Molecular and biological characterization of three *Citrus tristeza virus* candidate cross-protection sources

by
Jacoba Wilhelmina Lubbe
(28022981)

Supervisor: Prof. Gerhard Pietersen

Submitted in partial fulfillment of the requirements for the degree
Magister Scientiae Microbiology (MSc)
In the Faculty of Natural & Agricultural Science
University of Pretoria
Pretoria
March 2015

DECLARATION

I, Jacoba Wilhelmina Lubbe declare that the dissertation, which I hereby submit for the degree Magister Scientiae Microbiology (MSc) at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Signature: _____

Date: _____

ACKNOWLEDGEMENTS

Firstly I would like to thank the University of Pretoria and the Department of Microbiology and Plant Pathology for being able to complete my degree at the institution and for the provision of funding.

Thank you to my supervisor, Prof. Gerhard Pietersen for guidance and support during the completion of the degree.

Thank you to Citrus Research International for financial support and a special thanks to Dr. Fanie van Vuuren for all the advice and support, as well as Kobus Breytenbach for help with preparing the plants used in the biological indexing trials as well as his assistance and guidance during the inoculation of the plants.

Thank you to Dr. Martin Coetzee for giving support and knowledge during the phylogenetic analysis and thank you to Prof. Jacques Theron for help on molecular principles.

Thank you to all my lab mates for all the help and friendship through the years, whether it was by listening to a ranting session, giving advice, taking samples out of a thermocycler or just providing some laughs. David, Ronel, Kirsti, Jennifer, Megan, you made Lab 9-37 an excellent place to be.

Thank you to my other friends in the department, Marike, Vanessa and Gaby for all your friendship during my time on the 9th floor. Marike, a special thanks for your willingness to share your knowledge on phylogenetics and always helping with quick questions.

Pieter, thank you for all your support and patience, you have done a great job at keeping me sane. I love you with all my heart.

Thank you to my Mother and Father for your endless support, financially and emotionally, without you this work would have never come to light and I would have never been able to be where I am today.

Lastly, I would like to thank our Heavenly Father for blessing me with all the talents and the opportunities to make my dreams come true.

Opgedra aan my ouers, Cezar en Wilma Lubbe.

ABSTRACT

Citrus tristeza virus (CTV) is a RNA plant virus that infects the phloem cells of members of the family *Rutaceae*. CTV has a very important impact on the citrus industry worldwide and in South Africa especially so on grapefruit. CTV isolates can cause differing levels of severity of Tristeza disease, which can lead to quick decline as well as stem pitting and seedling yellows. Mild strain cross-protection is commonly used in South Africa to control the negative effects of the virus. This control mechanism is based on the super-infection exclusion principle where the presence of one specific genotype of CTV prevents the secondary infection of strains of the same genotype. This necessitates the characterization of CTV sources occurring within given citrus producing areas to know which genotypes to protect against, as well as the thorough characterization of potential cross-protection sources to ensure the specific genotypes that need to be protected against are present and to ensure that there are no strains within the source that would cause severe symptoms. The aim of this study was to characterize several sources of CTV which could potentially be used for cross-protection and at the same time to use and evaluate several methods for this. By doing next generation sequencing on an overlapping amplicon template of the 3' half of the genome it was found that the three Grape Fruit Mild Strain 12 sub isolates, GFMS 12-7, 12-8 and 12-9 mostly exists of a T68 genotype previously identified as CT-ZA3. Using immuno-captured virus particles as template, followed by the production of cDNA through the use of degenerate primers and random amplification of the DNA as well as a p33 gene amplicon for next generation sequencing, it was found that the New Venture 41/2 candidate mild source is a mixed source containing at least the VT, RB, B165 and HA16-5 genotypes. The B390/3 candidate mild source was characterized through biological indexing and was found to only produce mild symptoms on the hosts used in the trial. The virus population was also characterized through Sanger and Illumina sequencing of the p33 gene as well as using genotype specific RT-PCRs. The source is dominated by a Taiwan-Pum/SP/T1–like isolate which belongs to the RB genotype. Additionally a comprehensive phylogenetic analysis was performed on 45 published complete genomes of CTV where it was shown that 9 genotypes exist, namely VT, T36, RB, T30, B165, T68, HA16-5, T3 and A18. The best method for genotyping, as found to produce the phylograms most similar to the complete genome phylograms, was found to be by doing a Bayesian analysis on a concatenated dataset of three segments of the genome, namely ORF 1b, ORF 2 and ORF 5.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AMV	Avian Myeloblastosis Virus
BA	Bayesian
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
BSA	*Bovine serum albumin*
cDNA	Complimentary DNA
CP	Capsid protein
CPm	Minor coat protein
CREC	Citrus Research Education Center
CRI	Citrus Research International
CTV	*Citrus tristeza virus*
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dRNA	Defective RNA
dsRNA	Double stranded RNA
DTT	Dithiothreitol
ELISA	Enzyme linked immunosorbentassay
EtBr	Ethidium bromide
GFMS	Grapefruit Mild Strain
HEL	Helicase
HSP	Heat shock protein
IUPAC	International Union of Pure and Applied Chemistry
Kb	Kilo base
L	Litre
MEGA	Molecular Evolutionary Genetics Analysis
ml	Millilitre
ML	Maximum likelihood
mM	Millimolar

MT	Methyl transferase
NCBI	*National Center for Biotechnology Information*
NJ	Neighbour joining
ORF	Open reading frame
PBST	Phosphate buffered saline
PCR	Polymerase chain reaction
PRO	Papain-like protease
RB	Resistance breaking
RdRp	RNA dependant RNA polymerase
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SACIP	South African Citrus Improvement Program
SATs	Single aphid transmissions
SSCP	Single strand confirmation polymorphism
ssDNA	Single stranded DNA
ssRNA	Single stranded RNA
U	Units of enzyme
UTR	Untranslated region
µl	Microliter
µM	Micromolar

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Review of Literature

1.1 CTV structure and genome

The flexible virions of *Citrus tristeza virus* (CTV) are filamentous with a length of approximately 2 µm. CTV has only one RNA strand that is positive sense, making it a Group IV virus (Albiach-Marti *et al.*, 2010; Bar-Joseph *et al.*, 1989). The CTV genome consists of approximately 19 296 nucleotides with 12 open reading frames (ORF) as can be seen in Figure 1.1 (Karasev *et al.*, 1998). This makes CTV the largest known RNA virus (Vives *et al.*, 1999).

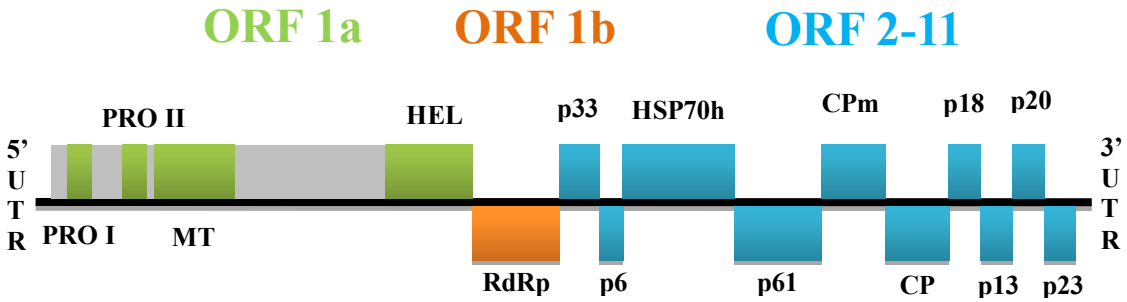

Figure 1.1: The genome of *Citrus tristeza virus*. UTR, untranslated region; PRO, papain-like protease; MT, methyl transferase; HEL, Helicase; RdRp, RNA dependant RNA polymerase; HSP70h, HSP70 homolog; CPm, minor coat protein; CP, major coat protein (Reproduced from Folimonova, 2012 and Moreno *et al*. 2008).

The proteins encoded by the genome range from 6 kDA to 400 kDA in size. ORF 1a is translated into a single 400 kDA polyprotein that is cleaved by two virus specific leader proteases to yield 4 proteins: methyltransferase, helicase and two papain-like proteases (Karasev, 2000). The RNA dependent RNA polymerase is translated from ORF 1b and was found to form oligomers in prokaryotic and eukaryotic experiment systems. The oligomerization is not co-translational and it was discovered that the binding site is within the N terminal of the protein (Çevik, 2013). ORF 3 encodes a hydrophobic protein (p6), ORF 4 encodes a homolog of the cellular heat shock related proteins of the HSP70 family and the protein encoded by ORF 5 (p61) was found to have some similarity to the HSP90 cellular heat shock protein family (Pappu *et al.*, 1994) and plays a role in virion assembly (Satyanarayana *et al.*, 2000). ORF 7 encodes the capsid protein gene (CP) which covers 95%

of the virion whereas ORF 6 (CPm) encodes a duplicate of the capsid protein, although it has diverged, that covers 5% of the virion at one end (Febres *et al.*, 1996). The protein encoded by ORF 10 (p20) is a component of inclusion bodies and plays a role in inducing systemic infection (Gowda *et al.*, 2000; Tatineni *et al.*, 2008). The protein coded by ORF 11, p23, has RNA binding properties and plays a role in symptom development and RNA silencing (Flores *et al.*, 2013). ORF 2 – 11 is expressed by translational frame-shifting and formation of subgenomic RNAs (Karasev, 2000). CTV has also been known to form defective RNAs (dRNAs) that can interfere with the replication of the virus. These particles can either suppress or enhance the development of disease symptoms (Karasev *et al.*, 1998).

The level of divergence of the 3' and 5' end of the genome differs. The 3' region of the VT and T36 isolates is more conserved, whereas the 5' end is much more divergent (Mawassi *et al.*, 1996). Hilf *et al*. (1999) found a relative consistent convergence of the genome between T3 and T30, while the comparison with T36 indicated a decrease in convergence at the 5' end. This implied that VT, T30 and T36 are different genotypes of CTV. Since then many other isolates have been found but not all of them have been assigned to genotypes. The genotypic groups of CTV is still mutable at this time; Folimonova (2012) and Harper (2013) described six CTV genotypes; T36, T3, T30, T68, VT and RB, while Biswas *et al*. (2012) described them as T36, T30, VT, B165, HA16-5 and RB.

1.2 CTV hosts and symptoms

CTV replicates in the phloem cells of a few species within the *Rutaceae* family (Moreno *et al.*, 2008). This includes most citrus species, varieties, hybrids and some relatives (Albiach-Marti *et al.*, 2000). *Passiflora* is the only host of CTV outside the *Rutaceae* family *(Bar-Joseph et al., 1989)*. Although CTV has been found to replicate in the protoplasts of *Nicotiana benthamiana,* it is not a natural host of CTV and cannot be spread by aphid vectors

to- and from it. This host system however has been utilized in molecular studies of the virus (Albiach-Marti *et al.*, 2010; Ambrós *et al.*, 2013; Bar-Joseph *et al.*, 2002; Gowda *et al.*, 2000; Navas-Castillo *et al.*, 1997; Satyanarayana *et al.*, 2000).

The movement of CTV within plants is limited and differs based on the susceptibility of the plant to the virus. In more susceptible plants, there is only a small portion of the phloem associated cells that are infected, limiting the long distance movement of CTV through sieve elements. Less susceptible plants have a much lower number of infected cells. There is also only a small amount of cell to cell movement in the susceptible plants since the site of infection only contains a small cluster of infected cells. Cell to cell movement of the virus in less susceptible plants is absent, since the site of infection only contains single infected cells (Folimonova *et al.*, 2008).

The severity and types of symptoms differ based on; the host plant species (scion and rootstock), strain of CTV, and environmental conditions (Broadbent *et al.*, 1996; Garnsey *et al.*, 1991; Sambade *et al.*, 2003; Van Vuuren, 2002). CTV has a lower titre in hot environments, leading to less severe symptoms, while it causes more severe symptoms in cooler environments where it has a higher titre (Lee *et al.*, 1988). One strain of CTV can cause different symptoms in two species of citrus under the same conditions and the severity may also differ (Garnsey *et al.*, 2005). There are three different sets of symptoms that can develop due to a CTV infection (Figure 1.2), namely Tristeza disease, Stem pitting and Seedling yellows (Moreno *et al.*, 2008).

The most devastating of the three, Tristeza disease, causes the collapse and death of the sieve tubes as well as the production of a large amount of non-functional phloem by the companion cells close to the bud union. This leads to the root system becoming reduced which deprives the tree of water and minerals. Tristeza disease causes quick decline which results in the wilting of trees leading to the complete collapse of the tree within a few weeks. Symptoms

include dull green or yellow leaves, shedding of the leaves, dieback of twigs and leaves that may be chlorotic. The fruit are small and pale colored, obliterating the market value. The chlorosis of the leaves resembles nitrogen deficiency symptoms. Sweet orange (*Citrus sinensis*), mandarin (*Citrus reticulata*), grapefruits (*Citrus paradisi*), kumquat (*Citrus japonica*) and limes (*Citrus aurantifolia*) propagated on sour orange (*Citrus aurantium*) or lemon (*Citrus limon*) rootstock are susceptible to this disease (Moreno *et al.*, 2008).

Figure 1.2: Symptoms induced by *Citrus tristeza virus;* (A and B) Decline in Star Ruby grapefruit, (C) Stem pitting in Star Ruby grapefruit, and (D) Small sized fruits from Star Ruby grapefruit trees.

Stem pitting does not kill the host, but causes severe losses. It is caused by the interruption of meristematic activity at areas of the cambium. This results in irregular growth with local depressions in these sites. Further symptoms of stem pitting include the limitation of radial growth, stunting of growth, thin, small, yellow foliage and fruits that are small with a low

juice content which are borne low on the tree. Acid limes, grapefruits, mandarins and some sweet orange varieties can be affected (Moreno *et al.*, 2008).

A tree showing seedling yellows symptoms may recover after a certain period of time. Symptoms include stunting, small pale or yellow foliage, a reduction in the root system and in some cases, a complete termination of growth. Seedling yellows disease mainly affects sour orange, grapefruit and lemon seedlings (Moreno *et al.*, 2008).

1.3 Vectors

CTV has a large number of aphid vectors, the most common being *Toxoptera citricida* and *Aphis gossypii*, but other species like *Aphis spiraecola*, *Toxoptera aurantii*, *Myzus persicae*, *Aphis craccivora* and *Uroleucon jacaea* have also been shown to transmit the virus (Moreno *et al.*, 2008). CTV is transmitted in a semi-persistent manner, in other words, it is only circulated in the foregut of the insect. The virus can be acquired if the aphid feeds for only a few seconds, and can then be transmitted after 4 to 6 hours for a period of 24 to 48 hours (Bar-Joseph *et al.*, 1989). The efficiency by which the virus is transmitted differs between aphid species and the life stages of the aphids. *T. citricida* can be 6 to 25 times more efficient in transmitting the virus than *A. gossypii* (Moreno *et al.*, 2008). For *T. citricida*, nymphs and apterous adults (wingless) are more efficient in transmitting the virus than alatae adults (winged). Although no difference could be identified between the efficiency of transmission of *A. gossypii* nymphs and adults, this aphid transmits different strains of CTV with different efficiency. The host plant was also found to play a role in the transmission of the virus by *T. citricida* and *A. gossypii* (Bar-Joseph *et al.*, 1989). *T. citricida* is the main vector of CTV in Asia, Australia, Sub-Saharan Africa, Central and South America and some Caribbean countries. *A. gossypii* is the main vector in the Mediterranean basin and areas in North America (Moreno *et al.*, 2008).

1.4 CTV recombination, evolution and the role of viral quasispecies

A CTV population consists of genetically related variants of the virus. This is called the quasispecies phenomenon and is common of RNA virus populations. This phenomenon is caused by the error prone RNA polymerase, large population sizes and short replication times of RNA viruses (Rubio *et al.*, 2001). Domingo *et al.*(1998) described quasispecies as a collection of mutant and recombinant viral genomes within a viral population, and predicted that an increase in the population titre will lead to an increase in mutations.

Many studies have provided evidence that recombination takes place in the CTV genome (Roy and Brlansky, 2010; Rubio *et al.*, 2001; Vives *et al.*, 2005), and that recombination can take place between different variants of CTV (Rubio *et al.*, 2001) resulting in both homologous and non-homologous recombination (Vives *et al.*, 2005). Weng *et al.* (2007) showed that recombination takes place frequently between genotypes present within a plant and studied the divergence of these recombinants within the host plant. More specifically, it was found that some CTV genomes have arisen due to recombination; including B165, a Stem pitting isolate from orange in India (Roy and Brlansky, 2010) and SY568, a Seedling yellows isolate from Spain (Vives *et al.*, 1999).

There are several reasons why it is more advantageous for a virus genome to be dynamic rather than static. Domingo *et al.*(1998) proposed that genetic variation caused by mutation and recombination within a genome may play a role in the adaption of the virus to changes in the environment. Recombination can also create genetic diversity and prevent the accumulation of deleterious mutations by regenerating functional genomes from genomes with deleterious mutations due to the error-prone RNA polymerase (Vives *et al.*, 2005).

The occurrence of recombination and mutations within the CTV genome would suggest that this virus has a high evolutionary rate, and this could impact on the control of CTV. However, the opposite is true. Albiach-Marti *et al.* (2000) found that pure CTV strains

separated in time and space, and subjected to different hosts, graft and aphid transmission passages were basically identical after several decades. Silva *et al.*(2012) used the capsid protein of CTV to determine the evolutionary rate of the virus. It was found that CTV is the most slowly evolving plant RNA virus known, and the evolutionary rate of CTV compares to the evolutionary rate of the slowest evolving animal RNA viruses.

1.5 Mixed CTV populations with mild and severe CTV strains

Different genotypes of CTV have been identified based on variation in the 5' gene region (Biswas *et al.*, 2012; Folimonova *et al.*, 2010). Since citrus plants are long-lived and repeated infections of CTV can take place through aphid transmission, mixed CTV infections can occur in one host plant (Rubio *et al.*, 2001). This creates the possibility of a CTV population containing many different genotypes and even defective RNAs (dRNAs). Most CTV sources (the complete CTV population as found in a host plant) are a mix of different genotypes and dRNAs, while others may have a principle genotype, with minor concentrations of other genotypes (Albiach-Marti *et al.*, 2000; Kong *et al.*, 2000). Černi *et al.* (2008) found that during transmissions, the dominant strain within the plant is transmitted to subsequent hosts and remained stable for five years in the sub-isolates. Minor strains transmitted were more dynamic and also influenced symptom expression. These authors speculate that interactions between the dominant strain and minor strains or even interactions between only the minor strains may influence the symptom development.

Strains within these genotypes are usually described as 'mild' or 'severe' based on the symptoms caused in a host plant. A CTV source is often referred to in these terms in spite of containing different strains and possibly different genotypes. Mild sources of CTV may contain a mixture of CTV strains, both mild and severe, but expresses mild symptoms due to the dominance of the mild strain within the population. The variation of the strains present is

influenced by host and environmental conditions (Albiach-Marti *et al.*, 1996). The strains within a CTV source may differ between glasshouse and field maintained sources, as well as sources maintained in different hosts. Albiach-Marti *et al.* (1996) suggested that this is due to super-infection of other CTV strains that occurs in the field, or due to increased replication of other components within the original CTV source.

1.6 Control of CTV

CTV can be controlled by placing affected areas under quarantine and eradicating infected trees to prevent further spread of the virus. Certification schemes are also often utilized where virus free budwood is propagated and inoculated with mild variants of the virus for cross-protection (Sambade *et al.*, 2003). The grafting of some cultivars onto resistant rootstocks is also used, although this control method is not always possible since the combination of cultivars may vary in susceptibility (Fulton, 1986). Numerous attempts have been made to create genetically modified citrus hosts resistant to the virus, but with limited success. Transformation of plants with variants of the 3' UTR and the p23 gene lowered CTV replication but did not provide full resistance (Ananthakrishnan *et al.*, 2007). Transformation attempts with the coat protein also did not provide any resistant plants (Muniz *et al.*, 2011). Over expression of recombinant single chain variable fragment antibodies against the p25 major coat protein in Mexican lime conferred 40-60% resistance for 4 years (Cervera *et al.*, 2010). Thus far, the most promising and commonly used control mechanism is mild strain cross-protection.

1.6.1 Mild strain cross-protection

Mild strain cross-protection is based on the super-infection exclusion or homologous interference principle. This is where the primary infection of a plant with a virus prevents the secondary infection of the same, or a closely related virus (Folimonova *et al.*, 2010). For

CTV, super-infection exclusion only occurs between variants of the same genotypes and not between variants of different genotypes (Folimonova *et al.*, 2010).

The mechanism of super-infection exclusion has not been elucidated. In other systems, the expression of coat proteins of the virus that first infected the plant, prevented the uncoating of the second infecting virus, or the genome of the second virus infecting the plant is degraded by RNA silencing mechanisms. In the case of CTV, even phloem-associated cells not infected with the primary virus are protected against infection. This means that the ability of protection spreads to uninfected cells (Folimonova *et al.*, 2010), and known mechanisms of super-infection exclusion do not provide an explanation for this phenomenon. Thus far it has been found that a specific protein, p33, is crucial for super-infection exclusion and without it, a virus cannot exclude super-infection by the same or a closely related virus (Folimonova, 2012).

Since more than one genotype of CTV can occur within a plant (Albiach-Marti *et al.*, 2000); and one strain of CTV can cause mild symptoms in one host, but severe symptoms in another (Fulton, 1986); and a mild strain of CTV can produce mild symptoms in one citrus producing area, but severe symptoms in another (Karasev *et al.*, 1998); there can be variability in the success of a mild strain cross-protection source. Environmental conditions and host species can influence the selection of genotypes within a plant (Albiach-Marti *et al.*, 1996), and this may cause a source of CTV to become more severe under certain circumstances. For example, a CTV source containing a mild strain of CTV that is dominant under certain conditions can be inoculated into a different host in other environmental conditions. Under these circumstances, the mild strain may become less dominant and the host may select for a more severe strain of CTV within the population, causing more severe symptoms (Van Vuuren, 2002) and leading to the breakdown of cross-protection.

This creates the need for homogenous cross-protection sources that only contain a mild variant of a genotype. If this can be obtained for every CTV genotype, a superior cross-protection source can be created that only contains mild forms of each genotype. Since the selection of CTV genotypes differ in hosts and environmental conditions (Albiach-Marti *et al.*, 1996), these cross-protection sources may need to differ based on cultivars and growing areas.

The method most commonly used to identify pre-immunizing CTV sources is based on the visual selection of mild strains. CTV sources are obtained from host trees that are possibly infected with CTV but do not show symptoms, or only shows mild symptoms. These CTV sources are considered as candidate pre-immunizing sources and will then be inoculated into other hosts to study the symptom expression. If only mild symptoms are still produced after the field trials, it will be used as a pre-immunizing source (Albiach-Marti *et al.*, 1996; Zanutto *et al.*, 2013).

Another method that has also been attempted is passing a CTV isolate through *Passiflora* to attenuate the virus. Protective isolates that have been created in this manner have shown good results in the field (Roistacher *et al.*, 1988).

1.7 Detection and characterization of CTV

Visual inspection of symptoms on sensitive host plants was used in the past to identify CTV infection. At that stage, characterization of the virus was based on biological indexing. Although effective in some cases, this is a time consuming procedure that takes months to determine if a plant is CTV free. Infection can also be missed if no symptoms are produced by the CTV strain in question on that specific host and in that specific environment. A further way to study CTV populations is to do single aphid transmissions to separate the population into strains through aphid transmission and observing symptoms produced in inoculated

plants. The production of antisera and monoclonal antibodies allowed for the detection of CTV with ELISA (enzyme linked immunosorbentassay) (Bar-Joseph *et al.*, 1979). Using the two monoclonal antibodies specific to the p25 coat protein, 3DF1 and 3CA5, all CTV strains known to date can be detected (Cervera *et al.*, 2010). An adaption on ELISA is the direct tissue blot immunoassay where stem or petiole prints are made on nitrocellulose membranes followed by binding of the antibodies and detection under a microscope (Garnsey *et al.*, 1993). The production of antibodies specific to CTV and the availability of sequence information to create primers specific to the virus made it possible to do immuno-capture in PCR tubes or micro-titre plates or even direct print capture on nylon membranes followed by reverse transcription polymerase chain reaction (RT-PCR) (Nolasco *et al.*, 1993; Olmos *et al.*, 1996). When cloning the amplicons and doing Sanger sequencing, phylogenetic analysis can be performed to group isolates into genotypes (Scott *et al.*, 2013). Nested RT-PCR has also been incorporated for the detection of CTV (Olmos *et al.*, 1999) and with the use of specific primers in RT-PCR to amplify markers in the 5' end of the genome, the virus can be detected and classified into different CTV genotypes (Hilf *et al.*, 2005). Multiplex PCR has also been developed to detect and differentiate between genotypes of CTV in one reaction (Roy *et al.*, 2010). Furthermore, real time RT-PCR systems have been used to detect and quantify CTV levels within plant and vector tissues (Bertolini *et al.*, 2008; Ruiz-Ruiz *et al.*, 2007; Saponari *et al.*, 2008). It was found that real time RT-PCR is more sensitive than ELISA, but ELISA still has the highest specificity (the proportion of true negatives that are correctly identified by the test) (Vidal *et al.*, 2012). Ananthakrishnan *et al.* (2010) designed a multiplex real time RT-PCR that allows not only for the detection of CTV, but also quantification, identification and genetic diversity assessment.

Single strand confirmation polymorphism (SSCP) is a very common method of differentiating between different genotypes of CTV (Roy and Brlansky, 2009; Sambade *et*

al., 2003). It relies on nucleotide differences between same sized DNA fragments to form different banding patterns when denatured and separated by gel electrophoreses (Luttig *et al.*, 2002). Some genotypes can however not be separated from each other with this technique (Gago-Zachert *et al.*, 1999). Restriction fragment length polymorphism (RFLP) also allows for the differentiation of different genotypes based on the banding patterns on an agarose gel. But with this method, a gene of interest (for example the coat protein gene) is subjected to RT-PCR and then the amplicons are digested with restriction enzymes to create different gel profiles when separated by gel electrophoresis (Gillings *et al.*, 1993).

Since the era of next generation sequencing dawned and the sequencing of full genomes became less labor intensive and more affordable, the characterization of viruses based on full genome sequencing is also possible. The problem with the sequencing of complete RNA viral genomes lies firstly in producing complimentary DNA (cDNA) to the RNA and secondly in the amplification of the DNA to have enough starting material without introducing a primer bias. A study based on HIV implemented random primers as well as oligo d(t) primers instead of specific primers to limit PCR bias (Willerth *et al.*, 2010). Rather than focusing on the complete genome, specific amplicons can also be sequenced and analyzed phylogenetically (Martínez *et al.*, 2012; Morroni *et al.*, 2013).

The characterization of CTV based on sequence data, encompassing genome regions and techniques used, differ tremendously among different studies making comparisons between studies impossible. The use of Neighbor joining trees to describe sequence relationships is often used, but some studies rather make use of maximum likelihood analyses, which relies on a different algorithm to create phylograms which can lead to differences in the apparent relationships between sequences. The genome segments used include different portions within ORF 1a as well as the coat protein, p23, p20 and complete genome sequences (Davino

et al., 2013; Harper *et al.*, 2010; Roy *et al.*, 2005; Rubio *et al.*, 2001; Sambade *et al.*, 2003; Scott *et al.*, 2013).

1.7.1 Biological indexing of CTV

The biological characterization of CTV can be done by means of inoculating indicator plants through graft transmission. Due to the complex and diverse nature of symptom expression by CTV, different index plants have to be used to observe different symptoms. Symptoms apparent on leaves, for example vein clearing, leaf cupping and chlorosis, as well as plant height reduction can be studied on West Indian lime and Mexican lime (Broadbent *et al.*, 1996; Garnsey *et al.*, 2005). Seedling yellows symptoms are apparent on Duncan grapefruit, Eureka lemon and Bittersweet Seville orange (Broadbent *et al.*, 1996). If the stem is peeled away above the point of inoculation at the final reading, sweet orange, West Indian lime and Duncan grapefruit can be used to observe the degree of stem pitting caused by the isolate (Broadbent *et al.*, 1996; Garnsey *et al.*, 2005). Lastly, quick decline symptoms can be seen when sweet orange is grafted on sour orange rootstocks (Broadbent *et al.*, 1996).

1.7.2 Single aphid transmissions (SATs)

When dealing with a CTV population containing multiple strains, aphid transmissions can be done with the hope of an aphid only picking up one viral strain and transmitting that strain to a subsequent host. In this way the procedure has a dilution effect and might need to be repeated multiple times in order to get only one strain of CTV within a host (Van Vuuren and Van Der Vyver, 2000). Although the technique is simple in theory, it can become quite difficult due to the low efficiency with which the virus is transmitted.

The transmission efficiency of SATs can be described as the ratio between the number of infected plants and the total number of aphid inoculated plants, and is usually fairly low (Broadbent *et al.*, 1996). The transmission efficiency of SATs is influenced by several

factors. The isolate of CTV can influence the efficiency, whereas some isolates are transmitted more readily than others (Lin *et al.*, 2002). For instance, severe isolates are transmitted with a higher efficiency than mild isolates. The original host plant from which the isolate was obtained and the receptor plant also plays a role (Broadbent *et al.*, 1996; Hermoso *et al.*, 1988; Lin *et al.*, 2002). For example, isolates from orange or mandarin are more readily transmitted than isolates obtained from grapefruit (Broadbent *et al.*, 1996). The environmental conditions during transmission and the morphological stage of the aphid can also influence the efficiency (Huang *et al.*, 2005). The transmission efficiency can be as low as 1% or 1.5% in some cases (Lin *et al.*, 2002), but can be as high as 55% in other cases (Broadbent *et al.*, 1996), but this varies greatly based on isolates of CTV as well as host and receptor plants used.

SAT studies have provided supporting evidence of the mixed population nature of CTV sources. Field isolates producing certain symptoms subjected to SATs resulted in sub-isolates which did not necessarily produce the same symptoms as the original source. For example, sub-isolates of sources that only showed mild symptoms on the original host, showed severe symptoms after SAT to new hosts (Broadbent *et al.*, 1996). Costa *et al.*(2010) found that CTV populations within sub-isolates differ both from other sub-isolates from the same source and from the original source. During SSCP studies, d'Urso *et al.* (2003) also found a difference in CTV populations before and after SATs. This makes SAT an excellent method to separate strains of CTV and to produce homogenous sources of the virus.

1.8 CTV in South Africa

CTV has had a dramatic influence on the citrus industry of South Africa, especially with regards to the production of grapefruit (Roistacher and Moreno, 1991). The presence of the most efficient vector of the virus, *T. citricida*, caused the natural spread of CTV in all citrus

growing areas within the country (Bar-Joseph *et al.*, 1989). When sour orange rootstock first became popular for propagation, it was thought that the lack in success of this rootstock within South Africa was due to incompatibility. It is now known that it was due to CTV (Roistacher *et al.*, 2010). The South African climate plays a role in the expression of symptoms and causes difficulty in moving cultivars across areas. Redbush grapefruit for example showed few symptoms in most areas of South Africa, but when introduced to Natal this cultivar developed severe stem pitting (Roistacher and Moreno, 1991). The South African Citrus Improvement Program was established to cope with the impact of CTV on the industry. Within this program it was decided in 1982 to create virus free plant material and to inoculate these with a mild and protective strain of CTV (Von Broembsen and Lee, 1988).

1.9 GFMS 12 as a cross-protection source

In searching for a cross-protection source of CTV, the Nartia strain was identified in an orchard in Wellington. Three of four trees planted in 1926 were still showing only mild symptoms after 50 years. Nartia, now known as Grapefruit Mild Strain 12 (GFMS 12), was subjected to glasshouse trails to compare it to other possible mild strains on indicator plants (Von Broembsen and Lee, 1988). GFMS 12 allowed good protection of Marsh grapefruit for several years, but did not show so much promise for Star Ruby Grapefruit (Van Der Vyver *et al.*, 2002). In later years, trees pre-immunized with GFMS 12 started showing more severe CTV symptoms (Van Vuuren *et al.*, 1991) and caused different symptoms in different hosts (Meyer *et al.*, 2005). Field trails done with this cross-protection source showed severe stem pitting in both Marsh and Star Ruby grapefruit, indicating it was no longer suitable for cross-protection (Van Vuuren and Van Der Vyver, 2000). Van Vuuren *et al.*(2000) produced sub-isolates of GFMS 12 and found that the sub-isolates were either milder or more severe than the original isolate. This confirmed that GFMS 12 consisted of more than one strain,

including a severe strain. From the sub-isolates it was found that 12-2 and 12-5 were less virulent and 12-3 was more virulent than the original source. The sub-isolates may still contain mixtures of the virus and will need to be characterized further. Scott et al. (2013) did further analysis on GFMS 12 and sub-isolates 12-7, 12-8 and 12-9. Based on biological indexing on seven different citrus hosts, GFMS 12 showed moderate vein clearing and stem pitting on Mexican lime and mild stunting and moderate stem pitting on Duncan grapefruit. Furthermore the three sub-isolates showed different symptoms, with 12-7 producing sweet on sour decline but milder symptoms than GFMS 12 on Mexican lime and Duncan grapefruit. Sub-isolate 12-8 did not show any stem pitting on grapefruit and 12-9 showed mild stem pitting and stunting on all indicators. While doing sequence analysis of clones for the A region and p23 gene, it was found that GFMS 12 contains at least RB, T30, VT and B165. The sub-isolates produced different results based on whether the A-region (ORF 1a, 5' end of genome) or the p23 gene (3' end of genome) was analyzed and it was proposed that recombinants are present. Sub-isolate 12-7 contained RB/VT-like and B165/VT-like recombinant strains, while 12-8 contained a B165/VT-like recombinant strain and 12-9 contained a mixture of VT-like and B165/VT recombinant strains. The B165/VT-like recombinant was later sequenced by Zablocki and Pietersen (2014). It was named CT-ZA and was found to be a variant of the T68 genotype.

1.10 New Venture 41/2 as a potential cross-protection source

In a 2004 to 2015 study by Citrus Research International (CRI), South Africa, budwood was collected from 108 grapefruit CTV infected trees not showing symptoms of CTV. The sources were established in a glasshouse and inoculated into virus free Mexican lime, to assess the severity of the CTV sources. After biological testing in glasshouse trials the most promising sources were subjected to field trials. Among these sources, one of the most

promising sources was one named New Venture 41/2. This CTV source was inoculated into Star Ruby trees planted in February 2007 at Bosveld Sitrus in the Letsitele area, as well as Marsh trees planted in March 2007 at Riverside in the Malelane area. The trees are currently still being evaluated for growth and stem pitting and although differences are noticeable, it is too early to draw conclusions. The trees will be evaluated annually until 2015 for growth, production, fruit size and tree health (Breytenbach *et al.*, 2014b; c).

During a previous study, the New Venture 41/2 source was characterized based on the 1a gene sequence (Lubbe, *unpublished results*) and was found to contain strains from the VT genotype of CTV.

1.11 The B390/3 CTV source

The B390/3 source was one of the isolates obtained when SATs were done from the Mouton CTV source (originally from South Africa) in Beltsville, USA. The SAT isolates obtained from the Mouton source were imported back to South Africa where they were subjected to further studies. The B390/3 source has been subjected to field trials where both Marsh and Star Ruby Grapefruit were inoculated and monitored for symptom expression. B390/3 performed very well and might be a good option as a cross-protection source based on the mild symptoms produced on the field grapefruit trees. This source was found to produce mild stem pitting that regressed the next season and trees infected with this source was amongst the trees that produced the highest amount of fruit. The source is still being tested in other field trials (Breytenbach *et al.*, 2014a; b; c).

1.12 References

Albiach-Marti, M.R., da Graça, J.V., van Vuuren, S.P., Guerri, J., Cambra, M., Laigret, F., and Moreno, P. 1996. The Effects of Different Hosts and Natural Disease Pressure on Molecular Profiles of Mild Isolates of *Citrus tristeza virus* (CTV). In: *Proceedings of the 13th Conference of the International Organisation of Citrus Virologists. (Graça, J.V., Moremo, P., and Yokomi, R.K.)* pp. 147-153. *IOCV*. Riverside, California.

Albiach-Marti, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., Lopez, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., and Dawson, W.O. 2000. Sequences of *Citrus tristeza virus* Separated in Time and Space Are Essentially Identical. *The Journal of Virology* 74(15), pp. 6856-6865.

Albiach-Marti, M.R., Robertson, C., Gowda, S., Tatineni, S., Belliure, B., Garnsey, S.M., Folimonova, S.Y., Moreno, P., and Dawson, W.O. 2010. The Pathogenicity Determinant of *Citrus tristeza virus* Causing the Seedling Yellows Syndrome Maps at the 3′-Terminal Region of the Viral Genome. *Molecular Plant Pathology* 11(1), pp. 55-67.

Ambrós, S., Ruiz-Ruiz, S., Peña, L., and Moreno, P. 2013. A Genetic System for *Citrus tristeza virus* Using the Non-Natural Host *Nicotiana benthamiana*: An Update. *Frontiers in Microbiology* 4(165).

Ananthakrishnan, G., Orbović, V., Pasquali, G., Ćalović, M., and Grosser, J. 2007. Transfer of *Citrus tristeza virus* (CTV)-Derived Resistance Candidate Sequences to Four Grapefruit Cultivars through *Agrobacterium*-Mediated Genetic Transformation. *In Vitro Cellular & Developmental Biology - Plant* 43(6), pp. 593-601.

Ananthakrishnan, G., Venkataprasanna, T., Roy, A., and Brlansky, R.H. 2010. Characterization of the Mixture of Genotypes of a *Citrus tristeza virus* Isolate by Reverse Transcription-Quantitative Real-Time PCR. *Journal of Virological Methods* 164(1–2), pp. 75-82.

Bar-Joseph, M., Che, X., Mawassi, M., Gowda, S., Satyanarayana, T., Maria, A., Ayllón, M.A., Albiach-Marti, M.R., Garnsey, S.M., and Dawson, W., O. 2002. The Continuous Challenge of *Citrus tristeza virus* Molecular Research. In: *Proceedings of the 15th Conference of the International Organisation of Citrus Virologists. (Duran-Vila, N., Milne, R.G., and Da Graça, J.V.)* pp. 1-7. *IOCV*. Riverside, California.

Bar-Joseph, M., Garnsey, S., Gonsalves, D., Moscovitz, M., Purcifull, D., Clark, M., and Loebenstein, G. 1979. The Use of Enzyme-Linked Immunosorbent Assay for Detection of *Citrus tristeza virus*. *Phytopathology* 69(2), pp. 190-194.

Bar-Joseph, M., Marcus, R., and Lee, R.F. 1989. The Continuous Challenge of *Citrus tristeza virus* Control. *Annual Review of Phytopathology* 27(1), pp. 291-316.

Bertolini, E., Moreno, A., Capote, N., Olmos, A., de Luis, A., Vidal, E., Pérez-Panadés, J., and Cambra, M. 2008. Quantitative Detection of *Citrus tristeza virus* in Plant Tissues and Single Aphids by Real-Time RT-PCR. *European Journal of Plant Pathology* 120(2), pp. 177-188.

Biswas, K., Tarafdar, A., and Sharma, S. 2012. Complete Genome Sequence of Mandarin Decline Citrus tristeza virus of the Northeastern Himalayan Hill Region of India: Comparative Analyses Determine Recombinant. *Archives of Virology* 157(3), pp. 579-583.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014a. Final Report: Cross-Protection of Marsh and Star Ruby Grapefruit Using Beltsville Sub-Isolates of Nartia Mild Strain, pp. 23. Citrus Research International, Nelspruit, South Africa.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014b. Progress Report: Cross-Protection of Star Ruby Using Beltsville Sub-Isolates of Nartia Mild Strain for the Orange River Valley, pp. 24. Citrus Research International, Nelspruit, South Africa.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014c. Progress Report: Cross-Protection of Marsh and Star Ruby by Using the Best Field Isolates Collected in the Different Grapefruit Production Areas of Southern Africa., pp. 105. Citrus Research International, Nelspruit, South Africa.

Broadbent, P., Brlansky, R.H., and Indsto, J. 1996. Biological Characterization of Australian Isolates of *Citrus tristeza virus* and Separation of Subisolates by Single Aphid Transmission. *Plant Disease* 80(3), pp. 329-333.

Černi, S., Ruščić, J., Nolasco, G., Gatin, Ž., Krajačić, M., and Škorić, D. 2008.Stem Pitting and Seedling Yellows Symptoms of *Citrus tristeza virus* Infection May Be Determined by Minor Sequence Variants. *Virus Genes* 36(1), pp. 241-249.

Cervera, M., Esteban, O., Gil, M., Gorris, M., Martínez, M., Peña, L., and Cambra, M. 2010. Transgenic Expression in Citrus of Single-Chain Antibody Fragments Specific to *Citrus tristeza virus* Confers Virus Resistance. *Transgenic Research* 19(6), pp. 1001-1015.

Çevik, B. 2013. The RNA-Dependent RNA Polymerase of *Citrus tristeza virus* Forms Oligomers. *Virology* 447(1), pp. 121-130.

Costa, A.T., de Carvalho Nunes, W.M., Zanutto, C.A., and Müller, G.W. 2010. Stability of *Citrus tristeza virus* Protective Isolates in Field Conditions. *Pesquisa Agropecuaria Brasileira* 45(7), pp. 693-700.

D'Urso, F., Sambade, A., Moya, A., Guerri, J., and Moreno, P. 2003. Variation of Haplotype Distributions of Two Genomic Regions of *Citrus tristeza virus* Populations from Eastern Spain. *Molecular Ecology* 12(2), pp. 517-526.

Davino, S., Willemsen, A., Panno, S., Davino, M., Catara, A., Elena, S.F., and Rubio, L. 2013. Emergence and Phylodynamics of *Citrus tristeza virus* in Sicily, Italy. *PLoS ONE* 8(6).

Domingo, E., Baranowski, E., Ruiz-Jarabo, C.M., Escarmis, C., Martin-Hernandez, A.M., and Saiz, J.C. 1998. Quasispecies Structure and Persistence of RNA Viruses. *Emerging Infectious Diseases* 4(4), pp. 521.

Febres, V., Ashoulin, L., Mawassi, M., Frank, A., Bar-Joseph, M., Manjunath, K., Lee, R., and Niblett, C. 1996. The p27 Protein Is Present at One End of Citrus tristeza virus Particles.

Flores, R., Ruiz-Ruiz, S., Soler, N., Sánchez-Navarro, J., Fagoaga, C., López, C., Navarro, L., Moreno, P., and Peña, L. 2013. *Citrus tristeza virus* p23: A Unique Protein Mediating Key Virus–Host Interactions. *Frontiers in Microbiology* 4, pp. 98.

Folimonova, S.Y. 2012. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology*. 10.1128/jvi.00310-12

Folimonova, S.Y., Folimonov, A.S., Tatineni, S., and Dawson, W.O. 2008. *Citrus tristeza virus*: Survival at the Edge of the Movement Continuum. *Journal of Virology*, pp. JVI.00515-00508.

Folimonova, S.Y., Robertson, C.J., Shilts, T., Folimonov, A.S., Hilf, M.E., Garnsey, S.M., and Dawson, W.O. 2010. Infection with Strains of *Citrus tristeza virus* Does Not Exclude Superinfection by Other Strains of the Virus. *The Journal of Virology* 84(3), pp. 1314-1325.

Fulton, R.W. 1986. Practices and Precautions in the Use of Cross Protection for Plant Virus Disease Control. *Annual Review of Phytopathology* 24, pp. 67-81.

Gago-Zachert, S., Costa, N., Semorile, L., and Grau, O. 1999. Sequence Variability in p27 Gene of *Citrus tristeza virus* (CTV) Revealed by SSCP Analysis. *Electronic Journal of Biotechnology* 2, pp. 3-4.

Garnsey, S., Permar, T., Cambra, M., and Henderson, C. 1993. Direct Tissue Blot Immunoassay (DTBIA) for Detection of *Citrus tristeza virus* (CTV). In: *Proceedings of the 12th Conference of the International Organisation of Citrus Virologists. (Moreno, A., Da Graça, J.V., and Timmer, L.W.)* pp. 39-50. *IOCV.* Riverside, California.

Garnsey, S.M., Civerolo, D.J., Gumpf, D.J., Paul, C., Hilf, M.E., Lee, R.F., Brlansky, R.H., Yokomi, R.K., and Hartung, J.S. 2005. Biological Characterization of an International Collection of *Citrus tristeza virus* (CTV) Isolates. In: *Proceedings of the 16th Conference of the International Organisation of Citrus Virologists. (Hilf, M.E., Duran-Vila, N., and Rocha-Peña, M.A.)* pp. 75-93. *IOCV.* Riverside, California.

Garnsey, S.M., Civerolo, D.J., Gumpf, R.K., Yokomi, R.K., and Lee, R.F. 1991. Development of a Worldwide Collection of CTV Isolates. In: *Proceedings of the 11th Conference of the International Organisation of Citrus Virologists. (Brlansky, R.H., Lee, R.F., and Timmer, L.W.)* pp. 113-120. *IOCV*. Riverside, California.

Gillings, M., Broadbent, P., Indsto, J., and Lee, R. 1993. Characterisation of Isolates and Strains of Citrus Tristeza Closterovirus Using Restriction Analysis of the Coat Protein Gene Amplified by the Polymerase Chain Reaction. *Journal of Virological Methods* 44(2–3), pp. 305-317.

Gowda, S., Satyanarayana, T., Davis, C.L., Navas-Castillo, J., Albiach-Martí, M.R., Mawassi, M., Valkov, N., Bar-Joseph, M., Moreno, P., and Dawson, W.O. 2000. The p20 Gene Product of *Citrus tristeza virus* Accumulates in the Amorphous Inclusion Bodies. *Virology* 274(2), pp. 246-254.

Harper, S., Dawson, T., and Pearson, M. 2010. Isolates of *Citrus tristeza virus* That Overcome *Poncirus trifoliata* Resistance Comprise a Novel Strain. *Archives of Virology* 155(4), pp. 471-480.

Harper, S.J. 2013. *Citrus tristeza virus*: Evolution of Complex and Varied Genotypic Groups. *Frontiers in Microbiology* 4, pp. 93.

Hermoso, D.A., Ballester-Olmos, J.A., Serra, P.J., and Fuertes, C. 1988. Differences in Transmission Efficiency of *Citrus tristeza virus* by *Aphis Gossypii* Using Sweet Orange, Mandarin or Lemon Trees as Donor or Receptor Hosts Plants. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Timmer, L.W., Garnsey, S.M., and Navarro, L.)* pp. 62-64. *IOCV*. Riverside, California.

Hilf, M.E., Karasev, A.V., Albiach-Marti, M.R., Dawson, W.O., and Garnsey, S.M. 1999. Two Paths of Sequence Divergence in the *Citrus tristeza virus* Complex. *Phytopathology* 89(4), pp. 336-342.

Hilf, M.E., Mavrodieva, V.A., and Garnsey, S.M. 2005. Genetic Marker Analysis of a Global Collection of Isolates of *Citrus tristeza virus*: Characterization and Distribution of CTV Genotypes and Association with Symptoms. *Phytopathology* 95(8), pp. 909-917.

Huang, Z., Rundell, P.A., Guan, X., and Powell, C.A. 2005. Evaluation of the Transmission of Different Field Sources of *Citrus tristeza virus* and the Separation of Different Genotypes by Single Brown Citrus Aphids. *Hortscience* 40(3), pp. 687-690.

Karasev, A.V. 2000. Genetic Diversity and Evolution of Closteroviruses. *Annual Review of Phytopathology* 38(1), pp. 293-324.

Karasev, A.V., Dawson, W.O., Hilf, M.E., and Garnsey, S.M. 1998. Molecular Biology of *Citrus tristeza virus*: Implications for Disease Diagnosis and Control. *ACTA Horticulturae* 472, pp. 333-350.

Kong, P., Rubio, L., Polek, M., and Falk, B.W. 2000. Population Structure and Genetic Diversity within California *Citrus tristeza virus* (CTV) Isolates. *Virus Genes* 21(3), pp. 139-145.

Lee, R.F., Garnsey, S.M., Marais, L.J., Moll, J.N., and Youtsy, C.O. 1988. Distribution of *Citrus tristeza virus* in Grapefruit and Sweet Orange in Florida and South Africa. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Timmer, L.W., Garnsey, S.M., and Navarro, L.)* pp. 33-38. *IOCV.* Riverside, California.

Lin, Y., Brlansky, R.H., and Powell, C.A. 2002. Inefficient Transmission of *Citrus tristeza virus* from Grapefruit by Single Brown Citrus Aphids. *Hortscience* 37(6), pp. 936-939.

Luttig, M., Van Vuuren, S.P., and Van Der Vyver, J.B. 2002. Differentiation of Single Aphid Cultured Sub-Isolates of Two South African *Citrus tristeza virus* Isolates from Grapefruit by Single-Stranded Conformation Polymorphism. In: *Proceedings of the 15th Conference of the International Organization of Citrus Virologists. (Duran-Vila, N., Milne, R.G., and Da Graça, J.V.)* pp. 186-196. *IOCV.* Riverside, California.

Martínez, F., Lafforgue, G., Morelli, M.J., González-Candelas, F., Chua, N.-H., Daròs, J.-A., and Elena, S.F. 2012. Ultradeep Sequencing Analysis of Population Dynamics of Virus Escape Mutants in RNAi-Mediated Resistant Plants. *Molecular Biology and Evolution*.

Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G., and Bar-Joseph, M. 1996. Unusual Sequence Relationships between Two Isolates of *Citrus tristeza virus*. *Journal of General Virology* 77(9), pp. 2359-2364.

Meyer, J.B., Van Vuuren, S.P., Luttig, M., Manicom, B.Q., and Da Graça, J.V. 2005. Strain Prevalence of *Citrus tristeza virus* Cross-Protecting Isolates Altered by Red Grapefruit Hosts. In: *Proceedings of the 16th Conference of the International Organization of Citrus Virologists. (Hilf, M.E., Duran-Vila, N., and Rocha-Peña, M.A.)* pp. 205-212. *IOCV.* Riverside, California.

Moreno, P., Ambrós, S., Albiach-Martí, M.R., Guerri, J., and Peña, L. 2008. *Citrus tristeza virus*: A Pathogen That Changed the Course of the Citrus Industry. *Molecular Plant Pathology* 9(2), pp. 251-268.

Morroni, M., Jacquemond, M., and Tepfer, M. 2013. Deep Sequencing of Recombinant Virus Populations in Transgenic and Nontransgenic Plants Infected with *Cucumber Mosaic Virus*. *Molecular Plant-Microbe Interactions* 26(7), pp. 801-811.

Muniz, F.R., De Souza, A.J., Stipp, L.C.L., Schinor, E., Freitas Jr, W., Harakava, R., Stach-Machado, D.R., Rezende, J.A.M., Mourão Filho, F.A.A., and Mendes, B.M.J. 2011. Genetic Transformation of *Citrus Sinensis* with *Citrus tristeza virus* (CTV) Derived

Sequences and Reaction of Transgenic Lines to CTV Infection. *Biologia Plantarum*. 10.1007/s10535-011-0195-3

Navas-Castillo, J., Albiach-MartÍ, M.R., Gowda, S., Hilf, M.E., Garnsey, S.M., and Dawson, W.O. 1997. Kinetics of Accumulation of *Citrus tristeza virus* RNAs. *Virology* 228(1), pp. 92-97.

Nolasco, G., De Blas, C., Torres, V., and Ponz, F. 1993. A Method Combining Immunocapture and PCR Amplification in a Microtiter Plate for the Detection of Plant Viruses and Subviral Pathogens. *Journal of Virological Methods* 45(2), pp. 201-218.

Olmos, A., Angel Dasí, M., Candresse, T., and Cambra, M. 1996. Print-Capture PCR: A Simple and Highly Sensitive Method for the Detection of *Plum Pox Virus* (PPV) in Plant Tissues. *Nucleic Acids Research* 24(11), pp. 2192-2193.

Olmos, A., Cambra, M., Esteban, O., Gorris, M.T., and Terrada, E. 1999. New Device and Method for Capture, Reverse Transcription and Nested PCR in a Single Closed Tube. *Nucleic Acids Research* 27(6), pp. 1564-1565.

Pappu, H.R., Karasev, A.V., Anderson, E.J., Pappu, S.S., Hilf, M.E., Febres, V.J., Eckloff, R.M.G., McCaffery, M., Boyko, V., Gowda, S., Dolia, V.V., Koonin, E.V., Gumpf, D.J., Cline, K.C., Garnsey, S.M., Dawson, W.O., Lee, R.F., and Niblett, C.L. 1994. Nucleotide Sequence and Organization of Eight 3′ Open Reading Frames of the Citrus Tristeza Closterovirus Genome. *Virology* 199(1), pp. 35-46.

Roistacher, C.N., Da Graça, J.V., and Müller, G.W. 2010. Cross Protection against *Citrus tristeza virus* - a Review. In: *Proceedings of the 17th Conference of the International Organisation of Citrus Virologists. (Hilf, M.E., Timmer, L.W., Milne, R.G., and Da Graça, J.V.)* pp. 1-27. *IOCV*. Riverside, California.

Roistacher, C.N., Dodds, J.A., and Bash, J.A. 1988. Cross Protection against Citrus Tristeza Seedling Yellows and Stem Pitting Viruses by Protective Isolates Developed in Greenhouse Plants. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Garnsey, S.M., Timmer, L.W., and Navarro, L.)* pp. 91-100. *IOCV*. Riverside, California.

Roistacher, C.N., and Moreno, P. 1991. The Worldwide Threat from Destructive Isolates of *Citrus tristeza virus* - a Review. In: *Proceedings of the 11th Conference of the International Organisation of Citrus Virologists. . (Brlansky, R.H., Lee, R.F., and Timmer, L.W.)* pp. *IOCV*. Riverside, California.

Roy, A., Ananthakrishnan, G., Hartung, J.S., and Brlansky, R.H. 2010. Development and Application of a Multiplex Reverse-Transcription Polymerase Chain Reaction Assay for Screening a Global Collection of *Citrus tristeza virus* Isolates. *Phytopathology* 100(10), pp. 1077-1088.

Roy, A., and Brlansky, R.H. 2009. Population Dynamics of a Florida *Citrus tristeza virus* Isolate and Aphid-Transmitted Subisolates: Identification of Three Genotypic Groups and Recombinants after Aphid Transmission. *Phytopathology* 99(11), pp. 1297-1306.

Roy, A., and Brlansky, R.H. 2010. Genome Analysis of an Orange Stem Pitting *Citrus tristeza virus* Isolate Reveals a Novel Recombinant Genotype. *Virus Research* 151(2), pp. 118-130.

Roy, A., Manjunath, K.L., and Brlansky, R.H. 2005. Assessment of Sequence Diversity in the 5′-Terminal Region of *Citrus tristeza virus* from India. *Virus Research* 113(2), pp. 132-142.

Rubio, L., Ayllon, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., and Falk, B.W. 2001. Genetic Variation of *Citrus tristeza virus* Isolates from California and Spain: Evidence for Mixed Infections and Recombination. *The Journal of Virology* 75(17), pp. 8054-8062.

Ruiz-Ruiz, S., Moreno, P., Guerri, J., and Ambrós, S. 2007. A Real-Time RT-PCR Assay for Detection and Absolute Quantitation of *Citrus tristeza virus* in Different Plant Tissues. *Journal of Virological Methods* 145(2), pp. 96-105.

Sambade, A., López, C., Rubio, L., Flores, R., Guerri, J., and Moreno, P. 2003. Polymorphism of a Specific Region in Gene p23 of *Citrus tristeza virus* Allows Discrimination between Mild and Severe Isolates. *Archives of Virology* 148(12), pp. 2325-2340.

Saponari, M., Manjunath, K., and Yokomi, R.K. 2008. Quantitative Detection of *Citrus tristeza virus* in Citrus and Aphids by Real-Time Reverse Transcription-PCR (Taqman®). *Journal of Virological Methods* 147(1), pp. 43-53.

Satyanarayana, T., Gowda, S., Mawassi, M., Albiach Martı, M.R., Ayllón, M.A., Robertson, C., Garnsey, S.M., and Dawson, W.O. 2000. Closterovirus Encoded Hsp70 Homolog and P61 in Addition to Both Coat Proteins Function in Efficient Virion Assembly. *Virology* 278(1), pp. 253-265.

Scott, K.A., Hlela, Q., Zablocki, O., Read, D., van Vuuren, S., and Pietersen, G. 2013. Genotype Composition of Populations of Grapefruit-Cross-Protecting *Citrus tristeza virus* Strain GFMS 12 in Different Host Plants and Aphid-Transmitted Sub-Isolates. *Archives of Virology* 158(1), pp. 27-37.

Silva, G., Marques, N., and Nolasco, G. 2012. The Evolutionary Rate of *Citrus tristeza virus* Ranks among the Rates of the Slowest RNA Viruses. *Journal of General Virology* 93(2), pp. 419-429.

Tatineni, S., Robertson, C.J., Garnsey, S.M., Bar-Joseph, M., Gowda, S., and Dawson, W.O. 2008. Three Genes of *Citrus tristeza virus* Are Dispensable for Infection and Movement Throughout Some Varieties of Citrus Trees. *Virology* 376(2), pp. 297-307.

Van Der Vyver, J.B., Van Vuuren, S.P., Luttig, M., and Da Graça, J.V. 2002. Changes in the *Citrus tristeza virus* Status of Pre-Immunized Grapefruit Field Trees. In: *Proceedings of the 15th Conference of the International Organization of Citrus Virologists. (Duran-Vila, N., Miller, W., and Da Graça, J.V.)* pp. 175-185. *IOCV.* Riverside, California.

Van Vuuren, S.P. 2002. Effects of *Citrus tristeza virus* Isolates on Two Tolerant Commercial Scions on Different Rootstocks in South Africa. In: *Proceedings of the 15th Conference of the International Organisation of Citrus Virologist. (Duran-Vila, N., Milne, R.G., and Da Graça, J.V.)* pp. 31-38. *IOCV.* Riverside, California.

Van Vuuren, S.P., Collins, R.P., and Da Graça, J.V. 1991. The Performance of Exotic *Citrus tristeza virus* Isolates as Preimmunizing Agents for Sweet Orange on Sour Orange Rootstock under Natural Disease Pressure in South Africa. In: *Proceedings of the 11th Conference of the International Organization of Citrus Virologists. (Brlansky, R.H., Lee, R.F., and Timmer, L.W.)* pp. 60-63. *IOCV.* Riverside, California.

Van Vuuren, S.P., and Van Der Vyver, J.B. 2000. Comparison of South African Pre-Immunizing *Citrus tristeza virus* Isolates with Foreign Isolates in Three Grapefruit Selections. In: *Proceedings of the 14th Conference of the International Organisation of Citrus Virologists. (Da Graça, J.V., Lee, R.F., and Yokomi, R.K.)* pp. 50-56. *IOCV.* Riverside, California.

Van Vuuren, S.P., Van Der Vyver, J.B., and Luttig, M. 2000. Diversity among Sub-Isolates of Cross-Protecting *Citrus tristeza virus* Isolates in South Africa. In: *Proceedings of the 14th Conference of the International Organisation of Citrus Virologists. (Da Graça, J.V., Lee, R.F., and Yokomi, R.K.)* pp. 103-110. *IOCV.* Riverside, California.

Vidal, E., Yokomi, R., Moreno, A., Bertolini, E., and Cambra, M. 2012. Calculation of Diagnostic Parameters of Advanced Serological and Molecular Tissue-Print Methods for Detection of *Citrus tristeza virus*: A Model for Other Plant Pathogens. *Phytopathology* 102(1), pp. 114-121.

Vives, M.C., Rubio, L., Lopez, C., Navas-Castillo, J., Albiach-Marti, M.R., Dawson, W.O., Guerri, J., Flores, R., and Moreno, P. 1999. The Complete Genome Sequence of the Major Component of a Mild *Citrus tristeza virus* Isolate. *Journal of General Virology* 80(3), pp. 811-816.

Vives, M.C., Rubio, L., Sambade, A., Mirkov, T.E., Moreno, P., and Guerri, J. 2005. Evidence of Multiple Recombination Events between Two RNA Sequence Variants within a *Citrus tristeza virus* Isolate. *Virology* 331(2), pp. 232-237.

Von Broembsen, L., and Lee, T.C. 1988. South Africa's Citrus Improvement Programme. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Timmer, L.W., Garnsey, S.M., and Navarro, L.)* pp. 407-416. *IOCV.* Riverside, California.

Weng, Z., Barthelson, R., Gowda, S., Hilf, M.E., Dawson, W.O., Galbraith, D.W., and Xiong, Z. 2007. Persistent Infection and Promiscuous Recombination of Multiple Genotypes of an RNA Virus within a Single Host Generate Extensive Diversity. *PLoS ONE* 2(9), pp. e917.

Willerth, S.M., Pedro, H.A.M., Pachter, L., Humeau, L.M., Arkin, A.P., and Schaffer, D.V. 2010. Development of a Low Bias Method for Characterizing Viral Populations Using Next Generation Sequencing Technology. *PLoS ONE* 5(10).

Zablocki, O., and Pietersen, G. 2014. Characterization of a Novel *Citrus tristeza virus* Genotype within Three Cross-Protecting Source GFMS12 Sub-Isolates in South Africa by Means of Illumina Sequencing. *Archives of Virology*, pp. 1-7.

Zanutto, C.A., Corazza, M.J., Nunes, W.M.d.C., and Müller, G.W. 2013. Evaluation of the Protective Capacity of New Mild *Citrus tristeza virus* (CTV) Isolates Selected for a Preimmunization Program. *Scientia Agricola* 70(2), pp. 116-124.

Chapter 2: Comprehensive phylogenetic analysis of *Citrus tristeza virus*

2.1 Introduction

Citrus plants are long-lived and aphids can repeatedly infect the same plant with different strains of CTV (Rubio *et al.*, 2001). This can lead to a CTV population within a plant that consists of many different genotypes of the virus. In many cases, the complete CTV population within the host plant (CTV source) can consist of a mix of strains from different genotypes and defective RNAs (dRNAs). Sometimes however a CTV source may consist of one genotype which is dominant with low levels of strains from other genotypes (Albiach-Marti *et al.*, 2000; Kong *et al.*, 2000). When transmissions are done from these sources, the subsequent sub-isolates may have a different composition to the original CTV source (Černi *et al.*, 2008).

Not only can a CTV population be diversified due to infection of plants by additional genotypes through aphids, but a virus population may also be shaped by internal processes. A CTV population is a quasispecies that consists of genetically related variants of the virus. This is caused by the error prone RNA polymerase, large population sizes and short replication times of a RNA virus (Rubio *et al.*, 2001). Since recombination also takes place in the CTV genome (Roy and Brlansky, 2010; Rubio *et al.*, 2001; Vives *et al.*, 2005), and due to the fact that recombination is not just limited to similar variants of the virus (Rubio *et al.*, 2001), a very diverse population can occur within a CTV source. However in spite of a high occurrence of recombination and mutations within the CTV genome, it is the most slowly evolving plant RNA virus known when based on the capsid protein gene sequence (Silva *et al.*, 2012).

Due to the great variation that can exist within a CTV population, it is necessary to characterize the genotypic composition of CTV sources in order to both identify genotypes present in a specific mild strain cross-protection source as well as to determine which genotypes are circulating and need to be protected against. There are several techniques that

can be used to characterize a CTV population. Single strand confirmation polymorphism (SSCP), which relies on nucleotide differences between same sized DNA fragments to form different banding patterns when denatured and separated by gel electrophoreses (Luttig *et al.*, 2002), is commonly used to differentiate between different genotypes of CTV. Some genotypes can however not be differentiated from each other with this technique (Gago-Zachert *et al.*, 1999). Restriction fragment length polymorphism (RFLP) also allows for the differentiation of different genotypes based on the banding patterns on an agarose gel. But with this method, a gene of interest (for example the coat protein gene) is subjected to RT-PCR and then the amplicons are digested with restriction enzymes to create different gel profiles when separated by gel electrophoresis (Gillings *et al.*, 1993).

Specific amplicons of the genome can also be Sanger sequenced directly or cloned and multiple clones sequenced, followed by phylogenetic analysis (Martínez *et al.*, 2012; Morroni *et al.*, 2013; Scott *et al.*, 2013). Next generation sequencing has made it more affordable and less labor intensive to sequence full genomes.

The characterization of CTV based on sequence data, encompassing genome regions and techniques used, differ tremendously among different studies making comparisons between these difficult. The use of Neighbor joining dendrograms is often used, but in some studies maximum likelihood analyses is used. The genome segments analyzed include different portions within ORF 1a as well as the coat protein, p23, p20 and complete genome sequences (Davino *et al.*, 2013; Harper *et al.*, 2010; Roy *et al.*, 2005; Rubio *et al.*, 2001; Sambade *et al.*, 2003; Scott *et al.*, 2013). ORF 1a is translated into a single 400 kDA polyprotein that is cleaved by two virus specific leader proteases to yield 4 proteins: methyltransferase, helicase and two papain-like proteases (Karasev, 2000). Only certain areas within the 9300 bp ORF 1a have been used for characterization. This fragment lies within the 5' half of the genome

which is the more divergent half of the genome between different strains (Mawassi *et al.*, 1996), and thus gives more variation in genotyping results.

The other commonly used proteins for characterization includes, but is not restricted to, the coat protein which is encoded by ORF 7 and covers 95% of the virion (Febres *et al.*, 1996), p20 which is encoded by ORF11 and acts as a silencing suppressor (Lu *et al.*, 2004) and p23 which is encoded by ORF 11 and may have several functions including acting as a RNA silencing suppressor (Lu *et al.*, 2004), playing a role in RNA binding (López *et al.*, 2000), enhancing systemic infection, viral movement (Fagoaga *et al.*, 2011) and symptom development (Fagoaga *et al.*, 2005; Ghorbel *et al.*, 2001). All of these genes are located in the 3' half of the genome which is the more conserved part of the genome.

With an increasing number of whole genome sequence data becoming available for CTV, it is important to have a strategy as how to analyze the large amount of data. Unfortunately there is no rule book with guidelines how these analyses should be done and this leads to a considerable variation in the results obtained. This makes it difficult to compare data from different studies and to get a full rounded view of the diversity of CTV. The aim of this study was to analyze different areas of the CTV genome using phylogeny to find which genes most closely approximate genotyping based on whole genome sequences.

Within literature, there are different uses for 'strain', 'isolate' and 'genotype'. In this study we consider a sample of CTV, as a 'source' which may contain many different variants of the virus. Once the sequence of these variants are characterized individually they are defined as an 'isolate', while they are referred to as a 'strain' if the biological properties have been determined, for instance, if it is known whether it is mild or severe form of the virus. A variant of the virus may therefore be referred to as either an isolate or a strain, and use of either term is based on the context of the discussion. When it is regarding biological properties 'strain' will be used and when it is regarding the composition of the genome

'isolate' will be used. 'Genotype' will refer to a group of isolates that are similar based on their RNA sequence.

2.2 Materials and methods

2.2.1 Sequences and alignment

All complete CTV genomes available on GENBANK (Table 1) were downloaded and used to extract all ORFs, the two UTRs, as well as the protein domains within ORF 1a (Table 2). This was done by aligning all of the genomes, including those already annotated in CLC Main Workbench 6 (CLC Bio, Denmark) and then selecting the specific areas of interest for all the genomes and creating new sequence lists. Gaps created during alignment were removed and sequences aligned using MAFFT online (Katoh *et al.*, 2002) with the "strategy" set to auto, "scoring matrix for amino acid sequences" set to BLOSUM62, "scoring matrix for nucleotide sequences" set to 200PAM/k=2, "align unrelated segments" set to try to align gappy regions anyway, "number of homologs" set to 50, "threshold" set to 1a-10 and "plot last hit threshold" set to score=39 (E=8.4e11).

For amino acid sequences, the aligned FASTA files were opened in BioEdit (Hall, 1999) and translated to amino acids. The alignment and reading frame was checked before saving the files in the necessary formats.

2.2.2 Nucleotide and amino acid substitution model selection

Nucleotide sequence alignments were analyzed in jModelTest version 2.1.3 (Darriba *et al.*, 2012) using the Akaike information criterion model. Models used for sequences are indicated in Table 3. ProtTest version 3.3 (Abascal *et al.*, 2005) was used to determine the best fit amino acid substitution models. This was also done using the Akaike information criterion model. These results are summarized in Table 3.

Table 2.1: List of whole genome sequences and their Genbank accessions used for analyses in this study.

Name	Accession Number	Description
AB046398	AB046398.1	CTV genomic RNA, complete genome, seedling yellows strain
AF001623	AF001623.1	CTAF001623 CTV, complete genome
AF260651	AF260651.1	CTV T30, complete genome
AY170468	AY170468.1	CTV, complete genome
AY340974	AY340974.1	CTV isolate Qaha from Egypt, complete genome
DQ151548	DQ151548.1	CTV strain T318A, complete genome
DQ272579	DQ272579.1	CTV from Mexico, complete genome
EU076703	EU076703.3	CTV isolate B165, complete genome
EU857538	EU857538.1	CTV strain SP, complete genome
EU937519	EU937519.1	CTV strain VT, complete genome
EU937520	EU937520.1	CTV strain T30, complete genome
EU937521	EU937521.1	CTV strain T36, complete genome
FJ525431	FJ525431.1	CTV isolate NZRB-M12, complete genome
FJ525432	FJ525432.1	CTV isolate NZRB-G90, complete genome
FJ525433	FJ525433.1	CTV isolate NZRB-TH28, complete genome
FJ525434	FJ525434.1	CTV isolate NZRB-TH30, complete genome
FJ525435	FJ525435.1	CTV isolate NZRB-M17, complete genome
FJ525436	FJ525436.1	CTV isolate NZ-B18, complete genome
GQ454869	GQ454869.1	CTV strain HA18-9, complete genome
GQ454870	GQ454870.1	CTV strain HA16-5, complete genome
HM573451	HM573451.1	CTV isolate Kpg 3, complete genome
JF957196	JF957196.1	CTV isolate B301, complete genome
JQ061137	JQ061137.1	CTV isolate AT-1, complete genome
JQ798289	JQ798289.1	CTV isolate A18, complete genome
JQ911663	JQ911663.1	CTV isolate CT14A, complete genome
JQ911664	JQ911664.1	CTV isolate CT11A, complete genome
JQ965169	JQ965169.1	CTV isolate T68-1, complete genome
JX266712	JX266712.1	CTV isolate Taiwan-Pum/SP/T1, complete genome
JX266713	JX266713.1	CTV isolate Taiwan-Pum/M/T5, complete genome
KC262793	KC262793.1	CTV isolate L192GR, complete genome
KC333868	KC333868.1	CTV isolate CT-ZA3, complete genome
KC517485	KC517485.1	CTV isolate FS674-T36, complete genome
KC517486	KC517486.1	CTV isolate FS701-T36, complete genome
KC517487	KC517487.1	CTV isolate FS703-T36, complete genome
KC517488	KC517488.1	CTV isolate FS577, complete genome
KC517489	KC517489.1	CTV isolate FS701-T30, complete genome
KC517490	KC517490.1	CTV isolate FL278-T30, complete genome
KC517491	KC517491.1	CTV isolate FS703-T30, complete genome
KC517492	KC517492.1	CTV isolate FS703-VT, complete genome
KC517493	KC517493.1	CTV isolate FL202-VT, complete genome
KC517494	KC517494.1	CTV isolate FS701-VT, complete genome
KC525952	KC525952.1	CTV isolate T3, complete genome
NC001661	NC001661.1	CTV, complete genome
U16304	U16304.1	CTU16304 CTV complete genome
Y18420	Y18420.1	CTV complete genome, isolate T385

Table 2.2: Regions of the CTV genome used for analyses and their products. The region description for fragment 1-15 indicates the position within the alignment file of the whole genomes of the reference isolates listed in Table 1. For fragment 16-22 the region indicates the position within ORF 1a or the p349 protein of a representative CTV isolate (accession number AE199958.1). UTR, Untranslated region. ORF, Open reading frame.

	Fragment	Region	Product
1	5' UTR	1..113	Untranslated
2	ORF 1a	114..9599	"p349"
3	ORF 1b	9390..10968	"RdRp"
4	ORF 2	11017..11928	"p33"
5	ORF 3	12003..12158	"p6"
6	ORF 4	12164..13948	"p65"
7	ORF 5	13872..15491	"p61"
8	ORF 6	15457..16179	"p27"
10	ORF 7	16279..16951	"p25"
11	ORF 8	16917..17420	"p18"
12	ORF 9	17454..17813	"p13"
13	ORF 10	17889..1844	"p20"
14	ORF 11	18523..19155	"p23"
15	3' UTR	19156..19433	Untranslated
16	DUF3648	1..131	"Unknown function"
17	DUF 3762	211..290	"Unknown function"
18	Protease II	397..486	"Peptidase C42"
19	DUF 3614	646..806	"Unknown function"
20	Protease I	882..970	"Peptidase C42"
21	Methyltransferase	1029..1367	"Vmethyltransf""
22	Helicase	2708..2976	"Viral_helicase"

Table 2.3: Nucleotide and amino acid substitution models used for data sets. UTR, Untranslated region. ORF, Open reading frame. GTR, Generalized time reversible (Tavaré 1986). TVM, transversion model. TPM, three parameter model (Kimura 1989). HKY, (Hasegawa, Kishino and Yano 1985). TrN, (Tamura and Nei 1993). TIM, Transition model (Posada 2003). JTT, (Jones 1992). FLU, (Dang 2010). HIVw/b, HIV within or HIV between (Nickle 2007). +I, invariable sites. +G, rate variation among sites. +f, frequencies. *ef*, equal frequencies. *uf*, unequal frequencies.

Fragment	Nucleotide substitution model	Amino acid substitution model
5' UTR	GTR+G	
ORF 1a	TVM+I+G	JTT+G+F
ORF 1b	GTR+I+G	JTT+G
ORF 2	TPM1*uf*+G	FLU+G+F
ORF 3	HKY+G	HIVb
ORF 4	TPM2*uf*+I+G	HIVb+G+F
ORF 5	TPM1*uf*+G	HIVw+G+F
ORF 6	TrN+I+G	HIVw+I+G+F
ORF 7	GTR+G	HIVb+G
ORF 8	HKY+I	HIVw+I+F
ORF 9	TIM3+I	FLU+G
ORF 10	TVM+G	JTT+G
ORF 11	HKY+G	FLU+G
3' UTR	TPM2*uf*+G	
DUF3648	TIM2*ef*+I	HIVw+G
DUF 3762	TPM3*uf*+G	HIVb
Protease II	TPM1uf+G	JTT+G
DUF 3614	TVMef+I+G	HIVb+G
Protease I	TPM1uf+G	FLU+G+F
Methyltransferase	TVM+I+G	HIVb+G+F

Helicase	TIM2+I	HIVb+G+F
Complete	TVM+I+G	JTT+I+G+F

2.2.3 Maximum likelihood analyses

Maximum likelihood analyses for all data sets (nucleotide and amino acid) were performed in PhyML version 3.0_360.500M (Guindon and Gascuel, 2003). Default parameters were used except for the branch support which was altered to a 1 000 bootstrap replications. Furthermore, the topology of dendrogram searching was based on a best of both approach (nearest-neighbor interchange (NNI) and sub-dendrogram prune and regraft (SPR)) and the appropriate substitution model was used (Table 2.3).

2.2.4 Bayesian analysis

The Bayesian analyses were done for data sets (all nucleotide alignments and an amino acid alignment of the complete genome) in MrBayes version 3.2.1 (Ronquist *et al.*, 2012) using appropriate substitution models (Table 2.3), with 4 Markov chains and at least 3 million generations and repeating the analysis twice. P files were analyzed in Tracer version 1.5 (Rambaut and Drummond, 2009) to determine the burn in value. After burn in, dendrogram files were summarized and dendrograms viewed in FigDendrogram version 1.3.1 (Rambaut, 2009). Remaining amino acid data sets were not analyzed in MrBayes since these datasets takes very long to process.

2.2.5 Neighbor joining analysis and pair wise distance calculations

Neighbor joining analysis for all data sets (nucleotide and amino acid) were performed in MEGA 6 (Tamura *et al.*, 2007). The p-distance model was used with a bootstrap analysis of 1000 replicates. Pair wise distance calculation was performed on the alignment file of all the complete genomes for both the nucleotide and amino acid data.

2.2.6 Concatenation of data

Based on the results of individual dendrograms, some datasets were concatenated and reanalyzed. A partition homogeneity test was performed on all combination data sets to test

whether the fragments combined could be analyzed under the same substitution model. If the test provide a P value > 0.05 the datasets can be combined. The combinations of datasets are described in Table 2.4, and all combinations were analyzed in MrBayes as partitioned files that allowed for the specification of nucleotide or amino acid substitution model for each partition.

Table 2.4: Fragments of the CTV genome combined for phylogenetic analysis. UTR, Untranslated region. ORF, Open reading frame.

Combination	Fragments
Combo 1	5' UTR, ORF 2, ORF 7, Protease I and DUF3614
Combo 2	5' UTR, ORF 2 and ORF 7
Combo 3	5' UTR and ORF 2
Combo 4	5' UTR and ORF 7
Combo 5	ORF 2 and ORF 7
Combo 6	Protease I and DUF3614
Combo 7	ORF 2, ORF 4, ORF7, Protease I, Protease II, DUF 3614 and Methyltransferase
Combo 8	ORF 2, ORF 4, and ORF 7
Combo 9	ORF 2 and ORF 4
Combo 11	ORF 4 and ORF 7
Combo 12	Protease I, Protease II, DUF3614 and Methyltransferase
Combo 13	DUF 3614 and Methyltransferase
Combo 14	ORF 2 and Methyltransferase
Combo 15	ORF 2 and DUF 3614
Combo 22	ORF 1b, ORF 2, ORF 5, DUF 3614, Protease II, Methyltransferase and Helicase
Combo 23	ORF 1b, ORF 2 and ORF 5
Combo 25	ORF 1b and ORF 5
Combo 26	ORF 2 and ORF 5
Combo 27	DUF 3614, Protease II, Methyltransferase and Helicase
Combo 28	ORF 2 and Helicase
Combo 29	DUF 3614 and Helicase

2.2.7 Pairwise Distance analysis

To compute the pairwise distances, the alignment file for all 45 complete genomes were analyzed in MEGA 6 using all default parameters. This was done for nucleotides and amino acids.

2.2.8 Detection of recombination

The alignment file of all complete genomes was analyzed using the RDP4 program (Heath *et al.*, 2006) to detect recombination. All methods available in the program were used to detect recombination signals, and these include RDP, GENECONV, BOOTSCAN/RESCAN, MAXCHI, CHIMAERA, SISCAN, 3Seq, LARD, Topal/DSS, PHYLPRO and VisRD. All

parameters were left on the default values, except for selecting “step down correction” under General and changing the VisRD window size to 50. All recombination events detected by three or more programs were investigated closer to establish if it was a true recombinant and whether the minor and major parents were identified correctly.

2.3. Results and discussion

The aim of this study was to find a method of phylogenetic analysis and a specific region(s) of the genome that can be used to classify isolates of CTV into different genotypes. To reach this goal, it was necessary to look at the complete genome dendrograms based on the three different methods used, and determine which method groups the isolates into different genotypes with the best branch support. From there one can seek out the genome fragment(s) that can be used in phylogenetic analysis to best approximate that of the whole genome which can then be used in future to determine the genotype of isolates without needing to sequence the entire genome.

In initial analyses, 8 isolates representing a member of each currently accepted CTV genotype were used as the focus points to compare the basic structure of the dendrogram. These representative isolates are EU937520 (T30), EU076703 (B165), EU937519 (VT), EU937521 (T36), FJ525431 (RB), GQ454870 (HA16-5), JQ965169 (T68) and KC525952 (T3). Branch support was used with each dendrogram to see how many different clades can be identified. Clades being potential genotypes. A bootstrap value lower than 700 (in the case of Maximum likelihood dendrograms) or 0.7 (in the case of Neighbor joining dendrograms) and a posterior probability lower than 0.9 (in the case of Bayesian dendrograms) was considered as too little evidence and the branch would then be ‘collapsed’, grouping all isolates in sub branches together in a single clade. Secondly the dendrograms were inspected to see where the 8 representative isolates were located, whether they resolved

into separate clades and how they related to other isolates. If the branch support was high enough to allow the distinct grouping of most of the accepted genotypes in different clades, the positions of the remaining 37 isolates were investigated and compared with dendrograms obtained using different algorithms.

The three dendrograms based on Maximum likelihood (ML), Bayesian (BA) and Neighbor joining (NJ) methods (Figure 2.1 - 2.3) for complete CTV genomes are similar to each other but have some differences. The close relationship between T30, T3 and VT are apparent in all three dendrograms, but only the Bayesian method allows for branch support that is high enough to retain the three isolates in different clades. Neither the NJ nor the ML dendrograms offer high enough branch support and hence the individual clades are collapsed. In all three dendrograms the RB and T36 isolates are closely related to each other, but form two distinct clades. T68, B165 and HA16-5 form individual clades on all the dendrograms. With BA and NJ, the relationship between the B165, T68 and HA16-5 clades appeared much closer than with ML. When analyzing the position of the other isolates within the dendrogram, their resulting positions were similar in each dendrogram, but with minor differences. In the ML dendrogram one of the isolates (KC333868) that group with the T68 clade within the other two dendrograms, forms a separate clade. Four isolates that group together in the NJ and ML dendrograms (JQ061137, AB046398, DQ151548 and JQ911664) and fall within the T3/T30/VT clade, group as individuals in the case of JQ061137 and AB046398, and together as a separate clade in the case of DQ151548 and JQ911664 within the BA dendrogram. All group near the T30, T3 and VT clades, but JQ061137 groups closer to the T36, RB, HA16-5, B165 and T68 clades. In all three dendrograms it is very apparent that there is a distinct separation between the RB and T36 clades and the T30, T3 and VT clades. When B165, T68 and HA16-5 are considered, their position in relation to these two parts of the dendrograms is more variable.

When investigating the dendrograms constructed from the complete genome amino acid alignments, the ML dendrogram can be discarded as the branch support is too low to support any conclusions. The branch support for the NJ dendrogram (Figure 2.4) is much better and individual clades can be identified. Although branch support for the separation between the VT and T30 isolates is low, one can distinguish between clades for T3, B165, HA16-5, T68, RB and T36. In the amino acid dendrogram, the HA16-5 and T68 isolates are more closely related to each other, whereas the B165 isolate is more closely related to the T3/T30/VT clade. The amino acid BA dendrogram yields a similar result (Figure 2.5), with the T30 and VT isolates again forming one clade due to insufficient branch support to separate them, but in this instance, the T3 isolate is also collapsed into this clade. One would expect the amino acid dendrogram of the entire genome to be a true representative of the relationship between the different viruses, since the differences among sequences would be definitive, since differences in amino acids would result in differences in characteristics of proteins and hence differences in the biological characteristics of the virus. Differences on the nucleotide level however do not necessarily reflect differences in the virus because of the degenerate nature of the genetic code.

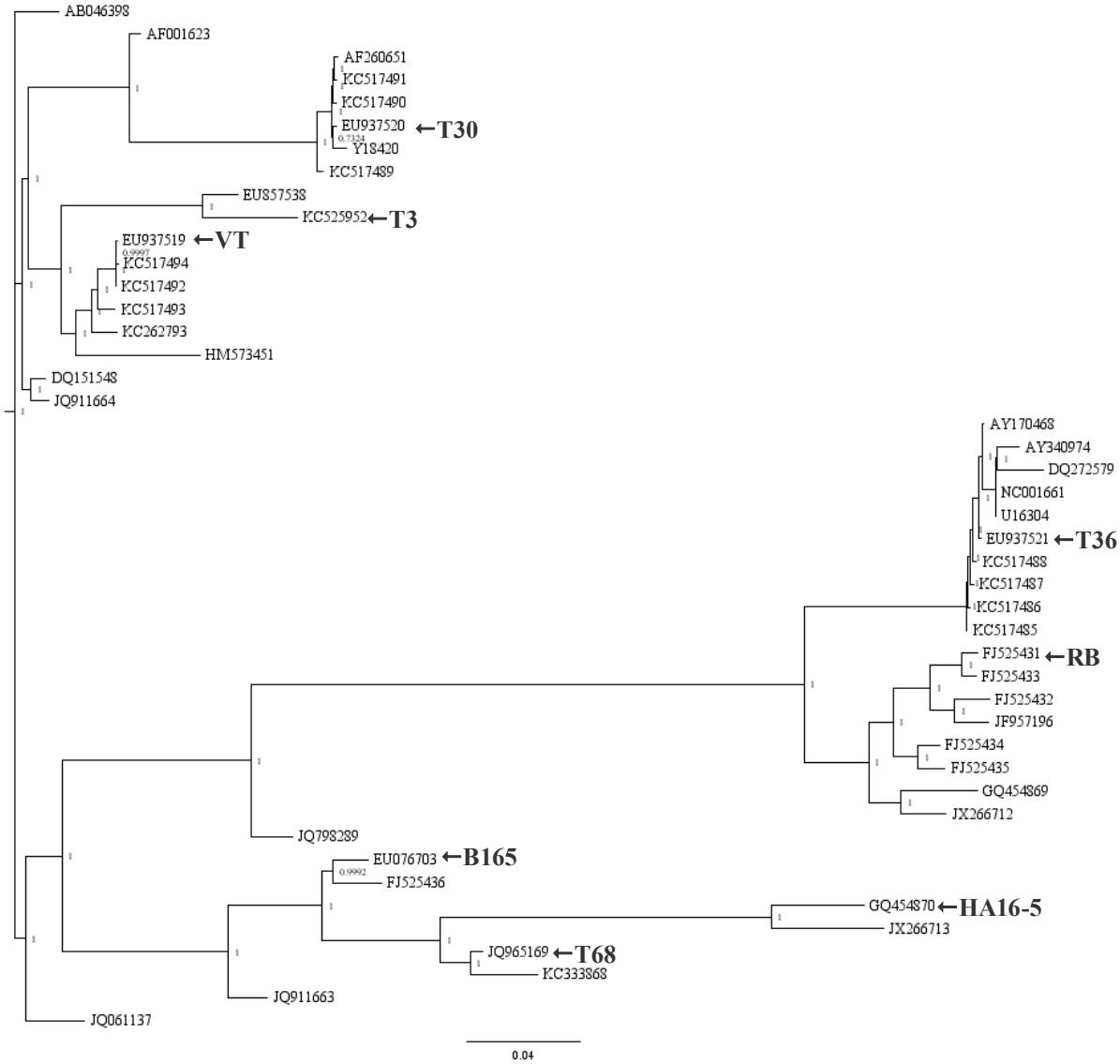

Figure 2.1: Phylogenetic dendrogram of the complete nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

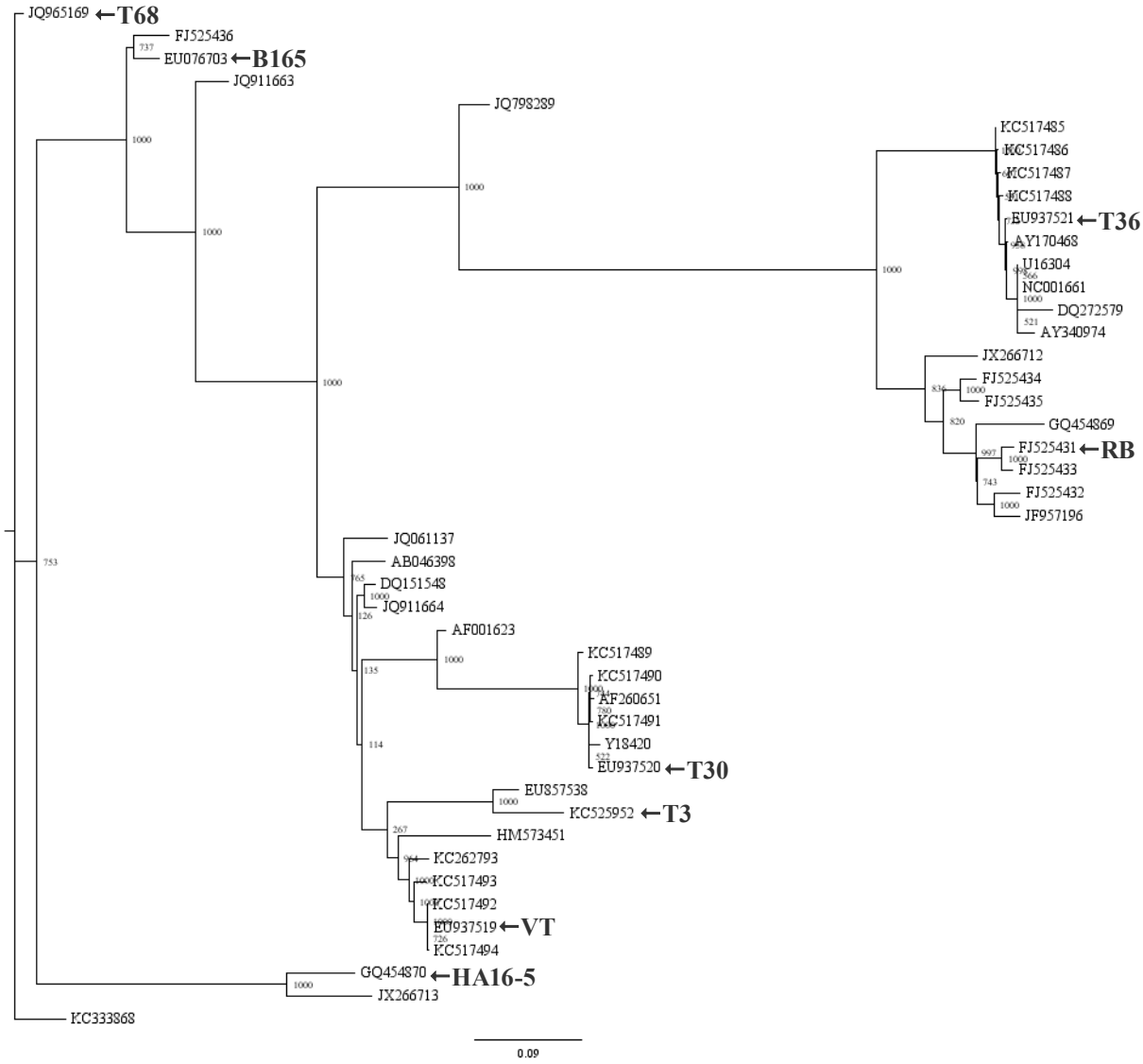

Figure 2.2: Phylogenetic dendrogram of the complete nucleotide sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

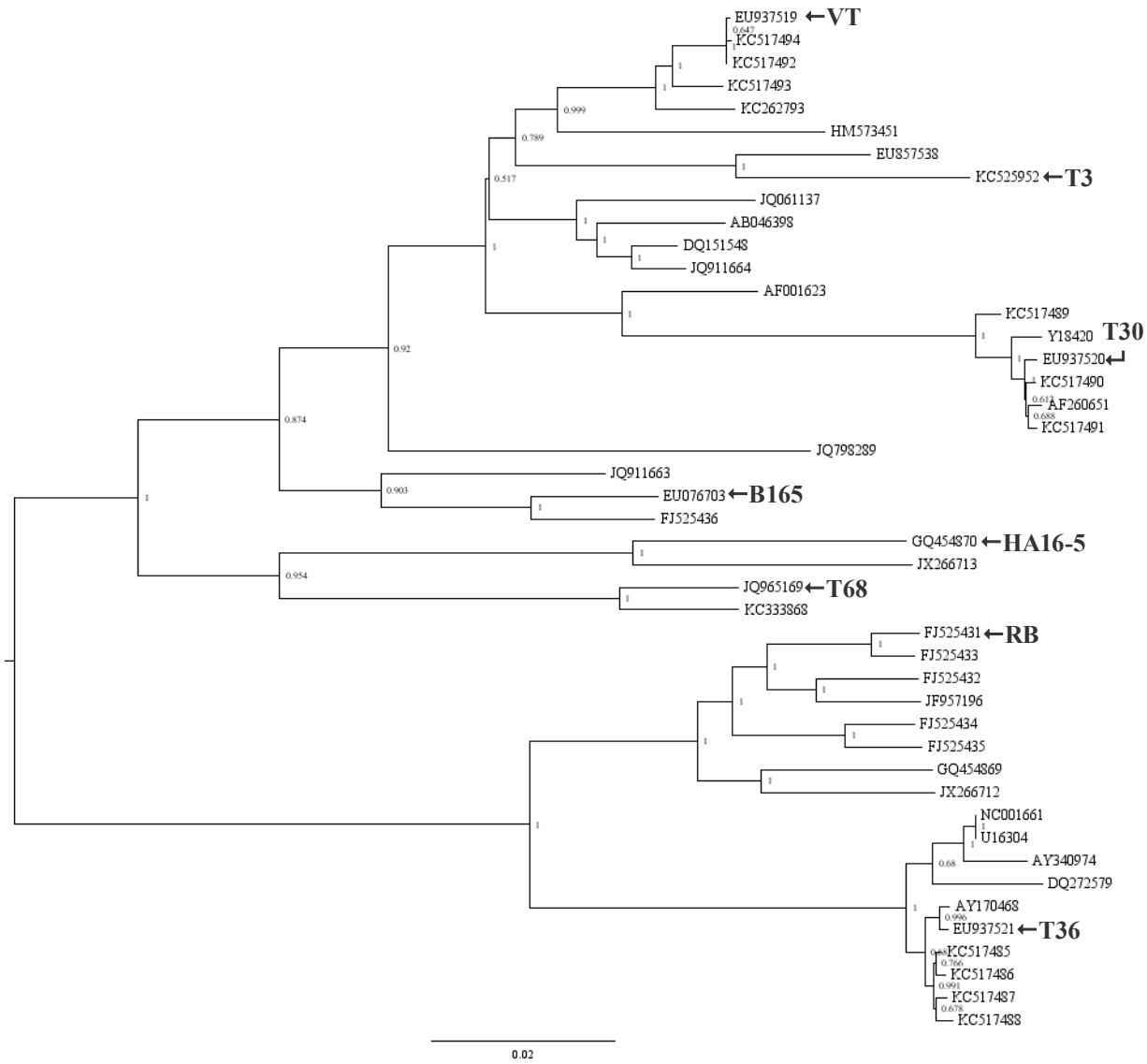

Figure 2.3: Phylogenetic dendrogram of the complete nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

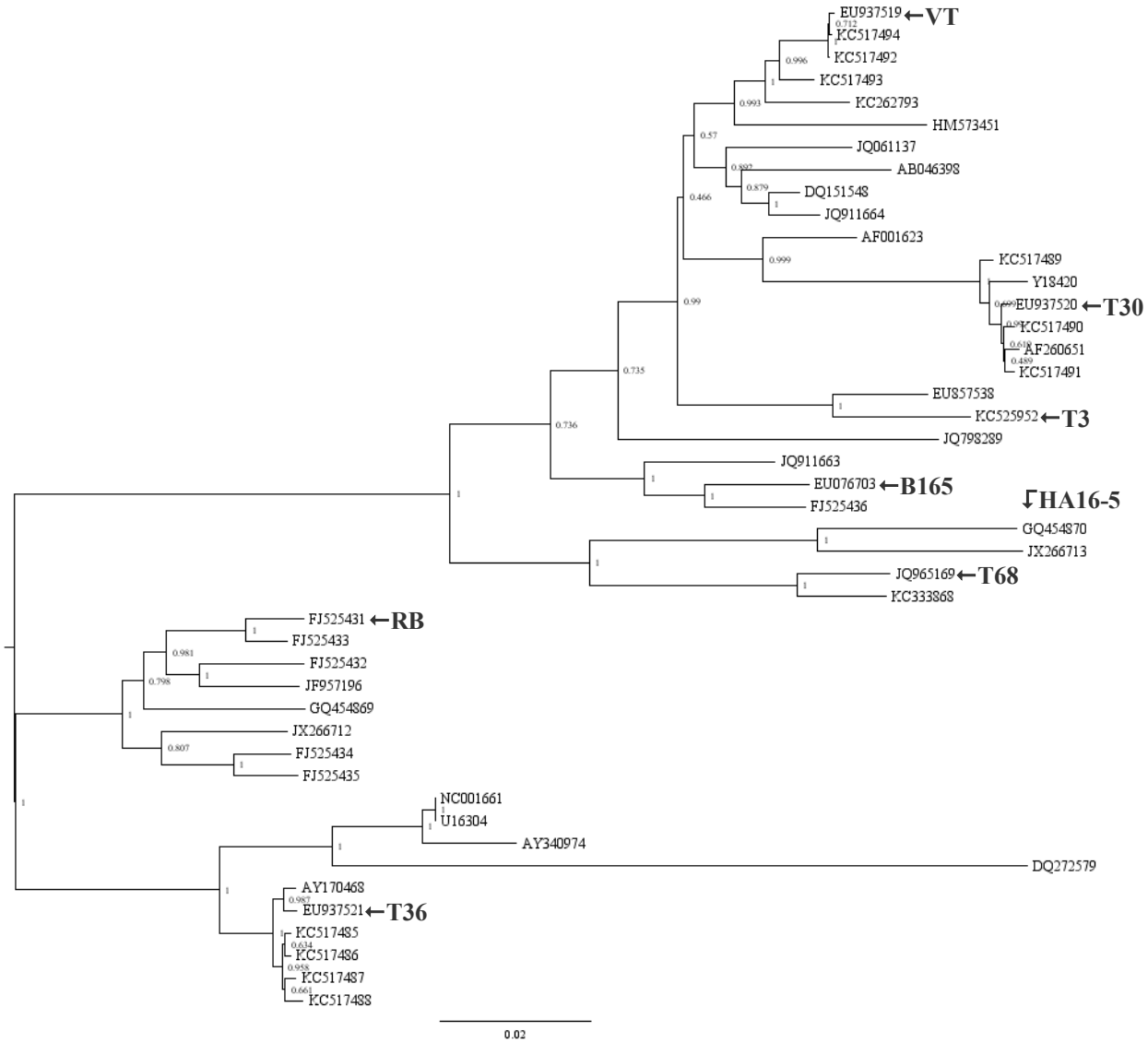

Figure 2.4: Phylogenetic dendrogram of the complete amino acid sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

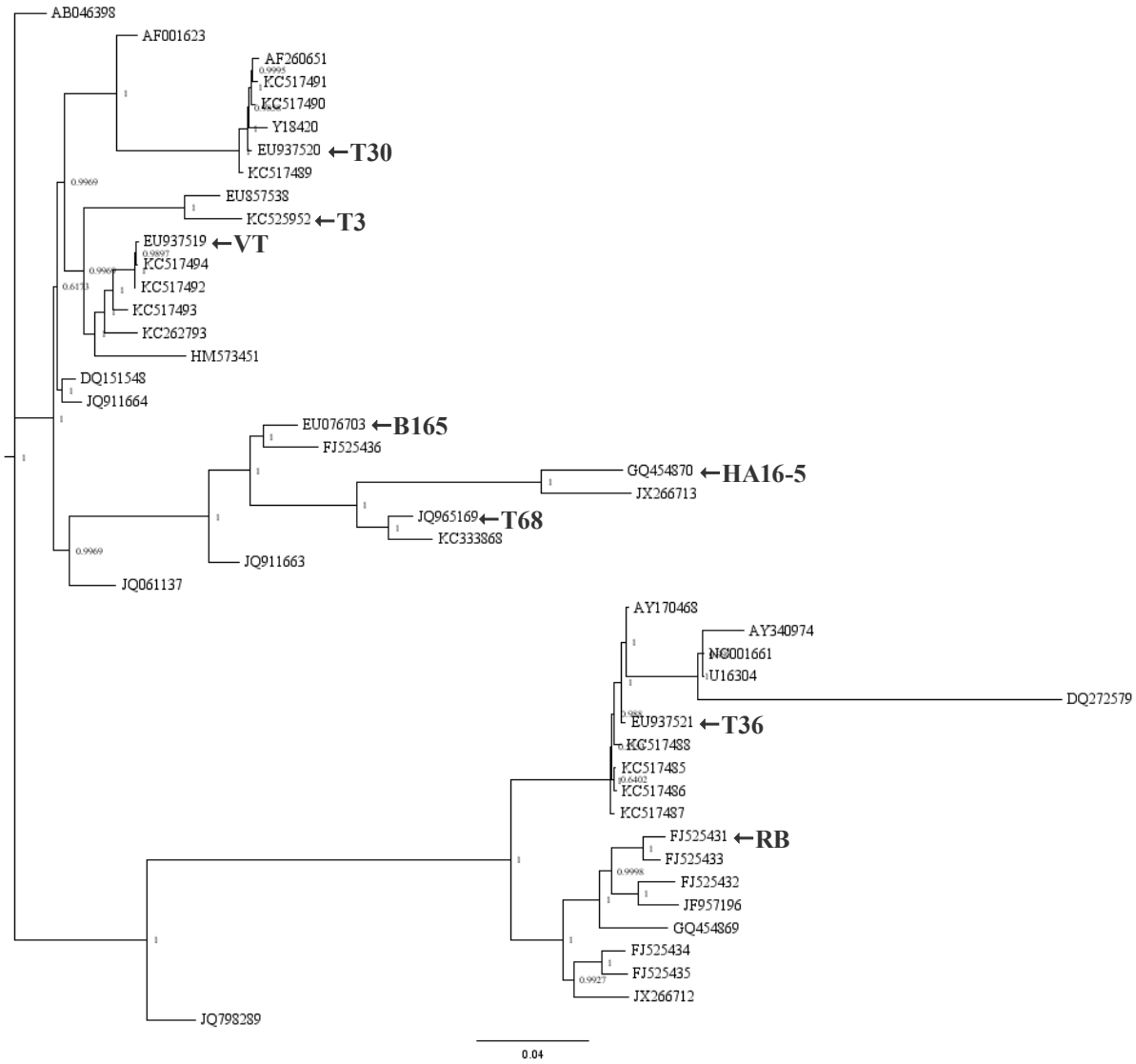

Figure 2.5: Phylogenetic dendrogram of the complete amino acid sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

If one accepts that the amino acid based phylogenies of the complete genome of CTV shows the most biologically relevant relationship between isolates, and given the difficulty of generating whole genome sequences of all sources, then it is clear that it is necessary to find a smaller, easily sequenced region of the genome which displays similar relationships, making it more feasible to use for genotyping.

Phylogenetic dendrograms constructed from various ORFs and protein domains were analyzed, and the number of well-supported individual clades determined and is summarized in Table 2.5. A more comprehensive summary on this analysis is provided in Table A.1 Appendix A. A summary of the positions of isolates within the dendrograms constructed from different gene fragments are shown in Table 2.6 and a more complete version of this table can be found in Table C.1, Appendix C. While the general topology of dendrograms constructed from the same data set using different phylogenetic analysis methods were similar in most cases, the branch support across the methods varied considerably.

Excluding the complete genome dendrograms, there were 101 dendrograms produced from the nominated segments of the genome. Of these, only 7 separated the representative accepted genotypes used from each other (except for some closely related isolates in some instances) and these are indicated in Table 2.5 with an *. Due to low branch support on some clades (less than 5 clades could be distinguished), 54 of the generated dendrograms were not analyzed further in this study. These included dendrograms constructed from the ORF 3, ORF 9, ORF 11 and DUF 3762 sequences which all lacked the branch support to group the isolates into different groups. This may be due to the low amount of nucleotide diversity between these segments of the genome. The remaining dendrograms were evaluated based on their individual morphology, and if found relevant, are discussed below.

Table 2.5: Summary of amount of clades obtained while doing phylogenetic analysis on the CTV genome. Fragments that provided useful dendrograms and that are discussed are bolded and underlined. Dendrograms separating most of the representative isolates used in the discussion are marked with a *.

	Number of clades				
	Nucleotide			Amino acids	
Fragment	**Maximum Likelihood**	**Neighbor Joining**	**Bayesian**	**Maximum Likelihood**	**Neighbor Joining**
Complete	9	10*	14*	3	11*
5' UTR	2	3	**8**	n/a	n/a
ORF 1a	**7**	**4**	**14***	3	8
ORF 1b	4	**7***	3	**6**	4
ORF 2	**15***	**10***	**11**	10	9
ORF 3	3	3	3	2	1
ORF 4	**10***	**5**	3	**7**	4
ORF 5	3	**8***	2	3	**6**
ORF 6	4	**10***	3	3	1
ORF 7	**6**	3	**6**	2	1
ORF 8	3	**7**	6	5	1
ORF 9	4	3	3	3	1
ORF 10	3	**6**	3	3	2
ORF 11	2	4	3	3	3
3' UTR	1	2	**10**	n/a	n/a
DUF 3648	4	**7**	2	2	**7**
DUF 3762	4	4	3	3	3
Protease I	5	**6**	5	5	4
DUF 3614	**8**	**6**	**7**	4	7
Protease II	5	**6**	3	5	1
Methyl-transferase	**10**	**6**	3	5	3
Helicase	2	**7**	2	3	**8**

Table 2.6: A summary of the position of isolates in the dendrograms constructed from different genome fragments in relevance to the representative isolates. Only dendrograms discussed in text are shown in the table, a more complete table can be found in Appendix C. Table shows positions of isolates in nucleotide (NT) and amino acid (AA) dendrograms from the three different methods used, Bayesian (BA), Maximum likelihood (ML) and Neighbor joining (NJ). Blank spaces indicate where isolate did not group with one of the representative isolates. Hierarchy of grouping of representatives is VT, T36, RB, T30, B165, HA16-5, T3, and T68.

	Complete					5' UTR	ORF 1a			ORF 1b		ORF 2		
	NT			AA		NT	NT			NT	AA	NT		
	BA	ML	NJ	ML	NJ	BA	BA	ML	NJ	NJ	ML	BA	ML	NJ
AB046398		VT	VT	VT	VT			VT	VT	VT	VT		VT	VT
AF001623		VT		VT	VT			VT	VT	VT	VT	T30	T30	T30
AF260651	T30	VT	T30	VT	VT	T30	T30	VT	VT	VT	VT	T30	T30	T30
AY170468	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
AY340974	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36		T36
DQ151548		VT	VT	VT	VT			VT	VT	VT	VT		VT	VT
DQ272579	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36		T36
EU076703	B165	B165	B165	VT	B165	B165	B165	B165	B165	B165	B165	B165	B165	B165
EU857538	T3	VT	VT	VT	T3		T3	VT	VT	VT	VT	16-5	16-5	16-5
EU937519	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
EU937520	T30	VT	T30	VT	VT	T30	T30	T30	T30	VT	VT	T30	T30	T30
EU937521	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
FJ525431	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
FJ525432	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
FJ525433	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
FJ525434	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
FJ525435	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
FJ525436	B165	B165	B165	VT	B165	B165	B165	B165	VT	B165	B165	B165	B165	B165
GQ454869	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
GQ454870	16-5	16-5	16-5	VT	16-5	T36	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5
HM573451	VT	VT	VT	VT	VT	VT	VT	VT	VT		VT	16-5	16-5	16-5
JF957196	RB	RB	RB	VT	RB	T36		RB	RB	RB	RB	RB	RB	RB
JQ061137		VT	VT	VT	VT	VT		VT	VT	VT	VT			B165
JQ798289				VT				VT						16-5
JQ911663			B165	VT		B165		VT	VT	VT	VT	B165	B165	B165
JQ911664		VT	VT	VT	VT			VT	VT	VT	VT		VT	VT
JQ965169	T68	T68	T68	VT	T68	T68	T68	T68	T68	B165	B165	B165	B165	B165
JX266712	RB	RB	RB	VT	RB	T36	RB	RB	RB	RB	RB	RB	RB	RB
JX266713	16-5	16-5	16-5	VT	16-5	T36	16-5	16-5	VT	16-5	16-5			
KC262793	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC333868	T68	T68	T68	VT	T68	B165	T68	T68	VT	B165	B165	VT	VT	VT
KC517485	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517486	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517487	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517488	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517489	T30	VT	T30	VT	VT	T30	T30	VT	VT	VT	VT	T30	T30	T30
KC517490	T30	VT	T30	VT	VT	T30	T30	VT	VT	VT	VT	T30	T30	T30
KC517491	T30	VT	T30	VT	VT	T30	T30	VT	VT	VT	VT	T30	T30	T30
KC517492	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC517493	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC517494	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC525952	T3	VT	VT	VT	T3	T3	T3	T3	T3	VT	VT	T3	16-5	16-5
NC001661	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36		T36
U16304	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36		T36
Y18420	T30	VT	T30	VT	VT	T30	T30	VT	VT	VT	VT	T30	T30	T30

Table 2.6 (continued): A summary of the position of isolates in the dendrograms constructed from different genome fragments in relevance to the representative isolates. Only dendrograms discussed in text are shown in the table, a more complete table can be found in Appendix C. Table shows positions of isolates in nucleotide (NT) and amino acid (AA) dendrograms from the three different methods used, Bayesian (BA), Maximum likelihood (ML) and Neighbor joining (NJ). Black spaces indicate where isolate did not group with one of the representative isolates. Hierarchy of grouping of representatives is VT, T36, RB, T30, B165, HA16-5, T3, and T68.

	ORF 4			ORF 5		ORF 6	ORF 7		ORF 8	ORF 10	3' UTR	DUF 3648	
	NT		AA	NT	AA	NT	NT		NT	NT	NT	NT	AA
	ML	NJ	ML	NJ	NJ	NJ	BA	ML	NJ	NJ	BA	NJ	NJ
AB046398	B165	VT	VT	B165	VT	B165			VT	VT		VT	T36
AF001623		T30	T30	T30	RB			B165	VT	VT	B165	VT	T36
AF260651	T30	T30	T30	T30	RB	T30	T36	T36	T36	T30	VT	T30	T36
AY170468	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	VT	VT
AY340974	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	VT	VT
DQ151548	B165	VT	VT	B165	VT	B165		B165	VT	VT	B165	VT	T36
DQ272579	T36	T36		T36	T36	T36	T36	T36	T36	16-5	T36	T36	VT
EU076703	B165	VT	VT	B165	VT	B165	B165	B165	VT	VT	B165	B165	B165
EU857538	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	T36
EU937519	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
EU937520	T30	T30	T30	T30	RB	T30	T36	T36	T36	T30	VT	T30	T36
EU937521	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
FJ525431	RB	RB	RB	RB	RB	RB	T36	T36	RB	RB	RB	RB	RB
FJ525432	RB	RB	RB	RB	RB	RB	T36	T36	RB	RB	RB	T36	RB
FJ525433	RB	RB	RB	RB	RB	RB	T36	T36	RB	RB	RB	RB	RB
FJ525434	RB	RB	RB	RB	RB	RB	T36	T36	RB	RB	RB	T36	VT
FJ525435	RB	RB	RB	RB	RB	RB	T36	T36	RB	RB	RB	T36	VT
FJ525436	B165	VT	VT	B165	VT	B165	B165	B165	VT	VT	B165	B165	B165
GQ454869	RB	RB	RB	16-5	RB	16-5	16-5	16-5	16-5	16-5	16-5	T36	RB
GQ454870	RB	RB	RB	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5
HM573451	T3	VT	RB	T3	VT	T3	16-5		16-5	VT	T3	VT	T36
JF957196	RB	RB	RB	RB	RB	RB	T36	T36	RB	RB	RB	T36	RB
JQ061137	B165	VT	VT	B165	VT	B165			VT	VT	B165	VT	T36
JQ798289		VT	RB							16-5		VT	T36
JQ911663	B165	VT	VT	B165	VT	B165			T3	VT	VT	B165	B165
JQ911664	B165	VT	VT	B165	VT	B165		B165	VT	VT	B165	VT	T36
JQ965169	B165	VT	VT	B165	VT	B165	T68	B165	VT	VT	B165	B165	B165
JX266712	RB	RB	RB	16-5	16-5	16-5	16-5		16-5	16-5	VT	T36	VT
JX266713				16-5	16-5	16-5	T36	T36	T36		VT	16-5	16-5
KC262793	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	T36
KC333868	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	B165	B165
KC517485	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	VT
KC517486	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	VT
KC517487	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	VT
KC517488	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36	VT
KC517489		T30	T30	T30	RB	T30	T36	T36	T36	T30	VT	T30	T36
KC517490	T30	T30	T30	T30	RB	T30	T36	T36	T36	T30	VT	T30	T36
KC517491	T30	T30	T30	T30	RB	T30	T36	T36	T36	T30	VT	T30	T36
KC517492	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	T36
KC517493	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	T36
KC517494	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	T36
KC525952	T3	VT	RB	T3	T3	T3	T3	T3	T3	VT	T3	VT	T3
NC001661	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	T36	VT
U16304	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	T36	VT
Y18420		T30	T30	T30	RB	T30	T36	T36	T36	T30	VT	T30	T36

Table 2.6 (continued): A summary of the position of isolates in the dendrograms constructed from different genome fragments in relevance to the representative isolates. Only dendrograms discussed in text are shown in the table, a more complete table can be found in Appendix C. Table shows positions of isolates in nucleotide (NT) and amino acid (AA) dendrograms from the three different methods used, Bayesian (BA), Maximum likelihood (ML) and Neighbor joining (NJ). Black spaces indicate where isolate did not group with one of the representative isolates. Hierarchy of grouping of representatives is VT, T36, RB, T30, B165, HA16-5, T3, and T68.

	Protease I	DUF 3614			Protease II	Methyltransferase		Helicase	
	NT	NT			NT	NT		NT	AA
	NJ	BA	ML	NJ	NJ	ML	NJ	NJ	NJ
AB046398	VT		VT		T3	VT		VT	VT
AF001623	VT	VT	VT	VT	VT	VT	VT	VT	VT
AF260651	T30	T30	VT	VT	VT	VT	VT	T30	T30
AY170468	T36	T36	T36	T36	T36	T36	T36	T36	T36
AY340974	T36	T36	VT	T36	T36	T36	T36	T36	T36
DQ151548	VT	VT	VT	VT	VT	VT	VT	VT	VT
DQ272579	T36	T36	VT	T36	T36	T36	T36	T36	T36
EU076703	B165	VT	VT	VT	VT	B165	B165	B165	B165
EU857538	VT	T3	VT	VT	T3	VT	VT	T30	T30
EU937519	VT	VT	VT	VT	VT	VT	VT	VT	VT
EU937520	T30	T30	VT	VT	VT	T30	VT	T30	T30
EU937521	T36	T36	T36	T36	T36	T36	T36	T36	T36
FJ525431	RB	RB	T36	RB	T36	RB	RB	RB	RB
FJ525432	RB	RB	T36	RB	T36	RB	RB	RB	RB
FJ525433	RB	RB	T36	RB	T36	RB	RB	RB	RB
FJ525434	T36	T36	T36	T36	T36	T36	T36	T36	
FJ525435	T36	T36	T36	T36	T36	T36	T36	T36	
FJ525436	B165	16-5	T68	T68	T68	B165	B165	B165	B165
GQ454869	RB	RB	T36	RB	T36	RB	RB	RB	RB
GQ454870	B165	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5
HM573451	VT	VT	VT	VT	VT	VT	VT	VT	VT
JF957196	RB	RB	T36	RB		RB	RB	RB	RB
JQ061137	VT	VT	VT	VT	VT	VT		VT	T30
JQ798289	VT	VT	VT	VT	T36		T36		
JQ911663	B165	16-5	T68	T68	T68	B165	B165	B165	B165
JQ911664	VT	VT	VT	VT	VT	VT	VT	VT	VT
JQ965169	B165	16-5	T68	T68	T68	B165	B165	B165	B165
JX266712	T36	T36	T36	T36	T36		T36	T36	
JX266713	B165	16-5	16-5	16-5	16-5	16-5		16-5	16-5
KC262793	VT	VT	VT	VT	VT	VT	16-5	VT	VT
KC333868	B165	16-5	T68	T68	T68	B165	VT	B165	B165
KC517485	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517486	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517487	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517488	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517489	T30	T30	VT	VT	VT	VT	VT	T30	T30
KC517490	T30	T30	VT	VT	VT	VT	VT	T30	T30
KC517491	T30	T30	VT	VT	VT	VT	VT	T30	T30
KC517492	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC517493	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC517494	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC525952	VT	T3	VT	VT	T3	T3	VT	T30	T30
NC001661	T36	T36	VT	T36	T36	T36	T36	T36	T36
U16304	T36	T36	VT	T36	T36	T36	T36	T36	T36
Y18420	T30	T30	VT	VT	VT	VT	VT	T30	T30

1) 5' UTR

In this genome region the only dendrogram with enough branch support to distinguish between some of the main isolates is the one produced with Bayesian analysis with the nucleotide sequences (Figure 2.6). This dendrogram groups the VT, T3 and T30 isolates in different clades. The B165 and T68 isolates are located on different branches, but are very closely related, however T36, RB and HA16-5 isolates all collapse into a single clade. There are two additional clades present, one containing AB046398 (which groups close or with the VT clade in the complete genome dendrograms) and JQ798289 (which groups individually in the complete genome dendrograms) and another clade containing AF001623, DQ15148 and JQ911664, all of which groups with or close to VT and T30 clades.

2) ORF 1a

Within this gene region all dendrograms resemble the complete genome dendrograms with each method used, but this is expected since ORF 1a constitutes almost half of the genome of CTV (Figure A.1 - A.3, Appendix A). These dendrograms only differ from those of the complete genome by the branch support which is lower in some instances. This fragment however cannot be considered as a viable standard for genotyping since it is very large (almost 9400 bp) and hence impractical to sequence for these type of studies.

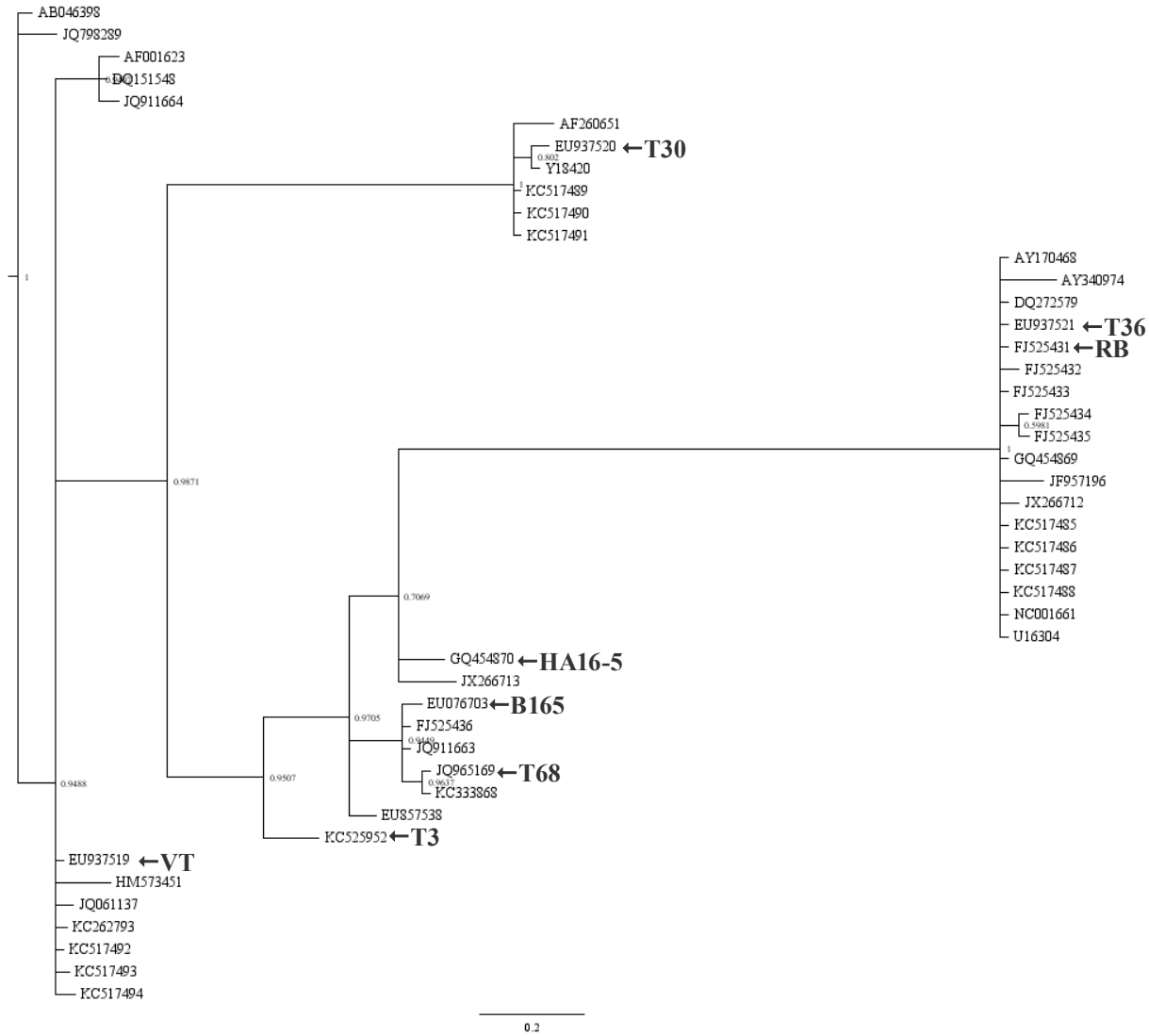

Figure 2.6: Phylogenetic dendrogram of the 5' UTR nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

3) ORF 1b

The dendrogram created from the nucleotide sequence with the NJ method (Figure 2.7) has too low branch support to separate the T3, T30 and VT isolates, and the T68 and B165 isolates forms one clade. Interestingly, Kpg3 (HM573451) groups separately, instead of with the VT isolate as in the complete genome dendrograms. Some isolates that grouped separate from the representative isolates group with these isolates in this dendrogram; AB046398, DQ151548, JQ911664 and JQ061127 group with the VT isolate

and JQ911663 that grouped closer to the T68 and B165 isolates, now groups within the clade containing the VT, T30 and T3 isolates. The same situations can be seen in the ML version of the amino acid data (Figure A.4, Appendix A).

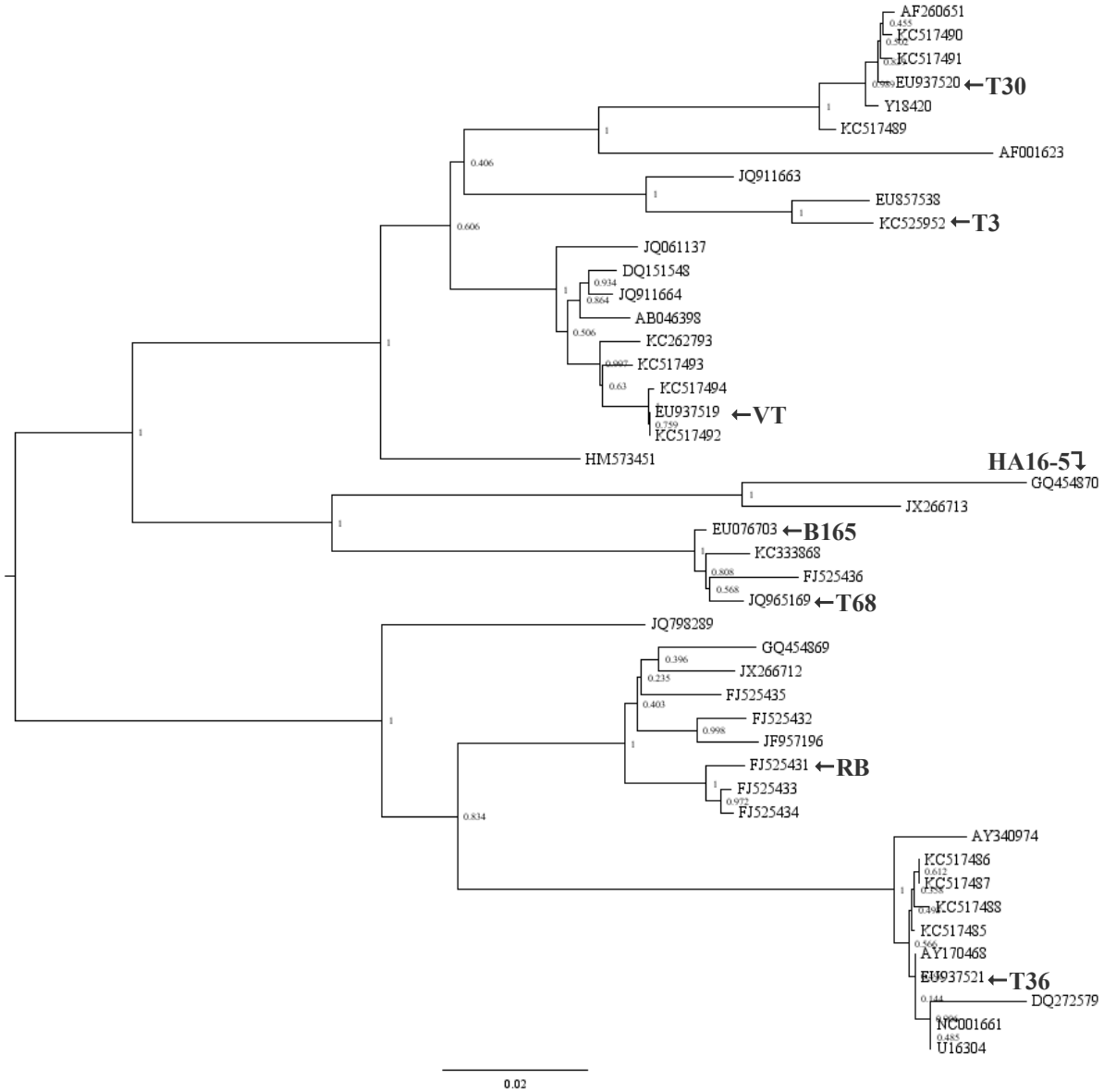

Figure 2.7: Phylogenetic dendrogram of the ORF 1b nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

4) ORF 2

Although dendrograms generated by all three algorithms for ORF 2 look similar, they differ significantly from the complete genome dendrograms. When using the nucleotide BA dendrogram as example (Figure 2.8, the other ORF 2 dendrograms are shown in Figure A.5 and A.6, Appendix A), it can be seen that the T30 and RB isolates are more closely related in this genome fragment whereas in other fragments the RB and T36 isolates are more closely related. The HA16-5 and T3 isolates forms one clade, and the B165 and T68 isolates also forms one clade. This is very interesting since ORF 2 encodes the p33 protein, which plays a role in cross-protection (Folimonova, 2012). In that study it was found that the p33 protein from a T36 (CTV9) isolate could prevent super-infection of that same isolate, but the p33 proteins from T68 and T30 could not prevent super-infection. It would be interesting to know exactly how similar these proteins should be to exclude super-infection, and for example, if strains from T30 and RB genotypes would be able to exclude super-infection of each other since their ORF 2 is so closely related to each other. The same goes for strains from the HA16-5 and T3 genotypes, as well as for the strains from the T68 and B165 genotypes. The CT-ZA3 (KC333868) isolate described in the Zablocki and Pietersen (2014) study, groups with the T68 isolate, but it is known to effectively be a VT/B165 genotype. With the p33 gene dendrogram (ORF 2), this isolate groups with the VT clade whereas the T68 isolate groups with the B165 clade. It would be very interesting to know whether CT-ZA3 can cross protect against strains form VT, B165 and/or T68 genotypes. This needs to be determined with further studies where constructs of the genome is created and p33 genes are swopped out. Although the dendrograms produced from this genome region differs considerably from the complete genome dendrogram, it still produces specific groups, although their relationship to each

other is different. Although these relationships are so different, they are very important to consider due to the importance of the gene product.

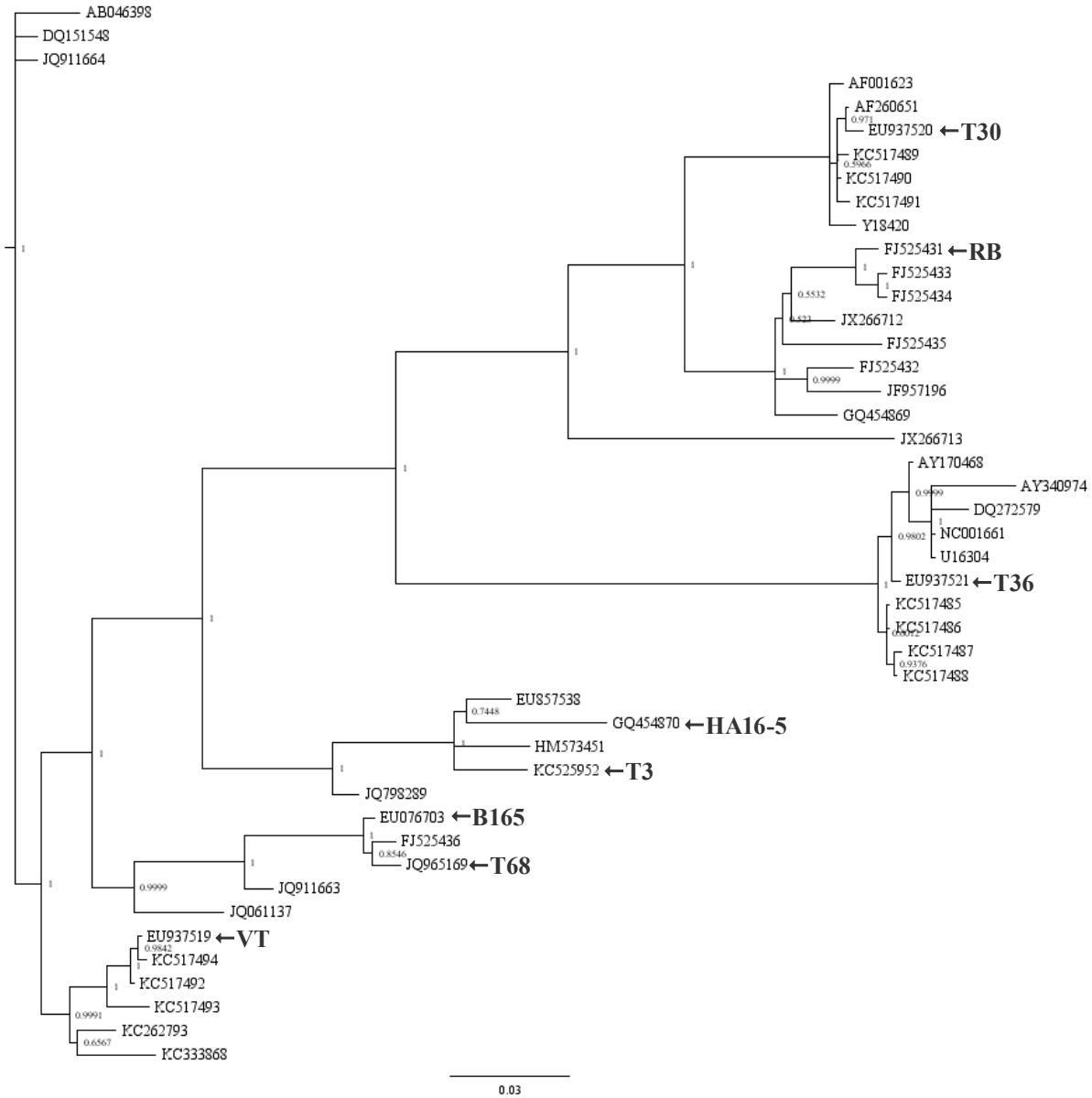

Figure 2.8: Phylogenetic dendrogram of the ORF 2 nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

5) ORF 4

Although different clades can be identified within this genome region, the positions of certain isolates in relation to other isolates are very different from their position in the

complete genome dendrograms. The T30 and T36 isolates are more closely related based on ORF 4 than based on other regions. The B165 and T68 isolates group with the VT clade due to low branch support and the HA16-5 isolates groups within the RB clade (Figure A.7 - A.9, Appendix A).

6) **ORF 5**

With ORF 5 nucleotide sequences, the dendrogram generated with ML and BA and the ML dendrogram for the amino acid sequence does not have high enough branch support to distinguish isolates into different clades. However, the NJ nucleotide and amino acid data dendrograms have much better branch support. In the nucleotide NJ dendrogram all representative isolates group individually, except for the B165 and T68 isolates which forms one clade as in many other dendrograms. In this dendrogram the T36 and T30 isolates are more closely related than the VT and T30 isolates, and the T3 isolate is more related to the VT clade than the T30 clade (Figure 2.9). The dendrogram produced from the amino acid data has the VT, B165 and T68 isolates all forming one clade due to lack of branch support. The T30 and RB isolates also groups together for the same reason (Figure A.10, Appendix A).

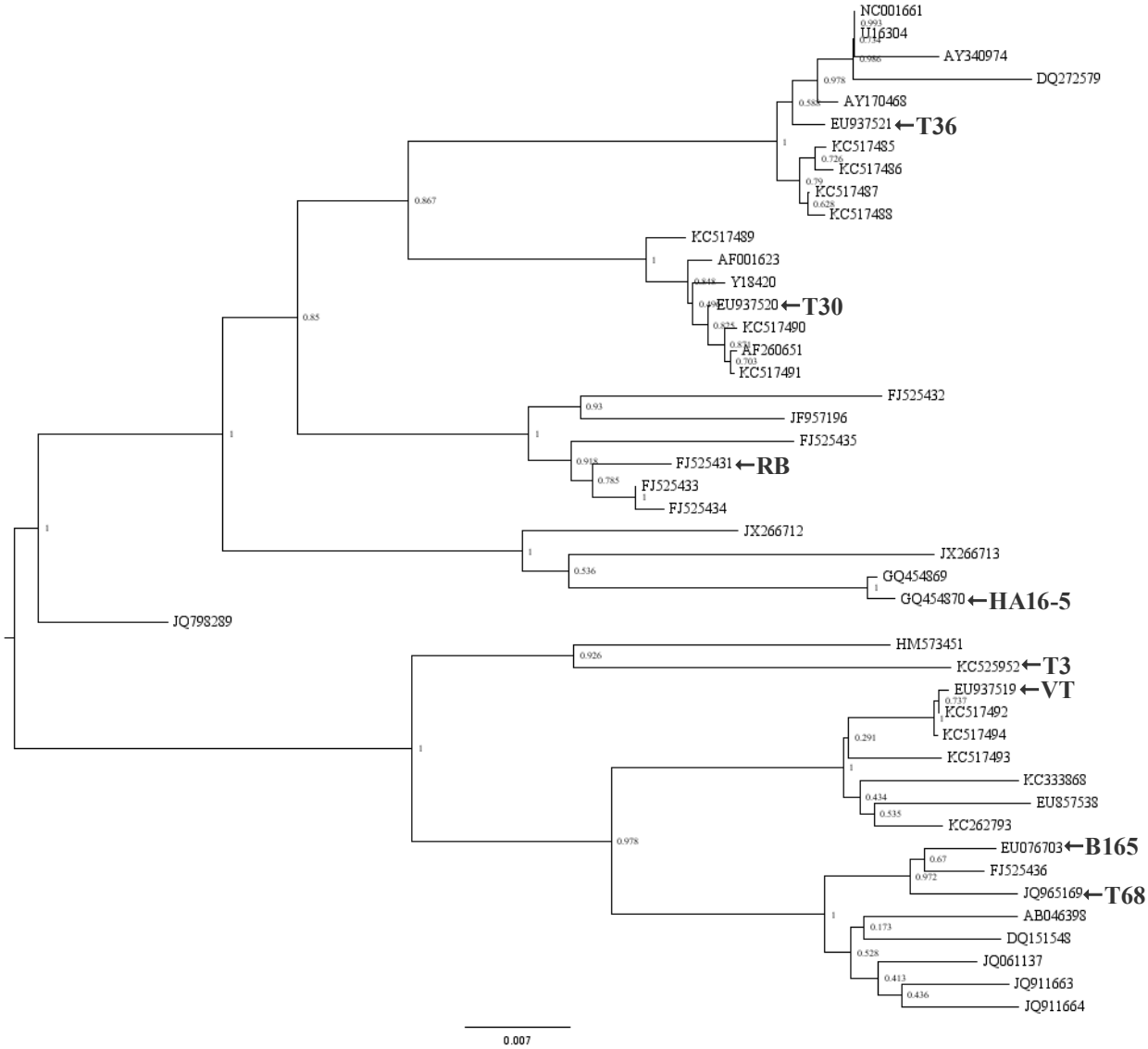

Figure 2.9: Phylogenetic dendrogram of the ORF 5 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

7) ORF 6

Within ORF 6 the NJ dendrogram has enough branch support to distinguish between all the representative isolates, however the B165 and T68 isolates forms one clade, but not as a result of low branch support (Figure 2.10). As in the dendrograms constructed from ORF 5, the T30 and RB isolates are more closely related than the RB and T36 isolates. There are some interesting clades in this dendrogram that are not similar to those within the complete genome tree. The JX061137 isolate which grouped close to the VT or T30

clades grouped on its own and two isolates that grouped within the RB clade, JX266712 and GQ454869 group within the HA16-5 clade.

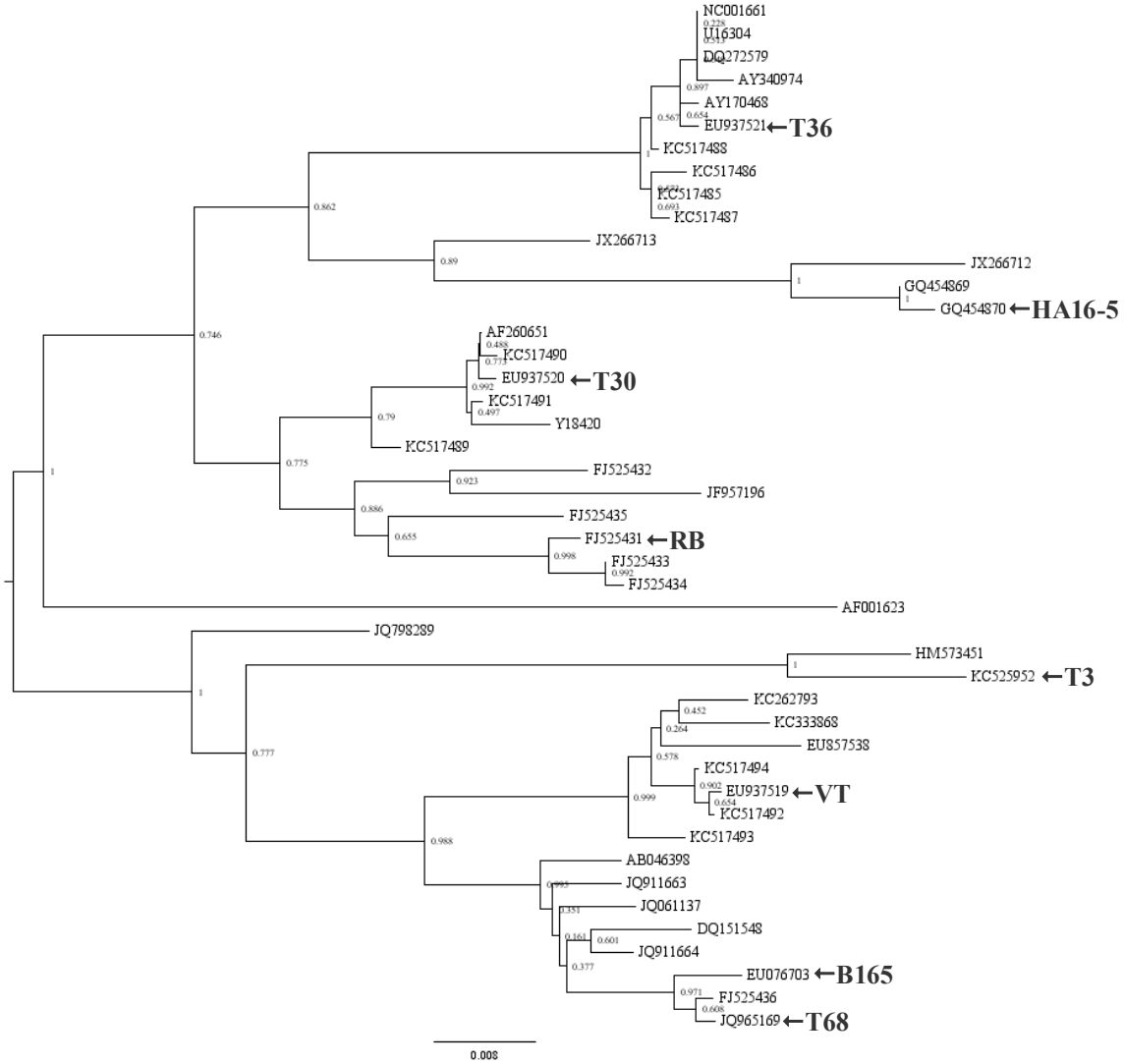

Figure 2.10: Phylogenetic dendrogram of the ORF 6 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

8) ORF 7

The ML and BA analysis for the nucleotide alignment of ORF 7 produced very similar results (Figure 2.11 and Figure A.11, Appendix A). Although some isolates group individually, the B165 and T68 isolates yet again formed one clade, also including

AF001623, DQ151548 and JQ911664 which grouped closer to the VT or T30 clades in the dendrograms based on the complete genome. Due to low branch support the T30, T36 and RB isolates collapse into one clade. AB046398, JQ061137 and JQ911663 forms a separate clade, although from the complete genome dendrograms AB046398 and JQ061137 seems to be more related to isolates in the VT or T30 clades, and JQ911663 is more related to the B165 isolate, but this isolate's position is close to the B165/T68 clade in the ORF 7 dendrogram.

9) ORF 8

The NJ dendrogram for ORF 8 (Figure 2.12) differentiates between VT, T3 and HA16-5 clades, but the B165 and T68 isolates still forms one clade including some isolates that grouped individually (but close to VT) in the complete genome dendrogram. The T36 and T30 isolates form one clade due to low branch support.

10) ORF 10

In the NJ dendrogram (Figure 2.13), the B165 and T68 isolates form one cluster, and due to the low branch support, the VT isolates also collapses with this clade. All other representative isolates group individually, with the T30 and RB clade being more related than the RB and T36 clade.

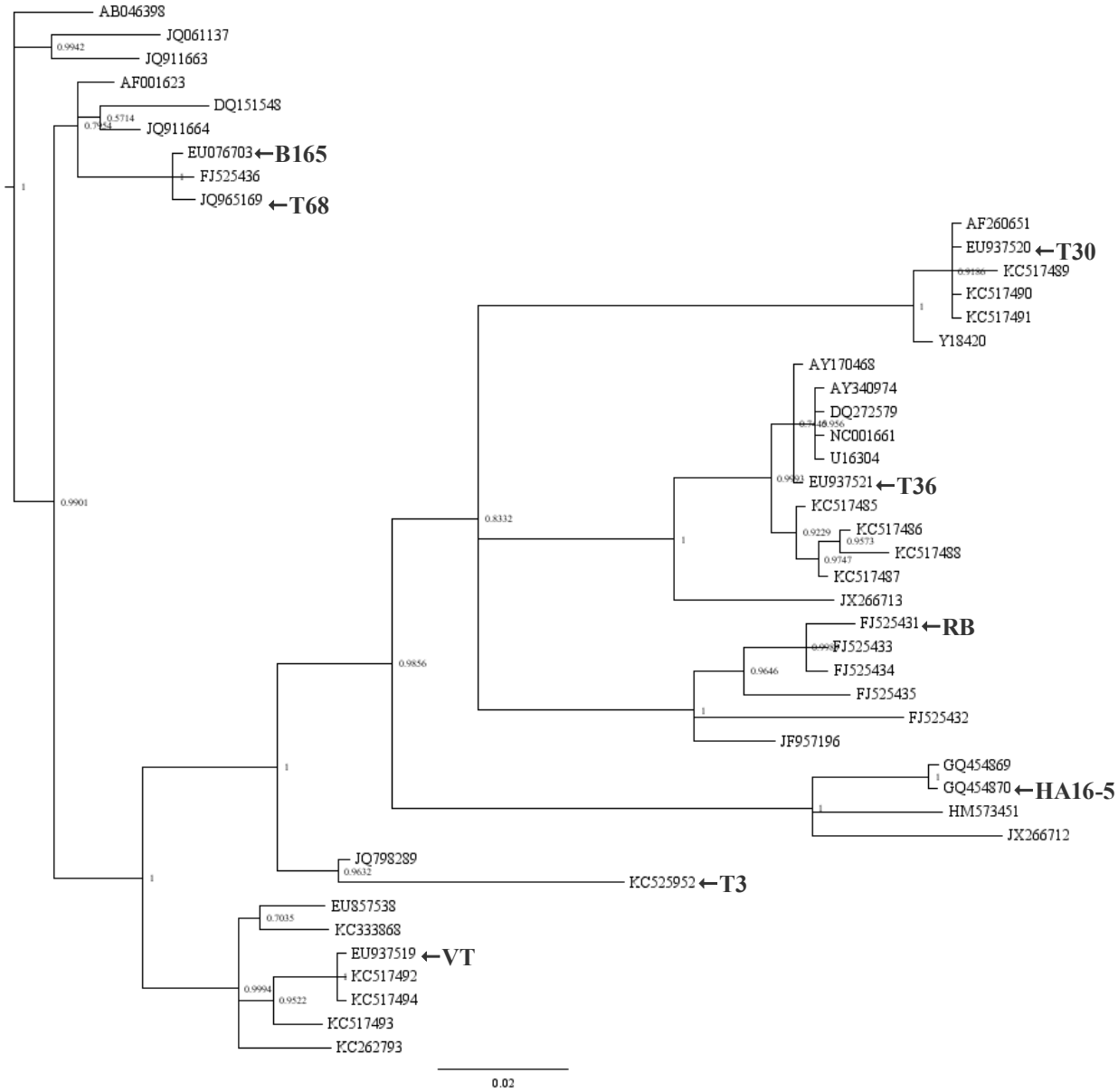

Figure 2.11: Phylogenetic dendrogram of the ORF 7 nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

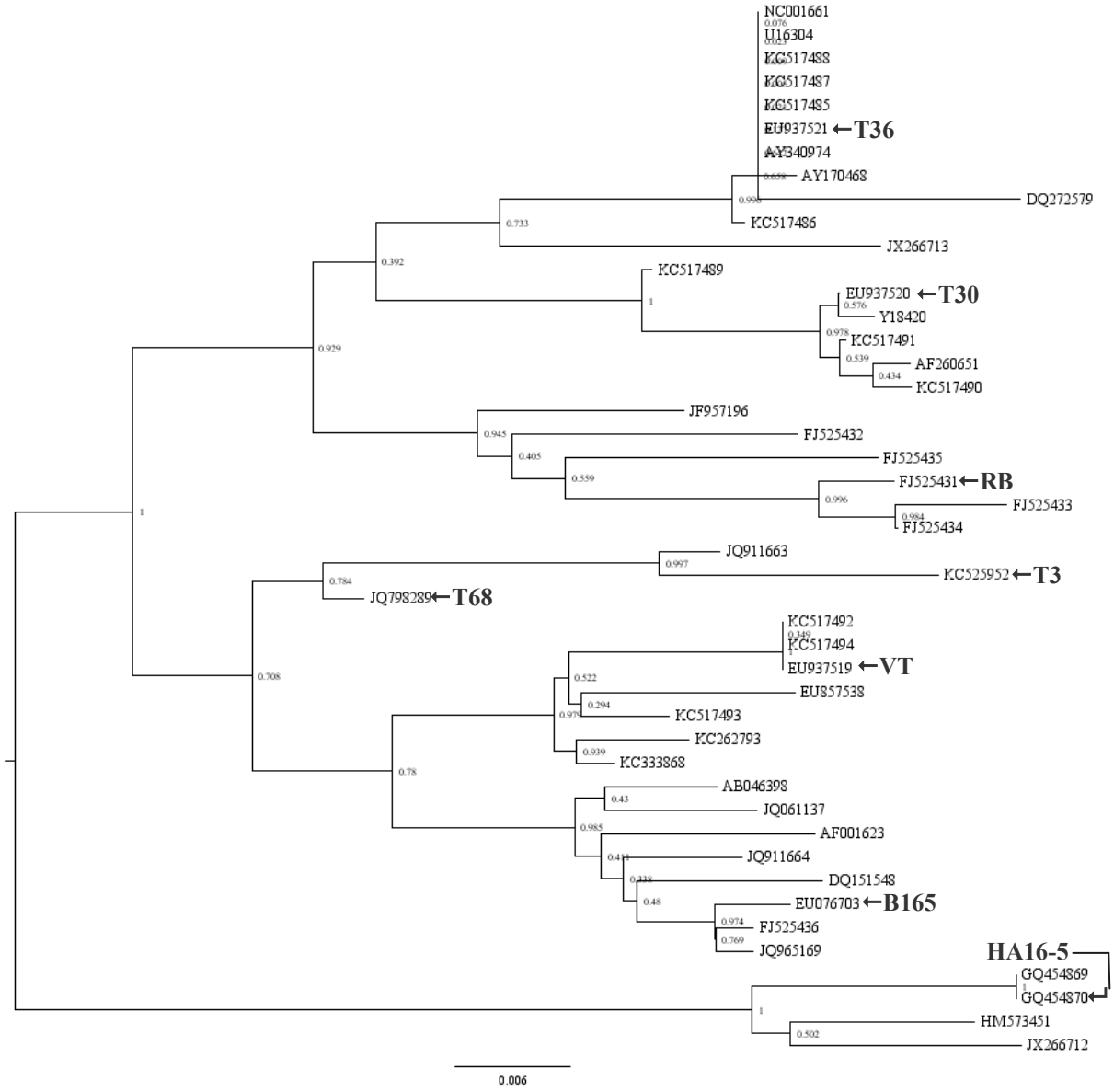

Figure 2.12: Phylogenetic dendrogram of the ORF 8 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

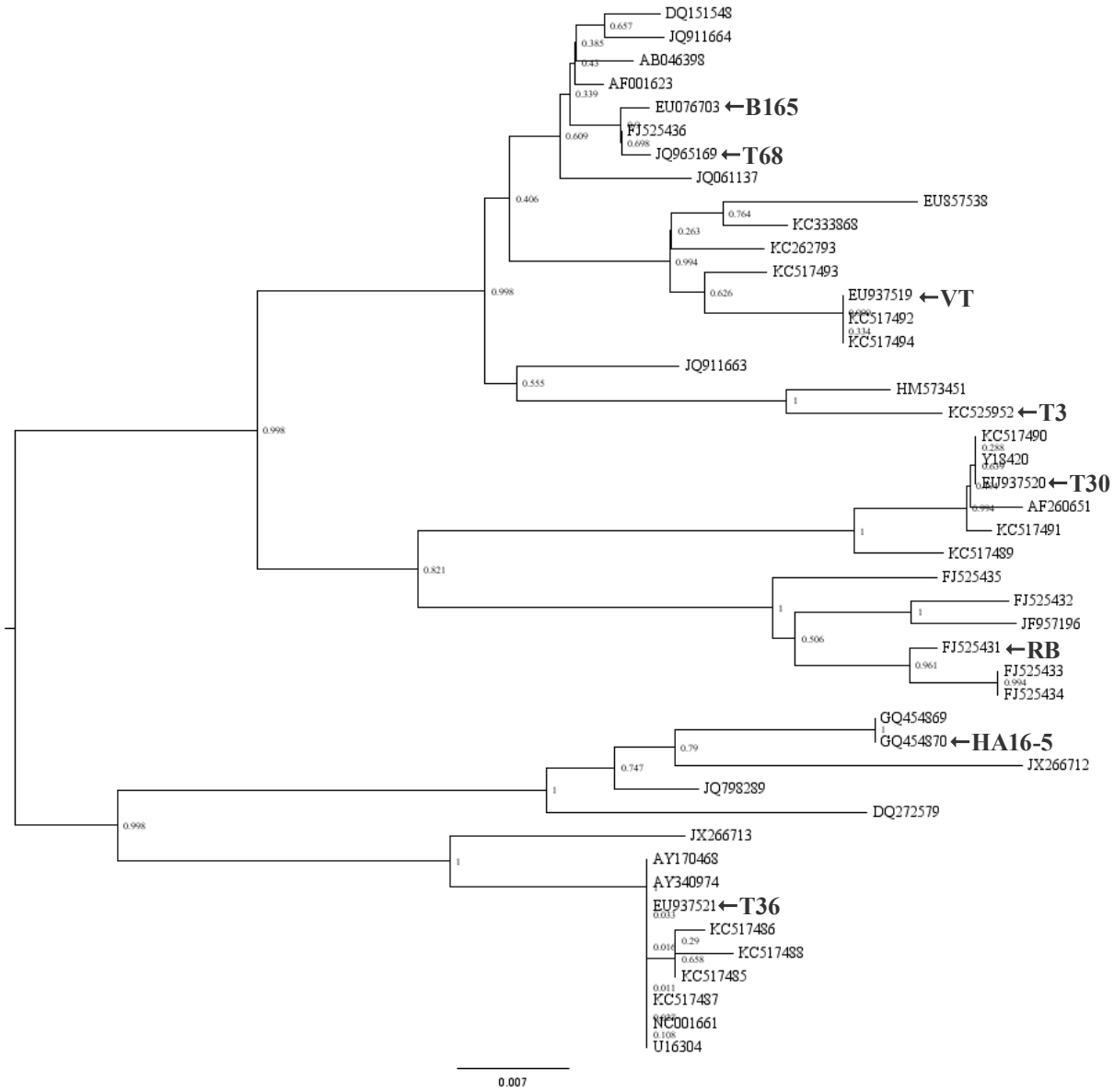

Figure 2.13: Phylogenetic dendrogram of the ORF 10 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

12) 3' UTR

In the BA dendrogram of the nucleotide sequences of the 3'UTR (Figure 2.14), the B165 and T68 yet again forms one clade, and due to low branch support the T30 and VT clades form one. The RB clade is more related to the VT/T30 clade than the T36 clade.

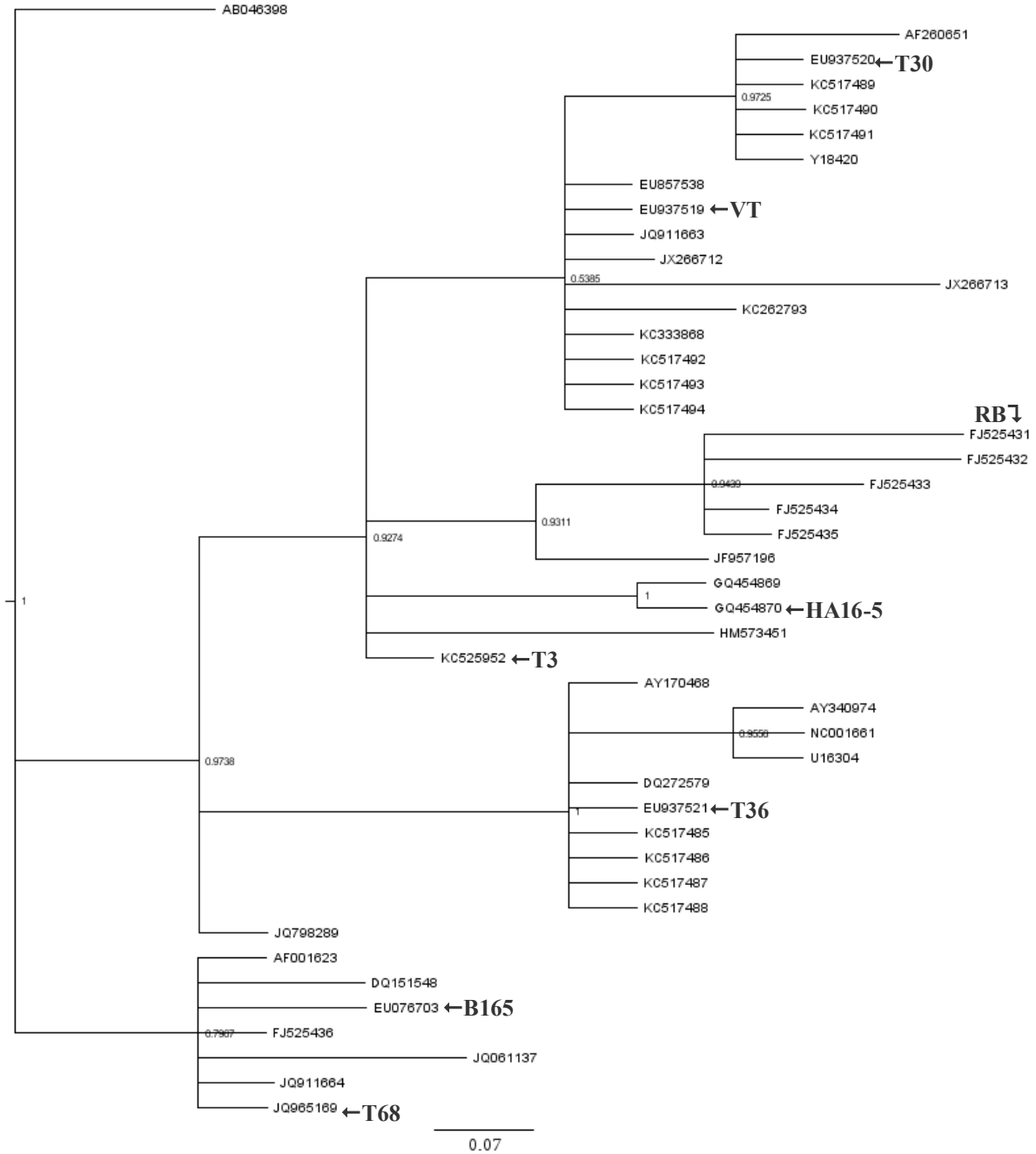

Figure 2.14: Phylogenetic dendrogram of the 3' UTR nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

13) DUF 3648

The NJ dendrograms for both the nucleotide (Figure A.12, Appendix A) and amino acid (Figure 2.15) alignments look very similar. In both dendrograms the B165 and T68 isolates form one clade, but the rest of the representative isolates form separate clades. The branch support in the nucleotide dendrogram is too low to keep RB and T36 in separate clades, but in the amino acid dendrogram there is sufficient branch support for these isolates being in different clades. The T36 and RB clades are closely related, the VT, T30 and T3 clades are more related and then the HA16-5 and T68/B165 clades are closely related with this clade being closer related to the VT, T30 and T3 clades than the RB and T36 clades.

14) Protease I

Dendrograms constructed from the Protease I alignment by the three algorithms used are very similar and the general topology of all of the dendrograms are similar to the complete genome dendrograms. In the nucleotide dendrogram for NJ (Figure 2.16), the VT and T30 isolates form one clade, as do the HA16-5, T68 and B165 isolates. The ML nucleotide dendrogram shows a similar picture, although the T3 isolate is included in the VT/T30 clade due to low branch support. Although the T68 and B165 isolates form one clade, the HA16-5 isolate is in a separate clade.

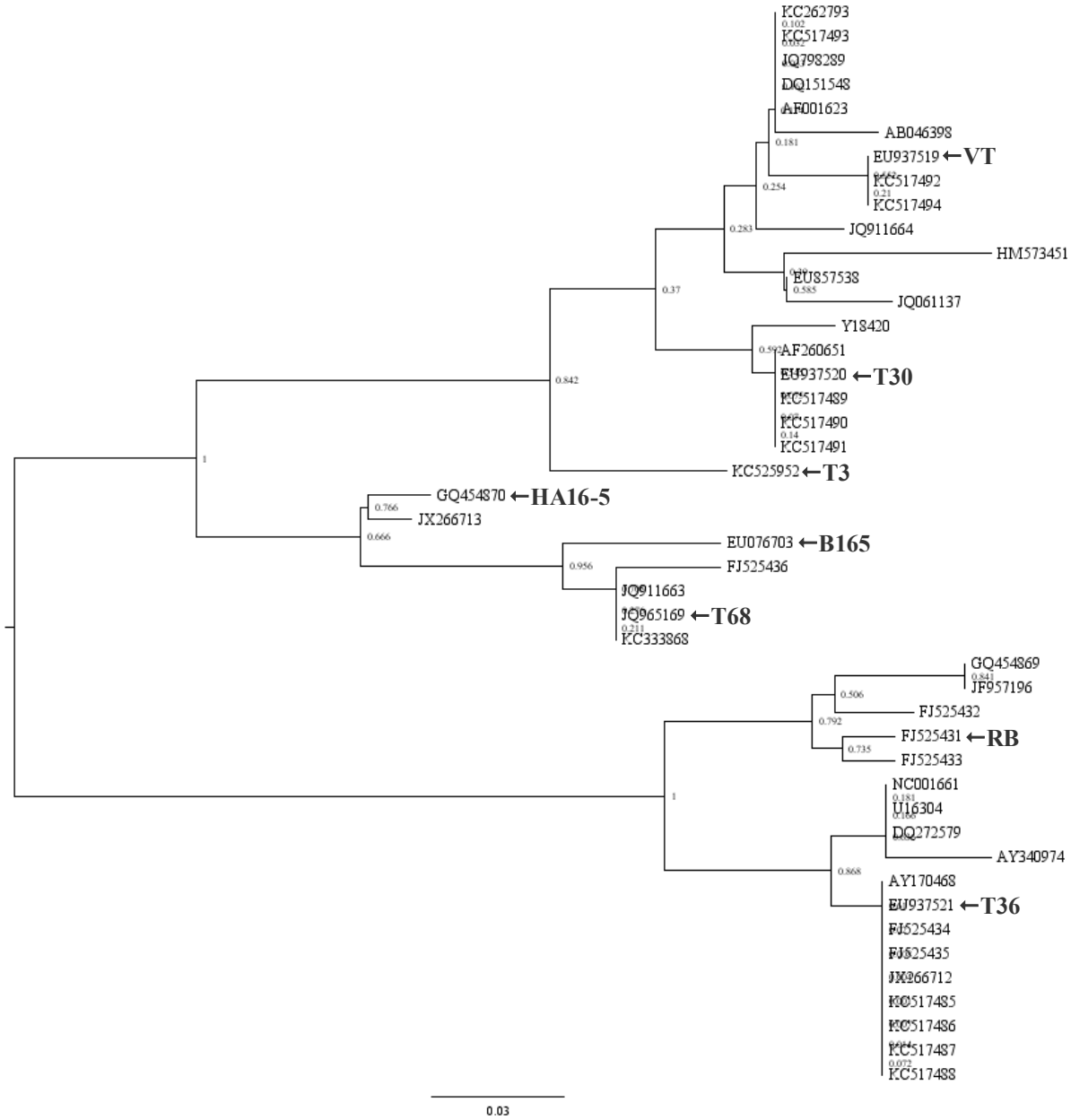

Figure 2.15: Phylogenetic dendrogram of the DUF 3648 amino acid sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

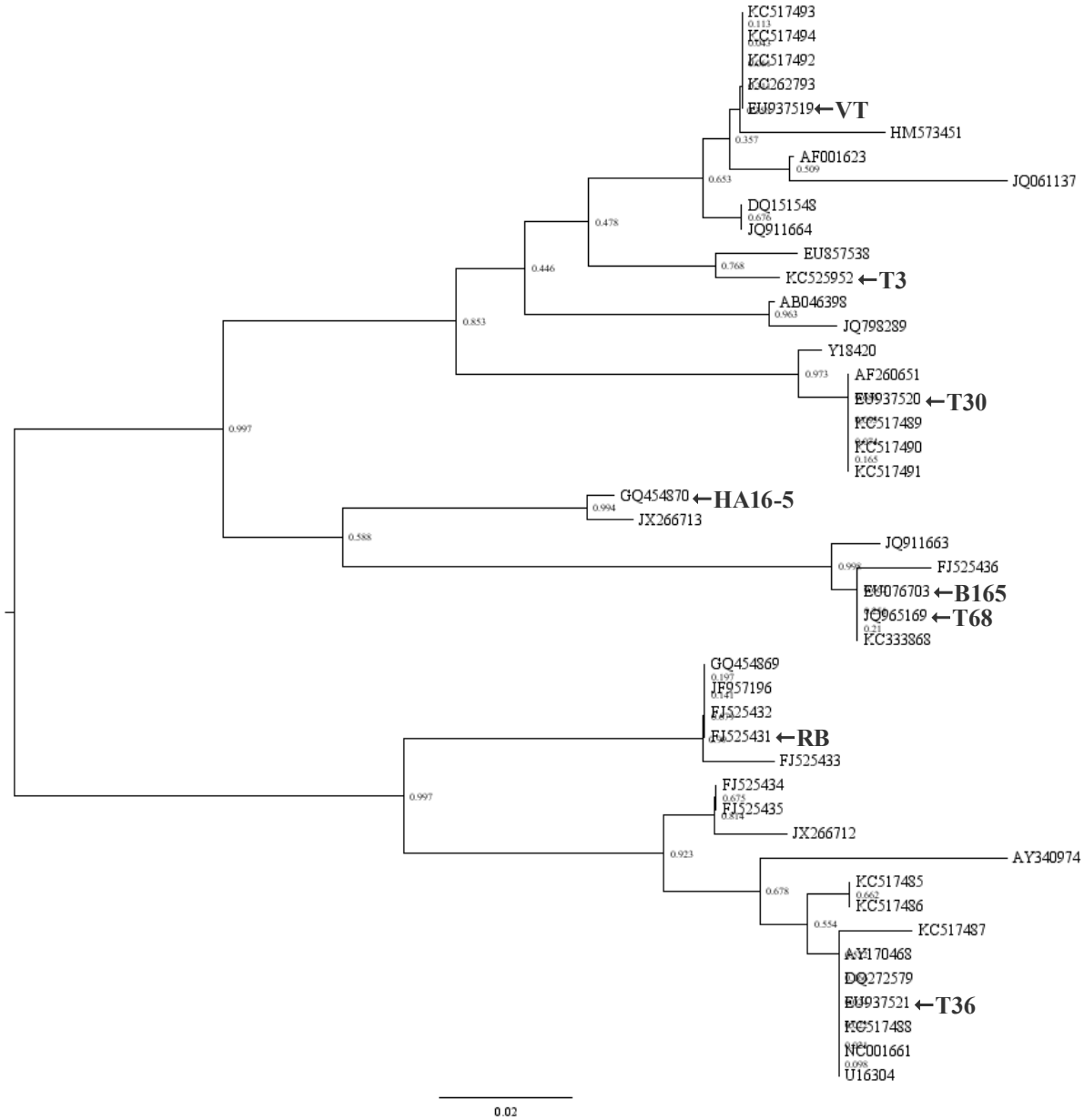

Figure 2.16: Phylogenetic dendrogram of the Protease I nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

15) DUF 3614

In the nucleotide BA dendrogram of this genome region (Figure 2.17), the T36 and RB isolates form separate clades, the B165 isolate groups with the VT clade, the T3 isolate groups with the T30 clade and due to low branch support, the T68 and HA16-5 isolates form one clade. This is somewhat different from previous dendrograms. For the

nucleotide ML dendrogram, the T30, B165 and VT isolates form one clade due to low branch support and the T68 and HA16-5 isolates form different clades.

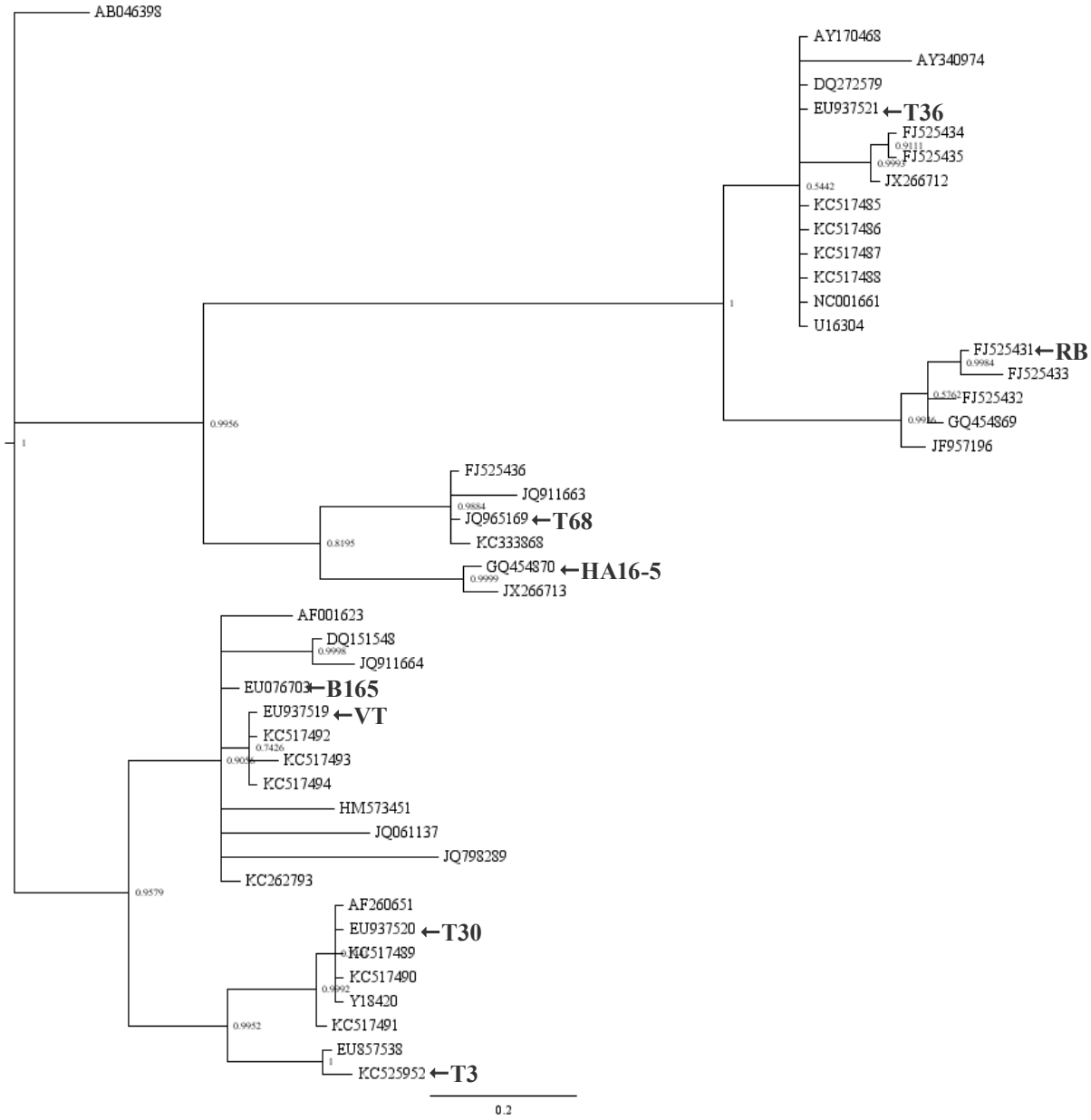

Figure 2.17: Phylogenetic dendrogram of the DUF 3614 nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

Although the representative T36 and RB isolates form two different clades, some isolates that grouped within the T36 clade in the complete genome dendrograms, now group

within the RB clade based on DUF 3614 (Figure A.13, Appendix A). For the nucleotide alignment analyzed with NJ the VT, B165, T3 and T30 isolates form one clade and the RB, T36, T68 and HA16-5 isolates all form different clades (Figure A.14, Appendix A).

16) Protease II

The NJ dendrogram (Figure 2.18) constructed from the nucleotide alignment shows the RB and T36 isolates forming one clade due to low branch support as well as the VT, B165 and T30 isolates forming one clade. The T3, HA16-5 and T68 isolates are all located in different clades.

17) Methyltransferase

For the ML nucleotide dendrogram based on the methyltransferase gene sequence the VT, T3 and T30 isolates cluster together due to low branch support, and the rest of the representative isolates form individual clades (Figure 2.19). As is the case with a few other dendrograms (e.g. DUF 3614), some isolates that clusters within the RB clade in the complete genome dendrogram group with the T36 clade. The same case exists with the NJ dendrogram of the nucleotide alignment (Figure A.5, Appendix A).

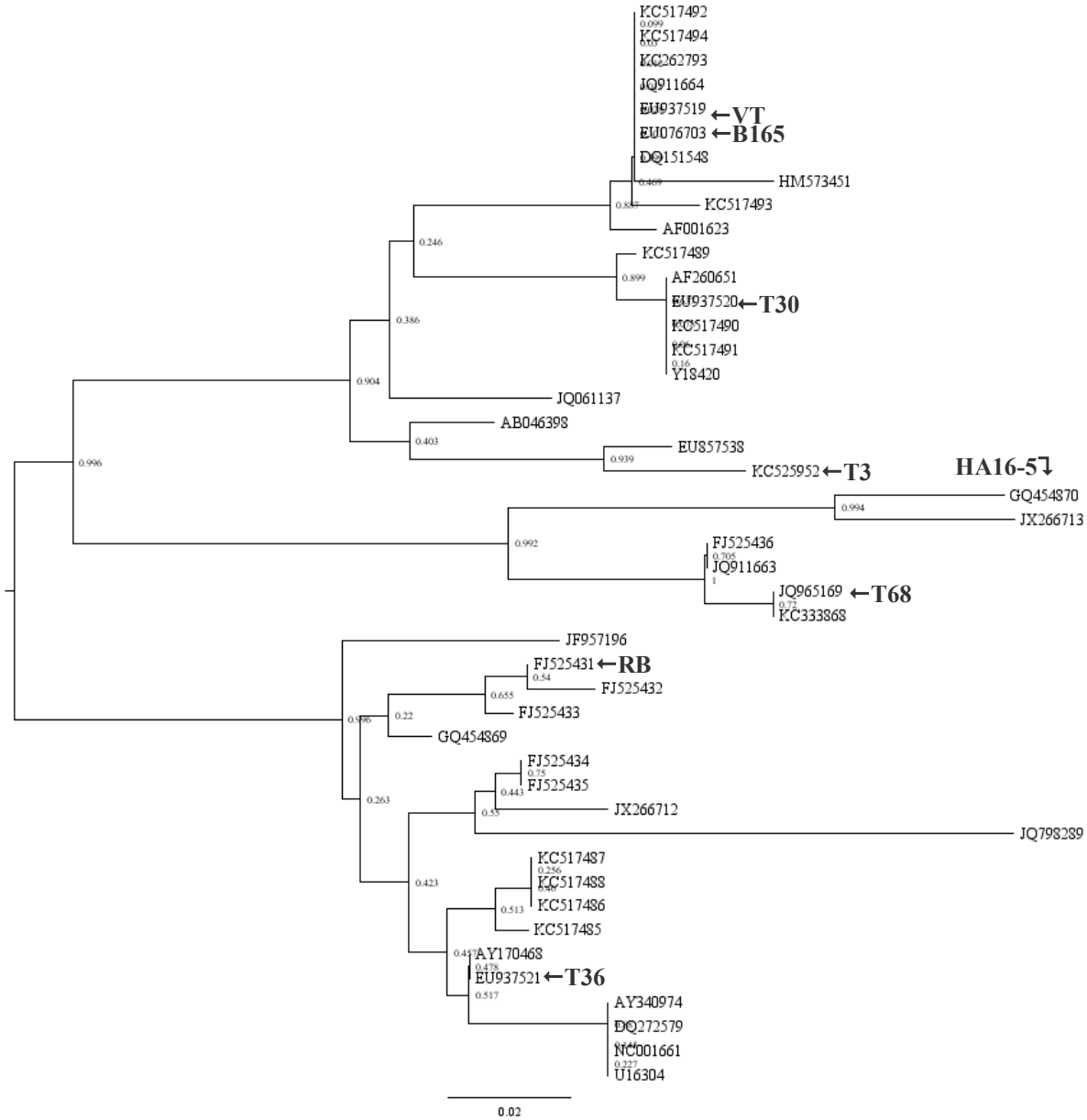

Figure 2.18: Phylogenetic dendrogram of the Protease II nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

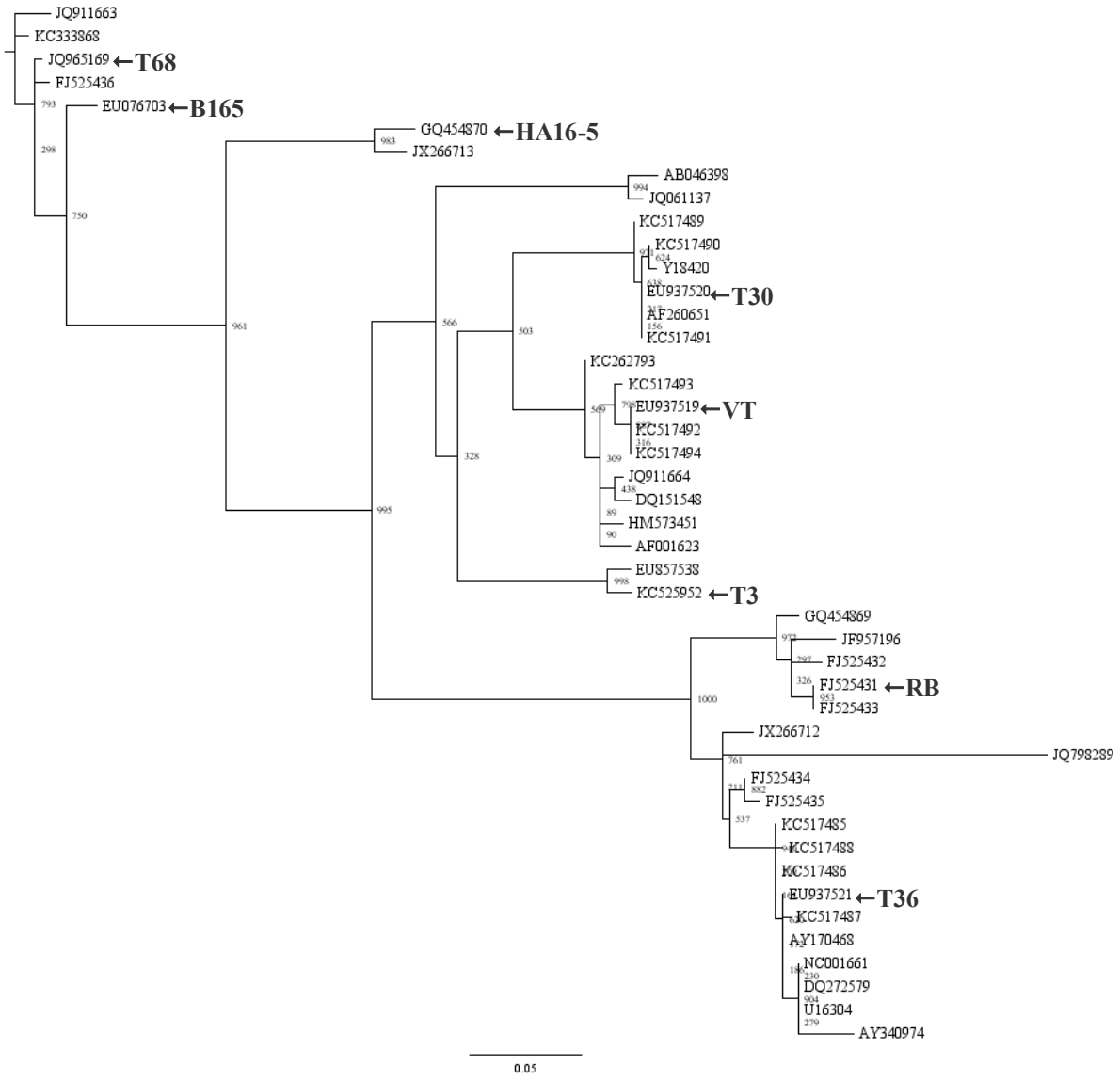

Figure 2.19: Phylogenetic dendrogram of the methyltransferase nucleotide sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

18) Helicase

The dendrograms constructed from both the nucleotide (Figure A.16, Appendix A) and amino acid (Figure 2.20) alignments using the NJ method for the helicase gene sequences both show similar sequence relationships. The T3 and T30 isolates collapse together in one clade due to low branch support and the B165 and T68 isolates form one clade. All other representative isolates group individually, with the same case as in the DUF 3614

and methyltransferase dendrograms where some isolates that group with RB in the complete genome dendrogram now group with T36.

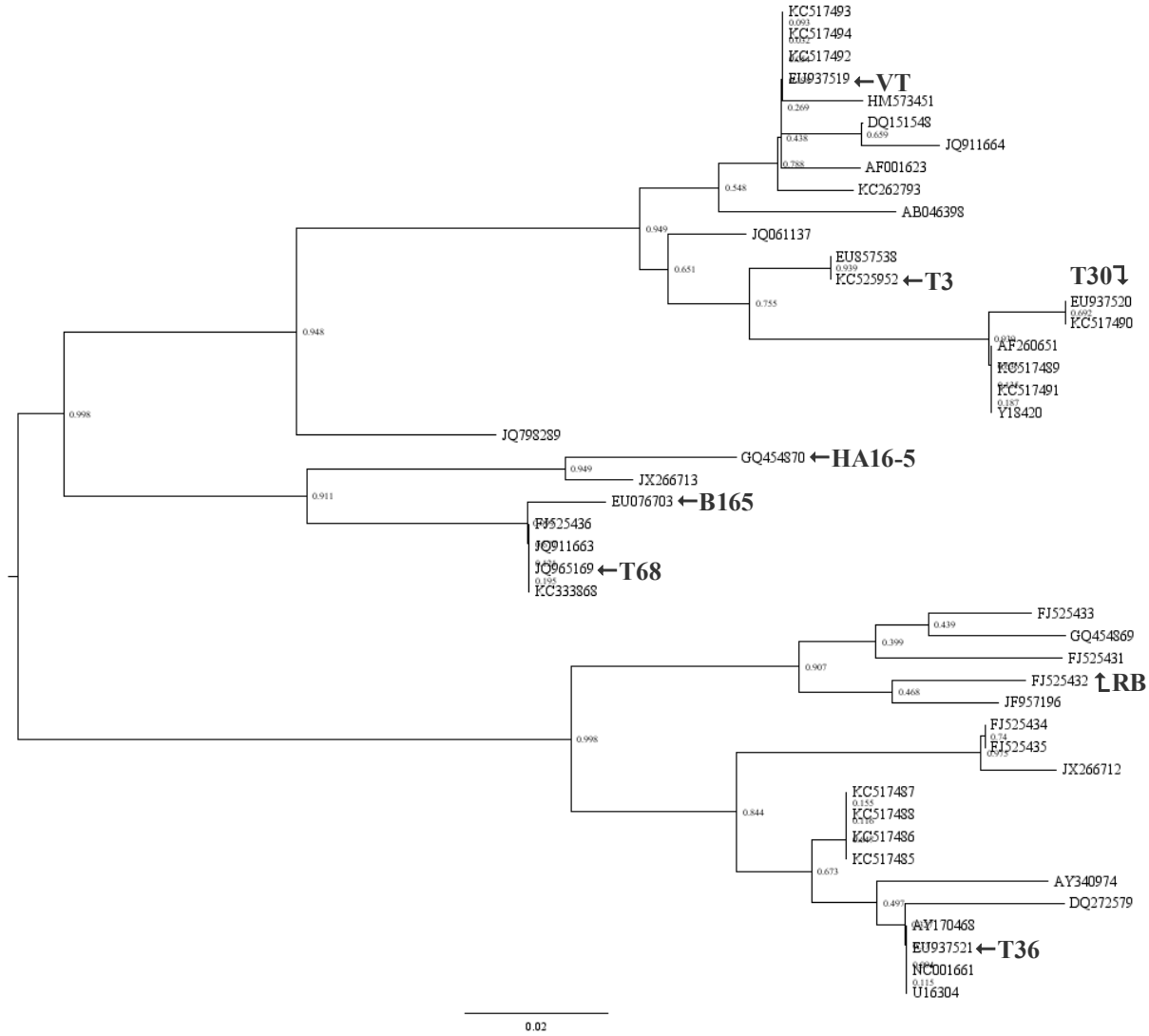

Figure 2.20: Phylogenetic dendrogram of the helicase nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

Since none of the individual fragment dendrograms gave results comparable with those obtained with dendrograms based on the complete CTV genomes, it was decided to combine some of the fragments that produced promising results. The partition homogeneity test for all concatenated datasets resulted in a P-value of 0.01 which indicated that these datasets could not be analyzed under the same amino acid substitution model and therefore had to be partitioned to allow different areas of the datasets to be analyzed with different substitution

models. Therefore all concatenated datasets where only analyzed in MrBayes where the data can be partitioned and each partition can be analyzed under a different nucleotide model. The combinations used and results obtained are shown in Table 2.4. Promising combinations will be discussed individually and a summary of the position of isolates in relevance to the representative isolates are shown in Table 2.7.

1) **Combination 3: 5' UTR and ORF 2**

When comparing this dendrogram (Figure 2.21) with the complete genome dendrogram, it seems very similar at first analysis. The T36, RB, T30 and VT isolates all form individual clades. The B165 and T68 isolates form one clade, as in many other dendrograms, but the T3 and HA16-5 isolates also cluster together. Some individual isolates cluster differently; JX266713 which usually groups with GQ45470 (HA16-5) now clusters individually, HM573451 which grouped near VT is now located within the T3/HA16-5 clade, JQ798289 which grouped alone is also within this cluster, JQ911663 which grouped alone is now within the B165/T68 clade, and lastly KC33868 which grouped with the T68 isolate now groups with the VT clade. The T30 clade is closer related to the T36 and RB clades than the VT clade, but this is expected, as with the original ORF 2 dendrogram the T36 and VT clades were closer related, whereas the RB and T30 clades where closer related.

Table 2.7: A summary of the position of isolates in the dendrograms constructed from combinations of different genome fragments with relevance to the representative isolates. Descriptions of the combinations used can be found in Table 2.4. Only dendrograms discussed in text are shown in the table, a more complete table can be found in Appendix C. Hierarchy of grouping of representatives is VT, T36, RB, T30, B165, HA16-5, T3, and T68.

	Complete					Combination								
	NT			AA										
	BA	ML	NJ	ML	NJ	3	7	8	15	22	23	25	27	28
AB046398		VT	VT	VT	VT									
AF001623		VT		VT	VT	T30	T30	T30	T30	T30	T30	T30	VT	T30
AF260651	T30	VT	T30	VT	VT	T30	T30	T30	T30	T30	T30	T30	T30	T30
AY170468	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
AY340974	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
DQ151548		VT	VT	VT	VT								VT	
DQ272579	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
EU076703	B165	B165	B165	VT	B165	B165	B165	B165	B165	B165	B165	B165	B165	B165
EU857538	T3	VT	VT	VT	T3	16-5			16-5	T3	T3	T3	T3	16-5
EU937519	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
EU937520	T30	VT	T30	VT	VT	T30	T30	T30	T30	T30	T30	T30	T30	T30
EU937521	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
FJ525431	RB	RB	RB	VT	RB	RB	RB	RB	RB	RB	RB	RB	RB	RB
FJ525432	RB	RB	RB	VT	RB	RB	RB	RB	RB	RB	RB	RB	RB	RB
FJ525433	RB	RB	RB	VT	RB	RB	RB	RB	RB	RB	RB	RB	RB	RB
FJ525434	RB	RB	RB	VT	RB	RB	RB	RB	RB	RB	RB	RB	T36	T36
FJ525435	RB	RB	RB	VT	RB	RB	RB	RB	RB	RB	RB	RB	T36	T36
FJ525436	B165	B165	B165	VT	B165	B165	B165	B165	B165	B165	B165	B165	T68	B165
GQ454869	RB	RB	RB	VT	RB	RB	RB	16-5	RB	RB	RB	RB	RB	RB
GQ454870	16-5	16-5	16-5	VT	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5	16-5
HM573451	VT	VT	VT	VT	VT	16-5	T3	T3	16-5	T3	T3		VT	16-5
JF957196	RB	RB	RB	VT	RB	RB	RB	RB	RB	RB	RB	RB	RB	RB
JQ061137		VT	VT	VT	VT			B165	B165	B165				
JQ798289				VT		16-5	T3	T3	16-5					16-5
JQ911663			B165	VT		B165	B165	B165	B165	B165	B165	T3	T68	B165
JQ911664		VT	VT	VT	VT								VT	
JQ965169	T68	T68	T68	VT	T68	B165	B165	B165	B165	B165	B165	B165	T68	B165
JX266712	RB	RB	RB	VT	RB	RB	RB	16-5	RB	RB	RB	RB	T36	T36
JX266713	16-5	16-5	16-5	VT	16-5					16-5	16-5	16-5	16-5	
KC262793	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC333868	T68	T68	T68	VT	T68	VT	B165	VT	VT	B165	B165	B165	T68	B165
KC517485	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517486	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517487	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517488	T36	T36	T36	VT	T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
KC517489	T30	VT	T30	VT	VT	T30	T30	T30	T30	T30	T30	T30	T30	T30
KC517490	T30	VT	T30	VT	VT	T30	T30	T30	T30	T30	T30	T30	T30	T30
KC517491	T30	VT	T30	VT	VT	T30	T30	T30	T30	T30	T30	T30	T30	T30
KC517492	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC517493	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC517494	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT	VT
KC525952	T3	VT	VT	VT	T3	16-5	T3	T3	16-5	T3	T3	T3	T3	16-5
NC001661	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
U16304	T36	T36	T36		T36	T36	T36	T36	T36	T36	T36	T36	T36	T36
Y18420	T30	VT	T30	VT	VT	T30	T30	T30	T30	T30	T30	T30	T30	T30

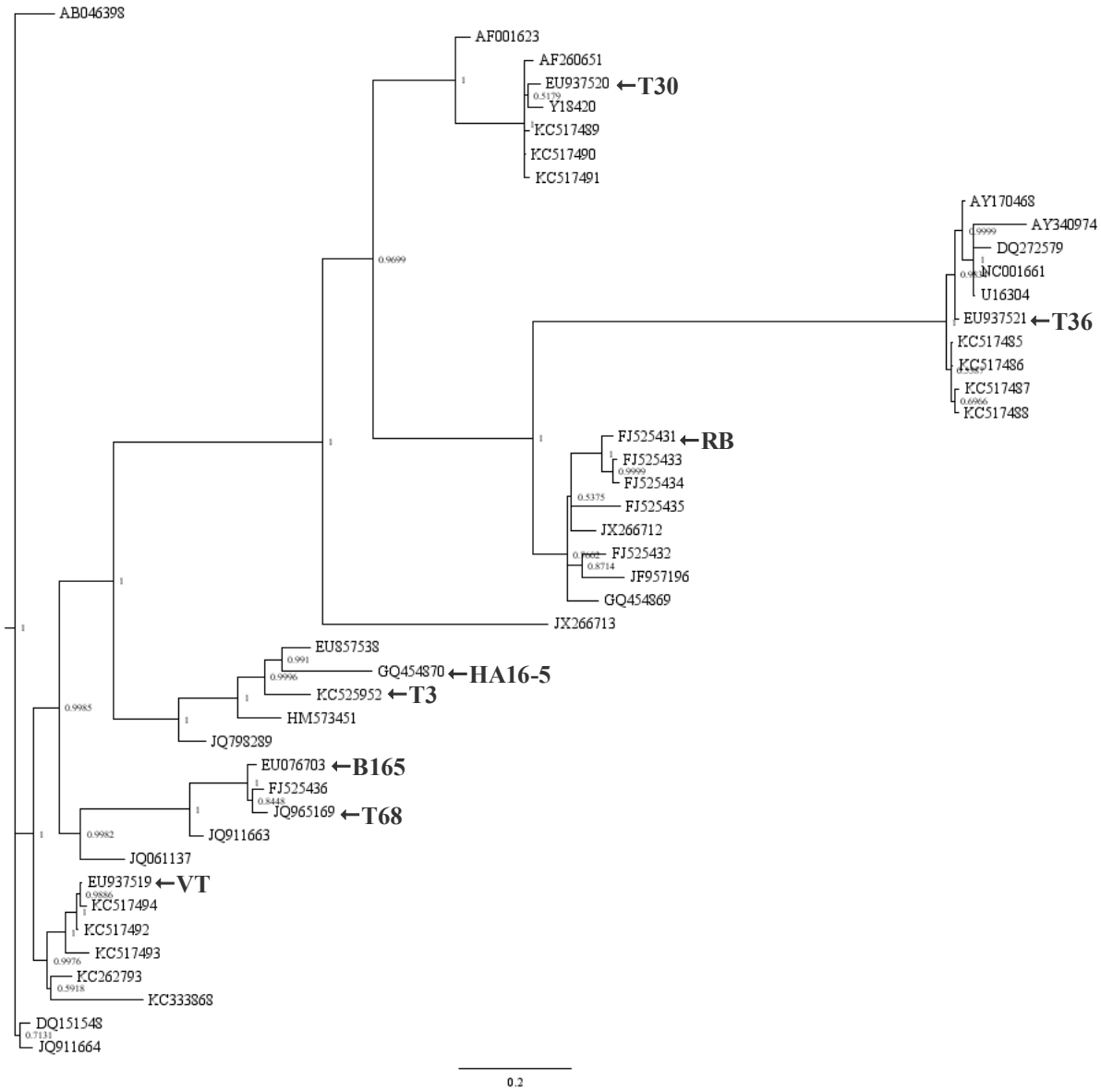

Figure 2.21: Phylogenetic dendrogram combining the 3' UTR and ORF 2 nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 3). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

2) Combination 7: ORF 2, ORF 4, ORF 7, Protease I & II, DUF 3614 and Methyltransferase

The combination 7 dendrogram (Figure 2.22) is very similar to the complete genome dendrogram, all representative isolates group individually except for the B165 and T68 isolates.

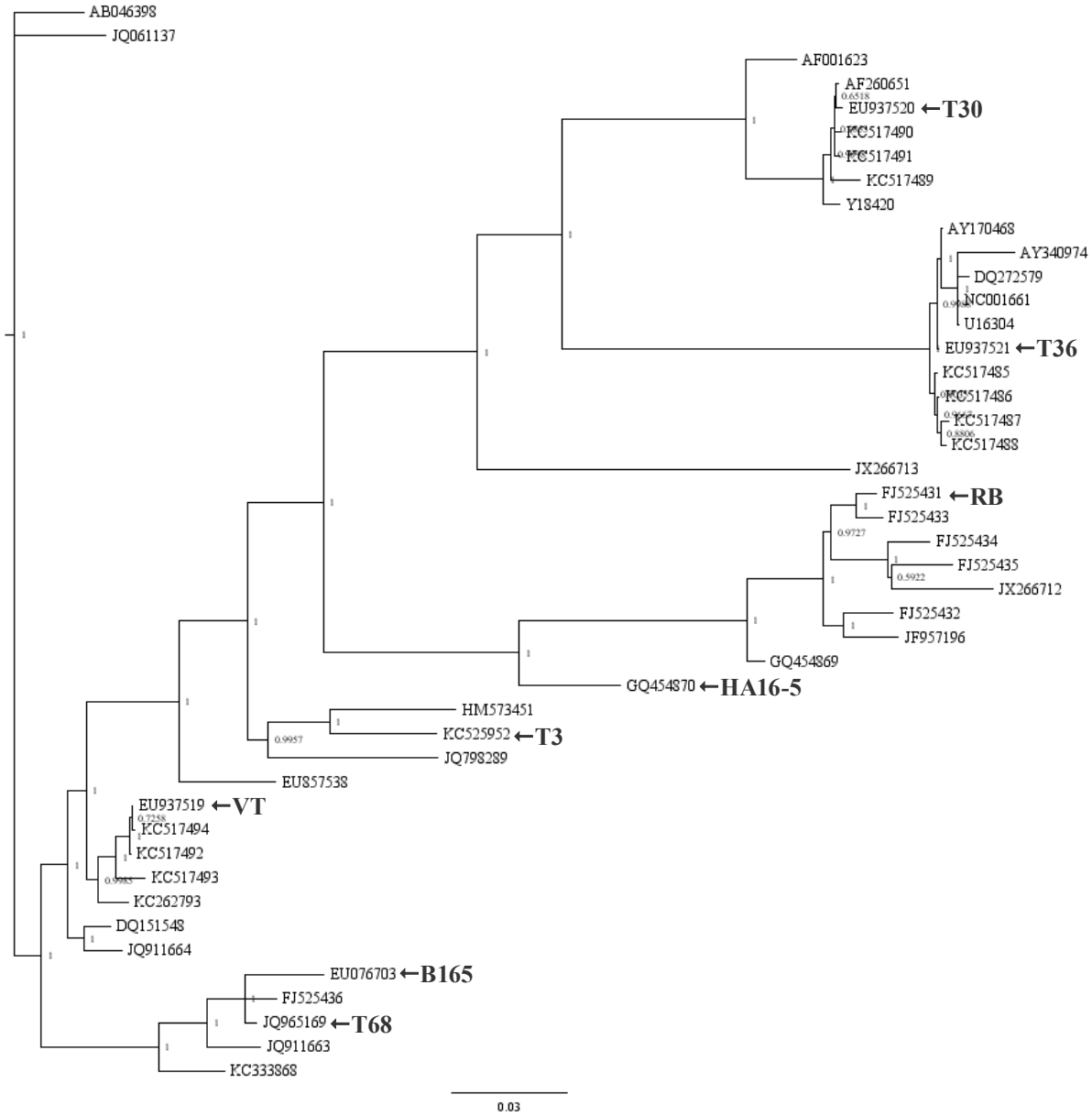

Figure 2.22: Phylogenetic dendrogram combining the ORF 2, ORF 4, ORF 7, Protease I & II, DUF 3614 and Methyltransferase nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 7). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

Again there are some individual isolates that group differently; JX266713 which usually groups with GQ45470 (HA16-5) now groups individually, HM573451 which grouped near the VT clade is now located within the T3 clade, JQ798289 which grouped alone is also within the T3 clade, EU857538 that grouped within the T3 clade now forms its own clade but it is still positioned near the T3 clade. JQ911663 which grouped alone is now

within the B165/T68 cluster. Although this combination of gene regions is a good representation of the existing genotypes, the large number of fragments requiring sequencing prevents it being feasible for genotyping.

3) Combination 8: ORF 2, ORF 4 and ORF 7

All representative isolates group individually except for the B165 and T68 isolates which formed one clade, as well as the HA16-5 isolate that clusters within the RB clade (Figure 2.23). There are some individual isolates that group differently to that within the whole genome dendrograms; JX266713 which usually groups with GQ45470 (HA16-5) now groups individually, HM573451 which grouped near the VT clade is now located within the T3 clade, JQ798289 which grouped alone is also within the T3 clade, EU857538 that grouped within the T3 clade is now alone but is still positioned near the T3 clade and JQ911663 and JQ061137 which grouped alone are now within the B165/T68 clade, and KC33868 which grouped within the T68 clade now clusters with the VT clade.

4) Combination 15: ORF 2 and DUF3614

Using this combination all representative isolates group individually except for the B165 and T68 isolates as well as the HA16-5 that clusters with the T3 clade (Figure 2.24). There are some individual isolates that group differently; JX266713 which usually groups with GQ45470 (HA16-5) now groups individually, HM573451 which grouped near the VT clade is now located within the T3 clade, EU857538 that grouped with the T3 isolate is now alone but is still positioned near the T3 clade, JQ911663 which grouped alone is now with the B165/T68 clade, and KC33868 which grouped with the T68 isolate now groups with the VT clade.

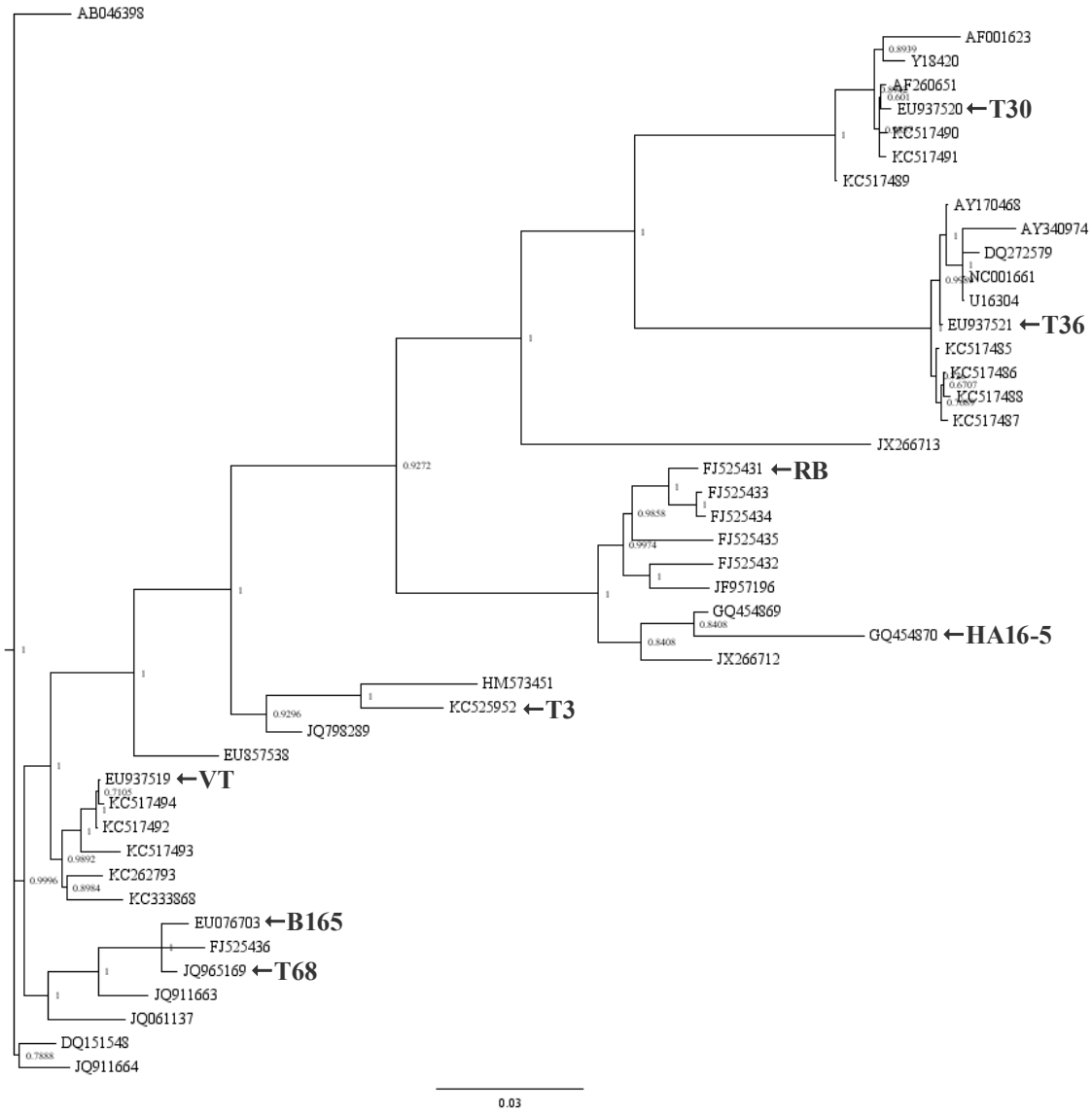

Figure 2.23: Phylogenetic dendrogram combining the ORF 2, 4 and 7 nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 8). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

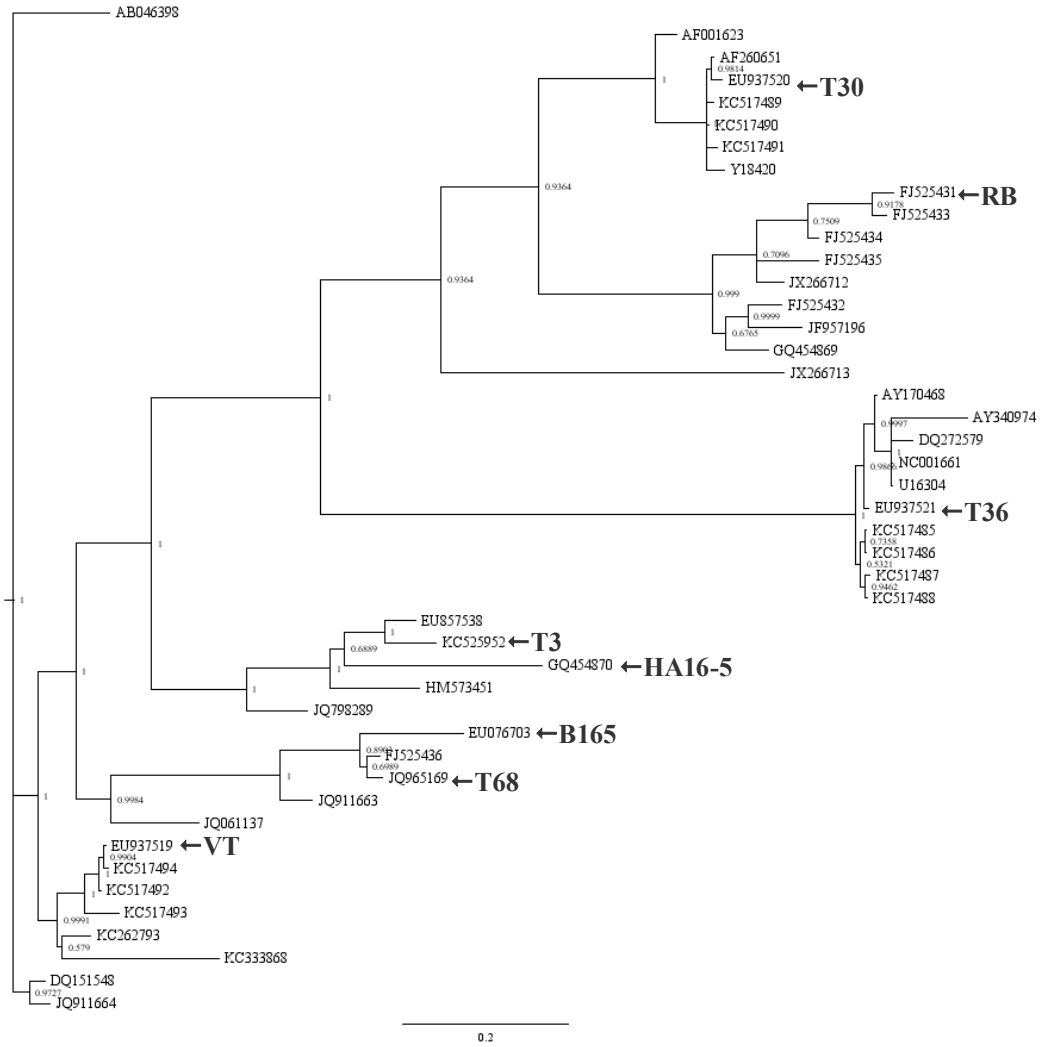

Figure 2.24: Phylogenetic dendrogram combining the ORF 2 and the DUF 3614 nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 15). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

5) Combination 22: ORF 1b, ORF 2, ORF 5, DUF 3614, Protease II, Methyltransferase and Helicase

This dendrogram is very similar to the complete genome dendrogram except for the B165/T68 clade (Figure 2.25). Although this combination produces a dendrogram very similar to the desired result, too many fragments were used and therefore it is not suitable for genotyping.

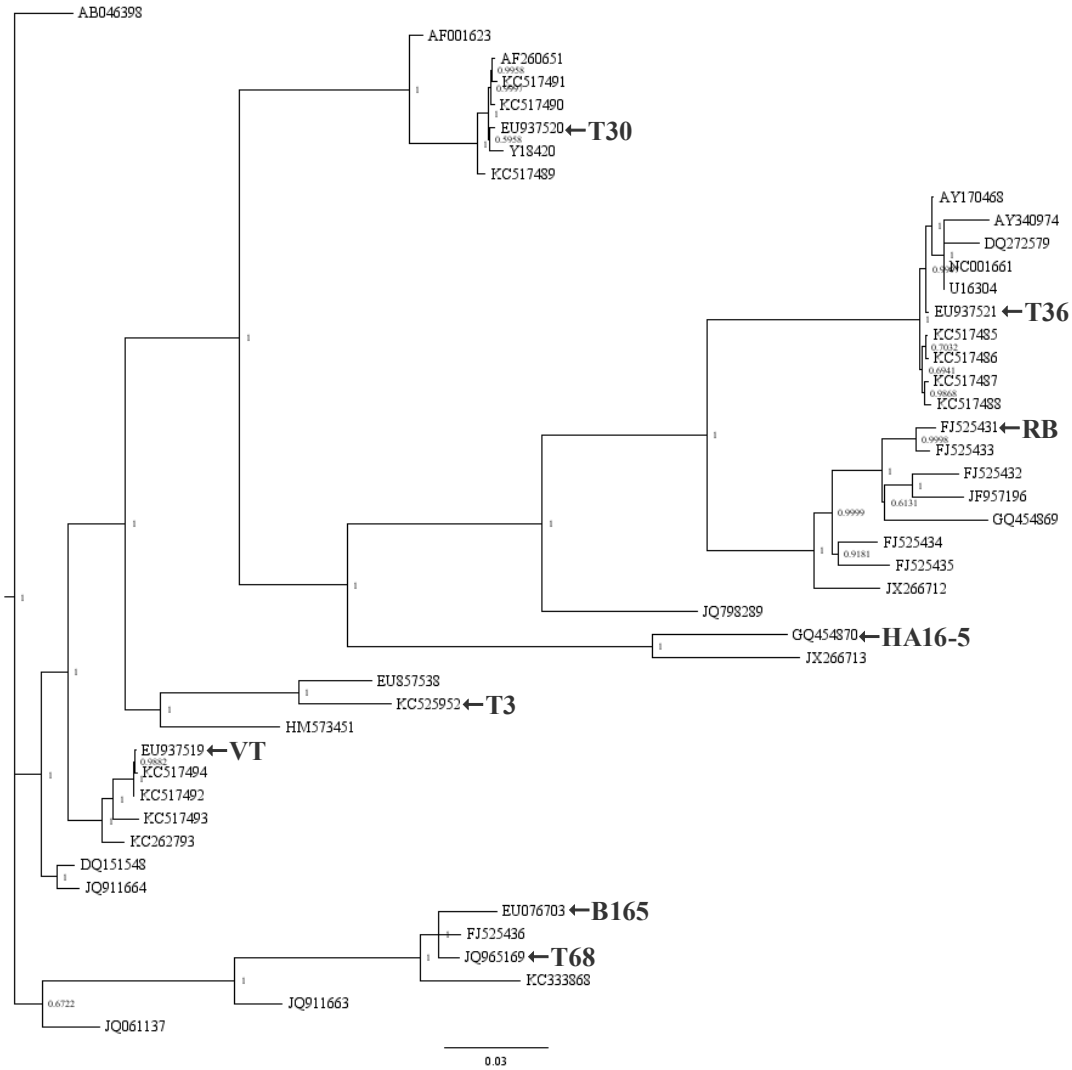

Figure 2.25: Phylogenetic dendrogram combining the ORF 2, 4 and 7 nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 22). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

6) Combination 23: ORF 1b, ORF 2 and ORF 5

This combination also produces a dendrogram very similar to the complete genome dendrogram (Figure 2.26). Again the T68 and B165 isolates form one clade and HM573451 groups closer to the T3 isolate than the VT clade. The T30 and VT clades are positioned further away from each other than in the complete genome dendrogram, with the T30 clade being closer to the T36 and RB clades, but this is a very promising combination.

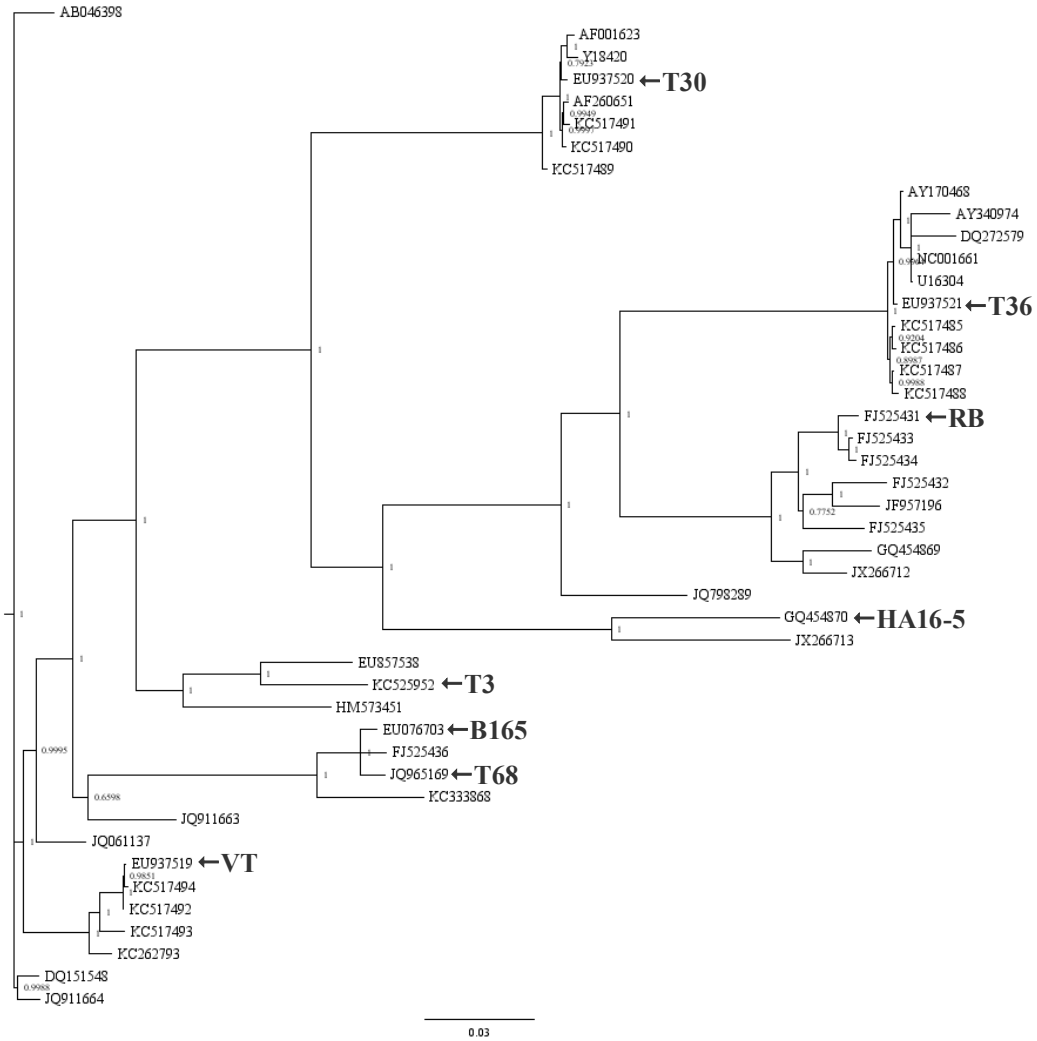

Figure 2.26: Phylogenetic dendrogram combining the ORF 1b, 2 and 5 nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 23). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

7) Combination 25: ORF 1b and ORF 5

This combination (Figure 2.27) also produces a dendrogram very similar to the complete genome dendrogram and is very similar to the dendrogram obtained from combination 23. Again the T68 and B165 isolates forms one clade. HM573451 groups individually but closer to the B165/T68 clade than the VT clade. The T30 and VT clades are positioned further away from each other unlike the complete genome dendrogram, with the T30 clade being closer to the T36 and RB clades, but this is also a very promising combination.

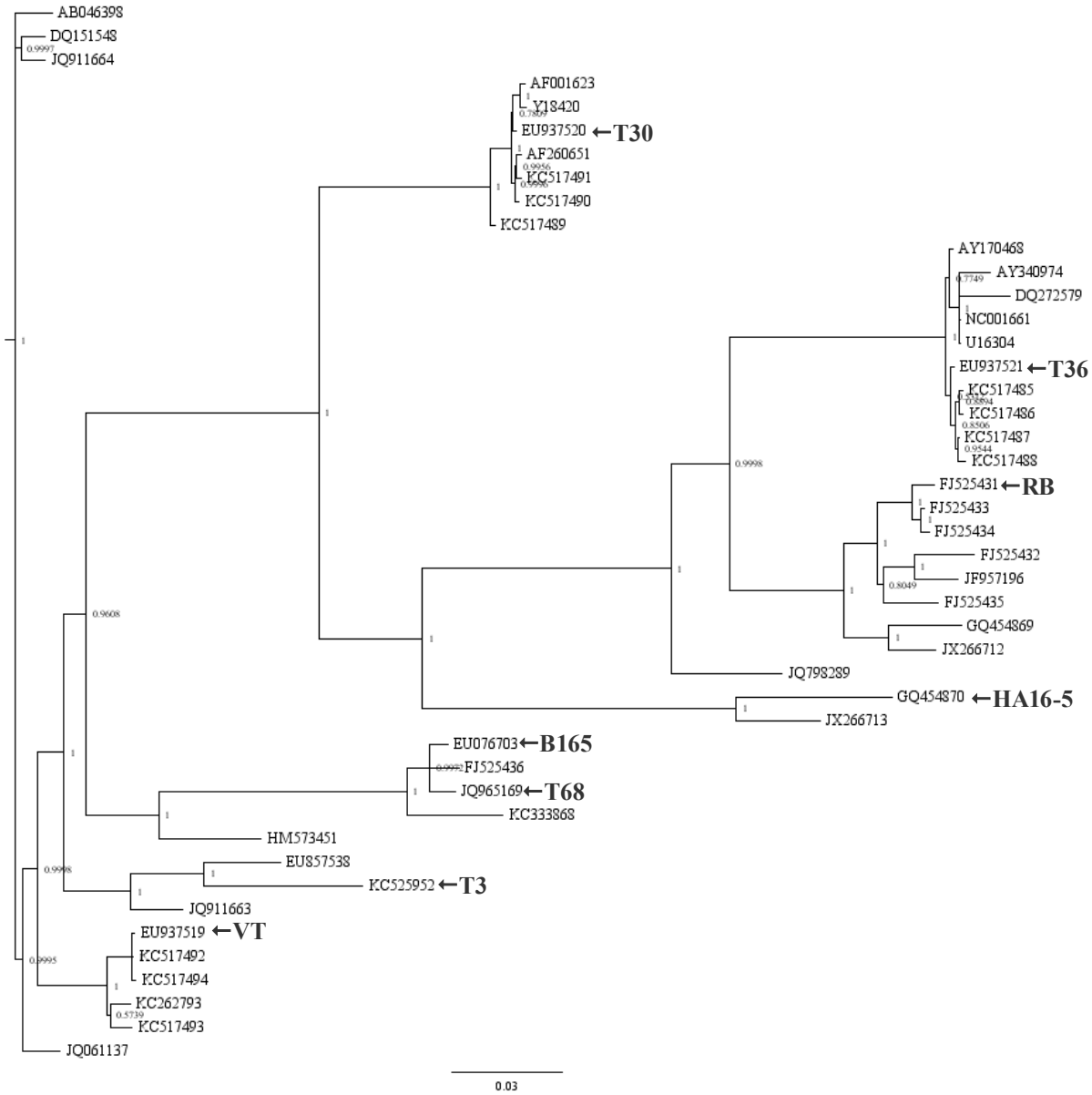

Figure 2.27: Phylogenetic dendrogram combining the ORF 1b and 5 nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 25). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

8) Combination 27: DUF3614, Protease II, Methyltransferase and Helicase

All the representative isolates group individually (Figure 2.28). Similar to the methyltransferase dendrogram, three of the isolates usually grouping within the RB clade now group with the T36 clade. The FJ525436 isolate that group with the B165 isolate in the complete genome dendrogram now group with the T68 isolate, which for a difference is not grouping with the B165 isolate, although the position of the B165 and T68 isolates within the dendrogram still shows a very close relationship. The T3 isolate groups individually but closer to the T30 clade than the VT clade. The AF001623 isolate that group with the T30 clade in the complete genome dendrograms now group with the VT clade, and the DQ151548 and JQ911664 isolates that formed their own group now clusters within the VT clade.

9) Combination 28: ORF 2 and Helicase

As was the case with the Helicase dendrogram (Figure 2.20), three of the isolates usually grouping with the RB clade now clusters closer to the T36 clade, but forms their own clade (Figure 2.29). The T3 isolate clusters with the HA16-5 isolate. Also within this clade is HM573451 that grouped closer to the VT clade and JQ798289 that grouped alone in the complete genome dendrogram. JQ911663 that also previously grouped alone, now clusters with the B165 and T68 isolates.

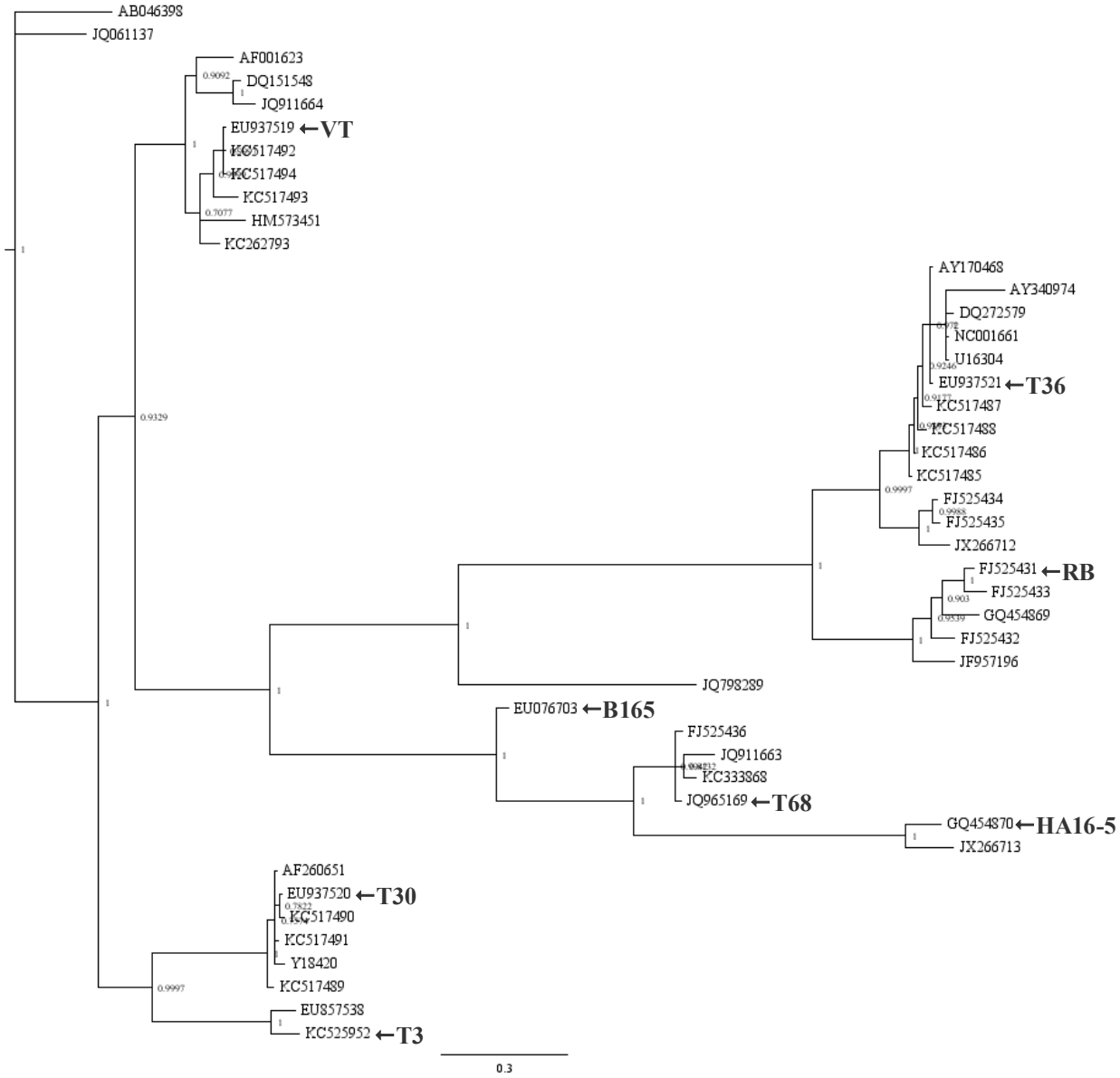

Figure 2.28: Phylogenetic dendrogram combining the DUF3614, Protease II, Methyl-transferase and Helicase nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 27). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

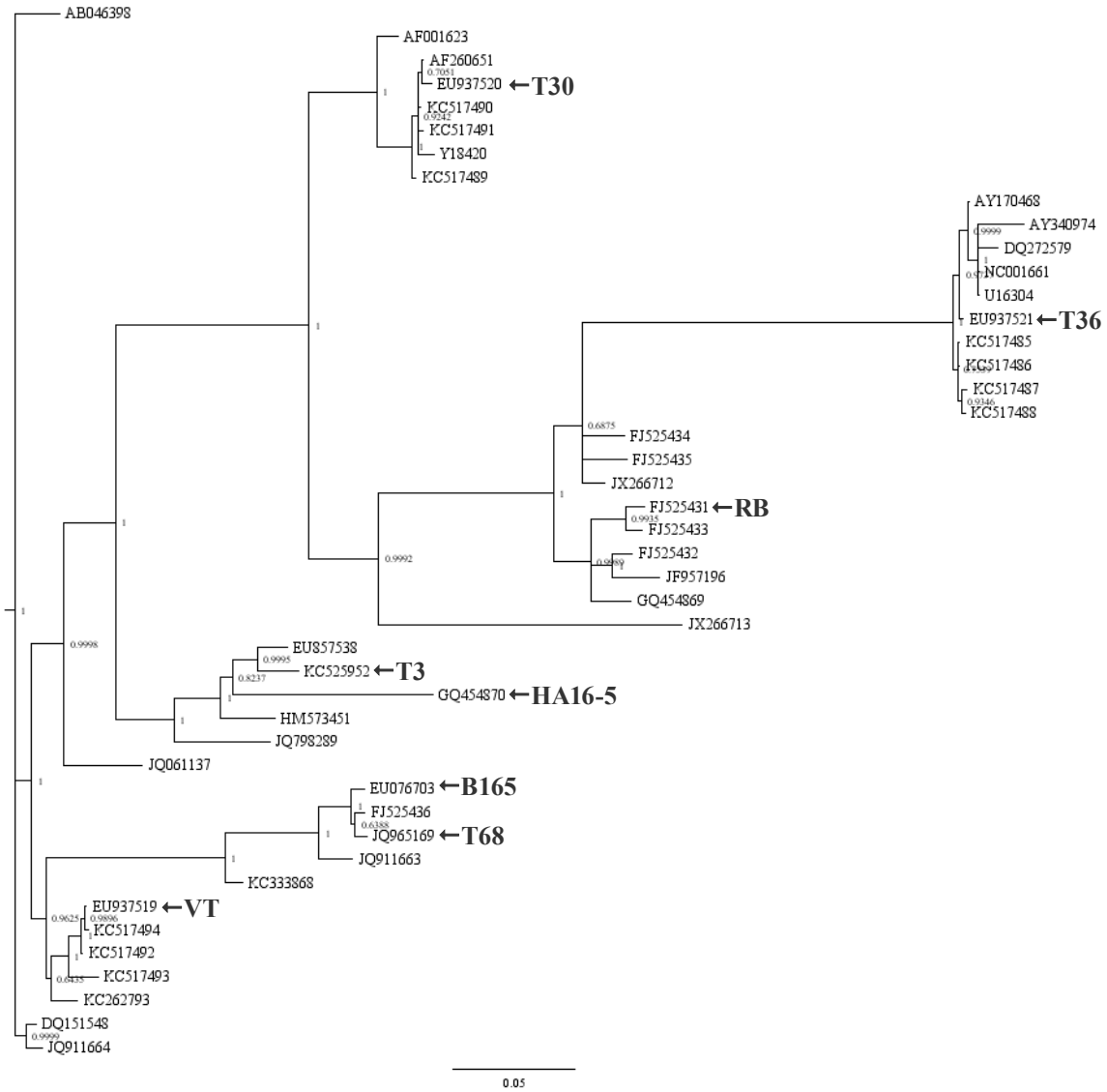

Figure 2.29: Phylogenetic dendrogram combining the ORF 2 and the Helicase nucleotide sequences of 45 CTV reference genomes based on Bayesian analysis (Combination 28). Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

Two combinations, 23 (ORF1b, ORF2 and ORF5) and 25 (ORF1b and ORF5), provided promising results, similar to the complete genome dendrogram but with minor differences. To determine which was better, these dendrograms were used in combination with the pairwise distance results (Table 2.8) to determine the most suitable combination. The pairwise distance analysis calculates the number of base differences between sequences. The same principle is used when NJ dendrograms are constructed, but in this

case the values were calculated to look at the amount of differences between isolates to see whether a trend can be observed.

While analyzing the results different threshold values were chosen based on visual inspection of when isolates will be grouped in different genotypes. When a threshold value of 0.06 is chosen, all the isolates can be classified into one of the reported genotypes, except for two, namely AF001623 and EU857538 (SP). AF001623 can either group with the VT or T30 genotype and EU857538 can either group with the VT or T3 genotype. According to the values, EU857538 is closer related to the T3 isolate (0.04) than the VT isolate (0.059). For AF001623 the situation is not so simple since values between it and both the VT and T30 isolates are very close together, ranging from 0.041 – 0.057 for VT isolates and 0.057 – 0.059 for T30 isolates. This would indicate that it represents an isolate belonging to the VT genotype, but if more T30 isolates become available, it might share a higher nucleotide similarity with the new isolate, and therefore a clear distinction cannot be made. One of the isolates, JQ798289 (A18) does not share a nucleotide distance value lower than 0.06 with any other isolate and this may indicate a new genotype. The lowest values are 0.076 and 0.084 with AB046398 and JQ061137, respectively. These two isolates are most closely related to other isolates from the VT genotype, but the distance is too great to be classified within this genotype. Therefore, JQ798289 may indicate a new genotype, A18. When using this very simplified approach one could conclude that there are thus a total of 9 CTV genotypes. These genotypes, together with the isolates grouping as part of them is proposed in Table 2.9.

Table 2.8: Nucleotide distance matrix between the 45 complete CTV genomes. Shaded cells indicate a distance of 0.06 or lower indicating a close relationship between the isolates. "16-5" indicates the HA 16-5 genotype.

JQ798289(A18)																																												
EU076703(B165)	0.13																																											
JQ911663(B165)	0.12	0.06																																										
FJ525436(B165)	0.13	0.03	0.06																																									
GQ454870(16-5)	0.17	0.16	0.16	0.16																																								
JX266713(16-5)	0.18	0.16	0.16	0.16	0.06																																							
FJ525434(RB)	0.18	0.22	0.22	0.22	0.21	0.22																																						
FJ525432(RB)	0.17	0.22	0.22	0.22	0.22	0.22	0.05																																					
FJ525433(RB)	0.18	0.22	0.22	0.22	0.21	0.22	0.03	0.03																																				
FJ525435(RB)	0.18	0.22	0.22	0.22	0.22	0.22	0.02	0.05	0.04																																			
JX266712(RB)	0.17	0.22	0.22	0.22	0.19	0.22	0.04	0.06	0.06	0.04																																		
JF957196(RB)	0.18	0.22	0.22	0.22	0.21	0.22	0.05	0.02	0.03	0.05	0.06																																	
GQ454869(RB)	0.17	0.22	0.22	0.22	0.18	0.22	0.06	0.05	0.05	0.06	0.04	0.05																																
FJ525431(RB)	0.18	0.22	0.22	0.22	0.21	0.22	0.03	0.03	0.01	0.04	0.06	0.03	0.05																															
KC525952(T3)	0.11	0.13	0.12	0.13	0.18	0.19	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22																														
EU857538(T3)	0.11	0.12	0.11	0.12	0.18	0.19	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.04																													
Y18420(T30)	0.13	0.15	0.14	0.15	0.19	0.18	0.21	0.21	0.21	0.21	0.22	0.21	0.21	0.21	0.11	0.11																												
AF260651(T30)	0.13	0.15	0.14	0.15	0.19	0.18	0.21	0.21	0.20	0.21	0.21	0.21	0.21	0.21	0.11	0.11	0.01																											
KC517491(T30)	0.13	0.15	0.14	0.15	0.19	0.18	0.21	0.21	0.20	0.21	0.21	0.21	0.21	0.21	0.11	0.11	0.01	0.00																										
KC517489(T30)	0.12	0.14	0.14	0.14	0.18	0.18	0.20	0.21	0.20	0.21	0.21	0.20	0.21	0.20	0.11	0.10	0.01	0.01	0.01																									
KC517490(T30)	0.13	0.15	0.14	0.15	0.19	0.18	0.21	0.21	0.20	0.21	0.21	0.21	0.21	0.21	0.11	0.11	0.01	0.00	0.00	0.01																								
EU937520(T30)	0.13	0.15	0.14	0.15	0.19	0.18	0.21	0.21	0.21	0.21	0.22	0.21	0.21	0.21	0.11	0.11	0.01	0.00	0.00	0.01	0.00																							
KC517486(T36)	0.17	0.22	0.22	0.23	0.23	0.21	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.23	0.22	0.21	0.21	0.21	0.20	0.21	0.21																						
KC517485(T36)	0.17	0.22	0.22	0.23	0.23	0.21	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.23	0.22	0.21	0.21	0.21	0.20	0.21	0.21	0.00																					
DQ272579(T36)	0.18	0.24	0.24	0.24	0.24	0.23	0.11	0.12	0.12	0.11	0.10	0.12	0.12	0.12	0.24	0.24	0.22	0.22	0.22	0.22	0.22	0.22	0.03	0.02																				
EU937521(T36)	0.17	0.22	0.22	0.23	0.23	0.21	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.23	0.23	0.21	0.21	0.21	0.21	0.21	0.21	0.00	0.00	0.02																			
KC517488(T36)	0.17	0.22	0.22	0.23	0.23	0.21	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.23	0.23	0.21	0.21	0.21	0.21	0.21	0.21	0.00	0.00	0.02	0.00																		
KC517487(T36)	0.17	0.22	0.22	0.23	0.23	0.21	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.23	0.22	0.21	0.21	0.21	0.20	0.21	0.21	0.00	0.00	0.02	0.00	0.00																	
NC001661(T36)	0.18	0.23	0.23	0.23	0.23	0.22	0.09	0.11	0.10	0.09	0.09	0.10	0.11	0.10	0.23	0.23	0.21	0.21	0.21	0.21	0.21	0.22	0.01	0.01	0.02	0.01	0.01	0.01																
AY340974(T36)	0.18	0.24	0.23	0.24	0.24	0.22	0.10	0.11	0.11	0.10	0.10	0.11	0.11	0.11	0.24	0.24	0.22	0.22	0.22	0.22	0.22	0.22	0.02	0.02	0.02	0.01	0.02	0.02	0.01															
AY170468(T36)	0.17	0.22	0.22	0.23	0.23	0.21	0.09	0.10	0.10	0.09	0.09	0.10	0.10	0.10	0.23	0.23	0.21	0.21	0.21	0.21	0.21	0.21	0.01	0.01	0.02	0.00	0.01	0.01	0.01	0.01														
U16304(T36)	0.18	0.23	0.23	0.23	0.23	0.22	0.09	0.11	0.10	0.09	0.09	0.10	0.11	0.10	0.23	0.23	0.21	0.21	0.21	0.21	0.21	0.22	0.01	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.01													
KC333868(T68)	0.17	0.08	0.09	0.08	0.13	0.13	0.22	0.23	0.22	0.23	0.23	0.22	0.23	0.22	0.17	0.15	0.19	0.19	0.19	0.18	0.19	0.19	0.23	0.23	0.24	0.23	0.23	0.23	0.23	0.24	0.23	0.23												
JQ965169(T68)	0.17	0.06	0.08	0.06	0.13	0.13	0.23	0.22	0.22	0.23	0.23	0.22	0.23	0.22	0.17	0.16	0.19	0.19	0.19	0.18	0.19	0.19	0.23	0.23	0.24	0.23	0.23	0.23	0.23	0.24	0.23	0.23	0.03											
KC517493(VT)	0.10	0.10	0.09	0.09	0.18	0.18	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.09	0.06	0.10	0.10	0.10	0.09	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.22	0.23	0.22	0.22	0.14	0.15										
KC517492(VT)	0.10	0.10	0.09	0.09	0.18	0.18	0.21	0.21	0.21	0.22	0.22	0.22	0.22	0.22	0.09	0.06	0.10	0.10	0.10	0.09	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.23	0.22	0.23	0.14	0.15	0.01									
KC517494(VT)	0.10	0.10	0.09	0.09	0.18	0.18	0.21	0.21	0.22	0.22	0.22	0.22	0.22	0.22	0.09	0.06	0.10	0.10	0.10	0.09	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.23	0.22	0.23	0.14	0.15	0.01	0.00								
EU937519(VT)	0.10	0.10	0.09	0.09	0.18	0.18	0.21	0.21	0.22	0.22	0.22	0.22	0.22	0.22	0.09	0.06	0.10	0.10	0.10	0.09	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.23	0.22	0.23	0.14	0.15	0.01	0.00	0.00							
DQ151548(VT)	0.09	0.10	0.08	0.10	0.18	0.18	0.21	0.21	0.21	0.21	0.22	0.21	0.22	0.21	0.09	0.07	0.10	0.10	0.10	0.10	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.22	0.23	0.22	0.22	0.15	0.14	0.04	0.04	0.04	0.04						
AF001623(VT)	0.09	0.12	0.10	0.12	0.18	0.17	0.20	0.20	0.20	0.20	0.21	0.20	0.21	0.20	0.10	0.09	0.06	0.06	0.06	0.06	0.06	0.06	0.20	0.20	0.21	0.20	0.20	0.20	0.21	0.21	0.20	0.21	0.17	0.17	0.06	0.06	0.06	0.06	0.04					
HM573451(VT)	0.10	0.11	0.11	0.11	0.17	0.19	0.22	0.22	0.22	0.22	0.21	0.22	0.21	0.22	0.08	0.09	0.11	0.11	0.11	0.10	0.11	0.11	0.23	0.23	0.24	0.23	0.23	0.23	0.23	0.24	0.23	0.23	0.17	0.17	0.05	0.05	0.05	0.05	0.06	0.07				
KC262793(VT)	0.10	0.10	0.10	0.09	0.18	0.18	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.09	0.07	0.10	0.10	0.10	0.10	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.23	0.22	0.23	0.14	0.15	0.02	0.02	0.02	0.02	0.04	0.06	0.05			
JQ911664(VT)	0.09	0.10	0.08	0.10	0.18	0.19	0.21	0.21	0.21	0.22	0.22	0.21	0.22	0.21	0.09	0.07	0.10	0.10	0.10	0.10	0.10	0.10	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.23	0.22	0.23	0.15	0.14	0.04	0.04	0.04	0.04	0.01	0.04	0.06	0.04		
JQ061137(VT)	0.09	0.10	0.08	0.10	0.18	0.19	0.22	0.21	0.22	0.22	0.22	0.22	0.22	0.22	0.09	0.08	0.11	0.11	0.11	0.10	0.11	0.11	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.23	0.22	0.23	0.15	0.15	0.05	0.05	0.05	0.05	0.03	0.06	0.07	0.05	0.03	
AB046398(VT/T30)	0.08	0.10	0.09	0.10	0.19	0.19	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.09	0.08	0.11	0.11	0.11	0.10	0.11	0.11	0.22	0.22	0.23	0.22	0.22	0.22	0.23	0.24	0.22	0.23	0.15	0.15	0.04	0.04	0.05	0.05	0.02	0.05	0.07	0.05	0.02	0.04

Table 2.9: Proposed classification of CTV isolates into genotypes based on pairwise distance analysis.

Accession Number	Description	Genotype
JQ798289.1	CTV isolate A18, complete genome	A18
EU076703.3	CTV isolate B165, complete genome	B165
FJ525436.1	CTV isolate NZ-B18, complete genome	
JQ911663.1	CTV isolate CT14A, complete genome	
GQ454870.1	CTV strain HA16-5, complete genome	HA 16-5
JX266713.1	CTV isolate Taiwan-Pum/M/T5, complete genome	
FJ525431.1	CTV isolate NZRB-M12, complete genome	RB
FJ525432.1	CTV isolate NZRB-G90, complete genome	
FJ525433.1	CTV isolate NZRB-TH28, complete genome	
FJ525434.1	CTV isolate NZRB-TH30, complete genome	
FJ525435.1	CTV isolate NZRB-M17, complete genome	
GQ454869.1	CTV strain HA18-9, complete genome	
JF957196.1	CTV isolate B301, complete genome	
JX266712.1	CTV isolate Taiwan-Pum/SP/T1, complete genome	
EU857538.1	CTV strain SP, complete genome	T3
KC525952.1	CTV isolate T3, complete genome	
AF260651.1	CTV T30, complete genome	T30
EU937520.1	CTV strain T30, complete genome	
KC517489.1	CTV isolate FS701-T30, complete genome	
KC517490.1	CTV isolate FL278-T30, complete genome	
KC517491.1	CTV isolate FS703-T30, complete genome	
Y18420.1	CTV complete genome, isolate T385	
AY170468.1	CTV, complete genome	T36
AY340974.1	CTV isolate Qaha from Egypt, complete genome	
DQ272579.1	CTV from Mexico, complete genome	
EU937521.1	CTV strain T36, complete genome	
KC517485.1	CTV isolate FS674-T36, complete genome	
KC517486.1	CTV isolate FS701-T36, complete genome	
KC517487.1	CTV isolate FS703-T36, complete genome	
KC517488.1	CTV isolate FS577, complete genome	
NC001661.1	CTV, complete genome	
U16304.1	CTU16304 CTV complete genome	
JQ965169.1	CTV isolate T68-1, complete genome	T68
KC333868.1	CTV strain CT-ZA3, complete genome	
AB046398.1	CTV genomic RNA, complete genome, seedling yellows strain	VT / T30
AF001623.1	CTAF001623 CTV, complete genome	VT
DQ151548.1	CTV strain T318A, complete genome	
EU937519.1	CTV strain VT, complete genome	
HM573451.1	CTV isolate Kpg 3, complete genome	
JQ061137.1	CTV isolate AT-1, complete genome	
JQ911664.1	CTV isolate CT11A, complete genome	
KC262793.1	CTV isolate L192GR, complete genome	
KC517492.1	CTV isolate FS703-VT, complete genome	
KC517493.1	CTV isolate FL202-VT, complete genome	
KC517494.1	CTV isolate FS701-VT, complete genome	

The results obtained in the pairwise distance analysis (summarized in Table 2.9) mostly supports the phylogenetic relationship of the isolates when using ORF1b, 2 and 5 (Combination 23) for the analysis. The minor differences that occur is that the T68 isolates (JQ965169 and KC333868) and the B165 isolates (EU076703, JQ911663 and FJ525436) group very closely together, and the Kpg3 isolate (HM573451) grouping closer to the T3 genotype than the VT genotype. For the T68 (JQ965169) isolate, the nucleotide distance to all three B165 isolates (0.06/0.07) is very close to the cut-off value of 0.06. But the KC333868 isolate shares a very small (0.03) nucleotide distance with JQ965169 (T68 representative) and a much higher nucleotide distance with the B165 isolates (0.07/0.08). The Kpg3 isolate (HM573451) shares a nucleotide distance of 0.08/0.07 with the T3 isolates and a nucleotide distance of 0.040.0 6 8with the VT isolates. This discrepancy may be explainable with a recombination analysis, but considering the very diverse nature of CTV, these discrepancies would be a very small compromise when looking at an efficient method for genotyping new isolates. Unfortunately recombination will always complicate the process, and using three genes to do genotyping is the best option when the entire genome cannot be analyzed.

Recombination analysis was performed on the 45 complete genomes and as expected multiple recombination events could be identified in most of the sequences based on the detection of the event by three or more algorithms. A total of 72 recombination events were obtained after analysis of the results, each occurring in one or many of the 45 genomes. We specifically focused on the six genomes that occurred as "anomalies" in the phylogenetic dendrograms. A graphic representation of these six genomes are shown in Figure 2.30, and supporting information is given in Table 2.10. For each recombination event multiple minor or major parents are given. The parent listed is not necessarily the actual parent, but usually just a sequence similar to the real parent (Martin, D.P., *personal*

communication). Starting at the B165 and T68 genotypes, it can be seen that JQ965169, KC333868 and JQ911663 all share recombination events 43 and 36 and all but JQ911663 share recombination event 51. These recombination events all include VT, T3 and T30 isolates, and the common recombination events might be why they group together in the combination dendrogram. EU076703 has an array of recombination events, but some (4 and 6) also include recombination with VT and T3 strains, and one of the recombination events (1) indicate recombination with a B165 strain that also support the fact that these strains all cluster together.

The recombination evidence for HM573451 shows that recombination event 48 introduced a small fragment from a VT isolate (minor parent) and the major parent is indicated as a T3 isolate. This may be the reason for Kpg3 grouping closer to T3 instead of VT in the combined phylogenetic dendrogram. AF001623 is the isolate that is very similar to isolates from the VT and T30 genotypes. Recombination events 5 and 26 indicate recombination between VT and T30 isolates, and this intermixing of isolates may be the reason for the similarity of the strain to both genotypes.

The recombination events discussed in this chapter is just the tip of the iceberg, and the entire picture is actually much more complicated and beyond the scope of this study. The complete table of recombination events, as well as all the phylogenetic dendrograms for the 45 genomes can be found in Appendix C.

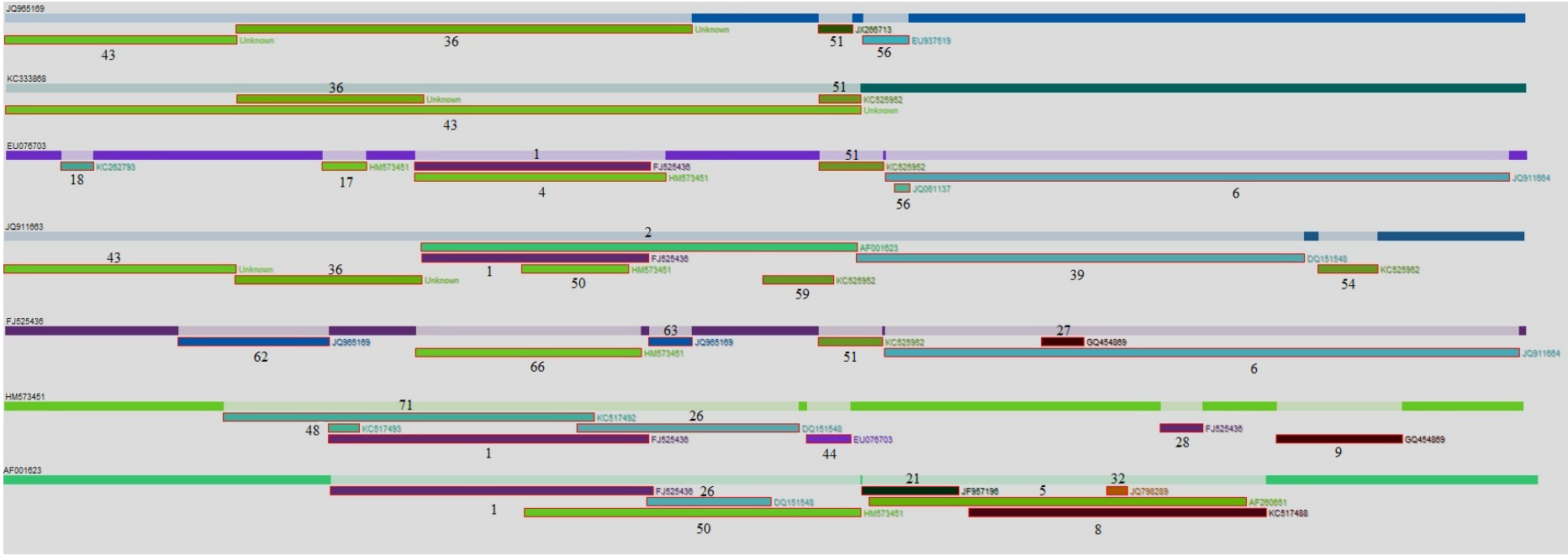

Figure 2.30: Recombination analysis of 7 CTV genomes (from the top JQ965169, KC333868, EU076703, JQ911663, FJ525436, HM573451 and AF001623) showing anomalies in the grouping with genotypes in either the Combination 23 phylogenetic dendrogram or during nucleotide distance calculations. Each genome is indicated as a rectangular strip. Additional strips under the genome indicate possible recombination events. For each recombination event multiple minor or major parents can be given. The parent listed is never the actual parent, but usually just a sequence similar to the real parent. Figure only indicative of overview of recombination events, please examine numbers that refer to recombination events described in Table 2.10.

Table 2.10: A description of recombination events shown in Figure 2.30. * indicates that the actual breakpoint position is undetermined (it was most likely overprinted by a subsequent recombination event). The minor parent is the parent contributing the smaller fraction of sequence and the major parent is the parent contributing the larger fraction of sequence. ? indicates where only one parent and a recombinant is in the alignment and the sequence listed as ? was used to infer the existence of a missing parental sequence.

Event Nr	Begin position	End position	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)	Event Nr	Begin position	End position	Recombinant Sequence(s)	Minor Parental Sequence(s)	Major Parental Sequence(s)
1	4148	8233	AF001623(VT/T30)	FJ525436(B165)	?(JQ965169(T68))	27	13261	13786	FJ525436(B165)	GQ454869(RB)	JQ911664(VT)
			EU076703(B165)		?(KC333868(T68))					FJ525432(RB)	AB046398(VT)
			HM573451(VT)							FJ525434(RB)	DQ151548(VT)
			JQ911663(B165)							FJ525435(RB)	EU857538(T3)
2	5336	10906	JQ911663(B165)	AF001623(VT/T30)	JQ965169(T68)					GQ454870(HA16-5)	EU937519(VT)
					EU076703(B165)					JF957196(RB)	JQ061137(VT)
					KC333868(T68)					JX266712(RB)	JQ911663(B165)
4	5235	8437	EU076703(B165)	HM573451(VT)	JQ965169(T68)						KC262793(VT)
				AF001623(VT/T30)	KC333868(T68)						KC333868(T68)
				AF260651(T30)							KC517492(VT)
				EU857538(T3)							KC517493(VT)
				EU937520(T30)							KC517494(VT)
				KC517489(T30)		28	14790*	15328	HM573451(VT)	FJ525436(B165)	KC525952(T3)
				KC517490(T30)						AB046398(VT)	
				KC517491(T30)						DQ151548(VT)	
				KC525952(T3)						EU076703(B165)	
				Y18420(T30)						JQ061137(VT)	
5	10982*	15740	AF001623(VT/T30)	AF260651(T30)	DQ151548(VT)					JQ911663(B165)	
				EU937520(T30)	AB046398(VT)					JQ911664(VT)	
				KC517489(T30)	JQ061137(VT)					JQ965169(T68)	
				KC517490(T30)	JQ911664(VT)					KC517493(VT)	
				KC517491(T30)		32	13972	14282	AF001623(VT/T30)	JQ798289(A18)	?(KC517493(VT))
				Y18420(T30)						HM573451(VT)	?(EU937519(VT))
6	11250	19344*	FJ525436(B165)	JQ911664(VT)	?(HM573451(VT))						?(JQ061137(VT))
			EU076703(B165)	AB046398(VT)							?(KC262793(VT))
				DQ151548(VT)							?(KC517492(VT))
				EU937519(VT)							?(KC517494(VT))
				JQ061137(VT)		36	2961*	8782*	JQ965169(T68)	?(AF260651(T30))	AB046398(VT)
				KC517492(VT)					JQ911663(B165)	?(EU937520(T30))	DQ151548(VT)
				KC517494(VT)					KC333868(T68)	?(KC517489(T30))	EU937519(VT)
8	12290	16466*	AF001623(VT/T30)	KC517488(T36)	KC525952(T3)					?(KC517490(T30))	JQ061137(VT)
				AY170468(T36)	HM573451(VT)					?(KC517491(T30))	JQ911664(VT)
				AY340974(T36)						?(KC525952(T3))	KC262793(VT)
				DQ272579(T36)						?(Y18420(T30))	KC517492(VT)
				EU937521(T36)							KC517493(VT)

				KC517485(T36)							KC517494(VT)
				KC517486(T36)		39	10907*	16617*	JQ911663(B165)	DQ151548(VT)	? (KC525952(T3))
				KC517487(T36)						EU937519(VT)	
				NC001661(T36)						JQ061137(VT)	
				U16304(T36)						JQ911664(VT)	
9	16278	17872	HM573451(VT)	GQ454869(RB)	AF001623(VT/T30)					KC262793(VT)	
				GQ454870(HA16-5)	AB046398(VT)					KC517492(VT)	
				JX266712(RB)	AF260651(T30)					KC517493(VT)	
					DQ151548(VT)					KC517494(VT)	
					EU076703(B165)	43	1*	10922*	KC333868(T68)	? (HM573451(VT))	KC517492(VT)
					EU857538(T3)				JQ911663(B165)		DQ151548(VT)
					EU937519(VT)				JQ965169(T68)		EU937519(VT)
					EU937520(T30)						JQ911664(VT)
					FJ525436(B165)						KC262793(VT)
					JQ061137(VT)						KC517493(VT)
					JQ798289(A18)						KC517494(VT)
					JQ911663(B165)	44	10268	10818*	HM573451(VT)	EU076703(B165)	JQ911664(VT)
					JQ911664(VT)					FJ525436(B165)	AB046398(VT)
					JQ965169(T68)					JQ965169(T68)	AF001623(VT/T30)
					KC262793(VT)					KC333868(T68)	AF260651(T30)
					KC333868(T68)						DQ151548(VT)
					KC517489(T30)						EU857538(T3)
					KC517490(T30)						EU937519(VT)
					KC517491(T30)						EU937520(T30)
					KC517492(VT)						JQ061137(VT)
					KC517493(VT)						JQ911663(B165)
					KC517494(VT)						KC262793(VT)
					KC525952(T3)						KC517489(T30)
					Y18420(T30)						KC517490(T30)
17	4048	4610	EU076703(B165)	HM573451(VT)	JQ965169(T68)						KC517491(T30)
				AB046398(VT)	JQ911663(B165)						KC517492(VT)
				AF001623(VT/T30)	KC333868(T68)						KC517493(VT)
				AF260651(T30)							KC517494(VT)
				DQ151548(VT)							KC525952(T3)
				EU857538(T3)							Y18420(T30)
				EU937519(VT)		48	4148*	4532	HM573451(VT)	KC517493(VT)	KC525952(T3)
				EU937520(T30)						EU937519(VT)	
				JQ061137(VT)						KC262793(VT)	
				JQ798289(A18)						KC517492(VT)	
				JQ911664(VT)						KC517494(VT)	
				KC262793(VT)		50	6613	10857*	AF001623(VT/T30)	HM573451(VT)	? (EU857538(T3))
				KC517489(T30)					JQ911663(B165)	FJ525436(B165)	?(KC525952(T3))

				KC517490(T30)		51	10409	11219*	EU076703(B165)	KC525952(T3)	JX266713(HA16-5)
				KC517491(T30)					FJ525436(B165)	EU857538(T3)	
				KC517492(VT)					JQ965169(T68)		
				KC517493(VT)					KC333868(T68)		
				KC517494(VT)		52	11257	13576	AB046398(VT)	EU937519(VT)	JQ061137(VT)
				KC525952(T3)						KC262793(VT)	
				Y18420(T30)						KC333868(T68)	
18	706	1114	EU076703(B165)	KC262793(VT)	JQ965169(T68)					KC517492(VT)	
				AB046398(VT)	FJ525436(B165)					KC517493(VT)	
				AF001623(VT/T30)	JQ911663(B165)					KC517494(VT)	
				AF260651(T30)	KC333868(T68)	54	16812*	17556	JQ911663(B165)	KC525952(T3)	FJ525436(B165)
				DQ151548(VT)							AB046398(VT)
				EU857538(T3)							AF001623(VT/T30)
				EU937519(VT)							DQ151548(VT)
				EU937520(T30)							EU076703(B165)
				HM573451(VT)							JQ061137(VT)
				JQ911664(VT)							JQ911664(VT)
				KC517489(T30)							JQ965169(T68)
				KC517490(T30)		56	10973*	11547*	JQ965169(T68)	EU937519(VT)	JQ061137(VT)
				KC517491(T30)						KC262793(VT)	
				KC517492(VT)						KC517492(VT)	
				KC517493(VT)						KC517493(VT)	
				KC517494(VT)						KC517494(VT)	
				KC525952(T3)		59	9704	10599*	JQ911663(B165)	KC525952(T3)	JQ061137(VT)
				Y18420(T30)						EU857538(T3)	JQ911664(VT)
21	10858*	12199*	AF001623(VT/T30)	JF957196(RB)	HM573451(VT)	62	2221*	4145	FJ525436(B165)	JQ965169(T68)	?(GQ454870(HA16-5))
				FJ525431(RB)	KC525952(T3)					KC333868(T68)	?(JX266713(HA16-5))
				FJ525432(RB)		63	8232	8782*	FJ525436(B165)	JQ965169(T68)	?(JX266713(HA16-5))
				FJ525433(RB)						KC333868(T68)	?(GQ454870(HA16-5))
				FJ525434(RB)		66	5258	8131*	FJ525436(B165)	HM573451(VT)	?(KC517489(T30))
				FJ525435(RB)							?(AF260651(T30))
				GQ454869(RB)							?(EU937520(T30))
				JX266712(RB)							?(KC517490(T30))
26	8160*	9722	AF001623(VT/T30)	DQ151548(VT)	?(Y18420(T30))						?(KC517491(T30))
			HM573451(VT)	AB046398(VT)	?(EU937520(T30))						?(Y18420(T30))
				HM573451(VT)	?(KC517489(T30))	71	2803	7536*	HM573451(VT)	KC517492(VT)	?(DQ151548(VT))
				JQ911664(VT)	?(KC517490(T30))					EU937519(VT)	?(JQ911664(VT))
					?(KC517491(T30))					KC262793(VT)	
										KC517493(VT)	
										KC517494(VT)	

2.4 Conclusion

There are differences in the relationship of the isolates based on the different fragments used and the technique applied. A similar study was performed on HIV where different segments of the genome was used for phylogenetic analysis and compared to the 'true' phylogeny (the actual transmission direction) of the isolates. It was found that the gene fragment used was more important than the phylogenetic method applied and that a combination of gene fragments produced the best results (Leitner *et al.*, 1996). In this study it was found that using ORF 1b (RNA dependent RNA polymerase), ORF 2 (p33 which plays a role in super-infection exclusion (Folimonova, 2012)) and ORF 5 (p61 which plays a role in virion assembly (Satyanarayana *et al.*, 2000)) can provide a dendrogram very similar to the complete genome dendrogram. A study by Harper (2013) suggested the use of domains within ORF 1a and ORF1b as these are the most divergent areas within the genome (Hilf *et al.*, 1999; Mawassi *et al.*, 1996). Using just the most divergent part of the genome (5' half) will lead to many different clades that does not necessarily relate to phenotype (biological characteristics) and using just the more conserved part of the genome (3' half) may lead to grouping of clades as one where there are differences in phenotypes. Using a combination of genes from both the divergent and conserved areas of the genome serves two functions: 1) getting a balanced view of the genome based on conserved and divergent areas and 2) to attempt to equalize the effects of recombination. If one constructs dendrograms for each separate segment and one combining the three, one could identify possible recombination within an isolate if it clusters drastically different in the four dendrograms. This will unfortunately only be possible if the recombination occurred in one of the three areas used. As the sequence data becomes available for more new isolates one may observe less diversity between different clades (genotypes) and genetic information may become a flow from the one genome to the other in a genetic continuum as it has been discussed for prokaryotes

(Doolittle and Zhaxybayeva, 2010). This can be seen for the isolates that can apparently cluster with two different genotypes, as is the case for isolate AB046395.

Throughout the study it could be seen that there is a close relationship between T3/T30/VT and RB/T36 and T68/B165. It would be interesting to know how this genotypic relationship relates back to the biological characteristics of the isolates, especially regarding cross-protection. In the case with the ORF 2 (p33 gene) dendrogram, the RB clade is closer related to the T30 clade than the T36 clade, as can be seen in multiple other dendrograms, including the complete genome dendrogram. With further research one might be able to determine whether this genotypic relationship also relates to the biological properties of the strains, in other words if RB isolates will be able to cross-protect against just other RB strains or also T30 or T36 strains. In the same scenario, it would also need to be studied if T3 and HA16-5 strains cross-protect against each other since these isolates cluster together. Another interesting scenario is the T68 and B165 isolates. These two isolates clusters together in many of the dendrograms, but in the complete genome dendrogram they form their own clades. In the ORF 2 dendrogram however, T68 groups with VT, so it will also need to be established if T68 cross-protects against VT or B168 strains or just T68 strains. Since the T68 and B165 isolates clusters together more than apart in different dendrograms, but it is separate in the complete genome dendrogram and pairwise distance analysis, the question arises if is it indeed two genotypes or actually just one. Harper (2013) described the B165 isolate (EU076703) as part of the T68 clade, but this was before the sequences data for the CT-ZA3 isolate (KC333686) became available that caused the T68 isolate (JQ965169) to cluster away from the B165 isolates. This shows that the classification of CTV isolates will never be static and as new data becomes available the picture will transform. This highlights the necessity to link molecular, biological and genetic data. With just one of the components

there is still many unanswered questions, but when combining all three, one will be able to extract very important information that can result in the better control of the virus.

The variability in dendrogram morphology based on different fragments used showed that the use of single fragments of the genome is not sufficient to distinguish between genotypes of the virus. The most likely explanation for this would be recombination. When recombination occurs within sequences, different fragments of the genomes have different evolutionary histories, leading to the dendrograms constructed from the different fragments to show these histories. Ignoring recombination in data sets (as we do when constructing a complete genome dendrogram) can cause various artifacts in the analysis, including underestimation of the time to the most recent ancestor and the amount of recent divergence. It can also cause the overestimation of the number of mutations (Schierup and Hein, 2000). For the purpose of this specific study these factors were not important as we just needed to distinguish between different clades and were not interested in the details of how these clades are related to each other. The high amount of recombination within viral RNA makes it very difficult to use the full capacity of phylogenetic analysis and for our purpose we simplify it tremendously.

This study yet again highlighted the very complicated nature of CTV. Almost every single gene used in analysis produced different results and many just lacked the nucleotide diversity to be able to produce enough branch support to allow for genotyping. It was also found that with every different technique used, different results could be obtained. This highlights the importance of using a standardized approach that would allow scientists to compare their data directly with previously published results. It is also important to remember that all information produced from this study is purely based on genetic information and many gaps still needs to be filled with biological data. It is not necessarily important to be able to fit all new isolates into perfectly set groups, but for control purposes it does help to have an idea of how new isolates are related to known isolates. Once the biological data becomes available

that identifies the pathogenicity factors of the virus and which genes control cross-protection, phylogenetic analysis can be based only on these factors. This would allow one to classify strains into mild and severe groups and to know which cross-protection strategy to use. But until this data becomes available, one can use gene fragments that produce similar phylogenetic results to the complete genome, and in this study it was found to be ORF 1b, ORF 2 and ORF 5. It is recommended that in instances where the entire genome cannot be sequenced, these three genes be sequenced and used to construct a phylogenetic dendrogram. It is important that the correct nucleotide model for each gene is tested and that these models be used during the analysis of the concatenated dataset in a program that allows for these specifications, for example Mr. Bayes (Ronquist *et al.*, 2012).

2.5 References

Abascal, F., Zardoya, R., and Posada, D. 2005. ProtTest: Selection of Best-Fit Models of Protein Evolution. *Bioinformatics* 21(9), pp. 2104-2105.

Albiach-Marti, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., Lopez, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., and Dawson, W.O. 2000. Sequences of *Citrus tristeza virus* Separated in Time and Space Are Essentially Identical. *The Journal of Virology* 74(15), pp. 6856-6865.

Černi, S., Ruščić, J., Nolasco, G., Gatin, Ž., Krajačić, M., and Škorić, D. 2008. Stem Pitting and Seedling Yellows Symptoms of *Citrus tristeza virus* Infection May Be Determined by Minor Sequence Variants. *Virus Genes* 36(1), pp. 241-249.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. 2012. jModelTest 2: More Models, New Heuristics and Parallel Computing. *Nature Methods* 9(8), pp. 772-772.

Davino, S., Willemsen, A., Panno, S., Davino, M., Catara, A., Elena, S.F., and Rubio, L. 2013. Emergence and Phylodynamics of *Citrus tristeza virus* in Sicily, Italy. *PLoS ONE* 8(6).

Doolittle, W.F., and Zhaxybayeva, O. 2010. Metagenomics and the Units of Biological Organization. *Bioscience* 60, pp. 102-112.

Fagoaga, C., Lopez, C., Moremo, P., Navarro, L., Flores, R., and Pena, L. 2005. Viral-Like Symptoms Induced by the Ectopic Expression of the p23 Gene of *Citrus tristeza virus* Are Citrus Specific and Do Not Correlate with the Pathogenicity of the Virus Strain *Molecular Plant-Microbe Interactions* 18(5), pp. 435-445.

Fagoaga, C., Pensabene-Bellavia, G., Moreno, P., Navarro, L., Flores, R., and Peña, L. 2011. Ectopic Expression of the p23 Silencing Suppressor of *Citrus tristeza virus* Differentially Modifies Viral Accumulation and Tropism in Two Transgenic Woody Hosts. *Molecular Plant Pathology* 12(9), pp. 898-910.

Febres, V., Ashoulin, L., Mawassi, M., Frank, A., Bar-Joseph, M., Manjunath, K., Lee, R., and Niblett, C. 1996. The p27 Protein Is Present at One End of Citrus tristeza virus Particles.

Folimonova, S.Y. 2012. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology*. 10.1128/jvi.00310-12

Gago-Zachert, S., Costa, N., Semorile, L., and Grau, O. 1999. Sequence Variability in p27 Gene of *Citrus tristeza virus* (CTV) Revealed by SSCP Analysis. *Electronic Journal of Biotechnology* 2, pp. 3-4.

Ghorbel, R., López, C., Fagoaga, C., Moremo, P., Navarro, L., Flores, R., and Pena, L. 2001. Transgenic Citrus Plants Expressing the *Citrus tristeza virus* p23 Protein Exhibit Viral-Like Symptoms. *Molecular Plant Pathology* 2(1), pp. 27-36.

Gillings, M., Broadbent, P., Indsto, J., and Lee, R. 1993. Characterisation of Isolates and Strains of Citrus Tristeza Closterovirus Using Restriction Analysis of the Coat Protein

Gene Amplified by the Polymerase Chain Reaction. *Journal of Virological Methods* 44(2–3), pp. 305-317.

Guindon, S., and Gascuel, O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* 52(5), pp. 696-704.

Hall, T.A. 1999. Bioedit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/Nt. *Nucleic Acids Symposium Series* 41, pp. 95-98.

Harper, S., Dawson, T., and Pearson, M. 2010. Isolates of *Citrus tristeza virus* That Overcome *Poncirus trifoliata* Resistance Comprise a Novel Strain. *Archives of Virology* 155(4), pp. 471-480.

Heath, L., van der Walt, E., Varsani, A., and Martin, D.P. 2006. Recombination Patterns in Aphthoviruses Mirror Those Found in Other Picornaviruses. *Journal of Virology* 80(23), pp. 11827-11832.

Hilf, M.E., Karasev, A.V., Albiach-Marti, M.R., Dawson, W.O., and Garnsey, S.M. 1999. Two Paths of Sequence Divergence in the *Citrus tristeza virus* Complex. *Phytopathology* 89(4), pp. 336-342.

Karasev, A.V. 2000. Genetic Diversity and Evolution of Closteroviruses. *Annual Review of Phytopathology* 38(1), pp. 293-324.

Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. 2002. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Research* 30(14), pp. 3059-3066.

Kong, P., Rubio, L., Polek, M., and Falk, B.W. 2000. Population Structure and Genetic Diversity within California *Citrus tristeza virus* (CTV) Isolates. *Virus Genes* 21(3), pp. 139-145.

Leitner, T., Escanilla, D., Franzén, C., Uhlén, M., and Albert, J. 1996. Accurate Reconstruction of a Known HIV-1 Transmission History by Phylogenetic Tree Analysis. *Proceedings of the National Academy of Sciences* 93(20), pp. 10864-10869.

López, C., Navas-Castillo, J., Godwa, A., Moremo, P., and Flores, R. 2000. The 23-kDA Protein Coded by the 3'-Terminal Gene of *Citrus tristeza virus* Is an RNA-Binding Protein. *Virology* 269, pp. 462-470.

Lu, R., Folimonov, A., Shintaku, M., Li, W.-X., Falk, B.W., Dawson, W.O., and Ding, S.-W. 2004. Three Distinct Suppressors of RNA Silencing Encoded by a 20-Kb Viral RNA Genome. *Proceedings of the National Academy of Sciences of the United States of America* 101(44), pp. 15742-15747.

Luttig, M., Van Vuuren, S.P., and Van Der Vyver, J.B. 2002. Differentiation of Single Aphid Cultured Sub-Isolates of Two South African *Citrus tristeza virus* Isolates from Grapefruit by Single-Stranded Conformation Polymorphism. In: *Proceedings of the 15th Conference of the International Organization of Citrus Virologists. (Duran-Vila, N., Milne, R.G., and Da Graça, J.V.)* pp. 186-196. *IOCV*. Riverside, California.

Martínez, F., Lafforgue, G., Morelli, M.J., González-Candelas, F., Chua, N.-H., Daròs, J.-A., and Elena, S.F. 2012. Ultradeep Sequencing Analysis of Population Dynamics of Virus Escape Mutants in RNAi-Mediated Resistant Plants. *Molecular Biology and Evolution*.

Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G., and Bar-Joseph, M. 1996. Unusual Sequence Relationships between Two Isolates of *Citrus tristeza virus*. *Journal of General Virology* 77(9), pp. 2359-2364.

Morroni, M., Jacquemond, M., and Tepfer, M. 2013. Deep Sequencing of Recombinant Virus Populations in Transgenic and Nontransgenic Plants Infected with *Cucumber Mosaic Virus*. *Molecular Plant-Microbe Interactions* 26(7), pp. 801-811.

Rambaut, A. 2009. Figtree: Tree Figure Drawng Tool. University of Edinburgh: Institute of Evolutionary Biology.

Rambaut, A., and Drummond, A. 2009. Tracer: MCMC Trace Analysis Tool. University of Edinburgh: Institute of Evolutionary Biology and University of Auckland: Department of Computer Science.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology* 61(3), pp. 539-542.

Roy, A., and Brlansky, R.H. 2010. Genome Analysis of an Orange Stem Pitting *Citrus tristeza virus* Isolate Reveals a Novel Recombinant Genotype. *Virus Research* 151(2), pp. 118-130.

Roy, A., Manjunath, K.L., and Brlansky, R.H. 2005. Assessment of Sequence Diversity in the 5′ -Terminal Region of *Citrus tristeza virus* from India. *Virus Research* 113(2), pp. 132-142.

Rubio, L., Ayllon, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., and Falk, B.W. 2001. Genetic Variation of *Citrus tristeza virus* Isolates from California and Spain: Evidence for Mixed Infections and Recombination. *The Journal of Virology* 75(17), pp. 8054-8062.

Sambade, A., López, C., Rubio, L., Flores, R., Guerri, J., and Moreno, P. 2003. Polymorphism of a Specific Region in Gene p23 of *Citrus tristeza virus* Allows Discrimination between Mild and Severe Isolates. *Archives of Virology* 148(12), pp. 2325-2340.

Satyanarayana, T., Gowda, S., Mawassi, M., Albiach Martı, M.R., Ayllón, M.A., Robertson, C., Garnsey, S.M., and Dawson, W.O. 2000. Closterovirus Encoded Hsp70 Homolog and P61 in Addition to Both Coat Proteins Function in Efficient Virion Assembly. *Virology* 278(1), pp. 253-265.

Schierup, M.H., and Hein, J. 2000. Consequences of Recombination on Traditional Phylogenetic Analysis. *Genetics* 156(2), pp. 879-891.

Scott, K.A., Hlela, Q., Zablocki, O., Read, D., van Vuuren, S., and Pietersen, G. 2013. Genotype Composition of Populations of Grapefruit-Cross-Protecting *Citrus tristeza virus* Strain GFMS 12 in Different Host Plants and Aphid-Transmitted Sub-Isolates. *Archives of Virology* 158(1), pp. 27-37.

Silva, G., Marques, N., and Nolasco, G. 2012. The Evolutionary Rate of *Citrus tristeza virus* Ranks among the Rates of the Slowest RNA Viruses. *Journal of General Virology* 93(2), pp. 419-429.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution* 24, pp. 1596-1599.

Vives, M.C., Rubio, L., Sambade, A., Mirkov, T.E., Moreno, P., and Guerri, J. 2005. Evidence of Multiple Recombination Events between Two RNA Sequence Variants within a *Citrus tristeza virus* Isolate. *Virology* 331(2), pp. 232-237.

Zablocki, O., and Pietersen, G. 2014. Characterization of a Novel *Citrus tristeza virus* Genotype within Three Cross-Protecting Source GFMS12 Sub-Isolates in South Africa by Means of Illumina Sequencing. *Archives of Virology*, pp. 1-7.

Chapter 3: Sequencing of four CTV sub-isolates through fragment amplification and Illumina sequencing

3.1 Introduction

The most important control mechanism for CTV in South Africa is mild strain cross-protection. The South African Citrus Improvement Program was established in 1973 to allow for cross-protection within the country (Von Broembsen and Lee, 1988). For CTV, cross protection is hypothesized to rely on the super-infection exclusion principle (Folimonova *et al.*, 2010) where the presence of one virus within the plant prevents a secondary infection of that same virus (Fulton, 1986). However, for CTV this only occurs between strains from the same genotype (Folimonova *et al.*, 2010). For example, if a plant is infected with a mild strain of the VT genotype, it is only protected against the secondary infection of strains from the VT genotype. It is not however protected from strains of the B165, T30, T68 and other genotypes. The exact mechanism of super-infection exclusion it not known yet, but thus far it has been established by Folimonova *et al*. (2012) that the p33 gene plays a role.

Since 1) more than one genotype of CTV can occur within a plant (Albiach-Marti *et al.*, 2000); 2) a given strain of CTV can cause mild symptoms in one host, but severe symptoms in another (Fulton, 1986); and 3) a mild strain of CTV can produce mild symptoms in one citrus producing area but severe symptoms in another (Karasev *et al.*, 1998); there can be variation in the success of a mild strain cross-protection source. Environmental conditions and host species can influence the selection of genotypes within a plant, and this may cause a source of CTV to display more severe symptoms under certain circumstances. For example, a CTV source containing a mild strain of CTV dominant under certain conditions can, when inoculated into a different host in other environmental conditions, become less dominant (Albiach-Marti *et al.*, 1996). The host may select for a more severe strain of CTV within the population, causing more severe symptoms (Van Vuuren, 2002) and leading to the breakdown of cross-protection.

This occurred with one of the first cross-protection sources used in South Africa, Grapefruit Mild Strain 12 (GFMS 12). It was identified in an orchard in Wellington where a tree had not shown any severe CTV symptoms despite 50 years of exposure to the virus. GFMS 12 was subjected to glasshouse trails on indicator plants to assess its suitability as cross-protection source (Von Broembsen and Lee, 1988), and was then used in the field. While the cross-protection source provided good protection of Marsh grapefruit for several years, it did not show as much promise for Star Ruby Grapefruit (Van Der Vyver *et al.*, 2002). In later years, trees pre-immunized with GFMS 12 started showing more severe CTV symptoms (Van Vuuren *et al.*, 1991) and caused different symptoms in different hosts (Meyer *et al.*, 2005). Field trails done with this cross-protection source showed severe stem pitting in both Marsh and Star Ruby grapefruit, indicating it was no longer suitable for cross-protection (Van Vuuren and Van Der Vyver, 2000). Van Vuuren *et al.*(2000) produced sub-isolates of GFMS 12 and found that different sub-isolates were either milder or more severe than the original isolate. This confirmed that GFMS 12 contained more than one strain, including a severe strain. Amongst the sub-isolates it was found that 12-2 and 12-5 were less virulent and 12-3 was more virulent than the original source. The sub-isolates may still contain mixtures of the virus and this makes further characterization necessary. Scott *et al.* (2013) did further analysis on GFMS 12 and sub-isolates 12-7, 12-8 and 12-9. Based on biological indexing on seven different citrus hosts, GFMS 12 showed moderate vein clearing and stem pitting on Mexican lime and mild stunting and moderate stem pitting on Duncan grapefruit. Furthermore the three sub-isolates showed different symptoms with 12-7 producing decline in sweet orange grafted on sour orange rootstock, but milder symptoms than GFMS 12 on Mexican lime and Duncan grapefruit. Sub-isolate 12-8 did not show any stem pitting on grapefruit and 12-9 showed mild stem pitting and stunting on all indicators. While doing sequence analysis of clones for the A region and p23 gene, it was found that GFMS 12

contains at least RB-, T30-, VT- and B165-like genotypes. The sub-isolates produced different results based on the a region (ORF 1a) and the p23 gene (3' end of genome) and it was proposed that recombinants are present. Sub-isolates 12-7, 12-8 and 12-9 all contained B165/VT-like recombinant strains.

The aim of this study was to characterize the 12-7, 12-8 and 12-9 GFMS 12 sub-isolates as well as an additional CTV source, B390/3, through whole genome sequencing using the Illumina platform (Illumina Inc, San Diego, California, USA). Within our laboratory, Mr. O. Zablocki also characterized these three sources with the Illumina sequencing technology, but using different template preparation techniques (Zablocki, 2013), including total RNA extraction, dsRNA extraction and immuno-capture followed by a random RT-PCR adapted from Roossinck *et al*. (2010). The technique used in this study relies on total RNA extraction followed by amplification of the complete genome in overlapping fragments and using this as template for Illumina sequencing.

3.2 Materials and methods

3.2.1 Viral sources

Viral sources used included the single aphid transmitted sub-isolates of Grapefruit Mild Strain 12, which are GFMS 12-7 (08-0010), GFMS 12-8 (08-2011) and GFMS 12-9 (08-0024), as well as the B390/3 source. The GFMS 12 sub-isolates used had been maintained in an insect free glasshouse (CRI@UP) in Mexican lime. The B390/3 source was used from the Mexican lime plant maintained in Nelspruit at Citrus Research International. This viral source was included in this chapter due to its potential as a cross-protection source undergoing biological indexing. Further background information and characterization of this source can be found in Chapter 5.

3.2.2 Develop one step RT-PCR for CTV

Total RNA extract was obtained with the GeneJet Plant RNA Purification Mini Kit (Thermo Scientific, Waltham, Masseuses, US) according to manufacturer's instructions. Of this extract, 10 µl was used as template in a RT-PCR reaction containing 5 U RNase Inhibitor, 4.2 U AMV reverse transcriptase (Roche diagnostics, Mannheim, Germany), 1x R Buffer, 2.5 mM $MgCl_2$, 2.5 U Taq Polymerase (Bioline, London, UK), 0.14 mM of a dNTP mix (Promega, Madison, Wisconsin, USA), 5 µl of a 1% bovine serum albumin (BSA) solution (Sigma, St. Louis, Missouri, USA), 1 µM of each forward and reverse primer and PCR grade water for a total volume of 50 µl. Primer pairs used are shown in Table B.1, Appendix B. Reaction conditions were as follows: 42°C for 60 min, 92°C for 2 min, 35 cycles of 92°C for 30 sec, 55°C for 45 sec, 72°C for 1 min and a final extension for 10 min at 72°C.

3.2.3 Primer design

Primers were designed using the CLC Main Workbench 6 (CLC Bio, Aarhus, Denmark) and were based on the 30 complete CTV genomes available on GENBANK (Benson *et al.*, 2005) at that stage. Accession numbers for these genomes are: AB046398.1, AF001623.1, AF260651.1, AY170468.1, AY340974.1, DQ151548.1, DQ272579.1, EU076703.3, EU857538.1, EU937519.1, EU937520.1, EU937521.1, FJ525431.1, FJ525432.1, FJ525433.1, FJ525434.1, FJ525435.1, FJ525436.1, GQ454869.1, GQ454870.1, HM573451.1, JF957196.1, JQ061137.1, JQ798289.1, JQ911663.1, JQ911664.1, JQ965169.1, NC001661.1, U16304.1 and Y18420.1. These sequences were aligned and a consensus sequence using the IUPAC (International Union of Pure and Applied Chemistry) nucleotide code was created. This sequence was used to identify regions in the genome that were most conserved and were suitable for the design of the degenerate primers that would amplify all genotypes. These primers are described in Table B.2, Appendix B.

3.2.4 RT-PCR for whole genome amplification

Total RNA was obtained from Mexican lime trees using the Isolate Plant RNA Mini Kit (Bioline, London, UK) according to manufacturer's instructions.

3.2.4.1 Region 1, 2, 3, 5 and 10

Region 1, 3 and 10 was amplified with primer pairs CTV 1F & 1R, 3F & 3R and 10F and 10.1R, respectfully. Amplification was done using the one step RT-PCR discussed in section 3.2.2, with cycle conditions adjusted as follows: 42°C for 60 min, 92°C for 2 min followed by 10 cycles of 92°C for 30 sec, 35°C for 45 sec and 68°C for 2 min. This was followed by 35 cycles of 92°C for 30 sec, 40°C for 45 sec and 68°C for 2 min. Final extension was at 68°C for 10 min.

Region 2 and 5 could be amplified with primer pairs CTV 2F & 2R and 5F & 5R respectfully, using the same protocol, but with the addition of 2.5 µl 1x dimethyl sulfoxide (DMSO).

3.2.4.2 Region 4 and 9

Region 4 and 9 was amplified with primer pair CTV 4F & 4R and 9F & 9R, respectfully, using a two-step RT-PCR. For the reverse transcription of RNA, 12 µl template was combined with 2 µM reverse primer and incubated for 15 min at 65°C, 10 min at 55°C and 5 min at room temperature. After primer annealing, 5 µl RT buffer, 10 U AMV reverse transcriptase, 5 U RNase inhibitor (Roche diagnostics, Mannheim, Germany) and 0.1 mM of a dNTP mix (Promega, Madison, Wisconsin, USA) was added and then incubated for 1 hour at 42°C. For the PCR, 10 µl of the cDNA was combined with 1x R Buffer, 2.5 mM $MgCl_2$, 2.5 U Taq Polymerase (Bioline, London, UK), 0.14 mM of a dNTP mix (Promega, Madison, Wisconsin, USA), 5 µl of a 1% BSA solution (Sigma, St. Louis, Missouri, USA), 1 µM of

each forward and reverse primer and PCR grade water for a total volume of 50 µl. PCR was performed using the same cycle conditions as in 3.2.4.1.

3.2.5 Illumina sequencing and data analysis

Regions 1 to 5 were pooled for each source in equivalent amounts as estimated by the intensity of expected product bands when run on a 1% agarose gel (Figure 3.1). The goal was to obtain similar amounts of amplicon for each fragment by comparing the amplicons' band intensity. For GFMS 12-7, 2 µl was used for region 2, 3 and 4, 3 µl was used for region 1, and 5 µl was used for region 4. For GFMS 12-8, 3 µl was used for region 2 and 5, 4 µl was used for region 3, and 5 µl was used for region 1 and 3. For GFMS 12-9, 2 µl was used for region 2, 3 and 5, and 3 µl used for region 1 and 4. For B390/3, 1 µl was used for region 2 and 3, 2 µl for region 1 and 5, and 3 µl for region 4. After pooling amplicons for all the sources, the final amount was made up to 40µl with PCR grade water for each source.

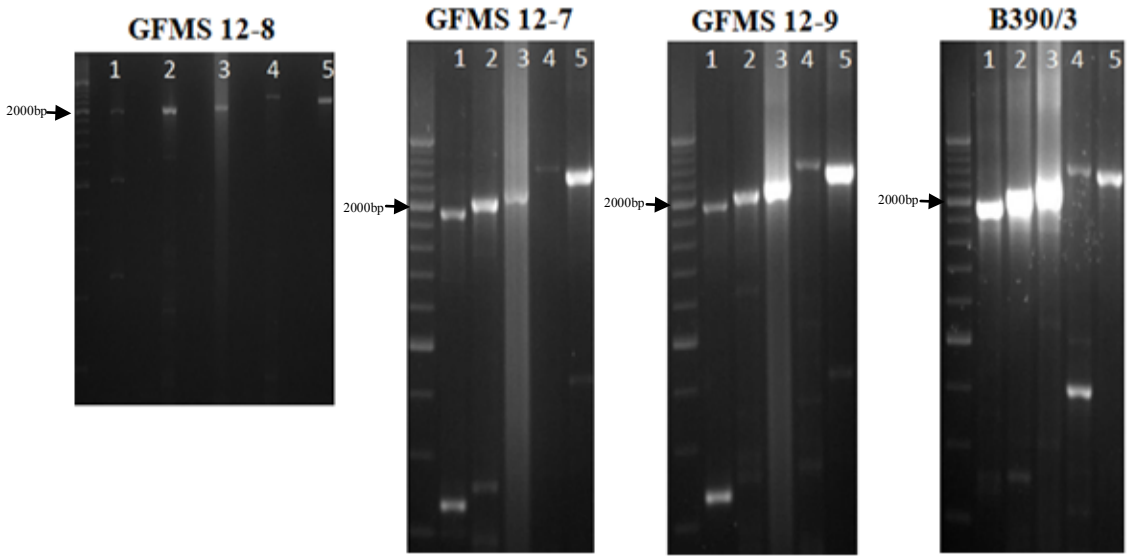

Figure 3.1: Agarose gel photographs of amplicons of region 1-5 for GFMS 12 sub-isolates and B390/3. Thermo Scientific O'RangeRuler 200 bp molecular marker included for size estimation.

The pooled samples were then sent for paired end Illumina sequencing on the MiSeq platform (Illumina Inc., San Diego, California, USA) at the Agricultural Research Council Bioinformatics platform at Onderstepoort, Pretoria.

Data sets were analysed on CLC Genomics Workbench 6 (CLC Bio, Aarhus, Denmark). Sequence reads were imported as paired end data and trimmed based on quality using the following parameters: “Ambiguous trim” yes, “ambiguous limit” 2, “quality trim” yes, “quality limit” 0.05, “use colorspace” no, “create report” yes, “also search on reversed sequences” no, “save discarded sequences” no, “remove 5' terminal nucleotides” no, “discard short reads” no, “remove 3' terminal nucleotides” no, “trim adaptor list” trim adapter library, “discard long reads” no, “save broken pairs” no. Sequences of the Nextera adapters, used during sequencing, were removed. Trimmed reads were mapped to a reference data set containing all 45 complete CTV genomes available on GENBANK at that stage. Accession numbers for these genomes are: AB046398.1, AF001623.1, AF260651.1, AY170468.1, AY340974.1, DQ151548.1, DQ272579.1, EU076703.3, EU857538.1, EU937519.1, EU937520.1, EU937521.1, FJ525431.1, FJ525432.1, FJ525433.1, FJ525434.1, FJ525435.1, FJ525436.1, GQ454869.1, GQ454870.1, HM573451.1, JF957196.1, JQ061137.1, JQ798289.1, JQ911663.1, JQ911664.1, JQ965169.1, JX266712.1, JX266713.1, KC262793.1, KC333868.1, KC517485.1, KC517486.1, KC517487.1, KC517488.1, KC517489.1, KC517490.1, KC517491.1, KC517492.1, KC517493.1, KC517494.1, KC525952.1, NC001661.1, U16304.1 andY18420.1. Mapping was done with settings as follows: “masking mode” no masking, “mismatch cost” 2, “insertion cost” 3, “deletion cost” 3, “length fraction” 1.0, “similarity fraction” 0.98, “global alignment” no, “auto-detect paired distances” yes, “non-specific match handling” ignore, “output mode” create reads track, “create report” yes, “collect un-mapped reads” no. Furthermore, *de novo* assembly was performed on all datasets using the following parameters: “mismatch cost” 2, “insertion cost” 3, “deletion cost” 3,

"length fraction" 1, "similarity fraction" 0.8, "add conflict annotations" no, "conflict resolution" vote, "non-specific matches" random, "minimum contig length" 200, "create full contigs" yes. All contigs larger than 500 bp were subjected to BLAST on the NCBI website (Altschul *et al.*, 1990) using CLC with the following parameters: "program" blastn, "match cost" 1, "mismap cost" 3, "existence cost" 5, "extension cost" 2, "expectation value" 10.0, "word size" 11, "mask lower case" no, "filter low complexity" yes, "maximum number of hits" 100, "limit by entrez query" all organisms, "database" nr. CTV related contigs mapped back to CTV genomes using parameters listed above, including a length fraction of 1, similarity fraction of 0.98 and non-specific match handling was set to 'ignore'.

3.3 Results and discussion

3.3.1 Develop one step RT-PCR for CTV

The one step RT-PCR was developed and tested using three primer sets to confirm the success of the protocol. For all three primer sets the correct product could be amplified with minimal non-specific amplification (Figure 3.2). The non-specific amplification products in the Modified A and PM reactions are at a low concentration, as judged by the faintness of bands on the agarose gel.

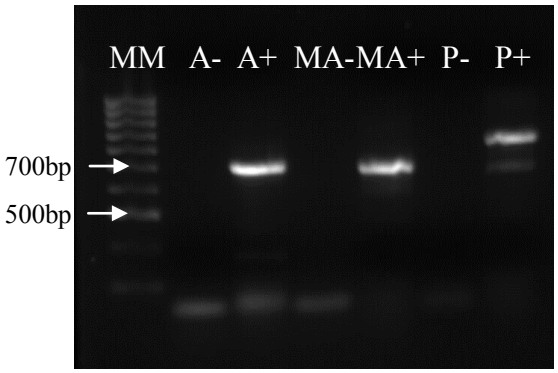

Figure 3.2: Agarose gel photograph of amplicons obtained with one step RT-PCR. Letters indicates primer sets used. – indicates negative control, + indicates positive control and MM indicates Bioline Hyper ladder IV molecular marker.

3.3.2 Primer design

Due to difficulties experienced during the conducting of the RT-PCR for region 6, 7, 8, 9 and 11, the primer sets were tested with PCR on the CTV9ΔGFP plasmid containing the complete T36 CTV genome (Folimonova, S.Y.). For this reaction, 1x R Buffer, 2.5 mM $MgCl_2$, 2.5 U Taq Polymerase (Bioline, London, UK), 0.14 mM of a dNTP mix (Promega, Madison, Wisconsin, USA), 1 µM of each forward and reverse primer and PCR grade water for a total volume of 50 µl was combined. Primer sets used were CTV 6.1F & 6.1R, 7.1F & 7R, 8.1F & 8.1R, 9.1F & 9.1R, 11.1F and 11R. Cycle conditions were 42°C for 60 min, 92°C for 2 min, 35 cycles of 92°C for 30 sec, 55°C for 45 sec, 72°C for 1 min and a final extension for 10 min at 72°C.

The correct sized products could be obtained for a11 the regions, although there was greater than expected non-specific amplification for region 11 (Figure 3.3). This proved that the primers were functioning correctly during amplification and that the problem was with the reverse transcription step.

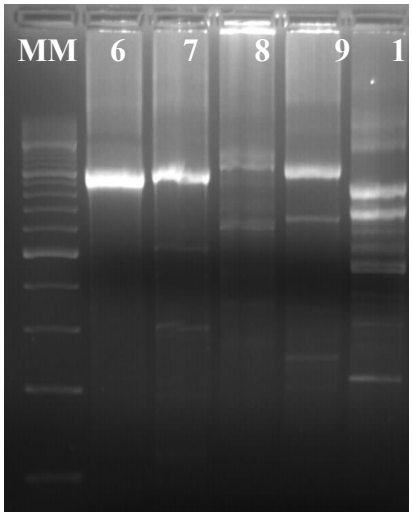

Figure 3.3: Agarose gel of products obtained from PCR of fragments 6, 7, 8, 9 and 11 on the CTV9ΔGFP plasmid containing the complete T36 CTV genome (MM – Thermo Scientific O'RangeRuler 200 bp molecular marker).

To obtain the best annealing temperature for the primers, the PCR was repeated but with a gradient of temperatures between 30°C and 55°C for the annealing step. From the agarose gel it could be seen that the ratio of correct product to non-specific amplification was the highest when using an annealing temperature of 55°C (Figure 3.4 - 3.8). This was fortuitous since a higher temperature will also limit the amount of secondary structure within the RNA.

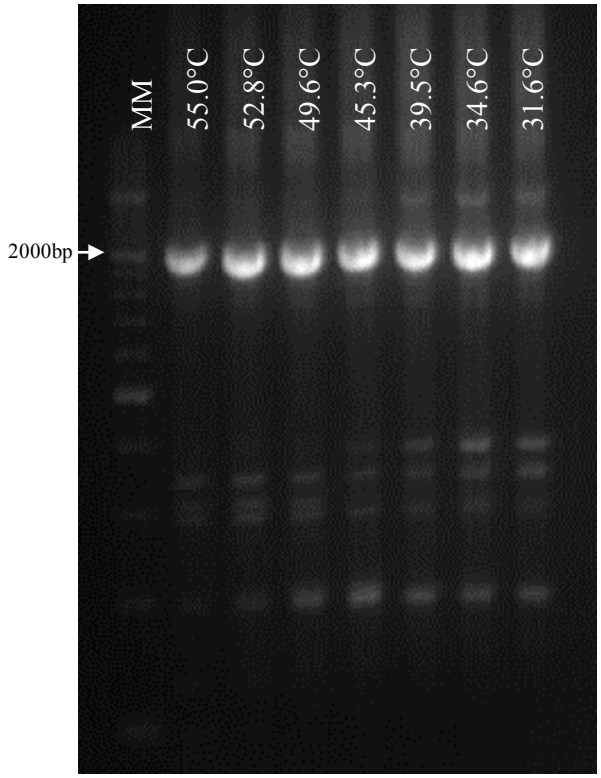

Figure 3.4: Agarose gel of products of PCR at various annealing temperatures for primer set 6 (MM – Thermo Scientific O'RangeRuler 200 bp molecular marker).

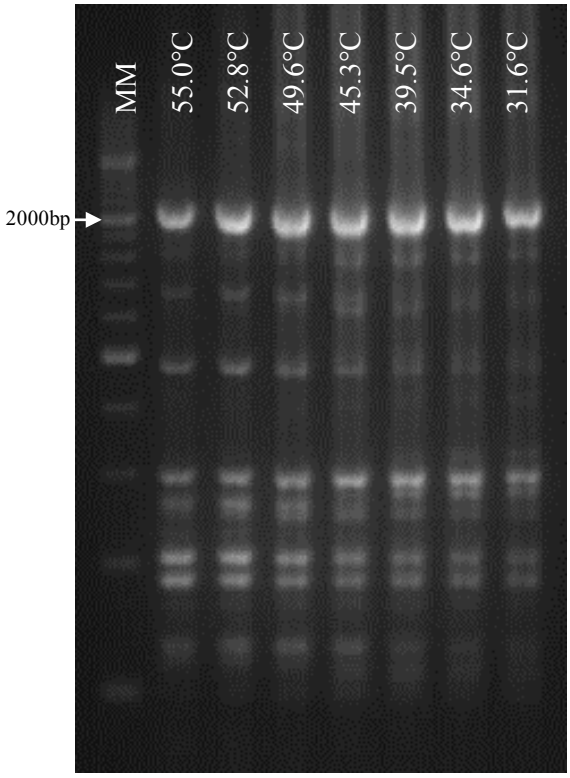

Figure 3.5: Agarose gel of products of PCR at various annealing temperatures for primer set 7 (MM – Thermo Scientific O'RangeRuler 200 bp molecular marker).

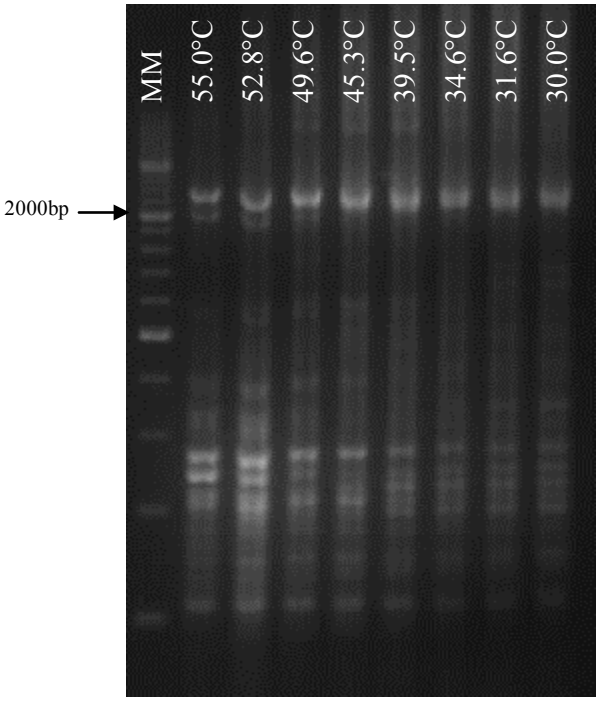

Figure 3.6: Agarose gel of products of PCR at various annealing temperatures for primer set 8 (MM – Thermo Scientific O'RangeRuler 200 bp molecular marker).

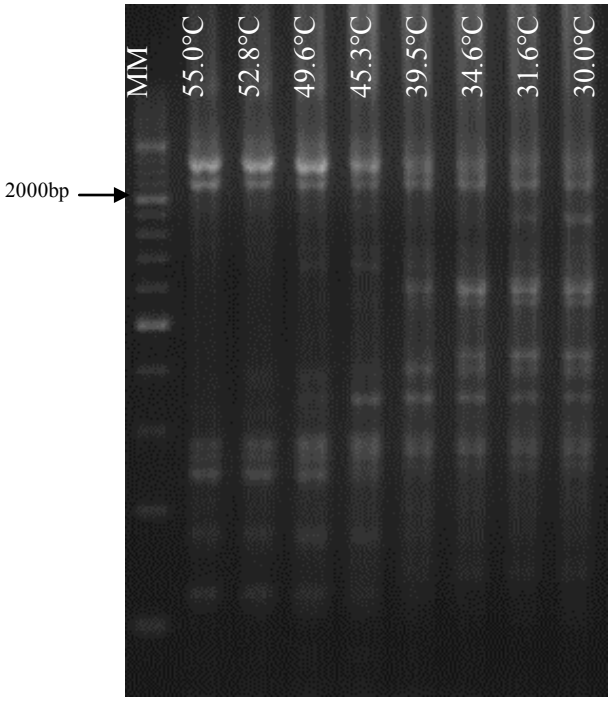

Figure 3.7: Agarose gel of products of PCR at various annealing temperatures for primer set 9 (MM – Thermo Scientific O'RangeRuler 200 bp molecular marker).

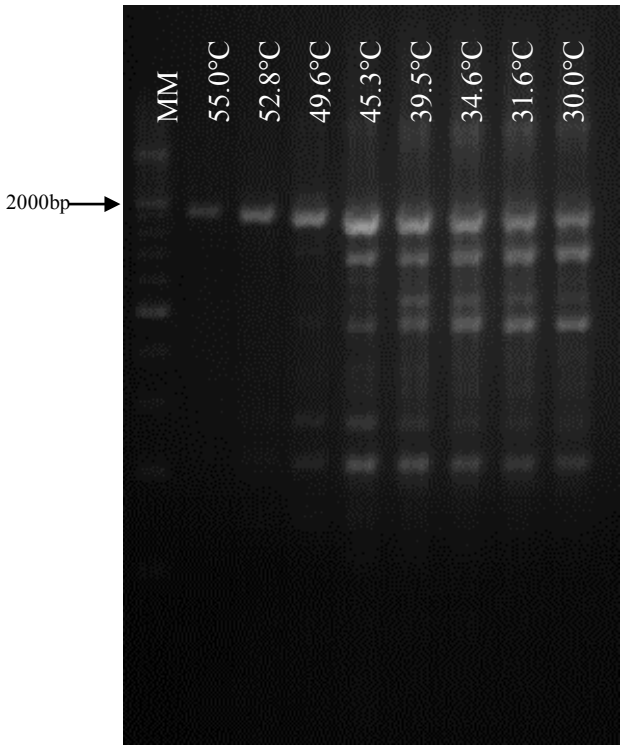

Figure 3.8: Agarose gel of products of PCR at various annealing temperatures for primer set 11 (MM – Thermo Scientific O'RangeRuler 200 bp molecular marker).

3.3.3 Amplification, troubleshooting and optimization

3.3.3.1 Region 1

Region 1 could be successfully amplified for all samples (Figure 3.9). Some additional bands could be seen for the GFMS 12 sub-isolates, but since a single product is not necessary for Illumina sequencing as with Sanger sequencing, the amplicons could be used as is.

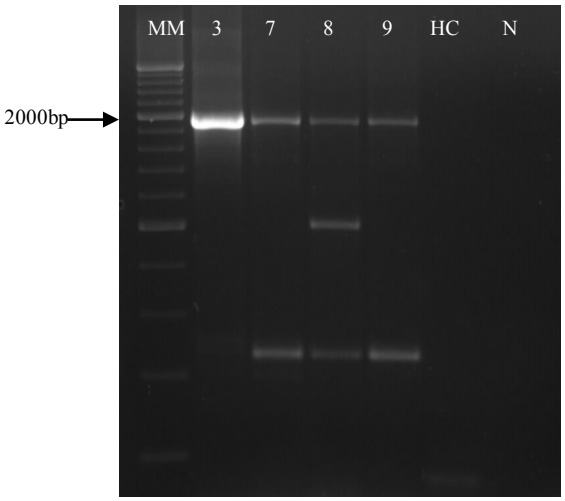

Figure 3.9: Agarose gel photograph of region 1 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.2 Region 2

The basic one step RT-PCR described in section 3.3.1 allowed amplification of the 2099 bp product for all samples but with very low yield for GFMS 12-7 and GFMS 12-8. The addition of DMSO increased the yield of the desired product, but also increased non-specific amplification (Figure 3.10). Since the non-specific amplification was much less than the desired product, the amplicons were still used for Illumina sequencing.

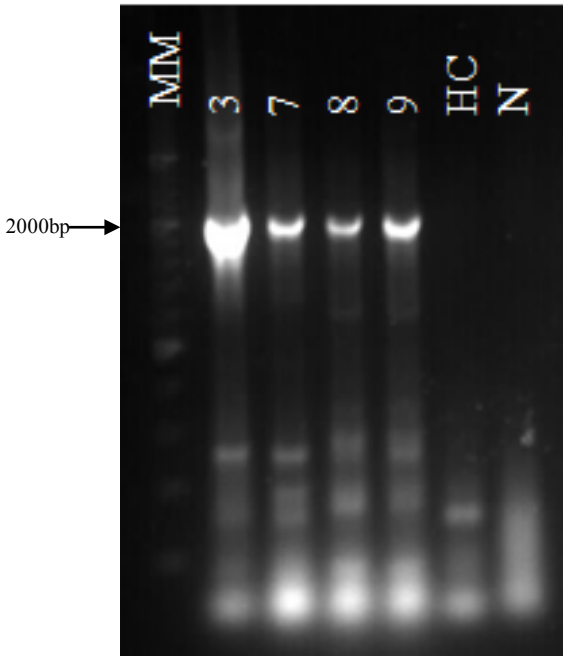

Figure 3.10: Agarose gel photograph of region 2 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.3 Region 3

The 2181 bp region 3 product was successfully amplified for all the samples with no non-specific amplification or amplification in the healthy plant control (Figure 3.11).

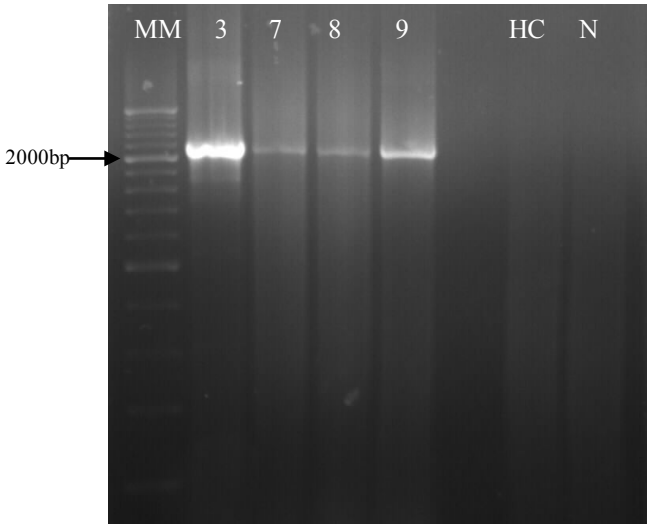

Figure 3.11: Agarose gel photograph of region 3 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp (MM).

3.3.3.4 Region 4

The one step RT-PCR amplified the expected 2483 bp product for all the samples, but the yield for GFMS 12-7 and GFMS 12-8 was very low. When a two-step RT-PCR approach was used, the correct product could be amplified for all the samples with a higher yield, except for GFMS 12-7. Although the band on the agarose gel for GFMS 12-7 was light, it should still be sufficient for next generation sequencing. An additional product of about 1000bp was also amplified in B390/3 (Figure 3.12), but not produced for any of the GFMS 12 sub-isolates. This suggests that there is a sequence variant of CTV present within the B390/3 source that is not present in any of the GFMS 12 sub-isolates.

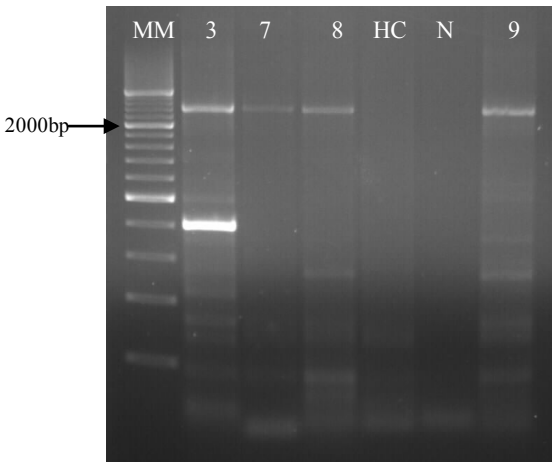

Figure 3.12: Agarose gel photograph of region 4 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.5 Region 5

As in the case with region 2, the expected 2455 bp product could be amplified for all samples with the basic one step RT-PCR protocol with the addition of DMSO (Figure 3.13). In both instances, the lack of amplification in the absence of additional DMSO might indicate that there is complex secondary structure in the specific area of the genome within these regions. The addition of DMSO most likely relaxes the structure to allow the polymerases to bind to the RNA/cDNA.

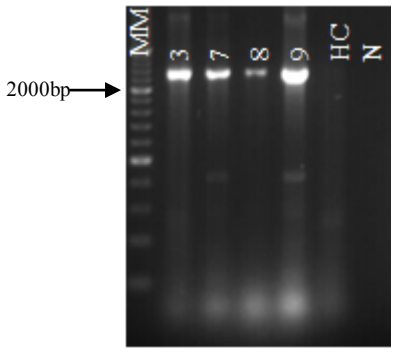

Figure 3.13: Agarose gel photograph of region 5 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.6 Region 6

The basic one step RT-PCR with additional DMSO to relax secondary structures failed to induce the amplification of the expected 2195 bp product for any of the samples except for B390/3. A smaller product of about 1900 bp was amplified for the GFMS 12-8 sub-isolates, as well as several non-specific products (Figure 3.14). Due to the high amount of non-specific amplification, two new primers were designed, 6.1F and 6.1R, using the original 45 complete CTV genomes. These new primers were tested with the one step RT-PCR with an increased $MgCl_2$ concentration of 4 mM, but no amplification could be obtained for any of the samples. Low yields of the expected 2085 bp product could be obtained for B390/3 when using a two-step RT-PCR with additional DMSO and increased $MgCl_2$. When DMSO was omitted but BSA and DTT included, the correct product was still produced for B390/3, but only non-specific amplification was obtained for the GFMS 12 sub-isolates. When a higher concentration of primers was used there was only non-specific amplification products were obtained for all the samples. To rule out RNA degradation during the primer annealing step of the reverse transcription, 5 U of RNase inhibitor (Roche diagnostics, Mannheim, Germany) was added. DTT and BSA were also included in the PCR with a $MgCl_2$ concentration of 3 µM. The correct product was still not obtained but this only increased non-specific amplification of smaller products. At this stage it was apparent that region 6 could not be amplified for the GFMS 12 sub-isolates.

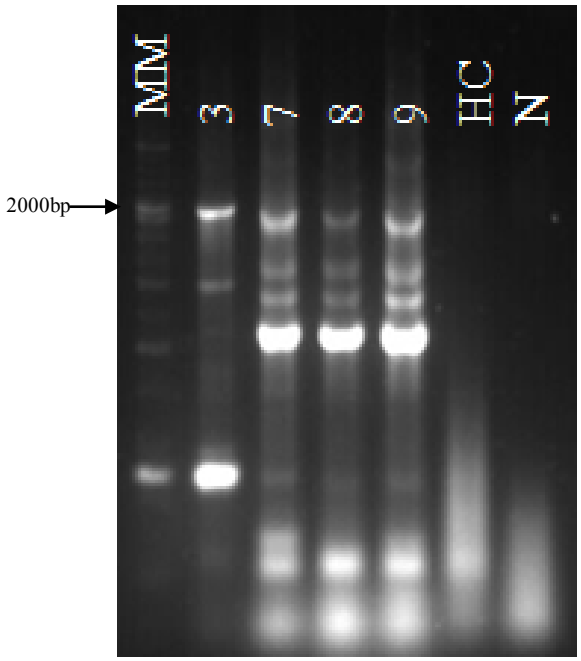

Figure 3.14: Agarose gel photograph of region 6 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.7 Region 7

When the standard one step RT-PCR was used, the expected 2331 bp product could not be amplified for any of the samples. Only non-specific amplification took place. The addition of DMSO did not change the size of products obtained. Due to the amount of non-specific amplification, a new forward primer was designed, 7.1F. With the new primer and the basic one step RT-PCR, the expected 2161 bp product was amplified for GFMS 12-7 and GFMS 12-9, but with very low yield. When a two-step RT-PCR was performed with DMSO and a $MgCl_2$ concentration of 4 mM, the correct product could be amplified for all the GFMS 12 sub-isolates but not for B390/3 (Figure 3.15). The yield of the products was very low and the non-specific amplification was a lot more than the desired products. A slightly lower $MgCl_2$ concentration and additional DTT only increased the non-specific amplification and reduced yield of the desired product. A higher primer concentration in a two-step RT-PCR with DMSO only produced non-specific amplification. Due to the inability to amplify region 7 for

B390/3 and the high amount of non-specific amplification for the GFMS 12 sub-isolates it was decided to exclude region 7 from the sequencing mix.

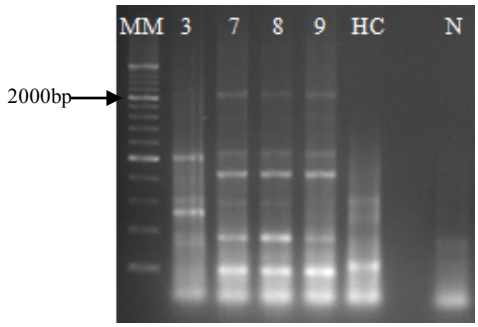

Figure 3.15: Agarose gel photograph of region 7 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.8 Region 8

Using the one step RT-PCR with additional DMSO, an unexpected 350 bp product was amplified. Due to the redesign of some primers in the 5' half of the genome new primers had to be designed for region 8 to still allow overlapping fragment amplification. The new primers, CTV 8.1F and 8.1R only produced non-specific amplification with a one-step RT-PCR. With a two-step RT-PCR and added DMSO and a $MgCl_2$ concentration of 4mM, a very small amount of the expected 2316 bp product could be amplified for the GFMS 12 sub-isolates, but not for B390/3. However, the amount of non-specific amplification was considerably more than the product (Figure 3.16). With variations of $MgCl_2$ and primer concentrations as well as variations in the additions of DMSO, BSA and DTT non-specific amplification was increased but amplification of the desired fragment was not improved. As a last resort the addition of random hexamers during the reverse transcription step was attempted, combined with an annealing temperature of 55°C during amplification, but this only lead to the amplification of multiple smaller products. Due to the several unsuccessful attempts to amplify region 8, the decision was made to not include this region in the sequencing mix.

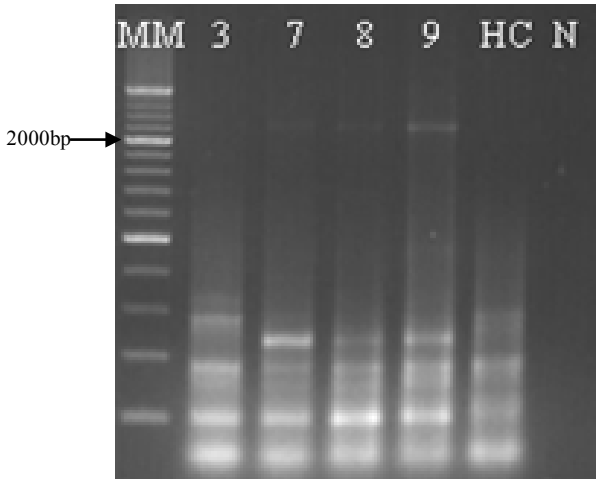

Figure 3.16: Agarose gel photograph of region 8 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.9 Region 9

Although a one-step RT-PCR with DMSO successfully amplified the 2304 bp product for all samples, the yield was very low for the GFMS 12 sub-isolates. A two-step RT-PCR allowed amplification of the region in all samples with limited non-specific amplification (Figure 3.17). To still allow the amplification of the genome in overlapping fragments, a new primer set had to be designed for region 9 with the redesign of the other 5' half primers. The new primers, CTV 9.1F & 9.1R only amplified non-specific products for all samples. Unsuccessful optimization attempts included increasing the primer concentration, increasing $MgCl_2$ concentration, the addition of DMSO, DTT, BSA or combinations thereof, the use of random hexamers in the reverse transcription step and the increase of the annealing temperature during PCR to 55°C. Since the region could not be amplified with the new primers, and the region amplified with the original primers did not overlap with other segments, this region was omitted from the sequencing mixture. It is important for the fragments to overlap since the objective of the study was to either sequence the whole genome in one Illumina run, or just sequence the region that can be assembled in order to test the method. The objective was not to sequence an entire novel genome, but test a method,

therefore sequence gaps would not be filled in later and regions not overlapping were not included.

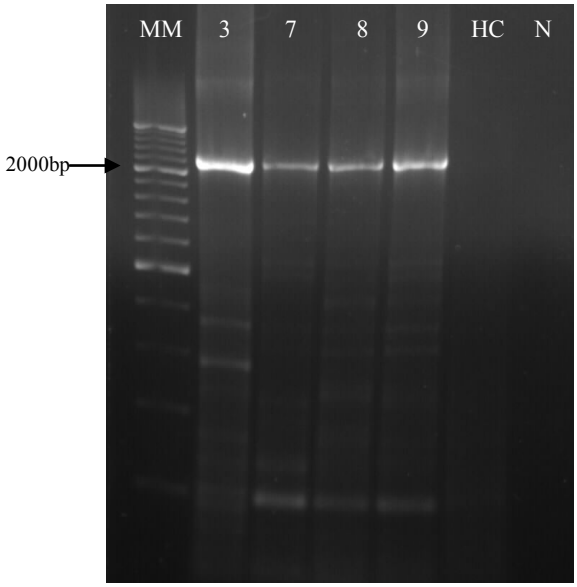

Figure 3.17: Agarose gel photograph of region 9 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.10 Region 10

The standard one step RT-PCR allowed amplification of the correct 2489 bp product in B390/3, but not the other samples. The addition of DMSO produced similar results, but with less amplicon for B390/3 and more non-specific amplification for the GFMS 12 sub-isolates. A two-step RT-PCR also only produced the correct product for B390/3 and non-specific amplification for the GFMS 12 sub-isolates. After the design of a new reverse primer, CTV 10.1 R, the correct 2301 bp product could be amplified for all the samples with only light non-specific amplification that was absent from the healthy control (Figure 3.17). Since no regions flanking region 10 could be amplified, it was excluded from the sequencing mixture.

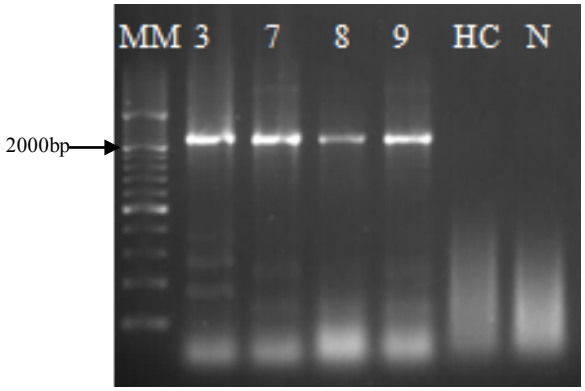

Figure 3.18: Agarose gel photograph of region 10 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.3.11 Region 11

Amplification of region 11 was not successful with the standard one-step RT-PCR with additional DMSO and a $MgCl_2$ concentration of 4mM. After the design of the new CTV 11.1 forward primer, the region could still not be amplified with either a one-step or two-step RT-PCR, irrespective of DMSO or DTT additions and fluctuations in primer and $MgCl_2$ concentrations. The largest product that could be amplified was a 900 bp product in GFMS 12-7 and 12-9 (Figure 3.18). After several futile attempts to amplify the 1872 bp product, it was decided to omit region 11 from the sequencing mixture.

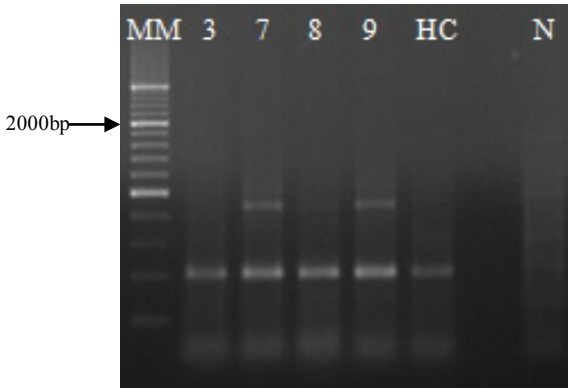

Figure 3.19: Agarose gel photograph of region 11 amplified for GFMS 12 sub-isolates (7, 8 & 9) and B390/3 (3), including healthy control (HC), negative control (N) and Thermo Scientific O'RangeRuler 200 bp molecular marker (MM).

3.3.4 Illumina sequencing

During reference assembly for the GFMS 12-7, 12-8 and 12-9 sources, most reads mapped to the CT-ZA3 genome (KC333868.1) (Figure 3.20 – 3.22), but minor reads mapping to the genomes from other genotypes, including B165 and VT, also occurred. When considering the consensus length (length of consensus sequence produced by mapping of reads to reference genome) of the reads that mapped to CT-ZA3, it was 12627 bp for GFMS 12-7, 11128 bp for GFMS 12-8, and 13784 bp for GFMS 12-9. This correlates with the cumulated amplified fragment of the genome which was approximately 11200 bp. The size difference may be due to some non-specific amplification within the 5' end of the genome due to the degeneracy of the primers. For some reference genomes, there were relatively few reads mapping to them but the consensus length of the mapped reads were quite high. In other instances some reference genomes had many reads mapping to relatively small regions in the genome. The first instance shows that a higher portion of the genome is covered, although the specific areas had only a few reads mapping to it. These included Taiwan-Pum/SP/T1 (JX266712.1) and genomes from the VT genotype. The coverage of mapped reads can be seen in Figure 3.24 – 3.27, with the maximum coverage value indicated between 'Consensus' and 'Coverage' in the figure. Note that the scale for each reference is different. For B390/3 most of the reads mapped to Taiwan-Pum/SP/T1 (JX266712.1), with a very high consensus length of 17008 bp (Figure 3.23). In Figure 3.26 it can be seen that mapping also occurred in the 5' half of the genome which was not specifically amplified as template. As with the GFMS 12 sub-isolates, this might be due to non-specific primer binding to the viral genome during amplification. A small amount of reads mapped to other genomes (Figure 3.22), all of which had a relatively low consensus length, but those with the highest values (1082 – 1890 bp) were all from the RB genotype. In Figure 3.23 – 3.26 it can be seen that the highest mapping of the sequence reads is where the regions amplified overlaps. This is expected since these

areas are sequenced from two different amplicons in each case, and this will also include the primers. Although the presence of the primes will increase the mapping in that region, the area mapped to is larger than the size of the primers themselves. Thus it is not just the primers that were sequenced, and it is expected that amplification took place more efficiently when initiated, but decreased further away from the primers.

The percentage of reads mapping back to all CTV genomes for 12-7, 12-8, 12-9 and B390/3 was 0.36% (of 2.5 million reads), 0.15% (of 2.8 million reads), 0.5% (of 2.1 million reads) and 2.25% (of 10.8 million reads), respectfully. The proportion of CTV-specific reads is very low when considering that CTV-specific amplicons were used as sequencing template. Higher mapping could be accomplished if parameters were set to less stringent conditions, but this resulted in fragments mapping incorrectly to areas within the genome that were not sequenced. The reason for this low mapping with amplicon sequencing is uncertain, but one reason may be sequencing of non-specific DNA seeing that a PCR clean-up was not performed. In future the samples will be gel purified to exclude any primers or remnant background amplification products. After *de novo* assembly, contigs that were greater than 500 bp in length were used to do a BLAST search. For GFMS 12-7, twelve contigs were related to CTV (largest 1699 bp); GFMS 12-8 returned 7 CTV contigs (largest 1801 bp); GFMS 12-9 had 9 CTV related contigs (largest 1921 bp); and B390/3 had 7 CTV related contigs (largest 3145bp).When using BLAST on the NCBI website, all CTV related GFMS 12 sub-isolate contigs were most closely related to CT-ZA3 (KC333868.1) and all CTV related B390/3 contigs blasted to Taiwan-Pum/SP/T1 (JX266712.1). When these contigs were mapped back to the database of complete CTV genomes obtained from GenBank, the same results were obtained. The mapping of these contigs to the reference sequences are shown in Figure 3.28.

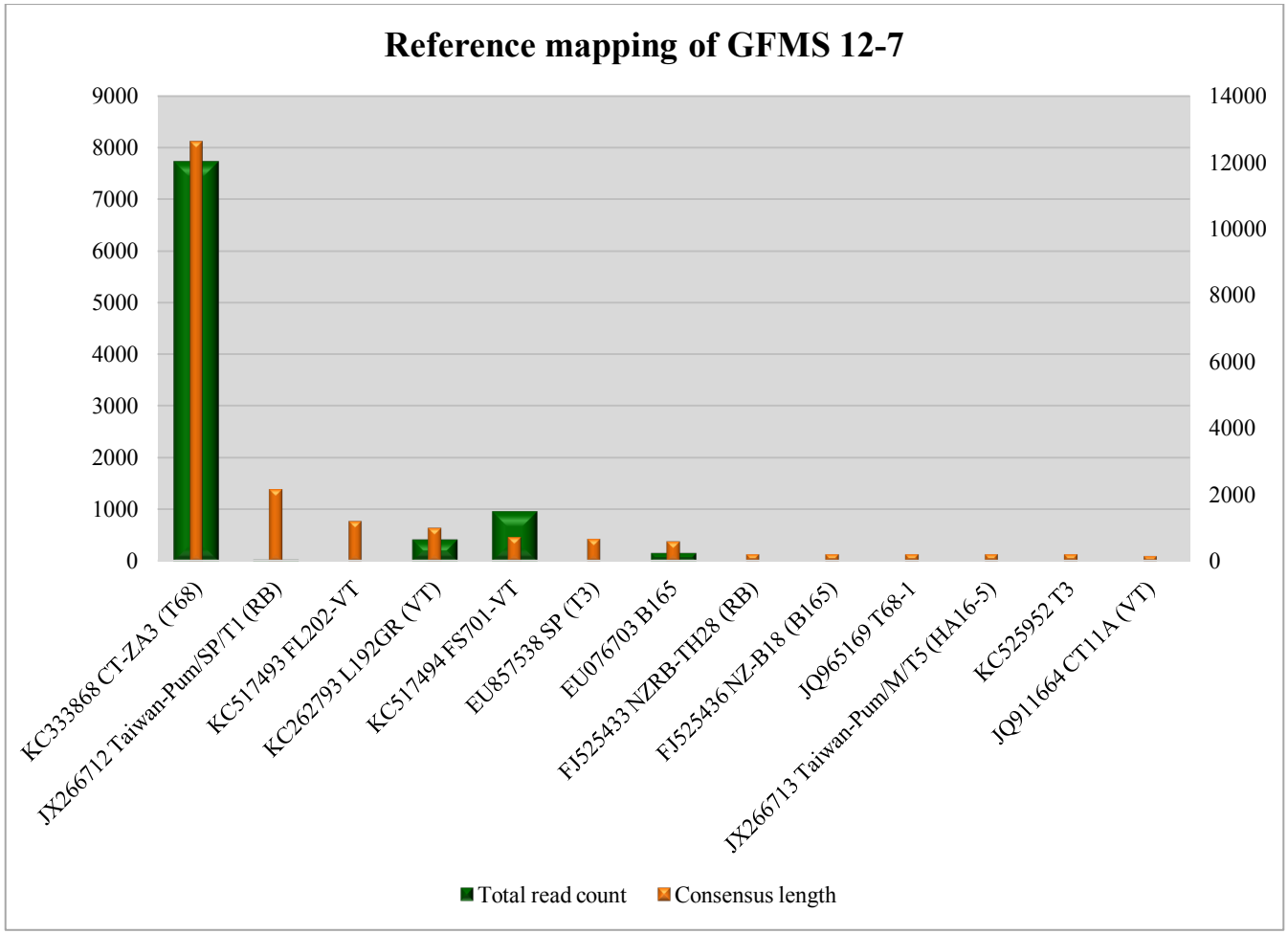

Figure 3.20: Reference mapping of GFMS 12-7 sequence reads against CTV genomes, showing total number of mapped reads and consensus length of the mapped reads in base pairs.

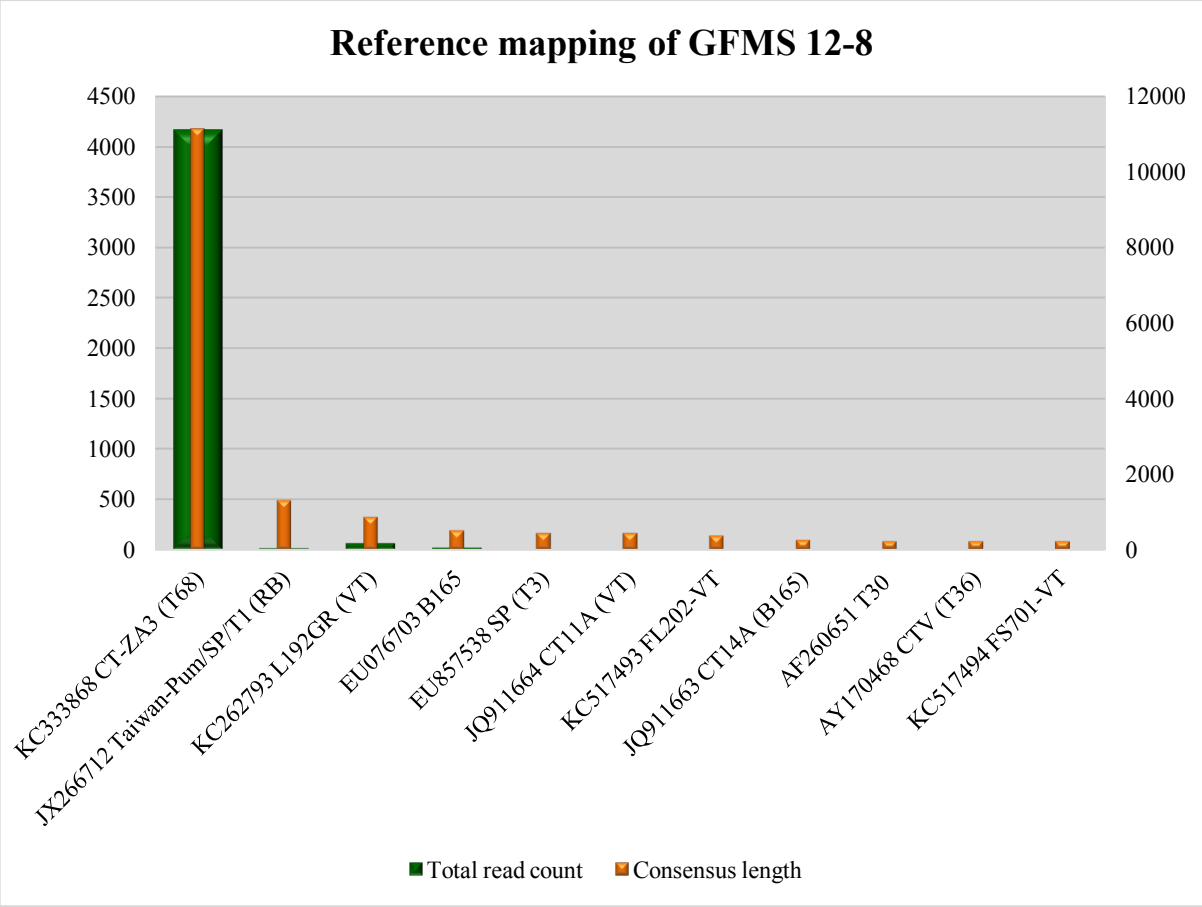

Figure 3.21: Reference mapping of GFMS 12-8 sequence reads against CTV genomes, showing total number of mapped reads and consensus length of the mapped reads in base pairs.

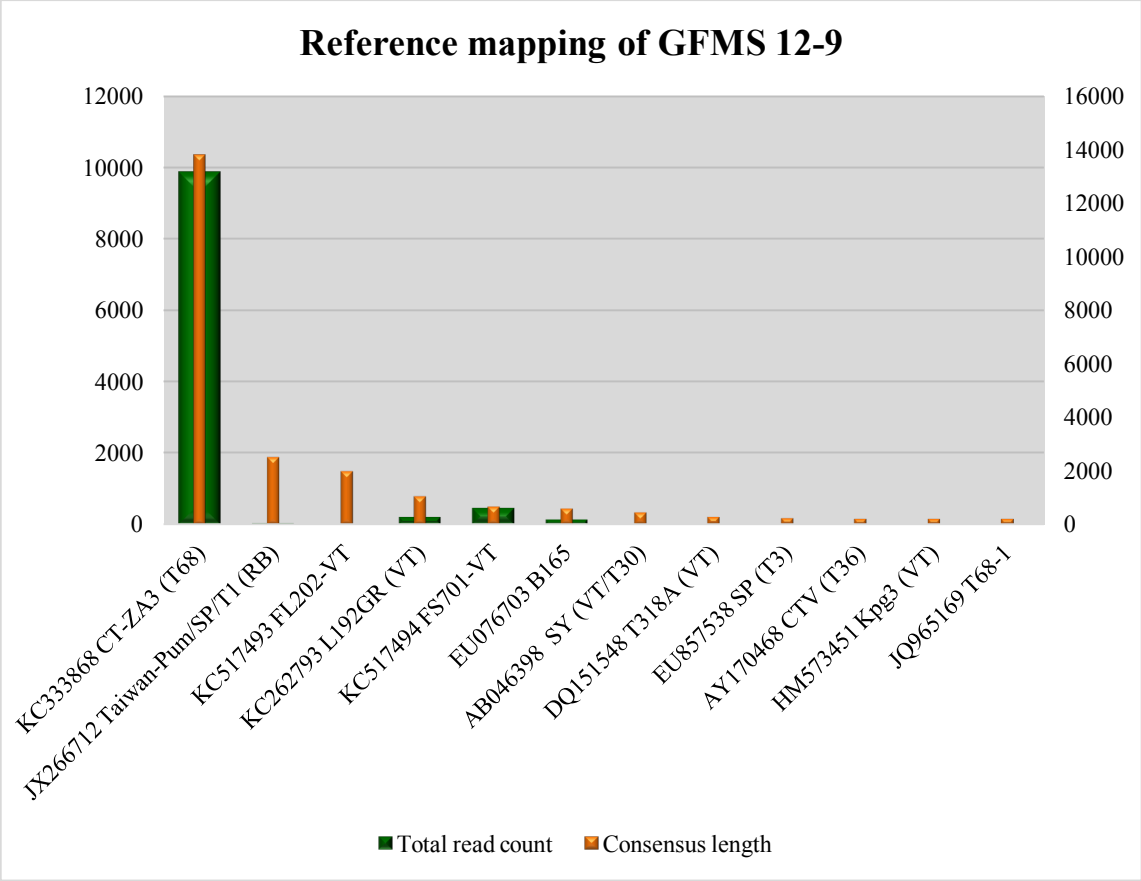

Figure 3.22: Reference mapping of GFMS 12-9 sequence reads against CTV genomes, showing total number of mapped reads and consensus length of the mapped reads in base pairs.

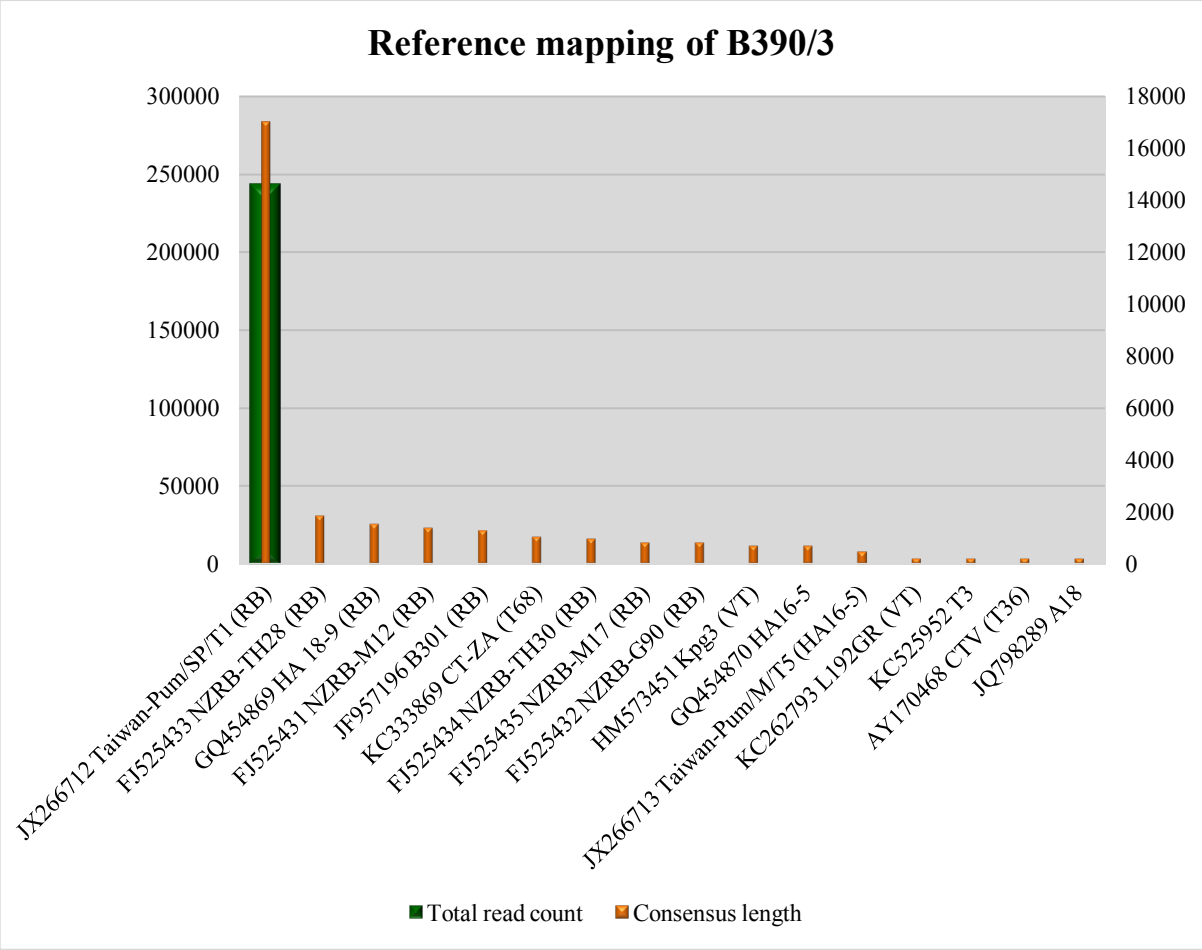

Figure 3.23: Reference mapping of B390/3 sequence reads against CTV genomes, showing total number of mapped reads and consensus length of the mapped reads in base pairs.

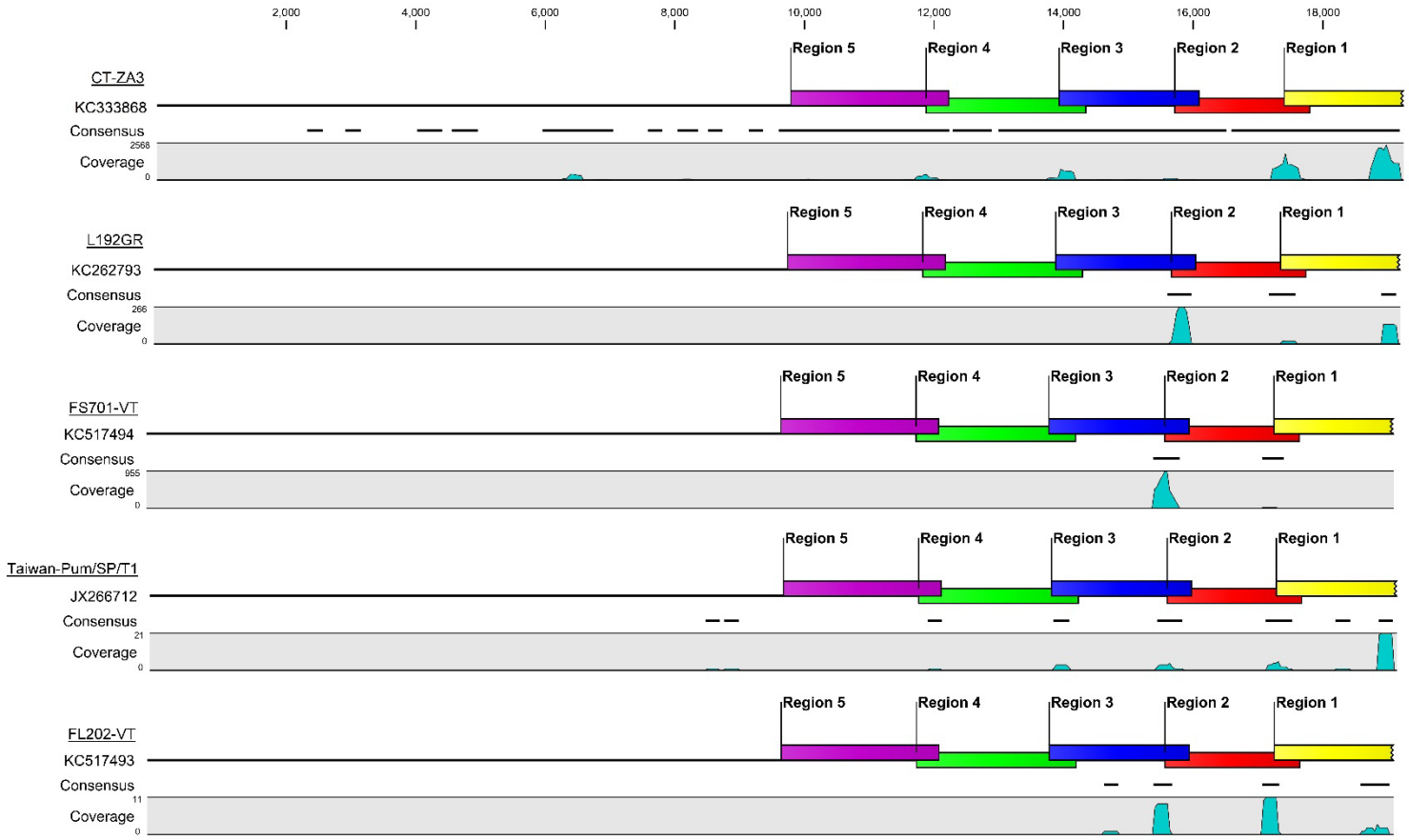

Figure 3.24: Mapping of GFMS 12-7 sequence reads to reference CTV genomes (line next to GenBank accession number), showing areas where a consensus sequence can be created (interrupted line under reference genome) and the coverage of these areas (graph under consensus sequence). Region 1-5 indicate the positions of template amplicons on the genome.

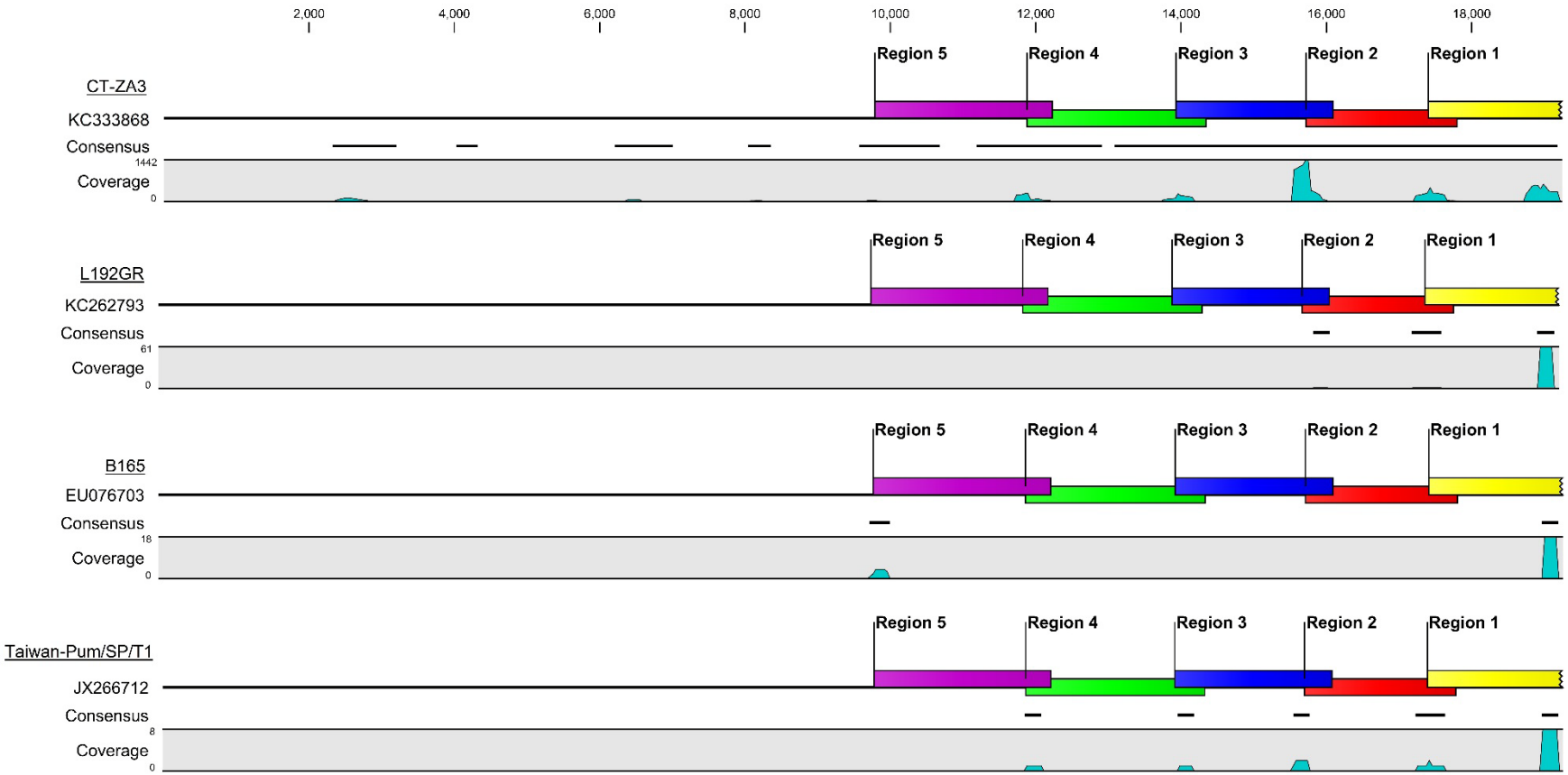

Figure 3.25: Mapping of GFMS 12-8 sequence reads to reference CTV genomes (line next to GenBank accession number), showing areas where a consensus sequence can be created (interrupted line under reference genome) and the coverage of these areas (graph under consensus sequence). Region 1-5 indicate the positions of template amplicons on the genome.

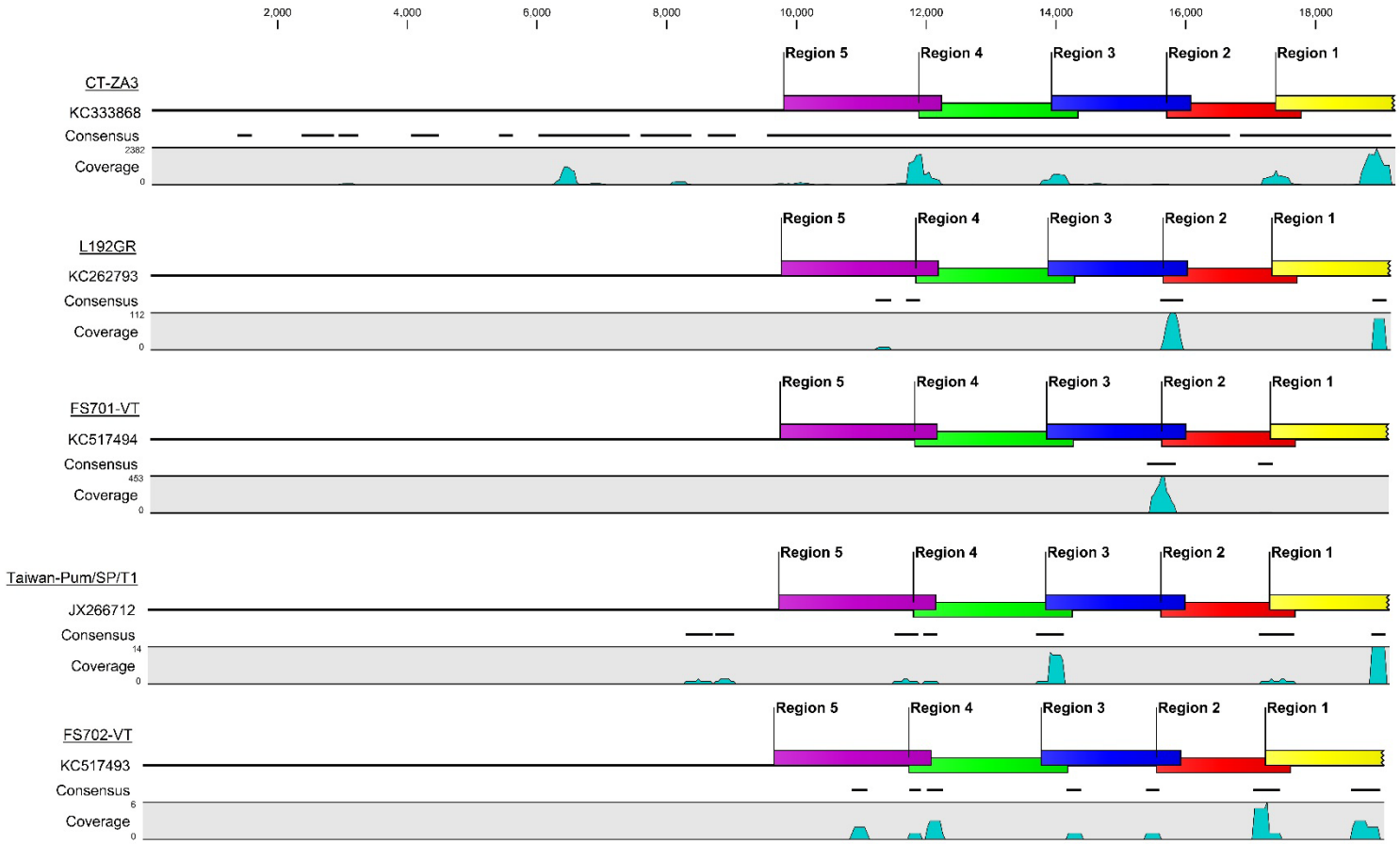

Figure 3.26: Mapping of GFMS 12-9 sequence reads to reference CTV genomes (line next to GenBank accession number), showing areas where a consensus sequence can be created (interrupted line under reference genome) and the coverage of these areas (graph under consensus sequence). Region 1-5 indicate the positions of template amplicons on the genome.

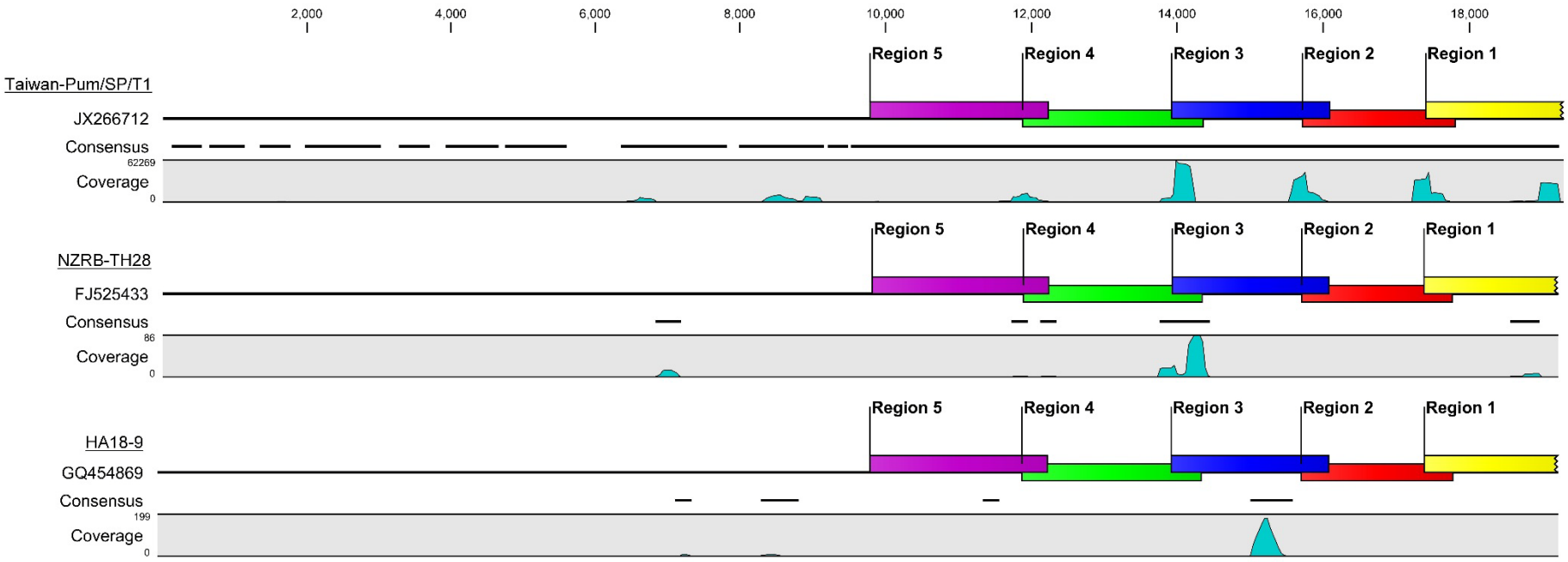

Figure 3.27: Mapping of B390/3 sequence reads to reference CTV genomes (line next to GenBank accession number), showing areas where a consensus sequence can be created (interrupted line under reference genome) and the coverage of these areas (graph under consensus sequence). Region 1-5 indicate the positions of template amplicons on the genome.

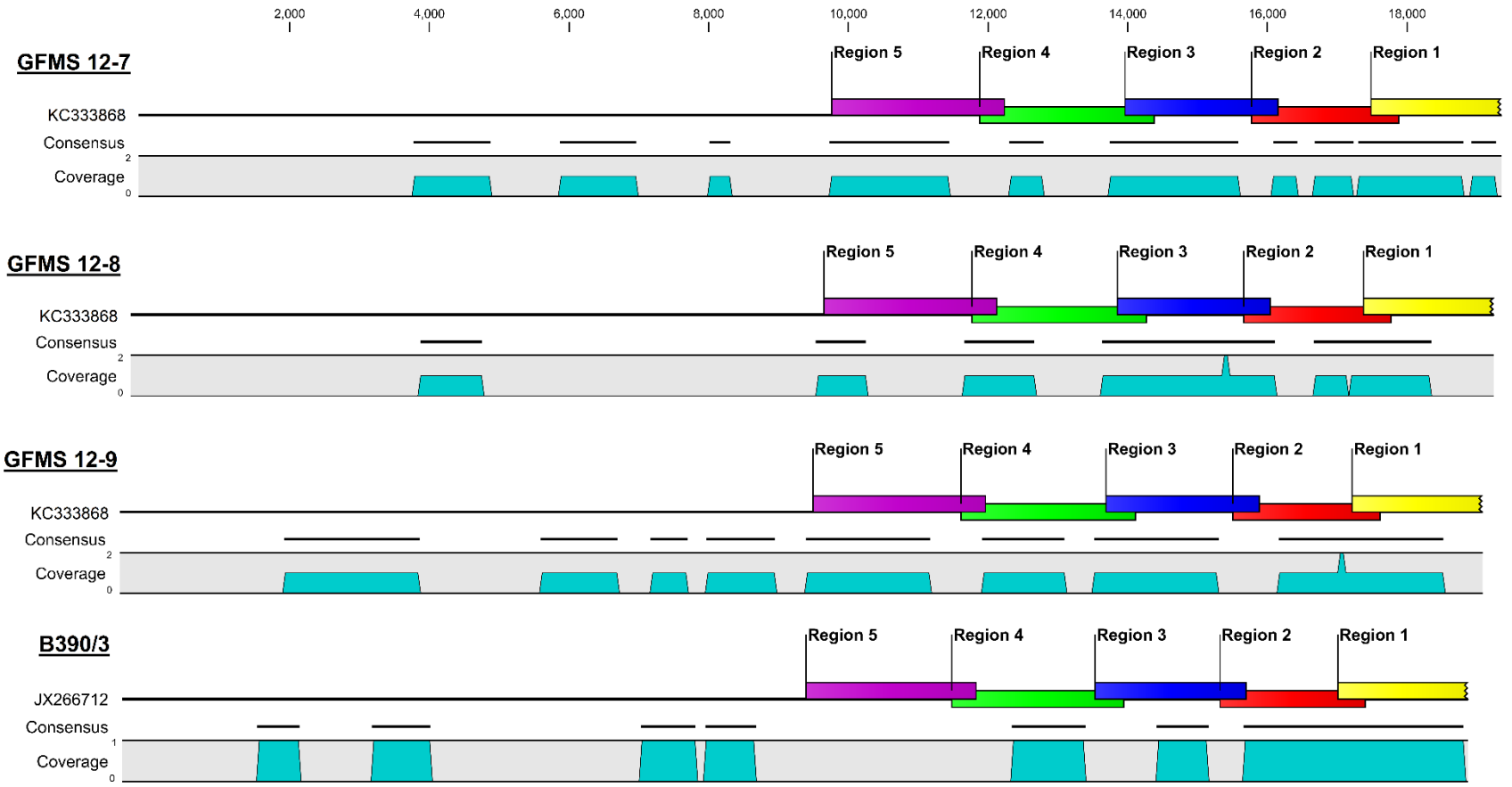

Figure 3.28: Mapping of contigs obtained through de novo assembly for GFMS 12 sub-isolates and B390/3 to CTV reference genomes (line next to GenBank accession number), showing areas where a consensus sequence can be created (interrupted line under reference genome) and the coverage of these areas (graph under consensus sequence). Region 1-5 indicate the positions of amplicons on the genome.

3.3.5 Comparison with previous work

Both the reference mapping and *de novo* results for all the GFMS 12 sub-isolates correlates with the work done by Mr. O. Zablocki, who used different methods of template preparation. Most sequence reads maps back to the CT-ZA3 genome which was the genome derived from GFMS 12-8, found to be the main constituent of all three sub-isolates (Zablocki, 2013). However, minor components detected in GFMS 12-7 differ slightly from that obtained in the Zablocki study, where most of the minor reads mapped to the 3' end of isolate SP, a T3 genotype (EU857538.1) and a VT genotype (EU937519.1) based on the total RNA extraction and viral particle templates. In this study the second and third most mapped reads were to VT genotypes (KC517494.1 & KC262793.1) rather than B165 (EU937519.1). This is most probably due to the fact that these two VT genomes were not available on GENBANK during the Zablocki study. Furthermore a member of the B165 genotype has a few reads mapping to it and a very low amount of reads map to SP (EU857538.1) and Taiwan-Pum/SP/T1 (JX266712.1) identified in the Zablocki study. Similarly the main component for GFMS 12-8 and 12-9 are the same in both studies, with slight differences of the minor components. These differences can be attributed to several factors. First of all, the template preparation between the two studies was different and this could have an influence on the genotypes found (Zablocki, 2013). Secondly this study only focused on the 3' half of the genome and it has been shown that genotyping results can differ for a source based on whether the 3' or the 5' half of the genome is used (Scott *et al.*, 2013). Thirdly, the fact that several additional CTV genomes have become available on GENBANK since the Zablocki study, also influence the BLAST results obtained. A strain may BLAST to a genome similar to it at a given time, but if a more closely related genome becomes available, it will then rather BLAST to the new genome.

During the Scott *et al*. (2013) study it was found that all three GFMS 12 sub-isolates contained a B165/VT recombinant. That study did not include the T68 genome (JQ965169.1) or the newly described CT-ZA3 genome (KC333868.1) and therefore was unable to identify these within the sources. It is possible that the B165/VT recombinant is indeed CT-ZA3 and therefore the results obtained in this study coincides with those obtained by Scott *et al*.(2013). To confirm this, sequences available from GFMS 12-7 in the Scott *et al*. (2013) publication was used to do a new phylogenetic analysis incorporating all parameters as described in the publication, using all 45 CTV genomes available currently. In the phylogram (Excerpt in Figure 3.29) it can be seen that one of the A region clones groups closely with CT-ZA3 (indicated with diamonds). The branch support values circled are too low and therefore all the sequences in the block constitute a single clade. These include two sequences from the 12-7 source and among others, the CT-ZA3 (KC333868.1) and T68 (JQ911663.1) reference genomes. This establishes that the results in this study coincide with the findings by Scott *et al*. (2013).

Interestingly enough the three sub-isolates did not show similar symptoms on indicator hosts (Scott *et al*., 2013), although their predominant strain present are the same and the minor strains are similar but in different ratios. This may indicate that strains present in minor concentrations can still play a role in symptom expression and there might even be interactions between strains that influence the symptoms observed, although this will have to be assessed for these specific sub-isolates in further studies of the CTV sources. In a previous study it was however found that although a predominant strain is transmitted to sub-isolates and stay stable over time, minor strains change over time and influence symptom development (Černi *et al*., 2008).

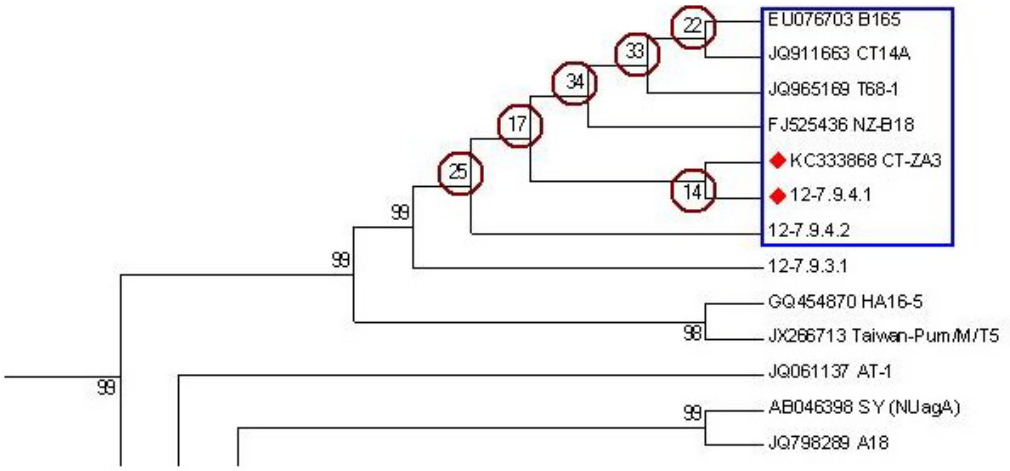

Figure 3.29: Excerpt of the Neighbor joining phylogenetic dendrogram showing the A region of CTV. Sequences included are all CTV genomes available on GENBANK as well as sequences of the GFMS 12-7 sub-isolate from the Scott *et al.* 2013 publication. A Jukes-Cantor model was utilized and Bootstrap values (1000 replicates) are indicated on branches. Diamonds indicate a region clones and block indicate which sequences collapse into one clade due to low branch support (circled).

3.4 Conclusion

The selective amplification of certain regions in various samples indicates the sequence differences between the genotypes present with the sources. This indicates that even when highly degenerate primers are used, certain genotypes can still be missed. The difficulty of amplifying the 5' half of the genome might be due to the secondary structure of the RNA or the cDNA after first strand synthesis. The flaw in the reaction is unlikely to be due to the primers since amplification of each region could be achieved from a plasmid containing the CTV genome. Even with the addition of agents that reduce secondary structure the full length regions could still not be amplified. With increased concentrations of these agents the non-specific amplifications increased, or at very high concentrations, no amplification was obtained. It is suspected that the higher concentrations of these agents prevented the binding of the primers to the template.

All the sources characterized in this study seem to have a predominant CTV genotype present with minor components of other genotypes. This is based on the 3' half of the genome seeing that the 5' half could not be amplified and sequenced. The GFMS 12-7, 12-8 and 12-9

sources all mainly consist of the CT-ZA3 (KC333868.1) genotype and B390/3 contains a strain of the RB genotype most closely related to Taiwan-Pum/SP/T1 (JX266712.1).

Although the results obtained coincides with those of the study done by Zablocki (2014), and Scott *et al*. (2013), this method of template preparation for Illumina sequencing is not recommended for future next generation sequencing as; firstly, we were unable to amplify the complete genome of CTV due to the secondary structure of the virus; secondly, overlapping fragments were amplified and the complete segment of the genome amplified was just over 11 000 bp, but the largest contig obtained from *de novo* assembly was just under 2 000 bp (although when mapping of the contigs was done back to a reference sequence, larger consensus sequences could be obtained); thirdly the amount of CTV-specific reads obtained was lower than that obtained by Zablocki (2013) when using dsRNA as template, and this was the main directive of this study.

Immuno-capture has been employed as a method for template preparation prior to Illumina sequencing but performed poorly in both mapping counts and *de novo* assembly (Zablocki, 2013). The technique was however not optimized in that study and it is possible that with further optimization might become a useful technique. No other method used previously on CTV has the same potential to enrich for the virus as much. If virus specific material only is sequenced, sample for analysis will be much lower, and this will lower the cost of the process. Therefore it might be advantages to look into the optimization of the technique in future studies.

3.5 References

Albiach-Marti, M.R., da Graça, J.V., van Vuuren, S.P., Guerri, J., Cambra, M., Laigret, F., and Moreno, P. 1996. The Effects of Different Hosts and Natural Disease Pressure on Molecular Profiles of Mild Isolates of *Citrus tristeza virus* (CTV). In: *Proceedings of the 13th Conference of the International Organisation of Citrus Virologists. (Graça, J.V., Moremo, P., and Yokomi, R.K.)* pp. 147-153. *IOCV*. Riverside, California.

Albiach-Marti, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., Lopez, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., and Dawson, W.O. 2000. Sequences of *Citrus tristeza virus* Separated in Time and Space Are Essentially Identical. *The Journal of Virology* 74(15), pp. 6856-6865.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215(3), pp. 403-410.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2005. GenBank. *Nucleic Acids Research* 33(Database Issue), pp. 34-38.

Černi, S., Ruščić, J., Nolasco, G., Gatin, Ž., Krajačić, M., and Škorić, D. 2008. Stem Pitting and Seedling Yellows Symptoms of *Citrus tristeza virus* Infection May Be Determined by Minor Sequence Variants. *Virus Genes* 36(1), pp. 241-249.

Folimonova, S.Y. 2012. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology*. 10.1128/jvi.00310-12

Folimonova, S.Y., Robertson, C.J., Shilts, T., Folimonov, A.S., Hilf, M.E., Garnsey, S.M., and Dawson, W.O. 2010. Infection with Strains of *Citrus tristeza virus* Does Not Exclude Superinfection by Other Strains of the Virus. *The Journal of Virology* 84(3), pp. 1314-1325.

Fulton, R.W. 1986. Practices and Precautions in the Use of Cross Protection for Plant Virus Disease Control. *Annual Review of Phytopathology* 24, pp. 67-81.

Karasev, A.V., Dawson, W.O., Hilf, M.E., and Garnsey, S.M. 1998. Molecular Biology of *Citrus tristeza virus*: Implications for Disease Diagnosis and Control. *ACTA Horticulturae* 472, pp. 333-350.

Meyer, J.B., Van Vuuren, S.P., Luttig, M., Manicom, B.Q., and Da Graça, J.V. 2005. Strain Prevalence of *Citrus tristeza virus* Cross-Protecting Isolates Altered by Red Grapefruit Hosts. In: *Proceedings of the 16th Conference of the International Organization of Citrus Virologists. (Hilf, M.E., Duran-Vila, N., and Rocha-Peña, M.A.)* pp. 205-212. *IOCV*. Riverside, California.

Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., ChavarrÍA, F., Shen, G., and Roe, B.A. 2010. Ecogenomics: Using Massively Parallel Pyrosequencing to Understand Virus Ecology. *Molecular Ecology* 19, pp. 81-88.

Scott, K.A., Hlela, Q., Zablocki, O., Read, D., van Vuuren, S., and Pietersen, G. 2013. Genotype Composition of Populations of Grapefruit-Cross-Protecting *Citrus tristeza*

virus Strain GFMS 12 in Different Host Plants and Aphid-Transmitted Sub-Isolates. *Archives of Virology* 158(1), pp. 27-37.

Van Der Vyver, J.B., Van Vuuren, S.P., Luttig, M., and Da Graça, J.V. 2002. Changes in the *Citrus tristeza virus* Status of Pre-Immunized Grapefruit Field Trees. In: *Proceedings of the 15th Conference of the International Organization of Citrus Virologists. (Duran-Vila, N., Miller, W., and Da Graça, J.V.)* pp. 175-185. *IOCV.* Riverside, California.

Van Vuuren, S.P. 2002. Effects of *Citrus tristeza virus* Isolates on Two Tolerant Commercial Scions on Different Rootstocks in South Africa. In: *Proceedings of the 15th Conference of the International Organisation of Citrus Virologist. (Duran-Vila, N., Milne, R.G., and Da Graça, J.V.)* pp. 31-38. *IOCV.* Riverside, California.

Van Vuuren, S.P., Collins, R.P., and Da Graça, J.V. 1991. The Performance of Exotic *Citrus tristeza virus* Isolates as Preimmunizing Agents for Sweet Orange on Sour Orange Rootstock under Natural Disease Pressure in South Africa. In: *Proceedings of the 11th Conference of the International Organization of Citrus Virologists. (Brlansky, R.H., Lee, R.F., and Timmer, L.W.)* pp. 60-63. *IOCV.* Riverside, California.

Van Vuuren, S.P., and Van Der Vyver, J.B. 2000. Comparison of South African Pre-Immunizing *Citrus tristeza virus* Isolates with Foreign Isolates in Three Grapefruit Selections. In: *Proceedings of the 14th Conference of the International Organisation of Citrus Virologists. (Da Graça, J.V., Lee, R.F., and Yokomi, R.K.)* pp. 50-56. *IOCV.* Riverside, California.

Van Vuuren, S.P., Van Der Vyver, J.B., and Luttig, M. 2000. Diversity among Sub-Isolates of Cross-Protecting *Citrus tristeza virus* Isolates in South Africa. In: *Proceedings of the 14th Conference of the International Organisation of Citrus Virologists. (Da Graça, J.V., Lee, R.F., and Yokomi, R.K.)* pp. 103-110. *IOCV.* Riverside, California.

Von Broembsen, L., and Lee, T.C. 1988. South Africa's Citrus Improvement Programme. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Timmer, L.W., Garnsey, S.M., and Navarro, L.)* pp. 407-416. *IOCV.* Riverside, California.

Zablocki, O. 2013. Unbiased, Next-Generation Sequencing for the Characterization of *Citrus tristeza virus* Populations. Dissertation submitted to the Faculty of Natural and Agricultural Sciences Department of Microbiology and Plant Pathology in partial fulfillment of the requirements for the degree Magister Scientiae (MSc) Microbiology. University of Pretoria.

Zablocki, O., and Pietersen, G. 2014. Characterization of a Novel *Citrus tristeza virus* Genotype within Three Cross-Protecting Source GFMS12 Sub-Isolates in South Africa by Means of Illumina Sequencing. *Archives of Virology*, pp. 1-7.

Chapter 4: Determination of the genotype composition of CTV mild strain cross-protection candidate New Venture 41/2 and isolation of component genotypes

4.1 Introduction

Citrus trees are long-lived and exposed to many aphids in the field, creating the opportunity for multiple infections with *Citrus tristeza virus* (CTV) and this can lead to one host being infected with multiple strains of this virus (Rubio *et al.*, 2001). The composition and dominance of strains within CTV sources are influenced by the host species as well as environmental conditions (Albiach-Marti *et al.*, 1996), hence differences may be observed in CTV derived from a single source, based on the host, geographical location of the host and whether the source is maintained under glasshouse or field conditions. This may be due to super-infection by other CTV strains within a field or due to the increased replication of some strains under specific conditions.

The complex nature of CTV populations complicates the mild strain cross-protection strategy which relies on the genotype specific super-infection exclusion principle (Folimonova *et al.*, 2010). This is where the presence of a specific genotype of a virus within a plant protects it from secondary infections of strains of the same genotype of that virus. Therefore the effectiveness of a cross-protection source may differ in an array of hosts and under different environmental conditions. If a secondary infection of a citrus plant occurs by a severe strain of a genotype that is not present within the cross-protecting source, the plant will not be protected and is likely to express severe symptoms which may lead to economic losses. Conversely, it is also possible that selection of specific strains within a mixed genotype cross-protection source can take place under different conditions, and therefore if a severe strain is present, cross-protection breakdown may occur even without additional secondary infections. This was the case with GFMS 12 which was found to consist of multiple strains, including a strain causing severe symptoms (Van Vuuren *et al.*, 2000).

The CTV genotypes circulating in an area needs to be thoroughly determined in order to know which genotypes need to be protected against. The requisite cross-protecting source

should contain mild sequence types of all genotypes present to allow complete protection of the plant against secondary infections within the field. As it is unlikely to find a natural source of the virus with these attributes, it is expected that an artificial source containing a blend of mild strains of every genotype will need to be created. To achieve this, one needs to find a pure source containing a single mild strain of a given genotype and test it on different indicator hosts to confirm it will not produce severe symptoms. Once a pure, mild strain for every genotype is established, all these strains can be introduced into single trees to see if cross-protection is accomplished when challenged with more severe CTV sources. It still needs to be established whether a genetically engineered cross-protection source of this type will be stable, since the possibility exists that the strains within the source may outcompete each other and that some may become remnants and no longer provide cross-protection.

To obtain a pure source of CTV is not a trivial task. When dealing with a CTV population containing multiple strains, single aphid transmissions (SATs) can be done with the hope of an aphid acquiring only one viral strain within the CTV population and transmitting that to a subsequent host. In this way the procedure has a bottleneck effect but might need to be repeated multiple times in order to get only one strain of CTV within a host (Van Vuuren and Van Der Vyver, 2000). Although the technique is simple in theory, it requires considerable replicates due to the low efficiency with which the virus is transmitted.

The transmission efficiency of SATs can be described as the ratio between the number of infected plants and the total number of single aphid inoculated plants, and is usually fairly low (Broadbent *et al.*, 1996). The transmission efficiency of SATs is influenced by several factors. The particular isolate of CTV can influence the efficiency, with some isolates being transmitted more readily than others (Lin *et al.*, 2002). For instance, severe isolates are often transmitted with a higher efficiency than mild isolates. The original host plant from which the isolate was obtained and the recipient plant also plays a role (Broadbent *et al.*, 1996;

Hermoso *et al.*, 1988; Lin *et al.*, 2002). For example, isolates from orange or mandarin are more readily transmitted than isolates obtained from grapefruit (Broadbent *et al.*, 1996). The environmental conditions during transmission and the morphological stage of the aphid can also influence the efficiency (Huang *et al.*, 2005). The transmission efficiency can be as low as 1% or 1.5% in some cases (Lin *et al.*, 2002) but may be as high as 55% in other cases (Broadbent *et al.*, 1996).

Due to the low efficiency of SATs, an alternative method can be used in the hope of diluting the virus and obtaining pure sources. This involves the mechanical transmission of the virus through either a contaminated blade or with bark extracts (Garnsey *et al.*, 1977). While the efficiency of transmission is very low (0.13% – 0.3%) it can serve as a good alternative to SATs as it is much less labor intensive and does not require the maintenance of an aphid colony.

Potential cross-protection sources need to be characterized to establish which genotypes are present, and sub-isolates of sources need to be characterized to check if they are homogenous. There exist many different mechanisms for the characterization of viral populations. When using gene specific primers to amplify certain regions of the genome care in the design of the primers needs to be taken to avoid the possibility of introducing amplification bias (Read and Pietersen, 2015) ,moreover this approach also only provides information on the specific gene amplified. Since recombination is common within the CTV genome (Roy and Brlansky, 2010; Rubio *et al.*, 2001; Vives *et al.*, 2005), targeting only a specific gene may provide misleading information regarding the genotype as recombination events occurring elsewhere in the genome would not be detected. This approach is however warranted when targeting a gene important for a specific biological function (Read and Pietersen, 2015), for example the p33 gene which has been shown to be involved in the genotype specificity of the super-infection exclusion mechanism of CTV (Folimonova, 2012). While useful for the specific

goal of identifying the genotypic variation within that specific gene, conclusions can only be made regarding that gene and not the actual genotype variation, which would require the sequence determination of the entire genome. To sequence the entire genome has become plausible with next generation sequencing on various platforms including Illumina sequencing (Zablocki and Pietersen, 2014). Various methods of template preparation prior to Illumina sequencing have been explored, including immuno-capture of virus particles (Zablocki and Pietersen, 2014). While immuno-capture did not produce the expected results, the undoubted potential of the technique suggests that it probably just required some optimization (Zablocki, 2013). Immuno-capture would allow the isolation of CTV particles with monoclonal antibodies, eliminating most of the unwanted RNA and enriching for the virus before RT-PCR. One of the aims of the current study is to optimize the immuno-capture technique and to overcome the hurdle of reverse transcribing and amplifying the single stranded genomic RNA (ssRNA) of CTV without using gene specific primers prior to Illumina sequencing.

The New Venture 41/2 source was identified as a useful candidate cross protecting source during a 2004 to 2015 study by Citrus Research International (CRI). Budwood was collected from 108 field grown grapefruit trees infected with CTV, but lacking any symptoms. The sources were established in a glasshouse and inoculated onto virus free Mexican lime, to assess the severity of the sources. After biological indexing in glasshouse trials the most promising sources were subjected to field trials. New Venture 41/2 was amongst the most promising of these sources. It was inoculated onto Star Ruby trees planted in February 2007 at Bosveld Sitrus in the Letsitele area, as well as Marsh trees planted in March 2007 at Riverside in the Malelane area. The trees are being evaluated for growth and stem pitting and although differences have already been seen between sources, it is too early to draw

conclusions. The trees will be evaluated until 2015 for growth, production, fruit size and tree health (Breytenbach *et al.*, 2014).

During a previous study, the variability of the 1a gene sequence of the New Venture 41/2 source was characterized (Lubbe, *unpublished results*). Based on the gene fragment characterized in that study it was found the source contained strains from the VT genotype. As it produces mild symptoms in field trials, it was decided to use this source for SATs and mechanical transmissions in an attempt to obtain a pure, mild source of VT.

4.2 Materials and Methods

4.2.1 Aphid colony

Aphids were collected at Braam Pretorius Street, Montana Park, Pretoria, from *Vepris lanceolata* on 21 June 2013 to use for establishing the aphid colony (Accession number 13-1001). The aphids were identified as *Toxoptera citricida* by Ian Millar at ARC-PPRI based on morphology, and based on sequencing of the mitochondrial cytochrome C oxidase subunit I gene (Inge Pietersen, *personal communication*) using the HCO2198 and LCO1490 primers (Vrijenhoek, 1994). The colony was established on Mexican lime seedlings planted on 17 January 2013 in an insect proof cage housed in a glasshouse. Greenhouse temperatures fluctuated considerably ranging from 4°C to 40°C, but on occasion even reached temperatures of around 50°C. The aphids were tested to be CTV free directly after collection through the sacrifice of representative individuals and seedlings were tested to be CTV free five months after planting using real time RT-PCR as described in 4.2.4.

4.2.2 Single aphid transmission

The New Venture 41/2 source tree used for aphid transmissions was a Mexican lime seedling (Accession number 13-1614) infected with CTV through bark strip inoculation of the New Venture 41/2 source tree (Accession number 11-0051) a year prior to transmissions. A

healthy seedling from the virus free *T. citricida* colony with multiple aphids feeding on it was pulled out of the soil and put next to the source tree to allow aphids to move of their own volition over *3-5* days to the CTV infected tree as the seedling wilted.

Aphids were removed from the CTV infected tree using an artist brush and put in a Petri dish to starve for 1-2 hours. Single aphids were then transferred to individual CTV free seedlings and aphid cages placed over them. Aphids were allowed to feed for at least 24 hours and then killed by crushing. These plants were kept under glasshouse conditions for 6 months before being tested for the presence for CTV with conventional RT-PCR. To minimize labor and reagent costs, leaves were collected from all the plants in a planting pot (n = 8 to 12), pooled, and subjected to RNA extraction and CTV detection. It was envisaged that once the group of plants from a pot tested positive, the individual plants within that pot would be tested separately. SAT-derived plants testing positive for CTV would be screened for CTV genotypes present using the method described in 4.2.6, and plants testing positive for single genotypes subjected to Illumina sequencing to confirm the purity of the source.

4.2.3 Mechanical transmission

Leaves of the New Venture 41/2 source tree (Accession number 11-0051) were macerated with the HOMEX 6 (Bioreba, Reinach, Basel-Landschaft, Switzerland) with 10 ml phosphate buffered saline (PBST) and the resulting sap was placed in 1.5 ml tubes. For the first 20 plants the scalpel was dipped in the plant sap, used to cut the seedling's stem to the phloem (10 - 20 cuts with a 45° angle) and then cleaned with bleach before repeating the process on the same seedling twice to do it a total of three times. For the next 20 plants the process was only done twice and on the next 20 plants only once. For the next 20 plants the scalpel was only dipped in the plant sap every second seedling, i.e. dip scalpel in sap and slash two seedlings (5 slices with 45° angle) before cleaning with bleach. For the last 20 plants the scalpel was only dipped in the sap once and then used to slash all the seedlings once.

4.2.4 One step real time RT-PCR for CTV detection in source plant and immune-capture optimization

CTV detection was done based on the method described by Bertolini *et al* (2008). 2.5 µl of the total RNA extract obtained with the GeneJET Plant RNA Purification Mini Kit (Thermo Scientific, Waltham, Massachusetts) was combined with 2 µl TaqMan® Universal PCR Master Mix, 4 U Roche RNase Inhibitor, 2 U Roche AMV Reverse Transcriptase (Roche Diagnostics, Germany), 0.5 µM 3'UTR1 forward primer, 0.5 µM 3'UTR2 reverse primer, 0.15 U 181T TaqMan® probe and RNase free water to a volume of 10 µl. The reaction was carried out in the LightCycler 1.5 (Roche Diagnostics, Germany) with 30 min at 48°C and 10 min at 95°C for reverse transcription, followed by 45 cycles of 15 sec at 95°C and 1 min at 60°C for amplification.

4.2.5 Conventional two step RT-PCR for CTV detection in SAT plants

Total RNA extract (2µl) obtained according to manufacturer's specifications with the GeneJET Plant RNA Purification Mini Kit (Thermo Scientific, Waltham, Massachusetts) was combined with 1µl of 10 µM Generic R primer (Table B.3, Appendix B) and 7 µl of nuclease free water and incubated for 3 min at 95°C and then 1 min on ice. After primer annealing, the RNA was added to 4 µl RT buffer, 0.21 U AMV reverse transcriptase (Roche Diagnostics, Germany), 0.5 U Ribolock RNase Inhibitor (Thermo Scientific, Waltham, Massachusetts), 1 mM of each dNTP, 10 mM DTT and nuclease free water to a final volume of 20 µl. Reverse transcription occurred at 42°C for 60 min followed by a 5 min incubation at 85°C to stop the reaction. Amplification was done by using 2 µl of synthesized cDNA, 10µl of the GoTaq Hot Start Green Master Mix (Promega, Madison, Wisconsin), 0.375 µM of each of the Generic forward and reverse primers (Table B.3, Appendix B) and 6.5 µl of nuclease free water for a final volume of 20 µl. Reaction conditions were as follows: 2 min at 94°C, 35 cycles of 20

sec at 94°C, 30 sec at 60°C and 20 sec at 72°C, finishing with an extension for 1 min at 72°C. Amplicons were visualized on a 1% agarose gel stained with 5 µl/L ethidium bromide (EtBr).

4.2.6 Genotype detection in CTV positive plants

Detection of CTV genotypes was based on genotype specific primers used in a RT-PCR. The reaction used for VT, RB, B165 and T36 detection is described in 4.2.5, and the reactions for the HA16-5, T30 and T3 detection used the same reaction as described in 4.2.5, except for the use of the DreamTaq Green PCR Master Mix 2X (Thermo Scientific, Waltham, Masseuses, US), instead of the GoTaq Green Master Mix. The annealing temperatures for the genotype specific primers (Table B.3, Appendix B) differed; annealing temperatures for primer pairs T36, T30, T3, VT, NZRB1 and NZRB2 was 60°C, B165 was 59°C and HA16-5 was 56°C.

4.2.7 Immuno-capture optimization for use as template for Illumina sequencing

In order to find the optimal dilution of antibody to use during immuno-capture, an experiment was performed to determine the saturation binding point of antibodies in the coating step. Five different dilutions (1:10000, 1:7500, 1:5000, 1:2500 and 1:1000 in coating buffer (1.59 g/L Na_2CO_3 and 2.93 g/L $NaHCO_3$ with a pH of 9.6)) of the anti CTV CREC 29 antibody (University of Florida, Citrus Research Education Center) were used to coat the 60 central wells of two ELISA plates. The plates were incubated overnight at 4ºC and then washed 3 times for 3 minutes with PBST buffer (8 g/L NaCl, 0.2 g/L KH_2PO_4, 1.15 g/L anhydrous Na_2HPO_4 and 0.2 g/L KCL (pH 7.4) with 1 ml/L Tween-20). As a positive control, 8 g of a mix of plant material infected with CTV (Accession numbers 08-0001, 08-0010 and 11-0051) was macerated with the HOMEX 6 (Bioreba, Reinach, Basel-Landschaft, Switzerland) in 20 ml coating buffer and then 200 µl of the resulting sap placed in the wells. As a negative control, 200 µl of plant sap originating from the maceration of 3.5 g healthy Mexican lime (Accession number 08-0029) in 18 ml coating buffer, was placed in wells designated as

negative controls. Buffer controls were also included. The plates were incubated overnight at 4ºC. After incubation the plates were rinsed 5 times with PBST for 3 minutes and then 200µl of a 1:20000 dilution of the goat anti CTV G604 antibody (University of Florida, Citrus Research Education Center) in conjugate buffer (500 ml/L PBST, 10 g/L PVP-40 and 1 g/L BSA) was added, incubated for 1 hour at 37ºC and then washed 3 times for 3 minutes with PBST buffer. For the conjugation step, 100 µl of a 1:30000 dilution rabbit anti goat antibody conjugated to alkaline phosphatase was placed in wells and incubated for 3 hours at 37ºC. A wash step consisting of 5 repeats of 3 minute incubations with PBST was performed. During the substrate reaction step 200 µl of a 0.6 mg/ml of para-nitrophenyl phosphate solution was added to the wells and after 10 minutes the first reading was done, followed by readings every 15 minutes with the Multiskan Go (Thermo Scientific, Vantaa, Finland).

In order to find the best buffer to use for virus release after immuno-capture, 200 µl of a 1:5000 dilution of CREC 29 in coating buffer was added to the 60 central wells of the ELISA plate and incubated overnight at 4ºC. A wash step of 3 minutes x 3 with PBST was carried out before 200 µl of the New Venture 41/2 sample (Accession number 11-0051) as prepared previously was added to 40 wells. As a negative control 200 µl of coating buffer was added to 20 wells. After 4 hours for incubation at 37ºC, a wash step of 10 minutes x 2 and 3 minutes x 10 was performed with PBST. For virus release 200 µl of each virus release buffer (Table 4.1) was placed into 3 cells, 2 with samples and one as negative control. The ELISA plate was covered with microplate adhesive film and wet paper towel to prevent evaporation and incubated at 94ºC for 10 minutes. After incubation the extracts were placed in 1.5 mL Eppendorf tubes. Some of the RNA for each buffer was used in a real time RT-PCR (described in 3.2.4) to test which buffer releases the most RNA, while the rest was stored at -80ºC for a few days until cDNA synthesis and amplification was done for Illumina sequencing.

Table 4.1: Virus release buffers tested. GES, buffer containing glycine, NaCl and EDTA. NaCl, sodium chloride. EDTA, Ethylenediaminetetraacetic acid. DMSO, dimethyl sulfoxide. TE, buffer containing TRIS and EDTA.

Buffer	Contents	Reference
Buffer 1	GES (0.1 m glycine pH: 9, 50 mm NaCl, 1 mm EDTA)	(Papayiannis *et al.*, 2010)
Buffer 2	GES + 0.5% Tween-20	
Buffer 3	GES + 0.5% Triton-X	
Buffer 4	GES + 0.5% DMSO	
Buffer 5	GES + Betaine (0.25 g/ml)	
Buffer 6	TE	
Buffer 7	TE + 0.5% Tween-20	
Buffer 8	TE + 0.5% Triton-X	
Buffer 9	TE + 0.5% DMSO	
Buffer 10	TE + Betaine (0.25 g/ml)	
Buffer 11	Type 2 analytical-grade water	
Buffer 12	Type 2 analytical-grade water + 0.5% Tween-20	
Buffer 13	Type 2 analytical-grade water + 0.5% Triton-X	
Buffer 14	Type 2 analytical-grade water + 0.5% DMSO	
Buffer 15	Type 2 analytical-grade water + Betaine (0.25 g/ml)	
Buffer 16	GES + 0.5% Tween-20 + Betaine (0.25 g/ml)	
Buffer 17	GES + 0.5% Triton-X + Betaine (0.25 g/ml)	
Buffer 18	GES + 0.5% Triton-X + 0.5% Tween-20 + Betaine (0.25 g/ml)	
Buffer 19	10mM Trizma HCI, 1.0% Triton X-l00	(Harju *et al.*, 2005)
Buffer 20	50mM Sodium Acetate pH 5.6 (Acetic acid used for pH adjust)	

4.2.8 Random amplification of RNA obtained from immuno-capture for Illumina sequencing

The New Venture 41/2 RNA obtained using immuno-capture and the TE + 0.5% Tween 20 virus release buffer which was found to be most promising during real time RT-PCR, was subjected to reverse transcription using a reverse primer mix containing a mix of the degenerate primers used in Chapter 2 and random hexamers (Table B.4, Appendix B). 2 µl of this reverse primer mix was combined with 2 µl RNA, 2 µl 10x BSA, 10µl 100% DMSO and 7 µl water. This mixture was incubated at 95℃ for 3 min and then left on ice for 1 minute for primer annealing. This 10 µl mix was combined with 1x RT Buffer, 25 U AMV Reverse Transcriptase, 10 U Protector RNase Inhibitor (Roche Diagnostics, Mannheim, Germany), 1 mM of a dNTP mix (KAPA Biosystems, Cape Town, South Africa), 10 mM DTT (Thermo

Scientific, Waltham, Masseuses, US), and PCR grade water to a total volume of 2 0 µl. Reverse transcription was performed at 42°C for 60 min, followed by a 85°C incubation for 5 min to inactivate the enzymes.

For amplification of the cDNA, 6 µl of the random cDNA mixture was used in a 60 µl PCR reaction containing 6 µl of both the forward and reverse primer mix, 30 µl GoTaq Hot Start Green Master Mix (Promega, Madison, Wisconsin) and 12 µl PCR grade water. Reaction conditions were as follows: 94°C for 2 min, followed by 60 cycles of 94°C for 20 sec, 55°C for 30 sec and 72°C for 20 sec. This was finished off with a final extension period of 1 min at 72 °C. Amplification products were visualized by running 10 µl on a 1% agarose gel stained with EtBr.

4.2.9 Confirmation of the presence of the CTV genome in the randomly amplified DNA

In order to confirm the presence of several different segments of the CTV genome in the randomly amplified cDNA, a PCR (i.e. no reverse transcriptase step) was conducted using several different primers to amplify different regions of the genome. The primers used are described in Table B.5, Appendix B, and target different regions of the genome, including the 3' end, 5' half and the center of the genome.

Amplification was done by using 1 µl of the randomly amplified cDNA, 10µl of the GoTaq Hot Start Green Master Mix (Promega, Madison, Wisconsin), 0.375 µM of each of the forward and reverse primers (Table B.5, Appendix B) and 6.5 µl of nuclease free water for a final volume of 20 µl. Reaction conditions were as follows: 2 min at 94°C, 35 cycles of 20 sec at 94°C, 30 sec at 60°C and 20 sec at 72°C, finishing with an extension for 1 min at 72°C. Amplicons were visualized on a 1.5% agarose gel stained with 5 µl/L EtBr.

4.2.10 Immuno-capture template preparation for Illumina sequencing

The template preparation for Illumina sequencing was done by immuno-capture and random cDNA synthesis followed by amplification. For the immuno-capture, 200 µl of a 1:5000 dilution of the CREC 29 antibody in coating buffer (1.59 g/L Na_2CO_3 and 2.93 g/L $NaHCO_3$ with a pH of 9.6) was added to the wells of the ELISA plate and incubated overnight at 4ºC. A wash step of 3 X 3 minutes with PBST (8 g/L NaCl, 0.2 g/L KH_2PO_4, 1.15 g/L anhydrous Na_2HPO_4 and 0.2 g/L KCL (pH 7.4) with 1 ml/L Tween-20) was carried out before 200 µl of the sample was added to the wells. The samples were prepared by macerating leaf petioles, midribs and bark shavings in coating buffer (2 ml buffer per 0.5 g sample) using the HOMEX 6 (Bioreba, Reinach, Basel-Landschaft, Switzerland). As a negative control 200 µl of coating buffer was added to 20 wells. After 4 hours of incubation at 37ºC, a wash step of 10 minutes x 2 and 3 minutes x 5 was performed with PBST. For virus release, 200 µl of virus release buffer (Buffer 7: TE Buffer with 0.5% Tween-20) was placed into the wells. The ELISA plate was covered with microplate adhesive film and wet paper towel to prevent evaporation and incubated at 65ºC for 10 minutes. Every minute the plate was vortexed for 3 seconds and placed back in the incubator. After virus release the random cDNA synthesis and amplification was performed as described in 4.2.8 followed by Illumina sequencing.

4.2.11 Amplification of p33 gene for Illumina sequencing

Total RNA (2 µl) from Mexican lime leaves infected with New Venture 41/2 (Accession number 11-0051) was extracted according to the manufacturers specifications with the GeneJET Plant RNA Purification Mini Kit (Thermo Scientific, Waltham, Massachusetts) and was combined with 1 µl of 10 µM 1 kb Univ-p33-R (Table B.3, Appendix B) and 7 µl of nuclease free water and incubated for 3 min at 95°C, 1 min on ice. After primer annealing, the RNA was added to 4 µl RT buffer, 0.21 U AMV reverse transcriptase (Roche Diagnostics, Germany), 0.5 U Ribolock RNase Inhibitor (Thermo Scientific, Waltham,

Massachusetts), 1 mM of each dNTP, 10 mM DTT and nuclease free water to a final volume of 20 µl. Reverse transcription occurred at 42°C for 60 min followed by a 5 min incubation at 85°C to stop the reaction. Amplification was done by using 4 µl of synthesized cDNA, 20 µl of the GoTaq Hot Start Green Master Mix (Promega, Madison, Wisconsin), 0.375 µM of each of the p33 forward and reverse primers (Table B.3, Appendix B) and 13 µl of nuclease free water for a final volume of 40 µl. Reaction conditions were as follows: 2 min at 94°C, 35 cycles of 20 sec at 94°C, 45 sec at 65°C and 1 min at 72°C, finishing with an extension for 10 min at 72°C. Amplicons were visualized on a 1 % agarose gel stained with 5 µl/L EtBr and the gel was purified using the NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, Düren, Germany) according to the manufacturer's instructions.

4.2.11 Illumina sequencing and data analysis

Immuno-captured and randomly amplified cDNA as well as the p33 amplicons were sequenced on the Illumina MiSeq platform (Illumina Inc., San Diego, California, USA) at the Agricultural Research Council Bioinformatics platform at Onderstepoort, Pretoria.

Data sets were analyzed on CLC Genomics Workbench 6 (CLC Bio, Aarhus, Denmark) with default parameters except where stated otherwise. Sequence reads were imported as paired end data and trimmed based on quality and sequences of the TruSeq adapters which was used during sequencing. Trimmed reads were mapped to a reference data set containing the 45 complete CTV genomes (for immuno-capture dataset) and p33 gene area (for amplicons dataset) available on GENBANK at that stage. Accession numbers for these genomes are: AB046398.1, AF001623.1, AF260651.1, AY170468.1, AY340974.1, DQ151548.1, DQ272579.1, EU076703.3, EU857538.1, EU937519.1, EU937520.1, EU937521.1, FJ525431.1, FJ525432.1, FJ525433.1, FJ525434.1, FJ525435.1, FJ525436.1, GQ454869.1, GQ454870.1, HM573451.1, JF957196.1, JQ061137.1, JQ798289.1, JQ911663.1, JQ911664.1, JQ965169.1, JX266712.1, JX266713.1, KC262793.1,

KC333868.1, KC517485.1, KC517486.1, KC517487.1, KC517488.1, KC517489.1, KC517490.1, KC517491.1, KC517492.1, KC517493.1, KC517494.1, KC525952.1, NC001661.1, U16304.1 and Y18420.1. Mapping was done with length fraction set to 0.5, similarity fraction set to 0.75 for the immuno-capture template and set to 0.9 for both parameters for the amplicons template. Non-specific match handling was set to 'random'. Furthermore, *de novo* assembly was performed on the immuno-capture dataset and all contigs larger than 500 bp were blasted on the NCBI website (Altschul *et al.*, 1990).

4.3 Results and discussion

4.3.1 Detection of CTV in SAT recipient plants

RT-PCR for CTV was performed on a pooled sample of petioles from all seedlings in a pot (n = 8 to 12). All plants (n = 500) tested negative for CTV. A number of factors may have influenced the outcome of this experiment. The high temperature in the glasshouse may have inhibited the replication of the virus, as it has been found that heat treatment can eliminate CTV from the host (Arif *et al.*, 2005). Numerous other CTV source plants within the glasshouse, growing under the same conditions however retained CTV, but in general they are single plants per pot and are older, larger trees. The seedlings used for the SATs and MTs, were 8-10 plants per pot and still small, making them more susceptible to heat stress due to the divided resources within the pot. The efficiency of SATs described in the literature is, in general, also very low. An efficiency of only 1% has been reported (Broadbent *et al.*, 1996) and was influenced by several factors. Lin *et al.*(2002) found that the specific isolate of CTV plays a role in the transmission efficiency. It was found that certain isolates could not be transmitted by SATs at all and that often severe isolates had a higher SAT efficiency than mild isolates. Although biological indexing was not performed on the New Venture 41/2

source, no severe symptoms could be observed on the original source (Accession 11-0051) in the three years observed.

4.3.2 Optimization of immuno-capture and downstream amplification for use as template for Illumina sequencing

Five different coating antibody concentrations were tested to see at which dilution the threshold of plate binding saturation is obtained. It was expected that with an increase of the antibody concentration, there would be a concomitant increase in the absorbance, showing that more antibodies were bound to the well and hence more virus particles could be captured. Furthermore, it was expected that the absorbance value would increase up to a point where a threshold is reached where after saturation of all binding sites by the antibodies occur and no further increase in absorbance values would be noted. This expected trend was observed with absorbance values having non-significant increases with every higher dilution up to 1:5000 from where it started to decrease (Figure 4.1). A 1:5000 dilution therefore represents the saturation threshold for coating antibodies and was thus used in further immune-capture attempts.

After establishing that the 1:5000 dilution of the coating antibody was the best for capturing the virus particles, it was necessary to determine which buffer was best suited for the release of the virus and by corollary the RNA. The release of the RNA is based on various properties of the solutions, such as pH, and also the heating of the plate and vortexing. Harju *et al.* (2005), conducted an ELISA for *Beet necrotic yellow vein virus*, and for positive wells a virus release buffer was added followed by the heating of the plate to 65°C. The extracts were then used directly for real time PCR. In view of the success of this, this method was therefore also used for virus release of CTV in this study, but with additional buffers also being assessed for their ability to release the viral RNA. These were mainly buffers used by

Papayiannis *et al*. (2010) when determining the most effective virus release buffers after spotting of *Cucurbit yellow stunting disorder virus* infected samples on nylon membrane prior to real time RT-PCR. In the Papayiannis *et al*. (2010) study it was found that buffers with Tween-20, Triton X and Betaine were the most efficient, probably due to the role these additives have to disrupt the binding to the membrane and by enhancing the cDNA synthesis.

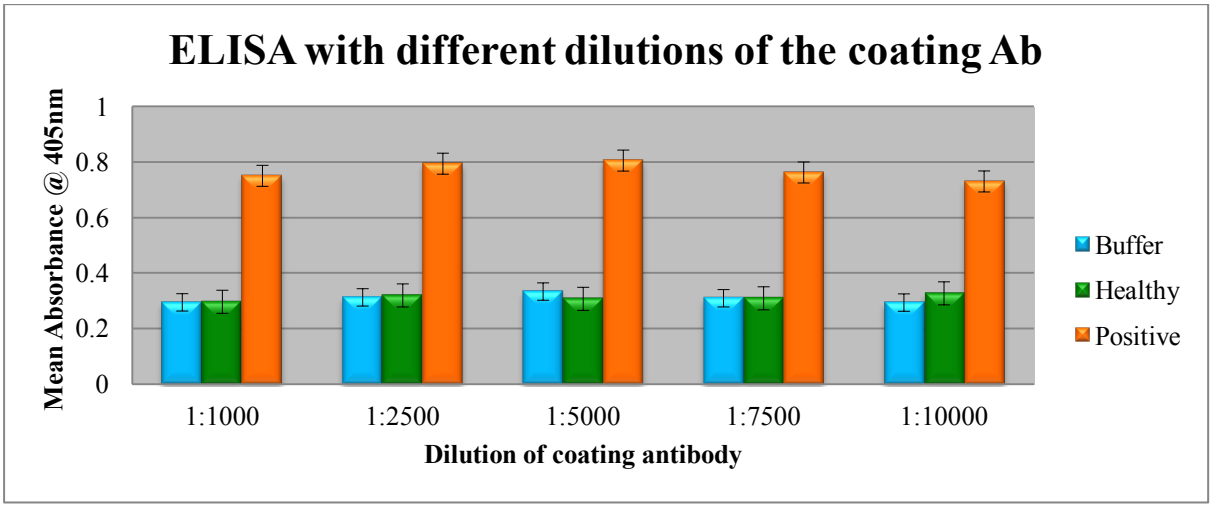

Figure 4.1: Mean ELISA absorbance values at 405mm for 8 replicates for different dilutions of coating antibody used. Error bars indicate standard deviation of the replicates.

The real time RT-PCR in this study showed a low level of florescence for the negative samples after a certain number of cycles, and this may be due to primer or probe degradation late in the cycling procedure. Threshold cycle (C_t) values below 27 were interpreted as positive. The lower the C_t value, the less amplification cycles were needed for the sample to cross the threshold and be deemed positive. This indicates a higher amount of virus material and therefore the buffers producing the lowest C_t values was taken as the most efficient. Amongst the virus release buffers used, the TE based buffer had the lowest C_t value within real time RT-PCR, indicating that either the highest amount of starting material was present or the lowest amount of inhibition occurred during the protocol. Either possibility makes this release buffer the most suitable for downstream processes. Although there was some variation between the replicates, the buffer with the lowest average C_t value was buffer 7

which contained TE + 0.5% Tween-20 (Table 4.2). It was therefore decided to use TE buffer with added Tween-20 as a virus release buffer.

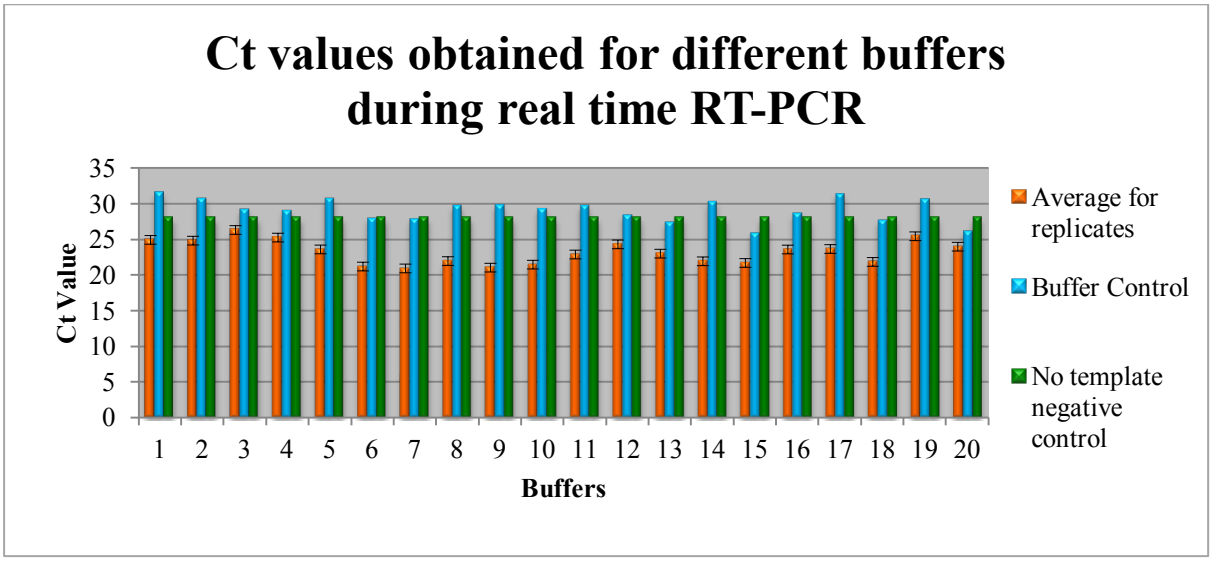

Figure 4.2: Average (n =2) real time RT-PCR results for testing of virus release buffers using a known positive as well as a buffer control and no template negative control. Error bars indicate standard deviation between C_t values of samples. Buffer 1: GES (0.1 m glycine pH: 9, 50 mm NaCl, 1 mm EDTA), Buffer 2: GES + 0.5% Tween-20, Buffer 3: GES + 0.5% Triton-X, Buffer 4: GES + 0.5% DMSO, Buffer 5: GES + Betaine (0.25 g/ml), Buffer 6: TE, Buffer 7: TE + 0.5% Tween-20, Buffer 8: TE + 0.5% Triton-X, Buffer 9: TE + 0.5% DMSO, Buffer 10: TE + Betaine (0.25 g/ml), Buffer 11: Type 2 analytical-grade water, Buffer 12: Type 2 analytical-grade water + 0.5% Tween-20, Buffer 13: Type 2 analytical-grade water + 0.5% Triton-X, Buffer 14: Type 2 analytical-grade water + 0.5% DMSO, Buffer 15: Type 2 analytical-grade water + Betaine (0.25 g/ml), Buffer 16: GES + 0.5% Tween-20 + Betaine (0.25 g/ml), Buffer 17: GES + 0.5% Triton-X + Betaine (0.25 g/ml), Buffer 18: GES + 0.5% Triton-X + 0.5% Tween-20 + Betaine (0.25 g/ml), Buffer 19: 10mM Trizma HCI, 1.0% Triton X-l00, Buffer 20: 50mM Sodium Acetate pH 5.6 (Acetic acid used for pH adjust).

These results showed the most effective options for conducting the immuno-capture step to enrich for CTV RNA and were therefore used on New Venture 41/2 (Accession number 11-0051) with a view to doing Illumina sequencing. The RNA extracts obtained were subjected to reverse transcription using 11 CTV specific degenerate primers, designed as described in Chapter 3 for the amplification of the entire genome, which was conducted prior to this study. Possibly due to the secondary structure of the RNA, a number of the PCRs failed and only half of the genome could be amplified. In the application intended in this chapter these primers were used in conjunction with random hexamers to create cDNA of the RNA genome as it is postulated that the addition of many different primers when the RNA secondary

structure is relaxed by heating to 95°C would allow the binding of the primers over the entire genome prior to reverse transcription. For the current application it was not necessary to obtain full length cDNA transcripts, but many different lengths cDNA spanning the entire genome would suffice.

The cDNA fragments obtained were then subjected to a PCR including the same reverse primers used in the reverse transcription, but in addition also the complementary forward counterparts along with random hexamers to randomly amplify the cDNA. Exponential amplification of fragments would not necessarily occur; therefore many cycles are included to maximize the amount of DNA obtained. When this protocol was tested on a total RNA extract, a bright smear could be seen when the product was run on an agarose gel. When applying the protocol to a CTV RNA enriched immuno-capture extract, a very light smear was obtained with some distinct bands (Figure 4.3). The lighter smear is expected since many plant RNA components within total RNA extracts would have been removed following CTV specific immuno-capture.

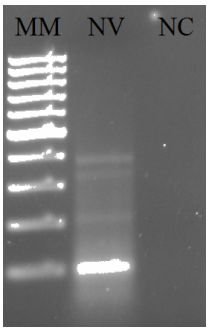

Figure 4.3: Agarose gel showing random amplification technique on New Venture 41/2 immuno-captured RNA, showing Bioline Hyper ladder IV molecular marker (MM), New Venture 41/2 (NV), and the negative control (NC).

To confirm successful cDNA synthesis of the whole CTV genome, prior to conducting costly Illumina sequencing, several CTV specific PCRs were performed to test whether cDNA to different regions of the genome were generated. It was not expected to get a positive reaction for all primer pairs, since the area between any two primer binding sites may not be available on a single cDNA transcript. Nine primer sets were used (Table B.5, Appendix B), including;

a) the generic CTV primer set which targets the 5' end, amplifying a fragment within ORF 1a, b) 3' CTV VT primers targeting the 3' end with part of the p23 gene, c) the PM50 and PM51 primers that amplifies almost the same area as the 3' CTV VT primer set, and d) the Univ-p33 primer set amplifying ORF 2 located in the center of the genome. Furthermore, five genotype specific primer pairs designed by Stewart (2006) were used. These primers amplify regions within the 5' half of the genome. Although these primers are VT genotype specific, we had previously established that New Venture 41/2 contained at least a VT 1a fragment and the VT p23 gene (Lubbe, *unpublished results*). The results for these reactions are shown in Figure 4.4. The generic primer set as well as the 3'CTV VT primer set yielded the correct sized fragments and confirmed that, at least, these areas of the 3' and 5' end of the CTV genome were present. The PM50 and PM51 primers could not amplify the fragment from the DNA although the target is very similar to the target for the 3'CTV VT primers. This may be an indication as to how variable the amplified fragments are and that one will not necessarily obtain amplicons for all primer sets, as even though these areas may have all been converted to cDNA, they may be present on different fragments. The p33 gene could also not be amplified, probably due to its size (1000 bp), resulting in a high probability that the expected amplicon may occur over various fragments of cDNA. Three out of the six fragments could be amplified from the New Venture 41/2 genome using the VT specific primers. These were the fragments amplified by the VT2, VT3 and VT5 primer sets. At this stage it was decided that enough of the CTV genome could be identified to justify subjecting the sample to Illumina sequencing.

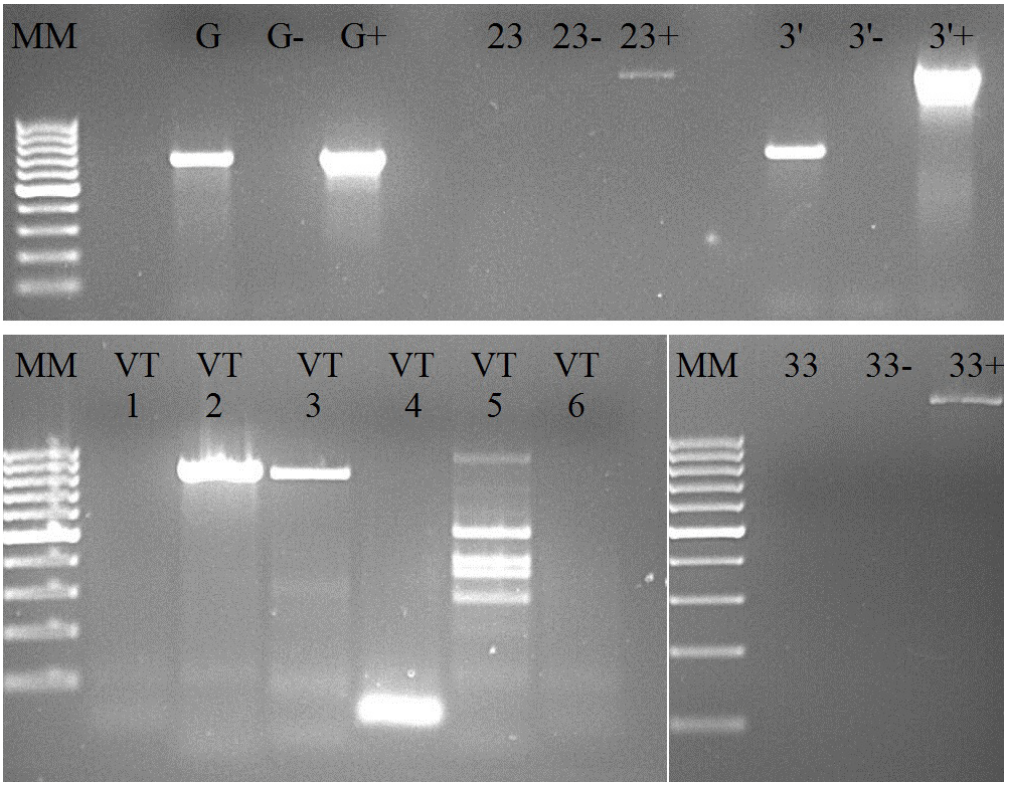

Figure 4.4: Agarose gel photographs showing amplicons obtained from generic primer set (G), PM51 and PM51 primer set (23), 3' CTV VT primer set (3'), VT genotype marker primer sets (VT1-6) and Univ-p33 primer set (33), including positive (+) and (-) controls for G, 23, 3' and 33 primer sets and Thermo Scientific O'GeneRuler 100 bp DNA ladder (MM). Positive control used was the CTVΔ9GFP plasmid (Folimonova, F) causing the large size of positive controls for 23 and 3', since the GFP protein is also amplified.

Reverse transcription and random PCR using only random hexamers were also attempted, but this yielded too little amplification products to sequence. This may be due to the fact that minimal, if any, exponential amplification is possible with only random primers.

The technique used by Zablocki and Pietersen (2014) based on an article by Roossinck *et al.* (2010) could not be applied in this instance since the technique only allows for amplification of dsRNA while the immuno-capture rather specifically enriches for genomic ssRNA.

Another approach, successfully used by several other authors to amplify small amounts of RNA (Chenchik *et al.*, 1996; Edwards *et al.*, 1991; Reddy *et al.*, 2002; Troutt *et al.*, 1992; Zhang and Chiang, 1996) was therefore also assessed. This protocol relies on the production of cDNA from RNA with either specific primers or an oligo-dT primer (for poly(A) tailed RNA) followed by the ligation of an adapter to the cDNA. This is followed by amplification of the fragment between the primer binding area used for reverse transcription and the

annealed oligomer. This is difficult with CTV as it does not have a poly(A) tail to allow the reverse transcription of the RNA with an oligo-dT primer, therefore a specific primer directed at a sequence conserved amongst CTV genotypes near the 5' end of the genome would have to be used in order to not introduce bias in the amplification of any genotypes. Due to the secondary structure of the CTV RNA it is difficult to produce even a 2000 bp amplicon from the genome, and producing a full length cDNA transcript of CTV has yet to be done successfully in our laboratory. To overcome these difficulties, the RNA was therefore transcribed with the mixture of reverse primers described previously (Table B.4, Appendix B) as well as random hexamers to produce single stranded cDNA that could then be used for ligation (described under 4.2.8).

Ligation of both a 3' adapter and 5' adapter to the different cDNA fragments was attempted so that amplification could be done using primers directed at these adapters and thereby avoiding the primer bias which would have occurred using CTV specific primers. These would unavoidably have been directed at sequences with some sequence variability amongst CTV genotypes. For the protocol to work, the 3'adapter must be 5' phosphorylated and have a 3' C3 spacer to allow ligation to the 3'end of the cDNA. The introduction of a C3 spacer prevents the ligation of the adapter in the incorrect orientation, self-ligation or self-circularization. The 5' adapter sequence does not require any modification as it already contains a 3' OH group to allow ligation to the 5'end of the cDNA but lacks the 5' P, therefore preventing ligation of the adapter in the incorrect orientation, self-ligation or self-circularization. The 5' adapter also functions directly as a primer in the PCR needing only another primer, complementary to the 3' adapter for successful amplification. These oligomers are described in Table B.6, Appendix B. A few different variations on the protocol were attempted, all of which were unsuccessful, but described in the interest of completeness,

and in order for future researchers on CTV to expand on or modify in order to achieve success.

1) The first attempt entailed a two-step ligation followed by amplification, but with purification steps after each ligation to get rid of unligated oligos. The ligation was done using T4 RNA ligase (Thermo Scientific, Waltham, Massachusetts) with an adapter to template ratio of 6:1, using 1x of the supplied buffer and 10 U of the ligase in a 10 µl reaction. Between ligations the product was purified using the Wizard SV Clean-up System (Promega, Madison, Wisconsin) according to manufacturer's instructions and between steps the concentration of the ssDNA was determined using a NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, Massachusetts). This was followed by amplification using the 3' A primer and 5' adapter at a concentration of 0 3 7 5µM, 2 µl cDNA, 1 x GoTaq Ho t Start Green Master Mix (Promega, Madison, Wisconsin) and PCR grade water to a final volume of 20 µl. The only product observed on an agarose gel after amplification was primer dimers, and small products that could be adaptors ligated to each other.
2) In a variation of the methodology described in (1), an extra purification step was incorporated after the 5' adapter ligation. The only product observed on an agarose gel after amplification was primer dimers, and small products that may be adaptors ligated to each other.
3) To test whether too much product might have been lost during purification, the entire protocol was repeated without any purification steps, but this still only resulted primer dimers, and small products that could be adaptors ligated to each other.
4) The reaction was also repeated by adding 1 µg BSA to prevent the adhesion of reagents to tube walls, to stabilize the enzymes during temperature cycles and to increase the yield of the reaction through binding to inhibitors (Thermo-Fisher-

Scientific-Inc, 2012). Even with the addition of BSA only primer dimers and small products were produced.

5) To test if the process will be more efficient with shorter pieces of cDNA, a DNase I digestion was performed using 0.0002 U of the DNase I (Thermo Scientific, Waltham, Massachusetts) (diluted in 10x reaction buffer with $MgCl_2$) and 8.5 µl cDNA in a total reaction volume of 10 µl. The reaction was stopped by adding 1 µl 50 M EDTA. The reaction was purified using the NucleoSpin® PCR Clean-Up system (MACHEREY-NAGEL, Düren, Germany). The ligation was tested by doing the 3' and 5' ligation separately as well as simultaneously. Again only very small product smears could be obtained.
6) T4 RNA ligase reactions are very inefficient and it has been found that the addition of PEG can increase the efficiency of the reaction by 67% (Tessier *et al.*, 1986), therefore 25% PEG 8000 and 0.1 mg/µl BSA was added to the ligation reaction using 2x reaction buffer. This still resulted in only very small product smears.

All the ligation reactions mentioned were tested at room temperature, 22°C and 37°C, and reaction times that were tested were 4 hours, 16 hours and 24 hours.

The failure of the reaction could be due to the inefficiency of the enzyme, the length of the products that was used as template and the secondary structure of the cDNA. Since single stranded DNA was used, fragments might bind to each other, preventing amplification after ligation of adapters.

4.3.3 Illumina sequencing of the original New Venture 41/2 source following particle enrichment by immuno-capture

Just over 4 million reads were obtained with Illumina sequencing. When doing reference mapping, only a very small percentage (1.23%) of the reads mapped back to the CTV

reference genomes. A much higher value was expected since the immuno-capture was used for the purpose of enriching for CTV. It remains possible however many other RNAs could have been present during the random cDNA synthesis and amplification since ELISA is quite a crude technique. Some of the CTV RNA might also have been damaged or lost during the process of virus release by heating. Although we could detect CTV cDNA within the sample, it might still have been present as very small amounts. The real time RT-PCR with which the buffers were tested is a very sensitive method and it is possible that the amount of CTV RNA after release was insufficient for downstream processes and sequencing, while still being detectable with real time RT-PCR.

While the CTV specific reads constituted only a small portion of the total reads obtained, one could still use the data to make valuable conclusions. From the mapping data it could be seen that this is a multiple genotype source (Figure 4.5).

The majority of reads mapped to the Kpg3 strain (accession HM573451) which groups within the VT genotype. Furthermore, a considerable number of reads mapped back to strains from the Resistance Breaking (RB) genotype, and some also mapped to the T3 isolate. Illumina results in Chapter 3 demonstrated that for relative homogenous sources the consensus length was high for the single strain to which most reads mapped. Conversely in this case we see that for many of the reference sequences the consensus lengths are quite high, confirming the high heterogeneity of this source. It is clear that it was not just small fragments of other genomes that was being sequenced, which may be suggestive of recombination events within a single genotype, but rather a larger region of a second complete genome present being sequenced. Visualization of the distribution of reads mapping to different genomes can be seen in Figure 4.6.

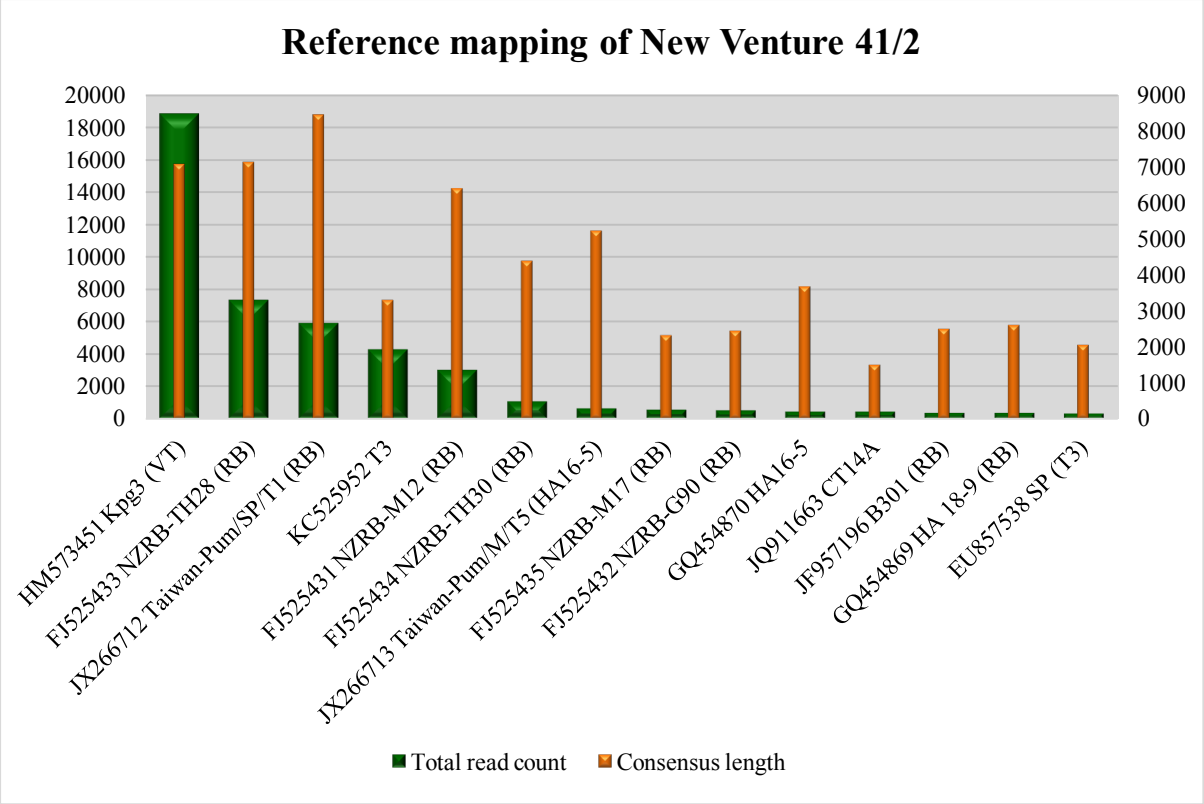

Figure 4.5: Reference mapping of a CTV New Venture 41/2 source, enriched for virus particles by immuno-capture template sequence reads against CTV genomes, showing total number of mapped reads and consensus length of the mapped reads in base pairs.

After *de novo* assembly, contigs greater than 300 bp were subjected to BLAST against the GenBank sequence database. Sixteen of the 1217 contigs returned as CTV related, ranging in size from 300 bp to 598 bp. Seven of the CTV related contigs, ranging from 312 bp to 534 bp were most closely related to the Kpg3 strain (HM573451.1) while one other contig (598 bp) was most similar to the FS701-VT (KC517494.1), both being part of the VT genotype. There were five contigs that were similar to strains within the RB genotype, three being similar to NZRB-TH28 (FJ525433.1) and Taiwan-Pum/SP/T1 (JX266712.1) and NZRB-M12 (FJ525431.1), respectfully. All these contigs range from 300 bp to 335 bp. Two other contigs of about the same size were most similar to the Taiwan-Pum/M/T5 (JX266713.1) strain which forms part of the HA16-5 genotype. The last contigs of 508 bp were most similar to the CT14A (JQ911663.1) isolate, which belongs to the B165 genotype.

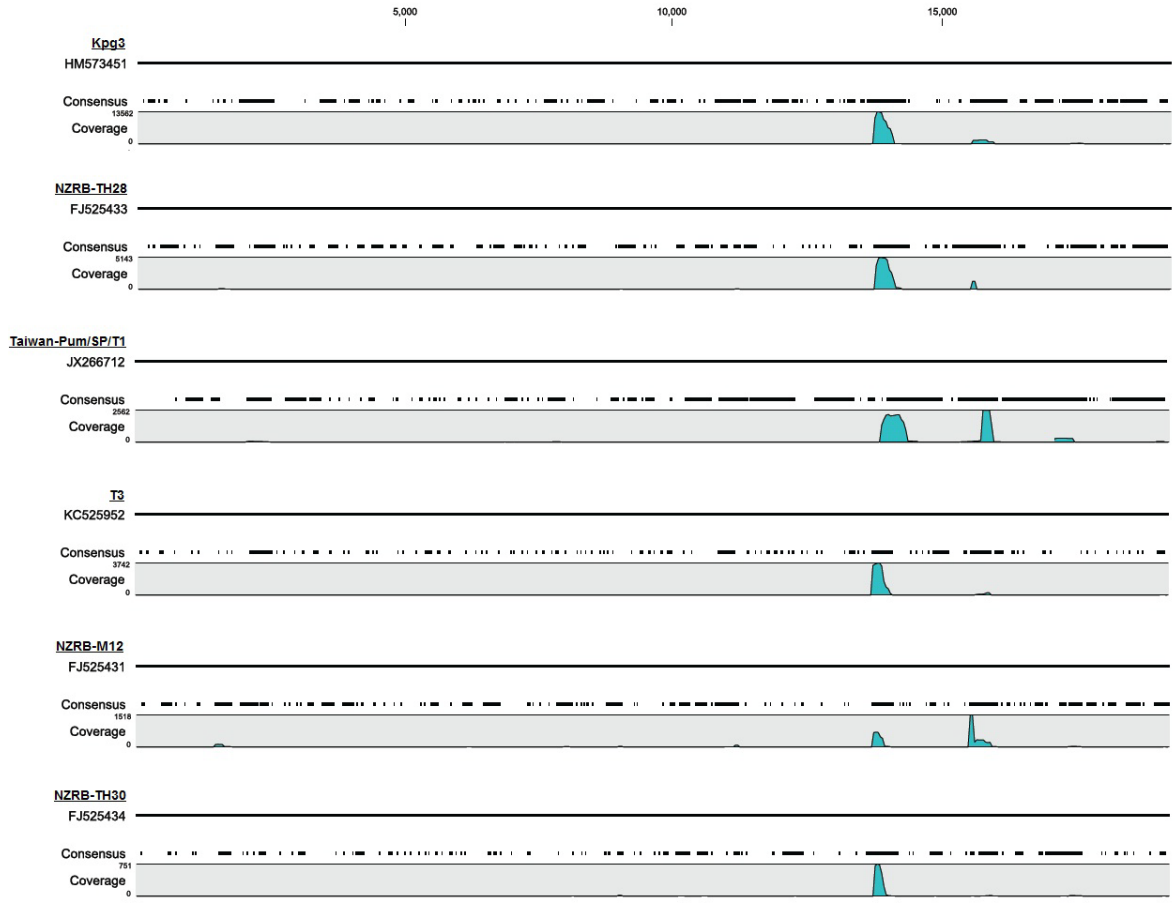

Figure 4.6: Mapping of New Venture 41/2 immuno-capture template sequence reads to reference CTV genomes (line next to GenBank accession number), showing areas where a consensus sequence can be created (interrupted line under reference genome) and the converge of these areas (graph under consensus sequence).

4.3.4 Genotype detection for New Venture 41/2

Genotype detection in the New Venture 41/2 (11-0051) source revealed that it consists of the T36, VT, RB, B165 and HA16-5 genotypes. The published T36 primers also amplify RB isolates and hence the positive result obtained with these primers might be due to the presence of either T36 primers or RB, and was reassessed by Illumina sequencing.

4.3.5 Illumina results for sequencing of p33 gene from New Venture 41/2

During reference mapping to the reference strains, 185 776 (25.8%) of the total of 719 490 reads mapped to the CTV references. This is higher than all other methods used so far, but still low when considering that these were amplicons being sequenced which were gel purified to eliminate all unspecific amplification.

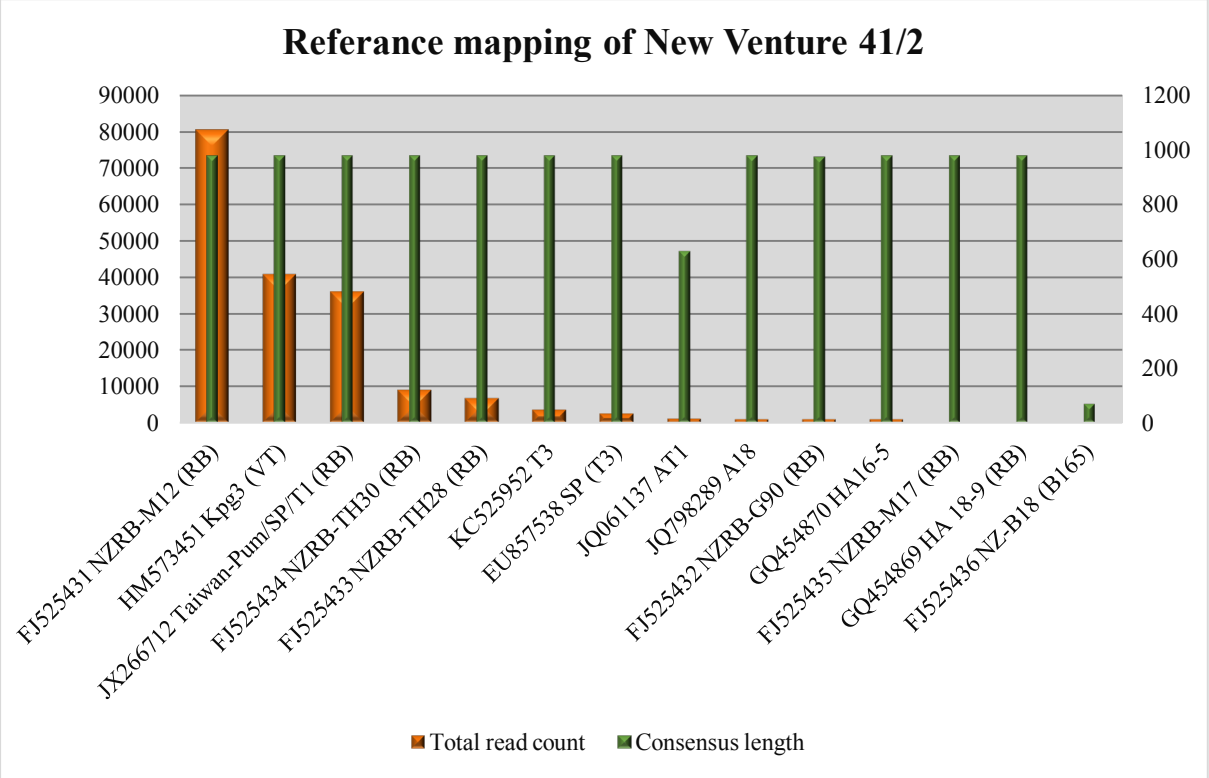

Figure 4.7: Reference mapping of New Venture 41/2 amplicon template sequence reads against the CTV p33 area, showing total number of mapped reads (primary axis) and consensus length of the mapped reads (secondary axis) in base pairs.

The majority of CTV-specific reads (43.2%) mapped to a strain within the RB genotype, followed by reads mapping to the VT genotype (22%) and then to another isolate within the RB genotype (19.4%). Other strains to which small numbers of reads mapped which cover the entire 977 bp of the reference length, belong to the RB, T3, HA 16-5 and A18 genotypes. Reads mapping to less than the entire expected 977 bp of references were not considered indicative of the presence of that genotype and are discarded from further analysis.

4.4 Conclusion

Unfortunately the SATs and mechanical transmissions were unsuccessful possibly due to the low efficiency of the technique or the possibility that the viruses present are inherently poorly transmitted by aphids. Therefore it may be necessary to repeat this with even more replicates. In a transmission study done by Lin *et al*. (2002), of the 2120 SATs done, only 31 tested positive, showing that a large number of transfers is needed for transmission. This is however a very labor intensive and time consuming technique, making numbers like that improbable for this study. In view of the inability to obtain SAT isolates of New Venture 41/2, the original population was characterized to determine if it consisted of a single or multiple genotypes.

Although different methods used to determine the genotypic composition of the New Venture 41/2 source provided different results, it is very clear that this is a mixed source containing multiple genotypes. The genotype specific primers which focus on the 5' end of the genome showed this source contains VT, RB, B165, HA16-5 and possibly T36 genotypes as part of its population. The presence of VT, RB, B165 and HA16-5 could be confirmed with Illumina sequencing. This confirmation was clearer for the VT and RB genotypes than B165 and HA16-5. Reference mapping only indicated VT and RB to be dominant within the source with only a few of reads mapping to HA16-5 (2% for IC template and 0.5% for p33 amplicon) although the full consensus sequence was covered in the case of the p33 amplicon. A *de novo* contig of the immune-capture sequence data blasted to an isolate that is most closely related to the B165 genotype isolates. Although some reads mapped to T36, the read count and consensus length was very low for this genotype. If not for the RT-PCR results it would not have been considered to be present within the source. The positive reaction of the T36 primers is most probably due to the ability of the T36 primers to also amplify the RB strains. The establishment of a threshold regarding the presence/absence of a genotype

remains vexing as reads from the Illumina data mapped back to every one of the reference genomes, but PCR results only confirm the presence of 4 different genotypes. This may be explained by the fact some parts of the genome are conserved amongst strains (Mawassi *et al.*, 1996) to the extent that the mapping of small fragments occurs to all of these conserved regions. The BLAST results from the contigs obtained from the immuno-capture enriched template with *de novo* assembly also confirmed the presence of RB and VT, in addition to the HA16-5 genotype, and a contig similar to the CT14A strain which forms part of the B165 genotype. The Illumina data obtained when using amplicons as template showed a complete consensus length for reads mapping to the T3 and A18 genotypes. The presence of T3 could not be confirmed with the immuno-capture enriched genomic RNA data or the RT-PCR. The A18 genotype could be seen in both Illumina datasets, but it was not possible to confirm its presence with RT-PCR as there is no primers specific to this genotype available yet. In essence, the data generated from the three different methods do confirm the presence of RB, VT and B165 within the source, but minor strains are not consistently detected by all the methods used. This illustrates the different results that can be obtained when using different methods for analysis and shows again how complicated the analysis of CTV genotypes are.

Immuno-capture enriched genomic RNA template did not yield the high quality results expected as the number of CTV specific reads received was still only a small component of total reads obtained. This method of enrichment is therefore not recommended for future use. Although the immuno-capture itself is a very efficient way of enriching for the virus, the inefficient conversion of the RNA to DNA and the amplification of the DNA without specific primers may be the problem. It might be more valuable in systems where the virus being sequenced has a poly(A) tail that can be used for cDNA synthesis, or is not as diverse and would allow the use of specific primers without introducing bias. When comparing the results to those obtained from the Zablocki and Pietersen (2014) study, the dsRNA extraction

followed by amplification based on the Roossinck *et al.* (2010) study, it is found that the immuno-capture still produced a lower amount of CTV data. It is also unsatisfactory that a larger *de novo* sequence could not have been assembled.

As expected using amplicons as template for Illumina sequencing produced more reads mapping to CTV than with the immuno-capture particle enrichment template. Unfortunately however, amplicon templates only allow conclusions limited to the specific gene fragment amplified, and no information on whether recombination took place in any other place of the virus genome and hence if the gene sequenced is representative of the entire genome. While it is not known whether the p33 gene is the only gene playing a role in cross-protection, it does appear as though it is responsible for the genotype specificity of super-infection exclusion (Folimonova, 2012). This makes it particularly important to ascertain the variability with regards this gene region of circulating CTV strains with a view to identifying potential cross-protection sources, as is the case in this study.

4.5 References

Albiach-Marti, M.R., da Graça, J.V., van Vuuren, S.P., Guerri, J., Cambra, M., Laigret, F., and Moreno, P. 1996. The Effects of Different Hosts and Natural Disease Pressure on Molecular Profiles of Mild Isolates of *Citrus tristeza virus* (CTV). In: *Proceedings of the 13th Conference of the International Organisation of Citrus Virologists. (Graça, J.V., Moremo, P., and Yokomi, R.K.)* pp. 147-153. *IOCV*. Riverside, California.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215(3), pp. 403-410.

Arif, M., Ibrahim, M., Ahmad, A., and Hassan, S. 2005. Elimination of Citrus Tristeza Closterovirus from Citrus Bud-Wood through Thermotherapy. *Pakistan Journal of Botany* 37(2), pp. 423.

Bertolini, E., Moreno, A., Capote, N., Olmos, A., de Luis, A., Vidal, E., Pérez-Panadés, J., and Cambra, M. 2008. Quantitative Detection of *Citrus tristeza virus* in Plant Tissues and Single Aphids by Real-Time RT-PCR. *European Journal of Plant Pathology* 120(2), pp. 177-188.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014. Progress Report: Cross-Protection of Star Ruby Using Beltsville Sub-Isolates of Nartia Mild Strain for the Orange River Valley, pp. 24. Citrus Research International, Nelspruit, South Africa.

Broadbent, P., Brlansky, R.H., and Indsto, J. 1996. Biological Characterization of Australian Isolates of *Citrus tristeza virus* and Separation of Subisolates by Single Aphid Transmission. *Plant Disease* 80(3), pp. 329-333.

Chenchik, A., Diachenko, L., Moqadam, F., Tarabykin, V., Lukyanov, S., and Siebert, P. 1996. Full-Length cDNA Cloning and Determination of Mrna 5'and 3'ends by Amplification of Adaptor-Ligated cDNA. *BioTechniques* 21(3), pp. 526-535.

Edwards, J.B.D.M., Delort, J., and Mallet, J. 1991. Oligodeoxyribonucleotide Ligation to Single-Stranded Cdnas: A New Tool for Cloning 5' Ends of mRNAs and for Constructing cDNA Libraries by in Vitro Amplification. *Nucleic Acids Research* 19(19), pp. 5227-5232.

Folimonova, S.Y. 2012. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. *Journal of Virology*. 10.1128/jvi.00310-12

Folimonova, S.Y., Robertson, C.J., Shilts, T., Folimonov, A.S., Hilf, M.E., Garnsey, S.M., and Dawson, W.O. 2010. Infection with Strains of *Citrus tristeza virus* Does Not Exclude Superinfection by Other Strains of the Virus. *The Journal of Virology* 84(3), pp. 1314-1325.

Garnsey, S.M., Gonsalves, D., and Purcifull, D.E. 1977. Mechanical Transmission of *Citrus tristeza virus*. *Phytopathology* 67(8), pp. 965-968.

Harju, V.A., Skelton, A., Clover, G.R.G., Ratti, C., Boonham, N., Henry, C.M., and Mumford, R.A. 2005. The Use of Real-Time RT-PCR (Taqman®) and Post-Elisa Virus Release for the Detection of *Beet Necrotic Yellow Vein Virus* Types Containing RNA 5 and Its Comparison with Conventional RT-PCR. *Journal of Virological Methods* 123(1), pp. 73-80.

Hermoso, D.A., Ballester-Olmos, J.A., Serra, P.J., and Fuertes, C. 1988. Differences in Transmission Efficiency of *Citrus tristeza virus* by *Aphis Gossypii* Using Sweet Orange, Mandarin or Lemon Trees as Donor or Receptor Hosts Plants. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Timmer, L.W., Garnsey, S.M., and Navarro, L.)* pp. 62-64. *IOCV.* Riverside, California.

Huang, Z., Rundell, P.A., Guan, X., and Powell, C.A. 2005. Evaluation of the Transmission of Different Field Sources of *Citrus tristeza virus* and the Separation of Different Genotypes by Single Brown Citrus Aphids. *Hortscience* 40(3), pp. 687-690.

Lin, Y., Brlansky, R.H., and Powell, C.A. 2002. Inefficient Transmission of *Citrus tristeza virus* from Grapefruit by Single Brown Citrus Aphids. *Hortscience* 37(6), pp. 936-939.

Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G., and Bar-Joseph, M. 1996. Unusual Sequence Relationships between Two Isolates of *Citrus tristeza virus*. *Journal of General Virology* 77(9), pp. 2359-2364.

Papayiannis, L.C., Hunter, S.C., Iacovides, T., and Brown, J.K. 2010. Detection of *Cucurbit Yellow Stunting Disorder Virus* in Cucurbit Leaves Using Sap Extracts and Real-Time Taqman® Reverse Transcription (Rt) Polymerase Chain Reaction (PCR). *Journal of Phytopathology* 158(7-8), pp. 487-495.

Read, D.A., and Pietersen, G. 2015. Genotypic Diversity of *Citrus tristeza virus* within Red Grapefruit, in a Field Trial Site in South Africa. *European Journal of Plant Pathology*, pp. 1-15.

Reddy, M.K., Nair, S., and Sopory, S.K. 2002. Global Amplification of cDNA from Limiting Amounts of Tissue. *Molecular Biotechnology* 22(3), pp. 223-230.

Roossinck, M.J., Saha, P., Wiley, G.B., Quan, J., White, J.D., Lai, H., ChavarrÍA, F., Shen, G., and Roe, B.A. 2010. Ecogenomics: Using Massively Parallel Pyrosequencing to Understand Virus Ecology. *Molecular Ecology* 19, pp. 81-88.

Roy, A., and Brlansky, R.H. 2010. Genome Analysis of an Orange Stem Pitting *Citrus tristeza virus* Isolate Reveals a Novel Recombinant Genotype. *Virus Research* 151(2), pp. 118-130.

Rubio, L., Ayllon, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., and Falk, B.W. 2001. Genetic Variation of *Citrus tristeza virus* Isolates from California

and Spain: Evidence for Mixed Infections and Recombination. *The Journal of Virology* 75(17), pp. 8054-8062.

Stewart, K.A. 2006. Strain Differentiation of *Citrus tristeza virus* Isolates from South Africa by PCR and Microarray. Dissertation submitted to the Faculty of Natural and Agricultural Sciences Department of Microbiology and Plant Pathology in partial fulfillment of the requirements for the degree Magister Scientiae (MSc) Microbiology. University of Pretoria.

Tessier, D.C., Brousseau, R., and Vernet, T. 1986. Ligation of Single-Stranded Oligodeoxyribonucleotides by T4 RNA Ligase. *Analytical Biochemistry* 158(1), pp. 171-178.

Thermo Fisher Scientific Inc. 2012. Product Information Bovine Serum Albumin (BSA), Molecular Biology Grade. http://www.thermoscientificbio.com/uploadedFiles/Resources/b14-product-information.pdf.

Troutt, A.B., McHeyzer-Williams, M.G., Pulendran, B., and Nossal, G. 1992. Ligation-Anchored PCR: A Simple Amplification Technique with Single-Sided Specificity. *Proceedings of the National Academy of Sciences* 89(20), pp. 9823-9825.

Van Vuuren, S.P., and Van Der Vyver, J.B. 2000. Comparison of South African Pre-Immunizing *Citrus tristeza virus* Isolates with Foreign Isolates in Three Grapefruit Selections. In: *Proceedings of the 14th Conference of the International Organisation of Citrus Virologists. (Da Graça, J.V., Lee, R.F., and Yokomi, R.K.)* pp. 50-56. *IOCV*. Riverside, California.

Van Vuuren, S.P., Van Der Vyver, J.B., and Luttig, M. 2000. Diversity among Sub-Isolates of Cross-Protecting *Citrus tristeza virus* Isolates in South Africa. In: *Proceedings of the 14th Conference of the International Organisation of Citrus Virologists. (Da Graça, J.V., Lee, R.F., and Yokomi, R.K.)* pp. 103-110. *IOCV*. Riverside, California.

Vives, M.C., Rubio, L., Sambade, A., Mirkov, T.E., Moreno, P., and Guerri, J. 2005. Evidence of Multiple Recombination Events between Two RNA Sequence Variants within a *Citrus tristeza virus* Isolate. *Virology* 331(2), pp. 232-237.

Vrijenhoek, R. 1994. DNA Primers for Amplification of Mitochondrial Cytochrome C Oxidase Subunit I from Diverse Metazoan Invertebrates. *Molecular Marine Biology and Biotechnology* 3, pp. 294-299.

Zablocki, O. 2013. Unbiased, Next-Generation Sequencing for the Characterization of *Citrus tristeza virus* Populations. Dissertation submitted to the Faculty of Natural and Agricultural Sciences Department of Microbiology and Plant Pathology in partial fulfillment of the requirements for the degree Magister Scientiae (MSc) Microbiology. University of Pretoria.

Zablocki, O., and Pietersen, G. 2014. Characterization of a Novel *Citrus tristeza virus* Genotype within Three Cross-Protecting Source GFMS12 Sub-Isolates in South Africa by Means of Illumina Sequencing. *Archives of Virology*, pp. 1-7.

Zhang, X.-H., and Chiang, V.L. 1996. Single-Stranded DNA Ligation by T4 RNA Ligase for PCR Cloning of 5 -Noncoding Fragments and Coding Sequence of a Specific Gene. *Nucleic Acids Research* 24(5), pp. 990-991.

Chapter 5: Biological indexing of the B390/3 CTV source and viral population characterization

5.1 Introduction

There are many different approaches to control the spread of *Citrus tristeza virus* (CTV), for example placing affected areas under quarantine, the eradication of infected trees, grafting of scions onto resistant rootstock and the chemical control of vectors (Fulton, 1986; Moreno *et al.*, 2008). Once the virus is common within an area, mild strain cross-protection can be used to prevent severe and economically crippling symptoms of the virus (Moreno *et al.*, 2008).

Mild strain cross-protection was introduced in South Africa during the initiation of the South African Citrus Improvement Program in 1973 (Von Broembsen and Lee, 1988) and probably relies on the principle of super-infection exclusion (Folimonova *et al.*, 2010). This is where the presence of one virus within a plant prevents the secondary infection of that same virus. Hence if all citrus can be inoculated with a strain of the virus that only produces mild symptoms, severe economic losses can be prevented since the plants are protected against the more severe strains. However, in the case of CTV, super-infection exclusion only occurs between strains of the virus from the same genotypes (Folimonova *et al.*, 2010). For example, a VT infected plant is only protected against infection by other VT strains, but a strain of B165 would still be able to infect the plant. To complicate matters even more, more than one genotype of the virus can occur within a host (Albiach-Marti *et al.*, 2000), since multiple vector transmissions can occur in the field. This makes it necessary to pre-immunize citrus with a source of the virus that contains mild strains of multiple genotypes. Furthermore, the severity of the symptoms caused does not just depend on the strain of the virus but also on the host and environmental conditions. A given strain can cause mild symptoms in one host, but more severe symptoms in another (Karasev *et al.*, 1998). The host species and environmental conditions also influence the selection of genotypes within a host (Albiach-Marti *et al.*, 1996) and can lead to a CTV source that only produced mild symptoms under certain circumstances to produce severe symptoms when transmitted to another host or exposed to different

environmental conditions. All these factors influence the success of a cross-protection source and can lead to cross-protection breakdown as was the case with GFMS 12, one of the first mild strains used for cross-protection in South Africa (Van Vuuren, 2002).

CTV has a very complicated nature where mixed infections are common within a plant due to the host being long-lived and the occurrence of repeated infections through aphid transmission, the quasispecies nature of the virus, and the propensity of recombination between strains (Roy and Brlansky, 2010; Rubio *et al.*, 2001; Vives *et al.*, 2005). This makes it very important to characterize viral populations thoroughly as pre-immunizing sources before using them for cross-protection. The method most commonly used to identify pre-immunizing CTV sources is based on the visual selection of CTV infected trees with mild or no symptoms. These CTV sources are considered as candidate pre-immunizing sources and are subjected to biological indexing in greenhouses and ultimately in field trials. If only mild symptoms persist after the field trials, the source is considered suitable for use in pre-immunizing citrus (Albiach-Marti *et al.*, 1996; Zanutto *et al.*, 2013).

The biological characterization of CTV is done by inoculating indicator plants through graft transmission. Due to the complex and diverse nature of symptom expression by CTV, different index plants are used to observe different symptoms. Symptoms apparent on leaves, for example vein clearing, leaf cupping and chlorosis, as well as plant height reduction can be studied on West Indian lime and Mexican lime (Broadbent *et al.*, 1996; Garnsey *et al.*, 2005). Seedling yellows symptoms are apparent on Duncan grapefruit, Eureka lemon and Bittersweet Seville orange (Broadbent *et al.*, 1996). If the stem is peeled away above the point of inoculation at the final reading, sweet orange, West Indian lime, Mexican lime and Duncan grapefruit can be used to observe the degree of stem pitting caused by the isolate (Broadbent *et al.*, 1996; Garnsey *et al.*, 2005). Lastly, quick decline symptoms can be seen when sweet orange is grafted on sour orange rootstocks (Broadbent *et al.*, 1996). Garnsey *et*

al. (1987) proposed a bio-characterization index where symptoms are scored on a scale from 1 – 3, where the index in certain hosts are multiplied with a weight factor according to economical impact. For example, Mexican lime has weight factor of 1, Sweet orange on Sour orange 2, Sour orange 3, Duncan grapefruit 4 and Madam Vinous sweet orange 5.

Biological indexing of CTV sources is very important and is unlikely to be replaced within the foreseeable future, however it should be supported by additional molecular characterization of the virus to ascertain which genotypes occur in the pre-immunizing source and hence that the biological properties are associated with these genotypes. In this study, the B390/3 source was subjected to biological indexing as well as molecular genotyping through RT-PCR and population sequencing. This was to determine; a) whether different hosts affect the viral population, for example by allowing a genotype which may be minor, or not even detected in some hosts, to replicate to significant levels in others, and b) whether the source is pure and therefore suitable for cross-protection.

5.2 Materials and methods

5.2.1 The B390/3 viral source

The B390/3 source was one of the isolates obtained when single aphid transmissions (SATs) were done from the Mouton CTV source (which originated in SA) in Beltsville, USA. The SAT isolates obtained from the Mouton source were imported back to South Africa where they were evaluated for their usefulness as pre-immunizing sources. The B390/3 source has been subjected to field trials where both Marsh and Star Ruby grapefruit were inoculated and monitored for symptom expression (Breytenbach *et al.*, 2014a; b; c). Although two of the three field trials are still in progress, B390/3 is performing very well thus far, and appears to be a good candidate pre-immunizing source (Breytenbach *et al.*, 2014a).

5.2.2 Inoculation of indicator hosts with B390/3

Bark strips were cut from the B390/3 source tree housed in the insect free greenhouse of CRI, Nelspruit, in order to graft-inoculate ten citrus hosts which were obtained through the citrus certification scheme and are therefore free of CTV. These hosts were; Mexican lime, Sour orange, Duncan grapefruit, Madam Vinous sweet orange, Palmer Navel sweet orange on sour orange rootstock, Midknight Valencia orange, Star Ruby grapefruit, Esbal clementine, Eureka lemon and Miho Wase mandarin. The Mexican lime, Sour orange, Duncan grapefruit and Madam Vinous sweet orange were grown from seed and maintained on own roots whereas the remaining hosts were grafted on rough lemon rootstock except for the Palmer Navel sweet orange that was grafted on sour orange rootstock (Breytenbach, *personal communication*). This selection of hosts were chosen based on a combination of the standardized method for evaluation of biological properties proposed by Garnsey *et al.*(1987) and citrus cultivars commonly used in South Africa (Van Vuuren, *personal communication*). For each host, 3 replicates and 3 negative controls were used. These plants were kept under greenhouse conditions for twelve months to evaluate symptom expression. Greenhouse temperatures ranged from 4°C to 40°C, but did reach temperatures of around 50°C on a few instances.

5.2.3 Symptom evaluation

Symptoms were evaluated on a monthly basis. Symptoms evaluated included; growth rate, vein clearing, leaf cupping and chlorosis each month and stem pitting at the final reading. To allow maximal shoot growth, all side branches were removed every month to train trees to a single leader. The main shoot was then measured and growth rate (cm of growth per days passed) was calculated. The vein clearing, leaf cupping, chlorosis and stem pitting was analyzed based on the Garnsey *et al*. (1987) scale of 1 to 3, where 1 is very mild symptoms and 3 is severe. The stem pitting was observed only at the last measurement since the bark needs to be pulled away from the stem to observe the symptoms.

5.2.4 CTV detection

RNA was obtained by doing extractions from leaf petioles and midribs using the GeneJet Plant RNA Purification Kit (Thermo Scientific, Waltham, Masseuses, US) according to manufacturer's instructions 2 µl of RNA, 1 µl of 10 mM Generic R primer (Table B.3, Appendix B) and 7 µl of water were first mixed and incubated at 95°C for 3 min and then left on ice for 1 minute for primer annealing. This 10 µl mix was combined with 1x RT Buffer, 25 U AMV Reverse Transcriptase, 10 U Protector RNase Inhibitor (Roche Diagnostics, Mannheim, Germany), 1 mM of a dNTP mix (KAPA Biosystems, Cape Town, South Africa), 10 mM DTT (Thermo Scientific, Waltham, Masseuses, US), and PCR grade water to a total volume of 20 µl. Reverse transcription was performed at 42°C for 60 min, followed by a 85°C incubation for 5 min to inactivate the enzymes. For the PCR, 2 µl of the cDNA was combined with 1x GoTaq Hot Start Green Master Mix (Promega, Madison, Wisconsin, USA), 0.375 µM of both the forward and reverse generic primer and PCR grade water for a total reaction volume of 20 µl. Reaction conditions were as follows: 94°C for 2 min, followed by 35 cycles of 94°C for 20 sec, 58°C for 30 sec and 72°C for 20 sec. This was finished off with a final extension period of 1 min at 72 °C. Amplification products were visualized by running 10 µl on a 1% agarose gel stained with EtBr.

5.2.5 Genotype detection

Detection of CTV genotypes was based on genotype specific primers used in RT-PCR (Roy *et al.*, 2010). The reaction used is described in 5.2.4, but annealing temperatures for genotype specific primers (Table B.3, Appendix B) varied. Annealing temperatures for primer pairs T36, T30, T3, VT, NZRB1 and NZRB2 was 60°C, B165 was 59°C and HA16-5 was 56°C.

5.2.6 Amplification and sequencing of the p33 gene

Total RNA extract (2µl) from leaves of representative samples from the biological indexing study that was obtained according to manufacturers specifications with the GeneJET Plant RNA Purification Mini Kit (Thermo Scientific, Waltham, Massachusetts) was combined with 1 µl of 10 µM Generic R primer (Table B.3, Appendix B) and 7 µl of nuclease free water and incubated for 3min at 95°C, 1 min on ice. After primer annealing, the RNA was added to 4µl RT buffer, 0.21 U AMV reverse transcriptase (Roche Diagnostics, Germany), 0.5 U Ribolock RNase Inhibitor (Thermo Scientific, Waltham, Massachusetts), 1 mM of each dNTP, 10 mM DTT and nuclease free water to a final volume of 20 µl. Reverse transcription occurred at 42°C for 60 min followed by a 5 min incubation at 85°C to stop the reaction. Amplification was done by using 4 µl of synthesized cDNA, 20µl of the GoTaq Hot Start Green Master Mix (Promega, Madison, Wisconsin), 0.375 µM of each of the Generic forward and reverse primers (Table B.3, Appendix B) and 13 µl of nuclease free water for a final volume of 40 µl. Reaction conditions were as follows: 2 min at 94°C, 35 cycles of 20 sec at 94°C, 45 sec at 65°C and 1 min at 72°C, finishing with an extension for 10 min at 72°C. Amplicons were visualized on a 1 % agarose gel stained with 5 µl/L EtBr and then gel purified using the NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, Düren, Germany) according to manufacturer's instructions. 25 µl of the purified gel product from the original B390/3 source and the Mexican Lime 12-0011 sample was sent for paired end Illumina sequencing on the MiSeq platform (Illumina Inc, San Diego, California, USA) at the Agricultural Research Council Bioinformatics platform at Onderstepoort, Pretoria. Re-amplification of the purified product was done exactly as described for the original p33 PCR, by using 1 µl of purified gel product. The amplified products were purified using Exonuclease and FastAP® (Fermentas, Maryland, USA) according to the manufacturer's instructions to use for Sanger sequencing. The sequencing reaction contained 2.25 µl 5x sequencing buffer (Applied

Biosystems, Foster City, CA, USA), 0.75 µl of 2 µM Univ-p33 forward primer, 1 µl of 2.5x terminator mix v3.1 (Applied Biosystems, Foster City, CA, USA), 4 µl nuclease free molecular grade water and 2 µl of template DNA. The reaction conditions was 94 °C for 1 min for initial denaturation, followed by 30 cycles of 94 °C for 10 sec, 50 °C for 5 sec and 60 °C for 4 min. The sequencing was done using an ABI Prism® 3130XL Genetic Analyser (Applied Biosystems, Foster City, CA, USA) at the African Centre for Gene Technologies, Automated Sequencing facility, Department of Genetics, University of Pretoria, South Africa.

5.2.7 Sanger sequence data analysis

Sequences obtained were opened in Chromas Lite 2.1.1 (Technelysium Pty Ltd, South Brisbane, Australia) to verify base pairs and to trim ends. Sequences were aligned using MAFTT online (Katoh *et al.*, 2002) and the best fit nucleotide model was tested in jModelTest 2.1.3 (Darriba *et al.*, 2012). Bayesian Analysis was performed on MrBayes 3.2.1 (Ronquist *et al.*, 2012). Phylogenetic trees were viewed on FigTree 1.3.1 (Rambaut, University of Edinburg).

5.2.8 Illumina sequence data analysis

Data sets were analyzed on CLC Genomics Workbench 6 (CLC Bio, Aarhus, Denmark) with default parameters except where otherwise stated. Sequence reads were imported as paired end data and trimmed based on quality and sequences of the TruSeq adapters which was used during sequencing. Trimmed reads were mapped to a reference data set containing the p33 region of all 45 complete CTV genomes available on GENBANK at that stage. Accession numbers for these genomes are; AB046398.1, AF001623.1, AF260651.1, AY170468.1, AY340974.1, DQ151548.1, DQ272579.1, EU076703.3, EU857538.1, EU937519.1, EU937520.1, EU937521.1, FJ525431.1, FJ525432.1, FJ525433.1, FJ525434.1,

FJ525435.1, FJ525436.1, GQ454869.1, GQ454870.1, HM573451.1, JF957196.1, JQ061137.1, JQ798289.1, JQ911663.1, JQ911664.1, JQ965169.1, JX266712.1, JX266713.1, KC262793.1, KC333868.1, KC517485.1, KC517486.1, KC517487.1, KC517488.1, KC517489.1, KC517490.1, KC517491.1, KC517492.1, KC517493.1, KC517494.1, KC525952.1, NC001661.1, U16304.1 and Y18420.1. Mapping was done with length fraction set to 0.9, similarity fraction set to 0.9 and non-specific match handling was set to 'ignore'.

5.3 Results and discussion

5.3.1 CTV detection

When the RT-PCR was performed to confirm the success of inoculation of the virus, all healthy control indicator hosts tested negative and all inoculated host plants tested positive. A number of the plants however died due to physiological stress induced by extreme heat when the cooling system failed. These included; one Mexican lime healthy control, one virus inoculated and one healthy control for Duncan grapefruit, two virus inoculated and two healthy controls for Madam Vinous sweet orange, one Miho Wase mandarin healthy control, one virus inoculated Eureka lemon and one Esbal clementine healthy control. Except for Madam Vinous sweet orange, all other host groups still retained at least one healthy control and more than one inoculated plant and the trial could still continue. Although symptoms on Madam Vinous are recorded in results, this host was not used to draw conclusions regarding the mildness of the B390/3 source.

5.3.2 Symptom evaluation

B390/3 did not elicit severe symptoms on any of the hosts evaluated (Figure 5.4). Chlorosis was present in some of the hosts, but these generally were on both the inoculated and the healthy control hosts, and were probably due to environmental factors. Plants showing

symptoms absent in the relevant healthy control included; Mexican lime with very mild chlorosis and leaf cupping (Garnsey value 0.5 – 1) and very mild stem pitting (Garnsey value 0.5 – 1) in two of the three replicates, and Midknight Valencia orange with mild chlorosis (Garnsey value 0.5) although the plants recovered from symptoms after two months. Based on the weighted Garnsey scale, B390/3 has an index value of 1, and hence is considered a mild source. Figure 5.1 and 5.2 shows the leaf cupping and stem pitting seen in Mexican lime and Figure 5.3 shows the summary of symptoms observed in the hosts.

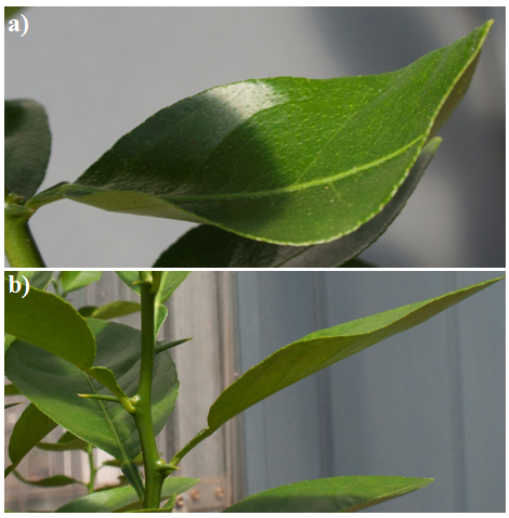

Figure 5.1: Symptoms of B390/3 on indexing plants a) Leaf cupping in Mexican lime 12-1009 b) Leaves of healthy control Mexican lime 12-1014.

No stunting occurred in any of the hosts. Due to the heat stress some of the shoots died and new shoots had to be measured in some plants, although the original shoots could still be measured in some plants. This prevented statistical comparisons of the shoot length between inoculated plants and healthy controls. Although not statistical valuable, the shoot length was still investigated in plants (Figure 5.4 - 5.5).

Figure 5.2: Symptoms of B390/3 on indexing plants a) Mild stem pitting in Mexican lime 12-1009 b) Stem of healthy control Mexican lime 12-1011.

There is no distinct trend of non-infected plants having longer shoots than infected plants. In most cases (Duncan grapefruit, Madam Vinous sweet orange, Palmer Navel sweet orange, Star Ruby grapefruit, Midknight Valencia sweet orange, Eureka lemon and Esbal clementine), the replicates for non-infected and infected plants had varying shoot lengths, with no trend of the non-infected plants having longer shoots than the infected plants or vice versa, but some individuals having longer shoots than others irrespective of being infected or not. The growth rates for all the plants were calculated based the amount of growth in a period of time measured as the increase in length of the shoot divided by the number of days passed (Figure 5.6 – 5.7), and this was used in the statistical analysis. As expected there were growth spurts resulting from new flushes and in most instances these occur simultaneously in infected and non-infected plants, or separated by short periods. When comparing the average growth rate for inoculated plants to healthy controls per cultivar, it was found that there was no significant difference between them. This was calculated based on a ANOVA test of variance in which it was found the *F* value (0.762) was smaller than the *F crit* value (5.12), indicating that the means of growth rate of the infected and healthy plants do not differ significantly. This indicates that B390/3 does not cause stunting in the cultivars tested.

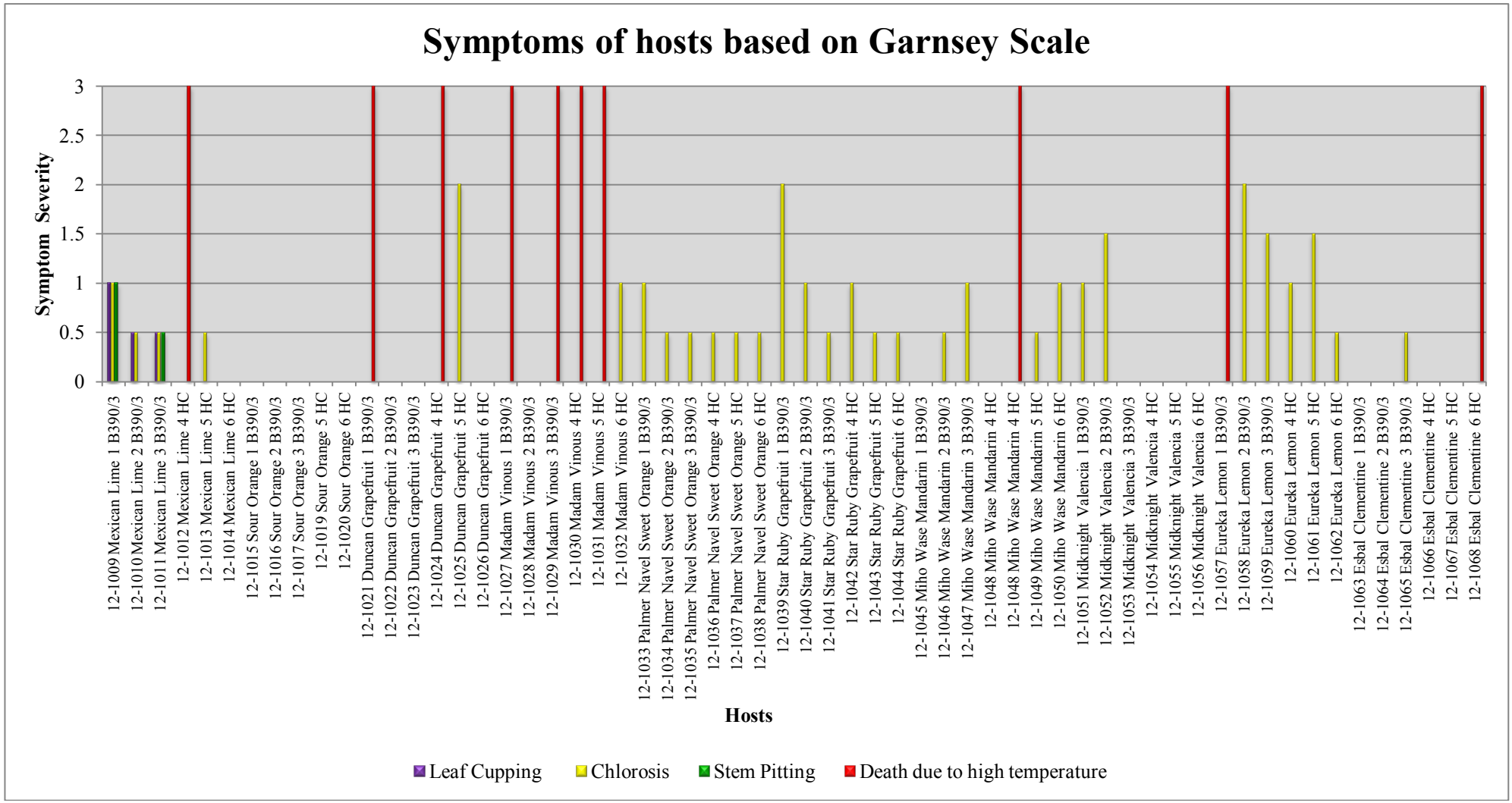

Figure 5.3: Summary of symptoms in hosts in biological indexing trial. Symptoms was evaluated based on the Garnsey scale where 0 is no symptoms and 3 is severe symptoms.

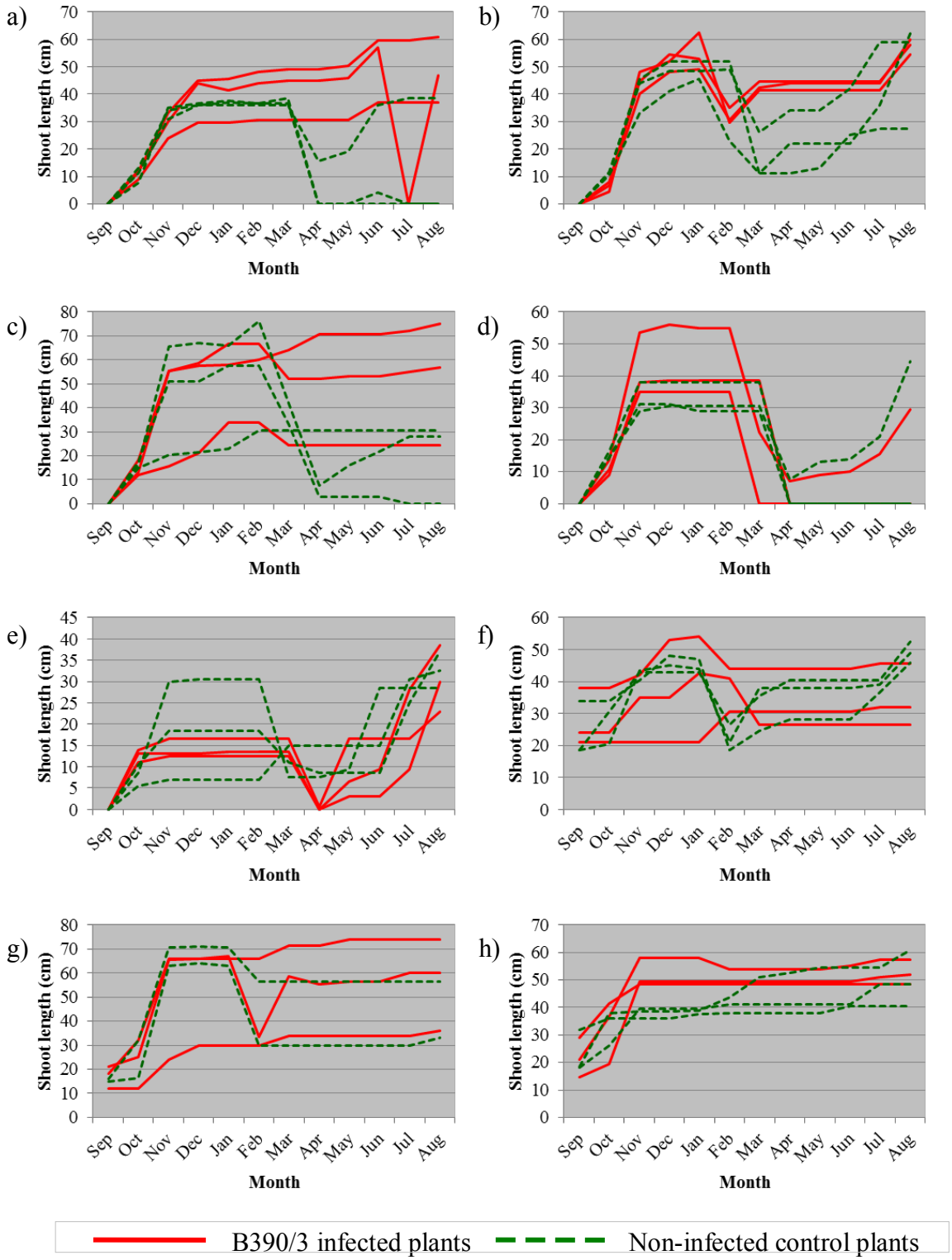

Figure 5.4: Line graphs showing the shoot length of B390/3 infected plants and non-infected plants of different citrus cultivars. A decrease in shoot length is indicative of the death of the measured shoot and the forced measurement of a new shoot. a) Mexican lime b) Sour orange c) Duncan grapefruit d) Madam Vinous sweet orange e) Palmer Navel sweet orange f) Star Ruby grapefruit g) Miho Wase mandarin h) Midknight Valencia sweet orange.

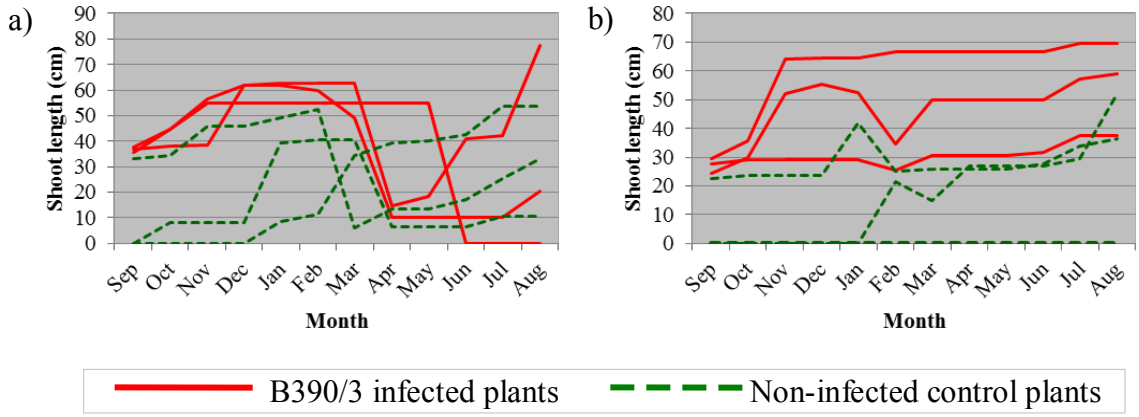

Figure 5.5: Line graphs showing the shoot length of B390/3 infected plants and non-infected plants of different citrus cultivars. A decrease in shoot length is indicative of the death of the measured shoot and the forced measurement of a new shoot. a) Eureka lemon b) Esbal clementine.

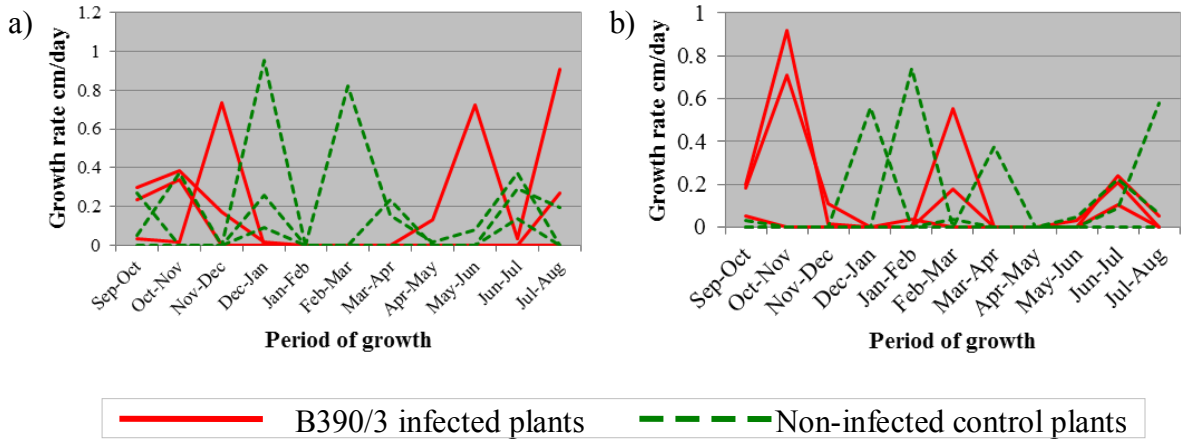

Figure 5.6: Line graphs showing the growth rates of B390/3 infected plants and non-infected plants of different citrus cultivars. a) Eureka lemon b) Esbal clementine.

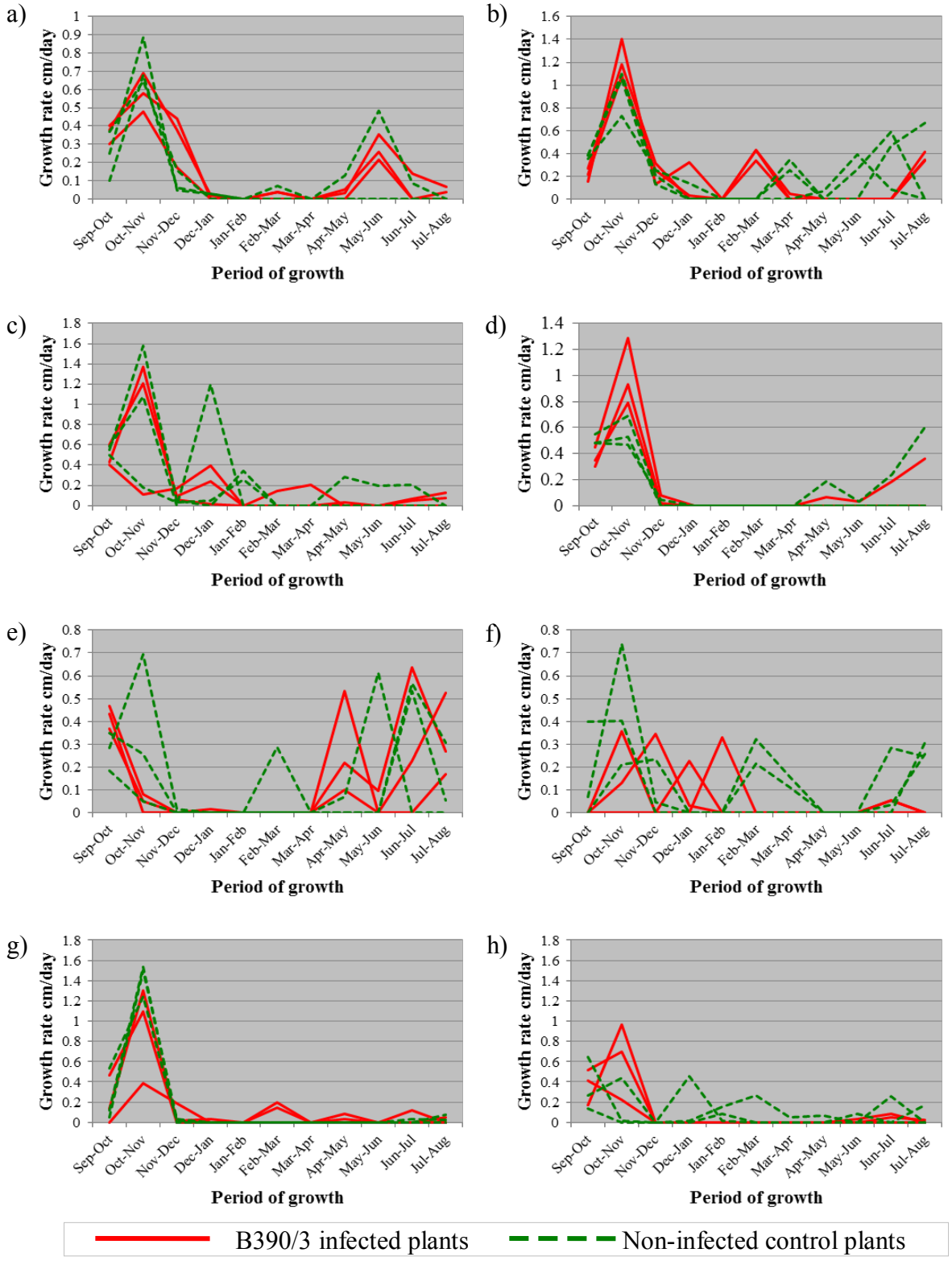

Figure 5.7: Line graphs showing the growth rates of B390/3 infected plants and non-infected plants of different citrus cultivars. a) Mexican lime b) Sour orange c) Duncan grapefruit d) Madam Vinous sweet orange e) Palmer Navel sweet orange f) Star Ruby grapefruit g) Miho Wase mandarin h) Midknight Valencia sweet orange.

5.3.3 Genotype detection with RT-PCR

The genotypes identified in inoculated plants are described in Table 5.1. The Resistance Breaking (RB) genotype was found in all plants and B165 was found in only one Mexican lime host and the original B390/3 source. Although the T36 test yielded amplicons in most of the hosts, it is most probably due to the known cross-reactivity of the T36 primers with the RB genotype. This was resolved by Illumina sequencing.

Table 5.1: Genotypes detected in biological indexing plants for the B390/3 source, as well as the genotypes detected in the original source. Strong, medium or light positive is based on the intensity of the band seen on the agarose gel

Tree	Generic	NZRB2	T36	B165	VT	T30	T3	NZRB1	HA16-5
12-1009 Mexican Lime 1 B390/3									
12-1010 Mexican Lime 2 B390/3									
12-1011 Mexican Lime 3 B390/3									
12-1015 Sour Orange 1 B390/3									
12-1016 Sour Orange 2 B390/3									
12-1017 Sour Orange 3 B390/3									
12-1021 Duncan Grapefruit 1 B390/3									
12-1022 Duncan Grapefruit 2 B390/3									
12-1023 Duncan Grapefruit 3 B390/3									
12-1027 Madam Vinous 1 B390/3									
12-1028 Madam Vinous 2 B390/3									
12-1029 Madam Vinous 3 B390/3									
12-1033 Palmer Navel Sweet Orange 1 B390/3									
12-1034 Palmer Navel Sweet Orange 2 B390/3									
12-1035 Palmer Navel Sweet Orange 3 B390/3									
12-1039 Star Ruby Grapefruit 1 B390/3									
12-1040 Star Ruby Grapefruit 2 B390/3									
12-1041 Star Ruby Grapefruit 3 B390/3									
12-1045 Miho Wase Mandarin 1 B390/3									
12-1046 Miho Wase Mandarin 2 B390/3									
12-1047 Miho Wase Mandarin 3 B390/3									
12-1051 Midknight Valencia 1 B390/3									
12-1052 Midknight Valencia 2 B390/3									
12-1053 Midknight Valencia 3 B390/3									
12-1057 Eureka Lemon 1 B390/3									
12-1058 Eureka Lemon 2 B390/3									
12-1059 Eureka Lemon 3 B390/3									
12-1063 Esbal Clementine 1 B390/3									
12-1064 Esbal Clementine 2 B390/3									
12-1065 Esbal Clementine 3 B390/3									
B390/3 Original Source (CRI)									

Strong positive Medium positive Light positive Negative Dead

5.3.4 Sanger sequencing

As suggested by the genotype specific PCR results, all sequences were very similar and grouped close to the Taiwan-Pum/SP/T1 isolate (JX266712) which is part of the RB genotype (Figure 5.8). Only one sequence showed a bit more diversity but also had a lower

quality score after sequencing, and this might be due to it being a mixed source of RB and B165. This was resolved with Illumina sequencing.

5.3.5 Illumina sequencing

The p33 gene were chosen for Illumina sequencing as p33 amplification and sequencing was already preformed routinely in our laboratory (Read, *personal communication*) and at the time of this study the results within Chapter 2 suggesting a combination of different genes was not yet available. For the B390/3 source, 197 028 (19.6%) of the 1 002 868 total reads obtained mapped to the CTV references and for the B390/3 inoculated Mexican lime sample, 79 266 (36.7%) of the 215 864 total reads mapped to CTV references. For both samples, most of the reads (92.1% for B390/3 and 94.1% for the Mexican lime sub-isolate of B390/3) mapped to the Taiwan-Pum/SP/T1 isolate (JX266712) which forms part of the RB genotype (Figure 5.9 and 10). Negligible numbers of reads (3 – 0.1 % of CTV specific reads), albeit of the entire amplicon consensus length, mapped to other genotypes. We are unsure whether these represent sequences of these genotypes present at extremely low incidences within the viral population or whether they represent artefacts of the technique. No T36 specific reads were obtained, confirming that the T36 positive result of PCR was due to the cross-reaction of the primers to the RB genotype which was shown to occur in the source by Illumina sequencing. The presence of B165 could also not be confirmed in either of the samples. This is similar to the case with the New Venture 41/2 source that was tested in Chapter 4. The RT-PCR gave a positive result, but the amount of reads that mapped to the B165 isolate was negligible. As the genotype specific PCR primers amplify a different region to the p33 gene used as template for Illumina sequencing this may reflect differences in the genome due to recombination, or the cross-reaction of the B165 primers to an unknown isolate. To resolve this the entire genome must be sequenced.

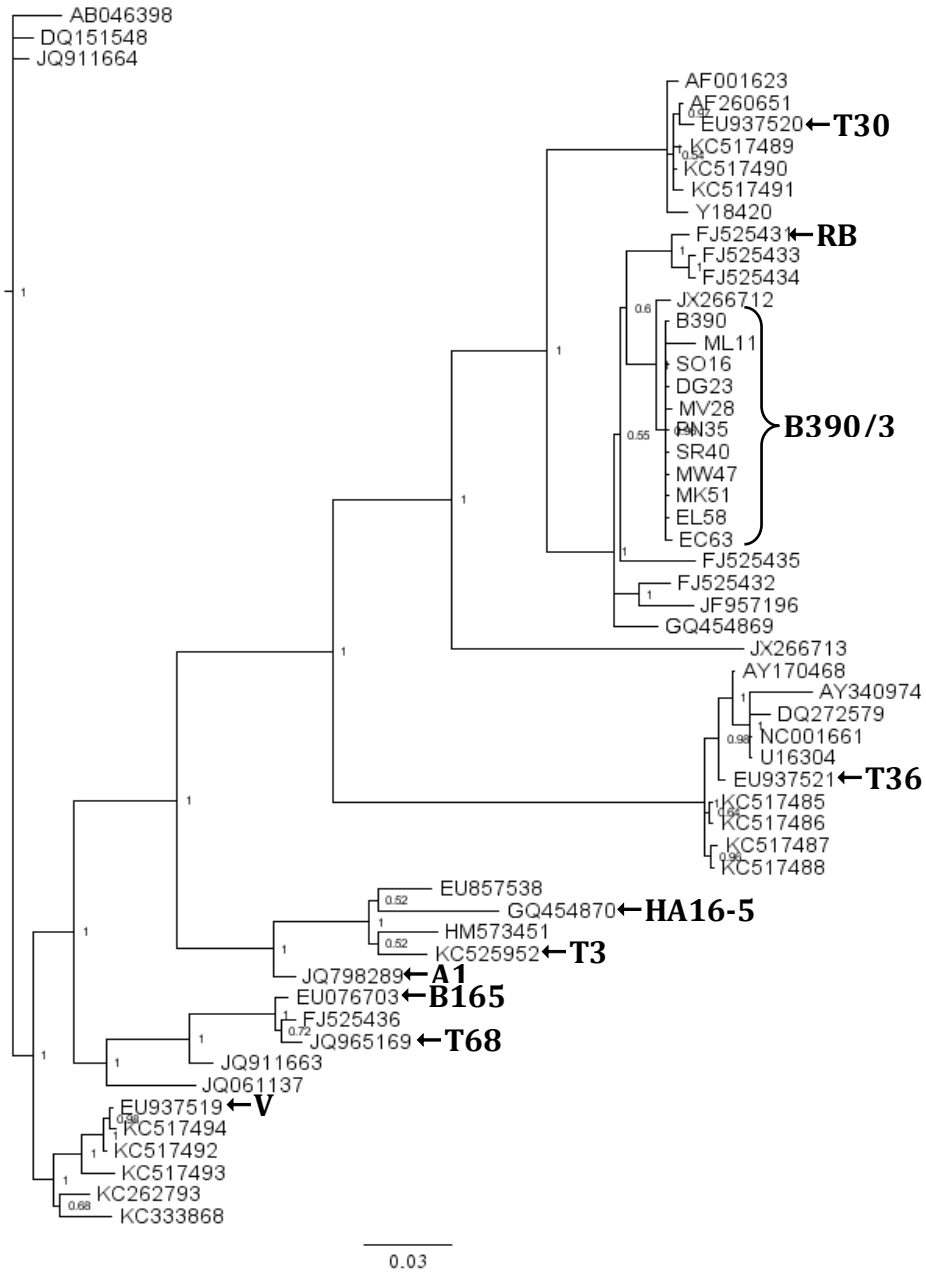

Figure 5.8: Phylogenetic dendrogram of the p33 gene sequence of the B390/3 biological indexing sources and the nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Amplicon sequences are indicated through the host name and last two values in their accession numbers. Mexican lime (ML), Sour orange (SO), Duncan grapefruit (DG), Madam Vinous sweet orange (MV), Palmer navel sweet orange (PN), Star Ruby grapefruit (SR), Miho Wase mandarin (MW), Midknight Valencia sweet orange (MK), Eureka lemon (EL) and Esbal Clementine (EC).Values at nodes show branch support based on posterior probability.

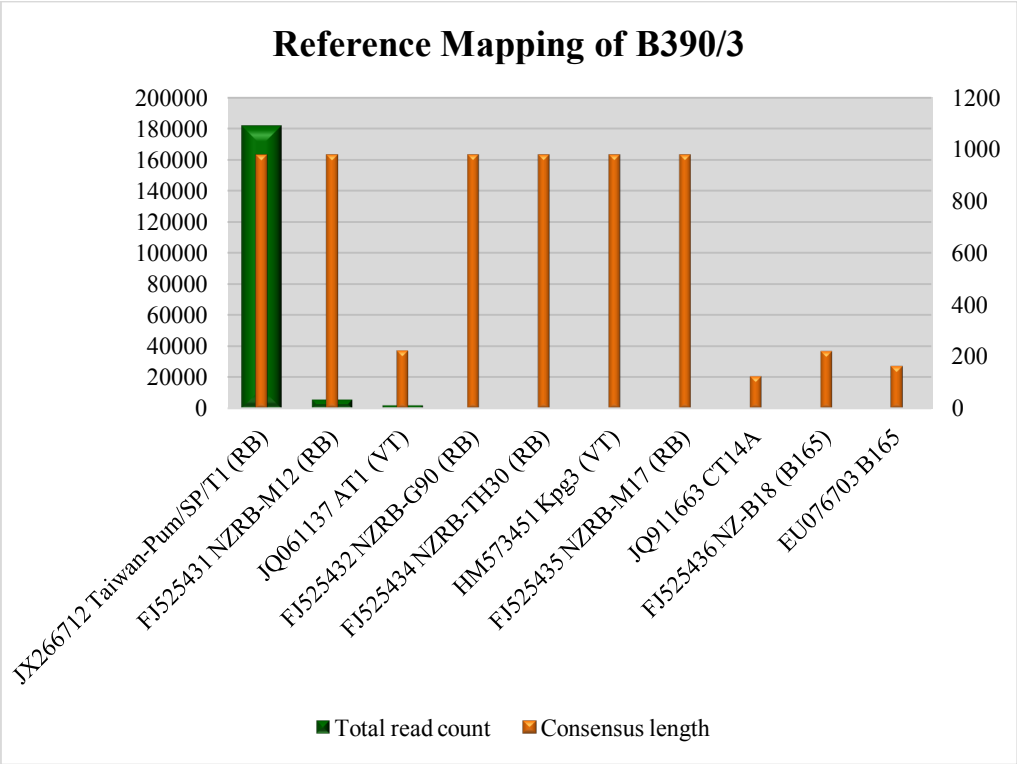

Figure 5.9: Reference mapping of B390/3 sequence reads against CTV references, showing total number of mapped reads (primary axis) and consensus length of the mapped reads (secondary axis) in base pairs.

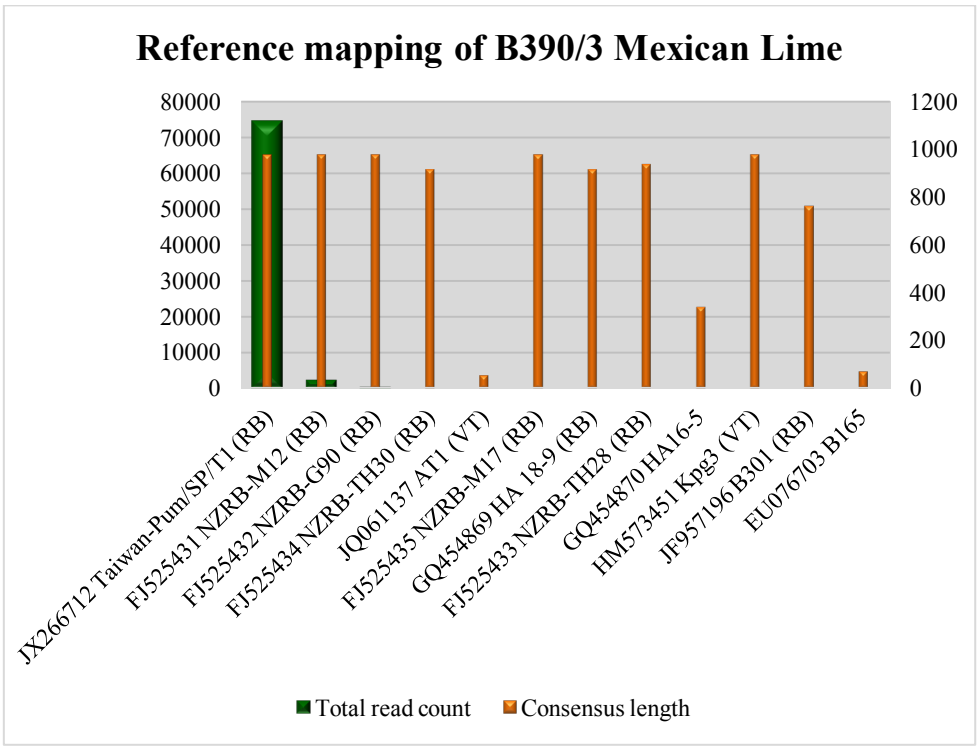

Figure 5.10: Reference mapping of B390/3 Mexican lime sequence reads against CTV references, showing total number of mapped reads (primary axis) and consensus length of the mapped reads (secondary axis) in base pairs.

5.4 Conclusion

CTV candidate mild protecting source B390/3 did not produce severe symptoms on either the Garnsey *et al*. (1987) set of indicators (Mexican lime, Sour orange, Duncan grapefruit, Palmer Navel sweet orange on sour orange rootstock, Midknight Valencia orange), nor on any of the citrus varieties commonly planted in South Africa (Midknight Valencia orange, Star Ruby grapefruit, Esbal clementine or Eureka lemon), under glasshouse conditions. Only mild chlorosis, leaf cupping and stem pitting were observed in Mexican lime which is expected as this is the most sensitive host to *Citrus tristeza virus*. The viral population of B390/3 can thus be considered mild.

The B390/3 source appears to contain a viral population dominated by, based on the p33 gene, a Taiwan-Pum/SP/T1–like isolate (JX266712.1) belonging to the resistance breaking genotype. A number of other genotypes may also be present in the population but these generated very few reads in Illunina next generation sequencing and we are not sure if they do not possibly represent artifacts of the technique. The RB genotype isolate was transmitted to all hosts, but the B165 genotype (if it is in fact present) was only transmitted to one of the three Mexican lime hosts. The low transmission rate may be due to the low titre of the B165 genotype in the original source (as judged by the low intensity band on the agarose gel). The presence of B165 could also not be confirmed with either Illumina sequencing (including results from Chapter 3) or direct Sanger sequencing. The lack of evidence for the presence of B165 using the direct Sanger sequencing data is expected since this technique only detects the most predominant genotype within a mixture. The lack of evidence during the Illumina sequencing was however unexpected. Presumably even if an isolate is present in very low concentrations it would be observed when reference mapping is done. The number of reads mapping to B165 were very low. A number of other genotypes, not detected with the genotype specific PCRs, also yielded read numbers in the same order of magnitude. It

remains unclear whether these genotypes actually do occur within the source. The most probable explanation would be the occurrence of recombination and since these different techniques use different areas of the genome (Genotype specific RT-PCR: ORF 1a, 5' end of genome; Illumina and direct Sanger sequencing: p33, center of genome) they are not directly comparable. This is similar to the situation that occurred during the Scott *et al.* study (2013) where sequencing produced different results for the 3' half and the 5' half of the genome for the same isolate. It was later discovered that the isolate was indeed a recombinant (Zablocki and Pietersen, 2014). The other option is that the p33 primers are detecting unknown isolates not detected with the genotype specific PCRs.

The lack of severe symptoms in the different host would make the B390/3 source suitable for cross-protection, but data from field trials conducted by CRI will also need to be considered. Since the source mainly consists of the RB genotype, if it is used as a cross-protection source it will only protect against strains from this genotype, and the plant will be vulnerable to strains from other genotypes (Folimonova *et al.*, 2010). This source would however be useful if it is used in combination with other mild single genotype sources to create a mixture of genotypes to be used as a cross-protection source.

5.5 References

Albiach-Marti, M.R., da Graça, J.V., van Vuuren, S.P., Guerri, J., Cambra, M., Laigret, F., and Moreno, P. 1996. The Effects of Different Hosts and Natural Disease Pressure on Molecular Profiles of Mild Isolates of *Citrus tristeza virus* (CTV). In: *Proceedings of the 13th Conference of the International Organisation of Citrus Virologists. (Graça, J.V., Moremo, P., and Yokomi, R.K.)* pp. 147-153. *IOCV*. Riverside, California.

Albiach-Marti, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., Lopez, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., and Dawson, W.O. 2000. Sequences of *Citrus tristeza virus* Separated in Time and Space Are Essentially Identical. *The Journal of Virology* 74(15), pp. 6856-6865.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014a. Final Report: Cross-Protection of Marsh and Star Ruby Grapefruit Using Beltsville Sub-Isolates of Nartia Mild Strain, pp. 23. Citrus Research International, Nelspruit, South Africa.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014b. Progress Report: Cross-Protection of Star Ruby Using Beltsville Sub-Isolates of Nartia Mild Strain for the Orange River Valley, pp. 24. Citrus Research International, Nelspruit, South Africa.

Breytenbach, J.H.J., Cook, G., and Van Vuuren, S.P. 2014c. Progress Report: Cross-Protection of Marsh and Star Ruby by Using the Best Field Isolates Collected in the Different Grapefruit Production Areas of Southern Africa., pp. 105. Citrus Research International, Nelspruit, South Africa.

Broadbent, P., Brlansky, R.H., and Indsto, J. 1996. Biological Characterization of Australian Isolates of *Citrus tristeza virus* and Separation of Subisolates by Single Aphid Transmission. *Plant Disease* 80(3), pp. 329-333.

Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. 2012. jModelTest 2: More Models, New Heuristics and Parallel Computing. *Nature Methods* 9(8), pp. 772-772.

Folimonova, S.Y., Robertson, C.J., Shilts, T., Folimonov, A.S., Hilf, M.E., Garnsey, S.M., and Dawson, W.O. 2010. Infection with Strains of *Citrus tristeza virus* Does Not Exclude Superinfection by Other Strains of the Virus. *The Journal of Virology* 84(3), pp. 1314-1325.

Fulton, R.W. 1986. Practices and Precautions in the Use of Cross Protection for Plant Virus Disease Control. *Annual Review of Phytopathology* 24, pp. 67-81.

Garnsey, S.M., Civerolo, D.J., Gumpf, D.J., Paul, C., Hilf, M.E., Lee, R.F., Brlansky, R.H., Yokomi, R.K., and Hartung, J.S. 2005. Biological Characterization of an International Collection of *Citrus tristeza virus* (CTV) Isolates. In: *Proceedings of the 16th Conference of the International Organisation of Citrus Virologists. (Hilf, M.E., Duran-Vila, N., and Rocha-Peña, M.A.)* pp. 75-93. *IOCV*. Riverside, California.

Garnsey, S.M., Gumpf, D.J., Roistacher, C.N., Civerolo, E.L., Lee, R.F., Yokomi, R.K., and Bar-Joseph, M. 1987. Toward a Standardized Evaluation of the Biological Properties of *Citrus tristeza virus*. *Phytophylactica* 19(2), pp. 151-157.

Karasev, A.V., Dawson, W.O., Hilf, M.E., and Garnsey, S.M. 1998. Molecular Biology of *Citrus tristeza virus*: Implications for Disease Diagnosis and Control. *ACTA Horticulturae* 472, pp. 333-350.

Katoh, K., Misawa, K., Kuma, K.i., and Miyata, T. 2002. MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform. *Nucleic Acids Research* 30(14), pp. 3059-3066.

Moreno, P., Ambrós, S., Albiach-Martí, M.R., Guerri, J., and Peña, L. 2008. *Citrus tristeza virus*: A Pathogen That Changed the Course of the Citrus Industry. *Molecular Plant Pathology* 9(2), pp. 251-268.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice across a Large Model Space. *Systematic Biology* 61(3), pp. 539-542.

Roy, A., Ananthakrishnan, G., Hartung, J.S., and Brlansky, R.H. 2010. Development and Application of a Multiplex Reverse-Transcription Polymerase Chain Reaction Assay for Screening a Global Collection of *Citrus tristeza virus* Isolates. *Phytopathology* 100(10), pp. 1077-1088.

Roy, A., and Brlansky, R.H. 2010. Genome Analysis of an Orange Stem Pitting *Citrus tristeza virus* Isolate Reveals a Novel Recombinant Genotype. *Virus Research* 151(2), pp. 118-130.

Rubio, L., Ayllon, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., and Falk, B.W. 2001. Genetic Variation of *Citrus tristeza virus* Isolates from California and Spain: Evidence for Mixed Infections and Recombination. *The Journal of Virology* 75(17), pp. 8054-8062.

Scott, K.A., Hlela, Q., Zablocki, O., Read, D., van Vuuren, S., and Pietersen, G. 2013. Genotype Composition of Populations of Grapefruit-Cross-Protecting *Citrus tristeza virus* Strain GFMS 12 in Different Host Plants and Aphid-Transmitted Sub-Isolates. *Archives of Virology* 158(1), pp. 27-37.

Van Vuuren, S.P. 2002. Effects of *Citrus tristeza virus* Isolates on Two Tolerant Commercial Scions on Different Rootstocks in South Africa. In: *Proceedings of the 15th Conference of the International Organisation of Citrus Virologist. (Duran-Vila, N., Milne, R.G., and Da Graça, J.V.)* pp. 31-38. *IOCV*. Riverside, California.

Vives, M.C., Rubio, L., Sambade, A., Mirkov, T.E., Moreno, P., and Guerri, J. 2005. Evidence of Multiple Recombination Events between Two RNA Sequence Variants within a *Citrus tristeza virus* Isolate. *Virology* 331(2), pp. 232-237.

Von Broembsen, L., and Lee, T.C. 1988. South Africa's Citrus Improvement Programme. In: *Proceedings of the 10th Conference of the International Organisation of Citrus Virologists. (Timmer, L.W., Garnsey, S.M., and Navarro, L.)* pp. 407-416. *IOCV*. Riverside, California.

Zablocki, O., and Pietersen, G. 2014. Characterization of a Novel *Citrus tristeza virus* Genotype within Three Cross-Protecting Source GFMS12 Sub-Isolates in South Africa by Means of Illumina Sequencing. *Archives of Virology*, pp. 1-7.

Zanutto, C.A., Corazza, M.J., Nunes, W.M.d.C., and Müller, G.W. 2013. Evaluation of the Protective Capacity of New Mild *Citrus tristeza virus* (CTV) Isolates Selected for a Preimmunization Program. *Scientia Agricola* 70(2), pp. 116-124.

Appendix A: Phylogenetic trees and recombination analysis

Table A.1: Summary of all phylogentic trees compiled in Chapter 2. When BS (Bootstrap value) or PP (Posterior probability) is shown it indicates that although the isolates grouped in separate clades, the branch support for these clades where too low to support separate clustering. reps* = Representative isolates for possible genotypes.

	Nucleotide						Amino acids				
Fragment	**Maximum Likelihood**		**Neighbor Joining**		**Bayesian**		**Maximum Likelihood**		**Neighbor Joining**		
	Nr of genotypes	**Separation of reps*? Which not?**	**Nr of genotypes**	**Separate reps*? Which not?**	**Nr of genotypes**	**Separate reps*? Which not?**	**Nr of genotypes**	**Separate reps*? Which not?**	**Nr of genotypes**	**Separate reps*? Which not?**	**Avg. Genotypes**
Complete	9	No, VT, T30 and T3 group together (BS)	10	Yes, all but VT and T3 (BS)	14	Yes	3	No, most group together (BS)	11	Yes	9.4
5' UTR	2	No, most group together (BS)	3	No, most group together (BS)	8	No, T36 and RB group together and T68 and B165 groups together	n/a	n/a	n/a	n/a	4.3
ORF 1a	7	No, VT, T30 and T3 group together (BS)	4	No, VT, T3, T30, B165, HA16-5 and T68 group together (BS)	14	Yes	3	No, most group together (BS)	8	No, VT, T3 and T30 group together (BS)	7.2
ORF 1b	4	No, most group together (BS)	8	Yes, all but T3 and T30 (BS)	3	No, most group together (PP)	6	No, T68 and B165 together (BS), T30, T3 and VT together (BS)	4	No, VT and T3 together (BS), HA16-5, B165 and T68 together (BS) and RB and T36 together (BS)	5
ORF 2	15	Yes, but some strains group with different genotypes or further away from other groups than for instance in the Bayesian complete tree	10	Yes, all but T68 and B165 group together. T3 is closer to HA16-5 than T30 and VT	11	No, B165 and T68 group together and HA16-5 and T3 groups together. RB groups closer to T30 than T36, and T30 is very far from VT.	10	No, T68 and B165 together (BS), CT-ZA3 with VT, T3 and HA16-5 together (BS)	9	No, T68 and B165 group together, HA16-5 and T3 group together (BS).	11
ORF 3	3	No, most group together (BS)	3	No, most group together (BS)	3	No, most group together (PP)	2	No, most group together (BS)	1	No, most group together (BS)	2.4
ORF 4	10	Yes, but some strains group with different genotypes than for e.g. in the Bayesian complete tree	5	No, VT, B165, T68 and T3 group together (BS), and HA16-5 and RB group together	3	No, most group together (PP)	7	No, Ha16-5, RB and T3 together (BS), VT, T68 and B165 together (BS)	4	No, B165, T68, VT, RB and HA16-5 together (BS)	5.8

ORF 5	3	No, most group together (BS)	9	Yes, but B165 and T68 group together	2	No, most group together (PP)	3	No, most group together (BS)	6	No, T30 and RB together (BS) and VT, B165 and T68 is together (BS)	4.6
ORF 6	4	No, most group together (BS)	10	Yes, but B165 and T68 group together	3	No, most group together (PP)	3	No, most group together (BS)	1	No, most group together (BS)	4.2
ORF 7	6	No, RB, T30 and T36 group together (BS), B165 and T68 is also together,	3	No, most group together (BS)	6	No, T30, T36 and RB group together (PP) and B165 and T68 group together	2	No, most group together (BS)	1	No, most group together (BS)	3.6
ORF 8	3	No, most group together (BS)	7	No, T36 and T30 group together (BS) and B165 and T68 group together	6	No, T30, T36 and RB group together (PP) and B165 and T68 group together	5	No, most group together (BS)	1	No, most group together (BS)	4.4
ORF 9	4	No, most group together (BS)	2	No, most group together (BS)	3	No, most group together (PP)	3	No, most group together (BS)	1	No, most group together (BS)	2.6
ORF 10	3	No, most group together (BS)	6	No, B165 and T68 group together, as well as VT (BS), some of the RB strains group with T36	3	No, most group together (PP)	3	No, RB, T30. VT. T68, B165 and T3 group together (BS)	2	No, most group together (BS)	3.4
ORF 11	2	No, all strains group together (BS) except for the FJ RB strains, they are separate from the rest. Only JX12 and GQ69 which usually group with RB are now grouping with HA16-5, i.e., everything else.	4	No, most group together (BS)	3	No, most group together (PP)	3	No, most group together (BS)	3	No, most group together (BS)	3
3' UTR	1	No, most group together (BS)	2	No, most group together (BS)	10	No, T30 and VT group together (PP) and B165 and T68 group together	n/a	n/a	n/a	n/a	4.3

DUF 3648	4	No, most group together (BS)	7	No, B165 and T68 group together and some of the RB strains group with T36	2	No, most group together (PP)	2	No, most group together (BS)	7	No, T36 and T30 together (BS) and RB and VT is separate, but some RB are with VT	4.4
DUF 3762	4	No, most group together (BS)	4	No, most group together (BS)	3	No, most group together (PP)	3	No, most group together (BS)	3	No, most group together (BS)	3.4
Protease I	5	No, VT, T3 and T30 group together (BS)	6	No, VT and T3 group together (BS), B165 and T68 group together as well as HA16-5 (BS)	5	No, T36, RB, B165 and T68 group together (PP)	5	No, T3, T30, RB, T36 and HA16-5 group together (BS)	4	No, most group together (BS)	5
DUF 3614	8	No, T30, B165 and VT group together (BS)	6	No, VT, B165, T3 and T30 group together (BS)	7	No, T68 and HA16-5 group together (PP) and B165 and VT group together. Many of the RB strains group with T36	4	No, T3, T30, B165, RB, T36 and VT group together (BS)	7	No, B165, VT, HA16-5 and T68 group together (BS), RB is split in two, with one grouping closer to T36	6.4
Protease II	5	No, RB and T36 are separate bur some RB strains group with T36, B165, VT, T30 and T3 group together (BS)	6	No, VT and B165 group together, as well as T30 (BS), RB and T36 group together (BS)	3	No, most group together (PP)	5	No, T36 and RB group together (BS), T3, B165, VT and T30 together (BS)	1	No, most group together (BS)	4
Methyl-transferase	10	No, T30, VT and T3 group together (BS), RB and T36 separate bur some RB strains group with T36	6	No, VT and T30 group together (BS) and B165 and T68 group together	3	No, most group together (PP)	5	No, T30, RB, T36 and VT group together (BS), also T68 and B165 together (BS)	3	No, most group together (BS)	5.4
Helicase	2	No, most group together (BS)	7	No, T3 and T30 group together (BS) and B165 and T68 group together	2	No, most group together (PP)	3	No, most group together (BS)	8	No, T3 and T30 group together (BS), RB is split in two, and the second group is closer to T36	4.4

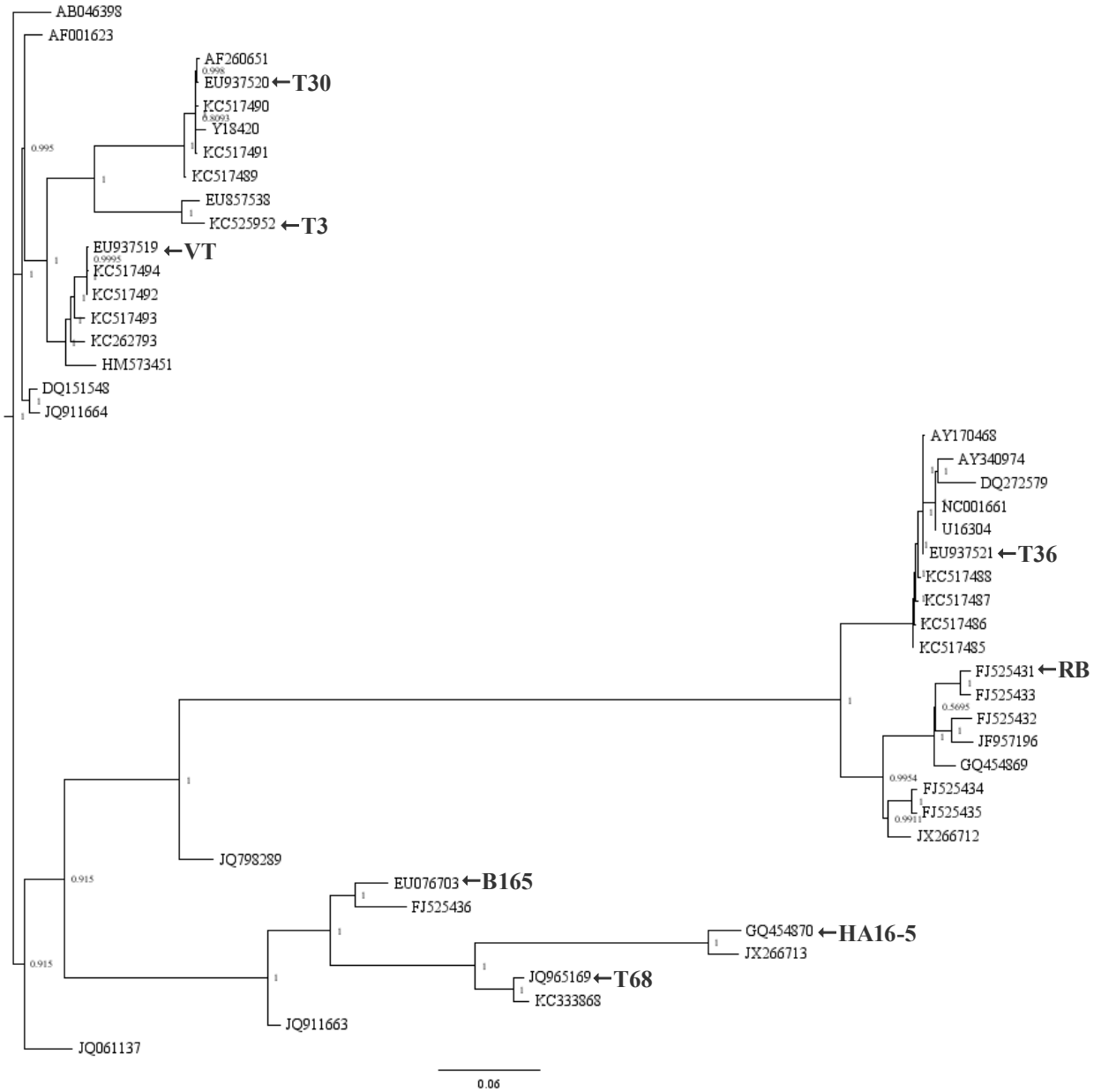

Figure A.1: Phylogenetic dendrogram of the ORF 1a nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

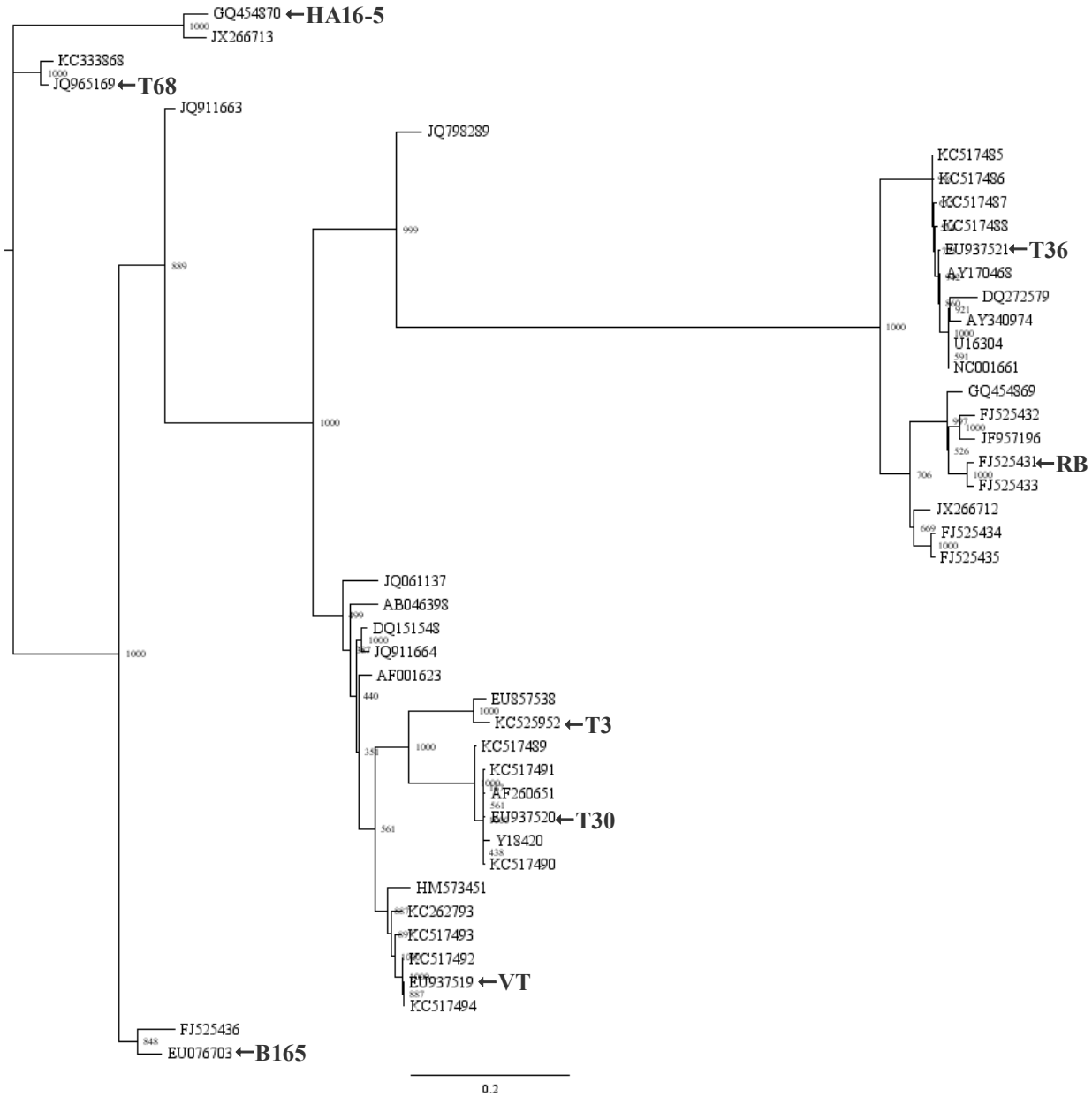

Figure A.2: Phylogenetic dendrogram of the ORF 1a nucleotide sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

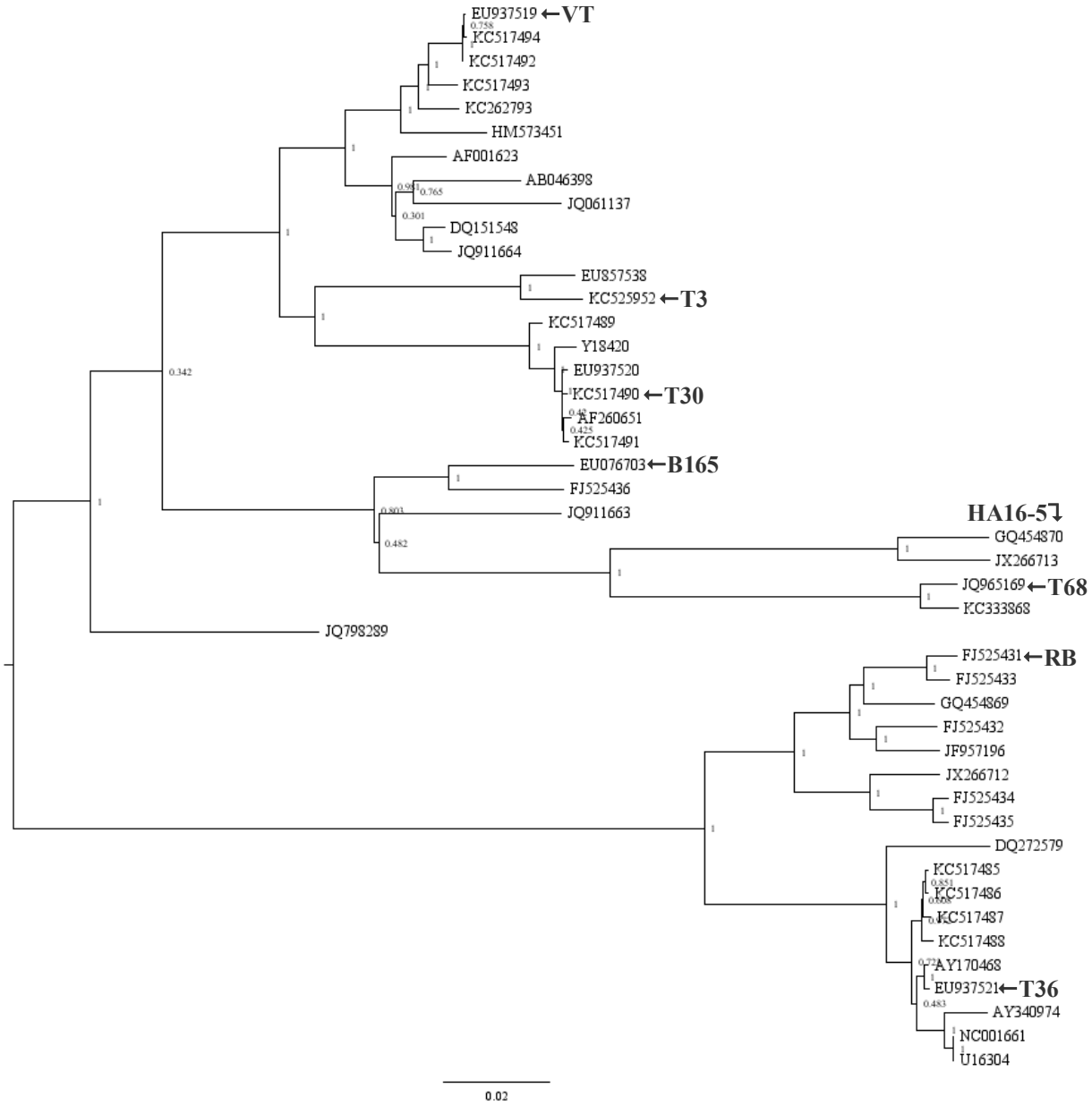

Figure A.3: Phylogenetic dendrogram of the ORF1a nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

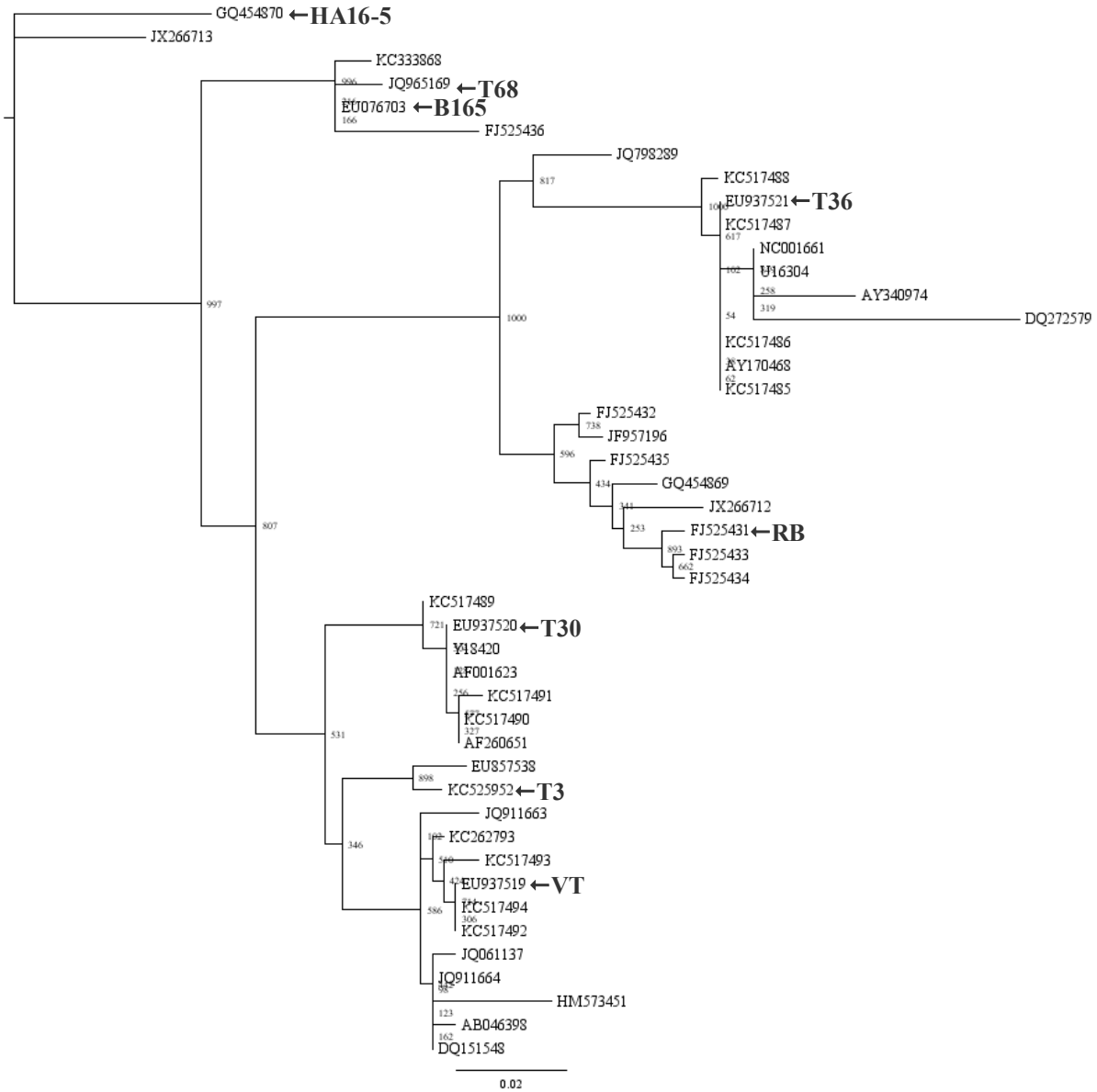

Figure A.4: Phylogenetic dendrogram of the ORF 1b amino acid sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

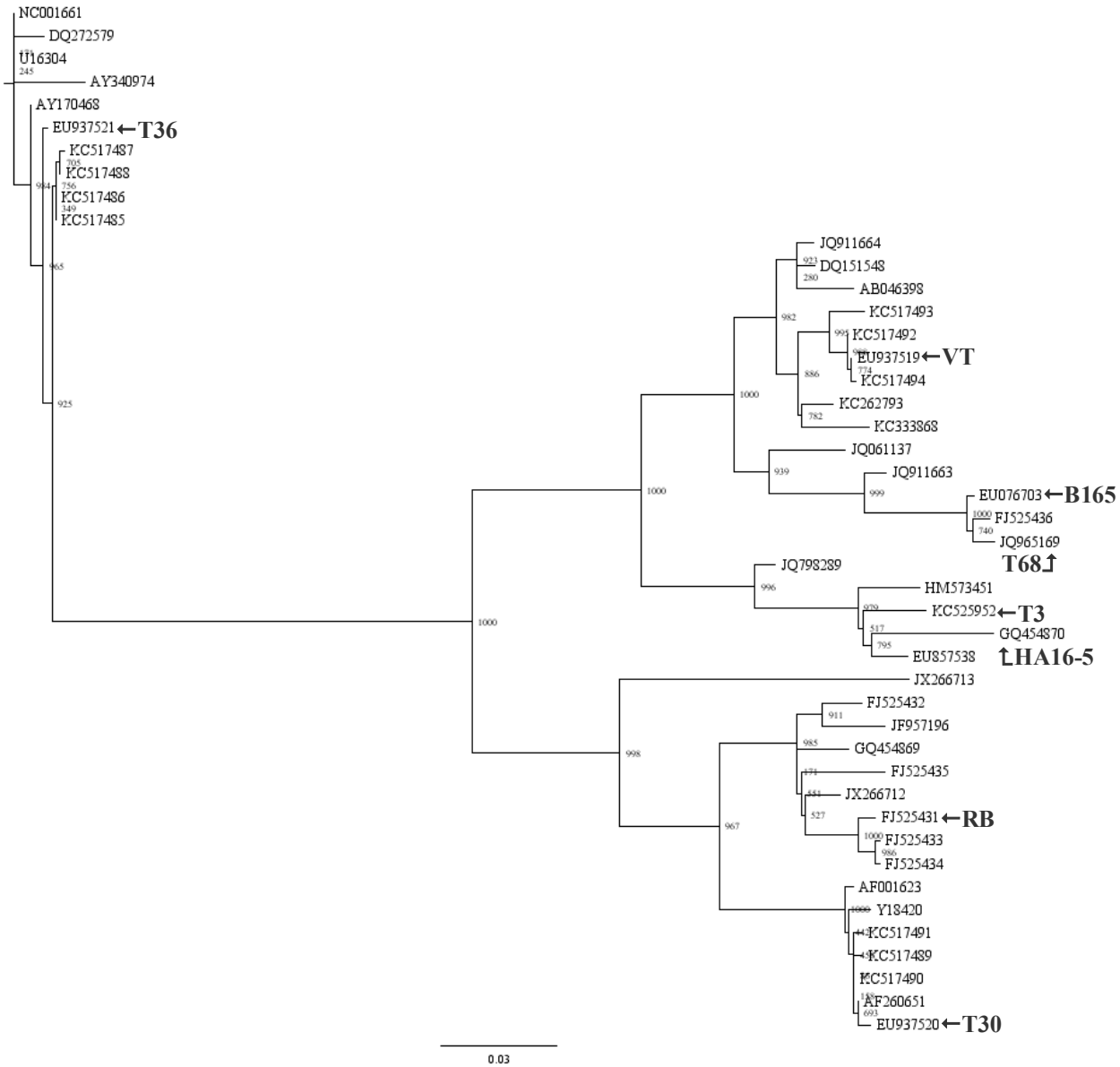

Figure A.5: Phylogenetic dendrogram of the ORF 2 nucleotide sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

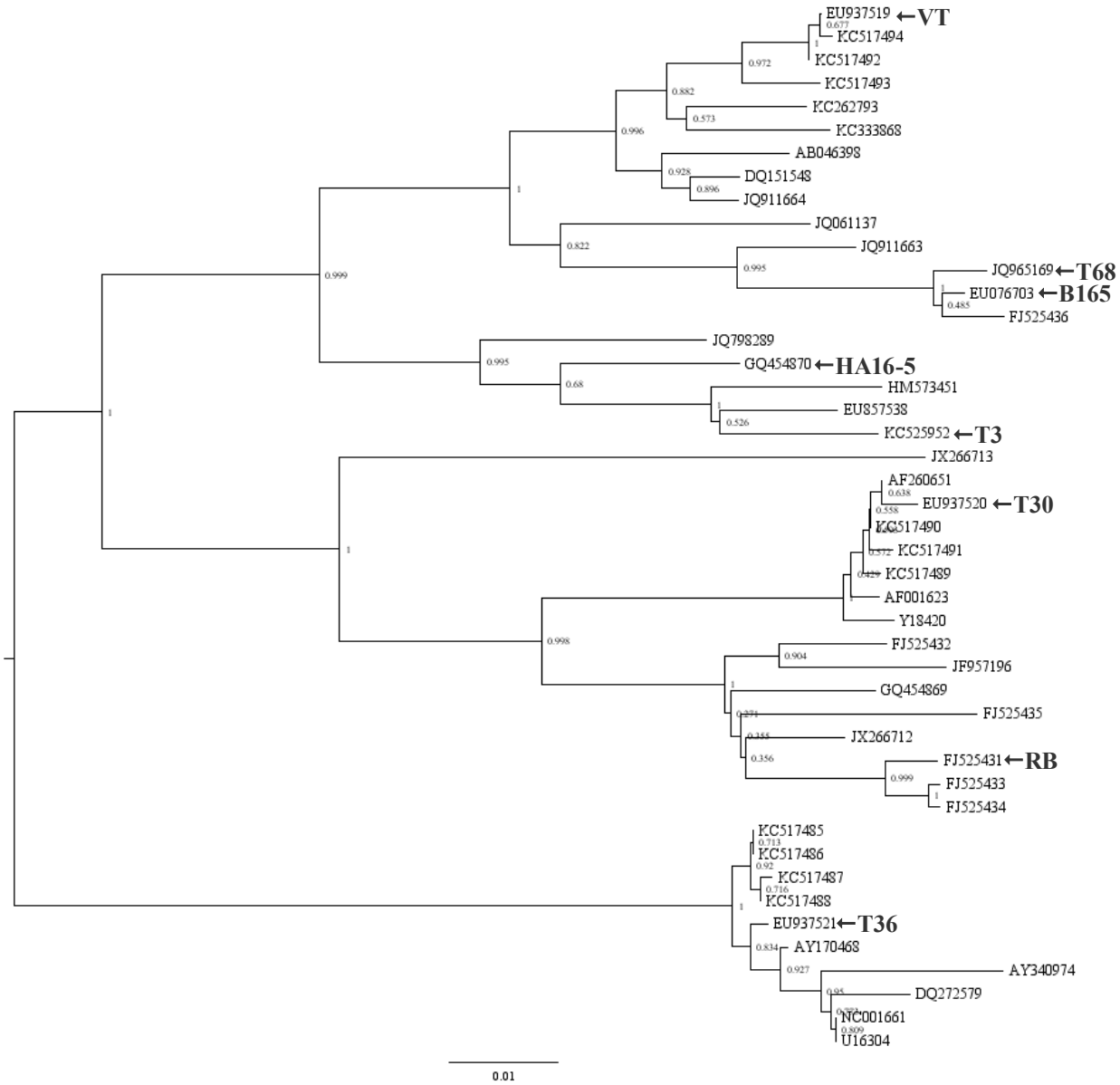

Figure A.6: Phylogenetic dendrogram of the ORF 2 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

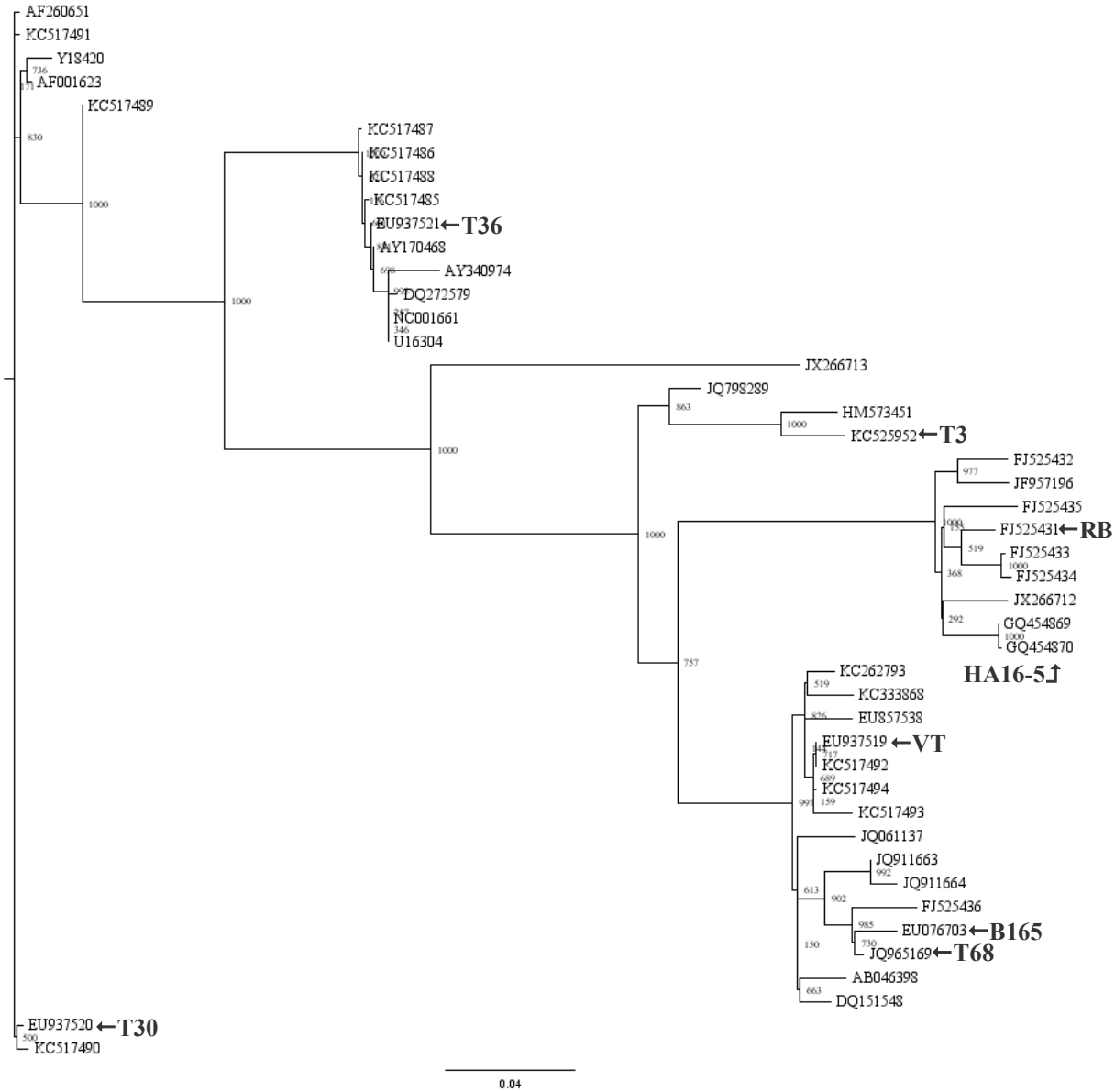

Figure A.7: Phylogenetic dendrogram of the ORF 4 nucleotide sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

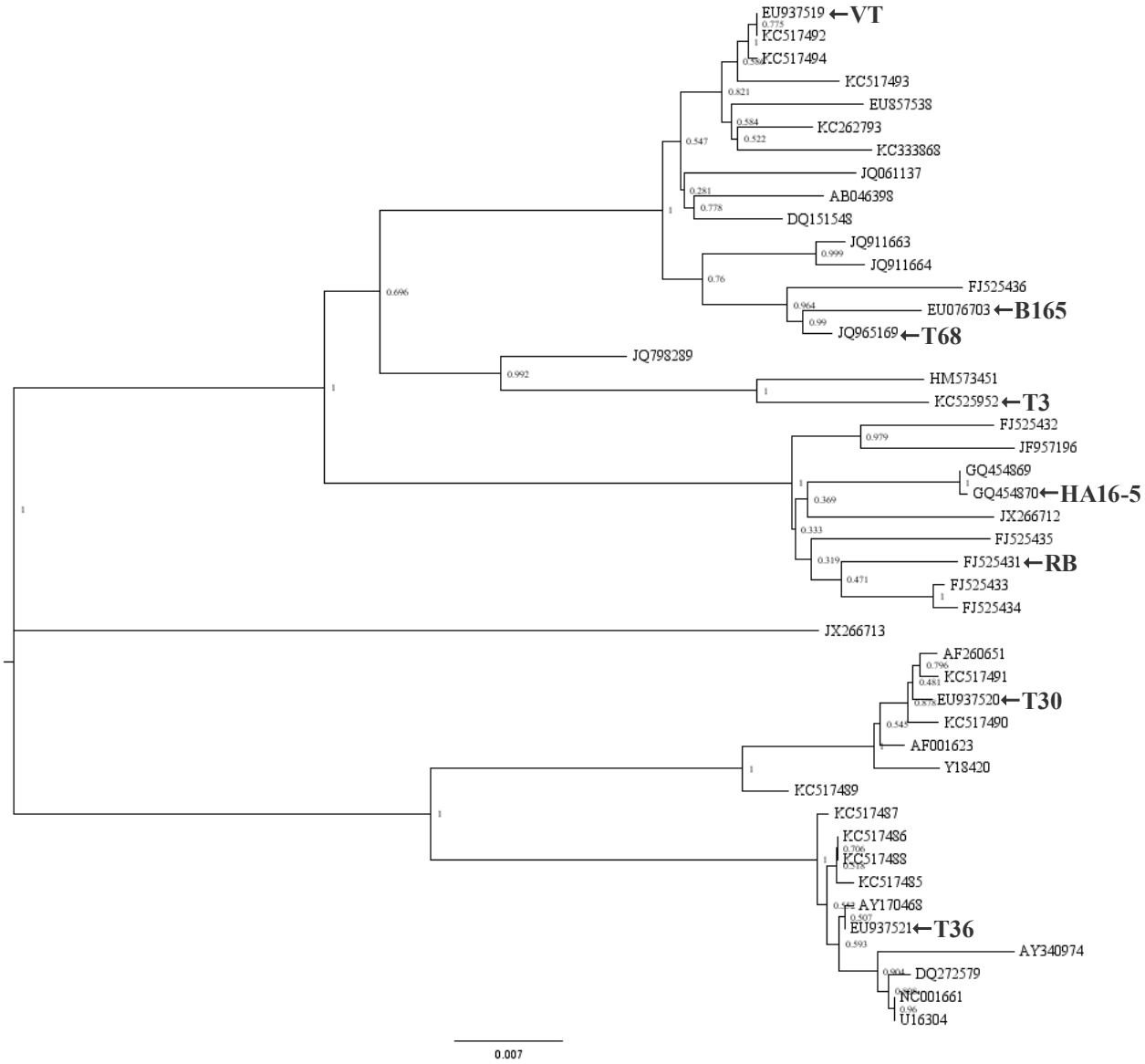

Figure A.8: Phylogenetic dendrogram of the ORF 4 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

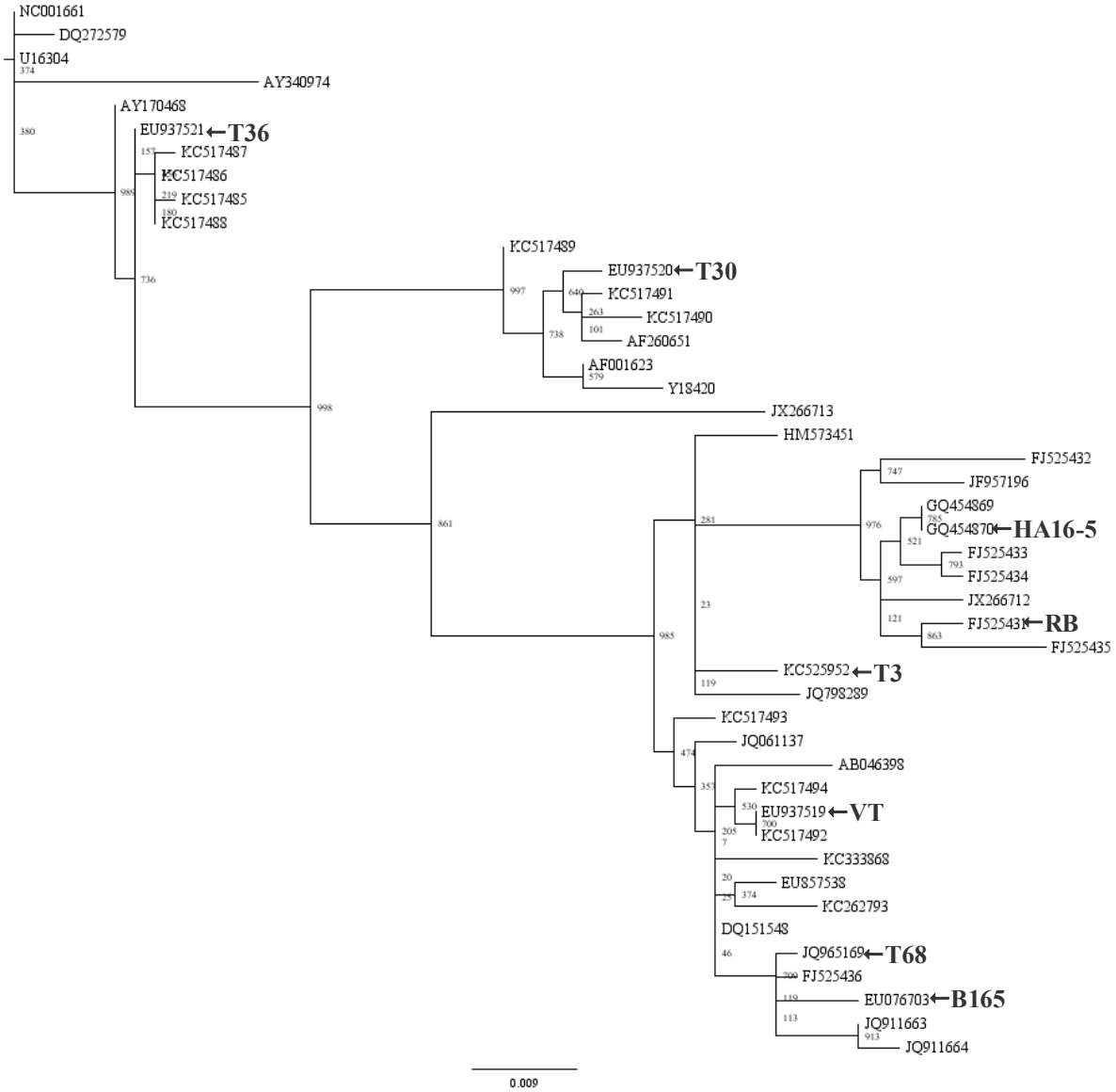

Figure A.9: Phylogenetic dendrogram of the ORF 4 amino acid sequence of 45 CTV reference genomes based on Maximum likelihood. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability (1000 replicates were used).

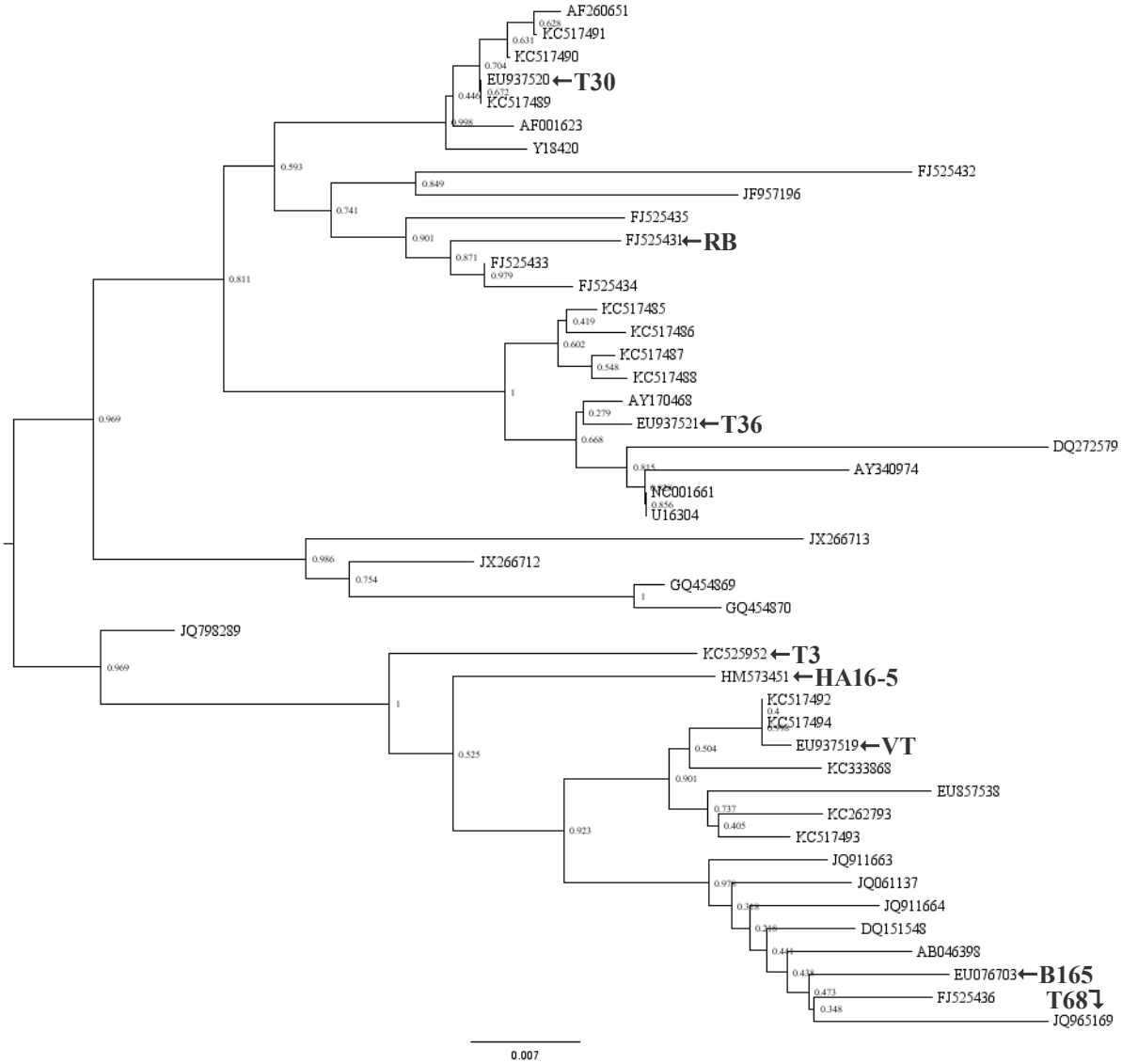

Figure A.10: Phylogenetic dendrogram of the ORF 5 Amino acid sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

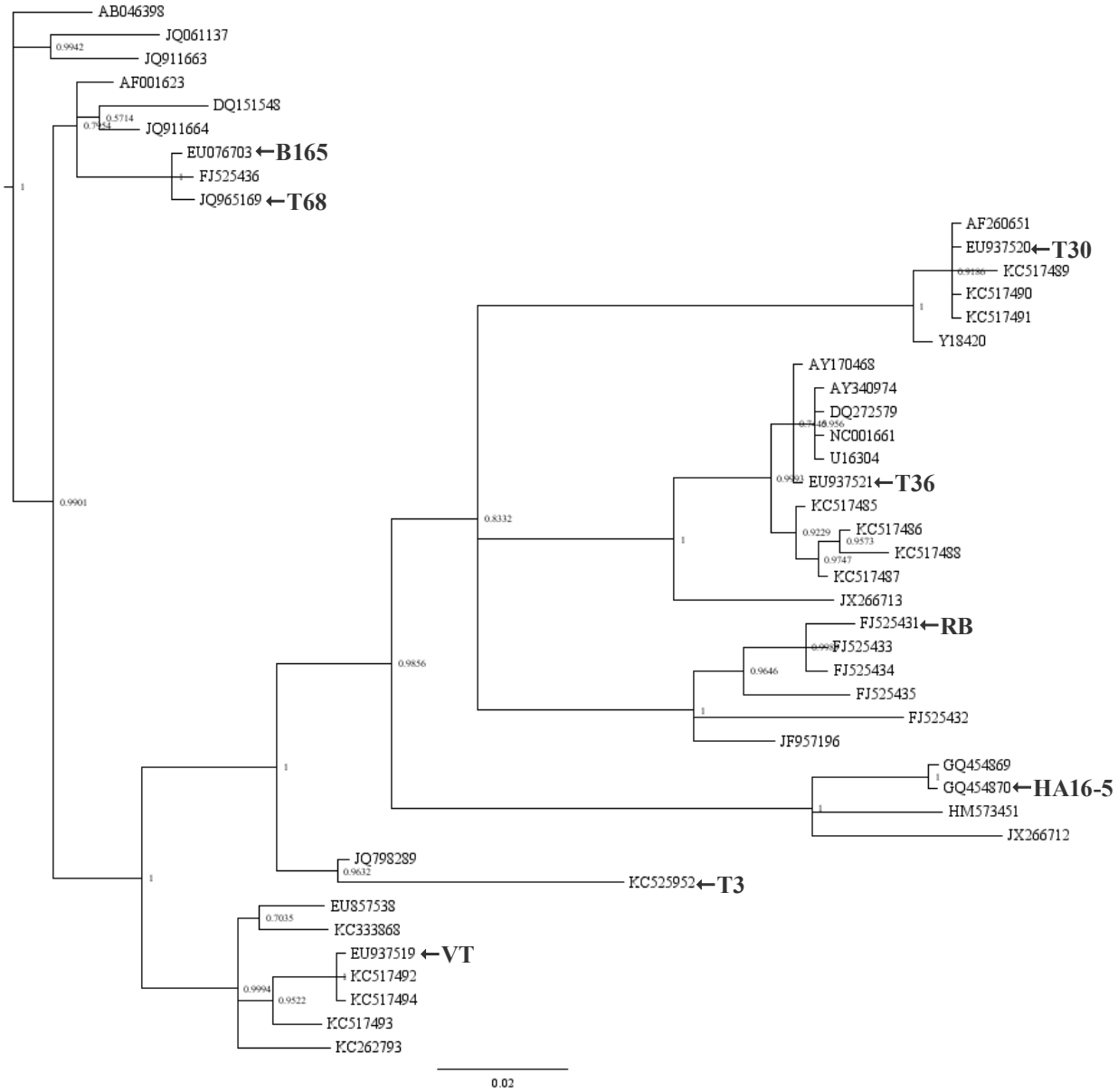

Figure A.11: Phylogenetic dendrogram of the ORF 7 nucleotide sequence of 45 CTV reference genomes based on Bayesian analysis. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on posterior probability.

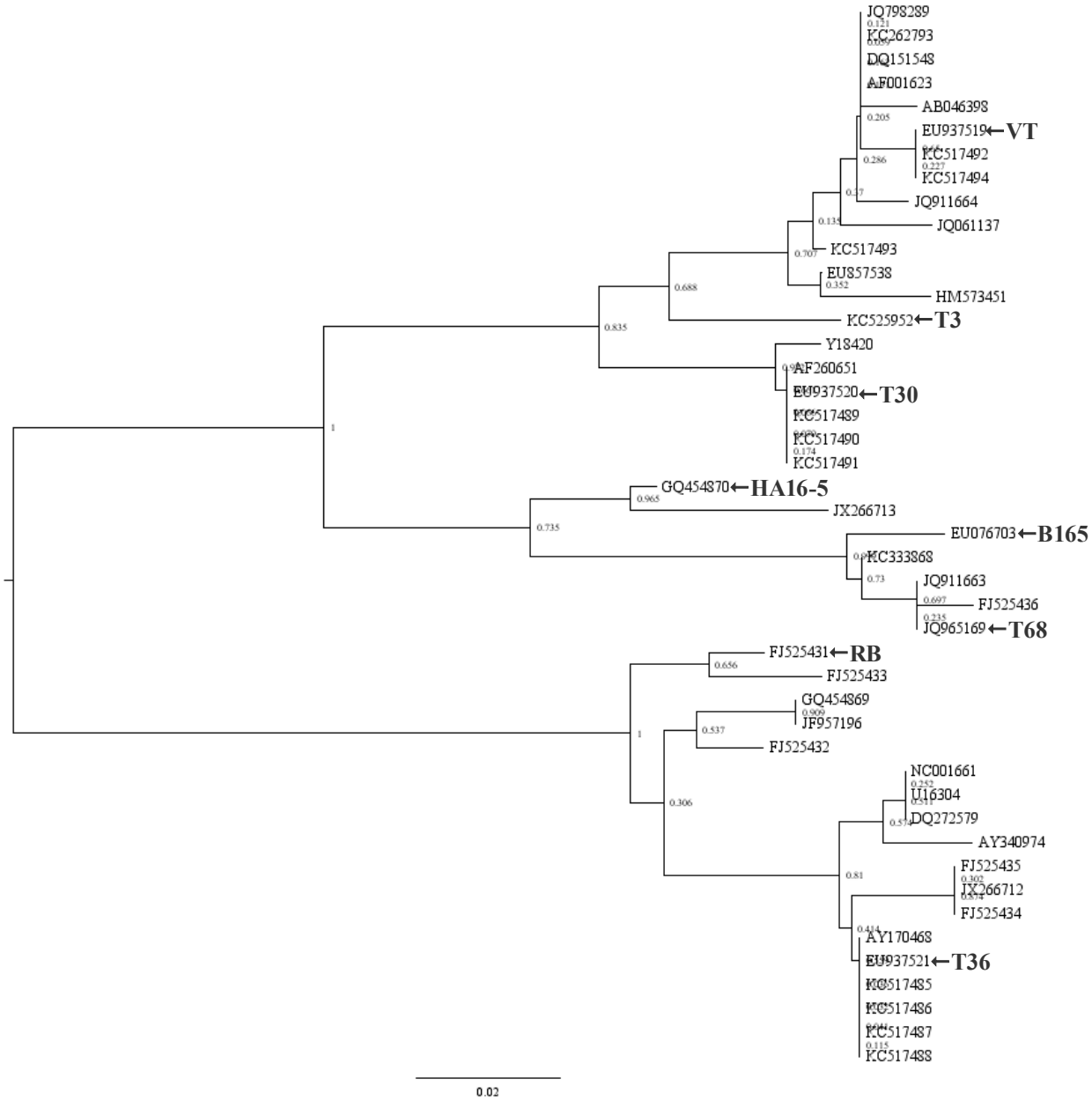

Figure A.12: Phylogenetic dendrogram of the DUF 3648 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

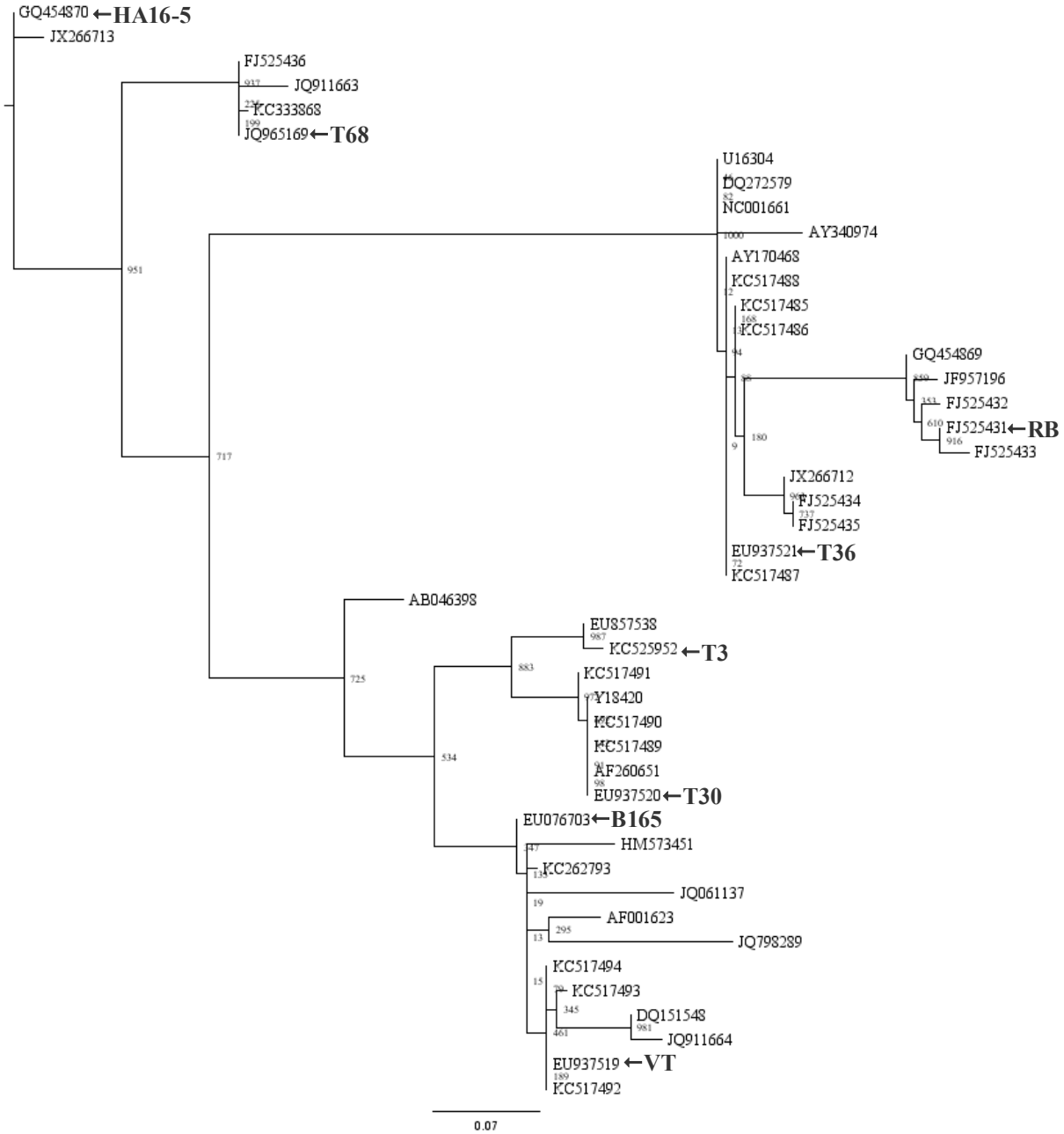

Figure A.13: Phylogenetic dendrogram of the DUF 3614 nucleotide sequence of 45 CTV reference genomes based on a Maximum likelihood approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

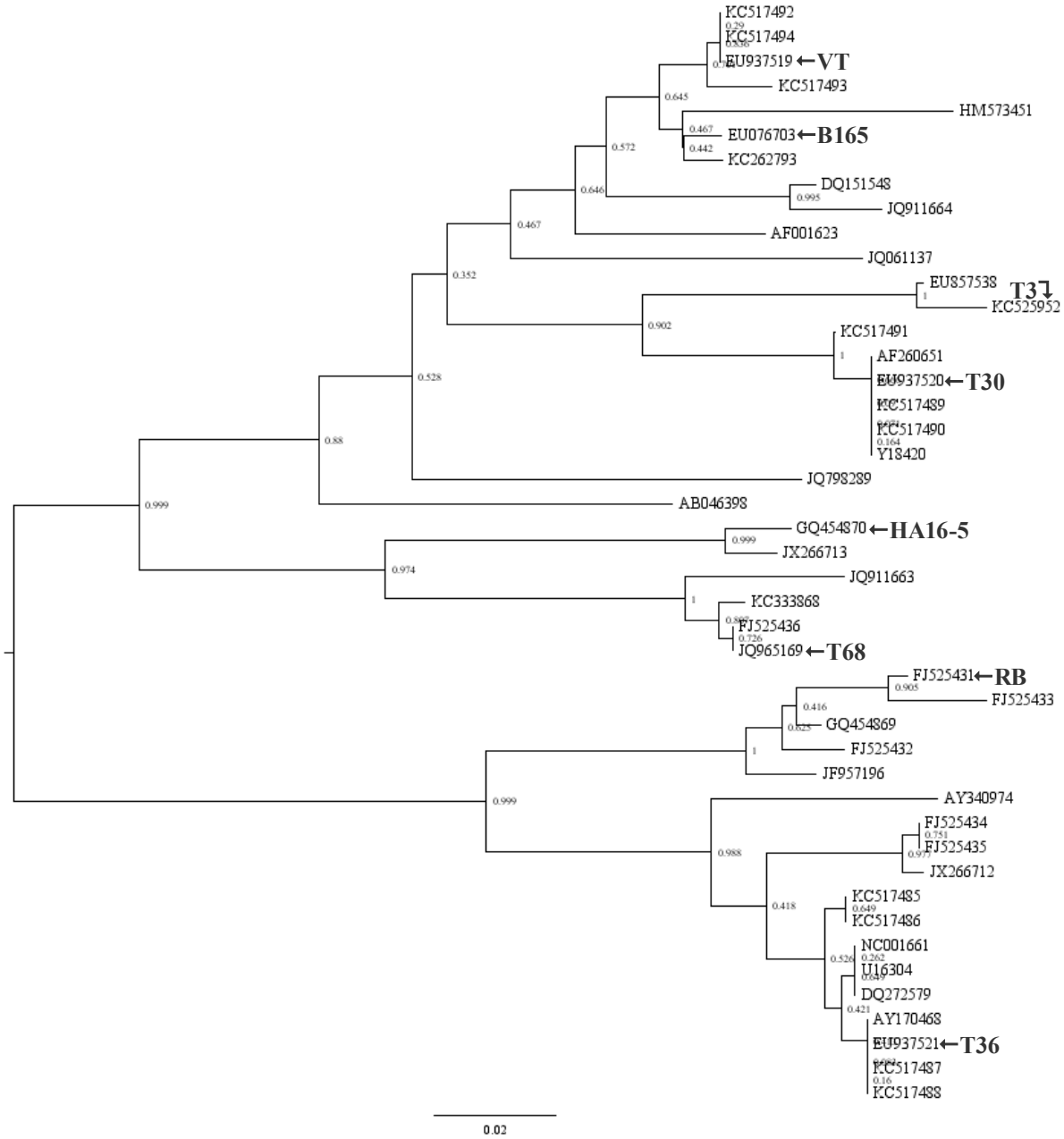

Figure A.14: Phylogenetic dendrogram of the DUF 3614 nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

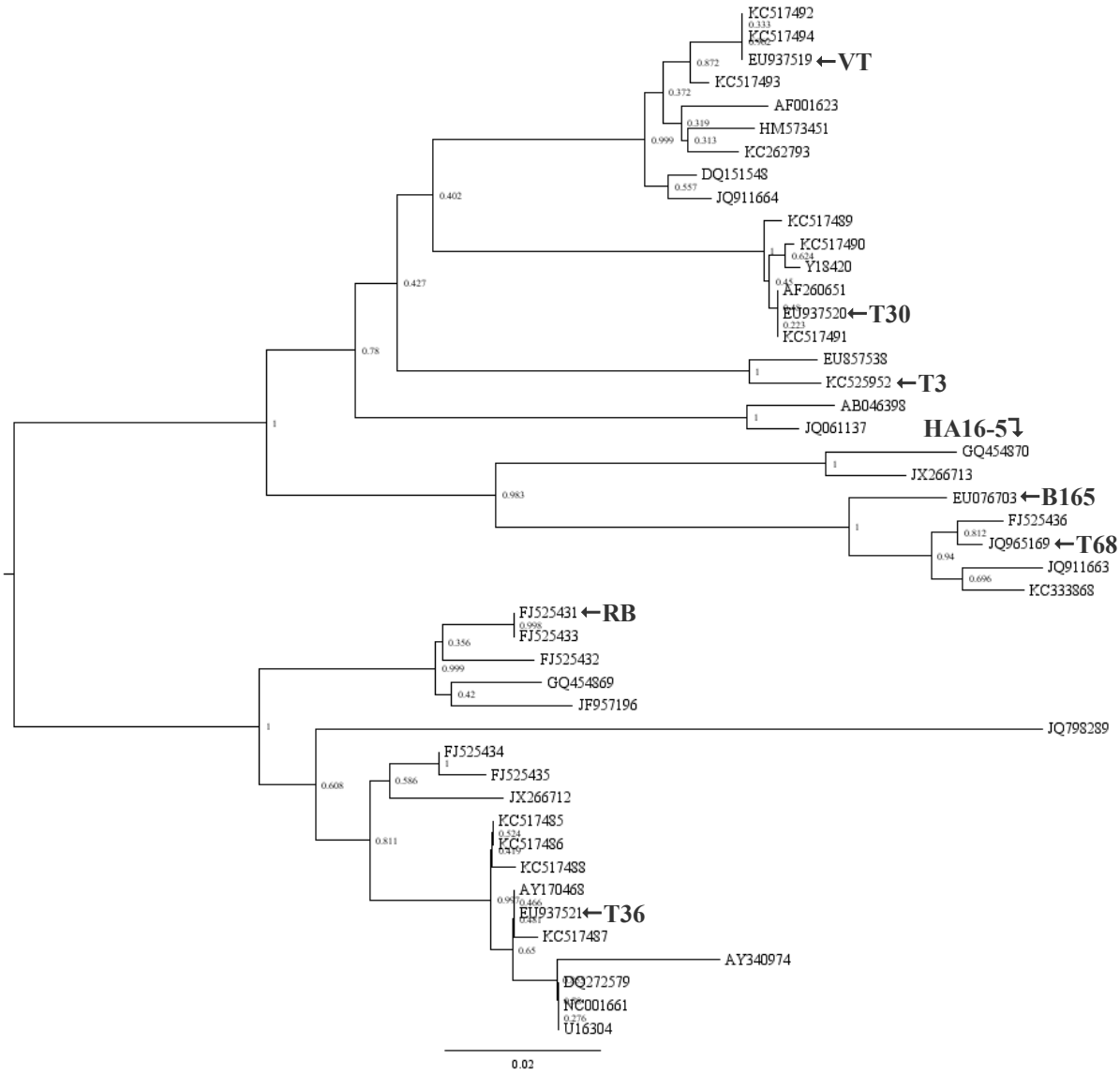

Figure A.15: Phylogenetic dendrogram of the methyltransferase nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

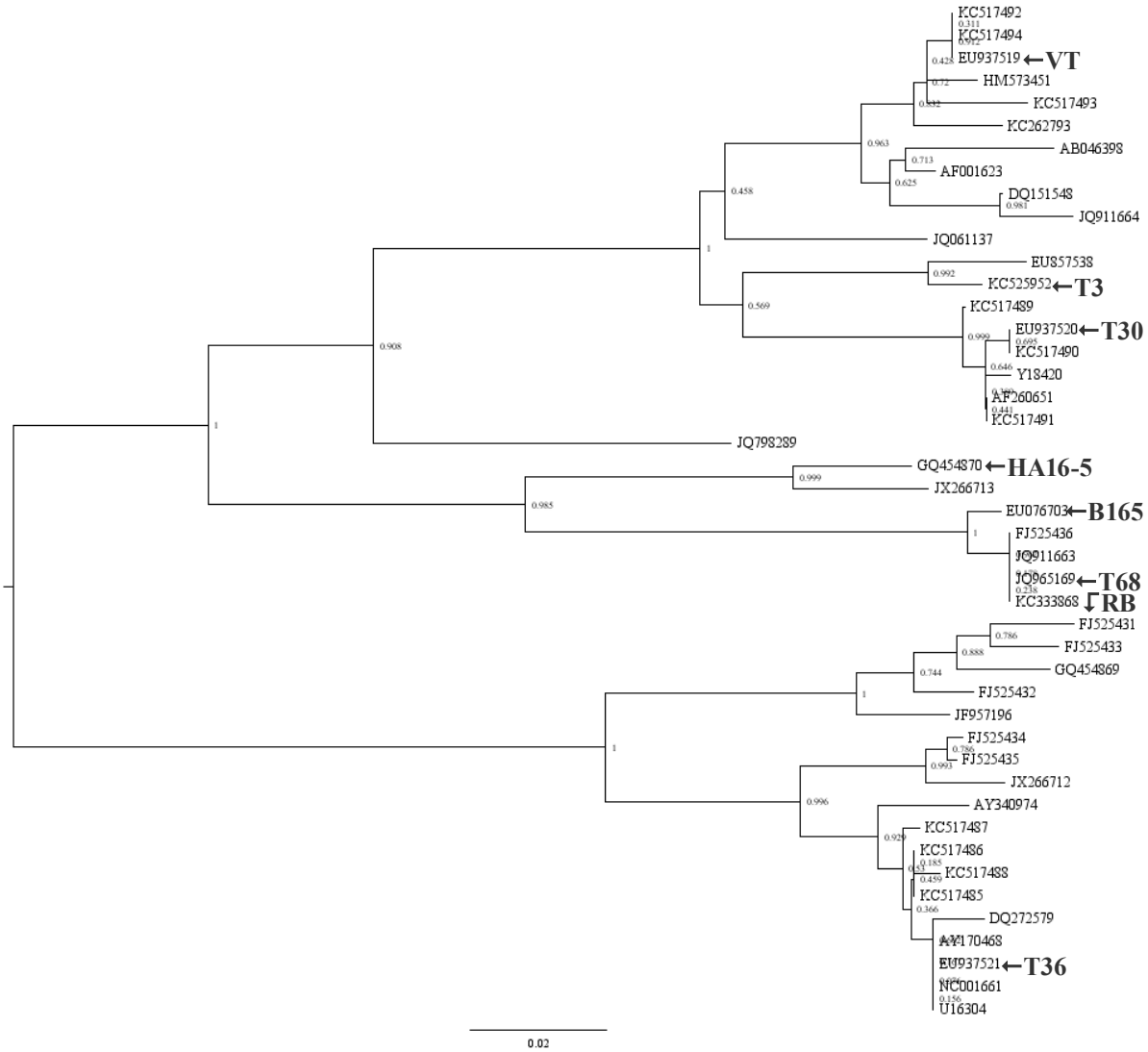

Figure A.16: Phylogenetic dendrogram of the helicase nucleotide sequence of 45 CTV reference genomes based on a Neighbor joining approach. Isolates are identified by their accession numbers as found on GenBank and sequence names of isolates chosen as representatives to analyze dendrograms are indicated. Values at nodes show branch support based on bootstrapping (1000 replicates were used).

Appendix B: Primers used in polymerase chain reactions

Table B.1: Primer pairs used to test one step RT-PCR. IUPAC nucleotide code used, where Y=C/T, R=A/G, K=G/T, V=A/C/G and B=C/G/T.

Primer	Nucleotide Sequence (5' to 3')	Target	Reference
A-Forward	ACGTGTTCGTGAAACGCGG	A region (ORF1a)	Rubio *et al.* 2001
A-Reverse	GTCGATAACTCGACAAACGAGC		
Modified A-Forward	ACGTGTTYRYGAARCGYGG		Modified from Rubio *et al.* 2001
Modified A-Reverse	VBCKRTAACYCGAYCGAYARACGRGC		
PM50	ACTAACTTTAATTCGAACA	p23 gene	Sambade *et al.* 2003
PM51	AACTTATTCCGTCCACTTC		

Table B.2: Primers designed to amplify the complete CTV genome in overlapping fragments. IUPAC nucleotide code used, where Y=C/T, R=A/G, K=G/T, W=A/T, V=A/C/G and B=C/G/T. * Alignment of 45 CTV genomes used for primer design in Chapter 3.

Name	Short Name	Expected Fragment Length	Location on Aligned Sequence*	Sequence (5' to 3')
CTV Primer Pair 1 Fwd	CTV 1 F	1940	(17397,17414)	ATCTCGTCRCTTTGTTTA
CTV Primer Pair 1 Rev	CTV 1 R		(19354,19372)	CTTCTTTGGTTCACRCATA
CTV Primer Pair 2 Fwd	CTV 2 F	2099	(15709,15727)	AYAARACGAAARCGGARGA
CTV Primer Pair 2 Rev	CTV 2 R		(17790,17807)	AAGGYAARAGCGAWGGRA
CTV Primer Pair 3 Fwd	CTV 3 F	2181	(13922,13939)	GGGGAAGYGATTTGGAAA
CTV Primer Pair 3 Rev	CTV 3 R		(16085,16102)	GACGCTCGAAGRATRATR
CTV Primer Pair 4 Fwd	CTV 4 F	2483	(11871,11888)	GTTTGYGTTTTAGTRGTK
CTV Primer Pair 4 Rev	CTV 4 R		(14334,14353)	RGTTTTGTAAGYWTCTATYT
CTV Primer Pair 5 Fwd	CTV 5 F	2455	(9782,9799)	BGYGGARGARCARATYWC
CTV Primer Pair 5 Rev	CTV 5 R		(12219,12236)	AACTCWGARGRYGYAGCC
CTV Primer Pair 6 Fwd	CTV 6 F	2195	(8843,8860)	TYGGRAARTCRAGYGTRT
CTV Primer Pair 6 Rev	CTV 6 R		(11020,11037)	TCGCGRARGCAAACATCY
CTV Primer Pair 7 Fwd	CTV 7 F	2331	(6849,6866)	YRGTCCWTYRTTYCCRCC
CTV Primer Pair 7 Rev	CTV 7 R		(9164,9179)	GWMGCRGTMTYRTCGT
CTV Primer Pair 8 Fwd	CTV 8 F	1791	(5657,5673)	RTTYCYTCVCTWCCYAY
CTV Primer Pair 8 Rev	CTV 8 R		(7432,7447)	RRYAYCTYWTSGGSAR
CTV Primer Pair 9 Fwd	CTV 9 F	2304	(3773,3788)	YGGYGARYTGYTGGAY
CTV Primer Pair 9 Rev	CTV 9 R		(6061,6076)	SSCRTTHCCACGAAGA
CTV Primer Pair 10 Fwd	CTV 10 F	2489	(1488,1503)	GGGRCYTACACHTTTG
CTV Primer Pair 10 Rev	CTV 10 R		(3961,3976)	CTNARYGRHACCATYT
CTV Primer Pair 11 Fwd	CTV 11 F	1913	(12, 27)	MAAATTCAHCYKBWYY
CTV Primer Pair 11 Rev	CTV 11 R		(1910, 1925)	RCACAWVTCRTCRAAR
CTV Primer Pair 6.1 Fwd	CTV 6.1 F	2085	(8672,8687)	TYCACCGRAAYGAYYT
CTV Primer Pair 6.1 Rev	CTV 6.1 R		(10742,10757)	ACRTACCAACCYCTRA
CTV Primer Pair 7.1 Fwd	CTV 7.1 F	2161	(7013,7028)	AYTTYGARCARATSGR
CTV Primer Pair 7 Rev	CTV 7 R		(9164,9179)	GWMGCRGTMTYRTCGT
CTV Primer Pair 8.1 Fwd	CTV 8.1 F	2316	(5410,5428)	CTATGGARRTRGGRWCRAA
CTV Primer Pair 8.1 Rev	CTV 8.1 R		(7710,7725)	WCATMARYGRRGCYYT
CTV Primer Pair 9.1 Fwd	CTV 9.1 F	2326	(3290,3305)	GAGTGRGARYCARCAR
CTV Primer Pair 9.1 Rev	CTV 9.1 R		(5600,5615)	CDGARAAYARRGAHGA
CTV Primer Pair 10 Fwd	CTV 10 F	1873	(1488,1503)	GGGRCYTACACHTTTG
CTV Primer Pair 10.1 Rev	CTV 10.1 R		(3767,3782)	RTCCARCARYTCRCCR
CTV Primer Pair 11.1 Fwd	CTV 11.1 F	1872	(54,69)	RGGAHYYGGWRTARRT
CTV Primer Pair 11 Rev	CTV 11 R		(1910, 1925)	RCACAWVTCRTCRAAR

Table B.3: Generic and genotype specific primers for CTV detection.

Primer	Nucleotide Sequence (5' to 3')	Product Size	Reference
Generic F	ATGGACGACGARACAAAGAAATTGAAGA	672	Roy *et al.* 2010
Generic R	TCAACGTGTGTTRAATTTCCCAAGCT		
T36 F	TTCCCTAGGTCGGATCCCGAGTATA	836	
T36 R	CAAACCGGGAAGTGACACACTTGTTA		
B165 F	GTTAAGAAGGATCACCATCTTGACGTTGA	510	
B165 R	AAAATGCACTGTAACAAGACCCGACTC		
T3 F	GTTATCACGCCTAAAGTTTGGTACCACT	409	
T3 R	CATGACATCGAAGATAGCCGAAGC		
VT F	TTTGAAAATGGTGATGATTTCGCCGTCA	302	
VT R	GACACCGGAACTGCYTGAACAGAAT		
T30 F	TGTTGCGAAACTAGTTGACCCTACTG	206	
T30 R	TAGTGGGCAGAGTGCCAAAAGAGAT		
NZRB1 F	AGTGGTGGAGATTACGTTG	646	Cook, *unpublished*
NZRB1 R	TACACGCGACAAATCGAG		
NZRB2 F	CGGAAGGGACTACGTGGT	662	
NZRB2 R	CGTTTGCACGGGTTCAATG		
HA16-5 F	TAGGAAGGGTCACTGCCCTGACA	658	
HA 16-5 R	GTAAGTATCTAAAACCAGGAG		

Table B.4: Primer mixes used for random cDNA synthesis and random PCR. For each primer mix, 10 µl of the 100 µM stock was combined to get an 8.3 mM final concentration of each primer. IUPAC nucleotide code used, where Y=C/T, R=A/G, K=G/T, W=A/T, V=A/C/G, B=C/G/T, N=A/G/C/T. * Alignment of 45 CTV genomes used for primer design in Chapter 3.

	Names of primers used	Short Name	Location on Aligned Sequence	Sequence (5' to 3')
Forward primer mix	CTV Primer Pair 1 Fwd	CTV 1 F	(17397,17414)	ATCTCGTCRCTTTGTTTA
	CTV Primer Pair 2 Fwd	CTV 2 F	(15709,15727)	AYAARACGAAARCGGARGA
	CTV Primer Pair 3 Fwd	CTV 3 F	(13922,13939)	GGGGAAGYGATTTGGAAA
	CTV Primer Pair 4 Fwd	CTV 4 F	(11871,11888)	GTTTGYGTTTTAGTRGTK
	CTV Primer Pair 5 Fwd	CTV 5 F	(9782,9799)	BGYGGARGARCARATYWC
	CTV Primer Pair 6.1 Fwd	CTV 6.1 F	(8672,8687)	TYCACCGRAAYGAYYT
	CTV Primer Pair 7.1 Fwd	CTV 7.1 F	(7013,7028)	AYTTYGARCARATSGR
	CTV Primer Pair 8.1 Fwd	CTV 8.1 F	(5410,5428)	CTATGGARRTRGGRWCRAA
	CTV Primer Pair 9.1 Fwd	CTV 9.1 F	(3290,3305)	GAGTGRGARYCARCAR
	CTV Primer Pair 10 Fwd	CTV 10 F	(1488,1503)	GGGRCYTACACHTTTG
	CTV Primer Pair 11.1 Fwd	CTV 11.1 F	(54,69)	RGGAHYYGGWRTARRT
	Random hexamers	$(N)_6$	n/a	NNNNNN
Reverse primer mix	CTV Primer Pair 1 Rev	CTV 1 R	(19354,19372)	CTTCTTTGGTTCACRCATA
	CTV Primer Pair 2 Rev	CTV 2 R	(17790,17807)	AAGGYAARAGCGAWGGRA
	CTV Primer Pair 3 Rev	CTV 3 R	(16085,16102)	GACGCTCGAAGRATRATR
	CTV Primer Pair 4 Rev	CTV 4 R	(14334,14353)	RGTTTTGTAAGYWTCTATYT
	CTV Primer Pair 5 Rev	CTV 5 R	(12219,12236)	AACTCWGARGRYGYAGCC
	CTV Primer Pair 6.1 Rev	CTV 6.1 R	(10742,10757)	ACRTACCAACCYCTRA
	CTV Primer Pair 7 Rev	CTV 7 R	(9164,9179)	GWMGCRGTMTYRTCGT
	CTV Primer Pair 8.1 Rev	CTV 8.1 R	(7710,7725)	WCATMARYGRRGCYYT
	CTV Primer Pair 9.1 Rev	CTV 9.1 R	(5600,5615)	CDGARAAYARRGAHGA
	CTV Primer Pair 10.1 Rev	CTV 10.1 R	(3767,3782)	RTCCARCARYTCRCCR
	CTV Primer Pair 11 Rev	CTV 11 R	(1910, 1925)	RCACAWVTCRTCRAAR
	Random hexamers	$(N)_6$	n/a	NNNNNN

Table B.5: Primers used to test for the presence of genome fragments of CTV in randomly amplified cDNA.

Primer	Nucleotide Sequence (5' to 3')	Target area	Reference
Generic F	ATGGACGACGARACAAAGAAATTGAAGA	Within ORF1a	Roy *et al.* 2010
Generic R	TCAACGTGTGTTRAATTTCCCAAGCT		
PM50	ACTAACTTTAATTCGAACA	p23 gene	Sambade *et al.* 2003
PM51	AACTTATTCCGTCCACTTC		
3'CTV VT	CGATATGCGGCCGCTGGACCTATGTTGGCCCC	p23 gene and 3'UTR	Read, *unpublished*
3'CTV VT inner	GTGTGCGTGGATTGTGGTAG		
p33 Univ F	GATGTTTGCCTTCGCGAGC	p33 gene	
1kb Univ-p33-R	CCCGTTTAAACAGAGTCAAACGG		
VT-1(-)	GGTAGGGTCTACTCGTTTCAT	Within ORF1a	Steward, 2006
VT-1(+)	GTACCCTCCGGAAATCACG		
VT-2(-)	TTCGCCTGCGCAGCTGCT	Within ORF1a	
VT-2(+)	AACGGCATGGTGCTCTCCTGTT		
VT-3(-)	AGAATCAGGCAAACGCCC	Within ORF1a	
VT-3(+)	CAGGTGAGAATTCTCCATCGT		
VT-4(-)	GGGTCGCCTGAAAAGGCCCGT	Within ORF1a	
VT-4(+)	ATCTACTGTTTTAACAATTCACG		
VT-5(-)	TTCGCGCTACACGTGTTA	Within ORF1a	
VT-5(+)	TGGTTTGCGGGCCGGGCG		
VT-6(+)	ACGCGTTTCGGTATGTTGTA	Within ORF1a	
VT-6(-)	TGGAACAAAAGGACTATC		

Table B.6: Primers and adapters used for adapter-ligated PCR.

Oligo name	Nucleotide Sequence (5' to 3')
3'Adapter	GATGCCACGAGGTCGACAGTGC Modification: 5' Phosphorylation / 3' C3 Spacer
5' Adapter	GTACACTCGCAGCAGGCTCACG
3' A Primer	GCACTGTCGACCTCGTGCATC

References:

Roy, A., Ananthakrishnan, G., Hartung, J.S., and Brlansky, R.H. 2010. Development and Application of a Multiplex Reverse-Transcription Polymerase Chain Reaction Assay for Screening a Global Collection of *Citrus tristeza virus* Isolates. *Phytopathology* 100(10), pp. 1077-1088.

Rubio, L., Ayllon, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., and Falk, B.W. 2001. Genetic Variation of *Citrus tristeza virus* Isolates from California and Spain: Evidence for Mixed Infections and Recombination. *The Journal of Virology* 75(17), pp. 8054-8062.

Sambade, A., López, C., Rubio, L., Flores, R., Guerri, J., and Moreno, P. 2003. Polymorphism of a Specific Region in Gene p23 of *Citrus tristeza virus* Allows Discrimination between Mild and Severe Isolates. *Archives of Virology* 148(12), pp. 2325-2340.

Stewart, K.A. 2006. Strain Differentiation of *Citrus tristeza virus* Isolates from South Africa by PCR and Microarray. Dissertation submitted to the Faculty of Natural and Agricultural Sciences Department of Microbiology and Plant Pathology in partial fulfillment of the requirements for the degree Magister Scientiae (MSc) Microbiology. University of Pretoria.

Appendix C: Additional resources