The detection of *Citrus tristeza virus* genetic variants using pathogen specific electronic probes

Tracey Jooste

Thesis presented in partial fulfilment of the requirements of the degree of Master of Science in the Faculty of AgriSciences at Stellenbosch University

Supervisor: Dr H. J. Maree

Co-supervisors: Dr M. Visser and Prof J. T. Burger

March 2017

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2017

Abstract

Citrus tristeza virus (CTV), a complex pathogen of citrus spp., is endemic to South Africa and has been responsible for great losses locally and internationally. CTV causes severe stem pitting in grapefruit, which forms an important sector of South Africa's citrus production and export market. The limited understanding of CTV's ability to cause severe disease in one host while no symptoms in another restricts the implementation of effective management strategies. The conservation of plant biosecurity relies on the rapid identification of pathogenic organisms including viruses. While there are many molecular assays available for the detection of plant viruses, these are often limited in their ability to test for multiple viruses simultaneously. However, with next-generation sequencing (NGS) based metagenomic analysis it is possible to detect multiple viruses within a sample, including low-titre and novel viruses, at the same time. Conventional NGS data analysis has computational limitations during contig assembly and similarity searches in sequence databases, which prolongs the time required for a diagnostic result. In this study, an alternative targeted method was explored for the simultaneous detection of 11 recognised citrus viruses in NGS data using electronic probes (e-probes). E-probes were designed, optimised and screened against raw, unassembled NGS data in order to minimise the bioinformatic processing time required. The e-probes were able to accurately detect their cognate viruses in simulated datasets, without any false negatives or positives. The efficiency of the e-probe based approach was validated with NGS datasets generated from different RNA preparations: dsRNA from 'Mexican' lime infected with different CTV genotypes, dsRNA from field samples, as well as small RNA and total RNA from grapefruit infected with the CTV T3 genotype. A set of probes were made publically available that is able to accurately detect CTV in NGS data irrespective of which genotype the plants are infected with. The results were confirmed by performing *de novo* assemblies of the high quality read datasets and subsequent BLAST analyses. This sequence based detection method eliminates the need for NGS data assembly, ultimately reducing the virus-detection turnaround time.

Opsomming

Citrus tristeza virus (CTV), 'n komplekse patogeen van sitrusspesies, is endemies aan Suid-Afrika en doen verantwoording vir groot verlies op beide plaaslike sowel as internasionale vlak. CTV veroorsaak terselfdertyd noodlottige en ernstige stam-uitputting in pomelo's, wat 'n belangrike sektor van Suid-Afrika se sitrusproduksie- en uitvoermark vorm. Die beperkte begrip van CTV se vermoë om ernstige siektetoestande in een gasheer te veroorsaak, terwyl geen simptome in ander gashere voordoen nie, beperk die implementering van effektiewe bestuurstrategieë. Die behoud van plant-biosekuriteit maak staat op die spoedige identifisering van patogeniese organismes, met virusse daarby ingesluit. Terwyl daar menigte molekulêre toetse vir die opsporing van plantvirusse beskikbaar is, blyk dit dat hierdie juiste toetse dikwels beperkte vermoë toon om gelyktydig vir veelvuldige virusse te toets. Nietemin, met volgende-generasie volgordebepaling (NGS) gebaseerde metagenomiese analise, is dit moontlik om veelvuldige virusse terselfdertyd binne 'n monster op te spoor, insluitend lae titer- en onbekende virusse. Konvensionele NGS data analise beskik oor rekenaar beperkinge tydens die samestelling van "contigs" sowel as ooreenkoms soektogte in volgorde gebaseerde databasisse, wat gevolglik die tyd wat versoek word vir 'n diagnostiese resultaat, verleng. In hierdie studie word 'n alternatiewe geteikende metode ondersoek vir die gelyktydige opsporing van 11 sitrus virusse in NGS data deur die gebruik van elektroniese probes, bekend as "e-probes". Hierdie "e-probes" was ontwerp, optimaliseer en gekeur binne onverwerkte NGS data om sodoende die bioinformatiese prosesseringstyd wat vereis word, te minimaliseer. Die "e-probes" was in staat om hul verwante virusse in gesimuleerde datastelle akkuraat op te spoor, sonder enige onwaar negatiewes of positiewes. Die doeltreffendheid van die "e-probe" gebaseerde benadering was bekragtig deur NGS datastelle wat versamel is vanuit verskillende RNS voorbereidings: dsRNS vanuit 'Meksikaanse' lemmetjie besmet met verskillende CTV genotipes; dsRNS vanuit veld-monsters sowel as sRNS en RNS vanuit pomelo's besmet met die CTV T3 genotipe. 'n Stel van elektroniese probes was binne openbare domeine beskikbaar gestel wat in staat is om CTV in NGS data akkuraat op te spoor, ongeag van watter genotipe die plante mee besmet is. Die resultate was bevestig deur "*de novo*" samestellings van hoogstaande kwaliteit datastelle en daaropvolgende BLAST-analises uit te voer. Hierdie volgorde-gebaseerde opsporingsmetode elimineer die behoefte vir *de novo* samestellings van NGS data en verminder gevolglik die virusopsporing-omkeertyd.

Acknowledgements

I would like to extend my sincere gratitude and appreciation towards the following people for their innumerable support and contributions during this study:

- Prof. Johan T. Burger, for his supervision, direction and support throughout the study, and for allowing me to form part of his exceptional research group.

- My supervisor Dr. Hans J. Maree, for outstanding leadership, intellectual input, recommendations, and his continued motivation, patience and guidance.

- Dr Marike Visser for her contributions towards the study, assistance with the bioinformatics, critical reading of manuscripts, and the pleasant team work.

- Glynnis Cook for establishing the plant material used in this study, suggestions, and for allowing and assisting in sample collections.

- Dr. Rachelle Bester for assistance with data analysis and the generation of graphics for manuscripts.

- The Agricultural Research Council (ARC) for their financial support.

- Citrus Research International (CRI), for research funding and sample material.

- THRIP, for research funding.

- The financial assistance of the National Research Foundation (NRF) towards the research is hereby acknowledged. Opinions expressed and conclusions drawn, are those of the authors and not necessarily to be attributed to the NRF.

- Stellenbosch University, for providing the resources to complete this research.

- Lundi Korkie, for her ongoing support, encouragement and friendship.

- Colleagues and friends in the Vitis laboratory (SU), for creating an entertaining and supportive working environment.

- My parents, sister, son and friends for endless love, support and reassurance.

- My Heavenly Father.

Dedicated to my beloved son.

Table of contents

List of figures

List of tables

List of abbreviations

aa	Amino acid
ABI	Applied Biosystems
ARC-BP	Agricultural Research Council Biotechnology Platform
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BWT	Burrows-Wheeler Transform
cDNA	Complementary DNA
CiLRV	Citrus leaf rugose virus
CiLV-C	Citrus leprosis virus, cytoplasmic type
CLBV	Citrus leaf blotch virus
CP	Coat protein
CPm	Minor capsid protein
CPsV	Citrus psorosis virus
CRI	Citrus Research International
CTAB	Cetyltrimethylammonium bromide
CTLV	Citrus tatter leaf virus
CTV	*Citrus tristeza virus*
cv	Cultivar
CVV	Citrus variegation virus
CYMV	Citrus yellow mosaic virus
CYVCV	Citrus yellow vein-clearing virus
DAS-ELISA	Double antibody sandwich ELISA
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
DOI	Digital object identifier
dsRNA	Double-stranded RNA
EDNA	E-probe diagnostic nucleic acid analysis
EDTA	Ethylenediamine tetra-acetic acid
ELISA	Enzyme-linked immunosorbent assay
EtBr	Ethidium Bromide
EtOH	Ethanol
GB	Gigabase
HRM	High-resolution melt

HSP70h	Heat shock protein 70 homologue
ICRSV	Indian citrus ringspot virus
ICTV	International Committee on Taxonomy of Viruses
IF	Immunofluorescence
kb	Kilobase
mRNA	Messenger RNA
MSD	Mock sequence dataset
NCBI	National Centre for Biotechnology Information
NGS	Next-generation sequencing
NRF	National Research Foundation
nt	Nucleotide
nts	Nucleotides
OD	Optical density
ORF	Open reading frame
OS	Operating system
PCR	Polymerase Chain Reaction
PVP	Polyvinylpyrrolidone
qPCR	Quantitative PCR
RAM	Random access memory
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RO-H$_2$O	Reverse Osmosis water
RT-PCR	Reverse-transcription Polymerase Chain Reaction
SDS	Sodium dodecyl sulfate
SDV	Satsuma dwarf virus
sgRNA	Sub-genomic RNA
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
spp	Species
SSCP	Single-stranded conformation polymorphism
ssRNA	Single-stranded RNA
STE	Sodium/ Tris/ EDTA
SU	Stellenbosch University
TAE	Tris-acetic acid/ EDTA
Taq	*Thermus aquaticus* DNA polymerase
Tris-HCL	2-amino-2-(hydroxymethyl)-1,3-propanediol-hydrochloric acid

| U | Unit |
| UTR | Untranslated region |

Chapter 1: Introduction

1.1 General Introduction

The international trade market has facilitated the enormous gain in the citrus industry and in 2015, has grown in excess of 100 million tonnes of citrus. This was generated by the top 15 producing countries which is inclusive of South Africa (http://www.cga.co.za). Exporting to over 60 countries worldwide, South Africa has been the second largest exporter of fresh citrus since 2006, while it is ranked 11th on the production list. South Africa is the third largest producer of grapefruit and the largest exporter (http://www.indexmundi.com).The grapefruit sector of the native industry is therefore largely focussed on cultivation for commercial processing (39.8%) and subsequent export (59.2%), with a mere 1% of the produce being sold locally (http://www.citrusresourcewarehouse.org.za). Losses experienced because of plant pathogens leads to a reduction in yield and cultivar sustainability, which ultimately places the industry under tremendous threat.

One of the most devastating and complex viral pathogens of citrus species locally and worldwide is the closterovirus, *Citrus tristeza virus* (CTV) (Bar-Joseph *et al.* 1989; Moreno *et al.* 2008). *Citrus tristeza virus* is endemic in southern Africa and has been responsible for great losses by causing a disease called "tristeza" or quick decline when citrus cultivars are established on sour orange rootstocks (McClean, 1956; Moreno *et al.* 2008). The South African citrus industry experienced major constraints with this rootstock since the initiation of the industry and moved to use of CTV tolerant rootstocks such as rough lemon and later also trifoliate hybrid types. Despite the local industry's use of less sensitive rootstocks, the virus is still a limiting factor in the production of sensitive citrus types such as grapefruit. Our limited understanding of how CTV can cause severe disease in one host species and cause no symptom expression in another, complicates the implementation of effective management strategies.

To minimize losses in the local citrus industry due to CTV, the South African Citrus Improvement Scheme (CIS) implemented cross-protection using mild CTV sources to reduce the effect of challenges by endemic severe CTV strains. However, cases of breakdown within the strategy have occurred, therefore driving the need for a better understanding of viral sources to address this protection breakdown in grapefruit specifically. With current scientific research being majorly focused on targeting the spread of Huanglongbing (HLB) disease, also known as citrus greening, an even larger opening in

1

research aiming to give insight to virus interactions within plants infected with CTV has been created.

With the latest advances in molecular technology, it is now possible to study the mechanisms of cross-protection in more detail by aiming to elucidate CTV interactions. The aforementioned also allows for the assessment of recombination occurring between CTV genotypes as well as whether this influences the exclusion mechanism proposed in literature. With the full comprehension of the latter, it will be possible to pre-inoculate trees with specific or combinations of CTV genotypes that will confer tolerance to a wider range of virulent CTV genotypes.

1.2 Aims and Objectives

The aim of this study was firstly to detect different CTV genotypes in 'Mexican' lime and grapefruit using a metagenomic, high-throughput next-generation sequencing (NGS) approach. Secondly, to explore an alternative bioinformatic approach for the detection of CTV in NGS data, using virus-specific e-probes. In order to achieve these aims, the following objectives were set out:

- To extract high quality double-stranded RNA and total RNA from citrus plants infected with different CTV isolates.
- To submit the extracted RNA to NGS.
- To characterise CTV source plants through conventional bioinformatic analyses of the metagenomic NGS data, which include read mapping and *de novo* assemblies.
- To design specific e-probes for CTV detection according to the bioinformatic pipeline, EDNA (E-probe Diagnostic Nucleic acid Analysis).
- To evaluate the e-probe based detection system for CTV across different sequence library types.
- To design e-probes for the detection of 10 additional citrus-infecting viruses of economic importance and assess their effectiveness with simulated sequence data.

1.3 Research Outputs

This study contributed to the following publications, conference proceedings and poster presentations.

1.3.1 Publications.

- **Jooste, T. L**., Visser, M., Cook, G., Burger, J. T., and Maree, H. J. *In silico* probe-based detection of citrus viruses in NGS data. Phytopathology. *Under review.*

  The bioinformatic pipeline described in this study was submitted to Phytopathology.

1.3.2 Conference proceedings

- **Jooste, T.L**., Visser, M., Cook, G., Burger, J.T., and Maree, H.J. Citrus virus detection in NGS data using e-probes.

  Presentation on the bioinformatic pipeline described in this study delivered by Dr. H.J. Maree at the 20th Conference of the International Organization of Citrus Virologists (IOCV). Chongqing, China. 10-15 April 2016.

- **Jooste, T.L**., Visser, M., Cook, G., Burger, J.T., and Maree, H.J. 2016. Detection of citrus viruses in next-generation sequencing data using e-probes.

  Dr. H.J. Maree presented the bioinformatic pipeline described in this study at the 9th Citrus research symposium, South Africa, 21-24 August 2016.

1.3.3 Posters

- **Jooste, T.L**., Visser, M., Cook, G., Burger, J.T., and H.J. Maree. Detection and differentiation of CTV genetic variants using metagenomic next-generation sequencing.

  The bioinformatic pipeline described in this study contributed to a poster presented by Ms T.L. Jooste at Virology Africa, Cape Town, South Africa. 30 November to 3 December 2015.

1.4 References

Bar-Joseph, M., Marcus, R., Lee, R. F. 1989. The continuous challenge of *citrus tristeza virus* control. Annu. Rev. Phytopathol. 27:291-316.

McClean, A. P. D. 1956. Tristeza and stem-pitting diseases of citrus in South Africa. FAO Plant Protection Bulletin. 88-94.

Moreno, P., Ambross, S., Albiach-Marti, M. R., Guerri, J., Pena, E. 2008. *Citrus tristeza virus*: A pathogen that changed the course of the citrus industry. Mol. Plant Pathol. 9:251-268.

1.5 Internet sources

Citrus Growers' Association of Southern Africa (CGA). 2016. Annual report. [Online]. Available: http://www.cga.co.za/Page.aspx?ID=3207 [2016, November 25].

Citrus Growers' Association of Southern Africa (CGA). 2016. Key Industry Statistics for Citrus Growers. [Online]. Available: http://www.citrusresourcewarehouse.org.za/home/document-home/information/cga-key-industry-statistics/3610-cga-key-industry-statistics-2016/file [2016, November 25].

U.S Department of Agriculture (USDA). Commodity production by country. 2015. [Online]. Available: http://www.indexmundi.com/agriculture/. [2016, November 25].

Chapter 2: Literature Review

2.1 Introduction

As fruit bearing, woody plants in the family *Rutaceae*, members of the genus *Citrus* are important commodities when it comes to international trade. The production of natural and hybrid citrus cultivars have increased significantly over the past decades, with commercial varieties such as grapefruit, lemons, limes, oranges and tangerines grown in over 140 countries worldwide. Consumer preferences, widespread cultivation, a decrease in storage related diseases, and subsequent product affordability are some of the factors contributing to this increase.

While Brazil is one of the world's leading citrus producing countries, countries in the northern hemisphere are responsible for producing more than 70% of the world's citrus (FAO 2016). The consumption of fresh citrus fruit has been known to aid in improving consumer health as it serves as a source of carbohydrates and various other nutrients in urbanised and rural areas. However, a third of the world's citrus industry is focussed on cultivation for commercial processing and subsequent export (FAO 2016). The latter involves the use of citrus fruit for the production of juice, dried and canned products, oils as well as flavouring agents. Orange juice specifically, accounts for over 80% of all processed citrus products, and is merchandised globally as frozen concentrates, in order to reduce transport and storage cost (FAO 2016).

Like many other economically important fruit crops, citrus is susceptible to various pathogens and pests that threaten citrus industries globally. These pathogens belong to multiple taxonomical groups and often lead to the occurrence of diseases that negatively affect the plant's productivity and/or crop characteristics. Syndromes resulting from viral infections specifically can be extremely severe as they are generally irremediable and management strategies mostly rely on early detection and subsequent eradication of the infected plant material. This threat serves as the driving force behind the magnitude of research devoted to plant virus studies in order to prevent the occurrence and spread of diseases.

Amongst the viral pathogens of citrus cultivars, infection with the largest member from the family *Closteroviridae*, *Citrus tristeza virus* (CTV), remains the most detrimental (Karasev 2000, Dolja *et al.* 2006, Moreno *et al.* 2008). The severe symptoms induced by this RNA

virus translates into massive economic losses for citrus production in many parts of the world, including South Africa (McClean, 1956; Moreno *et al.* 2008). Other devastating viruses include *Citrus yellow vein clearing virus* (CYVCV), *Citrus psorosis virus* (CPsV), and *Indian citrus ringspot virus* (ICRSV) amongst others (Sharma *et al.* 2007, Loconsole *et al.* 2012, Moreno *et al.* 2015). These viruses are spread across distinct families and cause citrus degeneration through leaf bleaching, bark scaling or graft union incompatibility.

The recent advances in next-generation sequencing (NGS) technologies provide an unbiased, powerful approach for plant virus detection that is sensitive enough to identify novel viruses as well as divergent variants of known viruses. Furthermore, coupling NGS to metagenomics, as proposed in this study, allows the user to establish a complete profile of all viruses within a given sample, in a manner that is less time consuming than conventional techniques. Although it is still too expensive to use for routine virus detection, the use of NGS has to date enabled a deeper understanding of viral biodiversity and consequent disease etiology (Beerenwinkel and Zagordi 2011, MacDiarmid *et al.* 2013).

2.2 Citrus tristeza virus (CTV)

2.2.1 Taxonomy

As the causative agent of a variety of damaging syndromes in citrus, CTV is one of eleven species in the genus *Closterovirus* (family *Closteroviridae*) (Karasev 2000, Martelli *et al.* 2002, Folimonova *et al.* 2010). This family of viruses has grown since its establishment and currently consist of four genera that are characterised based on viral genome type and the type of vector used for transmission (Martelli *et al.* 2002, Al Rwahnih *et al.* 2012). The first two genera, *Ampelovirus* and *Closterovirus* encompass monopartite viruses, whereas those with bipartite genomes are included in the *Crinivirus* genus (Martelli *et al.* 2012). Insect vectors such as mealybugs, aphids, and whiteflies, respectively generally transmit members of the previously listed genera (Dolja *et al.* 1994, Dolja *et al.* 2006, Folimonova *et al.* 2010). The most recent genus added to the family *Closteroviridae*, *Velavirus* is made up of members for which no vectors are known yet (Al Rwahnih *et al.* 2012, Melzer *et al.* 2013).

2.2.2 Morphology and genome organisation

Viruses that are included in the genus *Closterovirus*, have capillaceous particles that are flexuous in nature, ranging in length from 1,250 to 2,200 nm (Agranovsky *et al.* 1995, Martelli *et al.* 2002). The polar, non-enveloped particles of CTV specifically are 2,000 nm long and

comprise of two capsid proteins (CP) that display helical symmetry (Bar-Joseph *et al.* 1972, Agranovsky *et al.* 1995, Tian *et al.* 1999). Closteroviruses have a linear genome that consists of a single-stranded (ss), positive sense RNA (Martelli *et al.* 2002). With a genome of approximately 19.3 kb, CTV is the largest RNA virus known to infect plants (Pappu *et al.* 1994, Karasev *et al.* 1995, Bar-Joseph *et al.* 2002).
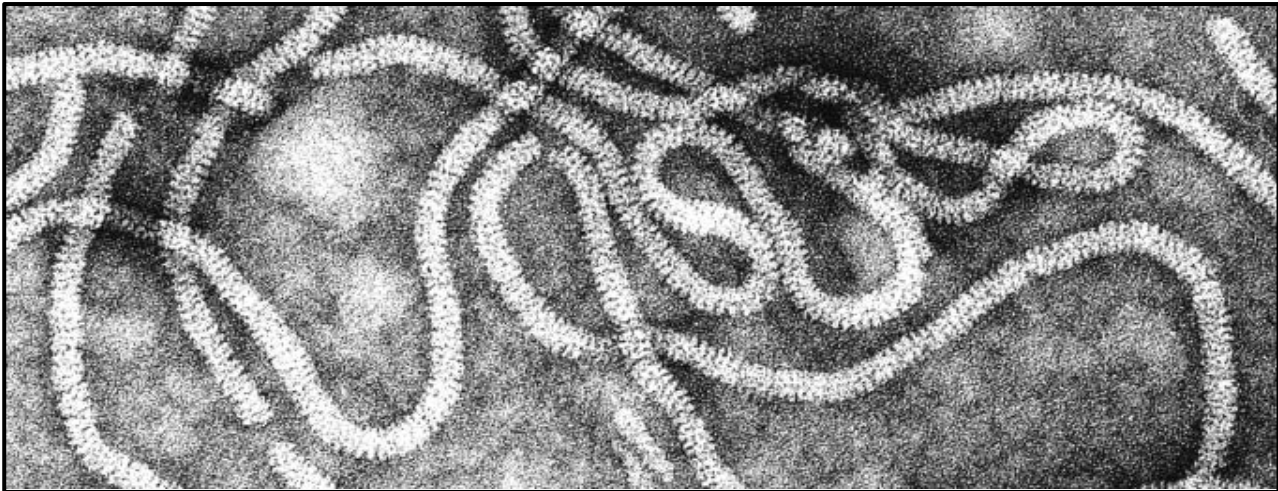
Figure 2.1. Negative contrast electron micrograph of virions of citrus tristeza virus (CTV) (Agranovsky 2013).

The nucleotide sequence of several CTV isolates revealed that the RNA genome is arranged into two untranslated regions (UTRs), one at each terminus, which encloses 11 open reading frames (ORFs) that encode at least 17 proteins (Figure 2.2) (Karasev *et al.* 1995, Vives *et al.* 1999, Yang *et al.* 1999, Flores *et al.* 2013). The two ORFs on the 5′ end of the genome, ORFs 1a and 1b, are expressed from genomic RNA and encode proteins that make up the replicase complex (Karasev *et al.* 2005, Dolja *et al.* 2006, Moreno *et al.* 2008, Melzer *et al.* 2010). The large polyprotein (approximately 349 kDA), encoded by ORF 1a, is composed of a helicase-like, a methyltransferase-like, and two papain-like protease conserved domains (Karasev *et al.* 2005, Folimonova *et al.* 2010). Open reading frame 1b on the other hand, encodes an RNA-dependent RNA polymerase that is translated via a +1 frameshift as the first nucleotides of this open reading frame overlaps with ORF 1a (Folimonova *et al.* 2010, Folimonova *et al.* 2013, Harper *et al.* 2013). The 5′ and 3′ untranslated regions along with ORFs 1a and 1b are the only essential components for virus replication. The 3′ half of the genome, consisting of the remaining 10 ORFs, is expressed by subgenomic (sg) RNAs, and encodes additional proteins that are involved in viral particle construction and movement (Pappu *et al.* 1994, Dolja *et al.* 2006, Moreno *et al.* 2008). Amongst the aforementioned proteins are the major and minor coat proteins (CP and CPm), as well as a heat shock protein HSP70 homolog (p65) which is conserved amongst viruses in the family

Closteroviridae (Satyanarayana *et al.* 2000). Other ORFs also encode RNA silencing suppressor proteins (p20, p23 and p25) (Lu *et al.* 2004, Cheng *et al.* 2015); and a small hydrophobic transmembrane peptide (p6) whose homolog in *Beet yellow virus* (BYV) has been reported to be involved in virus movement (Dolja *et al.* 2006, Tatineni *et al.* 2008). Oddly, most of the trees infected with CTV contain mutant RNAs, otherwise known as defective RNAs that comprise of selected segments of the 5′ and 3′ sequences of the viral genome only (Mawassi *et al.* 1996, Tatineni *et al.* 2008).
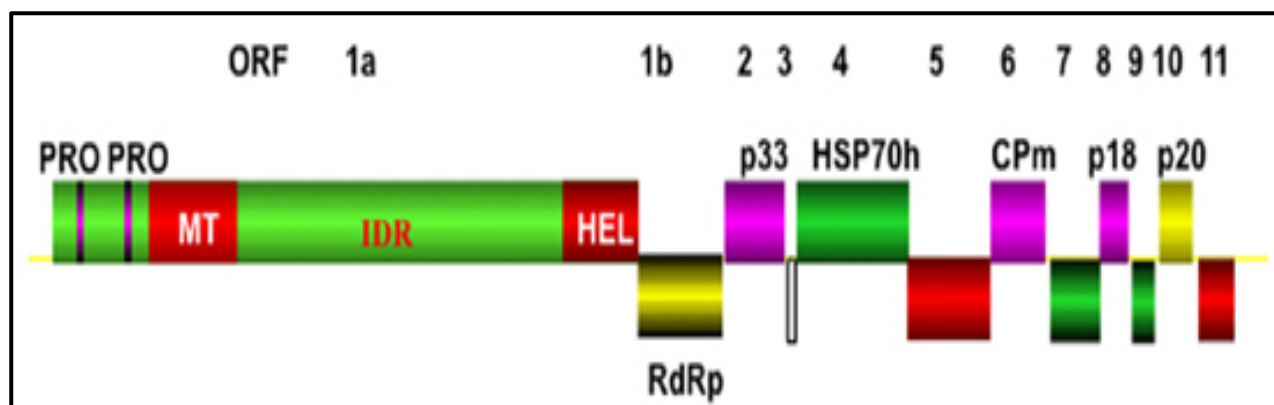
Figure 1.2. Schematic representation of the CTV genome displaying the 11 open reading frames and their corresponding encoded proteins. PRO, papain-like proteases; MT, methyl transferase-like domain; IDR, large interdomain region; HEL, helicase-like domain; RdRp, RNA-dependent RNA polymerase domain; HSP70h, analog to heat shock protein; CPm and CP, minor and major coat proteins (Dawson *et al.* 2013).

2.2.3 Genome variability and genotypes

Variations in symptom severity and aphid transmissibility observed during the first CTV outbreaks suggested the presence of numerous divergent isolates that are biologically and genetically distinct (McClean 1963, Satyanarayana *et al.* 1999, Kong *et al.* 2000, Hilf *et al.* 2005). Earlier studies attempting to resolve the diversity of CTV isolates involved classification based on phenotype as well as the use of monoclonal antibodies (Permar *et al.* 1990, Gillings *et al.* 1993). However, with the introduction of sequencing came the application of techniques with the ability to identify and characterise distinct isolates based on sequence identity (Permar *et al.* 1990, Moreno *et al.* 1990, Pappu *et al.* 1993). These techniques targeted the CP, the p23 protein, the 5′ UTR, and various regions of genomic RNA (ORF 1a/1b) and included restriction fragment length polymorphism (RFLP) analysis and single-strand conformation polymorphism (SSCP) analysis (Permar *et al.* 1990, Gillings *et al.* 1993, Rubio *et al.* 1996, Sambade *et al.* 2002). Phylogenetic analysis initially clustered CTV isolates into three groups namely severe stem pitting (SP) and seedling yellows (SY) inducing isolates; mild non-SP and non-SY isolates; and intermediate isolates (Karasev *et*

8

al. 1995, Mawassi *et al.* 1996, Albiach-Martí *et al.* 2000, Suastika *et al.* 2001). There are currently 60 complete sequences of CTV isolates available that have been categorised into seven distinct genotypic groups or strains, based on sequence evaluations across the entire genome (Folimnova *et al.* 2010, Dawson *et al.* 2013, Harper 2013). The genotypes defined as RB, T3, T30, T36, T68, VT, and the recombinant HA16-5 share an average identity of approximately 85.1% throughout the genome and amino acid identities in the range of 73.4 and 92.1% for ORF 1a specifically (Folimnova *et al.* 2010, Harper 2013). Further research comparing ORF 1a sequences of the different genotypes revealed that nucleotide identities between isolates belonging to the same strain range between 94.2 and 99.4%, and that T3, T30, and VT isolates are more similar (identities ranging from 89.4-90.3%) to each other than those belonging to T36 and T68 strains (Moreno *et al.* 2008, Dawson *et al.* 2013). The fact that CTV isolates are for the most part homologous in the 3′ half of the genome has led to the use of replication genes at the 5′ terminal for standardised genetic differentiation. Factors such as recombination and the occurrence of mixed infections with isolates from multiple genotypes however continue to convolute the classification of newly sequenced isolates.

2.2.4 Viral replication and expression of ORFs

As previously mentioned, CTV RNA is expressed in a manner similar to other positive-stranded RNA viruses through three processes, including the breakdown of proteins into smaller polypeptides; ribosomal frameshifting; and the construction of an array of subgenomic (sg) RNAs (Hilf *et al.* 1995. Gowda *et al.* 2001, Moreno *et al.* 2008). The mode in which the viral genome replicates consists of different phases and is shared amongst other members in the family *Closteroviridae*, such as the type member BYV (Dolja *et al.* 2006). The replication cycle is initiated upon disassembly of the virion in order to expose genomic RNA, and followed by translation of the proteins in the replicase complex. Translation of the remaining proteins occurs secondary to the aforementioned, as they are involved in downstream processes, at a later stage in the cycle. The viral genome is then replicated within cytoplasmic compartments through an RNA-dependent RNA polymerase (RdRp), yielding a double-stranded replicative form (Moreno *et al.* 2008). This double-stranded RNA intermediate contains a negative RNA strand, complementary to the positive viral RNA, and guides the formation of new virions. The expression of the ORFs situated at the 3′ end of the genome is independently controlled and assisted by the translation of sgRNAs. The latter is synthesised only once replication has commenced and differs from

9

genomic RNA in terms of length and 5′ end composition, which comprises deletions of portions of viral RNA (http://www.expasy.org/viralzone/). Subgenomic RNAs generally express structural or movement proteins, and are not considered part of the viral genome since they lack signals required for encapsidation into mature virions (http://www.expasy.org/viralzone/). Individual controller components are responsible for coordinating sgRNA production concerning timing and abundance, as they also interrupt host defence mechanisms by facilitating the translation of RNA silencing suppressors (Navas-Castillo *et al.* 1997, Dolja *et al.* 2005). Upon completion of viral particle assembly, movement proteins are responsible for mediating the cell-to-cell spread of virions throughout the plant.

2.2.5 Symptoms

Evaluations of a broad range of viruses revealed that CTV induced the largest number of recognisable host responses upon infection (Hilf *et al.* 2005, Moreno *et al.* 2008, Dawson *et al.* 2013). These responses are influenced by a combination of host and viral features including the infected citrus variety, the rootstock used for propagation of the variety, as well as the particular strain (or mixture of strains) of CTV (Moreno *et al.* 2008, Harper *et al* 2013, Folimonova 2013). As the viral infection is restricted to phloem tissue, the disease symptoms induced normally correlate with alterations in the structure and function of the phloem (Yokomi 2009, Folimonova *et al.* 2010). Besides for the symptoms used in greenhouse diagnostics such as vein clearing, leaf curling and stunting of young seedlings; CTV causes four major host reactions or syndromes namely: quick decline, stem pitting, seedling yellows, and no symptom expression (Bar-Joseph and Dawson 2008, Moreno *et al.* 2008).

2.2.5.1 Quick decline (QD)

The first disease and historically the most detrimental, CTV-induced decline (*tristeza*), destroys grapefruit, mandarin and sweet orange cultivars grown on sour orange (*Citrus aurantium*) rootstocks (Moreno *et al.* 2008). This man-made disease was established by grafting infected material onto sour orange rootstocks in an attempt to eliminate "root rot" (Dawson *et al.* 2013). The virus causes death of scion cultivars, grafted onto the rootstock by promoting phloem necrosis that renders the bud union incompatible (Figure 2.3 A and B). However, no phenotypic symptoms are observed when sour orange trees are produced using the sour orange rootstock (Garnsey *et al.* 2000, Yokomi 2009). The time required for symptom expression can vary between progressing over a few years to complete tree death within only a number of days post virus infection (quick decline). The devastating impact this

disease has had on citrus industries worldwide, especially citrus growing areas in Florida, resulted in the use of CTV-tolerant rootstocks. These alternate rootstocks were however more prone to root pathogens and performed subpar under certain soil conditions (McClean 1974, Dawson *et al.* 2013). Thus, creating the need for potential control strategies such as cross protection, that would allow growers to utilise the favoured sour orange rootstock without experiencing losses.

2.2.5.2 Stem pitting (SP)

The second syndrome, stem pitting (SP), is induced by selected CTV strains and causes significant problems for the cultivation of commercial citrus cultivars, irrespective of the rootstock used for propagation (Folimonova 2013, Dawson *et al.* 2013). Unlike decline, stem pitting does not lead to tree death; it does however have economic impact by substantially reducing fruit size and yield in sensitive cultivars such as acid lime, grapefruit and sweet orange (Garnsey *et al.* 2005, Yokomo 2009). Stem pitting is characterised by the presence of cavities (indented areas) referred to as pits that can be visualised by removing the bark of the tree (Figure 2.3 C and D). These pits represent areas on the stem where viral replication interfered with cambium differentiation, resulting in disrupted phloem and xylem development (Moreno *et al.* 2008, Tatineni and Dawson 2012, Folimonova 2013). This disease phenotype appears to be a common phenomenon amongst a wide range of virus-infected perennial plants, however the underlying mechanism of stem pit formation is yet to be elucidated. In regions of Australia, Brazil and South Africa, SP strains are endemic and continue to be one of the main factors restricting the production of severely sensitive citrus cultivars. The latter can be overcome by shifting the production focus towards that of varieties that display tolerance toward SP isolates, or by employing a means of "pre-immunisation" using mild CTV strains (Timmer *et al.* 2000, Dawson *et al.* 2013; Folimonova 2013).

2.2.5.3 Seedling yellows (SY)

Numerous CTV isolates induce a "seedling yellows" (SY) reaction that is unique to citrus cultivars such as grapefruit, lemon and sour orange during the seedling stage (Yokomi 2009; Harper 2013). The absence of this reaction in other citrus varieties suggests the involvement of host factors in addition to CTV genomic elements affecting viral pathogenicity (Yokomo 2009, Dawson *et al.* 2013). Symptoms associated with SY range from mild leaf chlorosis and growth reduction (Figure 2.3 E and F), to severely chlorotic (almost white), stunted young leaves after infection and complete cessation of growth (dwarfing) (Moreno *et al.*

2008, Albiach-Marti *et al.* 2010). Plants can occasionally recover from this syndrome and generate a new growth with symptomless leaves (Wallace and Drake 1972; Albiach-Marti *et al.* 2010). The occurrence of seedling yellows in plants has often also been associated with the presence of more severe CTV strains responsible for the formerly mentioned host syndromes (Yokomi 2009). Compared to the two previously mentioned syndromes, seedling yellows is not as abundant or economically important, but it is much easier to employ as a glasshouse assay.

2.2.5.4 No symptoms

The final CTV-induced host response in citrus is the absence of any disease symptoms in nearly all varieties, even those grafted onto rootstocks susceptible to quick decline (QD). This state of equilibrium is observed when the virus evolves with the host, despite of the fact that the virus may be present in high titres. The mild CTV isolates resulting in the aforementioned have been used effectively in cross protection strategies in Florida and South Africa (Dawson *et al.* 2013). However, the asymptomatic nature of host plants infected with these viral isolates poses a new threat by creating an ideal opportunity for the distribution of infected material to other citrus growing areas.
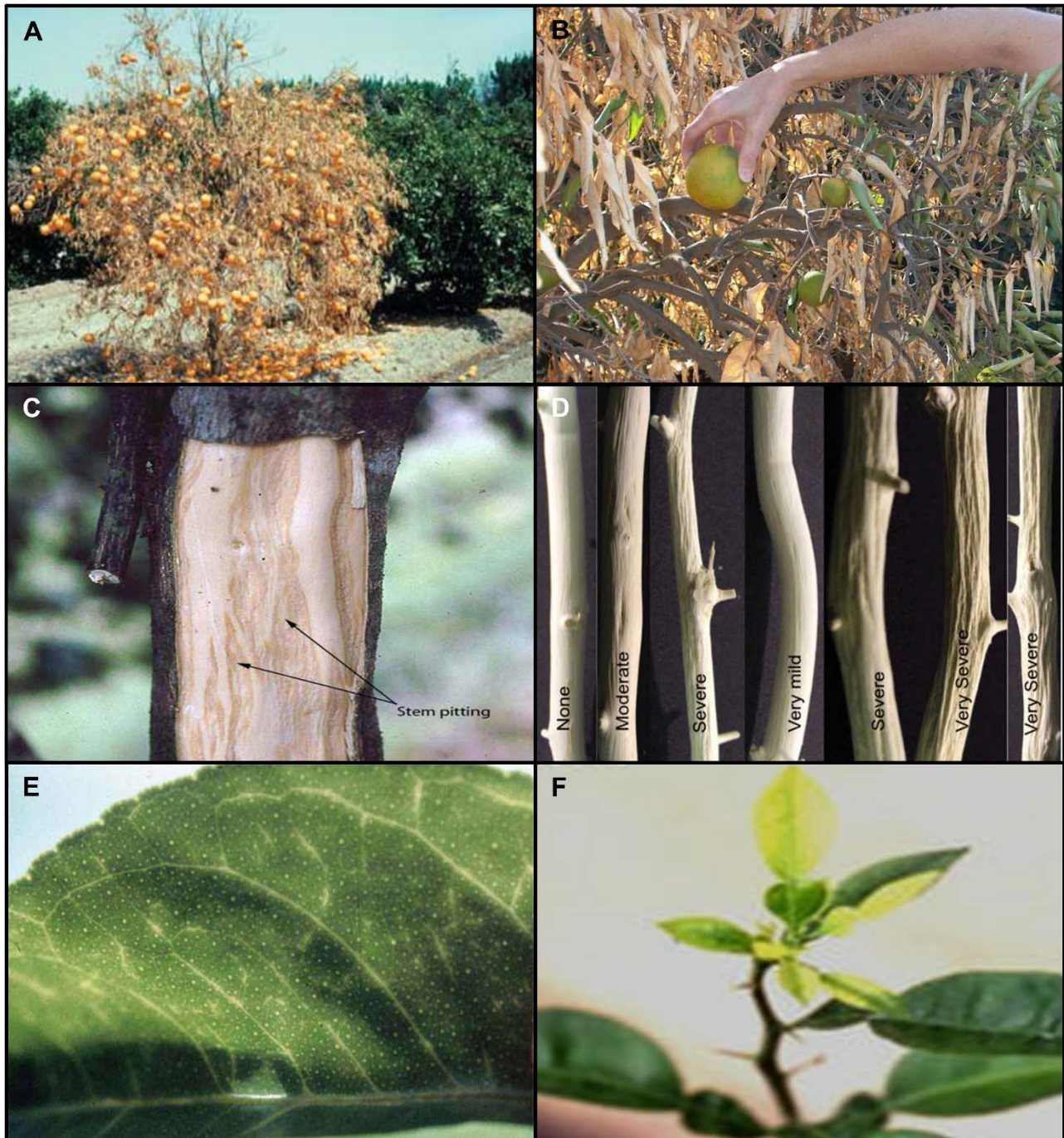
Figure 2.2. Decline, stem pitting and seedling yellows syndromes induced by Citrus tristeza virus (CTV) in different varieties and scion/rootstock combinations. **A)** and **B)** Quick decline syndrome in a sweet orange tree propagated on sour orange rootstock in comparison with non-decline neighbouring trees (dark green colour) (http://idtools.org/id/citrus/diseases/). **C)** Bark and **D)** stems of *Citrus macrophylla* infected with different CTV variant combinations showing the degrees of stem pitting (Dawson *et al.* 2013). **E)** Chlorotic veins of CTV-infected Mexican lime leaves (http://idtools.org/id/citrus/diseases/). **F)** Development of seedling yellows syndrome (SY) in CTV infected sour orange plants (Albiach-Marti *et al.* 2010).

2.2.6 Transmission

The dispersal of CTV over long distances predominantly occurs through graft transmission or the use of infected plant material for the propagation of new citrus trees (Moreno *et al* 2008, Yokomi 2009, Dawson *et al.* 2013). This was however circumvented in the past, as restrictions with large-scale shipping led to citrus plants being solely transported in the form of seeds, since CTV is not seed-borne. Under field conditions, the virus is spread locally, from tree to tree, by several aphid species in a semi-persistent manner (Bar-Joseph and Lee 1989, Brlansky *et al.* 2003). Aphids are insect vectors that transmit viruses by feeding on the sap of the phloem tissue in plant hosts (Wooton 1998). CTV is acquired by aphids within 5 min of feeding time, after which the vector is capable of retaining the virus for 24 – 48 hours (Raccah *et al.* 1976, Moreno *et al.* 2008, Yokomi 2009). The ability of a particular aphid species to transmit the virus efficiently is dependent on a number of factors including the number of aphids involved, the CTV isolate population, the variety of the citrus donor and receptor plants, as well as environmental conditions (Roistacher and Moreno 1990, Cambre *et al.* 2000, Marroquín *et al.* 2004).
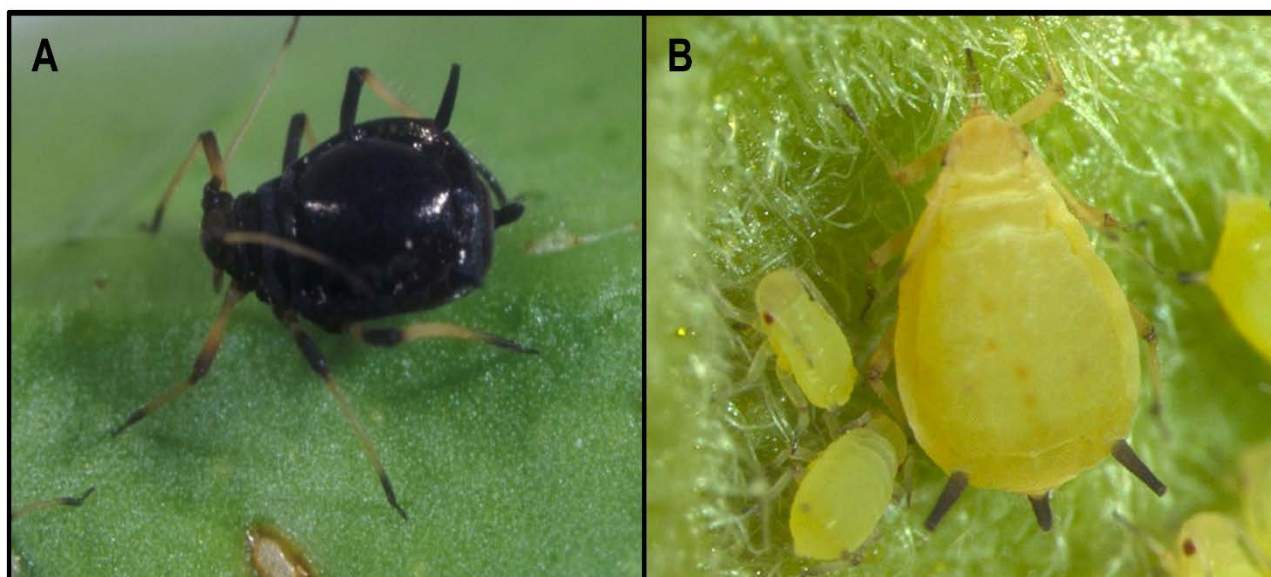
Figure 2.3. Images of citrus feeding aphids **A)** *Toxoptera citricida* (brown citrus aphid) and **B)** the melon aphid (*Aphis gossypii*) (http://idtools.org/id/citrus/pests/).

Amongst the citrus-feeding aphid species Toxoptera citricida (Kirkaldy), commonly known as brown citrus aphid (Figure 2.4 A), is the most efficient and frequent transmitter of CTV (Brlansky *et al.* 2003, Moreno *et al.* 2008). This aphid species has the ability to transmit most CTV isolates, including those that cause severe stem pitting and quick decline (Yokomi 2009). Second to the brown citrus aphid is the melon aphid (*Aphis gossypii*) (Figure 2.4 B)

and even though it has a host range that is not as broad as *T. citricida*, it has been responsible for secondary spread of the virus in citrus growing regions of North America (Cambra *et al.* 2000, Backus and Bennett 2009). Other less efficient aphid vectors of CTV that have been found to inhabit citrus intermittently include *Toxoptera aurantii* and the spirea aphid, *Aphis spiraecola* (Patch) (Moreno *et al.* 2008, Yokomi 2009). *Citrus tristeza virus* has also been transmitted experimentally to unaffected plants by *Cuscuta subinclusa* (dodder) as well as mechanically, by slash-inoculations with concentrated viral extracts (Roistacher 1991, Dawson *et al.* 2013).

2.2.7 Disease management

The prevalence and secondary transmission of CTV throughout citrus producing areas can be ascribed to the interaction between the virus isolate, host plant and any insect vectors present. The consideration of these elements and the implementation of multiple control measures are therefore required to manage the associated disease effectively. Many countries make use of preventative measures such as certification and quarantine programs to provide virus-free plant material for propagation and in doing so, prevent the introduction of CTV into citrus growing regions. These programs depend largely on the use of reliable and sensitive techniques for early virus detection (Constable *et al.* 2010). However, once the virus is present in an orchard, the removal of infected trees along with constant surveillance is recommended. This type of elimination scheme is only effective if the infection is localised to a few trees and the occurrence of natural vectors are limited. When eradication is unpractical, the use of CTV resistant rootstocks or scions can be implemented, especially to prevent the occurrence of quick decline symptoms. The latter has proven to be extremely successful in Asia where they have reverted to only growing citrus varieties that are tolerant to severe CTV isolates (Yokomi 2009). The control of CTV-induced stem pitting however, is more challenging as it influences both the rootstock and the citrus variety grafted onto it. Currently, the only way to protect economically important citrus cultivars against CTV isolates that cause severe stem pitting is to pre-inoculate them with a mild CTV isolate. Numerous viruses have displayed the phenomenon termed "cross protection" also referred to as "pre-immunisation" or "mild strain protection" since it was first observed between strains of *Tobacco mosaic virus* (McKinney 1929). Briefly, cross protection entails the inoculation of a plant with a mild isolate of a virus in order to protect it from any losses it may experience during a secondary infection with a more severe isolate of the same virus (Gonsalves and Garnsey, 1989; Foliminova, 2013). Although it has been shown to be mostly successful, the exact mechanism has not been fully resolved. Since the first commercial

15

manipulation of mild strain cross protection in citrus with the aim of protecting trees against severe CTV-associated stem pitting, it has been responsible for conserving productivity in citrus growing regions where severe CTV isolates and vectors such as the brown citrus aphid are prevalent (Grant and Costa 1951, Moreno *et al.* 2008). The South African Citrus Improvement Scheme (CIS) also implemented cross protection using mild CTV sources to minimise losses in the local citrus industry due to CTV diseases. However, cases of breakdown within the strategy have occurred and the local industry has funded research to address this protection breakdown in grapefruit specifically, as it comprises an important sector of the citrus production aimed at the export market (Van Vuuren *et al.* 1993). Protective isolates are normally selected from field trees of the same cultivar, which has been growing for years (vigorous trees) with mild or no symptom expression. These plants are assumed to be protected from the infection and there is a continuing search for usable protecting isolates in order to compensate for the changes in CTV populations in the environment as new genotypes or variants of CTV enter (Roistacher *et al.* 1993; Folimonova 2013). Feasible control strategies for CTV include reducing the population of vectors (aphids) in the area through chemical control and exploiting transgenic approaches to establish CTV resistant plants.

2.3 Citrus psorosis virus (CPsV)

The single-stranded (ss), negative-sense RNA virus, *Citrus psorosis virus*, is the type member of the only genus in the family *Ophioviridae*, *Ophiovirus* (Martín *et al.* 2005, Achachi *et al.* 2015, Moreno *et al.* 2015). Members of this genus share a unique "kinked" virion morphology resembling a coil (Milne *et al.* 1996, Velázquez *et al.* 2010) and include five other recognised species, namely: *Freesia sneak virus*, *Lettuce ring necrosis virus*, *Mirafiori lettuce big vein virus*, *Ranunculus white mottle virus*, and *Tulip mild mottle mosaic virus* (Achachi *et al.* 2014, Moreno *et al.* 2015). The genome of CPsV consists of three encapsulated RNAs, ranging from approximately 1,400 to 8,200 nts, and a coat protein (CP) of between 48 and 50 kDa in size (Naum-Onganía *et al.* 2003, Martín *et al.* 2005, Velázquez *et al.* 2010). *Citrus psorosis virus* has been conjectured to be associated with, psorosis, the first graft-transmissible disease in citrus (Moreno *et al.* 2015). One of the most characteristic symptoms of the disease is the scaling of the bark of the trunk and branches. Other symptoms include the accumulation of brownish gum (Figure 2.5 A) and blotches on the wood beneath the exposed bark (Roistacher *et al.* 1993, Martín *et al.* 2004, Moreno *et al.* 2015). Based on the degree of symptom expression the disease has been categorised into

two types, psorosis A (PsA) and psorosis B (PsB). The latter is the more severe syndrome, leading to the occurrence of chlorotic spots in old leaves as well as indentations on the fruit (Figure 2.5 B) (Achachi *et al.* 2015). Due to the detrimental impact psorosis has had on the citrus industry worldwide, certification schemes have been put in place to prevent the spread of the disease through budwood (Roistacher *et al.* 1993, Zanek *et al.* 2006).

2.4 Citrus tatter leaf virus (CTLV)

Characterisation of the monopartite, positive sense RNA genome of CTLV directed its inclusion in the genus *Capillovirus* (family: *Betaflexiviridae*), along with the type species, *Apple stem grooving virus* (ASGV) and *Cherry virus A* (CVA) (Tatineni *et al.* 2009, Komatsu *et al.* 2012). Sequence analysis of different CTLV strains revealed that the virus shares significantly high homology with ASGV and is consequently discerned as a citrus-infecting isolate thereof (Ohira *et al.* 1995, Martelli *et al.* 2007, Song *et al.* 2016). This graft transmissible virus is associated with the occurrence of "bud union disorder", similar to that seen with *Citrus leaf blotch virus* (CLBV) infection (Roistacher 1991, Hailstones *et al.* 2000, Song *et al.* 2009). In addition to leaf bleaching and deformation (tatter) (Figure 2.5 C), other symptoms attributed to CTLV infection range from restricted growth and graft incompatibility to severe decline symptoms and tree death within a minimum of five years (Roistacher 1991, Osvaldo *et al.* 2002, Song *et al.* 2015). The disease typically remains latent when citrus cultivars are propagated on their own roots, manifesting only upon grafting these trees onto rootstocks originating from trifoliate orange or any of its hybrids (Miyakawa and Ito 2000, Lovisolo *et al.* 2002).

2.5 Citrus variegation virus (CVV)

Variegation disease affects citrus globally inducing moderate to severe symptoms depending on the virus strain and citrus cultivar combination. The first strain of the contributing virus leads to the occurrence of acute infectious variegation in cultivars such as *C. medica* (citron) and *C. limon* (lemon). Plants infected with this strain of CVV usually have smaller, corrugated leaves that display different levels of chlorosis (Bennani *et al.* 2002, Abou Kubaa *et al.* 2015). The second, less severe strain, causes crinkly, bent leaves to occur without affecting the colour or size of the leaves (Desjardins and Bov´e 1980, Bennani *et al.* 2002). *Citrus variegation virus* belongs to the genus *Ilarvirus* within the family *Bromoviridae*, sharing serological similarities with members of the same genus such as *Asparagus virus 2* (AV-s) and *Citrus leaf rugose virus* (CiLRV) (Roossinck *et al.* 2005). The

17

positive-sense, ssRNA virus genome is tripartite in nature and transmissible between hosts through mechanical approaches (Roistacher 1991, Loconsole *et al.* 2009, Abou Kubaa *et al.* 2015).

2.6 Citrus yellow mosaic virus (CYMV)

Citrus yellow mosaic virus is the causative agent of citrus yellow mosaic diseases and has been preliminary classified as a member of the genus *Badnavirus* (family: *Caulimoviridae*) (Huang and Hartung 2001, Baranwal *et al.* 2003, Ghosh *et al.* 2014). This double-stranded DNA virus has a genome that is circular in nature of approximately 7500 bp, and shares homology with virus species including *Banana streak virus* (BSV); *Beetle vine yellow mottle virus*; *Cacao swollen shoot virus* (CSSV) and *Fig badnavirus* (Baranwal *et al.* 2005, Johnson *et al.* 2012). To date, cases of citrus yellow mosaic disease has been restricted to areas in India where it was initially described in sweet oranges in 1975, spreading to include Acid lime, Rangpur lime and Pumelo cultivars. As is the case with many other citrus-infecting viruses, CYMV can be transmitted to multiple cultivars through grafting, mechanical inoculation and by means of natural occurring vectors such as aphids and mealybugs (Baranwal *et al.* 2003, Ghosh *et al.* 2014). The name of the disease is founded on the chlorotic pattern observed on the leaves of infected plants (Figure 2.5 D), which may be accompanied by yellow mottling along the veins (Ahlawat *et al.* 1996). As a result, trees infected with CYMV produce fruit with reduced levels of ascorbic acid and experience an overall decrease in fruit production.

2.7 Citrus leaf rugose virus (CiLRV)

As a member of the genus *Ilarvirus* (subgroup 2) in the family *Bromoviridae*, CiLRV is serologically related to other viral species in this genus and has a genome that consists of three single, positive sense single-strand RNA molecules (Garnsey 1975, Scott and Ge 1995). Even though there are multiple genomic and biological similarities between CiLRV and a member of the same genus, CVV, they can be separated from one another without difficulty based on the different symptoms they induce. Cases of cross protection between these two viruses has also been observed when citron plants immunised with CiLRV were exposed to a secondary CVV infection (Garnsey 1975, Scott *et al.* 1995). CiLRV-infection is characterised by the rugose or wrinkling symptoms induced in Mexican lime, which forms the basis of its name. Other symptoms include the flecking of leaves and extensive growth inhibition in Eureka lemon and grapefruit, respectively (Garnsey 1975). This mechanically

transmitted virus was first discovered in Florida and is known to infect a wide variety of citrus hosts including Duncan grapefruit (*C. paradisi*), Eureka lemon (*C. limon*) and Mexican lime (*C. aurantifolia*) (Garnsey 1975, Dawson 2010).

2.8 Citrus leaf blotch virus (CLBV)

Citrus leaf blotch virus, previously known as *Dweet mottle virus* (DMV), is not only the type species but also currently the only member of the genus *Citrivirus* (family: *Betaflexiviridae*) (Vives *et al.* 2001, Hajeri *et al.* 2010, Adams *et al.* 2012). The virus has a monopartite genome that consists of a coat protein (~ 41 kDA) and a linear ssRNA (positive-sense) molecule of 8747 nts, making it comparable with members of the *Trichovirus* genus (Galipienso *et al.* 2001, Vives *et al.* 2001, Hernández-Rodríguez *et al.* 2016). Infection is associated with the occurrence of Dweet mottle disease that causes speckling of leaves in Dweet tangor[1] and the formation of pits in the stems of Etrog citron (*Citrus medica*) (Galipienso *et al.* 2000, Vives *et al.* 2008). In addition to *Citrus tatter leaf virus* (CTLV) and *Citrus tristeza virus* (CTV), CLBV has also been found to be linked to the manifestation of "bud union disorder" in citrus cultivars grafted onto trifoliate (including hybrids) rootstocks (Vives *et al.*2002, Hajeri *et al*, 2010). Bud grafting is one of the most common horticultural techniques and is often seen as the preferred method for citrus propagation, therefore complications that could result in the possible rejection of the graft could negatively affect the citrus industry (Guerri *et al.* 2004, Hernández-Rodríguez *et al.* 2016). The virus was initially detected in Spain, in Nagami kumquat (*Fortunella margarita*) plants, and has since then been reported in countries throughout the world including Australia; Florida; Italy; Japan; New Zealand; and recently Cuba (Vives *et al.* 2002, Galipienso *et al.* 2004, Hernández-Rodríguez *et al.* 2016).

2.9 Citrus leprosis virus C (CiLV-C)

Citrus leprosis is one of the most detrimental viral diseases, diminishing the productivity and life span of citrus plants in South and Central America, particularly Argentina and Brazil (Bastianel *et al.* 2010, Roy *et al.* 2013, Garita *et al.* 2014). The most prominent symptoms include necrotic lesions on the fruit, leaves, and stems (Figure 2.5 E and F) of affected trees that may at times be confused with that of citrus canker (Rodrigues *et al.* 2003, Locali-Fabris *et al.* 2006). Damaged fruit and leaves often drop from the tree prematurely, ultimately

---

[1] Mediterranean Sweet orange and Dancy tangerine hybrid

resulting in severe tree decline (Rodrigues *et al.* 2003). The severity of the disease varies amongst citrus cultivars, with sweet oranges being the most sensitive, while lemons are almost resistant (Bastianel *et al.* 2010, Garita *et al.* 2014). Three viral species with distinctive morphological characteristics namely, *Citrus leprosis virus* cytoplasmic type (CiLV-C); *Citrus leprosis virus* cytoplasmic type 2 (CiLV-C2); and *Citrus leprosis virus* nuclear type (CiLV-N), have been reported to be associated with the manifestation of leprosis (Locali *et al.* 2003, Locali-Fabris *et al.* 2006). These viruses are vectored by mites belonging to the genus *Brevipalpus* and can be transmitted to citrus plants experimentally through grafting and sap inoculations (Colariccio *et al.* 1995, Rodrigues *et al.* 2003, Bastianel *et al.* 2010). The genome sequence of the most prevalent virus, CiLV-C, revealed that it is made up of two, single-stranded, positive sense RNA components, leading to its classification as the type member of a newly accepted genus, *Cilevirus* (Locali-Fabris *et al.* 2006, Pascon *et al.* 2006, Locali-Fabris *et al.* 2012).

2.10 Indian citrus ringspot virus (ICRSV)

Sequencing data revealed that ICRSV is a filamentous virus with a single-stranded, message-sense RNA genome that is roughly 7.5 kb long (Rustici *et al.* 2002, Hoa and Ahlawat 2004). Although ICRSV is comparable to members of other genera in the family *Flexiviridae* complex, distinct differences lead to it being classified as the type member of a separate genus, *Mandarivirus* (family: *Alphaflexiviridae*) (King *et al.* 2012). The disease associated with ICRSV was first described in California (1968) and subsequently spread globally, where it has been responsible for serious losses to the citrus industry in India. The virus is known to affect one of the countries more essential fruit crops, the Kinnow mandarin (hybrid), by reducing the overall yield and quality of the fruit (Byadgi and Ahlawat 1995, Thind *et al.* 1997, Sharma *et al.* 2007). Infected trees exhibit psorosis-like symptoms that include the formation of chlorotic spot on the leaves; bleaching of the veins; and necrotic spots on mandarin fruit (Hoa and Ahlawat 2004). In more severe cases, these symptoms are accompanied with those of tree decline, ultimately rendering the tree unproductive (Lore *et al.* 2001, Sharma *et al.* 2007).

2.11 Citrus yellow vein clearing virus (CYVCV)

Yellow vein clearing disease (YVCD) mostly affects citrus leaves, inducing yellow vein clearing, leaf deformation, and crinkling symptoms that may lead to a yield decline of nearly 20% (Chen *et al.* 2014). This emergent viral disease was first observed in Pakistan during

the late 1980s, affecting economically important lemon and sour orange cultivars (Catara *et al* 1993). Subsequently, it has been found in China, India and Turkey, severely compromising the production of commercial citrus species including sweet orange and grapefruit (Alshami *et al.* 2003, Zhou *et al.* 2013). The occurrence of YVCD has recently been attributed to *Citrus yellow vein clearing virus* (CYVCV), a 7.5 kb positive sense RNA virus that has been characterised as a definitive member of the genus *Mandarivirus* (family: *Alphaflexiviridae*) (Önelge *et al.* 2011, Loconsole *et al.* 2012). This virus is transmitted to citrus and herbaceous plants, such as the common bean, through grafting, mechanical inoculation or with an aphid vector (Ahlawat and Pant 2003, Önelge *et al.* 2011). Although, the transmission of CYVCV through seeds has not yet been reported; a study by Zhou *et al.* (2015) demonstrated the presence of CYVC in seed tissues regardless of the fact that none of the progeny plants were infected.

2.12 Satsuma dwarf virus (SDV)

SDV was conditionally characterised as a member of the genus *Nepovirus* within the *Secoviridae* family (subfamily: *Comovirinae*), sharing symptomology similarities with viruses such as *Citrus mosaic sadwavirus* and *Navel orange infectious mottling virus* (Karasev *et al.* 2001). Further evaluations of its bipartite, positive sense, ssRNA genome however, lead to SDV being classified as the type member of the new genus, *Sadwavirus* (family: *Secoviridae*) (ICTV 2014). The graft transmissible pathogen, SDV, causes satsuma dwarf disease, negatively influencing the cultivation of satsuma mandarins (*Citrus unshiu*) and sweet oranges in areas of China, Japan and Turkey (Azeri 1973, Chi *et al.* 1991, Iwanami *et al.* 2001). Infection with SDV severely stunts tree growth, leading to the production of fruit with reduced accumulation of sugars and increased acidity, ultimately resulting in an overall decline in tree vitality and yield. In assessing the natural spread of the disease, soil transmission was implicated while no insect vectors have been identified (Kusano *et al.* 2007).

Figure 2.4. A) Bark scaling and gumming of a sweet orange, characteristic of psorosis A (PsA) (Moreno *et al.* 2015). **B)** Discoloration affecting wood below the bark lesions as a result of Citrus psorosis virus (Moreno *et al.* 2015). **C)** Leaf symptoms caused by citrus tatter leaf in citrange (http://www.ipmimages.org/browse/). **D)** Acid lime leaves showing mosaic symptoms upon graft-inoculation with *Citrus yellow mosaic virus* (Ghosh *et al.* 2014). **E)** Close-up of necrotic lesions on fruit and **F)** the green part of a branch of sour orange trees infected with *Citrus leprosis virus* (http://idtools.org/id/citrus/diseases/).

22

2.13 Virus detection

2.13.1 Current detection assays

The rapid and accurate identification of plant viruses during early stages of infection, accompanied by continued monitoring, is essential for effective disease management. Over the years, a number of techniques that allow for sensitive, specific and rigorous virus detection have been made available. The most basic approach in viral disease diagnostics is the visual observation of symptoms and subsequent confirmation using electron microscopy. This method requires highly skilled individuals and often involves the use of indicator plants, making it ineffective in instances where disease symptoms overlap or a delay in symptom expression is experienced. Presently, plant viruses are routinely detected with serological techniques such as enzyme-linked immunosorbent assay (ELISA); and nucleic acid amplification based methods such as PCR or RT-PCR.

Since its development in the 1970's, ELISA has become one of the most prevalent and versatile serological approaches for virus detection in plants (Engvall and Perlmann. 1971, Clark and Adams 1977, Ward *et al.* 2004. Boonham *et al.* 2014). The method involves fixing specific antibodies to a microtitre plate in order to detect viral antigens within the sample of interest. Advancements to the assay allows for the use of either polyclonal or monoclonal antibodies (Naidua and Hughes 2001, Boonham *et al.* 2014). The effortless implementation of ELISA has permitted the design of several forms of the technique, of which double antibody sandwich (DAS) ELISA is the most widely used (Koenig and Paul 1982). Although these variations differ in the manner in which they detect the antigen-antibody complex, they all employ the same underlying mechanisms (Koenig and Paul 1982). The DAS variant of the technique has been successfully used for the rapid and efficient detection of multiple citrus-infecting viruses, including CTV (Hancevic *et al.* 2012). However, despite the magnitude of advantages ELISA has for high throughput virus screening, it still lacks the flexibility and sensitivity that certain nucleic acid amplification based approaches provide (Ward *et al.* 2004, Boonham *et al.* 2014).

Amongst alternative detection methods that focus on identifying viral nucleic acids within a given sample, polymerase chain reaction (PCR) based techniques are the most commonly used and widely adapted (O'Donnell 1999, Boonham *et al.* 2014). These techniques rely on the use of complementary primers to target a specific genomic region of viral DNA for subsequent exponential amplification (Mullis *et al.* 1986, O'Donnell 1999, Ward *et al.* 2004, Bexfield and Kellam 2011). As most of the viruses infecting commercial citrus cultivars have

RNA genomes, a variation of the technique, known as reverse-transcription (RT) PCR, is employed which entails converting genomic viral RNA into complementary DNA (cDNA) ahead of primer annealing and amplification (Ward *et al.* 2004). In addition to the increased levels of sensitivity and specificity that accompany PCR based methods, they also allow for the simultaneous detection of multiple viruses or virus variants by making use of several primer pairs. Although the latter is more economical in terms of reaction costs, the experimental design and optimisation remains more challenging than microarray analyses (Ward *et al.* 2004, Mumford *et al.* 2006). The fragments resulting from PCR amplification often require further sequencing as gel-based visualisation alone does not provide irrefutable evidence for the presence or exclusion of a particular virus (Schaad and Frederick 2002, Boonham *et al.* 2014). Further improvements to traditional PCR or RT-PCR protocols came with the introduction of real time PCR, which allows for the absolute or relative quantitation of the target viral DNA or RNA concentration (Feng *et al.* 2008). This method detects and assesses target nucleic acid amplification throughout each successive cycle using a built-in fluorometer along with DNA binding dyes or fluorescent probes (Ward *et al.* 2004).

Many of the known viruses are routinely identified using serological or nucleic acid based assays as they have become more effective with improvements in multiplexing and enrichment processes in order to remove host nucleic acids. These techniques however all require prior knowledge regarding the target virus, restricting their ability to identify novel viruses or uncharacterised variants of previously known viruses (Schaad *et al.* 2003; Espach, 2013, Stobbe *et al.* 2013). The application of next-generation sequencing (NGS) provides a powerful alternative to traditional virus detection techniques and has to date, enabled a deeper understanding of viral biodiversity.

2.13.2 Virus detection through next-generation sequencing (NGS)

The study of nucleic acids obtained directly from the environment, also denoted as metagenomics or environmental genomics, provides an approach that is less biased than conventional pathogen discovery methods (Hugenholtz and Tyson 2008). This approach has the potential to identify all the organisms present in an environmental sample. It initially relied on the use of molecular techniques such as cloning and capillary-based sequencing (Hugenholtz and Tyson 2008, Mokili *et al.* 2012). Although metagenomic studies have been used to explore the microbial diversity in several different environments, the methodology involved can be costly and time consuming. The advent of high-throughput sequencing

however, has transformed traditional metagenomics and allowed for much more detailed, large-scale investigations.

Concerning virome studies, coupling NGS to metagenomics as applied in this study, provides an approach that is sensitive enough to detect novel and known viruses, and also has the ability to distinguish between different strains of virus species (Al Rwahnih *et al.* 2009, Koonin and Dolja 2012, Mokii *et al.* 2012). Additionally, massive parallel sequencing allows for the detection of viral agents that are present in low titres, as well as the assessment of virus quantity using the associated read count (Mardis 2008, Al Rwahnih *et al.* 2009). Next-generation sequencing has significantly advanced plant virology in terms of disease etiology and hitherto numerous approaches for the detection of plant viruses have been published (Al Rwahnih *et al.* 2011, Boonham *et al.* 2014, Massart *et al.* 2014). These approaches however, all vary in the sequencing technology enforced and the nucleic acid used as starting material.

Next-generation sequencing entails the sequencing of genetic material with the aid of ubiquitous adaptors and most technologies constitute three main steps: 1) preparation of the sample library; 2) clonal amplification; and 3) high-throughput sequencing (Mardis 2008, Massart *et al.* 2014). On-going advancements in the different sequencing platforms available, focuses on streamlining the library preparation and amplification steps in the hopes of reducing PCR errors and shortening the overall sequencing run time (Massart *et al.* 2014). A number of commercially accessible NGS platforms have been used for plant virus detection, each differing in the cost, fundamental sequence chemistry exploited and read lengths produced. Amongst these sequencing technologies, Illumina, Ion torrent, Roche 454, and PacBio are generally the most favoured (Zhang *et al.* 2011, Liu *et al.* 2012, Massart *et al.* 2014).

Although the sequencing platform used is imperative in studying plant viral populations, continued research has shown that the nucleic acid purification or enrichment strategy used plays a more crucial role (Massart *et al.* 2014). The use of total RNA as a starting material has been successful for the detection and characterisation of plant viruses in a number of studies (Adams *et al.* 2009, Al Rwahnih *et al.* 2009, Kreuze *et al.* 2009). While this approach disregards host genomic DNA, it becomes inadequate when dealing with viruses that are present at a low titre as a large amount of the sequence reads generated are derived from host RNA (Adams *et al.* 2009, Boonham *et al.* 2014). Additional improvements to minimise

host sequence contamination involves enriching the mRNA component through poly-A selection or ribo-depletion in order to select for non-plant RNA (Adams *et al.* 2009).

Another strategy frequently used in the case of positive or negative sense ssRNA viruses is the extraction of double-stranded RNA (dsRNA) molecules generated as an intermediate-product during the replication cycle of the virus (Adams *et al.* 2009, Al Rwahnih *et al.* 2009). This method has been extremely effective in reducing host contamination as endogenous RNA seldom form considerable amounts of double-stranded molecules (Coetzee *et al.* 2010, Al Rwahnih *et al.* 2011, Roossinck *et al.* 2011). The sequencing and assembly of small RNAs to study plant viruses has become particularly favoured since its establishment (Kreuze *et al.* 2009). This approach gives insight into the plants' viral defence system by sequencing small RNA molecules (21-24 nts in length), comparable to the target virus, generated by the plant RNA interference mechanisms (Mlotshwa *et al.* 2008, Boonham *et al.* 2014). However, the shortened read length makes the assembly of full genomes rather challenging specifically in plants infected with closely related viruses (Boonham *et al.* 2014, Massart *et al.* 2014).

2.14 Bioinformatic approaches to virus detection in NGS data

Next-generation sequencing, otherwise known as high-throughput sequencing, produces large amounts of sequence data, and is therefore subjected to computational limitations when it comes to analysing and interpreting the data (Lui *et al.* 2012). As not all NGS platforms generate the same read lengths and output formats, no generic bioinformatic tool can be utilised. Along with the tools made available by the different NGS platforms, a considerable number of computer algorithms and software have been developed specifically for managing the copious amounts of data generated.

In a conventional NGS virus detection approach, the raw sequence data will initially be pre-processed through the removal of adaptor sequences and the trimming of low quality bases (Massart *et al.* 2014). The quality trimming may be accompanied by a filtering step involving the removal of reads that are still below a specified overall quality threshold or relating to host sequences, in order to retain high quality viral reads (Mokili *et al.* 2012). The refined dataset is then used in downstream analyses that consist of either *de novo* assemblies and subsequent similarity screening, or read mapping to available reference genomes of viruses that are suspected to be present in the sample (Scholz *et al.* 2012, Massart *et al.* 2014).

De novo assemblies have become the benchmark for metagenomic studies, as they remain the only viable approach permitting the characterisation of unknown viruses (Massart *et al.* 2014). Most *de novo* assemblers function through fragmenting raw reads into smaller, overlapping segments referred to as k-mers. These k-mers are then joined using de Bruijn graphs in order to form longer contiguous sequences commonly known as contigs (Zerbino and Birney 2008, Scholz *et al.* 2012). The resulting contigs can be annotated using BLAST searches, against public or local databases (Kent 2002, Mokoli *et al.* 2012, Massart *et al.* 2014).

Read mapping on the other hand, entails aligning the NGS reads to a set of reference virus sequences in order to establish the viral status of the sample. Next-generation sequencing associated bioinformatics has evolved from previously being intricate and often requiring programming skills, to the development of more user-friendly packages such as the commercial packages, CLC Genomics Workbench and Geneious, as well as freely available tools. The latter can be executed either online, or on a local desktop or server, necessitating a command line interface, to assist with the analysis of exponentially large NGS datasets (Massart *et al.* 2014). Conventional NGS data analysis remains computationally demanding, especially during the assembly and similarity searching steps, extending the time involved in obtaining a diagnostic result.

Stobbe *et al.* (2013) developed a pipeline, called e-probe diagnostic nucleic acid analysis (EDNA), which is centred on utilising probes (short sequences) unique to target pathogens for subsequent similarity-based screening of raw NGS data. A near-neighbour comparison approach is utilised for the design of e-probes whereby the genome of the pathogen of interest is compared to that of a closely related relative. This approach was used to design probes for the detection of two viruses namely, *Bean golden mosaic virus* (BGYMV) and *Plum pox virus* (PPV), as well as a selection of bacteria and eukaryotic (fungi and oomycetes) pathogens (Stobbe *et al.* 2013). The assessment and verification of the EDNA pipeline was done using a series of simulated 454-pyrosequencing datasets, representing the selected pathogens in a general plant background (*Vitis vinifera*) (Stobbe *et al.* 2013). The pipeline was further validated with NGS data obtained from plants infected with the aforementioned viruses, and was also tested for its strain differentiation capability using simulated datasets containing different PPV strains (Stobbe *et al.* 2014). Although the application of this targeted approach for virus detection calls for a significant amount of bioinformatic expertise, it requires less computational resources, ultimately addressing one of the major constraints accompanying NGS data analyses (Stobbe *et al.* 2013). The e-

probe based detection system can be adapted for a variety of pathogens and hosts, offering multiple possibilities for diagnostics (Stobbe *et al.* 2014).

2.15 Conclusion

Citrus tristeza virus, has been responsible for tremendous losses to the local and international citrus industries. The complexity of CTV as a pathogen of citrus species cannot be overstated as it has been shown to display the largest number of distinct symptoms of any plant virus. Infection with CTV causes four recognised host responses that are influenced by the citrus species; the rootstock; the scion-rootstock combination; the genotype of CTV and/or the combination of CTV genotypes. Despite the local industry's attempts to avert quick decline, the virus still negatively affects the production of citrus cultivars such as grapefruit.

The importance of CTV to the citrus industry thus drives the need for a better understanding of viral sources, as well as sensitive and rapid early detection methods to facilitate certification schemes in preventing the spread of the virus. Recent improvements in sequencing technologies make it possible to study the complete virus composition within a given sample at any time. This in return has permitted detailed research regarding viral complexity as well as host responses upon infection.

Chapter 3: Materials and Methods

3.1 Plant material

The plant material was obtained from two greenhouse trials at Citrus Research International in Nelspruit as well as from the field. The first trial was performed on the biological indicator host 'Mexican' lime (*Citrus aurantifolia*) and the second on the commercial grapefruit (*Citrus paradisi*) variety 'Marsh'. Virus-free shoots of these varieties were grafted onto rough lemon rootstocks (*Citrus jambhiri*). After six months, CTV single variant sources of different genotypes were used to inoculate the 'Mexican' lime plants in the first trial (Table 3.1). Five plants (biological replicates) were inoculated for each CTV genotype. The grapefruit plants in the second trial were inoculated with CTV genotype T3 (Table 3.1). The plant material was harvested one year post inoculation and the bark was attained for subsequent extraction of nucleic acids. The bark obtained from the biological replicate plants from trial 1 was pooled to provide sufficient material. Six additional field samples were obtained from the Hoedspruit region, Limpopo Province, South Africa. Samples (1-10) from trial 1 and the field (14-19) were collected in 2014, while samples (11-13) from trial 2 were collected in 2015. The CTV infection status of the plants in the two greenhouse trials was confirmed by RT-PCR at the Citrus Research International, Nelspruit (Supplemental Table 1 A and B, Supplemental Protocol 1).

3.2 Nucleic acid extractions

3.2.1 Double-stranded RNA extractions

Double-strand RNA (dsRNA) was extracted from the pooled bark material of 'Mexican' lime plants (1-10) and the field samples (14-19) (Table 3.1) using an affinity chromatography method (Burger and Maree 2015). This method consists of two cycles of cellulose affinity chromatography using MN 2100 (Macherey-Nagel) cellulose powder with approximately twelve grams of plant bark material per extraction. The dsRNA suspension was subjected to RNase and DNase treatment using RQ1 RNase-free DNase (Promega) and T1 RNase (Roche) according to manufacturers' protocols. The quality and quantity of the dsRNA was evaluated through visualization on a 1% [w/v] TAE Agarose gel stained with ethidium bromide. All the dsRNA samples were selected for next-generation sequencing.

3.2.2 Total RNA extraction

Total RNA was extracted from three CTV-infected grapefruit plants (Table 3.1) using an adapted version of the low molecular weight (LMW) RNA extraction method described in Carra *et al.* (2007). The method involves extracting total RNA from 2 grams of bark material using a CTAB extraction buffer (2% [w/v] CTAB, 2.5% [w/v] PVP-40, 100 mM Tris-HCL [pH8], 2 M NaCl, 25 mM EDTA [pH8] and 3% [v/v] β-mercaptoethanol) and isopropanol precipitation. The PEG precipitation step was omitted and additional chloroform extractions were performed to remove polysaccharides.

Table 3.1. Virus infected plant material used in study

Sample number	Sample name	Source (CTV genotype)	Cultivar[b]
1	SS_RB1	B389-1 (RB1)[a]	MXL
2	SS_RB2	B389-4 (RB2)[a]	MXL
3	SS_LMS	LMS 6-6 (HA16.5)	MXL
4	SS_T68	GFMS 12-8 (T68)	MXL
5	SS_VT	Maxi x3 (VT)	MXL
6	T1_RB1	B390-5 (RB1)	MXL
7	T1_RB2	B389-4 (RB2)	MXL
8	T1_LMS	LMS 6-6 (HA16.5)	MXL
9	T1_T68	GFMS 12-8 (T68)	MXL
10	T1_VT	Maxi x 3 (VT)	MXL
11	T2_01[c]	T3	GF – M
12	T2_02[c]	T3	GF – M
13	T2_03[c]	T3	GF – M
14	UNK_01	Field samples - status unknown	Val – D
15	UNK_02	Field samples - status unknown	Val – D
16	UNK_03	Field samples - status unknown	Val – D
17	UNK_04	Field samples - status unknown	SO – MV
18	UNK_05	Field samples - status unknown	SO – MV
19	UNK_06	Field samples - status unknown	SO – MV

[a] RB1 and RB2 refer to different RB genotype strains that are differentially amplified with RB group primers (Cook *et al.* 2016)

[b] MXL, 'Mexican' lime, biological indicator host for CTV; GF – M, Grapefruit variety, 'Marsh'; Val – D, Valencia variety, 'Delta'; SO – MV, 'Madam Vinous' sweet orange indicator host for CPsV.

[c] Temporary sample name since each plant will ultimately represent two different sequence samples

3.3 Library preparation and next-generation sequencing

Library preparations for dsRNA of samples in trial 1 and the field (Table 3.2) were performed using an adapted Illumina TruSeq RNA sample preparation protocol (Burger and Maree 2015). These libraries were multiplexed and sequenced in a paired-end (2 x 125 nts) run on an Illumina HiSeq2500 platform at the Agricultural Research Council's Biotechnology Platform in Pretoria, South Africa. Total RNA from the trial 2 samples were used to prepare two separate types of sequencing libraries: an Illumina Small RNA TruSeq protocol was followed to allow for small RNA (sRNA) sequencing and a TruSeq Stranded mRNA protocol for whole transcriptome sequencing after ribosomal RNA depletion. Both libraries were sequenced on an Illumina HiSeq2500 at Fasteris (Geneva, Switzerland), with single reads (1x 50 nts) for sRNA libraries, and paired-end reads (2 x 125 nts) for transcriptome libraries.

Table 3.2. List of samples subjected to next-generation sequencing with their respective sequencing libraries

Sample number	Sample name	Library type[b]
1	SS_RB1	dsRNA (RNASeq)
2	SS_RB2	dsRNA (RNASeq)
3	SS_LMS	dsRNA (RNASeq)
4	SS_T68	dsRNA (RNASeq)
5	SS_VT	dsRNA (RNASeq)
6	T1_RB1	dsRNA (RNASeq)
7	T1_RB2	dsRNA (RNASeq)
8	T1_LMS	dsRNA (RNASeq)
9	T1_T68	dsRNA (RNASeq)
10	T1_VT	dsRNA (RNASeq)
11	T2_01[a]	sRNA
12	T2_02[a]	sRNA
13	T2_03[a]	sRNA
14	T2_04[a]	transcriptome
15	T2_05[a]	transcriptome
16	T2_06[a]	transcriptome
17	UNK_01	dsRNA (RNASeq)
18	UNK_02	dsRNA (RNASeq)
19	UNK_03	dsRNA (RNASeq)
20	UNK_04	dsRNA (RNASeq)
21	UNK_05	dsRNA (RNASeq)
22	UNK_06	dsRNA (RNASeq)

[a] The three grapefruit plants infected with CTV genotype T3 were each subjected to two different types of next-generation sequencing.

[b] All libraries were sequenced on an Illumina HiSeq2500 platform.

3.4 Conventional NGS data analysis

3.4.1 Sequence pre-processing

The sequence data generated for the dsRNA samples was evaluated using a conventional data analysis approach. The datasets were assessed individually for quality using FastQC v0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were processed to identify and remove sequencing adaptors, while retaining high quality bases using Trimmomatic v0.33 (Bolger *et al.* 2014). The trimming involved the removal of the first 10 bases from the 5' end of each read to get rid of potentially incorporated sequencing errors. Other Trimmomatic parameters included applying a minimum read length of 1 and the removal of bases with a Phred quality score below Q20 from the 3' end of the sequence reads. The use of a Q20 Phred score cut off, lead to a base calling accuracy of 99%, further improving the overall quality of datasets for subsequent analyses. In order to filter out host sequences present in the datasets, processed reads were mapped to the genome of *Citrus sinensis* (sweet orange) [AJPS00000000] as well as the *Citrus aurantifolia* ('Mexican' lime) mitochondria and chloroplast sequences. The mapping was done using CLC Genomics Workbench v8.0.3 (https://www.qiagenbioinformatics.com/) with parameters that specified the reference genomes would not be masked, default read alignment settings (mismatch cost: 2; insertion cost: 3; deletion cost: 3; length fraction: 0.7; similarity fraction: 0.7), with global alignment and random mapping of non-specific matches. The unmapped reads from each sample were collected and used in downstream analysis.

3.4.2 Assembly and homology searching

The trimmed and filtered reads were uploaded onto CLC Genomics Workbench v8.0.3 (https://www.qiagenbioinformatics.com/) as single or paired-end Illumina sequence files. Using CLC's assembly algorithm, the high quality reads were used to construct longer contig sequences. The *de novo* assembly occurred in two consecutive steps. The first step entailed the random assembly of the reads into contig sequences. In the final step the reads are mapped back onto the contig sequences, providing information regarding contig coverage which in turn can be used as an indication of the accuracy of the assembly. Importantly, the alignment of a read to a contig does not necessarily mean that this read was used in the assembly of the contig. Parameters for each *de novo* assembly included: a default (automatically calculated) word and bubble size; minimum contig length of 200 nts and the scaffolding of contigs. Reads were subsequently mapped back to the contigs with the

following parameters: global alignment; auto-detection of paired distances; and read alignment costs for a mismatch, insertion and deletion set at 2, 3, and 3, respectively. Furthermore, a length fraction of 0.5 and similarity fraction of 0.8 was used, and the update of contigs was selected. The resulting contigs with a minimum length of 250 nts and an average coverage above 5 were selected and firstly subjected to non-redundant blastn analysis using CLC's local BLAST function. The latter was performed against a custom database containing genome sequences of the six CTV genotypes used in the greenhouse trials, as well as the accessions of the additional 10 viruses explored in this study (match cost: 1, mismatch cost: 2, gap opening cost: 5, and gap extension cost: 2). Additionally the selected contigs were also blasted against the NCBI's non-redundant nucleotide database with CLC's NCBI blast+ option using default parameters as for the local BLAST. Only high confidence matches were evaluated and contigs were classified according to their highest identity. Figure 3.1 illustrates the workflow followed in processing and analysing the NGS data using the conventional pipeline.
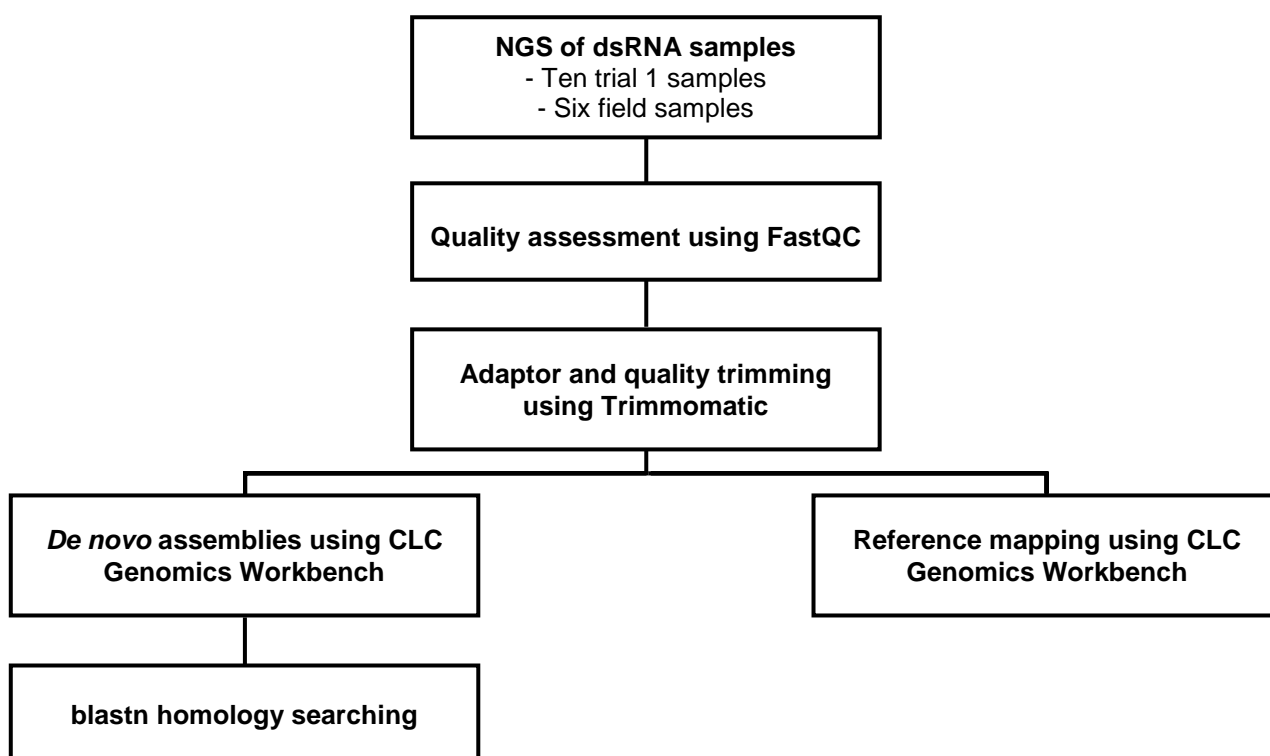
Figure 3.1. Conventional data analysis workflow used to evaluate NGS data obtained from the dsRNA samples. The workflow was implemented for the 16 sequence datasets individually.

3.5 E-probe based bioinformatics pipeline

3.5.1 Candidate e-probe design

To detect the presence of virus sequences in NGS data, short segments unique to the virus of interest (electronic probes) were identified. The latter was based on a near-neighbour comparison approach initially developed for the generation of probes used in microarrays (Vijaya Satya *et al.* 2008, Stobbe *et al.* 2013). Electronic-probes (e-probes) were designed for eleven citrus-infecting viruses, including CTV (Table 3.3), by aligning them to a closely related virus with Nucmer, a component of the Mummer software package (-g 0, -l 10 and -c 10) (Delcher *et al.* 1999, Kurtz *et al.* 2004). Other Nucmer parameters include using the maximum exact matches as alignment anchors, disabling the outward extension of alignments from their anchoring clusters and not simplifying the alignments in order to find inexact repeats. Virus genome sequences sharing similarity with near-neighbours were then removed using a custom Python script, retaining unique target sequences of a specified minimum length called candidate probes. As described in the EDNA pipeline, individual probe sequences with homo-oligomers longer than 4 nts were also removed from the candidate probe sets prior to optimisation and filtering steps (Stobbe *et al.* 2013).

3.5.2 Mock sequence dataset (MSD) construction

To optimise and evaluate the designed probes mock sequence datasets (MSDs) were generated containing genome sequences which belong to both the host and the pathogen. A variety of software products were evaluated for their ability to simulate Illumina HiSeq sequence data, including ART, GemSim, and MetaSim software (Richer *et al.* 2008, Huang *et al.* 2012, McElroy *et al.* 2012). The final datasets were simulated using ART software with the appropriate error model and read lengths, comparable to experimental data (Huang *et al.* 2012). The genome of *Citrus sinensis* (sweet orange) [AJPS00000000] was used as the host genome to provide background sequences within which virus-derived sequences could be present. The MSDs used for the optimisation of candidate e-probes (oMSDs) contained 10 000 simulated paired-end reads of 125 nucleotides in length. A series of oMSDs were generated to represent NGS data containing varying percentages of virus sequences. In accordance to Stobbe *et al.* (2013), a total of five different classes of oMSDs were constructed with different percentages of virus contribution: less than 1% was regarded as very low and 1% as low, 5, 10 and 25% were regarded as medium, medium-high and high respectively.

Table 3.3. E-probe design for eleven citrus infecting viruses

Target pathogen			Near-neighbour	
Virus	**Abbreviation**	**Accession**	**Virus**	**Accession**
Citrus tristeza virus	CTV	EU937519.1	*Grapevine leafroll associated virus 2*	NC_007448
Citrus psorosis virus	CPsV	NC_006314	*Mirafiori lettuce big-vein*	NC_004779
		NC_006315		NC_004781
		NC_006316		NC_004782
Citrus tatter leaf virus[a]	CTLV	JX416228	*Cherry virus A*	NC_003689
Citrus variegation virus	CVV	NC_009536	*Tobacco streak virus*	NC_003844
		NC_009537		NC_003842
		NC_009538		NC_003845
Citrus yellow mosaic virus	CYMV	NC003382	*Commelina yellow mottle virus*	NC_001343
Citrus leaf rugose virus	CiLRV	NC_003546	*Tobacco streak virus*	NC_003844
		NC_003547		NC_003842
		NC_003548		NC_003845
Citrus leaf blotch virus	CLBV	NC_003877	*Apple chlorotic leaf spot virus*	NC_001409
Citrus leprosis virus C	CiLV-C	NC_008169	*Orchid fleck virus*[b]	NC_009608
		NC_008170		NC_009609
Indian citrus ringspot virus	ICRSV	NC_003093	*Potato virus X*	NC_011620
Citrus yellow vein clearing virus - isolate CQ	CYVCV	NC_026592	*Indian citrus ringspot virus*	NC_003093
Satsuma dwarf virus	SDV	NC_003785	*Strawberry latent ringspot virus*	NC_006964
		NC_003786		NC_006965

[a] Isolate of Apple stem grooving virus.

[b] Near-neighbour selected, based on phylogenetic relationship described in Kondo *et al.* (2006).

3.5.3. Probe optimisation

Candidate probe sets with a range of minimum lengths (20, 40, 60, 80, 100, 120 and 140 nucleotides) were designed and screened against each oMSD using BLASTn, in order to establish the preferred length and e-value cut-off (Stobbe *et al.* 2013). As described by Stobbe *et al.* (2013) and for the purpose of this study, a hit can be defined as the instance where a probe is complimentary to a read (or parts thereof) present in an oMSD. A match on the other hand indicated the presence of a single probe within the entire oMSD, to the extent where a match consists of one or more hits. The optimal minimum probe length was determined by comparing the ratio of virus-specific hits to the overall observed hits (i.e. hits to virus as well as host sequences). The candidate probe sets of varying minimum lengths were also used to determine the e-value threshold. The latter was achieved by screening the candidate probe sets against all five categories of oMSDs using three different BLASTn e-value cut-offs (1×10^{-3}, 1×10^{-6} and 1×10^{-9}). Finally, candidate probes of optimal minimum length (20 nts and longer), were subjected to a filtering step against the online NCBI nt database in order to eliminate probes that are not specific to the virus of interest. The BLAST parameters used comprised of a gap cost of 5 and an extend cost of 2 as well as, reward 1; penalty -1; and a word size of 7 (Visser *et al.* 2016). Candidate probes that shared homology at an e-value of 1×10^{-3} (determined during optimisation) or lower with any species (including host), other than the target virus, were excluded from the final e-probe set. All the optimisation steps were performed using custom Python scripts and additionally resulted in the construction of a decoy set of probes for each of the eleven e-probe sets, by reversing the sequence of the final virus-specific e-probes.

3.5.4 E-probe based virus detection

Assigning a positive diagnostic call to samples of interest was centred on the presence of reads in sequence data that are homologous to e-probes for the target virus. Virus-specific e-probes and corresponding decoys were firstly screened against sequence data using the nucleotide blast function in Blast+ (Camacho *et al.* 2004). By means of the BLAST results, a score was allocated to every probe and decoy, depending on the number of hits, as well as the percentage coverage and e-value (Equation 1, in which n represents the number of a specific hit) (Stobbe *et al.* 2013).

Equation 1

$$\sum_{h=1}^{n}\{-logEval[h] * (\%cov.[h])\}$$

The scores of the e-probe sets were statistically correlated to that of the decoy sets. The normality (Shapiro-Wilk test) and population variance of the score data for the two sets was firstly determined. If the two datasets were normally distributed, either a student *t*-test (same variance) or a Welchs *t*-test (difference in variance) was performed. A non-parametric test (Wilcoxon Ranksum) was performed in the instance where the datasets were not normally distributed. A *P* value less than or equal to 0.05 indicated a positive call implying that the target virus for which the probes were designed is present, while a *P* value above 0.1 would indicate the absence of the specific virus (Stobbe *et al.* 2013). A third outcome was presented for samples with a *P* value less than or equal to 0.1 and greater than 0.05, since they were only "suspected" to be positive (Stobbe *et al.* 2013).

3.5.4.1 Detection using simulated datasets

In silico testing of the e-probe sets designed for the 11 viruses was performed with evaluation-MSDs (eMSDS) that were generated using ART Illumina simulation software (Huang *et al.* 2012) in the same way as previously described and contained both host and virus genome segments. The datasets used for probe evaluation consisted of 1 million paired-end reads, each 125 nts in length with a virus abundance level of 1%. Single infection datasets were constructed for each of the 11 viruses respectively. Multiple virus infection eMSDs were also constructed, one with all 11 viruses and the remaining sets excluding a different virus each time. Each eMSD served as a reference dataset against which virus-specific e-probe sets (including decoys) were screened using Blast+ (blastn) with an e-value set at 1×10^{-3} (determined during optimisation). Additionally, probe sets were evaluated for their ability to detect different isolates of each target virus species. GenBank accessions of full genomes of the different virus isolates were firstly used to construct eMSDs in the same way as previously described. Secondly, a diagnostic result was achieved by screening the simulated eMSDs with all 11 sets of probes and decoys.

3.5.4.2 Detection using NGS datasets

The raw data obtained from NGS of plant samples were screened with the designed target e-probe and corresponding decoy sets for all 11 viruses, using the nucleotide option in Blast+ (Camacho *et al.* 2004). In the case of paired-end data, the two read files were pooled to create one input dataset to be used for virus detection. Infection status was determined for all the plant samples in the same way as described for simulated data and confirmed by RT-PCR (Cook *et al.* 2016). Individual probes within the e-probe set designed for CTV detection were further evaluated for their performance with sRNA and transcriptome library types. The sequence data of three representatives of each sequence library type (sRNA and

transcriptome) was subjected to the random sampling of 1 million reads 100 times for each sample. The e-probes were screened against the 100 sub-sampled datasets and in each instance the number of times a specific probe hit was noted. The number of hits for a specific probe was averaged across the three samples from a specific dataset and plotted. To evaluate the influence of read length and the number of viral sequences present in each data set, additional sub-sampling experiments were conducted to assess the performance of the CTV probes. The influence of read length was evaluated by only using the first 23nt of the forward reads of the transcriptome datasets. The impact of the percentage of virus specific reads was evaluated by increasing the amount sRNA data sub-sampled to 5.7 million reads, to be comparable to the transcriptome data.
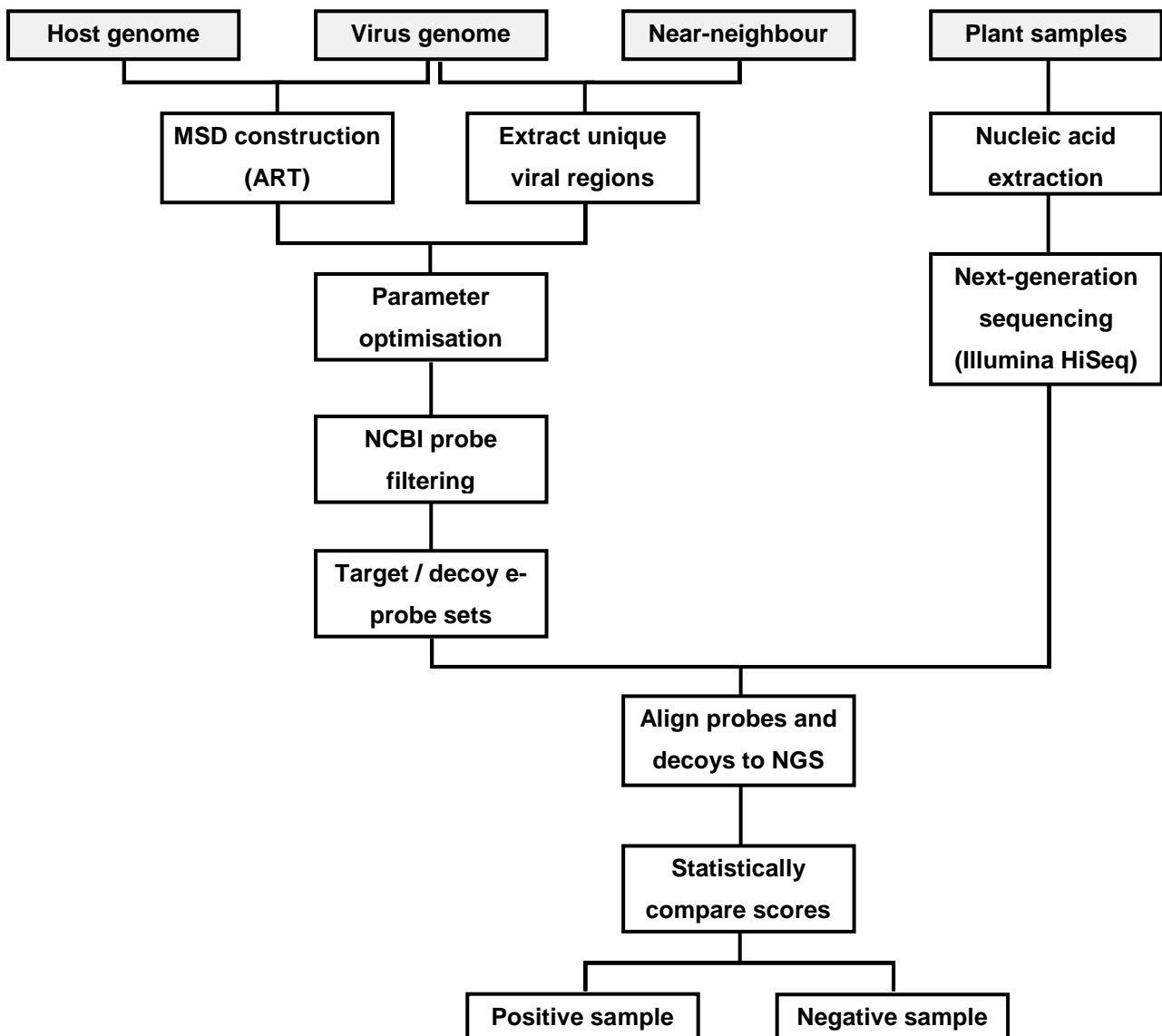
Figure 3.2. Experimental flow of virus specific e-probe design and screening against NGS data. This approached was followed for each of the 11 viruses, individually.

38

Chapter 4: Results and Discussion

4.1 Plant material and nucleic acid extractions

Since CTV goes through a dsRNA intermediate during replication, dsRNA was extracted from infected plants in trial 1, as well as the field samples, to enrich for CTV derived nucleic acids. The double-strandedness of the RNA and the fact that it is often extracted along with proteins, influences the use of spectrophotometry based quantification (Carmichael 2005). As a result the concentration of the dsRNA samples was estimated using gel electrophoresis (1% [w/v] TAE agarose gel) to compare the band intensity to that of the GeneRuler 1kb DNA ladder. The concentration of all the dsRNA samples was predicted to be above 30 ng and no significant differences regarding the quality was observed (results not show). Furthermore, total RNA was successfully extracted from the plants infected with CTV T3 genotype in trial 2 to investigate two additional sequencing library types. The concentration of the total RNA extracted from the three grapefruit plants ranged between 1,531 and 1,937 ng/µl.

4.2 Conventional data analysis

4.2.1 Sequence pre-processing

The datasets generated through next-generation sequencing (NGS) on an Illumina HiSeq 2500 platform are summarised in Table 4.1 and consisted of between 2,013,934 and 39,489,174 reads (approximately 64 gigabases of data in total). The variation observed in the data generated per sample could be ascribed to inaccuracies in the library quantification step, prior to multiplexing. Visualisation of the reads in FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) revealed adaptor contamination, over represented k-mers, and an uneven distribution of nucleotides at the 5′ end. The latter is due to the use of random hexamer primers during the reverse transcription step of the library preparation protocol employed by Illumina RNA-Seq experiments (Hansen *et al.* 2010). This consistent bias in nucleotide frequency is not seen in RNA-Seq experiments utilising alternative library preparation protocols that involves the use of oligo(dT) priming instead (Hansen *et al.* 2010). An improvement in the overall mean quality of the individual sets was observed after adaptor sequences were removed and reads were trimmed and filtered to retain only high quality sequences (Fig 4.1) using Trimmomatic (Bolger *et al.* 2014). Further filtering against citrus host genome sequences to enrich for viral

reads, lead to the removal of between 5.25% and 41.39% of the quality trimmed read pairs, per sample (Table 4.1). The large amount of data mapping to host sequences could indicate ineffective DNase and RNase treatment or host-derived nucleic acid contamination of dsRNA sample after treatment. In total, a substantial amount of sequencing data (between 53.05 and 97.13% of the read pairs) was lost due to insufficient quality, which could be ascribed to substandard library preparations and extraction protocols.

Table 4.1. Sequence data statistics of each sample, before and after processing. Reads remaining after trimming for quality and adaptor sequences were used as input for host filtering against *Citrus sinensis* sequences.

Sample name	Raw data (Gb)	Raw read pairs	Read pairs after quality trim (%)	Read pairs after host filter (%)	Total read pairs lost (%)
SS_RB1	4.37	7353840	3713777 (50.50)	2297196 (30.93)	68.76
SS_RB2	19.74	19744587	684438 (3.47)	566565 (41.39)	97.13
SS_LMS	4.82	4823971	511214 (10.60)	140186 (13.71)	97.09
SS_T68	2.94	5349502	4700591 (87.87)	1867258 (19.86)	65.09
SS_VT	5.94	10328814	9404680 (91.05)	4848914 (25.78)	53.05
T1_RB1	4.29	7678919	6080782 (79.19)	2037902 (16.76)	73.46
T1_RB2	2.82	4988052	3326471 (66.69)	1867500 (28.07)	62.56
T1_LMS	2.63	4664505	3623519 (77.68)	1023613 (14.12)	78.06
T1_T68	1.37	2431582	1808363 (74.37)	518396 (14.33)	78.68
T1_VT	1.39	2434556	928653 (38.14)	677303 (36.47)	72.18
UNK_01	4.36	4355434	1480236 (33.99)	155157 (5.24)	96.44
UNK_02	0.54	1006967	825850 (82.01)	260188 (15.75)	74.16
UNK_03	1.51	2383270	1718648 (72.11)	224763 (6.54)	90.57
UNK_04	6.39	10868594	8612797 (79.24)	2311182 (13.42)	78.74
UNK_05	0.57	912069	590937 (64.79)	201588 (17.06)	77.90
UNK_06	0.58	835430	519636 (62.20)	220527 (21.22)	73.60
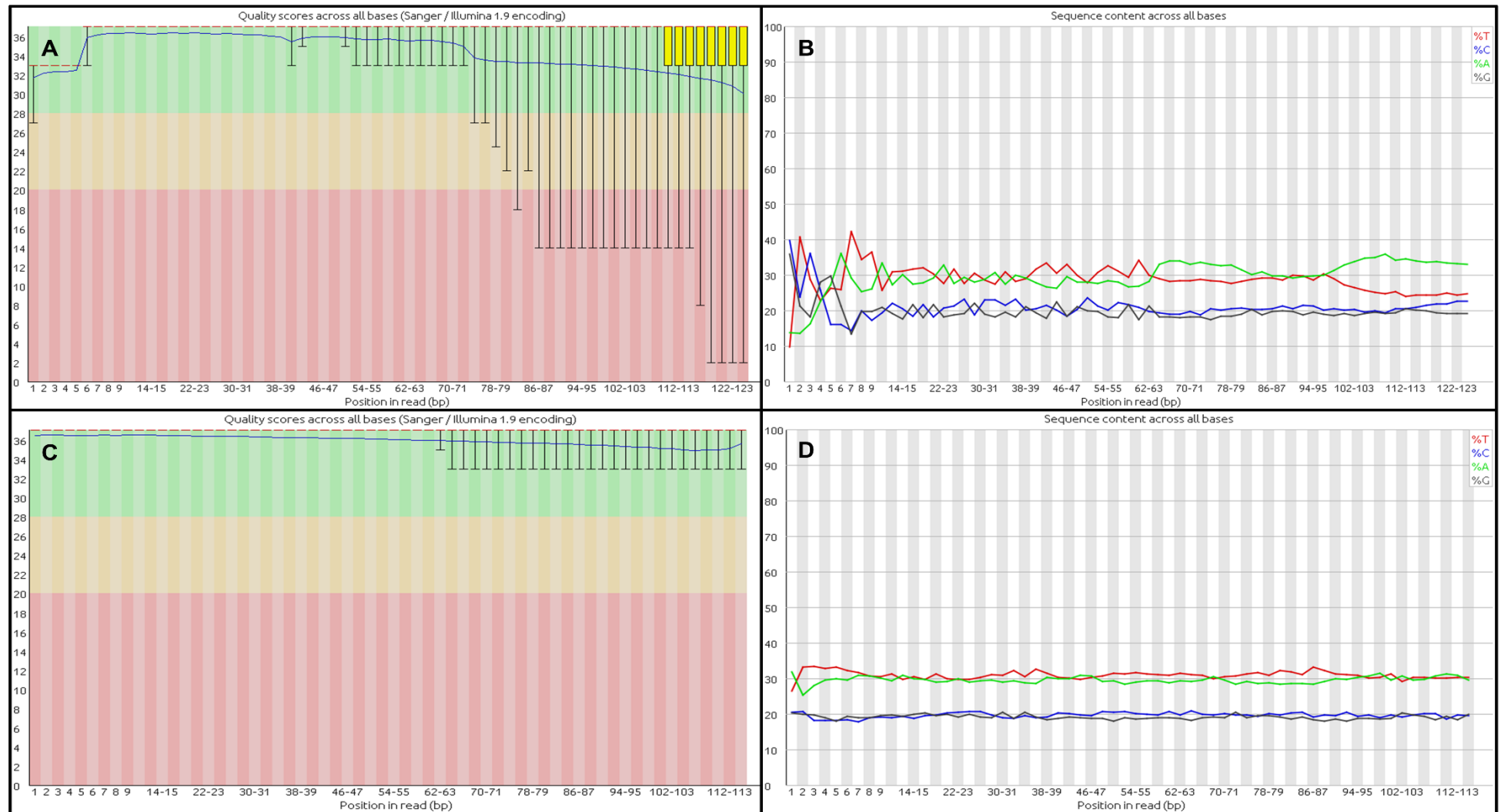
Figure 4.1. Graphical output generated in FastQC illustrating the quality of the sequence dataset of sample 1 (SS_RB1). **A)** The per base quality score distribution where the mean quality score is indicated by the blue line. **B)** The percentage nucleotide composition per base of the raw dataset. **C)** The improvement in quality of the sequence dataset after trimming and filtering. A Phred score of Q20 was used. **D)** After removing the first 10 bases from the 5' end with Trimmomatic's HEADCROP parameter, the uneven nucleotide distribution was no longer evident.

4.2.2 Assembly and homology searching

De novo assemblies were performed using CLC Genomics Workbench. As part of the two-step assembly process, following assembly, contigs were corrected by mapping the reads back to them. This allows for the removal of erroneous contigs, which might have been produced as assembly artefacts. In the instance where reads only map to a specific area on a contig, the uncovered regions were removed, leading to the occurrence of contigs shorter than the minimum length specified (<250 nts) (https://secure.clcbio.com/helpspot). The percentage of reads matched to contigs ranged from 48.29% to 99.02% (Table 4.2). Overall, 81,334 contigs with a total length of 25,459,468 nts were generated from approximately 39.82 million reads (Table 4.2). Compared to the remaining samples, samples SS_T68 and UNK_04 generated more or less 6 times as many contigs. No correlation was seen however, between the amount of data and the number of contigs constructed. The N50 value was used to assess the quality of the assemblies and ranged between 240 and 375 nucleotides, with an average N50 of 305 across all the samples. The highest N50 value obtained was 375 (sample UNK_02) implying that 50% of the 62.79 million bases that were assembled for this particular sample, was assembled into contigs that were >= 374 nts in length. An N50 of 375, although lower than what is expected for genome assembly, appears to be within the range observed in other viral metagenomic studies (Zerbino and Birney 2008, Yang *et al.* 2012, Vázquez-Castellanos *et al.* 2014). It is imperative to note that a large N50 is not necessarily more favourable since it would possibly suggest mis-assemblies and an increase in chimeric contigs (Mende *et al.* 2012, Vázquez-Castellanos *et al.* 2014). The variation observed in the assembly outputs of the individual samples could be due to differences in sequence quality or the fluctuating viral concentrations within each sample. In addition to the basic contig measurements, the average depth of coverage of contigs i.e. number of times on average a base within a contig was covered by the sequencing data, was also reviewed. Contigs with an average coverage greater than 5 and a minimum length of 250 nts were selected (34,872 sequences in total) for successive BLAST analysis.

BLAST searches were performed against a local virus database to allow for the genotyping of the CTV-infected samples, since the lack of variation across large areas of the CTV genome complicates a read mapping based approach. The generated local database contained sequences of the CTV genotypes investigated in this study, as well as a representative of the 10 additional citrus infecting viruses. The local BLAST results were supported with BLAST analysis against the NCBI's non-redundant nucleotide database

(default parameters). The blastn results indicated that 75.78% of the assembled reads originated from citrus-infecting viruses, illustrating the effectiveness of the viral nucleic acid enrichment strategy (dsRNA) used.

Table 4.2. Assembly statistics displaying contig measurements for each sample.

Sample name	Contig measurements				Reads matched (%)
	N50[a]	Average length	Contig count	Total length	
SS_RB1	281	304	2592	787255	3891815 (86.66)
SS_RB2	329	345	1412	487109	1115868 (99.02)
SS_LMS	315	329	1353	445387	251861 (92.22)
SS_T68	325	314	12258	3853179	2681730 (82.01)
SS_VT	336	318	5578	1775855	9225488 (95.29)
T1_RB1	359	343	3906	1340005	3805103 (95.27)
T1_RB2	240	267	6187	1654677	2754472 (75.42)
T1_LMS	250	286	7135	2043947	1222599 (62.04)
T1_T68	240	281	3274	920909	779720 (80.25)
T1_VT	266	294	1676	492859	1203774 (89.92)
UNK_01	320	332	8075	2681611	204002 (91.92)
UNK_02	375	367	2433	892918	276425 (55.03)
UNK_03	309	324	2730	884369	3048712 (97.48)
UNK_04	359	321	17371	5584605	4054016 (92.00)
UNK_05	319	320	2731	872875	209188 (54.98)
UNK_06	257	283	2623	741908	198763 (48.29)

[a] N50 value is a statistic comparable to the median of contig lengths, favouring longer contigs in a particular set. Given a set of contig sequences, the N50 can be defined as the length at which 50% of sequence bases are included into the contigs with a length that equals or exceeds the N50 length.

As expected, CTV was the most prevalent virus, accounting for 99.59% of the virus derived assembled reads. Furthermore, CTV was detected in all of the samples, while *Citrus psorosis virus* was detected solely in UNK_04, one of the field samples (Table 4.3). Samples T1_LMS6 and UNK_06 had the lowest number of CTV-derived reads, 20 and 30% respectively, in comparison to the 51-92% in the remaining 14 samples. This could be attributed to a number of factors including the variable sequence quality amongst the

43

samples or low virus titre. Several of the assembled reads, 24.22% in total, did not share any similarity to the virus sequences in the created database and as a result could not be identified in this manner.

Further characterisation of the contigs generated for the CTV-infected samples (Table 4.3) revealed that each sample contained contigs that aligned to every CTV genotype in the database. Additionally it is also evident that the amount of assembled reads sharing similarity to the VT representative (EU937519.1) is greater than for any other genotype. Samples SS_RB1, SS_RB2, SS_VT, SS_LMS, T1_RB1, and T1_RB2 were the only samples to largely align to only the appropriate genotype used as the inoculum source (highlighted in bold and grey). Whereas the remaining samples either contained assembled reads that were evenly spread out across the genotypes, or that predominantly aligned to a genotype other than the one it tested positive for (highlighted in bold). Although sample SS_LMS, representing infection with the recombinant cross protection isolate LMS 6 had more reads aligning to its own reference sequence, it also aligned significantly well to the VT sequence which could suggest mixed infection with both genotypes.

However, before the inference of a mixed infection status, the high level of sequence similarity amongst the different CTV genotypes and the level of recombination should be taken into account. Several members of VT, T68 and HA 16-5 genotypes alone have been reported to be involved in over 50 different recombination events (Harper *et al.* 2010, Melzer *et al.* 2010). Due to the high level of ORF 1a nucleotide identity (approximately 91%) between T3, T30, and VT species they have been considered to share a common ancestor (Harper *et al.* 2013). In light of this it would suggest the existence of at least two other CTV sources from which RB and T36, as well as HA16-5 and T68 could have descended, respectively (Harper *et al.* 2013). The complexity of the recombination events and the high level of homology within and between genotypes, could possibly elucidate the BLAST results observed for the samples that were presumed to be singly infected with a particular genotype. Furthermore, *de novo* assemblies performed in CLC with data from samples with mixed infections can result in the formation of chimeric contigs, particularly involving reads from the 3′ end of the genome. The overview generated with the *de novo* data might therefore not completely depict what is occurring in reality.

Table 4.3. Read distribution, per sample, across accessions after blastn analyses against local viral database.

Sample name	CTV genotype[a]							Other viruses
	RB1[b]	RB2 (B389-1)	RB2 (B390-5)	LMS6-1	T68	VT	T3	Psorosis[c]
SS_RB1	**3176979**	8219	39798	62	46400	160525	69930	.
SS_RB2	174117	33323	**213063**	118063	11678	16574	6443	.
SS_LMS	170183	5343	163394	**717426**	232869	512635	168681	.
SS_T68	1290	168	2478	1705	**203968**	11272	4160	.
SS_VT	62628	1745	56279	38258	73071	**8687844**	35376	.
T1_RB1	**1912623**	960	20199	14948	1507040	95284	23175	.
T1_RB2	7404	179	**2255069**	30605	45102	87630	29685	.
T1_LMS	56769	1037	37648	27143	72105	180527	**231215**	.
T1_T68	42263	208	11091	**421646**	28877	46513	13012	.
T1_VT	3477	237	11482	**531016**	15727	74644	335399	.
UNK_01	7721	5031	9488	**78351**	14253	21108	14767	.
UNK_02	6570	1553	17095	40479	29615	51628	27664	.
UNK_03	792398	3265	**1340819**	373815	47303	78827	239064	.
UNK_04	430292	185057	795764	623738	481101	618767	104679	126307
UNK_05	6475	2226	18910	12233	17697	**57010**	13537	.
UNK_06	2359	865	12904	6196	12647	**33027**	13325	.

[a] Bold entries represent the predominant CTV genotype assigned to a sample.

[b] Samples predominantly aligning to only the appropriate genotype used as the inoculum source, highlighted grey.

[c] Psorosis was the only other virus identified in one sample (UNK_04).

4.3 E-probe based bioinformatics pipeline

4.3.1 E-probe design and optimisation

Virus specific e-probes of varying minimum lengths (20, 40, 60, 80, 100, 120 and 140 nts) were generated for the eleven viruses individually by comparing each virus to a near neighbour from the same family (Table 3.3). In the case of *Citrus leprosis virus C*, which is still unassigned, *Orchid fleck virus* was used as a near neighbour based on their phylogenetic relationship reported in Kondo *et al.* (2006). The number of candidate probes generated per target virus, ranged between 57 and 211 (Table 4.4). The candidate probes were optimised for the minimum probe length and the e-value threshold required using oMSDs. In addition to calculating precision (the ratio of positive virus hits to total hits), the number of matches were recorded in each instance (Figure 4.2). For each virus, the number of matches and hits increased with the viral abundance in the oMSD. This is expected, since an increase in virus sequences within the dataset would increase the probability of an e-probe, or parts thereof, being represented in data. The number of matches were always greater than the number of hits received; suggesting that single probes recurrently generated multiple hits in an oMSD.

Table 4.4. Comparison of the number of e-probes generated across eleven citrus infecting viruses before and after BLAST filtering.

Target virus	Genome size (kb)	Candidate probes (>= 20	Final e-probes	Total probe length (nts)
CTV	19.25	211	209	10853
CPsV	8.19 (RNA1) 1.64 (RNA2) 1.45 (RNA3)	103	95	3407
CTLV	6.49	85	62	3095
CVV	2.31 (RNA1) 3.43 (RNA2) 2.91 (RNA3)	71	44	1695
CYMV	7.56	69	63	3784
CiLRV	2.29 (RNA1) 2.99 (RNA2) 3.40 (RNA3)	62	46	1378
CLBV	8.75	81	80	293
CiLV-C	8.74 (RNA1) 4.99 (RNA2)	130	119	6799
ICRSV	7.56	57	41	2417
CYVCV	7.53	72	37	3909
SDV	6.79 (RNA1) 5.34 (RNA2)	118	50	6278

During optimal minimum probe length optimisation, the candidate probe sets of different minimum lengths showed absolute (100%) precision for all the viruses, which was also seen, specifically for viruses, during the evaluation of the EDNA pipeline (Stobbe *et al.* 2013). As expected the number of matches was directly proportional to the number of probes and therefor indirectly proportional to the minimum length of the e-probes. The probe sets lacking the shorter probes (those with probes >=120 and >=140 nts) generated the lowest number (if any) of hits in the very low virus percentage oMSDs. The e-probe sets belonging to each virus with minimum length of 20 nucleotides were selected for e-value threshold determination, as it had the highest number of matches and still displayed precision of 100% when screened against oMSDs of varying viral concentrations (including those representing very low virus abundance levels). The selected minimum probe length (20 nts) additionally facilitates the generation of a larger number of probes in a set for screening purposes.

Different e-values were implemented to study the occurrence of potential false negative or positive results. No differences were seen since all of the viruses were detected in their respective oMSDs, at each e-value (10^{-3} – 10^{-9}), even in the very low viral abundance datasets. For all the viruses, the total number of matches was not affected by increased stringency (lower e-values) and no false positives were detected; however, the total number of hits was reduced at the lower e-values. Since no false positives were detected at an e-value cut-off of 1×10^{-3}, this value was selected for candidate specificity screening against NCBI, as well as for screening e-probe sets against NGS data. Larger, more complex pathogens, such as bacteria however may require the use of a more stringent e-value to reduce the potential occurrence of false positives (Stobbe *et al.* 2013).

NCBI filtering of the optimised candidate probes decreased the number of probes per virus resulting in the number of final e-probes ranging from 37 to 209 per virus (Table 4.4). The genome sizes for the viruses ranged from 1,447 to 19,251 bases and were compared to the number of e-probes designed (Table 4.4). No correlation was seen between the size of the genome and the probe number for all the target viruses.
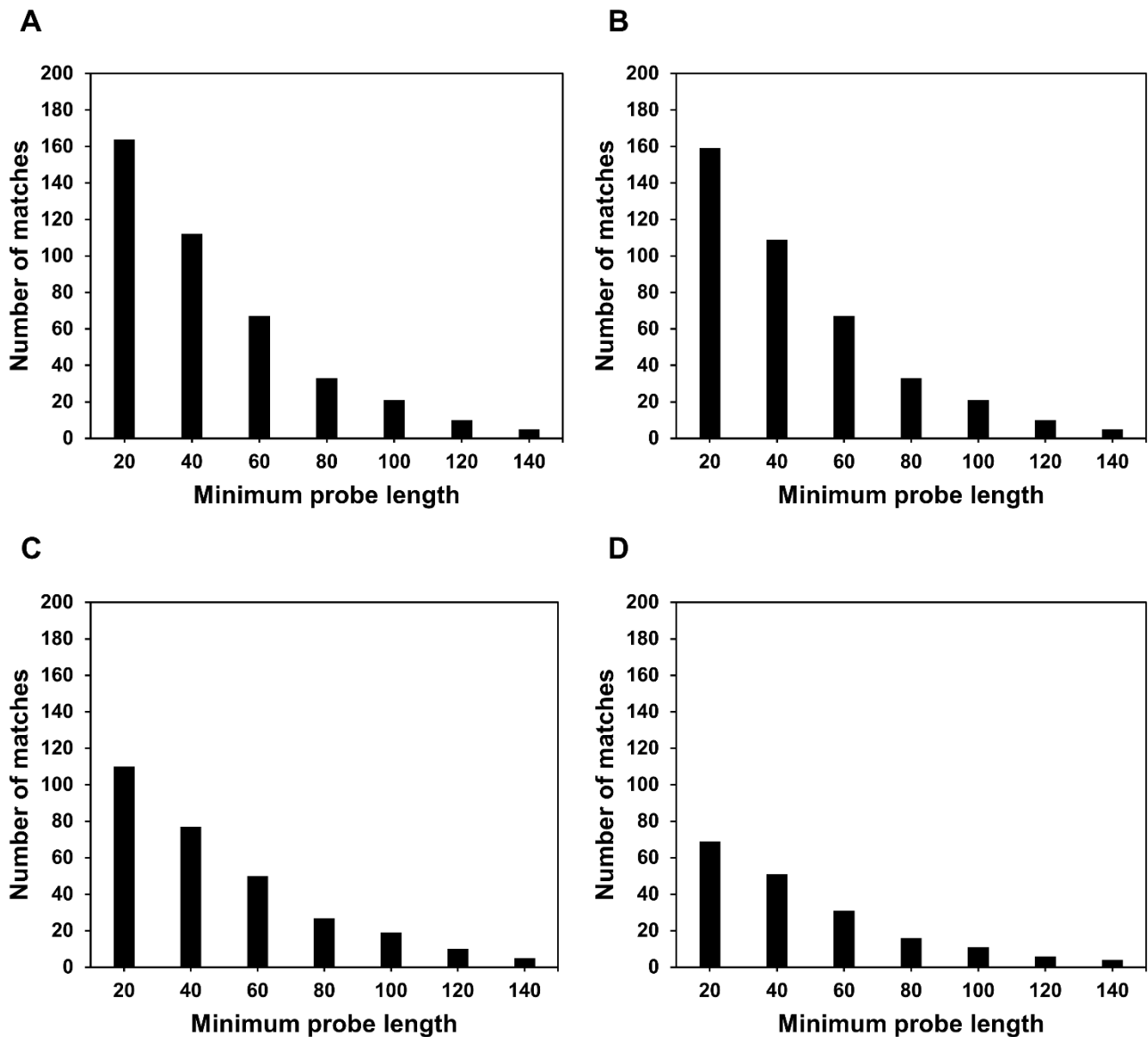
Figure 4.2. Probe length optimisation of CTV candidate probes. The number of positive matches obtained for each minimum probe length in oMSDs containing **A)** 15% (medium – high), **B)** 5% (medium), **C)** 1% (low), **D)** less than 1% (very low) viral reads. The profile obtained with the medium – high oMSD (15%) is identical to that obtained with the final oMSD category, very high (25%).

4.3.3 E-probe based virus detection

4.3.3.1 Detection using simulated datasets

The probe based detection approach was firstly evaluated using simulated datasets representing mixed infections with varying viral content. Each probe set could accurately detect the target virus/es in the evaluation-MSDs (eMSDs) producing a positive diagnostic result (Table 4.5). In the absence of a specific virus the eMSDs tested negative when screened using the e-probe set for that particular virus. Cross alignment between probe sets constructed for ICRSV, CYVCV, CiLRV, and CVV occurred where e-probe matches were detected in mixed infection eMSDs that did not contain the target virus (Table 4.5).

The 11 probe sets were additionally evaluated for their ability to detect different isolates of each target virus species using single infection eMSDs. These results displayed a similar trend as seen with mixed infection eMSDs, with each probe set designed from a particular accession being able to identify the target virus and all its corresponding isolates (Supplementary Table 1). This observation may however not be true for more genetically divergent virus species as observed in the case of GLRaV-3 in grapevine (Visser *et al.* 2016). Single infection eMSDs of the near neighbour accessions used during e-probe design were also simulated to aid in testing the specificity of the individual probe sets, and all tested negative for the presence of the 11 target viruses. CTV specific probes hit only, and was the only probes to hit, against CTV eMSDs (Supplementary Table 1). Once more probes designed for ICRSV, CYVCV, CiLRV, CVV and CLBV generated non-specific hits in simulated datasets constructed with the isolates of other viruses. These hits however did not generate a signal significant enough to shift the *P*-value below 0.01 and the final diagnostic result therefore remained negative (Supplemental Table 1). *Indian citrus ringspot virus* probe hits were observed in the eMSDs representing infection with CYVCV isolates (Supplemental Table 1), and vice versa. This reciprocal interaction was also seen between CiLRV and CVV.

Genome analysis of the positive sense RNA virus, CYVCV, revealed that its ORFs coincided with those of ICRSV regarding length, number, and position (Rustici *et al.* 2000, Loconsole *et al.* 2012). A study by Loconsole *et al.* (2012) estimated that the two viral genomes share a nucleotide identity of approximately 74% and an amino acid identity, for the coat protein gene specifically, of 97-98%. This along with phylogenetic analysis placing CYVCV on the same branch as ICRSV has led to the classification of CYVCV as a putative member of the *Alphaflexiviridae* genus, *Mandarivirus* (Loconsole *et al.* 2012). The cross alignment

observed with the probe sets of these two viruses however, cannot fully be explained by the previously mentioned similarities, since ICRSV was used as the near neighbour for CYVCV during probe design. The interaction between CiLRV and CVV however could possibly be as a result of the homology levels between these two Ilarviruses. Based on serological similarities, CiLRV and CVV have been suggested to originate from a common ancestor upon its transmission to cultivated citrus plants (Lovisolo 1993). The gene encoding the CP, a requirement for virus infection, of CVV is highly conserved and has been documented to share sequence similarities to that of CiLRV (between 62 and 63%) (Gonsalves and Fulton 1977, Scott and Ge 1995, Bennani *et al.* 2002).

Overall, no false positive results were observed. The non-specific matches, indicative of the detection limit of the pipeline, could potentially be improved by altering the e-value threshold or the minimum probe length used (Stobbe *et al.* 2013, Stobbe *et al.* 2014). Removing the probes generating the non-specific hits from the set is another option, although this would result in an overall decrease of hits to the target virus as well.

Table 4.5. E-probe based virus detection (*P* value indicated) of eleven citrus viruses in simulated NGS datasets (eMSDs) representing infection with multiple viruses.

Probe set	All viruses[a]	Except CTV[b]	Except CPsV	Except CTLV	Except CVV[c]	Except CYMV	Except CiLRV	Except CLBV	Except CiLV-C	Except ICRSV	Except CYVCV	Except SDV
CTV	7.25E-11		7.25E-11	7.25E-11	7.25E-11	7.25E-11	7.25E-11	7.25E-11	7.25E-11	7.25E-11	7.25E-11	7.25E-11
CPsV	1.10E-32	1.10E-32		1.10E-32	1.10E-32	1.10E-32	1.10E-32	1.10E-32	1.10E-32	1.10E-32	1.10E-32	1.10E-32
CTLV	7.62E-22	7.62E-22	7.62E-22		7.62E-22	7.62E-22	7.62E-22	7.62E-22	7.62E-22	7.62E-22	7.62E-22	7.62E-22
CVV	5.61E-16	5.61E-16	5.61E-16	5.61E-16	0.85	5.61E-16	5.61E-16	5.61E-16	5.61E-16	5.61E-16	5.61E-16	5.61E-16
CYMV	3.57E-22	3.57E-22	3.57E-22	3.57E-22	3.57E-22		3.57E-22	3.57E-22	3.57E-22	3.57E-22	3.57E-22	3.57E-22
CiLRV	1.03E-16	1.03E-16	1.03E-16	1.03E-16	1.03E-16	1.03E-16	0.72	1.03E-16	1.03E-16	1.03E-16	1.03E-16	1.03E-16
CLBV	9.21E-28	9.21E-28	9.21E-28	9.21E-28	9.21E-28	9.21E-28	9.21E-28		9.21E-28	9.21E-28	9.21E-28	9.21E-28
CiLV-C	1.50E-40	1.50E-40	1.50E-40	1.50E-40	1.50E-40	1.50E-40	1.50E-40	1.50E-40		1.50E-40	1.50E-40	1.50E-40
ICRSV	3.67E-15	3.67E-15	3.67E-15	3.67E-15	3.67E-15	3.67E-15	3.67E-15	3.67E-15	3.67E-15	0.57	3.67E-15	3.67E-15
CYVCV	3.03E-25	3.03E-25	3.03E-25	3.03E-25	3.03E-25	3.03E-25	3.03E-25	3.03E-25	3.03E-25	3.03E-25	0.77	3.03E-25
SDV	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	3.19E-40	

[a] Probe sets were able to accurately detect the target virus they were designed for (*P* value <= 0.05).

[b] No entry indicates the probes for the target virus did not hit in its absence.

[c] Entries highlighted in grey represent the instances where the probes for a specific virus generated hits in the eMSD, regardless of the absence of the virus.

4.3.3.2 Detection using NGS datasets

The e-probe based detection system was validated using the probe set constructed for CTV from the genome sequence of a VT genotype (EU937519) and employing it in searches of NGS data obtained from plant samples. This e-probe set positively detected CTV in sequence data generated from dsRNA of plants infected with the VT genotype or other CTV genotypes, as well as all six of the field samples, demonstrating that this probe set is able to detect CTV infection regardless of the genotype (Table 4.6). There was no instance where all 209 probes hit sequences within the plant sequence data, including the sample infected with a VT genotype. This can be due to the absence of genome areas represented by the specific probes in the NGS data or isolate variation since the samples were reportedly infected with VT Maxi x3 isolate. These datasets were also screened with the remaining ten probe sets, which revealed the presence of *Citrus psorosis virus* in one of the field samples in addition to CTV (Table 4.6). This result verified what was seen with the *de novo* assemblies and was further confirmed with end-point RT-PCR (result not shown).

To evaluate individual probe performance, three datasets, representing each of the two different sequencing libraries (sRNA and total RNA) were selected. The datasets were subjected to the random sampling of 1 million reads 100 times for each sample and screened with the CTV e-probe set (Figure 4.3). In both instances the CTV probe set was able to accurately detect the presence of the virus. However, the ability of each individual probe to hit virus-derived sequences 100% of time was markedly lower in the sRNA data compared to the transcriptome data (Figure 4.3). The coordinates of the probes on the genome were compared to their difference in performance across the two library types and no significant correlation was observed as the probes span the entire CTV genome (Figure 4.4).

Within the transcriptome dataset, approximately 70% of the total probes were detected in all 100 subsets (i.e. 100% of the time) while the remaining probes hit less than 20% of the time or not at all. This was not observed within the sRNA dataset, with only 3% of the probes hitting 100% of the time (Figure 4.3). More variation amongst individual probe performance is observed within the sRNA dataset with 43.5% of the probes not hitting at all for this library type. This variation in the performance of the two library types can be attributed to fundamental differences in the sequence data. The percentage of virus-derived sRNA sequences is dependent on the host defence response, unlike the virus-associated reads in the transcriptome data. The datasets used all contained the same number of sequence reads, but the percentages of virus-derived reads were 0.69% and 2.4% for sRNA and

transcriptome data, respectively. Increasing the sRNA subsample size to 5.7 million reads, to make the virus-derived data comparable to that of the transcriptome data, showed a similar decrease in probe performance to 1 million reads (Figure 4.3). This suggested that the percentage of virus-derived data in NGS datasets has a limited effect on the probe performance.

The influence of read length was evaluated by using only the first 23nt of transcriptome data reads. The probe performance with these shortened transcriptome reads showed a profile similar to the sRNA datasets' profile (Figure 4.3). This highlighted the significant impact of read length on the performance of the probes, with longer reads being more favourable. Probes that were not detected across both library types were not removed from the final probe set, because they might be homologous to a region on the virus genome that is not represented in this specific sequence data.

Table 4.6. E-probe based virus detection in NGS datasets of plant samples with virus specific probe sets for CTV, CPsV, CTLV, CVV, and CYMV.

Sample name	CTV[a]		CPsV[b]		Probe set CTLV		CVV		CYMV	
	Positive probes	*P* value	Positive probes	*P* value	Positive probes	*P* value	Positive probes	*P* value	Positive probes	*P* value
SS_RB1	**97/209**	**2.25E-16**	0/95	.	0/62	.	0/44	.	0/63	.
SS_RB2	**181/209**	**6.02E-53**	0/95	.	0/62	.	0/44	.	0/63	.
SS_LMS	**137/209**	**4.50E-31**	0/95	.	0/62	.	0/44	.	0/63	.
SS_T68	**116/209**	**9.66E-23**	0/95	.	0/62	.	0/44	.	0/63	.
SS_VT	**188/209**	**5.59E-57**	0/95	.	0/62	.	0/44	.	0/63	.
T1_RB1	**139/209**	**6.15E-32**	0/95	.	0/62	.	0/44	.	0/63	.
T1_RB2	**79/209**	**2.31E-11**	0/95	.	0/62	.	0/44	.	0/63	.
T1_LMS	**134/209**	**8.45E-30**	0/95	.	0/62	.	0/44	.	0/63	.
T1_T68	**112/209**	**2.62E-21**	0/95	.	0/62	.	0/44	.	0/63	.
T1_VT	**134/209**	**8.45E-30**	0/95	.	0/62	.	0/44	.	0/63	.
T2_HVK4	**98/209**	**1.34E-16**	0/95	.	0/62	.	0/44	.	0/63	.
T2_HVK5	**101/209**	**4.56E-19**	0/95	.	0/62	.	0/44	.	0/63	.
T2_HVK6	**125/209**	**2.78E-26**	0/95	.	0/62	.	0/44	.	0/63	.
T2_HVK16	**189/209**	**7.02E-59**	0/95	.	0/62	.	0/44	.	0/63	.
T2_HVK17	**175/209**	**2.71E-55**	0/95	.	0/62	.	0/44	.	0/63	.
T2_HVK18	**158/209**	**3.45E-41**	0/95	.	0/62	.	0/44	.	0/63	.
UNK_01	**189/209**	**1.44E-57**	0/95	.	0/62	.	0/44	.	0/63	.
UNK_02	**64/209**	**6.11E-08**	0/95	.	0/62	.	0/44	.	0/63	.
UNK_03	**196/209**	**8.98E-62**	0/95	.	0/62	.	0/44	.	0/63	.
UNK_04	**190/209**	**3.69E-58**	**53/95**	**3.09E-11**	0/62	.	0/44	.	0/63	.
UNK_05	**137/209**	**4.50E-31**	0/95	.	0/62	.	0/44	.	0/63	.
UNK_06	**105/209**	**6.41E-19**	0/95	.	0/62	.	0/44	.	0/63	.

[a] Bold entries indicate a positive diagnostic call (*P* value <= 0.05), while nonbold entries indicate no presence of the target virus.

[b] No *P* value generated if none of the probes for a specific virus hit.

Table 4.7. E-probe based virus detection in NGS datasets of plant samples with virus specific probe sets for CiLRV, CLBV, CiLV-C, ICRSV, CYVCV, and SDV.

Sample name	CiLRV[a] Positive probes	P value	CLBV[b] Positive probes	P value	CiLV-C Positive probes	P value	ICRSV Positive probes	P value	CYVCV Positive probes	P value	SDV Positive probes	P value
SS_RB1	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
SS_RB2	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
SS_LMS	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
SS_T68	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
SS_VT	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T1_RB1	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T1_RB2	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T1_LMS	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T1_T68	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T1_VT	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T2_HVK4	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T2_HVK5	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T2_HVK6	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T2_HVK16	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T2_HVK17	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
T2_HVK18	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
UNK_01	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
UNK_02	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
UNK_03	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
UNK_04	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
UNK_05	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.
UNK_06	0/46	.	0/80	.	0/119	.	0/41	.	0/37	.	0/50	.

[a] Bold entries indicate a positive diagnostic call (P value <= 0.05), while nonbold entries indicate no presence of the target virus.

[b] No P value generated if none of the probes for a specific virus hit.
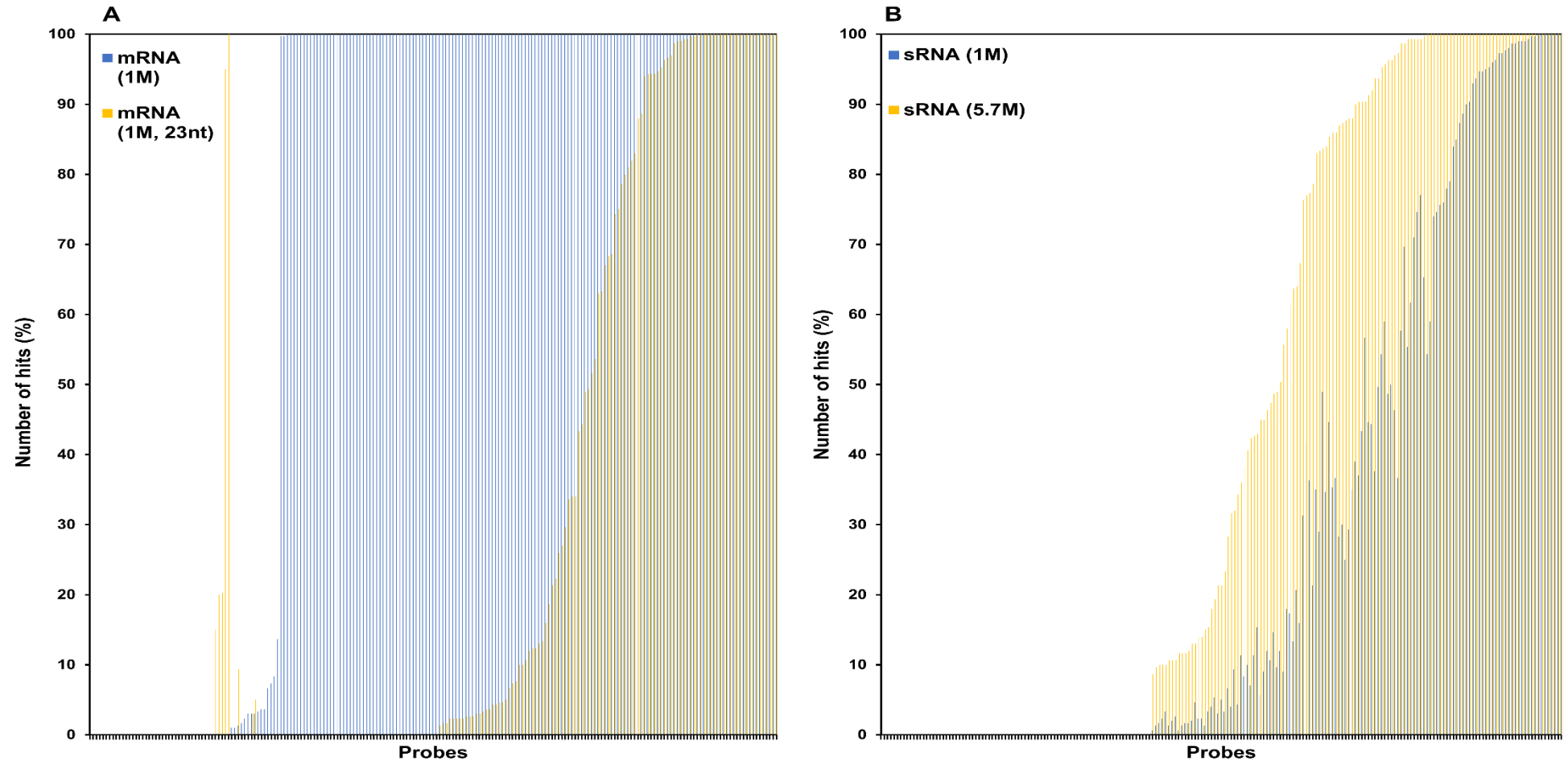
Figure 4.3. Performance evaluation of CTV e-probes across different library types. **A)** Illustration of probe performance for transcriptome data (subsample size of 1 million reads) with full-length paired reads of 125 nts compared to the first 23 nts of the forward reads of the same samples. Results are firstly arranged according to number of hits against the full-length transcripts (lowest to highest), followed by arrangement according to the number of hits against the trimmed reads (lowest to highest). **B)** Probe performance against sRNA data with a subsample size of 1 million compared to a subsample size of 5.7 million. Results are firstly arranged according to number of hits against the 1 million read-subsamples (lowest to highest), followed by arrangement according to the number of hits against 5.7 million read-subsamples (lowest to highest).

56
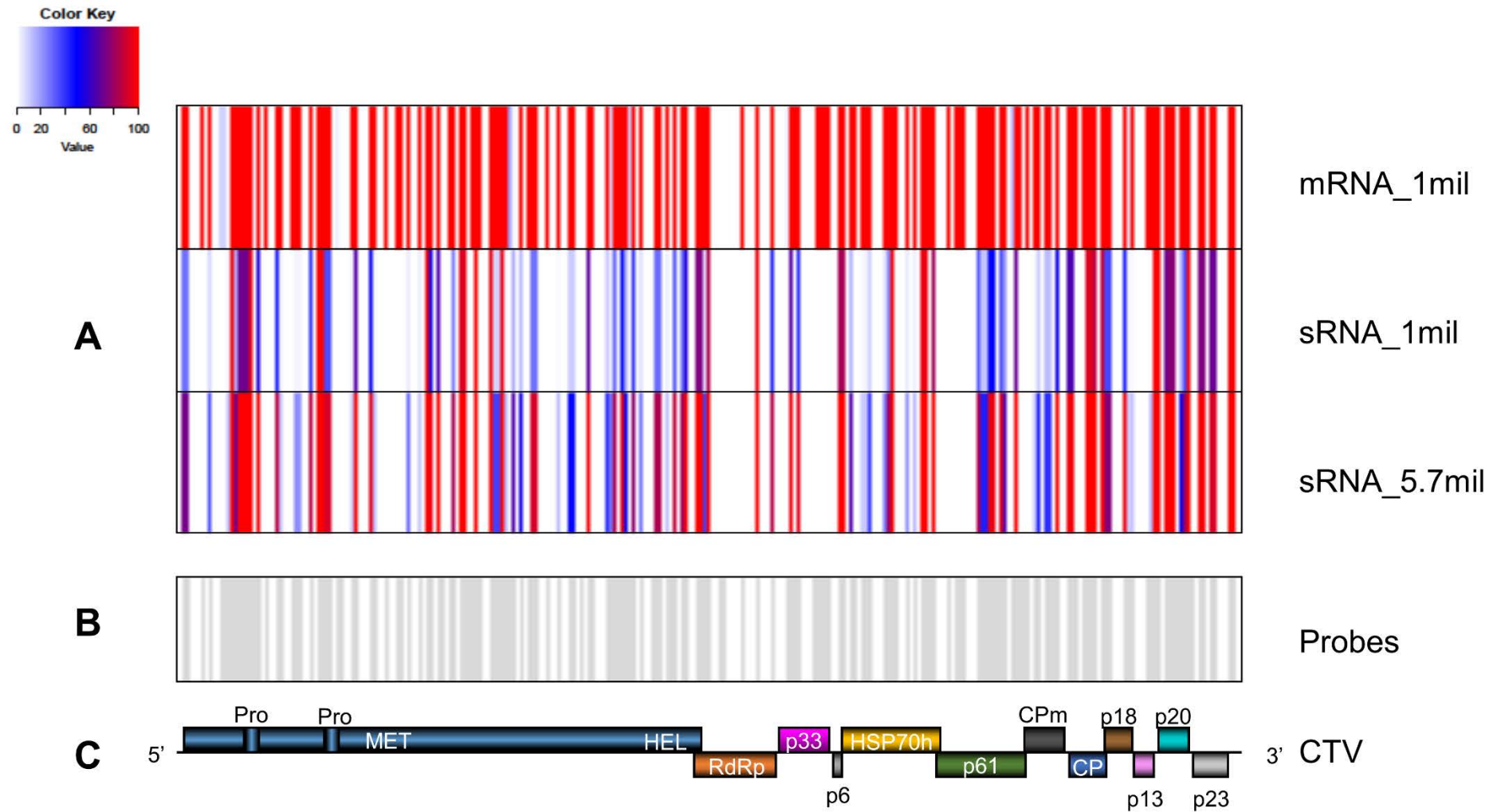
Figure 4.4. Evaluation of CTV probe performance across different library types by position in the CTV genome. **A)** Heat map of the number of times a specific probe hit in the four datasets. **B)** CTV genomic regions covered by the 209 e-probes, and **C)** CTV genome organisation with 112 ORF's.

Chapter 5: Conclusions and Prospects

Amidst the complex of graft-transmissible diseases of citrus that present a serious threat to local and international industries, are diseases caused by viruses. Infection with one such viral pathogen, *Citrus tristeza virus* (CTV), has led to the destruction of numerous citrus producing areas around the world. The number of distinct syndromes caused by CTV, as well as the array of factors influencing host response, complicates plant biosecurity procedures and reinforces the need for early detection. The advent of next generation sequencing (NGS) has transformed plant virology by providing an alternative approach to existing molecular assays, that is able to identify low-titre viruses, as well as novel viruses and virus variants (Al Rwahnih *et al.* 2009, Mokii *et al.* 2012). Improvements to NGS technologies have made it possible to simultaneously detect viruses in metagenomic samples, facilitating viral biodiversity and disease etiology studies. While there are multiple advantages to coupling NGS with metagenomics for virus detection, the exponential rate at which data is generated makes this approach exceedingly computationally intensive (MacDiarmid *et al.* 2013). The current study focussed on evaluating an alternative e-probe bioinformatic approach proposed by Stobbe *et al.* (2013) for its ability to detect multiple citrus-infecting viruses simultaneously in NGS data.

The EDNA pipeline utilises short, virus-specific sequences (e-probes) to screen NGS datasets and permits the user to define the amount of probes used, as well as the size of the sequence dataset. Previous corroboration of this approach entailed the use of a series of simulated datasets with representatives of a variety pathogens in a general plant background (*Vitis vinifera*) (Stobbe *et al.* 2013). In this study, e-probes were successfully developed for 11 citrus viruses of global economic importance by comparing each virus to a near-neighbour from the same family. The e-probes were optimised in terms of their minimum length and the E-value threshold used when querying a dataset, using a range of optimisation mock sequence datasets (oMSDs). The oMSDs consisted of genome sequences belonging to both the host and the target virus, differing in the level of virus derived reads present. No false positive hits or matches were generated during the optimisation steps for all oMSDs (including those containing less than 1% viral reads). The specificity of the e-probes was further improved by screening them against NCBI's nt database to eliminate e-probes that could potentially render false positives.

The e-probe sets were able to accurately detect their cognate viruses in simulated evaluation-MSDs (eMSDs), representing single and mixed infections, without any false negatives or positives. Additionally, the designed e-probes were also effective in detecting different isolates of the same citrus virus in simulated datasets containing all the target virus isolates available on GenBank. The lack of false positives or negatives observed for any of the viruses, indicate the specificity and sensitivity of the probe sets.

The validation of the e-probe based detection system was extended for CTV by employing the probe set constructed from the VT genotype (EU937519) genome sequence in searches against RNA-Seq datasets of CTV-infected plants and field samples. This e-probe set positively detected CTV in sequence data from dsRNA of plants infected with the VT genotype or other CTV genotypes, demonstrating that this probe set is able to detect CTV infection regardless of the genotype. The datasets were also screened with the remaining ten probe sets, which revealed the presence of CPsV in one of the field samples in addition to CTV. The latter was confirmed with end-point RT-PCR and coincided with the results obtained through *de novo* assemblies and BLAST analysis. The evaluation of individual CTV probe performance across two different NGS data types, sRNA and transcriptome, revealed that the probe set was able to accurately detect the virus, regardless of the type of the input dataset.

Future research should focus on investigating the performance of the remaining nine e-probe sets (CTV and CPsV excluded) in NGS datasets obtained from plants infected with these particular viruses. The latter would be beneficial in order to substantiate the results observed with the simulated datasets. Research into the generation of genotype specific CTV probes would also be of interest. Stobbe *et al.* (2014) investigated the design of strain specific probes for *Plum pox virus* (PPV), using PPV strains as near-neighbours and evaluated these probes in both simulated and NGS sequence data from infected plants. The level of recombination between CTV genotypes however complicates this process for CTV specifically, as most of the genotypes are homologous with at least one of the other genotypes across large areas of the genome.

The e-probed based bioinformatic pipeline shifts the focus of traditional metagenomic data analysis from attempting to establish a complete profile of all organisms within a given sample, to screening for multiple known viruses simultaneously. Searching for the presence of sequences specific to target viruses not only alleviates the computational workload but also reduces the time required to make a diagnostic call. This is particularly useful in agricultural sectors where products or plant material require specific viral tests prior to

imports. Our results show that an e-probe based assay for the detection of known viruses in a complex host such as citrus is not only effective with the use of simulated data, but also with different types of NGS data. The statistical tests employed add confidence to the diagnostic call, unlike the often subjective, user-based review process involved in conventional NGS approaches. Furthermore, the flexibility of this approach allows for it to be automated and adjusted for viruses infecting multiple alternate hosts (Visser *et al.* 2016). The probes generated, along with all the appropriate Python scripts used in this study are freely available with the user-friendly software, Truffle (https://sourceforge.net/projects/truffle).

Supplementary data

Supplemental Table 1.A. Species and strain-specific primer sequences used in a two-step RT-PCR to amplify Citrus tristeza virus (CTV) RNA as per Roy *et al.* (2010).

CTV strain	Polarity	Primer sequences from 5′ to 3′	Annealing temp. (°C)	Product size (bp)
B165 (T68)	Sense	GTT AAG AAG GAT CAC CAT CTT GAC GTT GA	59	510
	Antisense	AAA ATG CAC TGT AAC AAG ACC CGA CTC		
T3	Sense	GTT ATC ACG CCT AAA GTT TGG TAC CAC T	60	409
	Antisense	CAT GAC ATC GAA GAT AGC CGA AGC		
VT	Sense	TTT GAA AAT GGT GAT GAT TTC GCC GTC A	60	302
	Antisense	GAC ACC GGA ACT GCY TGA ACA GAA T		
T30	Sense	TGT TGC GAA ACT AGT TGA CCC TAC TG	60	206
	Antisense	TAG TGG GCA GAG TGC CAA AAG AGA T		

Supplemental Table 1.B. Species and strain-specific primer sequences used in a two-step RT-PCR to amplify Citrus tristeza virus (CTV) RNA as per Cook *et al.* (2016).

CTV strain	Polarity	Primer sequences from 5′ to 3′	Annealing temp. (°C)	Accession no. for nucleotide position	Nucleotide positions of primer	Product size (bp)
CTV generic	Sense	TCT GAT TGA AGT GGA CGG AAT AAG	62	NC_001661	19 019	157
	Antisense	GCT TAG ACC AAC GAG AGG ATA			19 155	
RB: group1[a]	Sense	AGT GGT GGA GAT TAC GTT G	60	FJ525433	1974	628
	Antisense	TAC ACG CGA CAA ATC GAG			2584	
RB: group2[b]	Sense	CGG AAG GGA CTA CGT GGT	60	FJ525434	1976	658
	Antisense	CGT TTG CAC GGG TTC AAT G			2615	
T36	Sense	GGT GTA AGG AAG CGT GTG TCG CAT TTA	66	NC_001661	5641	537
	Antisense	ACC TGC ACC GTC TAA CAA CAT CAT CG			6152	
HA16-5	Sense 1	TAG GAA GGG TCA CTG CCC TGA CA	56	GQ454870	2128	610
	Antisense	GTA AGT ATC TAA AAC CAG GAG			2717	
	Sense 2	CGA CAA GTG CAT TAC GTC TCA G			2563	176

[a] RB group 1 includes genotypes NZRB-TH28, NZRB-M12, NZRB-G90 and HA18-9.

[b] RB group 2 includes genotypes NZRB-TH30, NZRB-M17 and Taiwan-Pum/SP/T1.

Supplemental Protocol 1. Two step RT-PCR for the detection of CTV.

1. **cDNA synthesis:**
   - Performed using RevertAid H Minus Reverse Transcriptase (Thermo Fisher Scientific Inc., Waltham, MA, USA) with modifications to the manufacturer's instructions.
   - RNA template (0.5–1 µg total RNA) and antisense primer incubated at 65°C for 3 min and chilled on ice.
   - 40 units of RT enzyme and 10 units of RiboLock RNase Inhibitor (Thermo Fisher Scientific) used per reaction.
   - Reverse transcription performed at 50°C for 60 min followed by inactivation at 85°C for 5 min.

2. **PCR diagnostic**
   - Total reaction volume of 20 µl using the GoTaq® Hot Start Green Master Mix (Promega Corp., Madison, WI, USA) and 2 µl cDNA.
   - Cycling parameters:
     - 95°C for 3 min
     - 95°C for 20 s (35 cycles)
     - 30 s at primer annealing temperature
     - 72°C for 20 s
     - Final extension of 72°C for 5 min

Supplemental Table 2. E-probe based virus detection of eleven citrus viruses in simulated NGS datasets (MSDs) representing single infections of all available isolates.

Accession	Virus	Isolate	Probe set										
			CTV[a]	CPsV[b]	CTLV	CVV	CYMV	CiLRV	CLBV	CiLV-C	ICRSV	CYVC	SDV
KU358530	CTV	CA-RB-AT35	7.19E-12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU356770	CTV	CA-RB-AT25	3.41E-13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU361339	CTV	CA-VT-AT39	5.29E-63	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU361340	CTV	CA-RB-115	3.41E-13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU578007	CTV	CA-T30-AT4	4.64E-35	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU589212	CTV	CA-S1-L	5.43E-17	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU589213	CTV	CA-S1-L65	2.25E-16	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FJ525436	CTV	NZ-B18	5.57E-36	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JQ911664	CTV	VT-CT11A	5.59E-57	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DQ151548	CTV	VT-T318A	2.15E-56	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
AB046398	CTV	SY strain	1.63E-53	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC748392	CTV	SG29	1.18E-54	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JQ061137	CTV	VT-AT1	8.01E-52	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517492	CTV	VT-FS703	5.49E-70	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EU937519	CTV	VT	5.49E-70	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517494	CTV	VT-FS701	5.49E-70	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517493	CTV	VT-FL202	1.69E-65	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC262793	CTV	L192GR	9.14E-67	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
AF001623	CTV	unknown	8.12E-45	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
HM573451	CTV	VT-Kpg3	5.51E-48	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EU857538	CTV	SP	5.51E-48	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC525952	CTV	T3	2.92E-41	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EU076703	CTV	T68-B165	4.64E-35	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC333868	CTV	T68-CT-ZA3	3.81E-26	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JQ965169	CTV	T68-1	2.23E-22	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JQ911663	CTVV	T68-CT14A	7.40E-38	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EU937520	CTV	T30	1.05E-33	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
AF260651	CTV	T30	8.17E-33	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517491	CTV	T30-FS703	1.05E-33	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517490	CTV	T30-FL278	2.25E-32	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Accession	Virus	Isolate	Probe set										
			CTV[a]	CPsV[b]	CTLV	CVV	CYMV	CiLRV	CLBV	CiLV-C	ICRSV	CYVC	SDV
KC748391	CTV	Bau282	3.75E-34	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Y18420	CTV	T30-T385	3.75E-34	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517489	CTV	T30-FS701	1.91E-36	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JQ798289	CTV	A18	1.61E-35	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GQ454870	CTV	HA16-5	2.64E-17	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JX266713	CTV	Taiwan-Pum-M-T5	5.43E-17	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517487	CTV	T36-FS703	4.11E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
EU937521	CTV	T36	7.25E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517486	CTV	T36-FS701	2.31E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517485	CTV	T36-FS674	7.25E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
AY170468	CTV	unknown	1.27E-10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KC517488	CTV	FS577	2.31E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NC001661	CTV	unknown	2.21E-10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
U16304	CTV	unknown	3.81E-10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
AY340974	CTV	Qaha	3.81E-10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
DQ272579	CTV	Mexico	3.81E-10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FJ525435	CTV	NZRB-M17	1.29E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FJ525434	CTV	NZRB-TH30	2.31E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FJ525433	CTV	NZRB-TH28	1.29E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FJ525431	CTV	NZRB-M12	2.17E-12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JF957196	CTV	RB-B301	3.96E-12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
FJ525432	CTV	NZRB-G90	1.18E-12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
JX266712	CTV	Taiwan-Pum-SP-T1	7.25E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
GQ454869	CTV	RB-HA18-9	3.96E-12	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU883266	CTV	VT-Maxi3	3.69E-50	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU883267	CTV	LMS6-6	2.64E-17	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KU883265	CTV	RB-B390-5	4.11E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Noaccession	CTV	RB-B389-4	3.41E-13	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Noaccession	CTV	RB-B389-1	2.31E-11	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
KJ790175	CTV	Mac39	1.63E-53	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Accession	Virus	Isolate	Probe set										
			CTV[a]	CPsV[b]	CTLV	CVV	CYMV	CiLRV	CLBV	CiLV-C	ICRSV	CYVC	SDV
NC006314-16	CPsV	unknown	NA	1.10E-32	NA	NA	NA	NA	NA	NA	NA	NA	NA
JX416228	CTLV	Pk	NA	NA	7.62E-22	NA	NA	NA	NA	NA	NA	NA	NA
KC588948	CTLV	MTH	NA	NA	1.44E-05	NA	NA	NA	NA	NA	NA	NA	NA
KC588947	CTLV	XH	NA	NA	5.98E-17	NA	NA	NA	NA	NA	NA	NA	NA
FJ355920	CTLV	LCd-NA-1	NA	NA	5.98E-17	NA	NA	NA	NA	NA	NA	NA	NA
AY646511	CTLV	Kumquat-1	NA	NA	5.98E-17	NA	NA	NA	NA	NA	NA	NA	NA
JQ765412	CTLV	Shatangorange	NA	NA	1.59E-17	NA	NA	NA	NA	NA	NA	NA	NA
EU553489	CTLV	unknown	NA	NA	5.63E-05	NA	NA	NA	NA	NA	NA	NA	NA
NC001749	CTLV	ASGV	NA	NA	0.0002007	NA	NA	NA	NA	NA	NA	NA	NA
LT160740	CTLV	HPKu-2	NA	NA	0.0002007	NA	NA	NA	NA	NA	NA	NA	NA
KU198289	CTLV	Js2	NA	NA	0.0006539	NA	NA	NA	NA	NA	NA	NA	NA
LC143387	CTLV	FKSS2	NA	NA	2.88E-05	NA	NA	NA	NA	NA	NA	NA	NA
LC084659	CTLV	Matsuco	NA	NA	5.98E-17	NA	NA	NA	NA	NA	NA	NA	NA
HE978837	CTLV	clonep12	NA	NA	0.0002007	NA	NA	NA	NA	NA	NA	NA	NA
KJ579253	CTLV	YTG	NA	NA	3.36E-06	NA	NA	NA	NA	NA	NA	NA	NA
KF434636	CTLV	T47	NA	NA	0.0001075	NA	NA	NA	NA	NA	NA	NA	NA
JQ308181	CTLV	CHN	NA	NA	5.63E-05	NA	NA	NA	NA	NA	NA	NA	NA
KU605672	CTLV	HB	NA	NA	0.0002007	NA	NA	NA	NA	NA	NA	NA	NA
KR106996	CTLV	kfp	NA	NA	0.0011407	NA	NA	NA	NA	NA	NA	NA	NA
JX080201	CTLV	AC	NA	NA	0.0002007	NA	NA	NA	NA	NA	NA	NA	NA
JN701424	CTLV	HH	NA	NA	5.63E-05	NA	NA	NA	0.8914223	NA	NA	NA	NA
AB004063	CTLV	unknown	NA	NA	3.36E-06	NA	NA	NA	NA	NA	NA	NA	NA
AY596172	CTLV	PBNLSV	NA	NA	0.0006539	NA	NA	NA	NA	NA	NA	NA	NA
NC009536-8	CVV	unknown	NA	NA	NA	6.57E-16	NA	0.7194375	NA	NA	NA	NA	NA
NC011807-9	CVV	Asparagus-virus-2	NA	NA	NA	0.0275832	NA	NA	NA	NA	NA	NA	NA
NC003382	CYMV	unknown	NA	NA	NA	NA	3.57E-22	NA	NA	NA	NA	NA	NA
EU708317	CYMV	unknown	NA	NA	NA	NA	2.83E-17	NA	NA	NA	NA	NA	NA
DQ875213	CYMV	India	NA	NA	NA	NA	1.53E-14	NA	NA	NA	NA	NA	NA
FJ617224	CYMV	Nagri	NA	NA	NA	NA	1.94E-18	NA	NA	NA	NA	NA	NA
EU708316	CYMV	SOP	NA	NA	NA	NA	5.06E-13	NA	NA	NA	NA	NA	NA
EU489744	CYMV	AL	NA	NA	NA	NA	5.06E-13	NA	NA	NA	NA	NA	NA

Accession	Virus	Isolate	Probe set										
			CTV[a]	CPsV[b]	CTLV	CVV	CYMV	CiLRV	CLBV	CiLV-C	ICRSV	CYVC	SDV
EU489745	CYMV	PM	NA	NA	NA	NA	**1.61E-13**	NA	NA	NA	NA	NA	NA
JN006805	CYMV	SO-JNTU	NA	NA	NA	NA	**1.21E-19**	NA	NA	NA	NA	NA	NA
JN006806	CYMV	ROL	NA	NA	NA	NA	**4.64E-12**	NA	NA	NA	NA	NA	NA
NC003546-8	CiLRV	unknown	NA	NA	NA	0.8543296	NA	**1.43E-16**	NA	NA	NA	NA	NA
JX237459-256248-9	CiLRV	ATCC-PV-195	NA	NA	NA	NA	NA	**1.43E-16**	NA	NA	NA	NA	NA
NC003877	CLBV	unknown	NA	NA	NA	NA	NA	NA	**9.21E-28**	NA	NA	NA	NA
EU857540	CLBV	NZ-G78	NA	NA	NA	NA	NA	NA	**3.24E-25**	NA	NA	NA	NA
EU857539	CLBV	NZ-G18	NA	NA	NA	NA	NA	NA	**1.34E-24**	NA	NA	NA	NA
JN983456	CLBV	Actinidia-V20	NA	NA	NA	NA	NA	NA	**0.0016917**	NA	NA	NA	NA
JN983455	CLBV	Actinidia-V18	NA	NA	NA	NA	NA	NA	**0.0003865**	NA	NA	NA	NA
JN983454	CLBV	Actinidia-V1	NA	NA	NA	NA	NA	NA	**0.0016917**	NA	NA	NA	NA
JN900477	CLBV	Actinidia	NA	NA	NA	NA	NA	NA	**0.0006434**	NA	NA	NA	NA
FJ009367	CLBV	DMV-932	NA	NA	NA	NA	NA	NA	**1.79E-26**	NA	NA	NA	NA
DQ388512-3	CiLV-C	unknown	NA	NA	NA	NA	NA	NA	NA	**2.95E-39**	NA	NA	NA
DQ157465-6	CiLV-C	unknown	NA	NA	NA	NA	NA	NA	NA	**6.69E-40**	NA	NA	NA
KP336746-7	CiLV-C	SJRP	NA	NA	NA	NA	NA	NA	NA	**4.41E-15**	NA	NA	NA
NC008169-70	CiLV-C	unknown	NA	NA	NA	NA	NA	NA	NA	**1.50E-40**	NA	NA	NA
NC003093	ICRSV	unknown	NA	NA	NA	NA	NA	NA	NA	NA	**6.45E-15**	0.7735932	NA
HQ324250	ICRSV	Pune	NA	NA	NA	NA	NA	NA	NA	NA	**1.22E-13**	0.5650506	NA
NC026592	CYVCV	CQ	NA	NA	NA	NA	NA	NA	NA	NA	0.5684391	**3.91E-25**	NA
KP120977	CYVCV	RL	NA	NA	NA	NA	NA	NA	NA	NA	0.4469742	**1.74E-24**	NA
KP313241	CYVCV	PK	NA	NA	NA	NA	NA	NA	NA	NA	0.2539933	**2.24E-21**	NA
KP313242	CYVCV	YN	NA	NA	NA	NA	NA	NA	NA	NA	0.4469742	**1.74E-24**	NA
KT345342	CYVCV	IS	NA	NA	NA	NA	NA	NA	NA	NA	0.2539933	**1.36E-22**	NA
KT124646	CYVCV	HU	NA	NA	NA	NA	NA	NA	NA	NA	0.4469742	**1.74E-24**	NA
JX040635	CYVCV	Y1	NA	NA	NA	NA	NA	NA	NA	NA	0.1832479	**1.36E-22**	NA
NC003785-6	SDV	unknown	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	**3.19E-40**

[a] Bold entries indicate a positive diagnostic call (*P* value <= 0.05), while nonbold entries indicate no presence of the target virus.

[b] NA indicates no *P* value was generated as no probes hit, indicating a negative diagnostic call.

Bibliography

Abou Kubaa, R., Choueiri, E., El Khoury, M. I., and Djelouah, K. 2015. Identification and molecular characterization of *Citrus variegation virus* in Lebanon. J. Plant Pathol. 97: 391-403.

Achachi, A., Ait Barka, E., and Ibriz, M. 2014. Recent advances in *Citrus psorosis virus*. Virus Dis. 25:261-276.

Achachi, A., Curk, F., Jijakli, M. H., Gaboun, F., El Fahime, E., Soulaymani, A., El Ghilli, M., and Ibriz, M. 2015. Variability and genetic structure of a natural population of *Citrus psorosis virus*. Ann. Microbiol. 65:1195-1199.

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., Samuitiene, M., and Boonham, N. 2009. Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. Mol. Plant Pathol. 10:537-545.

Adams, M. J., Candresse, T., Hammond, J., Kreuze, J. F., Martelli, G. P., Namba, S., Pearson, M. N., Ryu, K. H., Saldarelli, P., and Yoshikawa, N. 2012. Family *Betaflexiviridae*. In: King, A. M. Q., Adams, M. J., Carstens, E. B., and Lefkowitz, E. J (eds). Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier-Academic Press, London. 920–941.

Agranovsky, A. A., Lesemann, D. E., Maiss,E., Hull, R., and Atabekov, J. G. 1995. "Rattlesnake" structure of a filamentous plant RNA virus built for two capsid proteins. *Proc. Natl. Acad. Sci. U.S.A.* 92:2470–2473.

Ahlawat, Y. S., and Pant, R. P. 2003. Major virus and virus-like diseases of citrus in India, their diagnosis and management. Annual Review of Plant Pathology 41:447-474.

Ahlawat, Y. S., Pant, R. P., Lockhart, B. E., Srivastava, M., Chakraborty, N. K., and Varma, A. 1996. Association of *badnavirus* with citrus mosaic disease in India. Plant Dis. 80:590-592.

Ahtawat, Y. S., Pant, R. P., Lockhart, B. E. L., Srivastava, M., Chakraborty, N. K., and Varma, A. 1996. Association of *badnavirus* with citrus mosaic disease in India. Plant Dis. 80: 590-592.

Al Rwahnih, M., Daubert, S., Golino, D., and Rowhani, A. 2009. Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology 387:395-401.

Al Rwahnih, M., Daubert, S., Úrbez-Torres, J. R., Cordero, F., and Rowhani, A. 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch. Virol. 156:397-403.

Al Rwahnih, M., Sudarshana, M. R., Uyemoto, J. K., Rowhani, A. 2012. Complete genome sequence of a novel *Vitivirus* isolated from grapevine. J. Virol. 86:9545.

Albiach-Martí, M. R., Mawassi, M., Gowda, S., Satyanayanana, T., Hilf, M. E., Shanker, S., Almira, E. C., Vives, M. C., López, 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. J. Virol. 74:6856–6865.

Albiach-Martí, 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. Mol. Plant Pathol. 11:55–67.

Alshami, A. A. A., Ahlawat, Y. S., and Pant, R. P. 2003. A hitherto unreported yellow vein clearing disease of citrus in India and its viral aetiology. Indian Phytopathol. 56:422-427.

Anthony Wooton. 1998. Insects of the World. Blandford. ISBN 0-7137-2366-1.

Azeri, T. 1973. First report of *satsuma dwarf virus* disease on Satsuma mandarin in Turkey. Plant Dis. Reptr. 57:149-153.

Backus, E. A., and Bennett, W. H. 2009. The AC–DC Correlation Monitor: New EPG design with flexible input resistors to detect both R and emf components for any piercing–sucking. J. Insect Physiol. 55:869–884.

Baranwal, V. K., Majumder, S., Ahlawat, Y. S., and Singh, R. P. 2003. Sodium sulphite yields improved DNA of higher stability for PCR detection of *Citrus yellow mosaic virus* from citrus leaves. J. Virol. Methods 112:153–156.

Baranwal, V. K., Singh, J., Ahlawat, Y. S., Gopal, K., and Charaya, M. U. 2005. *Citrus yellow mosaic virus* is associated with mosaic disease in Rangpur lime rootstock of citrus. Curr. Sci. 89:1596–1599.

Bar-Joseph, M., and Dawson, W. O. 2008. *Citrus tristeza virus*. In: Mahy, B. W. J., and van Regenmortel, M. H. V (eds). Encyclopedia of Virology: Third Edition Evolutionary Biology of Viruses, Elsevier-Academic Press, Amsterdam. 161–184.

Bar-Joseph, M., Loebenstein, G., and Cohen, J. 1972. Further purification and characterization of threadlike particles associated with the citrus tristeza disease. Virology 50:821–828.

Bar-Joseph, M., Marcus, R., and Lee. R. F. 1989. The continuous challenge of *citrus tristeza virus* control. Annu. Rev. Phytopathol. 27:291–316.

Bastianel, M., Noveli, V. M., Kubo, K. S., Kitajima, E. W., Bassanezi, R., Machado, M. A., and Freitas-Astúa, J. 2010. Citrus leprosis: centennial of an unusual mite-virus pathosystem. Plant Dis. 94:284-292.

Beerenwinkel, N., and Zagordi, O. 2011. Ultra-deep sequencing for the analysis of viral populations. Curr. Opin. Virol. 1:413-418.

Bennani, B., Mendes, C., Zemzami, M., Azeddoug, H., and Nolasco, G. 2002. *Citrus variegation virus*: molecular variability of a portion of the RNA 3 containing the coat protein gene and design of primers for RT-PCR detection. Eur. J. Plant Pathol. 108: 155-162.

Bexfield, N., and Kellam, P. 2011. Metagenomics and the molecular identification of novel viruses. Vet. J. 190:191-198.

Boonham, N., Kreuze, J., Winter, S., van der Vlugt, R., Bergervoet, J., Tomlinson, J., and Mumford, R. 2014. Methods in virus diagnostics: From ELISA to next generation sequencing. Virus Res. 186:20-31.

Brlansky, R. H., Damsteegt, V. D., Howd, D. S., and Roy, A. 2003. Molecular analyses of *citrus tristeza virus* subisolates separated by aphid transmission. Plant Dis. 87:397–401.

Burger, J. T., and Maree, H. J. 2015. Metagenomic next-generation sequencing of viruses infecting grapevines. Methods Mol. Biol. 1302:315-330.

Cambra, M., Gorris, M. T., Marroquín, C., Román, M. P., Olmos, A., Martínez, P. C., Hermoso de Mendoza, A. H., López, A., and Navarro, L. 2000. Incidence and epidemiology of *citrus tristeza virus* in the Valencian Community of Spain. Virus Res. 71:85–95.

Carra, A., Gambino, G., and Schubert, A. 2007. A cetyltrimethylammonium bromide-based method to extract low-molecular-weight RNA from polysaccharide-rich plant tissues. Anal. Biochem. 360:318-320.

Catara, A., Azzaro, A., Davino, M., and Polizzi, G. 1993. Yellow vein clearing of lemon in Pakistan. In: Moreno, P., da Graça, J. V., and Timmer, L. W (eds.). Proceedings of the 12th Conference of the International Organization of Citrus Virologist (ICOV), New Delhi, India. 364-367.

Chen, H. M., Li, Z. A., Wang, X. F., Zhou, Y., Tang, K. Z., Zhou, C. Y., and Zhao, X. Y. 2014. First report of *Citrus yellow vein clearing virus* on lemon in Yunnan, China. Plant Dis. 98:1747.

Chi, P. F., Gu, C. F., and Roistacher, C. N. 1991. Occurrence of *satsuma dwarf virus* in Zhejiang Province, China. Plant Dis. 75:242-244.

Clark, M. F., and Adams, A. N. 1977. Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. J. Gen. Virol. 34:475-483.

Coetzee, B., Freeborough, M. J., Maree, H. J., Celton, J. M., Reese, D. J. G., and Burger, J. T. 2010. Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. J. Virol. 400:157-163.

Colariccio, A., Lovisolo, O., Chagas, C. M., Galletti, S. R., Rossetti, V., and Kitajima, E. W. 1995. Mechanical transmission and ultrastructural aspects of citrus leprosis disease. Fitopatol. Bras. 20:208-213.

Cook, G., van Vuuren, S. P., Breytenbach, J. H. J., Burger, J. T., and Maree, H. J. 2016. Expanded strain-specific RT-PCR assay for differential detection of currently known *Citrus tristeza virus* strains: A useful screening tool. J. Phytopathol. doi:10.1111/jph.12454

Dawson, W. O., Garnsey, S. M., Tatineni, S., Folimonova, S. Y., Harper, S. J., and Gowda, S. 2013. *Citrus tristeza virus*-host interactions. Front. Microbiol. 4:8

Delcher, A. L., Kasif, S., Fleischmann, R. D., Peterson, J., White, O., and Salzberg, S. L. 1999. Alignment of whole genomes. Nucleic Acids Res. 27:2369-2376.

Delcher, A. L., Salzberg, S. L., and Phillippy, A. M. 2003. Using MUMmer to identify similar regions in large sequence sets. Curr. Prot. Bioinform. 00:10.3:10.3.1-10.3.18.

Desjardins, P. R., and Bov´e, J. M. 1980. Infectious variegation and crinkly leaf. In: Bov´e, J. M., and Vogel, R (eds). Description and Illustration of Virus and Virus-Like Diseases of Citrus. A Collection of Colour Slides (IOCV), Paris, I.R.F.A. SETCO-FRUITS.

Dolja, V. V., Karasev, A. V., and Koonin, E. V. 1994. Molecular biology and evolution of closteroviruses: Sophisticated build-up of large RNA genomes. Ann. Rev. Phytopathol. 32:261-285.

Dolja, V.V., Kreuze, J. F., and Valkonen, J. P. T. 2006. Comparative and functional genomics of closteroviruses. Virus Res. 117:38-51.

Engvall, R., and Perlmann, P. 1971. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. Immunochemistry. 8:871-874.

Espach, Y. 2013. The determination of mycoviral sequences in grapevine using next-generation sequencing. MSc thesis, Stellenbosch University.

Feng, J., Zeng, R., Chen, J. 2008. Accurate and efficient data processing for quantitative real-time PCR using a tripartite plant virus as a model. BioTechniques. 44:901-912.

Flores, R., Ruiz-Ruiz, S., Soler, N., Sánchez-Navarro, J., Fagoaga, C., López, C., Navarro, L., Moreno, P., Peña, L. 2013. *Citrus tristeza virus* p23: a unique protein mediating key virus–host interactions. Front. Microbiol. 4:98. doi: 10.3389/fmicb.2013.00098.

Folimonova, S. Y. 2013. Developing an understanding of cross-protection by *Citrus tristeza virus*. Front. Microbiol. 4:76. doi:10.3389/fmicb.2013.00076.

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. J. Virol. 84:1314-1325.

Francki, R. I. B., Fauquet, C. M., Knudson, D. L., and Brown, F. 1991. Classification and

Galipienso, L., Navarro, L., Ballester-Olmos, J. F., Pina, J. A., Moreno, P., and Guerri, J. 2000. Host range and symptomatology of a graft-transmissible pathogen causing bud union crease of citrus on trifoliate rootstocks. Plant. Pathol. 49:308-314.

Galipienso, L., Vives, M. C., Moreno, P., Milne, R. G., Navarro, L., and Guerri, J. 2001. Partial characterization of *Citrus leaf blotch virus*, a new virus from Nagami kumquat. Arch. Virol. 146:357-368.

Galipienso, L., Vives, M. C., Navarro, L., Moreno, P., and Guerri, J, 2004. Detection of *citrus leaf blotch virus* using digoxigenin-labeled cDNA probes and RT-PCR. Eur. J. Plant Pathol. 110:175–181.

Garnsey, S. M. 1975. Purification and Properties of *Citrus-Leaf-Rugose Virus*. Phytopathology 65:50-57.

Garnsey, S. M., Civerolo, E. L., Gumpf, D. J., Paul, C., Hilf, M. E., Lee, R. F., Brlanskyr, R. H., Yokomi, R. K., Hartung, J. S. 2004. Biocharacterisation of an international collection of *citrus tristeza virus* (CTV) isolates. In: Hilf, M. E. Duran-Vila, N., and Rocha-Peña, M. A (eds). Proceedings of the 16th Conference of the International Organization of Citrus Virology (IOCV), Riverside, California. 75-93.

Garnsey, S. M., Gottwald, T. R., Hilf, M. E., Matos, L., and Borbón, J. 2000. Emergence and spread of severe strains of *Citrus tristeza virus* isolates in the Dominican Republic. In: da Graça, J. V., Lee, R. F., and Yokomi, R. K (eds). Proceedings of the 14th Conference of the International Organization of Citrus Virologists (IOCV), Riverside, California. 57–68.

Ghosh, D. K., Bhose, S., Mukherjee, K., Aglave, B., Warghane, A. J., Motghare, M., Baranwal, V. K., and Dhar, A. K. 2014. Molecular characterization of *Citrus yellow mosaic badnavirus* (CMBV) isolates revealed the presence of two distinct strains infecting citrus in India. Phytoparasitica 42:681–689.

Gillings, M., Broadbent, P., Indsto, J., and Lee, R. 1993. Characterization of isolates and strains of *citrus tristeza closterovirus* using restriction analysis of the coat protein gene amplified by the polymerase chain reaction. J. Virol. Methods 44:305–317.

Gonsalves, D., and Fulton, R. W. 1977. Activation of *prunus necrotic ringspot* and *rose mosaic virus* by RNA 4 components of some ilarviruses. Virology 81:398–407.

Gonsalves, D., Garnsey, S. M. 1989. Cross protection techniques for control of plant virus diseases in the tropics. Plant Dis. 73:592-597.

Gosh, D. K., Boshe, S., Mukherjee, K., Aglave, B., Warghane, A. J., Motghare, M., Baranwal, V. K., and Dhar, A. K. 2014. Molecular characterization of *Citrus yellow mosaic badnavirus* (CMBV) isolates revealed the presence of two distinct strains infecting citrus in India. Phytoparasitica 42:681-689.

Gowda, S., Satyanayanana, T., Ayllón, M. A., Albiach-Martí, M. R., Mawassi, M., Rabindran, S., and Dawson, W. O. 2001. Characterization of the cis-acting elements controlling subgenomic mRNAs of *Citrus tristeza virus*: production of positive-and negative-stranded 3′-terminal and positive-stranded 5′ terminal RNAs. Virology 286:134–151.

Grant, T. J., and Costa, A. S. 1951. A mild strain of the tristeza virus of citrus. Phytopathology 41:114-122.

Guerri, J., Pina, J. A., Vives, M. C., Navarro, L., and Moreno, P. 2004. Seed transmission of *Citrus leaf botch virus*: implications in quarantine and certification programs. Plant. Dis. 88:906.

Hailstones, D. L., Bryant, K. L., Broadbent, P., and Zhou, C. 2000. Detection of *Citrus tatter leaf virus* with reverse transcription-polymerase chain reaction (RT-PCR). Australas. Plant Pathol. 29:240-248.

Hajeri, S., Ramadugu, C., Keremane, M., Vidalakis, G., and Lee, R. 2010. Nucleotide sequence and genome organization of *Dweet mottle virus* and its relationship to members of the family *Betaflexiviridae*. Arch. Virol. 155:1523–1527.

Hančević, K., Černi, S., Radić, T., and Škorić, D. 2012. Comparison of different methods for *Citrus tristeza virus* detection in Satsuma mandarins. J. Plant Diseases and Protection, 119:2-7.

Harper, S. J. 2013. *Citrus tristeza virus*: evolution of complex and varied genotypic groups. Front. Microbiol. 4:93. doi: 10.3389/fmicb.2013.00093.

Hernández-Rodríguez, L., Pérez-Castro, J. M., García-García, G., Ramos-González, P. L., Zamora-Rodríguez, V., Ferriol-Marchena, X., Peña-Bárzaga, I., and Riverend, L. B. 2016. *Citrus leaf blotch virus* in Cuba: first report and partial molecular characterization. Trop. Plant Pathol. 41:147-154.

Hilf, M. E., Karasev, A. V., Pappu, H. R., Gumpf, D. J., Niblett, C. L., and Garnsey, S. M. 1995. Characterization of *citrus tristeza virus* subgenomic RNAs in infected tissue. Virology 208:576–582.

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 95:909–917.

73

Huang, Q., and Hartung, J. S. 2001. Cloning and sequence analysis of an infectious clone of *Citrus yellow mosaic virus* that can infect sweet orange via Agrobacterium-mediated inoculation. J. Gen. Virol. 82: 2549-2558.

Huang, W., Li, L., Myers, J. R., and Marth, G. T. 2012. ART: a next-generation sequencing read simulator. Bioinformatics 28:593-594.

Hugenholtz, P., and Tyson, G. W. 2008. Microbiology: Metagenomics. Nature. 455:481-483. Iwanami, T. 2010. Properties and Control of *Satsuma Dwarf Viru*s. JARQ. 44:1-6. doi.org/10.6090/jarq.44.1.

Iwanami, T., Kondo, Y., Kobayashi, M., Han, S. S., and Karasev, A. V. 2001. Sequence diversity and interrelationships among isolates of satsuma dwarf-related viruses. Arch. Virol. 146:807-813.

J. Turk. Phytopathol. 32:53-55.

Johnson, A. M., Borah, B. K., Sai Gopal, D. V., and Dasgupta, I. 2012. Analysis of full-length sequences of two *Citrus yellow mosaic badnavirus* isolates infecting *Citrus jambhiri* (Rough Lemon) and *Citrus sinensis* L. Osbeck (Sweet Orange) from a nursery in India. Virus Genes 45:600–605.

Karasev, A. V. 2000.Genetic diversity and evolution of closteroviruses. Annu. Rev. Phytopathol. 38:293–324.

Karasev, A. V., Boyko, V. P., Gowda, S., Nikolaeva, O. V., Hilf, M. E., Koonin, E. V., Niblett, C. L. Cline, K., Gumpf, D. J., Lee, R. F., Garnsey, S. M., Lewandowski, D. J., and Dawson, W. O. 1995. Complete sequence of the *citrus tristeza virus* RNA genome. Virology 208:511–520.

Karasev, A. V., Han, S. S., and Iwanami, T. 2001. Satsuma dwarf and related viruses belong to a new lineage of plant picorna-like viruses. Virus Genes 23:45-52.

Koenig, R., and Paul, H. L. 1982. Variants of ELISA in plant virus diagnosis. J. Virol. Methods 5:113-125.

Kondo, H., Maeda, T., Shirako, Y., and Tamada, T. 2006. Orchid fleck virus is a rhabdovirus with an unusual bipartite genome. J. Gen. Virol. 87:2413-2421.

Kong, P., Rubio, L., Polek, M., and Falk, B. W. 2000. Population structure and genetic diversity within California *Citrus tristeza virus* (CTV) field isolates. Virus Genes 21:139–145

Koonin, E. V., and Dolja, V. V. 2012. Expanding networks of RNA virus evolution. BMC Biol. 10:54.

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., and Simon, R. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. Virology. 388:1-7.

Kurtz, S., Phillippy, A., Delcher, A. L., Smooth, M., Shumway, M., Antonescu, C., and Salzberg, S. L. 2004. Versatile and open software for comparing large genomes. Genome Bio. 5:R12.

Kusano, N., Hirashima, K., Kuwahara, M., Narahara, K., Imamura, T., Mimori, T., Nakahira, K., andTorii, K. 2007. Immunochromatographic assay for simple and rapid detection of *Satsuma dwarf virus* and related viruses using monoclonal antibodies. J. Gen. Plant Pathol. 73:66-71.

Laura Cristina Garita, L. C, Tassi, A. D., Calegario, R. F., Freitas-Astúa, J., Salaroli, R. B., Romão, G. O., and Kitajima, E. W. 2014. Experimental host range of *Citrus leprosis virus C* (CiLV-C). Trop. Plant Pathol. 39:043-055.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. 2012. Comparison of next- generation sequencing systems. J. Biomed. Biotechnol. doi:10.1155/2012/251364.

Locali, E. C., Freitas-Astúa, J., Souza, A. A., Takita, M. A. J., Antonioli, R., Kitajima, E. W, and Machado, M. A. 2003. Development of molecular tool for the diagnosis of leprosis a major threat to citrus production in the Americas. Plant Dis. 87:1317-1321.

Locali-Fabris, E. C., Freitas-Astúa, J., and Machado, M. A. 2012. Genus *Cilevirus.* In: King, A. M. Q., Adams, M. J., Carstens, E. B., Lefkowitz, E. J (eds.). Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier-Academic Press, San Diego, California. 1139-1142.

Locali-Fabris, E., Freitas-Astúa, J., Souza, A. A., Takita, M. A., Astúa-Monge, G., Antonioli-Luizon, R., Rodrigues, V., Targon, M. L. P. N., and Machado, M. 2006. Complete nucleotide sequence, genomic organization and phylogenetic analysis of *Citrus leprosis virus* cytoplasmic type. J. Gen. Virol. 87:2721-2729.

Loconsole, G., Fatone, M. T., and Savino, V. 2009. Specific digoxigenin-labelled riboprobes for detection of *Citrus psorosis virus* and *Citrus variegation virus* by molecular hybridization. J. Plant Pathol. 91: 311-319.

Loconsole, G., Onelge, N., Potere, O., Giampetruzzi, A., Bozan, O., Satar, S., De Stradis, A., Savino, V., Yokomi, R. K., and Saponari, M. 2012. Identification and characterization of *citrus yellow vein clearing virus*, a putative new member of the genus *Mandarivirus*. Phytopathology 102:1168-1175.

Loconsole, G., Önelge, N., Potere, O., Giampetruzzi, A., Bozan, O., Satar, S., De Stradis, A., Savino, V., Yokomi, R. K., and Saponari, M. 2012. Identification and characterization of *Citrus yellow vein clearing virus*, a putative new member of the genus *Mandarivirus*. Phytopathology 102:1168-1175.

Lovisolo, O. 1993. Agro-ecology and centres of origin of grafttransmissible diseases of citrus. Proceedings of the International Organisation of Citrus Virologists 12:406–441.

MacDiarmid, R., Rodoni, B., Melcher, U., Ochoa-Corona, M., and Roossinck, M. 2013. Biosecurity implications of new technology and discovery in plant research. PLoS Pathog. 9:e1003337.

Mardis, E. R. 2008. Next-generation DNA sequencing methods. Annu. Rev. Genomics Hum. Genet. 9:387-402.

Marroquín, C., Olmos, A., Gorris, M. T., Bertolini, E., Martínez, M. C., Carbonell, E. A., Hermoso de Mendoza, A. H., and Cambra, M. 2004. Estimation of the number of aphids carrying *Citrus tristeza virus* that visit adult citrus trees. Virus Res. 100:101–108.

Martelli, G. P., Agranovsky, A. A., Al Rwahnih, M., Dolja, V. V., Dovas, C. I., Fuchs, M., 2012. Taxonomic revision of the family *Closteroviridae* with special reference to the grapevine leafroll-associated members of the genus *Ampelovirus* and the putative species unassigned to the family. J. Plant Pathol. 94:7-19.

Martelli, G. P., Agranovsky, A. A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R. H. A., Dolja, V. V., Falk, B. W., Gonsalves, D., Jelkmann, W., Karasev, A. V., Minafra, A., Namba, S., Vetten, H. J., Wisler, G. C., Yoshikawa, N. 2002. The family *Closteroviridae* revised. Arch. Virol. 147:2039-2044.

Martín, S., López, C., García, M. L., Naum-Ongania, G., Grau, O., Flores, R., Moreno, P., and Guerri, J. 2005. The complete nucleotide sequence of a Spanish isolate of *Citrus psorosis virus*: comparative analysis with other ophioviruses. Arch. Virol. 150:167-176.

Massart, S., Olmos, A., Jijakli, H., and Candresse, T. 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Res. 118:90-96.

Mawassi, M., Mietkiewska, E., Gofman, R., Yang, G., and Bar-Joseph, M. 1996. Unusual sequence relationships between two isolates of *citrus tristeza virus*. J. Gen.Virol. 77:2359–2364.

McClean, A. P. D. 1963. The tristeza complex: Its variability in field-grown citrus in South Africa. S. Afr. J. Agric. Sci. 6:303–332.

McClean, A. P. D. 1974. The tristeza virus complex. In: Weathers, L. G., and Cohen, M (eds). Proceedings of the 6th Conference of the International Organisation of Citrus Virology (IOCV), Division of Agricultural Sciences, University of California, Richmond. 59-66.

McKinney, H. H. 1926.Virus mixtures that may not be detected in young tobacco plants. Phytopathology 16:893.

Melzer, M. J., Borth, W. B., Sether, D. M., Ferreira, S., Gonsalves, D., and Hu, J. S. 2010. Genetic diversity and evidence for recent modular recombination in Hawaiian *Citrus tristeza virus*. Virus Genes 40:111-118.

Mende, D. R., Waller, A. S., Sunagawa, S., Järvelin, A. I., Chan, M. M., Arumugam, M., Raes, J., and Bork, P. 2012. Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS One. 7:e31386.

Milne, R. G., Djelouah, K., García, M. L., Dal Bo, E., Grau, O. 1996. Structure of citrus ringspot-psorosis-associated virus particles: implications for diagnosis and taxonomy. In: da Graça, J. V., Moreno, P., and Yokomi, R. K (eds). Proceedings of the 13th Conference of the International Organization of Citrus Virologists (IOCV), Riverside, California. 189-197.

Mokili, J. L., Rohwer, F., and Dutilh, B. E. 2012. Metagenomics and future perspectives in virus discovery. Curr. Opin. Virol. 2:63-77.

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. Mol. Plant Pathol. 9:251-268.

Moreno, P., Guerri, J., and García, M. L. 2015. The psorosis disease of citrus: a pale light at the end of the tunnel. J. Cit. Pathol. iocv_journalcitruspathology_28860.

Moreno, P., Guerri, J., and Muñoz, N. 1990. Identification of Spanish strains of *citrus tristeza virus* (CTV) by analysis of double-stranded RNAs (dsRNA). Phytopathology 80:477–482.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Sym. 51:263-273.

Mumford, R., Boonham, N., Tomlinson, J., and Barker, I. 2006. Advances in molecular phytodiagnostics—new solutions for old problems. Eur. J. Plant Pathol. 116:1:19.

Naidua, R. A., and Hughes, J. d'A. 2001. Methods for the detection of plant virus diseases. In: Plant virology in sub-Saharan Africa. Proceedings of a conference organized by international institute of tropical agriculture, Ibadan, Nigeria. 233-260.

Naum-Onganía, G., Gago-Zachert, S., Peña, E., Grau, O., and García, M. L. 2003. *Citrus psorosis virus* RNA 1 is of negative polarity and potentially encodes in its complementary strand a 24K protein of unknown function and 280K putative RNA dependent RNA polymerase. Virus Res. 96:49-61.

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 228:92–97.

O'Donnell, K. 1999. Plant pathogen diagnostics: present status and future developments. Potato Res. 42:437-447.

Ochoa-Corona, F. M. 2011. Biosecurity, microbial forensics and plant pathology: education challenges, overlapping disciplines and research needs. Australas. Plant Path. 40:335-338.

Önelge, N., Satar, S., Elibüyük, O., Bozan, O., Kamberoğlu, M. 2011. Transmission studies on *Citrus yellow vein clearing virus*. Proceedings of the 18th Conference of the International Organization of Citrus Virologists (IOCV), Campinas, Brazil. 26-28.

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., Dolja, V. V., Koonin, E. V., Gumpf, D. J., Cline, K., 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 *citrus tristeza closterovirus* genome. Virology 199:35–46.

Pappu, H. R., Manjunath, K. L., Lee, R. F., and Niblett, C. L. 1993. Molecular characterization of a structural epitope that is largely conserved among severe isolates of a plant virus. Proc. Natl. Acad. Sci. U.S.A. 90:3641–3644.

Pascon, R. C., Kitajima, J. P., Breton, M. C., Assumpção, L., Greggio, C., Zanca, A. S., Okura, V. K., Alegria, M. C., Camargo, M. E., Silva, G. G. C., Cardozo, J. C., Vallmn, M. A., Franco, S. F., Silva, V. H., Jordão, Jr. H., Oliveira, F., Giachetto, P. F., Ferrari, F., Aguilar-Vildoso, CI., Franchiscini, F. J. B., Silva, J. M. F., Arruda, P., Ferro, J. A., Reinach, F., and Silva, A. C. R. 2006. The complete nucleotide sequence and genomic organization of *Citrus leprosis associated virus*, cytoplasmic type (CiLV-C). Virus Genes 32:289-298.

Permar, T. A., Garnsey, S. M., Gumpf, D. J., and Lee, R. F. 1990. A monoclonal antibody which discriminates strains of *citrus tristeza virus*. Phytopathology 80:224–228.

Raccah, B., Loebenstein, G., and Bar-Joseph, M. 1976. Transmission of *citrus tristeza virus* by melon aphid. Phytopathology 66:1102–1104.

Rodini, B. 2009. The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. Virus Res. 141:150-157.

Rodrigues, J. C. V., Kitajima, E. W., Childers, C. C., and Chagas, C. M. 2003. *Citrus leprosis virus* vectored by *Brevipalpus phoenicis* (Acari: Tenuipalpidae) on citrus in Brazil. Exp. Appl. Acarol. 30:161-179.

Roistacher, C. N. 1991. Graft-transmissible Diseases of Citrus: Handbook for Detection and Diagnosis. FAO, Rome, Italy.

Roistacher, C. N. 1993. Psorosis - A Review. In: Moreno, P., da Graça, J. V., and Timmer, L. W (eds). Proceedings of the 12th International Organisation of Citrus Virologists (IOCV) Conference, Riverside, California. 139–54.

Roistacher, C. N., and Moreno, P. 1990. The worldwide threat from destructive isolates of *Citrus tristeza virus* – A review. In: Brlansky, R. H., Lee, R. F., and Timmer, L. W (eds). Proceedings of the 11th Conference of the International Organization of Citrus Virologists (IOCV), Riverside, California. 7-19.

Roossinck, M. J., Bujarski, J., Ding, S. W., Hajimorad, R., Hanada, K., Scott, S., and Tousignant, M. 2005. Family *Bromoviridae*. In: Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., and Ball, L. A (eds). Virus Taxonomy: Eight Report of the International Committee on Taxonomy of Viruses. Elsevier-Academic Press, San Diego, California. 1049-1058

Roy, A., Stone, A., Otero-Colina, G., Wei, G., Choudhary, N., Achor, D., Shao, J., Levy, L., Nakhla, M. K., Hollingsworth, C. R., Hartung, J. S., Schneider, W. L., and Brlansky, R. H. 2013. Genome Assembly of *Citrus Leprosis Virus* Nuclear Type Reveals a Close Association with *Orchid Fleck Virus.* Genome Announc. 4:e00519–13. doi.org/10.1128/genomeA.00519-13.

Rubio, L., Ayllón, M. A., Guerri, J., Pappu, H., Niblett, C., and Moreno, P. 1996. Differentiation of *citrus tristeza closterovirus* (CTV) isolates by single-strand conformation polymorphism analysis of the coat protein gene. Ann. Appl. Biol. 129:479–489.

Rustici, G., Accotto, G. P., Noris, E., Masenga, V., Luisoni, E., and Milne, R. G. 2000. *Indian citrus ringspot virus*: A proposed new species with some affinities to potex-, carla-, fovea- and allexiviruses. Arch. Virol. 145:1895-1908.

Sambade, A., Rubio, L., Garnsey, S. M., Costa, N., Müller, G. W., Peyrou, M., Guerri, J., and Moreno, P. 2002. Comparison of the viral RNA populations of pathogenically distinct isolates of *Citrus tristeza virus*: application to monitoring cross protection. Plant Pathol. 51:257–265.

Sankaran, S., Mishra, A., Ehsani, R., and Davis, C. 2010. A review of advanced techniques for detecting plant disease. Comput. Electron. Agr. 72:1-13.

Satya, R. V., Zavaljevski, N., Kumar, K., and Reifman, J. 2008. A high-throughput pipeline for designing microarray-based pathogen diagnostic assays. BMC Bioiform. 9:185.

Satyanayanana, T., Gowda, S., Boyko, V. P., Albiach-Martí, M. R., Mawassi, M., Navas-Castillo, J., Karasev, A. V., Dolja, V., Hilf, M. E., Lewandowsky, D. J., Moreno, P., Bar-Joseph, M., Garnsey, S .M., and Dawson, W. O. 1999. An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. Proc. Natl. Acad. Sci. U.S.A. 96:7433–7438.

Satyanayanana, T., Gowda, S., Mawassi, M., Albiach-Martí, 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 278:253–265.

Schaad, N. W., and Frederick, R. D. 2002. Real-time PCR and its application for rapid plant disease diagnostics. Can. J. Plant Pathol. 24:250-258.

Schaad, N. W., Frederick, R. D., Shaw, J., Schneider, W. L., Hickson, R., Petrillo, M. D., and Luster, D. G. 2003. Advances in molecular-based diagnostics in meeting crop biosecurity and phytosanitary issues. Annu. Rev. Phytopathol. 41:305-324.

Scholz, C. F. P., Poulsen, K., and Kilian, M. 2012. Novel molecular method for identification of streptococcus pneumonia applicable to clinical microbiology and 16s RNA sequence-based microbiome studies. J. Clin. Microbiol. 50:968-1973.

Scott, S. W., and Ge, X. 1995. The complete nucleotide sequence of RNA3 of *citrus leaf rugose* and *citrus variegation* Ilarviruses. J. Gen. Virol. 76:957–963.

Sharma, S., Singh, B., Rani, G., Zaidi, A. A., Hallan, V. K., Nagpal, A. K, and Virk, G. S. 2007. In vitro production of Indian *citrus ringspot virus*-free plants of kinnow mandarin (*Citrus nobilis Lour* x *C. deliciosa Tenora*) employing chemotherapy coupled with Shoot Tip grafting. J. Cent. Eur. Agric. 8:1-8

Stobbe, A. H., Daniels, J., Espindola, A. S., Verma, R., Melcher, U., Ochoa-Corona, F., Garzon, C., Fletcher, J., and Schneider, W. 2013. E-probe diagnostic nucleic acid analysis (EDNA): A theoretical approach for handling next generation sequence data for diagnostics. J. Microbiol. Methods 94:356-366.

Stobbe, A. H., Schneider, W. L., Hoyt, P. R., and Melcher, U. 2014. Screening metagenomic data for viruses using the e-probe diagnostic nucleic acid assay. Phytopathology 104:1125-1129.

Suastika, G., Natsuaki, T., Terui, H., Kano, T., Ieki, H., and Okuda, S. 2001. Nucleotide sequence of *Citrus tristeza virus* seeding yellows isolate. J. Gen. Plant Pathol. 67:73–77.

Tatineni, S., and Dawson, W. O. 2012. Enhancement or attenuation of disease by deletion of genes from *Citrus tristeza virus*. J. Virol. 86:7850-7857.

Tatineni, S., Robertson, C. J., Garnsey, S. M., Bar-Joseph, M., Gowda, S., and Dawson, W. O. 2008. Three genes of *Citrus tristeza viru*s are dispensable for infection and movement throughout some varieties of citrus trees. Virology 376:297-307.

Tian, T., Rubio, L., Yeh, H. H., Crawford, B., and Falk, B. W. 1999. *Lettuce infectious yellows virus*: In vitro acquisition analysis using partially purified virions and the whitefly, *Bemisia tabaci*. J. Gen. Virol. 80:1111–1117.

Van Vuuren, S. P., Collins, R. P., and da Graça, J. V. 1993. Evaluation of *citrus tristeza virus* isolates for cross protection of grapefruit in South Africa. Plant Dis. 77:24–28.

Vázquez-Castellanos, J. F., García-López, R., Pérez-Brocal, V., Pignatelli, M., and Moya, A. 2014. Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. BMC Genomics 15:37.

Velázquez, K., Renovell, A., Comellas, M., Serra, P., García, M. L., Pina, J. A., Navarro, L., Moreno, P., and Guerri, J. 2010. Effect of temperature on RNA silencing of a negative-stranded RNA plant virus: *Citrus psorosis virus*. Plant Pathol. 59:982-990.

Visser, M., Burger, J. T., and Maree, H. J. 2016. Targeted virus detection in next-generation sequencing data using an automated e-probe based approach. Virology 495:122-128.

Vives, M. C., Galipienso, L., Navarro, L., Moreno, P., and Guerri, J. 2002. Characterization of two kinds of subgenomic RNAs produced by *citrus leaf blotch virus*. Virology 295:328–336.

Vives, M. C., Martín, S., Ambrós, S., Renovell, A., Navarro, L., Pina, J. A., Moreno, P., and Guerri, J. 2008. Development of a full-genome cDNA clone of *Citrus leaf blotch virus* and infection of citrus plants. Mol. Plant Pathol. 9:787–797.

Vives, M. C., Rubio, L., López, C., Navas-Castillo, J., Albiach-Martí, 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. J. Gen. Virol. 80:811–816.

Ward, E., Foster, S. J., Fraaije, B. A., and McCartney, H. A. 2004. Plant pathogen diagnostics: immunological and nucleic acid-based approaches. Ann. Appl. Biol. 145:1-16.

Wooley, J. C., Godzik, A., and Friedberg, I. 2010. A primer on metagenomics. PLoS Comput. Biol. 6:e1000667

Yang, X., Charlebois, P., Gnerre, S., Coole, M. G., Lennon, N. J., Levin, J. Z., Qu, J., Ryan, E. M., Zody, M. C., and Henn, M. R. 2012. *De novo* assembly of highly diverse viral populations. BMC Genomics 13:475.

Yang, Z. -N., Mathews, D. M., Dodds, J. A., and Mirkov, T. E. 1999. Molecular characterization of an isolate of *citrus tristeza virus* that causes severe symptoms in sweet orange. Virus Genes 19:11–142.

Yokomi, R. K. 2009. *Citrus tristeza virus.* In D'Onghia, A. M., Djelouah, K., and Roistacher, C. N (eds). *Citrus tristeza virus* and *Toxoptera citricidus*: a serious threat to the Mediterranean citrus industry, Series B Studies and Research, N° 65, Options Méditerranéennes, CIHEAM. 19-33.

Zanek, M. C., Pen˜a, E., Reyes, C. A., Figueroa, J., Stein, B., Grau, O., and Garcia, M. L. 2006. Detection of *Citrus psorosis virus* in the north western citrus production area of Argentina by using an improved TASELISA. J. Virol. Methods. 137:245–51.

Zerbino, D. R., and Birney, E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zhang, J., Chiodini, R., Badr, A., and Zhang, G. 2011. The impact of next-generation sequencing on genomics. J Genet Genomics 38:95-109.

Zhou, C. Y., Wang, X. F., Zhou, Y., and Liu, J. X. 2013. Occurrence and control citrus viral diseases in China. China Fruit News 30:70-72.

Zhou, Y., Chen, H. M., Wang, X. F., Li, Z. A., Tang, M., and Zhou, C. Y. 2015. Lack of evidence for seed transmission of *Citrus yellow vein clearing virus* despite its frequent detection in seed tissues. J. Plant Pathol. 97:1-3.

Internet sources

http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

http://www.crec.ifas.ufl.edu/academics/classes/PLP5115C/PDF/citrus_virus.pdf

http://www.expasy.org/viralzone/

https://secure.clcbio.com/helpspot)

https://www.qiagenbioinformatics.com/