylic acid (SA) biosynthesis (phosphofructose kinase 3, shikimate kinase 3, arogenate dehydratase, phenylalanine ammonia lyase, 4-coumarate:coA ligase), the TGA (TGACG MOTIF-BINDING PROTEIN 9) and WRKY 70 transcription factors among others. Genes involved in the antiviral RNA silencing pathway (DCL 2, argonaute 2, RNA-dependent RNA polymerase 1), silencing defective 5 (SDE5) were up regulated in the susceptible cultivar by up to 38.5 fold, yet there was no fold change of SDE5 in the resistant variety[21].

Non-detection of defence response genes in the resistant cultivar could be so because sampling at both studies was done at very late time points: 28 dag and 12 months after planting. This limitation, in part, is answered in this thesis, as samples were taken as early as 6 hag in UCBSV- and mock-inoculated samples.

5.3.1 There is limited or no correlation between RNAseq and QuantStudio analyses

So far, this thesis provides the first report of studies on cassava and CBSVs using QuantStudio qPCR platform. However, gene expression results drawn from RNAseq data of cassava infected with UCBSV have been validated using traditional qPCR techniques[3,21]. In one case four genes; a plamodesmata located protein 1 (PDLP1), calreticulin 3 (CRT3), calreticulin 1B (CRT1B) and heat shock protein 17.6 (HSP17.6) were analysed on qPCR. Based on RNAseq, the PDLP1 had 2.3 fold change between virus-inoculated and un-inoculated control in susceptible variety but was non-detected (NS) in the resistant variety. The CRT3 protein had NS fold change in susceptible variety and 1.8 fold change in resistant variety, while the CRT1B protein had NS expression in the susceptible variety and 2.0 fold change in the resistant variety. The HSP17.6 was non-detected in both susceptible and resistant varieties, using RNAseq. Unfortunately, expression values from qPCR were presented as absolute delta delta Ct values from qPCR[21], so, no fold change information was available to compare RNAseq and qPCR studies.

In another study, expression of NAC transcription factors in resistant and susceptible cassava varieties were analysed using RNAseq and qPCR technologies[3]. Six NAC encoding genes were studied using RNAseq and qPCR methods. Five out of the six genes were positively correlated between RNAseq and qPCR studies. The sixth gene was undetected at qPCR. In the susceptible variety, three of six NAC genes were positively correlated between RNAs7eq and qPCR technologies. Two genes, which had 2.2 and 3.6 fold change values in RNAseq were undetectable with qPCR. The other NAC gene, which was undetected at RNAseq scored 0.48 fold change with qPCR. In this thesis, three of these NAC genes (cassava gene id cassava4.1_011029m.g, cassava4.1_015961m.g and cassava4.1_023870, of cassava reference genome version 4.1[23]) were analysed. Results showed similar patterns expressions with highest fold change recorded in the resistant variety at 2 dag (Log2FC of 2.9, 2.2 and 1 respectively) and 5 dag (Log2FC of 5.1, 1.2 and 1 respectively). Altogether, these studies show different expression patterns with resistant and susceptible cassava varieties using RNAseq and qPCR studies.

In this thesis, the DEGs observed in the RNAseq study (Chapter 3) were analysed in Albert (susceptible variety) and Namikonga (resistant variety) over a time course. In the RNAseq study, samples were analysed at 6 hag, 1, 2, 5, 8, 45 and 54 dag. For the QuantStudio analysis, genes were analysed at 3, 5 and 7 dag. Twenty-seven genes common between RNAseq and QuantStudio studies were analysed at similar (5 dag in RNAseq and QuantStudio) or nearest (i.e. 2 dag versus 3 and 7 dag versus 8 dag) time points.

A comparative analysis of Albert and Namikonga using 27 defence-associated genes selected in Chapter 4 (Experiment 1) showed that at 2 and 3 dag, RNAseq and QuantStudio had no correlation in Albert because all 27 genes were not differentially expressed in Albert, hence, no fold change (or

Log2Foldchange) was available for comparison. For Namikonga, fold change data was available however; there was very low correlation (R^2 = 0.04) between RNAseq and QuantStudio results (Table 5.1). At 5 dag, Albert had negative (R^2 = -0.01) while Namikonga had very low (R^2 = 0.17) correlation (Table 5.2). This is the only common time point at which both RNAseq and QuanstStudio assays were compared in Albert and Namikonga. At 7 dag/ 8 dag, there was very low correlation (R^2 = 0.04) between RNAseq and QuantStudio assays of Namikonga, and none for Albert (Table 5.3)

Under normal circumstance, RNAseq and qPCR results should show high correlation. Poor correlation between RNAseq and qPCR studies is more common, rarely presented (if any) in plant-pathogen studies. For the first time with cassava, this thesis presents RNAseq and qPCR data where there is limited or no correlation of results. Limited correlation between the two methods may be due to technical variations in the normalisation processes. In RNAseq, global normalisation is performed with many genes, while in qPCR studies; normalisation is often performed by a limited number of reference “housekeeping” genes. In this study, two separate experiments were performed, each study having its owns set of reference genes. So, there is no doubt that results do not correlate because each study had its own unique point of reference.

The other reason for poor correlation is that majority of QuantStudio qPCR data were missing (shown as undetected signal). This lack of sufficient data was a major reason why only few time points (out of 20) were selected for downstream analysis. Even with the selected time points, several genes had limited or no data. Therefore, it was difficult to know whether the poor correlation was indeed a valid finding, or just a factor of missing data from QuantStudio experiments.

It also likely that both methods of gene expression had limited efficiency due to inefficiencies in primer design as one of the 56 QuantStudio primers did not amplify in all runs, including runs where all other genes amplified, or discrepancies during wet-laboratory procedures. During cDNA synthesis, some plates didn’t generate sufficient cDNA, prompting a repeat of this process from already freeze-thawed RNA. Beside all these, processed cDNA samples were transported (on dry ice) by courier from BecA (Nairobi, Kenya) to University of Pretoria (South Africa). This transit process alone could have led to further degradation of the cDNA.

Majority of the above limitations experienced with QuantStudio were not observed in RNAseq studies. The cDNA libraries for RNAseq were all prepared at BecA using the then 454-sequencing facility, a commercial unit where quality control was very high and the technical team were available to help, speeding the whole process. Coupled with RNAseq’s high sequencing depth, strictness ensured in preparing RNAseq libraries also contributed to the good quality of data generated. Nevertheless, a gene may still be undetected even in RNAseq study, as was reported in Anjanappa et al.[21].

Table 5.1: Gene expression values in RNAseq (at 2 dag) and QuantStudio (at 3 dag) in Albert and Namikonga: a comparative analysis

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based VirtualPlant*	Albert, RNAseq (2 dag)		Albert, QS (3 dag)		Namikonga, RNAseq (2 dag)		Namikonga, QS (3dag)	
			Log2FC	P-value	Log2FC	P-value	Log2FC	P-value	Log2FC	P-value
AIBJYVU	34154	Leucine-rich repeat (LRR) family protein	NA	NA	-0.6	0.207	0.9	0.05	0.5	0.023
AI89LJS	7501	Leucine-rich repeat (LRR) family protein	NA	NA	-0.8	0.039	3.57	0.00	NA	NA
AIQJCYG	978	Leucine-rich receptor-like protein kinase family protein	NA	NA	-1.0	0.174	NA	NA	0.4	0.007
AI5IQ05	29764	NB-ARC domain-containing disease resistance protein	NA	NA	-1.5	0.437	0.75	0.06	0.1	0.025
AICSW12	34172	NB-ARC domain-containing disease resistance protein	NA	NA	-2.4	0.652	0.20	NA	2.3	NA
AIWR3TL	19959	Late Embryogenesis Abundant 4-5	NA	NA	-6.7	0.002	2.30	NA	-3.6	0.810
AIGJRKP	11680	WRKY DNA-binding protein 70	NA	NA	2.1	0.004	-2.06	0.08	-1.1	0.024
AICSW11	8997	HSP70-interacting protein 1	NA	NA	-0.4	0.138	1.93	0.00	0.9	0.008
AI20UOO	2706	Chloroplast heat shock protein 70-2	NA	NA	0.4	0.012	NA	NA	1.9	0.009
AIT97G4	1607	Heat shock protein 70 (HSP 70) family protein	NA	NA	0.0	0.091	NA	NA	0.0	0.047
AI6RO7C	3340	Heat shock protein 70	NA	NA	-0.6	0.238	NA	NA	1.2	0.002
AIS09AX	18127	HSP20-like chaperones superfamily protein	NA	NA	0.3	0.222	4.47	0.01	1.8	0.009
AIX01ZT	20888	Trigger factor type chaperone family protein	NA	NA	1.1	0.028	2.12	0.00	1.4	0.011
AIBJYVT	8620	Trigger factor type chaperone family protein	NA	NA	1.0	0.023	1.91	0.00	NA	NA
AI0IYB8	1921	Chaperone protein htpG family protein	NA	NA	2.2	NA	1.52	0.00	1.2	0.004
AIRSA4P	18031	Chaperone DNAJ-domain superfamily protein	NA	NA	-0.4	0.212	3.19	0.00	1.1	0.013
AIWR3TK	1827	Chaperone protein htpG family protein	NA	NA	1.4	0.020	2.09	0.00	3.0	0.004
AIPAER9	14653	Chaperone DNAJ-domain superfamily protein	NA	NA	-1.1	0.191	3.89	0.00	0.1	0.070
AIX01ZS	1905	Chaperone protein htpG family protein	NA	NA	1.7	0.023	2.24	0.09	2.2	0.001
AI70NDK	4263	Chaperonin-60alpha	NA	NA	2.0	0.012	1.23	0.09	2.8	0.001
AILJJ9L	14410	Chaperonin 20	NA	NA	1.0	0.035	1.55	0.00	1.6	0.026
AIVI5ND	18353	Chaperone DNAJ-domain superfamily protein	NA	NA	1.4	0.038	2.41	0.00	0.6	0.002
AI70NDL	32811	GRAS family transcription factor	NA	NA	-1.0	0.191	2.88	0.00	-0.4	0.092

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based VirtualPlant*	Albert, RNAseq (2 dag)		Albert, QS (3 dag)		Namikonga, RNAseq (2 dag)		Namikonga, QS (3dag)	
			Log2FC	P-value	Log2FC	P-value	Log2FC	P-value	Log2FC	P-value
AIQJCYH	16750	GATA type zinc finger transcription factor family protein	NA	NA	-0.1	0.045	1.38	0.00	-0.4	0.012
AIHSPQX	11886	GATA transcription factor 5	NA	NA	0.8	0.032	2.16	0.00	-0.9	0.024
AIY9Z50	32811	GRAS family transcription factor	NA	NA	2.1	0.037	2.88	0.00	0.8	0.011
AI6RPBP	16601	Eukaryotic initiation factor 4E protein	NA	NA	0.1	0.077	0.40	NA	2.0	0.027

NA= data not available

Table 5.2: Gene expression values in RNAseq (at 5 dag) and QuantStudio (at 5 dag) in Albert and Namikonga: a comparative analysis

QuantStudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based VirtualPlant	Albert, RNAseq (2 dag)		Albert, QS (3 dag)		Namikonga, RNAseq (2 dag)		Namikonga, QS (3dag)	
			Log2FC	P-value	Log2FC	P-value	Log2FC	P-value	Log2FC	P-value
AIBJYVU	34154	Leucine-rich repeat (LRR) family protein	0.2	NA	0.7	0.017	1.8	0.015	2.1	0.050
AI89LJS	7501	Leucine-rich repeat (LRR) family protein	0.1	NA	-0.1	0.457	7.7	NA	3.1	0.023
AIQJCYG	978	Leucine-rich receptor-like protein kinase family protein	-0.7	NA	0.6	0.251	0.5	0.766	0.6	0.851
AI5IQ05	29764	NB-ARC domain-containing disease resistance protein	-0.5	NA	0.6	0.134	2.3	0.012	-0.2	0.303
AICSW12	34172	NB-ARC domain-containing disease resistance protein	0.6	NA	-0.3	0.488	1.8	0.048	0.7	0.946
AIWR3TL	19959	Late Embryogenesis Abundant 4-5	-0.1	NA	0.1	0.360	0.2	0.600	-3.9	0.000
AIGJRKP	11680	WRKY DNA-binding protein 70	-2.1	NA	-0.4	0.828	0.6	0.544	3.3	0.002
AICSW11	8997	HSP70-interacting protein 1	0.1	NA	1.4	0.005	3.3	0.015	0.6	0.738
AI20UOO	2706	Chloroplast heat shock protein 70-2	-0.4	NA	1.1	0.002	3.7	0.006	1.0	0.454
AIT97G4	1607	Heat shock protein 70 (HSP 70) family protein	0.1	NA	0.0	0.369	1.0	0.383	0.0	0.211
AI6RO7C	3340	Heat shock protein 70	0	NA	-0.8	0.909	2.3	0.001	-1.9	0.007
AIS09AX	18127	HSP20-like chaperones superfamily protein	0	NA	-0.1	0.241	14.3	0.003	0.2	0.114
AIX01ZT	20888	Trigger factor type chaperone family protein	0.1	NA	0.7	0.034	3.4	0.057	2.7	0.005
AIBJYVT	8620	Trigger factor type chaperone family protein	0.1	NA	1.8	0.009	1.0	0.013	2.1	0.095
AI0IYB8	1921	Chaperone protein htpG family protein	-0.1	NA	0.8	0.011	3.9	0.014	2.8	0.003
AIRSA4P	18031	Chaperone DNAJ-domain superfamily protein	-0.7	NA	1.3	0.009	3.8	0.030	2.3	0.009
AIWR3TK	1827	Chaperone protein htpG family protein	-0.5	NA	1.4	0.094	9.0	NA	3.5	0.004
AIPAER9	14653	Chaperone DNAJ-domain superfamily protein	0.4	NA	-0.1	0.554	4.0	0.070	2.0	0.052
AIX01ZS	1905	Chaperone protein htpG family protein	0.6	NA	1.0	0.021	1.0	NA	2.3	0.012
AI70NDK	4263	Chaperonin-60alpha	-0.2	NA	1.3	0.004	6.5	0.002	2.0	0.045
AILJJ9L	14410	Chaperonin 20	-0.5	NA	1.0	0.007	4.6	0.002	1.9	0.016
AIVI5ND	18353	Chaperone DNAJ-domain superfamily protein	0	NA	-1.6	0.669	17.4	0.009	-0.2	0.138

QuantStudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based VirtualPlant	Albert, RNAseq (2 dag)		Albert, QS (3 dag)		Namikonga, RNAseq (2 dag)		Namikonga, QS (3dag)	
			Log2FC	P-value	Log2FC	P-value	Log2FC	P-value	Log2FC	P-value
AI70NDL	32811	GRAS family transcription factor	-0.1	NA	1.2	0.080	3.2	0.008	-0.6	0.085
AIQJCYH	16750	GATA type zinc finger transcription factor family protein	0.3	NA	0.3	0.359	40.4	0.031	2.7	0.002
AIHSPQX	11886	GATA transcription factor 5	-0.9	NA	0.6	0.216	1.9	0.051	1.1	0.403
AIY9Z50	32811	GRAS family transcription factor	-0.1	NA	1.8	0.002	1.3	NA	1.8	0.081
AI6RPBP	16601	Eukaryotic initiation factor 4E protein	-0.5	NA	0.3	0.358	5.7	0.025	1.7	0.201

NA= data not available

Table 5.3: Gene expression values in RNAseq (at 8 dag) and QuantStudio (at 7 dag) in Albert and Namikonga: a comparative analysis

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based VirtualPlant*	Albert, RNAseq (2 dag)		Albert, QS (3 dag)		Namikonga, RNAseq (2 dag)		Namikonga, QS (3 dag)	
			Log2FC	P-value	Log2FC	P-value	Log2FC	P-value	Log2FC	P-value
AIBJYVU	34154	Leucine-rich repeat (LRR) family protein	1.77	0.00	NA	NA	NA	NA	-1.3	0.101
AI89LJS	7501	Leucine-rich repeat (LRR) family protein	2.59	0.00	NA	NA	NA	NA	-0.8	0.437
AIQJCYG	978	Leucine-rich receptor-like protein kinase family protein	1.55	0.00	NA	NA	NA	NA	-0.4	0.727
AI5IQ05	29764	NB-ARC domain-containing disease resistance protein	1.01	0.00	NA	NA	NA	NA	-0.7	0.466
AICSW12	34172	NB-ARC domain-containing disease resistance protein	1.36	0.03	NA	NA	NA	NA	0.3	0.572
AIWR3TL	19959	Late Embryogenesis Abundant 4-5	2.31	0.00	NA	NA	5.40	0.05	-3.7	0.003
AIGJRKP	11680	WRKY DNA-binding protein 70	2.08	0.01	NA	NA	NA	NA	-2.2	0.029
AICSW11	8997	HSP70-interacting protein 1	0.77	0.01	NA	NA	NA	NA	-0.1	0.889
AI20UOO	2706	Chloroplast heat shock protein 70-2	0.76	0.01	NA	NA	NA	NA	-0.1	0.903
AIT97G4	1607	Heat shock protein 70 (HSP 70) family protein	1.19	0.00	NA	NA	NA	NA	0.0	0.870
AI6RO7C	3340	Heat shock protein 70	0.98	0.02	NA	NA	NA	NA	-1.4	0.130
AIS09AX	18127	HSP20-like chaperones superfamily protein	NA	NA	NA	NA	NA	NA	0.4	0.409
AIX01ZT	20888	Trigger factor type chaperone family protein	0.85	0.05	NA	NA	NA	NA	-0.8	0.452
AIBJYVT	8620	Trigger factor type chaperone family protein	0.84	0.01	NA	NA	NA	NA	0.3	0.723
AI0IYB8	1921	Chaperone protein htpG family protein	1.82	0.00	NA	NA	NA	NA	1.0	0.199
AIRSA4P	18031	Chaperone DNAJ-domain superfamily protein	2.93	0.00	NA	NA	NA	NA	-1.0	0.336
AIWR3TK	1827	Chaperone protein htpG family protein	2.23	0.00	NA	NA	NA	NA	0.1	0.781
AIPAER9	14653	Chaperone DNAJ-domain superfamily protein	2.25	0.00	NA	NA	NA	NA	0.6	0.450
AIX01ZS	1905	Chaperone protein htpG family protein	1.00	0.00	NA	NA	NA	NA	1.1	0.264
AI70NDK	4263	Chaperonin-60alpha	1.17	0.00	NA	NA	NA	NA	0.9	0.232
AILJJ9L	14410	Chaperonin 20	1.17	0.00	NA	NA	NA	NA	0.7	0.299

Quantstudio code for gene	Gene ID (Prefix "cassava4.1_", suffix "m.g")	Gene annotation based VirtualPlant*	Albert, RNAseq (2 dag)		Albert, QS (3 dag)		Namikonga, RNAseq (2 dag)		Namikonga, QS (3 dag)	
			Log2FC	P-value	Log2FC	P-value	Log2FC	P-value	Log2FC	P-value
AIVI5ND	18353	Chaperone DNAJ-domain superfamily protein	1.81	0.00	NA	NA	-2.13	0.00	2.0	NA
AI70NDL	32811	GRAS family transcription factor	0.81	0.01	NA	NA	NA	NA	0.4	0.374
AIQJCYH	16750	GATA type zinc finger transcription factor family protein	NA	NA	NA	NA	NA	NA	2.6	0.023
AIHSPQX	11886	GATA transcription factor 5	1.71	0.00	NA	NA	NA	NA	1.1	0.409
AIY9Z50	32811	GRAS family transcription factor	0.81	0.01	NA	NA	NA	NA	0.3	0.652
AI6RPBP	16601	Eukaryotic initiation factor 4E protein	-0.4	NA	NA	NA	NA	NA	0.7	0.370

NA= data not available

5.4 Epilogue

The observations summarized in this chapter agrees with other studies that:

1. Namikonga and Kiroba are resistant to CBSD[1,2,24].
2. Albert and NDL06/132 are susceptible to CBSD[1,2,24].
3. Namikonga's resistance involves reduced replication and/or movement of UCBSV to other plant parts, shown by low detection levels in UCBSV-inoculated plants[2,3]. Virus titre was not measured in Kiroba (this study), so the role of virus titre in Kiroba's resistance could not be determined or compared with other studies where virus titre was measured in Kiroba.
4. This mode of resistance in Namikonga involves known plant defence genes expressed at specific time points and with defined levels of expression[25].
5. There is limited or no correlation between RNAseq and QuantStudio analyses.

These observations suggest that resistance to CBSD is a factor of virus multiplication and replication. Levels of virus load are inversely proportional to expression of defence genes in Namikonga, but in Albert, virus load is highest at time points with the lowest expression of defence genes. However, the studies above did not measure how long each virus particle lives inside its host, or how much damage is caused in the host (whether resistant or susceptible). Moreover, it has not been established whether or not these defence-associated genes translate to form functional proteins that actively regulate defence in the host. A study to determine these parameters could explain the underlying factors leading to Namikonga's resistance, which will help breeders in selection of high yielding, farmer-preferred, disease resistant cassava varieties[26].

5.5 Limitations of the study and future prospects

The study focused on analysis of five candidate CBSD resistant cultivars, of which two (Namikonga and Kiroba) were confirmed resistant. RNAseq was only performed on leaf samples of one resistant variety, Namikonga. It is not known if the transcriptome profiles observed in Namikonga would be similar in Kiroba because both varieties have different genetic backgrounds[27]. Preliminary findings from QuantStudio analysis (Chapter 4, this study) already show that both resistant varieties have non-identical gene expression profiles at different time points. Future studies should focus on profiling the transcriptomes of all CBSD resistant varieties under similar experimental conditions and timepoints. This will generate more informative data.

Only leaf samples of Namikonga were used in the transcriptome analysis for resistance to CBSD. Nevertheless, CBSD symptoms are known to manifest on leaves, stems and roots[1,12,28,29]. In fact, varieties are declared resistant or susceptible on the basis of symptoms expression in all three tissues. Therefore, drawing conclusions of resistance profiles from leaf samples alone is not representative of the whole signaling pathway leading to resistance in Namikonga. Future studies should consider profiling leaves, stems and roots from all CBSD resistant varieties in order to generate comprehensive results.

The whole study focused on analysis of resistance in Namikonga based on a single virus species, UCBSV. But, CBSD is caused by two virus species: CBSV and UCBSV[30,31]. Infection is normally caused by single or mixed infection by both virus species. Therefore, a transcriptome profiling of resistant varieties using both virus species as single and mixed infection will provide a better understanding of the nature and basis of resistance when cassava is challenged by single or mixed infection from both virus species.

An analysis of 12 new genomes alongside already known genomes confirmed evolution within both virus species[32]. Indeed, signatures of newer species were noticed in the study of 12+ genomes. This discovery calls for a more robust avenue for identifying new variants, strains and even species of the CBSD-causing viruses. With the advent of metagenomics technology, it is possible to perform surveillance studies of all cassava varieties and other CBSD hosts to identify new virus strains and species before they spread and cause epidemic level losses in cassava production.

Metagenomics approach has made it possible to diagnose plants with virus-like symptoms and asymptomatic plants to identify the infecting pathogen(s). This technology is particularly useful for identifying previously unknown strains of known pathogen species and new, emergent pathogens. The approach is to extract nucleotides (RNA or DNA) from target samples and perform deep-sequencing using next-generation or third generation technologies. Sequences obtained can be analyzed via taxonomic profiling and/or comparison to other already available pathogen sequences. This technology has prospects in aquaculture research[33], identification of chicken viruses[34] and human-based studies as seen with identification of food-borne diseases[35]. In plants, it has been applied to identify new viruses affecting common beans (*Phaseolus vulgaris* L.)[36,37], maize[38], and for development of infectious clones for plant

virology studies (reviewed by Fabio 2019[39]). For CBSVs, a metagenomics approach is essential considering that CBSVs recently re-emerged to affect nearly all cassava growing areas previously unaffected[40-43]. This re-emergence brought with it a new, previously unknown species UCBSV, and there are reports of rapid evolution within the genomes of both UCBSV and CBSV[32,44]. There is a high possibility that other species and strains of CBSD causing virus are still undiscovered, and therefore no breeding efforts are being taken to control it. Hence, the serious need for a metagenomics approach to rapidly identify and characterize new strains to develop prevention measures before the strains cause significant losses.

5.6 References

1 Kanju, E. E., Masumba, E., Masawe, M., Tollano, S., Muli, B., Zacarias, A., Mahungu, N., Khizzah, B., Whyte, J. & Dixon, A. Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmer and consumer preferences. Proceedings 13th ISTRC Symp (**2007**), 95-101.

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

3 Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PLoS One (**2014**), **9**, e96642.

4 Ogwok, E., Patil, B. L., Alicai, T. & Fauquet, C. M. Transmission studies with cassava brown streak Uganda virus (Potyviridae: Ipomovirus) and its interaction with abiotic and biotic factors in *Nicotiana benthamiana*. J Virol Methods (**2010**), **169**, 296-304.

5 Wagaba, H., Beyene, G., Trembley, C., Alicai, T., Fauquet, C. M. & Taylor, N. J. Efficient transmission of cassava brown streak disease viral pathogens by chip bud grafting. BMC Res Not (**2013**), **6**, 516.

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

7 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. & Rokhsar, D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. Nature Biotechnol (**2016**), **34**, 562-570.

8 Nzuki, I., Katari, M. S., Bredeson, J. V., Masumba, E., Kapinga, F., Salum, K., Mkamilo, G. S., Shah, T., Lyons, J. B., Rokhsar, D. S., Rounsley, S., Myburg, A. A. & Ferguson, M. E. QTL Mapping for pest and disease resistance in cassava and coincidence of some QTL with introgression regions derived from *Manihot glaziovii*. Fron Plant Sci (**2017**), **8**, 1168.

9 Anjanappa, R. B., Mehta, D., Maruthi, M. N., Kanju, E., Gruissem, W. & Vanderschuren, H. Characterization of Brown Streak Virus–Resistant Cassava. Mol Plant Microbe Interact (**2016**), **29**, 527-534.

10 Adams, I. P., Abidrabo, P., Miano, D. W., Alicai, T., Kinyua, Z. M., Clarke, J., Macarthur, R., Weekes, R., Laurenson, L., Hany, U., Peters, D., Potts, M., Glover, R., Boonham, N. & Smith, J. 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 Pathol (**2013**), **62**, 233-242.

11 Xu, J., Aileni, M., Abbagani, S. & Zhang, P. A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). Phytochem Anal. (**2010**), **21**, 395-398.

12 Rwegasira, G. M. & Rey, C. M. E. Relationship between symptoms expression and virus detection in cassava brown virus streak-infected plants. J Agric Sci (**2012**), **4**, 246-253.

13 Rwegasira, G. M., Rey, M. E. C. & Nawabu, H. Approaches to diagnosis and detection of cassava brown streak virus (Potiviridae: Ipomovirus) in field-grown cassava crop. Afr J F Agric Nutri Develop (**2011**), **11**, 4739-4756.

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

15 Sheen, J. Phosphorelay and transcription control in cytokinin signal transduction. Science (**2002**), **296**, 1650-1652.

16 Bauer, J., Reiss, K., Veerabagu, M., Heunemann, M., Harter, K. & Stehle, T. Structure–function analysis of *Arabidopsis thaliana* histidine kinase AHK5 bound to Its cognate phosphotransfer protein AHP1. Mol Plant (**2013**), **6**, 959-970.

17 Romir, J., Harter, K. & Stehle, T. Two-component systems in *Arabidopsis thaliana* – a structural view. Eur J Cell Biol (**2010**), **89**, 270-272.

18 Nicaise, V., German-Retana, S., Sanjuán, R., Dubrana, M., Mazier, M., Maisonneuve, B., Candresse, T., Caranta, C. & LeGall, O. The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus *Lettuce mosaic virus*. Plant Physiol (**2003**), **132**, 1272-1282.

19 Bai, S., Liu, J., Chang, C., Zhang, L., Maekawa, T., Wang, Q., Xiao, W., Liu, Y., Chai, J., Takken, F. L., Schulze-Lefert, P. & Shen, Q. H. Structure-function analysis of barley NLR immune receptor MLA10 reveals its cell compartment specific activity in cell death and disease resistance. PLoS Pathog (**2012**), **8**, e1002752.

20 Ashby, J. A., Stevenson, C. E., Jarvis, G. E., Lawson, D. M. & Maule, A. J. Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to *Pea seed-borne mosaic virus* in pea. PLoS One (**2011**), **6**, e15873.

21 Anjanappa, R. B., Mehta, D., Okoniewski, M. J., Szabelska-Beresewicz, A., Gruissem, W. & Vanderschuren, H. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. Mol Plant Pathol (**2018**), **19**, 476-489.

22 Shi, B., Lin, L., Wang, S., Guo, Q., Zhou, H., Rong, L., Li, J., Peng, J., Lu, Y., Zheng, H., Yang, Y., Chen, Z., Zhao, J., Jiang, T., Song, B., Chen, J. & Yan, F. Identification and regulation of host genes related to Rice stripe virus symptom production. New Phytol (**2016**), **209**, 1106-1119.

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

24 Ogwok, E., Alicai, T., Rey, M. E. C., Beyene, G. & Taylor, N. J. Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta*) plants. Plant Pathol (**2015**), **64**, 1235-1246.

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

26 Zimmermann, P., Hirsch-Hoffmann, M., Henning, L. & Gruissem, W. GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol (**2004**), **136**, 2621 - 2632.

27 Ferguson, M., Koga, T. M., Johnson, D. A., Koga, K. A., Hirsch, G. N., Lopez-Lavalle, L. A. B. & Messier, W. Identification of genes that have undergone adaptive evolution in cassava (Manihot esculenta) and that may confer resistance to cassava brown streak disease. AJB (**2015**), **14**, 96-107.

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

29 Hillocks, R. J., Raya, M. D., Mtunda, K. & Kiozia, H. Effects of brown streak virus disease on yield and quality of cassava in Tanzania. J Phytopathol (**2001**), **149**, 389-394.

30 Monger, W. A., Alicai, T., Ndunguru, J., Kinyua, Z. M., Potts, M., Reeder, R. H., Miano, D. W., Adams, I. P., Boonham, N., Glover, R. H. & Smith, J. The complete genome sequence of the Tanzanian strain of cassava brown streak virus and comparison with the Ugandan strain sequence. Arch Virol (**2010**), **155**, 429-433.

31 Winter, S., Koerbler, M., Stein, B., Pietruszka, A., Paape, M. & Butgereitt, A. Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. J Gen Virol (**2010**), **91**, 1365-1372.

32 Ndunguru, J., Sseruwagi, P., Tairo, F., Stomeo, F., Maina, S., Djikeng, A., Kehoe, M. & Boykin, L. M. Analyses of Twelve New Whole Genome Sequences of Cassava Brown Streak Viruses and Ugandan Cassava Brown Streak Viruses from East Africa: Diversity, Supercomputing and Evidence for Further Speciation. PLoS One (**2015**), **10**, e0141939.

33 Munang'andu, H. M. Environmental Viral Metagenomics Analyses in Aquaculture: Applications in Epidemiology and Disease Control. Front Microbiol (**2016**), **7**, 1986-1986.

34 Qiu, Y., Wang, S., Huang, B., Zhong, H., Pan, Z., Zhuang, Q., Peng, C., Hou, G. & Wang, K. Viral infection detection using metagenomics technology in six poultry farms of eastern China. PloS One (**2019**), **14**, e0211553.

35 Carleton, H. A., Besser, J., Williams-Newkirk, A. J., Huang, A., Trees, E. & Gerner-Smidt, P. Metagenomic Approaches for Public Health Surveillance of Foodborne Infections: Opportunities and Challenges. Foodborne Pathog Dis (**2019**), **16**, 474-479.

36 Mutuku, J. M., Wamonje, F. O., Mukeshimana, G., Njuguna, J., Wamalwa, M., Choi, S.-K., Tungadi, T., Djikeng, A., Kelly, K., Domelevo Entfellner, J.-B., Ghimire, S. R., Mignouna, H. D., Carr, J. P. & Harvey, J. J. W. Metagenomic Analysis of Plant Virus Occurrence in Common Bean (*Phaseolus vulgaris*) in Central Kenya. Front Microbiol (**2018**), **9**, 2939-2939.

37 Wainaina, J. M., Ateka, E., Makori, T., Kehoe, M. A. & Boykin, L. M. A metagenomic study of DNA viruses from samples of local varieties of common bean in Kenya. PeerJ (**2019**), **7**, e6465.

38 Wamaitha, M. J., Nigam, D., Maina, S., Stomeo, F., Wangai, A., Njuguna, J. N., Holton, T. A., Wanjala, B. W., Wamalwa, M., Lucas, T., Djikeng, A. & Garcia-Ruiz, H. Metagenomic analysis of viruses associated with maize lethal necrosis in Kenya. Virol. J (**2018**), **15**, 90.

39 Pasin, F., Menzel, W. & Daros, J. A. Harnessed viruses in the age of metagenomics and synthetic biology: an update on infectious clone assembly and biotechnologies of plant viruses. Plant Biotech J (**2019**), **17**, 1010-1026.

40 Alicai, T., Omongo, C. A., Maruthi, M. N., Hillocks, R. J., Baguma, Y., Kawuki, R., Bua, A., Otim-Nape, G. W. & Colvin, J. Re-emergence of cassava brown streak disease in Uganda. Plant Dis (**2007**), **91**, 24-29.

41 Amisse, J. J. G., Ndunguru, J., Tairo, F., Boykin, L. M., Kehoe, M. A., Cossa, N., Ateka, E., Rey, C. & Sseruwagi, P. First report of Cassava brown streak viruses on wild plant species in Mozambique. Physiol Mol Plant Pathol (**2019**), **105**, 88-95.

42 Bigirimana, S., Barumbanze, P., Ndayihanzamaso, P., Shirima, R. & Legg, J. First report of cassava brown streak disease and associated Ugandan cassava brown streak virus in Burundi. New Dis Rep (**2011**), **24**, 2044-0588.

43 Mulimbi, W., Phemba, X., Assumani, B., Kasereka, P., Muyisa, S., Ugentho, H., Reeder, R., Legg, J., Laurenson, L. & Weekes, R. First report of Ugandan cassava brown streak virus on cassava in Democratic Republic of Congo. New Dis Rep (**2012**), **26**, 2044-0588.

44 Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, M. A. & Boykin, L. M. Cassava brown streak virus has a rapidly evolving

genome: implications for virus speciation, variability, diagnosis and host resistance. Sci Rep (**2016**), **6**, 36164.

Appendix

6.1 Supplementary tables

Table S3.1: Laboratories where RNA sequencing was carried out. UCB=University of California, Berkeley; DOW-Dow Agrosciences, Indianapolis

Time point	Albert	Namikonga
6 hag	UCB	UCB
1 dag	DOW	DOW
2 dag	DOW	DOW
5 dag	DOW	DOW
8 dag	UCB	UCB
45 dag	UCB	DOW
54 dag	UCB	DOW

Table S3.2: Numbers of RNAseq reads that were mapped or unmapped to the cassava reference genome v4.1 (plant accession number AM560-2) compared between read-lengths and genotypes.

(a) Output with 50 bp read length

Variety	Mapped	Unmapped	Total Coverage	Mapped (%)	Unmapped (%)
Albert	19,744,151,034	4,754,316,730	24,498,467,764	81	19
Namikonga	11,793,050,691	2,760,319,897	14,553,370,588	81	19
Total	**31,537,201,725**	**7,514,636,627**	39,051,838,352		

(b) Output with 101 bp read length

Variety	Mapped	Unmapped	Total Coverage	Mapped (%)	Unmapped (%)
Albert	10,696,774,139	3,477,542,903	14,174,317,042	75	25
Namikonga	20,697,282,109	5,717,006,828	26,414,288,937	78	22
Total	**31,394,056,248**	**9,194,549,731**	**40,588,605,979**		

Mapped and unmapped reads were equally distributed in both varieties with 50 bp mRNA reads (a) but Namikonga mapped slightly better (78%) than Albert (75%) with 101 bp reads (b). Altogether, shorter reads mapped better for both varieties (a).

Table S3.3: Number of biological replicates for each cassava treatment used to identify differentially expressed genes using DESeq software.

Time-point	Albert		Namikonga	
	UCBSV-inoculated	Mock-inoculated	UCBSV-inoculated	Mock-inoculated
6 hag	3	3	3	2
1 dag	2	2	3	3
2 dag	3	2	3	2
5 dag	2	2	3	3
8 dag	3	2	3	3
45 dag	3	2	3	3
54 dag	3	3	3	2

Table S3.4: Number of enriched GO terms from up-regulated and down-regulated genes between UCBSV-inoculated and mock- inoculated cassava (varieties Albert and Namikonga) at each time point.

Time point	Albert						Namikonga					
	Up-regulated genes			Down-regulated genes			Up-regulated genes			Down-regulated genes		
	B	M	C	B	M	C	B	M	C	B	M	C
6 hag	6	2	0	34	10	0	1	9	7	0	0	0
1 dag	0	0	0	3	5	0	29	0	1	7	7	2
2 dag	0	7	0	-	-	-	32	2	0	0	11	1
5 dag	-	-	-	-	-	-	18	6	14	0	0	3
8 dag	-	-	-	0	0	0	46	24	2	23	8	5
45 dag	0	2	4	16	6	6	-	-	-	-	-	-
54 dag	-	-	-	-	-	-	44	22	0	7	36	6

GO enrichment is reported for the following categories: biological process (B), molecular function (M) and cellular component (C) identified using the *biomaps* tool of VirtualPlant 1.3 (2) queried on the cassava genome v4.1 (3). At some time points, there were no differentially expressed genes (represented by -) or there were no GO terms associated with any of the differentially expressed genes (represented by 0).

Table S3.5: Log2Foldchange and adjusted (FDR-corrected) P- values of two eIF4E genes in Albert and Namikonga at all sampled time points.

Time point	Albert				Namikonga			
	G016620m.g		G013223m.g		G016620m.g		G013223m.g	
	Log2FC	P-adj	Log2FC	P-adj	Log2FC	P-adj	Log2FC	P-adj
6 hag	-0.11	1	-0.17	1	0.30	1	0.22	1
1 dag	0.19	1	0.23	1	-0.13	1	0.01	1
2 dag	0.23	1	0.10	1	-0.19	0.860	-0.17	0.909
5 dag	0.44	1	0.41	1	0.72	0.020	0.63	0.062
8 dag	0.00	1	0.04	1	-0.19	1	-0.26	0.973
45 dag	-0.22	1	0.17	1	0.04	1	-0.07	1
54 dag	NA	NA	NA	NA	NA	NA	NA	NA

Log2foldchange (Log2FC) and level of significance (Adj p-values) of selected cassava4.1_016620m.g (G016620m.g) and cassava4.1_013223m.g (G013223m.g) genes in Albert and Namikonga at all sampled time points.

Table S3.6: Expression ratios in cassava variety Namikonga of 56 manually selected genes that belong to defence-related gene families and that showed significant differential expression in at least one treatment. Gene families included those encoding proteins with Leucine Rich Repeat motifs (LRR), NBARC-motifs, pathogenesis- related (PR) proteins, late embryogenesis abundant (LEA) proteins, heat shock proteins, chaperone proteins, elongation factors (eIF) and several transcription factors (WRKY, NAC, GATA, GRAS, NmrA).

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
LRR															
Cassava4.1_001048m.g	Leucine-rich repeat transmembrane protein kinase	0.1	>0.05	-0.4	>0.05	1.9	>0.05	-1.3	1.20E-05	0.2	>0.05	-0.5	>0.05	-2.3	>0.05
Cassava4.1_034154m.g	Leucine-rich repeat (LRR) family protein	0.0	>0.05	-0.2	>0.05	1.0	>0.05	1.8	2.24E-07	-0.1	>0.05	-0.2	>0.05	0.7	>0.05
Cassava4.1_007501m.g	Leucine-rich repeat (LRR) family protein	0.4	>0.05	1.2	>0.05	3.8	>0.05	2.6	7.26E-16	-0.1	>0.05	0.0	>0.05	0.2	>0.05
Cassava4.1_000978m.g	Leucine-rich receptor-like protein kinase family protein	1.0	>0.05	-0.1	>0.05	0.3	>0.05	1.5	2.08E-04	0.4	>0.05	-0.2	>0.05	0.4	>0.05
Cassava4.1_000765m.g	Leucine-rich repeat transmembrane protein kinase	0.7	>0.05	1.0	>0.05	0.8	>0.05	2.1	1.36E-02	1.1	>0.05	-0.7	>0.05	-0.1	>0.05
NBARC															

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_029764m.g	NB-ARC domain-containing disease resistance protein	-0.1	>0.05	-0.4	>0.05	0.6	>0.05	1.0	1.20E-03	0.2	>0.05	0.2	>0.05	-0.6	>0.05
Cassava4.1_022732m.g	NB-ARC domain-containing disease resistance protein	0.3	>0.05	-0.2	>0.05	0.1	>0.05	1.0	1.43E-03	0.4	>0.05	-0.1	>0.05	-0.8	>0.05
Cassava4.1_022172m.g	NB-ARC domain-containing disease resistance protein	-0.1	>0.05	-0.1	>0.05	0.7	>0.05	1.1	2.33E-03	0.1	>0.05	0.0	>0.05	-0.7	>0.05
Cassava4.1_001752m.g	NB-ARC domain-containing disease resistance protein	0.1	>0.05	-0.1	>0.05	0.6	>0.05	1.1	4.72E-02	0.1	>0.05	-0.4	>0.05	-1.0	>0.05
Cassava4.1_034172m.g	NB-ARC domain-containing disease resistance protein	0.4	>0.05	-0.4	>0.05	0.2	>0.05	1.4	3.68E-02	0.2	>0.05	0.0	>0.05	-1.0	>0.05
Cassava4.1_000058m.g	NB-ARC domain-containing disease resistance protein	0.2	>0.05	-0.2	>0.05	0.5	>0.05	1.1	2.92E-03	-0.2	>0.05	0.0	>0.05	-0.7	>0.05
PR															
Cassava4.1_012383m.g	Pathogenesis-related thaumatin superfamily protein	1.1	>0.05	0.2	>0.05	2.5	2.88E-02	3.5	2.48E-03	0.6	>0.05	0.4	>0.05	1.5	1.70E-02
Cassava4.1_011960m.g	Pathogenesis-related thaumatin superfamily protein	1.1	>0.05	0.5	>0.05	-2.2	>0.05	1.4	>0.05	2.3	1.62E-03	-0.8	>0.05	0.7	>0.05
LEA															

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_025676m.g	Late embryogenesis abundant protein, group 2	0.4	>0.05	1.0	>0.05	1.2	2.89E-02	1.4	3.32E-06	-0.1	>0.05	-0.3	>0.05	-0.1	>0.05
Cassava4.1_019959m.g	Late Embryogenesis Abundant 4-5	-1.0	>0.05	0.0	>0.05	2.3	>0.05	2.3	1.28E-05	5.4	>0.05	0.3	>0.05	-0.4	>0.05
WRKY															
Cassava4.1_014614m.g	WRKY DNA-binding protein 40	3.0	>0.05	-0.6	>0.05	-2.2	>0.05	4.0	4.54E-03	1.7	>0.05	-3.5	>0.05	-3.3	1.70E-07
Cassava4.1_011680m.g	WRKY DNA-binding protein 70	1.3	>0.05	-0.1	>0.05	-2.2	2.08E-02	2.1	1.16E-02	2.2	>0.05	-1.9	>0.05	-1.8	>0.05
Heat shock proteins															
Cassava4.1_033681m.g	Heat shock protein 90.1	0.6	>0.05	1.9	>0.05	4.7	>0.05	2.0	3.67E-02	-0.6	>0.05	-0.2	>0.05	-0.6	>0.05
Cassava4.1_034243m.g	Heat-shock protein 70T-2	-0.3	>0.05	1.1	>0.05	3.9	>0.05	0.0	>0.05	0.0	>0.05	0.3	>0.05	0.3	>0.05
Cassava4.1_003240m.g	Heat shock protein 70B	1.5	>0.05	2.2	>0.05	3.6	>0.05	-1.3	>0.05	1.1	>0.05	1.5	>0.05	0.9	>0.05
Cassava4.1_014648m.g	Heat shock protein 21	2.0	>0.05	2.0	>0.05	7.6	1.58E-01	1.6	>0.05	1.0	>0.05	-0.6	>0.05	-3.8	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_009750m.g	Heat shock transcription factor A2	-0.4	>0.05	1.3	>0.05	6.1	1.35E-03	7.3	2.78E-03	-0.5	>0.05	-1.4	>0.05	0.0	>0.05
Cassava4.1_010803m.g	Heat shock transcription factor A6B	0.7	>0.05	2.0	>0.05	5.4	6.80E-05	3.4	7.36E-04	-2.8	>0.05	-0.1	>0.05	0.0	>0.05
Cassava4.1_003144m.g	Heat shock protein 70 (HSP 70) family protein	1.0	>0.05	-0.2	>0.05	0.8	4.41E-02	2.1	>0.05	-0.6	>0.05	-0.7	>0.05	-0.3	>0.05
Cassava4.1_001300m.g	Heat shock protein 101	0.4	>0.05	2.0	>0.05	1.5	>0.05	-0.2	>0.05	0.8	>0.05	0.5	>0.05	1.4	>0.05
Cassava4.1_007779m.g	DNAJ heat shock N-terminal domain-containing protein	0.7	>0.05	1.6	>0.05	0.8	>0.05	1.2	3.97E-04	-0.8	>0.05	0.2	>0.05	0.1	>0.05
Cassava4.1_008997m.g	HSP70-interacting protein 1	0.4	>0.05	0.9	>0.05	1.8	1.53E-07	0.8	1.61E-02	0.0	>0.05	-0.4	>0.05	0.5	>0.05
Cassava4.1_001924m.g	Heat shock protein 89.1	0.3	>0.05	2.1	>0.05	1.3	>0.05	2.4	2.06E-14	-0.1	>0.05	-0.7	>0.05	-0.2	>0.05
Cassava4.1_002706m.g	Chloroplast heat shock protein 70-2	0.0	>0.05	0.8	>0.05	0.1	>0.05	0.8	6.52E-03	-0.5	>0.05	-0.2	>0.05	-0.1	>0.05
Cassava4.1_001607m.g	Heat shock protein 70 (HSP 70) family protein	0.2	>0.05	2.1	>0.05	0.9	>0.05	1.2	7.00E-06	-0.5	>0.05	-0.2	>0.05	0.2	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_003340m.g	Heat shock protein 70	0.9	>0.05	1.7	>0.05	0.9	>0.05	1.0	2.46E-02	0.1	>0.05	-0.1	>0.05	-0.7	>0.05
Chaperones															
Cassava4.1_018127m.g	HSP20-like chaperones superfamily protein	1.0	>0.05	1.8	>0.05	4.1	>0.05	1.3	>0.05	-0.4	>0.05	-0.5	>0.05	-0.9	>0.05
Cassava4.1_020888m.g	Trigger factor type chaperone family protein	-0.2	>0.05	2.0	6.39E-04	2.4	1.60E-04	0.8	>0.05	0.1	>0.05	0.3	>0.05	0.0	>0.05
Cassava4.1_008620m.g	Trigger factor type chaperone family protein	-0.2	>0.05	1.7	3.55E-03	2.0	1.79E-07	0.8	1.53E-02	0.1	>0.05	0.1	>0.05	-0.2	>0.05
Cassava4.1_001921m.g	Chaperone protein htpG family protein	0.1	>0.05	2.0	>0.05	1.4	1.82E-04	1.8	5.72E-12	-0.6	>0.05	-0.2	>0.05	-0.5	>0.05
Cassava4.1_018031m.g	Chaperone DNAJ-domain superfamily protein	0.2	>0.05	2.1	>0.05	3.2	2.18E-19	2.9	9.29E-28	0.4	>0.05	0.6	>0.05	0.4	>0.05
Cassava4.1_018200m.g	HSP20-like chaperones superfamily protein	0.0	>0.05	2.0	>0.05	5.7	>0.05	1.5	>0.05	0.9	>0.05	-0.4	>0.05	-1.7	>0.05
Cassava4.1_026342m.g	HSP20-like chaperones superfamily protein	2.7	>0.05	1.6	>0.05	8.6	>0.05	-1.1	>0.05	0.5	>0.05	-10.0	>0.05	-1.2	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_001827m.g	Chaperone protein htpG family protein	1.4	>0.05	0.2	>0.05	2.0	1.81E-07	2.2	2.78E-18	-0.9	>0.05	-1.0	>0.05	-0.2	>0.05
Cassava4.1_014653m.g	Chaperone DNAJ-domain superfamily protein	0.8	>0.05	1.6	>0.05	3.9	4.07E-25	2.3	3.15E-15	-0.4	>0.05	0.3	>0.05	-1.7	>0.05
Cassava4.1_001905m.g	Chaperone protein htpG family protein	0.4	>0.05	2.0	>0.05	1.9	>0.05	1.0	1.41E-03	0.4	>0.05	0.0	>0.05	-0.6	>0.05
Cassava4.1_004263m.g	Chaperonin-60alpha	0.2	>0.05	1.7	>0.05	1.7	3.27E-02	1.2	2.03E-05	-0.6	>0.05	-0.4	>0.05	-0.2	>0.05
Cassava4.1_014410m.g	Chaperonin 20	0.0	>0.05	1.3	>0.05	1.3	1.82E-03	1.2	5.13E-05	-0.1	>0.05	-0.3	>0.05	-0.1	>0.05
Cassava4.1_018353m.g	Chaperone DNAJ-domain superfamily protein	1.7	>0.05	1.3	>0.05	2.9	6.91E-03	1.8	1.71E-04	-2.1	1.90E-03	0.4	>0.05	1.8	>0.05
Cassava4.1_001905m.g	Chaperone protein htpG family protein	0.4	>0.05	2.0	>0.05	1.9	>0.05	1.0	1.41E-03	0.4	>0.05	0.0	>0.05	-0.6	>0.05

Transcription factors (NmrA, GATA and GRAS)

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_028589m.g	NmrA-like negative transcriptional regulator family protein	-0.4	>0.05	-0.2	>0.05	3.2	2.64E-04	3.5	8.63E-06	0.1	>0.05	0.5	>0.05	4.7	>0.05
Cassava4.1_032524m.g	NmrA-like negative transcriptional regulator family protein	-0.7	>0.05	1.2	>0.05	2.7	3.00E-03	2.2	3.63E-02	0.4	>0.05	0.6	>0.05	2.8	>0.05
Cassava4.1_032811m.g	GRAS family transcription factor	-0.1	>0.05	0.2	>0.05	3.2	4.52E-05	0.8	6.83E-03	0.0	>0.05	0.5	>0.05	-0.3	>0.05
Cassava4.1_016750m.g	GATA type zinc finger Transcription factor family protein	-0.2	>0.05	1.3	>0.05	2.0	>0.05	0.7	>0.05	-0.7	>0.05	-0.2	>0.05	-0.1	>0.05
Cassava4.1_011886m.g	GATA transcription factor 5	0.3	>0.05	0.3	>0.05	2.8	>0.05	1.7	1.70E-05	-0.8	>0.05	0.4	>0.05	0.4	>0.05
NAC transcription factors															
Cassava4.1_011029m.g	NAC domain containing protein 96	-1.5	>0.05	0.2	>0.05	1.0	>0.05	5.1	2.07E-06	0.1	>0.05	-2.0	>0.05	0.6	>0.05
Cassava4.1_015961m.g	NAC transcription factor-like 9	0.8	>0.05	1.3	>0.05	2.9	>0.05	1.0	>0.05	-0.6	>0.05	-5.8	>0.05	0.8	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue	Log2FC	Adj. Pvalue
Cassava4.1_023870m.g	NAC transcription factor-like 9	0.4	>0.05	1.0	>0.05	2.2	7.62E-04	1.2	>0.05	-1.1	>0.05	-2.7	>0.05	1.3	>0.05
eIF															
Cassava4.1_016601m.g	Eukaryotic initiation factor 4E protein	0.5	>0.05	-0.3	>0.05	0.4	>0.05	0.4	>0.05	-0.8	>0.05	-0.2	>0.05	0.3	>0.05
TF															
Cassava4.1_032811m.g	GRAS family transcription factor	-0.1	>0.05	0.2	>0.05	3.2	4.52E-05	0.8	6.83E-03	0.0	>0.05	0.5	>0.05	-0.3	>0.05

Log2foldchange (Log2FC) and level of significance (adj p-values) of selected 56 genes. Most significant genes, and respective time points of significance are highlighted red. Genes highlighted blue were not significant after FDR correction of p-values but were selected based on their substantial fold change. VirtualPlant annotation iS based on Arabidopsis genes as done on homology mapping using information provided in Phytozome.

Table S3.7: Expression ratios in cassava variety Albert of 56 manually selected genes that belong to defence-related gene families**.** These are the same 56 genes as shown in Table S3.6, but none of them was statistically significantly differentially expressed at any time point in Albert.

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
LRR															
Cassava4.1_001048m.g	Leucine-rich repeat transmembrane protein kinase	0.3	>0.05	-0.4	>0.05	0.1	>0.05	0.1	>0.05	-0.3	>0.05	2.6	>0.05	0.6	>0.05
Cassava4.1_034154m.g	Leucine-rich repeat (LRR) family protein	-0.9	>0.05	0.6	>0.05	0.6	>0.05	0.2	>0.05	-0.1	>0.05	-0.5	>0.05	-0.8	>0.05
Cassava4.1_007501m.g	Leucine-rich repeat (LRR) family protein	0.2	>0.05	-1.1	>0.05	0.2	>0.05	0.1	>0.05	0.2	>0.05	2.2	>0.05	-0.1	>0.05
Cassava4.1_000978m.g	Leucine-rich receptor-like protein kinase family protein	-0.8	>0.05	0	>0.05	-0.1	>0.05	-0.7	>0.05	0.3	>0.05	0.5	>0.05	-0.2	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_000765m.g	Leucine-rich repeat transmembrane protein kinase	-0.7	>0.05	-0.2	>0.05	-0.8	>0.05	0.2	>0.05	0	>0.05	-0.5	>0.05	-1.2	>0.05
NBARC															
Cassava4.1_029764m.g	NB-ARC domain-containing disease resistance protein	-0.1	>0.05	-0.6	>0.05	0.2	>0.05	-0.5	>0.05	0	>0.05	0.3	>0.05	0	>0.05
Cassava4.1_022732m.g	NB-ARC domain-containing disease resistance protein	-0.4	>0.05	-0.3	>0.05	0.1	>0.05	0.1	>0.05	0	>0.05	0	>0.05	0.1	>0.05
Cassava4.1_022172m.g	NB-ARC domain-containing disease resistance protein	-0.2	>0.05	-0.2	>0.05	0.3	>0.05	0.7	>0.05	0.1	>0.05	-0.3	>0.05	-0.2	>0.05
Cassava4.1_001752m.g	NB-ARC domain-containing disease resistance protein	-0.3	>0.05	-0.6	>0.05	0.2	>0.05	-1	>0.05	0	>0.05	0.5	>0.05	-0.1	>0.05
Cassava4.1_034172m.g	NB-ARC domain-containing disease resistance protein	-0.3	>0.05	-0.5	>0.05	0.5	>0.05	0.6	>0.05	0.2	>0.05	0.2	>0.05	0	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_000058m.g	NB-ARC domain-containing disease resistance protein	-0.3	>0.05	-0.4	>0.05	0.3	>0.05	-0.1	>0.05	0	>0.05	0.8	>0.05	-0.1	>0.05
PR															
Cassava4.1_012383m.g	Pathogenesis-related thaumatin superfamily protein	-0.3	>0.05	1.4	>0.05	0.4	>0.05	-3	>0.05	0.6	>0.05	1.7	>0.05	-0.5	>0.05
Cassava4.1_011960m.g	Pathogenesis-related thaumatin superfamily protein	-1.9	>0.05	-1	>0.05	-0.5	>0.05	0.3	>0.05	0.1	>0.05	-0.3	>0.05	-0.7	>0.05
LEA															
Cassava4.1_025676m.g	Late embryogenesis abundant protein, group 2	-0.1	>0.05	0.2	>0.05	-0.4	>0.05	-0.3	>0.05	-0.1	>0.05	-0.2	>0.05	-0.4	>0.05
Cassava4.1_019959m.g	Late Embryogenesis Abundant 4-5	-0.1	>0.05	0.4	>0.05	0.5	>0.05	-0.1	>0.05	-0.6	>0.05	-0.5	>0.05	1.1	>0.05
WRKY															

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_014614m.g	WRKY DNA-binding protein 40	-3.5	>0.05	-2.4	>0.05	-1.6	>0.05	-1.1	>0.05	0	>0.05	-0.2	>0.05	-1.6	>0.05
Cassava4.1_011680m.g	WRKY DNA-binding protein 70	-0.9	>0.05	-1.2	>0.05	-0.3	>0.05	-2.1	>0.05	0.5	>0.05	0	>0.05	-0.6	>0.05
Heat shock proteins															
Cassava4.1_033681m.g	Heat shock protein 90.1	-0.9	>0.05	1.1	>0.05	1	>0.05	0.9	>0.05	-0.5	>0.05	-1.3	>0.05	-0.1	>0.05
Cassava4.1_034243m.g	Heat-shock protein 70T-2	-0.3	>0.05	0.8	>0.05	1.1	>0.05	-0.4	>0.05	-0.8	>0.05	-1.2	>0.05	0.4	>0.05
Cassava4.1_003240m.g	Heat shock protein 70B	-0.1	>0.05	0.9	>0.05	1.6	>0.05	0.6	>0.05	-1.2	>0.05	0	>0.05	0.7	>0.05
Cassava4.1_014648m.g	Heat shock protein 21	0.1	>0.05	0.5	>0.05	1.4	>0.05	3.5	>0.05	-2.7	>0.05	-0.9	>0.05	2.8	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_009750m.g	Heat shock transcription factor A2	-1.1	>0.05	0.3	>0.05	-0.6	>0.05	-0.3	>0.05	0	>0.05	-1.2	>0.05	0.9	>0.05
Cassava4.1_010803m.g	Heat shock transcription factor A6B	1.9	>0.05	1.8	>0.05	2.2	>0.05	-0.1	>0.05	0.7	>0.05	-0.3	>0.05	0.9	>0.05
Cassava4.1_003144m.g	Heat shock protein 70 (Hsp 70) family protein	-0.8	>0.05	0.1	>0.05	-0.2	>0.05	-1	>0.05	0	>0.05	-0.1	>0.05	-0.3	>0.05
Cassava4.1_001300m.g	Heat shock protein 101	-0.5	>0.05	1	>0.05	0.2	>0.05	0.3	>0.05	0.1	>0.05	-0.8	>0.05	-0.6	>0.05
Cassava4.1_007779m.g	DNAJ heat shock N-terminal domain-containing protein	-0.7	>0.05	0.8	>0.05	0.3	>0.05	0	>0.05	0.2	>0.05	-0.5	>0.05	-0.2	>0.05
Cassava4.1_008997m.g	HSP70-interacting protein 1	-0.2	>0.05	0.3	>0.05	-0.3	>0.05	0.1	>0.05	-0.1	>0.05	-0.3	>0.05	0.1	>0.05
Cassava4.1_001924m.g	Heat shock protein 89.1	-0.3	>0.05	0.4	>0.05	-0.2	>0.05	0.3	>0.05	-0.3	>0.05	-0.9	>0.05	0	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_002706m.g	Chloroplast heat shock protein 70-2	-0.1	>0.05	0.4	>0.05	0.1	>0.05	-0.4	>0.05	0.1	>0.05	0	>0.05	-0.2	>0.05
Cassava4.1_001607m.g	Heat shock protein 70 (Hsp 70) family protein	-0.6	>0.05	0.8	>0.05	0.1	>0.05	0.1	>0.05	0.1	>0.05	-0.2	>0.05	-0.4	>0.05
Cassava4.1_003340m.g	Heat shock protein 70	-0.9	>0.05	0.8	>0.05	0.6	>0.05	0	>0.05	-0.2	>0.05	-0.6	>0.05	-0.6	>0.05
Chaperones															
Cassava4.1_018127m.g	HSP20-like chaperones superfamily protein	0.3	>0.05	0.4	>0.05	0.4	>0.05	0	>0.05	-0.4	>0.05	-0.4	>0.05	0.6	>0.05
Cassava4.1_020888m.g	Trigger factor type chaperone family protein	0.1	>0.05	0.3	>0.05	0.3	>0.05	0.1	>0.05	-0.1	>0.05	0.2	>0.05	0	>0.05
Cassava4.1_008620m.g	Trigger factor type chaperone family protein	0.3	>0.05	0.3	>0.05	0.2	>0.05	0.1	>0.05	0.2	>0.05	0.2	>0.05	0	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_001921m.g	Chaperone protein htpG family protein	0.2	>0.05	0.4	>0.05	-0.3	>0.05	-0.1	>0.05	-0.1	>0.05	-0.2	>0.05	0.1	>0.05
Cassava4.1_018031m.g	Chaperone DNAJ-domain superfamily protein	-0.2	>0.05	0.4	>0.05	0.5	>0.05	-0.7	>0.05	0	>0.05	0.8	>0.05	-0.1	>0.05
Cassava4.1_018200m.g	HSP20-like chaperones superfamily protein	0.4	>0.05	1	>0.05	2.1	>0.05	1.5	>0.05	-0.9	>0.05	-1.1	>0.05	0.5	>0.05
Cassava4.1_026342m.g	HSP20-like chaperones superfamily protein	2.2	>0.05	0.8	>0.05	4.5	>0.05	-0.6	>0.05	-1.5	>0.05	-10	>0.05	1.9	>0.05
Cassava4.1_001827m.g	Chaperone protein htpG family protein	-1	>0.05	0	>0.05	-0.6	>0.05	-0.5	>0.05	-0.1	>0.05	-0.2	>0.05	-0.3	>0.05
Cassava4.1_014653m.g	Chaperone DNAJ-domain superfamily protein	0.7	>0.05	0.8	>0.05	0.6	>0.05	0.4	>0.05	0.3	>0.05	0.8	>0.05	0.1	>0.05
Cassava4.1_001905m.g	Chaperone protein htpG family protein	0.4	>0.05	0.7	>0.05	-0.1	>0.05	0.6	>0.05	-0.2	>0.05	0	>0.05	0.2	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_004263m.g	Chaperonin-60alpha	0.1	>0.05	0.5	>0.05	-0.1	>0.05	-0.2	>0.05	-0.1	>0.05	0.1	>0.05	-0.1	>0.05
Cassava4.1_014410m.g	Chaperonin 20	0	>0.05	0.3	>0.05	-0.3	>0.05	-0.5	>0.05	-0.1	>0.05	-0.1	>0.05	0	>0.05
Cassava4.1_018353m.g	Chaperone DNAJ-domain superfamily protein	-2.1	>0.05	0.6	>0.05	0.6	>0.05	0	>0.05	0.2	>0.05	-0.2	>0.05	-0.1	>0.05
Cassava4.1_001905m.g	Chaperone protein htpG family protein	0.4	>0.05	0.7	>0.05	-0.1	>0.05	0.6	>0.05	-0.2	>0.05	0	>0.05	0.2	>0.05
Transcription factors (NmrA, GATA and GRAS)															
Cassava4.1_028589m.g	NmrA-like negative transcriptional regulator family protein	-2.1	>0.05	-1.4	>0.05	0.5	>0.05	-1.5	>0.05	0.2	>0.05	4.5	>0.05	-1.1	>0.05
Cassava4.1_032524m.g	NmrA-like negative transcriptional	-4	>0.05	-1.6	>0.05	-0.3	>0.05	-3.6	>0.05	0.7	>0.05	3.5	>0.05	0.1	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
	regulator family protein														
Cassava4.1_032811m.g	GRAS family transcription factor	0. vc4	>0.05	0.1	>0.05	0.6	>0.05	-0.1	>0.05	0.2	>0.05	-0.2	>0.05	0.6	>0.05
Cassava4.1_016750m.g	GATA type zinc finger transcription factor family protein	0.2	>0.05	0.2	>0.05	0.4	>0.05	0.3	>0.05	0.1	>0.05	0.2	>0.05	-0.1	>0.05
Cassava4.1_011886m.g	GATA transcription factor 5	0	>0.05	-0.5	>0.05	0.6	>0.05	-0.9	>0.05	0.2	>0.05	-0.1	>0.05	0.2	>0.05
NAC transcription factors															
Cassava4.1_011029m.g	NAC domain containing protein 96	-0.7	>0.05	-0.1	>0.05	0.4	>0.05	-2.4	>0.05	-0.5	>0.05	-0.4	>0.05	-0.7	>0.05
Cassava4.1_015961m.g	NAC transcription factor-like 9	0.6	>0.05	-1.4	>0.05	0.4	>0.05	0.4	>0.05	-3	>0.05	-2	>0.05	-3.2	>0.05

Gene ID	Gene annotation based VirtualPlant*	6 hag		1 dag		2 dag		5 dag		8 dag		45 dag		54 dag	
		Log2FC	Adj. Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue	Log2FC	Adj Pvalue
Cassava4.1_023870m.g	NAC transcription factor-like 9	0.3	>0.05	0.2	>0.05	0.6	>0.05	-1.9	>0.05	-1	>0.05	-0.5	>0.05	-0.4	>0.05
eIF															
Cassava4.1_016601m.g	Eukaryotic initiation factor 4E protein	-0.5	>0.05	0	>0.05	0.1	>0.05	-0.5	>0.05	-0.4	>0.05	-0.7	>0.05	-0.1	>0.05
TF															
Cassava4.1_032811m.g	GRAS family transcription factor	0.4	>0.05	0.1	>0.05	0.6	>0.05	-0.1	>0.05	0.2	>0.05	-0.2	>0.05	0.6	>0.05

Log2foldchange (Log2FC) and level of significance (Adj p-values) of selected 56 genes.

Table S4.1: Number of determined Ct values (after filtering out undetermined data) from Experiment 1

	Time point																		
	6 hag	**12 hag**	**1 dag**	**2 dag**	**3 dag**	**4 dag**	**5 dag**	**6 dag**	**7 dag**	**8 dag**	**45 dag**	**46 dag**	**47 dag**	**48 dag**	**49 dag**	**50 dag**	**51 dag**	**52 dag**	**53 dag**	**54 dag**
Determined [T]	122	1307	1473	1782	2186	162 5	2150	1484	1885	143 7	272	592	264	942	1181	794	575	644	1257	1373
%	4.5	48.6	54.8	66.3	81.3	60.5	80.0	55.2	70.1	53.5	10.1	22.0	9.8	35.0	43.9	29.5	21.4	24.0	46.8	51.1

[T] = Number of QuantStudio reactions that were scored as "determined" (row 1) and respective percentage (row 2) for each time point. Determined Ct values range from 16.7 to 33.8. **Yellow cells =** Time points were retained for DEG analysis. Green cells = Time points (in addition to those in yellow cells) used for selection of reference genes. hag: Hours after grafting. dag: Days after grafting

Table S4.2: Number of determined Ct values (after filtering out undetermined data) from Experiment 2

		Time point																		
	Before grafting (time 0)	**6 hag**	**12 hag**	**1 dag**	**2 dag**	**3 dag**	**4 dag**	**5 dag**	**6 dag**	**7 dag**	**8 dag**	**45 dag**	**46 dag**	**47 dag**	**48 dag**	**49 dag**	**50 dag**	**51 dag**	**52 dag**	**53 dag**	**54 dag**
Determined [T]	1611	2850	80	323 5	775	120	63	550	511	111 5	191 2	312	2903	744	37	2544	2546	3063	2827	2062	1833
%	47.9	84.8	2.4	96.3	23.1	3.6	1.9	16.4	15.2	33.2	56.9	9.3	86.4	22.1	1.1	75.7	75.8	91.2	84.1	61.4	54.6

[T] = Number of QuantStudio reactions that were scored as "determined" (row 1) and respective percentage (row 2) for each time point. Determined Ct values range from 16.7 to 33.8. **Yellow cells =** Time points were retained for DEG analysis. Green cells = Time points (in addition to those in yellow cells) used for selection of reference genes. hag: Hours after grafting. dag: Days after grafting

Table S4.3: Genes with at least two biological replicates (each for UCBSV- and mock-inoculated treatments) whose Ct values range between 16.7 and 33.8 in varieties Albert and Namikonga

Target ID (Experiment 1)	Selected for DEG analysis	Namikonga 3dag		Albert 3dag		Namikonga 5dag		Albert 5dag		Namikonga 7dag		Albert 7dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AI0IYB8	Y	✔	✖	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔
AI0IYB9	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI1RWIG	NS	✔	✖	✖	✔	✔	✖	✔	✔	✔	✖	✔	✖
AI1RWIH	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖	✔	✖
AI20UOO	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI20UOP	NS	✖	✔	✔	✔	✖	✔	✖	✖	✖	✖	✖	✖
AI20US1	NS	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI39SUW	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI39SUX	NS	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖
AI39SY9	NS	✖	✖	✖	✖	✖	✖	✖	✖	✖	✔	✖	✖
AI5IQ04	NS	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
AI5IQ05	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI5IQ5H	NS	✖	✖	✔	✔	✔	✖	✖	✖	✖	✖	✖	✖
AI6RO7C	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI6RO7D	NS	✖	✖	✔	✖	✔	✖	✖	✔	✔	✔	✖	✖

Target ID (Experiment 1)	Selected for DEG analysis	Namikonga 3dag		Albert 3dag		Namikonga 5dag		Albert 5dag		Namikonga 7dag		Albert 7dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AI6RPBP	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI70NDK	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI70NDL	Y	✔	✔	✔	✔	✔	✖	✔	✔	✔	✔	✔	✖
AI70NHX	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI89LJS	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AI89LJT	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AIAA0PL	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AIAA0PM	NS	✖	✖	✖	✖	✔	✖	✖	✖	✖	✖	✖	✔
AIBJYVT	Y	✖	✔	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔
AIBJYVU	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AICSW11	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AICSW12	Y	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AID1U79	NS	✔	✖	✔	✖	✔	✔	✔	✖	✖	✖	✔	✖
AID1U8A	NS	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
AIFATEH	NS	✖	✖	✔	✖	✔	✖	✖	✔	✔	✖	✔	✖
AIGJRKP	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

Target ID (Experiment 1)	Selected for DEG analysis	Namikonga 3dag		Albert 3dag		Namikonga 5dag		Albert 5dag		Namikonga 7dag		Albert 7dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AIHSPQX	Y	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖
AII1N3E	NS	✖	✖	✖	✖	✖	✖	✖	✔	✖	✖	✖	✖
AII1NW5	NS	✔	✖	✔	✔	✖	✔	✖	✔	✔	✔	✖	✔
AIKAL3D	NS	✔	✖	✔	✔	✔	✖	✔	✔	✔	✔	✔	✖
AILJJ9L	Y	✔	✔	✔	✔	✔	✖	✔	✔	✔	✔	✔	✖
AIMSIFT	NS	✔	✔	✖	✔	✔	✔	✖	✔	✔	✔	✖	✔
AIN1GL1	NS	✔	✔	✔	✔	✔	✖	✔	✖	✖	✖	✖	✖
AIPAER8	NS	✖	✖	✔	✔	✖	✖	✔	✖	✖	✖	✖	✖
AIPAER9	Y	✔	✔	✔	✔	✔	✔	✔	✖	✔	✔	✔	✖
AIQJCYG	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AIQJCYH	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖
AIRSA4O	NS	✔	✔	✔	✔	✔	✔	✖	✔	✔	✖	✔	✖
AIRSA4P	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AIS09AW	NS	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖	✔	✖
AIS09AX	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AIT97G4	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

Target ID (Experiment 1)	Selected for DEG analysis	Namikonga 3dag		Albert 3dag		Namikonga 5dag		Albert 5dag		Namikonga 7dag		Albert 7dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AIT97G5	NS	✔	✖	✖	✔	✔	✔	✖	✔	✖	✔	✖	✖
AIVI5NC	NS	✖	✖	✔	✖	✖	✖	✔	✖	✖	✖	✖	✖
AIVI5ND	Y	✔	✔	✔	✔	✔	✔	✔	✖	✔	✔	✔	✖
AIWR3TK	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖
AIWR3TL	Y	✔	✖	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
AIX01ZS	Y	✔	✔	✔	✔	✔	✔	✔	✔	✔	✖	✔	✖
AIX01ZT	Y	✔	✔	✔	✔	✔	✖	✔	✔	✔	✔	✔	✔
AIY9Z50	Y	✔	✖	✔	✖	✔	✖	✔	✔	✔	✔	✔	✔
AIY9Z51	NS	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖

Genes with fold change values (✔) and those without fold change values (✖) in Experiment 1. Genes without fold change value had no Ct data (data was undetermined or dropped at filtering stage) in mock- and/or UCBSV-inoculated samples. DEG = differential gene expression, Y = data used for DEG analysis, NS = data not used for DEG analysis

Table S4.4: Genes with at least two biological replicates (each for UCBSV- and mock-inoculated treatments) whose Ct values range between 16.7 and 33.8, in varieties Albert, Kiroba, Namikonga and NDL06/132

Target ID (Experiment 2)	Selected for DEG analysis	Namikonga 1dag		Kiroba 1dag		Mkombozi 1dag		^{N}NDL06/132 1dag		^{A1}Albert 1dag		Namikonga 51dag		Kiroba 51dag		Mkombozi 51dag		NDL06/132 51dag		^{A2}Albert 51dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AI0IYB8	NS	✔	✔	✔	✔	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AI0IYB9	NS	✔	✔	✖	✖	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI1RWIG	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI1RWIH	NS	✖	✖	✔	✔	✔	✔	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✖	✖
AI20UOO	NS	✔	✔	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AI20UOP	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI20US1	NS	✔	✔	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖
AI39SUW	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI39SUX	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI39SY9	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI5IQ04	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI5IQ05	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI5IQ5H	NS	✔	✔	✖	✖	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AI6RO7C	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI6RO7D	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI6RPBP	NS	✔	✔	✔	✔	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AI70NDK	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖

Target ID (Experiment 2)	Selected for DEG analysis	Namikonga 1dag		Kiroba 1dag		Mkombozi 1dag		^{N}NDL06/132 1dag		^{A1}Albert 1dag		Namikonga 51dag		Kiroba 51dag		Mkombozi 51dag		NDL06/132 51dag		^{A2}Albert 51dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AI70NDL	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI70NHX	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI89LJS	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AI89LJT	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIAA0PL	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AIAA0PM	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIBJYVT	NS	✔	✔	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AIBJYVU	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AICSW11	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AICSW12	NS	✖	✖	✖	✖	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AID1U79	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AID1U8A	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIFATEH	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIGJRKP	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIHSPQX	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AII1N3E	NS	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖
AII1NW5	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIKAL3D	NS	✔	✔	✔	✔	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AILJJ9L	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖

Target ID (Experiment 2)	Selected for DEG analysis	Namikonga 1dag		Kiroba 1dag		Mkombozi 1dag		^{N}NDL06/132 1dag		^{A1}Albert 1dag		Namikonga 51dag		Kiroba 51dag		Mkombozi 51dag		NDL06/132 51dag		^{A2}Albert 51dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AIMSIFT	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIN1GL1	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIPAER8	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIPAER9	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIQJCYG	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIQJCYH	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIRSA4O	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIRSA4P	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIS09AW	NS	✔	✔	✔	✔	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AIS09AX	NS	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIT97G4	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔		✖	✖
AIT97G5	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIVI5NC	NS	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✖	✖	✖	✖	✖	✖
AIVI5ND	NS	✔	✔	✔	✔	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AIWR3TK	NS	✔	✔	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AIWR3TL	NS	✔	✔	✔	✔	✔	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIX01ZS	NS	✔	✔	✔	✔	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖
AIX01ZT	Y	✔	✔	✔	✔	✔	✔	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✔	✖	✖
AIY9Z50	NS	✔	✔	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✔	✔	✔	✖	✖	✖

Target ID (Experiment 2)	Selected for DEG analysis	Namikonga 1dag		Kiroba 1dag		Mkombozi 1dag		^{N}NDL06/132 1dag		^{A1}Albert 1dag		Namikonga 51dag		Kiroba 51dag		Mkombozi 51dag		NDL06/132 51dag		^{A2}Albert 51dag	
		UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL	UL	LL
AIY9Z51	NS	✖	✖	✔	✔	✖	✖	✖	✖	✖	✖	✖	✖	✔	✔	✔	✔	✖	✖	✖	✖

Genes with fold change values (✔) and those without fold change values (✖) in Experiment 2. Genes without fold change value had no Ct data (data was undetermined or dropped at filtering stage) in mock- and/or UCBSV-inoculated samples. ^{N} = Ct data only available for UCBSV-inoculated samples at UL and LL, ^{A1}= Ct data only available for mock-inoculated samples at UL and LL, ^{A2} = Ct data only available for mock-inoculated samples at both UL and LL. DEG = differential gene expression, Y = data used for DEG analysis, NS = data not used for DEG analysis

Table S4.5: NormFinder stability values of candidate reference genes and the best NormFinder identified reference gene for Experiments 1 and 2

Gene ID	Stability value	
	Experiment 1	Experiment 2
AIT97G4[1]	**0.108**	0.04
AIBJYVU[2]	0.134	**0.019**
AI20UOO	0.141	NS
AIRSA4P	0.142	0.031
AI0IYB8	0.147	NS
AIX01ZS	0.15	NS
AI70NDL	0.153	0.023
AICSW12	0.158	NS
AI6RPBP	0.165	NS
AI5IQ05	0.172	NS
AICSW11	0.176	0.015
AIS09AX	0.181	NS
AIX01ZT	0.203	0.03
AIY9Z50	0.221	NS
AIQJCYG	0.223	0.03
AIWR3TL	0.229	NS
AIBJYVT	0.235	NS
AIPAER9	0.235	0.026
AIHSPQX	0.245	0.023
AILJJ9L	0.297	0.033
AI70NDK	0.307	NS
AIWR3TK	0.312	NS
AI6RO7C	0.322	NS
AIQJCYH	0.339	NS
AIGJRKP	0.402	NS
AIVI5ND	0.519	NS
AI89LJS	0.523	NS
AIN1GL1	NS	0.017
AI6RO7C	NS	0.017
AI39SUW	NS	0.019

Gene ID	Stability value	
	Experiment 1	Experiment 2
AI89LJT	NS	0.019
AI70NHX	NS	0.019
AI39SY9	NS	0.021
AIGJRKP	NS	0.022
AIMSIFT	NS	0.022
AIT97G5	NS	0.022
AI70NDK[2]	NS	**0.023**
AID1U8A	NS	0.024
AI1RWIG	NS	0.024
AIAA0PM	NS	0.025
AIRSA4O	NS	0.025
AIPAER8	NS	0.031

NS = No data, [1]= Gene with the lowest NormFinder stability value in Experiment 1, [2] = Genes with the lowest combined stability value (best combination of two genes) in Experiment 2. Genes [1] and [2] were selected as the references for DEG analysis in Experiments 1 and 2 respectively.

7.1 Supplementary figures

Fig. S2.1: Standard curve of serially diluted cDNA from UCBSV-infected samples. qPCR was performed using CBSDDR/CBSDDF2 primers. Detection limit (Ct 17.2 – 35.9)

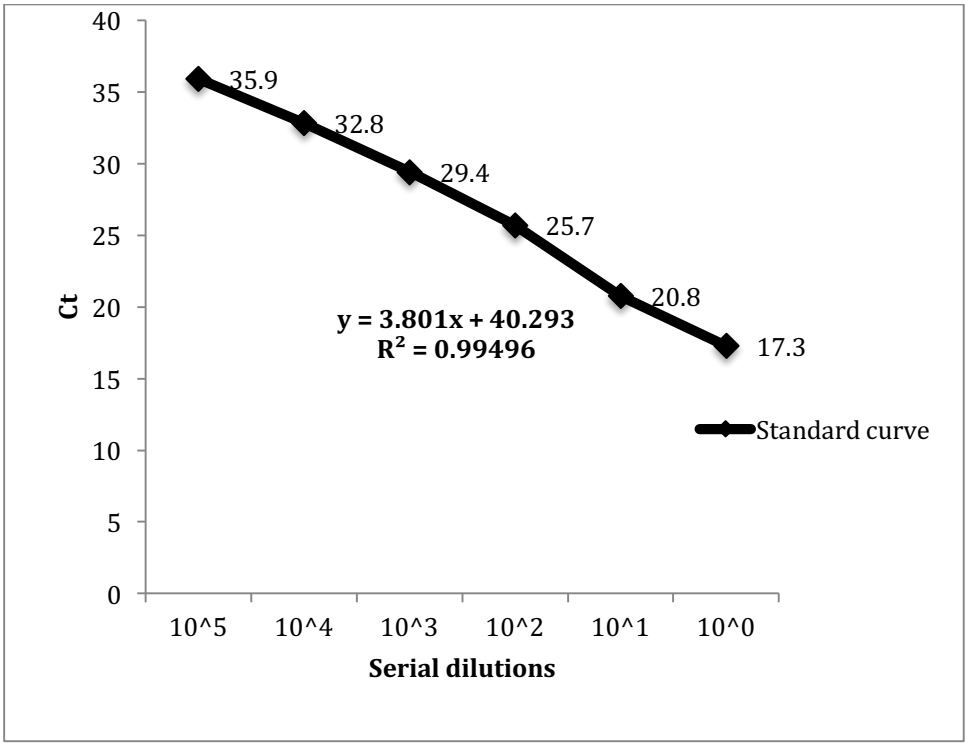


Fig. S3.1: Clustering of samples based on filtered and normalized RNAseq data, using Pearson's correlation model. Abbreviations: a=Albert, I=infected, h= hours, C=control, n=Namikonga.

Fig. S3.1a: Albert

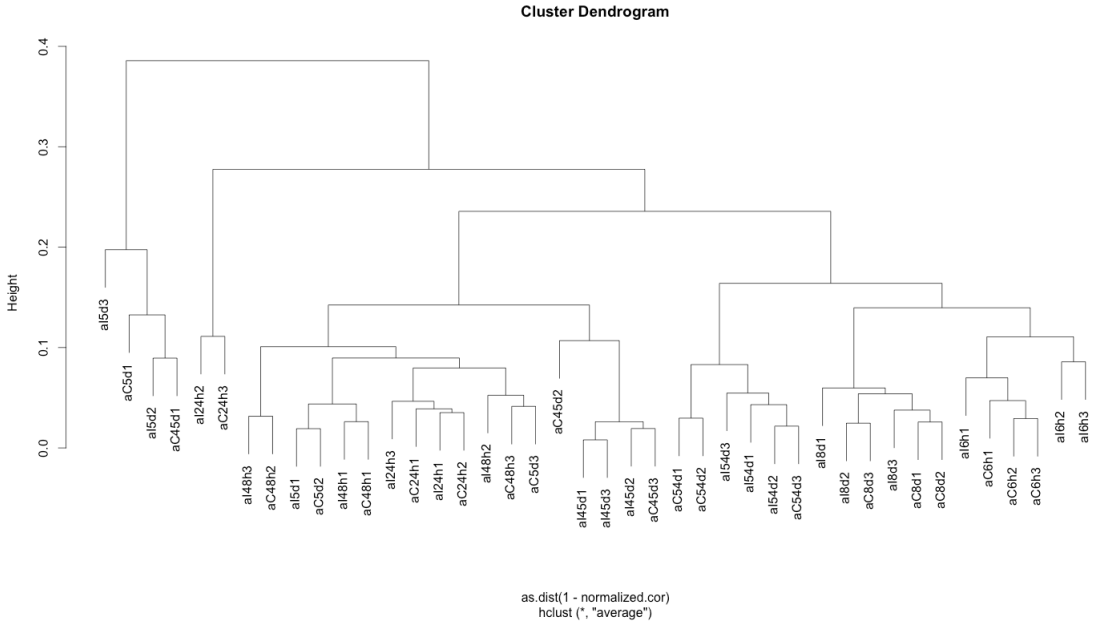


Fig. S3.1b: Namikonga

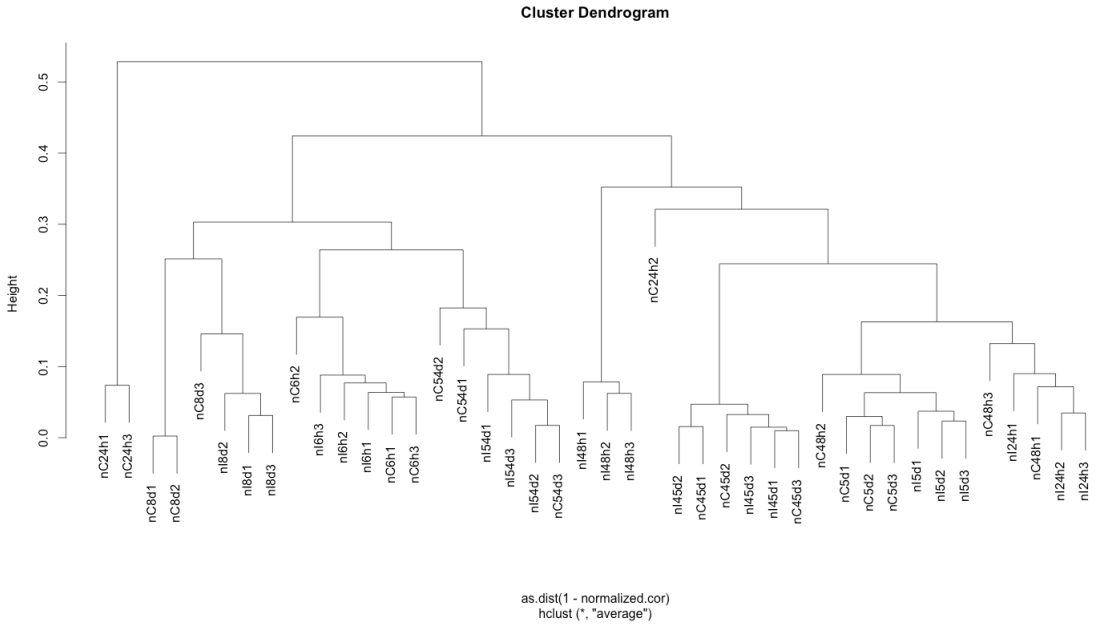


Fig. S3.2: Distribution of DESeq normalized reads for Albert (a) and Namikonga (b) after filtering, plotted using box plots. (Sample abbreviations: a=Albert, I=UCBSV-inoculated, h= hours, C= mock-inoculated, n=Namikonga, numbers 1, 2, 3 represent biological replicates 1, 2 and 3 respectively). Samples were taken at seven time points: 6 hours after grafting (6 hag), one day after grafting (1 dag), 2 dag, 5 dag, 8 dag, 45 dag and 54 dag.

Fig. S3.2a: Albert

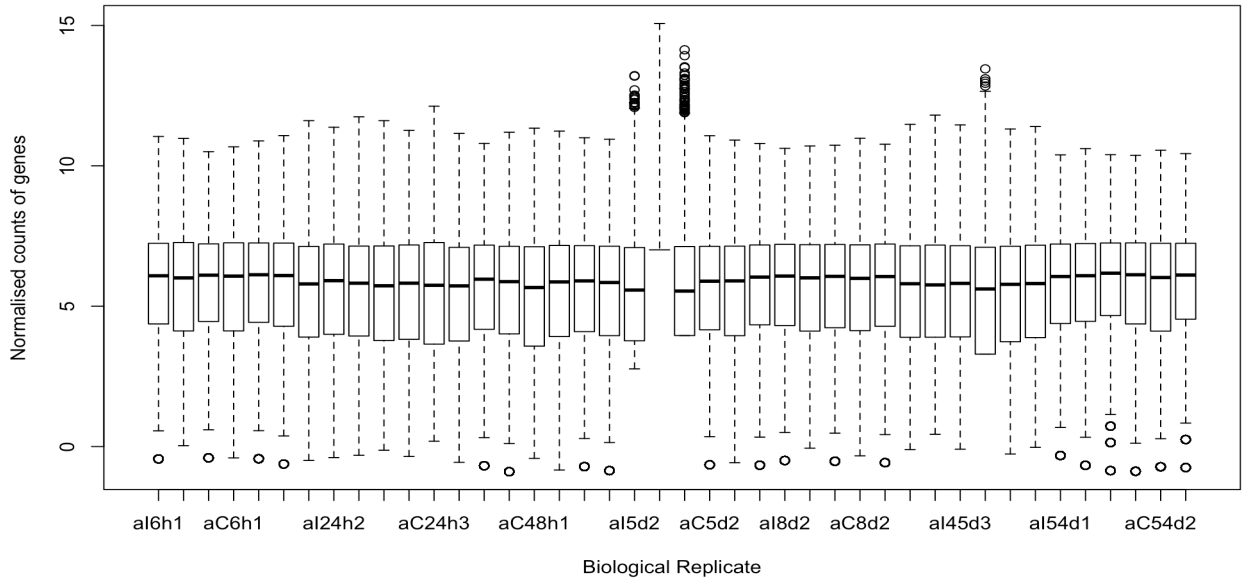


Fig. S3.2b: Namikonga

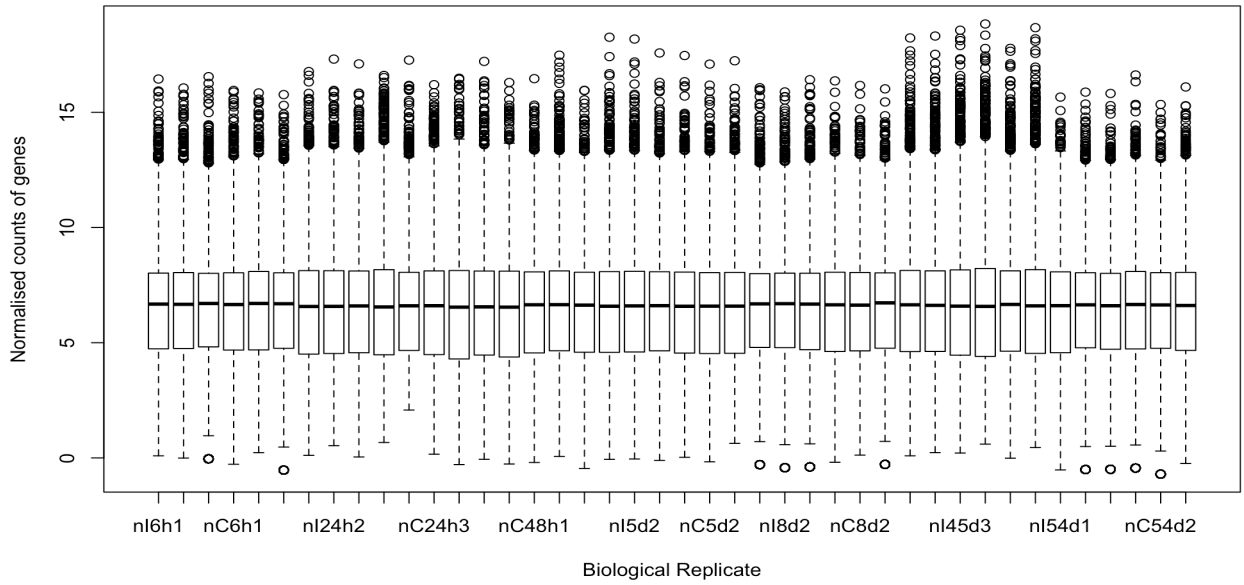
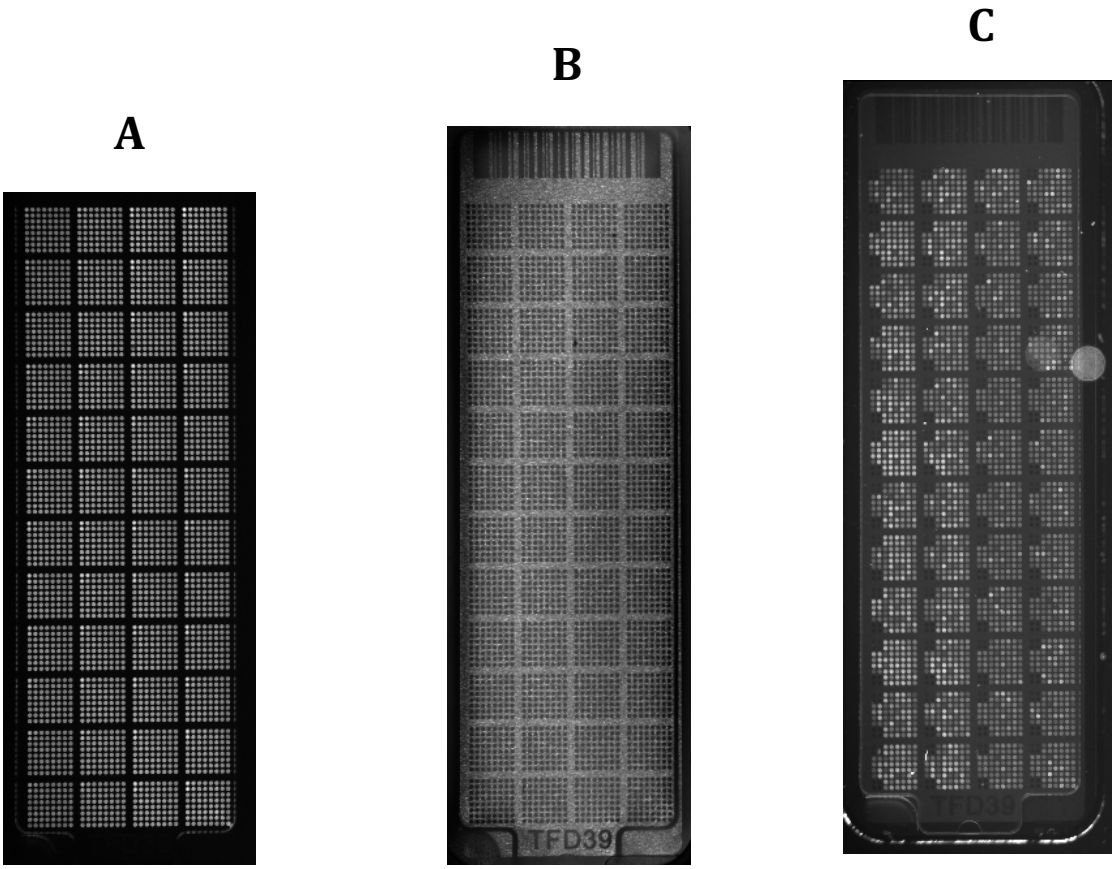


Fig. S4.1. Photographs of QuantStudio array chip to show that the chip was in perfect condition before amplification (A), all wells were loaded with required amount of cDNA (B) but not all wells had amplification (C).



A = before read, B = no leakage, C = Post amplification (showing some wells had no amplification).

8.1 Publication from results presented in this thesis

www.nature.com/scientificreports

SCIENTIFIC REPORTS

OPEN

Received: 7 December 2016
Accepted: 21 July 2017
Published online: 29 August 2017

A time series transcriptome analysis of cassava (*Manihot esculenta* Crantz) varieties challenged with Ugandan cassava brown streak virus

T. Amuge[1,2,3], D. K. Berger[2], M. S. Katari[4], A. A. Myburg[5], S. L. Goldman[4] & M. E. Ferguson[3]

A time-course transcriptome analysis of two cassava varieties that are either resistant or susceptible to cassava brown streak disease (CBSD) was conducted using RNASeq, after graft inoculation with Ugandan cassava brown streak virus (UCBSV). From approximately 1.92 billion short reads, the largest number of differentially expressed genes (DEGs) was obtained in the resistant (Namikonga) variety at 2 days after grafting (dag) (3887 DEGs) and 5 dag (4911 DEGs). At the same time points, several defense response genes (encoding LRR-containing, NBARC-containing, pathogenesis-related, late embryogenesis abundant, selected transcription factors, chaperones, and heat shock proteins) were highly expressed in Namikonga. Also, defense-related GO terms of 'translational elongation', 'translation factor activity', 'ribosomal subunit' and 'phosphorelay signal transduction', were overrepresented in Namikonga at these time points. More reads corresponding to UCBSV sequences were recovered from the susceptible variety (Albert) (733 and 1660 read counts per million (cpm)) at 45 dag and 54 dag compared to Namikonga (10 and 117 cpm respectively). These findings suggest that Namikonga's resistance involves restriction of multiplication of UCBSV within the host. These findings can be used with other sources of evidence to identify candidate genes and biomarkers that would contribute substantially to knowledge-based resistance breeding.

Cassava is among the six major crops of Africa, representing a staple food for >250 million people (FAO, 2010). Currently, production in parts of southern (Mozambique, Malawi and Angola), eastern (Uganda, Kenya, Tanzania, Rwanda and Burundi) and central Africa (D.R. Congo) is seriously affected by cassava brown streak disease (CBSD)[1,2]. At least two virus species cause CBSD: *Cassava brown streak virus* (CBSV) and *Ugandan cassava brown streak virus* (UCBSV)[3,4]. Here CBSVs is used to refer to both of these viruses. First reported in Tanzania, CBSD previously occurred at low levels primarily in coastal East Africa, Mozambique and around Lake Malawi and was thought to be restricted by altitude[5]. However in the early 2000s, CBSD had begun to spread around Lake Victoria, and by 2004, typical CBSD symptoms were widespread in farmers' fields in central Uganda. The disease has spread steadily since then as far as DR Congo and South Sudan and now, together with cassava mosaic disease, causes over US$1 billion losses in production annually in Africa[1,6,7].

Both CBSV and UCBSV are members of genus *Ipomovirus*, family *Potyviridae*[3,4,8]. Both genomes have particles measuring 650 nm with pinwheel inclusions in their cells, distinctive of *Potyviridae*[3]. The CBSV and UCBSV genomes are 9008 nucleotides (nt) and 9070 nt long, respectively[4,8]. Each genome contains a single open reading frame that is translated into 2912 amino acids (aa) for CBSV and 2012 aa for UCBSV. Like other members

[1]National Crops Resources Research Institute (NaCRRI), Namulonge, Uganda. [2]Department of Plant and Soil Sciences, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa. [3]International Institute of Tropical Agriculture (IITA), Nairobi, Kenya. [4]Center for Genomics and Systems Biology, New York University, New York, USA. [5]Genetics Department, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa. Correspondence and requests for materials should be addressed to M.E.F. (email: M.Ferguson@cgiar.org)

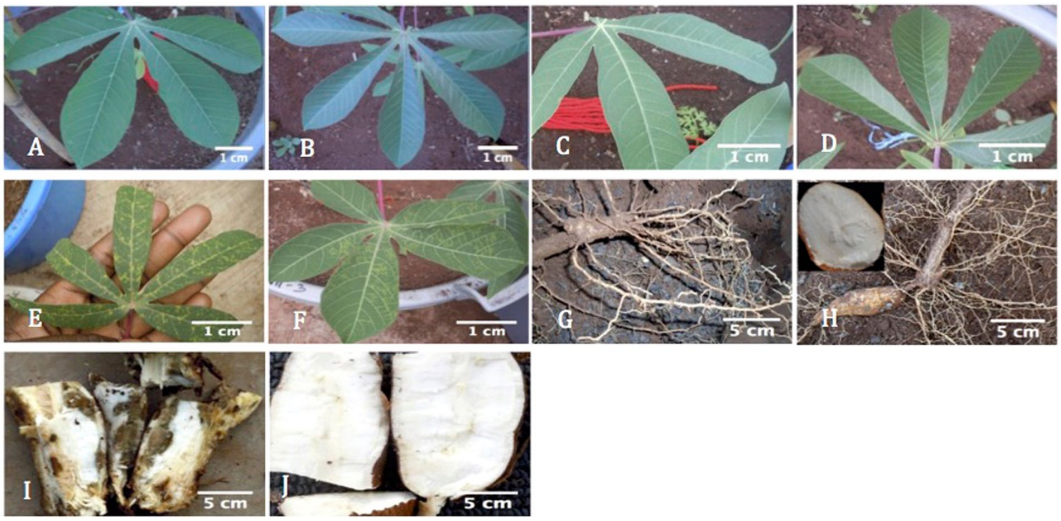


Figure 1. Leaf and root symptoms of UCBSV-inoculated versus mock-inoculated Albert and Namikonga plants during the experiment. (**A**) Albert, UCBSV inoculated, 8 dag; (**B**) Albert, mock inoculated, 8 dag; (**C**) Namikonga, UCBSV inoculated, 8 dag; (**D**) Namikonga, mock inoculated, 8 dag; (**E**) Albert, UCBSV inoculated, 54 dag; (**F**) Namikonga, UCBSV inoculated, 54 dag; (**G**) Albert root, UCBSV inoculated, 3 mag; (**H**) Namikonga root, UCBSV inoculated, 3 mag; (**I**) Albert root, UCBSV inoculated, 12 mag; (**J**) Namikonga root, UCBSV inoculated, 12 mag.

of *Potyviridae*, both genomes encode the proteins P1, P3, 6K1, CI, 6K, VPg, Nla-Pro, Nlb and CP, with a new Ham1-like (Ham1) protein[3,4,8]. Neither genome encodes the helper-component protein (HC-Pro) that is encoded by other members of the family *Potyviridae*[3,4,8].

(U)CBSV is transmitted naturally by *Bemisia tabaci*[5,9,10] whiteflies in a semi-persistent manner, with the spiraling whitefly (*Aleurodicus dispersus*) being a possible alternative vector[11,12]. The main form of disease spread in cassava fields in Africa is thought to be through virus-positive stem cuttings in this clonally propagated crop[13]. Mechanical transmission[14,15] and graft inoculation[9,16] are used in research studies.

Infected susceptible plants develop chlorosis along leaf veins, brown streaks on stems, and root necrosis (Fig. 1), with severe infection causing shoot dieback. Dual infection with both virus species is common in farmers' fields, although there are no reports of synergistic virus interaction[3,4,8,17], and both viruses cause similar symptoms, although those of CBSV tend to be more severe than UCBSV[18,19].

Currently, no known cassava variety has been reported to be immune to CBSVs, but varied levels of resistance or tolerance have been identified. Here, resistance is defined as the ability of the cassava variety to maintain a low virus load and show minimal shoot symptoms coupled with little or no root necrosis at harvest. Using this criteria, Namikonga (also known as Kaleso) was identified as resistant to CBSV from greenhouse experiments[20] and classified as tolerant to CBSVs based on field symptoms and virus load in Uganda[21]. Cassava breeding is a lengthy process, and disease response is influenced by genotype-by-environment interactions[22]. To shorten the breeding cycle and improve the accuracy of variety selection, breeders are implementing genomics-based approaches. An understanding of the resistance mechanisms involved, including biochemical pathways, and the identification of candidate genes and biomarkers would contribute substantially to knowledge-based genomics breeding, including marker-assisted selection (MAS).

The use of RNAseq (RNA sequencing) has enabled the high-throughput identification of new genes, exons and exon junctions, splice variants and promoter regions in sequenced transcriptomes[23–25]. Previously, in cassava, an RNAseq-based transcriptome analysis of CBSD-resistant and -susceptible cassava varieties infected with CBSV was conducted to identify genes putatively involved in disease resistance; however, the results were inconclusive[20].

In potyviruses, recessive resistance is known to involve >200 defense genes[26–28]. The majority of the cloned genes involve eukaryotic initiation factor (eIF) deployment[28–33]; however, other mechanisms involving defense response genes have been reported[34]. The eIF-mediated mechanism of resistance, also called 'passive'[28] or 'loss-of-susceptibility'[35] resistance, occurs when the host plant has a modified ribosomal protein sub-structure (usually eIF4E or eIF(iso)4E) so that the viral genome-linked (VPg) protein cap is unable to bind to the ribosome, thus preventing viral replication. An alternative mechanism of resistance to potyviruses was first identified in Arabidopsis plants infected with *Tobacco etch virus* (TEV) when a set of genes restricting the long-distance movement of TEV were cloned. These genes were named restricted TEV movement (RTM) genes. The RTM genes have since been found to cause resistance against other potyviruses including *Lettuce mosaic virus* (LMV) and *Plum pox virus* (PPV) by restricting the long-distance movement of virus particles[36]. Dominant resistance against potyviruses has been observed in pepper against *Potato virus Y* (PVY)[37] and in *Solanum lycopersicum* against TEV and *Pepper mottle virus* (PMV)[29]. This dominant resistance may be characterized by a hypersensitive response

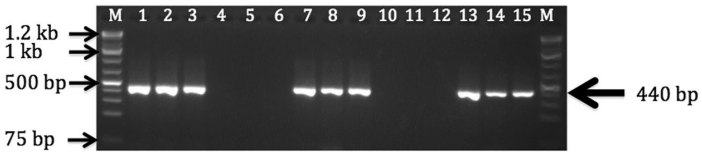


Figure 2. Diagnostic RT-PCR for UCBSV in the inoculum source NDL06/132, Albert and Namikonga plants at 3 mag. RNA from cassava plants from the UCBSV-inoculation experiment were amplified using the UCBSV diagnostic primers CBSDDF2/CBSDDR[54], and the RT-PCR products were separated by agarose gel electrophoresis (2%). Lanes 1–3: UCBSV inoculum source (variety NDL06/132); lanes 4–6: Albert, mock grafted with virus-negative scions of NLD06/132; lanes 7–9: Albert, UCBSV inoculated; lanes 10–12: Namikonga, mock grafted with virus-negative scions of NLD06/132; lanes 13–15: Namikonga, UCBSV inoculated and M: 1 Kb+ ladder.

(HR), as in tobacco resistant to PMV[38] and *Tobacco mosaic virus* (TMV)[39] and in potato resistant to both PVY and TEV[40]. The identification of potyvirus defense genes, such as those described above, motivated this study, as both CBSVs are potyviruses affecting cassava, a major food crop in sub-Saharan Africa.

The greater goal of this study was to define biomarkers for genomics-based breeding of CBSD-resistant cassava varieties to help address food insecurity in Africa. To improve our understanding of the mechanism of resistance and potentially identify candidate genes involved in resistance to UCBSV, DEGs were determined from a time-course experiment involving UCBSV-inoculated and mock-inoculated plants of two cassava varieties with contrasting responses to UCBSV infection, Albert and Namikonga. In addition, we tested the hypothesis that UCBSV accumulates at significantly lower rates in Namikonga compared to Albert because the induction of defense genes in Namikonga restricts virus replication[20,21].

Results

Symptoms of CBSD in leaves and roots. Characteristic CBSD symptoms were observed in UCBSV-inoculated plants and varied in magnitude by variety and time of observation. In the early sampling phase, no visible symptoms were observed on the leaves of either UCBSV-inoculated or mock-inoculated plants of both varieties (Fig. 1). In Albert, at the late sampling phase (54 dag), young leaves of UCBSV-inoculated plants showed chlorotic patterns along veins, expanding to form very large yellow areas (Fig. 1). At 12 months after grafting (mag), >50% of storage roots from UCBSV-inoculated Albert plants were necrotic, with a severity score of 4 (Fig. 1). Control Albert plants showed no disease symptoms on leaves and roots at 12 mag.

For Namikonga, leaf chlorotic spots (Fig. 1) were observed in the late sampling phase. At 54 dag, chlorosis covering 2–3 leaves per plant (score 2) was observed. At 12 mag, all storage roots were non-necrotic. The storage roots from mock-inoculated plants were also non-necrotic at 12 mag.

Detection of UCBSV in UCBSV-inoculated cassava using RT-PCR and RNAseq. In the early phase, all samples (UCBSV- and mock-inoculated) were negative (non-detectable) for UCBSV except at 6 dag, when one of the ten UCBSV-inoculated plants of both the Albert and Namikonga varieties were positive for UCBSV using the RT-PCR assay (data not shown). Seven out of ten UCBSV-inoculated Albert plants were consistently positive at the late time phase (data not shown). For the Namikonga variety, seven out of ten UCBSV-inoculated plants tested positive for at least one sampling point during the late time phase (data not shown). However, this was not as consistent as the results from the Albert variety.

To confirm the presence of UCBSV in UCBSV-inoculated plants, leaves were sampled at 3 mag, after sampling for RNAseq at early and late time phases. End-point RT-PCR with the UCBSV-specific primers CBSDDF2/CBSDDR was performed on these 3 mag samples of Albert and Namikonga and 7 of the 10 plants of each variety tested were positive for UCBSV, having a 440-bp RT-PCR fragment when resolved on a 2% agarose gel (Fig. 2). Mock-inoculated plants had no visible RT-PCR fragment on the same gel, confirming the absence of UCBSV. An RT-PCR fragment putatively diagnostic for UCBSV (approximately 440 nt) was gel purified and Sanger sequenced before being aligned to the full UCBSV genome sequences from GenBank. The best alignment (98% identity, 100% query cover, E value = 3e179) was obtained with accession KF878103.1 (annotated as UCBSV). The same RT-PCR fragment size was amplified from the UCBSV inoculum source (variety NDL06/132), which indicates that the same virus species was present in the inoculum and UCBSV-inoculated plants (Fig. 2). Ten plants of each variety and each treatment were confirmed to be negative for UCBSV by RT-PCR prior to grafting. At 3 mag, after samples for RNAseq had been taken and frozen from all 40 plants at all time points, RT-PCR was conducted to determine whether they were UCBSV positive or not. Sanger sequencing of the RT-PCR products was performed to confirm the identity of UCBSV. Three of the UCBS-positive plants of each variety were randomly selected along with three mock-inoculated plants of each variety and used as biological replicates for the RNAseq study.

The number of UCBSV reads detected in each RNAseq sample at each time point is given in Supplementary Data S1. UCBSV sequences were detected in the inoculated susceptible variety Albert at 45 dag and 54 dag (late sampling phase). At 45 dag, 733, 507 and 37 UCBSV read counts per million (cpm) were detected from each of the three UCBSV-inoculated biological replicates of Albert (Fig. 3). At 54 dag, the same samples had 1660, 940 and 80 UCBSV read cpm. Surprisingly, one of the three mock-inoculated biological replicates of Albert had six UCBSV reads at 45 dag, and at 54 dag, two replicates had two UCBSV reads each, and one replicate had a single

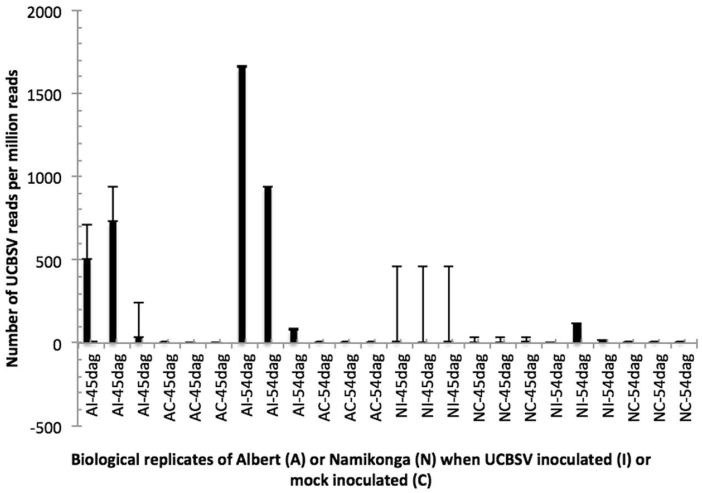


Figure 3. Number of UCBSV reads per million RNAseq reads retrieved from unmapped RNAseq reads of UCBSV inoculated (**I**) and control (**C**) samples from Albert (**A**) and Namikonga (**N**) varieties. In both varieties, UCBSV reads were only detected at 45 dag and 54 dag.

read. The surprising occurrence of UCBSV reads in mock-inoculated plants could be explained by contamination or the fact that the virus-negative plants host minute levels of UCBSV, which are not detectable by routine RT-PCR but are detectable with deep sequencing, such as RNAseq.

In Namikonga, very few UCBSV reads were detected in UCBSV-inoculated samples. The highest count (117 UCBSV sequences) was recorded in one biological replicate at 54 dag (Supplementary Data S1). This is 14 times lower than the highest count recorded in Albert at the same time point. The rest of the UCBSV-inoculated Namikonga samples had $\leq$10 UCBSV reads detected at any time point. For Namikonga, at 45 dag, one replicate from a mock-inoculated plant had one UCBSV sequence; the other two samples from mock-inoculated Namikonga plants had no detectable UCBSV sequence. The same was true for Namikonga at 54 dag.

Read depth and mapping to the cassava reference genome (v4.1). A total of approximately 1.92 billion raw reads were sequenced from 96 cDNA Namikonga and Albert libraries at eight time points with a coverage of approximately 20 million reads per library (Supplementary Table S1; Supplementary Data S2). All sequences have been deposited in NCBI's Sequence Read Archive (SRA) SRP with BioProject ID PRJNA360340.

After sequence quality control (QC), including filtering based on Phred score and trimming, reads from each library were mapped to the cassava reference genome using Tophat2[41]. The cassava reference genome (version 4.1) is derived from a partial inbred line, AM560-2, a third generation self of the Latin American variety MCOL-1505[42]. Albert had 29,755,234,095 mapped reads and 8,231,859,633 unmapped reads, while Namikonga had 32,490,332,800 mapped and 8,477,326,725 unmapped reads (Supplementary Data S3). Therefore, Albert had 78% of its reads mapped and 22% unmapped to the cassava reference genome, while Namikonga had 79% mapped and 21% unmapped reads (Table 1). The cassava reference genome is derived from a partial inbred line, AM560-2, a third generation self of the Latin American variety MCOL-1505[42]. Supplementary Table S2 contains a summary of the per variety statistics after QC. Altogether, reads from 40 out of the 48 libraries sequenced at Dow AgroSciences (101 bp) (Supplementary Table S2b) mapped at least 84% to the cassava reference genome. Four samples mapped at 62.3–78.7%, while the remaining four mapped at 24–36.3%. Among the 50-bp libraries sequenced at UC Berkeley (Supplementary Table S2a), one sample mapped poorly (3.9%), another was 59.4% mapped, ten samples mapped between 74.1–79.7% and the remaining 36 libraries were >80% mapped to the reference genome. Out of the 96 sequenced cDNA libraries, all libraries that mapped poorly (<40%) also failed to cluster with respective biological replicates (see section '*Clustering of samples using data from RNAseq reads mapped to cassava genome*'). These outlier libraries were removed from subsequent analyses.

HTseq-counts[43] was used to count the number of reads that aligned to gene models of the cassava reference genome (v4.1). More than 22,000 out of 33,000 genes in the cassava reference genome v4.1 were represented by sequence information from either variety. Supplementary Data S4 summarizes the reads that could not be counted by HT-seq (i.e. reads that did not map to any gene models in the reference genome or reads that mapped to more than one gene model).

Clustering of samples using data from RNAseq reads mapped to the cassava genome. Using Pearson's correlation, the normalized reads of samples from both Albert and Namikonga clustered by variety (Supplementary Figure S1). Respective biological replicates clustered together by time point, variety and treatment (UCBSV inoculated and mock inoculated), irrespective of the laboratory where samples had been sequenced. This was expected, as RNAseq samples sequenced from different laboratories are comparable provided that recommended

Time point	Mapping to cassava genome	Albert UCBSV inoculated	Albert Mock inoculated	Namikonga UCBSV inoculated	Namikonga Mock inoculated
Time zero	Mapped	—	4,218,021,147	—	3,996,224,040
	Unmapped	—	835,908,591	—	929,195,862
6 hag	Mapped	1,974,374,444	2,311,923,325	1,739,750,996	2,034,621,580
	Unmapped	465,562,984	558,691,687	373,228,177	416,264,064
1 dag	Mapped	1,782,476,393	1,606,541,112	1,751,611,492	1,415,764,268
	Unmapped	469,842,785	281,839,135	380,522,348	892,832,192
2 dag	Mapped	2,311,174,776	2,258,285,184	2,278,828,029	2,195,695,847
	Unmapped	425,576,288	411,150,949	494,433,874	648,333,549
5 dag	Mapped	1,152,362,044	900,243,552	2,223,072,359	1,809,373,734
	Unmapped	900,243,552	988,890,194	402,279,032	713,771,063
8 dag	Mapped	1,845,319,525	1,948,946,548	2,265,995,556	1,756,458,519
	Unmapped	352,441,534	373,483,677	493,741,019	547,890,775
45 dag	Mapped	1,701,360,044	1,258,543,240	2,269,455,275	2,155,716,341
	Unmapped	365,911,422	831,147,304	505,056,145	476,811,865
54 dag	Mapped	2,089,595,204	2,396,067,557	2,288,557,815	2,309,206,949
	Unmapped	473,446,221	497,723,310	545,859,978	657,106,782
Total	**Mapped**	**12,856,662,430**	**16,898,571,665**	**14,817,271,522**	**17,673,061,278**
	Unmapped	**3,453,024,786**	**4,778,834,847**	**3,195,120,573**	**5,282,206,152**

Table 1. Total statistics for RNAseq reads from cassava varieties Albert and Namikonga sampled at eight time points after mock or graft inoculation with UCBSV. The time points were time zero (before graft inoculation), 6 hag, 1 dag, 2 dag, 5 dag, 8 dag, 45 dag and 54 dag. The figures reflect the number of RNAseq reads that mapped or did not map (unmapped) to the cassava reference genome v4.1[42].

laboratory procedures are followed and sequence reads are filtered appropriately[44]. However, ten of 96 samples were outliers, and these were removed from further analyses (Supplementary Table S3). Seven of these removed outlier samples were from Albert (1dag_Alb_Inf_2, 5dag_Alb_Inf_1, 1dag_Alb_Ctl_3, 2dag_Alb_Ctl_3, 5dag_Alb_Ctl_2, 8dag_Alb_Ctl_1 and 45_dag_Alb_Ctl_1) (Supplementary Figure S1a), and three were from Namikonga (2dag_Nam_Ctl_2, 54dag_Nam_Ctl_3 and 6hag_Nam_Ctl_2) (Supplementary Figure S1b).

In Albert, once the outlier samples were removed and reads re-normalized, the median gene expression values were comparable across all time points (Supplementary Figure S2a). The treatments had a distinct median range of filtered, normalized reads cutting across UCBSV-inoculated and mock-inoculated treatments (approximately 5.8–6.2). The median of the gene expression values in the 1 dag and 45 dag samples was slightly lower (5.8–6.0) compared to other time points (6.0–6.2).

For Namikonga, the median of the re-normalized gene expression values was within the same range (approx. 6.5) (Supplementary Figure S2b), as was the number of outlier genes above the upper quartile range. A slightly higher number of outlier genes were recorded at 45 dag, and the reverse at 54 dag. Samples with medians below the lower quartile were only observed at 6 hag (one UCBSV-inoculated and one mock-inoculated sample), 8 dag (three UCBSV-inoculated and one mock-inoculated sample) and 54 dag (two UCBSV-inoculated and mock-inoculated samples). After checking the distribution of the data, DESeq[45] was used to identify DEGs (Supplementary Data S5) using at least two biological replicates per treatment at each time point (Supplementary Table S3).

Identification of DEGs between mock- and UCBSV-inoculated samples of susceptible (Albert) and resistant (Namikonga) varieties at different time points. To identify DEGs, UCBSV-inoculated samples were compared with mock-inoculated samples at each time point, per variety. We defined DEGs as those that were computed by DESeq to have a false discovery rate (FDR) for differential expression of less than 10% for a particular treatment comparison. We did not apply an additional filter of log2 fold change (log2FC) values, but DEGs with positive or negative log2FC values were classified as up-regulated ($\log2FC \geq 0$), or down-regulated ($\log2FC \leq 0$) for a treatment, respectively. In total, more genes were differentially expressed at particular time points in Namikonga (10,028) compared to Albert (688) (Fig. 4). It is unlikely that this is due to sequence differences between these varieties and the reference genome, since both had a similar proportion of unmapped reads (approximately 21%) (Table 1). For Namikonga, there were approximately equal numbers of up- and down-regulated genes at all time points, except 1 dag, when only 26% of the DEGs were up-regulated. For Albert, more than 75% of the DEGs were up-regulated at 1 dag and 2 dag, whereas there were more down-regulated genes at the other time points.

In Namikonga, the greatest differential gene expression (either down-regulated or up-regulated) occurred at the early time points, particularly 2 dag (3887 DEGs), 5 dag (4911 DEGs) and 8 dag (1438 DEGs) (Fig. 4). The number of DEGs in Albert in the early time phase was much lower than in Namikonga, with the highest being at 6 hag with only 92 DEGs. At the late time points, few genes were differentially expressed in either variety. Albert had 543 DEGs at 45 dag (which was the maximum for Albert) and none at 54 dag. Namikonga had no DEGs at 45 dag but 738 at 54 dag (Fig. 4). All statistically significant DEGs were further characterized by functional annotation and GO term enrichment.

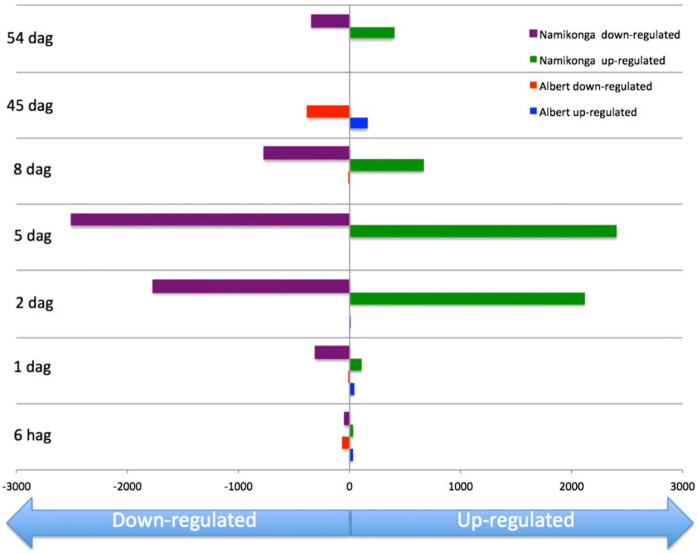


Figure 4. Number of DEGs between UCBSV-inoculated and mock-inoculated cassava varieties Albert and Namikonga at 6 hag, 1 dag, 2 dag, 5 dag, 8 dag, 45 dag and 54 dag. The DEGs were identified using the DESeq method[45].

GO term enrichment of DEGs. Over-represented GO terms were identified among the up-regulated and down-regulated genes separately at each time point and in each variety. The numbers of enriched GO terms for each treatment are presented in Supplementary Table S4. Namikonga had the largest number of DEGS at 2 dag (3887 DEGs) and 5 dag (4911 DEGs), which also corresponded to the largest number of over-represented GO terms (Supplementary Table S4). For this reason, these time points were the focus for further analysis, particularly the identification of enriched defense-related GO terms.

Enriched GO terms in Namikonga that are related to defense responses. Among the over-represented GO terms of Namikonga identified at 2 dag and 5 dag (Fig. 5), the following terms that are likely to be related to pathogen defense[33,46,47] were identified: translational elongation (GO:0006414), translation factor activity, nucleic acid binding (GO:0008135), ribosomal subunit (GO:0044391) and phosphorelay signal transduction (GO:0000160 and GO:0000156) were each represented by 12, 19, 13 and 18 (17 in GO:0000156) DEGs, respectively (Table 2). None of these defense-related GO terms were over-represented in Albert. Individual genes with these GO terms were significantly differentially expressed in Namikonga, but not in Albert. The expression of genes with these defense-related GO terms was further examined.

Translation initiation in the ribosomal subunit. The role of eIF4E genes in Potyvirus resistance has been demonstrated in many plants[28]. We identified two eIF4E genes (Cassava4.1_016620 m.g and Cassava4.1_013223 m.g) for which reads were mapped in our dataset. Neither gene was significantly differentially expressed in our experiment, except for Cassava4.1_016620 m.g in Namikonga at a single time point (5 dag) (log2FC = 0.72; adj. p-value = 0.02) (Supplementary Table S5). The eIF4E gene Cassava4.1_013223 m.g[20] had a log2FC of −0.17 and 0.63 at 2 dag and 5 dag in Namikonga. This gene was up-regulated by a log2FC of 0.1 and 0.4 in Albert at 2 dag and 5 dag, which was not statistically significant.

Although the data did not show transcriptional differences in eIF4E, there were three main enriched GO terms (GO:0006414, GO:0008135 and GO:0044391) associated with translation in Namikonga at the early time points. This indicates, in general, the potential role(s) of translation in CBSD resistance in Namikonga, and therefore individual genes are reported here. Within these GO terms, Namikonga had several families of translation factors, including eIF4E, EF1B, EF-Ts, 5A-1, eIF3 subunit 7, eIF2 subunit 1, 3B1, and IF2/1F5, each having one or two genes identified at 2 dag and/or 5 dag (Table 3). The two families of EF-T genes (5A-1 and GTP binding EF-Tu) had a log2FC $\leq$ 1 in both varieties, yet were statistically significant and up-regulated in Namikonga, in contrast to Albert. The EF1B gene was up-regulated in Namikonga, but the results were not significant in either variety. Within the same three GO terms, genes encoding ribosomal protein L10 and 60 S acidic ribosomal proteins were statistically significantly up-regulated in Namikonga, but non-significantly down-regulated in Albert (Table 3).

Phosphorelay signal transduction and response regulation. There were two GO terms associated with phosphorelay signal transduction: phosphorelay signal transduction system (GO:0000160) and phosphorelay response regulator activity (GO:0000156). The first GO term contained 18 genes that encoded histidine-related proteins and response regulators. All 18 genes were up-regulated in Namikonga (log2FC of 0.8–3.5), but in Albert, 12 genes

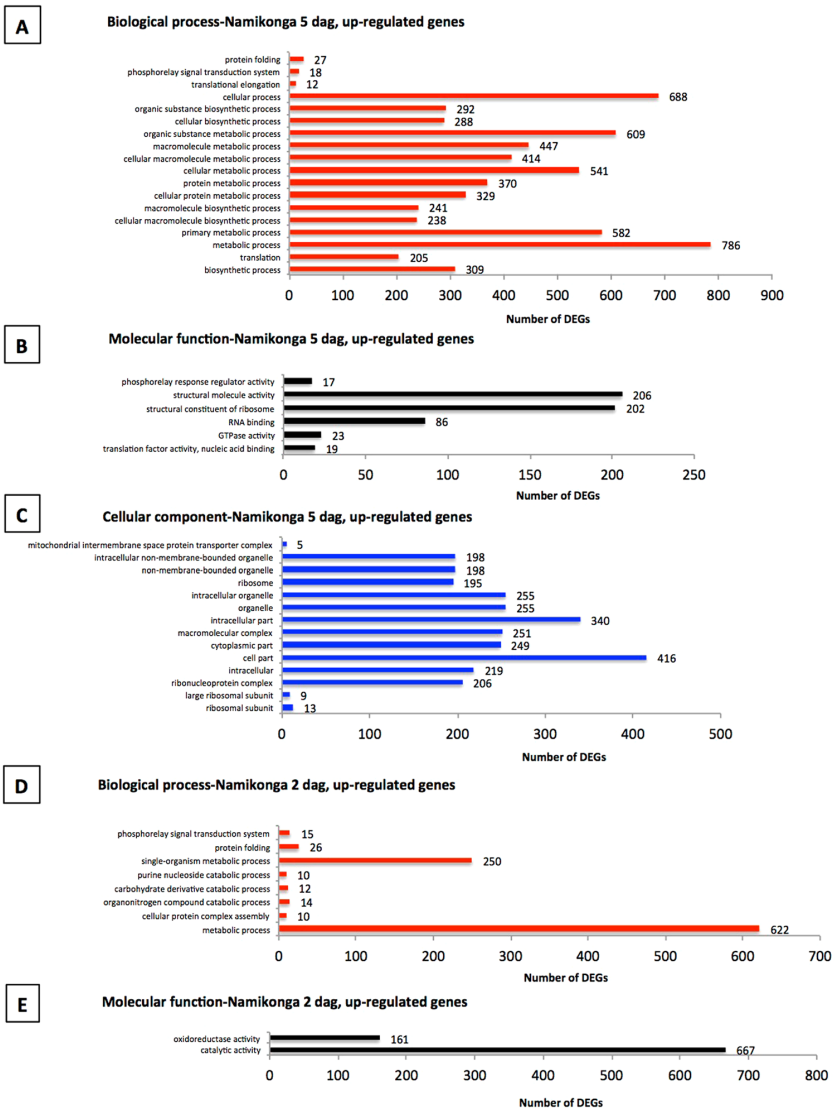


Figure 5. Over-represented GO terms for up-regulated genes in Namikonga at 2 dag and 5 dag. **A** = Over-represented GO terms of the category "biological process" in the Namikonga up-regulated genes at 5 dag; **B** = Over-represented GO terms of the category "molecular function" in the Namikonga up-regulated genes at 5 dag; **C** = Over-represented GO terms of the category "cellular component" in the Namikonga up-regulated genes at 5 dag; **D** = Over-represented GO terms of the category "biological process" in the Namikonga up-regulated genes at 2 dag; **E** = Over-represented GO terms of the category "molecular function" in the Namikonga up-regulated genes at 2 dag. Results are presented for over-represented GO terms with adjusted p-values in the 95% confidence interval.

were down-regulated, while the other six were expressed at very low levels (log2FC of 0.0–0.8). The other GO term, phosphorelay response regulator activity, had 17 genes that were expressed in the same pattern as the first GO term (GO:0000160) (Table 4).

Specific defense response genes are highly expressed in Namikonga at specific time points. Expression patterns (shown by log2FC) of 55 manually selected defense response genes were examined over the time course (6 hag, 1 dag, 2 dag, 5 dag, 8 dag, 45 dag and 54 dag) in Albert and Namikonga (Fig. 6). The genes were manually selected based on earlier studies in which these genes were shown to function in pathogen defense. The genes included those with leucine-rich repeats (LRRs) (Fig. 6A and B), those that have a nucleotide binding domain (NBARC) (Fig. 6C and D), pathogenesis-related (PR) proteins (Fig. 6E and F)[48], late embryogenesis

GO ID	Enriched GO terms found within the up-regulated genes of Namikonga at 2 dag	No. of DEGs	Adj. p-value
	Biological process		
GO:0043623	Cellular protein complex assembly	10	2.00E-02
GO:0006152	Purine nucleoside catabolic process	10	5.E-03
GO:1901136	Carbohydrate derivative catabolic process	12	2.E-02
GO:1901565	Organonitrogen compound catabolic process	14	1.E-02
GO:0000160	**Phosphorelay signal transduction system**	**15**	**2.E-02**
GO:0006457	Protein folding	26	5.E-03
GO:0044710	Single-organism metabolic process	250	1.E-02
GO:0008152	Metabolic process	622	1.E-02
	Molecular function		
GO:0016491	Oxidoreductase activity	161	3.E-02
GO:0003824	Catalytic activity	667	9.E-04
	Enriched GO terms found within the up-regulated genes of Namikonga at 5 dag		
	Biological process		
GO:0006414	**Translational elongation**	**12**	**3.E-03**
GO:0000160	**Phosphorelay signal transduction system**	**18**	**1.E-02**
GO:0006457	Protein folding	27	3.E-02
GO:0006412	Translation	205	3.E-68
GO:0034645	Cellular macromolecule biosynthetic process	238	1.E-59
GO:0009059	Macromolecule biosynthetic process	241	5.E-56
GO:0044249	Cellular biosynthetic process	288	3.E-43
GO:1901576	Organic substance biosynthetic process	292	1.E-42
GO:0009058	Biosynthetic process	309	5.E-40
GO:0044267	Cellular protein metabolic process	329	2.E-15
GO:0019538	Protein metabolic process	370	1.E-10
GO:0044260	Cellular macromolecule metabolic process	414	4.E-15
GO:0043170	Macromolecule metabolic process	447	3.E-11
GO:0044237	Cellular metabolic process	541	4.E-18
GO:0044238	Primary metabolic process	582	2.E-10
GO:0071704	Organic substance metabolic process	609	3.E-11
GO:0009987	Cellular process	688	6.E-13
GO:0008152	Metabolic process	786	1.E-04
	Molecular function		
GO:0000156	**Phosphorelay response regulator activity**	**17**	**5.E-03**
GO:0008135	**Translation factor activity, nucleic acid binding**	**19**	**1.E-02**
GO:0003924	Gtpase activity	23	5.E-03
GO:0003723	RNA binding	86	4.E-12
GO:0003735	Structural constituent of ribosome	202	3.E-78
GO:0005198	Structural molecule activity	206	3.E-73
	Cellular component		
GO:0042719	Mitochondrial intermembrane space protein transporter complex	5	3.E-02
GO:0015934	Large ribosomal subunit	9	4.E-03
GO:0044391	**Ribosomal subunit**	**13**	**3.E-04**
GO:0005840	Ribosome	195	4.E-63
GO:0043228	Non-membrane-bounded organelle	198	9.E-56
GO:0030529	Ribonucleoprotein complex	206	3.E-64
GO:0005622	Intracellular	219	3.E-19
GO:0044444	Cytoplasmic part	249	1.E-47
GO:0032991	Macromolecular complex	251	9.E-28
GO:0043226	Organelle	255	2.E-21
GO:0044424	Intracellular part	340	2.E-14
GO:0044464	Cell part	416	9.E-09

Table 2. GO term over-representation analysis of the up-regulated genes of Namikonga at 2 dag and 5 dag.

		2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
Gene ID	Gene annotation	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
GO term: Translational elongation (GO:0006414)					
Cassava4.1_015319 m.g	Translation elongation factor EF1B/ribosomal protein S6 family protein	0.0(1)	−0.7(1.3E-01)	0.2(1)	0.6(6.8E-02)
Cassava4.1_010349 m.g	Translation elongation factor Ts (EF-Ts), putative	−0.4(1)	−0.5(5.7E-01)	0.5(1)	1.0(1.3E-02)
Cassava4.1_018059 m.g	Eukaryotic elongation factor 5A-1	0.0(1)	0.5(3.1E-01)	−0.9(1)	0.9(6.0E-04)
Cassava4.1_018052 m.g	Eukaryotic elongation factor 5A-1	−0.2(1)	0.2(8.1E-01)	−0.1(1)	0.9(5.5E-03)
Cassava4.1_007378 m.g	GTP binding Elongation factor Tu family protein	−0.2(1)	0.6(3.1E-01)	−0.3(1)	1.7(4.5E-05)
Cassava4.1_011934 m.g	Ribosomal protein L10 family protein	−0.1(1)	−0.2(8.7E-01)	−0.4(1)	0.8(7.5E-03)
Cassava4.1_019599 m.g	60 S acidic ribosomal protein family	0.0(1)	−0.5(4.4E-01)	−0.6(1)	1.9(6.7E-10)
Cassava4.1_019305 m.g	60 S acidic ribosomal protein family	0.0(1)	−0.1(1.0E + 00)	−0.7(1)	0.9(3.7E-03)
Cassava4.1_020132 m.g	60 S acidic ribosomal protein family	0.1(1)	−0.2(8.2E-01)	−0.9(1)	0.7(1.5E-02)
Cassava4.1_019568 m.g	60 S acidic ribosomal protein family	−0.1(1)	−0.2(8.9E-01)	0(1)	0.6(4.4E-02)
Cassava4.1_019588 m.g	60 S acidic ribosomal protein family	0.0(1)	−0.1(9.4E-01)	−0.5(1)	0.8(4.2E-03)
Cassava4.1_018344 m.g	60 S acidic ribosomal protein family	0.0(1)	0.2(7.2E-01)	−0.7(1)	1.0(5.3E-04)
GO term: Translation factor activity, nucleic acid binding (GO:0008135)					
Cassava4.1_016620 m.g	Eukaryotic initiation factor 4E protein	0.2(1)	−0.2(8.6E-01)	0.4(1)	0.7(2.0E-02)
Cassava4.1_013223 m.g	Eukaryotic translation initiation factor 4E	0.1(1)	−0.2(9.1E-01)	0.4(1)	0.6(6.2E-02)
Cassava4.1_004498 m.g	Eukaryotic translation initiation factor 3 subunit 7 (eIF3)	−0.1(1)	−0.4(5.0E-01)	0.1(1)	0.8(1.0E-02)
Cassava4.1_033528 m.g	Eukaryotic initiation factor 3 gamma subunit family protein	−0.1(1)	0.9(5.8E-01)	1.7(1)	1.6(9.2E-02)
Cassava4.1_011050 m.g	Eukaryotic translation initiation factor 2 subunit 1	0.0(1)	−0.5(3.2E-01)	−0.6(1)	0.7(2.4E-02)
Cassava4.1_017333 m.g	Eukaryotic initiation factor 3 gamma subunit family protein	−0.2(1)	0.7(3.1E-01)	0.8(1)	1.5(1.0E-03)
Cassava4.1_018052 m.g	Eukaryotic elongation factor 5A-1	−0.2(1)	0.2(8.1E-01)	−0.1(1)	0.9(5.5E-03)
Cassava4.1_018059 m.g	Eukaryotic elongation factor 5A-1	0.0(1)	0.5(3.1E-01)	−0.9(1)	0.9(6.0E-04)
Cassava4.1_001203 m.g	Eukaryotic translation initiation factor 3 C	−0.2(1)	−0.5(3.2E-01)	0.3(1)	1.5(2.0E-07)
Cassava4.1_020990 m.g	Eukaryotic release factor 1–3	−0.1(1)	−0.4(5.7E-01)	−0.2(1)	0.6(7.8E-02)
Cassava4.1_002530 m.g	Translation initiation factor 3B1	0.1(1)	−0.1(9.7E-01)	−0.3(1)	0.8(1.3E-02)
Cassava4.1_002467 m.g	Translation initiation factor 2, small GTP-binding protein	0.0(1)	−0.3(7.4E-01)	0.6(1)	0.6(1.1E-01)
Cassava4.1_015319 m.g	Translation elongation factor EF1B/ribosomal protein S6 family protein	0.0(1)	−0.7(1.3E-01)	0.2(1)	0.6(6.8E-02)
Cassava4.1_007596 m.g	Translation initiation factor IF2/IF5	−0.1(1)	0.1(9.9E-01)	−0.4(1)	0.7(1.5E-02)
Cassava4.1_007700 m.g	Translation initiation factor IF2/IF5	0.1(1)	−0.6(1.7E-01)	−0.1(1)	0.6(4.3E-02)
Cassava4.1_010349 m.g	Translation elongation factor Ts (EF-Ts), putative	−0.4(1)	−0.5(5.7E-01)	0.5(1)	1.0(1.3E-02)

Continued

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
Cassava4.1_007378 m.g	GTP binding Elongation factor Tu family protein	−0.2(1)	0.6(3.1E-01)	−0.3(1)	1.7(4.5E-05)
Cassava4.1_018593 m.g	Nucleic acid-binding, OB-fold-like protein	0.2(1)	−0.6(3.0E-01)	0(1)	0.9(5.1E-02)
Cassava4.1_005116 m.g	Peptide chain release factor 2	−0.1(1)	1.2(4.5E-01)	1.6(1)	2.0(8.2E-04)
GO term: Ribosomal subunit (GO:0006414)					
Cassava4.1_018516 m.g	Translation protein SH3-like family protein	−0.1(1)	−0.2(8.7E-01)	−0.5(1)	0.5(9.9E-02)
Cassava4.1_016258 m.g	Ribosomal protein 5B	−0.2(1)	−0.4(4.7E-01)	0.2(1)	1.0(2.2E-04)
Cassava4.1_013486 m.g	Ribosomal protein l18e/L15 superfamily protein	−0.1(1)	0.5(4.8E-01)	−0.2(1)	0.8(8.6E-02)
Cassava4.1_015972 m.g	Ribosomal protein l1p/L10e family	−0.2(1)	0.3(6.2E-01)	−0.4(1)	1.1(7.9E-05)
Cassava4.1_015937 m.g	Ribosomal protein l1p/L10e family	−0.1(1)	−0.1(9.8E-01)	0.1(1)	0.8(3.6E-03)
Cassava4.1_009639 m.g	Ribosomal protein l1p/L10e family	0.3(1)	0.4(7.4E-01)	0.4(1)	2.0(9.5E-08)
Cassava4.1_009636 m.g	Ribosomal protein l1p/L10e family	0.4(1)	0.0(1.0E + 00)	0.3(1)	1.5(9.8E-05)
Cassava4.1_008473 m.g	Ribosomal protein l1p/L10e family	−0.1(1)	−0.8(8.2E-02)	0.4(1)	0.7(3.6E-02)
Cassava4.1_017425 m.g	Ribosomal protein l22p/L17e family protein	−0.2(1)	−0.2(8.3E-01)	−0.5(1)	1.1(8.8E-05)
Cassava4.1_019192 m.g	Ribosomal protein l22p/L17e family protein	0.0(1)	−0.4(5.2E-01)	−0.4(1)	1.2(2.0E-05)
Cassava4.1_012305 m.g	40 s ribosomal protein SA	−0.3(1)	−0.1(1.0E + 00)	1.1(1)	1.8(7.2E-08)
Cassava4.1_012175 m.g	40 s ribosomal protein SA	−0.3(1)	−0.2(8.8E-01)	0.6(1)	1.0(2.4E-03)
Cassava4.1_012280 m.g	40 s ribosomal protein SA	−0.3(1)	0.0(1.0E + 00)	1.7(1)	1.3(6.4E-04)

Table 3. Comparison of UCBSV-induced genes corresponding to the over-represented GO terms 'translational elongation', 'translation factor activity, nucleic acid binding', and 'ribosomal subunit' at 2 and 5 dag in Namikonga and Albert.

abundant (LEA) proteins (Fig. 6G and H), transcription factors WRKY; (NAM, ATAF and CUC (NAC)); NmrA; eIF; GATA; and GRAS) (Fig. 6I,J,O,P), chaperones (Fig. 6K and 6L) and heat shock proteins (HSP) (Fig. 6M and N). Both LRR and NBARC domains are structural units of the NOD-like receptors (NLRs), which are defense proteins whose structure contains a TIR or CC domain at the N-terminus, LRRs at the C-terminus and a centrally located NBARC domain[49]. Of these, seven gene families (LRR, NBARC, PR, LEA, Chaperone, HSP and TF) were significantly different between Albert and Namikonga, and the details are presented below.

LRR domain. In Namikonga, two (007501 and 001048) of the five selected LRR genes were up- regulated at 2 dag and four (007501, 000978, 034154 and 000765) were up-regulated at 5 dag. For Albert, only two genes (007501 and 001048) were up-regulated at a later time point, 54 dag.

NBARC domain. For Namikonga, all six selected NBARC genes were strongly up-regulated at 5 dag with similar profiles, suggesting their co-expression and co-regulation. For Albert, only a few genes were induced (034172 and 022172) at that same time point.

PR proteins (thaumatin-like superfamily, PR-5 family). Two PR genes (012383 and 011960) that belong to the PR-5 family[50–52], thaumatin-like superfamily, were analyzed in Namikonga and Albert. In Namikonga, gene 012383 was up-regulated at 2 and 5 dag, while gene 011960 had delayed induction. In Albert, however, gene 012383 was up-regulated at 1 dag and later at 45 dag.

LEA proteins. In Namikonga, gene 019959 was highly induced in Namikonga at 2, 5 and 8 dag, while gene 025676 was induced at 1, 2 and 5 dag but at a lower level than that of gene 019959. In Albert, however, both genes remained unchanged across all time points.

Chaperones. Of the 14 selected chaperones, gene 026342 was induced in both Namikonga and Albert at 2 dag but to a greater extent in Namikonga (log2FC = 18) compared to Albert (log2FC = 5). The same gene (026342) was repressed in both varieties at 45 dag. In Namikonga, several other chaperones were induced at 2 and 5 dag; however, gene 026342 showed the only interesting pattern in Albert, with the other genes remaining unchanged.

HSPs. In Namikonga, six (014648, 009750, 010803, 033681, 034243 and 003240) of the 14 selected HSPs were highly induced at 2 dag (log2FC = 3.6–7.5). The other eight genes were also induced at other early time points (6 hag, 1 dag, 2 dag and 5 dag) in Namikonga. For Albert, only two genes, 010803 (induced at 6 hag, 1 dag and 2 dag) and 003340 (induced at 45 dag), were interesting. The others remained unchanged.

TFs (WRKY, NmrA, GATA, GRAS and NAC). Of the five families of TFs examined, only two showed interesting patterns. The first, WRKY, had two genes (011680 and 014614) that were both repressed at 2 dag and induced at 5 and 8 dag in Namikonga, whereas only 011680 was induced in Albert at 8 dag. The other TF families (NmrA, GATA, GRAS and NAC) were induced at 2 and 5 dag in Namikonga only. In Albert, two genes (032524 and 028589) that encode NmrA proteins were induced at 45 dag, while the others remained unchanged at all other time points.

Other defense-related genes identified in an earlier study by Maruthi *et al.*[20]: A transcriptome analysis of Kaleso (identical to Namikonga[53]) and Albert infected with CBSV reported that three NAC transcription factors (011029, 015961 and 023870) and one elongation factor, eIF(iso)4E (016601) were over-expressed. The NAC transcription factors were over-expressed in CBSV-infected Kaleso by 27–139 times compared to CBSV-infected Albert[20], while the eIF was two-fold over-expressed in CBSV-infected Kaleso compared to Albert, in RPKM values. We examined the fold change of the count values of these genes in the present study (Supplementary Tables S6 and S7). Preliminary filtering dropped one NAC gene (016601) together with other genes whose expression was below the expression threshold (see the details of filtering low-quality genes in the methodology section "Filtering RNAseq reads"). The remaining two NAC genes were expressed as follows: In Namikonga, both NAC genes (011029 and 015961) were up-regulated at 2 dag (log2FC of 1.0 and 2.9) and 5 dag (log2FC of 5.1 and 1.0). In Albert, both genes were slightly up-regulated (log2FC of 0.4) at 2 dag. At 5 dag, one NAC gene was substantially down-regulated in Albert (log2FC = −2.4), and the other remained slightly up-regulated (log2FC = 0.4).

Discussion

CBSD is a major constraint to the production of the African staple food, cassava[1]. Two causative agents have been identified, the Potyviruses CBSV and UCBSV. There is an urgent need to identify molecular markers or biomarkers associated with resistance or tolerance to support the efficient breeding of new cassava varieties. Here, we used an RNAseq approach to identify DEGs between UCBSV-inoculated and mock-inoculated plants in a time-course experiment with two Tanzanian cassava varieties, Namikonga (CBSD resistant) and Albert (CBSD susceptible). The results indicated that Namikonga plants restricted disease progression, limited symptoms to the leaves and maintained a low virus load, while allowing normal root expansion (root bulking) without necrosis or constriction. In Albert, where viral loads were higher, infection with UCBSV caused substantial leaf chlorosis and root necrosis. The results also indicated that a strong resistance response is invoked in Namikonga from approximately 2 dag to at least 5 dag, reducing by 8 dag. Comparatively little differential gene expression was observed at the early time points (6 hag to 1 dag). The over-represented GO terms indicate the involvement of 'translation initiation' in the 'ribosomal subunit' and 'phosphorelay signal transduction and response regulation'. In addition, many genes with defense-related roles, including those with LRR domains, NLRs, PR proteins, LEA proteins and TFs are highly differentially expressed at these time points, indicating a complex response. Although some of these genes are up-regulated in Albert, significant up-regulation tends to occur much later, at 54 dag.

Low virus titer in Namikonga suggests that the defense mechanism controls virus replication. Typical foliar CBSD symptoms were observed on both Namikonga and Albert plants at the late sampling phase (54 dag), and at harvest, demonstrating that Namikonga is not immune to UCBSV and can be graft inoculated (Fig. 1). UCBSV-inoculated plants of both varieties tested positive for UCBSV (Fig. 1) using CBSDDR/CBSDDF2 primers[54]. No UCBSV sequences were identified in the Namikonga RNAseq reads, except at 45 and 54 dag. The high number of UCBSV sequences recovered from Albert at the early time points confirms that the virus multiplied at a higher rate in Albert than in Namikonga (Fig. 3). Susceptible cassava varieties have been shown to harbor a higher virus load than resistant varieties[21,55]. Low UCBSV titers have been reported in field samples of Namikonga that were naturally infected via the (U)CBSV vector, the whitefly[21,55]. We observed similar trends, as the UCBSV titer remained very low throughout the study and only increased at 54 dag, whereas that of Albert had doubled by 6 hag compared to virus levels in Namikonga at the same time. The slightly increased titer (117 UCBSV reads) at 54 dag in Namikonga might be a result of localized virus multiplication within the leaves of Namikonga, which was previously observed in another study[20], or may indicate the occurrence of a more general fluctuation in viral load, rather than loads remaining constantly low. There was an exponential increase in virus load in Albert at 1 dag, which reduced by 5 dag. Altogether, virus levels remained higher in Albert than in Namikonga up to 54 dag (1660 UCBSV reads), except at 5 dag, when both varieties had very low loads. This observation suggests that defense mechanism in Namikonga controls virus multiplication without necessarily eliminating the virus. Our data support earlier studies in which Namikonga was declared 'resistant' to CBSD[20].

Read mapping to the cassava reference genome. Over 75% of the raw reads mapped to genic regions of the cassava reference genome. Of the unmapped reads, 2% mapped to multiple gene models, irrespective of length. Other possible explanations for the unmapped reads could include sequence variation from SNPs between the cassava reference genome version and Namikonga and Albert. The cassava reference genome is from a partial inbred line (accession number AM560-2) derived by three generations of selfing from variety MCOL-1505 of Latin American origin[42], while Namikonga is a local variety derived from early cassava breeding work at Amani, Tanzania[56], and is known to contain approximately 14% *M. glaziovii* – *M. esculenta* hybrid genome[57]. Albert is a

Gene ID	Gene annotation	2 dag		5 dag	
		Albert	Namikonga	Albert	Namikonga
		Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)	Log 2 FC (Adj.Pv)
GO term: Phosphorelay signal transduction system (GO:0000160) and response regulator activity (GO:0000156)					
Cassava4.1_000395 m.g	Histidine kinase 2	0.1(1)	1.3(6.2E-04)	−0.9(1)	2.1(8.6E-14)
Cassava4.1_000859 m.g	CHASE domain containing histidine kinase protein	0.1(1)	2.6(3.3E-06)	−0.3(1)	2.1(1.6E-11)
Cassava4.1_001670 m.g	Histidine kinase 2	0.8(1)	4.7(1.6E-36)	0.1(1)	2.3(3.7E-18)
Cassava4.1_002119 m.g	Pseudo-response regulator 7	0.3(1)	1.8(1.0E-07)	−0.1(1)	0.8(3.8E-03)
Cassava4.1_002915 m.g	Response regulator 12	0.6(1)	1.5(1.7E-05)	0.1(1)	2.4(1.1E-10)
Cassava4.1_015446 m.g	Response regulator 9	0.8(1)	4.6(1.3E-04)	−1.5(1)	3.5(1.4E-14)
Cassava4.1_015815 m.g	Response regulator 9	−0.2(1)	4.0(6.9E-24)	−1.5(1)	1.8(2.3E-01)
Cassava4.1_018229 m.g	Histidine-containing phosphotransmitter 1	0.4(1)	1.0(3.0E-01)	−0.5(1)	1.4(4.1E-07)
Cassava4.1_018306 m.g	Histidine-containing phosphotransmitter 1	−0.4(1)	3.2(7.0E-07)	0.8(1)	−1.1(5.0E-01)
Cassava4.1_022288 m.g	Response regulator 9	0.9(1)	4.1(1.6E-01)	0.3(1)	2.5(8.7E-08)
Cassava4.1_022850 m.g	Pseudo-response regulator 7	0.2(1)	4.4(1.0E-14)	−0.5(1)	2.5(4.5E-08)
Cassava4.1_024410 m.g	Histidine kinase 5	−0.2(1)	3.8(1.9E-02)	2.6(1)	0.1(1.0E + 00)
Cassava4.1_025220 m.g	Pseudo-response regulator 3	0.7(1)	4.3(2.5E-16)	−0.5(1)	1.2(1.7E-03)
Cassava4.1_026975 m.g	Response regulator 3	0.4(1)	4.1(3.0E-03)	0.0(1)	3.5(1.3E-12)
Cassava4.1_033332 m.g	Pseudo-response regulator 9	0.4(1)	1.3(5.6E-04)	−0.7(1)	1.0(2.9E-04)
Cassava4.1_002375 m.g	Signal transduction histidine kinase, hybrid-type, ethylene sensor	−0.6(1)	−2.6(4.8E-06)	−0.2(1)	0.8(1.2E-02)
Cassava4.1_015797 m.g	Response regulator 5	−0.4(1)	1.5(5.0E-01)	−1.3(1)	1.3(1.6E-02)
Cassava4.1_023472 m.g	Response regulator 3	−0.7(1)	0.2(9.4E-01)	0.6(1)	1.4(3.8E-03)
Cassava4.1_027609 m.g	Response regulator 9	0.7(1)	0.2(9.6E-01)	−0.4(1)	1.8(2.3E-07)
Cassava4.1_027924 m.g	Signal transduction histidine kinase, hybrid-type, ethylene sensor	0.1(1)	0.2(8.3E-01)	0.8(1)	0.8(5.2E-02)
Cassava4.1_028820 m.g	Response regulator 2	0.2(1)	1.4(8.4E-02)	−0.2(1)	1.4(4.8E-02)

Table 4. Comparison of UCBSV-induced genes corresponding to the over-represented GO terms 'phosphorelay signal transduction system' and 'response regulator activity' at 2 and 5 dag in Namikonga and Albert.

local, pure *M. esculenta* variety from Tanzania[57]. These differences could explain the unmapped reads, as population structure does exist between cassava germplasms from South America and Africa[58].

GO term enrichment of DEGs. The most over-represented GO terms were identified at 2 dag and 5 dag in Namikonga, which corresponded to the largest number of DEGs (3887 and 4911, respectively, in Namikonga, compared to two and zero, respectively, in Albert). Of the enriched GO terms at these time points, those involved in plant defense included 'translation elongation factors' (containing eIF genes), 'ribosomal subunit' and 'phosphorelay signal transduction'. None of these GO terms were over-represented in Albert. This observation suggests that translation elongation, the ribosomal subunit and phosphorelay signal transduction have play key roles in the resistance of Namikonga to CBSD.

Translation elongation is among the defense responses in Namikonga. Some viruses utilize the host plant's translation factors to replicate within the host. The virus-encoded protein cap structure (VPg) covalently linked to the 5′ end of some viral genomes, including (U)CBSV, recognizes and binds to translation eIFs on ribosomal subunits. This delivers an RNA helicase to the 5′ region, bridges the mRNA to the ribosome and circularizes the mRNA, enabling viral replication[33,59]. Mutations in translation initiation factors can result in structural changes in the protein, likely preventing the interaction of viral RNAs or proteins with host factors and therefore restricting the multiplication or movement of the virus, indirectly rendering the host resistant to that virus. The involvement of eIF genes in passive host plant resistance through an altered virus-host interaction surface has been recognized for many years. Two of the eIFs most commonly implicated in passive host plant resistance are eIF4E and eIF4G and their isoforms eIFiso4E and eIFiso4G. This passive resistance has been widely demonstrated in Potyviruses (reviewed in refs 28, 33 and 60), the family to which UCBSV belongs[61]. eIF4E has been repeatedly identified as a naturally occurring, recessively inherited resistance locus in species such as *Pisum sativum* (*sbm1*, *wlv/cyv2*)[62] and

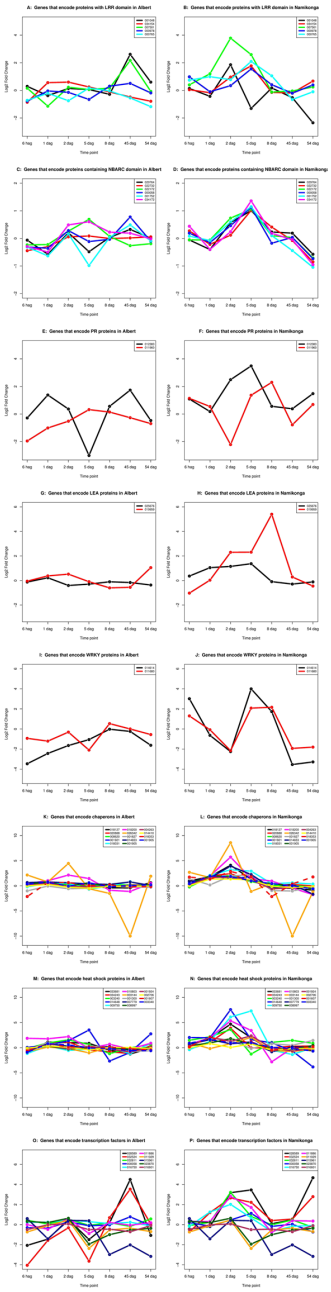


Figure 6. Selected DEGs that were expressed at significantly different levels in susceptible (Albert) and resistant (Namikonga) varieties at different time points (RNAseq data). Gene families include **A** and **B** = Leucine-rich repeat (LRR)-containing proteins, **C** and **D** = NBARC-containing proteins, **E** and **F** = Pathogenesis-related (PR) proteins, **G** and **H** = Late embryogenesis abundant (LEA) proteins, **I** and **J** = WRKY DNA, **K** and **L** = Heat shock proteins, **M** and **N** = Chaperones and **O** and **P** = Transcription factors.

Solanum habrochaites (*pot-1*)[31]. Alleles *pvr1* and *pvr1*^{2} from *Capsicum spp.*[32] confer resistance to the potyvirids TEV, PVY and PMV based on eIF4E. Similarly, over-expression of *pvr1* from *Capsicum* spp. confers resistance to TEV and PMV in *Solanum lycopersicum* based on eIF4E, and over-expression of *pvr1*^{2} confers resistance to PVY in *Solanum tuberosum*. The resistance gene *pvr6*, which functions with *pvr1* or *pvr1*^{2}, confers resistance to PVMV

and ChiVMV in *Capsicum* spp. based on eIF(iso)4E. In Arabidopsis, an elF4E gene confers resistance to melon necrotic spot virus (MNSV)[30]. Over-representation of the 'translation initiation factor' and 'ribosomal subunit' GO terms in Namikonga at 2 dag and 5 dag (Table 2) indicates the presence of an eIF-mediated resistance mechanism in the Namikonga defense response. In addition to these, the term 'phosphorelay signal transduction' was also enriched at 2 dag and 5 dag in Namikonga.

Phosphorelay signal transduction activates the defense response in Namikonga. The phosphorelay system is a complex interaction of multicomponent regulatory systems in which pathogen-triggered signals are transferred from host extracellular histidine- or histidine-containing proteins to intracellular regulatory proteins, which in turn trigger reactions that alter the plant's physiological patterns of defense[63–67].

More specifically, a phosphoryl group is transferred from a membrane-bound, conserved, sensor histidine kinase to an intracellular, receiver response regulator containing a conserved aspartate residue. The histidine kinases receive signals from extracellular sensors and transmit these signals via multi-step processes to cytoplasmic response regulators. The response regulators localize in the nucleus alone or in both the nucleus and the cytoplasm. The transfer of stimuli between the extracellular histidine kinases and intracellular response regulators occurs in a multi-step manner involving several plant proteins[63,64]. The response regulators in turn trigger a cascade of reactions that lead to pathogen defense and other physiological pathways.

Cytokinins, which are plant growth hormones, have been shown to receive and transmit signals in a manner similar to that of bacterial two-component phosphorelay signaling. In Arabidopsis, cytokinins have receptors (AHK2, AHK3 and AHK4)[68] whose role is similar to that of eukaryotic histidine kinases in the bacterial two-component system. These three proteins have extracellular cytokinin binding, cytoplasmic histidine transmitter and receiver domains, respectively[65,69]. Arabidopsis plants treated with cytokinins were resistant to *Pseudomonas syringae* pv. tomato DC3000 (Pst)[70]. Defense against Pst occurs via the salicylic acid (SA) pathway, which is triggered by cytokinins[71]. A receptor-like kinase (*Pseudomonas syringae* effector B) was shown to activate the expression of RPM1 (an NBS-LRR protein) against Pst in Arabidopsis[47]. The SA pathway confers resistance to the Potyviruses PVY in potato[72] and peanut mottle virus (PeMV)[73]. In tobacco, plants primed with cytokinins became tolerant to the strain PVY^{NTN} of potato virus Y (PVY)[66]. In this study involving cassava and UCBSV, both histidine kinases and regulatory proteins were among the over-expressed genes in Namikonga at 2 dag and 5 dag (Table 4). In addition to these GO terms and the constituent genes, other defense-related genes were significantly expressed in Namikonga compared to Albert.

A coordinated set of defense genes collectively confer resistance to CBSD in Namikonga. In addition to eIFs and possibly linked to phosphorelay signal transduction, genes that encode LRR- and NBARC-containing proteins, PR proteins, LEA proteins, TFs (WRKY, GRAS, GATA and NmrA), chaperones and HSPs were all highly up-regulated in Namikonga at 2 dag and 5 dag. These proteins have all been implicated in defense responses against viral, fungal and bacterial pathogens. Defense genes encoding proteins containing LRRs (Fig. 6A and B) and NBARCs (Fig. 6C and D) were substantially over-expressed at 2 dag and 5 dag in Namikonga, the latter time point being the peak of expression for most genes (Fig. 6). Other defense genes, including those encoding PR proteins (Fig. 6E and F), LEA proteins (Fig. 6G and H), TFs involved in pathogen defense (WRKY, NmrA, GATA, NAC and GRAS) (Fig. 6E,F, O,P), HSPs (Fig. 6M,N) and chaperones (Fig. 6K,L), were over-expressed at 2 dag and 5 dag. This suggests that Namikonga's resistance to UCBSV infection may be complex, involving several possibly interrelated strategies including NLR proteins with LRR and NBARC domains.

An NBS-LRR network contributes to Namikonga's resistance to UCBSV. In general, the NBARC genes were over-expressed at 5 dag in Namikonga and steadily decreased thereafter. In Albert, some genes were down-regulated and others over-expressed (but with lower log2FCs than in Namikonga) at 5 dag. However, three (000058, 001752 and 029764) of these genes were over-expressed at 45 dag in Albert. Maximum LRR gene expression was observed in Namikonga at 2 dag and 5 dag. When the LRR genes were maximally expressed in Namikonga, their expression was either normal or down-regulated in Albert. Two of these genes (001048 and 007501) were over-expressed at a much later time point (45 dag) in Albert, a time point when the expression of these genes in Namikonga had reduced to near normal.

Typically, LRRs, NBARCs and CCs or TIRs are sub-domains of the tri-modular NLR protein, the most studied type of defense gene[49,74,75]. Alone, LRR and NBARC domains can function in pathogen defense. Some NLR proteins require the presence of other 'helper' proteins to develop a resistance phenotype. The 'helper' is often an HSP[76,77]. In the case of Namikonga, synchronized timing for the up-regulation of these HSPs, chaperones and NLRs implies a synergistic interaction of the three defense proteins to orchestrate Namikonga's resistance to CBSD.

HSPs are structurally disordered proteins with diverse functions in regulatory, signaling and defense pathways. Alone, HSPs may function as defense proteins against biotic and abiotic stresses[78,79]. Some HSPs are also chaperones. Chaperones are mainly involved in RNA binding and protein folding[80,81], a role that is key in the functioning of elongation factors[46]. Their potential ability to independently confer resistance phenotypes in Namikonga cannot be ruled out.

During the peak over-expression of LRR, NBARC and HSP genes, LEA proteins were also over-expressed, reaching a maximum at 8 dag in Namikonga (Fig. 6H). LEA proteins are involved in adaptation to water stress[82] and are abundant in plant[82,83], fungi[84] and mammalian genomes[85–89]. As the name suggests, LEA proteins are produced during the late stages of embryogenesis, especially at seed development. They protect cells from desiccation[82], salinity[90] and extreme cold or hot temperatures[85] in a dose-dependent manner[88]. The expression of LEA

genes in Namikonga seems to ensure that the already infected plants maintain balanced cell solutes for normal functioning under the stressful conditions imposed by UCBSV.

Another type of defense proteins that were up-regulated in Namikonga at 2 dag and 5 dag are the PR proteins. The PR proteins are inducible defense-related proteins of varied molecular sizes (5 to 75 kDa)[91]. They were first discovered in tobacco resistant to infection by TMV[48, 51]. To date, PR proteins have been associated with resistance to several pathogens[48, 51]. The proteins are RNase and DNase active, implying that they function in defense by abolishing foreign RNA and DNA molecules in host plants. The PR proteins have antiviral properties[92–94], and resistance triggered by PR proteins causes an HR and programmed cell death at the infection site[92, 93]. They are critical for secondary metabolite biosynthesis, storage, phytohormones and ligand binding[94]. Their expression marks the peak of systemic acquired resistance (SAR) against pathogens[51]. In Namikonga, PR proteins were up-regulated at the peak of LRR, NBARC and HSP expression, but the role and extent of involvement of PR proteins in resistance to UCBSV is not clear.

At the same peak time points, five major groups of TFs (NAC, WRKY, GRAS, GATA and NmrA) were over-expressed in Namikonga. Although there is very little knowledge of the transcriptional networks that are activated in cassava plants in response to pathogens, the co-expression of these transcription factors in Namikonga with other defense genes in response to UCBSV inoculation suggests a role in defense.

The NAC (NAM, ATAF and CUC) superfamily of TFs is one of the largest TF families. It is found only in plants and was thus investigated here. The role of NAC proteins in defense is known in several plants. In Arabidopsis, a functional SA pathway is required for NAC-directed defense against viral pathogens such as TCV, but the jasmonic acid (JA) and ethylene (ET) pathways are not essential[95]. A NAC-mediated HR causes resistance in rice to rice blast disease[96], in maize to *Colletotrichum graminicola*[97] and in Arabidopsis to *Botrytis cinerea*, *Pseudomonas syringae* pv. *tomato* (Pst)[98], *Turnip crinkle virus*[95] and *Acidovorax avenae*[99]. In this study, NAC genes were differentially expressed at specific time points with other defense genes, implying that NAC has a role in defense. However, to date, no studies have focused on the role of NAC genes in cassava's defense against viral pathogens. Our data indicate that the induction of NAC genes in Namikonga at 2 dag and 5 dag contributes to NAC-mediated resistance to CBSD in Namikonga.

Another class of transcription factors, WRKY, was also up-regulated in Namikonga at 2 dag and 5 dag. The role of WRKY transcription factors in plant defense[100–102] and their structural features and functional network have been reviewed[103]. In rice, two alleles of a WRKY gene (OsWRKY45-1 and OsWRKY45-2) conferred resistance to *Magnaporthe grisea*, yet each allele encoded a different set of defense genes regulated by distinct promoter regions[104].

In this study, Namikonga had two differentially expressed WRKY genes at 2 dag and 5 dag. However, one WRKY was down-regulated at 2 dag, possibly to establish a network feedback loop to regulate their over-expression at 5 dag and 8 dag, respectively[104]. In some cases, SA and JA enhance the functionality of WRKY genes in defense. In rice, enhancing the levels of SA and JA production enabled OsWRKY45-1 to resist Xoo and Xoc, but OsWRKY45-2-mediated resistance to Xoo and Xoc required JA only[104]. Whether the SA levels enhanced the activity of the Namikonga WRKY genes against UCBSV was not examined in this study. There is a need to examine the roles of SA and JA in the Namikonga defense mechanism and if and how these phytohormones enhance the action of WRKY genes.

Concluding remarks. Based on these findings, a model for resistance to CBSD in Namikonga is proposed (Fig. 7). In this model, the challenge of Namikonga with UCBSV triggers an NBS-LRR signalling cascade, which subsequently induces a phosphorelay signal transduction pathway leading to defence. This is marked by induction of defence-related PR-proteins and stress response LEA, heat-shock and chaperone proteins, and a range of proteins associated with translation, including translation elongation factors. This may represent a mechanism in Namikonga of preventing efficient interaction of UCBSV with the host translational machinery, a known mechanism of resistance to potyviruses in other plants[31, 32, 61, 62, 105]. The outcome is that Namikonga shows reduced virus titre, limited leaf chlorosis and minimal (if any) root necrosis. In Albert, UCBSV's presence does not induce any recognizable defence pathway and virus titre increases unchecked (Fig. 7). The result is major leaf chlorosis and root necrosis. Our study confirms that resistance in Namikonga is associated with significant transcriptional re-programming soon after UCBSV challenge, in contrast to Albert in which the subterfuge of the virus appears to suppress defence gene expression (Fig. 7). Furthermore, our data provides a rich source of candidate alleles in the Namikonga genotype for testing hypothesis of resistance mechanisms by correlating gene expression with QTL mapping data, with the ultimate aim of developing robust biomarkers for cassava breeders to develop durable resistance to CBSD.

Methodology

Planting material sources. The study was conducted on Namikonga, a CBSD-resistant variety, and the susceptible variety Albert[21]. Namikonga is an interspecific hybrid between wild (*Manihot glaziovii* Müll. Arg.) and domesticated (*Manihot esculenta* Crantz) cassava, with an estimated 14% interspecific hybrid genome[57]. Albert is a farmer variety from southern Tanzania; it is susceptible to both virus species, showing characteristic CBSD-associated leaf chlorosis and severe root necrosis[21]. Virus-negative stem cuttings of Namikonga and Albert were obtained from plants maintained in a greenhouse at the Sugarcane Research Institute (SRI), Kibaha, Dar es Salaam, Tanzania. The source of inoculum, variety NDL06/132, was from a field trial at the SRI, Tanzania and was multiplied for scions in a greenhouse at BioSciences eastern and central Africa (BecA), Nairobi, Kenya. Experimental plants were tested for the presence or absence of both CBSV and UCBSV prior to grafting. For this, RNA was extracted from fresh leaf samples using a CTAB protocol, with the amount of starting extraction buffer material modified from 600 µl to 800 µl[106]. Respective second-strand cDNAs were synthesized using Superscript

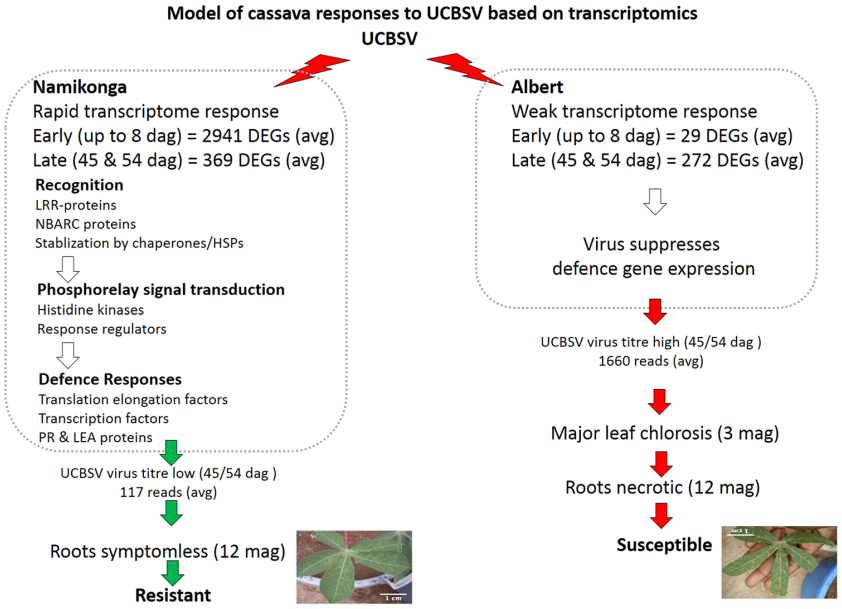


Figure 7. Model of cassava responses to UCBSV based on transcriptomics. DEGs = differentially expressed genes; avg = average; dag = days after grafting and mag = months after grafting.

III (Invitrogen, CA). End-point RT-PCR was performed using primers that distinguish both viruses simultaneously; namely, CBSDDR/F2[54] and CBSVF2/R8[107].

Establishment of the experiment, graft inoculation and sampling. The aim of this transcriptome time-series experiment was to determine the transcriptional signatures immediately after virus infection by grafting (early sampling) and several days before aerial symptom emergence (late sampling). Grafting is commonly used to transmit viruses in cassava, and was used here[4,16]. Artificial infectious clones were not available for inoculation, transmission using whiteflies was unreliable and mechanical transmission of CBSVs is only effective with model plants[15]. Initially 45 cassava cuttings of each variety were planted in large pots containing 3.85 kg gravel (construction grade) topped to 33.23 kg with forest soil and immediately watered with 3 liters of tap water. Thereafter, the plants were given 2 liters of tap water per week until harvest. At two months after planting, plants were ready for top grafting. Just prior to grafting a baseline leaf sample (time zero) was collected from the top leaves of all plants (see methodology below). Fifteen plants of Namikonga and Albert were grafted with scions positive for CBSV only, 10 plants of each variety were grafted with scions positive for UCBSV only and another 10 plants of each variety with scions positive for CBSV + UCBSV. Ten plants of each variety were grafted as controls with virus negative NDL06/132 scions. Samples were taken as described below from all grafted plants, and frozen until selection of biological replicates was made based on RT-PCR diagnostic results. An RT-PCR test at 3 mag (data not shown) confirmed that only one out of 15 grafted Namikonga plants was positive for CBSV, while seven of 10 plants each grafted with UCBSV or CBSV + UCBSV infected scions were positive for the respective virus. In Albert seven of the 15 plants grafted with CBSV infected scions were positive, and seven of 10 plants each infected with UCBSV or CBSV + UCBSV were positive. It is on the basis of successful graft transmission of UCBSV that this species was chosen for downstream studies. Three biological replicates for each of the two treatments (UCBSV-inoculated and mock-inoculated plants) of Albert and Namikonga were selected for RNAseq.

In this time-series experiment, samples were collected from the main experiment at 6 hag, 1 dag (i.e., 24 hag), 2 dag (i.e., 48 hag), 5 dag, 8 dag, 45 dag and 54 dag. The time points up to and including 8 dag constituted the "early sampling" and the remaining samples the "late sampling". The "late sampling" time points were selected to coincide with the period just before foliar symptom emergence. To determine this a number of indicator plants were grafted three weeks before the main experiment providing ample time to plan for pre-symptom emergence sampling. At each time point leaves were sampled from below the graft point to avoid sampling the leaves of the sprouting NDL06/132 scion. Typically both Albert and Namikonga plants have at least five lobes per leaf, each lobe weighing >400 g. Therefore, we sampled a different lobe of the same leaf at every different time point. Where lobes were fewer than needed to cover all successive time points, samples were taken from lobes on the leaf below in the same order. The samples were placed in pre-labeled aluminum foil envelopes and immediately frozen in liquid nitrogen, transferred to −80 °C and stored prior to RNA extraction. At harvest (12 mag), roots from all plants were scored for CBSD-related necrosis.

CBSD symptom scoring. UCBSV-inoculated and mock-inoculated plants were scored for above-ground symptoms during the late sampling phase (up to 54 dag), and additionally at harvest (12 mag) for symptoms in storage roots, according to standard 1–5 scales[108]. According to the leaf severity scale, 1 implies no symptoms, and 5 implies clear leaf chlorosis covering >75% of leaves and streaking on the stem with shoot die back. On the storage root severity scale, a score of 5 implies more than 50% root necrosis[7].

RNA extraction, cDNA synthesis and RNAseq. Frozen leaf samples were ground to a fine powder in liquid nitrogen and divided into three aliquots for (a) cDNA library synthesis, (b) virus diagnostics using RT-PCR and (c) a back-up. RNA was extracted using the Spectrum Plant Total RNA kit (Sigma-Aldrich) following the manufacturer's protocol. RNA integrity was confirmed using a Qubit fluorometer (Life Technologies). The RNA extraction, cDNA library synthesis, shipment and Illumina sequencing were performed 16 months apart for two batches of samples that were sequenced at two facilities, the University of California, Berkeley, CA, USA, and Dow AgroSciences, Indianapolis, IN, USA, due to funding availability. The same person in the same laboratory at BecA conducted the RNA extraction and cDNA library synthesis of both batches. Each batch contained both the mock-inoculated and UCBSV-inoculated samples from a genotype at a specific time point, as listed in Supplementary Table S1. The same shipping company was contracted to ship each batch of cDNA libraries to the sequencing facilities. The cDNA libraries were synthesized using the Illumina Truseq cDNA library synthesis kit (Illumina, San Diego, CA, USA) using set A indices for 50-bp sequencing at UC Berkeley and set B indices for 101-bp sequencing at Dow AgroSciences. Illumina Hi-Seq instruments were employed at both facilities, using in-house pre-sequencing preparations of cDNA library material and instrument run settings with the aim of generating 20 million reads of 50 bp (for UC Berkeley) or 101 bp (for Dow AgroSciences) in length for each biological replicate sample (Supplementary Table S1). Control and UCBSV-infected samples from both varieties obtained at 6 hag and 8 dag were sequenced at UC Berkeley. Control and infected Albert samples collected at 45 dag and 54 dag were also sequenced at UC Berkeley. For both varieties, control and UCBSV-infected samples collected at 1 dag, 2 dag and 5 dag, were sequenced at Dow AgroSciences. Dow AgroSciences also sequenced the Namikonga samples (mock- and UCBSV- inoculated) collected at 45 dag and 54 dag.

RNA extraction and cDNA synthesis for diagnostics. For RT-PCR virus diagnostics, aliquots of RNA samples from the Spectrum extractions described above were used, and cDNA was synthesized using Superscript III (Invitrogen). The presence of UCBSV was positively verified with a single band at 440 bp, and the absence of CBSV was verified using the primer sets described above with appropriate CBSV and UCBSV positive and negative controls[54] (Fig. 2).

Detection of UCBSV in RNAseq reads. The availability of RNAseq data from this experiment was exploited as an additional method to detect UCBSV RNA virus molecules in the cassava samples. To confirm successful graft inoculation with UCBSV and that control samples were UCBSV-negative, RNAseq reads from all 96 samples were mined for UCBSV sequences. To retrieve virus sequences from RNAseq reads, all five fully sequenced UCBSV genomes from NCBI were used as a reference to map the RNAseq reads of UCBSV-infected and mock-inoculated samples. The five UCBSV genomes were isolates HG965222.1, FN434109.1, FJ185044.1, HM181930.1 and NC_014791.1 and were combined to form one UCBSV reference file. RNAseq reads that failed to map to the cassava reference genome (unmapped reads) were each aligned to the UCBSV reference genome using Bowtie2[41]. The results were obtained from the alignment summary and transferred to Excel for plotting.

Data analysis. *Read quality analysis and mapping to the cassava reference genome.* Individual files received as batches of ≤four million reads (later combined to make one large file of 20 million reads) were analyzed for sequence quality using FastQC (FastQC)[109]. All reads had a Phred score above 20, meaning that 99% of the bases were accurately called. The first low-quality 10–13 bp of each read, which is typical for Illumina sequencing, were trimmed to 37 bp (for 50-bp reads sequenced at UC Berkeley) and 88 bp (for 101-bp reads sequenced at Dow AgroSciences) using FASTX_toolkit (fastx_trimmer)[110]. The trimmed files were mapped to the cassava reference genome v4.1 (plant accession number AM560-2)[42] using Tophat2[41], which runs Bowtie2 in the background. The default settings were used, allowing up to two mismatches per read. Allowing two mismatches for both 37 bp and 88 bp reads did not affect mapping, as reads that map to the same gene model are only counted once, irrespective of their length, taking into account the two allowed mismatches. This was tested and confirmed using randomly selected samples of 37 bp or 88 bp (data not shown). The mapped reads were de-duplicated using Picard's dedup function, correcting for any amplification bias caused during the PCR used in constructing the sequencing libraries. Mapped reads with gene models on the cassava genome were counted with HTseq_count[43], a python script widely used to count RNAseq reads[43]. Genes without gene models (counted altogether as no_feature) or matching more than one gene model (ambiguous) were removed before clustering and identification of DEGs. The HTseq_count command *no_feature* counts reads that are not aligned to any gene model on the reference genome, while the command *ambiguous* counts reads spanning an intersection of two gene models or aligned to more than one gene model (Supplementary Data S4).

Filtering RNAseq reads. To reduce artifacts and increase the statistical power for identifying DEGs, the counted reads were filtered in four steps. First, genes with zero count values were deleted, followed, second, by genes with very few reads (rowsum $\leq \log_2 5$) on the distribution curve (data not shown). Third, genes that passed stages (i) and (ii) but were below a set per time point minimum cut-off were deleted. The minimum cut-off per time point was calculated from row (per gene) median and standard deviation values (minimum cut-off = median − 2 * standard deviation). The fourth and last filtering was based on how close the biological replicates clustered by respective treatments, variety and time point. Outlier biological replicates were removed

from the downstream analysis, although at least two samples per treatment were retained. As was done for differential gene expression analysis, data for clustering were normalized using the *estimateSizeFactors* function of the DESeq package[45]. Data normalization was performed with combined gene counts from both Albert and Namikonga. To do this, only good-quality genes (genes that passed the four filtering stages described above) were used for the normalization analysis in both varieties. Clustering was performed using Pearson's algorithm. The algorithm clusters samples based on their covariance, providing a robust method of grouping samples as opposed to the distance-based Euclidean method. The filtering, clustering, selection of DEGs and plotting of data were performed using the R statistical package[111].

Identification of DEGs and cluster analysis. Compared to distribution-based methods and cuffdiff[112], negative binomial algorithms implemented by DESeq (edgeR and bayseq) provide a higher statistical power for identifying DEGs[43,113]. Using filtered reads, DESeq was applied to identify DEGs between control and treated samples of each variety and at each time point. In DESeq, each variety had a dataset for respective sampling time points (example: Albert_6hag). Each dataset contained filtered biological replicates for one time point and two treatments (UCBSV inoculated and mock inoculated). The function *newCountDataSet*, which imports gene counts with pre-defined conditions, was used to import samples under "infected" or "control", where "infected" identifies gene counts of UCBSV-inoculated samples and "control" identifies gene counts from mock-inoculated samples. Once imported, the samples were normalized by their respective library sizes. To perform 'normalization', the effective library size was determined using the *estimateSizeFactors* command. The per sample dispersion was then estimated using the *estimateDispersions* command, followed by differential expression analysis. Differential expression analyses were performed using the negative binomial algorithm with the command *nbinomTest*. After differential expression analysis, p-values were adjusted for multiple testing with the Benjamini-Hochberg procedure, which controls for the false discovery rate (FDR). For most RNAseq studies, a 10% FDR is the recommended limit for identifying significant DEGs[44,113]. We chose only to apply a FDR of 0.1 as the main filter to identify DEGs, and did not apply a further filter of Log2 fold change. Therefore genes for a treatment comparison with $\log2FC > 0$ and $FDR < 0.1$ were classified as up-regulated, whereas those with $\log2FC < 0$ and $FDR < 0.1$ were down-regulated. This was done to capture the maximum number of genes with differential expression across the time points.

Functional annotation, GO enrichment and network analysis of DEGs. The lists of DEGs were queried against the cassava genome database at VirtualPlant (http://www.virtualplant.org)[114] to determine functional annotations and GO enrichment. VirtualPlant is annotated using cassavaCyc data (functional categories) from Phytozome's version 4.1 of the cassava reference genome (plant accession number AM560-2) (26). VirtualPlant contains 30,666 cassava genes, 28,610 of which have predicted Arabidopsis orthologues that have protein-coding annotations. Thus, the functional annotations are primarily metabolic, as available in cassavaCyc (http://www.plantcyc.org/). The GO terms for cassava genes in VirtualPlant are allocated based on the GO terms of their corresponding predicted Arabidopsis orthologues in TAIR (www.Arabidopsis.org). VirtualPlant therefore harbors GO terms for 10,902, 15,368, and 4,546 cassava genes in the biological process, molecular function and cellular compartment categories, respectively. The manual analysis of defense genes focused on the time points with the highest number of DEGs.

References

1. Legg, J. P. *et al.* Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. *Virus Res* **159**, 161–170, doi:10.1016/j.virusres.2011.04.018 (2011).
2. Patil, B. L. *et al.* RNAi-mediated resistance to diverse isolates belonging to two virus species involved in cassava brown streak disease. *Mol Plant Pathol* **12**, 31–41, doi:10.1111/j.1364-3703.2010.00650.x (2011).
3. Mbanzibwa, D. R. *et al.* Genetically distinct strains of cassava brown streak virus in the Lake Victoria basin and the Indian Ocean coastal area of East Africa. *Arch Virol* **154**, 353–359, doi:10.1007/s00705-008-0301-9 (2009).
4. Winter, S. *et al.* Analysis of cassava brown streak viruses reveals the presence of distinct virus species causing cassava brown streak disease in East Africa. *J Gen Virol* **91**, 1365–1372, doi:10.1099/vir.0.014688-0 (2010).
5. Storey, H. H. Virus diseases of East African plants. VI-A progress report on studies of the disease of cassava. *E Afri Agric J* **2**, 34–39 (1936).
6. Alicai, T. *et al.* Cassava brown streak virus has a rapidly evolving genome: implications for virus speciation, variability, diagnosis and host resistance. *Sci Rep* **6**, doi:10.1038/srep36164 (2016).
7. Alicai, T. *et al.* Re-emergence of cassava brown streak disease in Uganda. *Plant Dis* **91**, 24–29, doi:10.1094/pd-91-0024 (2007).
8. Monger, W. A. *et al.* The complete genome sequence of the Tanzanian strain of cassava brown streak virus and comparison with the Ugandan strain sequence. *Arch Virol* **155**, 429–433, doi:10.1007/s00705-009-0581-8 (2010).
9. Maruthi, M. N. *et al.* Transmission of Cassava brown streak virus by *Bemisia tabaci* (Gennadius). *Phytopathology* **153**, 307–312 (2005).
10. Bock, K. R. Studies on cassava brown streak virus disease in Kenya. *Trop Sci* **34**, 134–145 (1994).
11. Mware, B. *et al.* Efficiency of cassava brown streak virus transmission by two whitefly species in coastal Kenya. *J Gen Mol Virol* **1**, 040–045 (2009).
12. Mware, B. *et al.* First record of spiraling whitefly in coastal Kenya: emergence, host range, distribution and association with cassava brown streak virus disease. *Int J Agric & Biol* **12**, 411–415 (2010).
13. Hillocks, R. J., Raya, M. D., Mtunda, K. & Kiozia, H. Effects of brown streak virus disease on yield and quality of cassava in Tanzania. *J Phytopathol* **149**, 389–394 (2001).
14. Lister, R. M. Mechanical transmission of cassava brown streak virus. *Nature* **183**, 1588–1589 (1959).
15. Ogwok, E., Patil, B. L., Alicai, T. & Fauquet, C. M. Transmission studies with cassava brown streak Uganda virus (Potyviridae: Ipomovirus) and its interaction with abiotic and biotic factors in *Nicotiana benthamiana*. *J Virol Methods* **169**, 296–304, doi:10.1016/j.jviromet.2010.07.030 (2010).
16. Wagaba, H. *et al.* Efficient transmission of cassava brown streak disease viral pathogens by chip bud grafting. *BMC Res Not* **6**, 516, doi:10.1186/1756-0500-6-516 (2013).
17. Adams, I. P. *et al.* 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 Pathol* **62**, 233–242, doi:10.1111/j.1365-3059.2012.02622.x (2013).

18. Mohammed, I. U., Abarshi, M. M., Muli, B., Hillocks, R. J. & Maruthi, M. N. The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. *Adv Virol.* **2012**, 795697, doi:10.1155/2012/795697 (2012).
19. Rwegasira, G. M. & Rey, C. M. E. Relationship between symptoms expression and virus detection in cassava brown virus streak-infected plants. *J Agric Sci* **4**, doi:10.5539/jas.v4n7p246 (2012).
20. Maruthi, M. N., Bouvaine, S., Tufan, H. A., Mohammed, I. U. & Hillocks, R. J. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. *PLoS One* **9**, e96642, doi:10.1371/journal.pone.0096642 (2014).
21. Kaweesi, T. *et al.* Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *Virol J.* **11**, 1 (2014).
22. Kawuki, R. S. *et al.* A breeding scheme for local adoption of cassava (*Manihot esculenta* Crantz). *J Plant Breed Crop Sci* **3** (2011).
23. Morozova, O., Hirst, M. & Marra, M. A. Applications of new sequencing technologies for transcriptome analysis. *Annu Rev Genomics Hum Genet.* **10**, 135–151, doi:10.1146/annurev-genom-082908-145957 (2009).
24. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628, doi:10.1038/nmeth.1226 (2008).
25. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63 (2009).
26. de Ronde, D., Butterbach, P. & Kormelink, R. Dominant resistance against plant viruses. *Front Plant Sci.* **5**, 307, doi:10.3389/fpls.2014.00307 (2014).
27. Robaglia, C. & Caranta, C. Translation initiation factors: a weak link in plant RNA virus infection. *Trends Plant Sci* **11**, 40–45 (2006).
28. Truniger, V. & Aranda, M. A. Recessive resistance to plant viruses. *Adv Virus Res* **75**, 119–231 (2009).
29. Kang, B. C., Yeam, I., Li, H., Perez, K. W. & Jahn, M. M. Ectopic expression of a recessive resistance gene generates dominant potyvirus resistance in plants. *Plant Biotech J* **5**, 526–536 (2007).
30. Nieto, C. *et al.* An eIF4E allele confers resistance to an uncapped and non-polyadenylated RNA virus in melon. *The Plant J* **48**, 452–462 (2006).
31. Ruffel, S. *et al.* A natural recessive resistance gene against *Potato virus Y* in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *The Plant J* **32**, 1067–1075 (2002).
32. Ruffel, S., Gallois, J. L., Lesage, M. L. & Caranta, C. The recessive potyvirus resistance gene pot-1 is the tomato orthologue of the pepper pvr2-eIF4E gene. *Mol Genet Genomics* **274**, 346–353 (2005).
33. Duan, C. G., Wang, C. H. & Guo, H. S. Application of RNA silencing to plant disease resistance. *Silence* **3**, 5, doi:10.1186/1758-907X-3-5 (2012).
34. Gonzalez-Ibeas, D., Cañizares, J. & Aranda, M. A. Microarray analysis shows that recessive resistance to watermelon mosaic virus in melon is associated with the induction of defense response genes. *MPMI* **25**, 107–118 (2012).
35. Lellis, A. D., Kasschau, K. D., Whitham, S. A. & Carrington, J. C. Loss-of-susceptibility mutants of *Arabidopsis thaliana* reveal an essential role for eIF (iso) 4E during potyvirus infection. *Curr Biol.* **12**, 1046–1051 (2002).
36. Cosson, P. *et al.* The RTM resistance to potyviruses in *Arabidopsis thaliana*: natural variation of the RTM genes and evidence for the implication of additional genes. *PLoS One* **7**, doi:10.1371/journal.pone.0039169 (2012).
37. Palloix, A., Ayme, V. & Moury, B. Durability of plant major resistance genes to pathogens depends on the genetic background, experimental evidence and consequences for breeding strategies. *New Phytol* **183**, 190–199, doi:10.1111/j.1469-8137.2009.02827.x (2009).
38. Huh, S. U. *et al.* Capsicum annuum transcription factor WRKYa positively regulates defense response upon TMV infection and is a substrate of CaMK1 and CaMK2. *Sci Rep* **5**, doi:10.1038/srep07981 (2015).
39. Wright, K. M. *et al.* Analysis of the N gene hypersensitive response induced by a fluorescently tagged *Tobacco mosaic virus*. *Plant Physiol* **123**, 1375–1385 (2000).
40. Hinrichs, J., Berger, S. & Shaw, J. G. A hypersensitive response-like mechanism is involved in resistance of potato plants bearing the Ry(sto) gene to the potyviruses *Potato virus Y* and *Tobacco etch virus*. *J Gen Virol* **79**, 167–176, doi:10.1099/0022-1317-79-1-167 (1998).
41. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).
42. Prochnik, S. *et al.* The cassava genome: current progress, future directions. *Trop Plant Biol* **5**, 88–94, doi:10.1007/s12042-011-9088-z (2012).
43. Anders, S., Pyl, P. T. & Huber, W. HTSeq - A Python framework to work with high-throughput sequencing data. *Biorxiv Preprint*, doi:10.1101/002824 (2014).
44. AC't Hoen, P. *et al.* Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nature Biotechnol* **31**, 1015–1022 (2013).
45. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106 (2010).
46. Gingras, A., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Annu Rev Biochem* **68**, 913–963 (1999).
47. Liu, J., Elmore, J. M., Lin, Z.-J. D. & Coaker, G. A receptor-like cytoplasmic kinase phosphorylates the host target RIN4, leading to the activation of a plant innate immune receptor. *Cell Host Microbe.* **9**, 137–146 (2011).
48. van Loon, L. C. & van Kammen, A. Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana tabacum* var. 'Samsun' and 'Samsun NN': II. changes in protein constitution after infection with *Tobacco mosaic virus*. *Virology* **40**, 199–211, doi:10.1016/0042-6822(70)90395-8 (1970).
49. Proell, M., Riedl, S. J., Fritz, J. H., Rojas, A. M. & Schwarzenbacher, R. The NOD-like receptor (NLR) family: a tale of similarities and differences. *PLoS One* **3**, e2119, doi:10.1371/journal.pone.0002119 (2008).
50. Liu, J. J., Sturrock, R. & Ekramoddoullah, A. K. M. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Rep* **29**, 419–436, doi:10.1007/s00299-010-0826-8 (2010).
51. van Loon, L., Rep, M. & Pieterse, C. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol* **44**, 135–162 (2006).
52. van Loon, L. C. & van Strien, E. A. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. *Physiol Mol Plant Path* **55**, 85–97 (1999).
53. Kanju, E., Mahungu, N., Dixon, A. G. O. & Whyte, J. Is resistance/tolerance to cassava brown streak disease associated with the zigzag stem trait. *Roots* **8**, 15–19 (2003).
54. Mbanzibwa, D. R. *et al.* Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by RT-PCR reveals wide distribution in East Africa, mixed infections, and infections in *Manihot glaziovii*. *J Virol Methods* **171**, 394–400, doi:10.1016/j.jviromet.2010.09.024 (2011).
55. Ogwok, E., Alicai, T., Rey, M. E. C., Beyene, G. & Taylor, N. J. Distribution and accumulation of cassava brown streak viruses within infected cassava (*Manihot esculenta*) plants. *Plant Pathol* **64**, 1235–1246 (2015).
56. Kanju, E. E. *et al.* Breeding cassava for brown streak resistance: regional cassava variety development strategy based on farmer and consumer preferences. *Proceedings 13th ISTRC Symp*, 95–101 (2007).
57. Bredeson, J. V. *et al.* Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nature Biotechnol* **34**, 562–570, doi:10.1038/nbt.3535 (2016).

58. Ferguson, M. E. *et al*. Identification, validation and high-throughput genotyping of transcribed gene SNPs in cassava. *Theor Appl Genet* **124**, 685–695 (2012).
59. Sonenberg, N. & Dever, T. E. Eukaryotic translation initiation factors and regulators. *Curr Opin Struct Biol* **13**, 56–63 (2003).
60. Kang, B. C., Yeam, I. & Jahn, M. M. Genetics of plant virus resistance. *Annu Rev Phytopathol* **43**, 581–621 (2005).
61. Maule, A. J., Caranta, C. & Boulton, M. I. Sources of natural resistance to plant viruses: status and prospects. *Mol Plant Pathol* **8**, 223–231, doi:10.1111/j.1364-3703.2007.00386.x (2007).
62. Ashby, J. A., Stevenson, C. E., Jarvis, G. E., Lawson, D. M. & Maule, A. J. Structure-based mutational analysis of eIF4E in relation to sbm1 resistance to *Pea seed-borne mosaic virus* in pea. *PLoS One* **6** (2011).
63. Bauer, J. *et al*. Structure–function analysis of *Arabidopsis thaliana* histidine kinase AHK5 bound to Its cognate phosphotransfer protein AHP1. *Mol Plant* **6**, 959–970, doi:10.1093/mp/sss126 (2013).
64. Romir, J., Harter, K. & Stehle, T. Two-component systems in *Arabidopsis thaliana* – a structural view. *Eur J Cell Biol* **89**, 270–272, doi:10.1016/j.ejcb.2009.11.007 (2010).
65. Sheen, J. Phosphorelay and transcription control in cytokinin signal transduction. *Science* **296**, 1650–1652, doi:10.1126/science.1071883 (2002).
66. Spoustová, P. *et al*. Tobacco susceptibility to *Potato virus* YNTN infection is affected by grafting and endogenous cytokinin content. *Plant Sci* **235**, 25–36, doi:10.1016/j.plantsci.2015.02.017 (2015).
67. West, A. H. & Stock, A. M. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci* **26**, 369–376, doi:10.1016/S0968-0004(01)01852-7 (2001).
68. To, J. P. C. & Kieber, J. J. Cytokinin signaling: two-components and more. *Trends Plant Sci* **13**, 85–92 (2008).
69. Mähönen, A. P. *et al*. Cytokinins regulate a bidirectional phosphorelay network in Arabidopsis. *Curr Biol.* **16**, 1116–1122 (2006).
70. Choi, J. *et al*. The cytokinin-activated transcription factor ARR2 promotes plant immunity via TGA3/NPR1-dependent salicylic acid signaling in Arabidopsis. *Dev Cell.* **19**, 284–295, doi:10.1016/j.devcel.2010.07.011 (2010).
71. Naseem, M. & Dandekar, T. The role of auxin-cytokinin antagonism in plant-pathogen interactions. *PLoS Pathog* **8**, e1003026 (2012).
72. Baebler, Š. *et al*. Salicylic acid is an indispensable component of the Ny-1 resistance-gene-mediated response against *Potato virus Y* infection in potato. *J Exp Bot* **65**, 1095–1109 (2014).
73. Kobeasy, M. I., El-Beltagi, H. S., El-Shazly, M. A. & Khattab, E. A. H. Induction of resistance in *Arachis hypogaea* L. against *Peanut mottle virus* by nitric oxide and salicylic acid. *Physiol Mol Plant Path* **76**, 112–118, doi:10.1016/j.pmpp.2011.07.005 (2011).
74. Maekawa, T. *et al*. Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. *Cell Host Microbe.* **9**, 187–199, doi:10.1016/j.chom.2011.02.008 (2011).
75. Marone, D., Russo, M. A., Laido, G., De Leonardis, A. M. & Mastrangelo, A. M. Plant nucleotide binding site-leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. *Int J Mol Sci* **14**, 7302–7326, doi:10.3390/ijms14047302 (2013).
76. Ito, M., Ohnishi, K., Hikichi, Y. & Kiba, A. Molecular chaperons and co-chaperons, HSP90, RAR1, and SGT1 negatively regulate bacterial wilt disease caused by *Ralstonia solanacearum* in Nicotiana benthamiana. *Plant Signal Behav* 0, doi:10.4161/15592316.2014.970410 (2014).
77. Zhong, L. *et al*. Chloroplast small heat shock protein HSP21 interacts with plastid nucleoid protein pTAC5 and is essential for chloroplast development in Arabidopsis under heat stress. *Plant Cell* **25**, 2925–2943, doi:10.1105/tpc.113.111229 (2013).
78. Mayer, F. L. *et al*. Small but crucial: the novel small heat shock protein HSP21 mediates stress adaptation and virulence in *Candida albicans*. *PLoS One* **7**, e38584, doi:10.1371/journal.pone.0038584 (2012).
79. Neta-Sharir, I., Isaacson, T., Lurie, S. & Weiss, D. Dual role for tomato heat shock protein 21: protecting photosystem II from oxidative stress and promoting color changes during fruit maturation. *Plant Cell* **17**, 1829–1838, doi:10.1105/tpc.105.031914 (2005).
80. Lashley, T. *et al*. Molecular chaperons, amyloid and preamyloid lesions in the BRI2 gene-related dementias: a morphological study. *Neuropathol Appl Neurobiol.* **32**, 492–504, doi:10.1111/j.1365-2990.2006.00747.x (2006).
81. Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* **18**, 1169–1175 (2004).
82. Hundertmark, M. & Hincha, D. LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. *BMC Genomics* **9**, 118 (2008).
83. Liu, Y. & Zheng, Y. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. *Biochem Biophys Res Commun* **311**, 325–332 (2005).
84. Shih, M. *et al*. Gene cloning and characterization of a soybean (*Glycine max* L.) LEA protein, GmPM16. *Plant Mol Biol* **56**, 689–703 (2004).
85. Gal, T., Glazer, I. & Koltai, H. An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. *FEBS Lett* **577**, 21–26 (2004).
86. Goyal, K., Walton, L. & Tunnacliffe, A. LEA proteins prevent protein aggregation due to water stress. *Biochem J* **388**, 151–157 (2005).
87. Hand, S. C., Jones, D., Menze, M. A. & Witt, T. L. Life without water: expression of plant LEA genes by an anhydrobiotic arthropod. *J Exp Zool* **307A**, 62–66 (2007).
88. Kikawada, T. *et al*. Dehydration-induced expression of LEA proteins in an anhydrobiotic chironomid. *Biochem Biophys Res Comm* **348**, 56–61 (2006).
89. Tunnacliffe, A., Lapinski, J. & McGee, B. A putative LEA protein, but no trehalose, is present in anhydrobiotic bdelloid rotifers. *Hydrobiologia* **546**, 315–321 (2005).
90. Bai, Y. *et al*. Isolation and functional characterization of a *Medicago sativa* L. gene, MsLEA3-1. *Mol Biol Rep* **39**, 2883–2892 (2012).
91. Sels, J., Mathys, J., De Coninck, B., Cammue, B. & De Bolle, M. Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol Biochem* **46**, 941–950 (2008).
92. Guevara-Morato, M. A., G de Lacoba, M., Garcia-Luque, I. & Serra, M. T. Characterization of a pathogenesis-related protein 4 (PR-4) induced in *Capsicum chinense* L3 plants with dual RNase and DNase activities. *J Exp Bot* **61**, 3259–3271 (2010).
93. Li, X. *et al*. A new pathogenesis-related protein, LrPR4, from *Lycoris radiata*, and its antifungal activity against *Magnaporthe grisea*. *Mol Biol Rep* **37**, 995–1001 (2010).
94. Park, C. J. *et al*. Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *Plant J: Cell Mol Biol* **37**, 186–198 (2004).
95. Donze, T., Qu, F., Twigg, P. & Morris, T. J. *Turnip crinkle virus* coat protein inhibits the basal immune response to virus invasion in Arabidopsis by binding to the NAC transcription factor TIP. *Virology* **449**, 207–214, doi:10.1016/j.virol.2013.11.018 (2014).
96. Nakashima, K. *et al*. Functional analysis of a NAC-type transcription factor OsNAC6 involved in abiotic and biotic stress-responsive gene expression in rice. *The Plant Journal* **51**, 617–630 (2007).
97. Voitsik, A. M., Muench, S., Deising, H. B. & Voll, L. M. Two recently duplicated maize NAC transcription factor paralogs are induced in response to *Colletotrichum graminicola* infection. *BMC Plant Biol* **13**, 85, doi:10.1186/1471-2229-13-85 (2013).
98. Liu, B. *et al*. Tomato NAC transcription factor SlSRN1 positively regulates defense response against biotic stress but negatively regulates abiotic stress response. *PLoS One* **9**, e102067, doi:10.1371/journal.pone.0102067 (2014).
99. Kaneda, T. *et al*. The transcription factor OsNAC4 is a key positive regulator of plant hypersensitive cell death. *The EMBO Journal* **28**, 926–936 (2009).

100. Banerjee, A. & Roychoudhury, A. WRKY proteins: signaling and regulation of expression during abiotic stress responses. *Scientific World J* **2015**, 17, doi:10.1155/2015/807560 (2015).
101. Chen, L. *et al*. The role of WRKY transcription factors in plant abiotic stresses. *Biochim Biophys Acta* **1819**, 120–128, doi:10.1016/j.bbagrm.2011.09.002 (2012).
102. Pandey, S. P. & Somssich, I. E. The role of WRKY transcription factors in plant immunity. *Plant Physiol* **150**, 1648–1655, doi:10.1104/pp.109.138990 (2009).
103. Eulgem, T. & Somssich, I. E. Networks of WRKY transcription factors in defense signaling. *Curr Opin Plant Biol* **10**, 366–371, doi:10.1016/j.pbi.2007.04.020 (2007).
104. Tao, Z. *et al*. A pair of allelic WRKY genes play opposite roles in rice-bacteria interactions. *Plant Physiol* **151**, 936–948, doi:10.1104/pp.109.145623 (2009).
105. Nicaise, V. *et al*. The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus *Lettuce mosaic virus*. *Plant Physiol* **132**, 1272–1282 (2003).
106. Xu, J., Aileni, M., Abbagani, S. & Zhang, P. A reliable and efficient method for total RNA isolation from various members of spurge family (Euphorbiaceae). *Phytochem Anal.* **21**, 395–398, doi:10.1002/pca.1205 (2010).
107. Abarshi, M. M. *et al*. 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. *J Virol Methods* **179**, 176–184 (2012).
108. Hillocks, R. J., Raya, M. & Thresh, J. M. The association between root necrosis and above-ground symptoms of brown streak virus infection of cassava in southern Tanzania. *Int J Pest Manage* **42**, 285–289 (1996).
109. Andrews, S. *FastQC: A quality control tool for high throughput sequence data*, Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ Accessed: 4th June, 2014 (2010).
110. Hannon, G. *FASTX-Toolkit v*.0.0.14. *Cold Spring Harbor Laboratory*, *Long Island*. Available at: http://hannonlab.cshl.edu/fastx_toolkit/ Accessed: 4th June, 2014 (2009).
111. R: A language and environment for statistical computing. (R foundation for statistical computing, Vienna, Austria. Retrieved from: http://www.R-project.org/ (2013).
112. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111, doi:10.1093/bioinformatics/btp120 (2009).
113. Rapaport, F. *et al*. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol* **14**, R95 (2013).
114. Katari, M. S. *et al*. VirtualPlant: a software platform to support systems biology research. *Plant Physiol* **152**, 500–515, doi:10.1104/pp.109.147025 (2010).

Acknowledgements

The authors would like to thank Illumina Inc., which provided a grant through the 'Greater Good Initiative' for the cDNA library preparation and sequencing of 48 samples (101 bp), conducted in collaboration with S.Rounsley (formerly with the University of Arizona and Dow AgroSciences, currently with Genus PLC) at Dow AgroSciences, BecA and IITA. The first set of cDNA libraries (48 samples, 50 bp) was sequenced at UC Berkeley, in D.Rokhsar's laboratory. We thank S.Prochnik and T.Mansfield for performing quality checks on the raw data for UC Berkeley and Dow AgroSciences, respectively. J.Bredeson provided two weeks of Unix training to TA. The experiment was designed in consultation with K.Denby from the University of Warwick. J. Harvey (formerly with Bioscience east and central Africa, currently working with Kansas State University) provided technical support on the identification of virus inoculum and isolation of RNA. Special thanks go to the Bill and Melinda Gates Foundation for funding this study under Project OPPGD1016, 'Biotechnology Applications to Combat Cassava Brown Streak Disease'.

Author Contributions

T.A. Conducted all field, screen-house (planted experimental plants, graft-inoculated samples, collected samples) and laboratory procedures (identified inoculum, isolated RNA and prepared cDNA libraries). Performed data analysis and wrote the initial draft of the manuscript. M.S.K. Led the data analysis component and trained and supervised T.A. in all aspects of RNAseq data analysis. D.K.B. Suggested appropriate data analysis and led discussions on data interpretation. Critically reviewed and extensively edited the manuscript. Provided University supervision. A.A.M. Provided critical comments, guidance, and University supervision. S.L.G. Compiled a cassava genome database at VirtualPlant. M.E.F. Principle Investigator. Designed the study, acquired funding, identified varieties, provided day-to-day supervision, and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-09617-z

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.