Use of infectious clones to gain insights into 

*Cassava brown streak virus* gene function

Katie Rose Tomlinson

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Doctor of Philosophy in the Faculty of Science

School of Biological Sciences

September 2018

80,767 words
Abstract

Cassava is the second most important food staple in terms of per capita calories consumed in Africa and offers unique climate change adaptation opportunities. Unfortunately, production in sub-Saharan Africa is currently constrained by cassava mosaic disease (CMD) and cassava brown streak disease (CBSD), which together are estimated to cause annual losses worth US$1 billion and adversely affect food security across the entire region. Compared with CMD, relatively little is currently understood about CBSD infections and there are no cassava varieties with high levels of CBSD resistance available to farmers. CBSD is caused by at least two potyviral species: Cassava brown streak virus and Ugandan cassava brown streak virus, collectively referred to as U/CBSVs. Little is known about U/CBSV gene function; to date only the silencing suppressor activity of the UCBSV P1 protein has been reported.

Viral infectious clones are highly valuable tools that can be used to screen for viral resistance and to gain fundamental insights into viral infections. Unfortunately, until recently the construction of U/CBSV ICs has been circumvented by sequence instability during propagation in Escherichia coli. In 2014 -16, two U/CBSV ICs were constructed at the University of Bristol, as part of a larger project related to this work. The main aim of this study was to test the tractability of these ICs and use them to characterise U/CBSV gene functions.

The UCBSV IC exhibited high levels of sequence instability during propagation in all tested E. coli strains. In contrast, the CBSV IC which contains introns, demonstrated sequence stability during propagation in the E. coli strains ccdB and TOP10. This permitted the insertion of GFP at two genome positions. N. benthamiana infections with CBSV_GFP ICs enabled visualisation of viral movement in planta.

U/CBSV genomes encode Ham1-like peptides with conserved ITPase pyrophosphohydrolase motifs, which have only been reported in Euphorbia ringspot virus. Eu/prokaryotic ITPase proteins selectively hydrolyze non-canonical nucleotide triphosphates to prevent their incorporation into nucleic acid and thereby reduce mutation rates. In vitro pyrophosphohydrolase assays, demonstrated that U/CBSV Ham1 proteins have significantly higher activities with non-canonical nucleotides compared with canonical nucleotides. However, no significant differences were found in the number of viral variants during N. benthamiana infections with full-length CBSV IC and a CBSV IC containing Ham1 deletion, indicating that CBSV Ham1 may not reduce viral mutation rates.

Additionally, mutation of the CBSV IC demonstrated that a Ham1 pyrophosphohydrolase motif is associated with the development of severe necrosis during N. benthamiana infections and a chimera IC, consisting of the CBSV genome with UCBSV coat protein replacement caused systemic N. benthamiana infections; raising the possibility for transcapsidation to occur during mixed U/CBSV infections in the field. These insights into CBSV gene function demonstrates the utility of U/CBSV ICs and should aid future studies into these important viruses.
Dedication and acknowledgements

First and foremost, I would like to thank my supervisors Dr. Andy Bailey and Prof. Gary Foster for their continued support throughout my PhD. I really appreciate the time and energy they have invested to improve my confidence, nurture my scientific abilities and provide opportunities that have expanded my horizons from the beyond the lab to the real world. I have loved being a part of the Molecular Plant Pathology group at the University of Bristol. I am especially grateful to Dr. Alice Banks, Dr. Sarah Nanyiti and Dr. Anna Tiley, who spent time training me at the start of my PhD. I also would like to thank my good friend Dr. Luis Pablo-Rodriguez, who taught me techniques, talked about viruses (a lot) and made working in the lab fun, especially the singing and dancing!

I am extremely thankful for my experiences visiting the National Crops Resources Research Institute in Uganda in 2016. I would like to thank Dr. Titus Alicai and his team who generously welcomed me into the Root Crops Research Programme during my internship. Seeing the fantastic work being done at NaCRRI has inspired me to continue working in this research area.

I would also like to thank my friend Everline Odrumuru and her lovely family, who kindly made me feel at home in Uganda.

I would like to give thanks to my parents Peter and Rosie Tomlinson. They have shown endless love and support right from an early age but also now as an adult. Their encouragement continues to be a great source of strength that I am extremely grateful for. I would also like to thank my sister Amy for her love and encouragement; I am so proud of you!

I have also had some incredible support from friends at the University of Bristol including Dora, Donald, Fiona, Noriane, Ashley and Francesca de la Torre, with whom I have shared good times over coffee, cake and beer. I would really like to thank Dr. Kelly Atkins who has been there through the highs and lows and never fails to make me smile. I am so excited to see your career move in a new direction and cannot wait to celebrate with you!

I would like to dedicate my thesis to my dear friend Nina Trott, who sadly passed away in 2017. I was very lucky to have grown-up under her friendship; her passion for life continues to inspire me.

Finally Joe, I cannot put into words how much I love you. These last few years have not always been easy, but you have always been there to listen, encourage and support me through even the hardest days. I am so excited to share the next stages of our lives together and to start some new adventures with you!
Author’s declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University’s Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate’s own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ………………………………………… DATE: …16/01/2019…………………………

Katie Rose Tomlinson
# Table of Contents

Chapter 1: Introduction ........................................................................................................... 1
  1.1 Global food security ........................................................................................................ 1
  1.2 Plant diseases ................................................................................................................. 2
  1.3 Plant viruses .................................................................................................................. 3
  1.4 Cassava .......................................................................................................................... 3
    1.4.1 Uses of cassava for food and industrial products ......................................................... 6
  1.5 Constraints on cassava production in Africa ................................................................. 6
    1.5.1 Bacterial diseases ...................................................................................................... 6
    1.5.2 Fungal diseases ......................................................................................................... 6
    1.5.3 Pests .......................................................................................................................... 7
      1.5.3.1 *Bemisia tabaci* ..................................................................................................... 7
    1.5.4 Viral diseases ............................................................................................................ 7
  1.6 Cassava mosaic disease ................................................................................................... 8
  1.7 Cassava brown streak disease ......................................................................................... 9
    1.7.1 CBSD epidemiology .................................................................................................... 10
    1.7.2 CBSD transmission and dispersal .............................................................................. 10
    1.7.3 Geographical distribution of U/CBSVs ....................................................................... 11
    1.7.4 CBSD symptoms ........................................................................................................ 11
    1.7.5 Responses of different cassava varieties to U/CBSVs .................................................... 12
    1.7.6 U/CBSV infections of herbaceous hosts ..................................................................... 13
    1.7.7 Quantification of U/CBSVs during infections ............................................................... 13
    1.7.8 Control of CBSD ........................................................................................................ 14
  1.8 Properties of plant viral genomes ................................................................................... 15
  1.9 Viral quasi-species and the critical error threshold ....................................................... 15
  1.10 *Potyviridae* family ..................................................................................................... 16
    1.10.1 *Potyviridae* genomes ............................................................................................ 17
    1.10.2 Genera in the *Potyviridae* family ........................................................................... 17
    1.10.3 Ipomoviruses ............................................................................................................. 18
      1.10.3.1 *Sweet potato mild mottle virus* ......................................................................... 19
      1.10.3.2 *Cucumber vein yellowing virus* ...................................................................... 20
      1.10.3.3 *Squash vein yellowing virus* ............................................................................ 20
      1.10.3.4 *Tomato mild mottle virus* .............................................................................. 20
      1.10.3.5 CBSV and UCBSV ........................................................................................... 20
1.11 Overview of the typical Potyviridae infection cycle .................................................. 22
  1.11.1 Virion disassembly, polyprotein translation and processing .......................... 22
  1.11.2 Viral genome replication ......................................................................................... 23
  1.11.3 Intra/intercellular and long-distance viral movement ........................................... 23
  1.11.4 Suppression of host silencing ................................................................................. 24
1.12 Functions of different proteins in other Potyviridae .................................................. 24
  1.12.1 P1 ......................................................................................................................... 25
  1.12.2 HC-Pro ................................................................................................................... 25
  1.12.3 P3 and P3N-PIPO ................................................................................................. 25
  1.12.4 6K1 and 6K2 ......................................................................................................... 26
  1.12.5 CI ....................................................................................................................... 26
  1.12.6 VPg ...................................................................................................................... 27
  1.12.7 Nla ..................................................................................................................... 27
  1.12.8 Nlb .................................................................................................................... 28
  1.12.9 CP ................................................................................................................ ...... 28
  1.12.10 5’ and 3’ UTRs ................................................................................................. 28
1.13 Plant-virus interactions ............................................................................................... 29
  1.13.1 Recessive resistance .............................................................................................. 29
  1.13.2 Dominant resistance ............................................................................................ 30
1.14 Viral sequence determinants ....................................................................................... 30
1.15 ICs as tool to study gene functions and breed for resistance ..................................... 30
1.16 Aims and objectives of this study ............................................................................... 31

Chapter 2: Materials and Methods ................................................................................. 33
2.1 Microbiological methods and strains ........................................................................... 33
  2.1.1 Strains .................................................................................................................... 33
    2.1.1.1 Escherichia coli ............................................................................................... 33
    2.1.1.2 Agrobacterium tumefaciens ........................................................................... 33
    2.1.1.3 Saccharomyces cerevisiae ............................................................................. 33
    2.1.1.4 U/CBSV strains ............................................................................................ 34
  2.1.2 Culture media ......................................................................................................... 34
    2.1.2.1 Bacterial media .............................................................................................. 34
    2.1.2.2 S. cerevisiae media ....................................................................................... 34
    2.1.2.3 Antibiotic selection ...................................................................................... 35
    2.1.2.4 Microbial culture and storage ..................................................................... 35
2.1.2.4.1  *E. coli* .......................................................... 35
2.1.2.4.2  *A. tumefaciens* .............................................. 35
2.1.2.4.3  *S. cerevisiae* ................................................. 35

2.2  Molecular methods ........................................................................ 36
  2.2.1.1  Small scale (miniprep) *E. coli* plasmid extraction.................. 36
  2.2.1.2  Medium scale (midiprep) *E. coli* plasmid extraction............. 36
  2.2.1.3  Large scale (maxiprep) *E. coli* plasmid extraction............... 36
  2.2.2  *S. cerevisiae* plasmid extraction ........................................ 37
  2.2.3  Plant RNA extraction ....................................................... 37
        2.2.3.1  RNA extraction from *Nicotiana* sp. using TRizol ............. 38
        2.2.3.2  RNA extraction from *N. benthamiana* using the E.Z.N.A. Plant RNA ...... 38
        2.2.3.3  CTAB RNA extraction from cassava ............................... 38
  2.2.4  Nucleic acid manipulation .................................................... 39
        2.2.4.1  Agarose gel electrophoresis ......................................... 39
        2.2.4.2  Agarose gel DNA purification ................................. 40
  2.2.5  Polymerase chain reaction .................................................... 40
        2.2.5.1  Primers ................................................................ 41
        2.2.5.2  Low fidelity PCR .................................................... 41
        2.2.5.3  Culture PCR .......................................................... 41
        2.2.5.4  High fidelity PCR .................................................... 41
        2.2.5.5  Reverse-transcription PCR ......................................... 42
        2.2.5.6  Quantitative reverse-transcription PCR ............................ 43
  2.2.6  Enzymatic restriction digest .................................................... 44
        2.2.6.1  Fast enzymatic restriction digest .................................... 44
        2.2.6.2  High-fidelity restriction digest ...................................... 44
  2.2.7  Cloning ........................................................................ 44
        2.2.7.1  Cloning PCR products into the pJET2.1 vector .................. 44
        2.2.7.2  Cloning PCR products into the POPINF vector ................. 45
        2.2.7.3  Infectious clone construction using yeast homologous recombination ... 45
  2.3  Bacterial transformation ............................................................... 46
        2.3.1  Preparation of electro-competent *E. coli* ........................... 46
        2.3.2  Preparation of electro-competent *A. tumefaciens* ................. 47
        2.3.3  Electroporation ................................................................ 47
        2.3.3.1  *E. coli* electroporation .............................................. 48
2.3.3.2 A. tumefaciens electroporation ................................................................. 48
2.4 Sequence analysis ............................................................................................. 48
  2.4.1 DNA sequencing .......................................................................................... 48
  2.4.2 Plasmid maps .............................................................................................. 48
  2.4.3 DNA sequence alignments .......................................................................... 49
  2.4.4 Phylogenetic sequence analysis .................................................................. 49
  2.4.5 Bio-informatic sequence analysis ............................................................... 49
2.5 Protein Analysis ................................................................................................. 50
  2.5.1 Protein expression and purification in E. coli ............................................. 50
  2.5.2 Quantification of U/CBSV Ham1 protein concentration .......................... 51
  2.5.3 SDS-PAGE protein analysis ...................................................................... 51
  2.5.4 Pyrophosphohydrolase assay ..................................................................... 52
    2.5.4.1 Calculation of phosphate concentration in pyrophosphohydrolase assays .... 53
2.6 Plant methods ................................................................................................... 54
  2.6.1 Plant maintenance ...................................................................................... 54
  2.6.2 Viral inoculations of plants ........................................................................ 54
    2.6.2.1 Agroinfiltration of plants ..................................................................... 54
    2.6.2.2 Mechanical inoculation with infected material ...................................... 55
    2.6.2.3 In vitro generation of viral transcripts ................................................. 55
2.7 Serological methods ........................................................................................ 56
2.8 Microscopy ........................................................................................................ 57
  2.8.1 Wide-field fluorescence microscopy .......................................................... 57
  2.8.2 Confocal fluorescence microscopy .............................................................. 58
2.9 Statistical analysis ........................................................................................... 58

Chapter 3: Tractability of two cassava brown streak disease infectious clones ........................................ 59
3.1 Introduction ........................................................................................................ 59
  3.1.1 Construction of Potyviridae ICs ................................................................. 59
  3.1.2 In vitro and in vivo transcription of ICs ....................................................... 60
  3.1.3 Genome sequence instability during Potyviridae IC propagation in E. coli .... 62
  3.1.4 Applications of plant viral ICs ................................................................... 62
    3.1.4.1 The use of plant viral ICs in resistance breeding ................................. 62
    3.1.4.2 The use of viral ICs to visualise plant infections ...................................... 63
    3.1.4.3 The use of plant viral ICs to characterise sequence determinants .......... 63
  3.1.5 Potential applications of U/CBSV ICs ...................................................... 64
3.1.6 Previous work at the UoB to construct U/CBSV ICs............................................. 64
  3.1.6.1 Construction and infectivity of the UCBSV Kikombe IC ................................. 65
  3.1.6.2 Construction and infectivity of the CBSV Tanza IC.......................................... 66
3.1.7 Hypothesis .............................................................................................................. 68
3.1.8 Aims ........................................................................................................................ 68
3.2 Results .......................................................................................................................... 69
  3.2.1 Sequence stability of the UCBSV Kikombe IC ......................................................... 69
    3.2.1.1 UCBSV Kikombe IC propagation in the E. coli strain α-select ....................... 69
    3.2.1.2 UCBSV Kikombe IC propagation in the E. coli strain DH5α ......................... 71
    3.2.1.3 UCBSV Kikombe IC propagation in the E. coli strain ccdB ......................... 72
  3.2.2 Sequence stability of the CBSV Tanza IC .............................................................. 72
    3.2.2.1 CBSV Tanza IC propagation in the E. coli strain C43 ................................. 72
    3.2.2.2 CBSV Tanza IC propagation in the E. coli strain ccdB ............................... 74
  3.2.3 Agroinfiltration of cassava with the CBSV Tanza IC ............................................. 75
  3.2.4 Insertion of a marker gene into the CBSV Tanza IC ............................................. 77
    3.2.4.1 Cloning design for construction of CBSV_Tanza_GFP1 ................................ 77
    3.2.4.2 PCR design for the construction of CBSV_Tanza_GFP1 ............................... 77
    3.2.4.3 Cloning of the CBSV_Tanza_GFP1 IC ......................................................... 80
    3.2.4.4 Verification of CBSV_Tanza_GFP1 construction ......................................... 81
    3.2.4.5 N. benthamiana infections with CBSV_Tanza_GFP1 ...................................... 82
    3.2.4.6 GFP expression during CBSV_Tanza_GFP1 infections .................................... 82
    3.2.4.7 RT-PCR detection of CBSV_Tanza_GFP1 during infections ......................... 83
    3.2.4.8 Symptom development during CBSV_Tanza_GFP1/2 infections ..................... 83
  3.2.5 Cloning design for the construction of CBSV_Tanza_GFP2 ..................................... 83
    3.2.5.1 PCR design for construction of CBSV_Tanza_GFP2 .................................... 85
    3.2.5.2 Cloning of the CBSV_Tanza_GFP2 IC ......................................................... 85
    3.2.5.3 Verification of CBSV_Tanza_GFP2 construction ......................................... 86
  3.2.6 Infections with CBSV_Tanza_GFP1, CBSV_Tanza_GFP2 and CBSV_Tanza ........... 87
    3.2.6.1 GFP expression during CBSV_Tanza_GFP1/2 infections ................................ 87
    3.2.6.2 RT-PCR detection of CBSV_Tanza_GFP1/2 during infections ....................... 88
    3.2.6.3 RT-PCR detection of viral transcripts with GFP deletions ............................. 89
    3.2.6.3 Quantification of CBSV_Tanza_GFP1/2 viral titers ...................................... 90
    3.2.6.4 Symptom development during CBSV_Tanza_GFP1/2 infections ................... 91
    3.2.6.5 Mechanical-inoculation with CBSV_Tanza_GFP1/2 ...................................... 91
3.2.6.6 Attempts to construct the CBSV_Tanza_iLOV IC ............................................. 92
3.2.7 Construction of a CBSV Tanza IC containing UCBSV coat protein replacement ........ 92
  3.2.7.1 Cloning design for construction of CBSV_Tanza_UCP ......................................... 93
  3.2.7.2 PCR design for construction of CBSV_Tanza_UCP .............................................. 94
  3.2.7.3 Cloning of CBSV_Tanza_UCP IC .................................................................. 96
  3.2.7.4 Verification of CBSV_Tanza_UCP construction ................................................. 96
3.2.8 N. benthamiana infections with CBSV_Tanza_UCP .............................................. 97
  3.2.8.1 Symptom development during CBSV_Tanza_UCP infections ............................... 98
  3.2.8.2 Quantitative assessment of symptom development ........................................... 99
  3.2.8.3 RT-PCR detection of CBSV_Tanza_UCP infections ........................................... 100
  3.2.8.4 Viral titers during CBSV_Tanza_UCP infections ............................................... 100
  3.2.8.5 Back-inoculation of N. benthamiana with CBSV_Tanza_UCP ......................... 101
  3.2.8.6 Attempts to infect cassava with CBSV_Tanza_UCP .......................................... 102
  3.2.8.7 Attempts to construct a UCBSV IC containing CBSV CP replacement .............. 102
3.3 Discussion .............................................................................................................. 103
  3.3.1 Sequence instability of the UCBSV Kikombe and CBSV Tanza ICs in E. coli .......... 103
  3.3.2 Agroinfiltration of cassava with CBSV_Tanza IC .................................................. 104
  3.3.3 Use of CBSV_Tanza_GFP1/2 ICs to visualise CBSV replication and movement .... 105
  3.3.4 Stability of GFP expression in CBSV_Tanza_GFP1/2 infections ............................ 105
  3.3.5 Use of the CBSV_Tanza_UCP IC to characterise sequence determinants ............. 107
  3.3.6 Conclusions .................................................................................................... 108

Chapter 4: Potential roles of U/CBSV Ham1 proteins during infection ..................... 109
4.1 Introduction .......................................................................................................... 109
  4.1.1 Non-canonical nucleotides and their deleterious effects ...................................... 109
    4.1.1.3 Generation of non-canonical nucleotides ........................................................ 111
  4.1.2 House-cleaning NTPase proteins ...................................................................... 113
    4.1.2.1 ITPases function to reduce mutations rates .................................................... 113
    4.1.2.2 ITPA mutagenic protection in E. coli .............................................................. 114
    4.1.2.3 ITPA mutagenic protection in yeast ................................................................. 114
    4.1.2.4 ITPA mutagenic protection in mammals ......................................................... 114
  4.1.3 Viral Ham1 proteins .......................................................................................... 115
    4.1.3.1 Euphorbia ringspot virus Ham1 protein .......................................................... 115
    4.1.3.2 U/CBSV Ham1 proteins .................................................................................. 115
  4.1.4 Insights into viral diversity using deep sequencing ............................................ 116

X
4.1.5 Previous work on U/CBSV Ham1s at the UoB .............................................. 117
  4.1.5.1 Investigations to detect U/CBSV Ham1 mutagenic protection .............. 117
  4.1.5.2 Effect of CBSV Ham1 on viral mutation rate ..................................... 117
  4.1.5.3 CBSV Ham1 association with symptom development and viral accumulation ................................................................. 117
4.1.6 Hypotheses .......................................................................................... 119
4.1.7 Aims ...................................................................................................... 119
4.2 Results ...................................................................................................... 120
  4.2.1 Bioinformatic analysis of U/CBSV Ham1 sequences ................................. 120
    4.2.1.1 Pyrophosphohydrolase motifs in U/CBSV Ham1 sequences ................ 120
    4.2.1.2 Phylogenetic relationships between U/CBSV Ham1s and ITPA homologs 120
  4.2.2 Construction of CBSV_mutHam .............................................................. 124
    4.2.2.1 Cloning design for the construction of CBSV_mutHam ......................... 125
    4.2.2.2 Cloning of CBSV_mutHam IC ........................................................... 126
    4.2.2.3 Verification of CBSV_mutHam IC construction .................................... 127
  4.2.3 Construction of CBSV_UHam with UCBSV Ham1 replacement .................. 128
    4.2.3.1 Cloning design for construction of CBSV_UHam IC ......................... 129
    4.2.3.2 PCR design for the construction of CBSV_Tanza_GFP1 ...................... 130
    4.2.3.3 Cloning of CBSV_UHam IC ............................................................... 130
    4.2.3.4 Verification of CBSV_UHam IC construction ..................................... 130
  4.2.4 Infections with CBSV Tanza ICs containing Ham1 manipulations ............ 132
    4.2.4.1 Confirmation of N. benthamiana infections with CBSV Tanza ICs .... 133
    4.2.4.2 Symptom development during infections with CBSV Tanza ICs ........ 133
    4.2.4.3 Quantification of viral titres during infections by TAS-ELISA ............. 135
    4.2.4.4 Quantification of viral transcripts during infections by qPCR ............. 136
    4.2.4.5 Exp. of N. benthamiana ITPA throughout infection with CBSV Tanza ICs. 137
  4.2.5 Potential role of CBSV Ham1 in reducing CBSV mutation rate ............... 139
    4.2.5.1 N. benthamiana infections used in deep sequencing analysis ............ 139
    4.2.5.2 RNA extractions from N. benthamiana infections ............................... 139
    4.2.5.3 Viral RT-PCR amplicon design for deep sequencing analysis ............. 140
    4.2.5.4 Amplification of viral RT-PCR amplicons .......................................... 140
    4.2.5.5 Deep sequencing of viral RT-PCR amplicons .................................... 141
    4.2.5.6 Reference sequence for alignment of deep sequencing reads ............. 141
    4.2.5.7 Bioinformatic analysis ................................................................. 141
  4.2.5.8 Exp. of N. benthamiana ITPA throughout infection with CBSV Tanza ICs. 137
  4.2.5.9 Deep sequencing of viral RT-PCR amplicons .................................... 141
  4.2.5.10 Reference sequence for alignment of deep sequencing reads ............. 141
alignment of deep sequencing reads to the reference sequence........ 142
4.2.5.9 In-depth read quality analysis .......................................................... 142
4.2.5.10 Low-frequency SNV calling from deep sequence reads .................. 143
4.2.5.11 Potential correlation between SNV no. and CBSV transcript abundance 145
4.2.5.12 Mapping SNVs to reference sequence ............................................ 146

4.3 Discussion .................................................................................................... 147
4.3.1 Bioinformatic analysis of CBSV Tanza and UCBSV Kikombe Ham1 sequences 147
4.3.2 \textit{N. benthamiana} infections with a range of CBSV ICs ......................... 148

4.3.2.1 \textit{N. benthamiana} infections with CBSV\textsubscript{mutHam} ....................... 148
4.3.2.2 \textit{N. benthamiana} infections with CBSV\textsubscript{UHam} ......................... 149
4.3.2.3 \textit{N. benthamiana} infections with CBSV\textsubscript{HKO} .......................... 150
4.3.2.4 Expression of \textit{N. benthamiana} ITPA in response to CBSV infection ......... 150
4.3.3 Effect of CBSV Ham1 on CBSV mutation rate .......................................... 151
4.3.4 Position of SNVs in reference sequence .................................................. 152
4.3.5 Conclusions ............................................................................................... 152

Chapter 5: Investigation into potential U/CBSV Ham1 pyrophosphohydrolase activities ...... 153

5.1 Introduction .................................................................................................... 153
5.1.1 ITPA enzymatic activities ........................................................................ 153
5.1.2 ITPA protein structures ........................................................................... 154
5.1.3 Model of ITPA substrate binding and pyrophosphohydrolase activity ........ 155
5.1.4 Hypotheses ............................................................................................... 156
5.1.5 Aims .......................................................................................................... 156

5.2 Results .......................................................................................................... 157
5.2.1 Structural similarity of CBSV Tanza and UCBSV Kikombe with hITPA ........ 157
5.2.2 Expression of the wild-type CBSV Tanza Ham1 proteins ......................... 158

5.2.2.1 Cloning design for the construction of the POPINF\_Cham\_WT plasmid ... 159
5.2.2.2 Construction of the POPINF\_Cham\_WT plasmid .................................. 160
5.2.2.3 Verification of the POPINF\_Cham\_WT plasmid construction .................. 160
5.2.2.4 Expression of wild-type CBSV Tanza Ham1 and purification ............... 161
5.2.2.5 Attempts to purify CBSV Tanza Ham1 protein for X-ray crystallography... 163
5.2.3 Expression and purification of UCBSV Kikombe Ham1 protein .................. 164
5.2.4 Pyrophosphohydrolase enzyme assays ................................................... 165

5.2.4.1 Phosphate concentration quantification ................................................ 166
5.2.4.2 CBSV Tanza Ham1 pyrophosphohydrolase activity with dITP .............. 167
5.2.4.3 Heat inactivation of CBSV Ham1 pyrophosphohydrolase activity .......... 168
5.2.4.4 Enzyme assay controls .................................................. 169
5.2.4.5 U/CBSV Ham1 activities with a range of nucleotides ....................... 169
5.2.4.6 Statistical analysis of U/CBSV Ham1 activities ................................ 170
5.2.5 Potential U/CBSV Ham1 activity with m7G cap triphosphate ........... 171
5.2.6 Involvement of SHR motif in CBSV Tanza Ham1 activity .................. 172
5.2.6.1 Construction and sequence verification of POPINF_CHam_Mut ........ 172
5.2.6.2 Expression of mutated CBSV Ham1 sequence in *E. coli* ............... 173
5.3 Discussion .............................................................................. 174
5.3.2 Similar CBSV and UCBSV Ham1 predicted protein structures and activities .... 175
5.3.3 Optimal dITP concentration for CBSV Tanza Ham1 dITPase activity ........ 176
5.3.4 Requirement of CBSV Tanza Ham1 protein structure for dITPase activity ...... 176
5.3.5 Future experiments to further investigate U/CBSV Ham1 activities ........ 177
5.3.6 Expression and purification of the wild-type CBSV Tanza Ham1 in *E. coli* .... 177
5.3.7 Expression and purification of mutant CBSV Tanza Ham1 protein in *E. coli*..... 178

**Chapter 6: General discussion** .................................................................. 180
6.1 Introduction ....................................................................................... 180
6.2 Tractability of the UCBSV Kikombe and CBSV Tanza ICs ..................... 181
6.3 Model hosts used to study U/CBSV infections ...................................... 184
6.4 Visualisation of CBSV Tanza replication and movement ......................... 184
6.5 U/CBSV sequence characterisation ...................................................... 185
6.6 Potential interactions between CBSV and UCBSV during mixed infections .... 187
6.7 U/CBSV Ham1 proteins ........................................................................ 188
6.8 Summary ........................................................................................... 192
6.9 Conclusions ....................................................................................... 193
6.10 Recommendations for future research ................................................ 193

References ................................................................................................... 194
Appendices .................................................................................................... 2121
List of Tables

Table 1.1 Viruses reported to infect cassava in Africa ......................................................... 8
Table 1.2: Description of genera in the *Potyviridae* family ............................................. 18
Table 2.1: Preparation of antibiotics used in this study ......................................................... 35
Table 2.2: Sequences of primers used in qPCR in this study ............................................... 43
Table 2.3: Nucleotide substrates used in pyrophosphohydrolase assays ............................... 53
Table 3.1: Primers used to amplify UCBSV Kikombe genome sections by PCR ...................... 69
Table 3.2: Primers used to amplify CBSV Tanza genome sections by PCR ............................. 73
Table 3.3: Sequence instability detected in CBSV_Tanza propagated in the *E. coli* C43 .......... 73
Table 3.4: Primers used to amplify PCR fragments to construct the CBSV_Tanza_GFP1 ........ 79
Table 3.5: Primers used to amplify PCR fragments used to construct CBSV_Tanza_GFP2 ...... 85
Table 3.6: Primers used in RT-PCR to detect GFP in infections with CBSV_Tanza_GFP1/2 ...... 89
Table 3.7: Primers used to amplify PCR fragments to construct CBSV_Tanza_UCP............... 95
Table 3.8: Symptom scoring system used to assess *N. benthamiana* infections ................. 99
Table 4.1: Primers used to amplify F1 - F2 with nucleotide sequence SAA mutation ............ 125
Table 4.2: Primers used to amplify PCR fragments to construct CBSV_UHam ..................... 129
Table 4.3: Infectious clones used to agroinfiltrate *N. benthamiana* .................................. 132
Table 4.4: Symptom development during *N. benthamiana* infections with CBSV ICs .......... 134
Table 4.5: Alignments of CBSV_Tanza and CBSV_HKO deep sequencing reads .................... 142
Table 4.6: Number of SNVs detected per million nucleotide of aligned sequencing reads ... 144
Table 5.1: Primers used to amplify CBSV Tanza Ham1 sequence by PCR ............................ 159
List of Figures

Figure 1.1: Estimated cassava production across sub-Saharan Africa in 2016 .......................... 4
Figure 1.2: Illustrative picture of cassava; cross-section of cassava roots and healthy appearing cassava leaves ................................................................. 5
Figure 1.3: Distribution of CMD across sub-Saharan Africa in the 1920s – 1990s ................. 9
Figure 1.4: Distribution of CBSD in coastal East Africa in the 1930s – 2000s ...................... 10
Figure 1.5: CBSD storage root necrosis, radial root constrictions, foliar chlorosis and brown streaks or lesions on stems ......................................................... 12
Figure 1.6: N. clevelandii plants after inoculation with a range of U/CBSV isolates .......... 13
Figure 1.7: Left: Schematic diagram of a potyvirus particle. Right: Negative contrast electron micrograph of Plum pox virus particles ........................................ 16
Figure 1.8: Genome organization of a typical potyvirus ..................................................... 17
Figure 1.9: Schematics of Ipomovirus genome structures .................................................... 19
Figure 1.10: Diagram of key processes that occur during a typical Potyviridae infection .... 22
Figure 2.1 Molecular weight ladders run by electrophoresis on 1% agarose gels ............. 40
Figure 2.2: SDS-PAGE band profile of PageRuler Protein Ladder ....................................... 52
Figure 3.1: Left: Schematic for the in vitro generation of infectious transcripts from an IC template. Right: Schematic for transformation of agroinfiltrated plant cells to enable in vivo transcription of viral ICs. ........................................ 61
Figure 3.2: Plasmid map of the UCBSV_Kikombe IC ...................................................... 65
Figure 3.3: N. benthamiana symptoms after inoculation with UCBSV_Kikombe IC transcripts at 21 dpi, compared to symptomless leaves of non-inoculated plant ....................... 66
Figure 3.4: Plasmid map of the CBSV_Tanza IC ......................................................... 67
Figure 3.5: Symptom development during Nicotiana sp. infections with CBSV_Tanza IC .... 67
Figure 3.6: PCR targeting eight UCBSV Kikombe genome sections ............................... 70
Figure 3.7: Sequence alignment of UCBSV Kikombe IC plasmid propagated in E. coli DH5α... 71
Figure 3.8: PCR on CBSV_Tanza propagated in the E. coli strain C43. .............................. 73
Figure 3.9: PCR targeting CBSV Tanza genome sections to verify sequence integrity ........ 74
Figure 3.10: RT-PCR detection of CBSV in cassava plants agroinfiltrated with CBSV_Tanza ... 75
Figure 3.11: An upper systemic cassava leaf showing chlorosis typical for CBSV infection after agroinfiltration with CBSV_Tanza IC at 42 dpi .............................. 76
Figure 3.12: RT-PCR detection of CBSV in one N. benthamiana plant after mechanical inoculation with CBSV_Tanza IC infected cassava material .................................. 76
Figure 3.13: Schematic for the construction of the CBSV_Tanza_GFP1 IC through homologous yeast recombination ............................................. 78
Figure 3.14: Schematic for the insertion of the GFP marker gene sequence between Ham1 and CP regions in the CBSV_Tanza IC to produce the CBSV_Tanza_GFP1 IC ........ 79
Figure 3.15: Schematic of the CBSV_Tanza_GFP1 IC genome with GFP sequence between the Ham1 and CP regions ..................................................... 80
Figure 3.16: Left: Gel purified PCR fragments used to construct CBSV_Tanza_GFP1. Right: Restriction digest of CBSV_Tanza IC with BamH1 used to construct CBSV_Tanza_GFP1. 80
Figure 3.17: Restriction digest of CBSV_Tanza_GFP1 plasmids to indicate that no significant sequence rearrangements had occurred during construction ....................... 81
Figure 3.18: PCR targeting CBSV_Tanza_GFP1 genome section to verify successful IC construction .................................................. 81
Figure 3.19: Confocal microscope images of GFP fluorescence in the epidermal cells of an inoculated N. benthamiana leaf at 5 dpi after agroinfiltration with the CBSV_Tanza_GFP1 IC and in the epidermal cells of an upper systemic N. benthamiana leaf at 15 dpi after agroinfiltration with CBSV_Tanza_GFP1 IC ........................................... 82
Figure 3.20: RT-PCR detection of GFP fragment in systemic leaves of *N. benthamiana* after agroinfiltration with CBSV_Tanza_GFP1 at 10 dpi.

Figure 3.21: Schematic for the insertion of the GFP marker gene sequence between the CP and 3’UTR regions of the CBSV_Tanza IC to produce the CBSV_Tanza_GFP2 IC.

Figure 3.22: Schematic of the CBSV_Tanza_GFP2 IC with GFP between the CP and 3’UTR.

Figure 3.23: Restriction digest of CBSV_Tanza_GFP2 plasmids with *Hind* III to indicate that no significant sequence rearrangements had occurred during construction.

Figure 3.24: PCR amplification of genome sections from two CBSV_Tanza_GFP2 plasmids to indicate the successful insertion of the GFP sequence into the CBSV_Tanza IC.

Figure 3.25: Confocal microscopy images of GFP fluorescence in the epidermal cells of *N. benthamiana* leaves and in the epidermal and mesophyll cells of systemic *N. benthamiana* leaves at 15 dpi after agroinfiltration with CBSV_Tanza_GFP1/2.

Figure 3.26: Wide-field microscope images of GFP expression in the vascular system and leaf lamina of systemic *N. benthamiana* leaves at 15 dpi after agroinfiltration with CBSV_Tanza_GFP1/2.

Figure 3.27: RT-PCR detection of the GFP in systemic leaves of *N. benthamiana* plants agroinfiltrated with CBSV_Tanza_GFP1/2.

Figure 3.28: Left: RT-PCR detection of viral transcripts containing GFP deletions in *N. benthamiana* agroinfiltrated with CBSV_Tanza_GFP1 at 15 dpi. Right: RT-PCR detection of viral transcripts containing GFP deletions in *N. benthamiana* agroinfiltrated with CBSV_Tanza_GFP2 at 15 dpi.

Figure 3.29: QPCR quantification of CBSV titers in *N. benthamiana* agroinfiltrated with CBSV_Tanza, CBSV_Tanza_GFP1, and CBSV_Tanza_GFP2 at 15 dpi.

Figure 3.30: *N. benthamiana* infections with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 were symptomless compared to infections with CBSV_Tanza.

Figure 3.31: ClustalW alignment of CBSV Tanka and UCBSV Kikombe CP sequences.

Figure 3.32: Schematic for the construction of the CBSV_Tanza_UCP.

Figure 3.33: Schematic for the replacement of the CBSV CP with UCBSV CP sequence in the chimera CBSV_Tanza_UCP IC.

Figure 3.34: Schematic of the CBSV_Tanza_UCP IC genome consisting of the CBSV Tanka genome with a UCBSV Kikombe CP replacement.

Figure 3.35: Left: Gel purified PCR fragments used to construct the CBSV_Tanza_UCP IC.

Figure 3.36: Restriction digest of CBSV_Tanza_UCP plasmids with *Hind* III to indicate that no significant sequence rearrangements had occurred during construction.

Figure 3.37: PCR targeting the six genome sections of CBSV_Tanza_UCP plasmids to verify successful construction.

Figure 3.38: Symptom development during *N. benthamiana* infections with CBSV_Tanza and CBSV_Tanza_UCP, compared to symptomless non-inoculated plant at 18 dpi.

Figure 3.39: Symptom severity scores throughout *N. benthamiana* infections with CBSV_Tanza and CBSV_Tanza_UCP ICs.

Figure 3.40: RT-PCR detection of UCBSV Kikombe CP in systemic leaf material from *N. benthamiana* agroinfiltrated with CBSV_Tanza_UCP at 14 dpi.

Figure 3.41: TAS-ELISA quantification of CBSV titers during *N. benthamiana* infections with CBSV_Tanza and CBSV_Tanza_UCP.

Figure 3.42: Symptom development during infections of *N. benthamiana* inoculated with CBSV_Tanza and CBSV_Tanza_UCP infected material.

Figure 4.1: Molecular structures of canonical nucleotides ATP and GTP and non-canonical nucleotides ITP and XTP.

Figure 4.2: Schematic for canonical and non-canonical base pairing.

Figure 4.3: Purine metabolism and generation of non-canonical nucleotides ITP and XTP.
Figure 4.4: Symptom development during CBSV_Tanza and CBSV_HKO infections of N. benthamiana at 15 dpi ................................................................. 118
Figure 4.5: Quantification of CBSV titres during N. benthamiana infections with CBSV_Tanza and CBSV_HKO by TAS-ELISA ........................................................................ 118
Figure 4.6: ClustalW alignment of CBSV Tanza and UCBSV Kikombe Ham1 amino acid sequences with ITPA homologs ........................................................................ 121
Figure 4.7: Phylogenetic relationships between CBSV Tanza and UCBSV Kikombe Ham1 with predicted ITPA homolog amino acid sequences ................................................................ 122
Figure 4.8: Phylogenetic tree of 19 CBSV and 29 UCBSV Ham1 nucleotide sequences ...................................................................................... 123
Figure 4.9: Schematic for the construction of the CBSV_mutHam IC .................................................. 124
Figure 4.10: Schematic for PCR site directed mutagenesis of CBSV Tanza Ham1 SHR pyrophosphohydrolase motif to SAA. ................................................................. 125
Figure 4.11: Gel purified PCR fragments used to construct CBSV_mutHam ........................................... 126
Figure 4.12: Restriction digest of CBSV_Tanza with Bam H1 used to construct CBSV_mutHam IC through homologous yeast recombination ................................................. 126
Figure 4.13: Restriction digest of CBSV_mutHam plasmids with Hind III to indicate that no significant sequence rearrangements had occurred during construction .......... 127
Figure 4.14: PCR targeting the unstable regions of the CBSV Tanza genome to indicate successful CBSV_mutHam IC construction ........................................................................ 127
Figure 4.15: Schematic for the construction of the chimeric CBSV_UHam IC ........................................ 128
Figure 4.16: Schematic showing the replacement of CBSV Tanza Ham1 sequence with UCBSV Ham1 sequence in the CBSV_UHam IC ................................................................. 129
Figure 4.17: Schematic of the CBSV_UHam IC, consisting of the CBSV Tanza genome with a UCBSV Ham1 replacement ..................................................................................... 129
Figure 4.18: Gel purified PCR fragments used to construct the CBSV_UHam IC .................................. 131
Figure 4.19: Restriction digest of CBSV_UHam plasmids to indicate that no significant sequence rearrangements had occurred during construction .................................. 131
Figure 4.20: PCR amplification of fragments targeting the UCBSV Ham1 replacement, from two CBSV_UHam plasmids, indicating successful construction ........................................... 131
Figure 4.21: PCR confirmation that A. tumefaciens cultures contained ICs: CBSV_Tanza, CBSV_mutHam, CBSVUHam and CBSV_HKO .................................................................. 132
Figure 4.22: RT-PCR amplification of CBSV Tanza CP fragment to confirm infection of N. benthamiana plants with the ICs: CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam at 10 dpi .......................................................... 133
Figure 4.23: TAS-ELISA quantification of CBSV titres in N. benthamiana infected with CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam ......................................................... 135
Figure 4.24: QPCR quantification of CBSV titres in N. benthamiana infected with ICs: CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam................................................................. 137
Figure 4.25: QPCR quantification of N. benthamiana ITPA gene expression during infection with CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam ..................................................... 138
Figure 4.26: RT-PCR confirmed that N. benthamiana plants were infected with CBSV_Tanza and CBSV_Ham1KO ......................................................................................... 139
Figure 4.27: CBSV Tanza genome showing the CP – 3’UTR region targeted by RT-PCR to produce amplicons from CBSV_Tanza and CBSV_HKO infected N. benthamiana plants .......... 140
Figure 4.28: Gel purified RT-PCR amplicons amplified from N. benthamiana plants infected with CBSV_Tanza and CBSV_HKO .................................................................................. 141
Figure 4.29: Qualimap coverage report of aligned reads from a representative sample (C61) mapped to the 1202 bp reference sequence ........................................................................ 143
Figure 4.30: Total number of SNVs detected per million nucleotides of aligned sequence reads from CBSV_Tanza and CBSV_HKO infections .................................................................. 144
Figure 4.31: Total number of transition SNVs detected per million nucleotides of aligned sequence reads from CBSV_Tanza and CBSV_HKO infections ........................................ 144
Figure 4.32: QPCR quantification of CBSV titres in N. benthamiana infected with CBSV_Tanza and CBSV_HKO ................................................................. 145
Figure 4.33: Mean frequency of SNVs in deep sequencing reads from CBSV_Tanza and CBSV_HKO infections of N. benthamiana. ........................................... 146
Figure 5.1: Relatively high pyrophosphohydrolase activity of purified human ITPA with non-canonical nucleotides: ITP, dITP and XTP compared to canonical nucleotides. ............... 154
Figure 5.2: Left: Protein structure of the human ITPA dimer without and with ITP bound in the binding cleft. Right: Model of the human ITPA substrate-binding pocket. .................... 155
Figure 5.3: Predicted protein structures of CBSV Tanza Ham1 and UCBSV Kikombe Ham1 based on the amino acid sequence alignment with the human ITPA protein ......................... 157
Figure 5.4: Mechanism for IPTG induced expression of recombinant proteins from POPINF plasmid in the E. coli strain BL21 (DE3) .............................................. 158
Figure 5.5: PCR amplification of the CBSV Tanza Ham1 sequence from the CBSV_Tanza IC template ......................................................................................... 159
Figure 5.6: Left: High fidelity Phusion PCR amplification of CBSV Tanza Ham1 sequence from CBSV_Tanza IC template. Right: Restriction digest of POPINF with Kpn I and Hind III to produce linear plasmid ................................................................. 160
Figure 5.7: Culture PCR to detect the presence of the CBSV Ham1 fragment in E. coli cultures to indicate successful construction of the POPINF_CHam_WT plasmid. ............... 161
Figure 5.8: AKTA traces of UV absorption from fractions eluted from the His-trap nickel chromatography column during CBSV Tanza Ham1 protein purification ...................... 162
Figure 5.9: SDS-PAGE gel electrophoresis of fractions eluted from the AKTA machine during CBSV Tanza Ham1 protein purification ..................................................... 162
Figure 5.10: SDS-PAGE gel electrophoresis of attempts to purify sufficient CBSV Tanza Ham1 protein for X-ray crystallography ...................................................... 163
Figure 5.11: SDS-PAGE analysis demonstrates that the expression and purification of CBSV Tanza Ham1 protein was unsuccessful in this attempt ........................................ 164
Figure 5.12: SDS-PAGE of purified UCBSV Kikombe Ham1 protein ........................................ 164
Figure 5.13 Schematic for reactions in the pyrophosphohydrolase enzyme assay .................. 164
Figure 5.14: Absorbance 655 nm values plotted against phosphate concentration (μM) of the phosphate standard ............................................................... 166
Figure 5.15: Incubation of 1.3 μg CBSV Tanza Ham1 protein with a range of dITP concentrations resulted in the release of phosphate demonstrating pyrophosphohydrolase activity. 167
Figure 5.16: Enzyme assay results from the heat inactivation experiment to test for loss of CBSV Tanza Ham1 pyrophosphohydrolase activity with dITP .................................. 168
Figure 5.17: Mean phosphate concentrations produced in three independent enzyme assay experiments with 1.3 μg CBSV or UCBSV Ham1 proteins incubated with 0.2 mM nucleotide substrates .................................................. 170
Figure 5.18 Phosphate concentrations produced in enzyme assay experiments with 1.3 μg CBSV or UCBSV Ham1 proteins incubated with 0.2 mM nucleotide substrates: mG triphosphate cap analog, dITP or dATP ........................................ 172
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full name</th>
</tr>
</thead>
<tbody>
<tr>
<td>°C</td>
<td>Degrees Celsius</td>
</tr>
<tr>
<td>6K1</td>
<td>Six kilodalton peptide 1</td>
</tr>
<tr>
<td>6K2</td>
<td>Six kilodalton peptide 2</td>
</tr>
<tr>
<td>ACMV</td>
<td>African cassava mosaic virus</td>
</tr>
<tr>
<td>AGO</td>
<td>Argonaute protein</td>
</tr>
<tr>
<td>AMP</td>
<td>Adenine monophosphate</td>
</tr>
<tr>
<td>AmpR</td>
<td>Ampicillin resistance gene</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>Analysis of covariance</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
</tr>
<tr>
<td>BCMV</td>
<td>Bean common mosaic virus</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>BYDV</td>
<td>Barley yellow dwarf virus</td>
</tr>
<tr>
<td>CaMV</td>
<td>Cauliflower mosaic virus</td>
</tr>
<tr>
<td>CaMV 35S</td>
<td>Cauliflower mosaic virus 35S promoter</td>
</tr>
<tr>
<td>CBSD</td>
<td>Cassava brown streak disease</td>
</tr>
<tr>
<td>CBSV</td>
<td>Cassava brown streak virus</td>
</tr>
<tr>
<td>CBSV_HKO</td>
<td>CBSV_Tanza infectious clone containing Ham1 deletion</td>
</tr>
<tr>
<td>CBSV_mutHam</td>
<td>CBSV_Tanza infectious clone containing a mutated Ham1 sequence</td>
</tr>
<tr>
<td>CBSV_UCP</td>
<td>CBSV_Tanza infectious clone containing the UCBSV Kikombe coat protein</td>
</tr>
<tr>
<td>CBSV_UHam</td>
<td>CBSV_Tanza infectious clone containing the UCBSV Kikombe Ham1</td>
</tr>
<tr>
<td>ccdB</td>
<td>E. coli toxin protein B</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CI</td>
<td>Cytoplasmic inclusion protein</td>
</tr>
<tr>
<td>CIYVV</td>
<td>Clover yellow vein virus</td>
</tr>
<tr>
<td>CMD</td>
<td>Cassava mosaic disease</td>
</tr>
<tr>
<td>CMGs</td>
<td>Cassava mosaic geminiviruses</td>
</tr>
<tr>
<td>CMV</td>
<td>Cucumber mosaic virus</td>
</tr>
<tr>
<td>CoRSV</td>
<td>Coffee ringspot virus</td>
</tr>
<tr>
<td>CP</td>
<td>Coat protein</td>
</tr>
<tr>
<td>CTAB</td>
<td>Cetyltrimethylammonium bromide</td>
</tr>
<tr>
<td>CTP</td>
<td>Cytidine triphosphate</td>
</tr>
<tr>
<td>CTV</td>
<td>Citrus tristeza virus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name/Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>CVYYV</td>
<td><em>Cucumber vein yellowing virus</em></td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxy adenosine triphosphate</td>
</tr>
<tr>
<td>DCL</td>
<td>DICER like</td>
</tr>
<tr>
<td>dCTP</td>
<td>Deoxy cytidine triphosphate</td>
</tr>
<tr>
<td>dGTP</td>
<td>Deoxy guanosine triphosphate</td>
</tr>
<tr>
<td>dHAPTP</td>
<td>Deoxy 6-hydroxyaminopurine triphosphate</td>
</tr>
<tr>
<td>dITP</td>
<td>Deoxy inosine triphosphate</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxy ribose nucleic acid</td>
</tr>
<tr>
<td>DPI</td>
<td>Days post inoculation</td>
</tr>
<tr>
<td>DRC</td>
<td>Democratic Republic of Congo</td>
</tr>
<tr>
<td>DSB</td>
<td>Double-strand DNA break</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>dTTP</td>
<td>Deoxy thymidine triphosphate</td>
</tr>
<tr>
<td>EACMV</td>
<td><em>East African cassava mosaic virus</em></td>
</tr>
<tr>
<td>eIF4E</td>
<td>Eukaryotic translation initiation factor 4E</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>EuRV</td>
<td><em>Euphorbia ringspot virus</em></td>
</tr>
<tr>
<td>F1 - 3</td>
<td>PCR fragments 1 - 3</td>
</tr>
<tr>
<td>g</td>
<td>Gram</td>
</tr>
<tr>
<td>g</td>
<td>Gravity</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GMP</td>
<td>Guanine monophosphate</td>
</tr>
<tr>
<td>GTP</td>
<td>Guanosine triphosphate</td>
</tr>
<tr>
<td>GUS</td>
<td>β-glucuronidase</td>
</tr>
<tr>
<td>Ham1</td>
<td><em>S. cerevisiae</em> gene controlling 6-N-hydroxyaminopurine sensitivity</td>
</tr>
<tr>
<td>HAP</td>
<td>6-hydroxyaminopurine</td>
</tr>
<tr>
<td>HC-Pro</td>
<td>Helper component protease</td>
</tr>
<tr>
<td>hITPA</td>
<td>Human ITPase protein</td>
</tr>
<tr>
<td>HPLC</td>
<td>High-pressure liquid chromatography</td>
</tr>
<tr>
<td>HR</td>
<td>Hypersensitive response</td>
</tr>
<tr>
<td>IC</td>
<td>Infectious clone</td>
</tr>
<tr>
<td>IgG</td>
<td>Polyclonal antibody</td>
</tr>
<tr>
<td>iLOV</td>
<td>The light, oxygen or voltage-sensing domain of the plant phototropin</td>
</tr>
<tr>
<td>IMP</td>
<td>Inosine monophosphate</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITP</td>
<td>Inosine triphosphate</td>
</tr>
<tr>
<td>ITPA</td>
<td>ITPase protein</td>
</tr>
</tbody>
</table>
Kb  Kilobase
kDa  Kilodalton
kV  Kilovolts
L  Liter
lacI  Lactose operon repressor
lacO  Lactose operon
LB  Luria broth
LSHR  Lethal systemic hypersensitive response
m  Meter
M  Molar
m7G cap  7-methyl guanosine cap
MAb  Monoclonal antibody
masl  Meters above sea level
Maxiprep  Large scale E. coli plasmid extraction
MES buffer  Morpholino ethane sulfonic acid buffer
Midiprep  Medium scale E. coli plasmid extraction
Miniprep  Small scale E. coli plasmid extraction
MSV  Maize streak virus
NCBI  National Center for Biotechnology Information
NiA  Nuclear inclusion A protease
NiB  Nuclear inclusion B RNA dependant RNA polymerase
nm  Nanometers
nt  Nucleotide
NTPase  Nucleoside-triphosphatase
OD  Optical density
ORF  Open reading frame
ORI  Origin of replication
P1  Protein 1 protease
P3  Protein 3
PABP  Poly A binding protein
PBS  Phosphate buffered saline
PCR  Polymerase chain reaction
PepMV  Pepino mosaic virus
PIPO  Pretty interesting Potyviridae ORF
PLDMV  Papaya leaf distortion mosaic virus
PLRV  Potato leafroll virus
POPINF_CHam_Mut  POPINF plasmid containing the mutated CBSV Ham1 sequence
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPINF_CHam_WT</td>
<td>POPINF plasmid containing the wild-type CBSV Ham1 sequence</td>
</tr>
<tr>
<td>PRRSV</td>
<td><em>Porcine reproductive and respiratory syndrome virus</em></td>
</tr>
<tr>
<td>PSbMV</td>
<td><em>Pea seed-borne mosaic virus</em></td>
</tr>
<tr>
<td>psi</td>
<td>Pounds per square inch</td>
</tr>
<tr>
<td>PVY</td>
<td><em>Potato virus Y</em></td>
</tr>
<tr>
<td>qPCR</td>
<td>Quantitative reverse-transcriptase PCR</td>
</tr>
<tr>
<td>R protein</td>
<td>Resistance protein</td>
</tr>
<tr>
<td>RAM-AP</td>
<td>Rabbit anti-mouse alkaline phosphatase</td>
</tr>
<tr>
<td>Rdgb</td>
<td>Rec-dependent growth B protein</td>
</tr>
<tr>
<td>Rdrp</td>
<td>RNA dependant RNA polymerase</td>
</tr>
<tr>
<td>RIN</td>
<td>RNA integrity number</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA induced silencing complex</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribose nucleic acid</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SAA</td>
<td>Serine-Alanine-Alanine mutated pyrophosphohydrolase motif</td>
</tr>
<tr>
<td>SAR</td>
<td>Systemic acquired resistance</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>Sodium dodecyl sulfate–polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SDW</td>
<td>Sterile deionised water</td>
</tr>
<tr>
<td>SHR</td>
<td>Serine-Histidine-Arginine pyrophosphohydrolase motif</td>
</tr>
<tr>
<td>SMV</td>
<td><em>Soybean mosaic virus</em></td>
</tr>
<tr>
<td>SNV</td>
<td>Single nucleotide variant</td>
</tr>
<tr>
<td>SOC</td>
<td>Super optimal broth with catabolite repression medium</td>
</tr>
<tr>
<td>SOS</td>
<td>Global response to DNA damage</td>
</tr>
<tr>
<td>SpDPV</td>
<td><em>Spinach deltapartitivirus</em></td>
</tr>
<tr>
<td>SPMMMV</td>
<td><em>Sweet potato mild mottle virus</em></td>
</tr>
<tr>
<td>SqVYV</td>
<td><em>Squash vein yellowing virus</em></td>
</tr>
<tr>
<td>SSA</td>
<td>Sub-Saharan Africa</td>
</tr>
<tr>
<td>SSB</td>
<td>Single-strand DNA break</td>
</tr>
<tr>
<td>STE buffer</td>
<td>Sodium chloride Tris-EDTA buffer</td>
</tr>
<tr>
<td>T-DNA</td>
<td>Transfer DNA</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate EDTA buffer</td>
</tr>
<tr>
<td>TAIR</td>
<td>The Arabidopsis Information Resource</td>
</tr>
<tr>
<td>Taq</td>
<td><em>Thermus aquaticus</em> DNA polymerase</td>
</tr>
<tr>
<td>TAS-ELISA</td>
<td>Triple Antibody Sandwich Enzyme Enzyme-Linked Immunosorbent Assay</td>
</tr>
<tr>
<td>TE buffer</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>TMV</td>
<td><em>Tobacco mosaic virus</em></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>tNOS</td>
<td>Nopaline synthase terminator</td>
</tr>
<tr>
<td>ToMMV</td>
<td>Tomato mild mottle virus</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>TuMV</td>
<td>Turnip mosaic virus</td>
</tr>
<tr>
<td>TVMV</td>
<td>Tobacco vein mottling virus</td>
</tr>
<tr>
<td>U/CBSVs</td>
<td>Collective term for CBSV and UCBSV isolates</td>
</tr>
<tr>
<td>UCBSV</td>
<td>Ugandan cassava brown streak virus</td>
</tr>
<tr>
<td>URA3</td>
<td>Orotidine 5-phosphate decarboxylase</td>
</tr>
<tr>
<td>UTP</td>
<td>Uridine triphosphate</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-violet</td>
</tr>
<tr>
<td>v/v</td>
<td>Volume/volume %</td>
</tr>
<tr>
<td>VPg</td>
<td>Viral genome linked protein</td>
</tr>
<tr>
<td>VRC</td>
<td>Viral replication complexes</td>
</tr>
<tr>
<td>vsiRNA</td>
<td>Virus-derived small interfering RNAs</td>
</tr>
<tr>
<td>w/v</td>
<td>Weight of solute in the total volume of solution %</td>
</tr>
<tr>
<td>WMV</td>
<td>Watermelon mosaic virus</td>
</tr>
<tr>
<td>X-Gal</td>
<td>5-Bromo-4-chloro-3-indolyl beta-D-galactopyranoside</td>
</tr>
<tr>
<td>XMP</td>
<td>Xanthine monophosphate</td>
</tr>
<tr>
<td>XTP</td>
<td>Xanthine triphosphate</td>
</tr>
<tr>
<td>YENB</td>
<td>Yeast extract nutrient broth</td>
</tr>
<tr>
<td>YEPD</td>
<td>Yeast extract peptone dextrose</td>
</tr>
<tr>
<td>YSDM</td>
<td>Yeast synthetic drop out media</td>
</tr>
<tr>
<td>ZYMV</td>
<td>Zucchini yellow mosaic virus</td>
</tr>
</tbody>
</table>
Chapter 1: Introduction

1.1 Global food security

Global food security is the condition in which all people, at all times, have physical, social and economic access to sufficient safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life (FAO, 2001). Unfortunately, global food security remains one of the biggest challenges facing humanity. In contrast to many Western countries, where obesity is now a major problem, access to food in many developing countries is failing to keep up with population growth (IFPRI, 2016). In 2016, approximately 11% of the global population was undernourished (FAO, 2017) and in 2017 famine occurred in south Sudan and food insecurity worsened in other parts of sub-Saharan Africa (SSA) and south-eastern and western Asia (FAO, 2017). In SSA, an estimated 153 million people suffered from severe food insecurity in 2014-15, accounting for 26% of the population above 15 years of age (FAO, 2017). Furthermore, food insecurity is set to worsen; by 2050 the global population is predicted to rise to 9.1 billion, there will be less land and resources available for food production and climate change will have unpredictable effects on farming (Godfray et al., 2010). In SSA, climate change scenario models predict that by 2050, temperatures will be consistently higher and there will fluctuations in rainfall that will thereby reduce crop yields and per capita calorie availability (Ringler et al., 2010). To meet the challenge of achieving global food security in a changing climate a myriad of innovations is required.

A key strategy is to reduce the gap between yields that are achieved under optimal conditions and yields that are achieved in the field (Godfray et al., 2010). In the latter half of the 20th century, significant increases in cereal productivity were achieved through the development and adoption of improved, high-yielding, early maturing varieties in Latin America and Asia (Evenson and Gollin, 2003). For instance, in China rice yields increased by 3.3 fold from 1961-2015 (FAOSTAT, 2018). However, during the same period, similar increases in crop productivity did not occur in SSA (Evenson and Gollin, 2003). This is partly due to a historic lack of scientific research to develop improved crop varieties specifically suited to SSA countries, as well as complex social, political, economic and climate factors (Evenson and Gollin, 2003; The World Bank, 2008). In recent years, improved crop varieties have been developed for the many agricultural environments in SSA, reaching approximate adoption rates of 70% for wheat, 45% for maize, 26% for rice, 19% for cassava, and 15% for sorghum by 2005 (Evenson and Pingali, 2010). According to farmer surveys, the adoption of improved varieties has doubled yields of
maize in Nigeria, rice in Ghana and cowpea in Uganda (AGRA, 2014). Cassava has received particular attention as a crop that could provide climate change adaptation opportunities in SSA that would improve food and economic security (Jarvis et al., 2012). Unfortunately, cassava production in SSA is mainly constrained by two viral diseases: cassava mosaic disease (CMD) and cassava brown streak disease (CBSD), which together are estimated to cause annual losses worth US$1 billion (IITA, 2014a) and adversely affect food security across the entire region (Patil et al., 2015). In contrast with CMD, relatively little is known about the fundamental biology of the viruses which cause CBSD (Tomlinson et al., 2017) and there are no cassava varieties with high levels of CBSD resistance currently available to farmers (Abaca et al., 2013). Research groups are now developing CBSD control strategies, which primarily involves breeding cassava varieties with high levels of CBSD and CMD resistance (Tomlinson et al., 2017). CBSD viral infectious clones (IC) are urgently needed to screen for CBSD resistance, unfortunately their construction has previously been prevented by sequence instability during plasmid propagation in Escherichia coli. Two CBSD viral ICs were recently constructed at the University of Bristol in Duff-Farrier and Mbwanzibwa et al., (unpublished) and Nanyiti (2016). This study describes investigations into their utility in terms of sequence stability during propagation and tractability to gain insights in viral gene functions during infection.

1.2 Plant diseases

A vast number of plant pathogens cause crop diseases; the major groups include fungi, bacteria, viruses, oomycetes, nematodes and parasitic plants (Strange and Scott, 2005). It is estimated that plant diseases cause losses of 10% of global harvests each year (Strange and Scott, 2005). Additional losses also occur due to postharvest deterioration, which can be a particular problem in developing countries that lack sufficient infrastructure to effectively store and transport produce (Agrios, 2005). Within social and political contexts, plant pathogens can cause famine and have had far reaching effects on the course of human history, as seen with the Irish potato famine in the 19th century (Fry, 2008) and Bengal famine in the 20th century (Padmanabhan, 1973). Insect pests also cause high yield losses, primarily through the destruction of plant material, consumption of photo-assimilates and transmission of pathogens (Oerke, 2006). Significant pathogens that are currently affecting food security include: wheat stem rust fungus (Puccinia graminis), which causes losses of up to 40% in countries across Africa and the Middle East (Pennisi, 2010). Worse still pests and pathogens are opportunists and so many will be able to rapidly to adapt to niches that are opened up through global trade and climate change (Chakraborty and Newton, 2011). Current measures to control plant pathogens
include the development and deployment of resistant crop varieties to farmers, which requires an understanding of both pathogen virulence and plant host resistance mechanisms (Strange and Scott, 2005).

1.3 Plant viruses
Viruses consist of nucleic acid genomes, which encode three or more proteins and are encapsidated in a protein coat. As obligate parasites, viruses can only reproduce inside host cells, where they manipulate host cellular machinery for viral reproduction. All viruses must direct the formation of at least three types of proteins: replication proteins that are essential for viral nucleic acid reproduction, structural proteins that form the coat protein shell and movement proteins that enable the movement of viral particles systemically and between cells. Viruses commonly rely on other organisms or the environment for their dissemination and transmission to new hosts (Gergerich and Dolja, 2006). There are approximately 1,000 known plant viruses (King et al., 2012), which can infect a wide range of crop and wild plant hosts and cause a number of important plant diseases that severely affect global food security (Scholthof et al., 2011). Next generation sequencing is now revealing that diverse plant viruses are found in wild plants from the tropics to Antarctica (Roossinck, 2015). Occasionally, plant viruses ‘jump’ from wild plant host into crop species. Indeed, plant viruses are estimated to cause half of reported emerging infectious diseases in plants (Anderson et al., 2004). Worryingly, more viral diseases are predicted to emerge in the future due to the effects of climate change, global trade of plant material and changes to vector dispersal (Jones, 2009).

1.4 Cassava
Cassava (Manihot esculenta Crantz, family Euphorbiaceae) also known as manioc, tapioca and yuca, is a widely grown tropical crop that can be cultivated on marginal soils and produces high yields even in unfavorable growing conditions (Bredeson et al., 2016). Cassava produces carbohydrate rich storage roots that serve as a staple food source for approximately 800 million people worldwide (FAO, 2013). In Africa, cassava is the second most important food staple in terms of per capita calories consumed (Nweke, 2004) that often serves as a fallback for subsistence farmers, as unlike other staple crops, cassava can withstand unpredictable rainfall, grows with minimal inputs, on marginal land and can be harvested throughout the year (Hillocks and Thresh, 2002). Critically, cassava can withstand high temperature and rainfall fluctuations and so offers unique climate change adaptation opportunities in Africa (Jarvis et
Cassava also contributes towards economic security as it can be sold in local markets and used in industry, helping to raise farmer income and alleviate rural poverty (Dixon et al., 2003).

The *Manihot* genus contains 98 species, which originate from South and Central America (Halsey et al., 2008). It is estimated that cassava was domesticated from wild populations of *M. esculenta* ssp. *flabellifolia* in the Amazon basin between 2,000 – 4,000 BC (Fauquet et al., 1990; Olsen and Schaal, 1999). In the 16th century, Portuguese navigators transported cassava to the west coast of Africa (Jones, 1959), and later to East Africa through Madagascar and Zanzibar (Jennings, 1972). Cassava was not widely grown across SSA until the 19th century, when its value as a food security crop was realized (Hillocks and Thresh, 2002). Cassava cultivation then spread across SSA in the 20th century (Hillocks and Thresh, 2002) and today Africa now produces over half of global cassava production (57%) (Bennett, 2015), with many SSA countries producing over 3 million tonnes a year (FAOSTAT, 2018) (Fig. 1.1). In most SSA countries, cassava is predominately grown by smallholder farmers. However, there are also some industrial plantations in Liberia, Nigeria and Togo. Nigeria is the world’s largest producer of cassava, while Thailand is the largest exporter of dried cassava (FAOSTAT, 2018).

![Figure 1.1: Estimated cassava production across sub-Saharan Africa in 2016 (FAOSTAT, 2018).](image-url)
Cultivated cassava can grow up to 7 m in height and produces tuberous roots in clusters of 4 – 8 at the base of the stem, as shown in Fig. 1.2 (Moore, 2005). In Africa, cassava is grown at altitudes ranging from 150 – 2000 meters above sea level (masl), at tropical latitudes between 30° north and 30° south of the equator, where the temperatures range from 18 - 35°C (Phillips, 1974). The ideal growth conditions for cassava are temperatures of 25 - 27°C, with abundant rainfall. However cassava shows a remarkable ability to tolerate annual rainfall as low as 500 mm or high as 5,000 mm (Moore, 2005). The crop is propagated almost entirely from vegetative stem cuttings, as pollination rates are often low and thus few seeds are produced; seed germination is also less than 50% (Moore, 2005). Stems are cut into 9 – 30 cm lengths and planted vertically at 8 – 15 cm underground (Moore, 2005). Roots can be harvested throughout the year; improved varieties typically reach maximum yields between 12 - 15 months after planting (El-Sharkawy, 2004). Unfortunately, cassava is prone to viral infections because propagated stems can carry viral pathogens that accumulate during multiple vegetative planting seasons (Legg et al., 2015).

Cassava is genetically diverse, with a huge number of varieties that have been bred and selected by farmers according to local preferences (Gillman and Erenler, 2009; Nakabonge et al., 2017). Bitter cassava varieties contain relatively high levels of cyanogenic glucosides, which are converted to cyanide upon digestion and so processing is required to ensure they are safe for human consumption (Cardoso et al., 2005). Genome sequencing has revealed that during the domestication of cassava, genes involved with metabolism, photosynthesis and starch accumulation have been positively selected whereas genes involved in cell wall biosynthesis and secondary metabolism, have been negatively selected (Wang et al., 2014).

Figure 1.2: A) Illustrative picture of cassava showing typical growth and tuberous roots ©Wilhelm Valder; B) Cross-section of cassava roots free from necrosis (Wikimedia: Amada44) and C) Healthy appearing cassava leaves (Wikimedia: CIAT, mrjohncummings).
1.4.1 Uses of cassava for food and industrial products

Storage roots are eaten fresh or processed into flour, which may be consumed by the farmer, sold in local markets or used in the production of several industrial products (Hillocks and Thresh, 2002). Cassava flour is a key ingredient in many African dishes, including enya asa bread in Uganda. In some areas leaves are also consumed, which are high in vitamins A and B. Cassava can be used as a raw material in the production of livestock feed, starch, alcohol, pharmaceuticals, bio-fuels and bio-polymers (Balat and Balat, 2009; Hillocks and Thresh, 2002). Cassava is a particularly attractive crop for bioethanol production as the low energy input requirements mean that an energy profit of 95% can be achieved (Jansson et al., 2009). Efforts to commercialize cassava production are underway in many SSA countries, which may help to alleviate rural poverty (Dixon et al., 2003; Kambewa, 2010).

1.5 Constraints on cassava production in Africa

The average fresh yield for cassava production in Africa is: 9.9 tonnes/ha, which is significantly lower than optimal yields of 15 – 40 tonnes/ha achieved under experimental conditions (Fermont et al., 2009). There are many reasons behind this, including restricted access to labor, poor soil quality, premature harvesting, limited inputs such as fertilizer, pesticides, herbicides and susceptibility to pests and diseases (Fermont et al., 2009). Cassava marketability is also reduced by rapid post-harvest physiological deterioration, which occurs from 24 – 72 hours after harvest (Zainuddin et al., 2017).

1.5.1 Bacterial diseases

Cassava production in Africa is affected by the bacterial pathogen: *Xanthomonas axonopodis* pv. *Manihotis*, which causes cassava bacterial blight associated with wilting, leaf blight, dieback and plant death and can cause yield losses of up to 75% depending on environmental conditions (Wydra and Verdier, 2002).

1.5.2 Fungal diseases

The fungal disease: cassava anthracnose caused by *Colletotrichum gloeosporioides* also affects cassava production through severe stem damage, wilting and dieback. High incidences have been reported in humid eco-zones of the Democratic Republic of Congo (DRC), Benin and Ghana (McCallum et al., 2017; Owolade et al., 2006; Wydra and Verdier, 2002).
1.5.3 Pests
Around 200 insect pests have been reported on cassava, most of which are found in Latin America, where they have co-evolved with the crop (Howeler et al., 2013). Insect pests cause mechanical damage through feeding on photo-assimilates and serve as vectors of viral pathogens (Howeler et al., 2013). The main pests on cassava in Africa are silverleaf whitefly (Bemisia tabaci), cassava mealybug (Phenacoccus manihoti), African root and tuber scale (Stictococcus vayssierei), cassava green mite (Mononychellus tanajoa) and nematodes (Melloidogyne spp.) (Nassar, 2007).

1.5.3.1 Bemisia tabaci
B. tabaci (Hemiptera, Aleyrodidae) is currently having a major impact on cassava production in Africa. B. tabaci directly damages cassava through feeding on photo-assimilates and the excretion of honeydew, which encourages sooty mold on lower leaves (Howeler et al., 2013). B. tabaci can feed on over 1,000 plant species and vectors over 300 viruses (Wosula et al., 2017), including CMD and CBSD causal viruses. B. tabaci forms a species complex containing at least 34 distinct species groups identified using mtCOI sequences (Polston et al., 2014). In SSA, specific B. tabaci species groups are designated SA1 – 5 and appear to have adapted to feed on cassava and undergone a period of rapid geographical expansion in the last twenty years (Omongo et al., 2012). Super-abundant numbers of whiteflies are now found on cassava across East Africa, which has likely been a driving factor in CMD and CBSD epidemiology (Alicai et al., 2007; Jeremiah et al., 2015; Legg et al., 2011). Ecological niche modeling has predicted, that with climate change, the potential distribution of CMD/CBSD carrying B. tabaci will spread over West, Central and southwestern coastal Africa, where cassava production is high and CBSD is currently absent (Herrera Campo et al., 2011). This means that monitoring and controlling B. tabaci is currently a major priority.

1.5.4 Viral diseases
At least 20 different viruses are known to infect cassava, nine of which have been reported in Africa, as shown in Table 1.1 (Calvert and Thresh, 2002). The viral diseases that currently cause the highest cassava losses in Africa are CMD and CBSD, which combined are estimated to cause annual losses worth US$1 billion (IITA, 2014a) and have a severe impact on security across SSA (Patil et al., 2015).
Table 1.1 Viruses reported to infect cassava in Africa (Calvert and Thresh, 2002).

<table>
<thead>
<tr>
<th>Species</th>
<th>Disease</th>
<th>Family, Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa cassava mosaic virus</td>
<td>Cassava mosaic disease</td>
<td>Geminiviridae, Begomovirus</td>
</tr>
<tr>
<td>East African cassava mosaic viruses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South African cassava mosaic virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East African cassava mosaic Cameroon virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East African cassava mosaic Malawi virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>East African cassava mosaic Zanzibar virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cassava brown streak virus</td>
<td>Cassava brown streak disease</td>
<td>Potyviridae, Ipomovirus</td>
</tr>
<tr>
<td>Ugandan cassava brown streak virus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cassava Ivorian bacilliform virus</td>
<td>Symptomless</td>
<td>Bromoviridae, Alfamovirus</td>
</tr>
<tr>
<td>Cassava Kumi viruses A, B and Q</td>
<td>No published information</td>
<td>No published information</td>
</tr>
<tr>
<td>Cassava common mosaic virus</td>
<td>Mosaics</td>
<td>Alphaflexiviridae, Potexvirida</td>
</tr>
</tbody>
</table>

1.6 Cassava mosaic disease

CMD is caused by at least nine distinct Cassava mosaic geminiviruses (CMG). CMGs have geminate particles (30 X 20 nm) that contain circular, single-stranded DNA, bipartite genome components: DNA A and DNA B (Alabi et al., 2011). CMD was first reported in the Usambara mountains of north-east Tanzania in 1894 (Warburg, 1894) and was present at relatively low levels across SSA until the late 1980s, when a the recombinant: East African cassava mosaic virus – Ugandan variant (EACMV-UG) emerged, causing a severe CBSD pandemic that swept across East Africa (Fig. 1.3). This resulted in wide-spread famine and forced many smallholder farmers to temporarily abandon the crop until cassava varieties with sufficient CMD resistance could be obtained (Legg et al., 2011). Since then, the severe CMD pandemic has continued to spread southwards through eastern DRC (Ntawuruhunga et al., 2010) and westward through central Cameroon (Szyniszewska et al., 2017). The pandemic now threatens cassava production in West Africa. CMD symptoms include stunted growth, reduced number and size of storage roots and distortion of leaves with chlorotic mosaic patterns (McCallum et al., 2017). CMD has been reported to cause tuber root losses of 82 - 90% in highly susceptible varieties (Legg et al., 2006; Owor et al., 2004). CMD is caused by at least 11 CMGs, that are in the Begomovirus genus of the Geminiviridae family (De Bruyn et al., 2016). Outside of Africa, CMGs are also found in India, Sri Lanka and Cambodia (Saunders et al., 2002; Wang et al., 2016). CMGs evolve rapidly in the field through frequent recombination during mixed infections and error prone replication, which has generated diverse recombinant viral isolates (Bernardo et al., 2013; McCallum et al., 2017).

CMGs are spread through planting infected material and persistent transmission by B. tabaci, which carry CMGs over long distances (Legg et al., 2011). Varieties with high levels of CMD resistance have been developed through traditional breeding and distributed to farmers, which
has helped to control the disease in several eastern and southern African countries (Fondong, 2017; McCallum et al., 2017). Unfortunately, these cultivars showed varying levels of CBSD susceptibility (Legg et al., 2006) and so deployment of these cultivars may have contributed to the increased distribution of CBSD in the field.

Figure 1.3: Distribution of cassava mosaic disease (CMD) across sub-Saharan Africa in the 1920s (red), outbreaks of severe CMD occurred in Uganda in the late 1980s (yellow) and a pandemic then spread in the 1990s (green) (Legg et al., 2011).

1.7 Cassava brown streak disease

CBSD is caused by at least two potyviral species: *Cassava brown streak virus* (CBSV) and *Uganda cassava brown streak virus* (UCBSV); collectively referred to as U/CBSVs (Monger et al., 2001a; Winter et al., 2010), which belong to the *Ipomovirus* genus of the *Potyviridae* family (Monger et al., 2001b) and are semi-persistently transmitted by *B. tabaci* (Maruthi et al., 2005).
1.7.1 CBSD epidemiology

Similarly to CMD, CBSD was first reported in north-east Tanzania in the 1930s (Storey, 1936) and until the early 2000s, the disease was reported to be largely restricted to low-altitude areas (<1000 masl) along coastal East Africa and in lakeshore districts of Malawi (Legg et al., 2011). In 2004, CBSD re-emerged inland, at altitudes above 1,000 masl in Uganda (Alicai et al., 2007). Subsequently CBSD has reached high incidences in Uganda, Tanzania and Kenya. The CBSD pandemic then spread across the entire the entire Great Lakes region (Legg et al., 2011) and has now reached the Comoros and Mayotte islands in the Indian Ocean (Azali et al., 2017; Roux-Cuvelier et al., 2014) and Burundi, Rwanda, DRC and Zambia in Central Africa (Fig. 1.4) (Bigirimana et al., 2011; FAO, 2013; Mulenga et al., 2018; Mulimbi et al., 2012). This ongoing westward spread severely threatens the major cassava growing areas of Central and West Africa.

![Figure 1.4: Distribution of CBSD in coastal East Africa in the 1930s (red). CBSD outbreaks occurred inland in Uganda in 2004 (yellow) and subsequently CBSD is reported to have spread to southern shores of Lake Victoria and western Democratic Republic of Congo. Reports in green circles have yet to be confirmed with laboratory diagnostic tests (Legg et al., 2011).](image)

1.7.2 CBSD transmission and dispersal

Studies have shown that CBSD is dispersed locally and over long distances through the transportation of infected planting material, whereas whiteflies disperse and amplify CBSD locally (McQuaid et al., 2017). As mentioned in section 1.5.3.1, the presence of super-abundant whiteflies appears to be a driving factor in CBSD epidemiology (Alicai et al., 2007; Jeremiah et al., 2015). Transmission studies with U/CBSVs in Tanzania have shown that they are transmitted semi-persistently, with whiteflies acquiring viruses within 5 – 10 mins, retaining them for up to 48 hours and transmitting them over relatively short distances of less than 17 m in a cropping season (Maruthi et al., 2017). Outside of cassava, U/CBSVs have been found in the wild tree species *M. glaziovii* in Tanzania; the potential impact of this and other reservoir hosts on CBSD epidemiology is unknown (Mbanzibwa et al., 2011a).
1.7.3 Geographical distribution of U/CBSVs

Winter et al., (2010) originally found that CBSD infected cassava samples from Mozambique and Tanzania contained CBSV, whereas samples from Kenya, Uganda, Malawi and northwestern Tanzania contained UCBSV. However it has recently been shown that CBSV and UCBSV are not limited to separate agro-ecological zones and both species are now found in Tanzania (Ndunguru et al., 2015), Uganda (Ogwok et al., 2016) and Rwanda (Munganyinka et al., 2018). Indeed, mixed U/CBSV infections are relatively common, making up 34 – 50% of tested infections in Kenya, Tanzania and Uganda (Kathurima et al., 2016; Mbanzibwa et al., 2011a; Ogwok et al., 2014). The potential for interactions between CBSV and UCBSV during mixed infections is largely unknown. Bud-grafting experiments have indicated that the prior presence of UCBSV increases CBSV infectivity in susceptible cassava varieties (Wagabwa et al., 2013).

1.7.4 CBSD symptoms

CBSD causes foliar feathery chlorosis on lower mature leaves, storage root necrosis and occasionally brown streaks or lesions on stems. CBSD is highly variable in terms of symptom severity, onset of symptom expression and parts of the plant affected, depending on viral strain, cassava cultivar, environmental conditions and the age of the plant upon infection. This variability can make diagnosis difficult for farmers (Nichols, 1950), who can be unaware that their crop is affected until they harvest storage roots (Legg et al., 2015). CBSD causes two main types of foliar symptoms: CBSV tends to cause feathery chlorosis along secondary veins, which eventually coalesce to form blotches and UCBSV tends to cause chlorotic mottling with no veinal association (Mohammed et al., 2012; Nichols, 1950; Winter et al., 2010). Brown streaks or lesions and dieback can also form on severely CBSD-infected cassava stems (Fig. 1.5). This symptom can be difficult to observe and is more obvious in highly susceptible varieties (Calvert and Thresh, 2002; Hillocks and Jennings, 2003). In roots CBSD can cause necrosis on the surface of the root, high levels of necrosis in the inner starchy tissue, radial constrictions and an overall reduction in the number and weight of roots (Fig. 1.5) (Hillocks et al., 2001). In the most sensitive varieties can reduce root weight by up to 70% with necrosis developing as early as 6 months post-planting (Hillocks et al., 2001). Root necrosis makes CBSD-infected roots largely inedible and unfit for human consumption. Ephraim et al., (2015) demonstrated that CBSD-infected roots of susceptible varieties contain 30% less amylose content and higher cyanide levels in the storage roots of susceptible varieties, compared with healthy roots.
1.7.5 Responses of different cassava varieties to U/CBSVs

Different cassava varieties respond very differently to infection by U/CBSVs, (as reviewed in Tomlinson et al., 2017); they produce a range of symptoms and are associated with varying viral loads at different time points of infection (Kaweesi et al., 2014). Sensitive varieties show severe shoot and root symptoms, whereas varieties with higher tolerance tend to express foliar symptoms but lack or exhibit mild root necrosis (Hillocks and Jennings, 2003). Varieties also vary in their responses to different U/CBSV isolates. For instance, the NASE3 variety is reported to resist UCBSV infection but remains susceptible to CBSV (Ogwok et al., 2016). CBSD symptom severity is not always correlated with viral load, as the NASE1 variety supports a high viral load but produces no foliar or root necrosis symptoms, whilst the NASE14 variety supports a low viral load but expresses severe root necrosis (Kaweesi et al., 2014)

Figure 1.5: Cassava brown streak disease (CBSD) storage root necrosis (A), radial root constrictions (B), foliar chlorosis (C) and brown streaks or lesions on stems (D) (Tomlinson et al., 2017).
1.7.6 U/CBSV infections of herbaceous hosts

In addition to the different foliar symptoms observed during CBSV and UCBSV infections of cassava, the two viral species are also associated with differential symptom development during infections of various herbaceous indicator plants. Mohammed et al., (2012) demonstrated that infections of the indicator host *Nicotiana clevelandii* with CBSV were associated with severe necrosis and stunting, whereas infections with UCBSV were less severe and lacked necrosis (Fig. 1.6). The viral sequences responsible for the differences in CBSV and UCBSV symptom development are unknown.

1.7.7 Quantification of U/CBSVs during infections

To detect the presence of U/CBSVs in cassava, reverse-transcription PCR (RT-PCR) is commonly performed with primers that amplify different size amplicons to indicate the presence of UCBSV or CBSV (Mbanzibwa et al., 2011a). The most widely used technique to quantify the abundance of U/CBSV viral transcripts in infected cassava tissue is quantitative RT-PCR (qPCR) (Ogwok et al., 2014; Shirima et al., 2017). QPCR has shown that CBSV tends to accumulate to higher levels than UCBSV during infections of cassava grown in the field and screen house (Shirima et al., 2017). It has also shown that CBSV tends to accumulate to higher levels in the roots than in aerial parts (stem and leaves), whereas UCBSV titers are higher in aerial parts in the roots (Ogwok et al., 2014). Similarly, CBSV also accumulates to higher levels than UCBSV during infections of herbaceous indicator plants (Mohammed et al., 2012) The viral sequences responsible for the differences in CBSV and UCBSV accumulation are currently unknown.
1.7.8 Control of CBSD

Since the re-emergence of CBSD, control efforts have focused on breeding cassava varieties for CBSD resistance (Kawuki et al., 2016). However, unfortunately there are no cassava varieties with a high level of CBSD resistance currently available to farmers (Abaca et al., 2013). Breeders have screened and hybridized diverse Tanzanian and Ugandan germplasm to develop CBSD tolerant varieties that exhibit relatively low levels of root necrosis of 12% compared with >80% in sensitive cultivars (Kawuki et al., 2016). Although tolerant cultivars typically develop reduced symptoms, they are susceptible to infection and therefore their deployment does not remove viral inoculum from the field.

Breeding cassava is notoriously difficult due to high heterozygosity and a challenging cross-pollination process (Ceballos et al., 2012). Breeding is further complicated by cultivars showing variation in CBSD resistance across different environments, which necessitates the testing of cultivars in different agro-ecological zones to ensure their resistance is stable (Tumuhimbise et al., 2014). Next generation sequencing is now being used to link diverse cassava genome sequences with desirable phenotypic traits, including CBSD and CMD resistance. In addition to traditional breeding, cassava has also been genetically engineered for CBSD resistance through the expression of UCBSV and CBSV coat protein (CP) sequences, which trigger post-transcriptional gene silencing of the corresponding viral sequences during infection (Ogwok et al., 2012; Yadav et al., 2011). Cassava cultivars expressing these viral sequences demonstrate resistance to both CBSV and UCBSV under field conditions with high disease pressure and through multiple vegetative propagation cycles (Odipio et al., 2014). However these cassava cultivars are not yet commercially available to farmers in SSA due to biosafety regulation issues (Taylor et al., 2016). In the absence of cassava varieties with high levels of CBSD resistance, there is a drive to provide farmers with virus-free, CMD resistant, CBSD tolerant planting material that would reduce field inoculum (Legg et al., 2017; McQuaid et al., 2015). Certified clean cassava seed systems have been piloted in Uganda and Tanzania (Legg et al., 2017; McQuaid et al., 2015). It is hoped that similar systems will now be established across SSA (Tumwegamire et al., 2018), as modelling has indicated that the distribution of virus-free planting material to several different farmers in a widespread area, with restricted trade can reduce CBSD dispersal and increase cassava yields (Legg et al., 2017). Once virus-free material has been distributed, it is vital that farmers are trained to effectively identify CBSD symptoms to enable sufficient rouging (Legg et al., 2017; McQuaid et al., 2015). The transportation of CBSD infected planting material to areas previously unaffected by CBSD, represents a major threat to cassava production. Surveillance projects across SSA are now tracking area-wide
changes in cassava viral diseases over time in an attempt to identify where control strategies should be targeted (IITA, 2016). I learnt about the CBSD control strategies being deployed during my PIPS internship at the National Crops Resources Research Institute in Uganda, as described in Appendix 1.1 – 4.

1.8 Properties of plant viral genomes

A wide range of viruses can infect plants with diverse genome compositions. Their genomes can be: single-stranded or double-stranded, composed of either DNA or RNA nucleic acid and be linear or circular. For example, African cassava mosaic virus (ACMV) is a single-stranded DNA virus from the Geminiviridae family (Patil et al., 2009); Cauliflower mosaic virus (CaMV) is a double-stranded DNA virus in the Caulimoviridae family (Franck et al., 1980) and Spinach deltapartitivirus (SpDPV) is a double-stranded RNA virus in the Partitiviridae family (Park and Hahn, 2017). Single-stranded RNA viruses, genomes are either positive-sense, whereby the genome is translated directly or negative-sense, whereby the genome is first transcribed by an RNA dependent RNA polymerase (Rdrp) protein into positive-sense genome before it is translated. For example Tobacco mosaic virus (TMV) is a positive-sense single-stranded RNA virus in the Virgaviridae family (Dawson et al., 1986) and Coffee ringspot virus (CoRSV) is a negative-sense, single-stranded RNA virus in the Rhabdoviridae family (Ramalho et al., 2014).

1.9 Viral quasi-species and the critical error threshold

Both plant and animal RNA viruses are characterised by having relatively high mutation rates, short generation times and large population sizes (Andino and Domingo, 2015). Viral RNA replication has a relatively low fidelity because compared with DNA polymerases, viral Rdrp have lower fidelity, lack proof-reading activity and there is no subsequent enzymatic base-excision repair of mis-paired RNA molecules (Duffy et al., 2008). This means that whereas DNA viruses are estimated to have mutation rates ranging from $10^{-6}$ – $10^{-8}$ substitutions per nucleotide per cell infection, RNA viruses have mutation rates ranging from $10^{-6}$–$10^{-4}$ (Sanjuán et al., 2010). This generates populations of closely related viral variants referred to quasi-species or viral swarms (Andino and Domingo, 2015). Natural selection then acts to positively select genome variants with enhanced fitness in a specific host environment and/or vector and negatively selects variants with reduced fitness (Andino and Domingo, 2015). Quasi-species diversity enables RNA viruses to rapidly adapt to changes in host environments and overcome host immune systems (Duffy et al., 2008; Holmes, 2003). The concept is also termed ‘survival
of the flattest’ whereby diverse viral populations are maintained that occupy a relatively large genetic landscape to enable rapid escape from host immune responses (Sanjuán et al., 2007; Wilke et al., 2001). Although there is considerable diversity in RNA viral populations, error rates are constrained by the need to conserve viral sequences required for multiple essential functions (Holmes, 2003). It has been reported that RNA viral mutation rates typically exist close to a critical error threshold, above which error catastrophe occurs whereby lethal numbers of deleterious mutations accumulate during each round of viral replication and there is a dramatic loss of viral viability (Holmes, 2003). RNA viruses must therefore balance maximizing sequence diversity to enhance adaptability whilst conserving sequence to maintain multiple key functions.

1.10 Potyviridae family

Most of the viruses which infect plants belong to the Potyviridae family; almost every crop and many wild plant species can be infected by at least one or multiple Potyviridae species (Berger, 2001). Potyviridae form flexuous-filamentous particles, ranging in 11 – 14 nm in diameter and 275 – 950 nm in length (Valli et al., 2015). Typically 2,000 single CP subunits encapsidate Potyviridae genomes (Shukla and Ward, 1989) (Fig. 1.7). Potyviridae infections are characterised by the presence of pin-wheel inclusion bodies that aggregate in the cytoplasm of infected cells (Shukla and Ward, 1989).

Figure 1.7: Left: Schematic diagram of a potyvirus particle consisting of helically arranged coat protein subunits with the N-terminal (large rectangle) and C-terminal (small rectangle) exposed on the surface (Shukla and Ward 1989). Right: Negative contrast electron micrograph of Plum pox virus particles. The bar represents 200 nm (Hull, 2014).
1.10.1 Potyviridae genomes

*Potyviridae* have single stranded, positive-sense, monopartite RNA genomes, except for Bymoviruses, which are bipartite (Berger, 2001). *Potyviridae* genomes range from 8 – 11 Kb in size and typically encode a single, large polyprotein, which is auto-catalytically cleaved by virally encoded proteases into 10 mature proteins that serve multiple functions during infection. An additional P3N-PIPO ORF is produced through transcriptional slippage in the N-terminus of the P3 protein (Chung et al., 2008; Olsper et al., 2015; Untiveros et al., 2016). The genome organization of typical potyvirus is shown in Fig. 1.8. *Potyviridae* genomes are polyadenylated at their 3’ terminals, while their 5’ terminals are bound to the viral genome linked protein (VPg) (Valli et al., 2015). *Potyviridae* have relatively small genomes and so encode proteins that serve multiple functions (Holmes, 2003). This size limitation is due the need to ensure stability of the RNA molecule, minimize negative effects of error-prone RNA replication, and protect against host defense mechanisms (Valli et al., 2015).

![Genome organization of a typical potyvirus encoding the following proteins: viral protein genome-linked (VPg); protein 1 protease (P1); helper component protease (HC-Pro); protein 3 (P3); pretty interesting *Potyviridae* ORF (PIPO); six kilodalton peptide 1 (6K1) and 2 (6K2); cytoplasmic inclusion (CI); nuclear inclusion A protease (NiA); nuclear inclusion B RNA-dependent RNA polymerase (NiB) and coat protein (CP), as well as the 5’ and 3’ untranslated regions (UTR) (Wylie et al., 2017).](image)

1.10.2 Genera in the Potyviridae family

There are almost 200 known *Potyviridae* species, which are assigned to eight genera according to their genome sequence/organization and vector specificity, as shown in Table 1.2 (Wylie et al., 2017). The current understanding is that viruses in the *Potyvirus* and *Macluravirus* genera are transmitted non-persistently by aphids (Andret-Link and Fuchs, 2005; Flasinski and Cassidy, 1998), ipomoviruses are transmitted semi-persistently by whiteflies (*B. tabaci*) (Andret-Link and Fuchs, 2005), paceviruses, tritomoviruses, and rymoviruses are transmitted semi-persistently by eriophyid mites (Stenger et al., 2005) and bymoviruses are transmitted persistently by the plasmidiophorid protist *Polymyxa graminis* (Kanyuka et al., 2003; Dombrovsky et al., 2014). The *Potyvirus* genera contains by far the most viral species, at 160 members (Wylie et al., 2017); significant members include the type member *Potato virus Y* (PVY), *Watermelon mosaic virus* (WMV) and *Turnip mosaic virus* (TuMV) (Hull, 2014).
Table 1.2: Description of genera in the Potyviridae family (Wylie et al., 2017).

<table>
<thead>
<tr>
<th>Potyviridae genus</th>
<th>No. of members</th>
<th>Virion length (nm)</th>
<th>Vector</th>
<th>Type member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brambyvirus</td>
<td>1</td>
<td>800</td>
<td>Unknown</td>
<td>Blackberry virus Y</td>
</tr>
<tr>
<td>Bymovirus</td>
<td>6</td>
<td>275 and 550</td>
<td>Protist</td>
<td>Barley yellow mosaic virus</td>
</tr>
<tr>
<td>Ipomovirus</td>
<td>6</td>
<td>750 – 950</td>
<td>Whiteflies</td>
<td>Sweet potato mild mottle virus</td>
</tr>
<tr>
<td>Macluravirus</td>
<td>8</td>
<td>650 – 675</td>
<td>Aphids</td>
<td>Maclura mosaic virus</td>
</tr>
<tr>
<td>Poacevirus</td>
<td>3</td>
<td>890</td>
<td>Mites</td>
<td>Triticum mosaic virus</td>
</tr>
<tr>
<td>Potyvirus</td>
<td>160</td>
<td>720 – 850</td>
<td>Aphids</td>
<td>Potato virus Y</td>
</tr>
<tr>
<td>Rymovirus</td>
<td>3</td>
<td>680 – 750</td>
<td>Mites</td>
<td>Ryegrass mosaic virus</td>
</tr>
<tr>
<td>Tritomovirus</td>
<td>6</td>
<td>680 – 750</td>
<td>Mites</td>
<td>Wheat streak mosaic virus</td>
</tr>
</tbody>
</table>

1.10.3 Ipomoviruses

The Ipomovirus genera currently contains six members, including: *Sweet potato mild mottle virus* (SPMMV), *Squash vein yellowing virus* (SqVYV), *Cucumber vein yellowing virus* (CVYV), CBSV, UCBSV, the Ethiopian and Israeli lines of *Tomato mild mottle virus* (ToMMV) (Wylie et al., 2017). Ipomoviral genome sizes range from around 9 – 11 Kb and encode polyproteins ranging from 2,900 – 3,000 amino acid in length (Dombrovsky et al., 2014). Relatively few studies have been performed to characterise the functions of ipomoviral genes, transmission mechanisms and interactions with plant hosts and vectors (Dombrovsky et al., 2014).

Ipomoviruses have diverse 5’ genome structures, which are characterised by the number of P1 proteins and the presence or absence of HC-Pro, as shown in Fig. 1.9 (Valli et al., 2018). SPMMV encodes a large P1 protein, which functions as a silencing-suppressor, whereas its HC-Pro does not have suppressor activity (Giner et al., 2010). ToMMVs share a similar genome organization with SPMMV in that they have a single but smaller P1 protein and contain a HC-Pro protein (Abraham et al., 2012). CVYV and SqVYV lack HC-Pro and encode two P1 proteins: P1a and P1b; P1b has silencing suppressor activity (Valli et al., 2008). U/CBSVs also lack HC-Pro and encode a single P1 protein, which shares sequence similarity with CVYV and SqVYV P1b proteins. The UCBSV P1 protein is reported to function as a silencing suppressor (Mbanzibwa et al., 2009a). U/CBSVs encode unique Ham1 proteins located between the N1b and CP peptides (Dombrovsky et al., 2014; Mbanzibwa et al., 2009a). Valli et al., (2007) have suggested that diverse ipomoviral genome structures may have arisen through a P1 gene duplication in an ancestral potyvirus, the two P1 (P1a and P1b) proteins then evolved separately and acquired different functions. The presence of both P1a and P1b can compensate the loss of HC-Pro in CVYV and SqVYV, whereas the presence of a single P1 protein and Ham1 proteins can compensate for the loss of HC-Pro in CBSV and UCBSV.
Figure 1.9: Schematics of *Ipomovirus* genome structures. Ipomoviruses have diverse 5’ structures: Sweet potato mild mottle virus (SPMMV) and Tomato mild mottle viruses (ToMMV) have single P1 proteins, Cucumber vein yellowing virus (CVYV) and Squash vein yellowing virus (SqVYV) have two P1 proteins: P1a and P1b, whereas Cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV) do not contain HC-Pro but have single P1 proteins and unique Ham1 proteins.

### 1.10.3.1 Sweet potato mild mottle virus

SPMMV is the type member of the *Ipomovirus* genus and can infect several plant species in the *Convolvulaceae* family, including sweet potato (*Ipomea batatas*). SPMMV has been detected in South Africa, Egypt, Indonesia, New Zealand, Peru, Kenya, Uganda and Tanzania (Dombrovsky et al., 2014). It has been suggested that East Africa may be the origin of SPMMV, where it is the third most prevalent virus found in sweet potato (Ateka et al., 2004; Mukasa et al., 2003; Tairo et al., 2004). Typical symptoms during sweet potato infections include mild leaf mottling and growth stunting (Plantwise Knowledge Bank). In addition to sweet potato, SPMMV is also found in 21 wild plants in the *Convolvulaceae* family from different agro-ecological areas of Uganda, including *Hewittia sublabata, Lepistemon owariensis* (Tugume et al., 2010).
1.10.3.2 **Cucumber vein yellowing virus**
CVYV infects plants in the *Cucurbitaceae* family, including cucumber (*Cucumis sativus*), squash (*Cucurbita pepo*), watermelon (*Citrullus lanatus*), bitter apple (*C. colocynthis*) and calabash (*Lagenaria siceraria*) (Dombrovsky et al., 2014). CVYV has been reported from Iran, Sudan, Cyprus, Israel, Jordan, Portugal, Spain and Turkey (EPPO, 2007). CVYV causes vein clearing symptoms and Jordanian isolates also induce severe stunting in cucumber and watermelon (Galipienso et al., 2012; Lecoq et al., 2000).

1.10.3.3 **Squash vein yellowing virus**
SqVYV also infects plants in the *Cucurbitaceae* family, including watermelon, squash, sweet melon (*C. melo*), crookneck pumpkin (*C. moschata*), and Arikara squash (*C. maxima*) (Dombrovsky et al., 2014). SqVYV causes vein yellowing in squash and vine collapse in watermelon (Adkins, 2015). SqVYV has been reported from Florida (Adkins et al., 2007; Akad et al., 2008), Indiana (Egel and Adkins, 2007), Puerto Rico (Acevedo et al., 2013) and Guatemala (Adkins, 2015). SqVYV has also been found in wild cucurbit species: *Momordica charantia* and *Melothria pendula* in Florida, which may serve as important reservoir hosts between growing seasons (Adkins et al., 2008).

1.10.3.4 **Tomato mild mottle viruses**
The Ethiopian tomato and Israeli aubergine lines of ToMMV have been recently assigned to the *Ipomovirus* genus (Wylie et al., 2017). ToMMV was first described in Yemen, where it causes plant stunting and leaf mottling on tomato and other solanaceous hosts (Walkey et al., 1994). ToMMV was then found to be widespread in Ethiopia and is commonly associated with PVY during mixed infections of tomato and solanaceous weeds (Hiskias et al., 1999). The Ethiopian line causes stunting and leaf motting in tomato and other solanaceous hosts (Walkey et al., 1994), whereas the Israeli line causes mild motting on leaves and fruit distortion and blistering in aubergine (Dombrovsky et al., 2013). The ToMMV genome shows high levels of sequence divergence from other members of the *Ipomovirus* genus (Abraham et al., 2012).

1.10.3.5 **CBSV and UCBSV**
The two distinct ipomoviral species: CBSV and UCBSV cause CBSD (Mbanzibwa et al., 2009b; Winter et al., 2010). U/CBSVs share unusual genome features in that they have a single P1 protein, lack HC-Pro and contain unique Ham1-like proteins (Mbanzibwa et al., 2009a).
Interestingly, the Potyvirus: Euphorbia ringspot virus (EuRV) also contains a Ham1-like sequence at the corresponding genomic position (Knierim et al., 2016). U/CBSV and EuRV Ham1 sequences share homology with the Maf/Ham1/ITPase family of proteins found across prokaryotes and eukaryotes, which hydrolyze non-canonical nucleotides, to prevent their incorporation in nucleic acid and thereby reduce mutation rates (Galperin et al., 2006). This has led to the hypothesis that U/CBSV Ham1 proteins may function reduce viral mutation rates (Mbanzibwa et al., 2009a).

The viral proteins required for U/CBSV vector transmission are currently unknown. In other Potyviridae, HC-Pro is involved with aphid transmission and is proposed to act as a ‘bridge’, according to the following mechanism: PTK motifs in the HC-Pro interact with DAG motifs in the viral CP and KIC motifs in HC-Pro bind to receptors in the aphid stylet (Valli et al., 2018). This mechanism was not previously thought to be used by U/CBSVs for vector transmission as they do not encode HC-Pro. However, Ateka et al., (2017) have recently identified the presence of the DAG and PTK motifs in CBSV CPs and the KIC motif in CBSV P1 sequences, which are not found in UCBSVs. This raises the possibility that the CBSV P1 protein may function to ‘bridge’ virion attachment between the CP and receptors in aphid stylets. It also raises the possibility that CBSV could be transmitted by aphids in addition to whiteflies.

Recently, whole genome sequence analysis has shown that U/CBSVs are diverse (Alicai et al., 2016; Mbewe et al., 2017; Ndunguru et al., 2015) and specific amino acid positions in the N-terminals of U/CBSV CP sequences appear to be under positive selection, which may be allowing adaptation to host environmental conditions and vector transmission mechanisms (Mbanzibwa et al., 2011b). Key areas CBSV and UCBSV genomes show relatively high levels of sequence divergence, including the P1, P3N-PIPO, Ham1 and CP regions (Winter et al., 2010), which may be responsible for symptom differences during CBSV and UCBSV infections (Mohammed et al., 2012; Nichols, 1950; Winter et al., 2010). Evidence for intraspecific recombination events have been identified within CBSV and UCBSV isolates but to date, there has been no evidence for interspecific recombination (Mbanzibwa et al., 2011b; Ndunguru et al., 2015). To date, the only characterised U/CBSV protein function is that the UCBSV P1 protein is reported to have silencing suppressor activity (Mbanzibwa et al., 2009a).
1.11 **Overview of the typical *Potyviridae* infection cycle**

The mechanisms that occur during U/CBSV infections are poorly understood. Research on other viruses has shown that the following processes occur during a typical *Potyviridae* infection: viral entry into host cells, viral particle disassembly, viral RNA translation and genome replication, interactions with host defense and viral counter defense strategies, virion assembly, movement between cells, systemic movement and vector acquisition, as shown in Fig. 1.10 (Mäkinen and Hafrén, 2014).

![Diagram of key processes that occur during a typical *Potyviridae* infection of a plant cell.](image)

**Figure 1.10**: Diagram of key processes that occur during a typical *Potyviridae* infection of a plant cell. Following viral entry, viral particles disassemble and viral genomes are translated by host ribosomes as large polyproteins and an additional P3N-PIPO ORF produced through transcriptional slippage (1); viral genomes are replicated by viral replication complexes (VRCs) in cytoplasmic vesicles (2); replicated viral RNA is recruited to plasmodesmata for cell-to-cell and systemic movement (3); or used as template for repeated rounds of translation and replication (4); host defence and viral counter defence mechanisms interact (5); viral RNA is encapsidated in coat protein subunits to produce virions which can be acquired by vectors (6) (Mäkinen and Hafrén, 2014).

1.11.1 **Virion disassembly, polyprotein translation and processing**

Vectors probe plant cells prior to feeding and release virions from their stylets into the cytoplasm, where viral particles disassemble (Mäkinen and Hafrén, 2014). During this process VPg remains covalently bound to the viral genome (Mäkinen and Hafrén, 2014; Oruetxebarria et al., 2001). Host mRNA is covalently bound by a 7-methylguanylate (m7G) cap, which is recognized by the cap-binding complex via an interaction with eIF4E, which co-ordinates the
attachment of ribosomal subunits to initiate translation (Gallie, 1991). *Potyviridae* genomes lack m7G caps, instead they are bound to the viral VPG protein and their 5’ and/or 3’UTRs contain secondary structures that enable recruitment of the host ribosomal components to initiate cap-independent translation (Zhang et al., 2015). Positive-sense viral genomes are immediately translated by host ribosomes to produce large (340 - 395 kDa) polyproteins, which are auto catalytically cleaved by viral proteases: P1, HC-Pro and Nla into ten mature proteins (Valli et al., 2015). Cleavage sequences have different rates of cleavage which leads to the production of precursor fusion proteins that also have specific functions (Adams et al., 2005).

1.11.2 Viral genome replication
Viral genome replication is performed by viral replication complexes (VRCs) protected inside cytoplasmic vesicles, which are derived from endoplasmic reticulum (ER) and chloroplast membranes. Vesicle formation is initiated by the viral 6K2 protein (Mäkinen and Hafrén, 2014). Viral RNA is recruited to these vesicles via interactions with secondary structures in the 3’UTR of the viral genome and the viral protein Nlb (Haldeman-Cahill et al., 1998). The current model for *Potyviridae* replication is that viral proteins, including: 6K2-Vpg-Nla, Nlb and CI as well as host proteins: eIF4E, poly(A) binding protein (PABP) and heat shock proteins are recruited to the vesicles and serve as components of VRCs (Mäkinen and Hafrén, 2014). Secondary structures in the viral genome 3’UTR interact with PABP to direct Nlb to the 3’UTR. Nlb uridylates Vpg to prime RNA synthesis, Nlb (Rdrp) then replicates the viral genome, while CI helicase unravels secondary structures in the viral RNA (Ivanov et al., 2014). Negative-sense viral RNA then serves as a template to produce progeny positive-sense genomes which are released into the cytoplasm and can be translated, translocated, encapsidated or degraded (Ivanov et al., 2014). During early stages of infection, VRC containing vesicles fuse with chloroplasts to form globular structures with specialized membranes that are reported to act as viral factories (Grangeon et al., 2012).

1.11.3 Intra/intercellular and long-distance viral movement
Vesicles containing VRC move along actin microfilaments of the cytoskeleton towards plasmodesmata (Harries and Ding, 2011). Infectious viral particles are trafficked through plasmodesmata via conical structures consisting of CI proteins to infect neighboring cells (Mäkinen and Hafrén, 2014). Potyviral proteins can increase the size exclusion limit (SEL) of plasmodesmata to enable the passage of infectious viral particles through plasmodesmata (Wolf et al., 1989). *Potyviridae* proteins associated with movement include P3N-PIPO, CI, CP,
HC-Pro and VPg (Mäkinen and Hafrén, 2014). To move systemically, viruses must move from mesophyll cells, through bundle sheath cells to vascular cells in the phloem, which requires compatible interactions between host and viral proteins (Nelson and van Bel, 1998). In the phloem, viruses are actively translocated from sieve elements into companion cells for replication and translation and then reloaded back into sieve elements (Cheng et al., 2000; Silva et al., 2002). Generally, viruses move in the direction of photo-assimilates, from mature leaves (source) upwards towards upper systemic (sink) leaves (Hipper et al., 2013). In systemic leaf cells, viral replication and translation continues and viral genomes are encapsidated in CP subunits to form mature virions, which are acquired by probing vectors (Hull, 2014).

1.11.4 Suppression of host silencing
During viral replication double-stranded RNA replicative intermediates are formed which can be recognized by the plant as a sign of viral infection and trigger activation of host silencing pathway (Hull, 2014). An overview of the silencing pathway is that double-stranded RNA is recognized and cleaved by DICER like RNAse III enzymes (DCL) into 21 – 24 bp small interfering RNAs (vsirNA), which are loaded in the RNA induced silencing complex (RISC) to guide cleavage and degradation of the corresponding viral target sequence by the RISC argonaute proteins. Plant host Rdrrps amplify vsirRNAs and host methylases stabilize vsirRNAs (Jiang et al., 2012). Viruses often encode silencing suppressors which target and interfere with various stages of the silencing pathway. Characterised Potyviridae proteins with silencing suppressor activities include HC-Pro and P1, which function by binding to vsirNA to prevent loading into RISC or interfere with vsirNA methylation (Ivanov et al., 2014). Silencing suppressors, such as HC-Pro can move systemically to suppress silencing over long-distances ahead of infection (Hull, 2014). The host silencing pathway is also involved with basic host biology and so viral suppressors which interfere with this pathway can lead to physiological changes and altered host gene expression, resulting in symptom development (Voinnet, 2005).

1.12 Functions of different proteins in other Potyviridae
To date the functions of U/CBSV proteins are poorly understood; only silencing suppressor activity for the UCBSV P1 protein has been reported (Mbanzibwa et al., 2009a). U/CBSV proteins share homology with proteins sequences from other members of the Potyviridae family that have been functionally characterised and so this homology can be used to predict potential functions. The functional activities of proteins in other Potyviridae are described below.
1.12.1 P1

*Potyviridae* P1 protein is a serine protease, which cleaves itself from the polyprotein at its C-terminus. P1 proteins are highly variable in size and sequence and intra and inter-species recombination events appear to be relatively common in P1 coding regions (Valli et al., 2015). Functions for *Potyviridae* P1 proteins include: an accessory factor during viral genome amplification (Verchot and Carrington, 1995), enhancement of viral translation (Martinez and Daros, 2014) and silencing suppression (Valli et al., 2008). Mechanisms for P1 silencing suppression include: the SPMMV P1 protein which binds directly to argonaute proteins via conserved WG/GW motifs and thereby suppresses RISC formation (Giner et al., 2010). The CVVV P1b protein binds directly to vsiRNAs to prevent their incorporation into RISC (Valli et al., 2008). The UCBSV P1 protein has been shown to have silencing suppressor activity (Mbanzibwa et al., 2009a). U/CBSV P1 proteins are most closely related to P1b proteins of CVYV and SqVYV, and so it is suggested that they may also bind to vsiRNA. In other *Potyviridae*, variations in the N-terminal of P1 sequences are host-range determinants and pathogenicity determinants (Valli et al., 2007). U/CBSV isolates have variable P1 sequences, which may be associated with differential symptom development and pathogenicity during cassava infection (Mbewe et al., 2017).

1.12.2 HC-Pro

Helper component proteinase (HC-Pro) proteins are typically 50 kDa in size and are involved with many functions, including: vector transmission, polyprotein processing, silencing suppression, movement and viral replication (Valli et al., 2018). HC-Pro proteins have cysteine proteinase activity at their C-terminals, which cleave HC-Pro co-translationally from the viral polyprotein (Carrington et al., 1989). HC-Pro proteins from members of the *Potyvirus* and *Rymovirus* genera have silencing suppression activities that target multiple steps in the silencing cascade, including binding and sequestering vsiRNA to prevent loading into RISC (Lakatos et al., 2006), interference with vsiRNA methylation (Ivanov et al., 2016) and induction of miRNA168 to downregulate AGO1 expression (Várallyay and Havelda, 2013). During aphid transmission HC-Pro is proposed to act as a ‘bridge’, linking the virion to the vector stylet via a reversible interaction (Govier and Kassanis, 1974). The *Bean common mosaic virus* (BCMV) HC-Pro can also increase the plasmodesmata size exclusion limit to facilitate the movement of viral RNA (Rojas et al., 1997). The ipomoviruses CVVV, SqVVV, CBSV and UCBSV lack HC-Pro; it is hypothesized that in these genomes other proteins have evolved to perform multiple functions that compensate for the loss of HC-Pro (Valli et al., 2007).
1.12.3 P3 and P3N-PIPO

P3 is one of the least characterised Potyviridae proteins and has not been assigned any clear function (Valli et al., 2015). Bioinformatic analysis of the TuMV P3 sequence lead to the identification of the P3N-PIPO ORF, which is present across Potyviridae species (Chung et al., 2008). The P3N-PIPO protein is produced through transcriptional slippage during viral genome replication at the ‘slippery’ leader sequence: G_{1-2}A_{6-7} in the N-terminal of the P3 sequence. The homopolymeric run of ‘A’s causes the viral RdRp to stutter and results in the insertion of an additional ‘A’ base in the nascent negative-sense RNA strand, causing a +2 frameshift during translation of the sequence (Olspert et al., 2015). This frameshift occurs relatively rarely, with an estimated frequency of 0.3 – 1.0%. The frameshift leads to the translation of the fusion protein: P3N-PIPO, which is essential for TuMV infection (Olspert et al., 2015) and required for intercellular movement of virions (Chung et al., 2008). Single amino acid variations in P3 sequences serve as symptom and virulence determinants in isolates of TuMV (Jenner et al., 2003), Clover yellow vein virus (ClYVV) (Choi et al., 2013) and Zucchini yellow mosaic virus (ZYMY) (Glasa et al., 2007). P3N-PIPO length can also vary between isolates of the same species due to alternative stop codons; these P3N-PIPO alleles are associated with different rates of intercellular movement in specific hosts (Hillung et al., 2013).

1.12.4 6K1 and 6K2

The 6K1 and 6K2 proteins are both 6 kDa in size. The function(s) of 6K1 are poorly understood, it may function as a P3-6K1 fusion protein, which is hypothesized to be involved with host range definition and pathogenicity (Valli et al., 2015). The 6K2 protein contains transmembrane domains and induces the formation of large secretory vesicles from the ER and chloroplast membranes. 6K2 then tethers VRC components to these vesicles to enable replication of viral RNA (Schaad et al., 1997).

1.12.5 CI

Cylindrical inclusion (CI) proteins are typically ≈70 kDa in size and form characteristic pin-wheel structures in the cytoplasm of infected cells (Valli et al., 2015). CI proteins are have multiple functions during viral replication, translation and short/long-distance movement (Sorel et al., 2014). This multi-functionality is achieved through interactions with many viral and host proteins, in multiple sub-cellular localisations (Sorel et al., 2014). CI functions in viral replication and localises to VRC containing vesicles (Cotton et al., 2009). CI proteins contain highly conserved helicase/ATPase activity at their N-terminals, which is proposed to bind to viral RNA.
and unwind secondary structures during viral replication (Deng et al., 2015). CI is also essential for cell-to-cell and systemic movement; the current model is that CI proteins are recruited and anchored to plasmodesmata by P3N-PIPO (Wei et al., 2010), CI self-interacts to form conical structures that bind VRC containing vesicles via interactions with 6K2 (Movahed et al., 2017) and interacts with 6K2 to traffic infectious viral particles through plasmodesmata (Sorel et al., 2014). Several virulence and symptom determinants have been mapped to isolates with CI variants (Sorel et al., 2014).

1.12.6 VPg

The VPg protein is typically 20-25 kDa in size and is covalently attached the 5' of viral RNA genomes (Valli et al., 2015). *Potyviridae* VPg proteins are highly disordered, with low sequence similarity and share only a small number of conserved motifs (Jiang and Laliberté, 2011). VPg is released from the polyprotein as part of precursor fusion products due to sub-optimal cleavage sequences, which have lower cleavage rates (Schaad et al., 1996). These VPg precursors serve different functions through their different sub-cellular localizations and participation in different host/viral protein interactions (Jiang and Laliberté, 2011). In susceptible hosts, VPg interacts with the eukaryotic translation initiation factor 4E (eIF4E) or its isoform (eIF4isoE) to initiate viral polyprotein translation. During viral genome replication, VPg is uridylated by NIb, which may act to prime viral RNA synthesis (Anindya et al., 2005; Puustinen and Mäkinen, 2004). The PVA VPg protein can also function as a silencing suppressor (Rajamäki et al., 2014) and perform phloem loading of PVA during systemic infections of *Solanum commersonii* (Dunoyer et al., 2004; Rajamäki and Valkonen, 2002).

1.12.7 Nla

The Nla protein is typically ≈26 kDa in size and forms crystalline inclusions in the nuclei of infected host cells (Rajamäki and Valkonen, 2009). Nla is the main *Potyviridae* protease and cleaves at least six peptides from the viral polyprotein co-translationally at seven highly conserved cleavage sequences (Adams et al., 2005). Different protein products are cleaved at different rates to enable the production of precursor fusion proteins (Adams et al., 2005). During infection, Nla shows dynamic sub-cellular localization, moving in and out of nucleus into the cytoplasm (Martinez and Daros, 2014). The function of Nla nuclear localization is unclear but may function to disrupt host gene expression, divert host nucleolar proteins to perform viral functions (Valli et al., 2015).
1.12.8  NiB

The NiB protein is typically ≈58 kDa in size and contains RdRp and RNA binding domains that function to replicate viral RNA genomes (Hong and Hunt, 1996). NiB localizes to both the nucleus and VRC cytoplasmic vesicles during infection (Valli et al., 2015). Recruitment of NiB to VRC vesicles occurs via interactions with Nla and VPg domains of the fusion 6K2-VPg-Nla product (Dufresne et al., 2008). Inside vesicles, NiB interacts with host proteins: eIF4E, PABP and heat shock proteins, to form functional VRCs (Thivierge et al., 2008). In vitro experiments have shown that NiB uridylates VPg, which may then prime viral RNA synthesis (Anindya et al., 2005). As with Nla, the potential functions of NiB nuclear localization are largely unknown (Valli et al., 2015). During TuMV infections of Arabidopsis thaliana, NiB interacts with SUMO-conjugating enzyme: SCE1 in both the cytoplasm and nucleus. This interaction may regulate NiB activity or affect the SUMOylation pattern of cellular proteins to create a favorable cellular environment for infection (Xiong and Wang, 2013).

1.12.9  CP

Potyviridae CPs are typically ≈30-36 kDa in size and self-assemble helically to encapsidate viral genomes in filamentous virions (Revers and García, 2015). CP N-terminals are highly variable in both sequence and length and are exposed on the surface of virions along with C-terminals. The CP central core is highly conserved and primarily involved with particle architecture and subunit self-interactions (Revers and García, 2015). CPs are multifunctional and can be involved with encapsidation, infectivity, transmissibility, virion movement, host adaptation, symptom expression, silencing suppression and interactions with host defense systems (Weber and Bujarski, 2015). This multi-functionality is created through interactions with many viral, host and vector factors. For instance, Potyviridae CPs can encode the highly conserved DAG motif, which is involved with HC-Pro mediated aphid transmission (Blanc et al., 1997). Potyviridae CPs be post-translationally modified through phosphorylation, glycosylation and ubiquitination, which can affect virion assembly and/or stability and may regulate whether a viral RNA genome is translated, replicated or moved during the infection process (Revers and García, 2015). CPs can also increase the plasmodesmata size exclusion limit during cell-to-cell movement of viral RNA (Rojas et al., 1997).

1.12.10  5’ and 3’ UTRs

The 5’ and 3’ untranslated regions (UTR) of Potyviridae genomes are highly variable in terms of length and sequence (Zhang et al., 2015); they contain secondary structures, which enable and
enhance viral genome replication and translation (Roberts et al., 2015; Zhang et al., 2015). For instance, the TEV 5’ UTR acts synergistically with the poly(A) tail at the 3’ of the viral genome to enhance cap-independent translation (Zhang et al., 2015).

1.13 Plant-virus interactions

In many interactions, plant viruses can have no apparent pathological effect on plant hosts and even provide a selective advantage. However in other interactions, viral infection leads to the development of disease symptoms (Palukaitis et al., 2008). Viruses manipulate hosts through altering host gene expression, interacting with host factors, competing for metabolic resources and interfering with cellular processes. Viral infections cause altered expression of genes involved a broad range of cellular processes, such as hormonal regulation, cell cycle control and endogenous transport of macromolecules (Palukaitis et al., 2008). This manipulation means that viruses cause a wide range of microscopic and macroscopic symptoms in their plant hosts and can cause significant yield losses (Hull, 2014). The most common symptoms associated with viral infection are leaf chlorosis and growth stunti

1.13.1 Recessive resistance

To establish infection, a virus must be able to replicate within the host plant cell and move systemically, which requires close, compatible interactions between viral and host proteins (Hull, 2014). When these interactions are not compatible, recessive resistance occurs, whereby the virus is unable to use host proteins to carry out essential processes (Hashimoto et al., 2016). This is illustrated by the requirement for the potyviral VPg protein to directly interact with the host eIF4(iso)E to initiate viral protein translation and cause infection (Hashimoto et al., 2016). For instance, naturally occurring eIF4(iso)E variants that are unable to bind VPg result in Capsicum resistance to TEV infection (Yeam et al., 2007). This interaction has been targeted as an anti-viral strategy. Overexpression of genetically engineered eIF4(iso)E mutations has been used to confer recessive resistance of Brassica rapa to TuMV (Kim et al., 2014). Similarly, the
CRISPR/Cas9 system has been used to generate eIF4(iso)E mutant cassava lines, which compared with wild-type plants display delayed and attenuated CBSD symptoms (Gomez et al., 2018).

1.13.2 Dominant resistance
Plants also encode dominant resistance (R) proteins, which recognize viruses directly through the presence of pathogen associated molecular patterns (PAMPs) or indirectly, through their effect on plant cells during infection called damage associated molecular patterns (DAMPs) (Jones and Dangl, 2006). The *N. glutinosa* N protein is a TIR-NB-LRR type receptor that recognizes the TMV replicase gene; this recognition triggers a signaling cascade that leads to a reactive oxygen burst and the hypersensitive response (HR) (Marathe et al., 2002). HR is associated with characteristic cellular changes, including the expression of defense related proteins, cell wall lignin and callose deposition, lipid peroxidation that lead to rapid cell death and limit the spread of the virus from the initial infection site (Mandadi and Scholthof, 2013). After the initial HR at the localized infection site, systemic acquired resistance (SAR) occurs whereby distant non-infected tissues show reduced pathogen susceptibility (Nicaise, 2014). Viruses are therefore under selective pressure to avoid R protein recognition to prevent HR and SAR induction (Hajimorad et al., 2003).

1.14 Viral sequence determinants
Viral sequences are associated with symptom development and serve as virulence or avirulence and host-range determinants. Between viral isolates of the same species, small sequence variations can be associated with drastic changes in symptom development. For instance, a single amino acid variant in the Chilean 2 *Pepino mosaic virus* (PepMV) is associated with the development of necrosis (Hasiów-Jaroszewska and Borodyńko, 2012). Examples of *Potyviridae* symptom and virulence determinants include the CI sequence of *Soybean mosaic virus* (SMV: Zhang et al., 2009) and the HC-Pro sequence of PVY (Tribodet et al., 2005). ICs have also been constructed of RNA viruses; a tobamovirus IC was used to map the region of the viral genome that is able to overcome N gene mediated resistance in *N. tabacum* (Padgett and Beachy, 1993).

1.15 ICs as tool to study gene functions and breed for resistance
Infectious clones consist of cloned DNA copies of viral genome sequences that have been inserted into bacterial plasmids. ICs are used to inoculate plants; they provide inoculum that is
genetically uniform and stable. This viral sequence consistency is critical during resistance breeding to accurately compare infection responses to the same viral inoculum between plant lines (Brewer et al., 2018). ICs can also be manipulated to characterise sequences involved with symptom development, virulence/avirulence, host-range, host interactions, movement, vector transmission. Chimeric ICs have been used to map viral sequences associated with different symptom phenotypes in isolates of *Maize streak virus* (MSV: Martin and Rybicki, 2002) and PepMV (Duff-Farrier et al., 2015). Marker gene sequences can also be inserted into ICs to track viral localization, replication and movement within the plant host (Dietrich, 2003).

CMG ICs have been used to gain understanding of CMD infections, including the rolling circle mechanism of viral replication (Saunders et al., 1991), synergism between African CMG viral species (Beachy et al., 2000) and the identification of satellite DNA sequences which alter CMG symptom development in cassava (Ndunguru et al., 2016). Critically CMG ICs have been used to characterise CMD infection responses in cassava breeding lines (Kuria et al., 2017), as well as levels of CMD resistance in transgenic cassava (Zhang et al., 2005). These advances have contributed towards the development and deployment of improved cassava varieties with strong CMD resistance across SAA (Fondong, 2017). This demonstrates how ICs are highly valuable tools in controlling and understanding viral plant diseases.

### 1.16 Aims and objectives of this study

ICs are urgently needed for U/CBSVs to characterise viral gene functions, viral interactions with hosts and vectors and to screen for broad CBSD resistance. Unfortunately, the construction of U/CBSV ICs has previously been prevented by sequence instability issues during plasmid propagation in *E. coli*. Two U/CBSV ICs were recently constructed by members of the Molecular Plant Pathology group at the University of Bristol. After their construction, these ICs needed to be verified for sequence stability during plasmid propagation in *E. coli* to ensure that they can be used as a consistent viral inoculum source. The ability for the U/CBSV ICs to cause infections in model hosts and cassava also needed to be tested, as these ICs should be able to efficiently cause infections that are identical to wild-type infections. Investigations into the utility of the U/CBSV ICs are described in Chapter 3.

ICs are valuable tools that can be used to gain insights into viral infection mechanisms, host interactions and for the characterisation of sequence determinants associated with symptom development. The function of U/CBSV genes are currently poorly understood; only the silencing suppressor activity of the UCBSV P1 protein has been reported (Mbanzibwa et al., 2009a). The U/CBSV Ham1 gene is of particular interest because this sequence has only been
reported in one other plant virus (EuRV) (Knierim et al., 2016) and is hypothesized to remove non-canonical nucleotides, thereby reducing viral mutations (Mbanzibwa et al., 2009a). In Chapters 4 and 5, the potential functions of U/CBSV Ham1 proteins are investigated through manipulation of the U/CBSV ICs. Finally, the U/CBSV genome sequences involved with symptom development are currently uncharacterised. In Chapters 3 and 4, the CP and Ham1 regions were investigated for their roles in symptom development during CBSV infections of *N. benthamiana*.

The overall aims of this study were:

- To verify the sequence stability of the CBSV Tanza and UCBSV Kikombe ICs for during plasmid propagation in *E. coli*.

- To investigate the abilities of the CBSV Tanza and UCBSV Kikombe ICs to cause infections in model hosts as well as cassava.

- To use U/CBSV ICs to visualize viral replication and movement *in-planta*.

- To use U/CBSV ICs to characterize viral sequences associated with symptom development.

- To investigate the potential function of U/CBSV Ham1 proteins to hydrolyze non-canonical nucleotides and thereby reduce viral mutation rate.
Chapter 2: Materials and Methods

This Chapter describes the general materials and methods used to generate the results in Chapters 3 – 5. All plastic and glassware, media and buffers were autoclaved at 121°C for 15 mins at 15 pound-force per square inch (psi). Sterile deionised water (SDW) was used to prepare solutions. Specific temperature sensitive reagents were filter-sterilised using 0.22 µm syringe filters (Millipore). Chemicals and media used in this study were molecular biology grade and purchased from Sigma-Aldrich or Thermo Fisher Scientific unless otherwise specified. Primers were purchased from Integrated DNA Technologies. Commercial kits were purchased from Thermo Fisher Scientific, Omega Bio-Tek, Takara Bio USA, Zymo Research and DSMZ.

2.1 Microbiological methods and strains

Microbial culturing was performed in a class II biosafety cabinet using a Bunsen burner and sterilised material. All work was performed under the DEFRA license No. S1045/197610/2.

2.1.1 Strains

2.1.1.1 Escherichia coli
Various *Escherichia coli* strains were used for the propagation of plasmids depending on their suitability for specific purposes, as outlined in Chapters 3 - 5. The DH5α strain (Thermo Fisher Scientific) was used for standard plasmid propagation. U/CBSV infectious clone sequence stability was tested in the following *E. coli* strains: DH5-α (Thermo Fisher Scientific), α-select (Bioline), C43 (Sigma Aldrich), One Shot ccdB survival (Thermo Fisher Scientific) and TOP10 (Thermo Fisher Scientific). The BL21 (DE3) strain (NEB) was used for viral protein expression. The genotypes for all *E. coli* strains used are provided in Appendix 2.1.

2.1.1.2 Agrobacterium tumefaciens
The *Agrobacterium tumefaciens* (*Rhizobium radiobacter*) strain LBA 4404 (Thermo Fisher Scientific) was used for the agroinfiltration of plants. The chromosomal background of LBA 4404 is TiArch5 conferring resistance to rifampicin. The strain harbors the Ti plasmid: pAL4404.

2.1.1.3 Saccharomyces cerevisiae
The *Saccharomyces cerevisiae* (yeast) strain YPH 499 was used for yeast homologous recombination.
2.1.1.4 U/CBSV strains

The wild-type CBSV Tanza strain (NCBI: MF975780) was isolated from *Nicotiana benthamiana* which had been mechanically inoculated with symptomatic cassava leaf material. The wild-type UCBSV Kikombe strain (NCBI: KY825166.1) was isolated from symptomatic cassava leaf material. Both strains originate from Mikocheni Rd, Tanzania and were supplied as freeze-dried leaf material, courtesy of Natural Resources Institute (NRI) (London-UK).

2.1.2 Culture media

All culture media were autoclaved prior to use.

2.1.2.1 Bacterial media

*Luria-Bertani (LB) Broth* was used to culture *E. coli* and *A. tumefaciens*. LB broth was prepared with 25 g of LB powder (Invitrogen), made up to 1000 ml with deionised water. **LB Agar** was used to grow *E. coli* and *A. tumefaciens* colonies. LB agar was prepared with 10 g of LB powder (Invitrogen) and 6 g agar, made up to 400 ml with deionised water. **Yeast extract nutrient broth** (YENB) was used to prepare electro-competent *E. coli* cells. YENB was prepared with 0.75% w/v yeast extract (Sigma Aldrich) and 0.8% w/v nutrient broth 2 (Sigma Aldrich), made up to 400 ml with deionised water. **Super optimal broth** with catabolite repression medium (SOC) was used during *E. coli* and *A. tumefaciens* transformation. SOC was prepared with 2% w/v tryptone, 0.5% w/v yeast extract, 10 mM sodium chloride, 2.5 mM potassium chloride, 10 mM magnesium chloride, 20 mM glucose, made up to 400 ml with deionised water.

2.1.2.2 *S. cerevisiae* media

- **Yeast Extract Peptone Dextrose** (YEPD) broth was used for the growth and maintenance of *S. cerevisiae*. YEPD was prepared with: 4 g yeast extract, 8 g peptone, 8 g glucose, 0.004% w/v adenine sulphate, made up to 400 ml with deionised water. **YEPD Agar** was prepared with YEPD and 6 g agar, made up to 400 ml with deionised water. **Yeast Synthetic Drop out Media** (YSDM) was used for the selection of *S. cerevisiae* transformants in yeast homologous recombination. The yeast strain YPH 499 is uracil deficient. The plasmids: pCAMBIA0380 and pYES2.1 that are used in this study contain the orotidine 5-phosphate decarboxylase (URA3) gene conferring uracil biosynthesis and enabling selection of transformants. YSDM broth was prepared with 0.68 g yeast nitrogen base, 2 g ammonium sulphate, 0.308 g yeast drop out (-ura) mix, made up to 390 ml with deionised water, after autoclaving 100 ml of filter-sterilised 20% glucose.
(w/v) was added. **YSDM Agar** was prepared the same as YSDM broth but with the addition of 8 g agar.

### 2.1.2.3 Antibiotic selection

To enable selection of transformants, culture media was supplemented with antibiotics. Antibiotics were purchased and prepared according to Table 2.1. The choice of antibiotic depended on the presence of antibiotic resistance genes on the plasmid used in bacterial transformation.

Table 2.1: Preparation of antibiotics used in this study.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Supplier</th>
<th>Dissolved in</th>
<th>Stock (mg/ml)</th>
<th>Working (µg/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>Sigma Aldrich</td>
<td>SDW</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Carbenicillin</td>
<td>Sigma Aldrich</td>
<td>SDW</td>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Sigma Aldrich</td>
<td>SDW</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>Melford</td>
<td>Methanol</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>Duchefa Biochemie</td>
<td>Ethanol</td>
<td>43</td>
<td>170</td>
</tr>
</tbody>
</table>

### 2.1.2.4 Microbial culture and storage

To obtain single *E. coli*, *A. tumefaciens* and *S. cerevisiae* colonies, glycerol stocks were streaked across LB agar plates using a flame-sterilised inoculating loop. Plates were stored at 4°C for up to four weeks. Glycerol stocks were used for long term microbial storage; 500 µl of culture was added to 500 µl of 20% v/v filter-sterilised glycerol in a cryovial, which was snap frozen in liquid nitrogen and stored at -80°C.

#### 2.1.2.4.1 E. coli

LB agar plates containing the appropriate antibiotics were inverted and incubated at 37°C overnight. Individual *E. coli* colonies were cultured in LB broth containing the appropriate antibiotic in an orbital shaker at 37°C and 200 rpm overnight.

#### 2.1.2.4.2 A. tumefaciens

LB agar plates containing the appropriate antibiotics were inverted and incubated at 28°C for 48 hours. Individual *A. tumefaciens* colonies were cultured LB broth containing the appropriate antibiotics in an orbital incubator at 28°C and 200 rpm for 48 hours.

#### 2.1.2.4.3 S. cerevisiae

YDSM agar plates were inverted and incubated at 28°C for 48 hours. Individual *S. cerevisiae* colonies were cultured in YEPD or YDSM in an orbital incubator at 28°C and 200 rpm for 48 hours.
2.2 Molecular methods

2.2.1 E. coli plasmid extraction

2.2.1.1 Small-scale (miniprep) E. coli plasmid extraction

A single E. coli colony was used to inoculate 10 ml of LB broth containing the appropriate antibiotic and incubated in a rotary incubator at 37°C and 200 rpm overnight. The GeneJET Plasmid miniprep kit (Thermo Fisher Scientific) was used to extract plasmids from 5 ml of culture following the manufacturer’s instructions.

2.2.1.2 Medium-scale (midiprep) E. coli plasmid extraction

To obtain more concentrated infectious clone plasmids in Chapter 3, a medium-scale plasmid extraction was performed from 50 ml E. coli cultures. A single E. coli colony was used to inoculate 5 ml of LB broth containing the appropriate antibiotic and incubated in a rotary incubator at 37°C and 200 rpm overnight. A 50 μl aliquot of this starter culture was used to inoculate 50 ml of LB broth containing the appropriate antibiotic. The culture was incubated at 37°C and 200 rpm until the OD₆₀₀ was 2 – 3. The GeneJET Plasmid midiprep kit (Thermo Fisher Scientific) was used to extract plasmid from 50 ml of culture following the manufacturer’s instructions.

2.2.1.3 Large-scale (maxiprep) E. coli plasmid extraction

In a second attempt to obtain more concentrated infectious clone plasmids in Chapter 3, a large-scale plasmid extraction was performed from 500 ml E. coli cultures, according to Green and Sambrook, (2001). The following solutions were prepared: STE buffer (10 mM Tris-HCl, 0.1 M NaCl, 1 mM EDTA - pH 8.0); Lysis solution 1 (50 mM Glucose, 25 mM Tris-HCl - pH 8.0, 10 mM EDTA); Lysis solution 2 (0.2 M NaOH, 1% w/v SDS); Lysis solution 3 - 50 ml (5 M Potassium acetate – pH 6.0, 11.5 ml Glacial acetic acid, water up to 50 ml); Lysozyme (10 mg lysozyme dissolved in 10 mL Tris-HCL - pH 8.0) and TE buffer (10 mM Tris-HCl - pH 8.0, 1 mM EDTA).

- A single E. coli colony was used to inoculate 30 ml of LB broth containing the appropriate antibiotic in a sterile 200 ml flask. The culture was incubated at 37°C with 200 rpm until the OD₆₀₀ was approximately 0.6.
- 25 ml of the culture was used to inoculate 500 ml of pre-warmed LB broth containing the appropriate antibiotic, in a sterile 2 L flask. The culture was incubated for 2.5 hours at 37°C with 200 rpm until the OD₆₀₀ was approximately 0.4
2.5 ml of 34 mg/ml chloramphenicol was added to the culture, to a final concentration of 170 μg/ml. The culture was incubated for 12 – 16 hours at 37°C and 200 rpm.

The culture was centrifuged at 2700 Xg for 15 mins at 4°C, the supernatant was decanted, the pellet was re-suspended in 200 ml of ice-cold STE buffer and centrifuged as before.

The supernatant was decanted, the pellet was resuspended in 10 ml of lysis solution 1 and 1 ml of freshly prepared 10 mg/ml lysozyme was added.

20 ml of freshly prepared lysis solution 2 was added and the contents were mixed by inversion and incubated for 10 mins at room temperature.

15 ml of ice-cold lysis solution 3 was added and the contents were mixed by inversion, placed on ice for 10 mins and centrifuged at 20,000 Xg for 30 mins at 4°C.

The supernatant was transferred to a new tube, 0.6X volume of isopropanol was added, the contents were mixed by gentle inversion at room temperature for 10 mins and centrifuged at 12,000 Xg for 15 mins at room temperature to pellet the plasmid DNA.

Isopropanol was removed with a pipette; the pellet was centrifuged as before, and the remaining isopropanol was drained off. The pellet was washed with 70% v/v ethanol and centrifuged as before. The supernatant was removed, and the pellet was left to air dry.

The pellet was dissolved in 800 μl of TE buffer. The crude plasmid was purified using the GeneJET Plasmid midiprep column (Thermo Fisher Scientific).

2.2.2 S. cerevisiae plasmid extraction

Recombinant plasmids were extracted from 3 ml of S. cerevisiae culture grown in YDSM. Plasmids were extracted using the Zymoprep Yeast Miniprep I kit (Zymo Research) according to the manufacturer’s protocol. Yeast plasmids (2 μl) were then transformed into electro-competent E. coli for propagation.

2.2.3 Plant RNA extraction

RNA was extracted from plants using a range of methods depending on the plant species and down-stream requirements. RNA was extracted from upper systemic leaves of infected plants. Approximately 0.2 g of leaf material was ground in a sterile pestle and mortar with liquid nitrogen into a fine power, which was immediately transferred into a sterile Eppendorf tube. The RNA concentration and purity of samples were calculated according to 260/280 nm
absorbance values that were measured with a NanoDrop ND-100 Spectrophotometer (Thermo Fisher Scientific). RNA integrity was visualised on a 1% w/v agarose gel.

2.2.3.1 RNA extraction from *Nicotiana* sp. using TRizol

RNA was extracted from *Nicotiana* plant species using TRizol (Thermo Fisher Scientific), according to the following protocol:

- 1 ml of TRizol was added to ground plant material and the sample was vortexed for 1 min. The sample was incubated at room temperature for 5 mins, 0.2 ml chloroform was added, and the sample was mixed by gentle inversion eight times.
- The sample was incubated at room temperature for 3 mins and centrifuged at 12,000 Xg for 15 mins at 4°C.
- The upper aqueous phase was transferred to a fresh Eppendorf tube containing 0.2 ml chloroform. The sample was mixed by gentle inversion eight times and centrifuged at 12,000 Xg for 15 mins at 4°C.
- The upper aqueous phase was transferred to a new Eppendorf tube containing 0.5 ml ice-cold isopropanol. The sample was incubated at room temperature for 10 mins and centrifuged at 12,000 Xg for 10 mins at 4°C.
- The isopropanol was removed, and the pellet was washed with 1 ml of 75% v/v ethanol. The tube was vortexed briefly and then centrifuged at 7500 Xg for 5 mins at 4°C.
- The pellet was air dried for 10 mins and then resuspend in 50 µl of DEPC treated water.

2.2.3.2 RNA extraction from *Nicotiana benthamiana* using the E.Z.N.A. Plant RNA

To extract RNA from *N. benthamiana* for use in the deep sequencing experiment in Chapter 4, the E.Z.N.A. Plant RNA kit (Omega Bio-Tek) was used according to manufacturer’s protocol with no modifications.

2.2.3.3 CTAB RNA extraction from cassava

RNA was extracted from cassava according to the Ndunguru et al., (2015) with minor modifications.

- CTAB buffer was freshly prepared and autoclaved (2X cetyltrimethylammonium bromide (CTAB), 100 mM Tris-HCl, 20 mM Ethylenediaminetetraacetic acid (EDTA), 1.4 M NaCl). Shortly before use, 700 µl aliquots of CTAB buffer were heated to 65°C and
5% v/v β-mercaptoethanol and 0.1 mg/ml proteinase K (Thermo Fisher Scientific) were added. The CTAB buffer aliquot was added to ground cassava material.

- The sample was vortexed and then incubated at 65°C for 30 mins, mixed gently by inversion every 10 mins and then left at room temperature for 10 mins.
- An equal volume of chloroform: isoamylalcohol (24:1) was added to the sample, which was mixed by gentle inversion for 10 mins and centrifuged at 13,000 Xg for 10 mins.
- The upper aqueous phase was transferred to a new Eppendorf tube and an equal volume of chloroform: isoamylalcohol (24:1) was added. The sample was mixed by inversion for 10 mins and centrifuged at 13,000 Xg for 10 mins.
- The upper aqueous phase was transferred to a new Eppendorf tube containing a 0.7 volume of ice-cold isopropanol. The sample was gently inverted and centrifuged at 14,000 Xg for 30 mins at 4°C to pellet the RNA.
- The supernatant was removed and 500 µl of 70% v/v ethanol was used to wash the pellet.
- The sample was centrifuged at 14,000 Xg for 10 mins, the ethanol was removed, and the pellet was left to air dry for 40 mins.
- The pellet was then resuspended in 200 µl DEPC treated water, an equal volume of 8 M LiCl was added and the sample was incubated at 4°C overnight.
- The sample was then centrifuged at 13,000 Xg for 30 mins at 4°C to re-pellet the RNA.
- The supernatant was removed and 500 µl of 70% v/v ethanol was used to wash the pellet by gentle tapping. The sample was centrifuged at 14,3000 Xg for 10 min, the ethanol was removed, and the pellet was left to air dry for 40 mins. The pellet was then resuspended in 50 µl DEPC treated water.

2.2.4 Nucleic acid manipulation

2.2.4.1 Agarose gel electrophoresis

Nucleic acid was analysed through agarose gel electrophoresis. A 1X TAE buffer was prepared with 40 mM Tris, 20 mM acetic acid and 1 mM EDTA. Products smaller than 300 bp were analysed on 2% w/v agarose gels, whereas products larger than 300 bp were analysed in 1% w/v agarose gels. Agarose gels were prepared by melting the appropriate amount of Type 1 molecular grade agarose (Bioline) in 1X TAE buffer. Melted agarose was left to cool to 50 - 60°C before the nucleic acid stain Midori Green Advance Nucleic Acid Stain (Bulldog Bio) was added at 1000X. Agarose was poured into a gel mold with a well-comb and left to solidify. Once
solidified, the well-comb was removed, and the gel was transferred to an electrophoresis tank containing 1 X TAE buffer. Gels were run at 120V for 30 – 40 mins, using the PowerPac Power Supply (Bio-Rad). To analyse molecular size of nucleic acid, a 5 µl aliquot of molecular weight ladder was added to the first lane. Two ladders were used, either the 1 Kb HyperLadder (Bioline) or Quick-Load Purple 2-Log DNA Ladder (NEB), depending on availability (Fig. 2.1). Gels were analysed under UV light using the ChemDoc Bio-Rad System and images were taken using the Quantity One 1D software (Bio-Rad).

2.2.4.2 Agarose gel DNA purification

DNA was purified from agarose gels using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific). Bands of interest were removed from the gel using a sterile razor, weighed and placed in a sterile Eppendorf tube. DNA was then purified according to the kit manufacturer’s instructions. To maximise the concentration of purified DNA, the elution buffer was heated to 65°C and the elution volume was reduced from 50 µl to 30 µl.

2.2.5 Polymerase chain reaction

Polymerase chain reactions (PCR) were performed in Mastercycler gradient (Eppendorf) or Prime Gradient Thermo-cycler (Techne) machines.
2.2.5.1 Primers

All primers were ordered from Integrated DNA Technologies at 100 µm stock concentration and were diluted to 10 µm with SDW prior to use in PCR. Primer sequences will be provided within the relevant results Chapters. The ‘annealing temperatures’ used during PCR thermocycles were calculated using the NEB online calculator: https://tmcalculator.neb.com.

2.2.5.2 Low fidelity PCR

For analytical PCR, e.g. culture PCR, low fidelity PCR was performed using Green DreamTaq DNA polymerase (Thermo Fisher Scientific). A 2X DreamTaq master mix was prepared containing 2 ml of 10X buffer, 400 µl dNTPs (40 µl of each 10 mM dNTP mixed thoroughly in 240 µl SDW), 400 µl DreamTaq DNA polymerase (5 U/µl), made up to 10 ml with SDW. For each PCR reaction, reagents were added as follows:

- 10 µl of 2X DreamTaq master mix
- 1 µl of 10 µm Forward primer
- 1 µl of 10 µm Reverse primer
- 1 µl (1 – 10 ng) of Template DNA
- SDW up to 20 µl

The following thermo-cycle conditions were used: initial denaturation at 94°C for 2 mins, followed by 35 cycles of (denaturation at 95°C for 30 secs, primer annealing temperature for 30 secs, 72°C for 1 min per 1 Kb) and a final extension of 72°C for 10 mins.

2.2.5.3 Culture PCR

To screen E. coli and A. tumefaciens cultures for the presence of transformant plasmids, culture PCR was performed. Taq PCR reactions were set up as above, containing 2 µl of culture as template. During thermo-cycling, the initial denaturation time at 95°C was increased from 2 mins to 10 mins.

2.2.5.4 High fidelity PCR

High fidelity PCR was performed when a high level of sequence fidelity was required, e.g. amplification of PCR fragments used in plasmid construction. PCR was performed using Phusion polymerase (Thermo Fisher Scientific). For each 20 µl reaction, reagents were added according to the manufacturers protocol, as follows:

- 4 µl of 5X Phusion buffer
- 0.4 µl of 10 mM dNTPs
- 1 µl of 10 µm Forward primer
- 1 µl of 10 µm Reverse primer
- 0.2 µl of Phusion DNA polymerase
- 1 µl (1 – 10 ng) of template DNA
- SDW up to 20 µl

The following thermo-cycle conditions were used: initial denaturation at 98°C for 30 secs, followed by 35 cycles of (denaturation at 98°C for 10 secs, primer annealing temperature for 30 secs, 72°C for 30 secs per 1 Kb) and a final extension of 72°C for 10 mins.

**2.2.5.5 Reverse-transcription PCR**

To identify the presence of U/CBSV in inoculated plants, reverse-transcription PCR was performed. Genomic DNA was first removed from plant RNA samples using DNase I (Thermo Fisher Scientific) according the manufacturer’s protocol. The reaction was set up as follows:

- 0.7 - 2 µg of total plant RNA
- 1 µl DNase I (1 U)
- 1 µl Reaction buffer containing MgCl₂
- Nuclease-free water up to 10 µl

The sample was incubated at 37°C for 1.5 hours. Then 0.5 µl 50 mM EDTA was added and the sample was heated to 65°C for 10 mins to terminate the reaction.

First strand cDNA was synthesised from RNA using the RevertAid First Strand cDNA synthesis kit (Thermo Fisher Scientific) according to the manufacturer’s protocol. The reaction was set up as follows:

- 1 µl of 100 µm Oligo d(T)₁₈ primer
- 10 µl of DNase I treated plant RNA

The sample was incubated at 65°C for 5 mins, spun down and then chilled on ice for 10 mins.

Then the following reagents were added:

- 4 µl of 5X Reaction buffer
- 0.5 µl of RiboLock RNase Inhibitor (20 U/µl)
- 1 µl of 10 mM dNTPs
- 1 µl of RevertAid M-MuLV RevertAid Transcriptase (20 U/µl).

The sample was incubated at 37°C for 60 mins and then heated at 70°C for 5 mins to terminate the reaction. Reverse-transcription PCR was performed using 2 µl of cDNA template.
Quantitative reverse-transcription PCR (qPCR) was performed to quantify transcript abundance relative to the abundance of an endogenous plant reference gene. Viral specific primers for qPCR were designed using the online software Primer 3 plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi). Parameters used to design optimal qPCR primers were: primer length of 20 bp, Tm of 60°C, GC content of 60% and amplified product size of 150 – 180 bp. The *N. benthamiana* F-BOX gene was used as the endogenous reference gene, as it is reported to show relatively stable expression during viral infections of *N. benthamiana* (Liu et al., 2012). All primers used in qPCR were validated for amplification efficiency using a 1:10 serial dilution of template; only primers with an $R^2$ >0.99 were used. Inspection of the qPCR amplification peaks confirmed single amplification peaks. The sequences of primers used in qPCR in this study are provided in Table 2.2.

### Table 2.2: Sequences of primers used in qPCR in this study.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’ – 3’</th>
<th>Amplicon size (bp)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>qPCR_CBSV_CP_Fw</td>
<td>ACTTCCTAGCCGAAGCACAA</td>
<td>163</td>
<td>CBSV coat protein</td>
</tr>
<tr>
<td>qPCR_CBSV_CP_Rv</td>
<td>GCACTAACATCCCAGGTAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qPCR_Fbox_Fw</td>
<td>GGCACTCAACAGCTCATTTCC</td>
<td>127</td>
<td><em>N. benthamiana</em> F-BOX (TAIR: At5g15710)</td>
</tr>
<tr>
<td>qPCR_Fbox_Rv</td>
<td>ACCTGGGAGGCATCCTGTTAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qPCR_ITPA_Fw</td>
<td>AAGGGTCTCCAGGTACTCA</td>
<td>123</td>
<td><em>N. benthamiana</em> predicted ITPA</td>
</tr>
<tr>
<td>qPCR_ITPA_Rv</td>
<td>TGTTAGGGCCCTCAGACCA</td>
<td></td>
<td>(SolGenomicsNetwork: Niben101Scf18106g00001.1)</td>
</tr>
</tbody>
</table>

Reactions were set up using Maxima SYBR Green/ROX (Thermo Fisher Scientific) following the manufacturer’s instructions, as follows:

- 7.5 μl SyberGreen
- 0.75 μl 10 mM Forward primer
- 0.75 μl 10 mM Reverse primer
- 2 μl Nuclease-free water
- 2 μl cDNA, diluted in nuclease-free water at a ratio of 1:2.

QPCR reactions were performed using a Stratagene MX3005 thermocycler. The following programme cycle was used: initial denaturation at 95°C for 10 mins, followed by 40 cycles of (denaturation at 95°C for 15 secs, primer annealing at 60°C for 30 secs and extension at 72°C for 30 secs). Data was acquired during the extension phase. The MxPro software calculated the threshold Ct values. Viral transcript abundance was calculated relative to the reference gene transcript abundance using the $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen, 2001).
2.2.6 Enzymatic restriction digest

2.2.6.1 Fast enzymatic restriction digest

Fast enzymatic restriction digests were performed to prepare vectors for cloning and to verify the integrity of plasmids. Plasmids were digested with specified Fast Digest restriction enzymes (Thermo Fisher Scientific) according to the manufacturer’s protocol, as follows:

- 5 µl – 20 µl of plasmid DNA
- 5 µl of Fast Digest Green Buffer
- 4 µl of Fast Digest enzyme
- Nuclease-free water up to 50 µl

Reactions were incubated at 37°C for 4 hours and then analysed by 1% w/v agarose gel electrophoresis.

2.2.6.2 High-fidelity restriction digest

NEB High-fidelity restriction enzymes were used during POPINF cloning in Chapter 5. Reactions were set up as follows:

- 20 µl of POPINF plasmid DNA
- 5 µl of NEB Buffer
- 1 µl of each NEB High-fidelity enzyme
- Nuclease-free water up to 50 µl

Reactions were incubated at 37°C for 12 – 16 hours and then analysed by 1% w/v agarose gel electrophoresis.

2.2.7 Cloning

2.2.7.1 Cloning PCR products into the pJET2.1 vector

To determine the sequence of PCR products, they were cloned into the pJET2.1 vector using the CloneJET PCR Cloning Kit (Thermo Fisher Scientific). PCR products were amplified with Phusion polymerase, gel purified using the GeneJET Gel Extraction Kit (Thermo Fisher Scientific) and ligated into pJET2.1 according to the following protocol:

- 50 – 60 ng of purified PCR product
- 0.5 µl of pJET1.2 blunt cloning vector
- 0.5 µl of T4 DNA Ligase (5U/µl)
- Nuclease free water up to 10 µl
Reactions were incubated at room temperature for 30 mins and then 2 µl of the reaction was used to transform electro-competent *E. coli*. Transformant colonies were cultured and plasmids were extracted. To confirm the presence of the insert in pJET1.2, PCR was performed on plasmids using the primers: pJET2.1Fwd (CGACTCAGTTTCCATACAGC) and pJET2.1Rev (AAGACATCGATTTTCCATACAGC). The pJET2.1Fwd/Rev primers were also used to sequence inserts in the pJET2.1 plasmid.

2.2.7.2 **Cloning PCR products into the POPINF vector**

To express CBSV Ham1 proteins in *E. coli*, the POPINF expression vector was used (OPPF-UK). The CBSV Ham1 sequence was amplified by PCR with Phusion polymerase and primers designed to contain 18–20 bp of POPINF sequence and 20–26 bp of insert sequence to enable recombination of the insert fragment into the POPINF vector. POPINF was linearized with high-fidelity restriction enzymes (NEB): *Kpn* I and *Hind* III at 37°C for 16 hours. Both the insert PCR fragment and linearized POPINF were gel purified and their concentrations were analysed using the NanoDrop ND-100 Spectrophotometer (Thermo Fisher Scientific). InFusion cloning (Takara Bio USA, Inc.) reactions were set up as follows:

- 100 ng of purified PCR product
- 100 ng of purified, digested POPINF vector
- 2 µl of 5X InFusion HD Enzyme pre-mix
- Nuclease free water up to 10 µl

Reactions were incubated for 15 mins at 50°C and then chilled on ice. 2 µl of the reaction was used to transform electro-competent *E. coli*. LB agar plates were prepared containing 100 mg/ml carbenicillin and a sterile spreader was used to cover the surface of plates with 40 µl of 0.1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG) and 120 µl of 20 mg/ml X-Gal substrate. Transformant *E. coli* were spread onto these plates and incubated at 37°C overnight. The presence of white colonies indicating disruption of the β-galactosidase (*lacZ*) gene during cloning, these colonies were cultured, and plasmids extracted. The plasmids were sequenced with the T7Fwd primer (TAATACGACTCAGTATAG) to confirm the presence of the insert sequence in the POPINF vector.

2.2.7.3 **Infectious clone construction using yeast homologous recombination**

Yeast homologous recombination was used to construct U/CBSV infectious clone plasmids, according to Raymond et al., (1999). Plasmid vectors were first digested with appropriate restriction enzymes. PCR fragments were designed to contain homologous sequences so that
once transformed into yeast, overlapping sequences would be recombined into the plasmid vector through yeast homologous recombination.

- An individual yeast colony was used to inoculate 10 ml of YEPD media and the culture was incubated at 28°C and 200 rpm overnight.
- This starter culture was used to inoculate 40 ml of YEPD media in a sterile conical flask and the culture was incubated for 4 – 5 hours at 28°C and 200 rpm.
- The culture was transferred to a sterile falcon tube and centrifuged at 2,000 Xg for 5 mins.
- The supernatant was decanted, and the cell pellet was resuspended in 25 ml of SDW and centrifuged as before.
- The supernatant was decanted, and the cells were resuspended in 1 ml of 0.1 M LiOAc. The mixture was transferred to a new sterile Eppendorf tube and centrifuged at 13,000 Xg for 15 secs.
- The pellet was resuspended in 400 μl of 0.1 M LiOAc and 50 μl aliquots were transferred to new Eppendorf tubes.
- The following reagents were added to the cells: 240 μl of 50% v/v filter-sterilised polyethylene glycerol (PEG), 50 μl of 2 μg/μl denatured salmon sperm DNA (boiled at 95°C for 5 mins then chilled on ice), 36 μl 1 M LiOAc and 34 μl DNA mix (7.5 μl PCR each DNA fragment, 5 μl digested plasmid, made up to 34 μl with SDW).
- The mixture was vortexed for 30 secs, incubated at 30°C for 30 mins and then 42°C for 30 mins.
- The mixture was centrifuged at 6000 Xg for 15 secs, resuspended in 500 μl SDW, 200 μl aliquots of the transformation mix were plated out onto YDSM agar plates and incubated at 28°C for 48 hours.

2.3 Bacterial transformation

2.3.1 Preparation of electro-competent E. coli

- An individual E. coli colony was used to inoculate 5 ml of YENB broth and the culture was incubated at 37°C and 200 rpm.
- Five flasks containing 50 ml YENB were each inoculated with 0.5 ml of the overnight culture and incubated at 37°C and 200 rpm until the OD600 was 0.5 – 0.9.
- The cultures were transferred to sterile falcon tubes and centrifuged at 6,000 Xg and 4°C for 20 mins.
• The supernatant was decanted, the pellet was resuspended in 10 ml of sterile water and then centrifuged at 6,000 Xg and 4°C for 10 mins.
• The supernatant was decanted, the pellet was resuspended in 20 ml of filter-sterilised 10% v/v glycerol and centrifuged at 6,000 Xg and 4°C for 10 mins.
• The supernatant was decanted, the pellet was resuspended in 1 ml of filter-sterilised 10% v/v glycerol and 50 μl aliquots were transferred to sterile Eppendorf tubes, which were snap frozen in liquid nitrogen.
• Competent cells were tested for transformation efficiency and stored at -80°C.

2.3.2 Preparation of electro-competent A. tumefaciens
• An individual A. tumefaciens colony was used to inoculate 5 ml of LB broth supplemented with 20 μg/ml rifampicin and cultured for 48 hours in an orbital shaker at 28°C and 200 rpm.
• An aliquot of 500 μl of culture was used to inoculate 50 ml LB broth supplemented with 20 μg/ml rifampicin and incubated until the OD600 was >0.5.
• The culture was transferred to a sterile falcon tube and centrifuged for 10 mins at 4,000 Xg at 4°C.
• The supernatant was decanted, the pellet was resuspended in 50 ml of filter-sterilised 10% v/v glycerol and centrifuged as before.
• The supernatant was decanted, the pellet was resuspended in 25 ml of filter-sterilised 10% v/v glycerol and centrifuged as before.
• The supernatant was decanted, the pellet was resuspended in 1 ml of filter-sterilised 10% v/v glycerol.
• 100 μl aliquots were transferred to sterile Eppendorf tubes and snap frozen in liquid nitrogen.
• Competent cells were tested for transformation efficiency and stored at -80°C.

2.3.3 Electroporation
Prior to transformation, electroporation cuvettes were soaked in 70% v/v ethanol, washed with SDW and dried in a laminar flow cabinet. Once dry, cuvettes were sterilised through 4 mins of UV radiation in the CL-1000 cross-linker (UVP) and then chilled on ice before use.
2.3.3.1  *E. coli* electroporation

- A 50 µl aliquot of electro-competent *E. coli* cells were transferred from -80°C and left to thaw on ice.
- 2 µl of low-salt plasmid was added to the cells and the mixture was transferred to a sterile cuvette.
- The cuvette was pulsed with 2.5 kV using a GenePulser electroporator (Bio-RAD). Then the mixture was immediately resuspended in 200 µl of SOC media and transferred to a new sterile Eppendorf.
- The transformant mixture was incubated at 37°C with 200 rpm for 30 mins for pCAMBIA0380-based plasmids containing kanamycin resistance genes or 1 hour for pYES2.1 plasmids containing ampicillin resistance genes.
- The mixture was spread onto LB agar plates containing appropriate antibiotic and incubated at 37°C overnight.

2.3.3.2  *A. tumefaciens* electroporation

- A 100 µl aliquot of electro-competent *A. tumefaciens* cells were transferred from -80°C and left to thaw on ice.
- 2 µl of low-salt plasmid was added to the cells and the mixture was transferred to a sterile cuvette.
- The cuvette was pulsed with 2.2 kV using a GenePulser electroporator (Bio-RAD). Then the mixture was immediately resuspended in 900 µl of SOC media and transferred to a sterile Eppendorf. The mixture was incubated at 28°C with 200 rpm for 3 hours.
- The mixture was spread onto LB agar plates containing 20 mg/ml rifampicin and 50 mg/ml kanamycin and incubated at 28°C for 48 hours.

2.4  Sequence analysis

2.4.1  DNA sequencing

DNA was sequenced by GATC biotech according to their sample preparation requirements. Sequence read quality was analysed using GENtle 2.0 software.

2.4.2  Plasmid maps

Plasmid maps were built in in Clone Manager suite 7, version 7.11 software (SE Central).
2.4.3 DNA sequence alignments

ClustalW alignments of two DNA sequences were performed throughout Chapters 3 – 5, using the Needle Pairwise Sequence Alignment tool (EMBL-EBI): https://www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html. A ClustalW alignment of multiple ITPA homolog protein sequences were performed in Chapter 5, using the ClustalW Omega Alignment tool (EMBL-EBI): https://www.ebi.ac.uk/Tools/msa/clustalo/.

2.4.4 Phylogenetic sequence analysis

A phylogenetic tree of ITPA homolog protein sequences was built in Chapter 4. First sequences related to U/CBSV Ham1 proteins were identified through tblastn searches of the NCBI database. The top five sequences with the highest e-values were downloaded. Protein sequences were then uploaded into MEGA6, aligned using ClustalW and a maximum likelihood phylogenetic tree was built using the Whelan and Goldman model (Whelan et al., 2001). To build a phylogenetic tree of viral Ham1 sequences in Chapter 4, the nucleotide sequences were uploaded into MEGA6, aligned using ClustalW and again a maximum likelihood phylogenetic tree was built using the Whelan and Goldman model. Trees were rooted using out-group sequences. All phylogenetic sequence analysis was performed in MEGA6.

2.4.5 Bio-informatic sequence analysis

Bioinformatic analysis of deep-sequencing reads generated from viral RT-PCR amplicons was performed in Chapter 5. FASTQ files generated from the MiSeq Illumina sequencing run were downloaded from the Illumina Basespace server and uploaded to the Partek Flow server. Adapter sequences were trimmed from the reads. The quality of the reads was assessed to confirm a consistent number of reads between samples and high-quality reads with phred scores above 30. Low quality reads with phred scores below 30 were removed. Reads were aligned to the reference sequence using the Bowtie 2-2.2.5, very-sensitive-local pre-sets and maximum read length of 800 bp. The quality of the alignment was assessed to identify the percentage of reads which aligned to the reference and coverage depth. Further analysis of the read depth and coverage of the aligned reads was performed using the Qualimap v2.2.1 software. The LoFreq statistical algorithm LoFreq (Wilm et al., 2012) was used to call low frequency single nucleotide variants (SNVs), with the assistance of Dr. Tom Batstone at the University of Bristol School of Biological Sciences.
2.5 Protein Analysis

2.5.1 Protein expression and purification in *E. coli*

In Chapter 5, the 6X-histidine tagged CBSV Ham1 protein was expressed in *E. coli* and purified according to the protocol below. The following buffers were prepared: Resuspension buffer – pH 7.5 (150 mM NaCl, 20 mM Tris-HCl - pH 7.5); Loading buffer – pH 7.5 (0.5 M NaCl, 20 mM Tris-base, 20 mM Imidazole, 10% v/v Glycerol); Elution buffer – pH 7.5 (0.5 M NaCl, 20 mM Tris-base, 1 M Imidazole, 10% v/v Glycerol and Storage buffer – pH 7.5 (0.5 M NaCl, 50 mM Tris-base, 1 mM dithiothreitol, 50 mM MgCl$_2$, 20% v/v Glycerol).

- Electro-competent *E. coli* BL21 (DE3) were transformed with 2 µl of POPINF plasmid containing the recombinant viral protein sequence. An individual transformant colony was used to inoculate 10 ml of LB broth containing 100 µg/ml carbenicillin and incubated at 37°C and 200 rpm overnight.
- This starter culture was used to inoculate 2 L of LB broth supplemented with 100 µg/ml carbenicillin and incubated at 37°C and 200 rpm until the OD$_{600}$ was approximately 0.4 ml of 1 M IPTG was added to the culture, which was then incubated at 37°C and 200 rpm for 18 hours.
- The culture was transferred to four sterile 500 ml centrifuge bottles and centrifuged at 6,000 Xg for 30 mins at 4°C. The supernatant was decanted, and each cell pellet was resuspended in 30 ml of resuspension buffer.
- The resuspension mixture was transferred two 50 ml falcon tubes and centrifuged at 6,000 Xg for 10 mins at 4°C. Each cell pellet was snap frozen in liquid nitrogen and stored at -80°C.
- The cell pellet was removed from -80°C and left to thaw on ice. The pellet was resuspended in 30 ml of loading buffer and a glass homogeniser was used to homogenise the sample.
- A French pressure cell press (Constant Systems) was used to lyse cells by passing them through a filter at 25 pki. The lysate was collected in falcon tubes and centrifuged at 18,000 Xg for 40 mins at 4°C. The supernatant was filtered using a 0.22 µm syringe filter (Millipore) into a new falcon tube.
- A 5 ml HisTrap FF Crude histidine-tagged protein purification column (GE Healthcare) was loaded with 12 ml of elution buffer to elute any non-specifically bound protein.
- The filtered cell lysate was loaded onto the column using a syringe at an approximate rate of 1 ml/min.
To elute the protein of interest, a Fast-Protein-Liquid Chromatography (ÄKTA) machine was used to set up an imidazole concentration gradient of 50% v/v loading to elution buffer.

1 ml fractions were collected, fractions corresponding to a peak in UV absorbance were analysed by SDS-PAGE to confirm the presence of a protein band with the expected size.

Fractions containing the protein were pooled into a 15 ml falcon tube and gently loaded into a SnakeSkin 10 kDa dialysis tube (Thermo Fisher Scientific). The sleeve was clipped at both ends and left to dialyse overnight in 4 L of storage buffer with gentle mixing using a magnetic stirrer.

The dialysed protein solution was transferred to a 15 ml falcon tube and centrifuged at 15,000 Xg for 15 mins. To increase the concentration of the protein, the sample was transferred to a 15 mL Vivaspin column (GE Healthcare) with a molecular weight cut-off of 10,000 KDa. The column was centrifuged at 4,500 Xg until the protein concentration was above 0.5 mg/ml.

2.5.2 Quantification of U/CBSV Ham1 protein concentration

The concentrations of U/CBSV Ham1 protein extractions were quantified using the Bradford Assay. Bovine Serum Albumin (BSA) protein was dissolved in the storage buffer at the stock concentration of 1 mg/ml. This was then used to set up standard 30 µl solutions containing a range of BSA concentrations: 0, 0.16, 0.32, 0.48, 0.64, 0.80, 0.96 mg/ml. 1.5 ml of Bradford reagent (Bio-Rad) was then added to these standards and tubes were left at room temperature for 20 mins to enable colour development. Three 100 µl aliquots of each standard were transferred to replicate wells of a 96-well microtitre plate and their absorbance (OD$_{595}$) was measured using an iMark Microplate Absorbance Reader microtiter plate reader (Bio-Rad). In Excel (Microsoft), absorbance values were plotted against the known BSA concentrations and linear regression was performed. The U/CBSV Ham1 protein samples were then diluted in storage buffer at the following volume ratios: 1:6, 1:2 and 1:0. Bradford reagent was added to these tubes, and their OD$_{595}$ absorbance values were measured as before. From the linear relationship determined with known BSA concentrations, the U/CBSV Ham1 protein concentrations could then be calculated from their OD$_{595}$ absorbance values.

2.5.3 SDS-PAGE protein analysis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to analyse protein samples. To prepare the protein samples for loading, 10 µl of loading
buffer, 1 μl of protein sample and 9 μl of water were mixed and heated at 95°C for 5 mins. A TruPAGE Precast Gel (Sigma Aldrich) was loaded with 10 μl of each prepared sample and 10 μl of PageRuler Protein Ladder (Thermo Fisher Scientific), as shown in Fig. 2.2. The tank was filled with fresh running buffer. The gel was run at 220V for 40 mins. The gel was stained with 20 ml InstantBlue Protein Stain (Sigma Aldrich) and analysed under UV light using the ChemDoc Bio-Rad System and images were taken using the Quantity One 1D software (Bio-Rad). The following buffers were prepared: **Loading buffer** (4% w/v SDS, 0.25 M Tris-HCl - pH 6.8, 20% v/v Glycerol, 0.004% v/v Bromophenol Blue, 10% v/v β-mercaptoethanol added fresh) and **10X Running buffer – pH 8.5** (25 mM Tris, 192 mM Glycine, 0.1% w/v SDS).

![Figure 2.2: SDS-PAGE band profile of PageRuler Protein Ladder (Thermo Fisher Scientific).](image)

### 2.5.4 Pyrophosphohydrolase assay

Pyrophosphohydrolase assays were performed in Chapter 5. Nucleotide substrates were supplied from manufactures, as outlined in Table 2.3. The PiColourLock kit (Innova Biosciences) was used to quantify phosphate concentration. The kit uses malachite dye that changes from orange to green in the presence of phosphate. Four negative controls were set up for each bioassay reaction, where: 1) substrate and buffer were added with bovine serum albumin protein, 2) substrate and buffer were added with no protein, 3) buffer only and 4) pure water only. To avoid phosphate contamination, ultrapure water (Invitrogen) was used in the preparation of all buffers and solutions.
• The reaction buffer was freshly prepared (50 mM Tris - pH to 8.5, 50 mM MgCl₂, 1 mM dithiothreitol) and heated at 37°C for 5 mins.

• 290 μl aliquots of the buffer were transferred to Eppendorf tubes and 1.3 μg of test protein, 0.1 units of Yeast pyrophosphatase (Thermo Fisher Scientific) and 0.2 mM nucleotide substrate were added to each.

• The reactions were incubated at 37°C for 20 mins and then reactions were stopped by the addition of 75 μl the PiColourLock kit Goldmix with accelerator. After 2 mins colour development was stabilised by the addition of 30 μl of the PiColourLock kit Stabiliser.

• Three 100 μl aliquots from each reaction were added to three replicate wells on a 96 well plate and their OD₆₅₅ absorbance values were measured using an iMark Microplate Absorbance Reader microtiter plate reader (Bio-Rad) with a 655 nm filter.

Table 2.3: Nucleotide substrates used in pyrophosphohydrolase assays.

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>dGTP</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>GTP</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>dCTP</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>CTP</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>dTTP</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>dATP</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>ATP</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>UTP</td>
<td>Invitrogen</td>
</tr>
<tr>
<td>dTTP</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>ITP</td>
<td>Sigma Aldrich</td>
</tr>
<tr>
<td>XTP</td>
<td>Sigma Aldrich</td>
</tr>
</tbody>
</table>

2.5.4.1 Calculation of phosphate concentration in pyrophosphohydrolase enzyme assays

The phosphate concentrations in pyrophosphohydrolase enzyme assays were calculated from their OD₆₅₅ values using a standard curve of known phosphate concentrations and linear regression. Phosphate concentrations of: 5, 15, 25, 35 and 85 μM were set-up using the phosphate standard supplied with the PiColourLock kit. 100 μl aliquots were added to three separate replicate wells on a 96 well plate and their OD₆₅₅ absorbance values were measured using an iMark Microplate Absorbance Reader microtiter plate reader (Bio-Rad) with a 655 nm filter. Linear regression was performed on these OD₆₅₅ values and known concentrations to generate a formula to calculate the phosphate concentration using OD₆₅₅ values.
2.6  Plant methods

2.6.1  Plant maintenance

*Nicotiana* sp. were grown in-house from seed in 10 cm pots containing William Sinclair multipurpose compost. Plants were inoculated at 3-4 weeks old and then grown in growth cabinets at 25°C with a 16 hours light and 8 hours dark cycle. Cassava plants were supplied by NRI (London, UK) and had been virus indexed to confirm the absence of CBSV and UCBSV. The “Colombian” variety was used which is susceptible to CBSD. Plants were grown in 16 cm pots containing William Sinclair multipurpose compost and grown in a greenhouse kept at approximately 28°C. To propagate the plants in-house, 20 cm stem cuttings were taken from green stems with sterile razors and planted in 10 cm pots containing William Sinclair compost and kept under a humidity dome for 10 days.

2.6.2  Viral inoculations of plants

Plants were inoculated with U/CBSV using three methods: 1) agroinfiltration with an *A. tumefaciens* culture containing *in vivo* viral infectious clones (IC), 2) mechanical inoculation with infected plant material and 3) mechanical inoculation with infectious transcripts, generated through *in vitro* transcription of viral ICs.

2.6.2.1  Agroinfiltration of plants

Plants were agroinfiltrated with *in vivo* viral ICs according to the protocol below. The following solutions were freshly prepared and filter-sterilised: **MES buffer** (10 mM morpholino ethane sulfonic acid - pH 5.7); **Acetosyringone** (20 mg acetosyringone (Sigma Aldrich) freshly dissolved in 1 ml dimethyl sulfoxide) and **Agroinfiltration buffer** (10 mM MgCl₂, 10 mM MES, pH 5.7, 150 µm acetosyringone).

- A single *A. tumefaciens* colony was used to inoculate 5 ml LB broth containing 20 µg/ml rifampicin and 50 µg/ml kanamycin and incubated at 28°C and 200 rpm for 24 hours.
- A 500 µl aliquot of this starter culture was used to inoculate 10 ml of LB broth containing 20 µg/ml rifampicin, 50 µg/mL kanamycin and 150 µm acetosyringone. The culture was incubated at 28°C and 200 rpm overnight.
- The culture was centrifuged at 6000 X g for 10 mins at 4°C. The cell pellet was resuspended in 10 mL of MES buffer and centrifuged as before.
- The pellet was resuspended in 10 ml of agroinfiltration buffer until the OD₆₀₀₅ was approximately 1.0.
• The culture was incubated for 5 hours at room temperature to activate Vir genes on the Ti plasmid.
• The agroinfiltration suspension was agroinfiltrated into plants using a sterile needle-less syringe to soak 1 – 3 ml of the culture into the abaxial leaf surface of the 3rd youngest leaf.
• Following agroinfiltration plants were kept in darkness for 48 hours before returning to the 16-hour light and 8-hour dark cycle at 28°C.
• The culture was incubated for 5 hours at room temperature to activate Vir genes on the Ti plasmid.
• The agroinfiltration suspension was agroinfiltrated into plants using a sterile needle-less syringe to soak 1 – 3 mL of the culture into the abaxial leaf surface of the 3rd youngest leaf.
• Following agroinfiltration plants were kept in darkness for 48 hours before returning to the 16-hour light and 8-hour dark cycle at 28°C.

2.6.2.2 Mechanical inoculation with infected material
Test plants were mechanically inoculated with U/CBSV infected plant material according to the protocol from Ogwok et al., (2010).
• The mechanical inoculation (potassium phosphate) buffer - pH 7.0 (0.01 M $K_2HPO_4$ 0.01 M $KH_2PO_4$, 0.2% w/v sodium sulphite, 0.01 M β-mercaptoethanol) was freshly prepared with SDW.
• U/CBSV infected plant material (2 – 4 g) was ground into a fine powder in a sterile pestle and mortar using liquid nitrogen. The powder was resuspended in 15 ml of freshly prepared ice-cold mechanical inoculation buffer and left to thaw.
• The 3rd youngest leaves of test plants were dusted with abrasive carborundum power (particle size 180 grit) and then the infected plant material was gently rubbed onto the adaxial surface. Plants were left for 10 mins to allow soaking before being washed with sterile water to remove the carborundum.

2.6.2.3 In vitro generation of viral transcripts
Attempts were made to generate infectious transcripts through in vitro transcription of the UCBSV Kikombe IC plasmid.
• The UCBSV Kikombe IC plasmid was linearized at the 3’ terminus of the viral genome sequence through digestion with the Fast Digest restriction enzyme Kpn I. The linear plasmid was gel purified using the GeneJET Gel Extraction kit (Thermo Fisher Scientific).
A NanoDrop ND-100 Spectrophotometer (Thermo Fisher Scientific) was used to measure its concentration.

- *In vitro* transcription reactions were set up using the Riboprobe SP6 polymerase transcription system (Promega) according the manufacturer’s instructions. Reactions contained: 10 µl Transcription Optimised 5 X Buffer; 5 µl of 100 mM dithiothreitol; 10 µl Recombinant RNAsin Ribonuclease inhibitor; 10 µl dNTPs (2.5 µl each of 10 mM rATP, rCTP, rUTP and rGTP); 5 µl of 5 mM Ribo m7G Cap analog; 16 µl Linearized plasmid template; 40 units of SP6 RNA polymerase; and nuclease free water up to 50 µl.

- Reactions were incubated at 37°C for 1 hour. After 1 hour, an additional 40 units of SP6 polymerase were added and the reactions were incubated for an additional hour to increase transcript yield.

- To inoculate test plants with these transcripts, 100 µl of the *in vitro* transcription reaction was mixed with 200 µl of freshly prepared ice-cold mechanical inoculation buffer. The sample was kept on ice during the inoculation process. The 3rd youngest leaves of test plants were dusted with abrasive carborundum power (particle size 180 grit), and the infection mixture was gently rubbed on the adaxial surface. Plants were left for 10 mins to allow soaking before being washed with sterile water to remove the carborundum.

### 2.7 Serological methods

Detection of U/CBSV coat proteins in *N. benthamiana* was performed using the Triple Antibody Sandwich Enzyme Linked Immunosorbent Assay (TAS-ELISA) manufactured by DSMZ. The following buffers were freshly prepared with SDW: **Coating buffer - pH 9.6** (15 mM Na₂CO₃, 35 mM NaHCO₃, 3 mM NaN₃; **Phosphate buffered saline (PBS) – pH 7.4** (0.14 M NaCl, 1.5 mM KH₂PO₄, 8.1 mM Na₂HPO₄, 2.7 M KCl, 3 mM NaN₃; **PBS-Tween (PBST)** (PBS + 0.5 ml Tween 20 (Sigma Aldrich) per litre); **Sample extraction buffer - pH 7.4** (PBST + 2% w/v PVP-15 polyvinylpyrrolidone (Sigma Aldrich)); **Blocking solution** (PBST + 2% w/v skimmed milk powder); **Conjugate buffer** (PBST + 2% w/v PVP + 0.2% w/v bovine serum albumin (BSA) (Sigma Aldrich)) and **Substrate buffer - pH 9.5** (100 mM Tris base, 100 mM NaCl, 5 mM MgCl₂).

- The third youngest leaf was sampled in clear plastic bags, weighed, snap frozen in liquid nitrogen and stored at -80°C.

- The samples were diluted in 1:20 (weight/volume) in sample extraction buffer and ground in their bags using a ball-bearing homogeniser.
• 20 μl of the polyclonal antibody (IgG) was diluted 10 ml of coating buffer and 200 μl of this mixture was added to a 96-well micro-titre plate. The plate was incubated at 37°C for 2 – 4 hours in a static incubator. The plate was washed with PBS-Tween three times.

• 200 μl of blocking solution was added to each well to block non-specific binding of proteins to antibodies. The plate was incubated at 37°C for 30 mins.

• The blocking solution was removed, and the wells were tapped dry onto tissue paper.

• 200 μl of each sample was added to duplicate wells. The plate was incubated at 4°C overnight and then washed three times with PBS-Tween.

• 40 μl of the Monoclonal antibody (MAb) was diluted in 10 ml of conjugate buffer and 200 μl of the mixture was added to each well. The plate was incubated at 37°C for 1 hour and then washed three times with PBS-Tween. The MAbs used are specific to U/CBSV coat proteins.

• 10 μl of Rabbit Anti-Mouse Alkaline Phosphatase (RAM-AP) was diluted in 10 ml of conjugate buffer and 100 μl of the mixture was added to each well. The plate was incubated at 37°C for 1 hour and then washed three times with PBS-Tween.

• 10 mg of p-nitrophenyl phosphate was dissolved in 10 ml of substrate buffer and 100 μl of the mixture was added to each well. The plate was incubated at 37°C for 3 hours.

• The optical densities of the wells were measured with an iMark Microplate Absorbance Reader microtiter plate reader (Bio-Rad) using the 405 nm filter. The measurements were collected using the Microplate Management Software (Bio-Rad).

• ELISA results were analysed in Excel (Microsoft). The positive control of CBSD-infected material was included provided by the kit manufacturer. Negative controls were included where no sample was added to the extraction buffer and samples from non-inoculated plants. A cut-off threshold was calculated with the formula: Cut-off threshold = (Mean OD405 values for non-inoculated controls + 3 standard deviations). The presence of U/CBSV in samples with OD405 values below this threshold could not be reliably detected.

2.8 Microscopy

2.8.1 Wide-field fluorescence microscopy

Green fluorescence was visualised in whole leaves using the Leica CL5 Fluorescence microscope with the GFP2 filter (480/40 nm excitation filter and 510 nm barrier filter). Photographs were taken using a digital camera (SPOT Imaging).
2.8.2 Confocal fluorescence microscopy

Approximately 4 mm X 4 mm sections of leaf were placed on a microscope slide in sterile water. A coverslip was place on top and a small drop of immersion oil was applied to the upper surface of the coverslip. A confocal TCS SP5 microscope (Leica) was used to visualise GFP fluorescence in individual cells at either 40X or 100X magnification, using the oil immersion objective. To visualise fluorescence at different depths of the leaf sample, x,y,z stacks were performed. Green fluorescence was measured through a detection channel set at: 490 – 505 nm and red fluorescence through a detection channel set at: 650 – 750nm. Images from the green channel were captured for presentation in Chapter 3.

2.9 Statistical analysis

The LoFreq variant-call was used to detect single nucleotide variants (SNVs) from next-generation sequencing data, with the assistance of Dr. Thomas Batstone in Chapter 4. All other statistical analyses were performed using the SPSS Statistics 24 (IBM) programme. To test whether the number of SNVs in viral genomes from different infection types were significantly different, t-Tests were performed in Chapter 4. To analyse U/CBSV Ham1 enzyme activity with canonical and non-canonical nucleotides, ANOVA and ANCOVA tests were performed in Chapter 5.
Chapter 3: Tractability of two cassava brown streak disease infectious clones

3.1 Introduction

Infectious clones (IC) are highly valuable tools that are used during plant breeding to select for disease resistance and to gain fundamental insights into viral infections. ICs can be manipulated to enable characterisation of viral sequences involved with symptom development, pathogenicity, host-range, host interactions, movement and vector transmission (Nagyová and Subr, 2007). To date challenges with genome sequence instability have frustrated the construction of U/CBSV ICs, which are urgently needed to screen cassava lines for CBSD resistance and to begin characterisation of viral gene functions. In previous work, a UCBSV IC and a CBSV IC were successfully constructed at the University of Bristol; these ICs were used to cause infections in plant hosts. In this Chapter, the tractability of these ICs is tested in terms of their sequence stability during propagation in *E. coli*, the ability to use the CBSV IC to infect cassava through agroinfiltration and to perform manipulations, such as marker gene insertion and chimera construction.

3.1.1 Construction of Potyviridae ICs

ICs are complementary DNA (cDNA) copies of viral genomes, maintained on plasmid vectors, which are used to inoculate plants with a genetically uniform source of viral inoculum (Brewer et al., 2018). The first *Potyviridae* virus IC to be constructed was for *Tobacco vein mottling virus* IC (TVMV) (Domier et al., 1989). Since then ICs have been constructed for a range of agronomically important *Potyviridae* viruses, such as *Potato virus Y* (PVY; Jakab et al., 1997), *Plum pox virus* (PPV; Maiss et al., 1992), *Tobacco etch virus* (TEV; Dolja et al., 1992), *Turnip mosaic virus* (TuMV; Sánchez et al., 1998) and *Zucchini yellow mosaic virus* (ZYMV; Gal-On et al., 1991). IC construction for *Potyviridae* viruses involves extracting RNA from infected plant material and performing reverse transcription with a primer, which binds to the 3’ of the viral genome to generate single stranded cDNA. As *Potyviridae* genomes contain poly(A) tail sequences at the 3’ end of their genomes, an oligo d(T) primer is often used to prime cDNA synthesis. The cDNA is used as template in PCR to amplify overlapping fragments that span the full-length of the viral genome. These PCR fragments are then combined to generate a full-length viral genome sequence, which is cloned into a plasmid expression vector to form the viral IC. Viral ICs can be used to generate infections either through mechanical inoculation with
in vitro transcribed viral transcripts or through transient plant transformation, followed by in vivo transcription.

3.1.2 In vitro and in vivo transcription of ICs

ICs which are transcribed in vitro contain polymerase promoter sequences at the 5’ of the viral genome; commonly used promoters are the bacteriophage SP6 polymerase (Fu and Taylor, 1993) or the T7 polymerase (Hemenway et al., 1990). The IC is linearized through a restriction enzyme digest, which cuts the plasmid at a 3’ site after the viral poly(A) tail to prevent run-off transcription of the plasmid sequence. The linearized IC is then used as template in an in vitro transcription reaction with the respective polymerase to generate infectious viral RNA transcripts (Boccard and Baulcombe, 1992; Hamilton and Baulcombe, 1989). Transcripts are often capped with 7-methylguanosine (m7G) caps to increase the stability of RNA transcripts (Schelle et al., 2001). Transcripts are used to mechanically inoculate plants to generate infections, according to the schematic in Fig. 3.1. The reagents required for this method are relatively expensive, the inoculation efficiency can be low due to degradation of the viral RNA transcripts and relatively large quantities of the IC plasmid are required for the in vitro transcription reaction.

In vivo ICs utilize *Agrobacterium tumefaciens* (updated name *Rhizobium radiobacter*) to transfer DNA (T-DNA) from an IC plasmid into the host nuclei for in vivo transcription by host RNA polymerases, according to the schematic in Fig. 3.1 (Tzfira et al., 2005). The T-DNA region in in vivo ICs is engineered to encode the full-length viral genome sequence. In vivo vectors typically contain the *Cauliflower mosaic virus* 35S (CaMV 35S) promoter sequence at the 5’ of the viral genome sequence, which drives high levels of expression. Meanwhile the nopaline synthase terminator (tNOS) terminates transcription of the viral genome sequence at the 3’. In vivo ICs are transformed into *Agrobacteria*, which are cultured and suspended in media containing acetosyringone to activate *vir* genes on the Ti plasmid required for T-DNA transfer (Gelvin, 2003). A needle-less syringe is used to infiltrate the suspension through the stomata of the abaxial leaf surface (Schöb et al., 1997). This method is less expensive than in vitro transcription and is not affected by RNA degradation. However, utility depends on the agroinfiltration efficiency of the target plant species, as several species, including cassava are challenging to agroinfiltrate. Previously, transient transformation of cassava has only been effectively achieved by using biolistic particle delivery (Franche et al., 1991), which was used to generate infections with *Cassava mosaic geminivirus* (CMG) ICs during resistance screening (Briddon et al., 1998).
Figure 3.1: **Left**: Schematic for the *in vitro* generation of infectious transcripts from an infectious clone (IC) template. *In vitro* ICs contain an RNA polymerase promoter, such as SP6 at the 5’ end before the viral genome sequence. The IC is linearized through restriction digest at the 3’ of the viral genome to prevent run-off transcription of the plasmid. Infectious, full-length viral transcripts are transcribed by bacteriophage RNA polymerases, such as SP6. Infectious transcripts are often capped with a 7-methylguanosine (m7G) cap to improve RNA stability. Transcripts are mechanically inoculated into plant cells and are then directly translated in the plant cell cytoplasm.

**Right**: Schematic for transformation of agroinfiltrated plant cells to enable *in vivo* transcription of viral ICs. ICs are agroinfiltrated into plant cells, the T-DNA containing the viral genome inserts into the host DNA and host RNA polymerases transcribe the viral genome. The CaMV 35S promoter drives expression of the viral genome and the nTOS terminator terminates transcription. Intron sequences are removed from the viral transcripts through splicing. Viral transcripts are exported from the nucleus into the cytoplasm for translation and replication.
3.1.3 Genome sequence instability during Potyviridae IC propagation in E. coli

To obtain high quantities of IC plasmids, they are usually propagated in E. coli. However, the propagation of Potyviridae ICs in E. coli is often associated with sequence instability (Desbiez et al., 2012) due to the expression of cytotoxic proteins from internal cryptic prokaryotic promoters in the viral genome (Jakab et al., 1997). This leads to sequence deletions and insertions that prevent the production of toxic proteins in E. coli but significantly change the viral genome sequence and so render the IC unusable. Previous studies have reported that the P3 and CI regions of the PVY genome are unstable (Chikh et al., 2011; Jakab et al., 1997). Many techniques have been deployed to overcome sequence instability challenges. The simplest technique is to test IC sequence stability in range of E. coli strains, as certain strains replicate ICs with higher stability (Al-Allaf et al., 2013; Boyer and Haenni, 1994; Jakab et al., 1997). If this is unsuccessful, more sophisticated cloning methods are required, including cloning the viral genome into two separate plasmids which are ligated and transcribed in vitro (Jakab et al., 1997). Alternatively site-directed mutagenesis can be performed on the cryptic promoter sequences (Chikh Ali et al., 2011), however this alters the viral genome sequence and so may affect infectivity. A fourth approach involves the insertion of plant introns into unstable regions of the viral genome sequence, which prevents the expression of toxic proteins in E. coli and thereby improves sequence stability. The IC plasmid is then transformed into plant cells and intron sequences are spliced out during in vivo expression of the viral genome sequence within the plant cell nucleus (Johansen, 1996). This technique has been used stabilize several ICs, including Pea seed-borne mosaic virus IC, (PSbMV; Johansen, 1996), PPV IC (López-Moya and García, 2000) and Papaya leaf distortion mosaic virus (PLDMV; Tuo et al., 2015).

3.1.4 Applications of plant viral ICs

3.1.4.1 Use of plant viral ICs in resistance breeding

ICs have many applications, firstly they are used during plant breeding to screen for and understand viral resistance. In the absence of ICs, to screen crop lines for viral resistance breeders use field inoculum, mechanical inoculation with infected plant material or grafting to infected plants (Lecoq et al., 2004). However these methods often have a low inoculation efficiency and result in researchers using locally adapted viral strains, which makes it difficult to compare resistance tested at different times and locations (Maruthi et al., 2005). Viral sequences can also change during in-planta propagation, making it difficult to reliably compare infections over time (Brewer et al., 2018). ICs are therefore highly valuable tools as they provide a stable, uniform viral genome sequence which can be used to repeatedly and efficiently infect
plants and enable accurate comparisons of infection responses (Brewer et al., 2018). ICs for CMGs were first developed in the 1980s (Stanley, 1983) have been repeatedly used for over 20 years to screen cassava lines for CMD resistance (Beyene et al., 2016). In addition to screening for resistance, ICs are also used to understand host resistance mechanisms. For example a Tomato chlorotic mottle virus (ToCMoV) IC was recently used to characterise differential proteome responses in resistant and susceptible tomato lines (Carmo et al., 2017).

3.1.4.2 Use of viral ICs to visualise plant infections

To visualise viral replication and movement during infections, ICs can be manipulated through the insertion of marker gene sequences. These ICs are used to infect plants and marker gene expression is monitored to track viral replication and movement in-planta. For example, the insertion of the marker gene beta-glucuronidase (GUS) into the TEV IC was used to track TEV spread in N. tabacum from inoculated epidermal cells to neighbouring epidermal and mesophyll cells, through phloem associated cells to the vascular tissue in systemic leaves (Dolja et al., 1992). These ICs can also be further manipulated through the deletion or mutation of viral genes to characterise their roles in viral replication and/or movement. This has demonstrated that the TMV coat protein (CP) is essential for viral movement during systemic infections (Takamatsu et al., 1987), the TEV CP is involved with cell-to-cell movement (Dolja et al., 1994) and that the HC-Pro protein is involved with long-distance movement (Cronin et al., 1995). More recently, Olspert et al., (2015) used a TuMV_GFP IC to identify the transcriptional slippage mechanism which enables expression of the P3N-PIPO protein. Finally, ICs for different viral species or strains can be tagged with different marker genes to visualise viral interactions during mixed infections (Dietrich and Maiss, 2003).

3.1.4.3 The use of plant viral ICs to characterise sequence determinants

As described above, viral ICs can be manipulated through mutations, deletions and/or substitutions to characterise gene functions and host responses associated with viral genome sequences. For instance, Kim and Palukaitis, (1997) identified through IC site directed mutagenesis that the cowpea hypersensitive response (HR) to Cucumber mosaic virus (CMV) is mediated by the CMV 2a polymerase. Meanwhile, Lee et al., (2005) used mutant Potato leafroll virus (PLRV) ICs to identify CP amino acids that are critical for virus assembly, stability, systemic infection of plants and aphid transmission. Chimeric ICs are also used to identify sequences responsible for differential symptom development and host responses caused by different viral
strains or species (Nagyová and Subr, 2007). For instance, Martin and Rybicki, (2002) constructed chimeric ICs with reciprocally exchanged genomic regions between a highly pathogenic *Maize streak virus* (MSV) isolate and three less pathogenic isolates to identify pathogenicity sequence determinants. Similarly, chimeric ICs have been used to identify pathogenicity determinants for *Pepino mosaic virus* (PepMV) (Duff-Farrier et al., 2015) and PVY (Faurez et al., 2012).

### 3.1.5 Potential applications of U/CBSV ICs

Once constructed and validated for utility, U/CBSV ICs would have many valuable applications in the fight against CBSD. Firstly, U/CBSV ICs are urgently needed to screen cassava lines for CBSD resistance. To date, there is no cassava cultivar with a high level of CBSD resistance available to farmers (Abaca et al., 2013). U/CBSV ICs would accelerate breeding efforts by providing breeders at different locations with the same stable, uniform inoculum that could standardise infections and improve comparisons between infections responses in different cassava lines. U/CBSV ICs could also be used to begin characterisation of viral sequences associated with symptom development, including root necrosis, pathogenicity, replication, movement, host-interactions and vector transmission, which are currently poorly understood. Chimeric ICs consisting of CBSV-UCBSV sequence replacements could be used to characterise the sequences associated with differences in symptom development and viral titers during CBSV and UCBSV infections (Mohammed et al., 2012; Winter et al., 2010). Finally, marker genes could be inserted into U/CBSV ICs to determine the direction and locations of systemic viral movement and replication and whether CBSV and UCBSV co-infect cells during mixed infections. The utility of U/CBSV ICs would depend on their inoculation efficiency, sequence stability during long-term storage, plasmid propagation and ease of performing manipulations. The use of U/CBSV ICs would also require access to facilities with sufficient biosafety containment and availability of required reagents (Brewer et al., 2018).

### 3.1.6 Previous work at the UoB to construct U/CBSV ICs

Two U/CBSV ICs were previously constructed at the University of Bristol; the UCBSV_Kikombe IC NCBI: KX753357.1) was constructed by Nanyiti (PhD thesis, 2016) and the CBSV_Tanza IC (NCBI: MG570022.1) was constructed by Duff-Farrier and Mbanzibwa et al., (unpublished). The construction of these ICs is described below.
3.1.6.1 Construction and infectivity of the UCBSV

The UCBSV_Kikombe IC was designed for in vitro transcription; it was constructed through RT-PCR amplification of eight overlapping genome fragments from UCBSV infected cassava material originating from Mikocheni Rd, Tanzania. These overlapping PCR fragments were recombined through homologous yeast recombination into a yeast adapted pYES2.1 vector with primers to insert the SP6 polymerase promoter at the 5’ of the UCBSV genome and Kpn I restriction site at the 3’, according to the protocol outlined in Duff-Farrier et al., (2015). The UCBSV Kikombe genome showed sufficient sequence stability to enable direct assembly of the full-length sequence in the E. coli strain α-select. The map for the UCBSV_Kikombe IC is provided in Fig. 3.2. The UCBSV_Kikombe IC was used to generate in vitro transcripts, which were mechanically inoculated onto N. benthamiana and cassava. N. benthamiana inoculated with UCBSV_Kikombe IC transcripts developed systemic mild mosaics, similar to infections with wild-type UCBSV Kikombe inoculum (Fig. 3.3). Cassava plants inoculated with UCBSV_Kikombe IC transcripts occasionally developed mild foliar chlorosis in upper systemic leaves at eight-weeks post-infection, similar to wild-type infections. Infections in both N. benthamiana and cassava were confirmed by RT-PCR amplicon sequencing. Once constructed, the sequence stability of the UCBSV_Kikombe IC during plasmid propagation in E. coli needed to be tested.
3.1.6.2 Construction and infectivity of the CBSV Tanza IC

Sequence instability was encountered during attempts to clone the full-length CBSV Tanza genome in *E. coli*. Six overlapping RT-PCR spanning the CBSV Tanza genome were amplified from CBSV infected cassava material originating from Mikocheni Rd, Tanzania. Sequence instability was detected in the P3, CI and NIb regions of the CBSV genome, which prevented the construction of an IC containing the full-length genome sequence. To overcome this, three plant intron sequences were inserted into the P3, CI and NIb regions, which enabled sufficient sequence stability for the full-length genome sequence to be cloned in yeast and propagated in the *E. coli* strain OverExpress C43. The full-length CBSV Tanza genome was cloned into a yeast adapted pCAMBIA0380 vector and included the CaMV 35S promoter and tNOS terminator to drive *in vivo* transcription. The map for the CBSV_Tanza IC is provided in Fig. 3.4.

To test the infectivity of the CBSV_Tanza IC, it was used to agroinfiltrate *N. benthamiana* and *N. clevelandii*. Infections of both *Nicotiana* sp. developed necrosis on inoculated leaves at 7 dpi and severe systemic chlorosis, wilting and necrosis at 14 dpi (Fig. 3.5). RT-PCR amplicon sequencing confirmed the presence of CBSV Tanza sequence in both infections and that introns had been removed from the viral transcripts. Further work however was then needed to test CBSV_Tanza IC sequence stability during plasmid propagation in *E. coli*, tractability for cloning manipulations and ability to inoculate cassava through agroinfiltration.
Figure 3.4: Plasmid map of the CBSV_Tanza IC consisting of the full-length CBSV Tanza genome inserted into the pCAMBIA0380 vector. Key features include: the kanamycin resistance gene (KanR), *E. coli* origin of replication (ori), yeast 2-micron ori, yeast selectable marker URA gene and the *in vivo* expression cassette with the *Cauliflower mosaic virus* (CaMV) 35S promoter and tNOS terminator. Introns 1–3 were inserted into P3, CI and NIb regions to improve sequence stability during plasmid propagation in *E. coli*.

Figure 3.5: Symptom development during *N. benthamiana* and *N. clevelandii* infections with CBSV_Tanza IC. *N. benthamiana* infections develop strong systemic chlorosis, wilting and necrosis 14 dpi (A), compared with symptomless non-inoculated plants (B). *N. clevelandii* infections develop chlorosis and necrosis on agroinfiltrated leaves at 7 dpi (C) and necrosis of upper systemic leaves at 14 dpi (D), whereas non-inoculated plants remain symptomless (data not shown).
3.1.7 Hypothesis
It was predicted that both the UCBSV Kikombe and CBSV Tanza ICs would exhibit sequence instability during plasmid propagation in *E. coli* and that the presence of introns in the CBSV_Tanza IC would result in higher sequence stability, compared with the UCBSV_Kikombe IC plasmid.

3.1.8 Aims

- To develop a protocol to agroinfiltrate cassava with the CBSV_Tanza IC plasmid.

- To manipulate both the UCBSV_Kikombe and CBSV_Tanza ICs through insertion of a marker gene sequence and to use these ICs to track viral replication and movement *in planta*.

- To construct chimera ICs consisting of the CBSV Tanza genome with UCBSV Kikombe gene replacements and reciprocal ICs consisting of the UCBSV Kikombe genome with CBSV Tanza gene replacements. These ICs could then be used to infect plants to initiate characterisation of viral sequences associated with differential UCBSV and CBSV symptom development.
3.2 Results

3.2.1 Sequence stability of the UCSV Kikombe IC

3.2.1.1 UCSV propagation in the E. coli strain α-select

To propagate the UCSV_Kikombe IC, the original IC plasmid was transformed into the E. coli strain α-select (for genotype see Appendix 2.1). Transformant colonies were cultured, and plasmids were extracted from 5 ml of bacterial culture using GeneJET plasmid miniprep kit (Thermo Fisher Scientific). To determine whether sequence rearrangements had occurred, the extracted plasmids were digested with XhoI, however due to low plasmid concentrations it was not possible to determine the band pattern. To test sequence integrity, PCR was performed targeting the eight UCSV Kikombe genome sections using primers in Table 3.1. This found that the correct sized genome sections were present. However, the total amount of plasmid extracted was low, typically around 120 ng, compared with around 300 µg in standard extractions. The minimum amount of plasmid required for template in in vitro transcription reactions using the Riboprobe in vitro transcription kit (Promega) is stated to be 1 µg. Despite the low amounts of the extracted UCSV_Kikombe IC plasmids, attempts were made to use them for in vitro transcription. However, gel electrophoresis failed to detect the presence of RNA in reactions where the linearized UCSV_Kikombe IC was used as template, compared with high levels of RNA detected in positive control reactions, using template provided with the kit. Furthermore, mechanical inoculation of N. benthamiana with these transcription reactions failed to develop infections. This indicated that due to the low plasmid DNA concentration, in vitro transcription had failed to generate sufficient transcripts to enable infection.

Table 3.1: Primers used to amplify UCSV Kikombe genome sections by PCR.

<table>
<thead>
<tr>
<th>Sec./region</th>
<th>Primer name</th>
<th>5’–3’ sequence</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1 =</td>
<td>UCSV_Kik_Sec1_Fw</td>
<td>CAAATTAGATCTATTATTAGTGACACTATGGAAAAATAAACA</td>
<td>2684</td>
</tr>
<tr>
<td>5’ UTR - Cl</td>
<td>UCSV_Kik_Sec1_Rv</td>
<td>CATATGTCATGATTGTAATTG</td>
<td></td>
</tr>
<tr>
<td>Section 2 =</td>
<td>UCSV_Kik_Sec2_Fw</td>
<td>ATTAGGACCACTACGAGTGAGCTATGG</td>
<td>1043</td>
</tr>
<tr>
<td>Cl</td>
<td>UCSV_Kik_Sec2_Rv</td>
<td>ACTGACCCCTATATCTCATCATCAC</td>
<td></td>
</tr>
<tr>
<td>Section 3 =</td>
<td>UCSV_Kik_Sec3_Fw</td>
<td>GATTITGGAACGCTGACACATGTTCAA</td>
<td>867</td>
</tr>
<tr>
<td>Cl</td>
<td>UCSV_Kik_Sec3_Rv</td>
<td>GAGTTACTCTTGAATCAATTAAATGAG</td>
<td></td>
</tr>
<tr>
<td>Section 4 =</td>
<td>UCSV_Kik_Sec4_Fw</td>
<td>AAAGAATACATTAGAGCAA</td>
<td>1227</td>
</tr>
<tr>
<td>Cl - Nla</td>
<td>UCSV_Kik_Sec4_Rv</td>
<td>GAGTTCAAGTTTCAGAAGCGAAACAGGAAACTAATC</td>
<td></td>
</tr>
<tr>
<td>Section 5 =</td>
<td>UCSV_Kik_Sec5_Fw</td>
<td>GCTCCATGGGATTTCAGAGAAGA</td>
<td>1040</td>
</tr>
<tr>
<td>VPG – Nlb</td>
<td>UCSV_Kik_Sec5_Rv</td>
<td>ACTAACCAGTTGAACACTGCTC</td>
<td></td>
</tr>
<tr>
<td>Section 6 =</td>
<td>UCSV_Kik_Sec6_Fw</td>
<td>GGATTTGAGCTGACATGAGATG</td>
<td>1491</td>
</tr>
<tr>
<td>Nla – Nlb</td>
<td>UCSV_Kik_Sec6_Rv</td>
<td>TCCAGAATGGAGACTACACAA</td>
<td></td>
</tr>
<tr>
<td>Section 7 =</td>
<td>UCSV_Kik_Sec7_Fw</td>
<td>GGGCGTCCGAAAAAGGGTTG</td>
<td>1395</td>
</tr>
<tr>
<td>Nlb – CP</td>
<td>UCSV_Kik_Sec7_Rv</td>
<td>GCTTTGGGAACCTCAATTGCACCCAAATCCTCA</td>
<td></td>
</tr>
<tr>
<td>Section 8 =</td>
<td>UCSV_Kik_Sec8_Fw</td>
<td>GGGCAATTGAGGTGCAAAATCC</td>
<td>805</td>
</tr>
<tr>
<td>CP – 3’ UTR</td>
<td>UCSV_Kik_Sec8_Rv</td>
<td>TTTTTTTTTTTTTTTTTTTTTTTTTTGGACTTTTTGTAAATGTATTTTCT</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.6: PCR targeting eight UCBSV Kikombe genome sections (Sec.) to verify the sequence integrity of two UCBSV_Kikombe IC plasmids after propagation in the E. coli strain α-select and extraction using the GeneJET Midiprep kit (Thermo Fisher Scientific). PCR fragments that were approximately 1 Kb larger than expected sizes amplified from plasmid one in genome section four and from plasmid two in genome section five, indicating sequence instability. Correct sized fragments amplified when the original UCBSV_Kikombe IC plasmid was used as template (lanes +), as described in Table 3.1 and no amplification occurred in negative controls (lanes -), when no template was added.

To obtain more concentrated UCBSV_Kikombe DNA, plasmids were extracted from a larger bacterial culture volume of 50 ml using the GeneJET plasmid midiprep kit (Thermo Fisher Scientific). Plasmids were extracted from two cultures, each yielded around 150 - 160 ng of DNA, which again was low compared with yields in standard plasmid extractions. PCR was performed to determine their sequence integrity, as shown in Fig. 3.6. PCR targeting section four, corresponding to CI – start of the Niα is expected to amplify a fragment of 1.2 Kb. However, a fragment of approximately 2.5 Kb amplified from plasmid 1. Similarly, PCR targeting section five, corresponding to genome region VPg – Niβ is expected to amplify a fragment of 1 Kb, whereas a fragment of approximately 2.5 Kb amplified from plasmid 2. This indicated that sequence insertions had occurred and that CI – Niα and VPg – Niβ genome regions may be unstable.

In a final attempt to extract sufficient UCBSV_Kikombe IC plasmid, a large-scale maxiprep plasmid extraction was performed according to Green and Sambrook, (2001). The protocol involved the growth of 500 ml bacterial cultures to an OD of 0.4. After this, chloramphenicol was added, which inhibits protein synthesis required for bacterial cell division and should result in a higher number of plasmids per bacterial genome (Begbie et al., 2005). PCR analysis targeting the eight genome sections revealed that a larger (2.5 Kb) than expected (1 Kb) PCR
fragment amplified for genome section five, which corresponds to the VPg – Nib region of the UCBSV Kikombe genome. This further demonstrated that the VPg – Nib region of the UCBSV Kikombe genome is associated with sequence instability during propagation in the E. coli strain α-select.

### 3.2.1.2 UCBSV Kikombe IC propagation in the E. coli strain DH5α

To test whether the UCBSV_Kikombe IC could be stably propagated in a different E. coli strain, the original plasmid was transformed into the strain DH5α (for genotype see Appendix 2.1). A transformant colony was cultured and plasmids were extracted from 5 ml cultures using the GeneJET plasmid miniprep kit (Thermo Fisher Scientific). To confirm the sequence integrity of the plasmid, eight sections of the UCBSV Kikombe genome were targeted by PCR with primers provided in Table 3.1. In the PCR targeting genome section five (VPg – Nib), a fragment of approximately 2.5 Kb amplified from the propagated plasmid template which is larger than the 1 Kb expected size and indicated that a sequence insertion had occurred. The region was Sanger sequenced in both the forward and reverse directions and sequence reads were aligned to the UCBSV_Kikombe IC map, which demonstrated that an insertion had occurred in the middle of the Nia region, as shown in Fig. 3.7. The extra sequence was used to BLAST the NCBI database and was found to have a high sequence identity (99%) with E. coli chromosomal genome DNA (NCBI: CP026085.1). This indicates that E. coli may have responded to sequence toxicity in the Nia region of the UCBSV_Kikombe IC through the insertion of this DNA sequence.

![Figure 3.7: Sequence alignment between the UCBSV_Kikombe IC map and the sequence read generated from a UCBSV Kikombe IC plasmid which had been propagated in the E. coli strain DH5α. Extra sequence was present in the propagated plasmid from position 10569 bp, which corresponds to the middle of the Nib region. The extra sequence was found to originate from E. coli and demonstrates sequence instability in this genome region.](image-url)
3.2.1.3 UCBSV Kikombe IC propagation in the *E. coli* strain ccdB

In a final attempt to extract sufficient quantities of the UCBSV_Kikombe IC, the original IC plasmid was transformed into the *E. coli* strain One Shot ccdB survival cells (for genotype see Appendix 2.1). Plasmids were extracted from four separate 5 ml cultures of using the miniprep kit (Thermo Fisher Scientific). To confirm sequence integrity, PCR was performed targeting the eight UCBSV Kikombe genome sections. This found that the CI – start of Nla region was missing in three of the plasmids and one plasmid contained an insertion in this area. Due to the high levels of sequence instability encountered during propagation of the UCBSV_Kikombe IC plasmid in all three of the tested *E. coli* strains, it was evident that sequence stabilisation was required, which was beyond the time constraints of this work.

3.2.2 Sequence stability of the CBSV Tanza IC

3.2.2.1 CBSV Tanza IC propagation in the *E. coli* strain C43

To test whether the CBSV_Tanza IC could be stably propagated in *E. coli*. The original IC plasmid was transformed into the *E. coli* strain C43 (for genotype see Appendix 2.1). CBSV_Tanza IC stability was tested in this strain because it was used to propagate the original the CBSV_Tanza IC during construction (Duff-Farrier and Mbanzibwa et al., unpublished) and is reported to tolerate expression of toxic proteins (Dumont-Seignovert et al., 2004). Six transformant colonies were cultured and plasmids were extracted using the miniprep kit (Thermo Fisher Scientific). To verify the sequence integrity of these plasmids, six sections of the CBSV Tanza genome were targeted by PCR with primers in Table 3.2. This demonstrated that the Nlb region of two of the plasmids contained a deletion, and four of the plasmids contained insertions (Fig. 3.8). These plasmids were sequenced in both forward and reverse directions; the results are shown in Table 3.3. Plasmids 2, 3, 4 and 6 were shown to contain extra sequence inserted into the same position in Nlb, which demonstrates that this region of CBSV Tanza genome is unstable during propagation in the *E. coli* strain C43. The inserted sequences were used to BLAST NCBI database and were found to have high levels of sequence identities with *E. coli* chromosomal and plasmid DNA (Table 3.3). The inserted sequences do not share high levels of sequence similarity with each other or with the inserted sequence identified in UCBSV_Kikombe IC.
Table 3.2: Primers used to amplify CBSV Tanza genome sections by PCR.

<table>
<thead>
<tr>
<th>Section/genome</th>
<th>Primer name</th>
<th>5′–3′ sequence</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1 = 5′UTR - P1</td>
<td>CBSV_Tanza_Sec1_Fw</td>
<td>ATGACAAACGATTCAGCTATTCAAGACTGGTCACTG</td>
<td>1085</td>
</tr>
<tr>
<td></td>
<td>CBSV_Tanza_Sec1_Rv</td>
<td>GTACATTTCAATCCTGTGGTAGTTCACTCCCAAGACTTCCTAA</td>
<td></td>
</tr>
<tr>
<td>Section 2 = P3 – 6K1</td>
<td>CBSV_Tanza_Sec2_Fw</td>
<td>TCATTTCACCTGAGTTGAGCGGCTGTC</td>
<td>665</td>
</tr>
<tr>
<td></td>
<td>CBSV_Tanza_Sec2_Rv</td>
<td>GGAGCTTCTCGCCCTGATTATTTGAGAGGCAAAGCGGTGATTGAGATGAAATGACCAATGGG</td>
<td></td>
</tr>
<tr>
<td>Section 3 = CI – VPg</td>
<td>CBSV_Tanza_Sec3_Fw</td>
<td>TGGACAGGACGCGGTTGATG</td>
<td>2296</td>
</tr>
<tr>
<td></td>
<td>CBSV_Tanza_Sec3_Rv</td>
<td>CCAGCCCTTCTCCAGAAGGGGG</td>
<td></td>
</tr>
<tr>
<td>Section 4 = VPg – Nla</td>
<td>CBSV_Tanza_Sec4_Fw</td>
<td>TCCAAGGAGGAGACGGTTG</td>
<td>899</td>
</tr>
<tr>
<td></td>
<td>CBSV_Tanza_Sec4_Rv</td>
<td>TTTCATCAATCTCCCTACGGGTTTCTTGCTGCTTGAATTTGCTATTAGC</td>
<td></td>
</tr>
<tr>
<td>Section 5 = Nlb</td>
<td>CBSV_Tanza_Sec5_Fw</td>
<td>GCCCAACCGGGAGAGGTGAAG</td>
<td>1708</td>
</tr>
<tr>
<td></td>
<td>CBSV_Tanza_Sec5_Rv</td>
<td>GCCATCACGCGAGGCAATG</td>
<td></td>
</tr>
<tr>
<td>Section 6 = Nlb – 3′UTR</td>
<td>CBSV_Tanza_Sec6_Fw</td>
<td>GCTCTGTTGTGAGGTTGATG</td>
<td>2640</td>
</tr>
<tr>
<td></td>
<td>CBSV_Tanza_Sec6_Rv</td>
<td>GAATTCCCGATCTAGAACA</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.8: PCR on six CBSV_Tanza ICs propagated in the E. coli strain C43, demonstrating sequence instability in the Nlb region of the CBSV Tanza genome. The expected size of the PCR fragment targeting the Nlb region (section 5) is 1.7 Kb, as shown in the positive control (lane +), when the original CBSV_Tanza IC was used as template. Two of the propagated plasmids contained sequence deletions (lanes 1 and 5), whereas four of plasmids contained sequence insertions (lanes 2, 3, 4 and 6).

Table 3.3: Sequence instability detected in CBSV_Tanza IC plasmids propagated in the E. coli strain C43.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Insertion/Deletion</th>
<th>Position in CBSV Nlb (bp)</th>
<th>Insertion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Deletion</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Insertion</td>
<td>1 – 529</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Insertion</td>
<td>339</td>
<td>99% sequence identity with E. coli APOBEC chromosome (CP006834.2)</td>
</tr>
<tr>
<td>4</td>
<td>Insertion</td>
<td>339</td>
<td>99% sequence identity with E. coli strain 6/14/6b plasmid plncF-MU4 (MF174860.1)</td>
</tr>
<tr>
<td>5</td>
<td>Deletion</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Insertion</td>
<td>339</td>
<td>99% sequence identity with E. coli strain cq9 plasmid unnamed1 (CP031547.1)</td>
</tr>
</tbody>
</table>
3.2.2.2 CBSV Tanza IC propagation in the *E. coli* strain *ccdB*

Sequence stability of the CBSV Tanza IC was then tested in the *E. coli* strain *ccdB* (for genotype see Appendix 2.1). The original IC plasmid was transformed into electrocompetent *ccdB* cells. Three transformant colonies were cultured and plasmids were extracted from three cultures using the GeneJET plasmid miniprep kit (Thermo Fisher Scientific). To confirm the sequence integrity, PCR was performed targeting the six CBSV Tanza genome sections. The correct sized fragments amplified for all sections (Fig. 3.9), which indicated sequence integrity and that the CBSV_Tanza IC could be stably propagated in this *E. coli* strain.

![PCR gel image showing amplification of CBSV Tanza IC genome sections.](image)

Figure 3.9: PCR targeting six CBSV Tanza genome sections (Sec.) to verify the sequence integrity of three CBSV_Tanza IC plasmids, after propagation in the *E. coli* strain *ccdB*. The correct sized fragments for all six sections amplified from three CBSV_Tanza IC plasmids to indicate sequence integrity. The correct sized fragments amplified from the original CBSV_Tanza IC (lane +) (Table 3.2) and no amplification occurred in negative controls (lanes -), which contained no template.
3.2.3 Agroinfiltration of cassava with the CBSV Tanza IC

Investigations were then performed to determine whether the CBSV Tanza IC could be used to inoculate cassava through agroinfiltration. Previous reports have found that cassava leaves are recalcitrant to agroinfiltration, due to the hydrophobic surface, which prevents agroinfiltration suspension buffer from soaking into the leaf (Díaz et al., 2014). To overcome this, several adaptations to the standard agroinfiltration protocol were tested, including using a sterile razor to slash cassava stems and then coating with filter paper soaked into agroinfiltration suspension or by injecting stems directly with agroinfiltration suspension. Both these approaches failed to produce infections. In a final approach, the third fully expanded cassava leaves were lightly dusted with carborundum powder and the agroinfiltration suspension was gently stroked onto the leaves. This was intended to enable the A. tumefaciens to enter wounded cassava leaf cells, as described in Andrieu et al., (2012). In addition, the surfactant Pluronic F-68 (Life Technologies) was added to the A. tumefaciens suspension at a concentration of 0.01% v/v to improve soaking of the A. tumefaciens suspension into the leaf, as described in Yang et al., (2006). After 10 mins, carborundum was washed from the leaves with sterile water and plants were returned to the greenhouse conditions at 28°C. This technique was tested in two repeat experiments, each with four replicate plants. Plants agroinfiltrated with the CBSV Tanza IC were regularly monitored for symptom development and sampled for subsequent viral detection. At 42 dpi, plants were tested for CBSV Tanza infections by RT-PCR; this indicated that one out of three plants in the first experiment and two out of four plants in the second experiment were infected (Fig. 3.10). RT-PCR amplicon sequencing confirmed the presence of CBSV Tanza sequence in cassava (Appendix 3.1).

**Figure 3.10:** RT-PCR detection of CBSV in cassava plants agroinfiltrated with the CBSV_Tanza IC. The CBSV_Tanza_Sec4_Fw/Rv (Table 3.2) primers weakly amplified the 899 bp fragment targeting the CBSV Tanza Nia region from one (lane 2) of the four cassava plants (lanes 1 - 4) agroinfiltrated in the first experiment and two (lanes 6 – 7) of the four cassava plants agroinfiltrated in the second experiments (lanes 5 – 8) at 42 dpi. Amplification occurred in positive controls from N. benthamiana agroinfiltrated with CBSV_Tanza (+1) and when the CBSV_Tanza IC plasmid was used as template (+2). No amplification occurred in the negative control when no cDNA was added to the PCR reaction (-).
Agroinfiltrated plants developed mild chlorosis on a small number of upper systemic leaves at 42 dpi (Fig. 3.11). To further confirm infections, systemic leaf material was used to mechanically inoculate four *N. benthamiana* plants. Out of these four plants, one developed mild systemic leaf curling without necrosis at 14 dpi. This symptom is not characteristic of CBSV infection of *N. benthamiana* and may indicate low viral titre or altered viral genome sequence in the back-inoculated CBSV inoculum. RT-PCR (Fig. 3.12) and amplicon sequencing detected the presence of CBSV sequence in the symptomatic plant (Appendix 3.2). This demonstrated that the CBSV_Tanza IC could be used to infect cassava, although the efficiency appears to be relatively low at 25 – 50%.

Figure 3.11: An upper systemic cassava leaf showing chlorosis typical for CBSV infection after agroinfiltration of the cassava plant with CBSV_Tanza IC at 42 dpi, compared to a symptomless leaf of a non-inoculated cassava plant (B).

Figure 3.12: RT-PCR detection of CBSV Tanza in one (lane 3) of four *N. benthamiana* plants (lanes 1 - 4) after mechanical inoculation with CBSV_Tanza IC infected cassava leaf material at 14 dpi. The CBSV_Tanza_Sec4_Fw/Rv (Table 3.2) primers were used to amplify an 899 bp fragment, targeting the CBSV Tanza Nla region. Amplification occurred in the positive control with cDNA generated from *N. benthamiana* agroinfiltrated with CBSV_Tanza (lane +). No amplification occurred in the negative control when no cDNA was added to the PCR reaction (lane -1) and when no RNA was added to the cDNA synthesis reaction (lane -2).
3.2.4 Insertion of a marker gene into the CBSV Tanza IC

Having established that the CBSV Tanza IC could be propagated in the *E. coli* strain *ccd*B with relatively high levels of sequence stability, the next aim was to test whether the CBSV Tanza IC could be manipulated through the insertion of a marker gene sequence. Previous studies have inserted marker gene sequences in ICs at the 5’ of *Potyviridae* genomes, between the P1 and HC-Pro sequences (Carrington et al., 1998) or at the 3’, between the Nlb and CP sequences (Dietrich, 2003). Beauchemin et al., (2005) found no detectable differences in marker gene expression between these two insertion sites in the TuMV IC. The genome position of between the Ham1 – CP was chosen for marker gene insertion so that marker protein translation would occur after most of the polyprotein sequence had been translated. The marker gene GFP was chosen as it has been widely used to visualise *in planta* viral movement and replication (Baulcombe et al., 1995; Beauchemin et al., 2005; Dietrich, 2003) and its visualisation does not involve the destruction of live plant material, as required for GUS staining (Bedoya et al., 2012). The GFP sequence used corresponds to the mGFP5 gene (NCBI: U87973.1), which is constitutively expressed at high-levels in the transgenic 16c line of *N. benthamiana* and localizes to the nucleus and cytoplasm (Ruiz et al., 1998).

3.2.4.1 Cloning design for construction of CBSV_Tanza_GFP1

The cloning plan was to construct the CBSV_Tanza_GFP1 IC through homologous yeast recombination. Three PCR fragments (F1, F2 and F3) were designed with overlapping sequences to enable recombination into *Bam* H1 digested CBSV_Tanza IC in yeast, according to the schematic in Fig. 3.13.

3.2.4.2 PCR design for the construction of CBSV_Tanza_GFP1

The three PCR fragments used to construct the CBSV_Tanza_GFP1 IC were amplified with primers in Table 3.4. Fragments 1 and 3 were amplified by PCR using the CBSV_Tanza IC as template, whereas fragment 2 was amplified using the pCAMBIA2300_GFP as template. The PCR fragments were designed as follows:

- **F1**: The 5’ end contains 30 bp of homologous sequence from before the first *Bam* H1 restriction site at position 7284 bp in the CBSV Tanza IC plasmid, it then encodes CBSV Tanza sequence to the end of Ham1, and the 3’ end contains 30 bp of homologous sequence with F2.
- **F2**: Encodes GFP sequence with I-D-V-Q/-A proteolytic cleavage sequences at the 5’ and 3’ ends.
• F3: The 5’ end contains 30 bp of homologous sequence with F2, the proteolytic cleavage sequence: I-D-V-Q-A, CBSV Tanza CP - 3’UTR sequence. At the 3’ end there is 30 bp of homologous sequence with CBSV Tanza IC from after the second Bam H1 restriction site at position 8961 bp.

Figure 3.13: Schematic for the construction of the CBSV_Tanza_GFP1 IC through homologous yeast recombination. The CBSV_Tanza IC was digested with Bam H1, which cuts at positions 7284 and 8961 bp to generate a 18.6 Kb fragment and a 1.7 Kb fragment (stage 1). The 18.6 Kb fragment and PCR fragments: F1 – 3 were then transformed then recombined in yeast (stage 2) and yeast plasmids were rescued into E. coli to form CBSV_Tanza_GFP1 consisting of the CBSV Tanza genome with GFP inserted between Nib – CP (stage 3).

To ensure release of the GFP peptide from the CBSV polyprotein, the Ham1 – CP proteolytic cleavage sequence of: isoleucine-aspartic acid-valine-glutamine/-alanine (I-D-V-Q/-A) was added to the 3’ of the GFP sequence through PCR amplification with primers containing the nucleotide sequence. During plant infections, marker gene sequences can be deleted from viral transcripts, which is predicted to occur through recombination (Arazi et al., 2001; Beauchemin et al., 2005; Dawson et al., 1989; Guo et al., 1998). These wild-type like viruses containing the
deletion can then outcompete viruses containing the marker gene. To reduce sequence homology between the two I-D-V-Q/-A cleavage sequences between Ham1 – GFP and GFP – CP, the cleavage sequence between GFP-CP was designed so that third base of the isoleucine, valine and glutamine codons were altered to an alternative nucleotide, according to the schematic in Fig. 3.14. The genome structure of CBSV_Tanza_GFP1 IC is shown in Fig. 3.15.

Table 3.4: Primers used to amplify three PCR fragments to construct the CBSV_Tanza_GFP1 IC. CBSV Tanza sequence is shown in black, GFP sequence in green and the sequence encoding the proteolytic cleavage: I-D-V-Q/-A is in purple.

<table>
<thead>
<tr>
<th>Primer</th>
<th>5' – 3' sequence</th>
<th>Size (bp)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV_GFP1_F1_Fw</td>
<td>GGAAGGGTTTGCTGAAATATAATGAAACCTG</td>
<td>2021</td>
<td>CBSV Ham1 – GFP</td>
</tr>
<tr>
<td>CBSV_GFP1_F1_Rv</td>
<td>CTCCAGTGAAGTTGCTGAAATATAATGAAACCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSV_GFP1_F2_Fw</td>
<td>GCAATGATAGAAGGAGAAGAATTTTCATGGAAGTT</td>
<td>723</td>
<td>GFP</td>
</tr>
<tr>
<td>CBSV_GFP1_F2_Rv</td>
<td>CTGCACATCGATTGTGATTTTCATCATCATG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSV_GFP1_F3_Fw</td>
<td>GATGAACATACAAAAATCGATGTGACGGCA ATTGACAGGAGATGAGATTGAAGGCTGAAA</td>
<td>1711</td>
<td>GFP - CBSV CP</td>
</tr>
<tr>
<td>CBSV_GFP1_F3_Rv</td>
<td>GGCTGGCTGGTGCCAGGATATTTTGCTGTAAN</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.14: Schematic for the insertion of the GFP marker gene sequence between Ham1 and CP regions in the CBSV_Tanza IC to produce the CBSV_Tanza_GFP1 IC. Three PCR fragments were amplified using the primers in Table 3.4. Fragment 1 (F1) at the 5' contains homologous sequence with the CBSV_Tanza IC from before the first Bam H1 restriction site and at the 3' end it has homologous sequence with fragment 2 (F2). F2 encodes the GFP sequence. Fragment 3 (F3) contains homologous sequence with GFP, it then encodes CBSV CP – 3'UTR sequence and at the 3' end it contains homologous sequence with CBSV_Tanza IC from after the second Bam H1 restriction site at 8961 bp. The regions with overlapping, homologous sequences were designed to be recombined in yeast. To enable release of the GFP peptide from the CBSV polyprotein the Ham1 – CP proteolytic cleavage sequence: I-D-V-Q/-A was also cloned onto the 3' end of the GFP sequence. To reduce nucleotide similarity between the cleavage sequences, the third base in the isoleucine, valine and glutamine codons was changed to an alternative nucleotide (red).
3.2.4.3 Cloning of the CBSV_Tanza_GFP1 IC

The CBSV_Tanza_GFP1 IC was constructed by recombining three overlapping PCR fragments into CBSV_Tanza IC through homologous yeast recombination. The three fragments were amplified by high-fidelity Phusion PCR and gel purified, as shown in Fig. 3.16. The GFP sequence was amplified by PCR from the plasmid template: pCAMBIA2300_GFP, which was constructed in Pablo-Rodriguez (2017). The plasmid contains GFP sequence that was amplified by RT-PCR from the transgenic N. benthamiana 16c line, which constitutively expresses GFP (Ruiz et al., 1998).

To prepare a plasmid backbone for cloning, a CBSV_Tanza IC plasmid which had previously been verified for sequence integrity, was digested with Bam H1, which generated two fragments of 18.6 Kb and 1.7 Kb (Fig. 3.16). The larger 18.6 Kb fragment consisting of the CBSV_Tanza IC but is missing Ham1 – CP sequence. This fragment and the three PCR fragments were gel purified and transformed into yeast according to Gietz et al., (2002). Ten individual yeast colonies on test plates were pooled and grown in YDSM liquid culture, plasmids were extracted and transformed into electrocompetent E. coli ccdB cells. The first four attempts to transform E. coli with yeast plasmids failed to produce any transformant E. coli colonies. On the fifth transformation attempt, a single transformant E. coli colony was generated.

Figure 3.15: Schematic of the CBSV_Tanza_GFP1 IC genome with GFP sequence between the Ham1 and CP regions.

Figure 3.16: Left: Gel purified PCR fragments used in the construction of CBSV_Tanza_GFP1 IC through homologous yeast recombination. Fragment 1 = 2021 bp (lane 1), fragment 2 = 723 bp (lane 2) and fragment 3 = 1711 bp (lane 3). Right: Restriction digest of CBSV_Tanza IC plasmid with Bam H1 generated fragments of: 18.6 Kb and 1.7 Kb. The 18.6 Kb fragment consists of the CBSV_Tanza IC but is missing the Ham1 – CP sequence. The 18.6 Kb fragment was gel purified and used to construct CBSV_Tanza_GFP1 IC through homologous yeast recombination.
3.2.4.4 Verification of CBSV_Tanza_GFP1 construction

To confirm whether the transformant *E. coli* colony contained the correct CBSV_Tanza_GFP1 IC, it was cultured, and the plasmid extracted. The plasmid was digested with the restriction enzymes *Hind* III and *Eco*RI in separate reactions, both of which produced the expected band pattern to indicate sequence integrity (Fig. 3.17). PCR was performed with primers in Table 3.2, which showed that all the CBSV Tanza genome sections are intact and that section 6 contains extra sequence corresponding to the GFP insertion (Fig. 3.18). To further confirm the insertion of the GFP sequence, the plasmid was Sanger sequenced using the Seq9_Fw primer, which targets the GFP insertion site between the Ham1 and CP sequences. This identified that the CBSV_Tanza_GFP1 IC plasmid contained the GFP sequence after the CBSV Ham1 sequence and that the sequence was mutation free (Appendix 3.3).

![Restriction digests](image1)

*Figure 3.17: Restriction digests of the CBSV_Tanza_GFP1 plasmid with *Hind* III and *Eco*RI produced the expected restriction digest band pattern to indicate successful construction. CBSV_Tanza_GFP1 digested with *Hind* III produced a band pattern of 1.0, 6.3, 4.0 Kb to indicate plasmid integrity (lane 2). CBSV_Tanza_GFP1 digested with *Eco*RI produced a band pattern of 1.6, 4.0 and 1.0 Kb (lane 4), compared with the band pattern of 1.6, 3.3 and 1.0 kb for the CBSV_Tanza IC (lane 4). Undigested CBSV_Tanza_GFP1 is shown in lane 1.*

![PCR targeting](image2)

*Figure 3.18: PCR targeting the six CBSV_Tanza_GFP1 genome sections (Sec.) to verify successful CBSV_Tanza_GFP1 IC plasmid construction (lanes 1) and CBSV_Tanza IC (lanes +). Section 6 includes the GFP insertion site; amplification of 3.4 Kb from the CBSV_Tanza_GFP1 IC plasmid (lane 1) is shown, compared with the amplification of 2.6 Kb from the CBSV_Tanza IC plasmid (lane +). No amplification occurred in negative controls (lanes -), which contained no template.*
3.2.4.5 *N. benthamiana* infections with CBSV_Tanza_GFP1

Infections were then set up to determine whether the CBSV_Tanza_GFP1 IC could be used to visualise CBSV replication and movement *in planta* and whether infections with CBSV_Tanza_GFP1 were comparable to unmodified CBSV_Tanza IC infections. The CBSV_Tanza_GFP1 IC plasmid was transformed into the *A. tumefaciens* strain LBA 4404. Transformant *A. tumefaciens* containing either CBSV_Tanza or CBSV_Tanza_GFP1 IC plasmids were cultured. PCR confirmed that the *A. tumefaciens* cultures contained the respective plasmids. Four *N. benthamiana* plants were then agroinfiltrated with either CBSV_Tanza or CBSV_Tanza_GFP1. Inoculated and systemic leaves were monitored for GFP expression using fluorescence and confocal microscopy throughout infections. Infections were repeated in three independent repeat experiments, which produced consistent results as described below.

3.2.4.6 GFP expression during CBSV_Tanza_GFP1 infections

During CBSV_Tanza_GFP1 infections of *N. benthamiana*, GFP fluorescence was observed by confocal microscopy in the nucleus and cytoplasm of epidermal and mesophyll cells of inoculated leaves at 5 dpi (Fig. 3.19). At 15 dpi, weak GFP expression was observed in the epidermal cells of upper systemic leaves (Fig. 3.19). Out of the four *N. benthamiana* plants that were agroinfiltrated with CBSV_Tanza_GFP1, GFP fluorescence was detected in plants three and four.

![Figure 3.19](image_url): Confocal microscope images of GFP fluorescence (490 – 505 nm) in the epidermal cells of an inoculated *N. benthamiana* leaf at 5 dpi after agroinfiltration with the CBSV_Tanza_GFP1 IC (A) and in the epidermal cells of an upper systemic *N. benthamiana* leaf at 15 dpi after agroinfiltration with CBSV_Tanza_GFP1 IC (B). Image B is an overlay of green (490 – 505 nm), red (650 – 750 nm) and bright-field channels.
3.2.4.7 RT-PCR detection of CBSV_Tanza_GFP1 during *N. benthamiana* infections

To confirm whether *N. benthamiana* plants were infected with CBSV_Tanza_GFP1, RT-PCR was performed on upper systemic leaves at 10 and 20 dpi with primers targeting GFP sequence (CBSV_GFP1_F2_Fw/Rv in Table 3.4). The GFP fragment amplified from plants three and four at 10 dpi, and plant three at 20 dpi (Fig. 3.20). Sanger sequencing confirmed that the RT-PCR amplicons at 10 dpi and 20 dpi contains GFP sequence (Appendix 3.4).

![RT-PCR detection of the GFP fragment (723 bp) in upper systemic leaves of *N. benthamiana* plants. The GFP fragment weakly amplified from two (lanes 7 – 8) of the four plants at 10 dpi (lanes 5 – 8) and one (lane 3) of the four plants (lanes 1 – 4) at 20 dpi. No amplification occurred from a non-inoculated plant (lane -1) or when no template was added to the PCR (lane -2). Primers used to amplify the GFP sequence were CBSV_GFP2_F2_Fw and CBSV_GFP2_F2_Rv, provided in Table 3.4.](image)

3.2.4.8 Symptom development during CBSV_Tanza_GFP1 and CBSV_Tanza IC infections

A comparison between infections with CBSV_Tanza_GFP1 and CBSV_Tanza, demonstrated that whereas all four *N. benthamiana* plants agroinfiltrated with the CBSV_Tanza IC developed infections with characteristic CBSV symptoms, infections with CBSV_Tanza_GFP were symptomless and infections could only be detected by RT-PCR in 50% of agroinfiltrated plants, indicating that the insertion of GFP at this position in the CBSV Tanza genome had reduced viral infectivity.

3.2.5 Cloning design for the construction of CBSV_Tanza_GFP2

Further cloning was then performed to test whether the insertion of GFP at a different genome position in the CBSV_Tanza IC would generate infections with symptoms comparable to unmodified CBSV Tanza infections. The genome insertion position of between the CP and 3’UTR was chosen, as this position is at the end of the polyprotein and so may reduce interference with polyprotein translation and processing and thereby restore infectivity. Release of GFP from the polyprotein at this position would only require proteolytic cleavage at the 5’ end of the GFP sequence, as this would be the final peptide to be translated. In addition, a different
proteolytic cleavage sequence was also used to test whether this would improve infectivity. The Nla – Nb cleavage sequence of: isoleucine-serine-valine-glutamine/-alanine (I-S-V-Q/-A) was added to the 5’ end of the GFP sequence. To reduce nucleotide sequence homology between the I-S-V-Q/-A cleavage sequences, cleavage sequence at the 5’ end of the GFP sequence was designed so that third base of the isoleucine, serine and valine codons were altered to alternative nucleotides. Again, the cloning plan was to construct the CBSV_Tanza_GFP2 IC through homologous yeast recombination, whereby three overlapping PCR fragments were recombined into Bam H1 digested CBSV_Tanza IC, according to the schematic in Fig. 3.21. The genome structure of CBSV_Tanza_GFP1 IC is shown in Fig. 3.22.

Figure 3.21: Schematic for the insertion of the GFP marker gene sequence between the CP and 3’UTR regions of the CBSV_Tanza IC to produce the CBSV_Tanza_GFP2 IC. PCR fragment 1 (F1): the 5’ end contains homologous sequence with the CBSV_Tanza IC from before the first Bam H1 restriction site, it then encodes CBSV Nb – CP sequence and at the 3’ it has the I-S-V-Q-A proteolytic cleavage sequence and homologous sequence with fragment 2 (F2). F2 encodes the GFP sequence. Fragment 3 (F3): the 3’ end contains homologous sequence with GFP, it then encodes CBSV 3’UTR sequence and at the 3’ contains homologous sequence with CBSV_Tanza IC from after the second Bam H1 restriction site. To reduce nucleotide similarity between the I-S-V-Q-A proteolytic cleavage sequences, the third bases in the isoleucine, serine and valine codons were changed to an alternative nucleotide (red).

Figure 3.22: Schematic of the CBSV_Tanza_GFP2 IC genome with GFP sequence between the CP and 3’UTR.
3.2.5.1 PCR design for construction of CBSV_Tanza_GFP2

The three PCR fragments used in the construction of the CBSV_Tanza_GFP2 IC were amplified with primers in Table 3.5. Fragments 1 and 3 were amplified using the CBSV_Tanza IC as template, whereas fragment 2 was amplified using the pCAMBIA2300_GFP as template. The PCR fragments were designed as follows:

- **F1**: The 5’ end contains 30 bp of homologous sequence from before the first Bam HI restriction site at position 7284 bp in the CBSV_Tanza IC plasmid, it then encodes CBSV Tanza sequence to the end of CP and at the 3’ end it contains the nucleotide sequence to encode the I-S-V-Q-A cleavage sequence and 30 bp of homologous sequence with F2.

- **F2**: Encodes the GFP sequence with I-S-V-Q-/A proteolytic cleavage sequences at the 5’ end.

- **F3**: The 5’ end contains 30 bp of homologous sequence with F2, it then encodes 3’UTR sequence and at the 3’ end there is 30 bp of homologous sequence with CBSV Tanza IC from after the second Bam HI restriction site at position 8961 bp.

Table 3.5: Primers used to amplify PCR fragments used to construct the CBSV_Tanza_GFP2 IC. CBSV Tanza sequence is shown in black, GFP sequence in green and the proteolytic cleavage: I-S-V-Q-/A is in purple.

<table>
<thead>
<tr>
<th>Primer</th>
<th>5’ – 3’ sequence</th>
<th>Size (bp)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV_GFP2_F1_Fw</td>
<td>GGAAGGGTTTGCTGAAATATAATGAACCTG</td>
<td>3154</td>
<td>CBSV CP – GFP</td>
</tr>
<tr>
<td>CBSV_GFP2_F1_Rv</td>
<td>GTGAAAAGTTCTTCTCCTTTACTGGCTTGAAACAGAGA TTCAATAGCAGCACCAGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSV_GFP2_F2_Fw</td>
<td>GCCAGTAAAGGGAGAAGAACTTTTCTCTGAGTTGTC</td>
<td>723</td>
<td>GFP</td>
</tr>
<tr>
<td>CBSV_GFP2_F2_Rv</td>
<td>CAAATTTTTTGTATAGTTCATCCGATGATACATTGTA ATCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSV_GFP2_F3_Fw</td>
<td>GATTACACATGGCATGAGTGAACTACATACAAAATTAA TTTGAAAGTCAGTTGAGGG</td>
<td>587</td>
<td>GFP – CBSV 3’UTR</td>
</tr>
<tr>
<td>CBSV_GFP2_F3_Rv</td>
<td>GGCTGGCTTGTTGCAAGATATTGTTGCTAAG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2.5.2 Cloning of the CBSV_Tanza_GFP2 IC

The CBSV_Tanza_GFP2 IC was constructed through recombining three overlapping PCR fragments into CBSV_Tanza IC through homologous yeast recombination. The three fragments were amplified by high-fidelity Phusion PCR and gel purified. The CBSV_Tanza IC was digested with Bam HI, which generated two fragments of 18.6 Kb and 1.7 Kb, as previously shown in Fig. 3.16. The larger 18.6 Kb fragment and three PCR fragments were gel purified and transformed into yeast according to Gietz et al., (2002). Ten individual yeast colonies on test plates were pooled and grown in YDSM liquid culture, plasmids were extracted and
transformed into electrocompetent *E. coli* ccdB cells. This transformation produced 10 transformant *E. coli* colonies.

### 3.2.5.3 Verification of CBSV_Tanza_GFP2 construction

To confirm whether transformant *E. coli* colonies contained the correct CBSV_Tanza_GFP2 IC plasmid, four colonies were cultured, and plasmids extracted. Restriction digest of all four plasmids with *Hind* III produced the expected band pattern to indicate that no significant sequence rearrangements had occurred (Fig. 3.23). PCR was then performed on the four plasmids with primers in Table 3.2, which demonstrated that all genome sections are intact and that plasmids 1 and 2 contain extra sequence in the sixth genome section, corresponding to the insertion of GFP (Fig. 3.24). To further confirm the insertion of the GFP sequence, plasmid 1 was Sanger sequenced using the Seq11_Fw primer, which targets the GFP insertion site between the CP and 3’UTR sequences. This identified that the CBSV_Tanza_GFP2 IC plasmid contains GFP sequence after the CBSV CP sequence and the sequence is mutation free (Appendix 3.5).

![Figure 3.23](image1.png)

**Figure 3.23:** PCR amplification of correct sized genome sections (Sec.) from two CBSV_Tanza_GFP2 IC plasmids, targeting the unstable regions of the CBSV Tanza genome and the GFP insertion site to indicate successful IC construction. The correct sized fragments amplified from plasmids 1 and 2 (lanes 1 - 2) and CBSV_Tanza IC (lanes +) to indicate sequence integrity of unstable genome sections: 4 = 0.9 Kb and 5 = 1.7 Kb. Section 6 (GFP insertion site); amplification of 3.4 Kb is shown from plasmids 1 and 2 (lane 1 - 2), compared with the amplification of 2.6 Kb from the CBSV_Tanza IC plasmid (lane +). This indicated successful insertion of the GFP sequence into the CBSV_Tanza IC. No amplification occurred in negative controls (lanes -), which contained no template.

![Figure 3.24](image2.png)

**Figure 3.24:** Restriction digest of four potential CBSV_Tanza_GFP2 plasmids with *Hind* III (lanes 1 - 4) produced expected banding pattern of 10.9, 5.6 and 4.1 Kb, which indicated that no significant sequence rearrangements had occurred.
3.2.6 Infections with CBSV_Tanza_GFP1, CBSV_Tanza_GFP2 and CBSV_Tanza

To test whether CBSV_Tanza_GFP2 could infect *N. benthamiana* and to compare infections with CBSV_Tanza_GFP1 and CBSV_Tanza IC, all three plasmids were used to agroinfiltrate *N. benthamiana*. Transformant *A. tumefaciens* containing CBSV_Tanza_GFP1, CBSV_Tanza_GFP2 and CBSV_Tanza IC plasmids were cultured and PCR confirmed the presence of IC plasmids. Four *N. benthamiana* plants were then agroinfiltrated with each IC. Inoculated and systemic leaves were monitored for GFP expression using wide-field and confocal microscopy. Infections were repeated in three independent repeat experiments, which produced consistent results as described below.

3.2.6.1 GFP expression during CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 infections

During *N. benthamiana* infections with both CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 ICs, fluorescence was observed by confocal microscopy in the epidermal and mesophyll cells of inoculated leaves at 5 dpi (Fig. 3.25). At 14 dpi GFP expression was observed in the vascular system and leaf lamina of upper systemic leaves (Fig. 3.26) and was visible in epidermal and mesophyll cells (Fig. 3.25). After 20 dpi, GFP was no longer visible in any of the inoculated plants. There was no obvious difference in the level of GFP expression in CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 infections. Green fluorescence was visible in three out of the four *N. benthamiana* plants agroinfiltrated with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2.

Figure 3.25: Confocal microscopy images of GFP fluorescence (490 – 505 nm) in the epidermal cells of *N. benthamiana* leaves agroinfiltrated with CBSV_Tanza_GFP1 (A) and CBSV_Tanza_GFP2 (B) at 5 dpi and in the epidermal (C) and mesophyll cells (D) of upper systemic *N. benthamiana* leaves at 14 dpi after agroinfiltration with CBSV_Tanza_GFP1 (C) and CBSV_Tanza_GFP2 (D).
3.2.6.2 RT-PCR detection of CBSV_Tanza_GFP1/2 during infections

To determine whether *N. benthamiana* plants were infected with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2, RT-PCR was performed on upper systemic leaves at 10 dpi with primers targeting GFP sequence (CBSV_GFP1_F1/2_Fw/Rv in Table 3.4-5). This GFP fragment amplified from all four plants agroinfiltrated with CBSV_Tanza_GFP1 and three of the four plants agroinfiltrated with CBSV_Tanza_GFP2 (Fig. 3.27). Sanger sequencing confirmed that the RT-PCR amplicons contained GFP sequence (Appendix 3.6).

![Figure 3.27: RT-PCR detection of the GFP fragment (723 bp) in upper systemic leaves of four *N. benthamiana* plants agroinfiltrated with CBSV_Tanza_GFP1 (lanes 1 – 4) and three of the four plants agroinfiltrated with CBSV_Tanza_GFP2 (lanes 5 – 7) at 10 dpi. Amplification of the GFP fragment occurred in the positive control PCR when the CBSV_Tanza_GFP1 IC was used as template (+). No amplification occurred when cDNA from a non-inoculated plant cDNA was used as template (not shown). PCR was performed with the CBSV_GFP1/2_F2_Fw/Rv (Table 3.4 - 5), which target the GFP fragment.](image-url)
3.2.6.3 RT-PCR detection of viral transcripts with GFP deletions

To determine whether viral transcripts with GFP sequence deletions could be detected, RT-PCR was performed on the upper systemic *N. benthamiana* leaf material at 15 dpi, using primers in Table 3.6. This detected GFP deletions in one of four plants agroinfiltrated with CBSV_Tanza_GFP1 (Fig. 3.28) and one of four plants agroinfiltrated with CBSV_Tanza_GFP2 (Fig. 3.28). RT-PCR amplicon sequencing confirmed complete, perfect GFP deletions in both CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 infections (Appendix 3.7 – 3.8).

Table 3.6: Primers used in RT-PCR to detect the presence or absence of GFP in *N. benthamiana* infections with the CBSV_Tanza and CBSV_Tanza_GFP1/2 ICs.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Primer sequence</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV_GFP1_del1_Fw</td>
<td>GTGAGATTGTGATGCCACGGGCACTGGAATT</td>
<td>CBSV_Tanza = 804</td>
</tr>
<tr>
<td>CBSV_GFP1_del1_Rv</td>
<td>CAACCTCCTCCTCATACGCTGTTCCATGGCCCTTAC</td>
<td>CBSV_Tanza_GFP1 = 1536</td>
</tr>
<tr>
<td>CBSV_GFP2_del2_Fw</td>
<td>GCTTTAGCTGCTCTGGATGATAATAAT</td>
<td>CBSV_Tanza = 914</td>
</tr>
<tr>
<td>CBSV_GFP2_del2_Rv</td>
<td>GAAAGGTTCCATGAAATATCTTGGTTTCTCA</td>
<td>CBSV_Tanza_GFP2 = 190</td>
</tr>
</tbody>
</table>

Figure 3.28: Left: RT-PCR detection of viral transcripts containing GFP deletions in one (lane 3) of the four *N. benthamiana* plants (lanes 1 – 4) agroinfiltrated with CBSV_Tanza_GFP1 at 15 dpi. PCR was performed with CBSV_GFP1_del1_Fw/Rv primers (Table 3.6) were used to target the GFP insertion site between the Ham1 and CP. Amplification of 1536 bp occurred in the positive control PCR when the CBSV_Tanza_GFP1 IC plasmid was used as template (lane +1) and amplification of 804 bp occurred when the CBSV_Tanza IC plasmid was used as template (lane +2). No amplification occurred with cDNA from a non-inoculated plant (lane -).

Right: RT-PCR detection of viral transcripts containing GFP deletions in one (lane 4) of the four *N. benthamiana* plants agroinfiltrated with CBSV_Tanza_GFP2 at 15 dpi. PCR was performed with CBSV_GFP2_del2_Fw/Rv primers were used (Table 3.6) to target the GFP insertion site between the CP and 3’UTR. Amplification of 914 bp occurred in the positive control PCR when CBSV_Tanza_GFP2 IC plasmid was used as template (lane +1) and amplification of 190 bp occurred when the CBSV_Tanza IC plasmid was used as template (lane +2). No amplification occurred with cDNA from a non-inoculated plant (lane -).
3.2.6.3 Quantification of CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 viral titers

To compare viral titers in *N. benthamiana* infections with the CBSV_Tanza_GFP1/2 with CBSV_Tanza, qPCR was performed on upper systemic leaf material at 15 dpi. The CBSV CP transcript abundance was calculated relative to the expression of the *N. benthamiana* F-box gene (TAIR: At5g15710), which has previously been validated to show relatively stable expression during viral infection of *N. benthamiana* (Liu et al., 2012). Relative transcript abundance was calculated using the $2^{\Delta\Delta Ct}$ method, according to Livak and Schmittgen, (2001). Results from the qPCR are shown in Fig. 3.29. This demonstrated that viral titers during CBSV_Tanza_GFP1/2 infections were extremely low compared to titers in CBSV_Tanza infections and the insertion of GFP into the CBSV Tanza genome significantly reduces viral infectivity. There may be slightly higher titers in CBSV_Tanza_GFP2 infections compared with CBSV_Tanza_GFP1, however further repeat experiments are required to verify this difference.

![Graph showing quantification of CBSV titers](image_url)

**Figure 3.29:** QPCR quantification of higher CBSV titers in systemic leaves of *N. benthamiana* plants agroinfiltrated with CBSV_Tanza, compared with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 at 15 dpi. The abundance of the CBSV coat protein transcript was compared to the abundance of the F-box reference gene. The $2^{\Delta\Delta Ct}$ method was used to calculate the relative CBSV transcript abundance in infected plants compared with three non-inoculated plants.
3.2.6.4  Symptom development during CBSV_Tanza_GFP1/2 infections

In comparison with severe necrosis which developed during *N. benthamiana* infections with CBSV_Tanza, infections with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 ICs that tested positive for CBSV infection by RT-PCR were symptomless (Fig. 3.30). This indicated that the presence of GFP in the CBSV Tanza genome significantly affects symptom development.

Figure 3.30: *N. benthamiana* infections with CBSV_Tanza_GFP1 (A) and CBSV_GFP2 (B) were symptomless compared to infections with wild-type CBSV_Tanza, which developed severe systemic necrosis at 14 dpi (C).

3.2.6.5  Mechanical-inoculation with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2

To test whether CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 infected *N. benthamiana* material was infectious, it was used in a mechanical inoculation experiment. Upper systemic leaf material was sampled from *N. benthamiana* infected with CBSV_Tanza_GFP1 and CBSV_Tanza_GFP2 at 15 dpi, as confirmed by RT-PCR. This material was used to mechanically inoculate five-week-old *N. benthamiana* and three-week-old cassava cuttings. Plants were monitored every 5 dpi for GFP expression using wide-field microscopy. Green fluorescence was not observed in inoculated or systemic leaves at 5, 10, 15, 20 or 30 dpi after mechanical inoculation and RT-PCR also failed to detect the presence of CBSV Tanza or GFP sequence at 10, 20 and 30 dpi after mechanical inoculation. This indicated that *N. benthamiana* material infected with CBSV_Tanza_GFP1/2 was not infectious. Whereas mechanical inoculation with wild-type CBSV_Tanza infected *N. benthamiana* successfully generated infections in *N. benthamiana* (Fig. 3.30).

91
3.2.6.6 Attempts to construct the CBSV_Tanza_iLOV IC

To test whether the insertion of a shorter marker sequence into the CBSV_Tanza IC would produce infections with higher viral titers, efforts were made to construct a CBSV_Tanza IC containing the shorter iLOV marker gene sequence (≈ 300 bp) between the CP – 3’UTR regions. The iLOV protein is derived from the ‘light, oxygen or voltage-sensing domain’ of the plant blue light phototropic receptor (Christie et al., 2012). It has previously been shown to outperform GFP as marker gene during TMV infections (Chapman et al., 2008). Unfortunately, efforts to construct the CBSV_Tanza_iLOV IC were unsuccessful. Yeast recombination was performed, and yeast plasmids were transformed into electro-competent E. coli ccdB, however, when plasmids were extracted from E. coli and analyzed by PCR, several of genome sections were found to be missing. It was evident that cloning CBSV_Tanza_iLOV would be challenging and would require alternative approaches that were beyond the time constraints of this project.

3.2.7 Construction of a CBSV Tanza IC containing a UCBSV coat protein replacement

After inserting the GFP marker gene, the next aim was to use the CBSV_Tanza IC to construct a chimera IC consisting of the CBSV Tanza genome with a UCBSV sequence replacement. This was performed to identify sequences associated with differential symptom development and viral accumulation during infections with CBSV Tanza and UCBSV Kikombe. Previously, CP sequence variants have previously been identified as symptom determinants for TMV (Saito et al., 1987), CMV (Takeshita et al., 2001) and PepMV (Duff-Farrier et al., 2015). To identify the levels of sequence identity between the CBSV Tanza and UCBSV Kikombe CP sequences, an alignment was performed as shown in (Fig. 3.31). This identified that overall the CP sequences share 72% nucleotide and 77% amino acid sequence similarity, that there is low sequence similarity at amino acid positions 1 – 120 and that the CBSV CP contains extra sequence, which is not found in the UCBSV sequence. These differences in the N’ terminals of CBSV and UCBSV CP sequences have been previously reported in Winter et al., (2010). Due to these CP sequence differences, it was hypothesised that this genome region may contain symptom determinants and so a chimera IC consisting of the CBSV Tanza genome with a UCBSV Kikombe CP sequence replacement was constructed.
The cloning plan was to construct the CBSV_Tanza_UCP IC through homologous yeast recombination. Three overlapping PCR fragments (F1, F2 and F3) were designed to be recombined into AarI and AscI digested CBSV_Tanza IC, according to the schematic in Fig. 3.32.

To ensure release of the UCBSV CP sequence from the CBSV Tanza polyprotein the proteolytic cleavage sequence was designed according to the schematic in Fig. 3.33. In the CBSV Tanza genome, the Ham1 and CP peptides are separated by the cleavage sequence: I-D-V-Q/-A, whereas in the UCBSV Kikombe genome they are separated by the sequence: I-D-V-Q/-V. The chimera was designed to have the I-D-V-Q/-V sequence between the CBSV Tanza Ham1 and the UCBSV CP sequences, which should enable cleavage. As the CP is the last encoded peptide before the 3’UTR sequence, there was no need to clone a proteolytic cleavage sequence at the 3’ end of the UCBSV Kikombe CP sequence. The genome structure of the CBSV_Tanza_UCP IC is shown in Fig. 3.34.

3.2.7.1 Cloning design for construction of CBSV_Tanza_UCP

The cloning plan was to construct the CBSV_Tanza_UCP IC through homologous yeast recombination. Three overlapping PCR fragments (F1, F2 and F3) were designed to be recombined into AarI and Ascl digested CBSV_Tanza IC, according to the schematic in Fig. 3.32.

To ensure release of the UCBSV CP sequence from the CBSV Tanza polyprotein the proteolytic cleavage sequence was designed according to the schematic in Fig. 3.33. In the CBSV Tanza genome, the Ham1 and CP peptides are separated by the cleavage sequence: I-D-V-Q/-A, whereas in the UCBSV Kikombe genome they are separated by the sequence: I-D-V-Q/-V. The chimera was designed to have the I-D-V-Q/-V sequence between the CBSV Tanza Ham1 and the UCBSV CP sequences, which should enable cleavage. As the CP is the last encoded peptide before the 3’UTR sequence, there was no need to clone a proteolytic cleavage sequence at the 3’ end of the UCBSV Kikombe CP sequence. The genome structure of the CBSV_Tanza_UCP IC is shown in Fig. 3.34.
3.2.7.2 PCR design for construction of CBSV_Tanza_UCP

The three PCR fragments used in the construction of the CBSV_Tanza_UCP IC were amplified with the primers in Table 3.7. Fragments 1 and 3 were amplified using the CBSV_Tanza IC as template, whereas the UCBSV CP sequence in fragment 2 was amplified using the UCBSV_Kikombe IC as template. The three PCR fragments were designed as follows:

- **F1**: The 5’ end contains 30 bp of homologous sequence from before the *AarI* restriction site at position 7423 bp in the CBSV_Tanza IC plasmid, it then encodes CBSV Tanza sequence from Nilb to the end of Ham1 and the 3’ end contains 30 bp of homologous sequence with F2.
- **F2**: Encodes the UCBSV Kikombe CP sequence.
- **F3**: At the 5’ end contains 30 bp of homologous sequence with F2, it then encodes CBSV 3’UTR sequence and at the 3’ end there is homologous sequence with CBSV_Tanza IC from after the *Ascl* restriction site at position 10695 bp.

**Stage 1: digest**

**Stage 2: yeast recombination**

**Stage 3: rescue into *E. coli***

Figure 3.32: Schematic for the construction of the CBSV_Tanza_UCP through homologous yeast recombination. The CBSV_Tanza IC was double digested with *AarI* and *Ascl*, to generate 3.2 Kb and 16.9 Kb fragments (stage 1). The 16.9 Kb fragment and PCR fragments: F1 - F3 were then recombined in yeast (stage 2) and yeast plasmids were rescued into *E. coli* to form the CBSV_Tanza_UCP IC consisting of the CBSV Tanza genome with the UCBSV CP sequence inserted between Ham1 - 3’UTR (stage 3).
Table 3.7: Primers used to amplify the three overlapping PCR fragments that were used to recombine UCBSV Kikombe CP sequence (red) into the CBSV_Tanza IC to generate the CBSV_Tanza_UCP IC.

<table>
<thead>
<tr>
<th>Primer</th>
<th>5' – 3' sequence</th>
<th>Size (bp)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV_UCP_F1_Fw</td>
<td>AGAGGAATTCATTATTGTGTGTCAGTGCGTGAAAGT</td>
<td>1790</td>
<td>CBSV NIb</td>
</tr>
<tr>
<td>CBSV_UCP_F1_Rv</td>
<td>GTCTCTGCTTCTATCTCCTCTGTGATTTTAAGACTTGAAACATCAATAAAGAAATCA</td>
<td>1103</td>
<td>UCBSV CP</td>
</tr>
<tr>
<td>CBSV_UCP_F2_Fw</td>
<td>GTCTTAAATCAGGAGGAGGATAGAAAGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSV_UCP_F2_Rv</td>
<td>TTATTCAATTGCGGCACCACTGTGACTGTGTCCTGTTGCGACTAACC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBSV_UCP_F3_Fw</td>
<td>TTACAGTGGTGGCGCAATTTGAATAAAATTATTTGAAAGT</td>
<td>572</td>
<td>CBSV 3'UTR</td>
</tr>
<tr>
<td>CBSV_UCP_F3_Rv</td>
<td>GGTGCCAGGATATATTG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.33: Schematic for the replacement of the CBSV CP with UCBSV CP sequence in the chimera CBSV_Tanza_UCP IC. Three PCR fragments were amplified using the primers in Table 3.7. Fragment 1 (F1) at the 5' end contains homologous sequence with the CBSV_Tanza IC from before the AarI restriction site in the CBSV_Tanza IC plasmid, it then encodes CBSV Tanza sequence from NIb to the end of Ham1 and at the 3' end it contains 30 bp of homologous sequence with fragment 2 (F2). F2 encodes the UCBSV Kikombe CP sequence. Fragment 3 (F3) at the 3' end contains homologous sequence with F2, it then encodes CBSV 3'UTR sequence and at the 3' end it contains homologous sequence with CBSV_Tanza IC from after the AscI restriction site at 10695 bp. The proteolytic cleavage sequence: I-D-V-Q-/V was inserted between the CBSV Ham1 and UCBSV CP peptides to enable release of the UCBSV CP from the CBSV polyprotein.

Figure 3.34: Schematic of the CBSV_Tanza_UCP IC genome consisting of the CBSV Tanza genome (blue) with a UCBSV Kikombe CP (red) replacement.
3.2.7.3 Cloning of CBSV_Tanza_UCP IC

The CBSV_Tanza_UCP IC was constructed by recombining three overlapping PCR fragments into CBSV_Tanza IC through homologous yeast recombination. The three fragments were amplified by high-fidelity Phusion PCR, as shown in Fig. 3.35. The CBSV_Tanza IC was digested with AarI and Ascl, which generated two fragments of 3.2 Kb and 16.9 Kb, as shown in Fig. 3.35. The larger 16.9 Kb fragment corresponds to the CBSV Tanza genome but without Nib – 3’UTR sequence. The 16.9 Kb digest fragment and three PCR fragments were gel purified and transformed into yeast according to Gietz et al., (2002). Transformant yeast were grown in YDSM liquid culture, plasmids were extracted and transformed into electrocompetent E. coli ccdB cells. This generated two transformant E. coli colonies.

3.2.7.4 Verification of CBSV_Tanza_UCP construction

To confirm whether the two transformant E. coli colonies contained the correct CBSV_Tanza_UCP IC plasmid, they were cultured, and plasmids were extracted. Restriction digest of the two plasmids with Hind III produced the expected band pattern to indicate that no significant sequence rearrangements had occurred (Fig. 3.36). The presence of the UCBSV CP replacement in both plasmids was confirmed by PCR using the CBSV_ UCP_F2_Fw/Rv primers (Table 3.7), which target the UCBSV CP sequence and amplified a fragment with the expected size (Fig. 3.37). PCR targeting all six genome sections demonstrated that the plasmids did not have significant sequence rearrangements (Fig. 3.37). To further confirm IC construction, plasmid one was Sanger sequenced using the Seq9_Fw primer, which targets the UCBSV CP replacement. This identified the presence of UCBSV CP sequence between the CBSV Ham1 and 3’UTR, in place of CBSV CP sequence (Appendix 3.9 – 3.10).
3.2.8 *N. benthamiana* infections with CBSV_Tanza_UCP

To test whether CBSV_Tanza_UCP can infect plants and whether the UCBSV CP replacement is associated with differential symptom development compared to CBSV_Tanza infections, *N. benthamiana* plants were agroinfiltrated with both plasmids. The *A. tumefaciens* strain LBA 4404 was transformed with the CBSV_Tanza_UCP IC plasmid. Four *N. benthamiana* plants were then agroinfiltrated with each IC. Plants were monitored for symptom development and sampled for viral detection and quantification. Infections were repeated in three repeat experiments, which produced consistent results as described below. Unfortunately it was not
possible to infect plants with UCBSV_Kikombe due to sequence instability of the IC described in section 3.2.1 that prevented inoculation with infectious transcripts, as well as a lack of viable UCBSV_Kikombe plant inoculum.

### 3.2.8.1 Symptom development during CBSV_Tanza_UCP infections

*N. benthamiana* plants agroinfiltrated with CBSV_Tanza_UCP developed systemic symptoms, demonstrating that the chimeric virus can replicate and move systemically. Differences in symptom development were observed during infections with CBSV_Tanza_UCP compared with CBSV_Tanza (Fig. 3.38). At 7 dpi, agroinfiltrated leaves in CBSV_Tanza infections developed necrosis, whereas CBSV_Tanza_UCP infections developed chlorosis. At 10 dpi, CBSV_Tanza infections developed severe systemic necrosis, whereas no systemic symptoms were present in CBSV_Tanza_UCP infections. At 14 - 18 dpi, CBSV_Tanza infections developed severe systemic necrosis that resulted in plant death, whereas CBSV_Tanza_UCP infections developed mild necrosis in upper systemic leaves, which became more severe at 18 – 21 dpi but did not result in plant death. *N. benthamiana* plants infected with either CBSV_Tanza or CBSV_Tanza_UCP showed similar levels of growth stunting compared to non-inoculated plants. This indicated that the CBSV CP may be associated with the development of more severe necrosis during CBSV infections of *N. benthamiana* and that the UCBSV CP may be able to complement the loss of the CBSV CP during CBSV_Tanza_UCP infections.

![Figure 3.38: Symptom development during N. benthamiana infections with CBSV_Tanza (A) and CBSV_Tanza_UCP (B), compared to a symptomless non-inoculated plant (C) at 18 dpi. Infections with CBSV_Tanza develop severe systemic chlorosis and necrosis, whereas infections with CBSV_Tanza_UCP develop milder necrosis in upper systemic leaves. N. benthamiana infected with CBSV_Tanza or CBSV_Tanza_UCP show similar levels of growth stunting compared to non-inoculated plants.](image)
3.2.8.2 Quantitative assessment of symptom development

To quantitatively compare symptom development during *N. benthamiana* infections with CBSV_Tanza_UCP and CBSV_Tanza, symptoms were assessed throughout infection using the scoring system in Table 3.8, adapted Ogwok et al., (2010). This demonstrated that compared with CBSV_Tanza infections, CBSV_Tanza_UCP infections show a delay in onset of symptom development and a reduction in symptom severity (Fig. 3.39). This indicated that the CBSV CP may be involved with development of severe necrosis during early infections.

<table>
<thead>
<tr>
<th>Severity score</th>
<th>Symptom description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No symptoms</td>
</tr>
<tr>
<td>2</td>
<td>Necrosis/chlorosis on agroinfiltrated leaf</td>
</tr>
<tr>
<td>3</td>
<td>Mild systemic necrosis/chlorosis</td>
</tr>
<tr>
<td>4</td>
<td>Severe systemic necrosis/chlorosis</td>
</tr>
<tr>
<td>5</td>
<td>Plant death</td>
</tr>
</tbody>
</table>

Table 3.8: Symptom severity scoring system used to quantitatively assess symptom development during *N. benthamiana* infections with CBSV_Tanza and CBSV_Tanza_UCP.

Figure 3.39: Symptom severity scores throughout *N. benthamiana* infections with CBSV_Tanza and CBSV_Tanza_UCP ICs. Infections with CBSV_Tanza develop necrosis and chlorosis in agroinfiltrated leaves at around 7 dpi, which becomes systemic at 10 dpi, severe necrosis/chlorosis develops around 14 dpi and plants die from infection at 18 – 21 dpi. Whereas infections with CBSV_Tanza_UCP do not develop symptoms until 10 dpi, when necrosis/chlorosis is observed in agroinfiltrated leaves, which becomes systemic around 14 dpi and more severe at 18 - 21 dpi but does not result in plant death.
3.2.8.3 RT-PCR detection of CBSV_Tanza_UCP infections

To confirm that *N. benthamiana* plants were infected with CBSV_Tanza_UCP viral transcripts, containing UCBSV Kikombe CP sequence, RT-PCR was performed on upper systemic leaf material at 14 dpi. The primers CBSV_UCP_F2_Fw/Rv primers (Table 3.7) were used, which detected the presence of the UCBSV Kikombe CP sequence in three out of the four-agroinfiltrated plants (Fig. 3.40). RT-PCR amplicon sequencing confirmed the presence of the UCBSV Kikombe CP sequence in these plants (Appendix 3.11).

![RT-PCR detection of UCBSV Kikombe CP in systemic leaf material sampled from three N. benthamiana plants (lanes 2 – 4) agroinfiltrated with the CBSV_Tanza_UCP IC at 14 dpi. The CBSV_UCP_F2_Fw/Rv primers (Table 3.7) were used, which specifically amplify the UCBSV Kikombe CP fragment = 1103 bp. Amplification of 1103 bp occurred in the positive control PCR when the CBSV_Tanza_UCP IC plasmid was used as template (lane +). No amplification occurred from a non-inoculated plant (lane -1) or when no template was added to the PCR (lane -2).]

3.2.8.4 Viral titers during CBSV_Tanza_UCP infections

To compare viral titers during *N. benthamiana* infections with CBSV_Tanza and CBSV_Tanza_UCP, TAS-ELISA was performed on upper systemic leaf material at 7, 14, 21 and 28 dpi (Fig. 3.41). This demonstrated that viral titers in CBSV_Tanza infections are higher at 7 dpi, peak at 14 dpi and then decrease at 21 – 28 dpi due to plant death. In comparison, viral titers in CBSV_Tanza_UCP infections are lower than CBSV_Tanza at 7 – 14 dpi and increase to higher levels at 21 – 28 dpi. This indicated that CBSV CP may be involved with relatively high levels of viral accumulation during early CBSV infection of *N. benthamiana.*
3.2.8.5 Back-inoculation of *N. benthamiana* with CBSV_Tanza_UCP infected material

To test whether CBSV_Tanza_UCP virions are infectious, a mechanical inoculation experiment was performed. *N. benthamiana* plants infected with CBSV_Tanza and CBSV_Tanza_UCP were sampled at 14 dpi and material from each was used to mechanically inoculate six *N. benthamiana* plants. Infections developed in two of the six plants inoculated with either CBSV_Tanza_UCP or CBSV_Tanza. The remaining plants were symptomless. Plants infected with CBSV_Tanza developed severe systemic chlorosis, strong necrosis and wilting by 10 dpi, whereas plants infected with CBSV_Tanza_UCP developed systemic necrosis and wilting at 14 dpi (Fig. 3.42). This demonstrated that the timing and severity of symptom development during CBSV_Tanza_UCP or CBSV_Tanza infections were similar in the first passage, when plants were agroinfiltrated with the ICs and in the second passage, when plants were mechanically inoculated with infected material. Therefore differences in symptom development between CBSV_Tanza and CBSV_Tanza_UCP were not dependent on inoculation method.
Figure 3.42: Symptom development during infections of *N. benthamiana* inoculated with CBSV_Tanza (A) and CBSV_Tanza_UCP (B) infected material, compared to symptomless non-inoculated plant (C) at 18 dpi. Plants infected with CBSV_Tanza developed severe systemic chlorosis, necrosis and wilting by 10 dpi, whereas plants infected with CBSV_Tanza_UCP developed systemic necrosis and wilting at 14 dpi.

3.2.8.6 Attempts to infect cassava with CBSV_Tanza_UCP

To determine whether the CBSV_Tanza_UCP transcripts could infect cassava, several inoculation attempts were performed. Firstly, attempts were made to agroinfiltrate cassava through mechanical damage of cassava leaves with carborundum followed by soaking with agroinfiltration suspension. However, agroinfiltrated cassava plants failed to develop infections as tested by RT-PCR on systemic leaf material sampled at 21 dpi and 3, 6, and 12 months post inoculation. In a second approach, *N. benthamiana* leaf material infected with CBSV_Tanza_UCP was used to mechanically inoculate cassava leaves, however this also failed to generate infections.

3.2.8.7 Attempts to construct a UCBSV IC containing a CBSV coat protein replacement

To further demonstrate that the CBSV CP is involved with the development of severe necrosis, ideally a UCBSV Kikombe IC with reciprocal CBSV Tanza CP replacement would be constructed. Unfortunately, it was not possible to construct this IC due to sequence instability associated with the UCBSV_Kikombe IC, which prevented the extraction of sufficient quantities of the IC plasmid required for cloning, as described in section 3.2.1.
3.3 Discussion

In this Chapter the UCBSV_Kikombe and CBSV_Tanza ICs, which were previously constructed at the University of Bristol were tested in terms of their utility. Firstly, they were tested for sequence stability during plasmid propagation in *E. coli*, which identified that both IC genomes contain unstable regions. Sequence instability issues with the UCBSV_Kikombe IC were encountered during propagation in three different *E. coli* strains. Whereas the CBSV_Tanza IC demonstrated higher sequence stability the *E. coli* strain *ccd*B, which is likely to be due to the presence of three introns. The CBSV_Tanza IC was also used to agroinfiltrate cassava; however, the inoculation efficiency was relatively low. Therefore, to enable the CBSV_Tanza IC to be distributed to and used by other researcher groups, further sequence stabilization and improvements to agroinfiltration will be required. The CBSV_Tanza IC was sufficiently stable to enable manipulations through both the insertion of the marker gene GFP and a UCBSV CP sequence replacement. The insertion of GFP enabled CBSV *in planta* replication and movement to be visualised, whereas the UCBSV CP replacement demonstrated that the CBSV CP is associated with the development of severe necrosis and relatively high levels viral accumulation early in infection of *N. benthamiana*, as assessed by TAS-ELISA. The results from these experiments are discussed below.

3.3.1 Sequence instability of the UCBSV Kikombe and CBSV Tanza ICs in *E. coli*

The UCBSV Kikombe and CBSV Tanza ICs were tested for sequence stability during propagation in a range of *E. coli* strains. Plasmid propagation of UCBSV_Kikombe IC in the DH5α strain was associated with sequence instability in the middle of the NiA region. Similarly, propagation in ccdB strain resulted in rearrangements in the VPg – NiA region. This suggests that the UCBSV Kikombe NiA region should be targeted for sequence stabilization in future work. Although other areas of the UCBSV Kikombe genome, such as the CI region may also be associated with instability, as reported in Nanyiti (2016). Interestingly, instability has not been found in this region of the CBSV Tanza genome, which suggests that instability issues may be CBSV/UCBSV species-specific. Propagating the UCBSV_Kikombe IC in the *E. coli* strain *α*-select did enable the extraction of plasmids without significant sequence rearrangements. However, the concentrations of these plasmids were extremely low, which prevented their use in *in vitro* transcription to inoculate plants or to perform cloning manipulations. Jakab et al., (1997) have also reported similar challenges when constructing a full-length PVY IC, where only low concentrations of plasmids without sequence rearrangements could be extracted from *E. coli*. Sequence instability was also found during propagation of the CBSV_Tanza IC in the *E. coli* strain
C43. In this case, it appears that the NiB region contains unstable sequence, which indicates that moving or extending the third intron to further disrupt NiB may permit higher sequence stability. Propagation of the CBSV_Tanza IC in the E. coli strain ccdB was relatively stable, as no major sequence rearrangements were detected by PCR. Therefore this E. coli strain should be used for the long-term storage and propagation of the CBSV_Tanza IC.

Future work to improve the sequence stability of U/CBSV ICs could include using in vitro cloning methods, which bypass plasmid construction in yeast and propagation in E. coli. For instance Gibson-assembly (Bordat et al., 2015) or In-Fusion PCR (Tuo et al., 2015) could be used to assemble overlapping PCR fragments to clone the full-length U/CBSV genomes in a plasmid. Once the IC plasmid has been constructed, it could be transiently transformed into plants through micro-projectile bombardment (Bordat et al., 2015) or agroinfiltration (Sun et al., 2017). This approach was recently used in Pasin et al., (2017), who used Gibson-assembly to clone a full-length UCBSV genome into a pLX based plasmid vector. This UCBSV IC was used to agroinfiltrate N. benthamiana was able to generate infections.

3.3.2 Agroinfiltration of cassava with CBSV_Tanza IC

To enable the use of the CBSV_Tanza IC in resistance breeding, it should be able to efficiently cause standardised infections in cassava. Several attempts were made to agroinfiltrate cassava; it was found that mechanical damage combined with adding a surfactant to the agroinfiltration suspension could be used to successfully agroinfiltrate and inoculate cassava with the CBSV_Tanza IC. Similar modifications were also used to successfully agroinfiltrate rice, which is also recalcitrant to agroinfiltration (Andrieu et al., 2012; Yang et al., 2006). Although this technique was able to generate infections with the CBSV_Tanza IC, the inoculation efficiency was relatively low, as CBSV could be weakly detected by RT-PCR in 25 – 50% of agroinfiltrated plants. Therefore, agroinfiltration of cassava would need to be further improved to enable use of the CBSV_Tanza IC. A number of further adaptations to the agroinfiltration could be tested, including lowering the growth temperature of agroinfiltrated plants from 28°C to 20°C and increasing the concentration of acetosyringone from 150 to 200 µm, which may improve the efficiency of T-DNA transfer (Andrieu et al., 2012). Alternatively, the virG gene on the pCAMBIA0380 vector of the CBSV_Tanza IC could be mutated to enable constitutive expression of vir proteins, which has been shown to increase the efficiency of T-DNA transfer to recalcitrant plants (Johnson et al., 2001). It is possible that despite optimisation efforts, agroinfiltration may continue to be an ineffective method for inoculating cassava with U/CBSV ICs. Similar low inoculation rates of 10% were also found when agroinfiltrating cassava with
CMG viral ICs, compared with 100% inoculation rates when cassava plants were inoculated with the same ICs using micro-projectile bombardment (Ariyo et al., 2006). Therefore, future work should also consider using microprojectile bombardment as an inoculation method for U/CBSV ICs.

3.3.3 Use of CBSV_Tanza_GFP1/2 ICs to visualise CBSV replication and movement

The CBSV_Tanza IC was successfully modified through the insertion of GFP in two positions; between the Ham1 – CP regions and between the CP – 3’UTR regions. This is the first report of a CBSV IC being modified in this way. *N. benthamiana* infections with both CBSV_Tanza_GFP1/2 ICs enabled visualisation of CBSV Tanza replication and movement. These infections demonstrated that CBSV replicates in the epidermal and mesophyll cells of inoculated leaves by 5 dpi and moves systemically through the vascular system to the leaf lamina of upper systemic leaves by 15 dpi. Similarly, Saggaf et al., (2018) have recently used immune-histochemical staining to show that during cassava infections, CBSV primarily localizes to the phloem and adaxial and abaxial epidermal cells along leaf blades.

The timing of visual CBSV_Tanza_GFP1/2 movement in *N. benthamiana* is slightly slower than observed during GFP tagged PVX infections of *N. clevelandii* (Baulcombe et al., 1995) and GUS tagged TEV infections of *N. tabacum* (Dolja et al., 1992), which may be due inherent differences in viral movement and replication and/or infection differences in the three *Nicotiana* sp. Alternatively, the insertion of the GFP marker gene may have a greater detrimental effect on CBSV infections than for PVX or TEV. *N. benthamiana* infections with both the CBSV_Tanza_GFP1/2 ICs generated similar levels of GFP expression visualised using fluorescence and confocal microscopy. This suggests changing the genome position for the insertion of GFP from between the CP – Ham1 to CP - 3’UTR did not lead to higher levels of GFP expression or viral infectivity.

3.3.4 Stability of GFP expression in CBSV_Tanza_GFP1/2 infections

During *N. benthamiana* infections with CBSV_Tanza_GFP1/2, GFP fluorescence was not observed after 20 dpi and viral titers were extremely low compared with CBSV_Tanza infections. This suggests that the presence of GFP in the CBSV genome significantly reduces viral infectivity, which may be due to interference with viral genome replication, translation, polyprotein processing or viral movement (Scholthof et al., 1996). Although GFP was detected by wide-field and confocal microscopy in CBSV_Tanza_GFP1/2 infected *N. benthamiana* leaves at 5 – 15 dpi, extremely low viral titers were detected by qPCR at 15 dpi. This may indicate that
GFP is expressed through CBSV_Tanza_GFP1/2 translation during early infection and is then retained at 15 dpi, by which point viral genome replication has significantly reduced. Low viral titers during CBSV_Tanza_GFP1/2 infections may also explain the lack of symptom development and the unsuccessful attempts to use N. benthamiana material from CBSV_Tanza_GFP1/2 infections to mechanically back-inoculate N. benthamiana and cassava. Similar reductions in symptom severity and viral titers were also shown when the Lettuce mosaic virus (LMV) and TuMV ICs were tagged with either GUS or GFP (Beauchemin et al., 2005; German-Retana et al., 2000). In this study, viral transcripts containing GFP deletions were also detected in 25% of plants agroinfiltrated with CBSV_Tanza_GFP1/2. Previous studies have also reported the deletion of marker genes from plant viral ICs during infections (Arazi et al., 2001; Beauchemin et al., 2005; Dawson et al., 1989; Guo et al., 1998). Marker gene deletions are likely to occur during replication of the viral genome through recombination. Once a deletion has occurred the wild-type like viral genome will have a selective advantage over viral genomes containing marker genes and so deletants can increase in the viral population over time (Scholthof et al., 1996). It would be interesting to track CBSV_Tanza_GFP1/2 viral titers throughout infections, to determine whether transcripts containing GFP deletions increase later in infection. Ideally, marker genes should show stable expression throughout infections and multiple passages. Chapman et al., (2008) have previously reported that compared with GFP (=700 bp), the insertion of the shorter iLOV marker gene sequence (= 300 bp) into the Tobacco mosaic virus (TMV) IC enables increased systemic TMV movement during infections of N. tabacum. It was therefore hypothesized that inserting iLOV into the CBSV_Tanza IC would improve viral infectivity and produce infections more comparable to wild-type in terms of symptoms and viral titers. Unfortunately, efforts to construct CBSV_Tanza_iLOV were circumvented by challenges with sequence stability during cloning. Another approach to reduce interference of the marker gene with CBSV infectivity would be to insert a short six amino acid epitope sequence into the CBSV_Tanza IC; antibodies labelled with fluorescein dyes could then be added which bind to the epitope sequence and thereby enable visualisation of the presence of CBSV (Snapp, 2005). In addition to using smaller marker gene sequence, future attempts to improve marker gene expression during CBSV_Tanza infectious could involve inserting at the 5’ insertion site between the P1 - P3 peptides. This insertion position enables higher infectivity of marker tagged LMV (Bordat et al., 2015). Once these marker gene issues have been addressed, there would be many valuable applications for a marker gene tagged CBSV_Tanza IC. These include the identification of viral sequences required for key viral functions, such as replication and movement, whether CBSV and UCBSV co-infect the same
cells during mixed infections and the locations of U/CBSV replication and movement during cassava infections.

3.3.5 Use of the CBSV_Tanza_UCP IC to characterise sequence determinants

In this Chapter, the first chimeric U/CBSV IC was successfully constructed. The CP region was selected because a sequence alignment between the CBSV Tanza and UCBSV Kikombe CP sequences revealed low sequence homology in N' region, including the presence of extra sequence in the CBSV CP. This low sequence homology is typical for the N' and C' terminals of *Potyviridae* CP, which are exposed on the virion surface and interact with a wide range of viral and host partners (Allison et al., 1985; Makarov and Kalinina, 2016). Differences in CP sequences serve as symptom determinants between TMV, CMV and PepMV isolates (Duff-Farrier et al., 2015; Saito et al., 1987; Takeshita et al., 2001). To test whether differences in symptom development during CBSV and UCBSV infections are associated with the CP regions the chimeric CBSV_Tanza_UCP was constructed, consisting of a CBSV Tanza genome with a UCBSV Kikombe CP replacement. Unfortunately, due to the sequence instability of the UCBSV Kikombe, it was not possible to construct an IC with the reciprocal replacement consisting of the UCBSV Kikombe genome with the CBSV Tanza CP.

The CBSV_Tanza_UCP IC was able to cause systemic infections in *N. benthamiana* in both agroinfiltration and mechanical inoculation experiments. Whereas *N. benthamiana* plants infected with CBSV_Tanza develop severe systemic necrosis and plants infected with UCBSV_Kikombe develop mild mosaics (Nanyiti PhD thesis, 2016), plants infected with CBSV_Tanza_UCP develop mild necrosis. This indicates that the CBSV CP sequence may be associated with the development of severe necrosis during CBSV infections of *N. benthamiana*.

In terms of viral accumulation, CBSV_Tanza titers are higher at 7 dpi, peak at 14 dpi and then rapidly decrease at 21 – 28 dpi due to plant death, whereas titers during CBSV_Tanza_UCP infections are lower at 7 – 14 dpi and steadily increase at 21 – 28 dpi. This indicates that compared with the UCBSV Kikombe CP, the CBSV Tanza CP may enable increased viral replication and/or movement early in infection and this in turn may lead to more severe symptoms, earlier in infection. Alternatively, the reduction in CBSV_Tanza_UCP viral titers during early infections may be due to a reduced efficiency of proteolytic cleavage of the UCBSV CP from the CBSV Tanza polyprotein, reduced ability for the UCBSV CP to interact with CBSV Tanza proteins and/or to encapsidate the viral genome for movement.

CPs are required for the systemic movement of nearly all plant viruses (Ivanov and Mäkinen, 2012) and so it seems likely that the UCBSV Kikombe CP is able to trans-encapsidate the CBSV
Tanza genome during systemic infections with CBSV_Tanza_UCP. It is not known whether U/CBSV trans-encapsidation occurs during mixed U/CBSV cassava infections in the field, which are relatively common in some areas (Kathurima et al., 2016; Mbanzibwa et al., 2011b; Ogwok et al., 2014). If U/CBSV trans-encapsidation does occur in the field, there may be implications for synergistic interactions, including altered vector transmission specificity and/or efficiency.

3.3.6 Conclusions
To conclude, this Chapter has identified regions of the UCBSV Kikombe and CBSV Tanza genomes which show sequence instability during IC plasmid propagation in *E. coli*. Further work is required to improve the sequence stability of the UCBSV_Kikombe IC before it can be used. Propagation of the CBSV_Tanza IC in the *E. coli* strain ccdB was sufficiently stable and so this strain should be used for its long-term storage and propagation. Construction of the CBSV_Tanza_GFP1/2 ICs enabled visualisation of CBSV *in-planta* replication and movement to be visualised. Once optimised for marker gene stability and infectivity, these ICs should serve as an important tool for understanding CBSV infection mechanisms. Similarly, the construction of the CBSV_Tanza_UCP IC enabled an initial investigation into U/CBSV symptom determinants; it was discovered that the CBSV Tanza CP is not associated with necrosis development during CBSV infections of *N. benthamiana*. To further investigate the development of necrosis during CBSV infections of *N. benthamiana*, modifications were then performed on the Ham1 region of the CBSV_Tanza IC, as described in Chapter 4.
Chapter 4: Potential roles of U/CBSV Ham1 proteins during infection

4.1 Introduction

Cellular metabolism produces harmful by-products, including non-canonical nucleotides, which interfere with conventional Watson-Crick nucleotide base pairing and thereby cause miscoding mutations. Attempts to repair these mutations can lead to DNA strand breaks, recombinations and chromosomal abnormalities (Simone et al., 2013). Prokaryotes and eukaryotes have evolved pyrophosphohydrolase proteins, which specifically remove these non-canonical nucleotides from cellular pools of nucleotides and so reduce the likelihood of incorporation into nucleic acid (Galperin et al., 2006). U/CBSVs encode Ham1 proteins, with conserved pyrophosphohydrolase motifs (Mbanzibwa et al., 2009). Aside from in U/CBSV Ham1s, these motifs have recently been identified in one other virus: *Euphorbia ring spot virus* (Knierim et al., 2016). To date the functions of these viral Ham1 proteins during infection are unknown. In this Chapter the potential roles of U/CBSV Ham1 proteins during infection of *N. benthamiana* are investigated.

4.1.1 Non-canonical nucleotides and their deleterious effects

Cells contain canonical deoxyribose and ribose nucleotide triphosphates: adenosine triphosphate: (d)ATP, cytosine triphosphate: (d)CTP, guanosine triphosphate: (d)GTP, thymidine triphosphate: TTP and uridine triphosphate (UTP). Canonical deoxyribose nucleotides are incorporated into DNA during DNA replication and ribose nucleotide triphosphates are incorporated into RNA during transcription, guided by base-pairing. However, cellular nucleotide pools also contain relatively low levels of non-canonical nucleotides, such as inosine triphosphate: (d)ITP and xanthine triphosphate (XTP). Estimate of relative canonical and non-canonical nucleotides cellular concentrations were reported in Bennett et al., (2009), who found that in glucose-fed *E. coli*, ITP exists at a concentration of $2.1 \times 10^{-4}$ mol l$^{-1}$ compared with GTP, which exists at $4.9 \times 10^{-3}$ mol l$^{-1}$. In addition, non-canonical nucleotides are also incorporated into nucleic acid by DNA polymerases with much lower efficiencies than canonical nucleotides. Toshinori Suzuki et al., (1998) reported that dXTP is incorporated opposite cytosine in DNA at a rate of 0.28% min$^{-1}$ µm$^{-1}$, compared with 1506% min$^{-1}$ µm$^{-1}$ for dGTP. Non-canonical nucleotides (d)ITP and (d)XTP resemble (d)ATP and (d)GTP respectively, except that their nitrogenous bases contain keto groups in place of amine groups,
as shown in Fig. 4.1. When (d)ITP or (d)XTP are incorporated into nucleic acid, they can serve as a GTP analogues and predominately base pair with (d)CTP but can also pair with (d)TTP, leading to cytosine to thymidine transition mutations (Fig. 4.2) (Simone et al., 2013). If a transition mutation is non-synonymous and alters the encoded amino acid sequence, this can lead to the translation of dysfunctional proteins, which negatively affect cellular processes (Simone et al., 2013). The presence of inosine and xanthine in DNA is recognised by components of the DNA repair pathway, such as endonucleases which excise the non-canonical nucleotides from DNA and thus create a single-strand break. If the DNA replication fork that is replicating in the other direction encounters the break, the replication fork collapses, resulting in a double-strand break. This can lead to deletions, translocations, recombinations and other forms of chromosomal instability (Budke and Kuzminov, 2010; Burgis et al., 2003; Simone et al., 2013). The presence of xanthine in DNA can also cause DNA polymerases to stall, inducing repair mechanisms that can lead to further mutagenesis (Simone et al., 2013). The presence of non-canonical nucleotides in RNA structural complexes such as ribosome or tRNA can also negatively affect translation and potentially cause further errors (Simone et al., 2013). Finally, many host proteins that are involved signaling are activated by ATP and GTP; it has been reported that ITP and XTP could inhibit these proteins (Simone et al., 2013).

**Figure 4.1:** Molecular structures of canonical nucleotides adenosine triphosphate (ATP) and guanosine triphosphate (GTP) and non-canonical nucleotides inosine triphosphate (ITP) and xanthine triphosphate (XTP).
4.1.1.3 Generation of non-canonical nucleotides

Non-canonical nucleotides such as (d)ITP and (d)XTP are produced during purine salvage or de novo synthesis particularly when cells are under oxidative stress, according to the schematic in Fig. 4.3 (Simone et al., 2013). During purine synthesis, inosine monophosphate (IMP) is the first intermediate, which can be used in the synthesis of adenosine monophosphate (AMP) or guanosine monophosphate (GMP). To produce AMP, host enzymes replace the 6-keto group in the nitrogenous base of IMP with an amine group. To generate GMP, host enzymes first oxidise position 2 in the nitrogenous base of IMP to produce a 2-keto group to form the intermediate: xanthine monophosphate (XMP). The 2-keto group in XMP is then replaced with an amine group to generate GMP. AMP and GMP are then sequentially phosphorylated by host kinases and reduced to form ATP and GTP respectively (Simone et al., 2013). Deoxy-nucleotides are formed through reduction of the 2′-hydroxyl group on the ribose of nucleoside-diphosphates, such as ADP and GDP by ribonucleotide reductase enzymes to form dADP and dGDP which are then phosphorylated to form dATP and dGTP. Although this is the predominant metabolic pathway, intermediates: IMP and XMP are also slowly phosphorylated to form (d)ITP and (d)XTP respectively (Agarwal et al., 1971; Simone et al., 2013). Reactive oxygen species can also oxidatively deaminate the (d)ATP 6-amine group and the (d)GTP 2-amine group to keto groups to generate (d)ITP and (d)XTP respectively (Simone et al., 2013), hence the mutagenic effect of oxidative stress.
Figure 4.3: Purine metabolism and generation of non-canonical nucleotides inosine triphosphate (ITP) and xanthine triphosphate (XTP). ITP can be generated through sequential phosphorylation of inosine monophosphate (IMP) or oxidative deamination of adenosine triphosphate (ATP). Whereas XTP is generated through sequential phosphorylation of xanthine monophosphate (XMP) or deamination of guanosine triphosphate (GTP) (Simone et al., 2013).
4.1.2 House-cleaning NTPase proteins

To protect against the deleterious effects of non-canonical nucleotides, eukaryotes and prokaryotes have evolved ‘house-cleaning’ enzymes, which selectively remove non-canonical nucleotides from the cellular nucleotide pool. There are four protein families which are reported to perform this nucleotide ‘house-cleaning’ function (Galperin et al., 2006):

- Nudix protein superfamily: includes mutT enzymes, which hydrolyse oxidised, mutagenic nucleotides such as 8-oxo-dGTP, 2-oxo-dATP and 8-oxo-dATP (Maki and Sekiguchi, 1992; Mishima et al., 2004).
- Trimeric dUTPases: encoded in virtually every bacterial, archaeal or eukaryotic genome; they function to selectively hydrolyse the mutagenic nucleotide dUTP (Larsson et al., 1996).
- α-NTP pyrophosphohydrolases: examples include dimeric dUTPases that are encoded by several bacterial and eukaryotic species, including several pathogens. Another example is the dCTPase protein encoded by the T4 coliphage, which is active against dCTP in addition to dUTP (Moroz et al., 2004; Snyder et al., 1976). Finally this superfamily includes the MazG-related proteins, which are predicted to act on non-canonical nucleotides such as 2-oxo-(d)ATP; a mutagenic, oxidative derivate of dATP (Moroz et al., 2005).
- ITPases (ITPA): encoded by bacteria, archaea or eukaryotes; they function to selectively hydrolyse non-canonical mutagenic nucleotides including (d)ITP and (d)XTP (Simone et al., 2013). ITPA proteins encode conserved pyrophosphohydrolase motifs, which are also found in U/CBSV Ham1 proteins; investigations into the potential functions of these proteins are the focus of Chapters 4 and 5.

4.1.2.1 ITPases function to reduce mutations rates

ITPA proteins hydrolyse the β-pyrophosphate bond in non-canonical nucleotide triphosphates, such as (d)ITP and (d)XTP to release a pyrophosphate (PPi) molecule and a nucleotide monophosphate. Nucleotide monophosphates such as IMP cannot be incorporated into nucleic acid and so are returned to the nucleotide metabolic pathway (Galperin et al., 2006). ITPA proteins contain highly conserved pyrophosphohydrolase motifs, including the Serine-Histidine-Arginine (SHR) motif, which in the human ITPA protein is involved with nucleotide binding and substrate specificity (Gall et al., 2013; Stenmark et al., 2007). Characterised ITPA enzymatic activities and structures are described in more detail in Chapter 5. ITPA homologs in E. coli, yeast and human are shown to protect against non-canonical purine nucleotide induced
mutations (Simone et al., 2013). Plant ITPA proteins remain poorly understood and are yet to be functionally characterised.

4.1.2.2 ITPA mutagenic protection in *E. coli*

In *E. coli* when the gene encoding the ITPA homolog *rec-dependent growth B* (Rdgb) protein is mutated, there are higher cellular levels of inosine and xanthine nucleotides, which leads to increased inter-chromosomal rearrangements and induction of SOS DNA repair mechanisms (Clyman and Cunningham, 1987). In *E. coli*, when the gene encoding the SOS DNA repair *recA* protein is also mutated, the *rdgB recA* double mutant strain shows a total loss of viability, as the RecA protein cannot repair DNA damage caused by increased levels of inosine and xanthine nucleotides (Budke and Kuzminov, 2006).

4.1.2.3 ITPA mutagenic protection in yeast

In yeast, mutations in the ITPA homolog gene *ham1* are hypersensitive to the mutagenic effects of the purine base analog 6-N-hydroxylaminopurine (HAP) (Pavlov, 1986). It is proposed that HAP is taken up from the culture media by yeast, phosphorylated and reduced by cellular enzymes to produce dHAP triphosphate (dHAPTP) (Galperin et al., 2006). To protect against dHAPTP incorporation into DNA, the yeast *Ham1* protein hydrolyses dHAPTP to dHAP mono phosphate. The yeast *Ham1* protein can also provide protection against the pyrimidine analog 5-Flourouracil, demonstrating a broader specificity that is not strictly limited to purine analogs (Carlsson et al., 2013).

4.1.2.4 ITPA mutagenic protection in mammals

In mice, when both alleles of the *Itpa* gene are knocked out in *Itpa* −/− mutant lines; mice demonstrate severe losses of viability and growth defects, especially in the heart and brain where ITPA expression is normally the highest (Behmanesh et al., 2005, 2009). *Itpa* −/− mice also have much higher levels of ITP in their erythrocyte nucleotide pool compared with wild-type mice (Behmanesh et al., 2009). In human cell lines, a sequence variant in the *Itpa* gene, which results in a Pro32Thr substitution, is associated with higher cellular levels of ITP and increased DNA strand breaks (Waisertreiger et al., 2010). Zamzami et al., (2013) also found that patients who are deficient in functional ITPA protein have a higher a mutation rate in their mitochondrial DNA.
4.1.3 Viral Ham1 proteins

There are only three plant viruses reported to encode Ham1-like proteins: CBSV, UCBSV (Mbanzibwa et al., 2009) and EuRV (Knierim et al., 2016). These three viruses belong to the *Potyviridae* family and naturally infect plant species in the *Euphorbiaceae* family. In CBSV, UCBSV and EuRV the Ham1 protein sequences are located between the NiB and coat protein sequences.

4.1.3.1 *Euphorbia ringspot virus* Ham1 protein

*Euphorbia ringspot virus* (EuRV) infects ornamental plants such as *Euphorbia milii*; it causes ring-shaped chlorotic spots on leaves, leaf and flower deformations and stunting (Bode and Lesemann, 1976). EuRV has been reported on ornamental plants in Australia, Thailand, the USA and Venezuela (Gibbs et al., 2008; Guaragna et al., 2004; Marys and Romano, 2011). *De novo* genome sequencing revealed that EuRV encodes a Ham1 protein, which typically shares 39% and 40% amino acid sequence identity with CBSV and UCBSV Ham1 sequences respectively (Knierim et al., 2016). EuRV is the only reported member of the *Potyvirus* genus to encode a Ham1 protein and as with ipomoviruses U/CBSV Ham1 proteins, its function(s) are also uncharacterised.

4.1.3.2 U/CBSV Ham1 proteins

Mbanzibwa et al., (2009) first identified the presence of Ham1-like peptides which contain conserved ITPA pyrophosphohydrolase amino acid motifs in the genomes of U/CBSVs isolates. The functions of U/CBSV Ham1 proteins are currently unknown. As U/CBSVs do not encode HC-Pro, which functions as a silencing suppressor in other *Potyviridae*, Mbanzibwa et al., (2009) performed an experiment to determine whether silencing suppressor activity could be detected in the UCBSV P1, P3 or Ham1 proteins. This demonstrated that the UCBSV P1 protein appears to have silencing suppression activity, whereas suppressor activity was not detected for P3 or Ham1. Given that no silencing suppressor activity could be detected for UCBSV Ham1 protein, they speculated that the presence of conserved ITPA pyrophosphohydrolase motifs, suggests that they may have this enzymatic activity and perhaps function to intercept non-canonical nucleotides and thereby reduce viral mutation rates (Mbanzibwa et al., 2009). As described in Chapter 1, section 1.9, RNA viruses typically have finely tuned mutation rates that exist close to a critical threshold to enable maximum viral genome diversity and adaptability, whilst conserving sequences that are required for multiple essential functions (Holmes, 2003).
Therefore perhaps, U/CBSVs may have acquired Ham1 sequences to further fine-tune their viral mutation rates.

Outside of their conserved pyrophosphohydrolase ITPA motifs, CBSV and UCBSV Ham1 proteins share a relatively low level of amino acid sequence similarity, at typically around 47%, compared with 85 – 89% for 6K1 proteins (Mbanzibwa et al., 2009; Winter et al., 2010). This low level of sequence similarity has led to the suggestion that perhaps CBSV and UCBSV Ham1 sequences may serve as determinants for differences in symptom development and viral titres that are characteristic for CBSV and UCBSV infections (Winter et al., 2010). Meanwhile in terms of evolution, the origin of U/CBSV Ham1 sequences is also unknown. U/CBSV Ham1 sequences may have been acquired from a plant host during infection. A single Ham1-like sequence may have integrated into an ancestral U/CBSV genome and then undergone significant divergence, or two separate Ham1-like integrations may have occurred in CBSV and UCBSV lineages (Monger et al., 2010).

4.1.4 Insights into viral diversity using deep sequencing

Next generation deep sequencing can be used to gain insights into the diversity of viral genomes during infections. RT-PCR is performed on infected plant material, and then a single amplicon is used to generate millions of sequencing reads, corresponding to the ‘cloud’ of viral variants that were present in the infected material. The sequence reads are aligned to a reference sequence and an algorithm is used to detect single nucleotide variants (SNVs) which occur at very low frequencies (Marston et al., 2013). This technique has demonstrated how HIV-1 diversity early in infection facilitates immune escape (Fischer et al., 2010). When detecting low frequency SNVs, it is important to reduce and account for technical errors introduced through processes such as PCR and UV-exposure during gel electrophoresis visualisation. PCR errors include single base pair substitutions, indels and recombination events (Potapov and Ong, 2017). To reduce PCR-associated errors, a high-fidelity polymerase such as Phusion can be used, which has a 50X lower error rate than Taq polymerase due to enhanced proof-reading exonuclease activity (Böhlke et al., 2000). Mutations are also introduced during PCR thermo-cycling, as high temperatures damage DNA and so reducing the number of thermo-cycles can reduce this artificial source of mutation (Potapov and Ong, 2017). In addition, when performing RT-PCR it is also important to use RNA with high RNA integrity, as this can significantly improve the quality of deep sequencing results (Gallego Romero et al., 2014; Kukurba and Montgomery, 2015).
4.1.5 Previous work on U/CBSV Ham1s at the UoB

4.1.5.1 Investigations to detect U/CBSV Ham1 mutagenic protection
To determine whether U/CBSV Ham1 proteins can protect yeast against the mutagenic nucleotide 5-Flurouracil, resistance assays were performed by Bunawan (PhD thesis, 2016). Wild-type yeast cells were transformed with pYES2 plasmids containing either: CBSV Tanza Ham1, CBSV Nampula Ham1, UCBSV Kikombe Ham1 or yeast Ham1 sequences. Over-expression of the Ham1 proteins was induced through addition of galactose to yeast culture media. Only yeast transformed with the pYES2-yeast Ham1 plasmid grew on plates and liquid media containing 10 µg/ml 5-Flurouracil, indicating resistance. By contrast, transformation with CBSV or UCBSV Ham1 pYES2 plasmids did not result in resistance and transformants were unable to grow in the presence of 10 µg/ml 5-Flururacil. This suggests that unlike the yeast Ham1 protein, CBSV and UCBSV Ham1 proteins are unable to protect against the mutagenic pyrimidine analog 5-Flururacil.

4.1.5.2 Effect of CBSV Ham1 on viral mutation rate
To determine whether the CBSV Ham1 protein does reduce viral mutation rates, Pablo-Rodriguez (PhD thesis, 2017) performed an experiment to investigate whether the transgenic expression of CBSV Nampula Ham1 sequence (NCBI: MG019914) in N. tabacum affected PVY or TMV mutation rates. Deep sequencing analysis detected no significant differences in the number of SNVs in PVY or TMV RT-PCR amplicons from infections of three Ham1 transgenic N. tabacum lines compared to wild-type infections at 14 dpi. This indicates that the transgenic expression of CBSV Nampula Ham1 protein in N. tabacum did not affect the mutation rate of PVY or TMV.

4.1.5.3 CBSV Ham1 association with symptom development and viral accumulation
To identify whether the CBSV Ham1 protein is essential for infection of N. benthamiana, Pablo-Rodriguez (PhD thesis, 2017) deleted the Ham1 sequence from the CBSV_Tanza IC to produce the CBSV_Tanza Ham1 knockout IC (CBSV_ HKO). Both CBSV_Tanza and CBSV_ HKO were then used to agroinfiltrate N. benthamiana. The CBSV_ HKO IC could infect N. benthamiana, however dramatic changes in symptom development and viral accumulation were observed. Whereas infections with CBSV_Tanza develop severe necrosis, chlorosis, stunting and plant death, infections with CBSV_HKO develop strong systemic leaf curling, chlorotic mottling, stunting and lack necrosis (Fig. 4.4). This demonstrates that the CBSV Ham1 sequence is not essential for N. benthamiana infection and is associated with necrosis development. Pablo-Rodriguez (PhD thesis, 2017) also demonstrated that viral titres during CBSV_HKO infections
were lower at 10 dpi compared to CBSV_Tanza. By 15 – 20 dpi, titres in CBSV_HKO infections are relatively high, whereas titers in CBSV_Tanza infections are low due to plant death (Fig. 4.5). This indicated that the CBSV Ham1 protein is also associated with relatively high levels of CBSV accumulation during early infection.

![Symptom development during CBSV_Tanza and CBSV_HKO infections of N. benthamiana at 15 dpi. CBSV_Tanza infections develop necrosis, chlorosis and stunting (A), whereas CBSV_HKO infections develop leaf curling and mottling (B), compared to a symptomless non-inoculated plants (C) from Pablo-Rodriguez (PhD thesis, 2017).](image)

![Quantification of CBSV titres during N. benthamiana infections with CBSV_Tanza and CBSV_HKO by TAS-ELISA. Viral titres at 10 dpi are higher in CBSV_Tanza infections, whereas at 15 – 20 dpi titres are higher in CBSV_HKO infections and low in CBSV_Tanza infections. This indicates that CBSV Ham1 is involved with relative high levels of CBSV accumulation during early infection from Pablo-Rodriguez (PhD thesis, 2017).](chart)

Figure 4.4: Symptom development during CBSV_Tanza and CBSV_HKO infections of N. benthamiana at 15 dpi. CBSV_Tanza infections develop necrosis, chlorosis and stunting (A), whereas CBSV_HKO infections develop leaf curling and mottling (B), compared to a symptomless non-inoculated plants (C) from Pablo-Rodriguez (PhD thesis, 2017).

Figure 4.5: Quantification of CBSV titres during N. benthamiana infections with CBSV_Tanza and CBSV_HKO by TAS-ELISA. Viral titres at 10 dpi are higher in CBSV_Tanza infections, whereas at 15 – 20 dpi titres are higher in CBSV_HKO infections and low in CBSV_Tanza infections. This indicates that CBSV Ham1 is involved with relative high levels of CBSV accumulation during early infection from Pablo-Rodriguez (PhD thesis, 2017).
4.1.6 Hypotheses

It was hypothesised that conserved pyrophosphohydrolase motifs found in CBSV and UCBSV Ham1 sequences may function to hydrolyse non-canonical nucleotides and thereby specifically reduce U/CBSV mutation rates. It is possible that during infection, plant hosts may ‘shut-off’ their own ITPA expression to increase cellular levels of non-canonical nucleotides and thereby increase viral mutation rates. If proven, this would be a fascinating insight into plant-virus interactions. It was speculated that perhaps CBSV and UCBSV Ham1 sequences were acquired through the integration of a plant ITPA mRNA sequence during infection and therefore viral Ham1 sequences should be related to plant ITPA homolog sequences. It was also hypothesised that the low levels of sequence identity between CBSV and UCBSV Ham1 sequences suggests that they may be associated with differential symptom development and viral accumulation during infection. Finally, it was speculated that the SHR pyrophosphohydrolase motif in the context of the CBSV Ham1 sequence be specifically associated with necrosis development during CBSV infections of *N. benthamiana*.

4.1.7 Aims

- To assess relationships between ITPA sequences, phylogenetic trees will be generated of ITPA homologs across five kingdoms and CBSV and UCBSV Ham1 sequences.

- To investigate the involvement of pyrophosphohydrolase SHR motif in CBSV symptom development and viral accumulation, the SHR motif in the CBSV_Tanza IC will be mutated to generate the mutant: CBSV_mutHam IC. *N. benthamiana* will then be agroinfiltrated with CBSV_Tanza and CBSV_mutHam ICs and infections compared.

- To investigate whether CBSV and UCBSV Ham1 sequences are associated with differential symptom development and viral accumulation, the chimeric CBSV_UHam IC will be constructed consisting of the CBSV Tanza IC with a UCBSV Kikombe Ham1 replacement. *N. benthamiana* will then be agroinfiltrated with CBSV_Tanza and CBSV_UHam and infections compared.

- To investigate whether CBSV Ham1 has an effect of CBSV mutation rate, a deep sequencing experiment will be performed to gain insights into CBSV genome diversity during *N. benthamiana* infections with CBSV_Tanza and CBSV_HKO, which lacks the Ham1 sequence.
4.2 Results

4.2.1 Bioinformatic analysis of U/CBSV Ham1 sequences

4.2.1.1 Pyrophosphohydrolase motifs in CBSV Tanza and UCBSV Kikombe Ham1 sequences

To identify the presence of conserved pyrophosphohydrolase motifs in CBSV Tanza and UCBSV Kikombe Ham1 amino acid sequences, their Ham1 sequences were aligned with ITPA homologs (Fig. 4.6). This identified conserved pyrophosphohydrolase motifs, including the SHR motif, which in the human ITPA protein is involved with nucleotide binding and substrate specificity (Gall et al., 2013; Stenmark et al., 2007). This motif was found at amino acid positions: Ser192, His193, Arg194 in the both the CBSV Tanza and UCBSV Kikombe Ham1 sequences.

4.2.1.2 Phylogenetic relationships between U/CBSV Ham1s and ITPA homologs

To identify sequence relationships between U/CBSV Ham1s and potential ITPA homologs, the CBSV Tanza and UCBSV Kikombe Ham1 sequences were used to TblastX search the NCBI database. For each respective blast search, the top 5 protein sequences with the lowest e-values are shown in Appendix 4.1. These sequences along with ITPA sequences from key plant, animal, fungal, animal, bacterial and archaeal species were aligned in ClustalW. A phylogenetic tree was generated using the Maximum Likelihood method and the Whelan and Goldman, (2001) model. The tree indicates that the CBSV Tanza, UCBSV Kikombe Ham1 sequences are not highly related to the plant, animal, fungal, bacterial or archaeal ITPA sequences that were used in this analysis. This is in contrast to the EuRV Ham1 sequences which is highly related to the bacterial RdgB sequence from E. coli. The tree also shows that the CBSV and UCBSV Ham1 sequences do not share a strong phylogenetic relationship with each other.

To investigate the relationships between the CBSV Tanza and UCBSV Kikombe Ham1 sequences with other U/CBSV Ham1s a phylogenetic tree was generated based on their nucleotide sequences (Fig. 4.8). U/CBSV Ham1 sequences were obtained from the Description of Plant Virus database. The tree was generated using nucleotide sequences and inferred using the Maximum Likelihood method based on the Jukes and Cantor, (1969) model. The tree shows that Ham1 sequences form two separate clades, according to the two separate UCBSV and CBSV species. The ability to define CBSV and UCBSV isolates as two separate species using only their Ham1 sequences further demonstrates the low sequence similarity of CBSV and UCBSV Ham1 proteins.
<table>
<thead>
<tr>
<th>CBSV_Tanza</th>
<th>1 VVDRSQPSNVAKREEEVTSKIRMGEAPIETVI1IAQLKKEPLC-P-GPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCBSV_Kikombe</td>
<td>1 TDKLGRGKPELRISEHDGVPQMKFFVTVNLGKALKYSL-GIA</td>
</tr>
<tr>
<td>EuV</td>
<td>1 ALED---------------------------------------------</td>
</tr>
<tr>
<td>H.sapiens</td>
<td>1 MAAS---------------------------------------------</td>
</tr>
<tr>
<td>S.cerevisiae</td>
<td>1 M------------------------------------------------</td>
</tr>
<tr>
<td>A.thaliana</td>
<td>1 MAA---------------------------------------------</td>
</tr>
<tr>
<td>M.esculenta</td>
<td>1 -------------------------------------------------</td>
</tr>
<tr>
<td>J.curcas</td>
<td>1 -------------------------------------------------</td>
</tr>
</tbody>
</table>

| CBSV_Tanza     | 50 ------IP--IIRSVKPYEEVSTGEE-IEKESRVVAEVLVGGY--- |
| UCBSV_Kikombe | 50 ------SD--VIAKNIDPVECGTFV-EEVRAQPLYKNTSYY--- |
| EuV           | 33 ------IV-LQVTPLFHEVSTGQ-E-IMMCKIGQFKLOT---    |
| H.sapiens     | 31 ------FPCTLVAKILGVESEPD-E-SIQQPLFQVQVCGVY--- |
| S.cerevisiae  | 27 VDVDNDKTHIE LINEALELQDCTLNAALAKGQAVAGGOTK--- |
| A.thaliana    | 25 TDKLGRGKPELRISEHDGVPQMKFFVTVNLGKALKYSL-GIA     |
| M.esculenta   | 22 ------IP--LRSKLDLESE-P-SEPE-DE--SKEARLALQVDG--- |
| J.curcas      | 34 ------IP--LRSKIIIEE-IEP-A-CKEARLAKKEVGH---    |

| CBSV_Tanza     | 88 IDECDFDALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| UCBSV_Kikombe | 88 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| EuV           | 71 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| H.sapiens     | 71 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| S.cerevisiae  | 75 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| A.thaliana    | 75 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| M.esculenta   | 60 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |
| J.curcas      | 72 IDECFNALNGE-EILRNPMEFEGILEYKLVEPQNYKM--LCVFAP |

| CBSV_Tanza     | 138 MNKVGDGDFI-IFKVLPRHANGPNSFGPGPIPH-EP-LNRRB-- |
| UCBSV_Kikombe | 138 MNKVGDGDFI-IFKVLPRHANGPNSFGPGPIPH-EP-LNRRB-- |
| EuV           | 117 YDGKTMIE-IVEAKSNE-VEEOKHNGFGPMQID-SKQGQ--- |
| H.sapiens     | 121 STGDSQFQKLFRGSQLS-AGCQDFGCP-MCQ-DGYQEL--- |
| S.cerevisiae  | 125 ADSSR-GEYH-FFQGITSRK-SSRPTPFCGSSPFFDSHKG--- |
| A.thaliana    | 125 SRGPGAEFL-TRKLMPS-AGRPTDFGSPFVDP-DGYQEL--- |
| M.esculenta   | 110 ALDAESEFY-TLKTMPK-RERKFQGPIPH-EP-DGYQEL--- |
| J.curcas      | 122 ALDPQDSFPI-TFQKLMPS-AGRPTDFGSPFVDP-DGYQEL--- |

| CBSV_Tanza     | 185 TEKOLISHRFRALSLVRDFLKDSSYFSFAKGVDRDDFFIDV-Q |
| UCBSV_Kikombe | 185 TEKOLISHRFRALSLVRDFLKDSSYFSFAKGVDRDDFFIDV-Q |
| EuV           | 163 PLKROVSHRAAALKRLQEVLRKGRGTQTVRH--------------- |
| H.sapiens     | 169 KABROVSHJRPELEEQLYEGFSA---------------------A |
| S.cerevisiae  | 172 KDKOLISHRFRKAFQPKLEYQND---FQDTDDGT          |
| A.thaliana    | 172 KERKOLISHRKYSLAVKSHFKAEAY-------------------|
| M.esculenta   | 157 KERKOLISHRKALLSVKDFAPSSG---------------------|
| J.curcas      | 169 KERKOLISHRFRALAVHDFAEAG---------------------|

Figure 4.6: ClustalW alignment of CBSV Tanza and UCBSV Kikombe Ham1 amino acid sequences with ITPA homologs. Highly (100% identity) conserved motifs are highlighted in black, highly conserved motifs predicted to be involved with pyrophosphohydrolase activity are highlighted in yellow (Burgis & Cunningham, 2007). Protein sequences were obtained from the NCBI database and include: CBSV Tanza Ham1 (NCBI: MG570022), UCBSV Kikombe Ham1 (NCBI: KX753357.1), *Euphorbia ringspot virus* Ham1 (NCBI: NC_031339.1), *Homo sapiens* ITPA (NCBI: AAH10138), *Saccharomyces cerevisiae* Ham1 (NCBI: CA89597.1), *Arabidopsis thaliana* ITPA (NCBI: NP_001328955.1), *Manihot esculenta* hypothetical MANES protein (NCBI: OAY28191.1) and *Jatropha curcas* ITPA (NCBI: XP_012077670.1).
Figure 4.7: Phylogenetic tree of relationships between CBSV Tanza (blue circle) and UCBSV Kikombe Ham1 (red circle) with predicted ITPA homolog amino acid sequences. The tree shows the CBSV Tanza and UCBSV Kikombe Ham1 sequences are not highly related to plant, fungal, animal, archaean or bacterial sequences, in contrast to the EuRV Ham1 sequence, which is highly related to the E. coli RdgB sequence. The tree also shows that the CBSV Tanza and UCBSV Kikombe Ham1 sequences are highly divergent. The tree was inferred using the Maximum Likelihood method based on the Whelan and Goldman model (Whelan et al., 2001) with 250 bootstrap replicates. The tree with the highest log likelihood (−4560.7312) is shown. Bootstrap values are provided for each branch. The tree is rooted using the outgroup sequence from M. jannaschii. Sequences were obtained from NCBI databases; accession numbers for each sequence are provided in the tree. The tree is drawn to scale where 1 = unit of evolutionary distance in terms of number of amino acids substitutions per site. Analyses were conducted in MEGA6.
Figure 4.8: Phylogenetic tree of 19 CBSV and 29 UCBSV Ham1 nucleotide sequences inferred using the Maximum Likelihood method based on the Jukes-Cantor model (Jukes et al., 1996) with 250 bootstrap replicates. The tree shows that CBSV and UCBSV Ham1 sequences form two separate clades, which contain the CBSV Tanza (blue circle) and UCBSV Kikombe (red circle) Ham1 sequences. The tree is drawn to scale where 0.5 = unit of evolutionary distance in terms of number of nucleotide substitutions per site. Analyses were conducted in MEGA6. Bootstrap values are shown for each branch. The tree was rooted using the outgroup EuRV Ham1 sequence.
4.2.2 Construction of CBSV_mutHam IC containing mutated pyrophosphohydrolase motif

To investigate the role of pyrophosphohydrolase motifs in CBSV infection, the SHR motif was targeted for mutagenesis in the CBSV_Tanza IC. Gall et al., (2013) have previously found that mutating the SHR motif to Serine-Alanine-Alanine (SAA) in the human ITPA protein resulted in a loss in ITPase activity. It is expected that mutating the positively charged histidine and arginine amino acids in the SHR motif to neutrally charged alanine alters electro-static interactions with nucleotide substrates and thereby abolishes pyrophosphohydrolase activity. The CBSV_mutHam IC clone was constructed through yeast homologous recombination. The cloning plan involved the PCR amplification of two overlapping fragments: fragment 1 and 2 with primers encoding the SHR to SAA mutation and were designed to be recombined into the CBSV Tanza IC yeast through homologous recombination of the overlapping sequences to form the mutant IC: CBSV_mutHam, according to the schematic in Fig. 4.9.

Stage 1: digest
Stage 2: yeast recombination
Stage 3: rescue into E. coli

Figure 4.9: Schematic for the construction of the CBSV_mutHam IC through homologous yeast recombination. The CBSV Tanza FL IC was digested with Bam H1, which cuts at positions 7284 and 8961 bp to generate a 18.6 Kb fragment and a 1.7 Kb fragment (stage 1). The 18.6 Kb fragment and PCR fragments: F1 and F2 were then transformed into yeast and overlapping sequences were recombined (stage 2) to form the mutant IC: CBSV_mutHam consisting of the CBSV Tanza with a mutated Ham1 sequence (stage 3).
4.2.2.1 Cloning design for the construction of CBSV_mutHam

The cloning design for the construction of CBSV_mutHam IC involved the amplification of PCR fragments 1 and 2, which encode the SAA mutation. The PCR fragments were designed to be recombined in yeast, according to the schematic in Fig. 4.10. The primers used to amplify fragments 1 and 2 are provided in Table 4.1. The PCR fragments were designed as follows.

- Fragment 1 (F1): at the 5’ end contains 30 bp of homologous sequence from before the first Bam H1 restriction site at position 7284 bp in the CBSV Tanza IC plasmid, it then encodes CBSV Tanza sequence from NiB – end of Ham1 with the SAA mutation and at 3’ end contains 30 bp of homologous sequence with fragment 2.
- Fragment 2 (F2): at the 5’ end contains 30 bp of homologous sequence with F1 sequence and encodes the SAA mutation, it then encodes CBSV Tanza coat protein - 3’UTR, and at the 3’ end contains 30 bp of homologous sequence with CBSV Tanza IC from after the second Bam H1 restriction site at position 8961 bp.

Figure 4.10: Schematic for PCR site directed mutagenesis of CBSV Tanza Ham1 SHR pyrophosphohydrolase motif to SAA. Two overlapping fragments (F1 and F2) were amplified using primers in Table 4.1, which encode the SAA mutation. The two fragments and Bam H1 digested CBSV Tanza IC were transformed into yeast and recombined to produce the CBSV_mutHam IC.

Table 4.1: Primers used to amplify F1 - F2 with nucleotide sequence encoding the SHR to SAA mutation (red).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’ – 3’</th>
<th>Size (bp)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>SHR_mut_F1_Fw</td>
<td>GAAATATAATGAAACCTGTTCGAGTGTTGGTTGGTAAAC</td>
<td>1916</td>
<td>NiB – Ham1</td>
</tr>
<tr>
<td>SHR_mut_F1_Rv</td>
<td>TCTTCTAAAAAGTCTCTCAGATGACACAGCCGAAAAAGCAGCATATCATATTTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHR_mut_F2_Fw</td>
<td>GAAGAATATGATATCTCTGCTGCCCTTCGGCTCTGTCATTAGTAGAGAGACTTTTTGAAGGAA</td>
<td>1810</td>
<td>Ham1 – 3’UTR</td>
</tr>
<tr>
<td>SHR_mut_F2_Rv</td>
<td>GGCGGAGCTGGTGTCGAGATATGATATCTCTGCTCTGTCATTAGTAGAGAGACTTTTTGAAGGAA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4.2.2.2 Cloning of CBSV_mutHam IC

To construct the CBSV_mutHam1 IC, fragments 1 and 2 were amplified by high-fidelity Phusion PCR using the CBSV Tanza IC as template. This generated PCR products with the expected sizes of 1916 bp for fragment 1 and 1810 bp for fragment 2, as shown in Fig. 4.11. Then the CBSV Tanza IC was digested with Bam H1, which cuts at positions: 7284 and 8961 bp and produces two fragments of 18.6 Kb and 1.7 Kb (Fig. 4.12). The linearized CBSV Tanza fragment (18.6 Kb) and PCR fragments 1 and 2 were gel purified and transformed into yeast according to Gietz et al., (2002). Ten individual yeast colonies on test plates were pooled and grown in YDSM liquid culture, plasmids were extracted and transformed into electrocompetent *E. coli* ccdB cells. However, this failed to generate transformant colonies and so the yeast plasmids were transformed into the *E. coli* strain TOP10, which is derived from the ccdB strain and was so predicted to be able to propagate the IC plasmid with similar levels of sequence stability. Transformation of TOP10 with recombinant yeast plasmids produced 50 colonies.

![Image of gel purification and restriction digest](image_url)

**Figure 4.11:** Gel purified PCR fragments used in the construction of CBSV_mutHam IC through homologous yeast recombination. Fragment 1 (1916 bp) encodes overlapping sequence with CBSV_Tanza IC at the 5’ end and F2 at the 3’ end, as well as the SAA mutation. Fragment 2 (1810 bp) encodes overlapping sequence with F1 at the 5’ end and the SAA mutation and overlapping sequence with CBSV Tanza IC at the 3’ end.

**Figure 4.12:** Restriction digest of CBSV_Tanza IC plasmid with Bam H1 generated an 18.6 Kb fragment and a 1.67 Kb fragment. The 18.6 Kb fragment consists of linearized CBSV_Tanza IC without the Nib-Ham1 sequence, this fragment was gel purified and used to construct CBSV_mutHam IC through homologous yeast recombination.
4.2.2.3 Verification of CBSV_mutHam IC construction

To confirm whether transformant *E. coli* contained the correct CBSV_mutHam IC, six colonies were cultured, and plasmids extracted and analysed. To verify IC plasmid integrity, the plasmids were digested with the restriction enzyme *Hind* III, which produced the expected band pattern and indicated that no significant sequence rearrangements had occurred (Fig. 4.13). To further confirm the integrity of the plasmids, PCR was performed targeting the six CBSV Tanza genome sections, with the primers provided in Chapter 3, Table 3.2. This amplified PCR products with the expected sizes for all six genome sections, including sections four and five, which were associated with CBSV Tanza genome sequence instability in Chapter 3 (Fig. 4.14). To confirm whether the plasmids contain a mutated SHR motif, three plasmids were sequenced using the primer: Seq_9, which targets the NIb-Ham1 region. All three plasmids were found to contain the mutagenized Ham1 sequence, encoding SAA instead of SHR (Appendix 4.2).

![Figure 4.13: Restriction digest of six plasmids extracted from *E. coli* with *Hind* III gave the expected sized bands of: 10271, 5602 and 4096 bp to indicate that no significant sequence rearrangements had occurred during the construction of the CBSV_mutHam IC.](image1)

![Figure 4.14: PCR targeting the unstable regions of the CBSV Tanza genome to indicate successful CBSV_mutHam IC construction. PCR was performed targeting unstable genome sections (Sec.) four and five with primers in Chapter 3, Table 3.2 This produced the expected sized fragments of 0.9 Kb for genome section four and 1.7 Kb for genome section from five of the CBSV_mutHam IC plasmids extracted from *E. coli* (lanes 1–4 and 6). This indicated that no significant sequence rearrangements had occurred in these genome regions during construction of the CBSV_mutHam.](image2)
4.2.3 Construction of CBSV_UHam with UCBSV Ham1 replacement

To test whether the CBSV Tanza and UCBSV Kikombe Ham1 proteins are associated with differential symptom development and viral accumulation during *N. benthamiana* infection, the chimeric CBSV_UHam IC was constructed consisting of the CBSV Tanza genome with a UCBSV Kikombe Ham1 sequence replacement. The CBSV_UHam IC was constructed through homologous yeast recombination. The cloning plan involved the PCR amplification of three overlapping fragments with primers in Table 4.2. These fragments, along with *Bam* H1 digested CBSV Tanza IC were designed to be recombined in yeast through homologous recombination of the overlapping sequences to form the chimeric CBSV_UHam IC, according to the schematic in Fig. 4.15.

Stage 1: digest

Stage 2: yeast recombination

Stage 3: rescue into *E. coli*

Figure 4.15: Schematic for the construction of the chimeric CBSV_UHam1 IC through homologous yeast recombination. CBSV Tanza FL IC was digested with *Bam* H1, which cuts at positions 7284 and 8961 bp to generate a 18.6 Kb and 1.7 Kb fragment (stage 1). The 18.6 Kb fragment and PCR fragments: F1, F2 and F3 were then transformed into yeast and overlapping sequences were recombined (stage 2) to form the chimeric CBSV_UHam IC consisting of CBSV Tanza genome with a UCBSV Kikombe Ham1 sequence replacement (stage 3).
4.2.3.1 Cloning design for construction of CBSV_UHam IC

The three PCR fragments used in the construction of the CBSV_UHam IC were amplified with the primers provided in Table 4.2. To ensure proteolytic release of the UCBSV Ham1 peptide from the CBSV Tanza polyprotein, cleavage sequences were designed so that CBSV Tanza Nib-Ham cleavage sequence: T-L-T-Y-V-V-D was maintained and fused to the 5’ end of the UCBSV Ham1 sequence, according to the schematic in Fig. 4.16. The genome structure for CBSV_UHam is shown in Fig. 4.17.

Table 4.2: Primers used to amplify three overlapping PCR fragments used to construct CBSV_UHam IC.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’ – 3’</th>
<th>Size (bp)</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCBSV_Ham1_F1_Fw</td>
<td>GAAATATAATGAACCTGTTCGAGTGGGGTTGGTAAAC</td>
<td>1336</td>
<td>CBSV Nib</td>
</tr>
<tr>
<td>UCBSV_Ham1_F1_Rv</td>
<td>CAGGCTTCTCTCTTCTCTCAAACTCTTTGTGTCACCACTTG TAATGCAATGTAACAT</td>
<td></td>
<td>UCBSV Ham1</td>
</tr>
<tr>
<td>UCBSV_Ham1_F2_Fw</td>
<td>ACAAAGAGTTGAGAGAAGAGAGAGAGGAGAGCTCTGAGAGATTGAGGCAAT</td>
<td>677</td>
<td>UCBSV Ham1</td>
</tr>
<tr>
<td>UCBSV_Ham1_F2_Rv</td>
<td>CTGCACATCAATTGTTAGCCACCTTTGCTTCTCTCTCTCTGTTTCAACCATC</td>
<td></td>
<td>UCBSV Ham1</td>
</tr>
<tr>
<td>UCBSV_Ham1_F3_Fw</td>
<td>GAAGGGGAAGGTGAGCTCTGCTCAATGATGATGAGGCAATGAGGAGGAGGAGGAGGAGATTG</td>
<td>1729</td>
<td>CBSV CP – 3’UTR</td>
</tr>
<tr>
<td>UCBSV_Ham1_F3_Rv</td>
<td>GGCTGGCTGTTGTCAGGATATATTGTGTCAGGGAAC</td>
<td></td>
<td>CBSV CP – 3’UTR</td>
</tr>
</tbody>
</table>

Figure 4.16: Schematic showing the replacement of CBSV Tanza Ham1 sequence (blue) with UCBSV Ham1 sequence (red) in the CBSV_UHam IC. To ensure proteolytic cleavage of UCBSV Ham1 sequence from the CBSV Tanza polyprotein the Nib – Ham1 protease cleavage sequence T-L-T-Y-V-V-D was fused to the start of the UCBSV Ham1 sequence: T-K-D.

Figure 4.17: Schematic of the CBSV_UHam IC, consisting of the CBSV Tanza genome (blue) with the UCBSV Ham1 (red) replacement.
4.2.3.2 PCR design for the construction of CBSV_UHam IC

The three PCR fragments used in the construction of the CBSV_UHam IC were designed as follows.

- **F1**: The 5' end contains 30 bp of homologous sequence from before the first *Bam* H1 restriction site at position 7284 bp in CBSV Tanza IC plasmid, it then encodes CBSV Tanza sequence to the end of NiB and contains 30 bp of F2 homologous sequence at the 3’ end.
- **F2**: Encodes the UCBSV Kikombe Ham1 sequence.
- **F3**: The 5’ end contains 30 bp of homologous sequence with F2, it then encodes CBSV Tanza coat protein - 3’UTR, and at the 3’ end there is 30 bp of homologous sequence with CBSV Tanza IC from after the second *Bam* H1 restriction site at position 8961 bp.

4.2.3.3 Cloning of CBSV_UHam IC

The chimeric CBSV_UHam IC was constructed through recombining three overlapping PCR fragments into CBSV Tanza IC through homologous yeast recombination. The three fragments were amplified by high-fidelity Phusion PCR, as shown in Fig. 4.18. The CBSV Tanza IC was digested with *Bam* H1 as previously shown in Fig. 4.12. Linearized CBSV Tanza (18.6 Kb) and three PCR fragments were gel purified and transformed into yeast according to Gietz et al., (2002). Yeast colonies grown in YDSM liquid culture, plasmids were extracted and transformed into electrocompetent *E. coli* ccdB cells. This failed to generate transformant colonies and so the yeast plasmids were transformed into the *E. coli* strain TOP10, as this strain was used to construct CBSV_mutHam. This transformation generated 50 transformant colonies.

4.2.3.4 Verification of CBSV_UHam IC construction

To confirm whether transformant *E. coli* contained the correct CBSV_UHam IC, six colonies were cultured, and plasmids extracted and analysed. To confirm the integrity of the plasmids, they were digested with *Hind* III, which produced the expected band pattern to indicate that no significant sequence rearrangements had occurred (Fig. 4.19). To further confirm successful construction of the CBSV_UHam IC, PCR was performed targeting the six CBSV Tanza genome sections, with the primers provided in Chapter 3, Table 3.2. This amplified products with the expected sizes for all six genome sections, indicating that no significant rearrangements had occurred. The correct sized PCR fragments also amplified to indicate successful Ham1 replacement (Fig. 4.20). Plasmids 1 and 2 were sequenced using the primer: Seq_9, which targets the UCBSV Ham1 replacement; both were found to contain the UCBSV Ham1 sequence (Appendix 4.3).
Figure 4.18: Gel purified PCR fragments used in the construction of the CBSV_UHam IC.

Figure 4.19: Restriction digest of three plasmids extracted from *E. coli* with *Hind* III. Plasmids in lanes 1, 2 and 6 produced the expected banding pattern of: 10271, 5602 and 4096 bp to indicate that no significant sequence rearrangements had occurred during the construction of CBSV_UHam IC.

Figure 4.20: PCR amplification of correct sized fragments targeting the UCBSV Ham1 replacement, from two plasmids extracted from TOP10 *E. coli* (lanes 1 – 2), indicating successful construction of CBSV_UHam IC. PCR 1 targets the CBSV Nb – UCBSV Ham1 region and was performed with primers: UCBSV_Ham1_F1_Fw and UCBSV_Ham1_F2_Rv (Table 4.2). PCR 2 targets UCBSV Ham1 – CBSV CP and was performed with UCBSV_Ham1_F2_Fw and UCBSV_Ham1_F3_Rv (Table 4.2). PCR fragments of the expected sizes of 2013 bp amplified for PCR 1 and 2406 bp for PCR 2 amplified from both plasmids. This indicated successful construction of the CBSV_UHam IC with UCBSV Ham1 replacement.
4.2.4 *N. benthamiana* infections with CBSV Tanza ICs containing Ham1 manipulations

*N. benthamiana* plants were then infected with the CBSV_mutHam and CBSV_UHam ICs to investigate the effects of the mutated pyrophosphohydrolase motif and UCBSV Ham1 replacement on symptom development and viral accumulation. To enable comparison of infections, plants were also agroinfiltrated with the CBSV_Tanza IC, containing the full-length genome and the CBSV_HKO IC, containing the Ham1 deletion. Descriptions of the ICs used in this experiment are provided in Table 4.3. *N. benthamiana* plants were infected with the ICs through agroinfiltration. The *A. tumefaciens* strain LBA 4404 was transformed with each IC and transformant colonies were cultured. To confirm that the cultures contained the ICs, PCR was performed, which confirmed the presence of ICs (Fig. 4.21). Ten *N. benthamiana* plants were agroinfiltrated with each IC. Plants were monitored for symptom development and systemic leaf material was sampled for subsequent viral detection and quantification.

Table 4.3: Infectious clones used to agroinfiltrate *N. benthamiana*.

<table>
<thead>
<tr>
<th>Infectious clone</th>
<th>Description</th>
<th>Constructed by</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV_Tanza</td>
<td>Full-length, unmodified CBSV Tanza genome</td>
<td>Duff-Farrier (unpublished)</td>
</tr>
<tr>
<td>CBSV_HKO</td>
<td>CBSV Tanza genome with Ham1 deletion</td>
<td>Pablo-Rodriguez, 2017</td>
</tr>
<tr>
<td>CBSV_UHam</td>
<td>CBSV Tanza with UCBSV Kikombe Ham1 replacement</td>
<td>This study</td>
</tr>
<tr>
<td>CBSV_mutHam</td>
<td>CBSV Tanza genome with pyrophosphohydrolase motif SHR mutated to SAA</td>
<td>This study</td>
</tr>
</tbody>
</table>

Figure 4.21: PCR confirmation that *A. tumefaciens* cultures contained ICs: CBSV_Tanza (lane 1), CBSV_mutHam (lane 2), CBSV_UHam (lane 3) and CBSV_HKO (lane 5). PCR targetting the Ham1 region was performed with the primers: CBSV_Ham1_KO_Fw (GGTATGCTGTGTTCTGGGCATCTGA) and CBSV_Ham1_KO_Rv (CGAGCTACTTGAGGAGCAGAGCC). Amplification of 963 bp occurred from cultures containing CBSV_Tanza, CBSV_mutHam and CBSV_UHam, whereas amplification of 285 bp occurred from the culture containing CBSV_HKO. Positive controls of CBSV_Tanza (lane +2) and CBSV_HKO (lane +1) IC plasmids are shown. No amplification occurred in the negative control (lane -).
4.2.4.1 Confirmation of N. benthamiana infections with CBSV Tanza ICs

To confirm that N. benthamiana plants were infected with each of the ICs, RT-PCR was performed on systemic leaf material of three N. benthamiana plants from each infection type at 10 dpi (Fig. 4.22). This confirmed that the agroinfiltrated N. benthamiana plants were infected. Weak RT-PCR amplification occurred from N. benthamiana plants infected with CBSV_mutHam, perhaps indicating low viral titres in these infections. RT-PCR targeting the entire Ham1 gene and amplicon sequencing on N. benthamiana plants infected with CBSV_mutHam, confirmed the presence of CBSV transcripts encoding the SAA Ham1 mutation (Appendix 4.4). Similarly, RT-PCR targeting the Ham1 region and amplicon sequencing confirmed the presence of CBSV transcripts containing the UCBSV Ham1 replacement in N. benthamiana plants infected with CBSV_UHam (Appendix 4.5).

![RT-PCR amplification of CBSV Tanza CP fragment (163 bp) to confirm infection of N. benthamiana plants with CBSV_Tanza (lanes 1 – 3), CBSV_HKO (lanes 4 – 6), CBSV_UHam (lanes 7 – 9) and CBSV_mutHam (lanes 10 – 12) at 10 dpi. No amplification occurred from a non-inoculated plant (lane -). PCR was performed with the primers: qPCR_CBSV_CP_Fw/Rv (Chapter 2, Table 2.1).](image)

4.2.4.2 Symptom development during N. benthamiana infections with CBSV Tanza ICs

N. benthamiana infected with the different ICs developed a range of symptoms as illustrated in Table 4.4. As previously shown, N. benthamiana infected with the CBSV_Tanza IC developed necrosis at 10 dpi and severe necrosis, chlorosis, stunting and plant death at 14 – 18 dpi. In contrast, infections with CBSV_HKO developed slight foliar mottling and leaf distortion of the upper leaves at 10 dpi and strong systemic leaf curling, chlorotic mottling and stunting at 14 – 18 dpi. Infections with CBSV_UHam developed similar leaf curling and mottling symptoms to CBSV_HKO but these symptoms were less severe and developed at 12 dpi rather than 10 dpi. Finally, infections with CBSV_mutHam developed no symptoms at 10 dpi. At 14 – 18 dpi, three of the ten plants agroinfiltrated with CBSV_mutHam developed symptoms similar to CBSV_HKO infections but with much milder systemic leaf curling and stunting. The remaining seven plants were asymptomatic. The symptom development described in the experiment was consistently observed across three independent repeat experiments.
Table 4.4: Symptom development during *N. benthamiana* infections with ICs: CBSV_Tanza, CBSV_HKO, CBSV_mutHam and CBSV_UHam.

<table>
<thead>
<tr>
<th>Infection type</th>
<th>Symptom photo at 18 dpi</th>
<th>Symptom description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-length, unmodified CBSV_Tanza genome (CBSV_Tanza)</td>
<td><img src="image1.png" alt="Symptom photo" /></td>
<td>10 dpi: Systemic mild necrosis, chlorosis and wilting. 14 dpi: Systemic severe necrosis, chlorosis and wilting. 18 dpi: High levels of systemic necrosis and chlorosis, stunting and plant death.</td>
</tr>
<tr>
<td>CBSV Tanza genome with Ham1 deletion (CBSV_HKO)</td>
<td><img src="image2.png" alt="Symptom photo" /></td>
<td>10 dpi: Slight foliar mottling and leaf distortion of upper leaves. 14 dpi: Strong foliar mottling and leaf distortion of upper leaves. 18 dpi: Strong systemic leaf curling, chlorotic mottling, stunting and no necrosis.</td>
</tr>
<tr>
<td>CBSV Tanza with UCBSV Kikombe Ham1 replacement (CBSV_UHam)</td>
<td><img src="image3.png" alt="Symptom photo" /></td>
<td>10 dpi: Distortion of upper systemic leaves. 14 dpi: Mild systemic leaf curling, chlorotic mottling, stunting but no necrosis. 18 dpi: Mild systemic leaf curling, chlorotic mottling, stunting and no necrosis.</td>
</tr>
<tr>
<td>CBSV Tanza genome with SHR mutated to SAA (CBSV_mutHam)</td>
<td><img src="image4.png" alt="Symptom photo" /></td>
<td>10 dpi: No symptoms. 14 dpi: Mild leaf curling of upper leaves. 18 dpi: Mild systemic leaf curling, chlorotic mottling, stunting and no necrosis.</td>
</tr>
<tr>
<td>Non-inoculated</td>
<td><img src="image5.png" alt="Symptom photo" /></td>
<td>No symptoms.</td>
</tr>
</tbody>
</table>
4.2.4.4 Quantification of viral titres during *N. benthamiana* infections by TAS-ELISA

To examine the effects of these CBSV Tanza Ham1 manipulations on viral accumulation, TAS-ELISA was performed on upper systemic leaf material sampled throughout infection. For each infection type, systemic leaf material was sampled from three plants at 5, 10, 12, 14 and 20 dpi. Relative viral titers in test samples were calculated as a percentage of viral titers in the positive control, consisting of CBSV infected material that was provided with TAS-ELISA (DSMZ) kit.

The TAS-ELISA results indicate that viral titres in *N. benthamiana* infected with CBSV_HKO are much higher at 12 dpi than titres in infections with CBSV_Tanza, CBSV_mutHam or CBSV_UHam (Fig. 4.23). Unfortunately, the OD values for most of the samples are below the cut-off threshold. Therefore TAS-ELISA was not sufficiently sensitive to accurately quantify viral titres in this experiment. The experiment was repeated and again, the OD values were too low to accurately quantify viral titres.

![Relative viral titre (%)](image)

**Figure 4.23:** TAS-ELISA quantification of CBSV titres in *N. benthamiana* infected with CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam. Relative titres are shown, which were calculated as a percentage of titre in the positive control supplied with the TAS-ELISA kit. Titres in CBSV_HKO infections were higher than CBSV_Tanza, CBSV_HKO, CBSV_mutHam and CBSV_UHam at 12 dpi. However, the OD values for most samples were below the cut-off threshold (35%) and so TAS-ELISA was not sufficiently sensitive to gain insights into viral accumulation in this experiment. Each result is the mean phosphate concentration from three replicate plant samples (n = 3) ± S.E.
4.2.4.4 Quantification of viral transcripts during *N. benthamiana* infections by qPCR

QPCR was performed to compare CBSV transcript abundance in *N. benthamiana* infections with the CBSV Tanza ICs, as this technique has a higher detection sensitivity than TAS-ELISA. The qPCR protocol described in Chapter 2, section 2.2.5.6 was followed. The CBSV coat protein transcript abundance was calculated relative to the expression of the *N. benthamiana* F-box gene (TAIR: At5g15710), which has previously been validated to show relatively stable expression during viral infection of *N. benthamiana* (Liu et al., 2012). The $2^{\Delta\Delta C_T}$ method was used calculate the relative abundance of the CBSV Tanza coat protein transcripts in infected plants throughout infection, according to Livak and Schmittgen, (2001). For each infection type, systemic leaf material was collected from three *N. benthamiana* plants at 7, 10, 14 and 18 dpi and systemic leaf material was sampled from three non-inoculated plants at 18 dpi. Primer sequences used in qPCR are provided in Chapter 2, Table 2.2.

Results from qPCR (Fig. 4.24) demonstrate that at 7 dpi, CBSV_Tanza transcripts are 10X higher than CBSV_HKO transcripts and 100X higher than CBSV_UHam transcripts, while CBSV_mutHam transcripts are extremely low. At 10 – 14 dpi, viral transcripts in CBSV_Tanza, CBSV_HKO and CBSV_UHam infections are at a similar level, while CBSV_mutHam transcripts remain low. By 18 dpi, CBSV_HKO transcripts increase dramatically to levels that are 10,000X higher than CBSV_Tanza. Viral transcripts in CBSV_UHam infections appear to increase from 7 - 10 dpi and remain relatively stable throughout infection, whereas viral transcripts remain low and steadily increase throughout CBSV_mutHam infections.

These results suggest that during early infection (7 dpi) CBSV_Tanza accumulates to much higher levels than in infections with viral transcripts that lack Ham1 (CBSV_HKO), have a mutated Ham1 (CBSV_mutHam) or have a UCBSV Ham1 replacement (CBSV_UHam). However, by 10 dpi, the lack of Ham1 (CBSV_HKO) or UCBSV Ham1 replacement (CBSV_UHam) does not significantly affect viral accumulation, compared with unmodified CBSV_Tanza infections. At 18 dpi, the reduction in CBSV_Tanza transcript abundance is likely to be due to the development of severe necrosis that limits viral accumulation. The ability for CBSV_HKO transcripts to accumulate to such high levels at 18 dpi is interesting. It is possible that as CBSV_HKO has a shorter genome sequence that can replicate its genome faster. Alternatively, perhaps CBSV_Tanza transcripts would also accumulate to similar levels but this is prevented by the induction of necrosis. The ability for CBSV_UHam to accumulate to similar levels to CBSV_Tanza may suggest that the UCBSV Ham1 is able to complement the loss of the CBSV Ham1 sequence. Finally mutating the SHR pyrophosphohydrolase in CBSV_mutHam has a large impact on viral accumulation, which could be due to: a loss or reduction in Ham1.
pyrophosphohydrolase activity or the mutated Ham1 protein could have a detrimental effect on viral processes. Overall these results suggest that CBSV Ham1 is involved with viral accumulation during early infection of *N. benthamiana*.

4.2.5 Expression of *N. benthamiana* ITPA throughout infection with CBSV Tanza ICs

It was hypothesised that perhaps plant hosts reduce expression of their ITPA genes in response to CBSV infection to thereby increase the viral mutation rate above a critical threshold that causes viral error catastrophe. To test this qPCR analysis was then performed on the same samples as section 4.2.6.4, to quantify the expression of the *N. benthamiana* ITPA gene throughout CBSV infection. To identify the *N. benthamiana* ITPA gene, the predicted *N. tabacum* ITPA sequence (NCBI: XM_016586575.1) was used to BLAST the draft *N. benthamiana* genome sequence on the SolGenomicsNetwork website. This identified a predicted ITPA gene with the sequence ID: Niben101Scf18106g00001.1, that shares 93% sequence identity with the *N. tabacum* sequence. The qPCR primers were designed to target the *N. benthamiana* ITPA sequence and are provided in Chapter 2, Table 2.2. The $2^{\Delta\Delta Ct}$ method was used to calculate the *N. benthamiana* ITPA transcript abundance relative to the F-box transcript abundance (Livak and Schmittgen, 2001). For each infection type, systemic leaf material was sampled from three plants at 7, 10, 14 and 18 dpi and from non-inoculated plants at 18 dpi. QPCR was performed on the samples used in section 4.2.6.4. Primer sequences used in qPCR are provided.
in Chapter 2, Table 2.2. The results are shown in Fig. 4.25. This appears to show that the expression of the predicted *N. benthamiana* ITPA gene remains relatively stable throughout infection with all four of the CBSV ICs until 18 dpi, when expression increases in CBSV_Tanza, CBSV_KO, CBSV_UHam infections. The reason for this increase in ITPA expression is unclear, it is possible that later by 18 dpi stress responses have been activated in *N. benthamiana* in response to CBSV_Tanza, CBSV_HKO and CBSV_UHam infection, which induce reactive oxygen bursts that are associated with the release of oxidative agents such as nitrous oxide that limit pathogen spread through the hypersensitive response (HR) but can also cause oxidative damage of canonical nucleotides and thereby cause host DNA mutations. Therefore perhaps *N. benthamiana* attempts to limit the accumulation of oxidatively damaged, non-canonical nucleotides by increasing ITPA expression, however this is currently speculation. To further study this, a future experiment could be performed to quantify the expression of ITPA throughout the life-span of *N. benthamiana* in the presence and absence of CBSV infection.

![Figure 4.25](image-url)

**Figure 4.25:** QPCR quantification of the predicted *N. benthamiana* ITPA gene during infection with CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam. The abundance of ITPA transcript was compared with the abundance of the F-box reference gene transcript. The $2^{\Delta\Delta Ct}$ method was then used to calculate the relative ITPA transcript abundance in infected plants compared with three non-inoculated plants.
4.2.5 Investigation into the potential role of CBSV Ham1 in reducing CBSV mutation rate

To investigate whether the CBSV Ham1 protein reduces CBSV mutation rate, a deep sequencing experiment was designed to compare the number of single nucleotide variants (SNVs) in deep-sequencing reads generated from viral RT-PCR amplicons amplified from *N. benthamiana* infections with CBSV_Tanza and CBSV_HKO, which lacks Ham1.

4.2.5.1 *N. benthamiana* infections used in deep sequencing analysis

Four replicate *N. benthamiana* plants were infected with CBSV_Tanza or CBSV_HKO. At 10 dpi, CBSV_Tanza infections developed severe systemic necrosis and wilting, whereas CBSV_HKO infections developed systemic leaf wrinkling and mottling. RT-PCR was performed to confirm plants were infected with the correct IC constructs (Fig. 4.26).

![RT-PCR gel](image)

Figure 4.26: RT-PCR confirmed that four replicate *N. benthamiana* plants were infected with CBSV_Tanza (lanes 1 – 4) and four replicate plants were infected with CBSV_Ham1KO (lanes 5 – 8). This confirmed that these infections could be used in subsequent deep sequencing analysis. RT-PCR was performed with the CBSV_Ham1_KO_Fw (GGTATGCTGTGTTCTGGGCATCTGA) and CBSV_Ham1_KO_Rv (CGAGCTACTTGAGGAGCACGAAGCC) primers. Positive control PCR reactions are shown where 963 bp amplified from CBSV_Tanza IC plasmid (+C6) and 285 bp amplified from the CBSV_HKO IC plasmid (+KO). Whereas no amplification occurred in the negative control PCR reaction where no template was added (-).

4.2.5.2 RNA extractions from *N. benthamiana* infections used deep sequencing analysis

To generate viral RT-PCR amplicons for deep sequencing, RNA was extracted from *N. benthamiana* plants infected with CBSV_Tanza and CBSV_HKO at 10 dpi using the E.Z.N.A Plant RNA extraction kit (Omega Bio-Tek). To ensure that RNA with a high purity and integrity was used, two separate RNA extractions were performed from each *N. benthamiana* plant. RNA concentration, purity and integrity were analysed using an RNA screen tape. For each plant, the RNA extraction with the highest RNA Integrity Number (RIN) was used as template in RT-PCR.
4.2.5.3 Viral RT-PCR amplicon design for deep sequencing analysis

The original experimental plan was to amplify four RT-PCR amplicons (2.5 Kb – 3 Kb), spanning the entire CBSV Tanza genome. Unfortunately, these large amplicons could not be reliably amplified for all samples. Therefore, the plan was changed to amplify a single RT-PCR amplicon targeting the CBSV coat protein to the start of 3’UTR (Fig. 4.27). Alicai et al., (2016) have previously shown that the N terminals of CBSV coat protein sequences are variable and so it was hypothesised that targeting this region may increase the probability of detecting low frequency SNVs.

![Diagram of CBSV Tanza genome showing the CP – 3’UTR (1202 bp) region targeted by RT-PCR to produce amplicons from CBSV_Tanza and CBSV_HKO infected N. benthamiana plants.](Image)

4.2.5.4 Amplification of viral RT-PCR amplicons used in deep sequencing analysis

To generate amplicons, RT-PCR was performed on 2 μg of RNA using oligo(dT)18 primer to generate single stranded cDNA, which was used as template in PCR to amplify the 1202 bp amplicon targeting the coat protein – 3’UTR region using the forward primer: CBSV_CP_Fw (ATTGACAAGGATGAGATTGAAGC) and the reverse primer: CBSV_3UTR_Rv (TCTTGGCTTCACAACAAAC). To ensure a minimal number of SNVs were introduced through PCR, high-fidelity Phusion polymerase was used and the number of PCR thermo-cycles was reduced from 35 to 30. For each cDNA sample, two RT-PCR reactions were set up and the two amplicons were pooled, and gel purified. Gel electrophoresis demonstrated that amplicons of the correct sized had been generated from each plant (Fig. 4.28). 200 ng of each purified RT-PCR amplicon were then given to the University of Bristol Genomics facility for deep sequencing.
4.2.5.5 Deep sequencing of viral RT-PCR amplicons

All following library preparation procedures were performed by the University of Bristol Genomics facility. Each viral RT-PCR amplicon was used to produce an individual library according to the TruSeq Nano DNA Library Prep protocol (Illumina). Amplicons were fragmented to an approximate size of 550 bp using Bioruptor sonication instrument (Diagenode) for 5 mins and run on a DNA screen tape to confirm their fragmentation. Fragments were then cleaned using the AmPure XP 0.8 clean-up procedure and two indexed adapters were ligated to paired ends of the fragments. PCR amplification was then performed to enrich for fragments containing both paired-end adaptors. Library sizes and purity was then validated using a DNA 1000 ScreenTape. Subsequently the samples were diluted to 10 pM with 25% PhiX spike-in and run in a MiSeq instrument using the MiSeq Reagent Kit v3 (2X 75 bp).

4.2.5.6 Reference sequence for alignment of deep sequencing reads

To confirm a consensus reference sequence that deep sequencing reads could be aligned to, the CBSV_Tanza and CBSV_HKO ICs were sent for Sanger sequencing. Both ICs were found to contain the same sequence in the 1202 bp region and so the same reference could be used to align both CBSV_Tanza and CBSV_HKO deep sequencing reads.

4.2.5.7 Bioinformatic analysis

Bioinformatic analysis was performed with the assistance of Dr. Thomas Batstone at the University of Bristol, School of Biological Sciences. FASTQ files generated from the MiSeq run were uploaded onto the Partek Flow server, which showed that there were over 1 million reads.
produced from each sample with expected read length of 76 bp (Table 4.5). The average read quality for the samples was high and consistent between samples with phred scores of 37.16 – 37.23. The Illumina adapter sequences were then trimmed from reads, which did not significantly affect the number or quality of reads. Low quality reads with phred scores of less than 30 were then trimmed, which removed 1.38% - 1.64% of reads from the total reads.

4.2.5.8 Alignment of deep sequencing reads to the reference sequence

Reads were then aligned to the 1202 bp reference sequence using Bowtie 2-2.2.5, very-sensitive-local pre-sets and maximum read length of 800 bp. A high proportion of reads aligned to the reference ranging from 91.44% - 94.14% with an approximate average coverage depth of 117,000 – 218,000 reads per base position (Table 4.5).

Table 4.5: Outputs from alignments of CBSV_Tanza (C6 1 – 4) and CBSV_HKO (KO 1 – 4) deep sequencing reads to the reference sequence.

<table>
<thead>
<tr>
<th>Sample</th>
<th>No. pre-aligned reads</th>
<th>Reads aligned to reference (%)</th>
<th>Average read depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>C61</td>
<td>1555692</td>
<td>93.44</td>
<td>175,910</td>
</tr>
<tr>
<td>C62</td>
<td>1504526</td>
<td>93.13</td>
<td>169,533</td>
</tr>
<tr>
<td>C63</td>
<td>1225506</td>
<td>94.13</td>
<td>138,314</td>
</tr>
<tr>
<td>C64</td>
<td>1287935</td>
<td>93.78</td>
<td>145,497</td>
</tr>
<tr>
<td>KO1</td>
<td>1155227</td>
<td>92.67</td>
<td>128,995</td>
</tr>
<tr>
<td>KO2</td>
<td>1068194</td>
<td>93.06</td>
<td>119,992</td>
</tr>
<tr>
<td>KO3</td>
<td>1002569</td>
<td>93.22</td>
<td>112,789</td>
</tr>
<tr>
<td>KO4</td>
<td>1906683</td>
<td>91.43</td>
<td>210,241</td>
</tr>
</tbody>
</table>

4.2.5.9 In-depth read quality analysis

To analyse the quality of the aligned sequencing reads Qualimap v2.2.1 software was used. An example of the coverage of aligned sequence reads across the reference sequence is shown in Fig. 4.29 and are provided for all samples in Appendix 4.6. There is a high read coverage at the 5’ or 3’ ends of the reference sequence and a normal distribution of read coverage around the centre. This is likely to have been caused by bias for fragmentation to occur in the middle of the 1202bp amplicons during library preparation. Once amplicons were fragmented in the middle, there would be a high proportion of ≈ 600 bp amplicons, consisting of the either the 5’ end – middle or middle – 3’ end of the original amplicon sequence. These amplicons would then be further enriched during size selection, leading to higher read coverage at the 5’ and 3’ ends.
Figure 4.29: Qualimap coverage report of aligned reads from a representative sample (C61) mapped to the 1202 bp reference sequence. There was higher coverage at the 5' and 3' ends of the fragments and a normal distribution of coverage in the centre, which may be due to bias in fragmentation around the middle of amplicon fragments during library preparation.

4.2.5.10 Low-frequency SNV calling from deep sequence reads

To detect the number of SNVs in deep sequencing reads, the statistical algorithm LoFreq was used, which models sequencing run-specific error rates to distinguish low frequency true variants (<0.5%) from sequencing errors (Wilm et al., 2012). LoFreq identified SNVs in aligned sequencing reads from each of the samples in Table 4.6. To compare the SNVs in sequences across samples with different number of aligned reads, the total number of SNVs per million nucleotides of aligned reads were calculated, as shown Fig. 4.30 – 31. The mean SNV number per million/nt of aligned CBSV_Tanza reads was 0.081 ±0.047, whereas the mean for CBSV_HKO reads was 0.046 ±0.013. This difference was not statistically significant as assessed by independent t-test: $t(6) = 1.398, p = 0.212$. The mean number of transition SNVs per million/no of aligned CBSV_Tanza reads was 0.05 ±0.0245, whereas the mean for CBSV_HKO reads was 0.0387 ±0.019. This difference was not statistically significant as assessed by independent t-test: $t(6) = 1.111, p = 0.309$. 

143
Table 4.6: Number of SNVs detected per million nucleotide of aligned sequencing reads.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total no. SNVs</th>
<th>No. of SNV per million/nt aligned reads</th>
<th>Total no. transition SNVs</th>
<th>No. transition SNV per million/nt aligned reads</th>
</tr>
</thead>
<tbody>
<tr>
<td>C61</td>
<td>5</td>
<td>0.045</td>
<td>4</td>
<td>0.036</td>
</tr>
<tr>
<td>C62</td>
<td>8</td>
<td>0.075</td>
<td>6</td>
<td>0.056</td>
</tr>
<tr>
<td>C63</td>
<td>13</td>
<td>0.149</td>
<td>8</td>
<td>0.092</td>
</tr>
<tr>
<td>C64</td>
<td>5</td>
<td>0.055</td>
<td>4</td>
<td>0.044</td>
</tr>
<tr>
<td>KO1</td>
<td>5</td>
<td>0.062</td>
<td>5</td>
<td>0.062</td>
</tr>
<tr>
<td>KO2</td>
<td>4</td>
<td>0.053</td>
<td>3</td>
<td>0.040</td>
</tr>
<tr>
<td>KO3</td>
<td>3</td>
<td>0.042</td>
<td>3</td>
<td>0.042</td>
</tr>
<tr>
<td>KO4</td>
<td>4</td>
<td>0.030</td>
<td>2</td>
<td>0.015</td>
</tr>
</tbody>
</table>

Figure 4.30: Total number of SNVs detected per million nucleotides of aligned sequence reads from CBSV_Tanza infections (C61–4) and CBSV_HKO infections (KO1–4). SNVs were detected using the LoFreq algorithm. There were no significant differences in the total number of SNVs: t(6) = 1.398, p = 0.212.

Figure 4.31: Total number of transition SNVs detected per million nucleotides of aligned sequence reads from CBSV_Tanza infections (C61–4) and CBSV_HKO infections (KO1–4). SNVs were detected using the LoFreq algorithm. There were no significant differences in the number of transition SNVs: t(6) = 1.111, p = 0.309.
4.2.5.11 Potential correlation between SNV number and CBSV transcript abundance

The number of SNVs in viral amplicons will be influenced by the number of times the viral genome has been replicated. Therefore, further analysis was performed to investigate whether there is a correlation between the number of SNVs in sequence reads and the abundance of CBSV transcripts in corresponding infections. QPCR was used quantify the abundance of CBSV transcripts in the RNA samples that were used in the deep sequencing analysis. QPCR was performed as previously described using primers in Chapter 2, Table 2.2 that target the CBSV coat protein and the *N. benthamiana* F-box reference gene. Results from qPCR are shown in Fig. 4.32. A one-way ANCOVA test was performed to compare the number of SNVs in sequencing reads from CBSV_Tanza and CBSV_HKO, whilst accounting for variance in viral transcript abundance. Again, there was no significant difference, $F(1, 0.008) = 1.423, p = 0.286$. This further indicated that the presence of Ham1 in CBSV_Tanza infections did not significantly reduce the number of SNVs in viral sequencing reads.

![Figure 4.32: QPCR quantification of CBSV titres in *N. benthamiana* infected with CBSV_Tanza (C61 – 4) and CBSV_HKO (KO1 – 4). The abundance of the CBSV coat protein transcript was compared to the abundance of the F-box reference gene. The $2^{\Delta\Delta Ct}$ method was then used to calculate the relative CBSV transcript abundance in infected plants compared with three non-inoculated plants](image-url)
4.2.5.12 Mapping SNVs to reference sequence

To determine whether SNVs occurred at a higher frequency at specific positions, SNVs were mapped to the CP – 3’UTR reference sequence (Fig. 4.33). SNVs at positions 274, 432 and 454 bp were the most frequent in CBSV_HKO samples, whereas SNVs at positions 324, 432, 454, 489 and 676 bp were the most frequent in CBSV_Tanza samples. SNVs at positions 432 (G→A) and 454 (T→C) were detected in all CBSV_Tanza and CBSV_HKO samples. Overall mean SNV frequencies were higher in CBSV_Tanza infection reads, especially the SNV at position 324 bp in the reference sequence, which occurred at relatively high frequencies of 2.17% and 2.63% in two of the CBSV_Tana samples. In both CBSV_Tanza and CBSV_HKO sequencing reads, SNVs were detected at high frequencies in the N’ terminal of the CBSV Tanza coat protein, which may indicate that this genome region is relatively variable during viral infections. This agrees with phylogenetic analyses that have demonstrated that the N’ terminals of U/CBSV coat protein sequences are divergent (Winter et al., 2010) and that positive selection in this region of the CBSV coat protein may enable host adaptation (Alicai et al., 2016a; Mbanzibwa et al., 2011b).

Figure 4.33: Mean frequency of SNVs in deep sequencing reads from CBSV_Tanza and CBSV_HKO infections of N. benthamiana.
4.3 Discussion

In this Chapter analysis was performed to identify phylogenetic relationships between CBSV Tanza and UCBSV Kikombe Ham1 proteins with ITPA homologs from a range of species and with other CBSV and UCBSV isolates. Next, the involvement of CBSV and UCBSV Ham1 proteins in symptom development and viral accumulation during infections of *N. benthamiana* was investigated through CBSV Tanza IC manipulations and infections. Manipulations included mutating the conserved Ham1 pyrophosphohydrolase motif SHR to SAA in the CBSV_mutHam IC, which is expected to abolish potential ITPAse activity (Gall et al., 2013). To investigate whether CBSV and UCBSV Ham1 sequences are associated with differential symptom development, the chimeric IC CBSV_UHam was constructed, which contains a UCBSV Ham1 sequence replacement. Then, qPCR was performed to test whether *N. benthamiana* alters expression of its predicted ITPA gene in response to CBSV expression. Finally, to determine whether CBSV Ham1 affects CBSV mutation rate, the number of SNVs in viral RT-PCR amplicons were compared between *N. benthamiana* infections with CBSV_Tanza and CBSV_HKO. The results from these experiments are discussed below.

4.3.1 Bioinformatic analysis of CBSV Tanza and UCBSV Kikombe Ham1 sequences

A sequence alignment of CBSV Tanza and UCBSV Kikombe Ham1 amino acid sequences with ITPA homologs identified the presence of conserved phosphohydrolase motifs, including the SHR motif, which in the human ITPA protein is involved with nucleotide binding and substrate specificity (Gall et al., 2013; Stenmark et al., 2007). The fact that CBSV and UCBSV genomes contain these ITPA phosphohydrolase motifs is intriguing and suggests there must be selective pressure to integrate and maintain them. A phylogenetic tree with ITPA homologs demonstrated that CBSV Tanza and UCBSV Kikombe Ham1 sequences do not share strong phylogenetic relationships with the plant, animal, fungal, bacterial or archaeal ITPA sequences that were included in this analysis. Therefore it is not possible to determine the origin of the U/CBSV Ham1 sequences. In contrast, the EuRV Ham1 sequence is highly related to the *E. coli* RdgB sequence, which may indicate that EuRV acquired this sequence from a co-infecting bacteria. This has previously been reported for the integration of proteins with Alkb domains into *Flexiviridae* viral genomes, which may originate from co-infecting bacterial pathogen(s) (Bratlie and Drabløs, 2005).

Phylogenetic analysis of Ham1 sequences from 19 CBSV and 29 UCBSV isolates further illustrated that CBSV and UCBSV Ham1s share a low sequence similarity, typically of around 55%. This indicates that CBSV and UCBSV Ham1 sequences have either undergone significant
sequence divergence from an ancestral Ham1 sequence or may have been integrated into CBSV and UCBSV genomes separately, as suggested in Monger et al., (2010).

### 4.3.2 N. benthamiana infections with a range of CBSV ICs

To gain insights into the roles of CBSV Tanza and UCBSV Kikombe Ham1 sequences in symptom development, two CBSV Tanza ICs were constructed that encode a mutated SHR to SAA mutated Ham1 sequence (CBSV_mutHam) and a UCBSV Kikombe Ham1 sequence replacement (CBSV_UHam). Attempts were made to transform yeast plasmids into the *E. coli* strain *ccdB*, as this strain had been able to replicate CBSV_Tanza IC with relatively high levels of sequence stability in Chapter 3. In this Chapter, transformations of *ccdB* failed to generate transformant colonies. Whereas transformation of the strain TOP10 did generate transformant colonies and was able to propagate plasmids with sufficient sequence stability. The *ccdB* strain is derived from TOP10 and so shares similar features (Appendix 2.1), which may explain their similar abilities to stably propagate CBSV_Tanza ICs.

#### 4.3.2.1 N. benthamiana infections with CBSV_mutHam

CBSV_mutHam infections of *N. benthamiana* demonstrated a low infectivity rate, with 30% of plants typically developing symptoms. The remaining plants were asymptomatic, and infections could not be detected by RT-PCR. Symptoms in CBSV_mutHam infections developed late in infection, at 18 – 21 dpi and resemble CBSV_KO infections in that they lack necrosis but instead develop systemic leaf distortion and mottling. This demonstrates that the SHR motif in the context of the CBSV Ham1 is somehow associated with necrosis development in *N. benthamiana*. There are many possible explanations for this. For instance, perhaps the development of necrosis is associated with CBSV Ham1 pyrophosphohydrolase activity, which requires a functional SHR motif. It is also possible that later in infection, *N. benthamiana* is able to recognise the CBSV Ham1 protein in a specific conformation that depends on the SHR motif and this recognition triggers the development of necrosis. Mutating the SHR motif to SAA may significantly affect electrostatic interactions; changing Ham1 protein conformation, resulting in a loss of recognition and necrosis development. Alternatively, the SHR motif may be involved with CBSV Ham1 protein folding and aggregation in cells, which may perturb cellular processes, leading to the development of necrosis. Finally, the SHR motif may be required for interaction(s) between the CBSV Ham1 protein and other CBSV or host proteins and these interactions result in necrosis development. Analysis of CBSV transcript abundance in CBSV_mutHam infections revealed that CBSV_mutHam titres remain relatively low but
appeared to increase slightly throughout infection. This may indicate that mutating the CBSV Ham SHR motif to SAA negatively affects CBSV replication and/or protein translation, which may be due to an inability for the CBSV Ham1 protein to fold correctly.

4.3.2.2 *N. benthamiana* infections with CBSV_UHam

CBSV_UHam infections of *N. benthamiana* also lack necrosis and develop similar symptoms to CBSV_KO infections with systemic leaf distortion and mottling. Nanyiti (PhD thesis, 2016) demonstrated that *N. benthamiana* plants infected with UCBSV Kikombe develop mild mosaics and so perhaps the presence of UCBSV Ham1 in CBSV_UHam infections is associated with a reduction in symptom severity, compared with necrotic CBSV_Tanza infections. This may indicate that CBSV Ham1 sequence is the determinant for necrosis development during CBSV infection of *N. benthamiana*. However, this requires further investigation. For example, a chimeric IC could be constructed consisting of the UCBSV Kikombe genome with a CBSV Tanza Ham1 replacement. If *N. benthamiana* infections with this IC develop necrosis it would further confirm that the CBSV Ham1 sequence is the determinant for necrosis development in this plant host. Interestingly, the UCBSV Ham1 protein also encodes a SHR motif. Therefore perhaps the UCBSV Ham1 protein folds into a different shape, which is not directly or indirectly associated with the development of necrosis. QPCR analysis of CBSV transcript abundance in CBSV_UHam *N. benthamiana* infections show that during early infection (7 dpi), CBSV_UHam accumulates to lower titres than both CBSV_Tanza and CBSV_KO. Other studies have reported that UCBSV tends to accumulate to lower levels than CBSV cassava infections (Kaweesi et al., 2014; Mohammed et al., 2012; Ogwok et al., 2014). Therefore, perhaps this reduction in CBSV_UHam titers at 7 dpi, reflects this reduced UCBSV accumulation. By 10 dpi CBSV_UHam titres are comparable to CBSV_Tanza and CBSV_KO and that CBSV_UHam transcripts remain relatively high at 14 – 18 dpi. This may suggest that later in infection, the UCBSV Ham1 protein sequence can complement the loss of the CBSV Ham1 sequence. It is also possible that the UCBSV Ham1 protein is not efficiently cleaved from the CBSV Tanza polyprotein. In an attempt to ensure that the UCBSV Ham1 peptide is cleaved from the CBSV Tanza polyprotein, the CBSV_UHam IC was designed so that the CBSV Nib – Ham cleavage sequence of L-T-Y-V-V-D was maintained. However, this resulted in the fusion of the amino acids: V-V-D to the start of the UCBSV Ham1 peptide, as shown in Fig. 4.16, which may interfere with polyprotein processing and/or interfere with protein functions. To further investigate whether the UCBSV Ham1 peptide is cleaved from the CBSV Tanza polyprotein, Western-blot analysis could be performed with UCBSV Ham1 specific antibodies.

149
4.3.2.3 *N. benthamiana* infections with CBSV_HKO

*N. benthamiana* infections with CBSV_HKO further confirmed results previously found in Pablo-Rodriguez (PhD thesis, 2017), that CBSV_HKO can systemically infect and accumulate to high titres in *N. benthamiana*. This demonstrates that the CBSV Ham1 protein is not essential for CBSV replication or movement in *N. benthamiana*. This is unusual because *Potyviridae* genomes are usually highly stream-lined and encode multi-functional proteins from limited genetic information (Valli et al., 2018). Therefore, the ability of CBSV_HKO to accumulate to such high titres when missing an entire protein is surprising. It is possible that U/CBSV Ham1 proteins function during vector transmission, which was not tested in this study. In addition, although CBSV_HKO can infect *N. benthamiana*, it is not known whether it can infect cassava. Further investigations to study this would indicate whether the CBSV Ham1 protein is an adaptation to Euphorbia hosts. Analysis of CBSV transcript abundance, shows that during early infection CBSV_Tanza accumulates to higher levels than CBSV_KO, at 10 – 14 dpi CBSV_KO reaches comparable titres to CBSV_Tanza and at 18 dpi CBSV_HKO increases to much higher levels than CBSV_Tanza. This shows that the CBSV Ham1 protein is likely to be involved with viral replication and/or movement during early infection. Later in infection CBSV_KO may be able to reach high titres due to lack of necrosis, which in CBSV_Tanza infections limits CBSV accumulation.

4.3.2.4 Expression of *N. benthamiana* ITPA in response to infection with CBSV ICs

Quantitative transcript abundance analysis appears to show that expression of the predicted *N. benthamiana* ITPA sequence is stable throughout infection with all of the CBSV ICs: CBSV_Tanza, CBSV_HKO, CBSV_UHam and CBSV_mutHam. This suggests that the hypothesis that *N. benthamiana* may down-regulate its ITPA gene in response to CBSV infection, to thereby increase CBSV mutation rate may be false. Although this response to CBSV may not occur in *N. benthamiana*, it is still possible that cassava may down regulate its ITPA gene(s) during U/CBSV, and this could be further investigated. It is also possible that the expression of other genes may be regulated to affect U/CBSV mutation rates during infection. For example, several studies in mammals have shown that expression of the APOBEC3 cytidine deaminase protein is up-regulated in response to viral infections, which results in increased viral mutation and reduced viral accumulation (Bonvin et al., 2006; Okeoma et al., 2007; Suspene et al., 2005). In plants, transgenic expression of the *A. thaliana* APOBEC3 protein in *N. bigelovii* plants was shown to increase mutation rate of CaMV and lower viral accumulation (Martín et al., 2017). The role of
deaminases in plant-virus interactions is currently largely unknown. It would be interesting to investigate whether deaminases are upregulated during U/CBSV infections of cassava.

4.3.3 Effect of CBSV Ham1 on CBSV mutation rate

In this study, the potential for CBSV Ham1 to reduce CBSV mutation rate was investigated by deep sequencing analysis of viral RT-PCR amplicons to compare the number of SNVs in *N. benthamiana* infections with CBSV_Tanza and CBSV_HKO. If CBSV Ham1 does function to reduce CBSV mutation rate, it is expected that there would be a higher number of SNVs in viral amplicons from CBSV_HKO infections compared with CBSV_Tanza infections. Deep sequencing analysis was performed on the 1202 bp amplicons, targeting the CBSV coat protein – 3’UTR region. Low-frequency SNVs were detected in sequence reads using the LoFreq algorithm. Statistical analysis found that there were no significant differences in the number of SNVs per million nt of aligned sequence reads in CBSV_Tanza and CBSV_KO infections, even when CBSV transcript abundance had been accounted for. Therefore, this experiment indicates that the CBSV Ham1 protein does not function to reduce CBSV mutation rate during infection of *N. benthamiana*. This agrees with Pablo-Rodriguez (PhD thesis, 2017), who demonstrated that the transgenic expression of CBSV Ham1 in *N. tabacum* did not significantly affect the number of SNVs in PVY or TMV amplicons. However, this is a preliminary experiment, which provides a snapshot of SNV diversity at a single time point. Future experiments could be performed to gain more insights into the potential role of CBSV Ham1 in reducing the CBSV mutation rate. For instance, perhaps 10 dpi was too early in infection to detect a significant difference in mutation rate between CBSV_Tanza and CBSV_HKO infections. Lafforgue et al., (2011) demonstrated that TuMV mutants accumulated to relatively high levels only after 10 - 15 passages through *A. thaliana*. Therefore, the number of SNVs could be measured at multiple time-points throughout infections with CBSV_Tanza and CBSV_HKO and through multiple passages. It is possible that SNV differences were not detected due to bias in fragmentation shearing at the middle of amplicon fragments during library preparation, which may have led to unequal coverage across the reference. Knierim et al., (2011) demonstrated that preparing deep sequencing libraries with amplicons smaller than 3,000 bp lead to unequal and lower read coverage. Therefore, increasing the size of target amplicon should reduce fragmentation bias and increase read coverage, which may increase probability of detecting SNV differences. Finally, the effect of CBSV Ham1 on CBSV mutation rate should be investigated during infection of cassava as it has been shown that viral mutation rates can depend on the host species
infected. For example, fidelity of CMV replication during infections of pepper is lower than during infections of tobacco (Pita et al., 2007).

**4.3.4 Position of SNVs in reference sequence**

The frequencies of SNVs across RT-PCR amplicons from both CBSV_Tanza and CBSV_HKO infections, demonstrated that there were higher frequencies of SNVs in the N’ - central region of the coat protein – 3’UTR sequence. The N’ terminals *Potyviridae* coat protein sequences tend to be highly variable and are associated with many functions, including systemic infection and host specificity (Andersen & Johansen, 1998; Ullah et al., 2003). Mbanzibwa et al., (2011) and Alicai et al., (2016) have used evolutionary modelling to predict amino acid positions N’ terminal and the central region (306 – 552 bp) of CBSV coat protein, which are likely to be under positive selection and may be enabling CBSV adaptation to cassava hosts. Therefore, perhaps the SNVs observed in the N’ – central coat protein region of CBSV_Tanza and CBSV_HKO amplicon sequences may enable enhanced adaptation of CBSV to *N. benthamiana*.

**4.3.5 Conclusions**

To conclude, this Chapter has identified that the CBSV Tanza Ham1 protein and specifically its conserved SHR pyrophosphohydrolase motif is associated with necrosis development and relatively high viral accumulation during early infection of *N. benthamiana*. We speculate that the SHR motif may be required for correct folding and functioning of the CBSV Ham1 protein, which is directly or indirectly associated with necrosis development in *N. benthamiana*. In contrast, the UCBSV Ham1 is not associated with necrosis in *N. benthamiana*. The UCBSV Ham1 protein may be able to complement the loss of CBSV Ham1 in CBSV_UHam IC infections but the presence of UCBSV Ham1 and does restore the development of necrosis. Deep sequencing analysis suggests that CBSV Ham1 does not significantly affect SNV number or frequency in CBSV genomes during *N. benthamiana* infection and therefore may not reduce CBSV mutation rate. To further study the potential functions of CBSV and UCBSV Ham1 proteins, pyrophosphohydrolase enzyme assays were performed, as described in Chapter 5.
Chapter 5: Investigation into potential U/CBSV Ham1 pyrophosphohydrolase activities

5.1 Introduction

As described in Chapter 4, the incorporation of non-canonical nucleotides into nucleic acid can cause mutation. Eukaryotic and prokaryotic organisms encode inosine triphosphatase (ITPA) proteins which specifically target non-canonical nucleotide triphosphates by hydrolysing their β-pyrophosphate bonds to release a pyrophosphate (PPI) molecule and a nucleotide monophosphate that cannot be incorporated into nucleic acid (Galperin et al., 2006) and so ITPA proteins function to reduce mutation rates (Kozmin et al., 1996). U/CBSV Ham1 proteins share relatively high levels of sequence similarity with the ITPA superfamily, however their potential pyrophosphohydrolase activities are currently uncharacterised. In this Chapter the pyrophosphohydrolase activities of CBSV Tanza and UCBSV Kikombe Ham1 proteins are investigated through in vitro enzyme assay experiments.

5.1.1 ITPA enzymatic activities

ITPA activity was first detected by Liakopoulou & Alivisatos (1964), who demonstrated that lysed erythrocytes have higher pyrophosphohydrolase activity with the non-canonical nucleotide ITP, compared with low activity with canonical nucleotides ATP, GTP and UTP. Subsequently Vanderheiden (1979) purified ITPA from rabbit red blood cells and found that it had higher pyrophosphohydrolase activity with the non-canonical nucleotides: ITP, dITP and XTP. ITPA proteins from a range of species have since been expressed in E. coli, purified and enzyme assays performed to characterise their substrate specificities. The protein purification process involves cloning the protein sequence into an expression plasmid vector, expressing the protein in E. coli and purifying the protein using a metal ion affinity chromatography column. During the assay, the protein is incubated with nucleotide triphosphate substrates. If the protein has pyrophosphohydrolase activity a monophosphate nucleotide and pyrophosphate is released. The reaction is coupled with the addition of yeast pyrophosphatase which enzymatically cleaves inorganic pyrophosphate into two phosphate molecules. The reaction is stopped, for example through the addition of trichloroacetic acid and the reaction products quantified. Usually phosphate concentration is quantified through the addition of dye which forms a coloured complex with inorganic phosphate and this colorimetric change is measured (Simone et al., 2013).
The first ITPA protein to be expressed in *E. coli* was Mj0226 from the Achaea thermophile *Methanococcus jannaschii* (Cho et al., 1999). Mj0226 activity was screened with a range of nucleotide substrates and high-pressure liquid chromatography (HPLC) was used to identify reaction products. Mj0226 was found to have the highest activity with non-canonical nucleotides XTP and ITP and relatively low activity with GTP, dGTP, ATP, CTP and TTP. Mj0226 NTPase activity was shown to be the most efficient in the presence of metal co-factors: Mg$^{2+}$ or Mn$^{2+}$ (Cho et al., 1999). The human ITPA (hITPA) protein was also shown to have a high specificity for non-canonical nucleotides: XTP, dITP and ITP compared with canonical nucleotides (Fig. 5.1) (Lin et al., 2001). The activity was greatly enhanced in the presence of Mg$^{2+}$ and the reducing agent dithiothreitol. Finally, the *E. coli* Rdgb protein was also shown to have high specificity for non-canonical nucleotides XTP and dITP compared to GTP (Burgis & Cunningham, 2007; Zheng et al., 2005).

Figure 5.1: Relatively high pyrophosphohydrolase activity of purified human ITPA with non-canonical nucleotides: ITP, dITP and XTP compared to canonical nucleotides: ATP, dATP, CTP, dCTP, UTP, dTTP, GTP and dGTP (Lin et al., 2001).

### 5.1.2 ITPA protein structures

X-ray crystallography has been used to determine several ITPA protein structures from a range of eukaryotic and prokaryotic species, including Mj0226 from *M. jannaschii* (Cho et al., 1999), human ITPA (Stenmark et al., 2007) and Maf from *Bacillus subtilis* (Minasov et al., 2000; Zheng et al., 2005). This has shown that ITPA proteins share common structural features; they are formed of two globular monomers, which bind each other through multiple hydrogen bonds and bridging water molecules. A nucleotide binding cleft is formed between N and C terminal lobes. The model of the hITPA protein structure is shown in Fig. 5.2.
5.1.3 Proposed model of ITPA substrate binding and pyrophosphohydrolase activity

ITPA binding clefts contain highly conserved amino acid residues, for instance the Serine-Histidine-Arginine (SHR) motif is found in all ITPA homologs and is predicted to be involved with nucleotide binding (Burgis and Cunningham, 2007; Cho et al., 1999; Savchenko et al., 2007; Stenmark et al., 2007; Zheng et al., 2005). In hITPA, His177 and Arg178 residues from the SHR motif are predicted to form hydrogen bonds with the 6-keto oxygen group of ITP (Fig. 5.2). Upon hITPA binding ITP, the N terminal lobe moves towards the C terminal lobe by 25°, closing the binding cleft (Stenmark et al., 2007). A water molecule is predicted to be co-ordinated at position Asp72 or Asn16 for nucleophilic attack of the pyrophosphate bond (Stenmark et al., 2007). The hITPA binding cleft is structured to specifically accommodate non-canonical nucleotides but has steric incompatibility for canonical nucleotides (Galperin et al., 2006). The proposed model is that the hITPA binding cleft can accommodate ITP and XTP keto-oxygen groups but not the amino groups of ATP and GTP (Stenmark et al., 2007). Although the 2-amino group of GTP could be accommodated in the binding cleft, there is a high free energy cost, making it energetically unfavourable (Galperin et al., 2006). The hITPA Phe-149 residue is predicted to be critical in excluding GTP from the binding cleft (Gall et al., 2013). Nucleotide ribose hydroxyl groups are not predicted to be involved with hITPA binding and point outwards towards the solvent, enabling hITPase activity with both deoxyribose and ribose nucleotides.

Figure 5.2: Left: Protein structure of the human ITPA (hITPA) dimer without (blue) and with ITP (pink) bound in the binding cleft. A yellow arrow shows the binding cleft of each monomer. Upon hITPA binding to ITP, the lower N terminal lobe moves towards the C terminal lobe to close the binding cleft. A co-ordinated water molecule then undergoes nucleophilic attack on the ITP β-pyrophosphate bond releasing inosine monophosphate and pyrophosphate (Stenmark et al., 2007). Right: Model of the hITPA substrate-binding pocket. The hypoxanthine base (Hx) of ITP is accommodated and bound through hydrogen bonds with conserved amino acids, including Ser176, His177 and Arg178, which form the SHR motif. The ITP ribose sugar (R) hydroxyl groups do not make strong contacts with hITPA and point outwards towards the solvent, enabling ITPase activity with both deoxyribose and ribose nucleotide triphosphates. The Phe149 residue is predicted to make the binding cleft unfavourable for (d)GTP binding. Image from Gall et al., 2013 adapted from Stenmark et al., 2007.
5.1.4 Hypotheses

CBSV and UCBSV Ham1 protein sequences share high levels of sequence identity with ITPA protein homologs and contain conserved pyrophosphohydrolase motifs such as SHR (Mbanzibwa et al., 2009). It was hypothesised that U/CBSV Ham1 proteins also function as pyrophosphohydrolases and target non-canonical nucleotides. Viral pyrophosphohydrolase activity may serve several potential in vivo functions, such as reducing viral mutation rate. Finally, due to the relatively low-level sequence identity between the CBSV Tanza and UCBSV Kikombe Ham1 proteins, it was hypothesised that these proteins may have evolved different substrate activities depending on the requirements of the two viral species.

5.1.5 Aims

The overall objective of this Chapter is to investigate the potential for U/CBSV Ham1 proteins to have pyrophosphohydrolases in vitro activities that are specific to non-canonical nucleotides.

- To clone the wild-type CBSV Tanza Ham1 protein sequence into an E. coli expression vector, express and purify CBSV Tanza Ham1 protein.

- To perform enzyme assays to detect potential CBSV Tanza Ham1 and UCBSV Kikombe Ham1 pyrophosphohydrolase activities with a range of nucleotide substrates.

- To clone a mutated CBSV Tanza Ham1 protein sequence into an E. coli expression vector, express in E. coli and purify. Perform enzyme assays to detect differences in pyrophosphohydrolase activities of wild-type and mutated CBSV Tanza Ham1 proteins.

- To express and purify wild-type CBSV Tanza Ham1 protein for use in X-ray crystallography for protein structural determination.
5.2 Results

5.2.1 Structural similarity of CBSV Tanza and UCBSV Kikombe with hITPA

To compare structural similarity of CBSV Tanza and UCBSV Kikombe Ham1 proteins with related proteins, the online SWISS-MODEL software (SIB, 2018) was used to build predictive protein models. The U/CBSV Ham1 amino acids sequences were aligned to related protein structures, which have structurally characterised using X-ray crystallography. The CBSV Tanza Ham1 and UCBSV Kikombe Ham1 sequences share the highest sequence identity of 52% and 55% respectively with hITPA (NCBI: NP_258412.1). The predictive protein models cover 83% and 80% of the CBSV Tanza and UCBSV Ham1 sequences respectively, indicating high levels of structural similarity with the hITPA (Fig. 5.3). The models predict that the conserved pyrophosphohydrolase SHR motif in both U/CBSV Ham1 proteins is situated in the binding cleft between the N and C terminal lobes of the protein monomers. Sequence homology alone is not direct evidence for a proteins structure and potential function, as proteins with relatively high sequence identity (up to 50%) can fold into highly different structures (Dalal et al., 1997).

![Predicted protein structures](image)

Figure 5.3: Predicted protein structures of the CBSV Tanza Ham1 protein (A) and UCBSV Kikombe Ham1 protein (B) based on the amino acid sequence alignment with the human ITPA protein (NCBI: NP_258412.1). The models were built using the SWISS-MODEL online software (SIB, 2018). Amino acid residues in the pyrophosphohydrolase SHR motif (circled in yellow) are predicted to be in the binding cleft between the N and C terminal lobes of the protein monomers.
5.2.2 Expression of the wild-type CBSV Tanza Ham1 proteins

To express the wild-type CBSV Tanza Ham1 sequence in *E. coli*, the Ham1 sequence was first cloned into the expression vector: POPINF (OPPF-UK). The POPINF plasmid contains the lac Operator (*lacO*) and enables high levels of protein expression, induced by the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG) to *E. coli* growth media. Upon IPTG induction, IPTG de-represses *lacO*, which leads to expression of the T7 polymerase from chromosomal DNA in the *E. coli* strain BL21 (DE3) (Fig. 5.4). Simultaneously, IPTG also de-represses the *lac* Operator in the POPINF plasmid, which allows T7 RNA polymerase to bind to the T7 promoter in POPINF and drive the expression of the recombinant protein sequence. POPINF also encodes a 6x Histidine residue tag (His6-tag), which is fused to the N’ terminal of the recombinant protein sequence so that the protein can be purified by nickel ion affinity column chromatography. The column contains nickel chelated by nitrilotriacetic acid (Ni-NTA), which binds the protein via the His6-tag with micromolar affinity. This means that at a specific imidazole concentration range, imidazole will out-compete the His6-tagged recombinant protein for nickel binding resulting in protein elution. Therefore an imidazole concentration gradient can be used to elute the recombinant protein from the nickel column (Berrow et al., 2007). To initiate translation of the recombinant protein, POPINF contains a methionine start codon immediately upstream of the His6-tag (Berrow et al., 2007).

![Mechanism for IPTG induced expression of recombinant proteins from POPINF plasmid in the E. coli strain BL21 (DE3).](image)

Figure 5.4: Mechanism for IPTG induced expression of recombinant proteins from POPINF plasmid in the *E. coli* strain BL21 (DE3). *E. coli* BL21 (DE3) encodes the T7 RNA polymerase in its chromosomal DNA, which is under the control of the *lac* Operator (*lacO*). In the absence of IPTG, *lacO* is repressed by *lacI* and so T7 RNA polymerase is not expressed and cannot transcribe the recombinant protein sequence on the POPINF plasmid. Once IPTG is added, IPTG binds to *lacI*, de-repressing *lacO* and enabling the expression of T7 RNA polymerase. T7 RNA polymerase can then bind to the T7 promoter in the POPINF plasmid, leading to the expression of the recombinant protein.
5.2.2.1 Cloning design for the construction of the POPINF_CHam_WT plasmid

To enable the expression of the CBSV Tanza Ham1 protein sequence in *E. coli*, the Ham1 sequence was cloned into the POPINF expression vector (OPPF-UK). The CBSV Tanza Ham1 sequence was amplified by PCR using the CBSV_Tanza IC as template and the POPINF_CBSVHam1_Fwd/Rev primers in Table 5.1. These primers target the 5’ end of the Ham1 sequence from the first amino acid (valine) of the Ham1 sequence following the cleavage site with Nib and at the 3’ end with the four amino acids of the cleavage sequence (I-D-V-Q) preceding the start of the CP, according to the schematic in Fig. 5.5. The primers also contain 20 – 22 bp of homologous sequence with the POPINF plasmid at their 5’ ends. This enables the integration of the insert gene into the POPINF vector. The cloning reaction uses the In-Fusion enzyme (Takara Bio), which progressively removes nucleotides at 3’ end of both insert sequence and POPINF vector to produce single-stranded overlapping sequences, which then spontaneously anneal through base-pairing, resulting in the integration of the insert gene into the POPINF vector. The POPINF vector contains the 6X histidine tag sequence at the 3’ end of the insert sequence.

![Figure 5.5: PCR amplification of the CBSV Tanza Ham1 sequence from the CBSV_Tanza IC template using the POPINF_CBSVHam1_Fwd primer, which targets the Ham1 sequence from the first amino acid (valine) preceding the protein cleavage sequence I-T-Y-V with Nib. Whereas the POPINF_CBSVHam1_Rev primer targets the end of Ham1 sequence, including the four amino acids: I-D-V-Q that form the cleavage sequence between Ham1 and CP.](image)

Table 5.1: Primers used to amplify CBSV Tanza Ham1 sequence by PCR. The forward and reverse primers contain homologous sequence (red) with the POPINF plasmid.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence 5’ – 3’</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>POPINF_CBSVHam1_Fwd</td>
<td>AAGTTCTGTGTTTCAAGGCCCAGGTGGACAGGTCTCAGCC</td>
<td>CBSV Ham1 (617 bp)</td>
</tr>
<tr>
<td>POPINF_CBSVHam1_Rev</td>
<td>ATGGTCTAGAAAGCCTTAGTTGAAACATCAATAAAGAATCAG</td>
<td></td>
</tr>
</tbody>
</table>
5.2.2.2 Construction of the POPINF_CHam_WT plasmid

To construct the POPINF_CHam_WT plasmid, the CBSV Tanza Ham1 sequence (617 bp) was amplified by high-fidelity Phusion PCR from the CBSV_Tanza IC template and gel purified (Fig. 5.6). The POPINF plasmid was digested with high-fidelity Kpn I and Hind III (NEB) overnight at 37°C (Fig. 5.6). This disrupts the β-galactosidase gene and enables blue-white screening of transformant E. coli colonies. An In-Fusion cloning reaction (Takara Bio) was set up with a 1:2 ratio of plasmid (ng): insert (ng) according to the protocol in Chapter 2, section 2.2.7.2. The In-Fusion cloning mix was used to transform electrocompetent TOP10 E. coli cells. The cells were plated out onto LB agar plates containing 100 μg/ml carbenicillin, 40 μl of 0.1M IPTG and 120 μl of 20 mg/ml X-Gal and incubated at 37°C overnight.

Figure 5.6: **Left:** High fidelity Phusion PCR amplification of CBSV Tanza Ham1 sequence (617 bp) from CBSV_Tanza IC template using POPINF_CBSVHam1_Fwd/Rev primers in Table 5.1 (lane 1). **Right:** Restriction digest of the POPINF plasmid with Kpn I and Hind III to produce the linear plasmid (5531 bp). The CBSV Tanza Ham1 PCR fragment was inserted into linearized POPINF using In-Fusion cloning to generate the POPINF_CHam_WT plasmid.

5.2.2.3 Verification of the POPINF_CHam_WT plasmid construction

A mixture of blue and white colonies grew on test plates; ten white colonies were cultured in liquid LB containing 100 μg/ml carbenicillin overnight. PCR was performed on these cultures with POPINF_CBSVHam1_Fwd/Rev primers and the correct sized fragment 617 bp amplified from all ten, indicating the presence of the Ham1 sequence in the POPINF plasmids (Fig. 5.7). Plasmids were extracted from three of the E. coli cultures and the correct sized CBSV Ham1 sequence (617 bp) amplified by PCR from all three. One of these plasmids was sent for Sanger sequencing using the T7_Fwd primer: GAGCCAACTACTAGCGGTCC, which targets the T7 promoter sequence in POPINF. This generated a 923 bp sequence read, which aligned to the POPINF_CHam_WT plasmid map at positions 2228 – 3150 bp, corresponding to the CBSV Tanza Ham1 insertion at: 2244 – 3168 bp (Appendix 5.1). This verified correct construction of the POPINF_CHam_WT plasmid and that the insert CBSV Ham1 sequence was error free.
Culture PCR to detect the presence of the CBSV Ham1 fragment (617 bp) in ten *E. coli* cultures (lanes 1–10) to indicate successful construction of the POPINF_CHam_WT plasmid. PCR was performed with POPINF_CBSVHam1_Fwd/Rev primers (Table 5.1). Amplification of the 617 bp fragment occurred when the positive control plasmid CBSV_Tanza IC was used as template (lane +) and no amplification occurred in the negative control PCR reaction when no template was added (lane -).

5.2.2.4 Expression of wild-type CBSV Tanza Ham1 and purification

To express the CBSV Tanza Ham1 protein in *E. coli*, the POPINF_CHam_WT plasmid was transformed into the protein expression *E. coli* strain BL21 (DE3). A transformant colony was cultured in 2 L of LB media containing 100 μg/ml carbenicillin until the OD reached approximately 0.4. Protein expression was then induced by the addition of IPTG according to Chapter 2, section 2.5.1. Cells were resuspended, homogenised, lysed and the resulting lysate was centrifuged and filtered before being loaded onto a 5 ml His-trap nickel column (GE Life Sciences) according to Chapter 2, section 2.5.1. The column was then washed with 20 ml of loading buffer to elute any non-specifically bound protein. The CBSV Tanza Ham1 was eluted from the column in 1.4 ml fractions using an imidazole concentration gradient of loading buffer (20 mM imidazole) to elution buffer (1 M imidazole) set up on an AKTA machine. Fractions B7 – B12 gave a relatively high UV absorption to indicate protein elution (Fig. 5.8). The separate fractions were run on an SDS-PAGE gel; confirming the purification of a protein of the expected size (25 kDa) with a relatively high purity (Fig. 5.9). These fractions were pooled and dialysed into the storage buffer at a 1:120 v/v ratio overnight according to Lin et al., (2001). The dialysed 5 ml protein sample was centrifuged in a 10 kDa Vivaspin centrifugal concentrator (Vivaproducts) to a final volume of 2 ml. The protein concentration was estimated to be 0.5 mg/ml using a Bradford assay according to the protocol described in Chapter 2, section 2.5.2.
Figure 5.8: AKTA traces of UV absorption (blue trace) from fractions eluted from the His-trap nickel chromatography column during CBSV Tanza Ham1 protein purification. A range of imidazole concentrations were produced through a gradient of % elution buffer (orange trace). Fractions B7 – B12 produced relatively high levels of UV absorption to indicate protein elution.

Figure 5.9: SDS-PAGE gel electrophoresis of pooled, dialyzed fractions (D), and separate fractions B7 – B12 that were eluted from the AKTA machine during CBSV Tanza Ham1 protein purification at a range of imidazole concentrations. The presence of a 25 kDa protein indicates successful purification of the CBSV Tanza Ham1 protein (25 kDa). The PageRuler ladder (Thermo Scientific) is shown in lane L.
5.2.2.5 Attempts to purify CBSV Tanza Ham1 protein for X-ray crystallography

Several attempts were made to purify sufficient CBSV Tanza Ham1 protein to be used for X-ray crystallography and subsequent protein structure determination. To produce crystals, the protein must have a high level of purity and at least 5 mg of the protein is required. The CBSV Tanza Ham1 protein purification above produced a 2 ml sample of 0.5 mg/ml, hence a total of 1 mg of protein, which was not sufficient for crystallography. Two further attempts were made to repeat the CBSV Tanza Ham1 protein expression and purification. The first attempt was carried out as previously described. SDS-PAGE analysis demonstrated that the extracted protein was of the correct size and had relatively high concentration and purity (Fig. 5.10). However, when the fractions containing the protein were pooled and dialysed overnight into the storage buffer, the protein had precipitated out of solution, which meant that it could not be used to form crystals.

![SDS-PAGE gel electrophoresis of attempts to purify sufficient CBSV Tanza Ham1 protein for X-ray crystallography](image)

Figure 5.10: SDS-PAGE gel electrophoresis of attempts to purify sufficient CBSV Tanza Ham1 protein for X-ray crystallography. Lanes correspond to pooled, dialyzed fractions (D), and separate fractions B2 – B13. A protein of approximately 25 kDa protein was successfully purified in the first attempt (left), however the protein precipitated out of solution during overnight dialysis into the storage buffer. The PageRuler ladder (Thermo Scientific) is shown in lane L.

A second attempt was carried out as previously described, SDS-PAGE analysis demonstrated that no protein of the correct size had been extracted and there were multiple bands of different sizes to indicate the presence of contaminant proteins (Fig. 5.11). Due to the difficulties in obtaining sufficient quantity of CBSV Tanza Ham1 protein, it was decided that protein structure determination through X-ray crystallography would require optimisation that was not possible within the time constraints of this project. Although insufficient for X-ray crystallography, the quantities of CBSV Tanza Ham1 protein which had been purified, were sufficient to perform enzyme assays to identify potential pyrophosphohydrolase activity.
5.2.3 Expression and purification of UCBSV Kikombe Ham1 protein

To determine whether the UCBSV Kikombe Ham1 protein also has pyrophosphohydrolase activity, the UCBSV Ham1 sequence was also cloned into the POPINF expression vector, expressed in *E. coli* and purified using the same method as previously described. A Bradford assay was determined the concentration of the purified UCBSV Ham1 protein to be 0.31 mg/ml. SDS-PAGE gel electrophoresis demonstrated that a protein of the correct size (25 kDa) had been purified to indicate the presence of the UCBSV Kikombe Ham1 protein and the sample has a relatively high purity, without significant contaminant protein bands (Fig. 5.12). Patrick Green, an MSc student at the University of Bristol, performed the cloning, expression and purification of the UCBSV Kikombe protein, under my supervision.

Figure 5.11: SDS-PAGE analysis of the second attempt shows multiple bands of different sizes and a lack of 25 kDa protein, which demonstrates that the expression and purification of CBSV Tanza Ham1 protein was unsuccessful in this attempt. The PageRuler ladder (Thermo Scientific) is shown in lane L.

Figure 5.12: SDS-PAGE gel electrophoresis of purified UCBSV Kikombe Ham1 protein. Lanes correspond to pooled, dialyzed fractions (D) and the PageRuler ladder (Thermo Scientific). A protein of approximately 25 kDa protein was successfully purified. The expression and purification of the UCBSV Ham1 protein was performed by the MSc student Patrick Green.
5.2.4 Pyrophosphohydrolase enzyme assays

To determine whether the CBSV Tanza and UCBSV Kikombe Ham1 proteins have pyrophosphohydrolase activity, an enzyme assay experiment was designed according the schematic in Fig. 5.13. If U/CBSV Ham1 proteins have pyrophosphohydrolase activity, they would hydrolyze the β-phosphate bond in the nucleotide triphosphate to produce a nucleotide monophosphate and a pyrophosphate molecule. The γ-phosphate bond in pyrophosphate would then be hydrolyzed in a coupled reaction by inorganic yeast pyrophosphatase to form two phosphate molecules. The production of phosphate (Pi) is measured through the addition of the malachite dye in the PiColourLock kit (Innova Biosciences), which changes colour from orange to green once bound to phosphate. The colour change was measured at OD = 655nm.

Figure 5.13: Schematic for reactions in the pyrophosphohydrolase enzyme assay. In reaction 1, U/CBSV pyrophosphohydrolase would hydrolyse the β-phosphate bond in the nucleotide triphosphate to produce a nucleotide monophosphate and pyrophosphate molecule. Then in reaction 2, yeast pyrophosphatase would hydrolyse the γ-phosphate bond in pyrophosphate to produce two phosphate molecules. Finally, in reaction 3, the presence of phosphate is detected colourmetrically by the addition of malachite dye, which changes colour from orange to green in the presence of phosphate.
5.2.4.1 Phosphate concentration quantification

To calculate the phosphate production from absorbance readings, a standard curve of OD \text{\textsubscript{655}} nm readings of known phosphate concentration were set up. A standard phosphate solution is supplied with the PiColourLock kit; this was used to set up tubes with a range of phosphate concentrations: 5, 15, 25, 35 and 85 μM and the malachite dye was added. For each concentration, 100 μl was transferred to a 96 well plate and the absorbance (OD \text{\textsubscript{655}} nm) values measured. The OD \text{\textsubscript{655}} nm absorbance readings were then plotted against known phosphate concentrations (Fig. 5.14). Linear regression was performed, and the equation of line was found to be: \( y = 0.0122x + 0.0418 \). Where \( y \) = Absorbance 655nm value and \( x \) = Pi concentration (μM). Rearrangement of this formula enables the calculation of phosphate concentration in test samples from their absorption values as follows:

\[
\text{Phosphate concentration} = \frac{(\text{OD}_{655} \text{ reading} - 0.0418)}{0.0122}.
\]

Figure 5.14: Absorbance 655 nm values plotted against phosphate concentration (μM) of the phosphate standard. Linear regression identified that the linear relationship to be \( y = 0.0122x + 0.0418 \). Rearrangement of this formula enables the phosphate concentration of test samples to be calculated from their absorption values.
5.2.4.2 CBSV Tanza Ham1 pyrophosphohydrolase activity with dITP

Preliminary enzyme assay experiments were performed to identify whether the CBSV Tanza Ham1 has pyrophosphohydrolase activity with the non-canonical nucleotide dITP and if so which substrate concentration is optimal. The preliminary experiment was set up with dITP at a concentration range of: 0.05, 0.2, 0.8 and 1.5 mM. A standard assay (300 μl) involved incubation of dITP (various concentrations) with 1.3 μg of CBSV Tanza Ham1 protein in the reaction buffer at 37°C for 20 mins according to Lin et al., (2001). Control assays were set up where no CBSV Tanza Ham1 protein was added. Samples were incubated at 37°C for 20 mins. The reactions were stopped by addition of 75 μl of PiColourLock Goldmix with accelerator and colour development was stabilised after 2 mins by the addition of 30 μl of stabiliser. For each 300 μl sample, three 100 μl aliquots were added to three separate wells in a 96-well plate. The absorbance (655 nm) values for the assays with no CBSV Tanza Ham1 were subtracted from absorbance values with CBSV Tanza Ham1. Absorbance values were then used to calculate phosphate concentration using linear relationship as previously described.

The assay results indicate that CBSV Tanza Ham1 does have pyrophosphohydrolase activity with dITP and phosphate release increases with dITP concentration from 0.05 – 0.2 mM and phosphate release decreases with higher dITP concentrations of 0.8 to 1.5 mM (Fig. 5.15). Therefore, the optimal substrate concentration of 0.2 mM was chosen for subsequent enzyme assays, as this was associated with relatively high levels of pyrophosphohydrolase activity.

![Figure 5.15: Incubation of 1.3 μg CBSV Tanza Ham1 protein with a range of dITP concentrations resulted in the release of phosphate demonstrating pyrophosphohydrolase activity. Phosphate release increased with dITP concentrations of 0.05 – 0.2 mM, whereas phosphate concentration decreased at dITP concentration of 0.8 - 1.5 mM dITP.](image-url)
5.2.4.3 Heat inactivation of CBSV Ham1 pyrophosphohydrolase activity

To confirm that the CBSV Ham1 pyrophosphohydrolase activity detected with dITP is dependent on the protein structural integrity, a heat inactivation experiment was performed whereby the CBSV Tanza Ham1 protein was subjected to different heat treatments of: 70°C – 10 mins, 95°C – 10 mins and 95°C – 1 hour. Enzyme reactions were set up as described above, with 1.3 μg CBSV Tanza Ham1 was incubated with 0.2 mM dITP at 37°C for 20 mins. The phosphate concentrations produced from the reactions are shown in Fig. 5.16. The incubation of active CBSV Tanza Ham1 with 0.2 mM dITP resulted in a phosphate concentration of: 137 μM, whereas heating CBSV Tanza Ham1 at 95°C for 10 mins – 1 hour resulted in 41 – 43% reduction in phosphate concentration, indicating inactivation of activity. This demonstrates that CBSV Tanza Ham1 activity with dITP is dependent on the Ham1 protein integrity.

Figure 5.16: Enzyme assay results from the heat inactivation experiment to test for loss of CBSV Tanza Ham1 pyrophosphohydrolase activity with dITP. Incubation of 0.2 mM dITP with active CBSV Tanza Ham1 protein (1.3 μg) resulted in a phosphate concentration of 136 μM. Heating the CBSV Tanza Ham1 at 95°C for 10 mins – 1 hour resulted in a 41 – 43% reduction in phosphate concentration, indicating inactivation of its pyrophosphohydrolase activity. Control assays were set where BSA protein (1.3 μg) was added, which produced a comparable phosphate concentration to assays where CBSV Tanza Ham1 had been heat inactivated, indicating that BSA could be used as a control for addition of protein to assay samples. Low background phosphate concentrations were found in negative controls: 1) containing 0.2 mM dITP in reaction buffer with the addition of 0.1 Units of yeast inorganic pyrophosphatase, 2) 0.2 mM dITP in reaction buffer and 3) water.
5.2.4.4 Enzyme assay controls

Negative control assay samples were tested to ensure that colour change in assay reactions containing CBSV Tanza Ham1 were due to genuine enzymatic activity, rather than artificial sources of phosphate. To control for the addition of protein to assay samples 1.3 μg of the protein standard: Bovine serum albumin (BSA) was incubated with 0.2 mM dITP. This produced a phosphate concentration of 39 μM, which is comparable to phosphate concentration of 50 μM in assays with heat inactivated CBSV Tanza Ham1 (Fig. 5.16). As there was limited U/CBSV Ham1 protein available, BSA was therefore used as a negative control for protein addition in subsequent assays. Low phosphate concentrations were also produced in other negative controls containing: 1) 0.2 mM dITP and 0.1 units yeast pyrophosphatase, 2) 0.2mM dITP in reaction buffer and 3) water. These controls confirmed that the increase in phosphate concentration during CBSV Tanza Ham1 incubation with dITP was not due to artificial sources of phosphate.

5.2.4.5 U/CBSV Ham1 activities with a range of nucleotides

To test and compare CBSV and UCBSV Ham1 pyrophosphohydrolase activities with a range of nucleotide substrates an experiment was set up whereby 1.3 μg of each protein was incubated with 0.2 mM nucleotides at 37°C for 20 mins. A range of nucleotides was tested including the non-canonical XTP and dITP and canonical nucleotides dGTP, GTP, dATP, ATP, dCTP, CTP, dTTP and UTP. As U/CBSV genomes are composed of RNA, both deoxyribose and ribose triphosphate nucleotides were tested to investigate potential differences in pyrophosphohydrolase specificity. The non-canonical nucleotide ITP was also tested. However, ITP was highly unstable and associated with high background levels of phosphate release in negative controls and so this substrate was not included in results. The absorbance values for the assays with BSA were subtracted from absorbance values with U/CBSV Ham1 proteins added. Three independent assays were performed. Phosphate concentrations were calculated from absorbance values according to the linear relationship previously described in section 5.2.4.1.

Mean phosphate concentrations produced in three independent experiments are shown in Fig. 5.17. The results show that CBSV and UCBSV Ham1 proteins have the highest activities with the non-canonical nucleotides XTP and dITP and relatively low activities with canonical nucleotides: UTP, dTTP, dATP, dCTP, CTP and ATP. Both UCBSV and CBSV Ham1 proteins are relatively active with the canonical nucleotides dGTP and GTP. Overall CBSV and UCBSV Ham1 appear to have similar activities. There does not appear to be a significant preference for deoxyribose or ribose nucleotides.
5.2.4.6 Statistical analysis of U/CBSV Ham1 activities

To determine whether U/CBSV Ham1 activities with non-canonical and canonical nucleotides are significantly different, one-way ANOVA tests were performed. There were no outliers in the data, as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. All data are normally distributed tested by Shapiro-Wilk’s test ($p > 0.05$), except for data from dGTP with CBSV Ham1 reactions. To account for this, data were transformed through inversion to ensure that they were normally distributed, this transformation did not affect the outcome of ANOVA results. For both CBSV and UCBSV data the assumption of homogeneity of variances was violated, as assessed by Levene’s test for equality of variances ($p < 0.002$). To account for this the post-hoc Games Howell one-way ANOVA was performed. The outcome from one-way ANOVA tests for CBSV and UCBSV Ham1 proteins are shown in

Figure 5.17: CBSV Tansa and UCBSV Kikombe Ham1 proteins have higher ITPase activities with non-canonical, mutagenic nucleotide triphosphates (NTP): XTP and ITP, compared with canonical NTPs. Purified CBSV Tansa and UCBSV Kikombe Ham1 proteins (1.3 μg) were incubated with 0.2 mM substrates at 37°C for 20 min in 50 mM Tris-HCl, pH 8.5, 1 mM DTT, 50 mM MgCl2, according to (Lin et al., 2001). Activity was measured colourimetrically. One-way Games Howell ANOVA analyses demonstrated that the CBSV Ham1 protein has significantly higher activities with the non-canonical nucleotide triphosphates XTP and dITP compared to canonical nucleotide triphosphates UTP, dTTP, dATP dCTP, CTP and ATP ($* = p < 0.05$) and that the UCBSV Ham1 protein has significantly higher activities with XTP and dITP compared to dTTP, dATP dCTP, and ATP ($* = p < 0.05$). Each result is the mean phosphate concentration from three separate experiments ($n = 3$) ± S.E.
Tables Appendix 5.2 – 5.3 respectively. This statistical analysis shows that incubation of CBSV Tanza Ham1 with non-canonical nucleotides XTP and diTP resulted in significantly higher phosphate concentrations than with canonical nucleotides UTP, dTTP, dATP, dCTP and ATP (p < 0.05). Meanwhile incubation of UCBSV Kikombe Ham1 with the non-canonical nucleotide diTP resulted in a significantly higher phosphate concentration than with canonical nucleotides dTTP and ATP (p < 0.05) and incubation with XTP resulted in higher phosphate concentration than with ATP (p < 0.05). Differences between diTP and XTP and the remaining canonical nucleotides were not significantly different (p > 0.05). Although CBSV and UCBSV Ham1 proteins both appear to have relatively highly levels of activity with dGTP and GTP, this was not significantly higher than their activities with the remaining canonical nucleotides (p > 0.05). A two-way ANOVA found that that the CBSV and UCBSV Ham1 proteins do not have significantly different activities to each other, $F(9, 40) = 0.661, p = 0.738$.

5.2.5 Potential U/CBSV Ham1 activity with m$^7$G cap triphosphate

The above assay experiment demonstrated that U/CBSV Ham1 proteins have relatively high levels of activity with the canonical nucleotides: dGTP and GTP. Unlike viral RNA, host mRNA has a 7-methylguanosine (m$^7$G) cap, which functions to regulate host mRNA nuclear export, prevent exonuclease degradation, promote translation and intron excision (Ramanathan et al., 2016). Several pathogenic bacteria encode pyrophosphohydrolases that degrade host mRNA through conversion of the 5’-terminal triphosphate to a monophosphate (Foley et al., 2015; Messing et al., 2009). This led to the hypothesis that perhaps the viral Ham1 proteins may target the m$^7$G triphosphate caps on host mRNA by hydrolyzing the β-pyrophosphate bond to produce m$^7$G monophosphate, which then cannot be used to cap host mRNA. This potential reduction in host mRNA capping could contribute towards reduced host mRNA processing and host gene expression. To test this hypothesis an enzyme assay experiment was performed according to the protocol previously described, containing 1.3 μg CBSV or UCBSV Ham1 proteins incubated with 0.2 mM: m$^7$G triphosphate cap analog (Promega), diTP and dATP. The absorbance values for the assays with BSA were subtracted from absorbance values with U/CBSV Ham1 proteins added. Phosphate concentrations were calculated according to the linear relationship previously described. The phosphate concentrations produced in the assay samples are shown in Fig. 5.18. This showed that both CBSV and UCBSV Ham1 proteins have negligible pyrophosphohydrolase activity with the m$^7$G triphosphate cap analog, compared with high levels of activity with diTP and low activity with dATP. This indicates that U/CBSV Ham1 proteins do not hydrolyze m$^7$G triphosphate caps on host mRNA.
After showing that the CBSV Tanza Ham1 protein has pyrophosphohydrolase activity, I decided to investigate whether this activity requires a functional SHR motif. Gall et al., (2013) have previously demonstrated that mutating the SHR motif to Serine-Alanine-Alanine (SAA) abolishes hITPA in vitro ITPase activity. To investigate the role of SHR motif in CBSV Ham1, the plasmid POPINF_CHam_Mut was constructed containing CBSV Tanza Ham1 sequence with a SHR → SAA mutated motif. This plasmid could then be used to express and purify the mutated CBSV Tanza Ham1 protein so that it could be used in enzyme assays to compare the pyrophosphohydrolase activities of mutated and wild-type CBSV Tanza Ham1 proteins.

### 5.2.6.1 Construction and sequence verification of POPINF_CHam_Mut

To mutate the SHR motif in the CBSV Tanza Ham1 sequence, the motif was first identified in the CBSV Tanza Ham1 sequence through nucleotide sequence alignments with related ITPA/Ham1 sequences as shown in Chapter 4, section 4.2.1. The POPINF_CHam_Mut was constructed using the same In-Fusion cloning method previously described for the construction
of POPINF_CHam_WT. High-fidelity Phusion PCR was performed to amplify the mutated Ham1 sequence (617 bp) with the POPINF_CBSVHam1_Fwd/Rev primers and the CBSV_mutHam IC plasmid as template. The CBSV_mutHam IC was constructed in Chapter 4, section 4.2.2 and already encodes the SHR → SAA Ham1 mutation. The PCR fragment containing the mutated CBSV Ham1 sequence was cloned into the POPINF plasmid, which had been digested with Kpn I and Hind III. Transformant colonies were screened for the presence of POPINF plasmids containing the mutated Ham1 sequence using blue/white screening. White colonies were cultured, and PCR was performed using the T7_Fwd and POPINF_CBSVHam1_Rev primers, which identified four cultures containing the correct size insert. Plasmids were extracted from two of these cultures and PCR was performed, which confirmed the presence of the correct sized insert. Two of these plasmids were sent for Sanger sequencing using the T7_Fwd primer, which targets the insertion site. One plasmid was found to contain the CBSV Tanza Ham1 sequence, including the SHR → SAA Ham1 mutation (Appendix 5.4). This verified the successful construction of the POPINF_CHam_Mut plasmid.

5.2.6.2 Expression of mutated CBSV Ham1 sequence in E. coli

To express the mutated CBSV Ham1 sequence in E. coli, the POPINF_CHam_Mut plasmid was transformed into E. coli strain BL21 (DE3). Transformant colonies were cultured in 2 L of LB media containing 100 μg/ml carbenicillin until the OD reached approximately 0.8, protein expression was then induced by the addition of IPTG according to Chapter 2, section 2.5.1. Unfortunately, several attempts to express and purify the mutated Ham1 protein were unsuccessful, as no protein of the correct size was eluted from the His-trap nickel column. This meant that the mutated CBSV Ham1 protein could not be used in enzyme assays to determine the requirement of the SHR motif for CBSV Ham1 pyrophosphohydrolase activity. Attempts to purify the mutated CBSV Tanza Ham1 protein were performed by Patrick Green (MSc student) under my supervision.
5.3 Discussion

In this Chapter the substrate activity of the CBSV Tanza and UCBSV Kikombe Ham1 proteins were investigated through \textit{in vitro} pyrophosphohydrolase enzyme assays. The respective viral Ham1 sequences were expressed in \textit{E. coli} and proteins of the correct size were extracted with sufficient purity and concentration to be used in enzyme assays. Enzyme assays involved incubating the viral Ham1 proteins with a range of nucleotide triphosphate substrates and measuring the release of phosphate colorimetrically to indicate pyrophosphohydrolase activity. This showed that both CBSV Tanza and UCBSV Kikombe Ham1 proteins have similar pyrophosphohydrolase activities that are specific for non-canonical nucleotides compared with canonical nucleotides. This activity is unusual in viruses and has been widely characterised in prokaryotic and eukaryotic ITPA proteins. This presents a fascinating new insight in U/CBSV viral evolution and potential host interactions during U/CBSV infections. Although this pyrophosphohydrolase activity has been demonstrated \textit{in vitro}, the \textit{in vivo} functions of these viral Ham1 proteins requires further investigation.

5.3.1 Higher U/CBSV Ham1 pyrophosphohydrolase activities with non-canonical nucleotides

Three independent enzyme assay experiments were performed to compare CBSV Tanza and UCBSV Kikombe Ham1 activities with a range of nucleotide triphosphate substrates. This showed that the viral Ham1 proteins do have pyrophosphohydrolase activities, and their activities are highest with the non-canonical nucleotides XTP and dITP, compared with relatively low activities with canonical nucleotides: UTP, dTTP, dATP, dCTP, CTP and ATP. It is likely that U/CBSV Ham1 proteins are also active with ITP, however high levels of non-enzymatic, chemical breakdown of ITP prevented this being tested. The pattern of U/CBSV Ham1 preference for non-canonical nucleotides over canonical is similar to activities of related homolog proteins including Mj0226 from \textit{M. jannaschii} (Cho et al., 1999) and human ITPA (Lin et al., 2001). CBSV and UCBSV Ham1 proteins show relatively high levels of activity with the canonical nucleotides dGTP and GTP. The two ITPA homologs: Mj0226 and hITPA also show the highest activity with dGTP and GTP compared with other canonical nucleotides (Cho et al., 1999; Lin et al., 2001). However, the relative U/CBSV Ham1 activities with dGTP and GTP are much higher than reported for Mj0226 and hITPA. This led to the hypothesis that perhaps the activity may be due to potential U/CBSV Ham1 activity with m$^7$G cap triphosphate that has a similar structure to dGTP and GTP and is required to cap host mRNA. However, an enzyme assay experiment demonstrated that U/CBSV Ham1 proteins were not active against a m$^7$G cap triphosphate analog and so do not appear to function in this way. An alternative hypothesis for
U/CBSV Ham1 activity with (d)GTP is that it may interfere with G-protein defense signaling. In plants, G-proteins are key components in signaling pathways that trigger defense responses to pathogens (Trusov and Botella, 2016). For instance, the Arabidopsis thaliana G-protein complex is involved with triggering programmed cell death in response to CMV and TuMV infection (Brenya et al., 2016). G-proteins require GTP to activate downstream effector molecules and so perhaps U/CBSV Ham1 proteins reduce cellular levels of GTP to limit this activation. Alternatively, U/CBSV Ham1 proteins may deplete cellular (d)GTP levels to reduce host DNA replication and expression. This potential function is found in other systems, for instance the bacterial T4 coliphage virus encodes a pyrophosphohydrolase protein, which is active against the canonical nucleotide dCTP. During genome replication, the phage utilizes the non-canonical nucleotide deoxy-5-hydroxymethylcytosine as an alternative to dCTP and so dCTPase activity functions to reduce dCTP availability that is required for host DNA replication (Snyder et al., 1976). Another explanation for the relatively high activity of U/CBSV Ham1 proteins with (d)GTP may be due to lower selective pressure to maintain the U/CBSV Ham1 binding cleft structure so that it is unfavourable to (d)GTP binding. UCBSV and CBSV Ham1 proteins did not show a detectably higher activity with deoxyribose compared with ribose triphosphate nucleotides, which was also found for Mj0226 and hITPA enzymatic activities (Cho et al., 1999; Lin et al., 2001). An explanation for this lack of preference was provided by Stenmark et al., (2007) who modelled the structure of the hITPA protein bound to ITP and demonstrated that the nucleotide ribose hydroxyl groups are orientated towards the solvent and are therefore not highly involved with binding. It should be noted that although these results demonstrate that CBSV and UCBSV Ham1 proteins have pyrophosphohydrolase activities with nucleotide triphosphate substrates in vitro, it remains unknown whether these proteins function as pyrophosphohydrolases in vivo, during U/CBSV infection and similarly their potential in vivo substrates are also unknown.

5.3.2 Similar CBSV and UCBSV Ham1 predicted protein structures and activities

Predictive models of the CBSV Tanza and UCBSV Kikombe Ham1 protein structures revealed that despite relatively low levels of sequence identity between the two viral Ham1 sequences (46%), they are both predicted to form similar structures to hITPA, with the conserved SHR motif is in the binding cleft between N and C terminal lobes. Overall the CBSV and UCBSV Ham1 proteins have similar substrate specificities; a two-way ANOVA found no significant differences in the activities of the two proteins, which is surprising as they share relatively low sequence identity of 46%. Explanations for this include that while pyrophosphohydrolase motifs have
been conserved, areas of the U/CBSV Ham1 sequences with low levels of identity may serve other functions, which are specific to CBSV or UCBSV. Alternatively, the divergent areas of the Ham1 sequences may have evolved through genetic drift of a common ancestral Ham1 sequence or the CBSV and UCBSV Ham1 sequences may have been acquired separately from two independent integrations of a host sequence. As pyrophosphohydrolase motifs are not widely found in plant viral genomes, perhaps separate acquisition in CBSV and UCBSV lineages is unlikely.

5.3.3 Optimal dITP concentration for CBSV Tanza Ham1 dITPase activity

A preliminary experiment was performed to identify the optimal dITP substrate concentration to use in enzyme assays with CBSV Tanza Ham1. This showed that the dITP concentration of 0.2 mM was associated with highest levels of phosphate release, whereas dITP concentrations of 0.8 mM and 1.5 mM were associated with lower phosphate release. There are several potential reasons for the lower phosphate release at higher dITP concentrations in these assays, including inhibition. ITPase activity of hITPA is also lower at ITP concentrations above 0.8 mM, which is suggested to be due to the non-enzymatic chemical break down of ITP into IDP, which is a strong ITPA inhibitor (Chern et al., 1969). Therefore, it is possible that at dITP concentrations above 0.8 mM, there may be higher levels of ITP chemical break down, resulting in a higher IDP concentration, which may have led to inhibition of CBSV Tanza Ham1 ITPase activity.

5.3.4 Requirement of CBSV Tanza Ham1 protein structure for dITPase activity

To test for the requirement for CBSV Tanza Ham1 protein structural integrity for its pyrophosphohydrolase activity with dITP a heat inactivation experiment was performed. This revealed that heating the CBSV Tanza Ham1 protein at 70°C for 10 mins was not sufficient to fully inactivate the protein, as this treatment was associated with relatively high levels of phosphate release compared with controls containing the control protein BSA. Heating the CBSV Tanza Ham1 protein at 95°C for 10 mins was sufficient to reduce phosphate release by 41% of active CBSV Ham1 level and was comparable to assays containing the control protein BSA. This demonstrates that CBSV Tanza Ham1 pyrophosphohydrolase activity with dITP is dependent on protein structural integrity which is denatured by heating at 95°C for 10 mins. This is a relatively high inactivation temperature, compared with hITPA which is denatured by heating to 53 – 57°C (Stepchenkova et al., 2009). This may indicate that CBSV Tanza Ham1 is
relatively stable at higher temperatures, which may be due to the need to preserve protein structural integrity at the high temperatures that cassava leaves are exposed to in the field.

5.3.5 Future experiments to further investigate U/CBSV Ham1 activities

Following on from these findings, additional further experiments would involve calculating the kinetic parameters of CBSV Tanza and UCBSV Kikombe Ham1 activities with a range of nucleotides, to identify values for catalytic turnover (Kcat) and binding specificity (Km). These experiments could be performed alongside ITPA homologs: hITPA, yeast Ham1 and E. coli Rdgb proteins to determine their relative pyrophosphohydrolase activities. HPLC analysis could be performed to confirm that nucleotide monophosphates are the primary products of U/CBSV Ham1 pyrophosphohydrolase activity with nucleotide trisphosphates. In addition an experiment could be performed to compare pyrophosphohydrolase activity in U/CBSV infected and non-infected plant cell lysates according to the method described in Ji et al., (2017). Ji et al., (2017) use an ITPA specific chimeric dinucleotide substrate, which replaces the pyrophosphate group of ITP with ATP and the release of ATP due to ITPase hydrolysis of the phosphate bond is detected via luciferase luminescence.

5.3.6 Expression and purification of the wild-type CBSV Tanza Ham1 in E. coli

In this Chapter several attempts were made to express and purify wild-type CBSV Tanza Ham1 protein in E. coli for X-ray crystallography and subsequent protein structure determination. Purification attempts resulted in protein samples, which were sufficiently concentrated for enzyme assays but not for X-ray crystallography, the reasons for this are described below. To proceed with crystallization a minimum of 5 mg of highly pure protein is required. In the first purification attempt, SDS-PAGE analysis showed that the purified protein is of the expected 25 kDa size to indicate the presence of the CBSV Tanza Ham1 protein. The 2 ml protein sample had a protein concentration of 0.54 mg/ml. There were only a few faint bands of sizes different to 25 kDa on the SDS-PAGE gel, which indicates a relatively pure sample with low levels of contaminant proteins. The total amount of protein purified from this extraction was approximately 1 mg, which although sufficient for enzyme analysis was not sufficient for crystallization. Two further attempts were made to purify sufficient CBSV Tanza Ham1 protein for crystallization. In the first attempt, protein with relatively high purity and concentration was extracted however during overnight dialysis into the storage buffer, the protein precipitated out of solution. Once precipitated, the protein could not be used for crystallization, as there will be non-homogenous protein aggregation, which would interfere with the formation of
uniform crystals. Proteins are unstable and precipitate out of solution when they lack the appropriate cofactors or stabilizing ligands or they are not in an optimal buffer (Vedadi et al., 2010). Therefore, to reduce precipitation, buffers with a range of conditions could be tested to identify an optimal storage buffer to ensure CBSV Tanza Ham1 solubility. In the second purification attempt SDS-PAGE analysis revealed a lack of a protein with the expected 25 kDa. It was subsequently found that this lack of expression may be due to growing the BL21 (DE3) culture from glycerol stock and that cultures should be inoculated with freshly transformed E. coli colonies. Due to time constraints, it was decided that X-ray crystallography was beyond the time constraints of this project. Future work to determine the structure of both CBSV and UCBSV Ham1 proteins using X-ray crystallography would determine whether these proteins form similar structures to related ITPA proteins. If so, I expect the U/CBSV Ham1 proteins to form a binding cleft between two globular lobes. It would be especially informative to produce crystals of these U/CBSV Ham1 proteins alone and in complex with nucleotide substrates, as this would enable identification of nucleotide binding sites, the structural basis behind substrate specificity and the conformational changes that occur upon substrate binding.

5.3.7 Expression and purification of mutant CBSV Tanza Ham1 protein in E. coli

In this Chapter, attempts were also made to express and purify the CBSV Tanza Ham1 sequence with a mutated SHR → SAA pyrophosphohydrolase motif so that pyrophosphohydrolase activities of wild-type and mutated protein could be compared in enzyme assays. The mutated CBSV Tanza Ham1 sequence was successfully cloned into the expression POPINF vector; verified through PCR and Sanger sequences. However, attempts to express the mutated Ham1 sequence in the E. coli strain BL21 (DE3) were unsuccessful. SDS-PAGE analysis of purifications demonstrated the absence of a protein band with the correct size, indicating that the mutated CBSV Tanza Ham1 sequence was not expressed. There are multiple potential reasons for this lack of expression. For instance, it is possible that mutating the positively charged Histidine and Arginine in the SHR motif to neutrally charged Alanine in the SAA motif may have affected electrostatic interactions required for proper folding of the CBSV Tanza Ham1 protein and the mis-folded protein may be toxic to E. coli. To test for this, future attempts to express and purify the mutated Ham1 sequence could be performed using the E. coli expression strain Lemo21 (DE3), which enables the expression of a toxic protein sequence to be tuned to a level that can be tolerated by the host (Wagner et al., 2008). Cell-free systems such as PURExpress (New England Biolabs) could also be used to synthesize mutated Ham1 sequence, which would circumvent toxicity issues (Shimizu et al., 2001). Finally, in addition to the SHR motif there are
other highly conserved amino acids associated with pyrophosphohydrolase activity, which could be mutated in the CBSV Tanza Ham1 protein sequence, which may permit expression of a mutated Ham1 sequence in *E. coli*.

**Acknowledgements:** I would like to thank Dr. Paul Race and his PhD students Carl Marsh and Laurence Mashio from the School of Biochemistry at the University of Bristol, who provided me with training in protein purification. I would also like to thank Patrick Green (MSc student) for his assistance with UCBSV Ham1 and mutant CBSV Ham1 protein purification.
Chapter 6: General discussion

6.1 Introduction

Food security is one of the biggest global challenges currently facing humanity. By 2050, it is predicted that the global population will increase to 9.1 billion, there will be less land available for food production and unpredictable impacts of climate change on farming (Godfray et al., 2010). In SSA, climate change scenario models predict that by 2050, temperatures will be consistently higher and rainfall will fluctuate, leading to reduced crop yields and hence jeopardising per capita calorie availability (Ringler et al., 2010). Historically, compared with cereals, orphan crops, such as cassava have not received high levels of scientific investment and attention, despite their importance for food and economic security in developing countries (Pingali, 2012). In recent years, cassava has been recognised as a ‘survivor’ crop, which can withstand high temperatures and fluctuations in rainfall and so offers unique climate change adaptation opportunities in Africa (Jarvis et al., 2012). In addition, the growing commercialisation of cassava production in several SSA countries is enabling farmers to raise incomes and reduce poverty (Dixon et al., 2003). Unfortunately, cassava production in SSA, is mainly limited by two viral diseases: CMD and CBSD, which together are estimated to cause annual losses worth US$1 billion (IITA, 2014a) and adversely affect food security in the entire region (Patil et al., 2015). Since the early 2000s, the CBSD pandemic has spread across many East and Central African countries (Legg et al., 2011) and so controlling this disease is now a major research priority. Unfortunately, cassava varieties with high levels of CBSD resistance are not currently available to farmers (Abaca et al., 2013) and relatively little is known about the fundamental biology of U/CBSVs and their interactions with hosts and vectors. U/CBSV ICs are urgently needed to screen diverse cassava germplasm for broad CBSD resistance and to characterise viral gene functions that can then be targeted in anti-viral strategies. Two U/CBSV ICs were previously constructed at the University of Bristol in Duff-Farrier and Mbanzibwa et al., (unpublished) and Nanyiti (2016). However, before other research groups can use these ICs, their utility needed to be tested in terms of sequence stability during plasmid propagation in E. coli, their ability to cause infections in cassava and their tractability to enable viral gene function characterisation. The results from these investigations are presented in Chapters 3 – 5 and are further discussed below.
6.2 Tractability of the UCBSV Kikombe and CBSV Tanza ICs

To enable the long-term use of U/CBSV ICs by different research groups, they need to fulfil several criteria. Firstly, they should show high levels of genome sequence stability during multiple rounds of plasmid propagation in *E. coli* so that the viral inoculum remains consistent over time. Secondly, they should be able to generate efficient, standardised infections. Finally, there should also be a range of U/CBSV ICs available, which represent the high diversity of U/CBSV genomes found in the field (Alicai et al., 2016b) so that breeders can screen for broad CBSD resistance, that is durable across different agro-ecological areas (IITA, 2014b; Kawuki et al., 2016). To date, there has only been one published report of successful U/CBSV IC construction (Pasin et al., 2017) and manipulation (Shan et al., 2018). Therefore, the U/CBSV ICs described in this study represent an important milestone in CBSD research and the lessons learnt during their construction, propagation and manipulation can be applied to the construction of other U/CBSV ICs, as well as *Potyviridae* ICs in general.

*Potyviridae* ICs are often associated with sequence instability during propagation in *E. coli*, which is predicted to be due to the presence of cryptic promoters that lead to the expression of cytotoxic proteins (Jakab et al., 1997). *E. coli* responds to sequence toxicity through the deletion or disruption of toxic sequences, which render the ICs unusable (Bedoya and Darós, 2010). In Chapter 3, the UCBSV Kikombe and CBSV Tanza ICs were tested for sequence stability during propagation in a range of *E. coli* strains. PCR analysis detected continued sequence instability in the middle of the Nla region of the UCBSV Kikombe genome and at the start of the Nlb region in the CBSV Tanza genome, despite the presence of three introns in the CBSV Tanza IC. Therefore, future work to stabilise the UCBSV Kikombe and CBSV Tanza ICs should focus on the Nla and Nlb regions respectively. These genome regions are not reported to be unstable in other *Potyviridae* ICs, as previous reports have identified sequence instability in the P3 and CI regions of the PVY and PPV genomes (Chikh Ali et al., 2011; Jakab et al., 1997; Maiss et al., 1992). Therefore, it is not possible to predict the viral genome regions where sequence instability may occur and so newly constructed *Potyviridae* ICs should be thoroughly screened for sequence instability throughout the entire viral genome.

Previous reports have shown that different *E. coli* strains replicate IC plasmids with differing levels of sequence stability (Al-Allaf et al., 2013; Boyer and Haenni, 1994; Jakab et al., 1997). This was also shown in Chapter 3, as propagation of the UCBSV Kikombe and CBSV Tanza ICs in different *E. coli* strains was associated with different levels of sequence stability. For instance, propagation of the UCBSV Kikombe IC in the *E. coli* strain α-select was relatively stable, however sufficient plasmid quantities could not be extracted. Higher quantities of the UCBSV Kikombe
IC plasmid could be extracted from the *E. coli* strains DH5α and ccdB strains, however this was associated with significant sequence instability. Similarly, propagation of the CBSV Tanza IC in the *E. coli* strain C43 was associated with sequence instability, whereas propagation in the strains ccdB and TOP10 was sufficiently stable, as shown in Chapters 3 and 4. The ability of specific *E. coli* strains to replicate the U/CBSV ICs with higher sequence stability could not be readily predicted from the *E. coli* genotypes. Therefore, this work further suggests that *Potyviridae* IC sequence stability should be tested in a range of *E. coli* strains to identify the most suitable strain for long-term storage and plasmid propagation.

To detect U/CBSV IC sequence instability, two diagnostic tests were performed in this study: 1) restriction enzyme digest band pattern and 2) PCR of overlapping viral genome sections. When ICs are digested with restriction enzymes, the band pattern can be compared to the expected band pattern if the genome sequence is intact. Bands that are the incorrect size indicate that sequence deletions or insertions have occurred. The strength of these bands indicates the proportion of IC plasmids containing sequence rearrangements. On the other hand, PCR is more sensitive and can detect rare ICs containing sequence rearrangements, which make-up a low proportion of the plasmids and may not significantly affect IC infectivity. Therefore, both tests are recommended to test the sequence integrity of other *Potyviridae* ICs.

The insertion of introns into viral ICs is widely used to improve viral genome sequence stability during propagation in *E. coli* (Desbiez et al., 2012; Johansen, 1996; López-Moya and García, 2000; Sun et al., 2017; Yang et al., 1998). The CBSV Tanza IC, which contains introns in the P3, CI and NLb genome regions also exhibited higher sequence stability during propagation in *E. coli*, compared with the intron-less UCBSV Kikombe IC, which is highly unstable. Therefore, this technique can be used to at least partially stabilise other U/CBSV ICs.

The U/CBSV ICs may also show sequence instability during propagation in *A. tumefaciens*, which was not assessed in this study. To totally avoid sequence instability during U/CBSV IC construction and propagation, the most efficient approach may be to use cell-free *in vitro* cloning techniques, such as Gibson-assembly (Bordat et al., 2015) or InFusion cloning (Tuo et al., 2015) to combine viral genome PCR fragments into IC plasmids containing full-length viral genomes. These plasmids could then be used to generate infectious *in vitro* viral transcripts for mechanical inoculation (Tuo et al., 2015) or transiently transformed into plant cells through micro-projectile particle bombardment (López-Moya and García, 2000).

Previous reports have shown that cassava can be challenging to agroinfiltrate due to the hydrophobic surface of cassava leaves (Ariyo et al., 2006; Díaz et al., 2014). In this project, the *in vivo* CBSV Tanza IC was used to agroinfiltrate cassava and successfully generated CBSV
infections, which were detectable by RT-PCR amplicon sequencing. Cassava leaves were successfully agroinfiltrated when the surfactant Pluronic F-68 (Life Technologies) was added to the agroinfiltration suspension to improve soaking and carborundum was used to mechanically damage cassava leaves to provide *A. tumefaciens* access to plant cells. The use of an abrasive to increase agroinfiltration efficiency has been patented (Giritch et al., 2010). The infection efficiency of the cassava agroinfiltration method used in this study was relatively low, with only 25 – 50% of agroinfiltrated plants developing infections. In contrast, infection efficiencies of 100% have been reached when grafting CBSV infected axillary buds onto non-infected cassava rootstocks (Wagaba et al., 2013) and when using micro-projectile bombardment to inoculate cassava with CMG ICs (Ariyo et al., 2006). The low infection rates of cassava agroinfiltration may be due to a low ability for *A. tumefaciens* to transform cassava cells or for the CBSV Tanza IC to be efficiently replicated and translated. Future work is needed to improve the efficiency of cassava agroinfiltration with U/CBSV ICs, before this can be reliably used by researchers. Future optimisation attempts could involve the use of a positive control plasmid, containing GFP on an *in vivo* expression cassette, to give more immediate results on agroinfiltration efficiency. Similar positive control plasmids have been used to optimise *Agrobacterium*-mediated production of transgenic cassava (Chauhan et al., 2015) and grapevine plants (Li et al., 2006). However, optimisation attempts may fail to improve cassava agroinfiltration efficiency and so alternative inoculation methods may be required. Micro-projectile bombardment has been widely used to efficiently infect cassava with CMG ICs (Ariyo et al., 2006; Briddon et al., 1998; Rothenstein et al., 2005) and so future work could test whether this technique could also be used to efficiently infect cassava with U/CBSV ICs. Ideally, all U/CBSV ICs used in comparative infection studies should use the same inoculation method, to avoid bias in different levels of infection efficiency.

It is important to note that the use of U/CBSV ICs to screen cassava lines at research institutes in African countries poses a highly significant biosafety risk (Brewer et al., 2018). In areas surrounding these institutes there are likely to be high numbers of cassava/alternative hosts and whitefly vectors, making the chances and consequences of containment breaches extremely high. Therefore, to limit this risk, U/CBSV ICs should only be used by highly-trained researchers at African institutes that adhere to strict biosafety containment measures and at institutes that are not in close proximity to potential plant hosts or vectors (Brewer et al., 2018).
6.3 Model hosts used to study U/CBSV infections

In this study the model host *N. benthamiana* was routinely used to study U/CBSV infections, as it is easily agroinfiltrated and rapidly develops symptoms. However, as CBSV causes severe necrosis during infections of *N. benthamiana* at 14 – 18 dpi, it was difficult to accurately compare infections. *N. benthamiana* is highly susceptible to viral infections due to a disruptive insertion in its Rdrp gene; an important component of host RNAi silencing of viral RNA (Bally et al., 2015). Ideally future U/CBSV sequence characterisation would involve infecting cassava, however due to the continued challenges of infecting cassava with *in vivo* U/CBSV ICs and the length of time before symptoms appear, model hosts are still required. Therefore, future work could identify a model Euphorbia host that develops U/CBSV infections that more closely resemble cassava infections.

6.4 Visualisation of CBSV Tanza replication and movement

In Chapter 3, the marker gene GFP was successfully inserted into the CBSV Tanza IC through yeast homologous recombination of overlapping PCR fragments and rescue of yeast plasmids into *E. coli*. GFP was inserted into two different positions in the CBSV genome: 1) between Ham1 – CP and 2) between CP – 3’UTR. Both these ICs were able to systemically infect *N. benthamiana* and GFP fluorescence was used to track CBSV replication and movement *in planta*. However, compared with the unmodified CBSV Tanza IC, infections with the CBSV Tanza GFP ICs were associated with reduced infectivity, low viral titers and GFP was deleted from a proportion of viral transcripts by 15 dpi. Therefore, the presence of GFP in the CBSV Tanza genomes appears to significantly reduce viral infectivity and so selection acts to delete GFP from the CBSV Tanza genome. Similar results have been reported in other studies. German-Retana et al., (2000) have shown that inserting marker genes into the *Lettuce mosaic virus* IC leads to reduced viral accumulation and pathogenicity and several studies have reported the deletion of marker genes from viral genomes during plant infections (Arazi et al., 2001; Beauchemin et al., 2005; Dawson et al., 1989; Guo et al., 1998). Previous studies have also found that marker genes can be retained within viral genomes when marker genes are inserted into a different genome position (German-Retana et al., 2000) or a shorter marker gene sequence is inserted (Chapman et al., 2008). Therefore, these strategies could be tested to improve the stability of marker gene expression during CBSV infections. It may also be beneficial to include a nuclear localisation signal sequence to target and retain the marker protein within the nuclei of plant cells (Chiu et al., 1996), as this compartmentalisation may
reduce interference with viral replication and or viral polyprotein processing in the cytoplasm. Once stabilised and successfully transformed into cassava, U/CBSV ICs containing marker genes could be used to identify the sites that U/CBSVs bind to whitefly stylets during vector transmission. Similarly, U/CBSV ICs containing marker genes could be used to determine whether UCBSV and CBSV co-infect cells during mixed infections, which would shed light on potential interactions between the two viral species.

6.5 U/CBSV sequence characterisation

*N. benthamiana* infections with CBSV typically develop severe necrosis and plant death, whereas UCBSV infections lack necrosis and develop mild mosaics (Mohammed et al., 2012). To characterise the viral genome sequences associated with this differential CBSV and UCBSV symptom development, two chimeric ICs were constructed. The first was constructed in Chapter 3 and consists of the CBSV Tanza genome with a UCBSV Kikombe CP replacement (CBSV_Tanza_UCP) and the second was constructed in Chapter 4 and consists of the CBSV Tanza genome with a UCBSV Kikombe Ham1 replacement (CBSV_UHam). Compared with CBSV Tanza, *N. benthamiana* infections with CBSV_Tanza_UCP also developed necrosis, suggesting that the CP genome region does not contain the necrosis sequence determinant. However necrosis development in CBSV_Tanza_UCP infections was milder and developed later in infection, suggesting that the CBSV CP may be associated with high levels of necrosis during early infection. The Ham1 region was then investigated. *N. benthamiana* infections with CBSV_UHam totally lacked necrosis but instead developed systemic leaf distortion and mottling, this indicates that CBSV Ham1 is a sequence determinant for the development of necrosis associated with CBSV infection of *N. benthamiana*.

To identify the specific region of the CBSV Tanza Ham1 associated with necrosis, the CBSV Tanza Ham1 SHR pyrophosphohydrolase motif was mutated to form the CBSV_mutHam IC, in Chapter 4. *N. benthamiana* infections with CBSV_mutHam did not develop necrosis, which suggests that the Ham1 is specifically associated with necrosis development. The mechanism by which CBSV Tanza Ham1 induces necrosis in *N. benthamiana* is not known. It is possible that the CBSV Tanza Ham1 protein or its activities are recognised by a resistance (R) protein in *N. benthamiana* and this recognition leads to an attenuated or delayed hypersensitive response (HR) that does not effectively limit replication to inoculated cells, leading to systemic viral movement so that later on in infection, once viral titers have surpassed a given threshold, lethal systemic hypersensitive response (LSHR) is induced that leads to plant death. Similar cases in other viruses include the P3 protein in two Soybean mosaic virus strains, which is recognised
by the Rsv1 protein leading to the induction of LSHR during soybean infections (Hajimorad et al., 2005). Alternatively, the CBSV Tanza Ham1 protein may perturb cellular functions and thereby indirectly lead to the development of necrosis in *N. benthamiana*.

To further confirm that CBSV Tanza Ham1 elicits necrosis development in *N. benthamiana*, ideally a reciprocal chimeric IC would be constructed, consisting of the UCBSV Kikombe genome with CBSV Tanza Ham1 replacement. However, this is not currently possible due to the sequence instability of the UCBSV Kikombe IC. It should be noted that CBSV Tanza Ham1 may not be associated with the development of root necrosis in cassava, as cassava infections with both CBSV and UCBSV develop root necrosis (Mohammed et al., 2012). In future work, it would be highly valuable to identify the U/CBSV sequences associated with the development of root necrosis during cassava infections.

In addition to symptom development, viral accumulation was also altered during *N. benthamiana* infections with the chimera ICs: CBSV_Tanza_UCP and CBSV_UHam. Quantification of viral titers using ELISA in Chapter 3, demonstrated that CBSV viral proteins peak at 14 dpi and then decrease after 21 dpi, whereas titers in CBSV_Tanza_UCP infections are lower than CBSV Tanza at 7 – 14 dpi and then increase to higher levels at 21 - 28 dpi. Similarly, qPCR in Chapter 4 demonstrated that CBSV_UHam transcript levels are lower than CBSV Tanza at 10 dpi, transcripts are at comparable levels in both infection types at 10 – 14 dpi and then CBSV_UHam transcripts increase to higher levels at 18 dpi. This indicates that both the CBSV CP and Ham1 may be required for high levels viral accumulation during early CBSV infections. However, this may also be due to reduced abilities for the UCBSV CP and Ham1 proteins to interact with CBSV proteins and/or be cleaved from the CBSV polyprotein. At later time points in infection, sufficient quantities of the UCBSV CP and Ham1 proteins may be present to compensate for the loss of CBSV CP and Ham1 proteins and thereby enable higher viral accumulation. Interestingly, U/CBSV genome analysis has only found evidence of intraspecific recombination between CBSV and UCBSV strains but not between the two species (Mbanzibwa et al., 2011b; Ndunguru et al., 2015). This may indicate that if UCBSV-CBSV genome chimeras are formed naturally through interspecific recombination during mixed infections of cassava in the field, they have reduced fitness and are not positively selected and maintained during infections. In contrast, CMGs ACMV and EACMV have undergone extensive interspecific recombination to generate highly pathogenic recombinant viruses (Beachy et al., 2000; Munoz et al., 1997). Alternatively, it is also possible that the cloning process itself may have reduced the infectivity of the chimera CBSV_Tanza_UCP and CBSV_UHam ICs, indeed all CBSV Tanza ICs which have been manipulated in this study demonstrated reduced
pathogenicity. This may imply that the genes manipulated do indeed play a specific role in pathogenicity. Perhaps the manipulations may negatively affect viral replication and/or polyprotein translation and processing and thereby indirectly lead to reduced pathogenicity. It is also possible that during cloning, point mutations were introduced to CBSV Tanza genome, in areas outside of the manipulated regions that negatively affect viral pathogenicity. Non-intentional mutagenesis has occurred during the construction of other viral ICs. For instance, Nielsen et al., (2003) reported that point mutations that were unintentionally introduced into the Porcine reproductive and respiratory syndrome virus (PRRSV) IC lead to reduced pathogenicity. To investigate whether non-intentional mutagenesis had occurred, future work should involve Sanger sequencing the entire viral genomes of the ICs constructed and propagated in this study.

6.6 Potential interactions between CBSV and UCBSV during mixed infections

Mixed U/CBSV infections of cassava are relatively common in the field, making up 34 – 50% of tested infections in Kenya (Kathurima et al., 2016), Tanzania (Mbanzibwa et al., 2011a) and Uganda (Ogwok et al., 2014). However the ability for UCBSV and CBSV to interact during mixed infections is currently unclear. Wagaba et al., (2013) demonstrated in a bud-grafting experiment that 100% of rootstocks of the cassava cultivar 6044 become infected, when UCBSV infected buds were grafted first, followed by grafting with CBSV infected buds two weeks later. Whereas simultaneous grafting with UCBSV and CBSV infected buds led to 67% of rootstocks becoming infected and grafting with CBSV infected buds, followed by UCBSV infected bud led to 17% of rootstocks becoming infected. This suggests that the prior presence of UCBSV increases CBSV infectivity and so perhaps UCBSV proteins are able to provide synergistic functions that increase CBSV replication and/or movement. For instance, this study suggests that the UCBSV Kikombe CP can transcapsidate the CBSV Tanza genome, as CBSV_Tanza_UCP was able to systemically infect N. benthamiana when inoculated through agroinfiltration and mechanical back-inoculation, in Chapter 2. If transcapsidation occurs during mixed U/CBSV infections of cassava in the field, this may have significant implications for vector transmission. Currently U/CBSV vector specificity is unclear, CBSV can be transmitted by whitefly (B. tabaci) in a semi-persistent manner, however relatively low transmission efficiencies of up to 53% have been shown when using 100 insects per plant, which does not match the rapid spread of CBSD observed in the field (Maruthi et al., 2017). Initially, aphid transmission was not considered possible for U/CBSVs, as they do not encode HC-Pro proteins, which in other Potyviridae are involved with aphid transmission. In the ‘bridge hypothesis’ model, the PTK motif in the HC-Pro
protein binds to the CP DAG motif and the KiTC motif in the HC-Pro protein binds to receptors in the aphid. This interaction acts as a bridge to allow the retention of virus particles in the aphid stylet or foregut (Valli et al., 2018). Recently, Ateka et al., (2017) have reported the presence of the highly conserved DAG and PTK motifs in CBSV CPs and KiC motif in CBSV P1s, which are not present in UCBSV genomes. This raises the possibility for CBSV to be transmitted by aphids in addition to whiteflies, which requires further investigation, as CBSD vector control may need to be adapted to include aphids. Additionally, if U/CBSV transcapsidation occurs in the field, UCBSV genomes encapsidated in CBSV CP may acquire aphid transmissibility. In other viruses, transcapsidation of Barley yellow dwarf viruses has been shown to occur during mixed infections in the field and under laboratory conditions, which broadens vector specificity of the Ny-Mav strain (Creamer and Falk, 1990). The ability for U/CBSV transcapsidation to occur during mixed infections could be tested in an experiment where cassava leaf material containing both UCBSV and CBSV is sampled and mono-clonal antibodies are used to bind UCBSV or CBSV CPs in ELISA. RNA could then be extracted from bound viral particles and RT-PCR performed to determine the species of the encapsidated viral genome.

6.7 U/CBSV Ham1 proteins

UCBSV, CBSV and Euphorbia ringspot virus (EuRV) are the only plant viruses reported to encode Ham1 proteins; their functions are currently unknown. The potential in vivo and in vitro functions of the U/CBSV Ham1 proteins were investigated in Chapters 4 and 5. In Chapter 4, the CBSV_HKO IC, which lacks a Ham1 sequence was able to systemically infect N. benthamiana. This is perhaps unexpected because RNA viruses typically have streamlined genomes, encoding a small number of proteins that serve multiple, indispensable functions (Holmes, 2003). Non-functional sequences are rapidly lost from viral genomes, as demonstrated by the deletion of GFP from CBSV_Tanza_GFP1/2 genomes during CBSV Tanza infections in Chapter 3. Willemsen et al., (2017) have also shown that exogenous sequences are only maintained in TEV genomes through multiple N. tabacum passages, if they provide advantageous functions for TEV. All reported U/CBSV genome sequences contain Ham1 sequences with highly conserved pyrophosphohydrolyase motifs (Alicai et al., 2016b). This suggests that U/CBSV Ham1 proteins must provide selectable functions during infection and/or transmission. It is possible that U/CBSV Ham1 proteins function specifically during cassava infections and/or vector transmission and may serve as a Euphorbia host adaptation. This would explain why the only other virus, EuRV reported to encode a Ham1 protein also infects
a Euphorbia host (Knierim et al., 2016). To test whether the CBSV Ham1 protein is required for replication in cassava, an experiment could be performed whereby *N. benthamiana* and cassava protoplasts are transformed with: 1) CBSV_Tanza IC, 2) CBSV_HKO IC or 3) empty vector, according to (Nguyen et al., 2010; Wu et al., 2017). The protoplasts would then be incubated to enable viral replication and after a given amount of time qPCR could be performed to compare viral transcript abundance. If CBSV_Tanza is able to accumulate in both *N. benthamiana* and cassava protoplasts, whereas CBSV_HKO can only accumulate in *N. benthamiana* protoplasts, this would indicate that the CBSV Ham1 protein is required for CBSV replication within cassava cells.

As U/CBSVs are not found in South America where cassava originates from, they must be native to Africa. Therefore, perhaps acquisition of Ham1 sequences enabled U/CBSVs to ‘jump’ host from a native plant species to cassava in Africa. To date, U/CBSVs have not been detected in any hosts that are native to Africa and so it is challenging to identify where U/CBSVs may have ‘jumped’ host from and what the ancestral viral genome sequences might be. There are examples whereby plant viruses appear to have acquired genes that extend the viral host range. For example, *Citrus tristeza virus* (CTV) appears to have acquired three non-conserved genes: p33, p18 and p13 that are expendable for systemic infections of two model citrus plant species but enable CTV to infect sour orange, lemon and calamondin plants (Tatineni et al., 2011). It is not known where U/CBSVs acquired their Ham1 sequences from; it has been hypothesised that they may originate from plant ITPA sequences (Mbanzibwa et al., 2009a).

However phylogenetic analysis in Chapter 4, revealed that the U/CBSV Ham1 sequences are not highly related to plant, fungal, bacterial, animal or archaeal ITPA sequences. Therefore it was not possible to determine the origin of the U/CBSV Ham1 sequences. The lack of similarity with sequences may indicate separate acquisition of the Ham1 sequences in CBSV and UCBSV lineages and significant sequence divergence. Interestingly, it appears that CMGs may have also co-opted host genes that enhance pathogenicity. Ndunguru et al., (2016) detected the presence of novel episomal DNA sequences in CMD infected cassava that enhance symptoms and share high levels of sequence similarity with cassava genes associated with RNA synthesis/processing and protein translation/movement. One of these DNA sequences appears to be encapsidated into virions and transmitted by whiteflies. Therefore, perhaps the co-option of host genes to serve viral functions is a common strategy used by cassava infecting viruses.

U/CBSV phylogenetic analysis in Chapter 4, also revealed that outside of conserved pyrophosphohydrolase motifs, CBSV and UCBSV Ham1 proteins share low nucleotide sequence similarity, typically of around 51 – 56% (Mbanzibwa et al., 2011b). This suggests that perhaps
U/CBSV Ham1 sequence integrations occurred separately in CBSV and UCBSV lineages, or that a single integration occurred in an ancestral virus, which has then undergone significant sequence divergence in CBSV and UCBSV species (Monger et al., 2010). Horizontal gene transfer and maintenance of the transferred sequences into plant RNA viral genomes is considered to be relatively rare (Tromas et al., 2014). Examples of this include the presence of AlkB domains in several Flexiviridae genomes (Bratlie and Drabløs, 2005), which repair RNA damage caused by oxidative methylation (van den Born et al., 2008). AlkB domains are predicted to reduce the accumulation of mutations in viral genomes during long-term Flexiviridae infections of woody plant species (van den Born et al., 2008). Similarly, U/CBSVs infect cassava for long periods of time, during multiple rounds of vegetative propagation and so perhaps their Ham1 proteins function to reduce the accumulation of mutations in viral genomes.

In humans, hITPA activity has been directly linked to reducing DNA mutation rate (Menezes et al., 2012; Waisertreiger et al., 2010). To determine whether the CBSV Ham1 protein functions to reduce the CBSV mutation rate, a deep-sequencing experiment was performed in Chapter 4, whereby *N. benthamiana* plants were infected with the CBSV_Tanza or CBSV_HKO ICs. Bioinformatic analysis of deep-sequencing reads detected no significant difference in the number of SNVs in viral RT-PCR amplicons between the two infection types, suggesting that the CBSV Ham1 protein does not function to reduce CBSV mutation rate. This agrees with previous work by Pablo-Rodriguez (2017) who found no significant difference in the number of SNVs in PVY or TMV RT-PCR amplicons generated from infections of wild-type *N. tabacum* and transgenic *N. tabacum* lines, expressing the CBSV Ham1. It is possible that U/CBSV Ham1 proteins do reduce viral mutation rates during cassava infections, but this function is not detectable during *N. benthamiana* infections.

To determine whether UCBSV and CBSV Ham1 proteins have pyrophosphohydrolase activities that can be detected *in vitro*, enzyme assays were performed in Chapter 5. Purified UCBSV Kikombe and CBSV Tanza Ham1 proteins were incubated with a range of substrate nucleotide triphosphates and the release of phosphate was measured to indicate pyrophosphohydrolase activity. This identified that both U/CBSV Ham1 proteins have pyrophosphohydrolase activities and that their activities are significantly higher with the non-canonical nucleotides XTP and dITP, compared with canonical nucleotide triphosphates. Similar activities have been reported for other ITPA proteins, including the Mj0226 from *M. jannaschii* (Cho et al., 1999) and hITPA from humans (Lin et al., 2001). Critically, this is the first time that U/CBSV Ham1 proteins have been shown to have pyrophosphohydrolase activities. It should be noted that although U/CBSV
Ham1 proteins display pyrophosphohydrolase activity with XTP and dITP in vitro, they may act on different or additional substrates in vivo. X-ray crystallography and protein structure determination is required to further confirm the enzymatic mechanism for U/CBSV Ham1 pyrophosphohydrolase activity with nucleotide substrates. Unfortunately, sufficient quantities of the CBSV Tanza Ham1 required for X-ray crystallography could not be expressed and purified from E. coli in this study, as described in Chapter 5.

This study found that U/CBSV Ham1 proteins have relatively high pyrophosphohydrolase activities with the canonical nucleotide (d)GTP, compared with characterised activities of other ITPA homologs (Cho et al., 1999; Lin et al., 2001). It is difficult to say whether U/CBSV Ham1 activity with (d)GTP serves a function during infection or whether the proteins have lost specificity due to reduced selection pressure to make (d)GTP binding unfavourable. It was predicted that the (d)GTP activity may indicate that U/CBSV Ham1 proteins target m7G caps on host RNA to reduce host gene expression, as several pathogenic bacteria use pyrophosphohydrolase activity to degrade host RNA in this way (Foley et al., 2015; Messing et al., 2009). To test this, an enzyme assay was performed where the UCBSV and CBSV Ham1 proteins were incubated with the M7G triphosphate cap analog, however no pyrophosphohydrolase activity was detected. There are other potential functions for U/CBSV Ham1 activity with (d)GTP, including interference with host G-protein defence signaling or reducing host DNA replication and expression.

To continue investigations into the potential in vivo roles of U/CBSV Ham1 proteins, it would be highly useful to determine their sub-cellular localisation during infections. For instance, localisation to cytoplasmic vesicles containing viral replication complexes would indicate roles in viral replication. The identification of U/CBSV Ham1 protein localization could be achieved by using specific labelled antibodies or using CBSV/UCBSV ICs where the Ham1 sequences are fused to fluorescent protein sequences, such as GFP. These ICs could be used to transiently transform plants, along with sub-cellular markers for cytoskeletal, endosome and plasma membrane proteins. This technique has been widely used to visualise the sub-cellular localisation of viral proteins in plant cells (Heinlein et al., 1998; Restrepo et al., 1990; Yuan et al., 2016). It would also be useful to perform protein-protein interaction experiments to identify host proteins that U/CBSV Ham1 proteins may interact with. This could provide further insights into their roles during infection. The host proteins could then be mutated to reduce U/CBSV infectivity. For instance, Gomez et al., (2018) have recently used CRISPR/Cas9 to mutate two cassava elf4E proteins, which interact with VPg during viral replication. CBSV
infections of these mutated lines developed delayed and attenuated CBSD aerial and root symptoms, indicating that mutating host factors may be a viable anti-viral strategy.

6.8 Summary

To summarise, unlike other staple food crops, cassava is able withstand high temperatures and fluctuations in rainfall and so offers unique climate change adaptation opportunities in Africa (Jarvis et al., 2012). Significant efforts to commercialise cassava production in several SSA countries should also help raise farmer income and alleviate rural poverty (Dixon et al., 2003). The use of cassava in biofuel production is particularly attractive as this could also provide a sustainable energy source (Jansson et al., 2009). However, before cassava can provide farmers with higher food, economic and fuel security in an unpredictable climate, significant efforts are needed to successfully control CMD and CBSD. A key part in the fight against CBSD will be to understand the fundamental biology of U/CBSV infections and transmission so that anti-viral strategies can be developed. In other cases, such as CMD viral ICs have been instrumental in the development and deployment of improved cassava varieties with high levels of CMD resistance (Fondong, 2017). To date, the construction of U/CBSV ICs has been circumvented by viral genome sequence instability during propagation in E. coli. Two U/CBSV ICs were recently constructed at the University of Bristol and were thoroughly tested in this study. Overall, this study found that that U/CBSV ICs continue to display high levels of genome sequence instability during propagation in E. coli and that agroinfiltration of cassava with the in vivo CBSV Tanza IC has a low inoculation efficiency. Despite these challenges, the U/CBSV ICs were used in this study to gain initial insights into the potential viral gene function and symptom determinants. Prior to this, only the gene silencing activity of the UCBSV P1 protein had been characterised (Mbanzibwa et al., 2009a). It was found that the CBSV Ham1 is associated with necrosis development during N. benthamiana infections and that U/CBSV Ham proteins have in vitro pyrophosphohydrolase activities with non-canonical nucleotides. These findings are just the beginning in understanding how U/CBSVs cause such devastating disease across SSA. Investments are now needed to overcome the continued challenges of constructing and propagating U/CBSV ICs so that they can be used to characterise U/CBSV infections and to develop cassava varieties with broad CBSD resistance that can be distributed to farmers for effective CBSD control.
6.9 Conclusions

From the investigations performed in this study, the following conclusions can be drawn:

• The UCBSV Kikombe IC displays significant sequence instability during plasmid propagation in E. coli and so sequence stabilisation is required before this IC can be used. Whereas the presence of introns in the CBSV Tanza IC enables higher sequence stability during propagation in the E. coli strains ccdB and TOP10.

• The CBSV Tanza IC can be used to agroinfiltrate and inoculate cassava, however further modifications to the cassava agroinfiltration protocol are required to improve transformation efficiency.

• CBSV Tanza CP may be involved with the induction of severe necrosis and high levels of viral accumulation during early N. benthamiana infections.

• CBSV Tanza Ham1 appears to be associated with the induction of necrosis development in N. benthamiana and may also be associated with high levels of viral accumulation during early N. benthamiana infections.

• The UCBSV Kikombe and CBSV Tanza Ham1 proteins have similar in vitro pyrophosphohydrolase activities, which are significantly higher with the non-canonical nucleotides XTP and dITP, compared with canonical nucleotides.

• The in vivo functions of UCBSV and CBSV Ham1 proteins during infection remain unclear but do not appear to reduce viral mutation rates during N. benthamiana infections.

6.10 Recommendations for future research

To gain further insights into U/CBSV infections, the following investigations are recommended:

• Overcome sequence stability issues to generate a set of stable U/CBSV ICs represent the diversity of U/CBSVs in the field and can be used to efficiently generate standardised infections in cassava.

• Use these U/CBSV ICs to screen diverse cassava germplasm for broad CBSD resistance.

• Investigate potential interactions between CBSV and UCBSV during mixed infections, particularly the potential for transcapsidation during vector transmission.

• Characterise the involvement of the CBSV CP DAG motifs in vector transmission and whether CBSV can be transmitted by aphids, in addition to whiteflies.

• Identify the in vivo functions of CBSV and UCBSV Ham1 proteins and the host proteins that they interact with to develop anti-viral strategies.
References


AGRA Planting the Seeds of a Green Revolution in Africa - World | ReliefWeb.


divergent geminivirus: evolutionary and taxonomic implications. Virus Res. 177, 35–45.


Ephraim, N., Yona, B., and Evans, A. (2015). Effect of cassava brown streak disease (CBSD) on cassava (*Manihot esculenta* Crantz) root storage components, starch quantities and starch quality


Gergerich, R.C., and Dolja, V. V. (2006). Introduction to plant viruses, the invisible foe. The Plant Health Instructor (APSnet).


IITA (2014b). SCP Updates Highlights of the progress of the “New Cassava varieties and Clean seed to Combat CBSD and CMD project (SCP).


Mukasa, S.B., Rubaihayo, P.R., and Valkonen, J.P.T. (2003). Incidence of viruses and virus like diseases of


Plantwise Knowledge Bank Mild mottle of sweet potato (Sweet potato mild mottle virus). CABI.


proteolytic processing, and genome amplification. J. Virol. 70, 7039–7048.


Potyviridae family: an enviable multitasking Helper Component that every virus would like to have. Mol. Plant Pathol. 19, 744–763.


Wylie, S.J., Adams, M., Chalam, C., Kreuze, J., López-Moya, J.J., Ohshima, K., Praveen, S., Rabenstein, F.,

Xiong, R., and Wang, A. (2013). SCE1, the SUMO-conjugating enzyme in plants that interacts with Nib, the RNA-dependent RNA polymerase of Turnip mosaic virus, is required for viral infection. J. Virol. 87, 4704–4715.


Appendices

Appendix 1.1: Blog-post article from ‘cassavavirus.wordpress.com’ on screening cassava for CBSD resistance in different agro-ecological areas of Uganda.

Taking a trip to the cassava field!

At the end of last week I was lucky enough to be invited on a trip to the field. I didn’t really know what to expect but was very excited to find out! The purpose of the trip was to collect data for the 5CP project to find out how different varieties of cassava respond to Cassava brown streak disease (CBSD) and Cassava mosaic disease (CMD) in different areas. We set off at 5.30am in the morning; the first stop was Lake Victoria to catch a ferry to the Sesse Islands. The team consisted of me, the driver (Bosco), research assistant (Gerald Adiga) and research technician (Joseph). Along the road, we saw several accidents, sadly a far too common occurrence in Uganda... Due to delays, the ferry was rammed, and by the time we arrived it was almost the evening. We raced to the agricultural school with the field trial. Here the team have planted blocks of 25 clean cassava varieties from five African countries and our job was to score them for disease symptoms. CBSD and CMD are not very common on the Sesse Islands, and so most of the plants were healthy.

After a night of drinking Guinness in a corner shop we headed out, again at 5.30am! This time we headed to the city of Mbarara in the western region. The drive was really beautiful, passing Lake Mburo National Park and mountains covered with matoke. Whilst scoring the cassava plants here we noticed a super abundance of whiteflies, which carry CBSD viruses. The weather had been particularly dry, allowing the whiteflies to breed like crazy. Fortunately, CBSD is also uncommon in this area and very few plants were diseased. The data from the 5CP project will help farmers to decide which cassava varieties offer the most protection against CBSD and CMD in their local areas; helping to protect them from the devastating yield losses caused by these diseases. Fun stuff On the way back we passed the equator line, and I got the chance to take some touristy photos. This week I also saw the Ndere dance troupe, who showcase the different dance and music styles from all over Uganda and other neighbouring countries. It was a lot of fun, some dances bore a weird resemblance to morris dancing and marching brass bands!
Breeding cassava for the next generation

Last week I helped to harvest and score cassava tubers for a breeding trial at the National Crops Resources Research Institute (NaCRRI). The trial is part of the NEXTGEN Cassava project which applies genetic techniques to conventional breeding and aims to produce new varieties with Cassava brown streak disease (CBSD) and Cassava mosaic disease (CMD) resistance. Why cassava and what’s the CBSD problem? Approximately 300 million people rely on cassava as a staple food crop in Africa. It is resilient to seasonal drought, can be grown on poor soils and harvested when needed. However cassava production is seriously threatened by CBSD, which can reduce the quality of tubers by 100% and is currently threatening the food security of millions of people.

Crossing cassava from around the world

Cassava varieties show a huge variation in traits including disease resistance. The NEXTGEN Cassava project has crossed 100 parent plants from Latin America with high quality African plants to produce new improved varieties, with higher levels of CBSD and CMD resistance. Crossing involves rubbing the pollen from one parent variety on to the female flower part (pistil) of the second parent variety to produce seeds.

Cutting back on time

The process is not easy. The complex heritability of traits in cassava means that many plants have to be screened to identify plants with the best traits. To cut down on this time, researchers from Cornell University sequenced the DNA from 2,100 seedlings and selected plants containing sequences linked to desirable traits. Screening for resistance These plants were transferred to field site in Namulonge, where there is a high level of CBSD, making it easier to spot resistant plants. After 12 months the tubers were dug up and cut into sections. Each root was scored for the severity of CBSD. Plants which show no disease symptoms have now been selected for the next stage of breeding. Eventually varieties will be tested for their performance at sites across Uganda and given to farmers for their feedback.

Time to harvest!

I helped to score and tag plants, it was hard work! I was impressed by the stamina of the workers who harvested from 8am until 3pm without a rest. I was struck by the mammoth task of breeding cassava for so many traits and by the programme manager Alfred Ozimati’s determination to get the work done as quickly as possible. Alfred is currently a PhD student at Cornell University; he kindly offered to answer these questions: What are the challenges of conventional breeding and how does sequencing help to address these? Alfred Ozimati is managing the breeding programme Typical conventional breeding cycle of cassava is 8-10 years before parents are selected for crossing. The sequencing
information allows a breeder to select parents early at the seedling stage, allowing more crossing cycles over time than conventional cassava breeding. With sequencing, the process of releasing varieties with improved CBSD and CMD resistance should take about 5 years. What are your long-term hopes for the project and the future of cassava breeding? We hope to use genomic selection routinely, to address any other challenges cassava as a crop of second importance to Uganda will face. And also, to take the technology to other East African, cassava breeding programs to faster address their major breeding constraints.
Appendix 1.3: Blog-post article from ‘cassavavirus.wordpress.com’ on using engineering cassava to confer CBSD resistance.

Using GM to fight cassava brown streak disease

Last week I helped plant a new confined field trial for genetically-modified (GM) cassava in western Uganda. The aim is to find how well the plants resist Cassava brown streak disease (CBSD). Before planting, the National Crops Resources Research Institute (NaCRRI) held discussions with people from the local government and farmers’ groups. It’s vital to engage the local community so that people are correctly informed and on-board with the project. There were certainly some very strange myths to debunk! Henry Wagaba (Head of Biosciences at NaCRRI) explained the huge losses caused by CBSD, which spoils tubers and can wipe out entire fields. CBSD is now the most devastating crop disease in Uganda and there are no resistant varieties currently available. To fight the disease, NaCRRI researchers have developed GM cassava plants, which show high levels of resistance to CBSD at sites in southern and central Uganda. This trial will test how the plants perform in the growing conditions in western Uganda. Work will also be carried out to cross the GM plants with farmer varieties to improve their growing and taste qualities.

I enjoyed getting stuck in and planting my first GM cassava! GM crops are a contentious topic in Uganda. The passing of a National Biotechnology and Biosafety law has stalled in Parliament for over three years due to disagreements. Currently GM technology is used for research on banana, cassava, maize, potato, rice and sweet potato. However these are not approved for human consumption. In nearby countries Kenya and Sudan, GM food products have been approved and many of these food products are imported into Uganda without regulation. It’s hoped the law will be passed soon to enable Ugandan farmers to reap the benefits of GM crops and protect against any potential risks.
Appendix 1.4: Blog-post article from ‘cassavavirus.wordpress.com’ on establishing a ‘clean’ cassava system in Uganda.

**Clean cassava to solve brown streak problem?**

Since arriving in Uganda, I’ve been learning a lot about the affects of Cassava brown streak disease (CBSD), which is devastating cassava production and threatening food security. The disease is spread by the whitefly insect, which picks up the virus from an infected plant and carries it to neighbouring healthy plants. Cassava plants are grown by planting stem cuttings in the ground, which go on to become new plants. If farmers use cuttings from infected plants, the new plants will also become infected. This is a big problem, as infected cuttings can be transported to new areas, spreading CBSD across large distances. What can be done? Tolerance Huge efforts are being put into a number of different solutions. These include breeding new cassava varieties, which are tolerant to CBSD. This is a very long and challenging process, as cassava plants also need to be resistant to Cassava mosaic disease (CMD) and have yield/taste properties which farmers and consumers prefer. The National Crops Resources Research Institute (NaCRRRI) has recently developed a new variety: NAROCASS1, which is tolerant to CBSD and resistant to CMD. This is now being used in areas where CBSD is particularly common and severe. Unfortunately, even tolerant cassava varieties can contain CBSD viruses and so it’s vital that farmers have access to clean cuttings.
Appendix 2.1: *Escherichia coli* genotypes used in this study.

<table>
<thead>
<tr>
<th>E. coli strain</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>DH5α</td>
<td>F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80lacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–</td>
</tr>
<tr>
<td>α-select</td>
<td>F - deoR endA1 recA1 relA1 gyrA96 hsdR17(rk -, mk + ) supE44 thi-1 phoA Δ(lacZYA argF)U169 φ80lacZΔM15λ -</td>
</tr>
<tr>
<td>C43(DE3)OverExpress</td>
<td>F– ompT gal dcm hsdS8(rB- m8-) ([DE3])</td>
</tr>
<tr>
<td>One Shot ccdB survival cells</td>
<td>F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2</td>
</tr>
<tr>
<td>TOP10</td>
<td>F[lacIq Tn10[tetR]] mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 deoR nupG recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL(StrR) endA1 λ-</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>F– ompT gal dcm lon hsdS8(rB–m8–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)</td>
</tr>
</tbody>
</table>
Appendix 3.1: Sanger sequence read generated in reaction using the CBSV_Tanza_Sec4_Fw primer (CTCAAGTGAAGAGACGGTTG) and the RT-PCR fragment amplified from a cassava plant that had been agroinfiltrated with the CBSV_Tanza IC at 42 dpi. The sequence aligns to the CBSV_Tanza IC map at positions 5580 – 5985 bp.

<table>
<thead>
<tr>
<th>CBSV</th>
<th>RT-PCR</th>
<th>Sequence Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>5551</td>
<td>1</td>
<td>AGTGGAGCTCGCATTTACTAAAGATGGCTCAAGTGAAGAGACGGTTGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5601</td>
<td>22</td>
<td>ATGGGTTTTCTCTGAGGAAGAGGCCGATGGAGACAAACAGGGAGTCCGGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5651</td>
<td>72</td>
<td>ATGGGTTTTCCCTGAGGAAGAGGCCGATGGAGACAAACAGGGAGTCCGGT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5701</td>
<td>122</td>
<td>TGTGCAGAAAAGATCAGAAGAAAATCAGAGTGTTGAAATACAAGTTGCGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5751</td>
<td>172</td>
<td>AACCGGAGACAACAAAACCCTATPGAGCATGTGTTGTTTGGACTTGGTAGG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5801</td>
<td>222</td>
<td>GCCCATTTAGGGACAAAAGGTATTAAACTGTGTTTCTCTCCATGGATCAAATG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5851</td>
<td>272</td>
<td>CGGTGACAATTTACCATTTGAGGACAGGAGGGTACTGAGCTGCTAGCTGTCG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5901</td>
<td>322</td>
<td>TGATATCGACTACTCGTGGACAGTTTGACTTTGGCCCATTGAGAAATATC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5951</td>
<td>372</td>
<td>AAGTGCAAAAAAGATAAAAAGATTATGATATAACTGTGTGTCATTGCCACG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 3.2: Sanger sequence read generated in reaction using the CBSV_Tanza_Sec6_Fw primer (GCTCTGTTTGCTGAGTG) and the RT-PCR fragment amplified from an *N. benthamiana* plant that had been mechanically inoculated with cassava plant material infected with CBSV_Tanza IC. The sequence aligns to the CBSV_Tanza IC map at positions 8149 – 8646 bp.

<table>
<thead>
<tr>
<th>CBSV</th>
<th>RT-PCR_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
<td>AGAGGATTGAAACATTGAGTATATGAGTCATTACATTGTGCTCAGCGTGA</td>
</tr>
<tr>
<td>8151</td>
<td>TGGCTTTTTCATTCAGAAGAGAGAATCGTCCATATTCTAGTTGCTCAGCGTGA</td>
</tr>
<tr>
<td>8201</td>
<td>GTGGAAAGGAGAGACGAGGTTATGAGAGACGTCCTCAATGCA</td>
</tr>
<tr>
<td>8251</td>
<td>GCTTACATCGAGAGCTTTGGTTATGATGATTTGATGACAGAGATTGAGAG</td>
</tr>
<tr>
<td>8301</td>
<td>GTATGCCTGTCTGCTGCAATCCACTTGAGATATCCACTGTTGG</td>
</tr>
<tr>
<td>8351</td>
<td>AAAAAACGGAGTTGAGAGGCTTTTACTTGGATGAATCACACTGACATCAAT</td>
</tr>
<tr>
<td>8401</td>
<td>GAGGAAATGGCTGATGAAATTTACCACACCACATCTTTCGAGAATTTACAT</td>
</tr>
<tr>
<td>8451</td>
<td>TGACTTACAAGTGTTGAGCACAGGGTCCTCAAGCATTCTGCTAAGAGAG</td>
</tr>
<tr>
<td>8501</td>
<td>AGGAAGATGTTACTAGTTAAATTCGGAATGGGATCGAAGCCACCAATTACA</td>
</tr>
</tbody>
</table>

RT-PCR_3

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>

RT-PCR_103

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>

RT-PCR_153

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>

RT-PCR_203

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>

RT-PCR_253

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>

RT-PCR_303

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>

RT-PCR_353

<table>
<thead>
<tr>
<th>CBSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>8101</td>
</tr>
<tr>
<td>8151</td>
</tr>
<tr>
<td>8201</td>
</tr>
<tr>
<td>8251</td>
</tr>
<tr>
<td>8301</td>
</tr>
<tr>
<td>8351</td>
</tr>
<tr>
<td>8401</td>
</tr>
<tr>
<td>8451</td>
</tr>
<tr>
<td>8501</td>
</tr>
</tbody>
</table>
CBSV  8551 TTTGTCACAGGAATGCAAAACACTGAAAGAAGTGAAACAAATCTTTGG
RT-PCR 403 TTTGTCACAGGGAATGCAAAACTGAAAGAAGTGAAACAAATCTTTGG
CBSV  8601 TCCCACTATTCCAATCATTTCTCGGAA
                    AGTTGATTTACCAGAACCACAAAG
RT-PCR 453 TCCCACTATTCCAATCATTTCTCGGAA
                     AGTTGATTTACCAGAACCACAG---
Appendix 3.3: Sanger sequence read generated in reaction using the Seq9_Fw primer (CCGTTAATGGAGAAGGAC) and the CBSV_Tanza_GFP1 IC. The sequence aligns to the CBSV_Tanza_GFP1 map at positions 9067 – 9902 bp, which corresponds to the GFP insertion at positions 9139 - 9867 bp (green).

<table>
<thead>
<tr>
<th>CBSV_GFP1</th>
<th>Seq9</th>
</tr>
</thead>
<tbody>
<tr>
<td>9051</td>
<td>1</td>
</tr>
<tr>
<td>9067</td>
<td>-</td>
</tr>
<tr>
<td>9081</td>
<td>-</td>
</tr>
<tr>
<td>9095</td>
<td>-</td>
</tr>
<tr>
<td>9101</td>
<td>35</td>
</tr>
<tr>
<td>9115</td>
<td>-</td>
</tr>
<tr>
<td>9129</td>
<td>-</td>
</tr>
<tr>
<td>9143</td>
<td>-</td>
</tr>
<tr>
<td>9151</td>
<td>-</td>
</tr>
<tr>
<td>9165</td>
<td>-</td>
</tr>
<tr>
<td>9180</td>
<td>-</td>
</tr>
<tr>
<td>9194</td>
<td>-</td>
</tr>
<tr>
<td>9201</td>
<td>-</td>
</tr>
<tr>
<td>9215</td>
<td>-</td>
</tr>
<tr>
<td>9230</td>
<td>-</td>
</tr>
<tr>
<td>9244</td>
<td>-</td>
</tr>
<tr>
<td>9251</td>
<td>-</td>
</tr>
<tr>
<td>9265</td>
<td>-</td>
</tr>
<tr>
<td>9280</td>
<td>-</td>
</tr>
<tr>
<td>9294</td>
<td>-</td>
</tr>
<tr>
<td>9301</td>
<td>-</td>
</tr>
<tr>
<td>9315</td>
<td>-</td>
</tr>
<tr>
<td>9330</td>
<td>-</td>
</tr>
<tr>
<td>9344</td>
<td>-</td>
</tr>
<tr>
<td>9351</td>
<td>-</td>
</tr>
<tr>
<td>9365</td>
<td>-</td>
</tr>
<tr>
<td>9380</td>
<td>-</td>
</tr>
<tr>
<td>9394</td>
<td>-</td>
</tr>
<tr>
<td>9401</td>
<td>-</td>
</tr>
<tr>
<td>9415</td>
<td>-</td>
</tr>
<tr>
<td>9430</td>
<td>-</td>
</tr>
<tr>
<td>9444</td>
<td>-</td>
</tr>
<tr>
<td>9451</td>
<td>-</td>
</tr>
<tr>
<td>9465</td>
<td>-</td>
</tr>
<tr>
<td>9480</td>
<td>-</td>
</tr>
<tr>
<td>9494</td>
<td>-</td>
</tr>
<tr>
<td>9501</td>
<td>-</td>
</tr>
<tr>
<td>9515</td>
<td>-</td>
</tr>
<tr>
<td>9530</td>
<td>-</td>
</tr>
<tr>
<td>9544</td>
<td>-</td>
</tr>
<tr>
<td>9551</td>
<td>-</td>
</tr>
<tr>
<td>9565</td>
<td>-</td>
</tr>
<tr>
<td>9580</td>
<td>-</td>
</tr>
<tr>
<td>9594</td>
<td>-</td>
</tr>
<tr>
<td>9601</td>
<td>-</td>
</tr>
<tr>
<td>9615</td>
<td>-</td>
</tr>
<tr>
<td>9630</td>
<td>-</td>
</tr>
<tr>
<td>9644</td>
<td>-</td>
</tr>
<tr>
<td>9651</td>
<td>-</td>
</tr>
<tr>
<td>9665</td>
<td>-</td>
</tr>
<tr>
<td>9680</td>
<td>-</td>
</tr>
<tr>
<td>9694</td>
<td>-</td>
</tr>
<tr>
<td>9701</td>
<td>-</td>
</tr>
<tr>
<td>9715</td>
<td>-</td>
</tr>
<tr>
<td>9730</td>
<td>-</td>
</tr>
<tr>
<td>9744</td>
<td>-</td>
</tr>
<tr>
<td>9751</td>
<td>-</td>
</tr>
<tr>
<td>9765</td>
<td>-</td>
</tr>
<tr>
<td>9780</td>
<td>-</td>
</tr>
<tr>
<td>9794</td>
<td>-</td>
</tr>
<tr>
<td>9801</td>
<td>-</td>
</tr>
<tr>
<td>9815</td>
<td>-</td>
</tr>
<tr>
<td>9830</td>
<td>-</td>
</tr>
<tr>
<td>9844</td>
<td>-</td>
</tr>
<tr>
<td>9851</td>
<td>-</td>
</tr>
<tr>
<td>9867</td>
<td>-</td>
</tr>
<tr>
<td>9881</td>
<td>-</td>
</tr>
<tr>
<td>9895</td>
<td>-</td>
</tr>
<tr>
<td>9901</td>
<td>-</td>
</tr>
<tr>
<td>CBSV_GFP1 9501</td>
<td>CAACAGGATCGAGCTTTAAGGGAAATCGATTICAAAGGGGAAACATCC</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>Seq9 435</td>
<td>CAACAGGATCGAGCTTTAAGGGAAATCGATTICAAAGGGGAAACATCC</td>
</tr>
<tr>
<td>CBSV_GFP1 9551</td>
<td>TCGGCCACAAGTTGGAATACAAACTACAATCCCACAAGCTATAATCATC</td>
</tr>
<tr>
<td>Seq9 485</td>
<td>TCGGCCACAAGTTGGAATACAAACTACAATCCCACAAGCTATAATCATC</td>
</tr>
<tr>
<td>CBSV_GFP1 9601</td>
<td>GCCGACAAGCAAAAAGACGCACTAAGCCAAACCTTCAAGGACCCGCAAC</td>
</tr>
<tr>
<td>Seq9 535</td>
<td>GCCGACAAGCAAAAAGACGCACTAAGCCAAACCTTCAAGGACCCGCAAC</td>
</tr>
<tr>
<td>CBSV_GFP1 9651</td>
<td>CATCGAAGACGGCGGGTGCAACTCGCTGATCATTATCAACAAAATACTC</td>
</tr>
<tr>
<td>Seq9 585</td>
<td>CATCGAAGACGGCGGGTGCAACTCGCTGATCATTATCAACAAAATACTC</td>
</tr>
<tr>
<td>CBSV_GFP1 9701</td>
<td>CAATGGCGATGGCGCTGCTsetupACAACTTACCTGTCACAA</td>
</tr>
<tr>
<td>Seq9 635</td>
<td>CAATGGCGATGGCGCTGCTsetupACAACTTACCTGTCACAA</td>
</tr>
<tr>
<td>CBSV_GFP1 9751</td>
<td>CATCTGCCCTTTTCAAGGATCCCCAAGAGAGACAACTGGCTCCCT</td>
</tr>
<tr>
<td>Seq9 685</td>
<td>CATCTGCCCTTTTCAAGGATCCCCAAGAGAGACAACTGGCTCCCT</td>
</tr>
<tr>
<td>CBSV_GFP1 9801</td>
<td>TCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTAT</td>
</tr>
<tr>
<td>Seq9 735</td>
<td>TCTTGAGTTTGTAACAGCTGCTGGGATTACACATGGCATGGATGAACTAT</td>
</tr>
<tr>
<td>CBSV_GFP1 9851</td>
<td>ACAAAATCGATGGCAGCAATGGACAAGGGGATGAGATTTGGAAGCTGAAATA</td>
</tr>
<tr>
<td>Seq9 785</td>
<td>ACAAAATCGATGGCAGCAATGGACAAGGGGATGAGATTTGGAAGCTGAAATA</td>
</tr>
<tr>
<td>CBSV_GFP1 9901</td>
<td>ACAAAATTTAAGGATGGTGGAGGAAACAAAGGACCGTAAACTCGAAG</td>
</tr>
<tr>
<td>Seq9 834</td>
<td>ACAAAATTTAAGGATGGTGGAGGAAACAAAGGACCGTAAACTCGAAG</td>
</tr>
</tbody>
</table>

222
Appendix 3.4: Sanger sequence read generated in reaction using the CBSV_GFP2_F2_Fw primer (GCAATGAGTAAAGGAGAAGAACTTTTCACTGGAGTT) and the RT-PCR amplicon amplified from upper systemic leaf material from *N. benthamiana* plant four agroinfiltrated with CBSV_Tanza_GFP1 at 20 dpi. The sequence aligns to the CBSV_Tanza_GFP1 map at positions 9139 – 9858 bp, which corresponds to the GFP insertion at positions 9139 – 9867 bp (green).

| CBSV_GFP1 | 9101 CAAGGGCGTTGATCCTGATTTCCTATTGATGTTCAAGCAATGAGTAAAA ||||| |
| CBSV_GFP1 | 9151 GGAAGAAGACCTTTTCACGGAGTTGTCCTCAATTGGAATTGATGGG |
| CBSV_GFP1 | 9201 TGATTTAATGGGCACAAATTTTGACAATACGCCCTAATGAGGAGCG |
| CBSV_GFP1 | 9251 CTTTTCACATTGCCAACCCTTGGCTACAACTTTCCAAGAGCCCAAGGC |
| CBSV_GFP1 | 9301 CCTGTTCTGGCCCAACCCTTGGCTACAACTTTCCAAGAGCCCAAGGC |
| CBSV_GFP1 | 9351 CTTTCAAGGATACCCAGATCATATGAAGCGGCACGACTTCTTTCAAGAGCCCAAGGC |
| CBSV_GFP1 | 9401 CACTGCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTTCAAGAGCCCAAGGC |
| CBSV_GFP1 | 9451 GGGAACTACAAGACGTGCTGAAGTCAAGTTTGAGGGAGACACCCTCGT |
| CBSV_GFP1 | 9501 CAACAGGATCGAAGGGATACCCAGATCATATGAAGCGGCACGACTTCTTTCAAGAGCCCAAGGC |

| RT_GFP4 | 1 -------------------------------------- |
| RT_GFP4 | 14 GGAAGAAGACCTTTTCACGGAGTTGTCCTCAATTGGAATTGATGGG |
| RT_GFP4 | 64 TGATTTAATGGGCACAAATTTTGACAATACGCCCTAATGAGGAGCG |
| RT_GFP4 | 114 CAACATACGGAAAAACTTTACCTTAAATTTGCACTACTGGAAAACCAAGGC |
| RT_GFP4 | 164 CCTGTTCTGGCCCAACCCTTGGCTACAACTTTCCAAGAGCCCAAGGC |
| RT_GFP4 | 214 GGAACCTGACGGATACCCAGATCATATGAAGCGGCACGACTTCTTTCAAGAGCCCAAGGC |
| RT_GFP4 | 264 CACTGCTGAGGGATACGTGCAGGAGAGGACCATCTTCTTTCAAGAGCCCAAGGC |
| RT_GFP4 | 314 GGAACCTGACGGATACCCAGATCATATGAAGCGGCACGACTTCTTTCAAGAGCCCAAGGC |
| RT_GFP4 | 364 CAACAGGATCGAAGGGATACCCAGATCATATGAAGCGGCACGACTTCTTTCAAGAGCCCAAGGC |
Appendix 3.5: Sanger sequence read generated in reaction using the **Seq11_Fw** primer (GTGGATAAAAGAAGCCAC) and the **CBSV_Tanza_GFP2** IC. The sequence aligns to the **CBSV_Tanza_GFP2** map at positions 10268 – 11116 bp, which corresponds to the GFP insertion at positions 10285 – 10999 bp (green).

<table>
<thead>
<tr>
<th><strong>CBSV_GFP2</strong></th>
<th>10251</th>
<th><strong>CGTATGC</strong></th>
<th></th>
<th><strong>CGTATGC</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Seq11</strong></td>
<td>1</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10301</td>
<td><strong>GCTGG</strong></td>
<td></td>
<td><strong>GCTGG</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>34</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10351</td>
<td><strong>AACGT</strong></td>
<td></td>
<td><strong>AACGT</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>84</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10401</td>
<td><strong>GGAAC</strong></td>
<td></td>
<td><strong>GGAAC</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>134</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10451</td>
<td><strong>CGGAC</strong></td>
<td></td>
<td><strong>CGGAC</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>184</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10501</td>
<td><strong>AACGT</strong></td>
<td></td>
<td><strong>AACGT</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>234</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10551</td>
<td><strong>GGAAC</strong></td>
<td></td>
<td><strong>GGAAC</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>284</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10601</td>
<td><strong>AACGT</strong></td>
<td></td>
<td><strong>AACGT</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>334</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
<tr>
<td><strong>CBSV_GFP2</strong></td>
<td>10651</td>
<td><strong>GGAAC</strong></td>
<td></td>
<td><strong>GGAAC</strong></td>
</tr>
<tr>
<td><strong>Seq11</strong></td>
<td>384</td>
<td>-----------</td>
<td></td>
<td>-----------</td>
</tr>
</tbody>
</table>
Appendix 3.6: Sanger sequence read generated in reaction using the CBSV_GFP2_F2_Fw primer (GCAATGAGTAAAGGAGAAGAAGTTTCTCAGTGAAGT) and the RT-PCR amplicon amplified from upper systemic leaf material from *N. benthamiana* plant two agroinfiltrated with CBSV_Tanza_GFP2 at 10 dpi. The sequence aligns to the CBSV_Tanza_GFP2 map at positions 10288 – 10958 bp, which corresponds to the GFP insertion at positions 10285 – 10999 bp (green).

<table>
<thead>
<tr>
<th>CBSV_GFP2</th>
<th>RT-PCR</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>10262</td>
<td>377</td>
<td>CTGCTATTGAAATCTCTGTTCAAGGCCAGTAAAGGAGAAGAAGTTTCTCAGTGAAGT</td>
</tr>
<tr>
<td>10312</td>
<td>409</td>
<td>GAGGTTGTCCCAATTTCTTGTTGAATTAGATGTTGATGTTAATGGGCACAA</td>
</tr>
<tr>
<td>10362</td>
<td>459</td>
<td>ATTTTCTGTCAGTGGAGAGGGTGAAGGTGATGCAACATACGGAAAACTTA</td>
</tr>
<tr>
<td>10412</td>
<td>509</td>
<td>CCCTTAAATTTATTTGCACTACTGGAAAACTACCTGTTCCATGGCCAACA</td>
</tr>
<tr>
<td>10462</td>
<td>559</td>
<td>CTTGTCACTACTTTTTATGGTGTTCAATGCTTTTCAAGATACCCAGA</td>
</tr>
<tr>
<td>10512</td>
<td>609</td>
<td>TCATATGAAGCGGCACGACTTTTCAAGAGCGCCATGCCTGAGGGATACG</td>
</tr>
<tr>
<td>10562</td>
<td>659</td>
<td>TGCAGGAGAGGACCATCCTTCTTCAAGAGCGCCATGCCTGAGGGATACG</td>
</tr>
<tr>
<td>10612</td>
<td>709</td>
<td>GCTGAAGTCAAGTTTGAGGGAGACACCCTCGTCAACAGGATCGAGCTTAA</td>
</tr>
<tr>
<td>10662</td>
<td></td>
<td>GGGAATCGATTTCAAGGAGACGGAACATCCCTCGGCCACAAAGTGAAAT</td>
</tr>
</tbody>
</table>
RT-PCR  759  GCCAATCTGATTCAAGGAGGACGGAAACATCCTCGGCCACAAGTTGGAAT
CBSV_GFP2  10712  ACACTACAAACTCCCAACAAGTGATACATCGACAGGCGACAAAGAA

RT-PCR  809  ACAACTACAACTCCCAACAAGTGATACATCGACAGGCGACAAAGAA
CBSV_GFP2  10762  GCCATCAAACCCCAACTTGAAGGCAACCCGCAACAATGGGACGCCCGGT

RT-PCR  859  GCCATCAAACCCCAACTTGAAGGCAACCCGCAACAATGGGACGCCCGGT
CBSV_GFP2  10812  GCAACTGCTGATCATTATCAACAAAATACTCCAATTGGCGAATGGCCCTG

RT-PCR  909  GCAACTGCTGATCATTATCAACAAAATACTCCAATTGGCGAATGGCCCTG
CBSV_GFP2  10862  FCTTTGAATACAGAACCATTCTGTCAACACAATCTCCCTTTCGAAA

RT-PCR  959  FCTTTGAATACAGAACCATTCTGTCAACACAATCTCCCTTTCGAAA
CBSV_GFP2  10912  GATCCCAACGAAAGAGAGACCATGTCCTTCTGGTATGCTACAGC

RT-PCR  1009  GATCCCAACGAAAGAGAGACCATGTCCTTCTGGTATGCTACAGC--
Appendix 3.7: Sanger sequence read generated in reaction using the CBSV_GFP1_del1_Fw (GTGAGATTGTGATGCCACGCGGACCAAATT) and the RT-PCR amplicon amplified from upper systemic leaf material from *N. benthamiana* plant four agroinfiltrated with CBSV_Tanza_GFP1 at 15 dpi. The sequence aligns to the CBSV_Tanza_GFP1 map at positions 8920 – 9138 bp, which corresponds to the end of the CBSV Ham1 sequence, the RT-PCR fragment is then missing the GFP sequence between positions 9139 – 9867 bp (green) and realigns at positions 9868 – 10149 bp corresponding to the start of the CBSV CP sequence. This demonstrates the detection of viral transcripts with a complete GFP deletion.

<table>
<thead>
<tr>
<th>CBSV_GFP1</th>
<th>RT-PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>8901</td>
<td>1</td>
</tr>
<tr>
<td>8951</td>
<td>36</td>
</tr>
<tr>
<td>9001</td>
<td>86</td>
</tr>
<tr>
<td>9051</td>
<td>136</td>
</tr>
<tr>
<td>9101</td>
<td>186</td>
</tr>
<tr>
<td>9151</td>
<td>224</td>
</tr>
<tr>
<td>9201</td>
<td>224</td>
</tr>
<tr>
<td>9251</td>
<td>224</td>
</tr>
<tr>
<td>9301</td>
<td>224</td>
</tr>
</tbody>
</table>
CBSV_GFP1  9351 hypertagGGGAGAGAACCCACAGTTCAGTGAAGAGAGGACCTGCCGGG
RT-PCR  224  

CBSV_GFP1  9401 CCTTTCAAGATACCCAGATCATATGAAGCGGCACGACTTCTTCAAGAGCG
RT-PCR  224  

CBSV_GFP1  9451 GGGAACTACAAGACACGTGCTGAAGTCAAGTTTGAGGGAGACACCCTCGT
RT-PCR  224  

CBSV_GFP1  9501 CAACAGGATCGAGCTTAAGGGAATCGATTTCAAGGAGGACGGAAACATCC
RT-PCR  224  

CBSV_GFP1  9551 TGGCCACAAAGTTGGAAATACAACCTACAACCTCCACAACGTATACACATCT
RT-PCR  224  

CBSV_GFP1  9601 GCCGACAAGCAAAAGAAGCGGCATCAAAGCCAACTTCAAGGACCGGAAACATCC
RT-PCR  224  

CBSV_GFP1  9651 CATCGAAGACGGCGCTGCAACTCGCTGATCATTATACAAAATACTC
RT-PCR  224  

CBSV_GFP1  9701 CAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCT
RT-PCR  224  

CBSV_GFP1  9751 CAATCTGCCCTTTCGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCT
RT-PCR  224  

CBSV_GFP1  9801 TCTTGAGTTTGTAAACAGCTGCTGGGATTACACATGGCATGGATGAACTAT
RT-PCR  224  

CBSV_GFP1  9851 ACAAAATCGATGTGCAACCAATTGACAAAGGATGAGATTAAGCTGAAATA

CBSV_GFP1  9901 ACAAAATTAAGGATGTGAGAAGGAGACCAACACTAGAACCTGAAAG

CBSV_GFP1  9951 CCCATTGGATCAAAGGAGGCTTCTGCTCTCAAGTAGCTGACTGTAATATG
RT-PCR 307 CCCATTGAATCAAAGAGGCTTCGTGCTCCTCAAGTAGCTCGTGTAAATG

CBSV_GFP1 10001 AACTACTTAAGCAACTGAAAGATGCTGGGATACAAACAAGTAAGAGGCCA

RT-PCR 357 AACTACTTAAGCAACTGAAAGATGCTGGGATACAAACAAGTAAGAGGCCA

CBSV_GFP1 10051 TGTGGAGAACCTGATGAAGGAGAAGTTGCTAGCCCAGAGTCAAGTGAAGA

RT-PCR 407 TGTGGAGAACCTGATGAAGGAGAAGTTGCTAGCCCAGAGTCAAGTGAAGA

CBSV_GFP1 10101 TGAGGAGCAACAAACAAACAAAGGAAAAGCGCCTATGGAATCACCTACTG

RT-PCR 457 TGAGGAGCAACAAACAAACAAAGGAAAAGCGCCTATGGAATCACCTACTG
Appendix 3.8: Sanger sequence read generated in reaction using the CBSV_GFP2_del_Fw primer (GACTTCCTAGCCGAAGCACAATTGTCACAAAGCACA) and the RT-PCR amplicon amplified from upper systemic leaf material from N. benthamiana plant four agroinfiltrated with CBSV_Tanza_GFP2 at 15 dpi. The sequence aligns to the CBSV_Tanza_GFP2 map at positions 10080 – 10284 bp, which corresponds to the end of the CBSV CP sequence, the RT-PCR fragment is then missing the GFP sequence between positions 10285 – 10999 bp (green) and realigns at positions 11000 – 11057 bp corresponding to the CBSV 3’UTR sequence. This demonstrates the detection of viral transcripts with a complete GFP deletion.

| CBSV_GFP2 | 10051 | ATTGAAAAACAGTTGATTTGATTTGACTTCTAGCCGAAGCCACAATTGTC |
| RT-PCR | 1 | ------------------------------------------CTAGCCGAAGCACAATTGTC |

| CBSV_GFP2 | 10101 | ACAAAAGCCAACTTGATATCAACATCAAATATTGCTGCTAATGTTGGTA |
| RT-PCR | 21 | ACAAAAGCCAACTTGATATCAACATCAAATATTGCTGCTAATGTTGGTA |

| CBSV_GFP2 | 10151 | GAAGTAAAACCTAGTTGCTTTAGCTGCTGCTAGTGATGTTGGTA |
| RT-PCR | 121 | GAAGTAAAACCTAGTTGCTTTAGCTGCTGCTAGTGATGTTGGTA |

| CBSV_GFP2 | 10201 | GTGGATAAAGGAAGGCACACTAGCCGGGATGGTTAGGCGACTAGCATAG |
| RT-PCR | 192 | GTGGATAAAGGAAGGCACACTAGCCGGGATGGTTAGGCGACTAGCATAG |

| CBSV_GFP2 | 10251 | CTACGCTGGCTGCTATTTGAAATCTCTGTTCAAAGCAAGTAAAGCCACAACG |
| RT-PCR | 171 | CTACGCTGGCTGCTATTTGAAATCTCTGTTCAAAGCAAGTAAAGCCACAACG |

| CBSV_GFP2 | 10301 | AACTTTCCTGAGGAGATTGTCACAAATCTTGGAATTAGGATGGTGATGTT |
| RT-PCR | 192 | AACTTTCCTGAGGAGATTGTCACAAATCTTGGAATTAGGATGGTGATGTT |

| CBSV_GFP2 | 10351 | ATGGGCAAAAAATTTTCTGFCAGGAGGGCTGGAAGGTTGATGCAACATA |
| RT-PCR | 192 | ATGGGCAAAAAATTTTCTGFCAGGAGGGCTGGAAGGTTGATGCAACATA |

| CBSV_GFP2 | 10401 | CGGAAAACCTTTACCCTAAATTCTTATTTGCACTACCTGGAAAAACTACCTGGT |
| RT-PCR | 192 | CGGAAAACCTTTACCCTAAATTCTTATTTGCACTACCTGGAAAAACTACCTGGT |

<p>| CBSV_GFP2 | 10451 | CATGGCACAACACTGGTCACTACTTTCTCTTTATGCTTTTCAAGCTTTTCA |
| RT-PCR | 192 | CATGGCACAACACTGGTCACTACTTTCTCTTTATGCTTTTCAAGCTTTTCA |</p>
<table>
<thead>
<tr>
<th>Gene</th>
<th>Start</th>
<th>Sequence</th>
<th>Method</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV_GFP2</td>
<td>10501</td>
<td>AGATACCCGATCTATGAAAGCCGAAAAGACCTCTCAACCAGGCCCACAGCC</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10551</td>
<td>TGGAGGAATACGTGGCAGGAGGACATCCCTCTTCAAAGGAGACCAGGGAC</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10601</td>
<td>ACAAACACCATGCAAGTCATGGGAGAGGACCCGGG</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10651</td>
<td>ATCGAGCTTAAGGGAATCGATTTCAAGGGAGGACGGAAACATCCTCGGGCA</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10701</td>
<td>CAAGTTGGAATAACACTACACCAATCCCAACGATATAATCATTGACC</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10751</td>
<td>AGCAAAAGAACGGCATCAAAGCCAACTTCAAGACCCG</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10801</td>
<td>GACGGCGGCTGCAAGCTGCTGATCATTATCAACAAAATACTCCAATTGG</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10851</td>
<td>CGATGGCCTGTCCTTTTACCAGACCTACATTACCTGTCCACACAATCTG</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10901</td>
<td>CCCTTTGAAAGATCCCAACGAAAAGAGAGACCACATGGTCCTTCTTGAG</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>10951</td>
<td>TTGTGCAACAGCAGCTGCTGGAATACACATGGCATGGGATTATA</td>
<td>RT-PCR</td>
<td>192</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>11001</td>
<td>TTAATTTGAAAGTCAAGTTTTGGTGAGTTTTAGTTAAGTTAGTTGTTTATATTTTT</td>
<td>RT-PCR</td>
<td>197</td>
</tr>
<tr>
<td>CBSV_GFP2</td>
<td>11051</td>
<td>GGTGTTGAAAGCAGATTTATATCGAGAACCTTCTCCATATCCCTTTGT</td>
<td>RT-PCR</td>
<td>247</td>
</tr>
</tbody>
</table>
Appendix 3.9: Sanger sequence read generated in reaction using the Seq9_Fw primer (CCGCTTAACCTGGAGAAGGAC) and the CBSV_Tanza_UCP IC. The sequence aligns to the CBSV_Tanza_UCP map at positions 9026 – 9968 bp, which corresponds to the UCBSV CP sequence replacement at positions 9139 – 10242 bp (purple).

<table>
<thead>
<tr>
<th>CBSV_UCP</th>
<th>Seq9</th>
</tr>
</thead>
<tbody>
<tr>
<td>9001</td>
<td>1</td>
</tr>
<tr>
<td>GCTGAGATGATGACCAGAGAAGAATATGATATCTCATCGATTTCGGGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9051</td>
<td>26</td>
</tr>
<tr>
<td>TCTGTCATTAGTGAAGACTTTTTITAAGGACTCTGAGCTATTTTCAGCTTTTG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9101</td>
<td>76</td>
</tr>
<tr>
<td>CAAAGGCGTTGATCGTATTTTTATAGATGTCTTTCAAGCTCCTTAACATCAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9151</td>
<td>126</td>
</tr>
<tr>
<td>GAGAGATACGAGACCTCTGCTTTAAAGAAGCTATGGAAAGATAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9201</td>
<td>176</td>
</tr>
<tr>
<td>TGGACCAACAAGAGAGCTTAGCTATCAAGTTTAGGAGCTAGGAGAGCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9251</td>
<td>226</td>
</tr>
<tr>
<td>CACAAGTTGACGCCATTAAATGTAATTCTCAGAAAATGAAAGAAAGA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9301</td>
<td>276</td>
</tr>
<tr>
<td>TGGAAAACAAGAGAGCCATTGGAGACAAGTATGACTGCGGAGATTGCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9351</td>
<td>326</td>
</tr>
<tr>
<td>GGATGATGATGATGATGATGAGAGAACAATTCCAGATCTGGGAAGAGAAGT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>9401</td>
<td>376</td>
</tr>
<tr>
<td>TGGATGAAGCCAGATAGTAAACAGTGATGGTATCAGAAAGCAGAATTTCC</td>
<td></td>
</tr>
</tbody>
</table>

234
Appendix 3.10: Sanger sequence read generated in reaction using the \textit{USeq10\_Fw} primer (GTGAACTGTATAGTGAATGGAACTAGTG) and the \textit{CBSV\_Tanza\_UCP} IC. The sequence aligns to the \textit{CBSV\_Tanza\_UCP} map at positions 9834 – 1068 bp, which corresponds to the \textit{UCBSV CP} sequence replacement at positions 9139 – 10242 bp (purple).
Appendix 3.1: Sanger sequence read generated in reaction using the RT-PCR amplicon amplified from upper systemic leaf material from an *N. benthamiana* agroinfiltrated with CBSV_Tanza_UCP at 14 dpi. The sequence aligns to the CBSV_Tanza_UCP map at positions 10091 – 10231 bp, which corresponds to UCBSV Kikombe CP sequence at positions 9139 – 10242 bp. This confirms that CBSV_Tanza_UCP can systemically infect *N. benthamiana*.
Appendix 4.1

Results from tblastn search of the NCBI data base using the CBSV Tanza Ham1 and UCBSV Kikombe Ham1 sequences. The five sequences with the highest e-values are shown.

<table>
<thead>
<tr>
<th>Query sequence</th>
<th>Species name</th>
<th>Classification (Kingdom/Family)</th>
<th>Annotation</th>
<th>Identity</th>
<th>e-value</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBSV Tanza Ham1</td>
<td>Jatropha curcas</td>
<td>Plant Euphorbiaceae</td>
<td>Predicted ITPA</td>
<td>57%</td>
<td>2e-73</td>
<td>XP_012077670.1</td>
</tr>
<tr>
<td></td>
<td>Hevea brasiliensis</td>
<td>Plant Euphorbiaceae</td>
<td>Predicted ITPA</td>
<td>56%</td>
<td>3e-73</td>
<td>XP_021644689.1</td>
</tr>
<tr>
<td></td>
<td>Ricinus communis</td>
<td>Plant Euphorbiaceae</td>
<td>Predicted ITPA</td>
<td>59%</td>
<td>3e-73</td>
<td>XP_015574043.2</td>
</tr>
<tr>
<td></td>
<td>Cucurbita maxima</td>
<td>Plant Cucurbitaceae</td>
<td>Predicted ITPA</td>
<td>58%</td>
<td>4e-73</td>
<td>XP_022980416.1</td>
</tr>
<tr>
<td></td>
<td>Erythranthe guttata</td>
<td>Plant Phrymaceae</td>
<td>Predicted ITPA</td>
<td>58%</td>
<td>6e-73</td>
<td>XP_012842913.1</td>
</tr>
<tr>
<td>UCBSV Kikombe Ham1</td>
<td>Basidiobolus meristosporus</td>
<td>Fungi Basidiobolaceae</td>
<td>Predicted ITPA</td>
<td>59%</td>
<td>1e-71</td>
<td>ORX99178.1</td>
</tr>
<tr>
<td></td>
<td>Acanthamoeba</td>
<td>Amoebozoa Acanthamoebida</td>
<td>Ham1 family</td>
<td>58%</td>
<td>5e-71</td>
<td>XP_004338375.1</td>
</tr>
<tr>
<td></td>
<td>Diversispora versiformis</td>
<td>Fungi Diversisporaceae</td>
<td>Hypothetical Glove_18g102 protein</td>
<td>58%</td>
<td>8e-70</td>
<td>RHZ89189.1</td>
</tr>
<tr>
<td></td>
<td>Heterostelium album</td>
<td>Amoebozoa Acytosteliaceae</td>
<td>Predicted ITPA</td>
<td>56%</td>
<td>5e-69</td>
<td>XP_020432479.1</td>
</tr>
<tr>
<td></td>
<td>Sesamum indicum</td>
<td>Plant Pedaliaceae</td>
<td>Predicted ITPA</td>
<td>57%</td>
<td>3e-68</td>
<td>XP_011073764.1</td>
</tr>
</tbody>
</table>

Appendix 4.2: Sanger sequence read generated in reaction using the Seq9_Fwd primer (CCGCTTAACTGGAGAAGGAC) and the CBSV_mutHam IC plasmid. The sequence aligns to the CBSV_Tanza map at positions 9026 – 9150 bp. Nucleotides in the SHR pyrophosphohydrolase motifs in the CBSV_Tanza map (yellow) have been mutated SAA in the CBSV_mutHam IC (turquoise). This verifies the successful construction of the CBSV_mutHam IC plasmid.
Appendix 4.3: Sanger sequence read generated in reaction using the UCBSV_Ham1_Fwd primer (GGCTTTTTCATTCCAAAGTTGAAGAAGA) and the CBSV_UHam IC plasmid. The sequence aligns to the CBSV_UHam map, including the UCBSV Ham1 sequence (8462 – 9139 bp) highlighted in green. This verifies the successful construction of the CBSV_UHam IC plasmid.

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8400</th>
<th>TGAGGAATGGCTGAGTGGAATCTTACCACCATCTTTGAGAAATTGTACA</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>342</td>
<td>TGAGGAATGGCTGAGTGGAATCTTACCACCATCTTTGAGAAATTGTACA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8450</th>
<th>TTGACTTACAAGTGCTGGACACAAAGGATTTGAGAGGAAGAGAGAGGCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>392</td>
<td>TTGACTTACAAGTGCTGGACACAAAGGATTTGAGAGGAAGAGAGGCT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8500</th>
<th>CAGTGAGAAATGGAAGCCATGATGGAATCCCAATATGCAAATGGAAGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>442</td>
<td>CAGTGAGAAATGGAAGCCATGATGGAATCCCAATATGCAAATGGAAGT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8550</th>
<th>CCCAGTGACATTTGTGACAGGTAATTTGGGGAAATTAGCAGAAGTGAAGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>492</td>
<td>CCCAGTGACATTTGTGACAGGTAATTTGGGGAAATTAGCAGAAGTGAAGT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8600</th>
<th>CATATCTGGCGATTCCAAGTGTGTTATATGCAAGAAACTAGTTATACCACX</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>542</td>
<td>CATATCTGGCGATTCCAAGTGTGTTATATGCAAGAAACTAGTTATACCACX</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8650</th>
<th>CAGTTGCAAAGCCTCGGCTTGAATAATGTAAGAAAGAAATGCCCAATTAGG</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>591</td>
<td>CAGTTGCAAAGCCTCGGCTTGAATAATGTAAGAAAGAAATGCCCAATTAGG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8700</th>
<th>AGTGAAGATGACTAAGATCCTCGGTGAGATCGTACTGTCTCCGTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>641</td>
<td>AGTGAAGATGACTAAGATCCTCGGTGAGATCGTACTGTCTCCGTA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CBSV_UHam</th>
<th>8750</th>
<th>TTAACGGCTTTTCAATGGCTTCCCGGGCAATACATTAAATGTTCTGCCAAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>UHam_RT-PCR</td>
<td>691</td>
<td>TTAACGGCTTTTCAATGGCTTCCCGGGCAATACATTAAATGTTCTGCCAAX</td>
</tr>
</tbody>
</table>
Appendix 4.4: Sanger sequence read generated in reaction using the pJET2.1Fwd primer (CGACTCACTATAGGGAGAGCGGC) and a pJET2.1 plasmid containing a cloned RT-PCR fragment, amplified from an *N. benthamiana* plant infected with CBSV_mutHam at 20 dpi. The RT-PCR sequence aligns to the CBSV_Tanza map at positions 8851 – 9027 bp. Nucleotides in the SHR pyrophosphohydrolase motifs in the CBSV_Tanza map (yellow) have been mutated SAA in the CBSV_mutHam IC infections (turquoise). This verifies that CBSV_mutHam can systemically infect *N. benthamiana*.

| CBSV_FL     | 8851 GCTCTCTGTATTTGCTTTTATGAATAAAAGTTGGTGATGATCTCTATAAT |
|            | Mut_Ham 360 GCTCTCTGTATTTGCTTTTATGAATAAAGTTGGTGATGATCTCTATAAT |
| CBSV_FL     | 8901 CTTTAAGGGTGTGTAAAGAGATGAGGATGATGCCACCGGAGACCAATT |
|            | Mut_Ham 410 CTTTAAGGGTGTGTAAAGAGATGAGGATGATGCCACCGGAGACCAATT |
| CBSV_FL     | 8951 CATTTGGGTGGGATCCAATTTTCCAGCCCTTAACTGGAGAAGGACATTT |
|            | Mut_Ham 460 CATTTGGGTGGGATCCAATTTTCCAGCCCTTAACTGGAGAAGGACATTT |
| CBSV_FL     | 9001 GCTGAGATGATGACCGAAGAGAAGAATATGATATCTCATCGATTTCGGGC |
|            | Mut_Ham 510 GCTGAGATGATGACCGAAGAGAAGAATATGATATCTCATCGATTTCGGGC |
| CBSV_FL     | 9051 TCTGTCATTAGTGAGACTTTTTTGAAGGACTCGACTTTCAGCTTTG |
|            | Mut_Ham 560 TCTGTCATTAGTGAGACTTTTTTGAAGGACTCGACTTTCAGCTTTG |
Appendix 4.5: Sanger sequence read generated in reaction using the UCBSV_Ham1_Fwd primer (GGCTTTTTTCAATCCCAAAGTTGAAGAAGA) and a pJET2.1 plasmid containing a cloned RT-PCR fragment, amplified from an N. benthamiana plant infected with CBSV_UHam. The sequence aligns to the CBSV_UHam map, including the UCBSV Ham1 sequence (8462 – 9139 bp) highlighted in green. This verifies that CBSV_UHam can systemically infect N. benthamiana.
Appendix 4.6: Qualimap coverage reports of aligned reads from CBSV_Tanza infections (C61 – 4) and CBSV_HKO infections (KO1 – 4) mapped to the 1202 bp reference sequence. There were higher coverages at the 5' and 3' ends of reference sequence and a normal distribution of coverage in the centre, which is suggested to be due to bias in fragmentation at the middle of the amplicon fragments.
Appendix 5.1: Sanger sequence read generated in reaction using the T7_Fwd primer and POPINF_CHam_WT plasmid. The sequence aligns to the POPINF_CHam_WT map at positions 2244 bp – 3168 bp, which corresponds to the insertion site of the CBSV Ham1 sequence (2390 bp – 3067 bp) and verifies successful POPINF_CHam_WT plasmid construction.

<table>
<thead>
<tr>
<th>POPINF_Map</th>
<th>WT_CBSV_Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>2201</td>
<td>TAATCCGGGACCTTTAATTCAACCCAACATATATTATAGTTAAATAA</td>
</tr>
<tr>
<td>2251</td>
<td>GAATTATTATCAAATCATTTGTATATTAATATTTAAATCTATTACTGTAAT</td>
</tr>
<tr>
<td>2301</td>
<td>TACATTTTTTTTACAATCAAAGGAGATATACCATGGCACAACCACATCACCA</td>
</tr>
<tr>
<td>2351</td>
<td>CATCACAGCAGCGGTCTGGAAGTTCTGTTTCAGGGCCCGGTGGTGGACAG</td>
</tr>
<tr>
<td>2401</td>
<td>GTCTCAGCCATCGAATGTTGCTAAGAGAGAGGAAGTTACTAGTAAAA</td>
</tr>
<tr>
<td>2451</td>
<td>TTCGGATGGGGATCGAAGCACCAATTACATTTGTCACAGGGAATGCACA</td>
</tr>
<tr>
<td>2501</td>
<td>AAACTGAAAGAAGTGAACAAATCCTTTGTCACCCACCTATTCAGGACCAA</td>
</tr>
<tr>
<td>2551</td>
<td>TCGGAAAGTTGATTTACCAGAACCACAAGGGGACAGTTGAGGAGATTATTA</td>
</tr>
<tr>
<td>2601</td>
<td>AAGAAAAAGCAGCAGTACCTGCTGAACATTGTTGGAGGGCCAGTTTCTTGTA</td>
</tr>
</tbody>
</table>
Appendix 5.2: Games-Howell one-way ANOVA tests to compare mean phosphate concentration in enzyme assay reactions with CBSV Tanza Ham1 incubated with the non-canonical nucleotides XTP and dITP and a range of canonical nucleotides.

<table>
<thead>
<tr>
<th>Protein</th>
<th>CBSV Ham1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-canonical nucleotides</td>
</tr>
<tr>
<td>Canonical nucleotides</td>
<td>Mean Pi (μM)</td>
</tr>
<tr>
<td></td>
<td>Mean Pi (μM)</td>
</tr>
<tr>
<td>dGTP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>GTP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>UTP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>dTTP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>dATP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>dCTP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>CTP</td>
<td>Difference (μM)</td>
</tr>
<tr>
<td>ATP</td>
<td>Difference (μM)</td>
</tr>
</tbody>
</table>
Appendix 5.3: Games-Howell one-way ANOVA tests to compare mean phosphate concentration in enzyme assay reactions with UCBSV Kikombe Ham1 incubated with the non-canonical nucleotides XTP and dITP and a range of canonical nucleotides.

<table>
<thead>
<tr>
<th>Protein</th>
<th>UCBSV Ham1</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-canonical nucleotides</td>
<td>XTP</td>
</tr>
<tr>
<td>Canonical nucleotides</td>
<td>Mean Pi (μM)</td>
<td>190</td>
</tr>
<tr>
<td>Mean Pi (μM)</td>
<td>dGTP</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.180</td>
</tr>
<tr>
<td>GTP</td>
<td>Difference (μM)</td>
<td>106</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.095</td>
</tr>
<tr>
<td>UTP</td>
<td>Difference (μM)</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.074</td>
</tr>
<tr>
<td>dTTP</td>
<td>Difference (μM)</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.051 *</td>
</tr>
<tr>
<td>dATP</td>
<td>Difference (μM)</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.052 *</td>
</tr>
<tr>
<td>dCTP</td>
<td>Difference (μM)</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.059</td>
</tr>
<tr>
<td>CTP</td>
<td>Difference (μM)</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.063</td>
</tr>
<tr>
<td>ATP</td>
<td>Difference (μM)</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>Sig. p value</td>
<td>0.050 *</td>
</tr>
</tbody>
</table>
Appendix 5.4: Sanger sequence read generated in reaction using the T7_Fwd primer and POPINF_Cham_Mut plasmid. The sequence aligns to the POPINF_Cham_WT map at positions 2245 bp – 2720 bp, which corresponds to the insertion site of the CBSV Ham1 sequence (2390 bp – 3067 bp). Nucleotides highlighted in yellow code for the wild-type nucleotide sequence encoding the SHR motif, nucleotides highlighted in turquoise code for the mutated SAA motif. This verifies the successful construction of the POPINF_Cham_Mut plasmid.

POPINF_Map 2351 CATCACAGCAGCGTCTGGAAGTTCTGTTTCAGGGCGTGTTGGACAG

MT_CBSV_Seq 1 ------------------------

POPINF_Map 2401 GTCTCAGCCATCGAATGTTGCTAAGAGAGAGGAAGAAGTTACTAGTAAAA

MT_CBSV_Seq 27 GTCTCAGCCATCGAATGTTGCTAAGAGAGAGGAAGAAGTTACTAGTAAAA

POPINF_Map 2451 TTCGGATGGGGATCGAAGCACCAATTACATTTGTACAGGGCGGCCACAA

MT_CBSV_Seq 77 TTCGGATGGGGATCGAAGCACCAATTACATTTGTACAGGGCGGCCACAA

POPINF_Map 2501 GCCGCTGAAAGCAGTGAAACAAATCTTTGGTCCCACTATTCCAATTTC

MT_CBSV_Seq 127 GCCGCTGAAAGCAGTGAAACAAATCTTTGGTCCCACTATTCCAATTTC

POPINF_Map 2551 TCGGAAAGTTGATTTACCAGAACCACAAGGGACAGTTGAGGAGATTATTA

MT_CBSV_Seq 177 TCGGAAAGTTGATTTACCAGAACCACAAGGGACAGTTGAGGAGATTATTA

POPINF_Map 2601 AAGAAAAGCAGCGATAGCTGCTGAAACTTTGGAGGGCAGTTCTTGTGA

MT_CBSV_Seq 227 AAGAAAAGCAGCGATAGCTGCTGAAACTTTGGAGGGCAGTTCTTGTGA

POPINF_Map 2651 GAAGACACGAGTTTTGTGGATAGCTCCTCTCCAGGGCCATA

MT_CBSV_Seq 277 GAAGACACGAGTTTTGTGGATAGCTCCTCTCCAGGGCCATA

POPINF_Map 2701 CATTAAGTGGTTTATGGGAGAGATTGATGGAATAAGTTGG

MT_CBSV_Seq 327 CATTAAGTGGTTTATGGGAGAGATTGATGGAATAAGTTGG
POPINF_Map 2751 TGGAACCATATCAGAATAAAATGGCTAGCGCTCTCTGTGTATTTGCTTTT

MT_CBSV_Seq 377 TGGAACCATATCAGAATAAAATGGCTAGCGCTCTCTGTGTATTTGCTTTT

POPINF_Map 2801 ATGAATAAGTTGTTGATGATCCTATAATCTTTAAGGGTGTTAAAGAGG

MT_CBSV_Seq 427 ATGAATAAGTTGTTGATGATCCTATAATCTTTAAGGGTGTTAAAGAGG

POPINF_Map 2851 TGAGATTGTGATGCCACGCGGACCAAATTCAGCGGGGGCAGCTCCAATTT

MT_CBSV_Seq 477 TGAGATTGTGATGCCACGCGGACCAAATTCAGCGGGGGCAGCTCCAATTT

POPINF_Map 2901 TCCAGCGCTTAACCTGGAGAAGGACATTTGCTGAGATGAGCGAAGAG

MT_CBSV_Seq 527 TCCAGCGCTTAACCTGGAGAAGGACATTTGCTGAGATGAGCGAAGAG

POPINF_Map 2951 AAGAATATGATATCTCATCGAATTTCCGGGCTCTGTCATTAGTGAGAGCTT

MT_CBSV_Seq 577 AAGAATATGATATCTGCTGCCATTTCCGGGCTCTGTCATTAGTGAGAGCTT

249