University of Bristol

MSc by Research

School of Bioscience

Investigating the function of *Cassava brown streak virus* and *Ugandan cassava brown streak virus* Ham1 proteins during infection

Author:
Patrick Green

Supervisors:
Dr. Andy Bailey,
Prof. Gary Foster

September 24th 2018

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master of Science by Research in the Faculty of Science

Word count:25,105

Abstract

Cassava is one of the most important sources of carbohydrates in sub-tropical regions of Africa. However, its annual production is continually devastated by viral diseases such as cassava brown streak disease (CBSD) and cassava mosaic disease (CMD). The presence of these diseases threatens the food and financial security of small-scale farming across the sub-tropics, where Cassava is often depended on as a drought resistant subsistence crop. Little is still understood about these diseases and their viral causes, thereby hampering resistance efforts.

The aim of this study was to characterise the function of the Ham1-like protein encoded by *cassava brown streak virus* (CBSV) and *ugandan cassava brown streak virus* (UCBSV), which cause CBSD. It has been proposed that this protein functions as a nucleotide triphosphate pyrophosphatase, due to its homology to cellular Inosine triphosphate pyrophosphatase (ITPase) proteins. It might therefore act to reduce the viral mutation rate during infection to preserve viral stability, which would be an unusual role for an RNA virus protein. To test this hypothesis, recombinant Ham1 proteins of CBSV and UCBSV were over-expressed in *E. coli* and purified by nickel-affinity chromatography, and their ITPase activity tested in a pyrophosphatase assay against a range of nucleotide substrates. It was demonstrated that both viral Ham1 proteins were active and preferred the mutagen-causing noncanonical nucleotides deoxy inosine triphosphate and xanthosine triphosphate. Therefore, Ham1 may act as ITPase proteins to protect the viral genome from mutation during infection.

To help understand why viruses benefit from encoding a Ham1-like protein, the role of host plant ITPase proteins was studied. To assess whether viruses require the presence of an ITPase protein for stable infections, the viral titres of the RNA virus *cucumber mosaic virus* (CMV) were studied in mutant *Arabidopsis* with abolished ITPase proteins. Systemic leaf CMV titres were found to be much lower in mutant plants versus WT *Arabidopsis*, suggesting ITPase proteins are somehow involved in virus stability and infectivity. However, the expression of WT *Arabidopsis* ITPase was found to not change in response to CMV infection. This provides the first proof that CBSV and UCBSV are indeed functional ITPase proteins, and that ITPase activity is important during viral propagation *in plantae*.

Dedications and Acknowledgements

I would like to thank Prof. Gary Foster and Dr Andy Bailey for allowing me to work in their group; without their supervision and expertise this project would not be possible. They have been instrumental in introducing me to scientific research and methodology. Secondly, I would like to thank Mrs Katie Tomlinson for her guidance through this work and her patience in teaching me. I would also like to thank Dr Ian Prosser, Mrs Helen Rees and other members of Lab 321 and the Biological Sciences department for their help and companionship throughout this work.

Lastly, I would like to thanks Ms Sally Dawson and Miss Alexandra Lindqvist Jones for their unwavering support and generosity, and to whom this thesis is dedicated.

Authors 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: ……

Patrick Green

Contents

List of Tables

List of Figures

List of Abbreviations

°C	Celsius degree
6k1	First kilodalton protein
6k2	Second kilodalton protein
6-MP	6-mercaptopurine
aa	Amino acid
AlkB	Alpha-ketoglutarate-dependent dioxygenase B
ATP	Adenosine triphosphate
AGO	Argonaut
AZA	Azathioprine
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BVY	Blackberry virus Y
CBSD	Cassava brown streak disease
CBSV	Cassava brown streak virus
cDNA	Complementary DNA
CI	Cylindrical inclusion body protein
CMD	Cassava mosaic disease
CMV	Cucumber mosaic virus
Col-0	Columbia-0 genetic background *Arabidopsis thaliana* plants
CP	Coat protein
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Ct	Crossing threshold
CTP	Cytosine triphosphate
CVYV	Cucumber vein yellowing virus
DAS-ELISA	Double antibody sandwich enzyme-linked immunosorbent assay
DNA	Deoxyribonucleic acid
DNase	Deoxy-ribonuclease
DPI	days post inoculation
dNTP	deoxynucleotide triphosphate
dsRNA	double strand RNA
EDTA	Ethylenediaminetetraacetic acid
eIF4E	eukaryotic translation initiation factor 4E
ER	Endoplasmic reticulum
EuRSV	Euphorbia ringspot virus
FAO	Food and Agriculture Organization
g	Acceleration of gravity
g	Grams
GFP	Green fluorescent protein
GTP	Guanine triphosphate
H_2O	Water
HAP	6-*N*-hydroxylaminopurine
HC-Pro	Helper component protease
ICTV	International Committee on Taxonomy of Viruses
IITA	The International Institute of Tropical Agriculture
ITPase	Inosine triphosphate pyrophosphatase
ITP	Inosine triphosphate
Kb	Kilo bases
L	Litre

LB	Luria-Bertani media
m	Meters
M	Molar
mRNA	Messenger RNA
NCBI	National centre for biotechnology information
NIa	Nuclear inclusion body protein A
NIb	Nuclear inclusion body protein B
NMP	Nucleotide monophosphate
nt	nucleotides
NTPs	Nucleoside triphosphates
ORF	Open reading frame
P1	First potyviral protein
P3	Third Potyviral protein
PCR	Polymerase chain reaction
P_i	Inorganic Phosphate
PIPO	Pretty Interesting Potyviridae ORF
PVY	Potato virus Y
RdRp	RNA dependent RNA polymerase
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNAi	RNA interference
RNase	Ribonuclease
rpm	Rounds per minute
rRNA	Ribosomal RNA
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Reverse transcription real-time polymerase chain reaction
SDW	Sterile deionised water
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
siRNA	small interference RNA
SNPs	Single nucleotide polymorphism
SPMMV	Sweet potato mild mottle virus
SqVYV	Squash vein yellowing virus
ss(+)RNA	single strand, positive sense RNA
TAE	Tris base, acetic acid, ethylenediaminetetraacetic acid
TB	Terrific broth
T-DNA	Transfer DNA
TMV	Tobacco mosaic virus
TomMMoV	Tomato mild mottle virus
tRNA	Transfer RNA
TuMV	Turnip mosaic virus
TTP	Thymine triphosphate
UV	Ultraviolet light
UTP	Uridine triphosphate
VPg	Viral protein genome-linked
VRC	Viral replication complex
w/v	Mass/volume percentage
YENB	Yeast extract nutrient broth
X-Gal	5-bromo-4-chloro-3-indoyl-b-D-Galactopyranoside
XTP	Xanthosine triphosphate

Chapter 1
Introduction

1.1 Cassava biology and cultivation

Cassava (*Manihot esculenta*, of the family *Euphorbiaceae*) is largely cultivated in tropical and subtropical regions for its tuberous storage root high in carbohydrate, which is depended on by 800 million people worldwide as a staple food (FAO, 2013). Cassava is a major staple crop in many developing countries, for example across the region of Sub-Saharan Africa where it is the second most important source of calories per capita (Nweke 2004). Typically growing to 7 m tall (*fig 1.1*) and reaching maturity in eight months under good conditions, it is typically grown between the latitudes of 30° north and south of the equator where it thrives under the 18-35°C temperature ranges and 500-5000mm annual precipitation (Moore and Lawrence, 2003). Cassava is particularly important for subsistence farmers in this region as it can tolerate periodic drought and marginal soils, making it an essential standby in times of drought and famine when other preferred crops fail. Furthermore, cassava is a perennial that can be grown and harvested throughout the year. Its hardiness and wide harvest window make it a valuable crop due to its productivity in terms of calories per unit land or per unit time for resource poor farmers (Hillocks and Thresh, 2002).

The starchy storage root can be processed by the grower into flour which can be consumed or sold at market, or at an industrial scale can be used to produce several food products from animal feed to ethanol biofuels. Through modelling, cassava has been suggested as a potential crop that may prove more resilient to climate change than other staple crops, as it's adaptions have great potential to shape future food security and production (Lobell et al, 2008).

Figure 1.1. A typical *Manihot esculenta* plant showing a cross section of the carbohydrate rich storage root. (from: http://stuartsbrasil.blogspot.com/2009/01/aipim-mandioca-manioc-pao-de-pobre.html)

Cassava was first domesticated from wild populations of *M. esculenta* in the Amazonian regions. Cassava was introduced to Africa in the 15th and 16th century by Portuguese traders, where it integrated well to agriculture in subtropical regions across Central and Eastern Africa (Olsen and Schaal, 1999). Nearly 270 million tonnes of cassava are produced annually worldwide, with 53% of this amount being produced in Africa (FAO, 2013). It is therefore one the most important non-grain crops, and a significant source of carbohydrate for people across the tropics of Africa, Asia, and Latin America in which it is grown, feeding approximately 800 million people in these regions (FAO, 2013).

Due to poor seed germination and potential outbreeding of desired traits, Cassava is commonly grown by vegetative propagation from cuttings (Ceballos et al, 2004). Cuttings are often referred to as 'stakes' which are typically 20cm long sections of the stem with multiple nodes, which are planted in loosened, moist soil after the danger of low temperatures has passed. Initial growth is relatively slow, making weed control important in early stages of development. Many farmers cover the soil surface with mulch during cassava establishment to prevent soil erosion and water runoff, as well as inhibit weed growth (FAO, 2013). Cassava can produce a crop with minimal inputs with roots ready to harvest from 6 months in some varieties. However, efficient cassava production in Africa is significantly constrained by abiotic, biotic, and crop management strategies (Fermont et al, 2009). Optimal yields at harvest are achieved from fields with good soil fertility levels and regular moisture availability. Poor soil fertility, water stress, ineffective weed management, and disease limit productivity even when optimal genotypes and farm management were used (Fermont et al, 2009). Under favourable conditions and commercial intensive farming settings, such as those found Nigeria, root harvests are typically over 20 t/ha. However, average yields under most subsistence farming systems across the subtropics are typically 9.8 t/ha. Surveys of average yield under farm management show that production could therefore be more than doubled through improved genotypes, crop nutrition, and water management. Cassava breeding should focus on developing varieties that are well-adapted to specific ecosystems and end-uses, as well as to produce good yields with minimal need for fertilizers and pesticides. Optimal farming strategies come from ensuring fertilizer application only during the first few months of growth to prevent excessive foliage production at the expense of storage root growth. Furthermore, by cutting the stems to be used as planting material only from plants with high root yields farmers can increase harvest size. The routine multiplication and distribution of disease-free planting material of improved varieties are essential for sustainable increase in cassava yields. Currently, limited infrastructure is in place, particularly for isolated small-hold and subsistence farmers.

Like all major crops cassava is susceptible to diseases can cause further reductions in crop yields. However, its cultivation by vegetative propagation makes the planting material vulnerable to pests and disease, such as bacterial blight, cassava mites, and viruses that cause CMD and CBSD. Cassava is especially susceptible to disease over successive generations of production, as it is often grown intensively as an annual crop without fallow years to break disease cycles. It is therefore important to ensure the initial planting material remains healthy and disease free. Aside from the use of pesticides, farmers can ensure planting material is a variety with tolerance or resistance to the most prevalent cassava diseases and pests (FAO, 2013). Furthermore, cuttings can be taken from mother plants that are free of disease symptoms and signs of pests and stakes soaked in hot water to kill pests or disease-causing organisms that might be present (FAO, 2013). However, the screening and conditions for the storage of viable stems for the next season is not always possible, leading to farmers acquiring stems of unknown quality from markets or neighbours, hampering disease prevention methods.

1.2 Viral diseases of cassava

Two viral diseases are a major factor in yield loss in large areas of Sub-Saharan Africa, cassava brown streak disease (CBSD) and cassava mosaic disease (CMD). CMD is caused by cassava mosaic geminiviruses (CMV), which are circular ssDNA viruses which are white fly transmitted. CMV causes systemic mosaic chlorosis and distortion of leaves and stunting of growth. CMD has been researched extensively due to their large impact on cassava crops in the 20th century and has been the subject of phytosanitary practises and resistance breeding programs by the International Institute of Tropical Agriculture in Nigeria (Patil et al, 2015). CBSD is caused by two closely related but distinct species of *Ipomoviruses*, Cassava brown streak virus (CBSV) and Ugandan cassava brown streak virus (UCBSV). Originally CBSD was thought to be caused by a single virus species, however analysis of the complete RNA genome sequences between viral samples taken from infected cassava across East Africa revealed that there are at least two distinct species which cause CBSD, CBSV and UCBSV, which share only 70% nucleotide identity (Winters et al, 2010). However, they both produce very similar symptoms in infected plants causing root rot and necrosis of the starchy tubers rendering the cassava inedible (Patil et al, 2015). CBSD is currently considered the biggest threat to food security in East Africa due to the estimated annual loss of US $1billion worth of crop in the region (IITA, 2014).

1.2.1 Cassava brown streak disease

CBSV and UCBSV belong to the family *Potyviridae* and are positive sense single stranded RNA viruses (Monger et al, 2001). CBSD was first reported in the 1930s in coastal lowlands (<1000m above sea level) of what is now Tanzania (Storey 1936). CBSD originally received limited scientific attention due

to its apparent geographical restriction and the greater threat of CMD to cassava productivity at that time (Hillocks and Jennings, 2003). However, since the early 2000's CBSD has started to spread at a rapid rate across parts of Eastern and Central Africa, posing a threat to agriculture across the region (Patil et al, 2015). Symptoms of CBSD in Cassava plants above ground include chlorosis along leaf veins, sometimes manifesting as patches of chlorosis in between primary veins, and occasionally brown streaks on the woody stems causing stem dieback in severe cases (*fig 1.2*) (Hillocks and Jennings, 2003). Root symptoms of the disease includes general radial restriction and a general reduction of the tubers' carbohydrate content, as well as dry-brown necrotic rot of the starchy tuberous root rendering the root inedible (*fig 1.2*) (Hillocks and Jennings, 2003; Nichols, 1950). The disease symptoms are dependent on the infecting viral isolate, cassava cultivar, stage of plant development, and environmental conditions. Thus, symptoms show large variability, often making the disease difficult to detect and therefore prevent (Patil, 2015). This variability, particularly in aerial symptoms, makes diagnosis hard, particularly due to their similarity to CMV infections and the symptomless nature of immature infected leaves (Hillocks et al, 1996). This often means infected stems get used for cuttings as farmers as unaware of the disease until roots are harvested, often later to reduce post-harvest spoilage (Patil, 2015).

CBSV can be mechanically transmitted into several herbaceous plant species to produce artificial infections, with the model host *Nicotiana benthamiana* being most commonly used to study the disease under controlled conditions (Mohammed et al, 2012). Comparison of symptoms between CBSV and UCBSV isolates on cassava and indicator hosts has shown that generally CBSV causes more

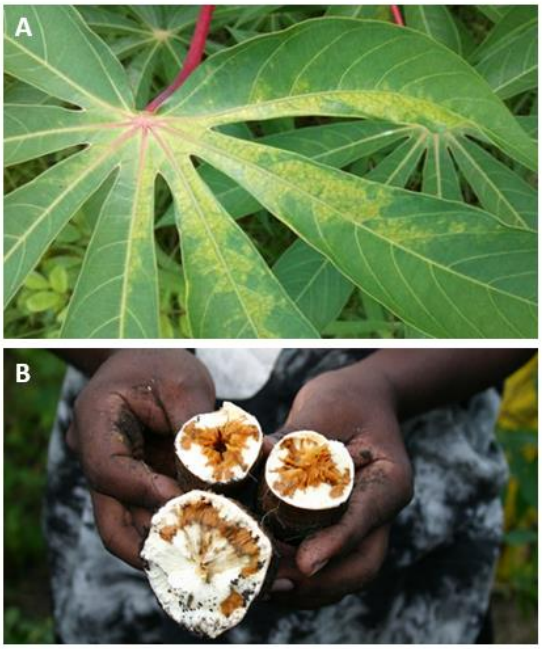

Figure 1.2. Characteristic symptoms of CBSD, (A) Feathery leaf chlorosis, (B) storage root necrosis (IITA, 2014)

severe root and foliar symptoms characterised by feathery chlorosis and root necrosis. Whereas UCBSV infections cause more mosaic and circular chlorotic blotches between lead veins in systemic leaves of infected plants (Mbanzibwa et al, 2011; Mohammed et al, 2012). Furthermore, CBSV was also shown to accumulate to higher viral titres than UCBSV in both Cassava and model plants (Ogwok et al, 2014; Kaweesi et al, 2014). High levels of sequence divergence between the genomes of the two viral isolates, such as in genes P1 (57% shared residues) and Ham1 (49% shared), may be the reason for differences in symptom severity and viral load (Winter et al, 2010).

1.3 Virus epidemiology

CBSV was first identified in the coastal areas of Tanzania in the late 1930's (Storey, 1936), in regions overlapping CMD affected areas (Patil, 2015). These reports described necrotic tuber rot and distinctive foliar chlorosis symptoms without the leaf distortion seen in CMD infected plants (Nichols, 1950). Original reports of CBSD observed that affected areas were localised to low-lying coastal and the Great Lakes region of Tanzania, with no reported cases above 1000m above sea level (Nichols, 1950). The disease was first reported in Uganda in 1945 due to possible introduction of infected cuttings from Tanzania (Nichols, 1950). Storey suggested a virus caused CBSD from evidence of disease transmission through grafted stems. This was further supported by the identification of 650nm viral particles by electron microscopy of CBSD infected *Nicotiana debneyi* (Bock, 1994). A partial sequence was obtained using RT-PCR from CBSD material in infected *N. benthamiana* and was shown to align most closely to coat proteins of *Sweet potato mild mottle virus* (SPMMV) an *Ipomovirus* virus of the *Potyviridae* family (Monger et al, 2001). Together this confirmed CBSV as the cause of CBSD.

In the early 1990s, there were increased reports of CBSD cases in countries across East Africa suggesting an outbreak of the virus from its original geographic localisation (Hillocks and Jennings, 2003). The rapid spread across eastern Africa also coincided with reports of CBSD infections at midaltitude levels (1000-1500m above sea level).

Although a cause for the increase in incidences and geographical range of CBSD have not been identified, it is likely that it is the combined result of an increase in populations of the whitefly vector, *B. tabaci*, alongside the distribution of infected cuttings for planting material, the use of susceptible genotypes bred for CMD resistance, and changing environmental conditions within the region (Legg et al, 1999; Jeremiah et al, 2015). *B. tabaci* was confirmed as a vector for CBSV, transmitting the virus from infected to healthy cassava plants under quarantined insectary and controlled greenhouse conditions by Maruthi et al (2005). It was further shown that *B. tabaci* transmit CBSV in a semi-persistent manner, like other *Ipomoviruses*, whereby the virus is held in the foregut of the insect for

no longer than 48 hours and is only transmitted over relatively short distances (<20m) between crops (Maruthi et al, 2016). Outbreaks of CBSD across East and Central Africa have coincided in the years following rises in *B. tabaci* numbers, giving further credit to whitefly vectors causing the increase mid-altitude CBSD outbreaks (Legg et al, 2011). However, whiteflies are only capable of transporting CBSD over a small range due to their short flight distances. Reports of CBSD from six countries around the great lakes are from independent hot-spots where it was previously absent, rather than spreading out in a front from an initial infection zone as is seen with CMD (Legg et al, 2011). The reason for these independent geographically separated infections is uncertain, however it is likely due to independent 'founder' infections from infected material transported from endemic regions, or from whiteflies introducing infections from local wild-relative hosts. The trade and movement of infected cassava cuttings and plant material between farmers across different localities causes the disease to be transported over greater distances than is likely through whitefly vectors, thereby leading to the emergence of isolated CBSD affected areas (Legg et al, 2011; Tomlinson et al, 2017).

CBSVs are only found within the regions of Africa discussed, and due to the relatively recent introduction of cassava to the continent, have therefore had limited time to evolve within cassava itself. Cassava originates from the Americas where there are no reports of CBSD infections. This suggests that CBSVs are opportunistic viruses that have evolved within an unknown species in a similar region of East Africa before transferring host and being capable of infecting and propagating with cassava crops (Monger et al, 2010). Therefore, there could be other plant hosts, such as wild *Euphorbia* relatives, present in hotspot areas which provide a source of viral inoculum contaminating cassava crops (Monger et al, 2010). No alternative host plant species have been detected from screening efforts (Legg et al, 2011), however more comprehensive studies with greater coverage are required to rule out other plant species as sources of CBSD infections.

1.4 CBSD resistance and management methods

The rapid spread of CBSD into East African highlands, and the involvement of two novel viruses, CBSV and UCBSV, has raised many concerns about combating the disease epidemic. Despite the growing research and understanding of the viral causal agents of CBSD, relatively little is understood about host-pathogen interactions. As such, studying the responses of different cassava varieties to the different strains of CBSVs has been a challenging task (Alicai et al, 2016). Currently no cassava cultivar is available to famers which exhibits high levels of CBSD (Tomlinson et al, 2017). The main approach has been to breed and disseminate CBSD resistant cassava varieties against one or both viral species, alongside studying and characterising the genomes of CBSV and UCBSV viral isolates. However, this has proved difficult due to the high genetic variability between cassava cultivars which possess

different characteristics and the challenging cross-pollination process involved in breeding (Tomlinson et al, 2017).

Breeding CBSD resistant cassava cultivars started in the 1930's at the Amani research station in Tanzania, though mainly focused on CMD resistance due to its prevalence and devastating affect across Africa, resistance to CBSD was often secondary or completely ignored (Jennings, 1960). Therefore, CMD resistant crops were widely distributed to affected areas, though these cultivars show varying susceptibility to CBSD. Breeding for CBSD resistant cultivars led to the challenge of selecting varieties that showed stable CBSD resistance across the spread effected ecological environments, as many cultivars varied depending on environmental factors (Hillocks and Jennings, 2003, Ceballos et al, 2004). One of the most successful cultivars developed in breeding programs has been the Kaleso variety in Kenya (Genetically identical to Namikonga variety in Tanzania) which is the product of backcrossing from *M. esulenta x M. glaziovii* (Patil, 2015). This variety, along with the NASE cultivars developed in Uganda, have been considered the best options available to farmers for CBSD resistance and have since been crossed into local varieties as well as being used for quantitative trait locus (QLT) detection using RNA-sequencing technologies to help identify CBSD resistance alleles for marker-assisted breeding approaches (Kaweesi et al, 2014; Maruthi et al, 2014). However, in field trials of eleven resistant cassava varieties revealed that none of the varieties are immune to either CBSV or UCBSV (Kaweesi et al, 2014). The different cassava genotypes did vary in symptom expression, in terms of the reduced severity of either shoot or root symptoms with the latter being more highly valued by breeders, as well as variation in viral titre, indicating a genotypic response to infection. Nonetheless mixed infections of both UCBSV and CBSV were found in all the varieties tested, with many showing a disparity between viral load and symptom severity making the quantification and selection of resistance difficult (Kaweesi et al, 2014). Although the cultivars remain susceptible to CBSD viruses, tolerant varieties which maintained a lower viral load and showed reduced root symptom development were evident. For example, NASE1 and Kiroba cultivars showed root necrosis in less than 20% of cassava tested and with greatly reduced severity (Kaweesi et al, 2014). Although these cultivars show the greatest resistance to CBSD and have therefore been adopted widely across East Africa, they do little to remove the source of viral inoculum to prevent the spread of disease and may allow the virus to overcome resistance.

Many initiatives such as the 'Great Lakes Cassava Initiative' funded by the Gates Foundation, have attempted to combat this by aiming to provide farmers with access to disease-free improved cassava varieties with the best combined resistance (Patil et al, 2015). This will provide widespread distribution of tolerant germplasm which will ensure initially disease-free crops, as well as helping prevent the spread of infected cutting material which has contributed to the spread of CBSD. As tolerant cultivars

retain viral inoculum in their stems, they must be subjected to extensive inspection during growth and at harvest before being selected for cuttings. Foundation stocks should be found to be free of leaf, stem and root symptoms before being collected for cuttings and distribution to farmers.

However, because of the variability of CBSD symptoms, early and accurate diagnosis of diseased crops remains a great challenge for farmers. Often the disease is only noticed once the necrotic tubers are harvested and cut, at which point most of the crop yield has often been lost and material for propagation collected, meaning management strategies to contain the disease have likely failed (Legg et al, 2011, Thresh et al, 1994). The difficulty of diagnosis based on leaf symptoms has meant that cuttings of infected stems are likely to have been distributed for use as planting material, leading to disease outbreaks which appear as hotspots in un-affected areas, which has contributed to the emergences of epidemic across regions of Africa (Martin et al, 2000; Miller et al, 2009). Therefore, due to the comparatively symptomless nature of CBSD and confusion with other diseases, diagnostic techniques such as reverse transcription polymerase chain reaction (RT-PCR) and real time RT-PCR have been developed to provide reliable detection of CBSV and UCBSV in screens of plant material before transportation for trade (Patil et al, 2015). These PCR based approaches use multiplexing primers to amplify target regions of the CBSVs from RNA isolated from leaf samples, allowing specific detection of CBSV and UCBSV viral strains (Abarshi et al, 2010). Furthermore, advances in sequencing technologies has led to viral identification by deep-sequencing of RNA extracts from infected material (Monger et al, 2010). However, their application in the field is rarely possible due to the associated costs and the requirement of centralized facilities, meaning screening for the disease in the field is not always possible for farmers. Therefore, development of effective diagnostic techniques and infrastructure which can be widely available used in Africa is vitally important for disease management.

Alongside breeding efforts, several strategies have been utilised to understand the plant transcriptional response to infection as well as genetic engineering of resistance (Reddy et al, 2009). Breeding processes in cassava have been restricted due to the crops highly heterozygous nature making transfer of desirable traits by backcrossing unfeasible, the difficult and cumbersome process of producing adequate quantities of recombinant seed, and the limited number of studies into the inheritance and genetic variability of agronomically relevant traits (Ceballos et al, 2004). Therefore, due to the difficult nature of breeding for resistance and crop quality, genetic engineering offers a more rapid and direct approach in understanding and thereby disrupting CBSD infections. For example, the use of CRISPR/Cas9 has already been used to give *Arabidopsis* resistance to the *Potyvirus* Turnip mosaic virus by introducing a deleterious point mutation into the gene for the eukaryotic translation initiation factor paralogue required for virus translation (Pyott et al, 2016). Cassava is an

ideal target for transgenic methods due to being grown by vegetative propagation. Therefore, transgenic cassava would lack traditional problems associated with GM technology such as pollen-based gene flow and outcrossing of characteristics.

One approach in cassava is the use of RNA interference (RNAi) to target UCBSV and CBSV infections (Patil et al 2015; Ogwok et al, 2012). This method involves transgenic expression of inverted repeats containing CBSD viral sequence, or artificial complimentary microRNA sequences, both of which activate the plant host's post-transcriptional gene silencing pathway producing short interfering RNA (siRNA) which targets complementary viral sequences for translational repression and degradation (Reddy et al, 2009; Ding and Voinnet, 2007), thereby providing resistance against viral RNA sequences. RNAi approaches against CBSD have been demonstrated in *N. benthamiana* where transgenic expression of the gene for UCBSV coat protein as an inverted repeat produced high levels of resistance in 85% of transgenic expressers to UCBSV isolates, as well as protection to some non-homologous CBSV isolates demonstrating RNAi cross-protection (Patil et al, 2011). Transgene expression and subsequent levels of siRNA expression had a positive correlation with host viral resistance, demonstrating the effectiveness of RNAi resistance by a dose-response relationship in *N. benthamiana* (Patil et al, 2011; Ogwok et al, 2012). A further benefit of transgenic approaches is that resistance has been shown to be stable in subsequent generations and propagation of stem cuttings through stable heritable expression of siRNAs (Patil et al, 2015).

Further regions of the CBSV/UCBSV genome are being researched as possible targets for RNAi, with the possibility of combined transgene expression or the targeting of common viral genomic elements leading to multi-viral resistance (Patil et al, 2015; Reddy et al, 2009). The main obstacle to genetic approaches have been the lack of available infectious clones of U/CBSV, the low efficiency of mechanical transmission to cassava for the study of virus/host interactions, poor whitefly transmission under controlled environmental conditions, as well as labour-intensive resistance screening methods (Maruthi et al, 2005; Wagaba et al, 2013). The development of improved cultivars has been hampered by lack of funding from the private sector due to the vegetative cloning of crops by farmers. With local governments and charities often lacking the technologies and knowledge required for large scale screening and genetic improvement of cassava. However, with increased research leading to more CBSD viral isolates being sequenced and the greater understanding of CBSD infections and host plant responses, molecular engineering of cassava varieties alongside existing strategies has considerable potential to benefit farmers across a range of agricultural scales in East Africa.

1.5 Molecular characterisation of CBSD

The viruses that cause CBSD are part of one of the largest and most abundant plant virus families - *Potyviridae*. Named after *Potato virus Y* (PVY), this viral family causes diseases in a vast range of crop and susceptible plant species making them globally significant pathogens. Due to the crop losses caused by these viruses they have drawn significant attention from the research community in-order to study their taxonomy, evolution, genome structure, and role of the viral proteins. The viral family consists of 8 genera, with most members of the *Potyviridae* having a monopartite positive-sense single stranded RNA genome, the exception being the genus *Bymovirus*. The viruses can be characterised as forming diagnostic cytoplasmic 'pinwheel' inclusion bodies composed of viral proteins. The viral RNAs are encapsulated by repeating coat protein (CP) units, with a typical potyvirus having about 2000CP subunits arranged helically around a single viral RNA genome.

Within the family, genera can be differentiated mainly based on the method of vector transmission, as well as their genome structure and the conserved amino acid sequence of several viral proteins. The positive-ssRNA genomes of *Potyviridae* range from 8.2 to 11kb in size, which are translated into a single polyprotein which is proteolytically auto-cleaved by viral-encoded proteases into mature proteins in equimolar ratios (Patel, 2015), additionally a +2-ribosomal frame-shift caused by an overlapping ORF leads to a further P3N-PIPO protein being produced. The large ORF of *Potyviridae* typically encodes 10 gene products, which in PVY from the *N*- to *C*- terminus are: P1, HCPro, P3, 6K1, CI, 6K2, VPg, Pro, NIb, and CP (fig 1.3). Most of the viral proteins have multiple functions in the replication cycle, processing, transcription, and infectivity of the virus, with some virus species having lost proteins and replaced their activity with others.

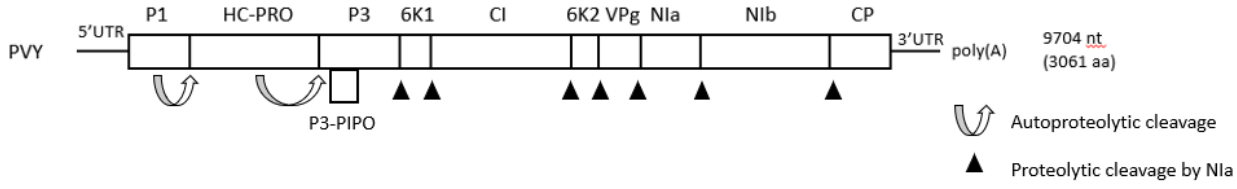

Figure 1.3 *Potyvirus* genus type species Potato virus Y (PVY) genome structure. PVY encodes 10 ORFs, plus the frame-shift protein P3-PIPO, which are translated into a single polypeptide. P1 and HC-Pro catalyse autoproteolytic cleavage of their respective C-terminal junctions, the remaining protein junctions are cleaved by trans-proteolytic cleavage by NIa-pro.

1.5.1 Potyvirus life cycle

In brief, the replication cycle of Potyviruses involves vector delivery and host entry, translation, replication, cell to cell movement, anti-viral defence, and finally encapsulation. As *Potyviruses* consist of positive sense single stranded RNA genomes, they can directly act as a template for replication or translation in the host cytoplasm after penetration and uncoating of the viral genomic RNA. Delivery of filamentous virions, containing viral RNA encapsulated in coat protein, into the host cell can be from an outside environment e.g. delivery by Aphid stylet, or by viral movement from a neighbouring cell. The starting point of Potyvirus infection cycle within a new host cell is the initial translation of viral proteins. Viral translation occurs in the endoplasmic reticulum, where the viral protein genome-linked (VPg) of potyviruses is covalently linked to the 5' terminus of the viral RNA interacts with and recruits the eukaryotic translation initiation factor 4E (eIF4E). The host factor eIF4E recognises the 7-methlyated guanine cap of host mRNA which leads to recruitment of the 40S ribosomal subunit onto the mRNA for translation. Therefore, recruitment of eIF4E by viral VPg primes translation of the viral genome using host ribosomes (Travert-Roudet et al, 2017). Once translated, the single polypeptide is proteolytically processed by three encoded viral proteases. The first encoded protein, P1, is a cis-acting serine protease which autoproteolytically cleaves the polypeptide at its respective C terminus. In *Potyviridae* members the P1 protein is extremely variable in size as well as some *Ipomoviruses* encoding two copies and like many viral proteins is multi-functional, having been shown to suppress host post-transcriptional gene silencing pathways (Valli et L, 2006; Kasschau and Carrington, 1998). The helper-component protein (HCPro), although multifunctional, also acts as a cysteine proteinase which cleaves the protein between its own C-terminus and P3 (fig 1.3). HCPro, alongside P1, is also involved in viral suppression of gene silencing by binding siRNAs, It's N-terminus has also been implicated in vector transmission in aphids and mites as well as cell to cell movement within the plant (Kasschau and Carrington, 1998). The third proteinase derives from the activity of the C-terminal part of nuclear inclusion protein (Nia-Pro), which has high specific activity in recognising the seven remaining polyprotein cleavage sites (fig 1.3).

Replication of the virus is carried out by NIb, a protein with homology to known RNA-dependent RNA polymerases. It is proposed to interact with the 3'UTR of the viral genome as well as a 6K2-VPg-Pro viral replication complex (VRC) alongside CI and HCPro viral (Li et al, 1997). The endoplasmic reticulum (ER) membrane of the host cell acts as the site of viral translation, from there viral RNA and proteins involved in replication are captured in VRCs which initiates replication. The formation of these replication vesicles is initiated by the viral protein 6K2 (Makinen and Hafren, 2014). From the ER, VRCs are trafficked to the chloroplast along the host cell actin cytoskeleton. At the chloroplast, the viral

vesicle fuses with the chloroplast membrane with the help of host vesicle trafficking machinery and the viral protein 6K2. VRCs anchored at the chloroplast membrane are proposed to act as replication factories, supported by the presence of dsRNA in the vesicles (Grangeon et al, 2012).

During the progression of infection, viral RNA is transferred from infected cells into neighbouring healthy cells. Along with HCPro, *Potyviridae* encode CI, CP, and the frameshift protein P3-PIPO, all of which are linked with viral movement activity. Plant viruses utilize plasmodesmata channels, used for plant cell communication, for their cell to cell movement by targeting and modifying the channels to move viral material through. Viral vesicles localise to the plasmodesmata where the CI protein forms transient conical structures which are anchored to plasmodesmata specific proteins by P3-PIPO (Grangeon et al, 2010; Wei et al, 2010) The viral RNA is partially coated by CP and its interaction with CI and the plasmodesmata facilitates the movement of viral RNA into neighbouring cells, where the cycle of replication and translation is repeated before further cellular movement and the spread of infection.

During viral infection the virus is subjected to plant defence responses which aim to stop their replication and movement in the plant. Viruses are targeted by host RNA-silencing mechanisms which target sequence specific viral RNA for degradation. The host RNA-induced silencing complex (RISC) containing the Argonaut protein interacts with viral siRNAs processed from Dicer protein cleavage of dsRNA replication intermediates to direct cleavage of homologous viral RNA. The viral siRNAs are able to spread systemically through plasmodesmata to direct RISC induced viral resistance in adjacent cells (Ding and Voinnet, 2007). Both HCPro and P1b are both potyviral suppressors of host RNA-silencing mechanisms through their proposed binding activity of ds-siRNAs (Valli et al, 2006; Makinen and Hafren, 2014), thereby sequestering viral RNA away from host RNA interference mechanisms (Kasschau & Carrington, 1998).

1.5.2 Ipomoviruses

Ipomoviruses are a genus of whitefly-transmitted (rather than aphid or mite transmitted) viruses of the family *Potyviridae*. The genus consists of six virus species; Sweet potato mild mottle virus (SPMMV), Cassava brown streak virus (CBSV), Ugandan cassava brown streak virus (UCBSV), Cucumber vein yellowing virus (CVYV), Squash vein yellowing virus (SqVYV) and Tomato mild mottle virus (TomMMoV). They are all characterised as ss(+)RNA viruses which form flexible filamentous rods of viral particles *in vivo*. Like other *Potyviridae,* Ipomoviruses encode a single polyprotein which is autoproteolytically cleaved by encoded proteases, though they possess different genome structures to other family members. For example, other *Potyviridae* often possess a single serine protease (P1) as well as HC-Pro, where the P1 protein acts as a serine-protease whilst HC-Pro acts as a

multifunctional protein contributing to silencing suppression, vector transmission, and viral accumulation (Patil et al, 2015). Whereas Ipomoviruses *Cucumber vein yellowing virus* (CVYV) and *Squash vein yellowing virus* (SqVYV), which lack HC-pro, have replaced its suppressor activity with two P1 serine proteinases (P1a & P1b), with P1b functioning as a suppressor of RNA silencing. Other genus members encode novels proteins, such as the presence of Ham1 in CBSV and UCBSV.

CBSV encodes a ~9003nt genome whereas the genome for UCBSV is 9069nt long, both of which are smaller than the genomes found in other *Ipomoviruses* such as SPMMV (10,818 nt), SqVYV (9,836 nt), and CVYV (9734 nt) (Mbanzibwa et al, 2009; Winter et al, 2010). Both U/CBSV encode polyproteins which are cleaved into the 11 mature proteins: P1, P3, 6K1, CI, 6K2, VPg, Nia, NIb, Ham1, and CP, plus the frameshift protein P3N-PIPO.

Although the genome structures of CBSV and UCBSV are similar, sharing 70% nucleotide identities (Winters et al, 2010), they differ from other *Ipomoviruses* in several key ways. Both CBSV and UCBSV genomes lack a HC-Pro component, but only contain a single P1 protein most closely related to the P1b proteins found in CVYV and SqVYV, though they only share ~30% protein sequence identity (*fig 1.3*). This high divergence in P1 protein is characteristic of *Potyviridae,* with the most closely related proteins containing a zinc finger and LXKA motifs which have been linked to RNA silencing suppression (Mbanzibwa et al, 2009, Valli et al 2006).

As well as a P1 protein, both CBSV and UCBSV genomes encode a further eight ORFs which are all characteristic of *Potyviridae* family genomes, which contain conserved sequences for the third protein (P3), two 6 kDa proteins (6K1 & 6K2), Cylindrical inclusion protein (CI), viral protein genome-linked (VPg), a viral proteinase (NIa-Pro), a replicase (NIb), and coat protein (CP) (fig 1.4). Which all shared high sequence identity with the proteins of other *Ipomovirus* genomes such as CVYV, and SqVYV (Chung et al, 2008). These proteins therefore are functionally similar to the homologous proteins found in other *Potyviridae* members already described.

1.5.3 Ham1 like protein

Both the CBSV and UCBSV genomes were found contain a 678 nt insertion found in the 3' proximal region between NIb and CP encoding the putative HAM1 sequence, which is a novel genome feature in the family Potyviridae and is particularly unusual due to the predominantly highly conserved 3' proximal region found in the other viral family members (fig 1.4) (Patil et al, 2015; Fauquet et al, 2005). The ORF is flanked by consensus proteolytic cleavage sites for the *Ipomovirus* viral proteinase NIaPro. Sequence and residue analysis of this sequence, although novel in *Potyviridae*, showed that it is homologous to the Maf/HAM1 family of proteins found in many prokaryotes and eukaryotes. The

Maf/HAM1 proteins found in other organisms are nucleoside triphosphate (NTP) pyrophosphatases which prevent the incorporation of non-canonical NTPS into DNA or RNA thereby reducing mutagenesis, suggesting it may perform a similar role for the viral genome of CBSV. A similar HAM1-like ORF in a corresponding genome position has also been reported in *Euphorbia* ringspot virus, which also belongs to *Potyviridae* (Knierim et al, 2017). The HAM1 encoding region is likely of cellular origin and suggests that members of the *Potyviridae* family can undergo genomic recombination with cellular RNAs, a process which has been characterised in the evolution of *Closteroviridae* family of ss(+)RNA viruses (Mbanzibwa et al, 2011).

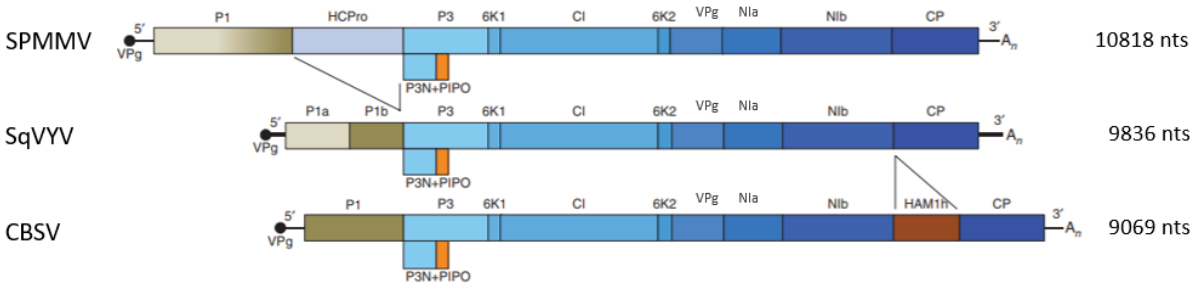

Figure 1.4. Example *Ipomovirus* genomes with the type member sweet potato mild mottle virus (SPMMV) (top), squash vein yellowing virus (SqVYV) (centre) and cassava brown streak virus (CBSV) (bottom). The ssRNA genome is shown as a line with the ORF translated into a polyprotein represented as a box, with vertical lines dividing individual gene products cleaved by proteinases. Additionally, the covalently linked VPg protein is shown as a solid circle at the 5′ -end, and the poly A tail at the 3′ -end indicated by A_n , as well as the frameshift protein P3N+PIPO. (Adapted from Valli et al, 2015).

1.6 CBSV and UCBSV Infections

Despite their similarities, CBSV and UCBSV have been shown to cause slightly different symptom severities in infected cassava plants. Plants with CBSV infections develop more severe root necrosis as well as more feathery chlorosis along leaf margins, whereas UCBSV infections are linked to more chlorosis between leaf vein margins which develops into blotches and a greatly reduced necrotic phenotype relative to CBSV in both *N. benthamiana* (Mohammed et al, 2012; Winter et al, 2010; Patil et al, 2015). Furthermore, CBSV shows higher viral titres in both infected cassava and indicator plants (Kaweesi et al, 2014; Mohammed et al, 2012). Pairwise comparison between UCBSV and CBSV isolates may explain differences in symptom severities, as the second half of the 5'-untranslated regions, P1, HAM1, and CP regions were shown to be the regions with the highest diversity between the viral species (Mbanzibwa et al, 2011; Ogwok et al, 2014). With P1 and HAM1 ORFs sharing only 59% and 49% amino acid residues respectively, whereas the diversity in CP sequences between CBSV and UCBSV isolates was lower with 80% similar amino acid identities (Winter et al, 2010; Mbanzibwa et al, 2011). The relatively low similarity of the HAM1 sequence between CBSV and UCBSV may suggest that both viruses acquired the gene independently from separate cellular origins, or alternatively, that the

HAM1 gene was acquired once from the same event but experienced different degrees of adaptive selection pressure between the two viral species leading to divergences in certain genome sequences between isolates (Mbanzibwa et al, 2009; Ogwok et al, 2014, Monger et al, 2010). In general, viruses of the family *Potyviridae,* and to some extent *Ipomoviruses* such as CVYV, show rather low genomic diversity (Mbanzibwa et al, 2011). Analysis of CBSV and UCBSV isolate HAM1 sequences by Mbanzibwa et al (2011) suggests that there were 33 aa residues under negative selection, corresponding to conserved residues in homologous eukaryotic and prokaryotic proteins. This suggests these residues may be key for the proteins' activity or structure in both eukaryotes and viruses. Furthermore, Mbanzibwa et al (2011), suggest that a few residues at either terminus of the protein were under positive selection in viral isolates of both CBSV and EuRSV. This may be due to the viral protein being expressed as part of one big polyprotein rather than as a single protein in cells, hence residues at the termini form part of proteolytic cleavage sites for excision of the protein and are therefore under greater selection pressure for optimal processing by viral proteases. Adaption of the HAM1 protein and the virus itself is important due to the cellular environment influencing protein stability and activity as well as the infectivity of the virus itself. Differences between UCBSV and CBSV sequences may be adaptions to infections in the different cassava varieties grown in different regions, and key to the spread of viral isolates and the degree of infectivity and symptom severity (Mbanzibwa et al, 2011; Patil and Fauquet, 2009).

Furthermore, the rate of genome evolution of CBSV and UCBSV populations may accelerate in the future due to the increasing geographical distribution and host-pathogen interactions in different host varieties applying greater selection pressure on genes, such as HAM1, leading to adaptive evolution via positive selection (Mbanzibwa et al, 2011). The genomic variation has also been studied within the viral species, with nucleotide identities ranging from 87-99% identical among UCBSV isolates and 79-95% among CBSV isolates (Alicai et al, 2016). This further source of diversity may be allowing CBSV to overcome host resistance responses which have been selected for in cassava breeding programs. RT-PCR has shown that mixed infections of both viral species frequently occur in geographical areas where the species overlap, for example in Kenya (Mbanzibwa et al, 2011). Although there is currently no evidence of CBSV and UCBSV isolates interacting, recombination and interaction effects between the species is entirely possible and have been demonstrated in other viral species such as CMD, where different viral strains act synergistically leading to increased viral titre and severity of symptoms (Tomlinson et al, 2017).

1.7 Investigating the function of Ham1 in CBSD

Analysis of the HAM1 amino acid sequence of CBSV and UCBSV shows that it is homologous to and shares conserved motifs with the HAM1 found in eukaryotes and prokaryotes, which belongs to the superfamily of ITPase pyrophosphatases (Mbanzibwa et al, 2009). The Maf/HAM1 family of proteins found in other organisms' function as nucleoside triphosphate (NTP) pyrophosphatases which cleave non-canonical nucleoside triphosphates, preventing their incorporation into DNA and RNA (Galperin et al, 2006; Mbanzibwa et al, 2009). Non-canonical nucleotides arise from oxidation or deamination of 'normal' canonical nucleotides thus becoming toxic by-products of normal cellular metabolism. If these abnormal nucleotides are incorporated into nascent DNA, they have the potential to cause mispairing of bases leading to an increased mutation rate due to transitional changes. Some non-canonical NTPS are produced during the *de novo* synthesis pathway of the canonical nucleotides (dATP, dTTP, dCTP, dGTP). The synthesis of ATP and GTP requires the precursor inosine monophosphate (IMP), which contains a hypoxanthine purine base which can be modified on either the 6-keto group to produce the ATP precursor AMP, or on the 2-keto group to produce the GTP precursor GMP. IMP can be directly phosphorylated to produce ITP, or ATP and GTP deaminated to produce ITP and XTP respectively (fig 1.5). As well as by-products of canonical nucleotide synthesis, non-canonical NTPs are also produced from the damage of nucleotides by reactive oxygen/nitrogen species under oxidative stress (Davies et al, 2012; Kamiya, 2003; Kamiya , 2004, Pang et al, 2012).

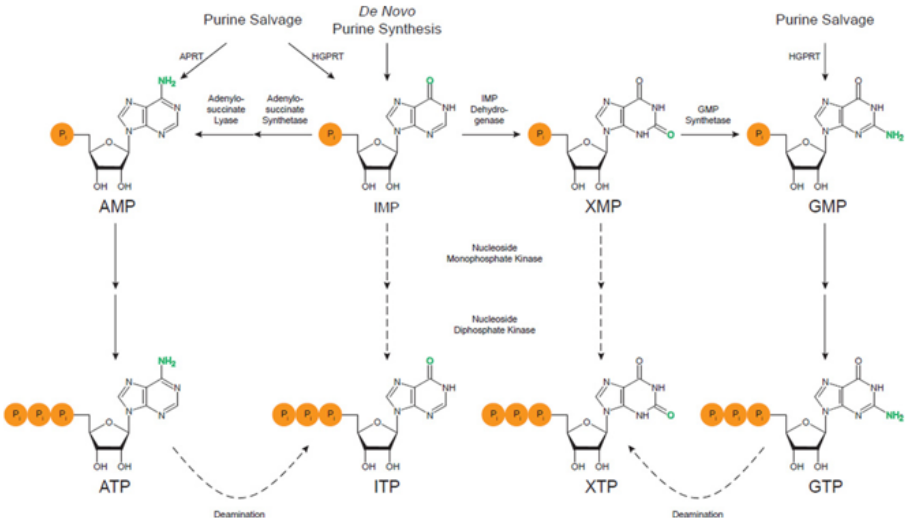

Figure 1.5 Purine synthesis and metabolism leading to the formation of ITP and XTP. The purine IMP is formed from *de novo* synthesis and can be converted to AMP or GMP. IMP and GMP can further be formed from free hypoxanthine and guanine bases, whilst AMP can be synthesised from adenine, in purine salvage pathways by phosphoribosyl transferases. AMP and GMP are sequentially phosphorylated to ATP and GTP by phosphate kinases, which may act on IMP and XMP directly to form ITP and XTP. Deamination of ATP and GTP amino groups to keto groups also forms the non-canonical nucleotides ITP and XTP. (adapted from Simone et al, 2013)

NTPs induce mutation due to their promiscuous base-pairing during replication. For example, dITP is incorporated opposite dCTP with approximately 1/3 of the efficiency of dGTP and can also base pair with TTP, though a hundred times less efficiently than dATP (Simone et al, 2013). Therefore, the Maf/HAM1 family of proteins which have activity on these non-canonical nucleotides act as 'housecleaning enzymes' to intercept and remove the abnormal non-canonical nucleotides from the available cellular pool (Galperin, 2006).

For example, the human enzyme ITP pyrophosphatase was characterised in 2001 and shown to have nucleotide binding properties that specifically hydrolysed Inosine triphosphate (ITP) and Xanthosine triphosphate (XTP) into their monophosphate forms (Lin et al, 2001). Patients with a decreased activity of the gene product of ITPase due to polymorphisms show enhanced sensitivity to several base analogues, such as azathioprine (AZA) the pro-drug of 6-mercaptopurine (6-MP), used as drug to prevent organ rejection after transplantation. ITPase deficiency results in an accumulation of inosine nucleotide (ITP), the drug 6-MP is activated through a 6-thio-IMP intermediate. Therefore, a toxic build-up of 6-thio-ITP is predicted in ITPase deficient patients treated with AZA (Marinaki et al, 2004). The homologous HAM1 gene was found in *E. coli* and yeast, Noskov et al (1996) showed that yeast (*Saccharomyces cerevisiae*) HAM1 gene reduces yeast cell susceptibility to the mutagenic purine base analogue 6-*N*-hydroxylaminopurine (HAP) by catalysing conversion of dHAPTP to dHAPMP. as a disruption of the HAM1 ORF leads to strong sensitivity to HAP (Noskov et al, 1996; Mbanzibwa et al, 2009). Takayama et al, (2007) further showed that purified recombinant yeast Ham1 detoxified abnormal purine and pyrimidine bases to their monophosphate derivatives and a pyrophosphatase group due to the lack of proof reading function in the viral replicase NIb.

Although it has an uncharacterized function in CBSV, suppression of NTP mutagens may provide an advantage to the virus under plant oxidative stress conditions observed in CBSD infected cassava leaves under high viral load (Hillocks and Jennings 2003; Mbanzibwa et al, 2009). For example, the human ITPase shows no difference in specificity for deoxy or ribose forms of NTPs and would therefore exclude them from both RNA and DNA precursor pools. Mutations caused by the incorporation of non-canonical NTPs into the viral genome could lead to changes in the codon sequence and mis-translation of viral proteins. Mutagenic NTPs could be introduced during viral replication leading to an accumulation of mutations in the population over several generations, due to a lack of proofreading ability in viral replicases to prevent their incorporation.

Mbanzibwa et al, (2009) demonstrated that that the HAM1-like protein in CBSV lacked suppression of host silencing activity in *N. benthamiana* under co-agroinfiltration with GFP gene and GFP specific hairpin RNA to induce host RNAi. However, HAM1 may perform a similar function to homologous

cellular ITPase proteins in preventing mutation during viral replication under oxidative conditions. The HAM1 gene is also found in Euphorbia Ringspot Virus (EuRSV) which also infects plants of the *Euphorbiaceae* family, of which cassava is a member. This suggests the presence of HAM1 in plant viruses might be an adaption to infections in woody perennial Euphorbia plants. It is likely present for a similar anti-mutagenesis function due to oxidative stress conditions in *Euphorbiaceae* under infection, which show an early leaf senescence phenotype under viral accumulation (Patil et al, 2015; Fauquet et al, 2005).

Although only the 3' end of the genome of EuRSV has been sequenced, it's physical characteristics of such as genome length, particle size, and vector species, appear to make it an atypical member of the genus *Potyvirus*. However, The HAM1 of EuRSV contains only 39% and 38% aa sequence similarity with the CBSV and UCBSV proteins respectively, lower than the similarity of U/CBSV Ham1 to eukaryotic Ham1 homologues, such as *Zea mays* (~50%) (Monger et al, 2010). Suggesting an ancient, or separate, acquisition event of Ham1 by these viruses. As little information exists about the pathology of this large plant family, the retained function of HAM1 is only speculative (Knierim et al, 2017). Although there are few reports of heterologous insertion sequences in plant viral genomes, such as the AlkB protein of *Flexiviridae*, stable gene expression has been shown for engineered potyvirus vectors for sequences inserted in the NIb/CP junction, which is where HAM1 is located in the U/CBSV genome (Arazi et al, 2001). The HAM1 gene of U/CBSV exists as a possible natural example of this recombination, making it an intriguing gene to study.

1.7.1 Viral mutation rates

RNA virus replication is associated with a high mutation rate leading to a random distribution of mutants, a process that has been linked to the structure and evolution of viral populations in a fitness environment (Drake and Holland, 1999). Increasing viral diversity within an infected host is generated by the progressive accumulation of mutations resulting in a group of related sequences that evolve together over time. This process produces a viral quasispecies of genetically distinct yet related genotypes (Anderson et al, 2004). It has been suggested that different RNA viruses achieve an optimal mutation rate during infection, allowing them to evolve and explore sequence space (Anderson et al, 2004; Kamp et al, 2003). Riboviruses have been measured as having a characteristic mutation rate per genome per replication of 0.76, compared to 0.2 for retroelements and 0.003 for DNA based viruses and microbes (Drake & Holland, 1999). This is in part due to the low fidelity of RNA virus replicases, a lack of error-correcting proof-reading activity compared to DNA polymerases, and lack of known replication repair mechanisms. The mutation frequency is so high that riboviruses can only accommodate a small increase in mutation rate before the viral population becomes susceptible to

mutagenesis. The genetic information encoded by the viral quasispecies can only be faithfully transmitted below a certain threshold. If the mutation rate exceeds this threshold, the viral population would not be able to maintain the fittest genotype due to persisting mutations reducing the number of viable progeny (Anderson et al, 2004). Small virus populations that have undergone bottleneck events are particularly vulnerable to these events which can lead to population collapse. Under this scenario *in planta*, the viral population would be compromised due to excessive mutations, leading to a decrease in infectivity and population extinction. It is therefore suggested that viruses maintain their genomes at a fidelity close to the error threshold, but an increase in the mutation rate due to the environment can push the population over this threshold reducing the populations' viability Anderson et al, 2004; Crotty et al, 2001).

This process has been demonstrated in poliovirus, which is related to *Potyviridae*, where the presence of ribavirin depletes the canonical GTP pool leading to a 10-fold increase in viral mutation rate leading to lethal mutagenesis (Crotty et al, 2001). Furthermore, the mutagenic nucleoside analogue 5-azacytidine increases the replication error rate of bacteriophage Qβ in aqueous solution leading to population extinction (Lazaro et al, 2008). Suppression of the mutation rate may be beneficial for stable long-term infections of slowly evolving viruses under mutagenic conditions, such as the high oxidative stress in infected plants (Patil, 2015). A similar strategy to prevent mutagenesis may be present in ss(+)RNA plant infecting viruses of the *Flexiviridae* and *Closteroviridae* family, many of which possess AlkB domains in virally expressed replicases. The AlkB domain possesses oxidative demethylase activity in vitro, with a preference for RNA substrates. Therefore, this domain may function to prevent nucleic acid methylation damage which causes base mis-pairing, thereby maintaining viral genome integrity. Blackberry virus Y (BVY), a member of the new *Potyviridae* genus *Brambyvirus*, contains an AlkB domain in its' P1 sequence, which shares several conserved residues with nucleotide recognition motifs found in vertebrate and bacterial homologous proteins (Van den Born et al, 2008). The BVY AlkB protein was shown to repair methylated ssRNA phage MS2 treated with methylation compound MMS when co-expressed in *E. coli* and was further shown to repair nucleic acid alkylation damage in vitro with substrate preference 10X higher for methylated RNA over DNA, which supports its suggested role in viral RNA repair (Van den Born et al, 2008).

Both the AlkB protein of BVY and HAM1 of CBSV and UCBSV have failed to suppress RNA silencing activity when co-expressed with dsRNA (Van den Born et al, 2008; Mbanzibwa et al, 2009; Pablo-Rodriguez, unpublished). Since both these viruses infected perennial, woody host plants, it may suggest they have acquired these cellular proteins as an adaption to maintain their genetic information with high fidelity under stressful environmental conditions during their infection cycles. Very little is known about the origin of unique viral proteins like HAM1 and AlkB and their role in plant

infections. Understanding their function in viral infections could be key to gaining an insight into virus-host interaction and how viruses adapt to infections in new host species.

1.8 Hypothesis and aims

CBSV and UCBSV are intriguing viruses to study due to their recent emergence and widespread affects as recent epidemics, as well as the presence of the unique viral encoded protein HAM1. Only three *Potyvirus* species (CBSV, UCBSV, and EuRSV) have been reported to encode homologous sequences to cellular Maf/Ham1 proteins. As these viral species all infect members of the *Euphorbiaceae* family it strongly suggests the presence of HAM1 is an adaption to infections in this plant family. Although CBSV and UCBSV possess diverse HAM1 sequences, their homology to cellular pyrophosphatases suggests they retain the function to intercept and hydrolyse mutagen causing NTPs such as ITP and XTP. Furthermore, previous *in planta* work by Pablo-Rodriquez (Unpublished) suggests CBSV Ham1 could function as an ITPase and reduce transitional changes in viral genomes. Transgenic *N. tabacum* expressing CBSV Ham1 were infected with PVY and TMV which do not encode any HAM1 like genes. 500bp regions of the viruses were then deep sequenced and infections compared between transgenic and wild-type *N. tabacum*. Comparison of SNPs showed a decrease in mutations within the viral regions sequenced in transgenic *N. tabacum* compared to wild-type plant infections, with relative HAM1 transgene expression correlating to a reduction in transitional changes in TMV and PVY genomes. Although only short viral regions were sequenced, this suggests a possible NTPase function for CBSV HAM1 and its possible role in viral stability during infection. Furthermore, infections of *N. benthamiana* with a CBSV Ham1 knockout virus produced different symptoms such as absence of systemic necrosis and longer periods of virus proliferation, whilst CBSV replication and movement was uninhibited (Pablo-Rodriquez, unpublished). Elucidation of this unique proteins' function may help to understand the evolution of these viruses and their interactions with host plants which causes devastating disease symptoms.

In this thesis, the activity of viral HAM1 proteins was assessed through the expression of UCBSV and CBSV HAM1 proteins and a subsequent nucleotide assay performed to evaluate their activity on canonical versus non-canonical nucleotides. Another approach to study the role of ITPase proteins in viral infections was to study the role of host plant copies during infections. As it is possible that viruses have acquired HAM1 to reduce the mutation rate in viral RNA genomes under host plant oxidative conditions or the variable expression of host nucleotide pyrophosphatase proteins. To explore whether host plant pyrophosphatases play a role in viral stability, the model plant *Arabidopsis thaliana* was used to compare viral titres during infection of a ss(+)RNA virus between plant wild type and host ITPase gene knockouts, in order to see if RNA viruses require host ITPase activity and are affected by

possible increased mutation rate. Furthermore, the expression of *Arabidopsis* ITPase was monitored during viral infection to see if it was modulated by the host as part of host plant defence against viral infection, which might reveal a previously undiscovered host defence pathway against viral infection.

Chapter 2
Materials and methods

2.1 Materials and standard procedure

All media, plastics, and glassware were sterilised by autoclaving at 121°C for 15 minutes at 15 p.s.i, unless sterilised by the manufacturer, and solutions were prepared with sterile deionised water (SDW) unless stated otherwise.

If autoclaving a solution was not appropriate, i.e. antibiotics, sterilization was performed using a 0.22 µm cellulose filter (Merk Millipore).

Chemicals and media used in this project were molecular biology grade, purchased from Thermo Fisher Scientific, Sigma-Aldrich, Melford, and VWR.

All statistical analysis was performed using IBM SPSS software platform.

Infected material and virus strains utilised during the project were made available from the University of Bristol Molecular Plant Pathology and Fungal Biology Group under DEFRA licence No. 51045/197610/2.

2.2 Culture media

2.2.1 Luria-Bertani (LB)

LB media was prepared using 25g of LB broth formula (Invitrogen) made up to 1 Litre with SDW

2.2.2 LB Agar (LBA)

LBA was prepared using 25g of LB broth with 15g of agar and made to 1 Litre with SDW

2.2.3 Super optimal broth with catabolite repression medium (SOC)

SOC media was prepared with 2% tryptone, 0.5% yeast extract, 10mM sodium chloride, 10mM potassium chloride, 10mM magnesium chloride, 20mM glucose, and made to 1 Litre with SDW

2.2.4 Terrific broth (TB)

TB media was prepared from 12g tryptone, 24g yeast extract, 4% glycerol, made to 900ml with SDW before addition of 100ml of phosphate buffer consisting of 0.17M potassium dihydrogen phosphate and 0.72M potassium dibasic phosphate.

2.2.5 Yeast extract nutrient broth (YENB)

YENB was prepared with 8.0g of nutrient broth and 0.5g yeast extract in 1L of SDW.

2.3 Antibiotic concentrations

Media were supplemented with antibiotics according to the working concentrations listed below. The antibiotic was used with the appropriate plasmid for the positive selection of transformed cells in both agar and broth media.

Table 1. Antibiotic concentrations used for plasmid selection

Antibiotic	Stock concentration (mg/ml)	Working concentration (µg/ml)
Ampicillin	50	50
Carbenicillin	100	100
Kanamycin	50	50

2.4 *Escherichia coli* cultures and transformation

E. coli strains BL21(DE3) (Thermo Fisher Scientific) and Top10 (Invitrogen) were normally streaked on an LBA plate and left to grow overnight at 37 °C before selection of a single colony for growth in 5-10ml of LB media at 37 °C overnight in an orbital shaker set at 180rpm. When culturing *E. coli* transformants, both medias were supplemented with a working concentration of antibiotic as listed above depending on the plasmid selectable marker

2.4.1 Preparation of *E. coli* electrocompetent cells

The desired *E. coli* strain was streaked on a LB agar plate without any antibiotic and incubated at 37°C overnight. A single colony was used to inoculate 5ml YENB media in a sterile universal tube which was incubated with shaking overnight. 0.5ml of this overnight culture was used to inoculate a sterile flask containing 50ml YENB, with a total of 5 flasks, which were incubated for 3-3.5 hours in a 37°C orbital shaker at 180rpm. The cultures were transferred to 50ml sterile sample tubes and centrifuged at 5000*g* for 20 minutes at 4°C. From this point, all procedures were performed on ice with chilled solutions. The supernatants were discarded, and pellets resuspended in 10ml of chilled SDW before combining the suspensions in 5 tubes with 20ml each. The suspensions were again centrifuged at 5000*g* for 10 minutes at 4°C before resuspension with 5ml SDW and combined into 2 tubes. The centrifugation was repeated, and pellets re-suspended in 20ml of chilled sterile 10% glycerol solution. The two tubes were centrifuged again, and pellets resuspended in 1ml of cold sterile 10% glycerol, with the final suspension being aliquoted to 50µl in sterile 1.5ml microcentrifuge tubes, before rapid freezing in liquid nitrogen and storage at -80°C.

2.5 Bacterial methods

2.5.1 *E. coli* transformation by electroporation

GenePulser electroporation cuvettes (electrode gap- 0.2cm) were sterilised in 75% ethanol before washing in SDW and dried. They were further sterilised by UV crosslinking in a UVP CL-1000 cross-linker for two minutes, then chilled on ice for a further 10 minutes.

An aliquot of 50µl electro-competent *E. coli* cells were thawed on ice along with a 30µl preparation of plasmid DNA. <2µl volume of plasmid DNA was added to the *E. coli* cells before transferring the preparation to a pre-chilled sterile cuvette which was electroporated using a BIO-RAD GenePulser electroporator at 2.5kV for 5msec. The cuvette was immediately chilled on ice before addition of 200µl of SOC media, the mixture was transferred to a sterile 1.5ml centrifuge tube and incubated at 37°C with shaking at 200rpm in an orbital shaker for 30 minutes. After incubation, 100 µl of the transformation mixture was plated onto LBA media containing the appropriate antibiotic for selection of the plasmid, followed by incubation overnight.

From the incubated plate, a single transformed colony was selected and used to inoculate 5ml of LB media supplemented with antibiotic in a sterile universal container. The mixture was incubated with agitation in the orbital shaker overnight, as previously, to propagate transformed cells for inoculating larger cultures or plasmid extraction.

2.5.2 *E. coli* plasmid extraction

After overnight incubation of a single transformant colony in 5ml LB, plasmids were extracted using the GeneJET Plasmid miniprep kit from Thermo Fisher Scientific. The overnight culture was harvested by centrifugation at 8000*g*. The supernatant was removed, and the cell pellet resuspended in 250µl of resuspension solution, before lysis of cells by inverting the suspension several times in 250µl of lysis solution. 350µl of neutralization buffer was then added and the solution applied to a GeneJET Spin Column by centrifuging for 1 minute. The column was washed and the flow through discarded before addition of 50µl elution buffer heated to 75°C to the column and centrifuging for 2 minutes to elute the plasmid. The flow through was collected and stored at -20°C until further use.

2.5.3 BL21 protein expression

Recombinant over-expression of CBSV HAM1 protein was achieved by inoculating 10ml of LB media with 100µg/ml carbenicillin with a single colony of BL21 *E. coli* transformed with Ham1 pOPINF plasmid construct. The culture was incubated overnight at 37°C with shaking before being used to inoculate 1L of LB supplemented with carbenicillin which was grown at 37°C in an orbital shaker until the optical

density of the culture at 600nm reached 0.4. Protein expression was induced with the addition of sterile IPTG to a final concentration of 1mM and the culture grown for a further 18 hours at 20°C with shaking. The culture was harvested by centrifugation at 5000*g* for 30 minutes at 4°C. The cell pellet was resuspended in 25ml Tris-buffered saline (150mM NaCl, 20mM Tris-HCl pH7.5) and transferred to a Falcon tube which was again centrifuged for 10 minutes and the resulting pellet frozen in liquid Nitrogen.

2.6 Protein Purification

The HAM1 protein was harvested from BL21 cells by resuspending and defrosting the cell pellet in 30ml Loading buffer before homogenizing the resuspension in a glass homogeniser on ice. The cells were then lysed using a cell disrupter (Constant systems) before centrifugation at 17,000rpm for 40min and 4°C to pellet cellular debris. The lysate was loaded onto a His-trap nickel column (GE Life Sciences) and set up on an AKTA machine with lines connected to elution and loading buffer. The column was washed with up to 20ml of Loading buffer containing 50mM Imidazole to elute any non-specific proteins before a 50% gradient of Elution (1M Imidazole) to Loading buffer was used to elute the protein in 1ml fractions. Fractions with a relatively high UV absorbance, indicating protein elution, were collected and pooled before running on an SDS-page gel to determine the presence of Ham1 protein and its purity. The protein solution was first concentrated using a 10KDa centrifugal concentrator (Vivaproducts) and then dialysed overnight in buffer.

Table 2. Buffers used for protein purification

Loading buffer	**Elution buffer**	**Dialysis buffer**
150mM NaCl	150mM NaCl	150mM NaCl
20mM Tris	50mM Tris	50mM $MgCl_2$
20mM Imidazole	1M Imidazole	1mM Dithiothreitol
10% Glycerol	10% glycerol	50mM Tris-HCl
Buffered to pH 7.5 with HCl	Buffered to pH 7.5 with HCl	10% glycerol
		Buffered to pH 7.5 with HCl

2.7 Plant methods

Arabidopsis seed lines containing a T-DNA insertion in plants with a Columbia-0 (Col-0) genetic background were ordered from the Nottingham *Arabidopsis* Stock Centre or were kindly provided by Dr Ioanna Kostaki of Bristol University. All soil was prepared by sieving and combining 3 parts soil to one-part sand before sterilisation in an oven at 60°C for 3 days.

2.7.1 T-DNA seed lines

Individual Transfer-DNA (T-DNA) lines with insertions at locus AT4g13720 were planted in 5X3 trays and grown under 16:8 h (day: night) light cycle days at 25°C in a growth cabinet.

2.7.2 Sucrose DNA extraction

Initially a sucrose preparation method was used for the rapid extraction of DNA from *Arabidopsis* based off Berendzen et al, (2005). Approximately 100mg of leaf tissue was collected in a sterile 1.5ml centrifuge tube to which 200µl of sucrose solution, containing 300mM NaCl, 300mM sucrose, 50mM Tris-HCl pH7.5, was added. The tissue sample was ground in the solution at room temperature using a micro-pestle before incubation of the homogenate at 100°C for 10 minutes. The cell debris was pelleted by centrifugation at 13,000*g* for 30 seconds, and the lysate removed without disturbing the pellet and stored on ice. 1-2µl of lysate was used as a DNA template per 25µl PCR reaction

2.7.3 Edwards DNA extraction

An adaption of Edwards (1991) genomic DNA extraction was used for larger scale processing of *Arabidopsis* mutants. 400 µl of extraction buffer containing 200mM Tris (pH 7.5), 250mM NaCl, 25mM EDTA, and 0.5% (w/v) SDS, was added to 100mg of leaf tissue. The samples were homogenised using sterile ball bearings in a tissue lyser, before centrifuging the preparations at 16,000*g* for 2 minutes. 300 µl of the suspension was transferred to a fresh 1.5ml centrifuge tube, before being mixed with an equal volume of isopropanol and incubated on ice for 2 minutes. The mixture was centrifuged for a further 5 minutes with the lysate being discarded and the resulting DNA pellet being washed in 100% Ethanol and left to dry at room temperature. The pellet was then resuspended in 100 µl of SDW, with 2 µl used for PCR reaction to validate the presence of the target mutation.

2.7.4 Viral Inoculation

Frozen *N. benthamiana* leaf tissue infected with CMV(Y) was used for non-purified mechanical inoculation of *Arabidopsis* leaves based on the method outlined in Ishikawa et al, (1991). 5g of frozen infected leaf tissue was homogenised in 10ml of 0.1M sodium phosphate buffer, before a 1:10 dilution

and dividing the solution into 1ml aliquots. *Arabidopsis* at 4 weeks post germination had a leaf which was the fourth-sixth to develop post germination brushed with carborundum powder. 1ml of the CMV solution was then applied to the leaf and stroked with a gloved finger. After several minutes the inoculated leaf was washed with SDW and the plant returned to the growth conditions stated previously.

2.7.5 DAS-ELISA

Double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) was performed on *Arabidopsis* leaf tissue to quantify the viral titre of Cucumber mosaic virus. A complete CMV kit (BIOREBA) containing antibodies and buffers was used according to the manufacturer's instructions. The CMV antibody provided was diluted 1:1000 in coating buffer and 200 µl added to each well of a 96 well ELISA plate and incubated at 4° overnight. Roughly 100mg of leaf tissue were homogenised using sterile ball bearings in a tissue-lyser before being vortexed with 750 µl of the extraction buffer provided. 200 µl of leaf solution was applied to the wells in triplicate and again incubated overnight. 200 µl of conjugated anti-CMV antibody (1:1000 dilution in conjugate buffer) was added to each well and the plate incubated for 5 hours at 30°C. 200 µl of para-nitrophenylphosphate (1mg in 1ml substrate buffer) was added to each well and incubated at room temperature in the dark for 1 hour before absorbance values were read at 415nm using a plate reader (Biorad Microplate Reader). Between each step the wells were washed 3 times with 200 µl of 0.05% PBS-Tween.

2.8 Molecular methods

2.8.1 Oligonucleotide primers

Oligonucleotides were designed using SnapGene software and ordered from Integrated DNA Technologies.

Table 3. Primers for InFusion cloning reaction of UCBSV in pOPINF plasmid

InFusion cloning primer	**Primer 5'-3'**
Forward	AAGTTCTGTTTCAGGGCCCGACAAAGGATTTGAGAGGAAG
Reverse	ATGGTCTAGAAAGCTTTACTGCACATCAATTGTTAGAGCCACCTTT

Table 4. Primers used for validation of *Arabidopsis* homozygous AT4g13720 locus mutant lines.

	Primers 5'-3'		
Plant line	**TDNA LB**	**Locus forward**	**Locus reverse**
SAIL_151	TAGCATCTGAATTTCATAACCAATCTCGATACAC	CATTAAATACTCGCGGTGAGAC	CGACACTATTACCTCTACTGCA
SAIL_1296	TAGCATCTGAATTTCATAACCAATCTCGATACAC	GTATGCAGAGATGGCAAAGG	GAATATAGAGAGGAGCAAGCACAG
GK-345B08	ATATTGACCATCATACTCATTGC	CATTCTTCTCCTGATTCTGTGC	ATCAGTTGGTCCTCTGGCTGG
Salk_072729	GCGTGGACCGCTTGCTGCAACT	CAGAAGCGGTAGTGTACGAGAGTC	GTGAACCTGAAGAGCAGCCA

Table 5. Primers used in RT-qPCR for *Arabidopsis* AT4g13720 and PDF2 reference gene

Gene primer pair		**Forward (5'-3')**	**Reverse (5'-3')**
AT4g13720	1	GGTCACGAAGGTCTGAACAA	TCATCTGTCTGGAACACATAGC
	2	GCTTCCTGGGCCATACATAAA	CACTGGATCCCATCCGAAATC
	3	GTGACGTTTGTGACTGGAAATG	TCCTCTGGCTGGAACTATCT
PDF2	1	TCATTCCGATAGTCGACCAAG	TTGATTTGCGAAATACCGAAC
	2	TGGAGAGTACGTGCTATAA	GAAGCGATACTGCACGAAGA
	3	AATGAGGCAGAAGTTCGGATAG	GACTGGCCAACAAGGGAATA

2.8.2 PCR

Polymerase chain reaction was performed with Dream Taq Polymerase (Thermo Fisher) in 20 µl reactions. Each reaction contained 10 µl of pre-made 2X master mix containing the Taq enzyme, buffer, $MgCl_2$, and dNTPs according to the manufacturer's instructions, as well as 1 µl each of forward and reverse primers (10 µM), 7 µl SDW, and 1 µl of the DNA template. A PCR cycle consisted of:

92°C for 2 min

X35 {
- 92°C for 30s
- 50-60°C for 30s (variable depending on Tm of the primers)
- 72°C for 30s per 500bp
}

72°C for 10 min

2.8.3 Electrophoresis

1X Tris-acetate-EDTA (TAE) buffer was used during electrophoresis and was prepared as follow: 40 mM Tris, 20 mM acetic acid, 1 mM EDTA.

Electrophoresis using agarose gels was performed to visualize and separate nucleic acids DNA and RNA according to their molecular size. The percentage of agarose gel was determined according to the size of the desired DNA fragment that was analysed. Agarose gels were typically prepared to 1% by w/v, dissolving 1 g of agarose (Type I, molecular biology grade, Bioline) per 100 ml of TAE buffer. The gel was made by heating the agarose/TAE solution and placing it on a casting tray with a comb to solidify. In order to visualize DNA molecules, 5 µl of Midori Green Advance Nucleic Acid Stain (NIPPON Genetics EUROPE) were added per 100 ml of prepared agarose gel. Reaction products containing nucleic acids were run on agarose gels at 120 volts for 30 minutes with samples being loaded inside the lanes created in the set gel by the comb. To determine the size of the DNA fragments on the gel, 5 µl of HyperLadder™ 1kb (Bioline) was run adjacently. The gel was then visualised under UV light on a ChemiDoc™ Bio-Rad System and imaged and recorded using the Qantity One® 1-D analysis software (Bio-Rad).

2.8.4 TRIzol RNA extraction

100mg of leaf tissue were taken at desired time points during plant infection, and the leaf material stored at -80°C. The frozen sample was homogenised in a tissue lyser and then the homogenate vortexed in 1ml of TRIzol reagent for 1 minute. The tissue sample was incubated in TRIzol for 5 minutes to allow complete dissociation of nucleoproteins from RNA. 0.2ml of chloroform was added and shaken for 15 seconds with a further incubation period of 3 minutes. The mixture was then centrifuged at 12,000*g* for 15minutes at 4°C. The resulting aqueous phase was separated and transferred to a new centrifuge tube, mixed gently with 0.5ml of chilled isopropanol and left to incubate for 10 minutes, before being centrifuged at 12,000*g* for 15minutes at 4°C to pellet the RNA. The supernatant was removed, and the resulting pellet washed with 1ml of 75% ethanol, before centrifuging again. The ethanol was removed, and the pellet left to air dry before resuspension in 50µl of SDW. The concentration of each sample was then measured using a Nanodrop spectrophotometer (Thermo Fisher) following manufactures instructions.

2.8.5 RT-PCR

Reverse transcription polymerase chain reaction (RT-PCR) was performed on RNA extracts to convert messenger RNA to cDNA using oligo dT primers. To remove dsDNA contamination from RNA extractions, 1 µl of DNase I and 1 µl of reaction buffer was added to 1 µg of RNA from each extraction

and made up to 10 µl with nuclease free water. The solution was incubated at 37°C for an hour before adding 1 µl of 50mM EDTA and incubating at 65°C for 10 minutes to terminate the reaction. The prepared RNA was then used as a template for reverse transcriptase. To each 1 µg RNA template sample, 1 µl of oligo dT_{18} primer was added along with 5X reaction buffer, RiboLock RNase Inhibitor, dNTP mix, and M-Mul-V Reverse Transcriptase supplied in a cDNA synthesis kit (Thermo Scientific). The reaction was incubated at 37°C for 60 minutes before terminating the reaction at 70°C for 5 minutes. The reaction was used in a PCR reaction with tissue specific genes to validate the presence of cDNA.

2.8.6 RT-qPCR

Reverse transcription quantitative PCR (RT-qPCR) was performed to determine quantitatively the relative expression of transcripts of interest in comparison with an endogenous reference gene. Reactions were performed using Maxima® SYBR Green/ROX (Thermo Fisher Scientific) and were carried out on a thermocycler Stratagene MX3005. PCR reactions were prepared to a final volume of 25 µl by adding the following reagents: up to 25 µl of endonuclease-free water, 0.4 µg of cDNA, 12.5 µl Maxima® SYBR Green/ROX, 2.5 µl specific primers at 10 mM. RT-qPCR cycles were set as follow: initial denaturation at 95 °C for 10 minutes, 40 cycles of: denaturation at 95°C for 30 seconds, annealing at 55 °C for 30 seconds and extension at 72 °C for 30 seconds. This was followed by a melting curve analysis for the primer performance measurement, doing a denaturation at 95 °C for 1 minute, with a continuous data collection from 55 °C to 95 °C. For the analysis of the crossing threshold (Ct) values. The threshold values for the SYBR Green were set by the MxPro software and the relative expression was calculated against the housekeeping gene using the 2-ΔΔCt method (Livak & Schmittgen, 2001)

2.8.7 Construction of pOPINF_UCBSVHAM1 plasmid

Primers were used to amplify the UCBSV Ham1 sequence in a PCR reaction using a previously constructed UCBSV infectious clone plasmid as a template, both supplied by Mrs Katie Tomlinson, PhD student Bristol University Molecular Plant Pathology Group. The UCBSV Ham1 sequence was amplified by high-fidelity Phusion PCR and run on a 1% Agarose gel to confirm its presence before purification using GeneJet gel extraction kit (Thermo Fisher). The pOPINF plasmid cloning vector was extracted from *E. coli* TOP10 overnight cultures and confirmed by enzyme restriction digest. Both the Ham1 template and pOPINF plasmid concentrations were measured using a Nanodrop spectrophotometer (Thermo Fisher) and an In-fusion cloning reaction performed with a 4:1 ratio of Ham1 insert: pOPINF plasmid. *E. coli* Top10 cells were transformed with the pOPINF_Ham1 vector from the In-fusion cloning reaction. The transformed cells were plated on agar plates containing 100mg/ml carbenicillin before

addition of 40 µl IPTG and 120 µl of X-Gal, the substrate for β-Galactosidase, for selection of transformed colonies by blue-white screening. Transformants were further confirmed by plasmid extraction from white colonies and PCR performed to detect the presence of a correct sized UCBSV Ham1 insert. Candidate plasmids were then sent for Sanger sequencing using a T7 complementary primer to determine the insertion sequence.

2.8.8 SDS-Page protein electrophoresis

SDS-Page was used to visualise BL21 recombinant expressed proteins by running the stained protein extract on 12% TruPAGE precast gels (Sigma). Protein gels were placed on a mini-Protean II vertical cell system (Bio-Rad) according to manufacturer's instructions. The inner container of the gel tank was filled to the top with running buffer (containing 25mM Tris-HCl, 200mM Glycine, and 0.1%w/v SDS) and the outside half-filled with the same buffer. 10 µl of protein samples were heated at 70 °C with the same volume 1x loading buffer (10% w/v SDS, 10mM DTT, 20% w/v glycerol, 0.2M Tris HCl pH6.8, 0.05% w/v Bromophenolblue) for 10 minutes and loaded in the gel. Gels were electrophoresed for 60 minutes at 130 volts powered by PowerPac™ Basic Power Supply (Bio-Rad). To determine the size of proteins in the fractions run, 10 µl of protein ladder pre-stained SDS-PAGE Standards Broad 7.1–209 kDa (Bio-Rad) was loaded adjacently. For the visualisation of protein, gels were stained and incubated with Coomassie brilliant blue staining solution, 0.1 g of Coomassie brilliant blue R250 in 40% methanol with 10% acetic acid and made up to 100 ml with SDW, with gentle agitation for up to 3 hours. Gels were then placed into de-staining solution, 40 % methanol with 10 % acetic acid and made up to 200 ml with SDW, for up to 24 hours with gentle agitation at room temperature. Finally, the gels were visualised under white light using a ChemiDoc™ Bio-Rad System and imaged and recorded using the Qantity One® 1-D analysis software (Bio-Rad).

2.8.9 Protein phosphatase assay

The bioassay buffer (50mM Tris-HCl pH8.5, 50mM $MgCl_2$, 1mM filter-sterilised DTT, made up to 50ml with Ultra-pure H_20) was prepared and pre heated to 37°C. 10mM stock solutions were made of nucleotides: dGTP, GTP, dCTP, CTP, TTP, dATP, ATP, UTP, dITP, ITP, and XTP. 0.2mM of each nucleotide substrate was incubated with 1.3µg of CBSV Ham1, UCBSV Ham1, and a BSA control. Each reaction was made up to 300µl with reaction buffer and incubated at 37°C for 20 minutes. PiColorLock Gold reagent (Innova) was prepared according to the manufactures instructions and 75 µl added to each reaction alongside 0.1U/µl of yeast Inorganic pyrophosphatase (Thermo scientific). 100 µl of reaction was then added to the wells of a micro titre plate in duplicate and the absorbance read at 655nm using a Biorad Microplate Reader.

Chapter 3
Results

3.1 Expression and functional analysis of viral Ham1 protein

Three viruses (CBSV, UCBSV, EuRSV) encode genomic proteins homologous to cellular ITPase pyrophosphatase. Although CBSV and UCBSV both encode HAM1 sequences, the amino acid identities shared between the two viral sequences is only about 48% (Winter et al, 2010). However, the presence of conserved pyrophosphatase motifs in both the CBSV and UCBSV HAM1 sequences suggest they both may act to hydrolyse non-canonical nucleotides such as ITPs and XTPs (Mbanzibwa et al, 2009). Therefore, to test this hypothesis, in this study both viral proteins were expressed and purified, and their pyrophosphatase activities were tested on a range of nucleotide substrates. Expression vectors containing the CBSV HAM1 sequence, and one containing a mutated SHR motif, had already been constructed. Therefore, an additional plasmid vector encoding the UCBSV HAM1 sequence was built. This allowed expression and purification protocols to be followed for all three homologous proteins, followed by a direct comparison of their relative yields and activities in a pyrophosphatase assay.

3.2 Construction of pOPINF_HAM1 plasmids

The expression of the viral Ham1 proteins in *E. coli* cultures requires a suitable expression vector. The plasmid pOPINF (Berrow et al, 2007) was chosen as it is compatible with ligation independent InFusion cloning system (Clontech) and achieves high copy number and expression in multiple hosts (fig 3.1). The plasmid is selectable by ampicillin resistance, allowing easy detection of correctly transformed cells on media supplemented with the antibiotic. The cloned insert is under *LacZ* control allowing induction of protein expression by addition of IPTG and contains an N-terminal His_6 fusion tag for Ni-affinity purification of the protein product.

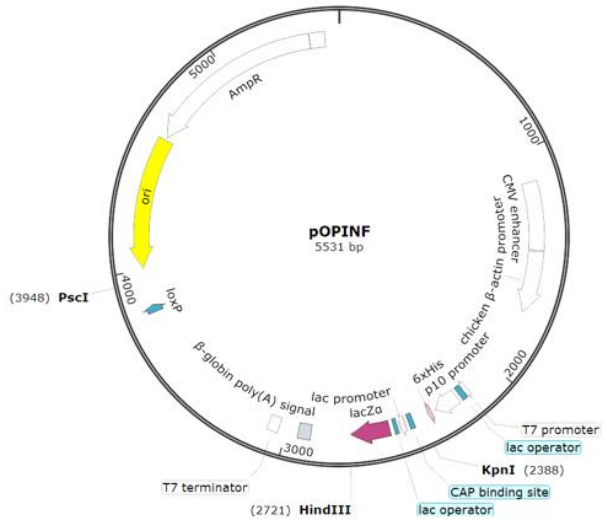

Figure 3.1 pOPINF vector map, showing restriction digest sites for cloning and vector confirmation, lac based promotor for protein induction, Ampicillin resistance gene, and 6x His tag for protein purification

Plasmids containing CBSV Ham1 and CBSV Ham1 mutant sequences (referred to as pOPINF_CHAM1 and pOPINF_CHAM1_mut respectively) were previously constructed by Mrs Katie Tomlinson. In brief, infectious clone (IC) constructs based on CBSV Tanza strain, pCambia0380_TanzaFULL+introns (Duff-Farrier et al., unpublished data) were used as a template to amplify the 678 bp HAM1 sequence. A mutated sequence was obtained using mis-matched PCR of the HAM1 sequence to produce a CBSV_mutHAM1 IC plasmid containing the desired SHR→SAA mutation, which was subsequently used as the template sequence for the pOPINF_CHAM1_Mut plasmid. A primer pair complimentary to the 5' and 3' ends of the HAM1 sequence was used to amplify both the WT CHAM1 and CHAM1_mut sequences from the corresponding IC plasmids in a high fidelity Phusion PCR reaction and the subsequent amplified sequences gel purified. The pOPINF vector was digested with *Kpn1* and *HindIII* restriction enzymes and was combined with both HAM1 inserts in two independent In-Fusion Cloning reactions with a 2:1 concentration ratio of insert: plasmid. The In-Fusion cloning mix was used to transform competent Top10 *E. coli* cells which were plated on carbenicillin selective LB agar plates. The plates were then incubated with 0.1M IPTG and 120ul of 5-bromo-4-chloro-3-indoyl-b-D-Galactopyranoside (X-Gal), the substrate for β-galactosidase encoded by the empty pOPINF vector. The *E. coli* cells which were successfully transformed with the correct pOPINF_CHAM1 constructs were then selected for by blue-white screening. The white *E. coli* colonies obtained were cultured and the presence of the correct sized insert confirmed firstly by PCR amplification of the insertion site. Two plasmids with the correct sized fragments were sent for Sanger sequencing across the insertion site using a T7-forward primer, with plasmid 2 shown to contain the correct SHR→SAA mutation in its HAM1 insertion sequence. Therefore, pOPINF_CHAM1_mut plasmid 2 was used for subsequent expression steps for the mutant sequence, alongside a confirmed pOPINF_CHAM1 plasmid for expression of the WT HAM1 sequence.

3.2.1 Construction of pOPINF_UHAM1 plasmid

A pOPINF plasmid containing the UCBSV Ham1 sequence (referred to as pOPINF_UHAM1) was constructed using the same method as previously described. The 678 bp UCBSV HAM1 sequence was amplified from a UCBSV Kikombe IC plasmid (supplied by Mrs Tomlinson) in two Taq polymerase reactions as test reactions to ensure proper amplification (fig 3.2). A high-fidelity Phusion polymerase reaction was used to obtain a fragment for cloning due to the polymerases' proof-reading activity (fig 3.3), using the primers:

pOPINF_UHAM1 Forward: 5'-AAGTTCTGTTTCAGGGCCCGACAAAGGATTTGAGAGGAAG-3'
pOPINF_UHAM1 Reverse: 5'-ATGGTCTAGAAAGCTTTACTGCACATCAATTGTTAGAGCCACCTTT-3'.

These primers contain a complimentary UCBSV HAM1 sequence (black) to amplify the 678bp HAM1 gene, and a pOPINF complimentary sequence (red) for insertion into the restriction digested pOPINF vector in an Infusion cloning reaction. Both primers have a Tm of 60°C based off the Ham1 specific sequence, as required for InFusion cloning.

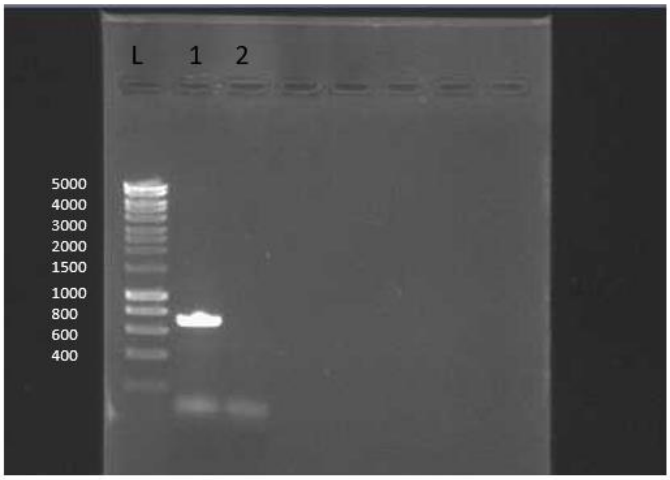

Figure 3.2 Taq Polymerase PCR of UCBSV HAM1 template (716bp) lane 1, lane 2 contains negative PCR control.

A single band containing the amplified HAM1 template was obtained and the negative control remained negative (fig 3.2). The band obtained from the Phusion polymerase reaction was purified using a GeneJET gel extraction kit (Thermo Fisher) using the manufacturers columns and instructions provided. 1/10th of the reaction volume was then run on a gel to confirm successful purification and the absence of contaminants (fig 3.3).

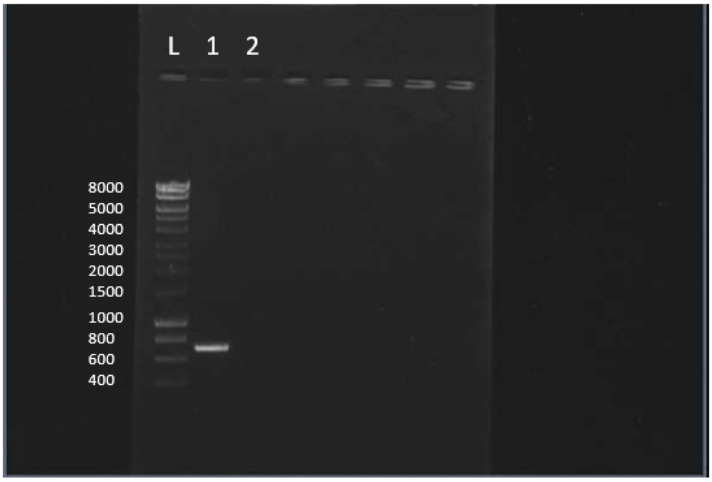

Figure 3.3 Phusion Polymerase PCR of 716 bp product of UCBSV HAM1 and pOPINF recombination sequences for InFusion cloning reaction (lane 1) and negative control PCR reaction (lane 2).

The plasmid pOPINF was obtained from glycerol stocks of Top10 *E. coli* containing the empty vector, which were plated on LB agar supplemented with ampicillin and the plasmid extracted using the GeneJET Plasmid miniprep kit. To check the pOPINF plasmid obtained, both a single and double restriction digest was performed on two separately extracted pOPINF plasmids, referred to as plasmid 1 and plasmid 2. Both reactions contained 1µl of each restriction enzyme and 5 µl of 5X fast digest buffer with 5 µg of pOPINF plasmid DNA, made up to 25 µl with SDW. The reactions were incubated for 15 minutes at 37°C and were stopped with 10 µl of 6X loading dye. The single digest reaction used *HindIII* to yield a predicted linearized fragment of 5,531 bp, whilst the double digest with *Kpn1* + *PciI* should give two fragments of 3971 and 1560 bps. Plasmid 1 gave correct sized fragments in both the single and double digest (fig. 3.4). However, plasmid 2 gave multiple bands in both reactions, which suggests that either it was not digested properly, it was not pOPINF, or it was simply contaminated with other DNA. Therefore, plasmid 1 was used in subsequent cloning reactions.

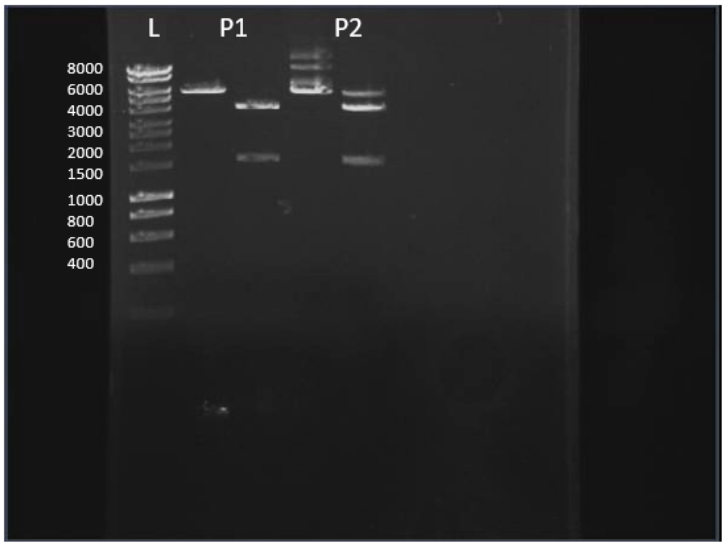

Figure 3.4 single digest of pOPINF plasmids 1 and 2 with *HindIII*, and double digest with *Kpn1* + *Pci1* restriction enzymes. Plasmid one gave correct 5531 bp product from linearization with *HindIII* and 3971 and 1560 bp products from the double restriction digest. Plasmid 2 produced multiple high molecular weight bands in both reactions suggesting it is not pOPINF or contaminated.

For the Infusion cloning reaction both the plasmid and UCBSV Ham1 fragment (hereon referred to as UHAM1) concentrations were measured using a Nanodrop spectrophotometer (Thermo Fisher) so a ratio of 4:1 of the template: plasmid could be used in the cloning step. The concentration of UHAM1 template was measured as 48 ng/µl whilst plasmid 1 was measured as 76 ng/µl. This gives a ratio of 6.3:1 (UHAM1: pOPINF) meaning 3.2 µl of template with 0.5 µl of linearized pOPINF plasmid was used

in a 10 µl cloning reaction alongside 2 µl of the Infusion enzyme for 15 minutes at 50°C. 2 separate cloning reactions were performed, both using plasmid 1 and the high fidelity amplified UHam1 fragment, to ensure a product was obtained.

Top10 *E. coli* cells were transformed with the cloning mixtures by electroporation followed by plating on LB agar supplemented with carbenicillin. Colonies were then incubated with IPTG and X-Gal as previously described for Blue-white screening. Abundant white colonies were obtained with both cloning reactions. The presence of pOPINF_UHAM1 plasmids was first confirmed using colony PCR for verify the UHAM1 insert. The pOPINF forward primer 5'-GACCGAAATTAATACGACTCACTATAGGG-3' was used in conjunction with the reverse primer 5'-ATGGTCTAGAAAGCTTTACTGCACATCAATT-3' which is complementary to UHAM1 and the pOPINF vector. The 941 bp product, confirming the correct insertion of the UHam1 sequence in the pOPINF vector, was obtained from a colony in the first cloning reaction, but the second colony sampled, and control, were both negative (fig. 3.5). The colony containing the UHAM1 insert was grown in an overnight culture supplemented with carbenicillin and the pOPINF vector extracted and sent for Sanger sequencing using the forward primer 5'-GACCGAAATTAATACGACTCACTATAGGG -3'. The sequenced pOPINF_UHAM1 was shown to contain the correct sequence for UCBSV Kikombe Ham1 which was cloned in-frame, the vector was therefore used in the subsequent protein expression steps.

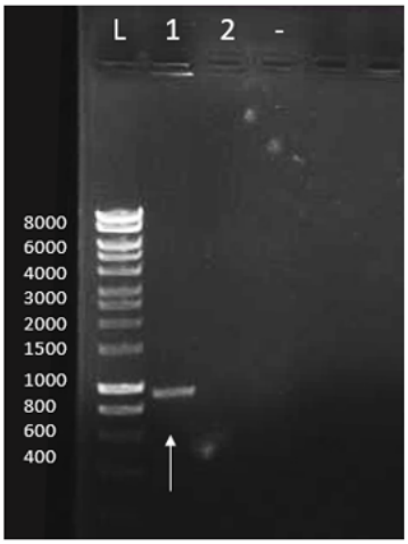

Figure 3.5 Colony PCR of transformed TOP10 cells with POPINF_UHAM1 clones 1 and 2 after blue-white screening. Top 10 cells transformed with pOPINF_UHam1 clone 1 (lane 1) gave correct 941bp product showing successful transformation with a pOPINF plasmid containing the UCBSV HAM1 sequence. Cells from the second cloning reaction (lane 2) failed to produce a PCR product meaning the correct pOPINF_UHAM1 plasmid is not present.

3.3 Ham1 protein expression

The BL21(DE3) strain of *E. coli* were used for expression viral Ham1 proteins as this strain is suited for high expression from T7-lacO promoter expression vectors, such as pOPINF. As in BL21(DE3) *E. coli* the T7 RNA polymerase required for gene transcription is produced from a λ-lysogen DE3 in the host bacterium, and its expression is under the control of the IPTG-inducible lac UV5 promoter.

The prepared plasmids pOPINF_UHAM1, pOPINF_CHAM1, and pOPINF_CHAM1_mut, were all transformed by electroporation into competent BL21 (DE3) *E. coli* cells. Transformed cells were grown in 5ml overnight cultures which was used to inoculate 2L cultures of LB media, both supplemented with Ampicillin. Protein expression from the 2L cultures was induced by addition of 1mM IPTG to allow T7 RNA polymerase expression and subsequent transcription of the HAM1 insert from the pOPINF plasmid. Protein expression was induced once the culture had reached an optical density of 0.4 and left overnight before harvesting the cells by centrifugation.

To extract the over-expressed Ham1 protein, the cell pellet was re-suspended and cells lysed in a cell disrupter (Constant systems) at 25 Kpsi, before loading the lysate in a His-trap nickel column (GE Life Sciences). The column was washed with 20ml of loading buffer to elute any bound non-specific proteins, before eluting the expressed HAM1 protein in 1 ml fractions with a 50% concentration gradient of loading buffer (20mM imidazole) to elution buffer (1M imidazole) using an AKTA machine. Fractions which caused a sharp increase in UV absorbance due to protein elution were collected and pooled. The collected fractions were then centrifuged in a 10KDa Vivaspin concentrator until they'd roughly decreased in volume by half, and then dialysed overnight in buffer (see chapter 2.6). 10 µl of resulting concentrated protein solution was then run on an SDS-PAGE gel, using the loading buffer as a negative control, to detect the presence of a HAM1 protein at 25KDa.

Initial attempts to express the WT CBSV Tanza Ham1 protein and CBSV Tanza Ham1 mutant (each successive extraction of CBSV Ham1 protein is hereby denoted as CHam1.1, CHam1.2 etc. and for CBSV Ham1 mutant protein as CHam1_mut.1, CHam1_mut.2 etc.) show several contaminants at higher molecular weights than the 25KDa expected for the Ham1 protein (figure 3.6). Furthermore, there a two concentrated low molecular weight protein bands, <18KDa, present in both samples. However, it is difficult to tell their actual size due to the compression of the migration front for low molecular weight proteins, suggesting the gel did not run correctly. There were no proteins evident in the loading buffer negative control, suggesting that the contaminants in the Ham1 purifications were due to non-specific bound proteins to the nickel column. Concentrated protein bands that might correspond to 25kDa overexpressed CBSV Ham1 was present in both proteins sample. However, the presence and

size of the bands are not clear due to the failure to separate low molecular weight proteins. Therefore, due to the high number of contaminants and the failure to run the gel correctly, both samples were repeated in a subsequent SDS-PAGE gel (fig 3.7).

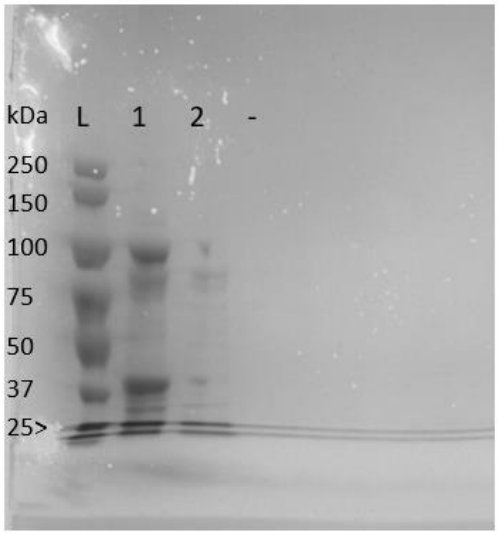

Figure 3.6 SDS-PAGE gel of CHam1.1 protein (lane 1) and CHam1_mut.1 protein (lane 2), alongside loading buffer control (-). CHam1.1 shows multiple contaminating protein bands which aren't present in CHam1_mut.1 purification. There is possibly an overexpressed 25kDa Ham1 protein band in both extractions though it is not clear due to compression of low molecular weight protein bands.

A subsequent SDS-PAGE gel with the Cham1.1 protein expression sample showed that a 25KDa protein, corresponding to the size expected for CBSV Tanza Ham1, was present at very low levels in the final elution volume with many high molecular weight protein contaminants present (fig 3.7 lane 1). The CHam1_mut.1 protein sample was run alongside, with the elution fraction showing virtually no protein traces on the gel. This lack of protein elution could be due to poor expression of mutant HAM1 in the culture or unfavourable conditions for protein binding and elution in the column as suggested by the lack of non-specific bound protein as is found in the CHam1.1 sample. Further attempts to express CHam1 (fig 3.7. lane 4) and CHam1_mut.2 (fig 3.7 lane 5) both showed very low Ham1 protein expression, with CHam1_mut.2 showing high amounts of contaminating protein. During dialysis, the Cham1_mut.2 changed from a colourless solution to orange, suggesting chelation of a metal by a protein in the elution volume, however the colour could not be quenched by addition of concentrated EDTA.

The Ham1 expression levels of both CHam1.2 and CHam1_mut.2 samples were deemed too low for it to be possible to separate and recover Ham1 protein from the contaminants present by size exclusion chromatography. However, UCBSV Ham1 was successfully purified with UHam1 (fig 3.7 lane 3) showing good protein expression at 25KDa and very low levels of contaminating proteins at other molecular weights.

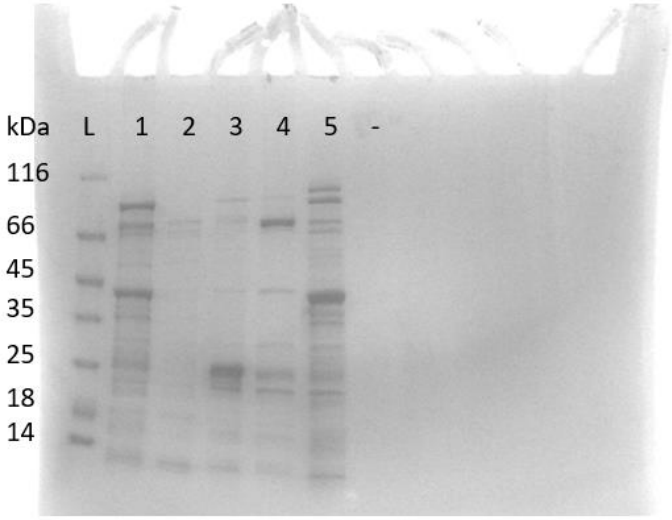

Figure 3.7 SDS-PAGE gel of separate Ham1 protein purifications. CHam1.1 protein (lane 1) and CHam1_mut.1 (lane 2), UHam1 (lane 3), CHam1.2 (lane 4), CHam1_mut.2 (lane 5), Loading buffer control (-).

Due to the difficulty of expressing and purifying wild type and mutant CBSV Ham1 proteins, the growth of BL21 cultures that were transformed with these vectors were measured 24 hours after protein induction. If Ham1 protein expression confers high metabolic strain and potential toxicity to BL21 *E. coli* cells, the transformed strains should grow at a much slower rate control cultures transformed with and empty pOPINF vector. The pOPINF_CHam1 and pOPINF_CHam1_mut constructs were compared to the control plasmid, all grown in 1L LB cultures at 37°C with gentle shaking. Protein expression was induced at an O.D of 0.8 to ensure higher initial *E. coli* biomass and therefore greater protein expression. The incubation temperature was reduced to 20°C to promote protein expression and folding. Prior to induction all three cultures grew at similar rates. However, upon IPTG protein induction at O.D 0.8, the pOPINF empty vector culture O.D more than doubled to a culture O.D of 1.8, whilst the pOPINF_CHam1 culture increased by an O.D of 0.27 over 24 hours, to an O.D of 1.1. The POPINF-CHam1_mut culture O.D increased by even less over 24 hours, only growing by 0.136 to an O.D of 0.97 (Fig 3.8). The 16% growth of the CHam1_mut culture compared to the 32% growth of the WT CHam1 protein culture suggests the protein mutation has a negative effect on *E. coli* growth and viability.

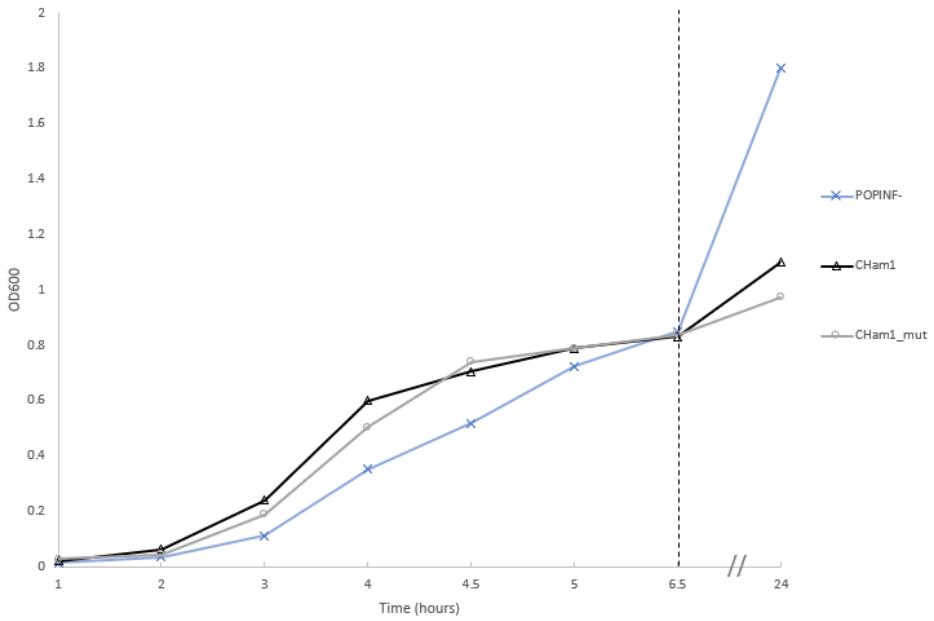

Figure 3.8. Growth of BL21 pOPINF_Ham1 cultures to measure metabolic strain of protein expression. Cultures were induced at O.D 0.8 (dotted line) and growth measured after 24 hours for CBSV Tanza Ham1 vector (black line), CBSV Ham1 SHR mutant (grey line), and pOPINF empty vector control (blue line).

Due to the negative effects of Ham1 expression on BL21 *E. coli* culture growth, the protein expression and purification stages were further optimised to increase protein yield. Firstly, Terrific broth (TB) was used as culture media, which was also supplemented with 1% w/v Glucose to prevent leaky protein expression from the POPINF Lac based promoters to allow optimal cell growth before protein induction with IPTG. Secondly, the induction temperature was lowered to 15°C as lowering the expression temperature (15-25°C) has been reported to improve the folding and solubility of recombinantly expressed proteins (Hannig and Makrides, 1998). At lower temperatures, cell metabolic processes slow down, and thus lead to reduced rates of cell growth, protein translation, and therefore reduced protein aggregation. Furthermore, *E. coli* cells containing pOPINF CBSV Tanza Ham1 and CBSV Ham1 mutant plasmids were cultured from the original frozen glycerol stocks.

The protein purification steps were also optimised by increasing the Imidazole concentration in the Loading buffer from 20mM to 50mM. Increasing the Imidazole in the initial buffer should decrease protein contaminants by reducing non-specific protein interactions with metal binding sites thereby improving target protein concentration and purity obtained in the elution fraction. Furthermore, the NaCl concentration in the dialysis buffer was increased to 300mM to improve protein solubility and activity, as white protein aggregates were evident after dialysis in some early protein purifications.

In the subsequent cultures of CHam1.3 and CHam1_mut.3 using the above changes, samples were taken at the different purification stages to see if change occurred in the protein bands (fig 3.9). Firstly,

samples of crude *E. coli* cell culture were taken from 2L batches of BL21 pOPINF_CHam1.3, pOPINF-, and pOPINF_CHam1_mut.3 (Fig 3.9 lanes 1,5, and 6 respectively). This was done by lysing the cells using repetitive freezing and thawing in liquid Nitrogen followed by treating with DNase1 to remove contaminating nucleotides. However, upon running samples on an SDS-PAGE gel very little protein was visible, therefore revealing little insight into HAM1 protein expression in the cultures (Fig 3.9 lanes 1,5, and 6).

Samples were then taken from the crude lysed cell extract after processing by the cell disrupter, and secondly, from the lysate after centrifugation of the lysed cells. These samples were run on the SDS-PAGE gel for analysis of CHam1.3 expression (fig 3.9 lanes 2 and 3) and for expression of CHam1_mut.3 (fig 3.9 lanes 7 and 8). The lysed Cham1.3 culture (fig 3.9, lane 2) shows a trace amount of a 25KDa protein which possibly corresponds to expressed WT CBSV Ham1. This faint 25KDa band is still present as a soluble protein in the lysate fraction after centrifugation of the lysed cell culture (fig 3.9, lane 3). A large band of 25KDa protein is present in the sample of the collected CHam1.3 fractions which showed increased UV absorption upon elution from the column (fig 3.9, lane 4). The presence of a 25KDa protein and the absence of other molecular weight protein contaminants in the fraction samples suggests successfully purified WT CBSV Tanza Ham1 protein.

No 25KDa band was present in the lysed culture and centrifuged lysate samples of CHam1_mut.3 (Fig 3.9 lanes 7 and 8), with only higher molecular weight *E. coli* cellular proteins present. A sample of pooled fractions of CHam1_mut.3 after elution, which showed no increase in UV absorption, is void of any protein bands (fig 3.9 lane 9). This shows that the increase in Imidazole to 50mM in the loading buffer prevented non-specific binding of contaminating proteins to the nickel column. As no noticeable mutant protein was expressed, it is difficult to tell if this concentration of imidazole would have prevented binding of the mutant protein to the column as well.

Finally, to see if the CHam1_mut.3 protein was present as an insoluble inclusion body, an extract from the culture cell pellet after cracking and centrifugation was mixed with 8M Urea to re-solubilise any cellular protein aggregates. However, no 25KDa protein was evident when this mixture was run on the gel (fig 3.9 lane 10), showing that the mutant protein was not expressed well in the culture (as was shown in fig 2.8) rather than being insoluble. Therefore, UCBSV Ham1 and CBSV Tanza Ham1 proteins were successfully expressed and purified allowing comparison of their pyrophosphatase activity. However, the CBSV Tanza Ham1 mutant protein was unable to be successfully expressed suggesting the mutation affected *E. coli* protein expression and cell viability.

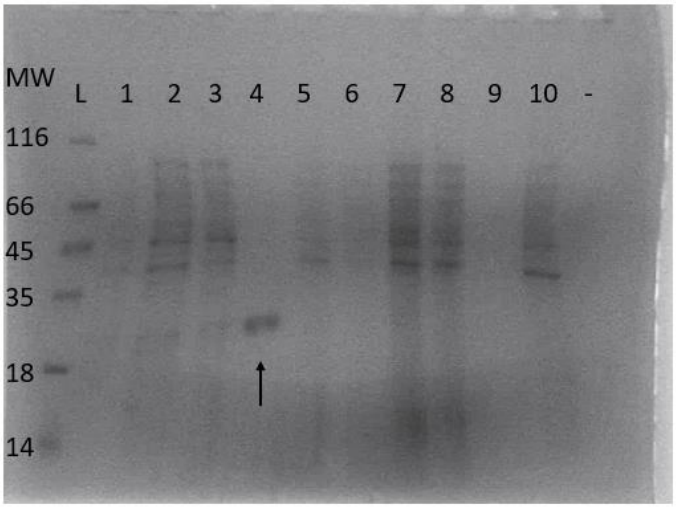

Figure 3.9 SDS-page gel of CHam1.3 protein purification stages. BL21 CHam1.3 culture crude extract (lane 1), CHam1.3 cell lysis extract (lane 2), CHam1.3 lysate extract after centrifugation (lane 3), CHam1.3 elution fraction (lane 4), BL21 POPINF- crude culture extract (lane 5), BL21 CHam1_mut.3 crude culture extract (lane 6), CHam1_mut.3 cell lysis extract (lane 7), CHam1_mut.3 lysate extract after centrifuge (lane 8), CHam1_mut.3 elution fraction (lane 9), BL21 CHam1_mut.3 cell pellet dissolved in 8M Urea (lane 10), Loading buffer control (-). Arrow indicates presence of 25KDa CHam1.3 band.

3.4 Measuring protein concentration

The concentration of Ham1 proteins was measured to ensure that equal amounts of protein was used in the bioassay to allow direct comparison of activity. Initially this was done using a Nanodrop spectrophotometer, however, this gave spurious repeat readings which appeared too high for the low expression of protein observed on the SDS-PAGE gels. Therefore, a Bradford assay was performed using a range of standard bovine serum albumin (BSA) concentrations up to 0.8 mg/ml (Fig 3.10). 5 µl of each protein standard was added to 250 µl of Bradford reagent, which was repeated for Ham1 protein samples in triplicate. The samples were incubated in the reagent at room temperature for 10 minutes before measuring the absorbance of the standards and samples at 595nm. The net absorbances of the standards were normalised to a blank sample and averaged before being plotted against protein concentration, allowing the protein concentration of Ham1 samples to be compared against the values of the standard curve (fig 3.10). This method gave a concentration of 0.313 mg/ml for the UCBSV Ham1 protein extract and 0.244 mg/ml for the CBSV Tanza Ham1 protein expressed. Although the data does not closely fit a linear trend through the origin as expected (R^2= 0.88), this is not a concern as the proteins were compared in the same assay and their relative concentrations can be accounted for in the pyrophosphatase assay.

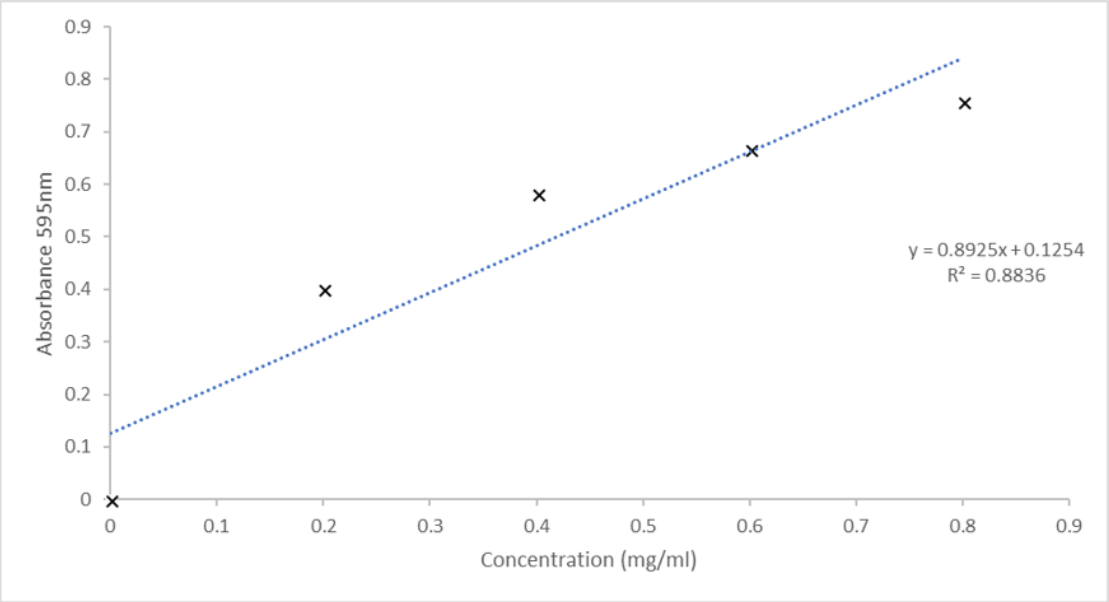

Figure 3.10. Bradford assay standard curve used to measure Ham1 protein concentrations. 5 BSA concentrations up to 0.8mg/ml were used to calibrate CBSV and UCBSV Ham1 protein concentrations using the linear equation displayed on the chart.

3.5 Analysis of U/CBSV Ham1 pyrophosphatase activity

Due to conserved amino acid sequences U/CBSV Ham1 is predicted to have pyrophosphatase activity (Mbanzibwa et al, 2009). If Ham1 is an ITPase family protein, the prime nucleotide to validate its activity is deoxyinosine triphosphate (dITP), due to the reported activity of homologous proteins with this nucleotide (Davies et al, 2012; Burgis and Cunnigham, 2006). An ITPase protein would cleave a pyrophosphate group from dITP, however this group is not detectable in a colourmetric assay. Therefore, a coupled assay was used with excess yeast inorganic pyrophosphatase which hydrolysed pyrophosphate to two inorganic phosphates (P_i), which can be detected by PiColorLock Goldmix (Innova) (fig 3.11). A PiColorLock assay uses the change in absorbance of the dye malachite green caused by the formation of phosphomolybdate complexes in the presence of P_i. An initial bioassay was performed to test both Ham1 protein activities in the assay conditions, and whether addition of an inorganic pyrophosphatase to the reaction mixture improved detection of Ham1 pyrophosphatase activity.

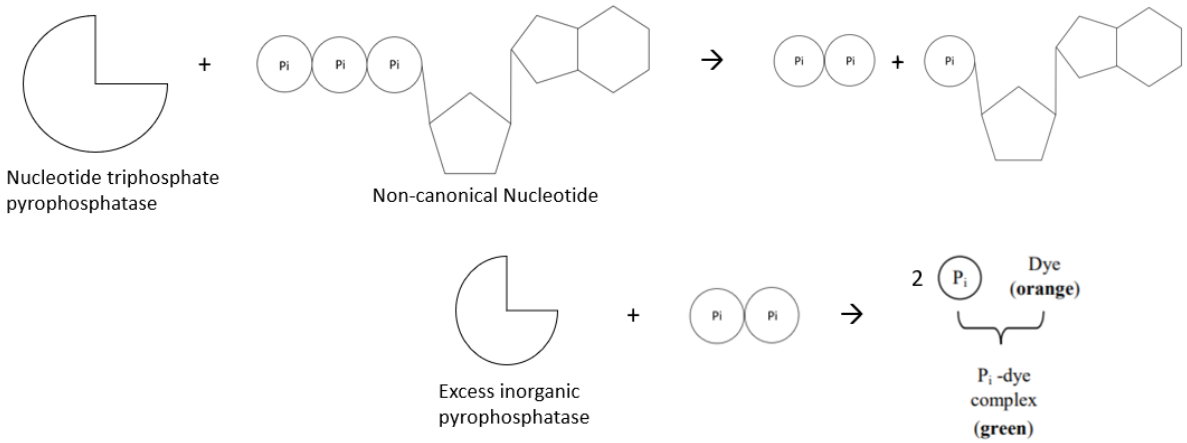

Figure 3.11 A diagram of protein triphosphate pyrophosphatase activity on non-canonical nucleotides. Ham1 NTP pyrophosphatase activity results in pyrophosphatase release which can be detected in a colorimetric assay by addition of excess inorganic pyrophosphatase and PiColorlock dye.

An assay with CBSV and UCBSV Ham1 with dITP demonstrated that the addition of the yeast inorganic pyrophosphatase greatly increased detectable phosphate release from viral Ham1 incubation with the nucleotide. Furthermore, incubation of the inorganic pyrophosphatase with reaction buffer and dITP in the absence of Ham1 did not cause significant detectable phosphate release compared to the reaction buffer with or without pyrophosphatase. This shows that the inorganic pyrophosphatase does not have any direct detectable activity with nucelotides or buffer, therefore any change in absorbance is due to Ham1 protein activity (fig 3.12). Furthermore, this shows detectable Ham1 activity under the conditions tested, allowing the comparison of UCBSV and CBSV proteins on a range of nucleotide substrates under these assay conditions.

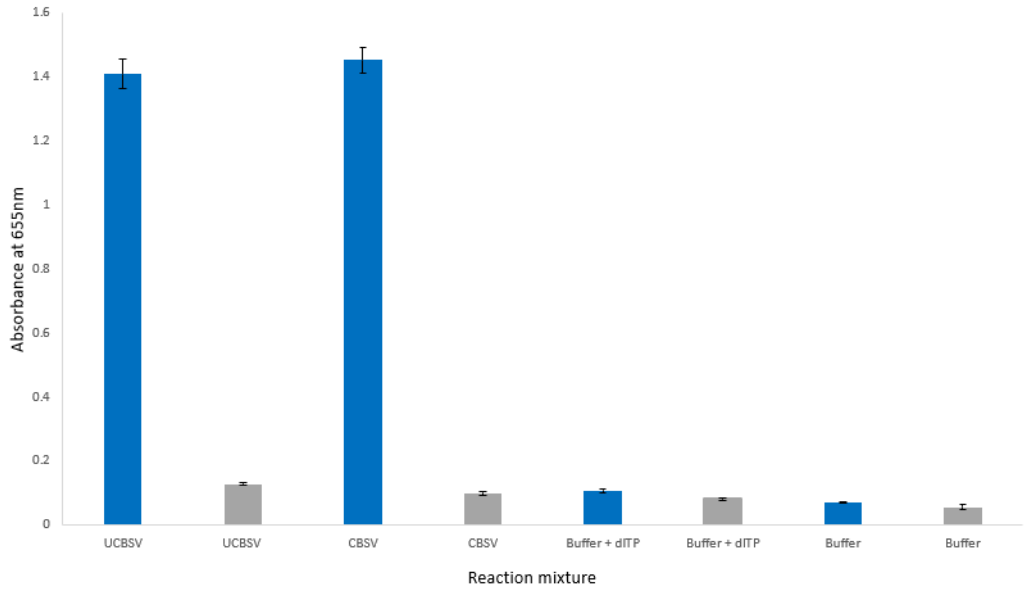

Figure 3.12. Phosphate bioassay testing the addition of PiColorLock Gold phosphate detection system on various reaction mixtures. Blue indicates the addition of 0.1 units of inorganic pyrophosphatase to the reaction whilst grey indicates it's absence. UCBSV and CBSV Ham1 proteins were incubated with dITP. Both protein show similar activity, and the addition of inorganic pyrophosphatase greatly increases the detection of phosphate release from protein activity but not with buffer controls.

The CBSV and UCBSV Ham1 were independently assayed with 10 different nucleotide substrates with triplicate repeats. 1.3 µg of protein was assayed with 0.2mM nucleotide, 75µl PiColorlock gold mix and made up to 300µl with reaction buffer. Reactions were incubated for 10 minutes at 37°C and the absorbance measured at 655nm. A 1.3 µg BSA control was incubated with each of the nucleotides in the in triplicate, with the mean enzyme free absorbance values subtracted from mean U/CBSV Ham1 values to account for any phosphate release due to protein interaction effects. Both U/CBSV Ham1 proteins were found to have the greatest activity with the non-canonical nucleotides XTP and dITP (fig 3.13). Activity of both proteins with the nucleotide ITP could not be measured due to the reaction forming a precipitate which, alongside a large background absorbance reading from non-specific phosphate release due to ITP instability, prevented accurate colorimetric measurement of viral Ham1 activity with ITP.

To determine whether a significant difference in U/CBSV Ham1 activities with non-canonical and canonical nucleotides was observed, a one-way ANOVA test was performed. Outliers were detected by box plots, with no data points lying 1.5 times the box length outside of the quartile range. The distribution of the data was tested by Shapiro-Wilk's test with most data normally distributed ($p>0.05$), except for data from dGTP incubation with CBSV Ham1 ($p>0.05$). The data for both CBSV and UCBSV showed un-equal variance, as assessed by Levene's test for equality of variances ($p<0.002$). Therefore, the post-hoc Games Howell one-way ANOVA assuming un-equal variance was performed.

Analysis by one-way ANOVA showed CBSV Tanza Ham1 had significantly higher activity when incubated with the non-canonical nucleotides XTP and dITP than with the canonical nucleotides UTP, dTTP, dATP, dCTP and ATP ($p < 0.05$). Similarly, UCBSV Kikombe Ham1 exhibited significantly greater pyrophosphatase activity with the non-canonical nucleotide dITP than with canonical nucleotides dTTP and ATP ($p < 0.05$), and incubation of UCBSV Ham1 with XTP showed significantly higher activity than with ATP ($p < 0.05$), with the remaining canonical nucleotide substrates not showing a significant difference to either dITP or XTP ($p > 0.05$). Both CBSV and UCBSV Ham1 proteins appear to cause relatively highly levels of orthophosphate release when incubated with dGTP and GTP nucleotide substrates, though Games Howell post-hoc analysis shows U/CBSV Ham1 activity with dGTP or GTP was not significantly higher than their activity with the other canonical nucleotide substrates tested ($p > 0.05$).

A two-way ANOVA was performed to account for differences in UCBSV and CBSV Ham1 protein activities with the nucleotides tested but found that the proteins do not have significantly different substrate activities, $F(9, 40) = 0.661$, $p = 0.738$.

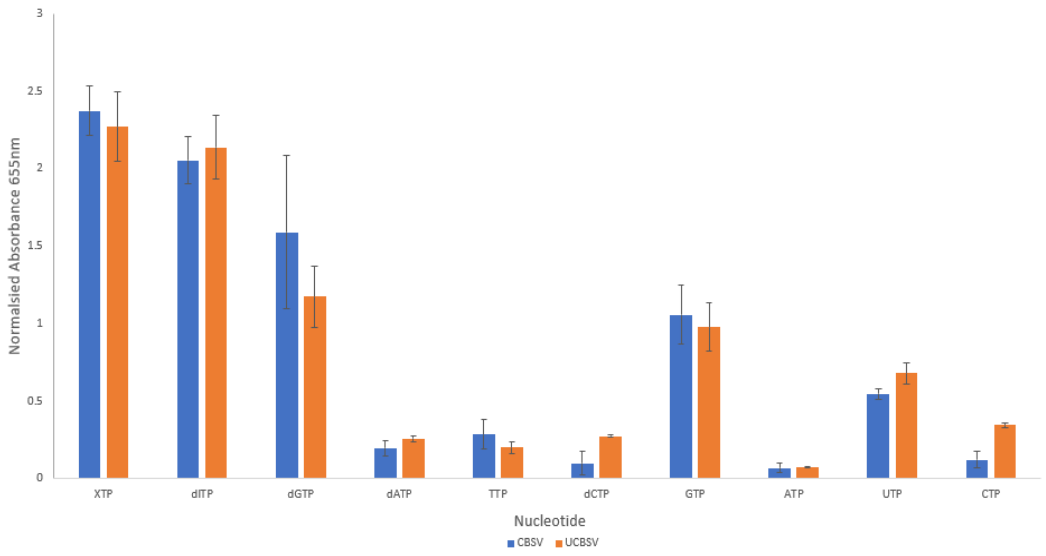

Figure 3.13. Phosphate bioassay testing activity of CBSV Ham1 protein (blue) and UCBSV Ham1 protein (Orange) on different nucleotide substrates. Both proteins had the greatest activity with the non-canonical nucleotides XTP and dITP. Analysis shows that incubation of CBSV Ham1 with non-canonical nucleotides XTP and dITP resulted in significantly higher phosphate concentrations than with canonical nucleotides UTP, dTTP, dATP, dCTP and ATP ($p < 0.05$). Incubation of UCBSV 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 ($p < 0.05$)

3.6 Investigation into potential U/CBSV Ham1 activity with M7G cap triphosphate

The phosphatase colorimetric assay (fig 3.13) demonstrated that U/CBSV Ham1 proteins have higher levels of activity with the canonical nucleotides dGTP and GTP relative to other canonical nucleotides. Unlike viral RNA, host mRNA has a 7-methylguanylate (m^7G) cap, which functions to promote cap-dependent translation, regulate host mRNA nuclear export, prevent exonuclease degradation, and promote exon excision (Ramanathan et al, 2016). The 5' cap of mRNA is derived from a guanine nucleotide which is methylated on the seventh position by a methyl transferase. The activity of U/CBSV Ham1 proteins with dGTP/GTP suggests that these proteins may function to hydrolyse M^7G triphosphate by cleavage releasing a pyrophosphate, similarly to the nucleotides previously discussed (fig 3.11). This would result in m^7G monophosphate which cannot be used to cap nascent mRNA. A reduction in host mRNA capping by viral Ham1 proteins could potentially cause reduced mRNA processing and gene expression, disrupting host plant cell metabolic processes.

To test this hypothesis, a colorimetric phosphatase assay was performed according to the protocol previously described, containing 1.3 µg CBSV or UCBSV Ham1 proteins incubated with 0.2 mM m^7G triphosphate cap analogue (Promega), and the nucleotides dITP and dATP. The absorbance values from negative control assays with BSA were subtracted from absorbance values from U/CBSV Ham1 protein assays to account for non-specific protein effects on phosphate release. It was found that both CBSV and UCBSV Ham1 proteins have minimal pyrophosphatase activity with the m^7G triphosphate cap analogue, compared with high levels of activity with dITP and low activity with dATP nucleotides.

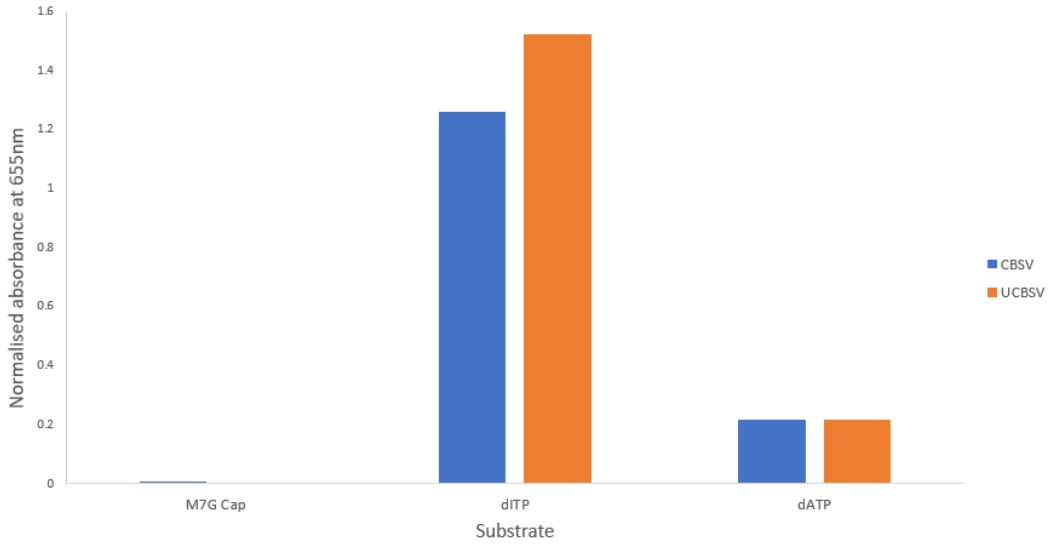

Figure 3.14. Absorbance values measured at 655nm produced in an enzyme assay with 1.3 µg CBSV or UCBSV Ham1 proteins incubated with 0.2 mM nucleotide substrates: M7G triphosphate cap analogue, dITP or dATP. Absorbance values from assays containing BSA were subtracted from absorbance values containing U/CBSV Ham1 proteins to account for phosphate release caused by protein addition

3.7 Investigating the interaction of viral infections with plant host Ham1

Many sequenced plant genomes such as *Zea mays*, *Oryza sativa*, *Sorghum bicolor*, *Triticum aestivum*, and *Arabidopsis thaliana* contain 1 or more ITPase genes but their functions are largely unexplored (Alexandrov et al, 2009; Yu et al, 2005; Paterson et al, 2009; Brenchley et al, 2012; Mayer et al, 1999). Without knowing what plant Ham1-like proteins do or whether they are involved in infections, it is difficult to understand the role of viral encoded Ham1. Therefore, the role of host plant ITPase proteins in response to viral infection was studied.

One hypothesis is that U/CBSV have acquired Ham1 proteins to reduce the mutation rate during replication of viral RNA genomes under oxidative conditions or the changeable expression of host plant nucleotide triphosphate pyrophosphatase proteins. Infections of Cucumber mosaic virus (CMV), which lacks a Ham1 protein, in *Arabidopsis thaliana* lines containing an ITPase protein mutation were studied to test the hypothesis that viruses in the presence of mutagens undergo a population 'error catastrophe' when the mutation rate exceeds a threshold level. Under this scenario *in planta*, the viability of the viral population would be reduced due to excessive mutations leading to a decrease in replication fidelity, viral infectivity and eventually population extinction (Anderson et al, 2004).

There is already evidence of plant viruses, such as BVY, containing AlkB domains in viral proteins to prevent methylation damage to the virus genomes (van den Born et al, 2008). It is therefore thought that U/CBSV contain the Ham1 protein to protect their genomes from incorporation of mutagen causing non-canonical nucleotides under oxidative infection conditions. However, the ability of proteins like Ham1 to protect an RNA genome from mutational damage has not been studied, largely due to the usually non-heritable nature of RNA compared to DNA genomes, which have been much more widely studied in this capacity.

To assess whether RNA viruses require host ITPase activity to maintain virus stability and infections, the model plant *Arabidopsis thaliana* was used to compare viral titres during infections with the ss(+)RNA virus Cucumber Mosaic Virus (CMV) between plant wild type and ITPase mutant plants. Viral titres were measured at a peak point in the infection cycle to see if the RNA virus population stability and replication were affected by a possible increase in the mutation rate cause by loss of host NTPase protein activity. Furthermore, the expression of *Arabidopsis* ITPase was monitored during a viral infection to see if it was modulated by the host as part of plant defence against viral infection. *Arabidopsis* was chosen as the plant to study due to the abundant mutant seed lines available, its quick reproduction cycle and ease of viral inoculation.

3.8 *Arabidopsis thaliana* seed lines

The locus tag for the closest Inosine triphosphate pyrophosphatase gene in *Arabidopsis* is AT4g13720, correspoding to a 206aa protein which shares 62% protein sequence homology to *Homo sapiens* ITPase (fig 3.15). *Arabidopsis* collections were searched for suitable mutant insertion lines at this locus. Seeds from SAIL_1296, SAIL_151, SALK_072729 and GABI-Kat 345B08 lines containing a T-DNA insertion, generated by vacuum infiltration of a Columbia (Col) plant with an *Agrobacterium tumefaciens* vector, were ordered from the Nottingham Arabidopsis Stock Centre with each line containing 100 seeds. The insertion of a T-DNA fragment into a plant host genome is caused by the transformation process of Agrobacterium infections. The Agrobacterium transfers a DNA fragment flanked by 25 bp border sequences (the T-DNA) from a modified tumour inducing (Ti) plasmid into the infected *Arabidopsis* genome at a random site, allowing large mutant catalogues to be produced. The T-DNA insert contains a known DNA sequence, the left border (LB) sequence, allowing primers designed from LB of the specific T-DNA vector to be used to isolate the genomic/T-DNA insertion junction. The genomic sequence adjacent to the insertion site can be mapped to the genome to precisely identify the chromosomal insert location for the *Arabidopsis* lines, and thus the gene disrupted.

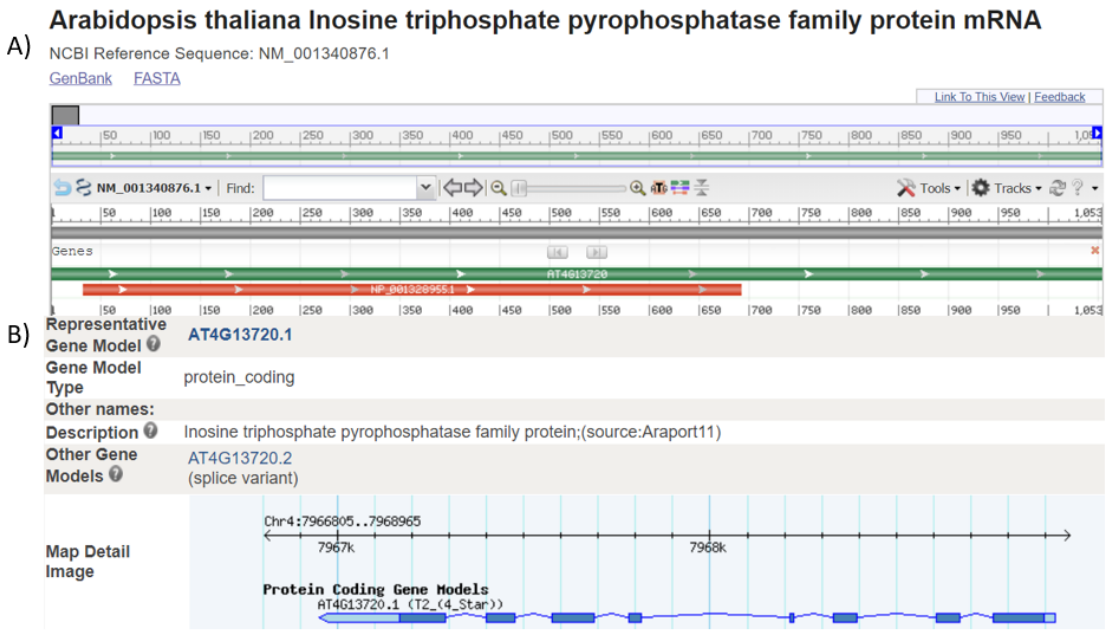

Figure 3.15. A) Blast nucleotide search for *Arabidopsis* Inosine triphosphate pyrophosphatase gene confirming AT4g13720 as the gene locus. B) View of locus details and protein model for the identification of T-DNA insertion lines in TAIR

3.9 Confirmation of T-DNA insertions in *Arabidopsis thaliana* seed lines

Seeds were initially grown in 3 parts soil to 1 part sand, sterilised in a 60°C oven for 72 hours, under a 16 hours light and 8 hours dark cycle at 25°C to maximise their growth rate. After 4 weeks of growth there was enough leaf material to begin DNA extraction for identification of homozygous ITPase mutants. PCR analysis was used to segregate plants containing homozygous insertions, which were then allowed to self-pollinate, and the seeds collected. The PCR reaction confirms homozygous gene-knockouts by T-DNA insertion by using a combination of a primer pair specific to the AT4g13720 genome sequence (hereby referred to as genomic primers) and a primer complimentary to the T-DNA left border (LB) sequence (table 4) (fig 3.16). No amplification of a PCR product with genomic primers suggests the presence of a T-DNA insert which due to its size disrupts the polymerase-based amplification of the target gene. This can be confirmed by amplification from a T-DNA LB primer and a complementary genomic primer across the insertion junction, with the correct sized PCR product confirming the presence of the T-DNA insertion at that locus. A PCR product in both reactions suggests the plant is heterozygous for the T-DNA insertion. For this experiment, complete elimination of the *Arabidopsis* ITPase gene expression was required so only plants which were homozygous for T-DNA insertions were selected and their seed collected.

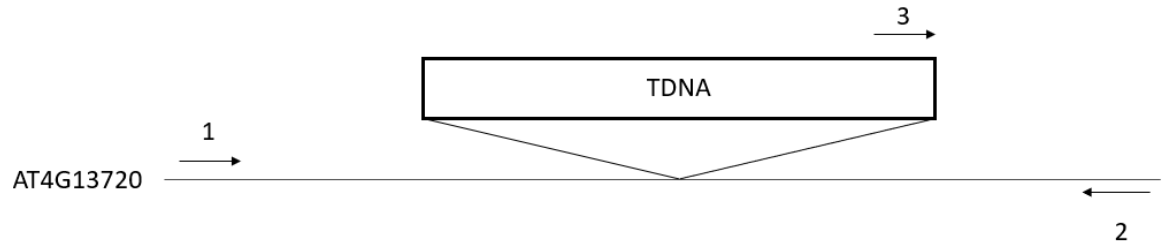

Figure 3.16 Illustration of the Arabidopsis HAM1 gene AT4G13720 with a TDNA insertion and primer for confirmation by PCR analysis. For homozygous mutant lines the primer pairs for amplification of the genomic Ham1 sequence (1 and 2) would be disrupted, and amplification across the T-DNA insertion border using an LB primer (3) and complimentary genomic (2) primer would yield a product confirming the correct insert is present.

Initially, a sucrose preperation method was used for rapid extraction of DNA based on Berendzen et al (2005). Approximately 100mg of fresh leaf tissue, equating to 1 large leaf or 2-3 smaller leaves, were placed in 200µl of a sucrose solution consisting of 50mM Tris HCl pH 7.5, 300mM NaCl, and 300mM sucrose. The solution was ground at room temperature using a micro-pestle before incubating the mixture at 100°C for 10 minutes and centrifuging at 13,000*g* for 30 seconds. The lysate was then removed into a sterile eppendorf and placed on ice until all samples were ready. 1-2µl of lysate was then used as a template per 25µl PCR reaction. Tests for the efficiency of DNA extraction by this method using genomic primer pairs revealed inconsistent and low quality template amplification (Fig 3.17). Therefore, a more thorough extraction method based on Edwards *et al*, (1991) was used.

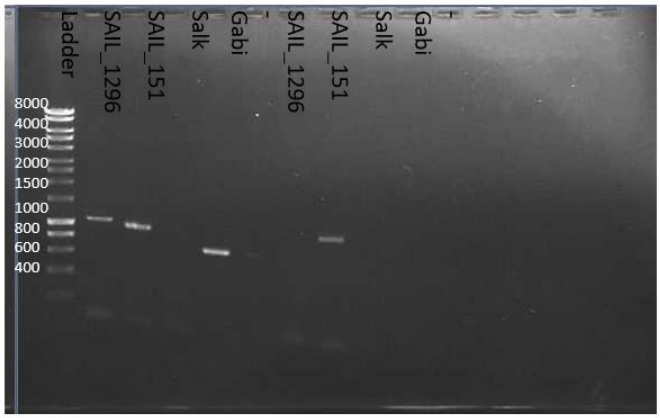

Figure 3.17 Test PCR of sucrose DNA extractions for different seed lines. Locus specific genomic primer pairs were used to check the quality of sucrose DNA extraction preparations of *Arabidopsis* seed lines: SAIL_1296, SAIL_151, Salk, and Gabi.

Edwards' DNA extraction method involved placing 100mg of leaf sample in a sterile eppendorf before grinding on ice with 400µl extraction buffer: 200mM Tris HCl pH7.5, 250mM NaCl, 25mM Ethylenediaminetetraacetic acid (EDTA), 0.5% w/v Sodium dodecyl sulfate (SDS). The mixture was vortexed for 5 seconds and left at room temperature until all preperations were ready. The samples were then centrifuged at 16,000*g* for 2 minutes, and the suspension transferred without disturbing the pellet. 300 µl of Isopropanol was added to the lysate before gentle mixing and left on ice for 2 minutes to precipitate the DNA. The preparations were again centrifuged for 5 minutes and the supernatant discarded. The pellet was washed with 70% Ethanol and left to dry at room temperature before resuspension in 100 µl of SDW. This was repeated for each seed line, and 2 µl of each preperation used in PCR reactions with genomic primers, which showed much more consistent amplification of target DNA (Fig 3.18). To maximise the processing ability of all seed lines, a tissue lyser with sterile ball bearings was used to grind up 20 samples at a time after freezing in liquid Nitrogen.

The ground samples were then vortexed with extraction buffer for up to a minute for extraction of DNA.

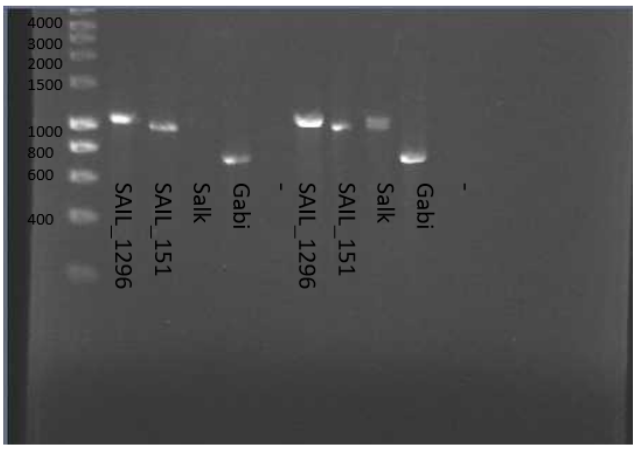

Figure 3.18 Test PCR of Edwards DNA extractions for different seed lines. Locus specific genomic primer pairs were used with 2µl of Edwards DNA preparations of *Arabidopsis* seed lines: SAIL_1296, SAIL_151, Salk, and Gabi, to check quality of DNA obtained.

Using Edwards' extraction method, 266 *Arabidopsis* were screened for homozygous T-DNA insertions in the AT4G13720 locus. Figure 3.19 demonstrates the PCR assays that were performed, firstly with AT4G13720 locus specific primer pairs, and also with a T-DNA LB primer and locus specific reverse primer pair. Amplification with only the locus specific primer pair indicates a WT plant without an insertion at this genomic site. Amplification of both the locus specific primer pair and T-DNA-locus primer pair indicates a heterozygous T-DNA insertion. For the purpose of this experiment complete disruption of *Arabidopsis* ITPase gene was required, meaning only homozygous plants identified by PCR amplification from the T-DNA and reverse locus primer were selected. In total, 16 homozygous T-DNA mutant *Arabidopsis* were identified by this method, which were re-analysed for confirmation (Fig 3.20). All mutants were identified from GABI-Kat seed lines which made up the majority of seeds planted, no homozygous Sail or Salk lines were identified by PCR. This may be due to the difficulty of clean DNA extraction from plant samples preventing efficient identification of T-DNA insertions, as can be seen in lanes with poor amplification of target DNA in Fig 3.19.

Mutant Arabidopsis showed no visible phenotype when compared to WT plants, with no reduction in germination or viability evident in homozygous plants. This suggests *Arabidopsis* Ham1 doesn't play a role in plant growth and morphology under the growth conditions used in this study.

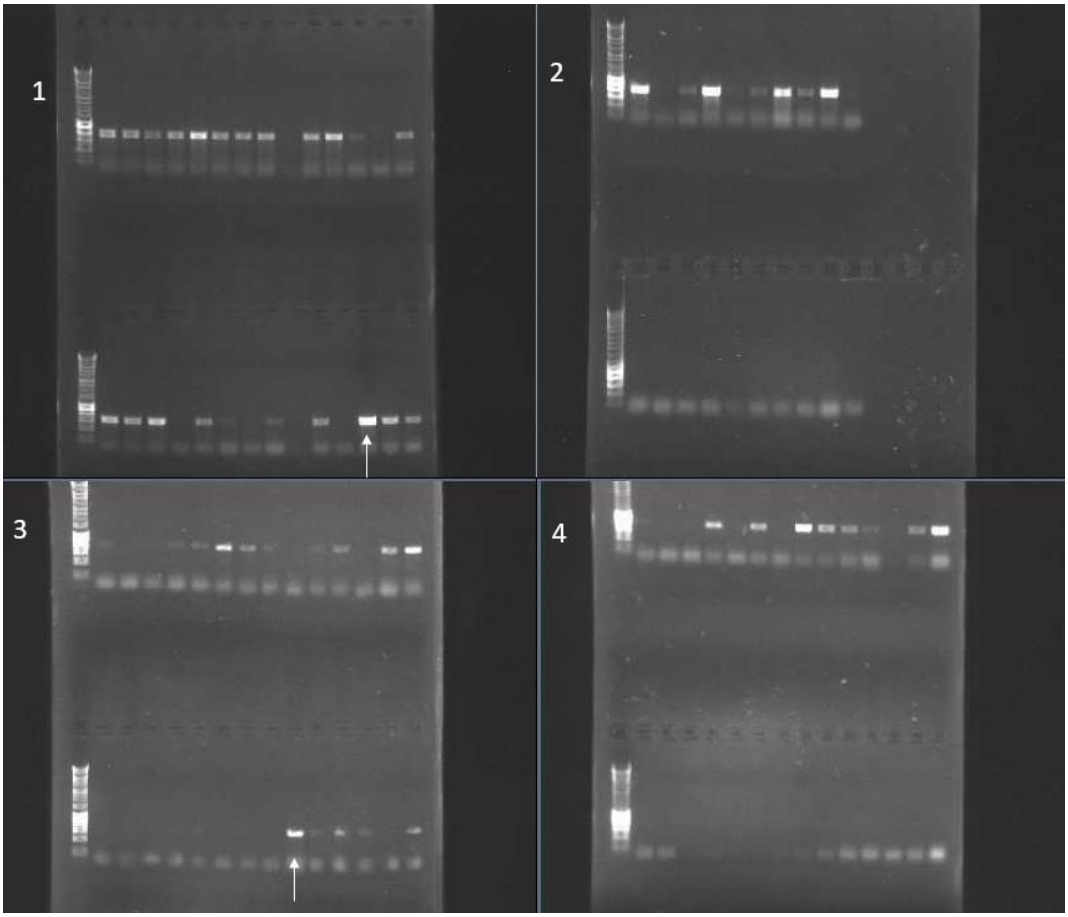

Figure 3.19. Example of 4 PCR assays to detect the presence of mutant *Arabidopsis* in GABI line of seedlings. Each assay is of 14 plants, with the top lanes testing amplification of locus specific primers for each plant, and the bottom lanes testing amplification from a T-DNA LB primer to a reverse locus primer. Homozygous T-DNA insertion plants, marked by white arrows, show no amplification with locus specific primer pairs but produce a PCR product from amplification across the T-DNA-locus border. Meaning successful T-DNA insertion in the *Arabidopsis* ITPase locus.

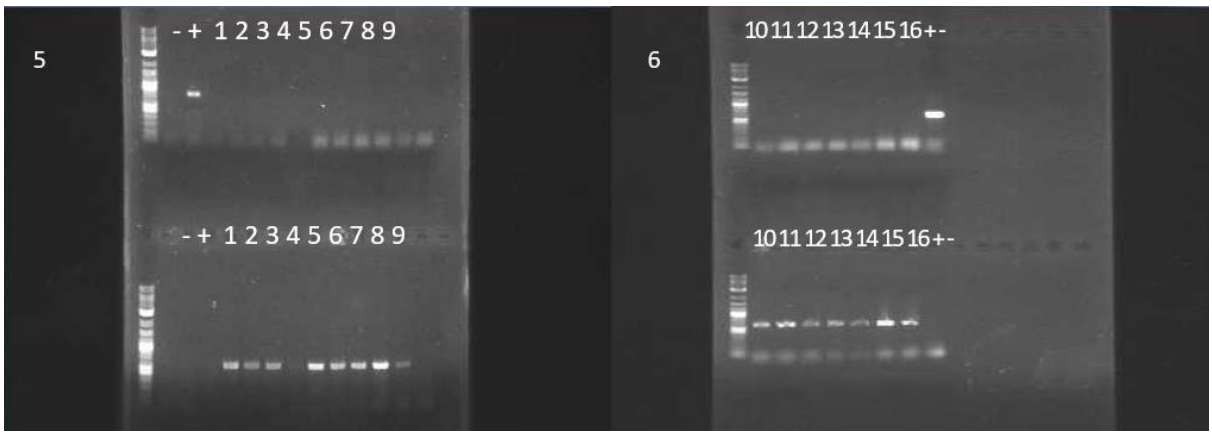

Figure 3.20. Secondary PCR assay of identified T-DNA insertion *Arabidopsis* mutants. The top lane is amplification of locus specific primers for each plant, and the bottom lane amplification from a T-DNA LB primer to a reverse locus primer. Homozygous T-DNA insertion plants show no amplification with locus specific primer pairs but show amplification of template across the T-DNA-locus border. 15 homozygous mutants were identified, Col-0 WT *Arabidopsis* were used as positive control for genomic primer pair amplification.

3.10 CMV *Arabidopsis* serological assay

To ascertain whether host plant nucleotide pyrophosphatases affect viral infection levels, *Arabidopsis* Ham1 WT and confirmed GK-345B08 homozygous mutant plants were infected with the +ssRNA virus CMV. At a peak point in infection, systemic leaf viral titres were measured by double antibody sandwich enzyme-linked immuno-sorbent assay (DAS-ELISA), and the levels of infection compared between plant groups. *Arabidopsis* is not reported as a host for U/CBSV, therefore the aternative RNA virus CMV was used. CMV is a member of the virus family *Bromoviridae,* and was chosen because of its broad host range, allowing it to infect *Arabidopsis* (Takashi et al, 1994), and also because it is a +ssRNA virus which lacks any pyrophosphatase motif in its genome. Little has been studied about the impact of nulceotide mutagens on heritable RNA molecules, such as the genomes of CMV and U/CBSV. Therefore, if pyrophosphatases are required to protect the integrity of these genomes from mutation causing nucleotides, we would expect viral titres and infectivity to be lower in mutant plants compared to WT. Both CMV and mutant *Arabidopsis* lack pyrophosphatase enzymes therefore potentially allowing the incorporation of mutation causing non-canonical nucleotides into viral genomes which could possibly cause viral population collapse.

20 *Arabidopsis* mutant seeds, from mutant plant 1 tested in fig 3.20, and 20 Col-0 WT seeds were planted in 3:1 soil to sand, with both plant groups grown in growth cabinets under short day cycles of 10 hours light to 14 hours dark. Shorts days were chosen for the production of larger leaves to allow ease of mechanical inoculation. To further confirm genotype of selected lines, two plants from each group were selected at random for DNA extraction and PCR analysis using the methods discussed above. Both Col-0 plants were negative for T-DNA and contained undisrupted HAM1 locus', whilst both mutant plants were positive for homozygous T-DNA insertions (Fig 3.21).

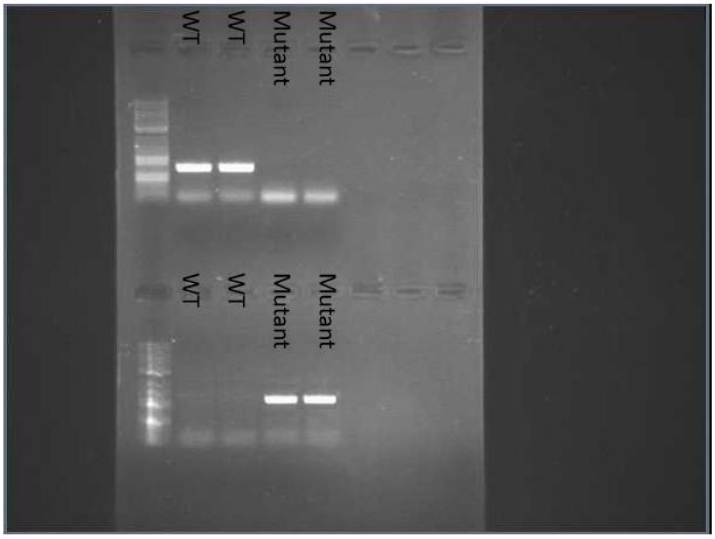

Figure 3.21 Final screening of PCR analysis of WT and mutant *Arabidopsis* for CMV serological assay. Amplification with locus specific primer pairs (top) and T-DNA and complementary reverse locus primer pair (bottom) confirm WT Col-0 and homozygous mutant plants.

CMV virus inoculum was initally propagated in *N.tabacum* and the leaves harvested and frozen 3 weeks post infection. To ensure as equal amount or viral material per plant, inoculation was based on methods in Ishikiwa (1991), whereby 5g of infected *N.tabacum* leaf was ground in 10ml of 0.1M Sodium Phosphate buffer until homogenised. This solution was diluted by a factor of 10 before dividing into 1.5ml aliquots. 28 days post sowing, the 5th rosette leaf of each *Arabidopsis* was inoculated by addition of carborundum powder and stroking with a gloved leaf. The leaf was then bathed in the 1.5ml CMV inoculum solution for 5 minutes before washing with SDW. Yoshii et al, (1998) measured CMV accumulation in *Arabidopsis* and showed that CP was most detectable in systemic leaves 10-14 days post inoculation. With upper rosette leaf parts (> fifth leaf) showing more rapid CMV accumulation than lower leaf parts (< fifth leaf). Therefore, leaves were samples from both upper and lower rosette parts as leaf age and position may effect viral titres measured. leaf samples were collected for DAS-ELISA with 8 leaves sampled from younger rosette leaves (> sixth leaf) and 8 from older leaves (< sixth leaf) for both mutant and WT plants at 10 days post mechanical inoculation (fig 3.22). Inoculated leaves were not possible to measure due to their dessication and death 2-3 days post inoculation (DPI).

The Col-0 ecotype of *A.thaliana* show very limited symptoms for CMV-Y infections (Takahashi, 1994; Yoshii et al, 1998). Only slight chlorosis and a stunting of growth were seen in a few plants at 10 DPI, however plants left to grow longer showing a delayed inflorescence phenotype. No difference in symptom severity was seen between Col-0 WT and HAM1 mutant plants.

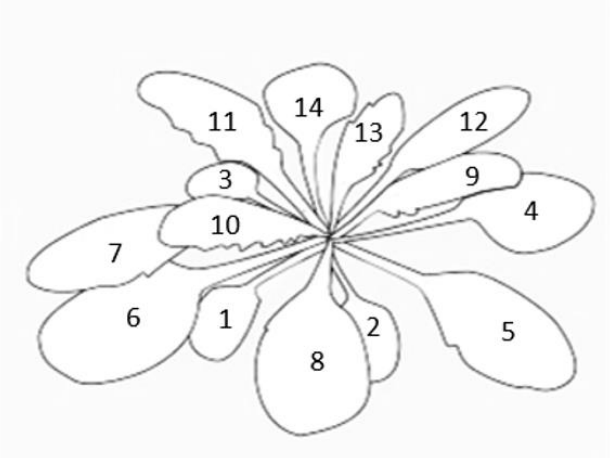

Figure 3.22 Diagram of *Arabidopsis* leaves used for inoculation and sampling At 28 days after sowing, when plants had 10 to 12 rosette leaves, they were inoculated with CMV on the fifth leaf The aerial parts of inoculated individuals were harvested separately at 10 days post inoculation. Separated by older leaves ,the fourth to sixth leaves, and younger leaves, the seventh to twelfth leaf, for analysis of viral titre by DAS-ELISA. Adapted from https://plantae.org/communicating-effectively-with-graphics/

To analyse viral titres in sampled leaf material, the CMV antibody provided in the complete ELISA kit was diluted 1:1000 in coating buffer and 200 µl added to each well of a micro-titre plate and incubated at 4° overnight for antibody absorbance to well walls. A standardized leaf punch was taken with the lid of an Eppendorf from individual leaf tissue sampled at 10 DPI. The leaf punches were homogenised using sterile ball bearings in a tissue-lyser before vortexing with 750 µl of the extraction buffer. 200 µl of leaf extraction mixture was then applied to the plate wells in triplicate, so readings for each sample could be averaged, and again incubated overnight. 200 µl of alkaline phosphatase conjugated anti-CMV antibody (1:1000 dilution in conjugate buffer) was added to each well and the plate incubated for 5 hours at 30°C. 200 µl of para-nitrophenylphosphate (1mg in 1ml substrate buffer) was added to each well and incubated at room temperature in the dark for 1 hour. The plate absorbance values were read at 415nm using a plate reader (Biorad Microplate Reader). Replicate absorbance values for each leaf sample were averaged and normalised by subtracting a mean absorbance value of a buffer control. Leaf sample mean absorbance values were then expressed as a percentage of the CMV positive control absorbance (fig 3.23). For statistical analysis, percentage data which may not follow a binomial distribution was variance-stabilised by Arcsine transformation [formula= Arcsine(sqrt(value/100))].

It was found that total WT leaves (M= 13.34 SD= 6.24) sampled had a significantly higher viral tighter than mutant leaves (M= 4.79 SD= 2.81); t (30) = 5.25, p=<0.0001. Furthermore, one-way ANOVA revealed a significant difference in viral titre between leaf age groups from WT and mutant leaves; F (3) = 20.15, p<0.0001. With Tukey's multiple comparison showing that WT 'new' leaves from upper

rosette parts had significantly higher viral levels than leaves from both lower WT leaves and both mutant leaf groups, p<0.002. Whilst lower WT leaves, older than the 5th inoculated leaf, had a significantly higher viral titre than their counterpart lower leaves from mutant plant, p=0.012 (Fig 3.23). These results suggest that CMV fails to maintain high systemic viral titres in the absence of plant host NTPase activity proteins.

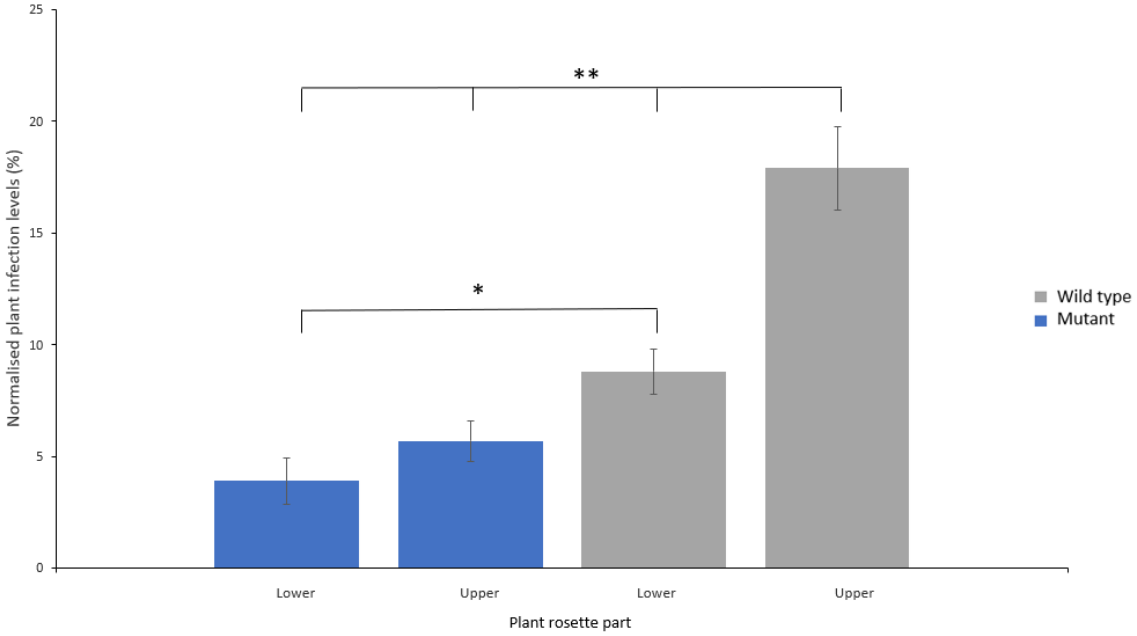

Figure 3.23 CMV infection levels in WT Col-0 (grey) and Ham1 mutant (blue) *Arabidopsis*, for both upper and lower rosette leaves. Viral titres are expressed as a normalised percentage of the positive control from ELISA absorbance data. ** indicates a significant difference where p<0.01, * indicates p= 0.012.

3.11 Expression of *Arabidopsis thaliana* HAM1 expression during viral infection

One hypothesis for the presence of U/CBSV Ham1 is that it is a compensation for the down regulation of host nucleotide pyrophosphatase expression under infection. The previously discussed ELISA data suggests +ssRNA viruses require expression of HAM1 to maintain high infection levels in host leaf tissue. A virus would therefore require its own genomic copy of a Ham1/maf protein if its host was capable of regulating its nucleotide pyrophosphatase expression as part of its plant defence response. To explore if HAM1 expression was regulated as part of host defence under viral infection, the expression of AT4G13720 was measured by real-time quantitative polymerase chain reaction (RT-qPCR) from RNA samples isolated from Columbia-0 ecotype plants inoculated with CMV.

Plants were again grown in 3:1 sand:soil at 25°C under 10:14 hours light:dark cycle to produce large rosette leaves for ease of inoculation. Plants at 4 weeks were inoculated with 1ml of homogenised CMV inoculum by abrasion with carborundum, as described previously, and 4 plants were mock inoculated for used as negative controls. 3 infected plants and a negative control plant had 4 leaves

sampled at time points 0, 3, 6, 9, and 12 DPI, and the samples frozen at -80 before extraction. Leaf samples were ground by vortexing with sterile steel ball-bearings and RNA extracted using TRIzol reagent (Thermo Fisher) following Rio et al (2010). The homogenised leaf samples were mixed with 1ml of TRIzol and incubated at room temperature for 5 minutes for protein denaturation and complete dissociation of nucleoprotein complexes. 0.2ml of chloroform was then added to the samples which were incubated for a further 3 minutes and centrifuged at 12,000*g* for 15 minutes at 4°C. This causes separation of RNA into the aqueous phase which was transferred to a fresh tube with 0.5ml isopropanol and precipitated by incubation on ice for 10 minutes. This was followed by further centrifugation at 12,000*g* for 10 minutes to pellet the RNA. The supernatant was discarded and the resulting pellet washed with 75% ethanol. After briefly centrifuging, the ethanol was discarded and the pellet left to air dry before resuspension in 50µl diethylpyrocarbonate- treated (DEPC) water (Thermo Fisher) to prevent RNase activity.

Samples were treated with DNaseI for the removal of DNA contamination from RNA preparations. Firstly, sample concentration was measured using a Nanodrop spectrophotometer, and volume equating to 1µg total RNA added to an RNase-free sterile tube. The RNA was then incubated with 1µl DNaseI enzyme and 1µl of 10x reaction buffer for 1 hour at 37°C, the reaction was stopped by addition of 1µl EDTA and 10 minute incubation at 65°C. The RNA sample quality was then checked for integrity and correct size distribution by electrophoresis on a 1% agarose gel (fig 3.24). RNA preperations were then used as a template for reverse transcriptase using the First Strand cDNA synthesis kit (Thermo Scientific). For each sample, 1 µg of total RNA was incubated with 1µl oligo (dT_{18}) primer for cDNA synthesis, 4 µl of 5x reaction buffer, 1 µl RiboLock RNase inhibitor, 2 µl of 10mM dNTP, and 2 µl of M-MuLV Reverse Transcriptase for 1 hour at 37°C.

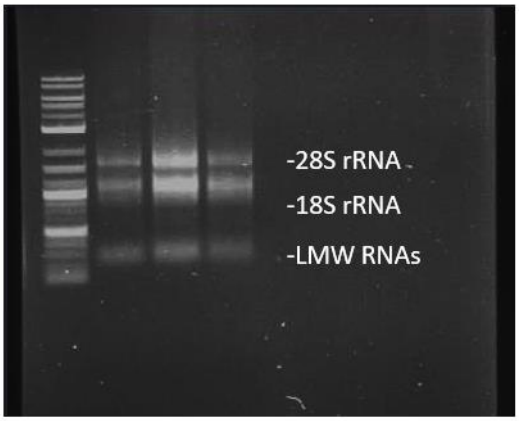

Figure 3.24 Example RNA extraction from 3 WT Col-0 plants on 1% agarose gel to check RNA quality and integrity. Bands visible are 28S and 18S rRNA and low molecular weight RNAs.

Reverse transcription of samples was checked by PCR amplification of the first strand cDNA. Primers for qPCR were designed and ordered using Integrated DNA Technologies OligoTool. For qPCR a reference gene is required for the normalisation of transcipt levels so any variation in the expression of HAM1 can accurately be determined. Lily et al (2011), validated the stability of 12 *Arabidopsis* genes in Col-0 ecotypes when infected with plant viruses such as CMV. Of traditional reference genes, protodermal factor 2 (PDF2), enoding a regulatory subunit of serine/threonine protein phosphotase 2A showed good stability across the experimental conditions. PDF2 was therefore chosen as the reference gene for normalisation of qPCR-derived HAM1 expression in CMV infected *Arabidopsis*.

Primer pairs for both HAM1 and PDF2 (table 5) were tested on *Arabidopsis* cDNA preperations and run on agarose gels to see transcipt amplification. Target amplification was obtained with all primers tested, and primer pair 1 chosen for HAM1 and primer pair 2 chosen for PDF2 for subsequent qPCR due to their similar Tm of 55°C and amplification size of 300bp (fig 3.25).

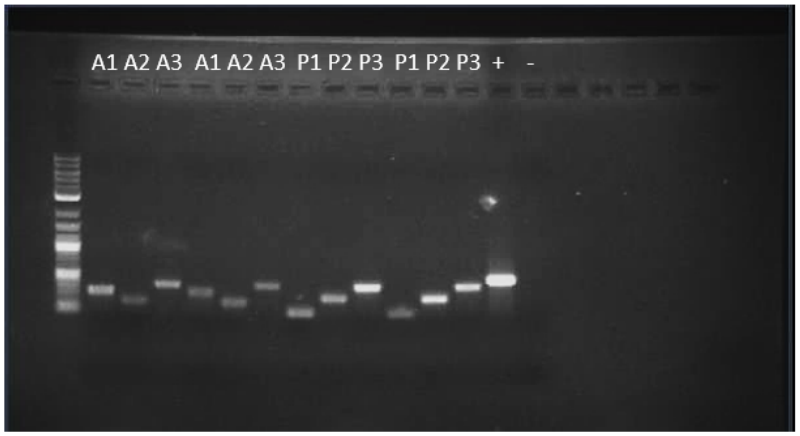

Figure 3.25 Amplification of WT *Arabidopsis* cDNA to test primer amplification of AT4g13720 and PDF2 transcripts. 3 primer pairs for amplification of AT4g13720 transcripts (A 1-3) and 3 primer pairs for PDF2 transcripts (P 1-3) were used alongside a GADPH positive control producing a 496bp PCR product (+).

Arabidopsis leaf samples were then amplified via conventional TaqPCR using PDF2 qPCR primer pair 1 to check the quality of cDNA obtained. Leaf samples from healthy and infected plants taken at time points at 0, 6, 9, and 12 days post inoculation mostly showed the amplification of correct HAM1 transcipt, although some samples showed none, or very low amplification of cDNA (figs 3.26, 27, 28). For example, leaf samples taken at 3 DPI (fig 3.27, T_3) showed very low quality cDNA amplification and possible DNA contamination due to the amplification of multiple bands. Analysis of 12 T=3 samples by gel electrophoresis showed the presence of degraded RNA which would have likely caused poor

conversion to cDNA (fig 3.29). Amplification by qPCR is more sensitive than Taq PCR, therefore samples from 0, 6, 9, and 12 DPI were used for expression analysis, however samples from day 3 were discarded due to the poor quality of cDNA obtained.

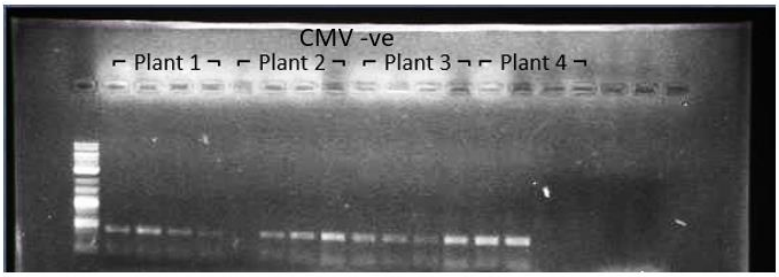

Figure 3.26 Amplification of WT uninoculated Col-0 *Arabidopsis* cDNA with PDF2 primer pair 1. 4 leaves were sampled from a healthy plant at each time point and the mean expression used to normalise expression in inoculated leaves relative to CMV negative plants.

Figure 3.27 Amplification of infected Col-0 *Arabidopsis* cDNA with PDF2 primer pair 1 by PCR to check cDNA quality. 4 leaf samples were collected from 3 infected plants at each time point: 0, 3, and 6 days post inoculation (time points labelled T_0, T_3, T_6).

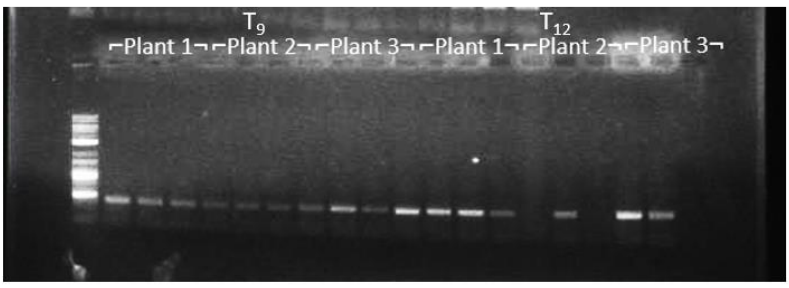

Figure 3.28 Amplification of infected Col-0 *Arabidopsis* cDNA with PDF2 primer pair 1 by PCR to check cDNA quality. 4 leaf samples were collected from 3 infected plants at each time points 9 and 12 days post inoculation (time points labelled T_9 and T_{12}).

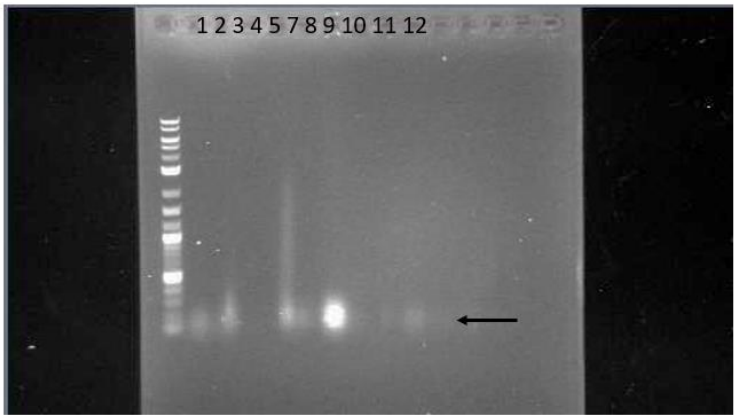

Figure 3.29 Gel electrophoresis of RNA from 12 leaf samples collected at T3 post inoculation. The samples RNA are considerably degraded with no ribosomal RNA bands visible.

To measure transcript amplification during qPCR, the fluorescent Maxima® SYBR Green/ROX (Thermo Fisher Scientific) was used following the manufacturer's instructions. Sample PCR reactions were prepared with 2µl cDNA in duplicate. Cycles for qPCR were performed in the thermocycler Stratagene MX3005, using the following conditions: initial denaturation at 95 °C for 10 minutes, 40 cycles of: denaturation at 95°C for 30 seconds, annealing at 55 °C for 30 seconds and extension at 72 °C for 30 seconds, with data collection for fluoresence at the end of each extension. This was followed by a denaturation at 95 °C for 1 minute, with a continuous data collection from 55 °C to 95 °C, for melting curve analysis (fig 3.30). The relative expression of HAM1 over the course of infection normalised to PDF2 and relative to un-inoculated Col-0 at time 0 was calculated using the 2-ΔΔCt method (Livak & Schmittgen, 2001).

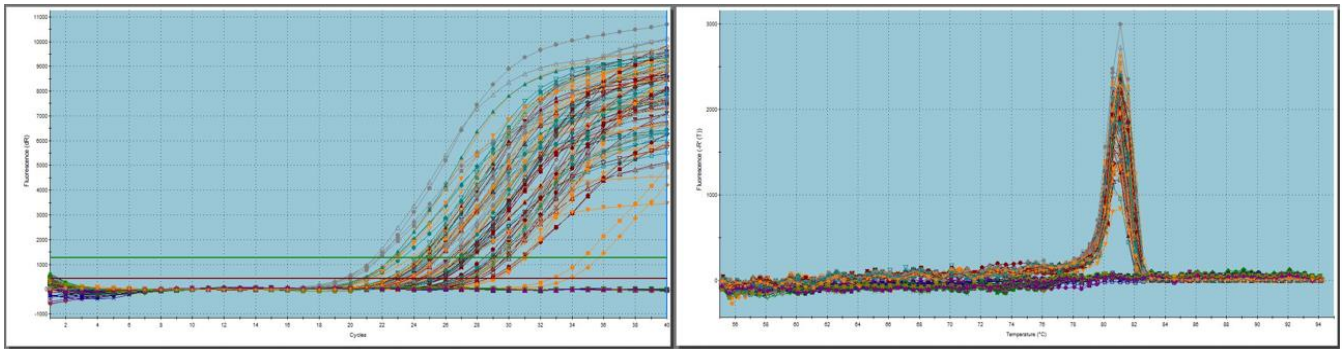

Figure 3.30 Example qPCR amplification plot (left) and dissociation curve (right) from HAM1 and PDF2 transcript amplification of cDNA from leaf samples taken at T0 and T6. The dissociation curve is used to check for amplification of a single product. The Amplification plot shows product accumulation by the increase in fluorescent signal over the 40 cycles for each sample, cycle threshold values (green line) are used to calculate the amount of transcript present in the initial sample and therefore the relative gene expression.

The Ct values calculated for HAM1 and PDF2 from replicate PCRs from the same sample were averaged between replicates. The mean Ct values were then used to calculate change in fold expression of HAM1 using the equation $2^{-\Delta\Delta Ct}$ where $\Delta\Delta CT = (Ct_{HAM1} - Ct_{PDF2})_{Tx} - (Ct_{HAM1} - Ct_{PDF2})_{T0}$. The relative fold change in the expression of HAM1 was then averaged across leaf samples at each time point and standard deviation calculated. The mean fold change of HAM1 in infected WT *Arabidopsis* over the 12 days measured suggests no change in expression with the equation for an averaged trendline showing a gradient of -0.0023 change in expression (fig 3.29). The large variance in data violated the conditions for homogenity of variance (Levenes test p=0.044). Analysis by one-way ANOVA assuming un-equal variance showed no singificant difference in HAM1 expression between the time points measured, (p=0.124), suggesting HAM1 expression is not regulated in response to viral infection in *Arabidopsis*.

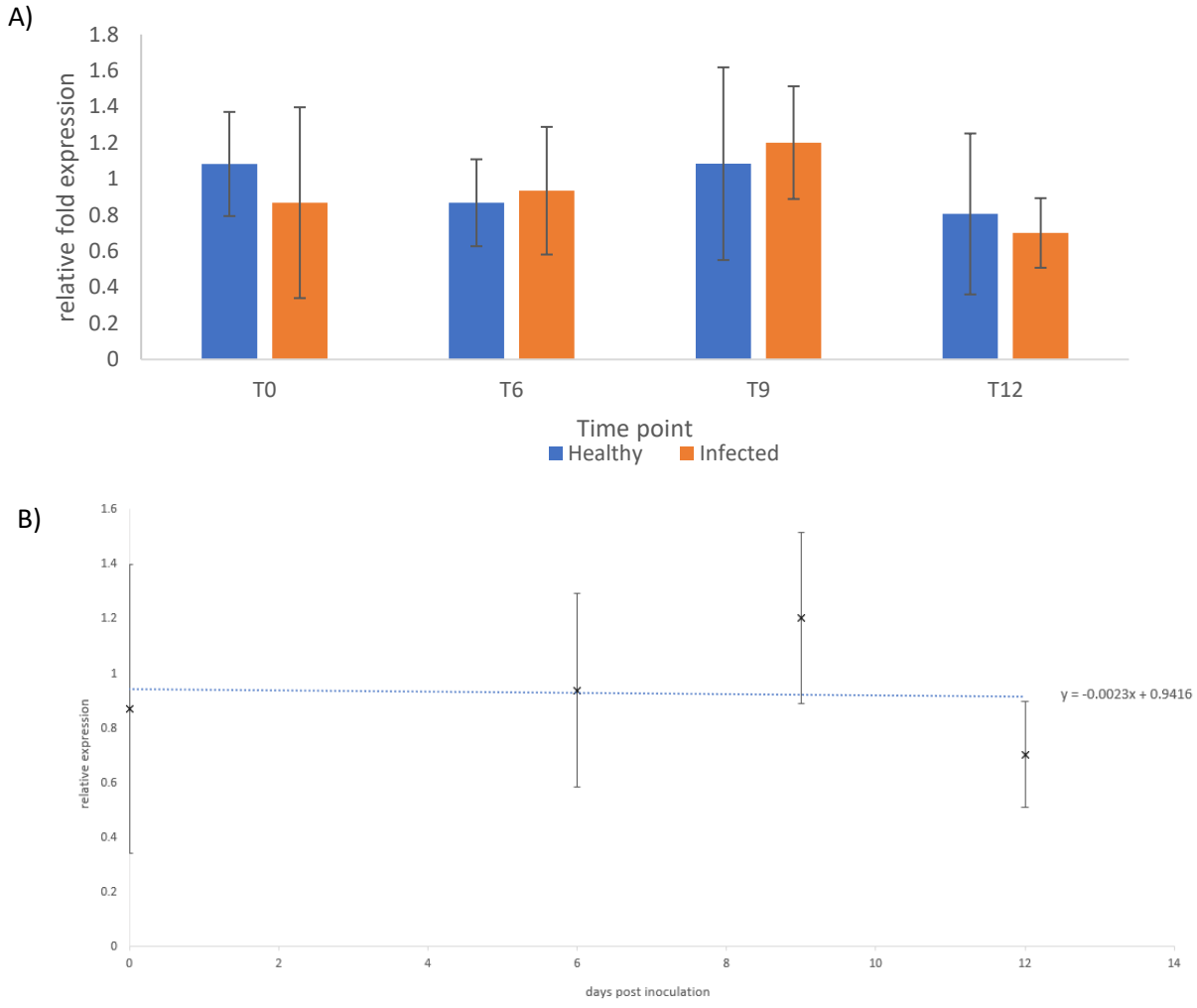

Figure 3.31 A) Mean *Arabidopsis* ITPase gene expression relative to PDF2 reference gene (+/- S.D) over the course of CMV infection in healthy and infected plants . B) Mean *Arabidopsis* ITPase expression relative to PDF2 transcript levels (+/- S.D), normalised to uninoculated plants at T0. An averaged trendline shows a change in expression of -0.0023 over 12 days p.i, suggesting no change in plant HAM1 transcripts during viral infection

Chapter 4
Discussion

In this thesis the hypothesis that CBSV and UCBSV Ham1 proteins perform the role of ITPase proteins during CBSD infections was studied. Viral genomes are prone to high mutation rates due to encoding low fidelity RNA polymerases which can cause an accumulation of mutations during replication which are not repaired by cellular mismatch repair pathways (Anderson et al, 2004). RNA viruses possess *in vivo* mutation rates in the range of 10^{-6} to 10^{-4} substitutions per nucleotide copied (Anderson et al, 2004), this means RNA viruses are susceptible to 'error catastrophe' during infection (Summers and Litwin, 2006). When the mutation rate exceeds a threshold, the faithful replication of the viral population breaks down, leading to randomly distributed mutant species. This increase in error rate is associated with a decrease in virus infectivity which can lead to population extinction (Cases-Gonzalez et al, 2008). CBSV infections have been shown to cause an early senescence phenotype in Cassava (Schippers et al, 2008) characterised by high oxidative stress which has been shown to increase the mutation rate in plants, which require repair mechanisms to protect the cell from oxidative damage (Boyko et al, 2006). Therefore, Ham1 pyrophosphatase proteins encoded by U/CBSV may play an important role in the reduction of mutagenesis for the long-term stability of virus RNA genomes in hosts.

The CBSV Ham1 protein has not been shown to interact with any other CBSV genomic protein or possess silencing activity in assays so far (Pablo-Rodriguez, unpublished; Mbanzibwa et al, 2009). Its homology to cellular ITPase proteins and the presence of conserved SHR motifs from this protein family suggests Ham1 may protect viral genomic integrity by preventing incorporation of non-canonical nucleotides such as ITP and XTP, which are found within the naturally occurring nucleotide pool. Ham1 could thereby confer an advantage to the virus under stressful conditions by acting as a repair mechanism to prevent deleterious mutations. Other viral proteins, such as AlkB, have previously been shown to act in a similar capacity to reduce the viral mutation rate. The AlkB protein, which has been found in many RNA viruses of the *Flexiviridae* family, functions by repairing alkylation damage to RNA genomes by oxidative demethylation, thereby protecting the virus from mutagenesis (Van Den Born et al, 2008). Homologous Ham1 proteins in *S. cerevisiae* and *E. coli* have also been shown to help protect the cell from purine base analogue mutagens such as 6-N-hydroxylaminopurine (Noskov et al, 1996; Pavlov et al, 1996). Yeast, murine, and human ITPase proteins have all been shown to have 100-fold preferences for the noncanonical purine nucleoside triphosphates XTP and ITP over other canonical nucleosides such as GTP (Burgis and Cunningham, 2007). This strongly suggests a conserved role in preventing mutagenesis from modified purine nucleotides for this protein family. Noncanonical purine nucleotides such as ITP and XTP are thought to arise endogenously through defects in metabolism during periods of intense metabolic or oxidative stress, where they readily incorporate into RNA and DNA (Shapiro et al, 1968, Pang et al, 2012). These stressful conditions have

been reported to be present in leaves of cassava during infection (Schippers et al, 2008), suggesting U/CBSV have acquired a homologous ITPase family protein from a cellular origin to protect viral genomic integrity by preventing lethal mutations from endogenous noncanonical purines which arise from the oxidative infection environment.

To test this hypothesis, both CBSV and UCBSV Ham1 proteins were expressed and their nucleotide triphosphate pyrophosphohydrolase activity tested against a range of nucleotides in an enzymatic assay. Despite poor expression in BL21 *E. coli,* both virus Ham1 proteins were successfully purified using nickel affinity chromatography. However, UCBSV Ham1 was purified at a lower culture density than CBSV Ham1, which eluted at a lower concentration with multiple contaminating proteins when expressed at the same culture O.D. This suggests UCBSV Ham1 was more easily expressed by BL21 *E. coli* cells, or more able to outcompete non-specific proteins for nickel binding sites, possibly due to its different polypeptide sequence (Kaur et al, 2018).

Analysis of CBSV Ham1 and CBSV mutant Ham1 *E. coli* cultures showed normal growth rates until protein induction with IPTG. Ham1 protein expression caused a large reduction in culture growth compared to the control culture containing an empty POPINF vector. This suggests CBSV WT and mutant Ham1 put a large metabolic strain on *E. coli* cells possibly due to toxic protein effects. Successful CBSV WT Ham1 purification was achieved using high culture density to ensure greater initial *E. coli* biomass and a low expression temperature of 15°C to increase the protein yield obtained. Despite this, the mutant Ham1 CBSV protein containing an SHR→SAA mutation in a conserved functional motif was not able to be purified, with no expression evident in culture fractions or as an insoluble cell deposit. The poor expression of the mutant protein suggests *E. coli* cell cultures expressing the mutant protein incur a greater metabolic strain than WT protein expression, potentially due to toxic effects from the mutation (Dumon-Seignovert et al, 2004). An alternate expression vector may allow for better expression in BL21 *E. coli* cells, such as targeted periplasmic protein expression or fusion with a highly expressed protein partner (Kaur et al, 2018).

The CBSV and UCBSV Ham1 protein activities were tested in a nucleotide assay using the PiColormix detection system to measure phosphate release from Ham1 pyrophosphatase activity. An inorganic yeast phosphatase was added in excess to the incubation mixture to cleave pyrophosphate groups released from Ham1 pyrophosphatase activity to two orthophosphates, thereby allowing detection of Ham1 activity in a colorimetric assay. The inorganic phosphatase was shown to increase the assays detection of Ham1 activity without contributing to nucleotide triphosphate cleavage. The inert protein BSA was used as a control for nucleotide phosphate release from non-specific protein effects and random background degradation. As predicted by homology to cellular ITPase proteins, both CBSV

and UCBSV had the highest activity with dITP and XTP non-canonical nucleotides, which were cleaved significantly more than TTP and dATP canonical nucleotides with both proteins. This is the first proof of such activity in a virus protein, as to date it has only been inferred from sequence similarity with cellular proteins.

Similar higher activity with ITP and XTP nucleotides have been reported in other Maf/Ham1 proteins such as *Homo sapiens* and *Methanococcus jannaschii* ITPase proteins, the latter of which displays little difference between deoxy- and ribonucleotide substrates (Hwang et al, 1999; Lin et al, 2001). This supports the hypothesis that hydrolysis of mutagen causing non-canonical nucleotides is one of the functions of Ham1 proteins during viral infection and could potentially act to protect the virus RNA genome.

Despite their low sequence similarity (48%), no significant difference was found between CBSV and UCBSV pyrophosphatase activity on the nucleotides tested. The Ham1 sequence of EuRSV represents an even more divergent sequence with only 34% similarity (Knierim et al, 2017), however the few conserved residues compromising the predicted nucleotide binding site found across U/CBSV, EuRV, and Eukaryotic Ham1 proteins (Mbanzibwa et al, 2009) and the results of the U/CBSV assay suggests a common function as NTP pyrophosphatases. Furthermore, U/CBSV and EuRSV infect species in the *Euphorbiaceae* plant family, suggesting Ham1 is an adaption to sustained infections in this plant family.

However, U/CBSV Ham1 proteins pyrophosphatase activity with ITP and XTP has only been demonstrated *in vitro*, viral Ham1 may target other nucleotide substrates during an *in vivo* infection. For example, both U/CBSV Ham1 proteins exhibited higher, though not significant, activity with GTP than other canonical nucleotides tested which is uncharacteristic of other ITPase family protein activities tested (Lin et al, 2001). U/CBSV Ham1 proteins broader substrate specificity may suggest it acts upon slightly different nucleotides during infection. One hypothesised cellular target was the m^7G cap of host mRNA transcripts which when cleaved would cause RNA degradation and reduced translation (Ramanathan et al, 2016) . For example, Pox viruses have been shown to encode mRNA de-capping enzymes contributing to host mRNA turnover (Walsh and Mohr, 2011). However, when incubated with an m^7G triphosphate cap analogue both U/CBSV Ham1 proteins showed negligible pyrophosphatase activity in comparison to the canonical and non-canonical nucleotides tested simultaneously. This suggests viral Ham1 protein do not target the 5' cap of mRNA *in vivo*. Therefore, the reason for the higher activity of U/CBSV Ham1 with GTP nucleotides compared to other homologous ITPase proteins is still unknown. This suggests viral Ham1 may have broader substrate

specificity than just non-canonical nucleotides and therefore a different, and as of present unknown, function *in vivo*.

Further evidence for the role of Ham1 proteins in protecting viral genomes came from analysis of CMV titres in *Arabidopsis* ITPase mutants. Of 266 plants screened by PCR analysis, only 15 homozygous mutants were identified. All homozygous mutants came from a few individual lines of GABI-Kat 345B08 plants, as was expected. No homozygous mutants were found in SAIL or SALK plant lines however they made up a minority of the seeds planted. The low number of mutants identified may be due to the low quality of DNA extracted for large scale analysis. More mutants may have been present but the difficulty of producing clean DNA samples may have prevented efficient identification of homozygous T-DNA insertions. When grown, ITPase homozygous mutant plants showed no noticeable hypermutable phenotype, such as reduced growth and delayed flowering, which suggests it is not a pivotal gene for *Arabidopsis* development and morphology. However, mutant *Arabidopsis* were grown under short day lengths to increase leaf rosette size for ease of inoculation. Therefore, the growth conditions may have been low-stress meaning a reduced non-canonical nucleotide pool due to reduced base-deamination by reactive species under oxidative stress (Davies et al, 2012). Similarly, mutations may have been occurring but at a low enough rate for DNA repair mechanisms to correct the damage (Boyko et al, 2006).

To explore the effects of host plant ITPase on viral infections, a serological assay was performed to measure viral titres in homozygous ITPase mutant versus WT *Arabidopsis*. It is unknown whether U/CBSV are capable of infecting *Arabidopsis,* therefore the virus CMV was chosen due to its large host range, including *Arabidopsis*, and is a ss(+)RNA virus like U/CBSV but does not encode a Ham1 protein. CMV infection levels were found to be much lower in ITPase mutant compared to WT *Arabidopsis* leaves. This suggests that without a functional nucleotide pyrophosphatase protein RNA virus infectivity is reduced, and the population potentially suffers error catastrophe due to an excessive mutation rate. Leaves of *Arabidopsis* produce a hypersensitive response associated with an increase in reactive oxygen species (Takahashi et al, 1994; Beneloujaephajri, et al 2013) which may have led to an increased mutation rate and loss of CMV infectivity in mutant plants. Plants possess cellular DNA repair mechanisms to repair damage, whilst CMV lacks genomic repair functions and would therefore incur cumulative harmful mutation affecting the populations viability. However, *Arabidopsis* infected with CMV-Y show virtually no symptoms (Takahashi et al, 1994), therefore lower viral titres did not correlate to changes in symptom severity. It would be interesting to see if host ITPase mutation leads to a reduction in disease symptoms in infections with another viral species; and to measure the effects of plant ITPase overexpression to see if this correlated to an increase in viral titres due to increased genomic stability in a dose-response relationship.

As these results suggest ITPase activity is required for RNA virus stability, plants may be under adaptive pressure to regulate the expression of ITPase when challenged with a virus. However, when the expression of ITPase was measured over 12 days in WT *Arabidopsis* infected with CMV, no change was observed. This suggests that *Arabidopsis* ITPase expression is not regulated as part of host defence, however, this may not be the case for Cassava ITPase proteins. Furthermore, Cassava ITPase proteins may have different activity, localisation, or function during infection. CMV showed lower viral titres in ITPase mutant plants possibly due to an increased rate of mutations, suggesting ITPase protein function is important for RNA virus stability. These results suggest that U/CBSV Ham1 proteins may benefit viral fitness during long term infections and may be an adaption to combat viral susceptibility to mutations or low ITPase levels in Cassava (Mbanzibwa et al, 2009; Schippers et al, 2008)

The study of mutant ITPase proteins in *E. coli* and Yeast which display increased recombination, mutagenesis, damage response, and HAP sensitivity phenotypes in mutants, suggests a role for pyrophosphatases in protecting DNA by preventing the incorporation of non-canonical purine nucleotides (Burgis and Cunningham, 2007; Noskov et al, 1996). Biochemical analysis of Human, Yeast and *Methanococcus jannaschii* Ham1 homologs suggests no difference in activity between deoxy- and ribonucleotides (Hwang et al, 1999; Burgis and Cunningham, 2007). Furthermore, dITP and ITP nucleotides have been shown to incorporate into both DNA and RNA respectively *in vitro* (Chanda et al, 1983), where they act as a GTP analogues causing transition mutations (Lin et al, 2001). Both UCBSV and CBSV Ham1 had a vigorous reaction with ITP forming a precipitate, which although it could not be measured accurately, suggests very high protein pyrophosphatase activity with the nucleotide. Both viral Ham1 proteins also recorded the highest activity with XTP in the phosphatase assay, which alongside ITP, are ribose nucleotides which may incorporate into RNA. Therefore, the results of the Ham1 nucleotide assay and CMV *Arabidopsis* ITPase mutant infections, means that the hypothesis that U/CBSV Ham1 proteins may act to protect the integrity of viral RNA genomes from mutagenic NTPs cannot be dismissed.

Further study of this unique viral proteins would require high enough expression to allow crystallography for structural comparison to other ITPase proteins and to decipher its enzymatic activity *in vivo*. For example, the ITPase protein of *M. jannaschii* has been reported as a homodimer (Hwang et al, 1999) whilst the ITPase of *E. coli* forms a different dimer, despite the homology of monomeric subunits (Zheng et al, 2005). Furthermore, Kinetic analysis of U/CBSV Ham1 protein activities with dITP, ITP, and XTP over a range of substrate concentrations would allow comparison of enzymatic activity to the values reported for human and *E. coli* ITPase proteins (Lin et al, 2001; Zheng et al, 2005). Allowing further insight into the homology and cellular origins of U/CBSV Ham1.

Furthermore, Gall et al, (2013) reported 7 amino acid residues, including the SHR motif, which were thought to interact directly with the nucleotide substrate as essential to human ITPase for pyrophosphatase activity and protection against non-canonical nucleotides. A CBSV Ham1 protein containing a mutated SHR motif was unable to be expressed and purified (fig 3.8, 3.9). However, expression and biochemical analysis of viral Ham1 with mutations at specific predicted conserved residues would allow identification of amino acid positions involved in nucleotide binding and comparison to essential residues and mechanisms for pyrophosphatase activity in other homologous cellular ITPase proteins.

RNA viruses typically encode a small number of proteins with multiple functions during viral transmission, infection, and replication (Mäkinen et al, 2014). Therefore, U/CBSV Ham1 may still have further intracellular roles which are unknown at present. Studying the localisation of Ham1 in the host cell during infection may help to ascertain further potential functions and interactions with host proteins that viral Ham1 may have. Therefore, to fully investigate the fitness benefits of viral incorporation of Ham1, infections must be followed in Cassava *in vivo*. The use of stable infectious clones for reproducible cassava infectious may help to identify the specific roles of viral Ham1 *in vivo* and could aid the screening of CBSD resistance germplasm.

Furthermore, whether Cassava regulates ITPase protein expression may be key to the benefits conferred to U/CBSV by encoding a Ham1 protein. The assay of CMV titres in ITPase knockout *Arabidopsis* showed that an absence of ITPase expression led to reduced viral titres, however *Arabidopsis* does not appear to regulate ITPase expression in response to viral infection. As reduced Ham1 activity limits viral titre, it may be an evolutionary adaption in some plants to support low ITPase protein expression. Whilst annuals like *Arabidopsis* might not do this, it may be more beneficial in long-lived plants such as cassava. Possible low or adaptive ITPase expression in cassava may be an evolutionary driver for U/CBSV to possess a Ham1 gene to maintain viral stability under long term infection. Therefore, a plant's ITPase gene may be a target for antiviral therapy, where inactivation of the plant's ITPase expression by methods such as RNAi or CRISPR-Cas9 could impart viral resistance to the host (Patil et al, 2011). These technologies have already been used to engineer potyviral resistance through transgenic expression of viral gene hairpin constructs giving resistance to homologous viruses in the case of RNAi (Leibman et al, 2011), and the targeting of host alleles required for potyviral translation with CRISPR-Cas9 (Pyott et al, 2016). Furthermore, the effectiveness of RNAi against CBSD by the transgenic expression of UCBSV and CBSV CP inverted repeats has already been demonstrated in cassava and tobacco (Wagaba et al, 2016; Patil et al, 2011). Cassava is an excellent crop for the development of transgenic resistance techniques due to the durability of resistance over vegetative

crop cycles. Therefore, the targeting of host or virus Ham1 genes may provide another avenue for the engineering of CBSD resistance in cassava crops.

Chapter 5
Bibliography

Abarshi, M., Mohammed, I., Wasswa, P., Hillocks, R., Holt, J., Legg, J., Seal, S. and Maruthi, M. (2010). Optimization of diagnostic RT-PCR protocols and sampling procedures for the reliable and cost-effective detection of Cassava brown streak virus. *Journal of Virological Methods*, 163(2), pp. 353-359.

Alexandrov, N. N., Brover, V.V., Freiden, S., Troukhan, M.E., Tatarinova, T.T., Zhang, H., Swaller, T.J., Lu, Y.P., Bouck, J,. Flavell, R.B., Feldmann, K.A. (2009). Insights into corn genes derived from large-scale cDNA sequencing. *Plant Molecular Biology*, 69(1–2), pp. 179–194.

Alicai, T., Ndunguru, J., Sseruwagi, P., Tairo, F., Okao-Okuja, G., Nanvubya, R., Kiiza, L., Kubatko, L., Kehoe, M. and Boykin, L., (2016). Characterization by next generation sequencing reveals the molecular mechanisms driving the faster evolutionary rate of Cassava brown streak virus compared with Ugandan cassava brown streak virus. *Scientific Reports,* 6(1), pp.36164.

Anderson, J., Daifuku, R. and Loeb, L. (2004). Viral Error Catastrophe by Mutagenic Nucleosides. *Annual Review of Microbiology*, 58(1), pp.183-205.

Arazi, T., S. H. Slutsky, Y. M. Shiboleth, Y. Wang, M. Rubinstein, S. Barak, J. Yang, and A. Gal-On. 2001. Engineering zucchini yellow mosaic potyvirus as a non-pathogenic vector for expression of heterologous proteins in cucurbits. *J. Biotechnol*. (87), pp.67–82.

Beneloujaephajri, E., Costa, A., L'Haridon, F., Metraux, J.P., Binda, M. (2013). Production of reactive oxygen species and wound-induced resistance in Arabidopsis thaliana against Botrytis cinerea are preceded and depend on a burst of calcium., *BMC plant biology*, 13, pp. 160.

Berendzen, K. Searle, I., Ravenscroft, D., Koncz, C., Batschauer, A., Coupland, G., Somssich, I.E., Ulker, B. (2005). A rapid and versatile combined DNA/RNA extraction protocol and its application to the analysis of a novel DNA marker set polymorphic between Arabidopsis thaliana ecotypes Col-0 and Landsberg erecta., *Plant methods*, 1(1), pp. 4.

Berrow, N. S. Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D.I., Owens, R. J. (2007). A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic acids research*, 35(6), pp. e45.

Bock, K. (1994). Studies on Cassava brown streak virus disease in Kenya. *Tropical Science*, 34(1), pp.134–145.

Boyko, A., F. Zemp, J. Filkowski, and I. Kovalchuk. 2006. Double-strand break repair in plants is developmentally regulated. *Plant Physiol*, (141), pp. 488– 497

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'Amore, R., Allen, A. M., … Hall, N. (2012). Analysis of the bread wheat genome using whole genome shotgun sequencing. *Nature*, 491(7426), pp. 705–710

Burgis, N.E. and Cunningham, R.P. (2007). Substrate specificity of RdgB protein, a deoxyribonucleoside triphosphate pyrophosphohydrolase. *The Journal of biological chemistry*, 282(6), pp. 3531–8.

Cases-González, C., Arribas, M., Domingo, E., Lazero, E. (2008) Beneficial Effects of Population Bottlenecks in an RNA Virus Evolving at Increased Error Rate, *Journal of Molecular Biology*, 384(5), pp. 1120–1129

Ceballos, H. Iglesias, C. Pérez, J, Dixon, A. (2004), Cassava breeding: opportunities and challenges, *Plant Molecular Biology*, 56(4), pp. 503–516.

Chanda, P.K., Roy, J. and Banerjee, A. K. (1983), In vitro synthesis of genome length complementary RNA of vesicular stomatitis virus in the presence of inosine 5'-triphosphate, *Virology*, 129(1), pp. 225–229.

Chung B. Y., Miller W. A., Atkins J. F., Firth A. E. (2008). An overlapping essential gene in the *Potyviridae*. *Proc Natl Acad Sci USA,* (105), pp. 5897–5902

Crotty, S., Cameron, C. and Andino, R. (2001). RNA virus error catastrophe: Direct molecular test by using ribavirin. *Proceedings of the National Academy of Sciences*, 98(12), pp.6895-6900.

Davies, O., Mendes, P., Smallbone, K., Malys, N. (2012), Characterisation of multiple substrate-specific (d)ITP/(d)XTPase and modelling of deaminated purine nucleotide metabolism., *BMB reports*, 45(4), pp. 259–64.

Ding SW and Voinnet O. (2007). Antiviral immunity directed by small RNAs. *Cell*, 130(3), pp. 413–426

Drake, J. W. and Holland, J. J. (1999), *Mutation rates among RNA viruses*. *Proc. Natl Acad. Sci. USA*, 96, pp. 13910-3

Dumon-Seignovert, L., Cariot, G. and Vuillard, L. (2004). The toxicity of recombinant proteins in Escherichia coli: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*, 37(1), pp.203-206.

Edwards, K., Johnstone, C. and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Research*, 19(6), pp.1349-1349.

FAO 2013 *Food and Agriculture Organisation Statistics Database*. http://faostat3.fao.org/faostat-gateway/go/to/home/E.

Fauquet C. M., Mayo M. A., Maniloff J., Desselberger U., Ball L. A. (2005). *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. London: Elsevier academic press

Fermont, A.M., van Asten, P. J., Tittonell, P., van Wijk, M. T., Giller, K. E. (2009). Closing the cassava yield gap: An analysis from smallholder farms in East Africa. *Field Crops Research,* 112(1), pp.24–36

Gall, A., Gall, A., Moore, A., Aune, M., Heid, S., Mori, A. and Burgis, N. (2013). Analysis of human ITPase nucleobase specificity by site-directed mutagenesis. *Biochimie*, 95(9), pp.1711-1721.

Galperin, M. Y., Moroz, O.V., Wilson, K.S., Murzin, A.G. (2006), House cleaning, a part of good housekeeping, *Molecular Microbiology*, 59(1), pp. 5–19.

Grangeon, R., Jiang, J., Wan, J., Agbeci, M., Zheng, H., & Laliberté, J. F. (2013). 6K2-induced vesicles can move cell to cell during turnip mosaic virus infection. *Frontiers in microbiology*, 4, pp. 351

Hannig, G. and Makrides, S. C. (1998), Strategies for optimizing heterologous protein expression in Escherichia coli, *Trends in Biotechnology*, 16(2), pp. 54–60.

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

Hillocks R.J. & Thresh, J.M. eds. (2002). Cassava: biology, production and utilization, Wallingford: CABI.

Hillocks, R.J., Raya, M. and Thresh, J.M. (1996) The association between root necrosis and above-ground symptoms of brown streak virus infection of cassava in southern Tanzania. *International Journal of Pest Managemen*t, 42, pp. 285–289

Hwang, K.Y., Chung, J.H., Kim, S.H., Han, Y.S., Cho, Y. (1999) Structure-based identification of a novel NTPase from *Methanococcus jannaschii*, *Nat Struct Biol*, 6(7), pp. 691-696.

IITA (2014). IITA Bulletin, Issue 2215. Nigeria: International Institute of Tropical Agriculture

Ishikawa, M., Obata, F., Kumagai, T., Ohno, T. (1991) Isolation of mutants of Arabidopsis thaliana in which accumulation of tobacco mosaic virus coat protein is reduced to low levels, *Molecular & General Genetics*, 230(1–2), pp. 33–38

Jennings D. L. (1960). Observations on virus disease of cassava in resistant and susceptible varieties. II. Brown streak disease. *Empire J Exp Agric,* 28, pp. 261–269

Jeremiah, S.C., Ndyetabula, I.L., Mkamilo, G.S., Haji S., Muhanna, M.M., Chuwa, C., Kasele, S., Bouwmeester, H., Ijumba, J.N., Legg J.P. (2015). The Dynamics and Environmental Influence on Interactions Between Cassava Brown Streak Disease and the Whitefly, *Bemisia tabaci*. *Phytopathology.* 105(5), pp. 646-655

Kamiya, H. (2003) Mutagenic potentials of damaged nucleic acids produced by reactive oxygen/nitrogen species: approaches using synthetic oligonucleotides and nucleotides: survey and summary. Nucleic Acids Res 31: 517– 531.

Kamiya, H. (2004) Mutagenicities of 8-hydroxyguanine and 2-hydroxyadenine produced by reactive oxygen species. Biol Pharm Bull 27: 475–479.

Kamp C, Wilke CO, Adami C, Bornholdt S. 2003. Viral evolution under the pressure of an adaptive immune system: optimal mutation rates for viral escape. *Complexity,* 8, p 28–33

Kasschau, K. D. and Carrington, J. C. (1998) A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing., *Cell*, 95(4), pp. 461–70

Kaur, J., Kumar, A. and Kaur, J. (2018). Strategies for optimization of heterologous protein expression in *E. coli*: Roadblocks and reinforcements. *International Journal of Biological Macromolecules*, 106, pp. 803-822.

Kaweesi T., Kawuki R., Kyaligonza V., Baguma Y., Tusiime G., Ferguson M.E. (2014). Field evaluation of selected cassava genotypes for cassava brown streak disease based on symptom expression and virus load. *Virol. J, 11*(1), pp. 216.

Knierim, D., Menzel, W. and Winter, S. (2017) Analysis of the complete genome sequence of euphorbia ringspot virus, an atypical member of the genus Potyvirus, *Archives of Virology*, 162(1), pp. 291–293

Legg, J. P. (1999). Emergence, spread and strategies for controlling the pandemic of cassava mosaic virus disease in east and central Africa. *Crop Protection*, 18(10), pp.627-637.

Legg, J. P., Jeremiah, S.C., Obiero, H.M., Maruthi, M.N, Ndyetabula, I., Okao, O.G, Bouwmeester, H., Bigirimana, S., Tata-Hangy, W., Gashaka, G., Mkamilo, G., Alicai, T., Kumar, P. (2011) Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa, *Virus Research*, 159(2), pp. 161–170

Leibman, D., Wolf, D., Saharan, V., Zelcer, A., Arazi, T., Yoel, S., Gaba, V. and Gal-On, A. (2011). A High Level of Transgenic Viral Small RNA Is Associated with Broad Potyvirus Resistance in Cucurbits. *Molecular Plant-Microbe Interactions*, 24(10), pp. 1220-1238.

Li XH, Valdez P, Olvera RE, Carrington JC. (1997), Functions of the tobacco etch virus RNA polymerase (NIb): subcellular transport and protein-protein interaction with VPg/proteinase (NIa). *Journal of Virology*. 71(2), pp. 1598-1607.

Lilly, S. T., Drummond, R.S, Pearson, M.N., MacDiarmid, R.M. (2011) Identification and Validation of Reference Genes for Normalization of Transcripts from Virus-Infected Arabidopsis thaliana, *Molecular Plant-Microbe Interactions MPMI*, 24(3), pp. 294–304.

Lin, S., McLennan, A.G., Ying. K., Wang, Z., Gu, S., Jin, H., Wu, C., Liu, W., Yuan, Y., Tang, R., Xie, Y., Mao Y. (2001), Cloning, expression, and characterization of a human inosine triphosphate pyrophosphatase encoded by the itpa gene., *The Journal of biological chemistry*, 276(22), pp. 18695–701.

Livak, K. J. and Schmittgen, T. D. (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2 C T Method, *METHODS*, 25, pp. 402–408

Lobell D. B., Burke, M., Tebaldi, C., Mastrandrea, M., Falcon, W., Naylor, R. (2008) Prioritizing Climate Change Adaptation Needs for Food Security in 2030. *Science*, 319(5863), pp. 607-610

Mäkinen, K. and Hafrén, A. (2014), Intracellular coordination of potyviral RNA functions in infection., *Frontiers in plant science*, 5, pp. 110.

Marinaki, A. M., Ansari, A., Duley, J.A., Arenas, M., Sumi, S., Lewis, C.M., Shobowale-Bakre, el-M., Escuredo, E., Fairbanks, L.D., Sanderson, J.D. (2004), Adverse drug reactions to azathioprine therapy are associated with polymorphism in the gene encoding inosine triphosphate pyrophosphatase (ITPase). *Pharmacogenetics*, 14(3), pp. 181–7

Martin R. R., James D., Levesque C. A. (2000). Impacts of molecular diagnostic technologies on plant disease management. *Annu Rev Phytopathol*, 38, pp. 207–239

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

Maruthi, M. N., Bouvaine, S., Tufan, H., Mohammed, I., Hollocks, R. (2014), Transcriptional Response of Virus-Infected Cassava and Identification of Putative Sources of Resistance for Cassava Brown Streak Disease, *PLoS ONE*. 9(5), pp. e96642

Maruthi, M., Jeremiah, S., Mohammed, I. and Legg, J. (2016). Virus-vector relationships and the role of whiteflies, *Bemisia tabaci*, and farmer practices in the spread of Cassava brown streak viruses. *Phytopathology*, 165(11-12), pp.707-717

Mayer, K. *et al.* (1999), Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*, *Nature*, 402(6763), pp. 769–777

Mbanzibwa D. R., Tian Y. PP., Tugume A. K., Patil B. L., Yadav J. S., Bagewadi B., Abarshi M. M., Alicai T., Changadeya W. (2011). Evolution of cassava brown streak disease-associated viruses. *J Gen Virol,* 92, pp. 974–987

Mbanzibwa D. R., Tian Y., Mukasa S. B., Valkonen J. PP. (2009). Cassava brown streak virus (*Potyviridae*) encodes a putative Maf/HAM1 pyrophosphatase implicated in reduction of mutations and a P1 proteinase that suppresses RNA silencing but contains no HC-Pro. *J Virol,* 83, pp. 6934–6940

Miller S. A., Beed F. D., Harmon C. L. (2009). Plant disease diagnostic capabilities and networks. *Annu Rev Phytopathol,* 47*,* pp. 15–38.

Mohammed I. U., Abarshi M. M., Muli B., Hillocks R. J., Maruthi M. N. (2012). The symptom and genetic diversity of cassava brown streak viruses infecting cassava in East Africa. *Advances in Virology*, 2012, pp. 1-10

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

Monger W. A., Seal S., Isaac A. M., Foster G. D. (2001). Molecular characterization of the *Cassava brown streak virus* coat protein. *Plant Pathol,* 50, pp. 527–534.

Moore LM, Lawrence JH (2003). Plant guide – Cassava: manihot esculenta Crantz. National Plant Data Centre, Baton Rouge, Louisiana and Pacific Islands, Mongmong, Guam.

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

Noskov, V., Staak, K., Shcherbakova, PP., Kozmin, S., Negishi, K., Ono, B., Hayatsu, H. and Pavlov, Y. (1996). HAM1, the gene controlling 6-N-hydroxylaminopurine sensitivity and mutagenesis in the yeast Saccharomyces cerevisiae. *Yeast*, 12(1), pp.17-29.

Nweke, F., 2004. New challenges in the cassava transformation in Nigeria and Ghana. International Food Policy Research Institute, 118.

Ogwok E, Ilyas M, Alicai T, Rey MEC, Taylor NJ. (2014), Comparative analysis of virus-derived small RNAs within cassava (*Manihot esculenta* Crantz) infected with cassava brown streak viruses. *Virus Research*, 215, pp. 1-11.

Ogwok E., Odipio J., Halsey M., Gaitan-Solis E., Bua A., Taylor N.J., Fauquet C.M., Alicai T. (2012), Transgenic RNA interference (RNAi)-derived field resistance to cassava brown streak disease. Mol. *Plant Pathol*. 13(9), pp. 1019–1031.

Olsen, K. M., & Schaal, B. A. (1999). Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*. *Proc Natl Acad Sci USA*, 96(10), pp. 5586–5591.

Pang, B., McFaline, J., Burgis, N., Dong, M., Taghizadeh, K., Sullivan, M., Elmquist, E., Cuningham, R., Dedon, P. (2012) Defects in purine nucleotide metabolism lead to substantial incorporation of xanthine and hypoxanthine into DNA and RNA, *Proc Natl Acad Sci USA*, 109(7), pp. 2319–2324

Paterson, A. H. *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses, *Nature*, 457

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

Patil B.L., Legg J.PP., Kanju E., Fauquet C.M. (2015), Cassava brown streak disease: a threat to food security in Africa. *J. Gen. Virol*. 96(5), pp. 956–968

Patil, B. and Fauquet, C. (2009). Cassava mosaic geminiviruses: actual knowledge and perspectives. *Molecular Plant Pathology*, 10(5), pp.685-701.

Pavlov, Y., Suslov, V., Shcherbakova, PP., Kunkel, T., Ono, A., Matsuda, A. and Schaaper, R. (1996). Base analog N6-hydroxylaminopurine mutagenesis in Escherichia coli: genetic control and molecular specificity. *Mutation Research*, 357(1-2), pp.1-15.

Pyott, D., Sheehan, E. and Molnar, A. (2016). Engineering of CRISPR/Cas9-mediated potyvirus resistance in transgene-freeArabidopsisplants. *Molecular Plant Pathology*, 17(8), pp.1276-1288.

Ramanathan, A., Robb, G. B., & Chan, S.-H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Research*, 44(16), pp. 7511–7526

Reddy D. V., Sudarshana M. R., Fuchs M., Rao N. C., Thottappilly G. (2009). Genetically engineered virus-resistant plants in developing countries: current status and future prospects. *Adv Virus Res,* 75, pp. 185–220

Rio, D., Ares, M., Hannon, G. and Nilsen, T. (2010). Purification of RNA Using TRIzol (TRI Reagent). *Cold Spring Harbor Protocols*, 2010(6), pp. 5439- 5439.

Schippers, J. H. M., A. Nunes-Nesi, R. Apetrei, J. Hille, A. R. Fernie, and PP. PP. Dijkwel. 2008. The Arabidopsis onset of leaf death5 mutation of quinolinate synthase affects nicotinamide adenine dinucleotide biosynthesis and causes early ageing. *Plant Cell*, 20, pp. 2909–2925.

Shapiro, R. and Pohl, S. H. (1968) 'Reaction of ribonucleosides with nitrous acid. Side products and kinetics', *Biochemistry*, 7(1), pp. 448–455.

Simone, PP. D., Pavlov, Y. I., & Borgstahl, G. E. (2013). ITPA (inosine triphosphate pyrophosphatase): from surveillance of nucleotide pools to human disease and pharmacogenetics. *Mutation Research*, 753(2), pp. 131-146

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

Summers, J. and Litwin, S. (2006), Examining the theory of error catastrophe. *Journal of virology*, 80(1), pp. 20–6.

Takahashi, H., Goto, N. and Ehara, Y. (1994). Hypersensitive response in cucumber mosaic virus-inoculated Arabidopsis thaliana. *The Plant Journal*, 6(3), pp.369-377.

Takayama, S., Fujii, M., Kurosawa, A., Adachi, N. and Ayusawa, D. (2007). Overexpression of HAM1 gene detoxifies 5-bromodeoxyuridine in the yeast Saccharomyces cerevisiae. *Current Genetics*, 52(5-6), pp.203-211.

Tavert-Roudet, G., Anne, A., Barra, A., Chovin, A., Demaille, C. and Michon, T. (2017). The Potyvirus Particle Recruits the Plant Translation Initiation Factor eIF4E by Means of the VPg covalently Linked to the Viral RNA. *Molecular Plant-Microbe Interactions*, 30(9), pp.754-762.

Thresh JM, Otim-Nape GW, Fabres G, Yaninek YS, Adipala E, 1994. Integrating the management of pests, weeds and diseases of cassava in Africa. *African Crop Science Journal,* 2, pp. 331–592

Tomlinson, K., Bailey, A., Alicai, T., Seal, S. and Foster, G. (2017). Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Molecular Plant Pathology*, 19(5), pp.1282-1294.

Valli A., Martín-Hernández A. M., López-Moya J. J., García J. A. (2006). RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing ipomovirus, a member of the family *Potyviridae* that lacks the cysteine protease HCPro. *J Virol,* 80, pp. 10055–10063

Valli, A., García, J. A. and López-Moya, J. J. (2015) 'Potyviridae', in *eLS*. Chichester, UK: John Wiley & Sons, Ltd, pp. 1–10.

van den Born, E., Omellchenko, M., Bekkelund, A., Leihne, V., Koonin, E.V., Dolja, V.V., Falnes, P. (2008). Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic acids research*, 36(17), pp. 5451–61.

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

Wagaba, H., Beyene, G., Aleu, J., Odipio, J., Okao-Okuja, G., Chauhan, R., Munga, T., Obiero, H., Halsey, M., Ilyas, M., Raymond, PP., Bua, A., Taylor, N., Miano, D. and Alicai, T. (2017). Field Level RNAi-Mediated Resistance to Cassava Brown Streak Disease across Multiple Cropping Cycles and Diverse East African Agro-Ecological Locations. *Frontiers in Plant Science*, 7, 2060

Walsh, D. and Mohr, I. (2011), Viral subversion of the host protein synthesis machinery, *Nature Reviews Microbiology*, 9(12), pp. 860–875

Wei, T., Zhang, C., Hong, J., Xiong, R., Kasschau, K. D., Zhou, X., & Wang, A. (2010). Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathogens*, 6(6), pp. 62.

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

Yoshii, M., Yoshioka, N., Ishikawa, M. and Naito, S. (1998). Isolation of an Arabidopsis thaliana mutant in which accumulation of cucumber mosaic virus coat protein is delayed. *The Plant Journal*, 13(2), pp. 211-219.

Yu, J. *et al.* (2005), The Genomes of Oryza sativa: A History of Duplications, *PLoS Biology*, 3(2), pp. 38.

Zheng, J., Singh, V. and Jia, Z. (2005). Identification of an ITPase/XTPase in Escherichia coli by Structural and Biochemical Analysis. *Structure*, 13(10), pp.1511-1520.