

Strain differentiation of *Citrus tristeza virus* isolates from South Africa by PCR and microarray.

by

KATHERINE ANNE STEWART

(21090999)

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae Microbiology (MSc.)

in the Faculty of Natural and Agricultural Science

University of Pretoria

Pretoria, South Africa

November 2006

Supervisor: Prof. G. Pietersen

I hereby declare that this thesis, except where indicated by a reference, is my own research and has not been submitted, in part or as a whole, for a degree at any other university.

Katherine Anne Stewart

Student Number: 21090999

Signature: _____

Date: _____



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to the following:

The University of Pretoria for the opportunity to complete my M.Sc. (Microbiology) degree and financial support;

Professor Gerhard Pietersen, my supervisor; for his guidance, assistance, friendship and academic input throughout my project, it is much appreciated;

Professor L.H. Nel, Professor L. Korsten, Professor D. Berger and Sanushka Naidoo for their academic inputs and advice;

Citrus Research International (CRI), for financial support with a prestigious bursary, project funding and the opportunity to do research through such a great organisation;

National Research Fund (NRF) for financial support with a Scare Skills bursary;

My colleagues in the Virology lab at the University of Pretoria for their friendship, encouragement and suggestions;

My father, grandparents and the rest of my family for their continued interest, love and encouragement throughout my studies;

Sean, my fiancé for his love, support and understanding. You have given me the strength to pursue my dreams;

The Lord, for His many blessings.

**In loving memory of my late Mother
Susan Virginia Stewart (1949-1986)**

The aim of this study was to characterize strains used in the cross-protection scheme in South Africa by establishing Polymerase Chain Reaction (PCR) systems aimed at differentiating the strains by targeting the conserved p23 gene and the variable 5' half of the *Citrus tristeza virus* (CTV) genome. Two cross-protecting sources GFMS 12 and GFMS 35; and eight single aphid sub-isolates were tested and classified into strain types or genotypes. An oligonucleotide microarray system was developed to differentiate T30 and T36 strains of CTV. The establishment and development of such tests will enable the South African Citrus Industry to better select mild strains for cross-protection and determine which strains are present in citrus growing areas so as to better understand the dynamics of the disease.

The first aim was to characterise the p23 gene of possible mild-strain cross-protection isolates in South Africa (RSA) and compare them to known isolates worldwide. Isolates were amplified with bi-directional RT-PCR, sequenced and phylogenetic analysis performed. The predicted amino acid sequences were compared for areas of possible variability for further strain differentiation. A bi-directional PCR developed by Sambade *et al.* (2003) was established that targets differences in amino acid positions 78-80 of the p23 gene and allows discrimination of isolates into mild, atypical and severe groups. The group designations of RSA isolates 390-3 and 390-5 were atypical; 390-4, 389-4 and 389-3 were mild; GFMS 35 had mild and atypical isolates; GFMS12, 12-7 and 12-9 had mild and severe isolates and; 12-5 was severe. The three main clusters on the phylogenetic tree confirmed the group designations of these isolates. Isolates in the atypical group were more diverse than ones in the mild or severe groups. There were 53 polymorphic sites within the amino acid sequences of p23 gene of the RSA and reference isolates, of which 4 distinct regions showed variability. The amino acid region 78-80 was confirmed as being very useful in grouping these isolates as mild, severe or atypical. The PCR system was robust, reproducible and has potential in the RSA Citrus industry as a screening tool in selecting mild strains for cross-protection and in detecting mixed strains in isolates.

The secondary aim was to establish the 23 primer pair PCR system developed by Hilf *et al.* (2000) to differentiate isolates as T36, T30, VT or T3 genotypes. Each isolate was tested with RT-PCR using 23 individually optimised genotype specific primer sets (Hilf *et al.*, 2000). The most common genotype detected was T30 and the least common was T3. The

GFMS 35, T30 plant and 389-3 isolates had a homogenous T30 genotype profile and isolate 12-5 had a VT genotype profile. The 389-4, 390-3 and 390-4 isolates had a predominantly T30 genotype profile and isolates 12-7 had a predominantly VT genotype profile. Isolate GFMS 12 had a mixed genotype profile indicative of a mixed infection while isolates 390-5 and 12-9 appeared to have mixed genotypes of VT, T30 and T36. Isolates 390-3 390-4 and 390-5 had no amplification within regions 4-7 and appear to be highly variable isolates or possible recombinants. The T3 genotype specific markers were found in region 2 of a few isolates and could be a cross-reacting primer set to the T30 genotype. It is useful for homogenous strains in determining the genotypes, molecular marker information, possible variability or recombination and for approving isolates for mild strain cross-protection. Potential drawbacks of the system include non-amplification of regions; cross-reacting primers; difficulty in optimising; and secondary structures. It was difficult to objectively draw conclusions if an isolated had mixed genotypes since mixed genotype amplifications were not consistently found in all regions targeted.

The third aim was to develop an oligonucleotide (oligo) microarray system to differentiate mild T30 and severe T36 strains. The 5' half of the CTV genome was Cy3 5'-end labelled and amplified. Oligos were designed to be T36-strain specific with a T_m above 60 °C and if possible a GC content above 65 %; and differed in amount and position of mismatches to strain T30. A standard operating protocol was set up by testing different labelling methods, hybridization mixes and washing steps. The array was tested using individual T30 and T36 strains as templates at 42, 52 and 60 °C. Experimental variation was quantified and normalised. The secondary structures of the hybridizing amplicons were determined by mfold (Zuker *et al.*, 2003). Some oligos were specific at 42 °C and others at 52 °C. The hybridization allowed a clear differentiation of strain T36 with 13 of the T36-specific oligos at their optimal hybridization temperature. A few oligonucleotides showed cross-hybridization to strain T30 and were not used in further analysis. Oligonucleotides with 21 % or more mismatches were successful oligos, whereas ones that had 18 % or fewer mismatches had cross-hybridization. Some oligos were modified to include Locked Nucleic Acid (LNA) instead of DNA in an attempt to increase specificity with two of them having increased specificity compared to the unmodified DNA oligonucleotides. The successful differentiation by hybridization to strain specific oligos opens paths for highly parallel, yet specific assays for strain differentiation of CTV strains and a more thorough insight into the future strains circulating in RSA.



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

LIST OF ABBREVIATIONS

A	adenine
AMV	Avian Myeloblastosis Virus
ATPase	Adenosine triphosphate enzymes
bp	base pair
BCA	Brown citrus aphid (<i>Toxoptera citricida</i>)
C	cytosine
cDNA	complementary Deoxyribose nucleic acid
CMV	Cucumber Mosaic virus
CP	Major Coat Protein
CPm	Minor Coat Protein
CTV	<i>Citrus tristeza virus</i>
Cy3	cyanine derived fluorescent dye
dG	free energy change
DHR	differential host reaction
DIG	Digoxigenin
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotriphosphate
D-RNA	Defective Ribonucleic Acid
dsRNA	Double stranded Ribose Nucleic Acid
DTBIA	direct tissue blot immunoassay
dUTP	deoxy uracil triphosphate
ELISA	enzyme-linked immunosorbent assay
F	Phenylalanine
G	guanidine
GAPS	Gamma Amino Propyl Silane
GC	Guanidine-Cytosine
GFMS	Grapefruit Mild Strains
gRNA	genomic Ribose Nucleic Acid
HEL	Helicase
HSP	Heat Shock Protein
ISIA	in situ immunoassay

ISIF	in situ immunofluorescence
kcal	kilocalorie
kDa	Kilodalton
LMS	Lime Mild Strains
M	Molar
MCA	Monoclonal Antibody
mg	milligram
MgCl₂	Magnesium Chloride
ml	millilitre
mM	millimolar
M-MLV	Moloney Murine Leukaemia Virus
mol	mole
MT	Methylase
MW	Molecular Weight
N/A	Not applicable
NCBI	National Centre for Biotechnology Information
ng	nanogram
nm	nanometre
nt	nucleotide
ORF	Open Reading Frames
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
RdRp	RNA dependent RNA polymerase
RFLP	Restriction Fragment Length Polymorphisms
RISA	radio—immunosorbent assay
RNA	Ribonucleic Acid
RSA	Republic of South Africa
RT	Reverse Transcription
SDS	Sodium dodecyl sulphate
sgRNA	subgenomic Ribonucleic Acid
SNP	single nucleotide polymorphism
SNR	signal to noise ratio
SP	Stem-Pitting
SSCP	Single Stranded Conformational Polymorphisms

SSC	sodium chloride and trisodium citrate
SSEM	serologically specific electron microscopy
SY	Seedling Yellows
T	thymine
Taq	Taq polymerase
Tm	Melting Temperature
TMV	Tobacco Mosaic Virus
U	Uracil
U	Units
UP	University of Pretoria
µg	microgram
µl	microlitre
µM	micro Molar
UTR	Untranslatable region
UV	Ultra-Violet
Y	Tyrosine

FIGURE NUMBER	PAGE
1A. CTV particles.....	6
1B. Electron microscope picture of CTV.....	6
2. The Genome organization of CTV.....	6
3A. Severe stem-pitting on a citrus tree trunk.....	25
3B Seedling Yellows symptoms.....	25
4. Schematic representation of the primer positions of the p23 gene.....	54
5. Gel electrophoresis photo of the conserved PCR of p23 gene.....	62
6. Gel electrophoresis photo of the bi-directional PCR: severe versus atypic....	63
7. Gel electrophoresis photo of the bi-directional PCR: severe versus mild.....	64
8. Multiple Alignment of predicted amino acid sequences of p23 protein.....	68
9. Unrooted phylogenetic tree of p23 gene of RSA and reference isolates.....	74
10. Detection of CTV by TAS-ELISA with CTV specific polyclonal antibodies.....	101
11. Detection of CTV in three different plant parts by TAS-ELISA.....	101
12. Genotype profiles of isolates 389-4 and 390-5.....	105
13. Genotype profiles of isolates 12-5 and 12-7.....	106
14. Genotype profiles of isolates 12-9 and GFMS 35.....	107
15. Genotype profiles of isolates GFMS 12 and T30.....	108
16. Genotype profiles of isolates 390-3, 389-4 and 390-4.....	109
17. 5'-Cy3-labelling and amplification of targets for microarray hybridization....	137
18. 5'-Cy3-labelling and amplification of targets for microarray hybridization....	137
19. The SNR 532 Averages of T36 specific oligonucleotides	147
20. The SNR 532 Averages of T36 specific oligonucleotides	148
21. The SNR 532 Averages of negative control oligonucleotides	149

22.	The SNR 532 Averages of CTV strain negative control oligos.....	149
23.	The SNR 532 Averages of conserved oligonucleotides.....	150
24.	The SNR 532 Averages of T30 specific oligonucleotides	150
25.	Hybridization results of the T36 strain at 52 °C and 42 °C on a.....	151
26.	Hybridization results of the T30 strain at 52 °C and 42 °C on a.....	152
27.	Secondary structures of CTV T36 & T30 strain DNA PCR products.....	155
28.	The p23 sequence of South African isolate 12-5.....	176
29.	The p23 sequence of South African isolate 12-7.....	176
30.	The p23 sequence of South African isolate 12-9.....	176
31.	The p23 sequence of South African isolate GFMS 12.....	177
32.	The p23 sequence of South African isolate 389-3.....	177
33.	The p23 sequence of South African isolate 389-4.....	177
34.	The p23 sequence of South African isolate 390-5.....	178
35.	The p23 sequence of South African isolate GFMS 35.....	178
36.	Multiple Nucleotide sequence alignment of the p23 gene region.....	179
37.	Gel electrophoresis photo of T30 1+/- molecular marker PCR.....	184
38.	Gel electrophoresis photo of T30 2+/- molecular marker PCR.....	184
39.	Gel electrophoresis photo of T30 3+/- molecular marker PCR.....	185
40.	Gel electrophoresis photo of T30 4+/- molecular marker PCR.....	185
41.	Gel electrophoresis photo of T30 5+/- molecular marker PCR.....	186
42.	Gel electrophoresis photo of T30 6+/- molecular marker PCR.....	186
43.	Gel electrophoresis photo of T30 7+/- molecular marker PCR.....	187
44.	Gel electrophoresis photo of T36 1+/- molecular marker PCR.....	187
45.	Gel electrophoresis photo of T36 2+/- molecular marker PCR.....	188
46.	Gel electrophoresis photo of T36 3+/- molecular marker PCR.....	188
47.	Gel electrophoresis photo of T36 5+/- molecular marker PCR.....	189

48.	Gel electrophoresis photo of T36 6+/- molecular marker PCR.....	189
49.	Gel electrophoresis photo of T36 7+/- molecular marker PCR.....	190
50.	Gel electrophoresis photo of VT 1+/- molecular marker PCR.....	190
51.	Gel electrophoresis photo of VT 2+/- molecular marker PCR.....	191
52.	Gel electrophoresis photo of VT 3+/- molecular marker PCR.....	191
53.	Gel electrophoresis photo of VT 4+/- molecular marker PCR.....	192
54.	Gel electrophoresis photo of VT 5+/- molecular marker PCR.....	192
55.	Gel electrophoresis photo of VT 6+/- molecular marker PCR.....	193
56.	Gel electrophoresis photo of T3 2+/- molecular marker PCR.....	193
57.	Gel electrophoresis photo of T3 3+/- molecular marker PCR.....	194
58.	Gel electrophoresis photo of T3 5+/- molecular marker PCR.....	194
59.	Gel electrophoresis photo of T3 6+/- molecular marker PCR.....	195
60.	Pairwise Alignment of the T30 (AF2606) and T36 (U16304).....	196

TABLE NUMBER	PAGE
1. Bi-directional primers for RT-PCR of the p23 gene.....	55
2. Reference isolates of p23 gene used in phylogenetic analysis.....	59
3. Graphical representation of the amplified products of p23 gene.....	65
4. Predicted amino acid sequence similarity (%) of the p23 gene.....	69
5. Differences in Amino acid residues of 4 regions of p23 gene.....	70
6. Intra-group & inter-group genetic diversity values for three groups.....	73
7. PCR primers of the four genotypes used for RT-PCR.....	98
8. Summarized results of the 23 genotype -specific PCR Primers.....	104
9. Designed and synthesized primers and oligos used in the.....	127
10. The layout of the CTV T30 versus T36 strain microarray slide.....	129
11. PCR primer sequences (Hilf <i>et al.</i> , 2000) used for PCR amplifica.....	131
12. PCR primers used for the verification of the DNA clone sequence.....	133
13. The SNR 532 averages and standard deviation results for oligos	140
14. The SNR 532 averages and standard deviation results for oligos	141
15. The SNR 532 averages and standard deviation results for oligos	143
16. The SNR 532 averages and standard deviation results for oligos	144
17. A description and function of the oligonucleotides used in the.....	146
18. Final Results of expected versus obtained.....	153
19. Summary of optimised PCR conditions for each of the 23 primer.....	183

	Page
DECLARATION	ii
ACKNOWLEDGEMENTS	iii
DEDICATION	iv
ABSTRACT	v
LIST OF ABBREVIATIONS	vii
LIST OF FIGURES	x
LIST OF TABLES	xiii
 CHAPTER 1. INTRODUCTION AND LITERATURE REVIEW	 1
1.1 SUMMARY	2
1.2 INTRODUCTION	3
1.3 CTV MOLECULAR CHARACTERISTICS	5
1.4 CTV STRAIN DIFFERENTIATION	8
1.4.1 INDICATOR PLANTS	9
1.4.2 SEROLOGICAL METHODS	9
1.4.3 DOUBLE-STRANDED RNA PROFILES	11
1.4.4 PCR BASED METHODS	12
1.4.5 SSCP	13
1.4.6 HYBRIDISATION ASSAYS	14
1.4.7 DNA SEQUENCING	14
1.5 GENOME BASED METHODS FOR STRAIN DIFFERENTIATION	15
1.5.1 PCR	15
1.5.2 MICROARRAYS	17

1.6	VECTOR	18
1.6.1	VECTOR INTRODUCTION	18
1.6.2	STRAIN SELECTION	19
1.6.3	FACTORS AFFECTING TRANSMISSIBILITY	20
1.7	SEQUENCE VARIANTS	21
1.8	QUASI SPECIES & STRAIN DOMINANCE	23
1.9	SEQUENCE DIVERGENCE	24
1.10	RECOMBINATION OF STRAINS	25
1.11	CROSS-PROTECTION	26
1.12	CROSS-PROTECTION SCHEME IN RSA	28
1.13	BREAKDOWN	29
1.14	ENVIRONMENTAL PRESSURE	31
1.15	HOST-VIRUS INTERACTION	31
1.16	CONTROL OF STRAIN MOVEMENT	33
1.17	REFERENCES	34
CHAPTER 2		
CHACTERIZATION OF THE P23 GENE AND STRAIN DIFFERENTIATION OF SOUTH AFRICAN CTV ISOLATES USING A BI-DIRECTIONAL RT-PCR SYSTEM & PHYLOGENETIC ANALYSIS		48
2.1	INTRODUCTION	49
2.2	MATERIALS & METHODS	51
2.3	RESULTS	60
2.4	DISCUSSION	75
2.5	REFERENCES	87

CHAPTER 3	91
ESTABLISHMENT OF 23 PCR PRIMER SYSTEMS TARGETING THE 5' END OF CTV FOR STRAIN DIFFERENTIATION OF RSA ISOLATES INTO 4 GENOTYPES	
3.1 INTRODUCTION	92
3.2 MATERIALS & METHODS	94
3.3 RESULTS	100
3.4 DISCUSSION	110
3.5 REFERENCES	118
CHAPTER 4	122
DEVELOPMENT OF AN OLIGONUCLEOTIDE MICROARRAY CHIP TO DIFFERENTIATE T30 AND T36 STRAINS OF CTV	
4.1 INTRODUCTION	123
4.2 MATERIALS & METHODS	125
4.3 RESULTS	135
4.4 DISCUSSION	155
4.5 ACKNOWLEDGEMENTS	167
4.6 REFERENCES	168
CONCLUDING REMARKS	170
APPENDIX 1	176
APPENDIX 2	183
APPENDIX 3	196

CHAPTER 1

INTRODUCTION & LITERATURE REVIEW

1.1 SUMMARY

The *Citrus tristeza virus* (CTV) causes varying degrees of symptoms from none to very severe. Severe symptoms are mainly the decline of trees, stem-pitting and seedling yellows. The effect of severe strains on the citrus industry worldwide has been devastating. The severity of CTV depends on parameters such as the host plant/cultivar; the vector; the strain of CTV and the environment. How these interact is still unknown, to produce varying symptoms. CTV isolates are frequently mixed populations of different genomic RNA populations and often also contain multiple defective RNAs. These strain mixtures occur primarily due to transmission by different aphid species and by infected budwood. *Toxoptera citricida*, the brown citrus aphid is the most efficient vector and has the ability to select for the more severe strains. There are also many factors affecting transmission efficiency, including the species of aphid, the plant source, temperatures and virus titer. CTV sequence variants have evolved through time by genetic drift; negative and positive selection; bottlenecks; by recombination; strain competition; the effect of quasi species and dominant strains.

Ways to differentiate strains have been developed and are constantly being updated and enhanced to provide more accurate answers. Indicator plants are used to index the symptoms and thereby give an isolate a biological designation, even though the symptoms are specific to the cultivar used, environment and strain mixture composition. Serological methods using polyclonal and monoclonal antibodies have been valuable in detection of CTV. MCA13 was developed to differentiate decline and non-decline inducing isolates. Serological techniques however are limited to recognizing characters of the coat protein and this alone is not always sufficient to detect different variants of CTV. Double-stranded RNA (dsRNA) profiles of isolates are valuable in classifying isolates to a specific pattern. However symptoms could not be correlated to dsRNA patterns. In the late 1990's the polymerase chain reaction (PCR) is a test to specifically target variability within isolates. Many PCRs were developed to differentiate between certain strains and viral genotypes. Since then many isolates have been fully or partially sequenced. This has provided a large amount of information as to the variability of the CTV genome. It was found that the 5' half of the genome is extremely variable whereas the 3' half is much more conserved. Single strand conformational polymorphisms (SSCP) have been used to detect single base mutations; however differences in patterns don't necessarily mean great differences in

nucleotide sequences. With nucleic acid hybridization severe strains could be detected with nine coat protein probes (Cevik *et al.*, 1995). More recently it has become important to target the whole genome when designing ways to accurately differentiate strains. Currently a whole range of PCR primers designed to target four main CTV genotypes across the 5' half of the CTV genome have shown a lot more promise. Oligonucleotide microarray technology could potentially be useful in detecting strain markers.

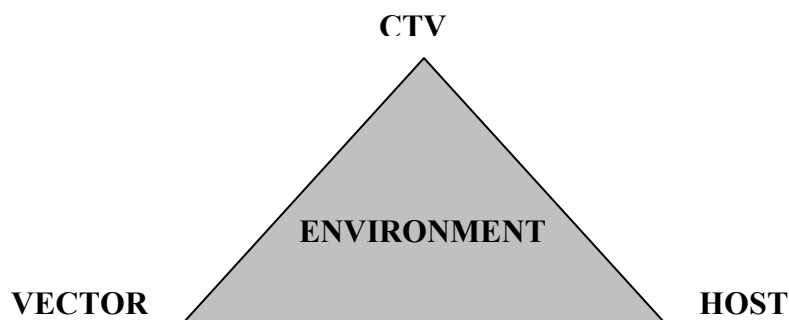
The control of CTV is mainly through eradication of severe strains, quarantine certification schemes and by mild-strain cross-protection. This form of control has been largely successful and has allowed a number of citrus industries to produce citrus economically again. There has however been cross-protection breakdown in plants. The exact reason for breakdown is not fully understood but strain separation, super-infection, strain dominance, recombination of strains, decreased viral replication and the effects of interactions of multiple or single strains and the cultivar have been suggested as possible reasons.

1.2 INTRODUCTION

The *Citrus tristeza virus* (CTV) is an aphid-borne closterovirus; and has ranked as one of the most important citrus diseases for the last sixty years (Bar Joseph *et al.*, 1989) and yet viral genetic basis of CTV disease is poorly understood. It is not yet possible to attribute particular symptoms on a particular citrus cultivar to specific viral sequences, nor is it known which sequences influence transmissibility by different aphid species. CTV causes multiple disease syndromes in citrus. CTV epidemics have caused the severe decline or death of millions of trees in many areas where sour orange (*Citrus aurantium*) was used as a rootstock. In other areas, the productivity of grapefruit and specific sweet orange varieties was considerably affected by CTV strains causing stem pitting decline irrespective of the rootstock. It is particularly a concern in South Africa where grapefruit production occurs for export.

The spread and movement of tristeza depend upon the distribution of infected budwood, the different species of vector and their abundance, strain of the virus present, the citrus variety infected and environmental temperatures (Roistacher *et al.*, personal communication). The disease follows the parameters of a typical disease triangle as shown

below. Yet how CTV, the vector, the host and the environment interact is still unknown. This is particularly complex with regards to different strains of CTV.



Aphid vectors will spread tristeza locally and internationally if they are carried on citrus plants or fruits (Garnsey *et al*, 2000). However, propagation of citrus with infected budwood is the most important means of domestic and international spread of tristeza (Garnsey *et al*, 2000).

Currently measures to control losses caused by CTV include quarantine systems to avoid introduction of exotic isolates; certification programmes to prevent spread; and cross-protection with mild strains. Eradication or cross-protection aimed at controlling severe variants of the virus requires accurate and reliable strain characterisation, and differentiation procedures. Another problem the citrus industry faces is mild strain cross-protection breakdown, where many possible causes for this have been suggested. This review aims at discussing all aspects relating to CTV strains and CTV caused disease, to form a better picture of the complexity of developing and implementing a CTV strain differentiation strategy.

1.3 CTV MOLECULAR CHARACTERISTICS

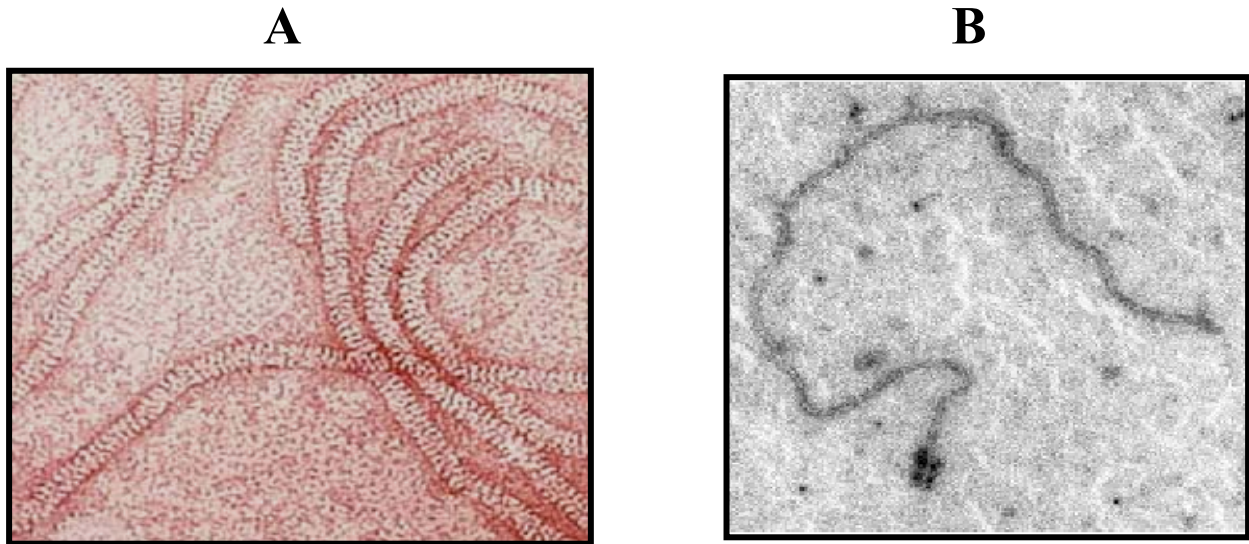
CTV is the largest RNA plant virus and belongs to the Closterovirus genus (Bar-Joseph *et al*, 1979) of the Closteroviridae family. The virions are flexuous, thread-like (Figs 1A and 1B) (Bar-Joseph *et al*, 1979), about 2,000 x 12 nm in size (Kitajima *et al*, 1963) and contain a non-segmented, positive-sense, single-stranded RNA genome (Kitajima *et al*, 1963). CTV is restricted to the phloem cells of *Citrus* species.

The complete sequences of 11 CTV genomes have been sequenced and submitted to the National Centre Biotechnology Institute (NCBI) GenBank Database. These sequences are

isolates T36 (AY170468, NC_001661 and U16304) and T30 (AF260651) from Florida (Albiach-Martí *et al.*, 2000b, Karasev *et al.*, 1995, Pappu *et al.*, 1994a), VT (U56902) from Israel (Mawassi *et al.*, 1996), SY568 (AF001623) from California (Yang *et al.*, 1999), T385 (Y18420) from Spain (Vives *et al.*, 1999), Qaha (AY340974) from Egypt (Abdelmaksoud, unpublished), T318A (DQ151548) strain from Spain (Ruiz-Ruiz *et al.*, 2006), a Mexican severe strain (DQ272579) (Quiroz, unpublished), and NUagA from Japan (AB046398) (Suastika *et al.*, 2001). The genomic RNAs (gRNAs) vary from 19226 to 19302 nucleotides (nt). The 3' half of the genome is relatively conserved with 90 % sequence identity. However the 5' half has less than 70 % sequence identity (Ayllón *et al.*, 2001). The only CTV sequences available from South Africa are of ORF 1 from three isolates.

The genome (figure 2) has the potential for encoding up to 19 proteins (Vives *et al.*, 1999). The genome is organized into 12 open reading frames (ORFs) (figure 2) and untranslated regions (UTRs) at both 5' and 3' termini (Karasev *et al.*, 1995, Pappu *et al.*, 1994b). Many RNA viruses such as CTV have cis-acting initiation of minus-stranded RNAs and subgenomic RNAs, initiation of virion assembly, regulation of gene expression, and probably movement within the host and induction of diseases (Miller *et al.*, 1998). López *et al.* (1998) analyzed the terminal sequences of 10 isolates of CTV representing three groups that differ by up to 56% in the 5' UTR, yet found that all were predicted to fold into almost identical structures consisting of two stem-loops. The remarkable conservation of these structures amid such variation in sequence suggested functionality, and later these structures were determined to be important for virus replication and virion assembly (Gowda *et al.*, 2003).

The two 5' end ORFs (ORF 1a and ORF 1b), are directly translated from gRNA and encode replication related proteins. The 5' terminal half has two presumed papain like proteases and the replication associated complex including RdRp (RNA dependent RNA polymerase), helicase (HEL), and methyltransferase (MT) domains. The RdRp domain is fused to the HEL and MT domains in a giant polyprotein via a +1 translational frameshift (figure 2) (Karasev *et al.*, 1995). But the ten 3' proximal ORFs 2-11 (Fig 2), are expressed from subgenomic RNAs (Hilf *et al.*, 1995).



Figures 1: (A) CTV particles, measuring 10 x 2,000 nm, at 80 000x magnification (Photographed by M.Bar-Joseph); and (B) Electron microscope picture of CTV (Niblett *et al*, 2000).

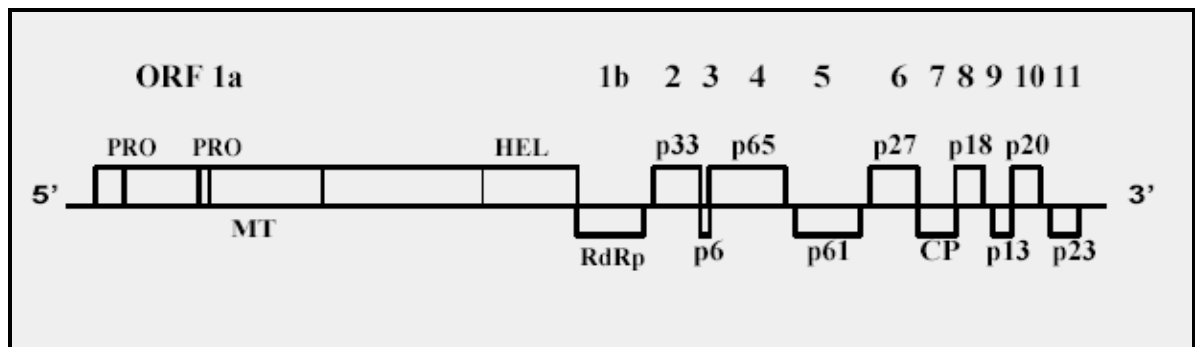


Figure 2: The Genome organization of CTV into 12 ORFs and depicting the proteins encoded (Karasev *et al.*, 1995).

ORF 3 encodes p6, a putative 6 kDa hydrophobic protein, postulated to be associated with the membrane (Karasev *et al.*, 1995). ORF 4 is a 65 kDa (p65) protein, a homologue of the heat-shock protein 70 (HSP70). The 61 kDa protein (p61) of ORF 5 may be involved in the assembly of multisubunit virions or in disarming the host defence response (Dolja *et al.*, 1994). p65 and p61 together with the two coat proteins (Febres *et al.*, 1994, Febres *et al.*, 1996) are involved in efficient virion assembly (Satyanarayana *et al.*, 2000) and are highly conserved. Assembly of the viral genome into virions is a critical process in the virus life cycle and crucial for the ability of the virus to move within the plant and to be transmitted horizontally to other plants.

The genome is encapsidated by a single type of coat protein. When purified CTV is analysed, two polypeptides are distinguished: the larger coat protein (CP) and the minor coat protein (CPm) (Lee *et al.*, 1988). ORF 7 encodes for the major CTV capsid protein of 25 kDa (CP or p25). ORF 6 encodes the related minor coat protein of 27 kDa (CPm or p27). Both forms of the protein react with CTV-specific antisera and have similar amino acid composition and proteolytic peptide maps (Lee *et al.*, 1988). ORF 10 encodes for the p20 (20 kDa) protein is the major component of these inclusion bodies of CTV infected cells (Gowda *et al.*, 2000). ORF 11 encodes the p23 (23 kDa) protein and is an RNA binding protein (Satyanarayana *et al.*, 2002).

The functions of ORF 2 encoding the p33 protein; ORF 8 encoding the p18 protein and ORF 9 encoding the p13 protein are presently unknown. It has been postulated that these proteins have evolved in CTV along with the expansion of the genome (Dolja *et al.*, 1994).

The defective RNAs (D-RNAs) are virion RNA molecules that contain the 5' and 3' termini of normal gRNA but are missing different internal portions (Mawassi *et al.*, 1995). D-RNAs are variable between isolates and are suggested to correlate to SY symptom expression (Yang *et al.*, 1999). CTV isolates are frequently mixed populations of different genomic RNA populations from which variants with distinct properties can be selected and often contain multiple D-RNAs that vary in size, abundance and sequence (Luttig *et al.*, 2002). A short 0.8 kb non-encapsidated single-stranded positive-sense RNA species was also found in infected plants (Mawassi *et al.*, 1995) and encompasses the 5' terminal part of the CTV genome (Mawassi *et al.*, 1995). The importance of this short piece of RNA will be discussed in a later section.

1.4 CTV STRAIN DIFFERENTIATION

A virus species might be defined simply as a collection of strains with similar properties. There are generally two different kinds of criteria for the discrimination of virus strains:

1. Structural criteria based on the properties of the virus particle itself and its components.
2. Biological criteria based on various interactions between the virus, its host plant and its vectors (Hull, 2002).

To designate CTV isolates as strains using the above criteria of Hull (2002) it becomes clear that the process of strain differentiation is complex. CTV strains are structurally

identical; they are composed of the same genes; same filamentous size and shape; and composed of genomes all in the range of 19.2 – 19.7 kb. However when biological criteria is exploited it is apparent that the CTV strains do not induce the same symptoms, are transmitted at varying degrees of efficiency from different species of aphid vectors and environmental conditions also play a role in strains performing differentially.

All methods of control need rapid procedures to detect CTV and differentiate between mild and severe isolates. CTV exists as a large number of distinct strains differing in biological properties and with different distributions in citrus producing countries (Gillings *et al.*, 1993). Rapid differentiation between mild and severe CTV strains is needed to aid studies on the epidemiology of tristeza, development of more effective cross-protection control strategies, and prevention of further propagation of severe CTV strains through budwood sources. The next sections describe all the current and future strain differentiation methods.

1.4.1 INDICATOR PLANTS

Isolates of CTV are traditionally differentiated by inoculating indicator plants and observing the symptoms (Huang *et al.*, 2004) also termed biological characterization. The indicator plant, Mexican lime (*Citrus aurantifolia*) is used and is also referred to as the seedling lime index. Plants are graft-inoculated with buds or bark patches from candidate trees and incubated in a temperature-controlled glasshouse at 18-25 °C. Symptoms such as vein clearing, leaf cupping and stem pitting symptoms vary, according to severity of the CTV isolate. A panel of five species combinations are also used for strain indexing (Garnsey *et al.*, 1987b). Another method to standardize the symptoms observed is to give a severity score (Garnsey *et al.*, 1987b) or to place the isolate in one of 11 reaction types based upon reactions to reference isolates in the world CTV collection (Lee *et al.*, 1994). Differential host reactions (DHR) on selected indicators and combinations of DHR and vector transmissibility have also been used (Balaraman *et al.*, 1980).

Bioindexing is however time-consuming (12-15 months), expensive (greenhouse space, personnel, management of pests etc.) and imprecise for strain differentiation (Cambra *et al.*, 1993). CTV isolates vary in the symptoms induced on citrus cultivar combinations under glasshouse conditions. Field symptoms of CTV are often more variable than those observed under controlled conditions (Cambra *et al.*, 1993). A CTV isolate that displays

mild symptoms when biologically indexed may contain severe strains that are expressed under certain conditions (Cambra *et al.*, 1993).

1.4.2 SEROLOGICAL METHODS

Polyclonal antiserum, specific for CTV, reacts with both mild and severe strains and several serological detection procedures have been developed from it. These include SDS-immunodiffusion (Brlansky *et al.*, 1984), direct tissue blot immunoassay (DTBIA) (Garnsey *et al.*, 1993), enzyme-linked immunosorbent assay (ELISA), dot-immunoblotting assay (Rocha-Peña *et al.*, 1991), radio-immunosorbent assay (RISA) (Rocha-Peña *et al.*, 1991) using ²⁵I serologically specific electron microscopy (SSEM) (Brlansky *et al.*, 1984), in situ immunofluorescence (ISIF) (Brlansky *et al.*, 1984), in situ immunoassay (ISIA) and a Western blot assay (Rocha-Peña *et al.*, 1991). ELISA (Bar-Joseph *et al.*, 1979a) is convenient but limited in sensitivity of detection especially with very low concentrations of CTV.

A major breakthrough was the development of CTV monoclonal antibodies (MCAs) (Vela *et al.*, 1986). Most of the monoclonal antibodies evaluated appear to be polyspecific, i.e. they recognize all CTV strains tested. However, the development of antibody, MCA-13, enabled differentiation of severe and mild strains from Florida using ELISA (Niblett *et al.*, 2000). Mild strains have tyrosine (Y) and severe strains have phenylalanine (F) at position 124 (Pappu *et al.*, 1993). The A or T in position 371 of the CP gene determines whether the amino acid will be Y or F, respectively, and MCA13 reacts only when F is present (Pappu *et al.*, 1993; Niblett *et al.*, 2000).

In Florida, the strong reactivity of MCA-13 with stunted trees confirmed the poor growth was associated with a CTV effect (Niblett *et al.*, 2000). MCA 13 reacts with most severe strains (Vela *et al.*, 1986) but it is impossible to say whether the sample contains a stem pitting or a decline strain or a mixture of severe strains (Niblett *et al.*, 2000). Serological methods have become a dependable tool though for CTV detection and isolate differentiation. A positive reaction with MCA 13 is commonly used as a first step for differentiation of decline-inducing and non-decline inducing isolates of CTV (Huang *et al.*, 2004). It is unclear if non-reactive samples were infected with a mild strain, non MCA-13 reactive severe strain or none of the above (Niblett *et al.*, 2000). A limitation of monoclonal antibodies (Cambra *et al.*, 1993) or peptide map analysis (Moreno *et al.*, 1990)

is that changes are only detected in the coat protein gene, which occupies just 3.5 % of the genome (Rubio *et al.*, 1996). Proteolytic digestion enabled differentiation between some of the Spanish isolates and between the Spanish isolates and those from Florida (Moreno and Lee, unpublished, cited by Rubio *et al.*, 1996). Serological analysis of peptide maps is a powerful technique for analyzing complex protein mixtures combining the discrimination properties of peptide analysis and the specificity of serological assays (using monoclonal and polyclonal antibodies) to differentiate isolates. However it is too cumbersome, expensive and takes several days for routine utilization in strain differentiation (Rubio *et al.*, 1996). And a major limitation is the ability to obtain sufficient amount of CP pure enough to prepare peptide maps.

1.4.3 DOUBLE-STRANDED RNA PROFILES

The electrophoretic patterns of the double stranded replicative RNA (dsRNA) forms of CTV were used to detect and differentiate strains of CTV (Moreno *et al.*, 1990). Minor dsRNAs are useful for making distinctions between strains of some RNA plant viruses (Dodds *et al.*, 1987a). CTV dsRNA profiles have also been used to discriminate between certain damaging forms of CTV (Dodds *et al.*, 1983, Dodds *et al.*, 1984) and a strong correlation between the ability to detect a 0.5×10^6 MW dsRNA and the biological designation of an isolate as either a SY type, or a SP isolate of grapefruit was observed (Dodds *et al.*, 1987a).

Moreno *et al.* (1990) examined the dsRNA patterns of 24 Spanish isolates representing a wide range of biological diversity. Seven distinct patterns were observed based on differences of the number and/or position of the dsRNA bands (Moreno *et al.*, 1990). Variations in dsRNA yield under favourable conditions may be due to strain differences in the rate of virus multiplication (Moreno *et al.*, 1990). Virus antigen titre and total dsRNA recovery was not always correlated with the intensity of symptoms induced on Mexican lime (Moreno *et al.*, 1990) and could not be correlated with any particular dsRNA pattern (Moreno *et al.*, 1990). Several strains showed host-induced variation of the dsRNA profile and in some cases seasonal variation (Moreno *et al.*, 1990). On the contrary to previous studies where a 0.5×10^6 MW band correlated to stem-pitting and seedling yellows isolates, Moreno *et al.* (1990) found several mild strains to have this band too. It becomes clear in the literature that the use of the word “mild” or “severe” for strains can cause a lot of confusion. This is especially important when one does not know if a plant is infected with a single type strain or a mixture and the effects of strain dominance and competition. Strain

mixtures occurring in isolates also further complicate the analysis. However dsRNA analysis of infected plants is easy to perform and can detect variations in different portions of the genome (Dodds *et al.*, 1987b).

1.4.4 PCR BASED METHODS

Polymerase Chain Reaction (PCR) is a molecular biological technique for amplifying selected sections of DNA by using two short DNA sequences (primers) that flank the beginning and end of sequence.

The 5'-UTR and 5'-proximal coding region of the CTV genome have been found to be highly polymorphic (Yang *et al.*, 1999). Sequencing of 58 DNA clones representing 15 virus isolates of this region showed that all sequences could be assigned to one of three types (Allyón *et al.*, 2001) established by López *et al.* (1998). RT-PCR was performed with sets of type-specific primers which differentiate 3 types: (I) T36 strain specific; (II) VT strain specific and (III) T317 stain specific (López *et al.*, 1998, Allyón *et al.*, 2001). None of the isolates yielded amplification of only type I or II alone, but in 19 of them only type III products were amplified and these caused only mild symptoms (Allyón *et al.*, 2001). Eight isolates contained type II and III sequences; 11 had type I and III sequences and 19 had all three types (Allyón *et al.*, 2001). Isolates causing stem-pitting generally contained type II sequences (Allyón *et al.*, 2001).

The amino acid difference in mild and severe strains at position 124 of the CP enabled the development of bi-directional PCR for differentiating CTV strains as well as detecting mixed infections (Cevik *et al.*, 1996). The cloning and sequencing of the CP gene was used for comparisons of CTV isolates with different geographic origin and/or biological properties (Pappu *et al.*, 1993). The CP primers used by Cevik *et al.* (1996) were used by Huang *et al.* (2004) with slight modifications to differentiate between non-decline (T30) and decline (T36) inducing isolates of CTV in infected trees as well as for detection of both isolate types in mixed infections (Huang *et al.*, 2004). However there was not a good correlation between field symptoms and the molecular tests. Some of the plants with decline symptoms were shown to be infected with non-decline isolates; this was attributed to significant citrus root weevil damage causing decline (Huang *et al.*, 2004). And other plants with non-decline symptoms were infected with decline-inducing isolates, possibly due to the fact that trees are infected for at least six years with decline-inducing isolates

before going into decline (Powell, unpublished cited by Huang *et al.*, 2004; Huang *et al.*, 2004).

PCR and RFLP (restriction fragment length polymorphism) were used to compare isolates (Rubio *et al.*, 1996., Gillings *et al.*, 1993). Based on the CP, a RFLP assay was developed using *HinfI* or *RsaI* restriction enzymes. Gillings *et al.* (1993) found that *HinfI* restriction enzyme digests of CP gene discriminated between strains, identifying seven patterns which were associated with specific biological activities. The findings of several RFLP types in individual reactions raised the question of mixtures of CTV strains (Gillings *et al.*, 1993). Roy *et al.* (2003) also performed RFLP studies on the CP gene and found that digestion with *HinfI* and *RsaI* revealed considerable polymorphisms between Indian and other exotic CTV isolates. This method meant that CTV strains could be differentiated without cloning and sequencing the CP gene (Niblett *et al.*, 2000). Some disadvantages of this method are that only nucleotide changes affecting restriction sites were detected but these may not result in alteration of the amino acid sequence. And the PCR products may need to be tested with many different restriction enzymes (Rubio *et al.*, 1996).

1.4.5 SINGLE STRAND CONFORMATIONAL POLYMORPHISM (SSCP)

This technique can be used to detect single base mutations in DNA between different individuals in a population. With ssRNA viruses, RT-PCR is performed on a specific region of the genome and the DNA is denatured and separated by non-denaturing polyacrylamide gel electrophoresis (PAGE). Different conformations are formed based upon the composition of the nucleotides in the strands which migrate at different distances from the origin. Research done on isolates of several geographical origins were compared for variations in the coat protein gene by analysis of SSCP and used to identify and distinguish strains of CTV (Rubio *et al.*, 1996). The banding patterns of mixed strains become very complex to visualize (Rubio *et al.*, 1996). And differences in SSCP patterns don't necessarily mean great differences in nucleotide sequence (Rubio *et al.*, 1996). It is inadequate for rapid comparisons of multiple samples (Niblett *et al.*, 2000). It was found that the same population of variants propagated on different host cultivars produced a different SSCP pattern profile indicating host selection of the predominant sequence variants in the population (Rubio *et al.*, 2000). This might help to explain biological diversity observed among CTV isolates. Sambade *et al.* (2002) performed SSCP analysis with genes p18, p13, p20 and p23 and found that mild isolates yielded two DNA bands suggestive of a more

predominant sequence variant, whereas virulent isolates contained more than two DNA bands. Cross-protecting sub-isolates were differentiated by SSCP of the coat protein gene (van Vuuren *et al.*, 2000).

1.4.6 HYBRIDISATION ASSAYS

Hybridisation of viral RNA with complementary DNA (cDNA) probes can potentially detect changes all along the genome (Lee *et al.*, 1987). Nucleic acid hybridisation was used originally to distinguish sequence variants by probes as well as to determine whether individual citrus plants were infected with more than one strain of CTV (Niblett *et al.*, 2000). Nine CTV strain discriminating CTV CP oligonucleotide probes have been developed based upon sequence analyses of the CP gene from diverse biological and geographical conditions (Cevik *et al.*, 1995, Nolasco *et al.*, 1997, Niblett *et al.*, 2000). These probes were used to detect severe strains in Portugal, Madeira and Florida.

A method based on non-isotopic hybridization was used to differentiate isolates using digoxigenin (DIG) labelled cDNA probes and different RNA targets (Narvaez *et al.*, 2000). It is rapid and sensitive and can be applied to many samples (Narvaez *et al.*, 2000). Hybridisation with cDNA is an easy and powerful method for detecting differences located in any genome region (Rosner *et al.*, 1986) but its use has been restricted in the past by the need for RNA purification and by radioactive probes which have a short life and safety hazards (Narvaez *et al.*, 2000). An advancement in nucleic acid hybridisation is microarrays and this will be discussed in another section.

1.4.7 DNA SEQUENCING

The nucleotide and the deduced amino acid sequences of various genes have been used to compare CTV isolates by various phylogenetic analysis but most times single or very few clones were sequenced from each isolate. CTV infected plants do not contain a unique genomic sequence but instead they have a population of variants usually clustered around one or more consensus sequences. The composition of this population may determine the pathogenic characteristics of CTV isolates and it is therefore important to characterize genomic populations for clues in identifying specific groups of isolates. Sequencing of multiple variants of one or more genes could be an accurate procedure to characterize the CTV isolates. Sequencing has its disadvantages in being too expensive and not appropriate

for routine differentiation. This has however been recently overcome with the development of the resequencing microarray chip (Xiong *et al.*, 2005) which can potentially sequence the complete genome of an isolate in two weeks (Xiong *et al.*, 2005). If this technology is accessible and affordable to the public it will become a very valuable tool in characterizing strains. It also becomes evident from the literature that different approaches should target the whole genome in one test to get a more accurate answer about the variability in different regions of the genome.

A limitation of SSCP, sequencing and RFLP is the need for the synthesis of cDNA by reverse transcription and PCR amplification (RT-PCR) and in this step primers may selectively amplify some RNA variants which are only minor components of the viral genomic RNA population (Narvaez *et al.*, 2000) or alternatively major components and therefore exclude the minor components.

1.5 GENOME BASED METHODS FOR STRAIN DIFFERENTIATION

1.5.1 PCR

A PCR based assay was developed to characterize CTV isolates by amplification of many sequence specific molecular markers (using sequence specific primer sets) derived from full genome sequences of VT, T3, T30 and T36 isolates (Hilf *et al.*, 1999, Hilf *et al.*, 2000, Hilf *et al.*, 2002, Hilf *et al.*, 2005). Hilf *et al.* (2000) found that the molecular markers for the 4 viral “genotypes” are conserved in global CTV populations. The technique can be used to determine the relatedness of isolates within and between regions, to track the movement of these genotypes in regions as well as to provide an initial measure of overall molecular variability of global populations (Hilf *et al.*, 2000). It can also be useful in understanding the population structure like the genotype prevalence of mixed infections, and selection of isolates for cross-protection (Hilf *et al.*, 2000). It was found that T3, T30 and T36 genotypes are not restricted to Florida and the VT genotype is not restricted to Israel (Hilf *et al.*, 2000). A high percentage of Indian isolates were VT genotypes (Roy *et al.*, 2003), a direct connection is unknown for this result (Hilf *et al.*, 2000).

In Florida it was found that the majority of the 400 isolates tested from the field were T30 and T36 genotypes, either singly or together (Hilf *et al.*, 2002). The VT genotype was found in some Meyer lemon trees and remains localised to this cultivar in Florida (Hilf *et al.*,

2002). It could be speculated that there is vector specificity for this genotype or that there are unspecified factors affecting spread of this genotype out of Meyer lemons (Hilf *et al.*, 2002). Biological indexing supported the observation that T36 genotype is associated with graft incompatibility on sour orange rootstock (Hilf *et al.*, 2002). It was mentioned that there is limited genetic variability of CTV currently present in Florida and genetically based assays will prove to be valuable in future control of exotic strain introductions. An interesting finding was that some T30 genotype isolates were positive by serological means for MC13 whereas the type isolate for T30 is MC13 negative (Hilf *et al.*, 2002). The T30 genotype is common worldwide and is said to be relatively genetically stable (Albiach-Martí *et al.*, 2000a). The T36 genotype is rarer worldwide and could suggest a more specific origin. T36 is MC13 positive; however some isolates that tested positive for T36 with genetic tests were negative for MC13 (Hilf *et al.*, 2002). This finding raises questions as to the use of serological tests to differentiate decline versus non-decline inducing isolates. It is clear from these findings that biological, serological and genetic based tests are valuable in achieving a complete picture of these CTV isolates.

In 2004, 21 Indian isolates were tested and 15 contained only the VT genotype, while the rest contained mixtures of T30 or T3 with VT, as well as T3 and VT genotypes (Roy *et al.*, 2004). One isolate had a mixture of the T36, T30 and T3 genotypes (Roy *et al.*, 2004). In 2005 it was found that 19 Indian isolates were VT specific, five were T30 specific and one was T36 specific (Roy *et al.*, 2005). A phylogenetic tree placed all the isolates into four distinct genogroups, VT, T36, T30 and the fourth represented by BAN-2 isolate (Roy *et al.*, 2005). It was noted that there was more sequence variability between nucleotides 1082-1484 than areas closer to the 5' end (Roy *et al.*, 2005). From this research it was hypothesized that three isolates in India are three naturally occurring variants that add to the complexity of the CTV populations in India (Roy *et al.*, 2005).

Multiplex and bi-directional PCRs have been developed, and the resulting gel patterns can be complex. Additionally, infected plants there is strain dominance which could lead to template bias in amplification. However recent work in the last few years has shown excellent progress in the optimization of complicated multi-primer PCRs as shown by Sambade *et al* (2003) with the classification of p23 gene into three groups: mild, severe or atypical. And Huang *et al* (2004) with the CP gene classified isolates as decline or non-

decline inducing. There are many more examples of different articles describing PCR methods to discriminate strains that are not covered here.

1.5.2 MICROARRAYS

The most limiting factor of all these methods is that only one pathogen/strain is detected per assay reliably. Hadidi *et al* (2001) predicted that detection and identification of many plant viruses would become a feasible strategy with DNA microarrays and would overcome all the downfalls of the methods discussed above (Hadidi *et al.*, 2004).

DNA microarrays exploit the feature of DNA complementarity through base pairing and the application of short DNA molecules (oligonucleotides) to a solid surface (Hadidi *et al.*, 2004). Very large numbers of DNA oligonucleotides can be applied to solid surfaces in ordered two-dimensional arrays by robotics, thus allowing parallel analysis of hybridization (Hadidi *et al.*, 2004). Each probe or oligonucleotide is specific for a DNA sequence of interest. It provides a medium for matching known and unknown nucleic acid samples based on base-pairing rules and automating the process of identifying the unknown. DNA microarrays have the potential to decrease the amount of labour required for sample preparation, time required for data analysis, total time required to run the assay and total cost. The main advantage is simultaneous detection and quantification of thousands of hybridization events with highly automated technology and a greater scope for miniaturization (Hadidi *et al.*, 2004). Current disadvantages could be the high cost of the machines needed to perform arrays; however costs have already started decreasing with an increased demand.

Mutations leading to new variants can only be detected by nucleic acid based technology (Deyong *et al.*, 2005). Microarrays would be an ideal supplement to solve the problem of differentiation after generic amplification (Deyong *et al.*, 2005). A system for microarrays was developed to detect and differentiate cucumber mosaic virus (CMV) serogroups and subgroups. The coat protein genes of 14 different isolates were amplified using Cy3-labelled generic but CMV specific primers (Deyong *et al.*, 2005). It was found that just five specifically selected oligonucleotides can result in differential hybridization and correct designation into serotype and subgroups (Deyong *et al.*, 2005). This clearly shows the potential application for strain differentiation based on oligonucleotides-based microarray technology. It was shown that short oligonucleotides are suitable as capture probes for the

discriminative hybridization of isolates differing by less than 8 % in an amplified PCR product larger than 700 bp (Deyong *et al.*, 2005). This shows the possible application of microarrays for CTV detection and strain differentiation.

1.6 VECTOR

1.6.1 VECTOR INTRODUCTION

Aphids transmit CTV in a semi-persistent manner, with no latent period; acquisition and inoculation periods being at least 30 minutes in some cases (Bar-Joseph *et al.*, 1989). The time for the aphid to tap into the phloem is essential for viral transmission. Aphids can remain viruliferous for at least 24 hours (Raccah *et al.*, 1976a) and can retain inoculativity during this period when transmitted to secondary hosts (Raccah *et al.*, 1976b).

Transmission occurs by species in genera *Aphis* and *Toxoptera* and examples include the Brown Citrus Aphid (BCA), *Toxoptera citricida* (Kirkaldy) the most efficient vector (Roistacher *et al.*, 1987), *Aphis gossypii* (Glover), *A. spiraecola* (Patch) and *T. aurantii* (Boyer de Fonscolombe). Costa and Grant (1951) showed that a single aphid of *Toxoptera citricida* could transmit the tristeza disease. *T. citricida* has been responsible for much of the natural spread of CTV in citrus-growing areas of South America, South Africa, Australia, USA and Asia (Costa *et al.*, 1951). The earliest detailed study about the biology and life history of BCA is by Symes (1924) where it was reported that there are as many as 30 generations per year and that the time for development to adult was 8-21 days. The BCA is anholocyclic (without sexual generations) and apterous (without wings) females produce nymphs. The BCA feeds on newly expanding shoots, leaves and flower buds of its host plant which are suitable for growth and reproduction for usually a period of only 3-4 weeks, depending on environmental conditions (Michaud, 1998). When the source of food diminishes and overcrowding occurs apterae develop wings, (alates) and fly away. The majority of alates do not fly far from their original colony (Michaud, 1998). BCA infestations tend to be endemic in citrus groves, surviving at low density on bits of asynchronous flush and root sprouts (Michaud, 1998). Long-range dispersal by alates is probably rare and movement is associated with human movement of infested material (Michaud, 1998). *T. citricida* nymphs and apterae transmitted CTV more efficiently than alatae whereas in the case of *A. gossypii* both mature and immature aphids and the alatae transmitted CTV at similar rates (Norman *et al.*, 1969).

The establishment of *T. citricida* in a new location is typically followed in 3-10 years by large scale outbreaks of CTV decline (Garnsey *et al.*, 2000). Whereas, in many areas a lag period of more than 30 years has been experienced between the early introduction of the virus and the natural spread by *A. gossypii* (Bar-Joseph *et al.*, 1978a). *Aphis gossypii* was found to be the primary vector in California even though it represented only 4% of the aphid population (Dickson *et al.*, 1951). Norman and Grant (1956) showed that *A. spiraecola* and *A. gossypii* would transmit CTV in Florida, also at low rates of transmission. On the other hand, transmission by the melon aphid, *Aphis gossypii* varied considerably on different occasions (Bar-Joseph *et al.*, 1977).

1.6.2 STRAIN SELECTION

A number of studies have shown that variant strains of CTV differ in their transmissibility by aphids (Raccah *et al.*, 1978). Various researchers have found that *T. citricida* is the more efficient vector compared to the other species of aphids (Celino *et al.*, 1966, Sharma *et al.*, 1989, Yokomi *et al.* 1991). Experimental transmission tests with CTV isolates showed that *T. citricida* is 6-25 times more efficient at CTV transmission than *A. gossypii* (Yokomi *et al.*, 1994). Balaraman *et al.* (1979) showed that severe strains were transmitted with a higher percentage with *T. citricida* than mild strains, the acquisition periods were shorter and retention periods were longer.

Increased transmissibility of certain CTV isolates by *A. gossypii* was inferred as the primary cause for the natural spread of the disease in Israel (Bar-Joseph *et al.*, 1978a). Early reports from California indicated rather low rates of CTV transmission by *A. gossypii* (Dickson *et al.*, 1951) although some SY-CTV isolates were transmitted efficiently (Martinez *et al.*, 1964). Extensive transmission trials (Roistacher *et al.*, 1984) during those last twenty years demonstrated the presence of both highly transmissible (10 %) and poorly transmissible CTV isolates in California (Roistacher *et al.*, 1984).

The efficiency of CTV transmissibility is affected by the species of aphid, by the source plant at acquisition feeding and the CTV isolate. It was mentioned in the previous section that different aphid species differ in their transmission efficiencies.

Cases of varying rates of transmission, probably resulting from mixed infections (Grant *et al.*, 1957) or from inefficient virus spread in the acquisition hosts were reported (Raccah *et al.*, 1978). It has been suggested that dominance of non-transmissible isolates, probably through cross protection, prevents the early establishment of the more transmissible mutants (Raccah *et al.*, 1978). It is very important to understand that the relative dominance of an isolate irrespective of its severity would probably be the most defining factor in the likelihood of being efficiently transmitted. The chances of an isolate in low percentage of a population of isolates being transmitted is statistically lower, except if that isolate has an inherent feature making it more efficiently transmitted.

CTV transmissibility, both by *T. citricida* and *A. gossypii*, was markedly affected by the source plant used for acquisition feeding (Bar-Joseph *et al.*, 1979). Varieties of sweet orange were more suitable for acquisition and more sensitive to infection when compared to grapefruits and lemon seedlings inoculated with three different CTV isolates (Roistacher *et al.*, 1984). Field observations in Florida, Spain, and Israel (Bar-Joseph, unpublished, cited by Roistacher *et al.*, 1984) indicated that only few grapefruit trees declined or became infected in areas where all sweet orange declined. Colonization experiments indicated that *A. gossypii* had no preference for the young sweet orange or grapefruit leaves but that there was a significant reduction (60%) in the number of aphids that fed on the lemon leaves (Roistacher *et al.*, 1984).

CTV quantification by ELISA indicated a good correlation between virus titer in these hosts and their efficiency as acquisition hosts (Roistacher *et al.*, 1984). This should not, however, be taken as a rule since certain plants (Bar-Joseph *et al.*, 1973) will behave as poor acquisition hosts even though they support high titers of CTV.

Single *T. citricida* transmitted CTV more efficiently than single *A. gossypii*, but groups of several *A. gossypii* were as efficient as *T. citricida* in transmitting certain CTV isolates

(Roistacher *et al.*, 1987). Virus spread has also been correlated with aphid population densities. The number of aphids flying to orange trees varies considerably according to location (Dickson *et al.*, 1956) and year (Racah *et al.*, 1987). Although several thousand *A. gossypii* are attracted annually to a citrus tree, only low annual infection rates were recorded (Bar-Joseph *et al.*, 1989).

In controlled temperature experiments, transmission rates of 60.8% and 12.2% were obtained from Madame Vinous sweet orange kept at a constant temperature of 22 °C and 31°C respectively. Reversing conditions from 31° to 22°C resulted in a significant increase in transmission (Bar-Joseph *et al.*, 1977). The decrease in transmission from plants kept at higher temperatures was correlated with a marked decrease in virus concentration (Bar-Joseph *et al.*, 1989).

1.7 SEQUENCE VARIANTS

There is evidence that CTV isolates frequently are populations from which variants with distinct properties can be selected and that also often contain multiple defective RNAs (DRNAs). A disease phenotype may, therefore result from a mixture of viral components that exist in different proportions and co-replicate in infected plants (Hilf *et al.*, 1999).

The distribution in the virus population of genetic variants generated by mutation or genetic exchange will depend on two major evolutionary processes: genetic drift and selection. Populations of most organisms may not be large enough to ensure that each variant will have progeny in the next generation, so random effects would occur during transmission of genetic traits to new generations; this random process is called genetic drift. Selection is a directional process by which variants that are fittest in a certain environment will increase their frequency in the population (positive selection), whereas variants less fit will decrease their frequency (negative or purifying selection).

Plant viruses usually cause persistent infections in their systemic hosts, and their populations can attain very large sizes within one infected plant (Hull, 2002). A large fraction of the population will consist of mutants that can not multiply (Hull, 2002). Population bottlenecks will also occur in different moments of the life history of the virus, such as each time a new host plant is infected, a new plant species/cultivar becomes a host, or a new geographical area is colonized. This effect results in a smaller diversity within a

population and in bigger diversities between populations (Hull, 2002). Random changes in the main genotype observed in passage experiments and aphid transmissions can be explained by these effects (Hull, 2002).

Selection results in a decrease of the population diversity, and may also cause an increased diversity between populations, if under different selection pressures. A first group of selection pressures can be associated with the maintenance of structural features of the virus. For instance, amino acids that have a role in the assembly and stabilization of nucleocapsids are conserved in the coat protein (CP) (Hull, 2002). Also, selection may have a role in the maintenance of both primary, secondary, and higher-order structures in the viral genome that are important for replication. A second group of selection factors may be associated with the host plant (Hull, 2002).

Evidence of host-associated selection also derives from the well-known phenomenon of overcoming resistance genes (Hull, 2002). Virulence, defined as the effect of a pathogen in decreasing the fitness of its host is an important property of pathogens that may be selected for and plays an important role in their evolution (Hull, 2002). The evolution of virulence in populations depends on trade-offs between increased virulence and decreased transmissibility and, consequently, the evolution of the virus population, depends on the population dynamics of its vectors (Hull, 2002).

It is well known that most characterized virus proteins are multifunctional, and suggest that virus proteins are under different selective constraints corresponding to various functions, and therefore are never optimized for just one of their functions. An extreme case of multiple functional constraints occurs in genomic regions with overlapping reading frames. The “focus of selection” changes according to environmental conditions (Hull, 2002).

The increase in fitness due to genetic exchange and the occasional appearance of more fit reassortants or recombinants may have an important role in virus evolution (Hull, 2002). A phenomenon mechanistically related to recombination is the generation of defective RNAs (Hull, 2002). Their effect, if any on the populations of CTV remains to be established. The strength of negative selection may be less in noncoding regions of viral genomes or in noncoding subviral replicating RNAs. Positive selection is considered to be

less frequent than negative selection and clearly acts with resistance-breaking isolates (Hull, 2002).

1.8 QUASI SPECIES & STRAIN DOMINANCE

Quasi-species are a type of population structure in which collections of closely related genomes are subjected to a continuous process of genetic variation, competition and selection. Typically, the distribution of mutants or variants is centred on one or several master sequences. The selection equilibrium is meta-stable and may collapse or change when an advantageous mutant appears in the picture. The scenario set above for quasi-species depends upon the complexity of the genetic information in the viral genome, the copy fidelity on replication of the genome and the superiority of the master sequence. Biologically quasi species could be defined by the phenotypic expression of the population, most likely dominated by that of the master sequence (Hull, 2002).

The generation of quasi species and the erratic nature of RNA transcription should have a considerable impact on the genomic variation, while cross-protection is said to lead to a higher rate of genomic conformity (Bar-Joseph *et al.*, 2002). In my mind this would depend on the interactions between mild forms and challenging severe strains, recombination events, reversion and possible mutations due to selective pressures if the cross-protective strains are to have a good genomic conformity.

Relative to cross-protection, there has been very little work done on strain dominance and the effects thereof. In my mind there are a few possible scenarios which should be considered. These are however just possibilities and must by no means be taken as fact. (1) It is still not known if severe strains tend to dominate the population of strains in an infected plant showing severe symptoms. There has however been a study to validate this theory, where plants in which the SSCP profile of a severe isolate became predominant showed stem-pitting and those where the mild SSCP profile corresponded to the mild isolate remained symptomless (Sambade *et al.*, 2002). (2) It could well be that mild strains dominate over severe strains in certain plants and that is why the symptoms appear mild, but under other conditions other cultivar's symptoms appear severe when the dynamic of dominance and/or host resistance changes. (3) Or, dominance of a particular strain might mean nothing. A severe strain will show symptoms no matter how insignificant the titer of

it is. And finally (4) it must also be considered that certain strains encoding RNA silencing suppressors could potentially increase in number to a undefined threshold where the plant can no longer protect itself against infection. This therefore might have nothing to do with the domination of certain strain types.

It could also be possible that certain sequence types, irrespective of symptom severity, could be better competitors and multiply at much higher rates and therefore indirectly dominate.

1.9 SEQUENCE DIVERGENCE

Genome comparison of the VT and T36 strains showed that they were highly conserved throughout the 3' terminal 8 400 nucleotides, but showed a gradual increase in sequence divergence toward the 5' terminus, changing from 97% sequence identity in the 3' untranslated region (UTR) to 47% sequence identity in the 5' UTR. The nucleotide sequence differences in the 5' region (30-40%) were greater than expected for strains of the same virus (<10% sequence divergence).

These findings raised many questions about CTV populations including: (i) genome asymmetry; (ii) existence of subgroups of CTV isolates; (iii) possibility of chimera strains from a typical CTV strain and a significantly different closterovirus; and (iv) possible ancestral CTV genotype evolution in an unequal or asymmetrical rate of divergence among progeny.

Hilf *et al.* (1999) demonstrated that comparison of the T3, T30 and VT strains showed a symmetrical distribution of nucleotide sequence identity in both 5' and 3' regions. However when T3, T30 and T36 were compared it showed a dramatic decrease in sequence identity in the 5' 11 kb region (Hilf *et al.*, 1999). It was concluded that the 5' region can serve as a measure of the extent of sequence divergence and can be used to define new groups and group members in the CTV complex (Hilf *et al.*, 1999). López *et al* (1998) also found similar results and reported that extensive variability exists in the 5' most 267 nucleotides.

Recombination, the exchange of DNA/RNA between genomes, plays an important role in evolution of living organisms. Previous research has shown evidence for possible CTV recombination with the finding that CTV-infected cells often contain defective RNAs, generated by non-homologous recombination (Ayllón *et al.*, 1999). Other evidence from incongruence of phylogenetic relationships in different genomic regions and full genome sequences having 90% nucleotide identity in both terminal regions, but over 99% identity in an approximately 6-kb central region suggesting homologous recombination (Vives *et al.*, 1999).

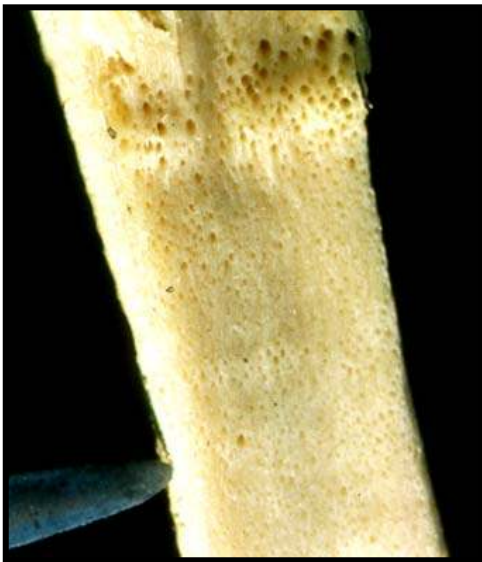
Multiple recombinations between two RNA sequence variants have been shown by Vives *et al.* (2005). It was shown that sequence variants of a natural CTV isolate revealed that its population was composed of three sequence types: (1) the most frequent type had $\geq 97.9\%$ nucleotide identity with the sequence predominant in severe CTV isolates from different origins; (2) a second variant, genetically close to the major component of several mild isolates had $\leq 85\%$ identity with the first; and (3) several variants (less than 4%) resulted from homologous recombination at one or more sites between sequences (1) and (2) (Vives *et al.*, 2005). Recombination sites had an AU-rich stretch of 8-89 nucleotides shared by both parental sequences, flanked by GC and AU-rich regions upstream and downstream respectively, a previously suggested hot-spot for homologous recombination (Vives *et al.*, 2005).

Extensive research has been reported on CTV recombination (Ayllón *et al.*, 1999, Mawassi *et al.*, 1995, Rubio *et al.*, 2001). It was suggested that homologous and nonhomologous recombination may be a frequent phenomenon in CTV and with RNA viruses; recombination has evolved as a way to regenerate functional genomes from others with deleterious mutations accumulated as a result of a lack of proofreading activity of RNA-dependent RNA polymerases (García-Arenal *et al.*, 2001). The main function of recombination in CTV (which has a large RNA genome) can be to act as a compensatory mechanism to offset accumulation of deleterious mutations. An additional benefit of recombination would be an increased genetic diversity and adaptability. Caution must be taken to avoid the introduction of exotic CTV isolates that might recombine and give rise to CTV isolates with new biological properties (Rubio *et al.*, 2001).

1.11 CROSS-PROTECTION

There are three economically devastating field symptoms caused by CTV:

- **Decline of trees:** Within months the canopy of a mature tree suddenly wilts and dies.
- **Stem pitting (SP):** Trees affected with CTV stem pitting strains decline, do not senesce, but have reduced fruit production (Fig 3A) and quality (Garnsey and Lee, 1988). There are varying degrees of mild to severe pitting of stems. Trees are typically stunted; have a bushy appearance; the leaves are chlorotic; the twigs are brittle and break easily (Garnsey and Lee, 1988).
- **Seedling Yellows (SY):** which causes losses in tree nurseries and is nicknamed the “glasshouse disease” and not a field symptom. The symptoms are leaf chlorosis (Fig 3B) and stunting of sour orange, grapefruit or lemon seedlings (Fraser, 1952).



Figures 3: (A) Severe stem pitting on a citrus tree trunk; (B) Seedling Yellows symptoms on a young Mexican lime seedling (Photographs courtesy of C. Roistacher).

Cross-protection is defined as the infection of a plant with a strain/s of a virus causing only mild disease symptoms (also known as the mild, attenuated, protecting, hypovirulent or avirulent strain) which may protect it from infection with severe strains (challenging strain). Mild protecting strains are selected from naturally occurring variants; from random mutagenesis or from site directed mutagenesis of severe strains. Plants might be purposely infected with a mild strain as a protective measure against severe disease. This was first reported by McKinney (1929), who observed that tobacco plants systemically

infected by a “green” strain of Tobacco Mosaic Virus (TMV) were protected from infection by another strain that induced yellow mosaic symptoms.

The selection of mild strains includes: (i) collecting budwood from symptomless field trees of susceptible varieties or from milder reacting isolates on indicator plants (Costa *et al.*, 1980); (ii) selecting mild isolates following heat treatment and vector transmission from plants that have recovered from SY infection; and by passage of SY and SP CTV isolates through *Passiflora spp* (Davino *et al.*, 1986). Experiments are performed to see the symptoms by inoculating the isolate onto different cultivar combinations and testing the isolate in the field and in the greenhouse. Certain parameters like stem-pitting, seedling yellows, growth, and virus titer are monitored by visual symptoms and ELISA tests. Sometimes these mild isolates are challenged with severe isolates available in CTV collections.

These trials could take years before isolates are cleared as suitable in the cross-protection scheme. Molecular tests to determine the general strain type would be very useful to the industry; together with greenhouse trials would ensure a much more solid programme. Certain strains do not display severe symptoms but later when strain, environmental or host factors change suddenly do. These strains might only be detected in the future through molecular based testing. It is however equally important that greenhouse trials are performed to see the effects of strain mixtures, cultivar variation, temperature and environmental differences etc, which can not be quantified with molecular tests. Possibly the best strategy would be to perform trials to establish possible mild strains biologically and then progress to testing only those isolates molecularly to verify their designation as mild. Possibly a genetic database could be used to genetically identify potential mild material without indexing.

Identified mild strains should have certain properties before it can be used for cross-protection as described by Hull (2002). For example:

They should induce milder symptoms in all cultivated hosts than any other isolates encountered; not alter the marketable properties of the crop products; give fully systemic infections and invade most, if not all tissues; be genetically stable and not give rise to severe forms; not be easily disseminated by vectors to limit unintentional spread to other crops; provide protection against the widest range of strains of the challenging virus. The

protective inoculum should be easy and inexpensive to produce, easily checked for purity, and accessible to farmers (Hull, 2002).

According to Roger Hull (2002) cross-protection could be worthwhile as a measure in very difficult situations; it is not to be recommended as a general practice, for the following reasons:

1. So-called mild strains often reduce yield by about 5-10%.
2. The infected crop may act as a reservoir of virus from which other more sensitive species or varieties can become infected.
3. The dominant strain of virus may change to a more severe type in some plants.
4. Serious disease may result from mixed infection when an unrelated virus is introduced.
5. The genome of the mild strain may recombine with that of another virus, leading to the production of a new virus.

CTV provides the most successful example for the use of cross-protection. In the 1920's after a devastating epidemic in Brazil, the successful application of cross-protection by inoculation with mild CTV isolates was employed and has been detailed by Costa and Muller (1980).

It was found that 75 % of the mild-strain protected trees had severe symptoms compared with about 85 % of the unprotected trees in Florida (Powell *et al.*, 2003). The use of the same isolates gave better protection of Ruby Red grapefruit on sour orange rootstock. Thus there are differences in the responses of the scion-rootstock combination, but it is also important to have a compatible mild strain. As mentioned before field symptoms differ compared to glasshouse ones and therefore glasshouse selected cross-protection isolates may differ in their performance in field conditions (van Vuuren *et al.*, 1993). It is therefore critically important to have molecular tests in place to make sure that severe strains are not present amongst these cross-protecting isolates. The search for improved strains is an ongoing process and countries are constantly employing the method.

It is still unknown what mechanism/s is responsible for cross-protection. It has been postulated that the mechanism is similar to that of RNA silencing. It is most likely that mild strain protection operates by the mild strain “priming” this defence system so that it

operates against the superinfecting severe strain (Matthews, 1991). With this in mind, it is important that the ‘priming’ sequence of the protecting mild strain is similar to that of the superinfecting severe strain. An analysis of the sequences of CTV isolates from different sites and collected at different times showed that the mild strains used in Florida and Spain were very similar to a diverse range of isolates (Albiach-Martí *et al.*, 2000b). However it is possible that in some instance of mild strain cross-protection, other mechanisms, such as competition for replication sites, operate (Matthews, 1991).

1.12 CROSS-PROTECTION SCHEME IN RSA

There are generally two different kinds of criteria available for the discrimination of virus strains:

1. Structural criteria based on the properties of the virus particle itself and its components.
2. Biological criteria based on various interactions between the virus, its host plant and its vectors (Hull, 2002).

In this study the term “strain” is defined as an isolate with a distinct symptom which relates to a mild or severe form. Traditionally the term “strain” has only been used with isolates that induce a specific symptom such as stem-pitting, seedling yellows, decline (classified as severe), or vein clearing (classified as mild). Since pathogenicity determinants have not been found in CTV, it is not possible to relate pathogenicity to a specific gene/s region. An “isolate” is defined as any plant/sample infected with CTV.

Within the Southern African Citrus Improvement Programme (SACIP) plants are cross-protected with a mixture of mild strains to mainly protect against severe stem-pitting strains. There are two mild isolates (comprised of multiple strains) used to cross-protect grapefruit, GFMS (Grapefruit Mild Strains) 12 and GFMS 35 (van Vuuren *et al.*, 1993). Isolate LMS 6 (lime mild strains) (van Vuuren *et al.*, 1993) contains a mild form of seedling yellows and is applied to cross-protect sweet orange, mandarin and lime trees. Results from field experiments showed that these isolates afforded good protection for several years in Marsh grapefruit and lime (van der Vyver *et al.*, 2002, van Vuuren *et al.*, 1993).

The grapefruit cultivars, Nel Ruby cross-protected with GFMS 12 and Star Ruby with GFMS 35 were afforded good protection (van Vuuren *et al.*, 2000). However certain cultivar/isolate combinations including GFMS 12 with Star Ruby, GFMS 35 with Nel Ruby

and LMS 6 with Marsh grapefruit all showed a high proportion of small fruit (van Vuuren *et al.*, 2000). Since virus free control plants also became naturally infected in the field but did not show small fruit it was thought that the problem could lay with strain shifts in certain grapefruit cultivar/mild CTV isolate combinations (van Vuuren *et al.*, 2000). From these findings it becomes clear that the host together with certain CTV strains interact in very specific ways to determine disease severity.

The mixture of strains in GFMS 12 was partially characterized by analysis of sub-isolates (12/1 – 12/9) obtained by single aphid transmissions. Two sub-isolates of GFMS 12 (12-2 and 12-5) were less virulent in Mexican lime and grapefruit hosts (van Vuuren *et al.*, 2000). Sub-isolate 12-3 which was more virulent than the original isolate and induced severe stem-pitting (van Vuuren *et al.*, 2000). It was hypothesized that this sub-isolate could become dominant under specific environmental conditions (van Vuuren *et al.*, 2000). However in South Africa it has been said that without CTV cross-protection, grapefruit production would be uneconomic (van Vuuren *et al.*, 1993).

1.13 CROSS-PROTECTION BREAKDOWN

There are factors such as different cultivars, strain separation, environment, multiple strain effects and super-infection by aphids etc. which affect the composition and/or the balance of strains within a host (Albaich-Martí *et al.*, 1996).

It was found that the same population of variants propagated on different host cultivars produced a different SSCP pattern profile of a particular gene indicating host selection of the predominant sequence variants in the population (Rubio *et al.*, 2000).

When plants are inoculated with complex isolates, strain separation can readily occur during systemic invasion (Moreno *et al.*, 1991). Hosts can influence the CTV strain balance as shown by passage through grapefruit, smooth Seville orange and Mexican lime (Moreno *et al.*, 1991). Permanent and non-permanent separation of CTV strains in an isolate after host passage or by stem-slash inoculation was also confirmed (Rubio *et al.*, 2000). Differences in SSCP patterns were found in different sectors of individual plants strongly suggesting uneven distribution of the CTV strains within the tree, possibly due to aphid introductions (van der Vyver *et al.*, 2002). This could aid breakdown by leaving certain parts of the plant susceptible to severe strains.

It was found that possible cross-protection breakdown occurred when the BCA (brown citrus aphid) was introduced into Florida (Powell *et al.*, 2003). Experimental areas with trees cross-protected with mild isolates showed detectable infection with decline inducing isolates of 13 % compared to unprotected trees of 67% in a 16-year period prior to BCA introduction (Powell *et al.*, 2003). However five years after BCA introduction, decline infections were 57, 81 and 71 % for trees protected with the three mild isolates respectively compared with 95% in unprotected trees (Powell *et al.*, 2003). This suggests that possibly the introduction of BCA accelerated the breakdown of cross-protection. How the BCA accelerated breakdown is not understood, nor are possible factors behind the breakdown. But since it is perceived that BCA selects for severe strains this might help explain the breakdown as being a bombardment of severe strains (Sharma *et al.*, 1989) or other more dominant strains could have changed the dynamics of the mild strains and how they interact.

In 2002 a study examined changes in the pre-immunized grapefruit trees in South Africa (van der Vyver *et al.*, 2002). Certain grapefruit cultivars pre-immunized with South African CTV isolates GFMS 12 and GFMS 35 exhibited changes in the level of protection by producing high percentages of small fruit (van der Vyver *et al.*, 2002). These isolates were biologically evaluated on CTV sensitive plants and examined by SSCP analysis (van der Vyver *et al.*, 2002). The biological data indicated that the cross-protecting isolates had not retained their original status since seedling yellows (CTV-SY) as well as severe stem-pitting was recorded (van der Vyver *et al.*, 2002). Other strains were also found in infected trees which were not part of the original pre-immunized strains (van der Vyver *et al.*, 2002). It was postulated that super-infection occurred by other CTV strains introduced by the BCA (van der Vyver *et al.*, 2002). There was no evidence in the SSCP patterns that segregation of strains within the original pre-immunizing isolates occurred. The SSCP patterns of the additional severe strains introduced did not correspond to the SSCP profiles of the strains within the cross-protecting isolates. It was concluded that changes occurred in the viral RNA populations within trees but did not necessarily indicate cross-protection failure (van der Vyver *et al.*, 2002). Failure or no failure, the point is still that cross-protection had broken down. In this case it is possible that changes in symptom expression were due to other strains interacting differently or because of dominance of a particular external strain.

It has been noted that an isolate which is a mixture of strains can harbor a severe strain which does not necessarily express any severe symptoms on a particular plant except when the severe strain is present alone (van Vuuren *et al.*, 2000). This was discovered during trials for cross-protecting strains. A shift in strain dominance within a CTV population can occur due to the influence of the host and/or environmental conditions, resulting in differential effects on a host. Therefore in certain instances a severe strain present in a mild isolate may become dominant with disastrous consequences. This has the potential to occur in South Africa with its dramatic variation of climatic conditions between different citrus producing areas.

It has been suggested that dominance of non-transmissible isolates by competition between virus strains, prevents the early establishment of the more transmissible mutants, and that later climatic and aging factors cause the breakdown of protection, and enable the establishment of the new, more highly transmissible mutants.

1.14 ENVIRONMENTAL PRESSURE

Factors such as temperature also play a role in altering the isolate composition or strain prevalence within and between trees of a certain cultivar (Broadbent *et al.*, 1991). A CTV isolate that displays mild symptoms when biologically indexed, may contain severe strains that can be expressed under certain conditions. This is a very important note to consider especially when evaluating strains in a greenhouse, as their performance might differ in the field. Differences in symptom expression of daughter trees, derived from the same mother source, planted in different climatic regions were previously reported (Broadbent *et al.*, 1991).

1.15 HOST-VIRUS INTERACTION

The symptom expression and susceptibility of different citrus species, hybrids and citrus relatives to CTV infection varies considerably. Tolerance and susceptibility are complicated by the different behavior of various CTV isolates on citrus (Moreno *et al.*, 1993). Many citrus species are CTV-tolerant e.g. *Citrus sinensis* (Sweet Orange), *C. paradisi* (grapefruits) and *Poncirus trifoliata* (Garnsey *et al.*, 1987a). CTV particles are present in the phloem of such plants even though it does not support detectable multiplication of the

virus and there are no symptoms caused by the virus. A pair of recessive genes has been linked to CTV susceptibility in trifoliate crosses (Yoshida, 1995).

Many citrus species are CTV-tolerant on their own or when grafted onto tolerant rootstocks. Tolerance varies according to CTV strain and varieties tolerant to some isolates may react to others (Bar-Joseph *et al.*, 1989).

CTV causes damaging effects in other cultivars. Seedlings of sour orange, lemon, grapefruit, and certain other varieties become severely chlorotic and stunted with SY isolates while sweet oranges, mandarins remain symptomless (Bar-Joseph *et al.*, 1989).

Another aspect to note when illustrating host-virus interactions is to take into consideration the state of susceptibility due to the individual plant's current health status and certain stress factors being exerted on the plant. This will account for possible differences between plants in a greenhouse and ones in a field as well as between plants in similar environmental conditions of the same cultivar. Also in greenhouse experiments researchers often use seedlings, which can add a factor of host genetic variability.

CTV isolates vary in the symptoms induced in a standard set of citrus cultivar combinations under glasshouse conditions. Hosts can influence CTV strain balance as shown by passage through grapefruit, Seville orange and Mexican lime (Jarupat *et al.*, 1988; Moreno *et al.*, 1993), and this separation could be permanent or non-permanent (Moreno *et al.*, 1993).

Pigmented grapefruits are more sensitive to stem-pitting symptoms than non-pigmented grapefruit (Marais *et al.*, 1996) and grapefruit also has been shown to influence the strain composition of CTV isolates (van Vuuren *et al.*, 2000).

Host-virus interactions between citrus and CTV are very complicated and I my aim here is to only illustrate a few examples of how the plant itself influences the effect of CTV. It is however very important to realize that the plant has a big role to play in how severe a particular strain could be and methods to differentiate strains must not exclude the different virus-host interactions.

1.16 CONTROL OF STRAIN MOVEMENT

Selective certification of budwood sources requires strain-specific detection methods. Presently, indexing by graft-inoculation directly to susceptible cultivars is the most reliable method, but regular testing of many budwood source trees for stem pitting by graft inoculation to indicators is expensive and takes time (Hadidi *et al.*, 2004). The monoclonal antibody MCA-13 does not differentiate most stem-pitting and decline inducing isolates of CTV, but is used as a presumptive screening tool to eliminate trees infected with stem pitting and decline from the certification programme in Florida. CTV programmes aimed at eradicating infected trees do so without knowledge of the severity of the strain they contain. Attempts have been made in the past in Israel to control a rapid decline inducing strain through continued removal and top grafting to varieties partly resistant to infection (Bar-Joseph *et al.*, 1989). As strain selective probes are developed for CTV, it should be easier to eliminate harmful strains and prevent exotic strains from entering RSA. This would reduce the cost of widespread tree removal and eliminate grower objections to removal of trees which pose little harm and may contain mild isolates capable of providing cross-protection.

Future work on controlling CTV includes the expression of the coat protein and/or other resistance genes in transgenic plants. This holds a lot of promise for CTV control, but currently this form of control is not being implemented and it might be a few years still until it is. It therefore remains a priority to make sure the cross-protection scheme in RSA is successful, by selecting and testing more superior mild strains and eliminating breakdown.

1.17 REFERENCES

- Albiach-Martí, M.R., Guerri, J., Cambra, J., Garnsey, S. M., Moreno, P. 2000a. Differentiation of *Citrus tristeza virus* isolates by serological analysis of p25 coat protein peptide maps. *J. of Virol Meth.* 88: 25-34.
- Albiach-Martí, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., López, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., Dawson, W.O. 2000b. Sequences of *Citrus tristeza virus* separated in time and space are essentially identical. *J. Virol.* 74: 6856–6865.
- Ayllón, M.A., López, C., Navas-Castillo, J., Mawassi, M., Dawson, W.O., Guerri, J., Flores, R., Moreno, P. 1999. New defective RNAs from *Citrus tristeza virus*: Evidence for a replicase-driven template switching mechanism in their generation. *J. Gen. Virology.* 80: 817-821.
- Ayllón, M.A., López, C., Navas-Castillo, J., Garnsey, S.M., Guerri, J., Flores, R., Moreno, P. 2001. Polymorphism of the 5' terminal region of *Citrus tristeza virus* (CTV) RNA: Incidence of three sequence types in isolates of different origin and pathogenicity. *Arch. Virol.* 146: 27-40.
- Balaraman, K., Ramakrishnan, R. 1979. Transmission studies with strains of tristeza virus on acid lime. *Z. Pflanzenkr. Pflanzen.* 86: 653-661.
- Balaraman, K., Ramakrishnan, R. 1980. Strain variation and cross protection in *Citrus tristeza virus* on acid lime. In “Proceedings of the 8th Conf. of the IOCV.” (S. M. Garnsey, L.W. Timmer, J.A. Dodds, Eds), pp. 60-75, IOCV, Riverside, California.
- Bar-Joseph, M., Loebenstein, G. 1973. Effects of strain, source plant and temperature on the transmissibility of *Citrus tristeza virus* by the melon aphid. *Phytopathology* 63: 7 16-20.
- Bar-Joseph, M., Raccach, B., Loebenstein, G. 1977. Evaluation of the main variables that affect *Citrus tristeza virus* transmission by aphids. *Proc. Int. Soc. Citric.* 3: 958-61, Univ. Fla., Gainesville, Florida Lake.

- Bar-Joseph, M. 1978. Cross-protection incompleteness: A possible cause for natural spread of *Citrus tristeza virus* after a prolonged lag period in Israel. *Phytopathology* 68: 1110-11.
- Bar-Joseph, M., Garnsey, S. M., Gonsalves, D. 1979a. The Closteroviruses: A distinct group of elongated plant viruses. *Adv. Virus Res.* 25:93-168.
- Bar-Joseph, M., Garnsey, S.M., Gonsalves, D., Moscovitez, M., Purcifull, D.E., Clark, M.E., Loebenstein, D. 1979b. The use of enzyme-linked immunosorbent assay for the detection of *Citrus tristeza virus*. *Phytopathology* 69: 190-194.
- Bar-Joseph, M., Marcus, R., R.F. Lee. 1989. The Continuous Challenge of *Citrus tristeza virus* Control. *Ann. Rev. Phytopathol.* 27: 291-316.
- Bar-Joseph, M., Che, X., Mawassi, M., Gowda, S., Satyanarayana, T., Maria, A., Ayllón, M.A., Albiach-Martí, M.R., Garnsey, S.M., Dawson, W.O. 2002. The continuous challenge of *Citrus tristeza virus* Molecular Research. In “Proceedings of the 15th Conf. of the IOCV.” (N. Duran-Vila, R.G. Milne, J.V. da Graca, Eds), pp. 1-7, IOCV, Riverside, California.
- Broadbent, P., Bevington, K.B., Coote, B.G. 1991. Control of stem-pitting of grapefruit in Australia by mild strain protection. In “Proceedings of the 11th Conf. of the IOCV.” (R.H. Brlansky, R.F. Lee, L.W. Timmer, Eds), pp. 64-70, IOCV, Riverside, California.
- Brlansky, R. H., Garnsey, S. M., Lee, R. F., Purcifull, D. E. 1984. Application of *Citrus tristeza virus* antisera in labelled antibody, Immunoelectron microscopical, and sodium dodecyl sulphate immunodiffusion tests. In “Proceedings of the 9th Conference of the International Organization of Citrus Virologists.” (S.M. Garnsey, L.W. Timmer, J. A. Dodds, Eds), pp. 337-342, IOCV, Riverside, California.
- Cambra, M., Camarasa, E., Gorris, M. T., Garnsey S.M., Gumpf, D. J., Tsai, M.C. 1993. Epitope diversity of *Citrus tristeza virus* (CTV) isolates in Spain. In “Proceedings of the 12th Conference of the International Organization of Citrus Virologists.” (P. Moreno, J V da Graca, L.W. Timmer, Eds), pp. 33-38, IOCV, Riverside, California.

Cambra, M., Olmos, A., Gorris, M.T., Marroquin, C., Esteban, O., Garnsey, S.M., Llauger, R., Batista, L., Peña, I., Hermoso de Mendoza, A. 2000. Detection of *Citrus tristeza virus* by print capture and squash capture-PCR in plant tissues and single aphids. In “Proceedings of the 14th Conf. of the IOCV.” (J. V. da Graca, R. F. Lee, R.K. Yokomi, Eds), pp. 42-49, IOCV, Riverside, California.

Celino, C. S., Panaligan, D.R., Molino, U.V. 1966. Studies on insect transmission of the tristeza virus in Philippines. *Philipp. J. Plant Ind.* 31: 89-93.

Cevik, B. 1995. Molecular differentiation of strains of *Citrus tristeza virus* using the coat protein gene sequences. M.Sc. thesis. University of Florida, Department of Plant Pathology, Gainesville.

Cevik, B., Pappu, S. S., Pappu, H. R., Bensch, D., Irey, M., Lee, R.F., Niblett, C. L. 1996. Application of bi-directional PCR to *Citrus tristeza virus*: Detection and strain differentiation. In “Proceedings of the 13th Conf. of the IOCV” (P. Moreno, J V da Graca and LW Timmer, Eds), pp. 17-24. IOCV. Riverside, CA.

Costa, A. S., Grant, T. J. 1951. Studies on transmission of the tristeza virus by the vector *Aphis citricidus*. *Phytopathology* 41: 105-13.

Costa, A. S., Muller, G. W. 1980. Tristeza controlled by cross protection, a U.S. – Brazil cooperative success. *Plant Dis.* 64: 538-41.

Davino, M., Catara, A. 1986. La tristeza degli agrumi. *Inf. Fitopatologico* 36: 9-18.

Deyong, Z., Willingmann, P., Heinze, C., Adam, G., Pfunder, M., Frey, B., Frey, J.E. 2005. Differentiation of cucumber mosaic virus isolates by hybridisation to oligonucleotides in a microarray format. *J. of Virol. Meth.* 123: 102-108.

Dickson, R. C., Flock, R. A., Laird, E. F. 1956. Citrus aphids and spread of tristeza. *Calif. Citrogr.* 41: 324-29.

- Dickson, R. C., Flock, R. A., Johnson, M. M. 1951. Insect transmission of citrus quick decline virus. *J. Econ. Entomol.* 44: 172-76.
- Dodds, J. A., Bar-Joseph, M. 1983. Double-stranded RNA from plants infected with closteroviruses. *Phytopathology* 73: 419-423.
- Dodds, J. A., Jarupat, T., Lee, J.G., Roistacher, C.N. 1987a. Effects of Strain, Host, Time of Harvest, and Virus Concentration on double-stranded RNA Analysis of *Citrus tristeza virus*. *Phytopathology* 77: 442-447.
- Dodds, J. A., Tamaki, S.J., Roistacher, C.N. 1984. Indexing of *Citrus tristeza virus* double-stranded RNA in field trees. . In “Proceedings of the 9th Conf. of the IOCV.” (S.M. Garnsey, L.W. Timmer, J. A. Dodds, Eds), pp. 327-329, IOCV, Riverside, California.
- Dodds, J. A., Jarupat, T., Lee, J.G., Roistacher, C.N. 1987b. Detection of strain specific double-stranded RNAs in citrus species infected with *Citrus tristeza virus*: A review. *Phytophylactica* 19: 131-137.
- Dolja, V. V., Karasev, A.V., Koonin, E.V. 1994. Molecular biology and evolution of closteroviruses. Sophicated build up of large RNA genomes. *Annu. Rev. Phytopathol.* 32: 261-285.
- Dreher, T.W. 1999. Functions of the 3'-untranslated regions of positive strand RNA viral genomes. Annual Reviews. *Phytopathology* 37: 151-174.
- Fawcett H.S. and Wallace J.M. 1946. Evidence of the virus nature of citrus quick decline. *Calif. Citrogr.* 32: 88-89.
- Febres, V.J., Pappu, H.R., Anderson, E.J., Pappu, S.S., Lee, R.F., Niblett, C.L., 1994. The diverged copy of the *Citrus tristeza virus* coat protein is expressed *in vivo*. *Virology* 201: 178–181.
- Febres, V.J. 1995. Ph.D. Thesis, pp. 104. University of Florida, Gainesville, FL.

Febres, V.J., Ashoulin, L., Mawassi, M., Frank, A., Bar-Joseph, M., Manjunath, K.L., Lee, R.F., Niblett, C.L. 1996. The p27 protein is present at one end of *Citrus tristeza virus* particles. *Phytopathology* 86: 1331–1335.

Fraser, L.R. 1952. Seedling yellows, an unreported virus disease in citrus. *Agric. Gazette. NSW* 63: 125-131.

García-Arenal, F., Fraile, A., Malpica, J.M. 2001. Variability and genetic structure of plant virus populations. Annual. Reviews. *Phytopathology* 39: 157-186.

Garnsey, S. M., Barret, H.C., Hutchison, D.J. 1987a. Identification of *Citrus tristeza virus* resistance in citrus relatives and its potential applications. *Phytophylactica* 19: 187-192.

Garnsey, S.M., Gumpf, D.J., Roistacher, C.N., Civerolo, E.L., Lee, R.F., Yokomi, R. K., Bar-Joseph, M. 1987b. Toward a standardized evaluation of the biological properties of *Citrus tristeza virus*. *Phytophylactica* 19: 151-157.

Garnsey, S.M., Permar, T.A., Cambra, M., Henderson, C.T. 1993. Direct tissue blot immunoassay for detection of *Citrus tristeza virus*. In “Proceedings of the 12th Conf. of the IOCV.” (P. Moreno, J V da Graca, L.W. Timmer, Eds), pp. 39-50, IOCV, Riverside, California.

Garsney, S.M., Timmer, L. W., Graham, J.H. 2000. Compendium of Citrus Diseases 2nd ed. APS Press. USA.

Ghorbel, R., López, C., Fagoaga, C., Moreno, P., Navarro, L., Flores, R., Peña, L. 2001. Transgenic citrus plants expressing the *Citrus tristeza virus* p23 protein exhibit viral-like symptoms. *Mol. Plant. Pathol.* 2: 27-36.

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

Gowda, S., Satyanarayana, T., Davis, C.L., Navas-Castillo, J., Albiach-Martí, M.R., Mawassi, M., Valkov, N., Bar-Joseph, M., Moreno, P., Dawson, W.O. 2000. The p20 gene product of *Citrus tristeza virus* accumulates in the amorphous inclusion bodies. *Virology* 274: 246–254.

Gowda, S., Satyanarayana, T., Ayllón, M.A., Moreno, P., Flores, R., Dawson, W.O. 2003. The conserved structures of the 5' nontranslated region of *Citrus tristeza virus* are involved in replication and virion assembly. *J. of Virol.* 317: 50-64.

Grant, T. J., Higgins, R. P. 1957. Occurrence of mixtures of tristeza virus strains in citrus. *Phytopathology* 47: 272-76.

Hadidi, A., Candresse T., 2001. Virus Detection-PCR. In: Maloy O.C., Murray, T.D. (eds). *Encyclopaedia of Plant Pathology*, pp. 1095-1100. John Wiley and Sons, Inc., New York, USA.

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

Hilf, M.E., Karasev, A.V., Albiach-Martí, M.R., Dawson, W.O., Garnsey, S.M. 1999. Two paths of sequence divergence in the *Citrus tristeza virus* complex. *Phytopathology* 89: 336-342.

Hadidi, A., Czosnek, H., Barba, M. 2004. DNA microarrays and their potential, Applications for the detection of plant viruses, viroids and phytoplasmas. *J. of Plant Path.* 86 (2): 97-104.

Hilf, M.E., Garnsey, S.M. 2000. Characterization and Classification of *Citrus tristeza virus* isolates by amplification of multiple molecular markers. In “Proceedings of the 14th Conf. of the IOCV.” (J. V. da Graca, R. F. Lee, R.K. Yokomi, Eds), pp. 18-27, IOCV, Riverside, California.

Hilf, M.E., Garnsey, S.M. 2002. *Citrus tristeza virus* in Florida: A synthesis of historical and contemporary biological, serological and Genetic data. In “Proceedings of the 15th Conf. IOCV.” (N. Duran-Vila, R.G. Milne, J. V. da Graca, Eds), pp. 13-20, IOCV, Riverside, California.

Hilf, M.E., Mavrodieva, V.A., 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.

Huang, Z., Rundell, P.A., Guan, X., Powell, C.A. 2004. Detection and Isolate Differentiation of *Citrus tristeza virus* in infected field trees based on Reverse Transcription-Polymerase Chain Reaction. *Plant Dis.* 88: 625-629.

Hughes, W., Lister, A. 1949. Lime disease in the Gold Coast. *Nature* 164: 880.

Hull, R. 2002. Matthews Plant Virology 4th edition. Academic Press. London. England.

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., Lewandowsky, D.J., Dawson, W.O. 1995. Complete sequence of the *Citrus tristeza virus* RNA genome. *Virology* 208: 511–520.

Kitajima, E. W., Silva, D.M., Oliveira, A.R., Muller, G.W., Costa, A.S., 1963. Electron microscopical investigations on tristeza. *Nature* 201: 1011-1012.

Lee, R.F., Bransky, R.H., Garnsey, S.M., Yokomi, R.K. 1987. Traits of *Citrus tristeza virus* important for mild strain cross protection of citrus: the Florida approach. *Phytophylactica* 19: 215-218.

Lee, R.F., Calvert, L. A. 1987. Polypeptide mapping of *Citrus tristeza virus* strains. *Phytophylactica* 19: 205-210.

Lee, R. F., Calvert, L. A., Nagel, J., Hubbard, J.D. 1988. *Citrus tristeza virus*: characterization of coat proteins. *Phytopathology* 78: 1221 – 1226.

Lee, R.F., Baker, P.S., Rocha-Peña, M.A. 1994. "The *Citrus tristeza virus* (CTV). An Introduction to Current Priorities, with Special Reference to the Worsening Situation in Central America and the Caribbean." International Institute of Biological Control (CAB International), Silwood Park, Ascot, Berks. UK.

López, C., Ayllón, M.A., Navas-Castillo, J., Guerri, J., Moreno, P., Flores, R. 1998. Sequence polymerisms in the 5' and 3' terminal regions of *Citrus tristeza virus* RNA. *Phytopathology* 88: 685-691.

López, C., Navas-Castillo, J., Gowda, S., Moreno, P., Flores, R. 2000. The 23-kDa protein coded by the 3'-terminal gene of *Citrus tristeza virus* is an RNA-binding protein. *Virology* 269: 462-470.

Lu, R., Folimonov, A., Li, W. X., Choi, Y.G., Shintaku, M., Falk, B.W., Dawson, W.O., Dind, S.W. 2003. *Citrus tristeza virus* genome encodes two distinct suppressors of RNA silencing. American Society of Virology Meeting, July 2003, Davis, (Abstract), W19-3.

Luttig, M., van Vuuren, S.P., van der Vyver, J. B. 2002. Differentiation of Single Aphid Sub-Isolates of two South African *Citrus tristeza virus* isolates from Grapefruit by single-Strand Conformation Polymorphism. In "Proceedings of the 15th Conf. of the IOCV." (N. Duran-Vila, R.G. Milne, J. V. da Graca, Eds), pp. 186-196, IOCV, Riverside, California.

Martinez, A. L., Wallace, J. M. 1964. Studies on transmission of the yellows components of citrus seedling yellows by *Aphis gossypii*. *Plant Dis. Repr.* 48: 131-33.

Mawassi, M., Karasev, A., Mietkiewska, E., Gafny, R., Lee, R.F., Dawson, W.O., Bar-Joseph, M. 1995. Defective RNA molecules associated with *Citrus tristeza virus*. *Virology* 208, 383-387.

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

- McClean, A. P. D. 1974. The tristeza virus complex. In “Proceedings of the 6th Conf. IOCV.” pp. 59-66, IOCV, Riverside, California.
- Medina, V., Peremyslov, V.V., Hagiwara, Y., Dolja, V.V. 1999. Subcellular localization of the HSP70-homolog encoded by Beet yellows closterovirus. *Virology* 260: 173-181.
- Michaud, J.P. 1998. A review of the Literature on Toxoptera Citricida (Kirkaldy) (Homoptera: Aphididae). *J. of Florida Entomol.* 81(1): 37-61.
- Miller, E.D., Plante, C.A., Kim, K-H., Brown, J.W., Hemenway, C.L. 1998. Stem-loop structure in the 5' region of Potato virus X genome required for plus-strand RNA accumulation. *J. of Mol. Biol.* 284: 591-608.
- Moreno, P., Guerri, J., Munoz, N. 1990. Identification of Spanish strains of CTV by analysis of dsRNA. *Phytopathology* 80: 477-482.
- Moreno, P., Guerri, J., Ballester-Olmos, J.F., Martinez, M.E. 1991. Segregation of *Citrus tristeza virus* strains evidenced by double-stranded RNA (dsRNA) analysis. In “Proceedings of the 11th Conf. of the IOCV.” (R.H. Bransky, R.F. Lee, L.W. Timmer, Eds), pp. 20-24, IOCV, Riverside, California.
- Moreno, P., Guerri, J., Ballester-Olmos, J.F., Fuertes-Polo, C., Albiach-Martí, M.R., Martinez, M.E. 1993. Separation and interference of strains from a citrus tristeza isolate evidenced by biological activity and double-stranded RNA (dsRNA) analysis. *Plant Pathol.* 42: 35-41.
- Moreira, S. 1942. Observacoes sobre a “tristeza” dos citrus ou “podridao das radículas”. *O. Biol.* 6: 269-72.
- Narváez, G., Skander, B.S., Ayllón, M.A., Rubio, L., Guerri, J., Moreno, P. 2000. A new procedure to differentiate *Citrus tristeza virus* isolates by hybridization with digoxigenin-labelled cDNA probes. *J. of Virol. Meth.* 85: 83-92.

Niblett, C.L., Genc, H., Cevik, B., Halbert, S., Brown, L., Nolasco, G., Bonacalza, B., Manjunath, K.L., Febres, V.J., Pappu, H.R., Lee, R.L. 2000. Progress on strain differentiation of *Citrus tristeza virus* and its application to the epidemiology of citrus tristeza disease. *Virus Res.* 71: 97-106.

Nolasco, G., Sequeira, Z., Bonacalza, B., Mendes, C., Torres, V., Sanchez, F., Urgoiti, B., Ponz, F., Febres, V.J., Cevik, B., Lee, R.F., Niblett, C.L. 1997. Sensitive CTV diagnosis using immunocapture, reverse transcriptional polymerase chain reaction and an exonuclease fluorescent probe assay. *Fruits* 52: 391-396.

Norman, P. A., Grant, T. J. 1956. Transmission of tristeza virus by aphids in Florida. *Proc. Fla. State Hortic. Soc.* 69:38-42.

Norman, P. A., Sutton, R. A. 1969. Efficiency of mature and immature melon aphids in transmitting tristeza virus. *J. Econ. Entomol.* 62: 1237-38.

Oberholzer, P.C.J., Matthews, I., Stiemie, S.F. 1949. The decline of grapefruit trees in South Africa. A preliminary report on so-called "stem pitting". *Sci. Bull. of the Dept of Agric. South Africa* No. 297.

Pappu, H.R., Pappu, S.S. Manjunath, K.L., Lee, R.F., 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. USA* 90: 3641-3644.

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.C., Garnsey, S.M., Dawson, W.O., Lee, R.F., Niblett, C.L. 1994a. Nucleotide sequence and organization of eight 3' open reading frames of the citrus tristeza closterovirus genome. *Virology* 199: 35-46.

Pappu, H.R., Pappu, S.S., Cambra, M., Moreno, P., Garnsey, S.M., Lee, R.F., Niblett, C.L. 1994b. Molecular characterization of a structural epitope that is largely conserved among severe isolates of a plant virus. *Proc. Natl. Acad. Sci. USA* 90: 3641-2644.

Pappu, H.R., Niblett, C.L., Lee, R.F. 1995. Application of recombinant DNA technology to plant protection: Molecular approaches to engineering virus resistance in crop plants. *World Journal MicroBiol. and Biotechm.* 11: 426-437.

Pappu, S.S., Febres, V.J., Pappu, H.R., Lee, R.F., Niblett, C.L. 1997. Characterization of the 3' proximal gene of the citrus tristeza closterovirus genome. *Virus Res.* 47: 51-57.

Powell, C.A., Pelosi R.R., Rundell, P.A., Cohen, M. 2003. Breakdown of Cross-Protection of grapefruit from decline inducing isolates of *Citrus tristeza virus* following introduction of Brown Citrus aphid. *Plant Disease* 87: 1116-1118.

Raccah, B., Loebenstein, G., Bar-Joseph, M. 1976a. Transmission of tristeza virus by the melon aphid. *Phytopathology* 66: 102-04.

Raccah, B., Loebenstein, G., Bar-Joseph, M., Oren, Y. 1976b. Transmission of Tristeza by aphids prevalent on Citrus and Operation of the Tristeza Suppression Programme in Israel. In "Proceedings of the 7th Conf. of the IOCV." (E.C. Calavan, Eds), pp. 47-49, IOCV, Riverside, California.

Raccah, B., Bar-Joseph, M., Loebenstein, G. 1978. The role of aphid vectors and variation in virus isolates in the epidemiology of tristeza disease. *Plant Disease Epidemiology*, ed. P. R. Scott, A. Bainbridge, pp. 221-27. Oxford, UK.

Raccah, B., Singer, S. 1987. Incidence and vector potential of the aphids which transmit *Citrus tristeza virus* in Israel. *Phytoparasitica* 19: 173-78.

Reed, J.C., Kasschau, K.D., Prokhnovsky, A.I., Gopinath, K., Pogue, G.P., Carrington, J.C., Dolja, V.V. 2003. Suppressor of RNA silencing encoded by Beet yellows virus. *Virology* 306: 203-209.

Roistacher, C. N., Bar-Joseph, M. 1984. Transmission of tristeza and seedling yellows tristeza by small population of *Aphis gossypii*. *Plant Dis.* 68: 494-96.

- Roistacher, C. N., Bar-Joseph, M. 1987. Aphid transmission of *Citrus tristeza virus*: a review. *Phytophylactica* 19: 103-67.
- Rocha-Peña, M.A., Lee, R.F. 1991. Serological technique for detection of *Citrus tristeza virus*. *J of Virol. Methods* 34: 311-331.
- Rosner, A., Lee, R.F., Bar-Joseph, M. 1986. Differential hybridisation with cloned cDNA sequences for detecting a specific isolate of *Citrus tristeza virus*. *Phytopathology* 76: 820–824.
- Roy, A., Ramachandran, P., Brlansky, R.H. 2003. Grouping and comparison of Indian *Citrus tristeza virus* isolates based on coat protein gene sequences and restriction analysis patterns. *Arch. Virol.* 148: 707-722.
- Rubio, L., Ayllón, M. A., Guerri, J., Pappu, H. R., Niblett, C. L., 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.
- Rubio., L., Guerri, J., Moreno, P. 2000. Characterization of *Citrus tristeza virus* isolates by single strand conformation polymorphism analysis of DNA complementary to their RNA population. In “Proceedings of the 14th Conf. of the IOCV.” (J. V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp. 12-17, IOCV, Riverside, California.
- Rubio, L., Ayllón, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., Falk, B. 2001. Genetic variation of *Citrus tristeza virus* isolates from California and Spain : Evidence for mixed infections and recombination. *J. of Virol.* 75: 8054-8062.
- Ruiz-Ruiz, S., Moreno, P., Guerri, J., Ambros, S. 2006. The complete nucleotide sequence of a severe stem pitting isolate of *Citrus tristeza virus* from Spain: comparison with isolates from different origins. *Arch. of Virol.* 151: 387-398.
- Sambade, A., Rubio, L., Garnsey, S.M., Costa, N., Müller, G.W., Peyrou, M., Guerri, J., Moreno, P. 2002. Comparison of viral RNA populations of pathogenically distinct isolates

of *Citrus tristeza virus*: application to monitoring cross-protection. *J. of Plant Pathol.* 51: 257-265.

Sambade, A., López, C., Rubio, L., Flores, R., Guerri, J., Moreno, P. 2003. Polymorphism of a specific region in gene p23 of *Citrus tristeza virus* allows discrimination between mild and severe isolates. *Arch. Virol.* 148: 2325-2340.

Satyanarayana, T., Gowda, S., Ayllón, M.A., Albiach-Martí, M.R., Rabindran, S., Dawson, W.O. 2002. The p23 protein of *Citrus tristeza virus* controls asymmetrical RNA accumulation. *J. of Virol.* 76: 473-483.

Sekiya, M. E., Lawrence, S. D., McCaffery, M., Cline, K. 1991. Molecular cloning and nucleotide sequencing of the coat protein gene of *Citrus tristeza virus*. *J. Gen. Virol.* 72: 1013-1020.

Sharma, S.R. 1989. Factors affecting vector transmission of *Citrus tristeza virus* in South Africa. *Zentralblatt für Mikrobiologie* 144: 283-294.

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

Symes, C.B. 1924. Notes on the black citrus aphid. *Rhodesia Agric Journal.* 11: 612-626.

Van der Vyver, J., van Vuuren, S.P., Luttig, M., da Graca, J.V. 2002. Changes in the *Citrus tristeza virus* status of the pre-immunized grapefruit field trees. In "Proceedings of the 15th Conf. IOCV, Riverside, California.

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

Van Vuuren, S.P., van der Vyver, J. B., Luttig, M. 2000. Diversity among sub-isolates of cross-protecting *Citrus tristeza virus* isolates in South Africa. In "Proceedings of the 14th Conf. of the IOCV." (J.V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp. 103-109, IOCV, Riverside, California.

Vela, C., Cambra, M., Cortes, E., Moreno, P., Miquet, J., Perez de San Roman, C., Sanz, A., 1986. Production and characterization of monoclonal antibodies specific for *Citrus tristeza virus* and their use for diagnosis. *J. Gen. Virol.* 67: 91-9.

Vives, M.C., Rubio, L., López, C., Navas-Castillo, J., Albaich-Martí, M.R., Dawson, W.O., Guerri, J., Flores, R., Moreno, P. 1999. The complete genome sequence of the major component of a mild *Citrus tristeza virus* isolate. *J.Gen.Virol.* 80: 811-816.

Vives, M.C., Rubio, L., Sambade, A., Mirkov, E., Moreno, P., Guerri, J. 2005. Evidence for multiple recombination events between 2 RNA sequence variants within a *Citrus tristeza virus* isolate. *Virology* 331: 232-237.

Xiong, Z., Barthelson, R., Weng, Z., Galbraith, D.W. 2005. Designing and testing of a *Citrus tristeza virus* (CTV) resequencing microarray. In “Proceedings of the 16th Conf. IOCV, 11-22. IOCV, Riverside, California.

Yang, G., Che, X., Gofman, R., Ben-Shalom. 1999. D-RNA molecules associated with sub-isolates of the VT strain of *Citrus tristeza virus* which induces different seedling yellows reactions. *Virus Genes* 19 (1): 5-13.

Yang, Z.N., Mathews, D.H., Dodds, J.A., Mirkov, T.E. 1999. Molecular characterization of an isolate of *Citrus tristeza virus* that causes severe symptoms in sweet orange. *Virus Genes* 19: 131–142.

Yokomi, R. K., Damsteegt, V.C. 1991. Comparison of citrus tristeza transmission efficacy by *Toxoptera citricidus* and *Aphis gossypii*. In: Peters, D.C., Webster, J.A., Choubler, C.S. Proceedings of Aphid-plant Interactions: Populations to molecules. Stillwater, Oklahoma.

Yokomi, R. K., Garnsey, S. M. 1987. Transmission of *Citrus tristeza virus* by *A. gossypii* and *A. citricola* in Florida. *Phytophylactica* 19: 169-72.

Yoshida, T. 1985. Inheritance of susceptibility of *Citrus tristeza virus* in trifoliate orange (*Poncirus trifoliate* Raf.) *Bull. Fruit Tree Res. Stn., Okitsu*, No. 12.

CHAPTER 2

**CHARACTERIZATION AND STRAIN DIFFERENTIATION
OF THE P23 GENE OF SOUTH AFRICAN CTV ISOLATES
USING A BI-DIRECTIONAL RT-PCR SYSTEM AND
PHYLOGENETIC ANALYSIS**

2.1 INTRODUCTION

The aim of this chapter was to establish a bi-directional PCR system developed by Sambade *et al.* (2003) that differentiates between mild, severe and atypical isolates based on differences at amino acid positions 78-80 of the p23 gene. To use this to molecularly differentiate 11 South African CTV infected sources and to assign each to a specific group (mild, severe or atypical). Finally to characterise the p23 gene of these sources by DNA sequencing and do phylogenetic comparisons with characterized sequences of Sambade *et al.* (2003). In this study, a bi-directional PCR is defined as having four primers: two outer conserved primers and two specific inner primers targeting two different strain sequences at one position, with the aim of amplifying two strains simultaneously and in opposite directions.

The p33, p13, p18 and p23 proteins are unique to CTV (Pappu *et al.*, 1994b; Karasev *et al.*, 1995) in comparison with other closteroviruses. ORF 11 encodes the p23 protein and is adjacent to the 3' UTR of the CTV RNA genome. It is an RNA binding protein suspected to play a regulatory role in the expression of other CTV genes (López *et al.*, 2000) and may serve as an indicator of disease severity (López *et al.*, 2000). This has significance for finding markers for strain differentiation. It has also been predicted to have a ribosome binding capacity (Dolja *et al.*, 1994) and was identified as the second CTV suppressor of post-transcriptional gene silencing (Lu *et al.*, 2003).

The p23 gene controls the asymmetrical accumulation of plus and minus RNA strands during replication (Satyanarayana *et al.*, 2002). This controlled by a “master switch”, a region including a RNA-binding and a zinc finger domain (Satyanarayana *et al.*, 2002). Its corresponding sub-genomic RNA (p23-sgRNA) is the second most abundant in infected plants (Hilf *et al.*, 1995). The p23 protein accumulates very early in cell infection (Navas-Castillo *et al.*, 1997) and down-regulates negative-stranded RNA accumulation which indirectly increases expression of the 3' genes (Ghorbel *et al.*, 2001).

The 3' half of the genome is relatively conserved amongst CTV isolates with 90 % sequence identity, compared to the 5' half with less than 70 % sequence identity when T30, VT and T385 are compared to T36 (Ayllón *et al.*, 2001). Analysis of the deduced amino acid sequence of the p23 gene from several CTV isolates of different biological properties and

geographical origins has shown a remarkable conservation of both the basic region and of the cysteine and histidine residues in the core of the proposed zinc-finger domain, suggesting their functional importance (Pappu *et al.*, 1997). A cluster dendrogram of the deduced amino acid sequences correlated with the biological properties of the isolates forming distinct groups of mild, quick decline or stem-pitting isolates (Pappu *et al.*, 1997). All mild and stem-pitting isolates formed distinctive groups (Pappu *et al.*, 1997). The T36 sequence, a quick decline strain was distinct in its sequence (Pappu *et al.*, 1997). The sequence clustering reflected the biological characteristics and suggested a possible role for p23 in determining the pathogenicity of a particular CTV isolate (Pappu *et al.*, 1997). The involvement of p23 in pathogenicity of a particular isolate may occur by p23 acting in concert with or by regulating the expression of the other disease severity determinants (Pappu *et al.*, 1997). Associated relationships between sequence data and biological properties were also found with the CP (Pappu *et al.*, 1993) and p27 (Febres *et al.*, 1995) proteins.

The damage caused by the virus is largely dependent on the isolate characteristics and on the cultivars infected. There are a few typical symptom patterns known: (1) Symptomless CTV isolates (Albertini *et al.*, 1988); (2) mild inducing vein clearing on Mexican lime; (3) decline and death of most citrus species propagated on sour orange; (4) stem-pitting, stunting, poor quality fruit and reduced yield; and (5) stunting and leaf yellowing (seedling yellows). The molecular determinants of all the symptoms are unknown. However transgenic Mexican limes over-expressing the p23 protein display symptoms identical to those caused by CTV in this host, and symptom appearance is associated with p23 accumulation (Ghorbel *et al.*, 2001).

In 2003, Sambade *et al.* reported that polymorphisms located at amino acid positions 78-80 of the 23 kDa protein allowed discrimination between mild and severe isolates. Eighteen CTV isolates of different geographic origin and pathogenicity characteristics were classified into 3 groups based on symptoms and differences at amino acid positions 78-80 of the p23 gene: 1) mild (vein clearing), 2) severe causing stem pitting and 3) an atypical group showing variable symptoms. The characteristic differences of each group's isolates were: mild Ala⁷⁸-Leu⁷⁹-Lys⁸⁰; severe Ala⁷⁸-Ser⁷⁹-Arg⁸⁰; and atypical Gly⁷⁸-Leu⁷⁹-Lys⁸⁰. These differences were incorporated into the design of the inner primers to detect each sequence type (Sambade *et al.*, 2003).

From this finding it is important to characterize the p23 gene of the South African sources and correlate the sequences of group (mild, severe and atypical) members assigned in previous work by Sambade *et al.* (2003). This will confirm if this strategy for CTV strain differentiation is appropriate for the South African Citrus Industry to track specific molecular markers in the future. It is also important in this study to use a CTV conserved gene such as p23 as a control in strain studies in further chapters in this thesis.

The mild strain cross-protecting source plants GFMS 12 (standard isolate for white grapefruit) and GFMS 35 (standard isolate for red grapefruit) have symptoms. Sub-isolates of these have been obtained by single aphid transmissions. Extensive glasshouse and field trials have been performed on these sub-isolates with different citrus selections and environmental conditions to ensure that isolates used in the cross-protection scheme exhibit good cross protecting traits such as biological activity, CTV multiplication and movement (van Vuuren *et al.*, 1993). Biological activity is monitored by visual symptoms; and multiplication and movement by ELISA tests. To understand cross-protection breakdown it is important to have molecular methods in place to monitor the specific strain markers in examining the dynamics in the field after exposure to natural infection. Such breakdown has been reported with some of the sub-isolates of GFMS 12 showing stem-pitting (van Vuuren *et al.*, 2000). It is therefore a necessity to be able to molecularly characterize these isolates before they are used in the Citrus Industry. It is also of great concern that if cross-protecting sources contain a severe component and are used in the industry, it will serve as a natural source for spreading severe strains around the country, and this technique may be used to obtain sources lacking the severe component.

2.2 MATERIALS & METHODS

2.2.1 CTV ISOLATES

A total of 11 CTV sources were used in this study and are all from single aphid transmissions performed to generate isolates with potential for mild strain cross-protection in South Africa. The single aphid transmissions produced sub-isolates (12-5, 12-7, and 12-9) that were obtained from a GFMS 12 source in Nelspruit, RSA by S.P. van Vuuren. The original source of the GFMS 12 strain mixture was a 78 year old grapefruit tree (2004) still in production referred to as Nartia A (S.P. van Vuuren., personal communication). GFMS 12 was used previously for cross-protection in all cultivars

(grapefruit, sweet oranges, mandarins) but is currently only used to protect white (Marsh, Nartia) grapefruit (*Citrus paradisi*) and pummelo (*C. grandis*) (S.P. van Vuuren., personal communication).

The three sub-isolates 390-3, 390-4 and 390-5 are derived from single aphid transmissions made from the Mouton isolate. This original isolate was mild but contained citrus viroids (personal communication, S.P. van Vuuren). Sub-isolates 389-3 and 389-4 were derived from the Nartia C isolate (GFMS 14) (S.P. van Vuuren., personal communication). The original isolates (Mouton and Nartia C) are not used as cross-protecting isolates (S.P. van Vuuren., personal communication). The sub-isolates are currently being tested in field experiments. The Mouton and Nartia C isolates were also from 78 year old grapefruit trees (2004) still in production (S.P. van Vuuren., personal communication). These five single aphid transmissions were prepared at the quarantine facility in Beltsville, USA and imported back to South Africa.

A plant infected with the T30 strain obtained from Nelspruit, RSA as well as two plants inoculated with the South African mild strain cross-protecting sources GFMS 12 and GFMS 35 were also used in this study. GFMS 35 is currently being used to cross-protect red grapefruit selections.

The clones used as positive controls (T36, T305, and T385) in this study were kindly supplied by Pedro Moreno. The T36 isolate induced variable seedling yellows reactions on sour orange or grapefruit and was assigned as an atypical group indicator control for this PCR system (Sambade *et al.*, 2003). The T305 isolate induced seedling yellows on sweet orange/sour orange and also variable stem-pitting on sweet orange and is the severe group control. The T385 isolate belongs to a collection kept at the Instituto Valenciano de Investigaciones Agrarias, Valencia, Spain and was characterized and assigned to the mild group for this PCR system (Sambade *et al.*, 2003)

2.2.2 TOTAL RNA EXTRACTION

Total RNA was extracted from Mexican lime plants infected with CTV sources using the SV Total RNA Isolation System (Promega, USA) according to the manufacturer's protocol. The total RNA was eluted with 25 µl of nuclease-free water and eluted through the column

twice. RNA samples were extracted from virus-free citrus plants to serve as a healthy citrus plant control.

2.2.3 cDNA SYNTHESIS

The reverse primer PM51 (IDT, USA) (Table 1) was used to prime cDNA synthesis and is specific for the p23 gene and is conserved in T36 (U16304), VT (U56902) and T385 (Y18420) isolates. The conserved and bi-directional PCRs were done with a two step RT-PCR (Sambade *et al.*, 2003) using the PM51 primed cDNA.

cDNA synthesis was performed as follows: The samples (3 µg total RNA extract) were heated at 65 °C for 15 minutes, 55 °C for 10 minutes and at room temperature for 5 minutes. The reaction mixture was added to a volume of 25 µl containing 50 pmol reverse primer PM51, 5U of RNAsin (Promega, USA), 10 U of *Avian myeloblastosis virus* reverse transcriptase (Roche, Germany), 1x AMV RT-Buffer (50 mM Tris-HCl, 8mM MgCl₂, 30 mM KCl, 1 mM DTT; pH 8.5) (Roche, Germany), 0.2 mM each of dATP, dCTP, dGTP, and dTTP. The contents were gently mixed then incubated at 47 °C for 1 hour on an Eppendorf Gradient Mastercycler (Eppendorf, Germany), after which 12 µl of nuclease-free water was added and the cDNA stored at -20 °C.

2.2.4 CONSERVED P23 GENE PCR

Each of the PCR reactions contained 5 µl of the cDNA reaction. A mastermix containing 100 pmoles of each of the primers PM50 and PM51 (IDT, USA) (Table 1, Figure 4); 2.5 mM MgCl₂; 1x Gotaq Buffer (Promega, USA); 2.5 U Gotaq (Promega, USA); 0.14 mM each of dATP, dCTP, dGTP and dTTP; 2 µg/µl BSA and made up to a total volume of 50 µl with Nuclease-free water. The contents were gently mixed. The thermal cycle conditions were 92°C for 2 min, 30 cycles of 92°C for 30 sec, 55°C for 45 sec, and 72°C for 1 min, then extension for 10 min at 72°C, and held at 4°C on an Eppendorf Gradient Mastercycler (Eppendorf, Germany).

cDNA from total RNA extract from a citrus virus-free plant sample was included as a healthy control. A negative control consisting of nuclease-free water instead of cDNA was included to control contamination and false positives. PCR products (5-25 µl) were analyzed by gel electrophoresis through 1.0% agarose (Whitehead Scientific) in sodium

borate electrophoresis buffer (5 mM disodium borate decahydrate, adjusted to pH 8.5 with boric acid), stained with ethidium bromide (0.5 µg/ml), and photographed under UV light.

2.2.5 BI-DIRECTIONAL PCR

The presence of each of the three types of CTV sequence variants of gene p23 was detected by a bi-directional PCR using two external primers (PM85 and PM86) and two overlapping internal primers of opposite polarity (PM82 and PM83, or PM82 and PM84) (Table 1, Figure 4). Primers PM85 and PM86 were designed by Sambade *et al.* (2003) based on two p23 regions conserved in most CTV isolates, whereas primers PM82 (VT isolate, severe group), PM83 (T385 isolate, mild group) and PM84 (T36 isolate, atypical group) were based on a short internal region which allows discrimination of CTV samples into the three different groups (mild, severe and atypical) (figure 4). All primers were synthesized by IDT (Coralville IA, USA). The reactions were performed as described above for the conserved PCR except each sample was tested using 50 pmoles of each of the primers PM85, PM86, PM82 and PM83 and again separately with primers PM85, PM86, PM82 and PM84. Virus-free samples and negative controls were used and samples were analysed as described previously. Each PCR was replicated in triplicates to ensure reproducibility.

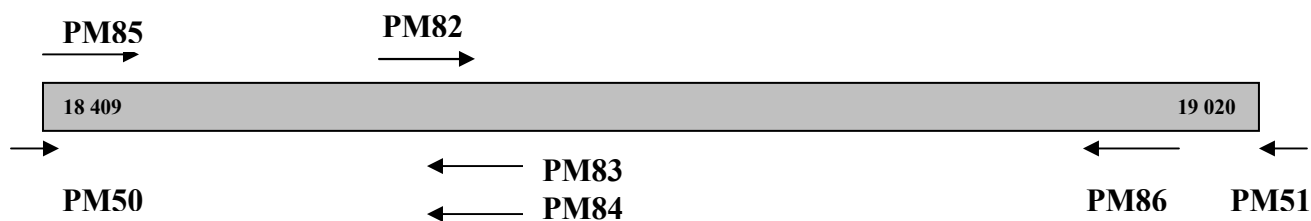


Figure 4: Schematic representation of the p23 gene with bi-directional primer sets PM50, PM51, PM85, PM86, PM82, PM83 and PM84 (not to scale).

Table 1:

Primers used for the RT-PCR amplification of sequence variants of the *p23* gene and developed by Sambade *et al.* (2003).

Target	Polarity	Primer	Primer sequence	Primer	Binding site *
CTV p23 gene (conserved area)	+	PM50	ACTAACTTTAATTCGAACA	PM 51	18347–18365
CTV p23 gene (conserved area)	-	PM 51	AACTTATTCCGTCCACTTC	PM 50	19026–19044
CTV p23 gene (conserved area)	+	PM 85	GGACAAACTTTTCGTTTCTGTGAACCTTTC	PM 86	18409–18437
CTV p23 gene (conserved area)	-	PM 86	GATGAAGTGGTGTTCACGGAGAACTC	PM85	18995–19020
CTV p23 gene (VT strain specific)	+	PM 82	AAACACGATAAGGCATCGAG	PM86	18544–18563
CTV p23 gene (T385 strain specific)	-	PM 83	CACTTACGTTTCAGTCTTGAGCG	PM 85	18588–18609
CTV p23 gene (T36 strain specific)	-	PM 84	CACTTACGTTTCAGTCTTCAAC	PM 85	18627–18647

The 697 bp sized PCR products (20 µl per well) amplified with the PM 50 and PM 51 primer set were excised from the 1% agarose gel at a minimal UV exposure time, and purified using a Wizard® SV gel purification kit (Promega, USA) according to the manufacturers instructions. In brief, 10 µl of membrane binding solution was added per 10 mg of excised gel and incubated at 50-65 °C for 10 minutes until the gel slice had melted. The DNA was bound to the silica membrane, contained within the spin basket, by centrifuging the mixture through the column at 13 000 g for 1 minute. The silica bound-DNA was washed three times with membrane wash solution, and eluted into 15 µl of nuclease free water. The purified DNA product was then quantified (1 µl) by using agarose gel electrophoresis as well as with a spectrophotometer (Nanodrop) to ensure the correct amount of template was added to the sequencing mix.

2.2.7 NUCLEOTIDE SEQUENCING

PCR Products were sequenced directly. Nucleotide sequencing was performed by using automated fluorescent sequencing. The nucleotide sequences of the selected samples were determined in both directions using the forward PM 50 or PM 85 and reverse PM 51 or PM 86 primers. The sequencing reactions were performed using ABI PRISM BigDye Primer Cycle Sequencing Kits. In short, 70 ng of purified PCR product was added to a reaction mixture containing 3.2 pmol of the primer (PM 85/86/50/51); 2 µl Big Dye Terminator Ready reaction premix v3.1 (2.5X) (PE Applied Biosystems, V3.1); 1 µl BigDye sequencing buffer (5X) (PE Applied Biosystems) and nuclease free water to a final volume of 10 µl. The reaction was performed in an Eppendorf Gradient Mastercycler (Eppendorf, Germany) with conditions: 94 °C for 1 minute, 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. The DNA was purified by the EDTA/NaOAc/EtOH method by adding 1 µl of 125 mM EDTA; 1 µl of 3M sodium acetate and 25 µl of 100 % non-denatured ethanol. The DNA products were recovered by incubating for 15 minutes at room temperature followed by centrifugation at 13 000 g for 30 minutes and washing the pellet with 100 µl of 70% ethanol. The samples were centrifuged at 13 000 g for 15 minutes and allowed to dry on the bench for 15 minutes. The completed reaction was submitted to a central University of Pretoria commercial sequencing facility for a sequencing run on an ABI 377 DNA sequencer (PE Applied Biosystems).

Phylogenetic analysis of the South African isolates was based on an alignment of the 627 bp nucleotide sequence of the complete p23 gene which included the basic region and zinc finger motif of the RNA-binding domain.

The eight South African isolates, as well as 18 reference sequences retrieved from GenBank (Table 2), were included in the analysis, in order to provide phylogenetic reference points. The reference p23 gene sequences included only isolates that were previously biologically characterised by Sambade *et al.* (2003). Other sequences available on Genbank to date for the p23 gene were not included in this study, since they were not necessarily published and their biological properties are unknown. A phylogenetic tree was constructed in order to address aspects of groupings according to mild, severe or atypical determinants of the p23 gene. The tree included sequences obtained from isolates with typical mild, severe and atypical symptoms of Sambade *et al.* (2003) and the three main group references are: T36 (atypical), T385 (mild) and T305 (severe). This analysis was conducted in order to demonstrate the relationship that exists between CTV isolates obtained from RSA, and CTV isolates obtained from different geographical regions around the world (Sambade *et al.*, 2003). This tree was further also used to demonstrate the relationship of RSA isolates to the previously established mild, severe and atypical groups of Sambade *et al.* (2003).

Ideally this tree would have been constructed with the inclusion of more local sequence data from mild strain cross-protection sources but these were the only single aphid transmissions available to be tested. There has been no study to date describing the genetic variation of the p23 gene of isolates from South Africa and consequently other sequence data from viral variants circulating in South Africa are not available for inclusion in this analysis.

Vector-NTI (Invitrogen, USA) was used to overlap the forward and reverse sequences and compile complete sequences. Sequence identity was determined based on comparison to the National Centre for Biotechnology Information (NCBI) BLAST server (<http://www.ncbi.nlm.nih.gov/>). In each case the derived sequences were trimmed using the BioEdit sequence alignment editor (v7.0, Tom Hall, Isis Pharmaceuticals, Inc 1997-2004), and alignments were carried out by using the ClustalW function, which is

incorporated into the BioEdit software (Thompson *et al.*, 1994; V7.0, Tom Hall, Isis Pharmaceuticals, Inc 1997-2004). The predicted amino acid sequences were determined by DNAMAN for all of the nucleotide sequences included in the study. The isoelectric point of amino acid residues were determined by the EMBL www Gateway to Isoelectric point service provided online (<http://www.embl-heidelberg.de/cgi/pi-wrapper.pl>).

The MEGA2 software package (Kumar *et al.*, 1993) was used to estimate nucleotide distances (distance matrix) between pairs of sequences using the Jukes and Cantor method and also for phylogenetic and bootstrap analyses using an unrooted Neighbourhood-joining tree (NJ) (Kumar *et al.*, 1993). A Neighbourhood-joining tree (NJ) was constructed using the methods of Saitou and Nei (1987). Positions with gaps were excluded from the analysis. The branching order of the tree was evaluated by using bootstrap analysis of a 1000 pseudoreplicate datasets, with a random seed generator number of 64238 (Swofford, 1993). Bootstrap values of more than 70% were generally regarded as providing evidence for a phylogenetic grouping. The graphical output for the 50% majority rule consensus trees was obtained by using the MEGA3 tree explorer (version 3.1, Kumar, Tamura and Nei 2004). Intra-group and inter-group genetic nucleotide distances were estimated with the MEGA programme.

Table 2:

The reference (Sambade *et al.*, 2003) and RSA isolates of the p23 gene used for phylogenetic analysis from different geographical origins.

Isolate	Clone number / Aphid transmission	Genbank Accession	Molecular Grouping (Sambade <i>et al.</i> , 2003)	Symptoms
T32	23-1	AJ579766	Mild	Mild to moderate
T55	23-12	AJ579764	Mild	Mild to moderate
T300	23-1	AJ579763	Mild	Mild to moderate
T312	23-10	AJ579765	Mild	Mild to moderate
T385	23-1	AJ579762	Mild	Mild to moderate
T346	23-13	AJ579768	Mild	Mild to moderate
T305	23-16	AJ579776	Severe	Seedling yellows & stem pitting ^a
Barao B	23-1	AJ579775	Severe	Seedling yellows & stem pitting ^b
C269-6	-	AY750750	Severe	Seedling yellows & stem pitting ^b
Cald-CB	23-15	AJ579778	Severe	Seedling yellows & stem pitting ^a
T388	23-18	AJ579777	Severe	Seedling yellows & stem pitting ^a
VT	23-1	AJ579773	Severe	Seedling yellows, mild stem pitting ^a
K	-	AJ579767	Atypical	Atypical – Symptomless
C270-3	-	AY750739	Atypical	Atypical - Stem-pitting ^b
Galego 50	23-1	AJ579769	Atypical	Atypical (seedling yellows)
C268-2	23-6	AY750748	Atypical	Atypical (seedling yellows/stem pits) ^b
T36	23-1	AJ579772	Atypical	Atypical – decline
12-5	12-5	-	-	Mild to moderate ^c
12-7	12-7	-	-	Mild to moderate ^c
12-9	12-9	-	-	Mild to moderate ^c
389-3	389-3	-	-	Mild to moderate ^d
389-4	389-4	-	-	Mild to moderate ^d
390-3	390-3	-	-	Mild to moderate ^d
390-4	390-4	-	-	Mild to moderate ^d
390-5	390-5	-	-	Mild to moderate ^d
GFMS 12	-	-	-	Mild to moderate ^c
GFMS 35	-	-	-	Mild to moderate ^c

^a Stem-pitting observed on Duncan Grapefruit and Sweet Orange

^b Stem-pitting observed on Duncan Grapefruit

^c Bio-indexing (van Vuuren *et al.*, 2000b) of stem-pitting on & growth on Mexican lime & Marsh Grapefruit

^d Bio-indexing (Breytenbach, personal communication) of stem-pitting & growth on Mexican lime

2.3 RESULTS

2.3.1 CLASSIFICATION OF ISOLATES BY BI-DIRECTIONAL PCR

The primer set PM50 and PM51 (Sambade *et al.*, 2003) was used as a first step screening PCR to amplify sequences conserved in VT (U56902), T385 (Y18420), and T36 (U16304) (Figure 5). All South African isolates yielded the expected 697 bp DNA fragment except isolates T30 and 390-4. Each bi-directional PCR was created using two external primers (PM 85 and PM86), and two group-specific internal primers of opposite polarity (PM82 and PM83, or PM82 and PM84) (Table 1 and Figure 4).

PCR amplification with primers PM85/PM86 and PM82/PM86 yields DNA fragments of 612 bp and 450 bp respectively, whereas amplification with primers PM85/PM83 or PM85/PM84 yield a fragment of 239 bp. Therefore, RT-PCR amplification with primers PM85, PM82, PM83 and PM86 permit detection of sequences characteristic of the mild (products of 612 and 239 bp), the severe (products of 612 and 450 bp) or both groups (products 612, 239 and 450 bp). Likewise, PCR amplification with primers PM85, PM82, PM84 and PM86 permitted detection of sequences characteristic of the severe (products of 612 and 450 bp), the atypical (products of 612 and 239 bp), or both groups (products 612, 239 and 450 bp). In both bi-directional PCRs the following were used as group controls: T36 (atypical), T305 (severe) and T385 (mild).

PCR amplification of the 11 RSA isolates using primers PCR amplification using PM85, PM82, PM84 and PM86 to differentiate between atypical and severe isolates generated (Figure 6 and Table 3): (i) DNA fragments of 612 and 239 bp with isolates 390-3, 390-5 and the T36 control of the atypical group; (ii) isolates 389-4 and T385 DNA clone had a 612 bp fragment of the mild group and; (iii) the DNA control T305 for the severe group amplified the 612 and 450 bp fragments. Isolates that did not group with the expected group classification system were: (iv) isolates GFMS 12, 12-5, 12-7, 12-9 which only amplified one of the two expected fragments of the severe group of 450 bp but not the 612 bp fragment; (v) isolate GFMS 35 only produced one of the two atypical fragments of 239 bp and not the 612 bp fragment and (vi) isolates T30, 389-3 and 390-4 did not amplify anything, only smears of undefined size. There were no mixed groupings in this PCR.

On the other hand, PM85, PM82, PM83 and PM86 to differentiate between mild and severe isolates (Figure 7 and Table 3) yielded some interesting results: (i) isolates GFMS 35, 389-4, and the mild group control T385 clone yielded two DNA fragments of 612 and 239 bp; (ii) only the DNA control clone T305 of the severe group yielded two DNA fragments of 612 and 450 bp and; (iii) Isolates 390-3, 390-5 and the atypical group control clone T36 of the atypical group yielded only a 612 bp fragment. Isolates that did not group with any group classification system were: (iv) isolates GFMS12, 390-4 and 389-3 which only amplified one of the two mild group fragments of 230 bp and not the 612 bp fragment; (v) isolates 12-7 and 12-9 amplified both the 450 and 239 bp fragments of the mild and severe groups but not the 612 bp fragment; (vi) isolate 12-5 amplified one of the two expected fragments of the severe group of 450 bp but not the 612 bp fragment; and (vii) isolate T30 did not amplify anything. The virus free plant and the buffer control showed no amplification.

In both bi-directional PCRs the virus free plant and the buffer control showed no amplification. The gel photographs depicting these amplifications and the summarized representation of these results are shown in figures 6-7 and table 3 respectively.

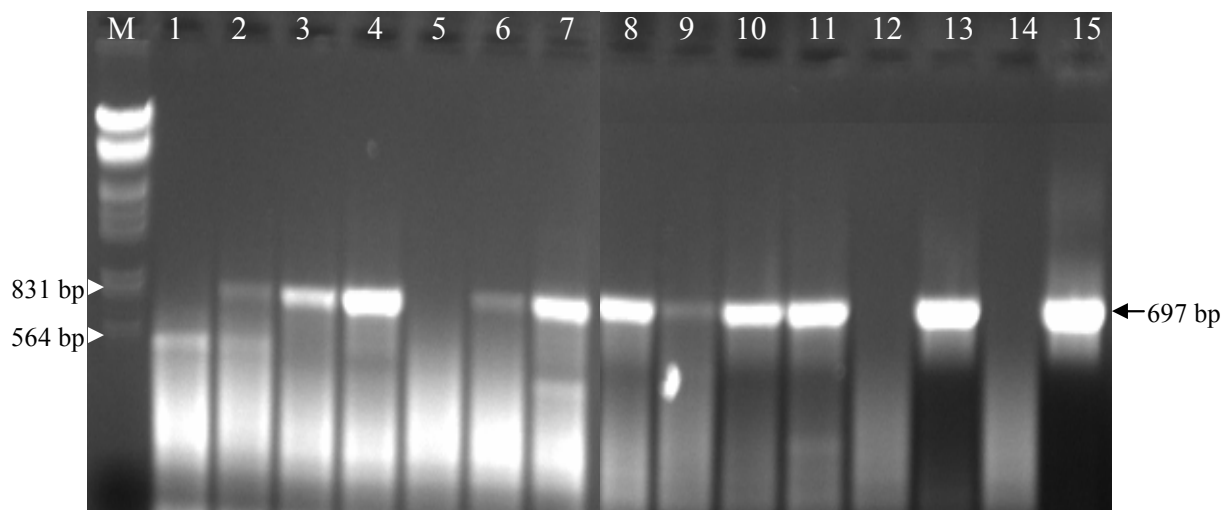


Figure 5: 1 % Gel electrophoresis photo of the conserved PCR of p23 gene from 11 CTV isolates using primers PM50 and PM51. The expected product is 697 bp. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA), (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 390-4; (6) 389-3; (7) 390-5; (8) 12-7; (9) 12-5 (10) 12-9 (11) GFMS 12; (12) Virus free; (13) T36 DNA clone; (14) Buffer control; and (15) GFMS 12 control cDNA.

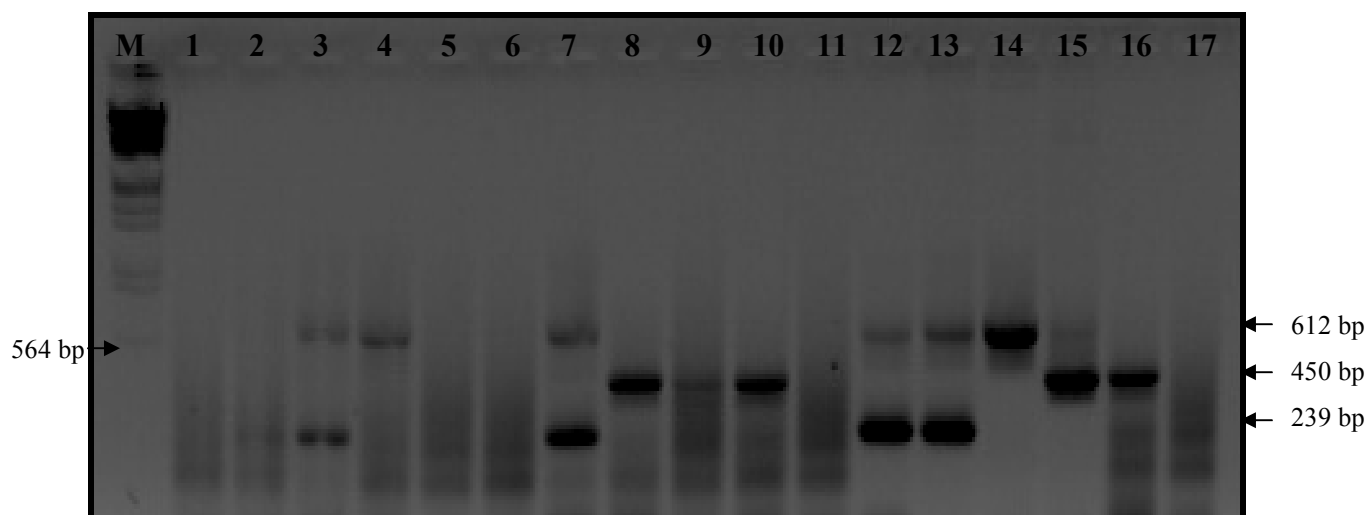


Figure 6: 1 % Gel electrophoresis photo of the bi-directional PCR of p23 gene from 11 CTV isolates using primers PM 85, 86, 82 and 84 to distinguish between severe and atypical strains. The possible products sizes are 612 bp, 450 bp, 239 bp. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA); (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 390-4; (6) 389-3; (7) 390-5; (8) 12-7; (9) 12-5 (10) 12-9 (11) Virus free; (12) T36 DNA clone (13) T36 DNA clone; (14) T385 DNA clone; (15) T305 DNA clone; (16) GFMS 12; and (17) Buffer control.

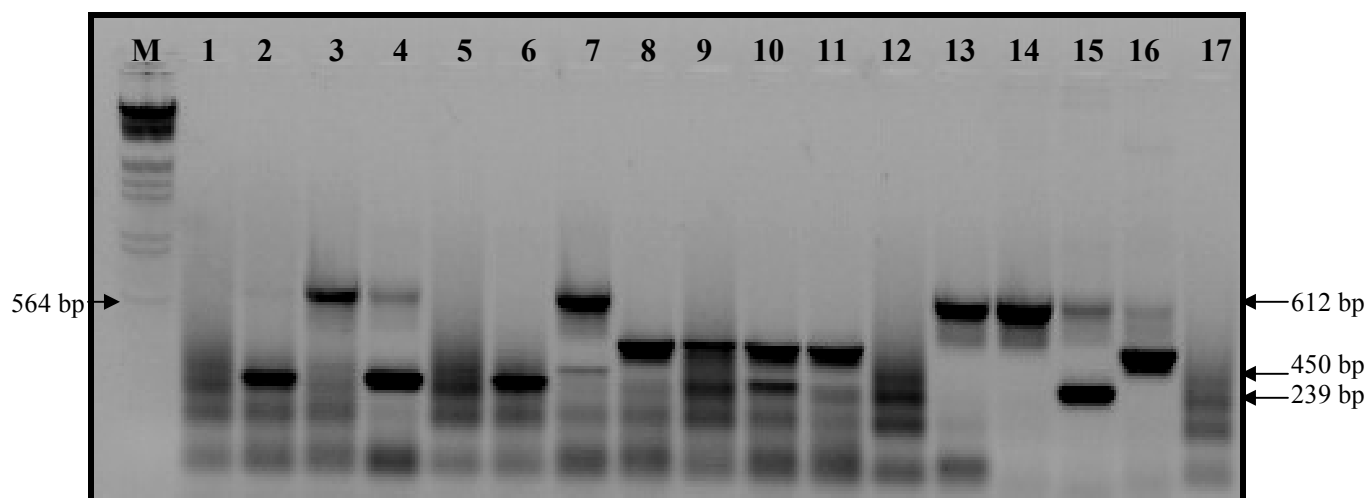


Figure 7: 1 % Gel electrophoresis photo of the bi-directional PCR of p23 gene from 11 CTV isolates using primers PM 85, 86, 82 and 83 to distinguish between severe and mild strains. The possible products sizes are 612 bp, 450 bp, 239 bp. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA), (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 390-4; (6) 389-3; (7) 390-5; (8) 12-7; (9) 12-5 (10) 12-9 (11) GFMS 12; (12) Virus free; (13) T36 DNA clone; (14) T36 DNA clone; (15) T385 DNA clone; (16) T305 DNA clone; and (17) Buffer control

Table 3:

A graphical representation of the amplified products for each isolate, using the primers PM 85, 86, 82 and 84 (severe versus atypical); and primers PM 85, 86, 82 and 83 (severe versus mild). The mild (T385), severe (T305) and atypical (T36) group controls are also included.

Isolate	PM 85, 86, 82, 83 (Severe vs. Mild)			PM 85, 86, 82, 84 (Severe vs. Atypical)		
	612 bp	450 bp	239 bp	612 bp	450 bp	239 bp
T30						
GFMS35						
GFMS12						
390-3						
390-4						
390-5						
389-3						
389-4						
12-5						
12-7						
12-9						
+ Control: T36 clone						
+ Control: T385 clone						
+ Control: T305 clone						
- Control: Virus free						
- Control: Buffer						

- + represents a positive control
- represents a negative control

Figure 8 shows the predicted amino acid sequences of the p23 gene of the 8 RSA isolates and 18 reference sequences from Sambade *et al.* (2003). A multiple amino acid sequence alignment of the p23 gene of the RSA isolates and other isolates of different geographic origins were compared in figure 8. All sequences were translated to their predicted amino acid sequences and all sequences had no unexpected stop codons. There were 53 polymorphic amino acid positions (approximately 25 % of the complete gene). There was great variability at positions 24-29, 50-54 and 78-80, as was also found by Sambade *et al.* (2003). There also appeared to be great variability at position 125-129. The summarised results of these positions are depicted in Table 5.

Isolates 12-9, 12-7, 12-5 and GFMS 12 had Glu²⁴, Lys²⁶ and Ser²⁹ as with other previously characterized severe isolates (VT, Cal-CB, T388, c269-6, Barao and T305). However isolates K, T36, and Galego-50; which are all atypical isolates also had this composition. Most reference mild isolates (T385, T32, T55, T300 and T312) had Lys²⁴, Glu²⁶ and Lys²⁹. The RSA isolates GFMS 35, 389-3 and 389-4 had Arg²⁴, Glu²⁶ and Lys²⁹, the same as the reference isolate T346 from the mild group. Isolate 390-5 had Ala²⁴, Arg²⁶ and Ser²⁹, the same as other atypical reference isolates 270-3 and 268-2.

In the region 50-54 RSA isolates 390-5, GFMS 35, 389-3 and 389-4 had Val⁵⁰, Thr⁵³ and Asn⁵⁴, the same as all the reference isolates from the mild group (T346, T385, T32, T55, T300 and T312) previously established and an atypical reference isolate c270-3. Other atypical reference isolates did not fall into any defined amino acid pattern and included: Galego-50 (Ile⁵⁰, Tyr⁵³ and Asn⁵⁴); isolate K (Ile⁵⁰, Ala⁵³ and Ser⁵⁴); and isolate c268-2 (Val⁵⁰, Thr⁵³ and Ser⁵⁴). RSA isolates GFMS12, 12-9, 12-7, 12-5 and all the severe reference sequences (Cal-CB, T305m T388, VT, c269-6 and Barao) had Ile⁵⁰, Asn⁵³ and Ser⁵⁴, as well as the atypical isolate T36.

In the region 78-80 isolates GFMS 35, 389-3 and 389-4 had Ala⁷⁸, Leu⁷⁹ and Lys⁸⁰, as described as the group defining amino acid region for the mild group (Sambade *et al.*, 2003) including mild reference isolates (T346, T385, T32, T55, T300 and T312). Isolates GFMS 12, 12-5, 12-7 and 12-9 had Ala⁷⁸, Ser⁷⁹ and Arg⁸⁰, as described as the group defining amino acid region for the severe group (Sambade *et al.*, 2003) including severe reference

isolates (T305, Barao, Cal-CB, T388, VT and c269-6). Isolate 390-5 had Gly⁷⁸, Leu⁷⁹ and Lys⁸⁰, as described as the group defining amino acid region for the atypical group (Sambade *et al.*, 2003) including atypical reference isolates T36, c270-3, K and c268-2. The only exception was Galego-50, an atypical reference isolate with Gly⁷⁸, Leu⁷⁹ and Arg⁸⁰.

And lastly the region 125-129 showed variability among isolates. Isolates GFMS 35, 389-3 and 389-4 had Glu¹²⁵, Leu¹²⁸ and Tyr¹²⁹; the same as the reference isolate T385 from the mild group. Many mild (T346, T55, T312 and T300) and atypical (270-3, Galego-50, T36, and c268-2) reference isolates had Asp¹²⁵, Met¹²⁸ and Tyr¹²⁹ as well as the RSA isolate 390-5. The only exception was a severe isolate, T305 which had this composition. Isolate T32 from the mild group had a unique composition of Asp¹²⁵, Ser¹²⁸ and Tyr¹²⁹. Isolates GFMS 12, 12-5, 12-7 and 12-9 had Asp¹²⁵, Met¹²⁸ and His¹²⁹, the same as the severe reference isolates (VT, 269-6, Barao, Cal CB and T388). Interestingly isolate K from the atypical group also had this composition.

The pairwise comparisons of amino acid sequence similarities of RSA isolates and group controls of Sambade *et al.* (2003) are shown in Table 4. There was a high degree of homology in amino acid sequences between: 12-5, 12-7, 12-9 and GFMS12 (100 %); 12-5, 12-7, 12-9, GFMS12 and the T305 severe group control (99 %); 389-3 and 389-4 (100%); and 389-3, 389-4 and GFMS35 (99%). The 389-3, 389-4 and GFMS35 isolates in comparison to the T385 mild group control ranged from 98.6- 98.1 %. There was a low degree of homology (95.2-93.8 %) between the following three groups of isolates: (1) 389-3, 389-4 and GFMS35; (2) 12-5, 12-7, 12-9 and GFMS12; and (3) 390-5. The T36 control was not particularly similar to any of the RSA isolates. The lowest sequence homology was found between 390-5 and the atypical control T36 of 92.9 %. There was a total range of 100-93.8 % similarity at the amino acid level (Table 4) among the RSA isolates in this study. And a range of 97.1-94.7 % similarity between the standards of the three control groups (T305, T385 and T36).

Figure 8

Multiple Alignment using ClustalW of predicted amino acid sequences of the p23 protein. Regions described in the text are boxed. The basic residues and zinc finger motif residues of the RNA binding domain are indicated in bold with boxes. For each position, identical residues are indicated by dots. Locations of the primers used in the PCR amplifications are indicated.

CGGACAAACTTTCGTTTCTGTGAACCTTTC

PM 85

	10	20	30	40	50	60	70	80
Consensus	MDDTSGQTFV	SVNLSDESNT	ASTEVKTVSS	EADRLEFLRK	MNPFIDALI	RKTSYQGARF	RARIIGVCVD	CGRKHKALK
T312 prot			K.EN.K	...D...	...V..V	...N...		
T346 prot			R.EN.K		...V..V	...N...		
T385 prot	...I....		K.EN.K	...I...	...V..V	...N...		
t32 prote			K.EN.K		...V..V	...N...		
t55 prote	..N.....		K.EN.K		...V..V	...N...		
t300 prot			K.EN.K	...D...	...V..V	...N...		
389-3			R.EN.N	...L...	...V..V	...N...		
389-4			R.EN.N	...L...	...V..V	...N...		
GF35			R.EN.K	...L...	...V..V	...N...		
268-2 pro	..N.....		A.R...	...D...	...L...V			G..
270-3 pro	I		A.R...		...V..V	...N...		G..
Galego-50	I	..D....	L		...T...N			..N.G.R
k protei	..N.....			...T...	...A...			G..
T36 prote	..N.....	..T...	...P...	...D..Q.	...I...N			G..
390-5			A.R...	...S...	...V..V	...K...		G..
Barao pro	I		...A...		...N...			..SR
Cal-CB pr	I				...N...			..SR
T305 prot	I				...N...			..SR
T388 prot	I				...N...			..SR
Vt protei	I		...L		...I...			..SR
269-6 pro	I		...A...		...N...			..SR
12-5	I		...I...		...N...			..SR
12-7_Tran	I		...I...		...N...			..SR
12-9	I		...I...		...N...			..SR
GF12	I		...I...		...N...			..SR

24-29
50-54 (Basic Region)
78-80 (Zinc Finger motif)

	90	100	110	120	130	140	150	160
Consensus	TERKCKVMNT	QSQNEVAHML	MHDPVKYLNK	RKARAFSNAE	MFAIDLVMYT	KERQLAVDLA	AEREKTRLAR	RHPMRSPEET
T312 prot								..I....
T346 prot			...R...		...T...	...K...	...A...	
T385 prot					...E..L			..I....
t32 prote					...S...			..I....
t55 prote								..I....
t300 prot					...T...			..I....
389-3					...E..L			..I....
389-4					...E..L			..I....
GF35					...E..L			..I....
268-2 pro						...K...		
270-3 pro			...SE	G..		...K...N.T		
Galego-50			..N..E	G..		...K.S..N.		
k protei					...H...			
T36 prote				...I...		...I...		
390-5			...R...	G..		...K...N.		
Barao pro					...H...			
Cal-CB pr					...H...			
T305 prot					...H...			
T388 prot				...E...	...H...			..A...
Vt protei					...H...			
269-6 pro			...I...		...H...			
12-5	A...				...H...			
12-7_Tran	A...				...H...			
12-9	A...				...H...			
GF12	A...				...H...			

125-129

Figure 8 continued

	170	180	190	200	210
Consensus	PEHYKFGMTA	KAMLPDINAV	DVGDNEDTSS	EYPVSLSVSG	GVLREHHFI*
T312 prot			E.....		
T346 prot		EF.....	G.....		
T385 prot			E.....		
t32 prote		T.....	E.....		
t55 prote			E.....		
t300 prot			E.....		
389-3			E.....		*
389-4			E.....		
GF35			E.....		*
268-2 pro		V.....EL.....	G.....		
270-3 pro			V.....	I.D.....	
Galego-50		T.....		N.....	
k protei	T.....	V.....	E.....		
T36 prote	..Y.....R.....		E.....		
390-5	..L.....I.....	E.....N.....		D.....	*
Barao pro					
Cal-CB pr					
T305 prot					*
T388 prot					
Vt protei		N.....I.....			
269-6 pro					
12-5	I.....	I.....			*
12-7 Tran	I.....	I.....			*
12-9	I.....	I.....			*
GF12	I.....	I.....			*

CTCAAGAGGCACTTGTGGTGAAGTAG
 PM 86

Table 4:

Predicted amino acid sequence similarity (%) of the p23 gene of CTV from South Africa

	GF12-5	GF12-7	GF12-9	389-3	389-4	390-5	GFMS12	GFMS35	T305*	T385*	T36 *
GF12-5	-	100.0	100.0	95.2	95.2	93.8	100.0	95.2	99.0	95.2	96.2
GF12-7		-	100.0	95.2	95.2	93.8	100.0	95.2	99.0	95.2	96.2
GF12-9			-	95.2	95.2	93.8	100.0	95.2	99.0	95.2	96.2
389-3				-	100.0	94.3	95.2	99.0	95.7	98.1	95.2
389-4					-	94.3	95.2	99.0	96.2	98.6	95.7
390-5						-	93.8	94.3	94.3	93.8	92.9
GFMS12							-	95.2	99.7	95.2	96.2
GFMS35								-	95.7	98.6	95.2
T305 *									-	96.2	97.1
T385*										-	94.7
T36 *											-

T305 (AJ579772): severe group; T36 (AJ579776): atypical group; T385 (AJ579762): mild group. These are published sequences that are indicators for CTV strain groups in previous work (Sambade *et al.*, 2003) and are used here as reference strain sequences for comparison.

Table 5:

Difference in Amino acid residues of 4 different regions of variability of the reference and RSA isolates. RSA isolates are shown in bold.

Isolate	Region 24,26,29	Region 50,53,54	Region 78,79,80	Region 125, 128, 129
Mild				
T346	REK	VTN	ALK	DMY
T385	KEK	VTN	ALK	ELY
T32	KEK	VTN	ALK	DSY
T55	KEK	VTN	ALK	DMY
T300	KEK	VTN	ALK	DMY
T312	KEK	VTN	ALK	DMY
GFMS35	REK	VTN	ALK	ELY
389-3	REK	VTN	ALK	ELY
389-4	REK	VTN	ALK	ELY
Atypical				
C270-3	ARS	VTN	GLK	DMY
Galego-50	EKS	ITN	GLR	DMY
K	EKS	IAS	GLK	DMH
T36	EKS	INS	GLK	DMY
C268-2	ARS	VTN	GLK	DMY
390-5	ARS	VTN	GLK	DMY
Severe				
Cal-CB	EKS	INS	ASR	DMH
T305	EKS	INS	ASR	DMY
T388	EKS	INS	ASR	DMH
VT	EKS	ITS	ASR	DMH
C269-6	EKS	INS	ASR	DMH
Barao	EKS	INS	ASR	DMH
GFMS12	EKS	INS	ASR	DMH
12-5	EKS	INS	ASR	DMH
12-7	EKS	INS	ASR	DMH
12-9	EKS	INS	ASR	DMH

The 697 bp RT-PCR amplified products were sequenced for each RSA isolate and trimmed to contain 627 bp of the complete p23 gene (209 amino acids). There were no gaps or insertions in these sequences. The DNA sequences of the p23 gene of isolates GFMS 12, GFMS 35, 389-3, 389-4, 390-5, 12-5, 12-7 and 12-9 are in Appendix 1 (figures 28-35).

The p23 gene DNA sequence of isolate 12-7 was 100 % homologous to isolate 12-5. The p23 gene DNA sequences of isolates 12-9 and GFMS 12 had a 96-97 % nucleotide similarity to the cognate region of many defective RNA sequences (AJ579773; U35120; AY206452). Isolates 389-3, 389-4 and GFMS 35 had a 98 % nucleotide similarity to the cognate regions of isolates 464-2 (AY995566); 464-1 (AY995565); 425 (AY995564); and 420-1 (AY995563) of unpublished work submitted to Genbank of single aphid and graft transmissions of a naturally infected tree in California, USA, where no biological characterization records are available. The DNA sequence of the p23 gene of isolates GFMS 35, 389-4 and 389-3 were very similar. The DNA sequence of the p23 gene of isolate 390-5 had a 98 % nucleotide identity to clones of isolate C315-14 (AY962359, AY962361, AY962351). The isolate comes from a population of strains of a grapefruit isolate selected for pre-immunisation assays in Argentina in 2005 (Iglesias, unpublished data). No data is available on the biological characterization of this isolate.

A multiple sequence alignment of the p23 gene of the RSA isolates and reference isolates of Sambade *et al.* (2003) from different geographic origins were compared in Appendix 1, figure 36. Phylogenetic analysis of the 8 RSA isolates and the 18 biologically characterized reference isolates of Sambade *et al.* (2003) of the complete p23 gene was done. The accession numbers of the sequences of the reference isolates, their biological properties and the strain classification according to Sambade *et al.* (2003) are depicted in Table 2. The tree showed two groups of isolates that appeared clearly separated: group 1 & 3 (figure 9). Group 2 did not appear to be clearly separated from the other groups (figure 9). Reference isolates within the first group (T305, T388, Cal-CB, Barao, 269-6 and VT) were previously characterized as severe inducing seedling yellows and stem-pitting (table 2). The isolates 12-5, 12-7, 12-9 and GFMS 12 from RSA grouped with these severe isolates. Reference isolates in the second group (K, 270-3, T36, Galego-50 and 268-2) showed variable pathogenicity characteristics (table 2) including the symptomless isolate K and

others inducing seedling yellows or even stem-pitting in grapefruit (270-3). Isolate T36 and K were more distantly related to the other atypical isolates. The isolate 390-5 from RSA grouped together with the atypical isolates. Reference isolates in the third group (T346, T55, T32, T385, T300, and T312) showed only mild to moderate symptoms (table 2) in Mexican lime and sometimes decline on sweet orange. The isolates 389-3, 389-4 and GFMS35 from RSA all grouped together with these mild reference isolates.

Within the severe clade (figure 9), RSA isolates 12-9 and GFMS 12 clustered very closely to the severe VT isolate. RSA isolates 12-5 and 12-7 which are 100 % identical clustered closely to severe isolates T388 and T305. The group 1 sequences were distinctly separated into 2 subgroups consisting of: (A) 12-9, GFMS12 and VT; and (B) 12-5, 12-7, T388, T305, Cal-CB, Barao, and 269-6. Within the atypical clade (group 2) the isolate 390-5 clustered most closely with c268-2 and together with c270-3 and Galego-50 formed a separate subgroup. Isolates T36 and K formed their own branches distinctly separate from the other atypical isolates. Within the mild clade (group 3) of isolates there were three distinct sub-branches. RSA isolates GFMS35, 389-3 and 389-4 clustered together in a subgroup on their own and isolates T312, T300, T32, T385 and T55 formed another subgroup. These two subgroups were more closely related to each other than the mild isolate T346 which formed its own subgroup.

Table 6 summarizes the intra-group genetic diversity (average nucleotide distance between two isolates of the same group selected randomly) and the inter-group genetic diversity (average nucleotide distance between two isolates, one of each group, selected randomly). A score of zero would indicate a 100% homology. The intra-group diversity of groups 1 and 3 were two to five fold (0.019 and 0.043 respectively) smaller than their inter-group diversity (0.094). Therefore the isolates of group 1 are similar to each other, as with group 3 isolates. The isolates of group 1 and group 3 are distinctively different from each other.

In contrast, intra-group and inter-group values of group 2 compared with groups 1 and 3 were similarly high, in agreement with the topology of the phylogenetic tree. This indicates that isolates are more distantly related within group 2 (0.106) and compared to groups 1 (0.094) and 3 (0.108).

The inter-group diversities of groups 1 and 2 and of groups 1 and 3 were equally high (0.094), however it was higher between the group 2 and 3 (0.108). This was unexpected and in contrast to previous work by Sambade *et al.* (2003), where there the most diverse groups were 1 and 3.

Table 6:

Intra-group and inter-group genetic diversity values estimated for mild, severe and atypical CTV isolates. Nucleotide distances were calculated with the MEGA program.

	Group 3 (Mild)	Group 1 (Severe)	Group 2 (Atypical)
Group 3 (Mild)	0.043 (4.3 %)	0.094 (9.4 %)	0.108 (10.8 %)
Group 1 (Severe)		0.019 (1.9 %)	0.094 (9.4 %)
Group 2 (Atypical)			0.106 (10.6 %)

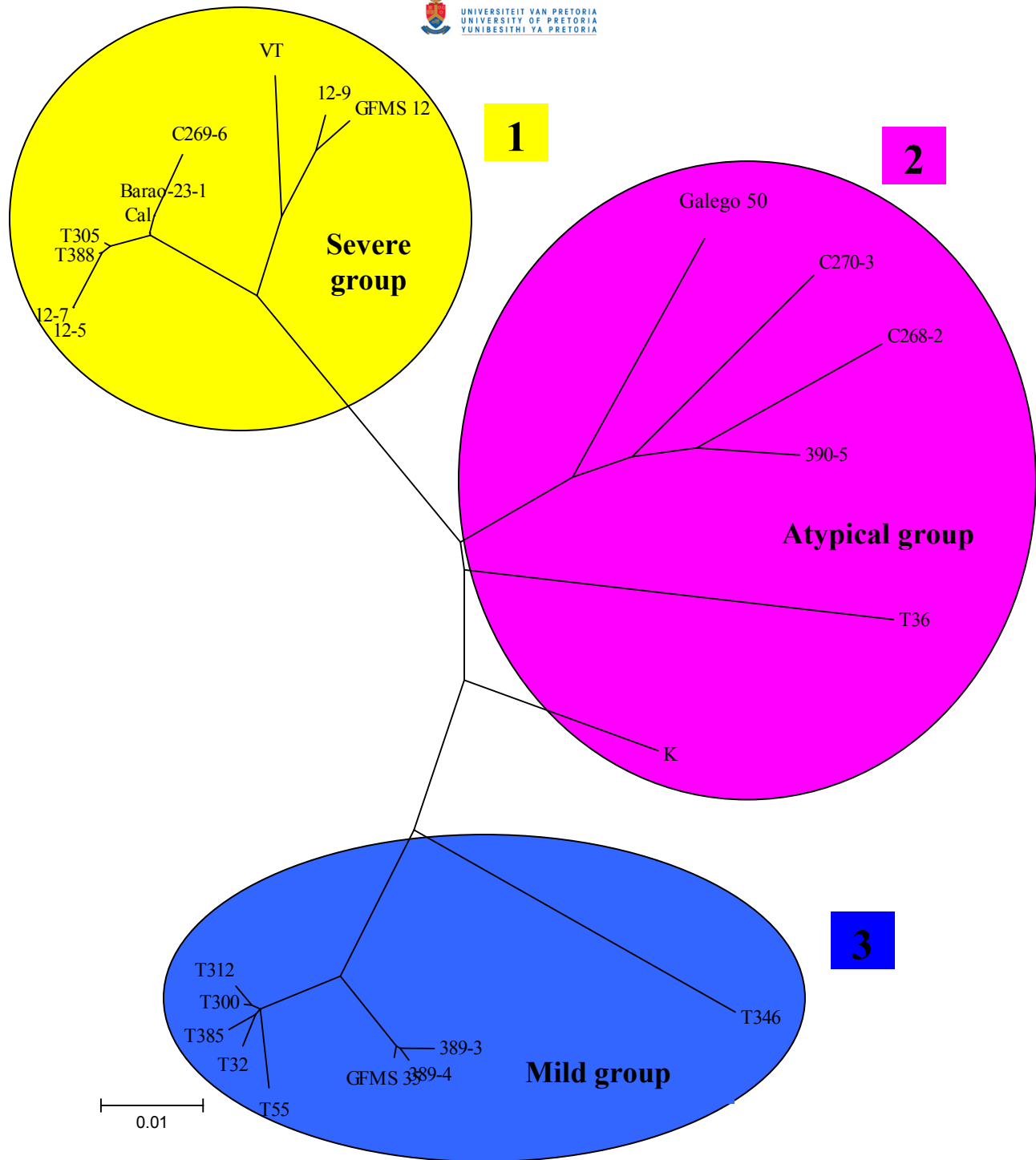


Figure 9: An unrooted phylogenetic tree obtained by the neighbour joining method with the nucleotide sequences of the p23 gene of RSA and other reference CTV isolates.

2.4 DISCUSSION

The isolates 389-3, 389-4, 390-3, 390-4, and 390-5 have been evaluated in the greenhouse and growth and stem-pitting per cm² on Mexican lime indicator plants recorded (Breytenbach, unpublished). Six months after bud inoculation growth and stem pitting were recorded. The sub-isolates all had stem-pitting values of below 0.2 stem pits per cm², growth of 700-980 mm and contained a high viral titer (Breytenbach, unpublished). Based on these evaluations these sub-isolates appear mild but it does not mean that they have cross-protecting ability (Breytenbach, unpublished). Work on these sub-isolates to determine their ability to cross-protect against severe strains is still in progress (Breytenbach, unpublished), and are currently not being used in the cross-protection scheme (van Vuuren., personal communication). These sub-isolates were prepared by single aphid transmissions. In theory this process selects for only one strain, but in reality a few strains could still be present after single aphid transmissions. These sub-isolates could therefore still have a heterogeneous population of strains, and this must be considered when analyzing molecular results and comparing them to biological indexing.

In 2000 biological and molecular characterization of single aphid transmissions from a cross-protecting source referred to as GFMS12 were performed (van Vuuren *et al.*, 2000b). It was found that 12-5 had significantly less stem-pitting in Mexican lime plants than the original GFMS12 (mixture of unknown strains). There were no differences in the growth of the plants infected with the sub-isolates compared to those with the original GFMS12 source (van Vuuren *et al.*, 2000b). The sub-isolates used in this study showed the following results: (12-5) had 5.1 and 0 pits per cm² and 500 and 513 mm growth in Mexican lime and Marsh grapefruit respectively; (12-7) had 8.8 and 18.1 pits per cm² and 530 and 493 mm growth in Mexican lime and Marsh grapefruit respectively; and (12-9) had 10.2 and 5.4 pits per cm² and 458 and 462 mm growth in Mexican lime and Marsh grapefruit respectively (van Vuuren *et al.*, 2000b). Some sub-isolates showed significantly higher stem-pitting and were not used in further cross-protection trials (van Vuuren *et al.*, 2000b). All the sub-isolates were amplified and tested with RFLP of the CP gene and SSCP. Isolate 12-7 gave similar RFLP bands to the original GFMS12 isolate, while all the other sub-isolates differed. 12-5, 12-7 and 12-9 all had a different RFLP patterns (van Vuuren *et al.*, 2000b). Isolates 12-5 and 12-9 contained one or two bands which were absent in the original. The SSCP technique revealed no differences between the CP genes of the original isolate and

the sub-isolates. It was reported that none of these molecular techniques gave well defined results, and none of the RFLP profiles corresponded to the RFLP groups described by Gillings *et al.* (1993). Either these sub-isolates contained sequences very divergent to ones tested before or these sub-isolates contain more than one strain (van Vuuren *et al.*, 2000b). Because of the importance of this on molecular tests performed in this study further serial single aphid transmissions were proposed to separate these individual strains (van Vuuren *et al.*, 2000b). A need for more sensitive molecular differentiation tests to differentiate these GFMS12 sub-isolates was expressed (van Vuuren *et al.*, 2000b). These sub-isolates could still contain a severe strain which could become dominant in different climatic conditions and in different hosts and cause segregation and changes in dominance of strains with a dramatic effect on the plant. It was previously reported that GFMS 12 pre-immunized in Star Ruby Grapefruit performed poorly and contained a severe strain (Marais *et al.*, 1996) resulting in GFMS35 being introduced to pre-immunize red grapefruit selections (van Vuuren *et al.*, 2000b).

In 2000 a study on the performance of accepted grapefruit immunizing isolates showed that GFMS12 performed poorly in Marsh and Star Ruby and GFMS 35 performed badly in Marsh and Nel Ruby (van Vuuren *et al.*, 2000a). The poor performance was attributed to either the breakdown of protective abilities or segregation of different strains present in each isolate induced by host and/or environmental conditions or by a strain already present becoming dominant under these conditions (van Vuuren *et al.*, 2000a). The poor performance was not attributed to a natural introduction of a severe strain as virus free control plants retained good production and crop value suggesting that cross-protection breakdown was not the cause (van Vuuren *et al.*, 2000a).

2.4.1 CLASSIFICATION OF ISOLATES BY BI-DIRECTIONAL PCR

The bi-directional PCRs showed that isolates with single differences at amino acid positions 78-80 could be detected in this PCR system developed by Sambade *et al.* (2003), some isolates however showed decreased PCR effectiveness with faint bands, which makes analysis complicated. Cloning was not performed in this study to ensure that the predominant sequence variant was characterised as done by Sambade *et al.* (2003) as this would be impractical for routine testing of samples, and in this study it is important to detect any sequence variants (mixed or alone) not just the predominant ones. Even though PCR could possibly select the more predominant sequence it can not be guaranteed. As

with all PCRs it is common to have some sort of PCR bias, and in this case it could be either due to a predominance of a certain strain/variant and/or a bias towards the amplification of a more stable primer-template annealing interaction with no association to the dominant strain. A more stable primer-template interaction could be based on thermodynamics, complementarity or the presence of secondary structures which of these parameters possibly responsible for PCR bias here remains to be seen.

Isolates were categorized into their respective groupings based on these PCR systems established by Sambade *et al.* (2003). The summary of the group designations of the isolates based on both bi-directional PCR amplifications (PM 85, PM 82, PM83, PM 86 and PM 85, PM 82, PM84, PM 86) are described below. Isolates 390-3 and 390-5 show the exact predicted amplifications of atypical group indicator T36. Isolates 390-4 and 389-3 yield the 239 bp band of mild strains but both lack the 612 bp amplicon, suggesting the PM86 primer is not annealing optimally and possibly has sequence differences at this position. Isolate 389-4 follows the same amplification pattern as the mild group control T385.

GFMS 35, a population of different strains used in the cross-protection scheme showed a mixed pattern of mild and atypical bands. According to this system it does not seem to have a severe component, or was not at detectable levels. Testing of single aphid transmissions from this isolate would help to validate this. Isolates GFMS 12, 12-7 and 12-9 have mixed infections of severe and mild isolates, but all lack a 612 bp amplicon in both PCR systems. Detection of mixed strains/variants in isolates 12-7 and 12-9 is of concern since these are sub-isolates from single aphid transmissions and were assumed to be single strains. van Vuuren *et al.* (2000) showed by biological indexing that isolates 12-5, 12-7 and 12-9 were mild but also had concerns that some of the sub-isolates were still mixed strains and under certain circumstances a severe component could become dominant and cause cross-protection breakdown.

Isolate 12-5 belongs to the severe group based on the presence of a 450 bp amplicon in the PM 85, 86, 82, 83 PCR system but lacks the 612 bp conserved amplicon whereas the severe control T305, had both expected amplicons. However, 12-5 was shown to be mild according to bioindexing (van Vuuren *et al.*, 2003). The severe form could become dominant under certain circumstances. Unusually, isolate 12-5 does not appear have any mild components.

It might be that the titer of mild variants was too low to detect with PCR or that a bias against amplifying the mild variants exists.

Isolates described above were indexed as mild but contain mild, atypical and severe components. Mild symptoms are not just a result of which strains are present but by more complex, undefined strain and host dynamics. The effects of unknown strain dominance; varying host responses; or strain interactions could be possible causes for this symptom expression. PCR based methods targeting other areas of the CTV genome will help to understand the strain variants present in these isolates. From this study isolates GFMS12, 12-5, 12-7 and 12-9 have atypical and/or severe forms potentially threatening the cross-protection programme.

Isolate T30 did not yield any amplicons even though primers PM 50/51 and PM85/86 are 100 % complementary to the T30 strain sequence (AF260651). ELISA results presented in Chapter 3 show that the T30 isolate has detectable CTV titers, as also shown by CTV amplifications of different parts of the genome. Therefore the lack of amplicons of isolate T30 can not be ascribed to sub-detectable titers.

Amplification of the 612 bp fragment of the PM85 and PM86 primer set with both of the PCR systems appears inconsistent. Sambade *et al.* (2003) does not report this but it is evident in figures published by them (see reaction of isolates Cal-CB, C- and C270-3). In their study, isolates yielded various levels of amplification evident from images of gel electrophoresis. Certain isolates also had faint 612 bp amplicon bands but very distinct bands of the group specific primers, indicating that presence of faint bands is not due to insufficient template. Similar observations were also made during this study.

Either PM85/86 primer/s had lost functionality/denatured or were not optimally annealing to certain isolates due to possible DNA sequence differences. All the controls (T385, T305 and T36) amplified the 612 bp fragment and therefore had viable primer/s. Primers PM85 and 86 are conserved for sequences T305, T36 and T385 (Sambade *et al.*, 2003). To amplify the internal 450 bp and 239 bp amplicons of the three groups both external primers PM85 and 86 are needed in these bi-directional PCRs. Therefore primers PM 85 and 86 are annealing optimally since the two internal amplicons are present in RSA isolates.

The third possibility is that the internal shorter fragments (239 or 450 bp) are preferentially amplified over the larger fragment of 612 bp. This could be due to primers PM 85 and 86 not annealing as optimally as the internal primers PM 82, 84 and 83 or a general preference for the amplification of the smaller amplicon. Both possibilities lead to a bias in the amplification of the smaller amplicon/s and an exhaustion of the reaction components. This was observed when the internal primers are not complementary to the target (e.g. isolates 390-3 and 390-5), and then larger 612 bp fragment is amplified preferentially.

More importantly the classification of strains into groups mild, severe and atypical is based solely on differences at amino acid positions 78-80 (Sambade *et al.*, 2003), which is exploited by internal primers PM 82, 83 and 84. Therefore amplification of 612 bp amplicon from primers PM 85 and 86 is not needed for strain differentiation but functions as the supporting primers in the bi-directional PCRs.

2.4.2 PREDICTED AMINO ACID SEQUENCE ANALYSIS

Multiple sequence alignment of the predicted p23 amino acid sequences showed four regions of increased variability for possible differentiation of isolates between amino acid positions: 24-29; 50-54; 78-80 and 125-129. Sambade *et al.* (2003) had found three regions of interest: positions 24-29 and 50-54 separated mild isolates from others, whereas positions 78-80 allowed discrimination between the three groups (Sambade *et al.*, 2003). Positions 50-54 and 78-80 are part of a p23 domain involved in binding RNA in a non-specific manner (López *et al.*, 2000). The region 50-54 is the basic region of the RNA binding domain (López *et al.*, 2000) and the region between 78-80 is the zinc-finger motif of the RNA binding domain (López *et al.*, 2000). The area between amino acid residues 98-171 did not affect RNA binding (López *et al.*, 2000) and has no known function.

There was increased variability at amino acid positions 24, 26 and 29. Sambade *et al.* (2003) found that isolates could be separated into severe and mild isolates with the composition Glu²⁴, Lys²⁶, Ser²⁹ and Lys²⁴, Glu²⁶, Lys²⁹ respectively. These differences affect the isoelectric point, which is 9.94 in mild isolates and 6.97 in severe isolates. Isolates 12-9, 12-7, 12-5 and GFMS 12 had Glu²⁴, Lys²⁶ and Ser²⁹ in common with characterized severe isolates (VT, Cal-CB, T388, c269-6, Barao and T305). However atypical isolates: K,

known to be symptomless but is a mixed infection of varying severity (Brlansky *et al.*, 2003); T36, causing quick decline; and Galego-50 showing atypical seedling yellows and stem-pitting, also had this composition. The amino acid at position 29 was incorrectly described in Sambade *et al.* (2003) as Val²⁹ which should have been Ser²⁹ for the severe group description. None of the presumed mild RSA isolates had the position 24, 26, 29 amino acid composition of the mild group of Sambade *et al.* (2003) where most mild isolates (T385, T32, T55, T300, T312) had Lys²⁴, Glu²⁶ and Lys²⁹. However RSA isolates GFMS 35, 389-3 and 389-4 had Arg²⁴, Glu²⁶ and Lys²⁹, the same as the reference isolate T346 from the mild group, which clusters slightly separate from the majority of isolates. There might be differences in the level of mildness of mild isolates in RSA compared to other known mild isolates from around the world, even though their sequences all cluster together (see section on Phylogenetic analysis). Isolate 390-5 had Ala²⁴, Arg²⁶ and Ser²⁹, the same as other atypical reference isolates 270-3 and 268-2. It seems as though this site is not reliable for differentiation of mild versus atypical groups since there are two different compositions for the biologically mild isolates and also two different compositions for the atypical isolates. The severe strains however seem to all have Glu²⁴, Lys²⁶ and Ser²⁹ which could potentially be used as a marker for severe variants even though there were some atypical isolates with this too. It is important for future molecular marker discovery that there were no mild isolates with this severe composition. The composition of the severe and mild isolates shows that all three positions differ and it could be hypothesized that this represents a major split and is possibly an ancient divergence of CTV variants into distinct groups which, with the strong modification of the isoelectric point, could result in functional differences, but would need to be validated with more isolates in future studies. It is also interesting that even though the atypical group had two different compositions for these sites, none of these are present in the mild isolates. The role of atypical strains in strain variability remains unclear but in the amino acid region 24-29 the two compositions suggests two main divergent paths: (1) isolates K, Galego-50 and T36 were the same as severe isolates (Glu²⁴, Lys²⁶, Ser²⁹) and (2) isolates c270-3, c268-2 and 390-5 had a completely unique and divergent composition (Ala²⁴, Arg²⁶ and Ser²⁹) not found in any other isolates of any of the mild and severe groups. This could represent a split in divergence of isolates with an undefined severity and variable symptoms that are not regarded as a conventional severe isolate. This hypothesis will be further expanded on with the phylogenetic tree discussion in the next section. The isoelectric point for Ala²⁴, Arg²⁶ and Ser²⁹ was 11.05, more similar to the mild of 9.94 than the severe of 6.97. Interestingly

the RSA isolate 390-5 consisted of the unique atypical composition of unknown significance.

Positions 50-67 is the basic region of the p23 gene consisting of several basic residues of the RNA binding domain (López *et al.*, 2000). The amino acid positions 50, 53 and 54 of RSA isolates 390-5, GFMS 35, 389-3 and 389-4 were Val⁵⁰, Thr⁵³ and Asn⁵⁴, the same as all the reference isolates from the mild group (T346, T385, T32, T55, T300 and T312), as well as the atypical reference isolate c270-3. RSA isolates GFMS12, 12-9, 12-7, 12-5 and all the severe reference sequences (Cal-CB, T305m T388, c269-6 and Barao) had Ile⁵⁰, Asn⁵³ and Ser⁵⁴, as well as T36 from the atypical group. The only exception within the severe isolates was VT, with position 53 being Tyr instead of Asn. There were six different compositions in the atypical group of isolates, of which three are unique to this group: (1) isolates c270-3 and 390-5 the same as the mild group isolates (Val⁵⁰, Thr⁵³ and Asn⁵⁴) ; (2) isolate Galego-50 with a difference at position 50 (Ile⁵⁰, Tyr⁵³ and Asn⁵⁴) relative to the mild isolates; (3) isolate c268-2 with a difference at position 54 (Val⁵⁰, Thr⁵³ and Ser⁵⁴) relative to the mild isolates; (4) isolate T36 the same the severe group isolates (Val⁵⁰, Thr⁵³ and Asn⁵⁴); and (5) isolate K with a difference at position 53 (Ile⁵⁰, Ala⁵³ and Ser⁵⁴) relative to the severe isolates. Again there seems to be a lot of variability among atypical isolates which raises the question as to where these isolates actually fit in. Of interest is that all the isolates of the three groups either had Val or Ile at position 50 and either Ser or Asn at position 54. Position 53 seemed to be variable among isolates of the severe and atypical groups and appears to discriminate between severe and mild strains. This could represent a major split/divergence of specific sequences. This region could also be very useful as a strain molecular marker.

Within amino acid position 68-86 is the putative zinc-finger motif of the RNA binding domain (López *et al.*, 2000). At positions 78, 80 and 81 isolates GFMS 35, 389-3 and 389-4 had Ala⁷⁸, Leu⁷⁹ and Lys⁸⁰, expected of the mild group (Sambade *et al.*, 2003) of isolates T346, T385, T32, T55, T300 and T312. Isolates GFMS 12, 12-5, 12-7 and 12-9 had Ala⁷⁸, Ser⁷⁹ and Arg⁸⁰, typical of the severe group (Sambade *et al.*, 2003) such as isolates T305, Barao, Cal-CB, T388, VT and c269-6. Isolate 390-5 had Gly⁷⁸, Leu⁷⁹ and Lys⁸⁰, as for members of the atypical group (Sambade *et al.*, 2003) with isolates T36, c270-3, K and c268-2. The only exception was atypical isolate Galego-50 with Gly⁷⁸, Leu⁷⁹ and Arg⁸⁰, which was incorrectly documented as T36 in the study by Sambade *et al.* (2003). The

amino acid residues in this region seem to follow a distinct pattern: (1) at position 78 it is either Ala for severe and mild or Gly for atypical isolates; (2) at position 79 it is either Leu for mild and atypical or Ser for severe isolates; and lastly (3) at position 80 it is either Lys for mild and atypical or Arg for severe (including atypical Galego-50) isolates. It appears from this region that the mild and atypical group are most similar (one amino acid difference) followed by the severe and mild groups (with two amino acid changes) and then by the severe and atypical groups having all three amino acids as different. The occurrence of Galego-50 with a unique composition is most similar to the atypical group with just one difference in the composition.

Concurring with Sambade *et al.* (2003) this region allows discrimination between the mild, severe and atypical groups and is the area from which the primers in the bi-directional PCR were developed (as discussed before). The three groups can clearly be separated based on this region; however the actual involvement of each group in symptom expression can not be concluded from this. More research should be done to document the effects of the 78-80 position on symptom expression. The amino acid residues involved in RNA binding are conserved and include a few basic residues; and Cys and His that coordinate the Zn ion.

Lastly, amino acid region 125-129 of unknown function, showed variability among isolates. The mild group of reference isolates were divided into three sequence types, however the majority of isolates had Asp¹²⁵, Met¹²⁸ and Tyr¹²⁹ (T346, T55, T312 and T300). Most atypical isolates also had this composition including RSA isolate 390-5 and one severe isolate T305. RSA isolates GFMS 35, 389-3 and 389-4 had Glu¹²⁵, Leu¹²⁸ and Tyr¹²⁹; the same as the reference isolate T385 from the mild group. This sequence was not found in any of the other isolates, and this could imply that these RSA isolates have a unique composition at this region found in only a few other isolates. Isolate T32 from the mild group had a unique composition of Asp¹²⁵, Ser¹²⁸ and Tyr¹²⁹. Isolates GFMS 12, 12-5, 12-7 and 12-9 had Asp¹²⁵, Met¹²⁸ and His¹²⁹, the same as the majority of severe reference isolates (VT, 269-6, Barao, Cal CB and T388). Interestingly isolate K from the atypical group also had this composition. It appears from this area that there are two main sequence types: the mild Asp¹²⁵, Met¹²⁸ and Tyr¹²⁹; and severe Asp¹²⁵, Met¹²⁸ and His¹²⁹ that differ by only one amino acid. The only other unique sequences were found in two isolates (T385 and T32) in the mild group. In this region it is unclear where the atypical isolates fit in but it appears that they either group with a mild or severe sequence composition, in this case most

grouped with the mild isolates. This area might also be a useful to target as severe strain molecular markers in future work on strain differentiation.

It appears from this work that the areas 24-29, 50-54, 125-129 can potentially be used as molecular marker sites to differentiate mild and severe strains. There have been a few exceptions though to the given pattern produced by severe and mild isolates, and therefore these sites should not be used separately but as a whole to get a better picture of what an isolate contains and from there possibly assign it to a defined group. The area 78-80 seems to be a more suitable area to assign isolates into mild and severe groups but more importantly to place isolates with no apparent defined symptom type into another group called atypical. However isolates in this atypical group seem to all be very different and it seems unlikely that this region alone would account for such different symptom expression. It is however plausible that the region 78-80 could interact with other regions on the genome to produce defined symptoms. However any functional differences discussed in this study of the p23 gene might not relate to symptom expression at all. This site is however a step forward to understanding possible interactions among strains and their symptoms. This area of the genome together with other similar areas in other genes hold a lot of promise in the future for the design of an oligonucleotide microarray chip to completely differentiate strains and for expansion work on the current microarray chip developed in this study (see chapter 4). The p23 gene is highly conserved and differences in DNA sequence could represent possible differences in function, which would not necessarily be the case in highly variable regions such as the ones targeted in Chapter 3.

2.4.4 PHYLOGENETIC ANALYSIS

The complete p23 genes of eight RSA CTV isolates were sequenced. These sequences are not representative of all the strains/variants in each sample, but only a certain variant. Phylogenetic analysis of the p23 gene of the eight RSA isolates and 18 biologically indexed and sequenced isolates of Sambade *et al.* (2003) supported the grouping pattern of Sambade *et al.* (2003). All CTV isolates clearly separated into three groups: mild (group 3); severe (group 1); and the atypical group (group 2) consisting of isolates showing variable pathogenicity and sequences less closely related to one another than to those sequences within the mild and severe groups (intra-group divergence). All the reference sequences in the atypical group induced seedling yellows except isolate K. However, isolate K has been shown to contain a mixture of variants some of which cause seedling yellows and/or stem-

pitting in grapefruit (Brlansky *et al.*, 2003). There are no records of T36 causing any stem-pitting only decline and mild seedling yellows in Duncan grapefruit and sour orange. The isolates from RSA: 12-5, 12-7, 12-9 and GFMS 12 grouped with the severe isolates; 390-5 grouped together with the atypical isolates; and 389-3, 389-4 and GFMS35 grouped together with the mild isolates. These groupings were supported by the intra-group and inter-group diversity scores. With group 2 sequences more highly diverse compared to group 1 and group 3 sequences. On the phylogenetic tree group 2 isolates do not seem to group clearly together.

Within the severe group RSA isolates 12-9 and GFMS 12 clustered very closely to the severe VT and isolates 12-5 and 12-7 clustered closely to T388 and T305 severe isolates. Isolates in the severe group can be separated into two sub-groups on the tree: (A) 12-9, GFMS12 and VT; and (B) 12-5, 12-7, T388, T305, Cal-CB, Barao, and 269-6. The RSA severe isolates cluster with one of these two main sequence types: those similar to isolates VT or T305.

The sequences of isolates 12-7 and 12-5 are identical and could be from a single aphid transmission in which the exact strain was selected from the mixture of strains of GFMS12 or is a different variant with a conserved p23 gene sequence. Isolates 12-9 and GFMS 12 also had a high homology with defective RNA sequences (D-RNAs). D-RNAs are variable between isolates and are suggested to correlate to Seedling yellows (SY) symptom expression (Garnsey *et al.*, 2000). Isolate GFMS 12 is a mixture of strains used as a mild-strain cross-protecting source. Sub-isolates made from this source have shown varying degrees of stem-pitting (not included in this study) (van Vuuren *et al.*, 2000).

The severe variants found in these RSA isolates could be dominant sequence types which could be easily transmitted via aphids and easily replicated in the host to reach high titers or are a more ideal template for primer annealing. Interestingly with isolates 12-7, 12-9 and GFMS 12 the severe component was selected over the mild form by PCR bias. Such findings are important for studies on strain dynamics. The bio-indexing reported for the RSA isolates did not show severe symptoms and therefore no differences in their pathogenicity or type of symptom expressed. It could be that severe components under standard conditions are not expressing any severe symptom and is only expressed under certain conditions. Possibly the dynamics of the strains within these isolates result in mild symptoms and if the dynamic changes possible severe symptoms could result. It could

also mean that these determinants of “severe” may not always be true and that there is no correlation between composition of amino acid positions 78-80 and symptoms.

Within the atypical group RSA isolate 390-5 clustered together with c268-2. These isolates together with c270-3 from Argentina and Galego-50 formed a separate subgroup in the atypical group (group 2) and induce seedling yellows (SY) and stem-pitting (SP). The sequence of isolate 390-5 had a high homology to isolates C315-14 and c271-8 which were not included in this study because of their unknown biological symptoms. Clones of isolate C315-14 are from a grapefruit source with a population of strains selected for pre-immunisation assays in Argentina (Iglesias, unpublished). Isolate c271-8 was also from Argentina. There could be a correlation between RSA isolate 390-5 and isolates from Argentina. Bio-indexing of 390-5 has shown only mild symptoms. It is also possible that a strain exists in 390-5 capable of producing seedling yellows and stem-pitting under certain conditions or under differential strain dynamic pressure. However it could also be that these aren't true determinants for SY and SP. Isolates T36 and K formed their own branches distinctly separate from the other isolates but still within the atypical group and appear to be quite different from the other atypical isolates and from one another. There is however no definite correlation that can be made between the sequences and the symptoms of these atypical isolates and to their positions in the tree. However it does seem that isolates capable of inducing SY and SP have clustered together and isolate T36 which causes decline and K (generally known to be symptomless) have all clustered separately according to possibly their different symptoms. This point would have to be further exploited with more sequence from isolates with known symptoms.

Within the mild group (group 3) of isolates there were three distinct sub-groups. RSA isolates GFMS35, 389-3 and 389-4 clustered together in a subgroup on their own and isolates T312, T300, T32, T385 and T55 formed their own subgroup. These two subgroups were more closely related than either was to the last mild isolate T346 which formed its own subgroup more distant to the other two. Like isolate GFMS 12, GFMS 35 is a mixture of strains used in the cross-protection scheme. The sequences of RSA isolates 389-3, 389-4 and GFMS35 had a high homology. Isolates 389-3, 389-4 and GFMS 35 sequences were most similar to isolates 464-2, 464-1, 425 and 81P, originally from transmissions of a naturally infected tree in California (Roy., unpublished). The biological properties of these

isolates have not been documented and were therefore not included in this phylogenetic study.

It appears that the RSA isolates are unique compared to other mild isolates and have formed their own sub-group. It is possible that the mild strains in South Africa have evolved separately from mild strains from other countries, possibly due to different evolutionary pressures and levels of these pressures. A question that arises from this observation is: Are all mild strains capable of cross-protecting trees with the same efficiency? Maybe there are differences within mild strains that make some more suitable as cross-protecting isolates and possibly a certain sequence type mild strain might be a better competitor and prevent segregation with other strains and influence the dynamics of strains in a more effective cross-protective manner as shown by (van Vuuren *et al.*, 2000) . Maybe only certain mild strains have the potential to recombine with more severe strains as was explored by Vives *et al.* (1999). All these points would be important in selecting mild strains suitable for the cross-protection scheme.

Previous work done by Sambade *et al.* (2003) failed to detect recombination within sequences closely related to the core of each group: T32, T55, T300, T312 and T385 (mild group); T305, T388, C269-6 and Cald-CB (severe group); and C270-3 and C268-2 (atypical group). While recombination was shown for: mild group (T346); severe (VT) and atypical (T36, Galego-50 and K). This finding supported the existence of three sequence types in the isolates studied and also explained the topology of the phylogenetic tree, in which the sequences more distantly related to the core of each group are recombinant. Recombination in other regions of the CTV genome has been reported (Rubio *et al.*, 2001, Vives *et al.*, 1999). This finding has relevance to the RSA isolates 12-9 and GFMS12 which show a high similarity to the VT isolate, a recombinant. This raises concerns that such isolates have the potential to recombine and possibly become more severe with devastating effects. These isolates could also show varying degrees of symptoms under different host and environmental pressures and become quite unpredictable in cross-protection.

There was a good correlation between designating RSA isolates 390-4, 389-3, 389-4 as mild; 12-5 as severe; and 390-3, 390-5 as atypical in groups based on the PCR results which targeted differences at amino acids 78-80 and phylogenetic analysis of the complete p23 gene. RSA isolates 12-7, 12-9 and GFMS 12 had mixed mild and severe strains; and

GFMS 35 had mixed atypical and mild strains. However for routine differentiation of strains based on the p23 gene it would only be necessary to use the PCR system since DNA sequencing and phylogenetic analysis would be too cumbersome and expensive. It was very useful however to characterise the RSA isolates in determining the variability and correlation to other sequences around the world and to compare the PCR results with the DNA sequence analysis. The p23 gene however only represents a small portion of the complete genome and therefore this system should be used in conjunction with other systems targeting other regions of the genome to get a complete picture of an isolate's sequence type before assuming any definite answer as to the strain type.

2.5 REFERENCES

- Albertini, D., Vogel, R., Bové, C., Bové, J.M. 1988. Transmission and preliminary characterization of *Citrus tristeza virus* strain K. In "Proceedings 10th Conference of the International Organization of Citrus Virologists" (L.W. Timmer, S.M. Garnsey, L. Navarro, Eds), pp 17-21, IOCV. Riverside, California..
- Ayllón, M.A., López, C., Navas-Castillo, J., Garnsey, S.M., Guerri, J., Flores, R., Moreno, P. 2001. Polymorphism of the 5' terminal region of *Citrus tristeza virus* (CTV) RNA: Incidence of three sequence types in isolates of different origin and pathogenicity. *Arch Virol* 146: 27-40.
- Brlansky, R.H., Damsteegt, V.D., Howd, D.S., Roy, A. 2003. Molecular analyses of CTV sub-isolates separated by aphid transmission. *Plant Disease* 87: 397-401.
- Che, X., Mawassi, M., Bar-Joseph, M. 2002. A novel class of large and infectious defective RNAs of *Citrus tristeza virus*. *Virology* 298: 133-145.
- Dolja, V. V., Karasev, A.V., Koonin, E.V. 1994. Molecular biology and evolution of closteroviruses. Sophisticated build up of large RNA genomes. *Annu. Rev. Phytopathol.* 32: 261-285.
- Febres, V. J. 1995. Ph.D. Thesis, p 104, University of Florida, Gainesville, FL.

Garsney, S.M. Timmer, L.W., Graham, J.H. 2000. Compendium of Citrus Diseases 2nd ed. APS Press. USA.

Ghorbel, R., López, C., Fagoaga, C., Moreno, P., Navarro, L., Flores, R., Pena, L. 2001. Transgenic citrus plants expressing the *Citrus tristeza virus* p23 protein exhibit viral-like symptoms. *Mol. Plant Path.* 2: 27-36.

Gillings, M., Broadbent, P., Indsto, J., Lee, R.F. 1993. Characterisation 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.

Hilf, M.E., Garnsey, S.M. 2000. Characterization and Classification of *Citrus tristeza virus* isolates by amplification of multiple molecular markers. In “Proceedings 14 th Conference of the International Organization of Citrus Virologists” (J.V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp 18-27, IOCV. Riverside, California.

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

Hilf, M.E., Karasev, A.V., Albiach-Marti, M.R., Dawson, W.O., Garnsey, S.M. 1999. Two paths of sequence divergence in the *Citrus tristeza virus* complex. *Phytopathology* 89: 336-342.

Hilf, M.E., Garnsey, S.M. 2002. *Citrus tristeza virus* in Florida: A synthesis of historical and contemporary biological, serological and Genetic data. In “Proceedings 15 th Conference of the International Organization of Citrus Virologists” (N. Duran-Vila, R.G. Milne, J.V. da Graca, Eds), pp 13-20, IOCV. Riverside, California.

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., Dawson, W.O. 1995. Complete sequence of the *Citrus tristeza virus* RNA genome. *Virology* 208: 511-520.

Kumar, S., Tamura, K., Nei, M. 1993. MEGA. Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park.

López, C., Navas-Castillo, J., Gowda, S., Moreno, P., Flores, R. 2000. The 23-kDa protein coded by the 3'-terminal gene of *Citrus tristeza virus* is an RNA-binding protein. *Virology* 269: 462-470.

Marais, L.J., Marais, M.L., Rea, M. 1996. Effect of tristeza stem-pitting on fruit size and yield of Marsh grapefruit in Southern Africa. In "Proceedings 13 th Conference of the International Organization of Citrus Virologists" (J.V. da Graca, P. Moreno, R.K. Yokomi, Eds), pp 163-167. IOCV. Riverside, California.

Mawassi, M., Mietkiewska, E., Hilf, M.E., Ashoulin, L., Karasev, A.V., Gafny, R., Lee, R.F., Garnsey, S.M., Dawson, W.O., Bar-Joseph, M. 1995. Multiple species of defective RNAs in plants infected with *Citrus tristeza virus*. *Virology* 214: 264-268.

Pappu, S.S., Febres, V.J., Pappu, H.R., Lee, R.F., Niblett, C.L. 1997. Characterization of the 3' proximal gene of the citrus tristeza closterovirus genome. *Virus Res.* 47: 51-57.

Pappu, H.R., Pappu, S.S., Cambra, M., Moreno, P., Garnsey, S.M., Lee, R.F., Niblett, C.L. 1994b. Molecular characterization of a structural epitope that is largely conserved among severe isolates of a plant virus. *Proc. Natl. Acad. Sci. USA* 90: 3641-2644.

Ruiz-Ruiz, S., Moreno, P., Guerri, J., Ambros, S. 2006. The complete nucleotide sequence of a severe stem pitting isolate of *Citrus tristeza virus* from Spain: comparison with isolates from different origins. *Arch. of Virol.* 151: 387-398.

Sambade, A., López, C., Rubio, L., Flores, R., Guerri, J., Moreno, P. 2003. Polymorphism of a specific region in gene p23 of *Citrus tristeza virus* allows discrimination between mild and severe isolates. *Arch. Virol.* 148: 2325-2340.

Satyanarayana, T., Gowda, S., Ayllón, M.A., Albiach-Martí, M.R., Rabindran, S., Dawson, W.O. 2002. The p23 protein of *Citrus tristeza virus* controls asymmetrical RNA accumulation. *J. Virol.* 76: 473-483.

Thompson, J. D., Higgins, D.J., Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

van Vuuren, S.P., Collins, R.P. 1993. Evaluation of *Citrus tristeza virus* isolates for cross-protection of grapefruit in South Africa. Plant Disease 77: 24-28.

van Vuuren, S.P., van der Vyver, J.B. 2000. Comparison of South African pre-immunizing *Citrus tristeza virus* isolates with foreign isolates in three grapefruit selections. In “Proceedings 14 th Conference of the International Organization of Citrus Virologists” (J.V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp 50-56, IOCV. Riverside, California..

van Vuuren, S.P., van der Vyver, J.B. Luttig, M. 2000. Diversity among sub-isolates of cross-protecting *Citrus tristeza virus* isolates in South Africa. In “Proceedings 14 th Conference of the International Organization of Citrus Virologists” (J.V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp 103-110, IOCV. Riverside, California.

CHAPTER 3
ESTABLISHMENT OF 23 PCR PRIMER SYSTEMS
TARGETING THE 5' END OF CTV FOR STRAIN
DIFFERENTIATION OF RSA ISOLATES INTO 4
GENOTYPES

3.1 INTRODUCTION

Isolates of CTV have been traditionally characterised and differentiated from one another by the comparison of symptoms that develop in inoculated assay plants. Decline and mild strains can be differentiated using an ELISA test with monoclonal antibody MCA13 (Permar *et al.*, 1990). Most molecular characterization of isolates has focused upon capsid protein gene sequences and 5'-terminal sequences but genome fragments in different parts of the genome have increasingly been used to classify CTV isolates by PCR amplifications, sequencing, SSCP, hybridisation, and other techniques. However the comparison of sequences of a single gene or region may reflect differences for just that region, but this region is unlikely to be representative of the entire genome. It would be better to compare complete genomes to obtain a better understanding of the molecular relatedness of isolates. It is however impractical to sequence the complete genome of the 19 kb CTV genome for routine analysis. An alternative is to target as many distinct molecular markers across the genome as possible.

The comparison of fully sequenced genomes revealed that CTV isolates are most divergent in the 5' proximal eleven kilobases of the genome and much more conserved in the 3' terminal eight kilobases of the genome (Ayllón *et al.*, 2001). The 5' half of the genome consists of a 107 nucleotide 5' UTR and the p349 polyprotein which consists of domains for: two cysteine papain-like leader proteases (P-Pro), a methyltransferase (MT), and a helicase (HEL). These proteins are part of ORF1a and are replication related proteins.

In this study a “genotype” is a genetic constitution (the genome) of a viral particle. The genotype is distinct from its expressed features (phenotype). This is different from a “strain” which is dependent on the biological symptom.

Studies assessing the sequence diversity in the 5' terminal region of CTV from India found that 19 isolates were a VT genotype, five were a T30 genotype and one was a T36 genotype (Roy *et al.*, 2005). Phylogenetic analysis of the sequenced products revealed 4 groups: VT, T30, T36 and a fourth represented by BAN-2 (Roy *et al.*, 2005). More variability was found between nucleotides 1082 and 1484 than between 1-50 or 697-1105 (Roy *et al.*, 2005). Variability observed by López *et al.* (1998) in the 5' UTR and ORF1a separated sequences into three groups, which represented the sequences of T36, T317 and VT. Some isolates

yielded clones that were placed into multiple groups illustrating the presence of different sequence variants in infected plants (López *et al.*, 1998). Four conserved regions were found within the 284 nucleotides in the 5' terminal region and were used in the development of four sets of primers (Ayllón *et al.*, 2001). The results confirmed the presence of three types of CTV populations from many countries from which the samples were taken.

Molecular markers developed by Hilf *et al.* (2000, 2002) are currently being used for CTV isolate classification and are derived from the 5' divergent regions of the genomes of Florida isolates T3, T30, T36 and of the Israeli isolate VT. The sequence specific molecular markers create a specific marker pattern termed the isolate genotype designation. The specific genotype designation (T3, T30, T36 and VT) is assigned to the sample based on the presence of RT-PCR genotype-specific amplifications of those markers. A few scientific papers have described studies using the molecular marker PCR systems of Mark Hilf (Hilf *et al.*, 1999; 2000; 2002; Roy *et al.*, 2003; 2004; 2005) and have been intensively optimised over the last few years. The four genotypes are: mild T30 (AF260651) isolate; the severe T36 (U16304) isolate causing quick decline; and severe VT (U56902) isolate causing seedling yellows and mild stem-pitting. The origin of the T3 isolate is unclear and is not typical of field isolates generally found in Florida, USA (Hilf, personal communication). T3 is believed to have been introduced into Florida through illegal importing channels (Hilf, personal communication). T3 is typical of field isolates in Brazil and Asian countries and it has been hypothesized that T3 was introduced into Brazil in mandarin propagations, which were probably asymptomatic. This theory also suggests a possible Asian origin. Sequence and hybridization data indicated that T3 was distinct from T30, T36 and VT genotypes.

In this study the aim was to establish and optimise the previously developed 23 primer sets (Hilf *et al.*, 2000) targeting eight different regions of the 5' half of the genome for the four specific genotypes (T3, T30, T36 and VT) and to determine the level of sequence variability in isolates from South Africa. This process assigns isolates a specific genotype(s).

3.2 MATERIALS & METHODS

3.2.1 CTV ISOLATES

A total of 11 CTV sources were used in this study and all are from single aphid transmissions performed to generate isolates with potential for mild strain cross-protection in South Africa (van Vuuren, Breytenbach, unpublished). The single aphid transmissions produced sub-isolates (12-5, 12-7, and 12-9) obtained from a GFMS 12 source in Nelspruit, RSA by S.P. van Vuuren. The original source of the GFMS 12 strain mixture was a 78 year old grapefruit tree (2004) still in production and referred to as Nartia A, Worcester, Western Cape (S.P. van Vuuren, personal communication). GFMS 12 has been used previously for cross-protection in all cultivars (grapefruit, sweet oranges, mandarins) but is currently only used to protect white (Marsh, Nartia) grapefruit (*Citrus paradisi*) and pummelo (*C. grandis*) (S.P. van Vuuren., personal communication).

The three sub-isolates 390-3, 390-4 and 390-5 are single aphid transmissions made from the Mouton isolate. This original isolate was mild but contained citrus viroids (S.P. van Vuuren, personal communication). Sub-isolates 389-3 and 389-4 were derived from the Nartia C isolate (GFMS 14) (S.P. van Vuuren., personal communication). The original isolates (Mouton and Nartia C) currently are not used as cross-protecting isolates (S.P. van Vuuren, personal communication) and are currently being tested in field experiments. The Mouton and Nartia C isolates were also from 78 year old grapefruit trees still in production Worcester, Western Cape (S.P. van Vuuren, personal communication). These five single aphid transmissions were prepared at the quarantine facility in Frederick, Maryland, USA and imported back to South Africa.

A plant infected with the reference T30 strain obtained from Nelspruit, RSA as well as two plants inoculated with the South African mild strain cross-protecting sources GFMS 12 and GFMS 35 were also used in this study. GFMS 35 is currently being used to cross-protect red grapefruit selections. Full length clones of T30 and T36 were kindly supplied by Dr. S. Gowda, USA. The VT number 40 clone consisting of 5' end up to 11.8 kb of the genome was supplied by Dr. O. Batuman and Dr. M. Bar-Joseph (Israel). The T3 R1-5 clone consisting of 1-11833 nucleotides was supplied by J.Fisher, USA. These DNA clones T36, T30, VT and T3 were used as positive controls in this study.

3.2.2 TAS-ELISA

The TAS-ELISA (Triple Antibody Sandwich Enzyme linked Immunosorbent Assay) was performed to determine the CTV titers using CTV polyclonal antisera CTV 1052 (from isolate T36 whole, unfixed, purified virus) with the methods outlined by Rocha-Peña and Lee (1991). The virus specific antiserum and the antibodies were kindly supplied by Richard Lee. Sterile flat bottom (Immulon) 96-well microtiter plates were incubated with CTV 1052 polyclonal antiserum (1:5000) in coating buffer (1.59 g/l Na_2CO_3 ; 2.93 g/l NaHCO_3 at pH 9.6) for 4 hours at 37°C. Between each incubation step, plates were washed three times for three minutes each with phosphate-buffered saline with Tween 20 (8g/l NaCl; 0.2 g/l KH_2PO_4 phosphate; 1.15 g/l Na_2HPO_4 ; 0.2 g/l KCl at pH 7.4, with 0.1% Tween 20). From each scion, fresh bark was collected near fresh-mature leaves and cut into 1mm sections and approximately 0.25 g from each test sample was pulverized in 5ml coating buffer for 1 min using a tissue homogenizer. The resultant sap for each sample was added to triplicate test wells on the antibody coated microtiter plates. Incubation for antigens was at 4°C overnight. The detecting antibody, G604 (E.coli expressed coat protein of B-227, a severe isolate of CTV from India) in conjugate buffer (PBST plus 2% polyvinylpyrrolidone-40, and 0.2% bovine serum albumin) at a dilution of 1:20,000, was added and incubated at 37°C for 4 hours. Rabbit anti-goat antibody IgG conjugate with alkaline phosphatase (Sigma A-4187) was added at a 1:30 000 dilution in conjugate buffer and incubated for 3 hours at 37°C. The substrate, 4-nitrophenyl phosphate hexahydrate was added to coating buffer and the hydrolyzed enzyme substrate extinction values were collected at an absorbance of 405 nm during the reaction, using a Multiskan Ascent V1.24 plate reader. The data represent three separate duplicated experiments with uncoated wells, CTV infected citrus controls and extraction buffer controls included in each plate.

For the quantitation of the CTV titer in different plant components: TAS-ELISA was carried out as described above except samples were collected from infected plants in three different parts: (1) The midribs and petioles of fresh fully expanded leaves from the last growth flush; (2) fresh and dark green bark from new twigs; and (3) brown bark from old lignified stems.

Total RNA was extracted from Mexican lime plants infected with CTV isolates as described above using the SV Total RNA Isolation System (Promega, USA) according to the manufacturer's protocol. The total RNA was eluted with 25 µl of nuclease-free water and eluted through the column twice. RNA samples were extracted from virus-free citrus plants to serve as a healthy citrus plant control. Samples extracted with this method are depicted in Appendix 2 (table 19).

3.2.4 IMMUNOCAPTURE

Thin walled 0.2 µl PCR tubes were incubated with 200 µl of CTV1052 polyclonal antiserum (1:5000) in coating buffer (1.59 g Na₂CO₃; 2.93 g NaHCO₃ at pH 9.6) for 4 hours at 37°C. The tubes were washed three times for three minutes each with phosphate-buffered saline with Tween 20 (PBST) (8g NaCl; 0.2 g KH₂PO₄ phosphate; 1.15 g Na₂HPO₄; 0.2 g KCl at pH 7.4, with 0.1% Tween 20). From each scion, fresh bark was collected near fresh-mature leaves and cut into 1mm sections and approximately 0.25 g from each test sample was pulverized in 5ml coating buffer for 1 min using a tissue homogenizer. The resultant sap for each sample was added to the antibody coated tubes. Incubation was at 4°C overnight. The tubes were washed with PBST as described above and used directly for cDNA synthesis. Samples extracted with this method are depicted in Appendix 2 (table 19)

Duplicate tubes of each sample were made and the final steps of a standard TAS-ELISA were performed directly in the tube to confirm that virions were successfully captured.

3.2.5 cDNA SYNTHESIS

CTV specific cDNA was made with a two-step RT-PCR method using 3 µg of the total RNA extract from each CTV sample. Reverse primers listed in Table 7 were used to make cDNA and are specific for one of the four strain genotypes (T30, T36, T3, and VT).

Samples were heated at 65 °C for 15 minutes, 55 °C for 10 minutes and then at room temperature for 5 minutes. This was added to 25 µl of reaction mixture containing 50 pmol a reverse primer (IDT, USA) (Table 7), 5U of RNasin (Promega, USA), 10 U of Avian *myeloblastosis virus* reverse transcriptase (Roche, Germany), 1x AMV RT-Buffer (50 mM Tris-HCl, 8mM MgCl₂, 30 mM KCl, 1 mM DTT; pH 8.5) (Roche, Germany), 0.2 mM each of

dATP, dCTP, dGTP, and dTTP. The contents were gently mixed then incubated at 47°C for 1 hour on an Eppendorf Gradient Mastercycler (Eppendorf, Germany), after which 12 µl of nuclease-free water was added and the cDNA stored at -20°C.

3.2.6 PCR & GEL ELECTROPHORESIS

The 23 primer sets were all designed by Mark Hilf and developed based on currently available sequences of T30 (AF260651); T36 (U16304) and VT (U56902) on the NCBI database. The T3 sequence is presently not available on the NCBI Database but has been sequenced by Mark Hilf.

The 23 primer sets used in this study were all individually optimised (Appendix 2, table 19). Previous work to optimise the primer sets (Hilf, personal communication) was used as a guideline. Each of the PCR reactions contained 5 µl of the cDNA reaction. A master mix of 100 pmoles of each of the forward and reverse primers (IDT, USA) (Table 7); 1.5-2.0 mM MgCl₂ (Appendix 2, table 17); 1x Gotaq Buffer (Promega, USA); 2.5 U Gotaq (Promega, USA); 0.14 mM each of dATP, dCTP, dGTP and dTTP and 2 µg/µl BSA in 50 µl total volume was made. The contents were gently mixed. The thermal cycle conditions were 92°C for 2 min, 25-35 cycles (depending on system, Appendix 2, table 17) of 92°C for 30 sec, 45-55°C for 45 sec, and 72°C for 1 min, then extension for 10 min at 72°C, and held at 4°C on an Eppendorf Gradient Mastercycler (Eppendorf, Germany).

Each PCR reaction was duplicated. cDNA from a citrus virus-free plant sample was included as a healthy control and to check any non-specific amplification of plant components. A negative control consisting of nuclease-free water instead of cDNA was included as a control to assess contamination and false positives. PCR products (5-50 µl) were analyzed by gel electrophoresis through 1.0% agarose (Whitehead Scientific) in sodium borate electrophoresis buffer (5 mM disodium borate decahydrate, adjusted to pH 8.5 with boric acid), stained with ethidium bromide (0.5 µg/ml), and photographed under UV light.

Table 7

PCR primer sequences of the four different strain genotypes (T30, T36, VT, and T3)
(Hilf *et al.*, 2000) used for RT-PCR amplification.

T30 Sequence Specific Primer Pairs				
Isolate	*Position	Sense	Sequence	Product(bp)
T30-1(+)	22	+	GTATCTCCGGAGCTCGATC	570
T30-1(-)	592	-	CAGTAGGGTCAACTAGTTTGC	
T30-2(+)	792	+	TACGGCTTGGTGCTCTGAGGCC	843
T30-2(-)	1635	-	ACGCCTGCGAACCGCCGAC	
T30-3(+)	2267	+	TGGATGAGGGTTCTTCACCAC	824
T30-3(-)	3091	-	AGAATCGGGCAAAGCTTT	
T30-4(+)	3766	+	GTTTACTGCTTTAACAATTCGGC	530
T30-4(-)	4296	-	GTATCTCCTGAAAAGGCAGCC	
T30-5(+)	4731	+	TGGTTTACGTGCCGGAGCCGC	845
T30-5(-)	5576	-	TCGCACTATACGTGATCA	
T30-6(+)	6012	+	GCGTGTCTTAGTGTATCGCA	733
T30-6(-)	6745	-	TGGAAACAAAGGACTGTT	
T30-7(+)	7356	+	CAGCGTTCAAGTTTGCGTTA	913
T30-7(-)	8269	-	GACACTTTAGTACAATAATC	
VT Sequence Specific Primer Pairs				
VT-1(+)	19	+	GTACCCTCCGGAAATCACG	564
VT-1(-)	583	-	GGTAGGGTCTACTCGTTTCAT	
VT-2(+)	783	+	AACGGCATGGTGCTCTCCTGTT	834
VT-2(-)	1617	-	TTCGCCTGCGCAGCTGCT	
VT-3(+)	2246	+	CAGGTGAGAATTCTCCATCGT	824
VT-3(-)	3070	-	AGAATCAGGCAAACGCCC	
VT-4(+)	3745	+	ATCTACTGTTTTAACAATTCACG	530
VT-4(-)	4275	-	GGGTCGCCTGAAAAGGCCCGT	
VT-5(+)	4710	+	TGGTTTGCGGGCCGGGCG	842
VT-5(-)	5552	-	TTCGCGCTACACGTGTTA	
VT-6(+)	5985	+	ACGCGTTTCGGTATGTTGTA	733
VT-6(-)	6718	-	TGGAACAAAAGGACTATC	

Table 7 (continued)

T36 Sequence Specific Primer Pairs				
T36-1(+)	68	+	AGCCTTTAAGCTCTAATA TT	594
T36-1(-)	662	-	ACCAAGTCGGCTGTTTCGTC	
T36-2(+)	855	+	AAACTGATTTCTCCACTCAG	763
T36-2(-)	1618	-	ACAATCGAGCCAGGAACACTG	
T36-3(+)	2323	+	CTTCTTTTAACTCGACAAGGA	739
T36-3(-)	3062	-	TGTGATTATCAGGGAGTTA	
T36-5(+)	6040		CGGTAGCGTGTTGTGAGTACG	718
T36-5(-)	6760		GACGAAGAAACCTCGTTCGAC	
T36-6(+)	7370	+	ATTCCCTAATTCTAGGGCTC	597
T36-6(-)	7967	-	CTTTCTATCGAAACCTGCGAC	
T36-7(+)	8405	+	GTATGTTACCGATGCAGCTTC	668
T36-7(-)	9073	-	CTAGTAAGTACGGAATAAACC	
T3 Sequence Specific Primer Pairs				
T3-2 (+)	962	+	GTGTTGAGGTCCCGAGCGTC	652
T3-2 (-)	1614	-	GATCGAGACGGTTTAGAGATG	
T3-3 (+)	3708	+	GTTCGGTGGAGTTGGACGTTA	934
T3-3 (-)	4663	-	TCGTCCGAACCGTCACCGTCTG	
T3-5 (+)	6133	+	TCCTTTGCCATCAATTGTATCAC	397
T3-5 (-)	6530	-	CACGTGGAAAGTTCCACGACG	
T3-6 (+)	8847	+	GGGTCGGACTAAAGCAGTA	649
T3-6 (-)	9496	-	GGCCATAGCCTTACCGAAATC	

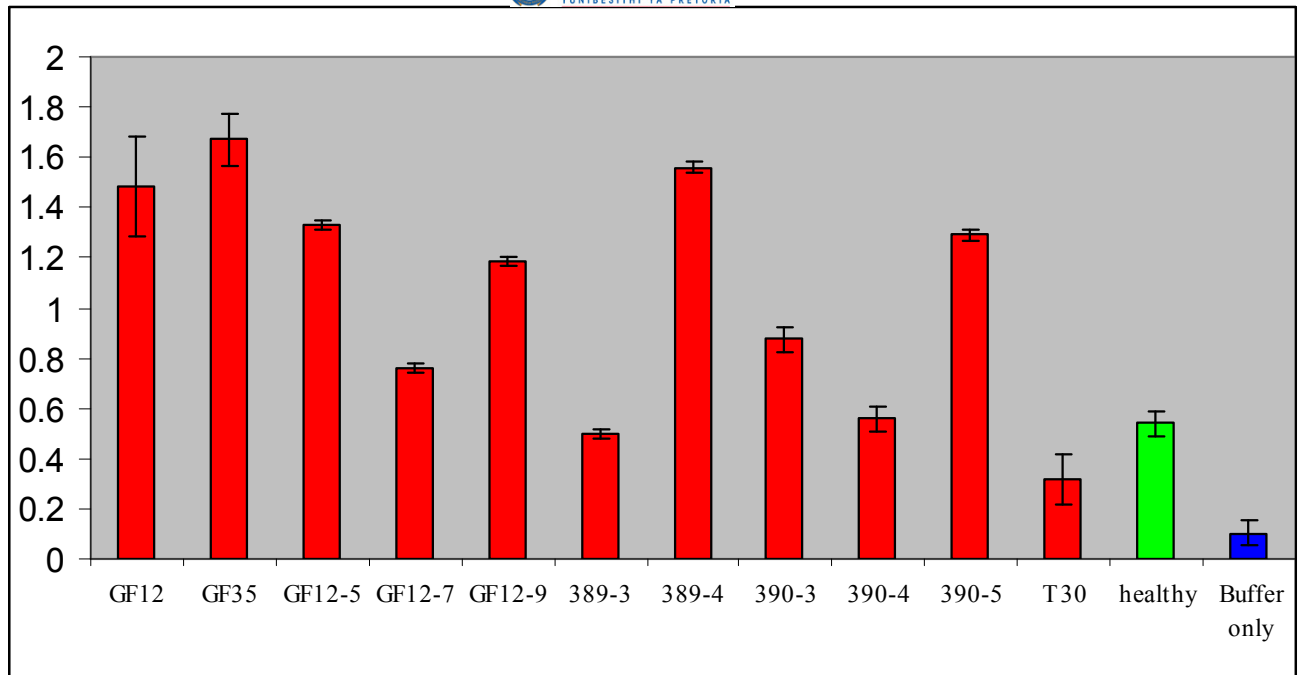
* Nucleotide position on each reference sequence T30 (AF260651); T36 (U16304); VT (U56902) and T3.

3.3 RESULTS

3.3.1 ELISA

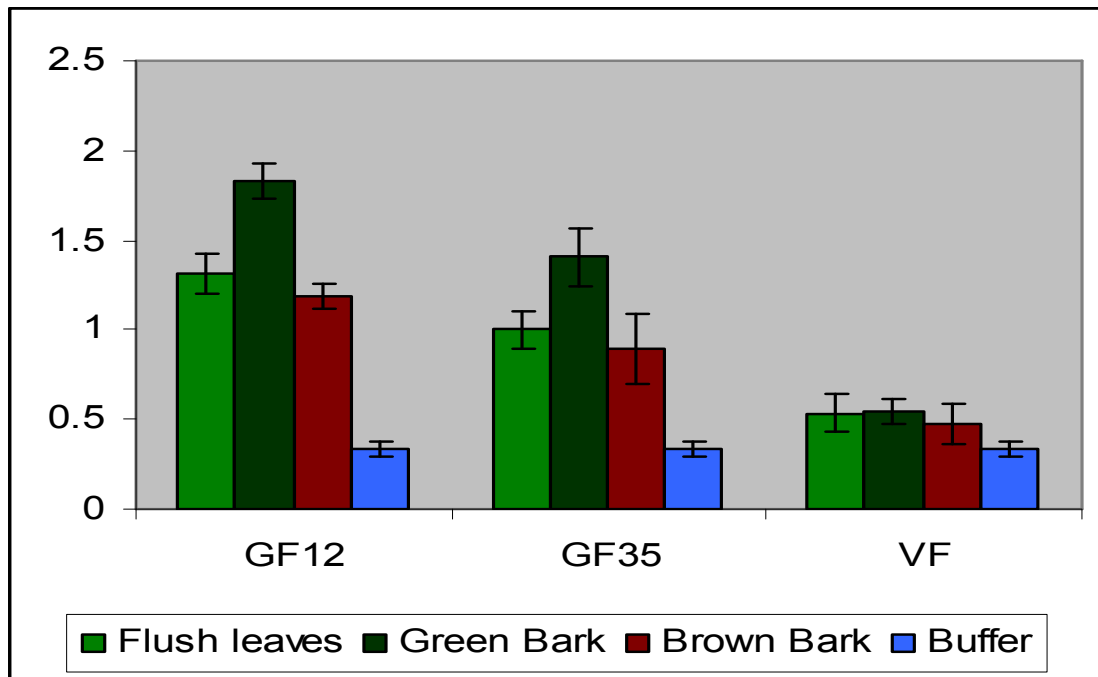
Eleven RSA isolates were tested using the TAS-ELISA system with CTV polyclonal antibody 1052 (figure 10). The samples were tested in triplicate and had acceptable variation. Isolates GFMS12, GFMS35, 12-5, 12-9, 389-4 and 390-5 were positive with an absorbance value 2-3 x that of the healthy control. Isolates 12-7, 389-3, 390-3, 390-4 and T30 had values similar to those of the healthy control. The negative, virus free and uncoated well controls had low values of between 0.1-0.5. GFMS 12 and GFMS 35 were used as positive controls as they have been inoculated with the cross-protective mild strains, showed mild symptoms and are positive for CTV.

Three plant components (newly-flushed leaves, brown lignified bark and green bark) of two isolates were tested with TAS-ELISA to determine which plant part had the overall highest titer for use in RNA extractions /RT-PCR use (Figure 11). The samples were tested in triplicate for greater reliability. Variability amongst replicates was acceptable. The highest absorbance values were obtained with green bark of isolates GFMS 12 and GFMS35 of 1.8 and 1.4 respectively (Figure 11), while the lowest values were with brown lignified bark (1.2 and 0.9 respectively). Absorbance values of newly-flushed leaves were intermediate (1.3 and 1.0 respectively). As expected little difference between plant parts were obtained for the virus free and buffer controls.



Standard error bars are depicted above each sample bar.

Figure 10: CTV Detection of 11 RSA isolates by TAS-ELISA with CTV specific polyclonal antibodies.



Standard error bars are depicted above each sample bar.

Figure 11: Titer determination in three different plant parts of two isolate by TAS-ELISA with CTV specific polyclonal antibodies.

3.3.2 PCR RESULTS

The molecular profiles of each of the 11 South African (RSA) isolates were obtained using the 23 primer sets developed by Hilf *et al.* (2000) and were compared to the standard genotypes of isolates T36 (Florida, decline), T30 (Florida, mild isolate), T3 (Florida, severe isolate) and VT (Israel, severe isolate). The 23 genotype specific primer sets encompass the 5' half of CTV genome. There are eight specific areas targeted within this region.

The amplified products were of the expected product size (Table 7). The gel photos representing amplifications with the genotype specific primer sets of the genotypes T30, T36, VT and T3 are depicted in Appendix 2, figures 37-59). The genotype-specific positive controls T30, VT, T36 and T3 successfully yielded all their expected products. No cross-amplifications of genotypes occurred. The virus free and buffer controls did not have any amplification. A summary of the amplifications of each genotype are shown in Table 8. The summary of the RSA isolate's molecular profiles from amplification/s in regions 1-8 are depicted in Figures 12-16 and briefly described below:

- Region 1 is from nucleotides (nt) 19-583 of the 5'-UTR and part of the p349 gene. There are three genotype specific primer sets: T30 1+/1- (22-592 nt); VT 1+/1- (19-583 nt) and T36 1+/1- (68-662 nt).
- Region 2 is from 783-1614 nt of the 5' end of the p349 gene. There are four genotype specific primer sets: T30 2+/2- (792-1635 nt); VT 2+/2- (783-1617 nt); T36 2+/2- (855-1618 nt) and T3 2+/2- (962-1614 nt).
- Region 3 is from 2246-3062 nt of the p349 gene. There are three genotype specific primer sets: T30 3+/3- (2246-3070 nt); VT 3+/3- (2267-3091 nt); and T36 3+/3- (2323-3062 nt).
- Region 4 is from 3708-4663 nt of the central part of the p349 gene. There are three genotype specific primer sets: T30 4+/4- (3766-4296 nt); VT 4+/4- (3745-4275 nt); and T3 3+/3- (3708-4663 nt).
- Region 5 is from 4710-5552 nt of the central part of the p349 gene. There are two genotype specific primer sets: T30 5+/5- (4731-5576 nt) and VT 5+/5- (4710-5552 nt).

- Region 6 is from 5985-6530 nt of the 3' half of the p349 gene. There are four genotype specific primer sets: T30 6+/6- (6012-6745 nt); VT 6+/6- (5985-6718 nt); T36 5+/5- (6040-6760 nt) and T3 5+/5- (6133-6530 nt).
- Region 7 is from 7356-8269 nt of the 3' end of the p349 gene. There are two genotype specific primer sets: T30 7+/7- (7356-8269 nt) and T36 6+/6- (7370-7967 nt).
- Region 8 is from 8405-9496 nt of the 3' end p349 gene. There are two genotype specific primer sets: T3 6+/6- (8847-9496 nt) and T36 7+/7- (8405-9073 nt).

The molecular profiles of isolates 389-3 and GFMS 35 and T30 isolates were identical. Isolate 12-7 was a mixture of T36 and VT genotype molecular markers. Isolates 12-9, GFMS 12 and 390-5 were a mixture of T30, T36 and VT genotype molecular markers. Isolate 390-3 was a mixture of T30 and T36 genotype molecular markers. Isolates 390-4 and 389-4 were a mixture of T30 and T3 genotype molecular markers. The 12-5 isolate contained only the VT genotype molecular markers and the T30, 389-3 and GFMS 35 isolates contained only the T30 genotype molecular markers. The most common genotype was T30 and the least common was T3.

Table 8:

Summarized results of the 23 specific PCR primers sets depicting each isolate's pattern.

Marker	CTV ISOLATE														
	12-5	12-7	12-9	GFMS12	GFMS35	T30	389-3	389-4	390-3	390-4	390-5	T3*	T36*	T30*	VT*
T30-1	0	0	0	1	1	1	1	1	1	0	0	0	0	1	0
T30-2	0	0	0	1	1	1	1	1	0	1	0	0	0	1	0
T30-3	0	0	1	1	1	1	1	0	1	1	1	0	0	1	0
T30-4	0	0	1	1	1	1	1	0	1	0	1	0	0	1	0
T30-5	0	0	0	1	1	1	1	1	0	1	0	0	0	1	0
T30-6	0	0	0	0	1	1	1	1	0	1	0	0	0	1	0
T30-7	0	0	0	1	1	1	1	1	1	0	0	0	0	1	0
T36-1	0	1	1	1	1	0	0	0	0	0	1	0	1	0	0
T36-2	0	1	1	1	0	0	0	0	0	0	1	0	1	0	0
T36-3	0	0	1	1	0	0	0	0	1	0	1	0	1	0	0
T36-5	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
T36-6	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
T36-7	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0
T3-2	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0
T3-3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
T3-5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
T3-6	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
VT-1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	1
VT-2	1	1	1	0	0	0	0	0	0	0	1	0	0	0	1
VT-3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1
VT-4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
VT-5	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1
VT-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

* depicts DNA used as positive controls in PCRs.

1 - depicts a positive amplification with the respective primer set.

0 - depicts no amplification with the respective primer set

Figure 12:

The genotype classification of isolates 384-4 and 390-5 across regions 1-8 of the CTV genome. The 5' half of the CTV genome is represented by the thick black line. The BLUE blocks represent T30 genotype specific amplicons; PINK blocks represent VT genotype specific amplicons; YELLOW blocks represent T36 genotype specific amplicons; and GREEN blocks represent T3 genotype specific amplicons. CLEAR blocks indicate no amplification.

Nucleotides Region Isolate	19-583 1	783-1635 2	2246-3091 3	3708-4663 4	4710-5552 5	5985-6530 6	7356-8269 7	8405-9496 8
389-4	T301+/1- VT 1+/1- T361+/1-	T302+2- VT 2+/2- T36 2+2- T3 2+2-	T303+3- VT 3+/3- T36 3+3-	T30 4+4- VT 4+/4- T3 3+3-	T30 5+5- VT 5+/5-	T30 6+6- VT 6+/6- T36 5+5- T3 5+5-	T30 7+7- T36 6+6-	T36 7+7- T3 6+6-
390-5	T301+/1- VT 1+/1- T361+/1-	T302+2- VT 2+/2- T36 2+2- T3 2+2-	T303+3- VT 3+/3- T36 3+3-	T30 4+4- VT 4+/4- T3 3+3-	T30 5+5- VT 5+/5-	T30 6+6- VT 6+/6- T36 5+5- T3 5+5-	T30 7+7- T36 6+6-	T36 7+7- T3 6+6-

Figure 13:

The genotype classification of isolates 12-5 and 12-7 across regions 1-8 of the CTV genome. The 5' half of the CTV genome is represented by the thick black line. The BLUE blocks represent T30 genotype specific amplicons; PINK blocks represent VT genotype specific amplicons; YELLOW blocks represent T36 genotype specific amplicons; and GREEN blocks represent T3 genotype specific amplicons. CLEAR blocks indicate no amplification.

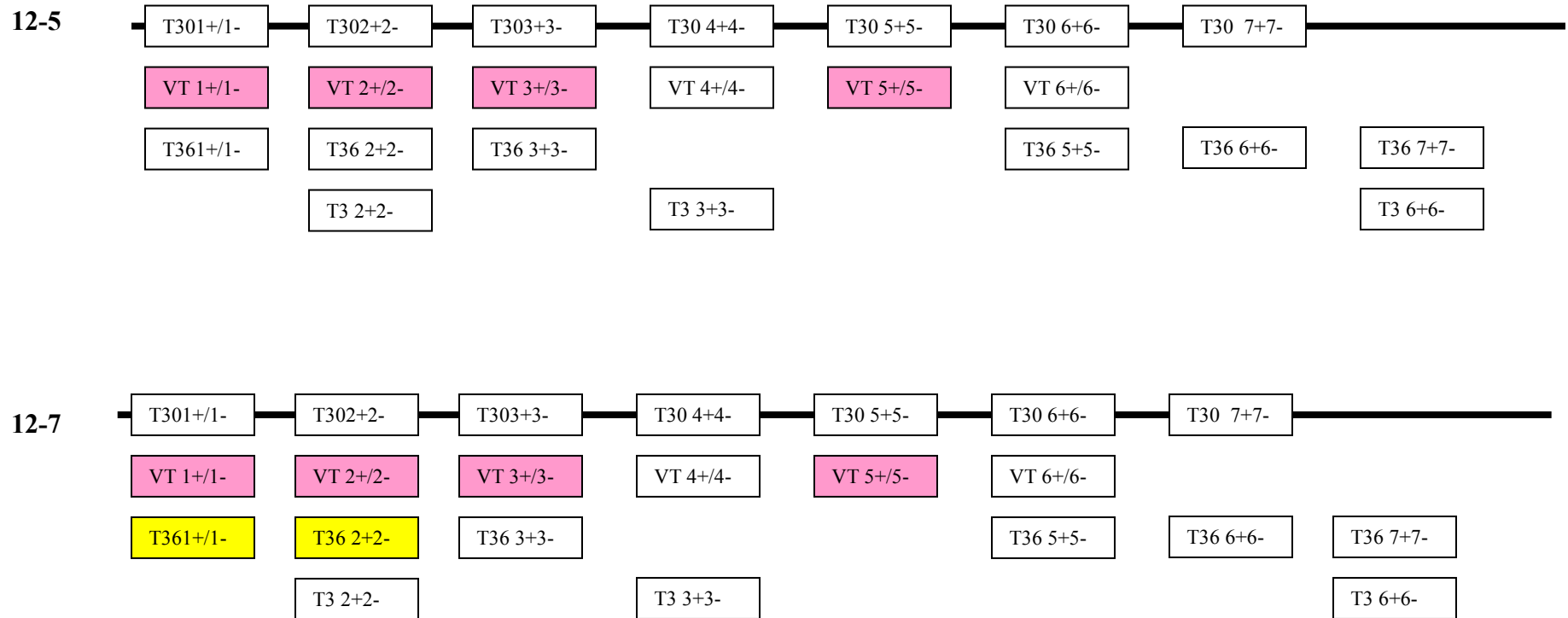


Figure 14:

The genotype classification of isolates 12-9 and GFMS35 across regions 1-8 of the CTV genome. The 5' half of the CTV genome is represented by the thick black line. The BLUE blocks represent T30 genotype specific amplicons; PINK blocks represent VT genotype specific amplicons; YELLOW blocks represent T36 genotype specific amplicons; and GREEN blocks represent T3 genotype specific amplicons. CLEAR blocks indicate no amplification.

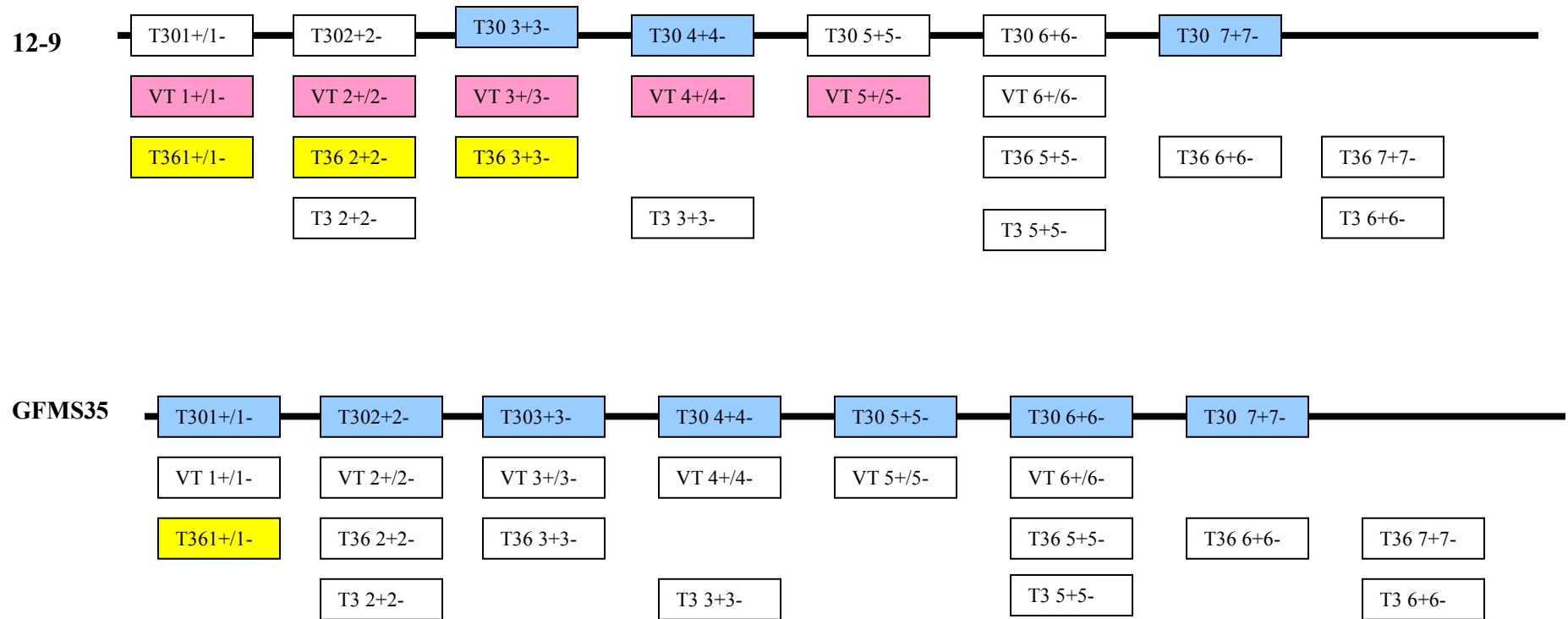


Figure 15:

The genotype classification of isolates GFMS 12 and T30 across regions 1-8 of the CTV genome. The 5' half of the CTV genome is represented by the thick black line. The BLUE blocks represent T30 genotype specific amplicons; PINK blocks represent VT genotype specific amplicons; YELLOW blocks represent T36 genotype specific amplicons; and GREEN blocks represent T3 genotype specific amplicons. CLEAR blocks indicate no amplification.

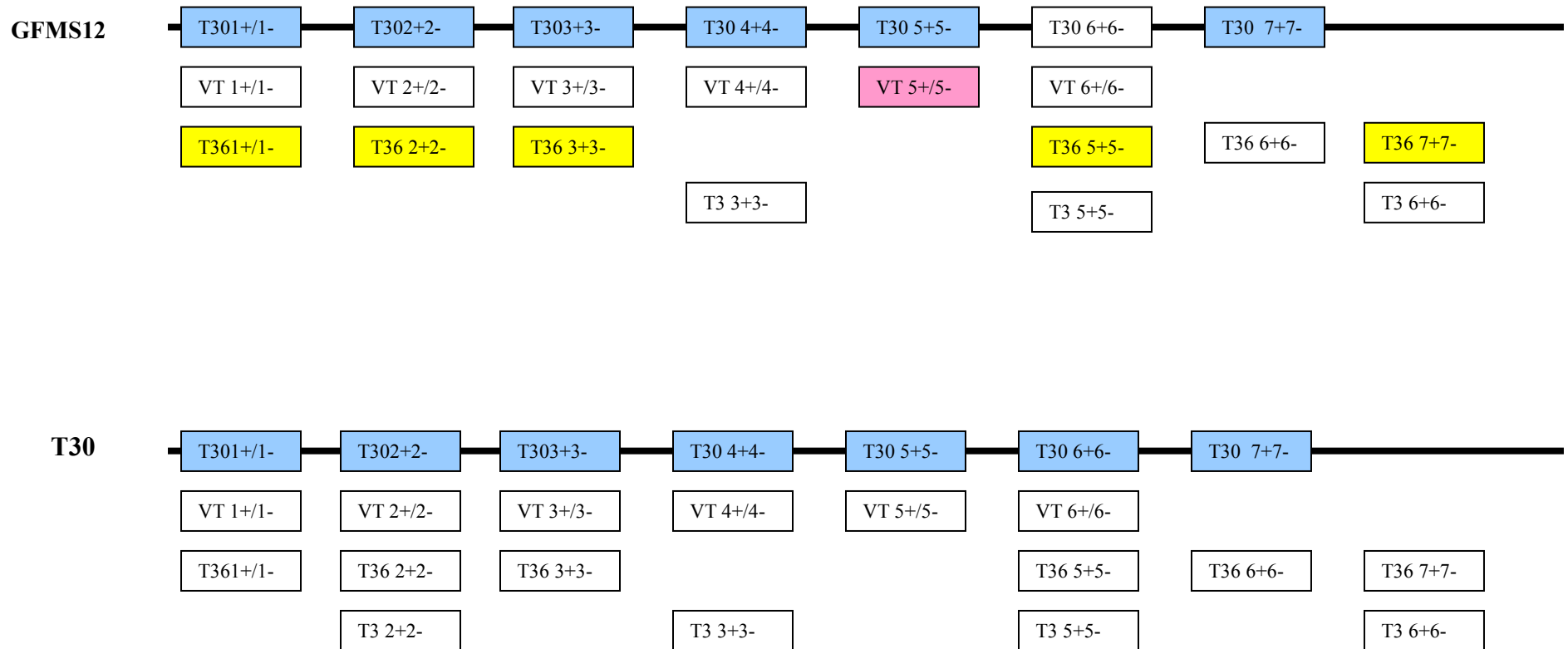
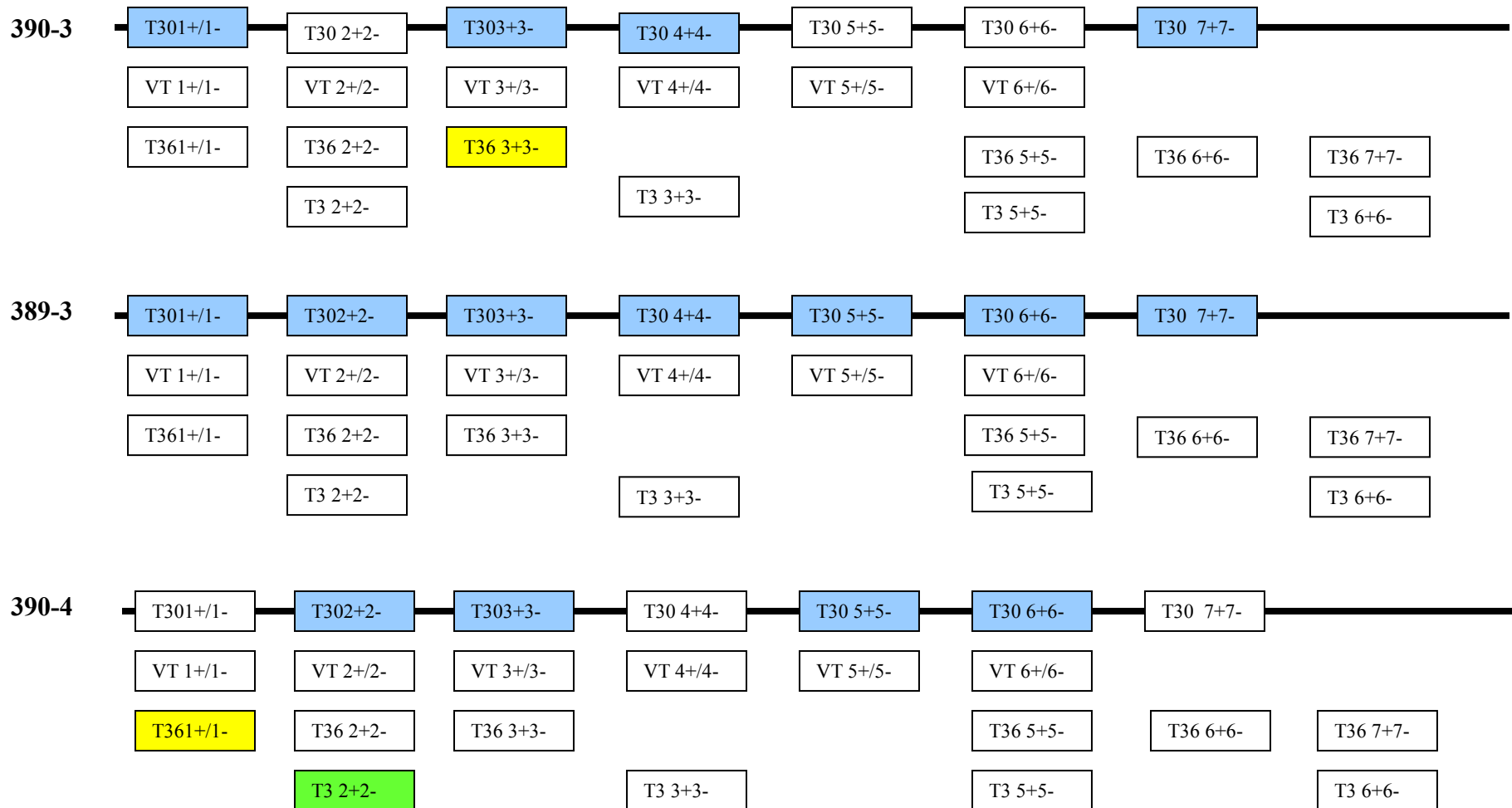


Figure 16:

The genotype classification of isolates 390-3, 389-4 and 390-4 across regions 1-8 of the CTV genome. The 5' half of the CTV genome is represented by the thick black line. The BLUE blocks represent T30 genotype specific amplicons; PINK blocks represent VT genotype specific amplicons; YELLOW blocks represent T36 genotype specific amplicons; and GREEN blocks represent T3 genotype specific amplicons. CLEAR blocks indicate no amplification.



3.4 DISCUSSION

3.4.1 SEROLOGICAL DETECTION

Factors that can influence the sensitivity and reliability of the assay include quality of antibodies, preparation and storage of reagents, incubation time and temperature, selection of appropriate parts of plant samples, time of year collected and the use of suitable extraction buffers (McLaughlin *et al.*, 1981). Another difficult aspect of ELISA is setting a threshold to score a positive and a negative result. Positive and negative controls are included in each assay to define a threshold for differentiating between “infected” and “uninfected” samples. In this study the ELISA methodology was performed and evaluated according to the guidelines set out by Richard Lee (USDA-ARS, Riverside, CA), where a good plate will show OD₄₀₅ values below 0.050 for the healthy and buffer only wells, and values over 1.00 for the positive controls. Samples were considered CTV positive if the average OD value was more than 0.500.

Viruses are known to be unevenly distributed in many host plants (Adams, 1978; Kolber *et al.*, 1982; Dahal *et al.*, 1998). This makes the sampling strategy critical for virus detection. Of the three plant components (fully flushed leaves, brown lignified bark and green bark) tested with ELISA, green bark had a 1.4-1.6 times higher titer compared to brown lignified bark and green fully flushed leaves. In this study sampling from fresh green bark of the 11 RSA samples was performed to ensure the highest titer for optimal RNA extraction. This is vital in this study where the sensitivity of detecting genotype/strain specific variants in a PCR reaction is not well defined, especially since it would not be known before testing what percentage of a certain variant is found in a particular sample. If the variant had a low titer and it affected amplification, a false negative result could occur. It is critical to ensure that the RNA extraction contains as much CTV RNA as possible and this was at least ensured by sampling the plant component with the highest titer. In a study performed to optimise dsRNA recovery it was found that the quantity of dsRNA was generally similar to that determined for antigen titer of CTV in all the hosted tested (Dodds *et al.*, 1987). Fresh green bark was therefore the preferred plant component to sample from (Dodds *et al.*, 1987).

The 11 RSA isolates tested with TAS-ELISA and CTV polyclonal antibody 1052 had varying absorbance values. Values higher than those of the controls indicated possible positive

detection of CTV and a correlation to the titer of CTV in these samples. The isolates GFMS12, GFMS35, 12-5, 12-9, 389-4 and 390-5 showed a 2.2-3.1 times higher absorbance value above the virus free control threshold and therefore a high titer of CTV. Isolates 12-7 and 390-3 were only slightly higher than the virus free control threshold. Isolate 390-4 and 389-3 were similar to the virus free control threshold. Isolate T30 had a 1.7 times lower absorbance than the virus free control. Isolates 390-4, 389-3 and T30 are possibly not infected with CTV or are at a level undetectable serologically. The negative/buffer control, virus free control and uncoated well controls had similar low values indicating that there was no or negligible cross reaction with plant or any other non-specific components and no problems with buffer preparation and antiserum viability. Candresse *et al.* (1998) and other researchers have shown that viral nucleic acid-based techniques like dot-blot hybridization assays and polymerase chain reaction are more sensitive than any other methods including serological ones. CTV at low titers might be enough to detect with more sensitive molecular tests like PCR. With most of the ELISA non reactive samples mild vein clearing symptoms were observed indicating possible CTV infection. Previous studies have shown that there was no correlation observed between symptom severity and antigen accumulation or dsRNA pattern intensity (Moreno *et al.*, 1990).

It is also possible that the strains infecting some of these low CTV titer samples could have poor replication and not reach high titres, possibly due to the inherent strain, host or environmental factors. Serological methods are useful as a first step screening method to detect CTV. Due to ELISA's adaptability, sensitivity and economical use of reagents, it is used in a wide range of situations, especially to test a large number of samples in a relatively short period of time.

3.4.2 OVERALL GENOTYPE CLASSIFICATION

The genotype classification of each isolate is summarised (Figures 12-16) as:

- Cognate regions of isolate **12-5** amplified with only VT genotype specific markers and therefore have a genotype profile similar to VT. There were however two regions of non-amplification including VT markers in region 4.
- Cognate regions of isolate **12-7** amplified with primers of two genotypes, VT and T36. The T36 genotype was found in the first two regions and could indicate possible sequence differences in the 5' end; non-specific amplification; or a region of

recombination. Unexpectedly region 4 and 6 again did not amplify with VT specific primers. Overall this isolate appears to be predominantly a VT genotype with possible sequence differences at the 5' end or a mixed infection with an isolate similar to but not identical to T36.

- Cognate regions of isolate **12-9** amplified with primers of three genotypes, VT; T30; and T36 and appear to have mixed genotypes. The VT genotype was more predominantly found throughout the regions with the T36 genotype located on the 5' end. T30 was found scattered in different regions targeted. Region 3 had all three of these genotypes which seems unlikely to have occurred from one strain and more so from a strain mixture. This isolate was previously thought to be a strain mixture (van Vuuren *et al.*, 2000). Possibilities for this occurrence could include PCR bias or optimal amplification of certain genotype primer sets only and not all. And raises concerns over the possible dominance of strains.
- Isolate **GFMS 12** is a mixture of strains used in the cross-protection scheme. Most cognate regions amplified with primers of genotypes, T36 and T30. In the central 5' half one VT genotype amplicon was found. The most consistent amplification of T30 and T36 genotype markers suggests these two genotypes being present.
- Isolate **GFMS 35** is a mixture of strains used in the cross-protection scheme. GFMS 35 has not shown any significant cross-protection breakdown in the industry. Cognate regions of isolate GFMS 35 amplified with primers of genotype T30. Thus could also indicate that GFMS 35 is not a mixed infection, as previously thought. This could be why GFMS 35 is a good mild cross-protecting source. The T30 genotype strains could be dominant and in high titer and therefore amplify preferentially.
- Cognate regions of isolate **389-3** amplified with primers of only genotype T30. This isolate has the potential to be used as a mild cross-protective isolate in RSA based on its mild T30 genotype designation with this molecular system. Biologically parameters such as replicative ability, influence on plant growth, symptoms and experimental challenge with severe strains would validate its potential for use.

- Cognate regions of isolate **T30** were used as a control for the T30 genotype and displayed a typical T30 marker profile with all of the expected T30 markers amplified. This finding confirms the genotype of this plant to be used in the future for experiments where a T30 strain is needed as a reliable control.
- Cognate regions of isolate **389-4** amplified with primers of the T30 genotype. Two regions did not have amplification where it was expected to amplify the T30 genotype. These two regions could represent an area of variation with slight mutations changing the sequences targeted by the primer preventing primer annealing. Region 2 had genotypes T30 and T3. This could represent cross-specific genotype amplification as found by Roy *et al.* (2004) and a possible area to be improved with the primer design. This isolate could possibly be used for cross-protection.
- Cognate regions of isolate **390-3** amplified with primers of two genotypes, T30 and T36 with four positive T30 genotype regions and one T36 genotype region. There was no amplification in three of the seven regions targeted by T30 primers. It is possible that this isolate has more sequence variation than expected and could represent a new sequence variant. Overall the isolate seems to be a T30 genotype with regions of variation not detectable with any genotype-specific primer sets.
- Cognate regions of isolate **390-4** amplified predominantly with genotype T30 and one T3 genotype positive region. Amplification of the T3 genotype could be from cross-specific genotype amplification, as discussed previously. Regions 1, 4 and 7 capable of amplifying the T30 genotype were not amplified. Overall the isolate had a predominantly T30 genotype.
- Cognate regions of isolate **390-5** amplified with primer sets of three different genotypes across the regions targeted and appears to be a mixed infection or atypical. Four of the regions were positive for the T36 genotype; two each for VT and T30 genotypes. The regions positive for VT were regions 1 and 2 in the 5' area. The two regions positive for T30 were located next to each other from about 2246-4663 nt, which is quite a large area. It is not clear whether this isolate is a mixed infection. This isolate might be a recombinant or large variability in most regions including

the three regions not amplified. Overall it seems the T36 genotype was more prevalent with 50 % of the regions positive including more conserved regions like region 6.

In summary isolates 389-4, 390-3, 389-3, 390-4 and GFMS 35 had predominantly T30-like genotypes while isolates 12-5 and 12-7 had VT-like genotypes. There were many isolates with non-standard genotype profiles and it becomes important to analyse each isolate separately as well as to notice any general trends with the isolates. With available data it is impossible to state whether these non-standard profiles are solely because of mixed infections or inherent variability of the strain/variant, recombination events or cross-reacting primers. This would only be objectively determined if single aphid transmissions or cloning were done and thereafter sequencing of an isolate to determine what the true population of strains are within these isolates. Isolates (389-3, 389-4, 390-3, 390-4, 390-5, 12-5, 12-7 and 12-9) (van Vuuren *et al.*, 2000) originated from single aphid transmissions and therefore have a greater probability of possessing one strain or sequence type. Multiple serial aphid transmissions of an isolate are needed to try to provide a more homogenous population of strains.

The mild strain cross-protecting isolates GFMS 12 and GFMS 35 were expected to have multiple strains and possibly multiple genotypes as they originated from a naturally infected plant with a population of CTV variants. These isolates however were regarded as mild until studies showed that severe strains existed within GFMS 12 (van Vuuren *et al.*, 2000). Sub-isolates were made by single aphid transmissions and some were used in this study (12-5, 12-7 and 12-9). These were prepared to try and reconstitute a mixture of strains without the severe component. These sub-isolates however were thought to still be mixtures of strains (van Vuuren *et al.*, 2000). Isolates GFMS 12, 12-9 and 390-5 did not fit a defined genotype of Hilf *et al.* (2000) and thus appear to be a group of unexpected mixtures of genotypes VT, T30 and T36. There was a consistent finding of T30, VT and T36 genotypes within the different regions targeted. Isolates which are mixed infections are expected to have different genotypes amplified throughout the targeted regions. Citrus plants can possibly be infected with multiple genotypes of CTV and proper identification is important for disease management.

Some isolates appeared to have mixed genotypes but not all genotype specific markers were consistently amplified to classify the isolate as a mix of genotypes. Isolate 390-4 had a mix of T3/T30 genotypes in region 2 but T30 in most of the other regions; isolate 390-3 had the T36 genotype in region 3 but T30 genotypes for the rest of the regions; isolate 389-4 had a T3 genotype at region 2 but T30 genotypes for the rest of the regions; and isolate 12-7 had the T36 genotype in the regions 1 and 2 but VT for most of the other regions. A pattern of incomplete amplification at regions on the 5' end suggests possible sequence diversity and/or cross-amplification of genotypes from slight mutations in the primer site/s. There was a more representative profile of the isolate's genotype from the other regions more distant from the 5' terminal end.

Recent work using multiple molecular marker primer pairs developed by Hilf *et al.* (2000) in regions of the CTV genome of Indian isolates showed that amplifications were not always genotype-specific (Roy and Bransky, 2004). The T3 genotype specific primer amplified the product of some isolates that had more sequence similarity with the T30 genotype rather than with the T3 genotype (Roy and Bransky, 2004). This may have occurred with isolates 389-4 and 390-4 with mixed T3/T30 genotypes in region 2 and T30 in the other regions. The T3 genotype has been described as unique and has not been commonly found in other areas of the world, its origin is unclear (Hilf, personal communication). Isolate GFMS 12 yielded a one VT amplicon. Isolate 390-3 had one T36 amplicon from region 3 and four T30 amplicons. These could represent instances of cross-genotype amplification with primers falsely detecting the incorrect genotype. These amplifications could have occurred from variability within this region where a slight mutation could result in a primer site resembling another genotype more specifically than it should.

Possibilities for non-amplifications in regions are:

(1) Possible secondary structures preventing the annealing of the primer to template in cDNA and/or PCR steps. DNA plasmid clones, used as positive controls and infected plant RNA samples are very different templates. DNA clones are already in DNA form and no cDNA synthesis is needed, so they are therefore only PCR controls. Problems arising from the synthesis of cDNA from RNA plant samples can not be established. Problems preventing successful cDNA synthesis could be the effects of secondary structures or inhibitory components in the plant. Inhibition from cDNA and/or PCR steps would have

resulted in the lack of amplification of all products in every region and this was not the case. Other indirect possibilities for amplification failure could have been reagent viability, human error and RNase/DNase contamination; however isolates were tested in duplicate and gave the same results. Secondary structures remain a possibility in the failure of certain isolates from amplifying certain regions.

(2) Variation can account for amplification failures if there is a slight mutation within the primer site/s or widespread variation relating to a new undefined genotype or sequence variant. Amplification failure in regions targeting the intended genotype could represent sequence changes in primer sites to a novel sequence which does not serve as a template to any of the genotype specific primers. This is particularly evident in regions 1-4, at the 5' end of the CTV genome. Mawassi *et al.* (1996) and Albaich-Marti *et al.* (2000) showed that sequences of different strains of CTV revealed sequence diversity at the 5' end. The 5' end of the genome is highly variable and much more diverse than expected for strains of the same virus (Gowda *et al.*, 2003). There is more variability between nucleotides 1082 and 1484 than between 1-50 or 697-1105 (Roy *et al.*, 2005). Amplification failure in sequential regions could be due to sequence diversity or possibly recombination. Isolates with varying regions of no amplification could represent sequence changes widespread along the genome and are possible sequence variants of a genotype.

Problems with primer site sequence changes, stringency conditions, strain mixtures, recombination events, amplification failures and cross-specific genotype amplification are difficult to differentiate from one another and can not be confirmed from these results. There were no infected plant controls available for the four genotypes (T36, T30, T3 and VT) to accurately eliminate such possibilities. These problems are a drawback in trying to establish a system like this and to accurately assign a genotype to isolates. Sequencing of amplified products will help to clarify these problems but would be impossible to perform routinely in the industry on all isolates. However sequencing of amplicons from problematic regions could help to eliminate region specific problems for example secondary structures and cross-amplification.

Isolates which are mixtures of different genotypes are expected to have all of the genotype markers consistently amplified. This would result in more than one genotype specific amplicon arising per region. However strain dominance, strain competition and PCR bias

could influence this expected result. Isolates with recombinant genomes could have regions amplified which are specific for a genotype. This would result in some regions displaying one genotype and other regions another. Areas on the CTV genome derived from recombination would have a greater possibility of displaying such results. It is also possible that primer sequences not 100 % homologous but similar enough to the template could still anneal and amplify, resulting in a false positive. The stringency conditions could therefore not be as optimal as needed.

As with the work of Hilf *et al.* (2000), it was found that T30-like, T36-like and VT-like genotypes are not just restricted to Florida and Israel but also exist in South Africa. The 23 primer pairs developed by Mark Hilf have been useful in determining an isolate's possible genotype profile. However it becomes a challenge with isolates that have some but not all markers characteristic of a standard genotype. Even a single base mutation on a specific primer could prevent the amplification of the selected region. This is especially so if the selected area is within a highly variable region, for example the CTV 5' end half of the genome. The system produces confusing results when multiple genotype markers are found in an isolate.

A considerable amount of time went into optimising every step from total RNA extraction or immunocapture of the virions to the conditions for each PCR. Immunocapture PCR resulted in greater sensitivity, and problems encountered with RNA extractions and inhibitors of RT-PCR were minimized. The 23 primer set system allowed a better understanding of the isolates variability than from a more conserved region like the previous study on the p23 gene where the analysis is related to possible function. In previous work full genome sequences of the Hepatitis B virus (HBV) differing by more than 8 % were classified as separate genotypes (Song *et al.*, 2006). The primers developed by Hilf *et al.* (2000) for this system only constitute about 1.5 % of the T30 genome; 1.2 % of the VT genome; 1.3 % of the T36 genome and 0.9 % of the T3 genome. It is unclear if half of the genome is appropriate for complete strain or genotype differentiation.

Hilf *et al.*, (2000) mentions that there was no direct correlation between the biological characterization and the assigned molecular genotype. In this study isolates were indexed as mild but show variable genotype patterns not necessarily expected of mild isolates.

Biological indexing alone is an inappropriate way of determining strain types. This could result in an inaccurate picture of an isolate's complete genetic complexity.

Genotypes depicted in this study are unlikely to represent the molecular variability of indigenous CTV populations in South Africa as only a few mild strains were tested. It does however give a good indication as to possible genotypes occurring in South Africa and in strains used for mild strain cross-protection. This PCR based system of 23 primer sets can be a very useful tool in studying the distribution of genotypes; to track their movement; to give a better understanding of the CTV population structure and in the selection of isolates for cross-protection. As more sequences become available, the number and specificity of the molecular markers can be increased to provide more accurate answers.

3.5 REFERENCES

Adams, A.N. 1978. The detection of plum pox virus in *Prunus* species by enzyme-linked immunosorbent assay (ELISA). *Ann. of Appl. Biol.* 90: 215-221.

Albiach-Martí, M.R., Mawassi, M., Gowda, S., Satyanarayana, T., Hilf, M.E., Shanker, S., Almira, E.C., Vives, M.C., López, C., Guerri, J., Flores, R., Moreno, P., Garnsey, S.M., Dawson, W.O. 2000. Sequences of *Citrus tristeza virus* separated in time and space are essentially identical. *J. Virol.* 74: 6856–6865.

Ayllón, M.A., López, C., Navas-Castillo, J., Garnsey, S.M., Guerri, J., Flores, R., Moreno, P. 2001. Polymorphism of the 5' terminal region of *Citrus tristeza virus* (CTV) RNA: Incidence of three sequence types in isolates of different origin and pathogenicity. *Arch. Virol.* 146: 27-40.

Ayllón, M.A., López, C., Navas-Castillo, J., Mawassi, M., Dawson, W.O., Guerri, J., Flores, R., Moreno, P. 1999. New defective RNAs from *Citrus tristeza virus*: Evidence for a replicase-driven template switching mechanism in their generation. *J. Gen. Virol.* 80: 817-821.

Candresse, T. M., Cambra, S., Dallot, M., Lanneau, M., Ascenio, M.T. Gorris, F., Revers, G., Macquaire, A., Olmos, D., Boscia, J.B., Quinot, Dunez, J. 1998. Comparison of monoclonal

antibodies and polymerase chain reaction assays for the typing of isolates belonging to the D and M serotypes of plum pox potyvirus. *Phytopathology* 88: 198-204.

Dahal, G., Hughes, J. A., Thottappilly, G., Lockhart, B.E.L. 1998. Effect of temperature on symptom expression and reliability of banana streak badnavirus detection in naturally-infected plantain and banana (*Musa* spp.). *Plant Dis.* 82: 16-21.

Dodds, J. A., Jarupat, T., Lee, J.G., Roistacher, C.N. 1987. Detection of strain specific double-stranded RNAs in citrus species infected with *Citrus tristeza virus*: A review. *Phytophylactica* 19: 131-137.

Dreher, T.W. 1999. Functions of the 3'-untranslated regions of positive strand RNA viral genomes. *Ann. Rev. Phytopathology.* 37: 151-174.

Gowda, S., Satyanarayana, T., Ayllón, M.A., Moreno, P., Flores, R., Dawson, W.O. 2003. The conserved structures of the 5' nontranslated region of *Citrus tristeza virus* are involved in replication and virion assembly. *J. of Virol.* 317: 50-64.

Hilf, M.E., Karasev, A.V., Albiach-Marti, M.R., Dawson, W.O., Garnsey, S.M. 1999. Two paths of sequence divergence in the *Citrus tristeza virus* complex. *Phytopathology* 89: 336-342.

Hilf, M.E., Garnsey, S.M. 2000. Characterization and Classification of *Citrus tristeza virus* isolates by amplification of multiple molecular markers. In "Proceedings of the 14th Conference of the International Organization of Citrus Virologists." (J. V. da Graca, R. F. Lee, R.K. Yokomi, Eds), pp. 18-27, IOCV, Riverside, California.

Hilf, M.E., Garnsey, S.M. 2002. *Citrus tristeza virus* in Florida: A synthesis of historical and contemporary biological, serological and Genetic data. In "Proceedings of the 15th Conference of the International Organization of Citrus Virologists." (N. Duran-Vila, R.G. Milne, J. V. da Graca, Eds), pp. 13-20, IOCV, Riverside, California.

Kolber, M., Nemeth, M., Szentivanyi, P. 1982. Routine testing of English walnut mother trees and group testing of seeds by ELISA for detection of cherry leaf roll virus infection. *Acta Horticulturae* 130: 161-172.

López, C., Ayllón, M.A., Navas-Castillo, J., Guerri, J., Moreno, P., Flores, R. 1998. Sequence polymerism in the 5' and 3' terminal regions of *Citrus tristeza virus* RNA. *Phytopathology* 88: 685-691.

Mawassi, M., Karasev, A., Mietkiewska, E., Gafny, R., Lee, R.F., Dawson, W.O., Bar-Joseph, M. 1995. Defective RNA molecules associated with *Citrus tristeza virus*. *Virology* 208, 383-387.

McLaughlin, M.R. Barnett, O.W., Burrows, P.M., Baum, R.H. 1981. Improved ELISA conditions for detection of plant viruses. *J. of Virol. Meth.* 3: 13-25.

Miller, E.D., Plante, C.A., Kim, K-H., Brown, J.W., Hemenway, C.L. 1998. Stem-loop structure in the 5' region of Potato virus X genome required for plus-strand RNA accumulation. *J. of Mol. Biol.* 284: 591-608.

Moreno, P., Guerri, J., Munoz, N. 1990. Identification of Spanish strains of CTV by analysis of dsRNA. *Phytopathology* 80: 477-482.

Permar, T.A., Garnsey, S.M., Gumpf, D.J., Lee, R.L. 1990. A monoclonal antibody that discriminate strains of *Citrus tristeza virus*. *Phytopathology* 80: 224-228.

Roy, A., Brlansky, R.H. 2004. Genotype classification and molecular evidence for the presence of mixed infections in Indian *Citrus tristeza virus* isolates. *Arch. Virol.* 149: 1911-1920.

Roy, A., Manjunath, K.L., Brlansky, R.H. 2005. Assessment of sequence diversity in the 5'-terminal region of *Citrus tristeza virus* from India. *Virus Res.* 113 (2): 132-142.

Roy, A., Ramachandran, P., Brlansky, R.H. 2002. Molecular characterisation of Indian *Citrus tristeza virus* isolates. *Phytopathology* 92: 71.

Roy, A., Ramachandran, P., Brlansky, K.H. 2003. Grouping and comparison of Indian citrus tristeza virus isolates based on coat protein gene sequences and restriction analysis patterns. *Arch. Virol.* 148: 707-722.

Song, Y., Dai, E., Wang, J., Liu, H., Zhai, J., Chen, C., Du, Z., Guo, Z., Yang, R. 2006. Genotyping of hepatitis B virus (HBV) by oligonucleotide microarray. *J. of Mol. and Cell. Probes* 20: 121-127.

Van Vuuren, S.P., van der Vyver, J. B., Luttig, M. 2000. Diversity among sub-isolates of cross-protecting *Citrus tristeza virus* isolates in South Africa. In “Proceedings of the 14th Conference of the International Organization of Citrus Virologists.” (J.V. da Graca, R.F.Lee, R.K. Yokomi, Eds), pp. 103-109, IOCV, Riverside, California.

Vela, C., Cambra, M., Cortes, E., Moreno, P., Miquet, J., Perez de San Roman, C., Sanz, A., 1986. Production and characterization of monoclonal antibodies specific for *Citrus tristeza virus* and their use for diagnosis. *J. Gen. Virol.* 67: 91-9.

CHAPTER 4

DEVELOPMENT OF AN OLIGONUCLEOTIDE MICROARRAY CHIP TO DIFFERENTIATE T₃O AND T₃6 STRAINS OF CTV

4.1 INTRODUCTION

Microarrays are one of the new emerging methods in plant virology currently being developed by various laboratories. Microarrays or DNA chips were initially designed for the study of gene expression or single-nucleotide polymorphism (SNP) profiles. Today they have become a standard tool for both molecular biology research and diagnostics. The principle of microarrays is the hybridization of fluorescently labelled sequences (targets) to detect their complementary short sequences (oligonucleotides) spotted on a solid surface. Tens of thousands of DNA probes can be spotted in a defined configuration onto a glass microscopic slide forming the chip. Recently, with progress on functional genomic techniques, microarrays have been developed into high-throughput methods with the ability to detect many targets simultaneously and are ideal for genotyping or strain differentiation (Song *et al.*, 2006).

Recently some studies have been published on microarray-based detection of plant viruses. Identification of potato viruses in single or mixed infections has been reported (Boonham *et al.*, 2003). A plant virus chip was developed for the detection of cucurbit-infecting Tobamovirus (Lee *et al.*, 2003). Good detection selectivity was attained for two potato viruses (PVS and PLRV) (Bystricka *et al.*, 2003). A system was developed for the detection and differentiation of Cucumber mosaic virus (CMV) serogroups and subgroups using the coat protein gene (Deyong *et al.*, 2005). In that study different isolates were hybridized against a set of five different serotype and subgroup specific 24-mer oligonucleotides (oligos). The results showed that the method allowed a clear differentiation of the isolates into serogroups 1 and 2, and in addition assigned nine out of the ten serogroups 1 isolates correctly into subgroups 1a and 1b (Deyong *et al.*, 2005). This differentiation was not possible by RFLP analysis with the restriction enzyme *MspI* (Deyong *et al.*, 2005).

Citrus tristeza virus (CTV) is one of the most economically devastating diseases of citrus worldwide. Certain strains of CTV (e.g. T36 and VT) cause symptoms that are difficult to control and can be destructive. Their severity depends frequently on a complex relationship between virus strains, hosts, vectors and the environment. It is therefore a prerequisite to have methods in place to accurately identify strains for effective and quick control of the disease. The entire genomic sequences of nine CTV isolates are now available in public

databases and researchers can design CTV-specific primers for PCR or oligonucleotides for microarrays.

The necessity to apply highly multiplexed methods for the detection and differentiation of viruses and other plant pathogens grows with the tendency of global production of crops and plants which are imported and exported under increasingly stringent quarantine regulations. Methods should be: highly sensitive and specific, affordable, reproducible and safely and rapidly performed. The best sensitivity is offered by PCR/RT-PCR, with these PCRs gaining speed, sensitivity and safety when performed under real-time conditions. Differentiation after general amplification can either be done by RFLP analyses, SSCP, or by hybridization with specific probes allowing further differentiation into species or even pathotypes (Letschert *et al.*, 2002). It has been shown for Tobamoviruses (Letschert *et al.*, 2002) and a Potyvirus (Boonham *et al.*, 2002) that mutations leading to new pathotypes can often only be detected by nucleic acid-based methodology. However, RT-PCR techniques on their own cannot detect virus mutants or related strains with significantly variable sequences. The diagnostic application of microarrays would be an ideal supplement to solve the problem of differentiation after generic amplification.

The primary aims of this study were to:

1. Assess the feasibility and potential use of oligo microarrays in the differentiation of CTV strains, using CTV T30 and T36 strains as models, in a pilot study.
2. To design oligos differentiating the strains by targeting the T36 strain.

The secondary aims were to:

1. To set up an oligonucleotide microarray standard protocol by testing different oligonucleotide design strategies; labelling methods; hybridization buffers; and effective controls (Section 4.3.1).
2. To test slides with strains T30 and T36 individually on the chip at 42, 52 and 60 °C (Section 4.3.2).
3. To test the effect of Locked Nucleic acids (LNA) on specificity of oligos.

In this study a novel CTV genotyping method using a low-density oligonucleotide microarray is described by firstly setting up an operational array and then testing the system with the two CTV strains.

4.2 MATERIALS & METHODS

4.2.1 OLIGONUCLEOTIDE DESIGN & PREPARATION

The oligonucleotides were designed from the 5' region of the CTV genome to distinguish between T30 (AF260651) and T36 (AY170468) strains (refer Table 9). The two strains were aligned using ClustalX 1.8 software and areas of variability were identified. The Pairwise sequence alignment is shown in Appendix 3, Figure 60. ArrayDesigner™ was used to design the best probe of 18-25 bp from an input of 70mer segments from the T36 strain complete genome (U16304). The programme scored the oligonucleotides based on the parameters: GC content; T_m; free energy and many other structural properties such as repeats, secondary structures and hairpins to find the best probes. The specificity of designed oligonucleotides was assessed by comparing them to sequences on Genbank using the BLASTN 2.2 function. A local database was set up on the university's local server using the standalone WWW BLAST service of all the CTV sequences available on GenBank. Where the e-value threshold was set at below 10⁻⁵ to prevent cross-homology non-target strains. Additional parameters were set as follows:

- Self dimer maximum dG: default value = -6 kcal/mol (free energy of the most stable self-dimer that is acceptable).
- Run/repeat maximum length: default value = 4 bp/dinucleotide. Probes with single (e.g. AAAAA) runs or dinucleotide (ATATATATAT) repeats of length greater than 4 are discarded.
- Hairpin maximum dG: default value = -3 kcal/mol (free energy of the most stable hairpin that is acceptable).
- T_m 55-65 ° C and a GC content preferably above 60 % for each oligonucleotide

Oligonucleotides were designed to be 100 % homologous to the targeted T36 strain (AY170468). These T36 specific oligonucleotides differed: (1) 1-20 %; (2) 21-40 %; and (3) above 41 % sequence identity when compared to the T30 strain. Mismatches were designed to be in four different positions on oligonucleotides namely (1) scattered; (2) centre; (3) on 3' end and (4) on the 5' end. There were two conserved oligonucleotides designed to be 100 % homologous to both strains. External controls included were a Rabies virus (Rab) and two West Nile virus (WNV) specific oligonucleotides. Oligonucleotides were also assessed

for possible cross-homology to (a) other parts of the CTV genome and (b) other non-CTV sequences on Genbank. Oligos suitable for use in this chip were ordered (Inqaba, Pretoria) with cartridge purification and desalting to remove any shorter strands and salts remaining from the synthesis.

4.2.2 LNA BASED OLIGONUCLEOTIDES

Three oligos were modified (refer Table 9) to include LNA bases together with DNA bases instead of only DNA bases (You *et al.*, 2006). Oligo 264-T36 was modified at base 12 and 20 to be LNA guanines and designated as 264LNA-T36. Oligo 230-T36 was modified at base 2 and 20 to be LNA guanines and renamed as 230LNA-T36. Oligo 40-T36 was modified at base 16 to be a LNA guanine and renamed as 40LNA-T36. These LNA modified oligos were synthesized, cartridge purified and desalted (Inqaba, Pretoria).

4.2.3 MICROARRAY FABRICATION & DESIGN

Each oligonucleotide was prepared and pipetted into a 384 well microplate (GE Health Care, Buckinghamshire, UK). The synthesized probes were suspended in 50 % DMSO to a final concentration of 20 μ M and spotted onto Corning-GAPS II slides (Corning, UK). This was done at the University of Pretoria Microarray facility using an Array Spotter Generation III (Molecular Dynamics) and Microarray Pen Assembly DMSO pins (Molecular Dynamics, USA). DMSO was added to all the blank wells and spotted. After spotting, slides were stored in a dessicator until use. Primers designed for PCR of specific molecular markers of genotypes T36, T30, VT and T3 (Hilf *et al.*, 2002) were also included on the Microarray slide. The microarray layout is shown in Table 10.

4.2.4 CTV ISOLATES

The full length DNA clones of T30 and T36 (refer Chapter 3) were kindly supplied by Dr. S. Gowda, Univ of Florida, USA. The confirmation of the presence of the T30 and T36 CTV strains was confirmed by PCR as described in chapter 3 and by sequencing to verify the strain type (refer to section 4.3.6).

Oligonucleotide Name	Sequence 5'→3'	Length	Tm	GC %	Position	Specificity
02-T36	TGAGACGAGCTGACACTGC	19	55.2	57.9	610-629	T36
06-T36	ATCACGGAAGCAGAGGGAAG	20	55.1	55	6688-6708	T36
40-T36	TGTCGAAACTCAGAGGAAGCT	21	54.9	47.6	129-140	T36
57-T36	GGATTGGATGGTTACGGCTT	20	53.6	50	2766-2786	T36
59-T36	AAGGTTCCGTGCTGCGTC	18	55.8	61.1	6742-6760	T36
64-T36	CATCTGGCGTTGTTTCGTCC	19	54.7	57.9	414-433	T36
82-T36	ACCTTCACACTGCGTTTCGT	19	55.3	52.6	1024-1043	T36
123-T36	TTGTGCGTGGTGAATCTTCG	20	54.9	50	7745-7765	T36
228-T36	CTGCTGTCGCTGAGAAAGTG	20	55	55	6611-6631	T36
230-T36	TGGAATCAGGTCTCGTTAAGGT	22	54.9	45.5	2650-2672	T36
264-T36	GCGATTACTTGGTGACCTTGAA	22	55.1	45.5	7792-7813	T36
268-T36	TTGAGGGCAGATGTTGTGGT	20	55.1	50	6424-6444	T36
Con 1	GGTTTCTCCCCGCAAGTG	18	62.18	61.11	951-969	Conserved for all CTV strains
Con 2	TAAAGGCTTATCTTGTTTCGC	20	56.3	40	1433-1453	Conserved for all CTV strains
264LNA-T36	GCGATTACTTGGTGTGACCTTGAA	22	57.1	45.5	7792-7813	LNA T36
230LNA-T36	TGGAATCAGGTCTCGTTAAGGT	22	57.1	45.5	2650-2672	LNA T36
40LNA-T36	TGTCGAAACTCAGAGGAAGCT	21	58.7	47.6	129-140	LNA T36
WNV-prM-Fbac	CGCGGATCCGTGACCCTCTCGACCT	27	68.7	70.3	N/A	West Nile Virus External Control
WNV I	AACCTCGCAGATGTGCGC	18	58.8	61	N/A	West Nile Virus External Control
RabMF	GATAAAATGAACTTTCTACGTAAG	24	47.8	29	N/A	Rabies Virus External Control
T36-1+	AGCCTTTAAGCTCTAATATT	20	46.0	30.0	68-88	T36
T36-2+	AAACTGATTTCTCCACTCAG	20	49.8	40	855-895	T36
T36-3+	CTTCTTTTAACTCGACAAGGA	21	50.0	38.0	2323-2344	T36
T36-5+	CGGTAGCGTGTTGTGAGTACG	21	57.9	57.1	6040-6061	T36
T36-6+	ATTCCCTAATTCTAGGGCTC	20	51	45	7370-7390	T36
T36-7+	GTATGTTACCGATGCAGCTTC	21	53.7	47.6	8405-8426	T36

Oligonucleotide Name	Sequence 5→3	Length	T _m (°C)	GC %	Position	Specificity
T30-1+	GTATCTCCGGAGCTCGATC	19	54.2	57.8	22-42	T30
T30-2+	TACGGCTTGGTGCTCTGAGGCC	22	63.8	63.6	792-814	T30
T30-3+	TGGATGAGGGTTCTTCACCAC	21	56.8	52.3	2267-2288	T30
T30-4+	GTTTACTGCTTTAACAATTCGGC	23	53.0	39.1	3766-3789	T30
T30-5+	TGGTTTACGTGCCGGAGCCGC	21	65.1	66.6	4731-4752	T30
T30-6+	GCGTGTCTTAGTGTATCGCA	20	54.8	50	6012-6032	T30
T30-7+	CAGCGTTCAAGTTTGCGTTA	20	54.0	45	7356-7376	T30
T3-2+	TGTTTGAGGTCCCGAGCGTC	20	60.5	65.0	962-982	T3
T3-3+	GTTCCGGTGGAGTTGGACGTTA	21	57.0	52.3	3708-3729	T3
T3-5+	TCCTTTGCCATCAATTGTATCAC	23	53.6	39.1	6133-6156	T3
T3-6+	GGGTCGGACTAAAGCAGTA	19	53.8	52.6	8847-8862	T3
VT-1+	GTACCCTCCGGAAATCACG	19	55.3	57.8	19-38	VT
VT-2+	AACGGCATGGTGCTCTCCTGTT	22	61.5	54.5	783-805	VT
VT-3+	CAGGTGAGAATTCTCCATCGT	21	54.3	47.6	2246-2267	VT
VT-4+	ATCTACTGTTTTAACAATTCACG	23	49.3	30.4	3745-3768	VT
VT-5+	TGGTTTGCGGGCCGGGCG	18	67.7	77.7	4710-4728	VT
VT-6+	ACGCGTTTCGGTATGTTGTA	20	54.5	45.0	5985-6005	VT

Table 9:
Designed and synthesized primers and oligonucleotides used in the microarray design.

Table 10:

The layout of the CTV T30 versus T36 strain microarray slide.

2	B	6	B	40	B	57	B	E1	B	E2	B	230L	B	Co1	B	A1	B	A2	B	A3	B	A5	B	A6	B	A7	B	40L	B	B	B
2	B	6	B	40	B	57	B	E1	B	E2	B	230L	B	Co1	B	A1	B	A2	B	A3	B	A5	B	A6	B	A7	B	40L	B	B	B
2	B	6	B	40	B	57	B	E1	B	E2	B	230L	B	Co1	B	A1	B	A2	B	A3	B	A5	B	A6	B	A7	B	40L	B	B	B

59	B	64	B	82	B	123	B	Co2	B	E3	B	40L	B	264L	B	B1	B	B2	B	B3	B	B4	B	B5	B	B6	B	B7	B	B	B
59	B	64	B	82	B	123	B	Co2	B	E3	B	40L	B	264L	B	B1	B	B2	B	B3	B	B4	B	B5	B	B6	B	B7	B	B	B
59	B	64	B	82	B	123	B	Co2	B	E3	B	40L	B	264L	B	B1	B	B2	B	B3	B	B4	B	B5	B	B6	B	B7	B	B	B

228	B	230	B	264	B	268	B	C2	B	C3	B	C5	B	C6	B	230L	B	D1	B	D2	B	D3	B	D4	B	D5	B	D6	B	B	B
228	B	230	B	264	B	268	B	C2	B	C3	B	C5	B	C6	B	230L	B	D1	B	D2	B	D3	B	D4	B	D5	B	D6	B	B	B
228	B	230	B	264	B	268	B	C2	B	C3	B	C5	B	C6	B	230L	B	D1	B	D2	B	D3	B	D4	B	D5	B	D6	B	B	B

Each small block represents a feature/spot on the slide. All features/spots are in triplicate. This whole layout is also replicated on the right hand side of the slide (not shown here). External controls: E1=WNVI; E2=RabMF and E3= WNV-prM-Fbac. Oligos designed by Hilf *et al.* (2002) are depicted here as A1-3 and A5-7 represents T36-1+ to 7+; B1-7 represents T30 1+ to 7+; C2-6 represents T3 2+ to 6+ and D1-D6 represents VT 1+ to 6+. L = LNA modified oligos. Conserved oligos = Co1 and Co2. Blanks with DMSO = B.

4.2.5 CY3 LABELLING & AMPLIFICATION OF TARGET DNA

PCR products were labelled using a Cy3 end labelled reverse primer in the PCR reaction as described by Deyong *et al* (2005). The primers used to amplify and label segments within the 5' region of the region of the genome were previously described (Hilf *et al.*, 2002). On average two oligonucleotides were designed within each amplified DNA fragment. Post-PCR Klenow labelling with the Cy3 dye and PCR based F3-UTP and Cy3-UTP labelling methods were also attempted.

Each of the PCR reactions contained 1 µl of the positive DNA clone (T30 or T36). PCR was conducted with 100 pmoles of each of the forward primers and Cy3 end labelled reverse primers (IDT, USA) (Table 11); 1.5-2.0 mM MgCl₂; 1x Gotaq Buffer (Promega, USA); 2.5 U Gotaq (Promega, USA); 0.21 mM each of dATP, dCTP, dGTP and dTTP; 2 µg/µl BSA; 0.1 % Triton-X-100; 10 mM mercaptoethanol and made up to a total volume of 50 µl with nuclease-free water. The contents were gently mixed. The thermal cycle conditions were 94 °C for 5 min, 40 cycles of 94 °C for 30 sec, 45-55 °C for 45 sec, and 72 °C for 1 min, then extension for 2 min at 72 °C, and held at 4°C on an Eppendorf Gradient Mastercycler (Eppendorf, Germany). Refer to Appendix 2, table 19 for specific conditions (optimal T_m and MgCl₂ concentrations) of individual primer sets. A positive PCR control included the use of primers without any Cy3-labelled ends (data not shown). A negative control consisting of nuclease-free water instead of DNA was included to show contamination and false positives. PCR products (5-25 µl) were analyzed by gel electrophoresis through 1.0 % agarose (Whitehead Scientific) in sodium borate electrophoresis buffer (5 mM disodium borate decahydrate, adjusted to pH 8.5 with boric acid), stained with ethidium bromide (0.5 µg/ml), and photographed under UV light. Amplicons were not purified since the reverse and forward primers were not complementary to any oligonucleotides present on the slide and to prevent any loss of labelled DNA through the process. Previously differentiation of CMV was also performed in this way (Deyong *et al.*, 2005). The concentration and amount of Cy3-dye present within a PCR tube was quantified using a Nanodrop spectrophotometer (Nanodrop Technologies, USA). Labelled products were shielded from excessive exposure to light to prevent photo-bleaching and Cy3 degradation. Tris-based buffers and reagents above pH 8.0 were avoided to prevent Cy3 degradation.

The sequences of the expected PCR products were used to determine if any secondary structures were present using the web based Mfold programme available on <http://www.bioinfo.rpi.edu/applications/mfold> (Zuker., 2003).

Table 11:

PCR primer sequences (Hilf *et al.*, 2000) used for PCR amplification of the T30 and T36 strains targeted by the microarray chip.

Isolate	Position	Strain Specific Primer Pairs		Product (bp)
		Sense	Sequence	
T30-1(+)	22	+	GTATCTCCGGAGCTCGATC	1613
T30-2(-)Cy3	1635	-	ACGCCTGCGAACCGCCGAC	
T30-3(+)	2267	+	TGGATGAGGGTTCTTCACCAC	824
T30-3(-)Cy3	3091	-	AGAATCGGGCAAAGCTTT	
T30-6(+)	6012	+	GCGTGTCTTAGTGTATCGCA	733
T30-6(-)Cy3	6745	-	TGGAAACAAAGGACTGTT	
T30-7(+)	7356	+	CAGCGTTCAAGTTTTCGTTA	913
T30-7(-)Cy3	8269	-	GACACTTTAGTACAATAATC	
T36-1(+)	68	+	AGCCTTTAAGCTCTAATA TT	1550
T36-2(-)Cy3	1618	-	ACAATCGAGCCAGGAACACTG	
T36-3(+)	2323	+	CTTCTTTTAACTCGACAAGGA	739
T36-3(-)Cy3	3062	-	TGTGATTATCAGGGAGTTA	
T36-5(+)	6040	+	CGGTAGCGTGTTGTGAGTACG	718
T36-5(-)Cy3	6760	-	GACGAAGAAACCTCGTTCGAC	
T36-6(+)	7370	+	ATTCCCTAATTCTAGGGCTC	597
T36-6(-)Cy3	7967	-	CTTTCTATCGAAACCTGCGAC	

* refers to the amplicons in the same sequence location on the CTV genome. There are 5 regions ranging from the 5' end to the centre of genome.

DNA amplicons from the 5' end of the genome of T30 and T36 clones were amplified and sequenced to confirm their identity. The 843 and 594 bp sized PCR products amplified with the forward primer T36 1+/T30 2+ and reverse primer T36 1-/T30 2- (Table 12) respectively were excised from the 1% agarose gel at the minimum UV exposure time, and purified using a MiniElute Gel Extraction kit (QiaGen, USA) according to the manufacturer's protocol and eluted into 15 µl of nuclease free water. The purified DNA product was then quantified (1 µl) by using agarose gel electrophoresis as well as with a Nanodrop spectrophotometer (Nanodrop Technologies, USA) to ensure the correct amount of template was added to the sequencing mix. Each sequencing reaction was performed by using automated fluorescent sequencing. The nucleotide sequences of the selected samples were determined in both directions using the aforementioned forward and reverse primers.

The sequencing reactions were performed using ABI PRISM BigDye Primer Cycle Sequencing Kits. In short, 100 ng of purified PCR product was added to a reaction mixture containing 3.2 pmol of the primer (T30 2+/T30 2-/T36 1+/ T30 1-); 2 µl Big Dye Terminator Ready reaction premix v3.1 (2.5X) (PE Applied Biosystems, V3.1, USA); 1 µl BigDye sequencing buffer (5X) (PE Applied Biosystems) and nuclease free water to a final volume of 10 µl. The reaction was performed in an Eppendorf Gradient Mastercycler (Eppendorf, Germany) with conditions: 94 °C for 1 minute, 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes. The DNA was purified by adding 1 µl of 125 mM EDTA; 1 µl of 3M sodium acetate (pH 4.6) and 25 µl of 100 % non-denatured ethanol. The DNA products were recovered by incubating for 15 minutes at room temperature followed by centrifugation at 13 000 g for 30 minutes and washing the pellet with 100 µl of 70% ethanol. The samples were centrifuged at 13 000 g for 15 minutes and allowed to dry on the bench for 15 minutes. The completed reaction was submitted to a sequencing facility at the University of Pretoria and run on an ABI 377 DNA sequencer (PE Applied Biosystems).

Table 12:

PCR primers used for the verification of the DNA clone sequence of strains T30 and T36 (Hilf *et al.*, 2000) used for PCR amplification.

Isolate	Position	Sense	Sequence	Product(bp)
T30-2(+)	792	+	TACGGCTTGGTGCTCTGAGGCC	843
T30-2(-)	1635	-	ACGCCTGCGAACCGCCGAC	
T36-1(+)	68	+	AGCCTTTAAGCTCTAATA TT	594
T36-1(-)	662	-	ACCAAGTCGGCTGTTTCGTC	

4.2.7 MICROARRAY HYBRIDIZATION

The Cy3-labelled amplicon targets were hybridized to immobilised oligonucleotide probes. Equal concentrations of PCR products from each strain were pooled and were denatured for 10 minutes at 96 °C, cooled on ice for 3-5 minutes and then kept on ice until hybridization. The slides were pre-treated with a solution of preheated 3.5 x SSC, 0.2 % SDS, 1 % BSA (Roche, Germany) made up to 50 ml with double distilled water. Incubation of 20 minutes at 60 °C was followed by rinsing three times with water for a few seconds. Hybridization was conducted in a custom made hybridization chamber overnight. Two different hybridization mixes were tested:

- (1) 100 µl of hybridization mix (Deyong *et al.*, 2005) containing 1 µg of purified product, 60 µl 20 x SSC, 4 µl 50 x Denhardt's solution (Sigma, USA), 2 µl 10 % SDS and ultrapure double distilled H₂O (Millipore, USA).
- (2) 50 µl of hybridization mix (D. Theron standard protocol at UP microarray facility) containing 1 ug of DNA, 25 µl formamide, 12.5 µl microarray hybridisation solution version 2 (product code RPK0325) (GE healthcare, USA) and ultrapure double distilled H₂O (Millipore, USA).

This was deposited under a 24 x 60 mm cover glass (Marienfeld). The DNA of the T30 and the T36 strains were hybridized separately at 42 °C, 52 °C and 60 °C overnight (15-17 hours) in a water bath and followed by a washing step. After the Deyong *et al* (2005) hybridization mix was selected different washing steps were tested. Two different washing step methods were tested:

1. three washes under low stringency at 42 °C (6 min in 2x SSC and 0.2 % SDS, 2 min in 0.2 x SSC and 0.2 % SDS, and 2 min in 0.075 x SSC) (Deyong *et al.*, 2005) and
2. six washes consisting of three washes at 42 °C (4 min in 1x SSC and 0.2 % SDS, and 4 min in 0.1x SSC and 0.2% SDS repeated) and three washes at room temperature (1 min in 0.1x SSC performed three times) (Danie Theron standard lab protocol at UP microarray facility).

4.2.8 DATA ANALYSIS

After hybridization the slides were analysed in a Genepix Personal 4000B microarray scanner (Molecular Devices, USA) at a wavelength 532 nm with a photomultiplier tube gain between 800-1000 to avoid overexposure. Data was analysed using Genepix 5.1 software programme. Spots that had low fluorescence intensity or poor morphology were “flagged” by the software during scanning; the remaining spots were then manually flagged if they appeared misshapen. All flagged spots were excluded from further analysis. The mean and standard deviation of the SNR (Signal-to-noise ratio) 532 of the replicates of each oligo were calculated. Spots with a SNR 532 of above 3 were regarded as a positive hybridisation. The SNR 532 is defined as the signal intensity median of the spot minus the background median signal and divided by the variation in the background. SNR is important for determining the confidence with which one can quantify a signal peak of a given value.

SNR Formula:

$$\text{SNR}_{532} = (\text{F}_{532} \text{ Median} - \text{B}_{532} \text{ Median}) / \text{B}_{532} \text{ Standard Deviation}$$

The % Coefficients of Variation (CV) compares data from different random variables to establish which show greater variance. This was calculated for replicates of each oligo as follows:

$$\% \text{ CV} = \text{Standard Deviation of data set} / \text{Mean of data set} \times 100$$

The closer the CV is to zero the less the variance. A cut-off of 20 % was set between closely positioned replicates.

A student t-test was performed to establish if there was any significant difference between the hybridization results of T30 and T36 targets at 42 °C and 52 °C. The t-test function on Microsoft Excel was used. The probability (P) result indicates whether two samples are likely to have come from the same two underlying populations that have the same mean.

Therefore the closer the P-value to zero, the more significant the difference between the hybridization results of T30 and T36.

4.3 RESULTS

4.3.1 CREATING A MICROARRAY PROTOCOL

The microarray protocol was assessed by testing different experimental strategies and adjusting each step to create a protocol that was most effective using the target T36 strain. Some parameters were pre-set as a set standard such as:

- The oligonucleotides were spotted in triplicate underneath each other to assess local hybridization variance.
- Two conserved probes were designed and arranged at different positions on the microarray slide to act as a positive control for both strains. Should any step of the protocol (efficiency of the Cy3 signal, hybridization step and washing stringency) have failed, no hybridization will be detected by these spots. Furthermore, the signal intensity of these spots could also be used for evaluation of the distribution of the hybridization buffer on the slides. Both conserved spots showed positive results, but with relatively low SNR 532 nm averages (data shown in next section). This result was unexpected as the probes are highly conserved between these two strains.
- Negative controls were included to assess cross-hybridization effects and prevent type-I-errors (false positives) of which some showed potential for use in a future array

To achieve an ideal experimental strategy with regards to signal intensity and reproducibility, various steps including the labelling method, hybridisation buffer constituents; and washing method were tested (as described individually below). The labelling methods were first tested with the standard microarray protocol and thereafter the different hybridization mixes were compared and finally the different washing strategies were tested and compared. When the microarray slides were hybridized with the T36 strain targets from PCR 5' Cy3-end labelling, with a Denhardt's based hybridization buffer and washed as described; more effective hybridization and signals were obtained (figures 25,26).

A. LABELLING AND AMPLIFICATION OF cDNA CLONES

The results of Cy3-5' end labelling by amplification of the eight different fragments (Table 11) used as targets for the microarray hybridization step are depicted in figures 17 and 18. The labelling method that was the most efficient and reproducible was 5' Cy3-end labelling in comparison with post-PCR Klenow labelling with the Cy3 dye, F3-UTP labelling and Cy3-UTP labelling (results not shown). With post-PCR Klenow labelling the concentration of Cy3 dye and therefore incorporation was extremely low with less than 0.02 pmol/ μ l Cy3 dye per reaction. Whereas Cy3-5' end labelling resulted in increased Cy3 dye concentrations of between 0.4-1.2 pmol/ μ l per amplification reaction, where 20-30 ng/ μ l of DNA product concentration was suitable. F3-UTP and Cy3-UTP labelling by incorporation in PCR was not successful in amplifying any amplicons and was therefore not considered for further use.

The Cy3 end labelling with PCR was the preferred method of labelling and amplicons comprised of different regions (Table 11) of the 5' end of the CTV genome of strains T30 and T36. The oligonucleotides within each region are depicted in Table 17. It was assumed that the majority of fragments were successfully labelled since the reverse primer was 5' end labelled and would have to be incorporated to produce the correct amplicon size. It is possible that with the commercial synthesis of the Cy3 end labelled reverse primers some fragments escaped labelling. The reverse primer was chosen to be synthesized with the Cy3 dye molecule on the 5' end to be able to label the reverse complement (negative sense) strand for hybridization to the oligonucleotides synthesized as the positive sense strand (complement).

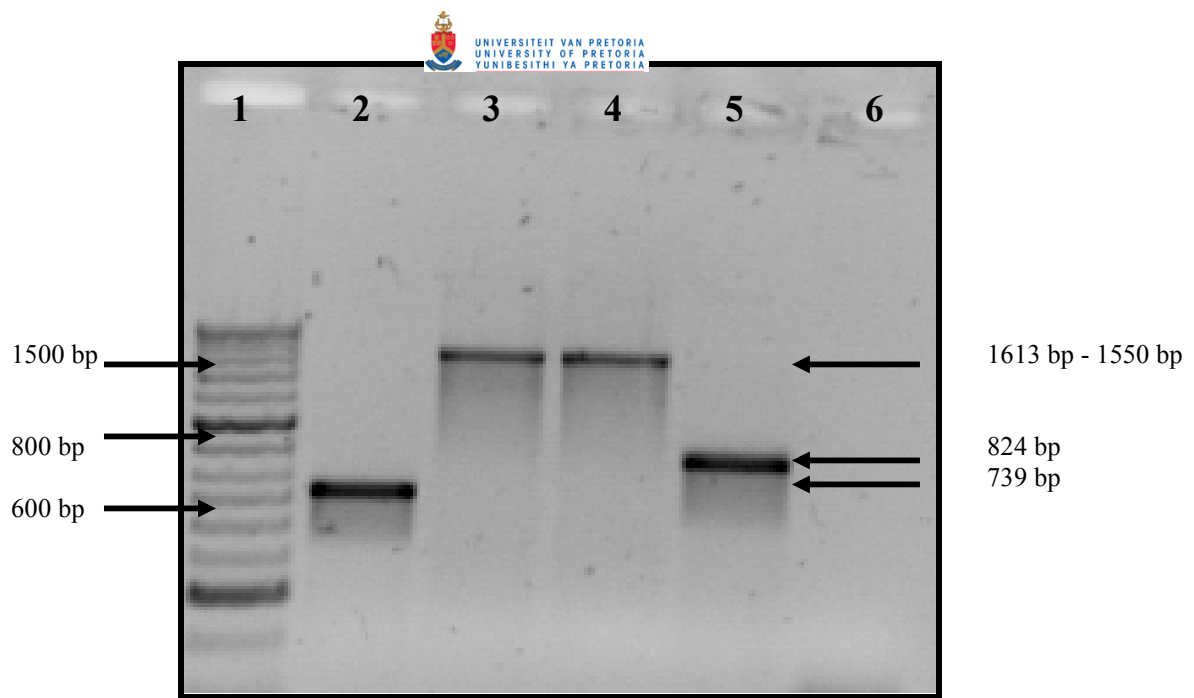


Figure 17: 5'-Cy3-labelling by amplification of targets for microarray hybridization. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) T36 3+/- of 739 bp; (3) T30 1+/- of 1613 bp; (4) T36 1+/- of 1550 bp (5) T30 3+/- of 824 bp; and (6) Negative control.

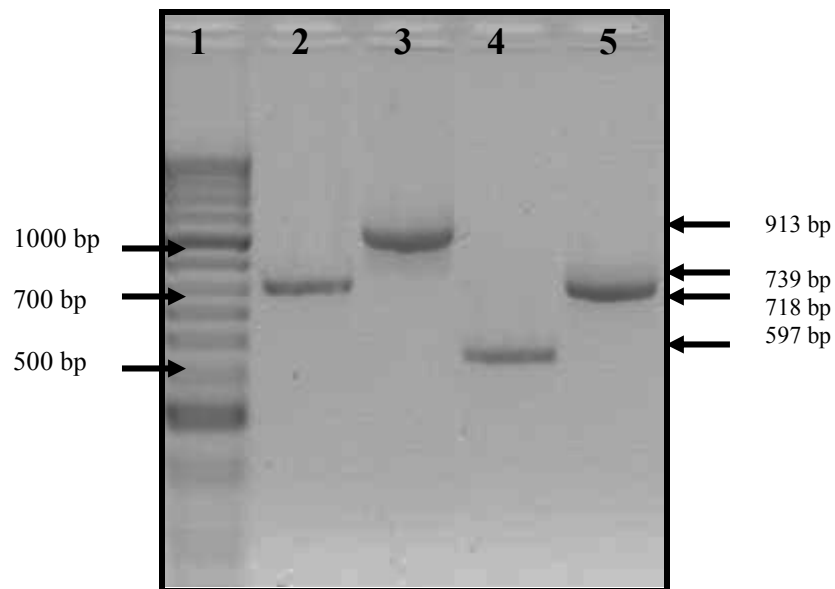


Figure 18: 5'-Cy3-labelling by amplification of targets for microarray hybridization. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) T36 5+/- of 739 bp; (3) T30 7+/- of 913 bp; (4) T30 6+/- of 733 bp and (5) T36 3+/- of 718 bp.

The hybridization mix containing Denhardt's solution described by Deyong *et al* (2005) showed greater signal intensity than the hybridization mix described by D. Theron and was therefore selected as the hybridisation mix of choice.

The washing step method of Deyong *et al* (2005) proved to be less stringent and the array showed signals for spots whereas the method of D. Theron proved to be too stringent and consequently few spots showed any signal (both methods described in Materials and Methods). The Deyong *et al.* (2005) method was therefore selected for future use in testing the array.

4.3.2 TESTING STRAINS T36 AND T30

A. FLAGGING AND FILTERING OF FEATURES

Various steps in the microarray experiment could induce experimental variation not related to the true variation that exists between the sequences of different CTV strains. Variation induced by (a) replicates of each oligonucleotide probe and (b) by replicates on different sides of the array was determined. The standard deviation and average of the SNR₅₃₂ of the set of replicates for each oligonucleotide was calculated. The variation induced between the replicates of each of the oligonucleotides were determined by comparing the average signal to noise ratio (SNR)₅₃₂ intensity against the standard deviation of the SNR₅₃₂ values by determining the % CV. Oligos with high % CV had possible replicate variation. The oligos with very high % CV were revisited to address any possible problems resulting in bad spots caused by possible slide or spotting defects or background fluorescing particles inducing unexpected high signals. Spots with increased signals due to experimental artefact/background that were above the rest of the replicate spots were flagged and removed from the data set.

The level of variability between replicates in different positions on the array was determined by comparing the average of the replicates on the left hand side of the slide and the average of all the features on the right hand side of the slide for any significant differences caused by defects on the slide or uneven distribution of hybridisation buffer. The variation on 4 slides, each representing a different hybridising strain target at 42 °C or 52 °C is described below:

1). With the hybridization of strain T30 at 42 °C (slide code 13257418) there were three oligos which showed unacceptably high interspot variability with %CV above 70% between all the replicates: *RabMF*; *40LNA-T36*, *T30-1+*; *T30-2+*; *T30-3+*, *T30-4+*, *T30-5+* and *VT 1+* (Table 13).

Oligos *59-T36*, *82-T36* and *VT-3+* had large spatial variance between the two sides of the slide, with the left side having a much higher average for the triplicate replicates than the right side. Oligos *T3-3+* and *T3-6+* had a much higher average for the triplicate replicates on the right side compared to the left hand side. The other oligos had relatively low %CVs between replicates and can be reproducibly used in further analysis in this study. There does seem to be an overall high amount of background noise possibly influencing the %CV on this slide.

2) With the hybridization of strain T30 at 52 °C (slide code 13257548) many oligos had high interspot variance (%CV) above 70 % between replicates and this can be attributed to a lot of background noise (Table 14). Even though the %CV of most oligos was above acceptable levels it was uniform across the slide. The effect of high background on each spot was assessed by ensuring that the mean signal intensity (F532) was twofold higher than the background means (B532). Each oligo will be considered separately on its particular values for further use in this study.

Oligos *T30 7+*, *VT 1+* and *VT 5+* had large spatial variance (%CV) between the two sides of the slide, with the left side having a much higher average for the triplicate replicates than the right side. Oligos *6-T36*, *40-T36* and *57-T36* had a much higher average for the triplicate replicates on the right side compared to the left hand side.

Table 13:

The SNR 532 averages and standard deviation results for each oligonucleotide after filtering hybridized against strain T30 at 42 °C.

Oligonucleotide	Ave± Std Dev	%CV	Left Ave ± Std Dev	%CV left	Right Ave± Std Dev	%CV right
2-T36	1.8± 1	55%	1.7±0.6	35%	1.7±0.9	57%
6-T36	7.9±2.3	29%	8.6±1.6	19%	7.3±2.7	35%
40-T36	1.8±1.2	67%	1±0.6	60%	2.5±1.8	72%
57-T36	1.5± 0.9	60%	0.6±0.4	67%	2.6±1.4	54%
59-T36^a	7.1± 0.6	8%	13.5±1	7%	0.8±0.13	11%
64-T36	64.9± 7.2	11%	80±6.9	8.6%	49±7.6	15%
82-T36^a	1.9± 0.7	39%	3.5±1.3	37%	0.2±0.1	62%
123-T36	2.9± 1.5	51%	3.6±1.7	47%	2.2±1	45%
228-T36	2± 0.9	45%	1.9±1.1	69%	2.4±0.6	25%
230-T36	4.2± 1.3	31%	4.2±1.4	33%	4.1±1.4	34%
264-T36	2.4± 0.9	38%	1.6±0.5	33%	3.08±0.7	21%
268-T36	3.1± 1.5	48%	2±0.6	29%	4.2±1.3	31%
Con1	2.6± 1.7	65%	3± 1.8	61%	2.3±1.4	63%
Con2	12.4 ± 1.6	12%	13±3	23%	11.9±0.1	0.8%
WNVI	2± 0.8	40%	2.5±0.8	32%	1.4±0.7	50%
WNV-prM	2.8± 1.2	43%	2.7±1.3	49%	2.9±1.2	41%
RabMF	1.4± 1.0	72%	0.3±0.0	3%	2.1±0.7	33%
40LNA-T36	1± 0.7	70%	1.1±0.7	60%	0.9±0.6	64%
230LNA-T36	4.7± 1.7	35%	4.2±1.7	40%	5.1±1.7	33%
264LNA-T36	2± 0.4	20%	2.2±0.5	22%	2±0.15	7%
T3 2+	4.6± 0.8	17%	4.7±0.6	12%	4.5±0.9	20%
T3 3+^a	4± 3.7	93%	0.5±0.5	100%	7.1±0.9	14%
T3 5+	3.5± 0.9	25%	4.2±0.8	19%	2.8±0.2	8%
T3 6+^a	5.1± 3.3	65%	7.6±1.5	20%	2.6±2.6	100%
T36 1+	1.2± 0.25	20%	1.6±0.3	21%	0.8±0.2	19%
T36 2+	0.7± 0.1	14%	0.5± 0.04	7%	0.8± 0.1	11%
T36 3+	2.1± 0.7	33%	3±0.4	11%	1.2±1.0	86%
T36 5+	2.3± 0.5	25%	1.2±0.3	25%	3.3±0.8	24%
T36 6+	1.0± 0.4	40%	1.0±0.3	26%	1±0.5	50%
T36 7+	5.1± 1.0	19%	6.9±1.1	14%	3.2±0.8	24%
T30 1+	0.6± 0.5	83%	0.8±0.6	75%	0.5±0.65	130%
T30 2+	1.3± 0.8	62%	1.9±0.5	26%	0.7±0.3	42%
T30 3+	0.3± 0.6	200%	0.7±0.5	71%	0.1±0.18	180%
T30 4+	0.8± 0.6	75%	1.04±0.4	38%	0.6±0.3	50%
T30 5+	0.7± 0.6	86%	1.1±0.3	27%	0.4±0.5	125%
T30 6+	1± 0.4	40%	1.4±0.4	27%	1.5±0.4	26%
T30 7+	5.4± 2.2	40%	5.8±3.3	57%	4.9±0.3	5%
VT 1+	1.8± 1.5	83%	1±0.5	50%	2.7±1.8	66%
VT 2+	3.3± 1.2	36%	2.4±0.8	34%	4.2±0.5	12%
VT 3+^a	2.4± 2	83%	4.2±0.3	7%	0.6±0.6	100%
VT 4+	1.8± 1.2	67%	2.7±0.9	34%	0.9±0.4	44%
VT 5+	29± 14.8	51%	23±6.6	28%	34±20.6	60%
VT 6+	6± 2.9	48%	4.9±1.4	29%	6.2±3.9	62%
Water	1.9± 0.3	16%	N/A	N/A	N/A	N/A
Averages		49%		54%		44%

^a Oligos highlighted as bold indicate high variance between left and right sides of the slide

Table 14:

The SNR 532 averages and standard deviation results for each oligonucleotide after spot filtering hybridized against strain T30 at 52 °C.

Oligonucleotide	Ave±SD	%CV	Left Ave±SD	%CV left	Right Ave±SD	%CV right
2-T36	0.9±0.7	76%	0.7±0.9	128%	1.3±0.3	23%
6-T36^a	2.4±2	83%	0.7±0.5	68%	4±1.4	35%
40-T36^a	3.2±2.8	88%	0.8±0.5	63%	5.6±1.4	25%
57-T36^a	3.2±3.4	106%	1.5±1	67%	5±4.2	84%
59-T36	0.6±0.2	31%	0.4±0.09	20%	0.8±0.27	33%
64-T36	1.2±0.6	50%	1.3±0.7	53%	0.8±0.01	13%
82-T36	1.2±1.5	125%	2.6±1.3	52%	0.3±0.2	65%
123-T36	7.8±3	38%	7.4±4.1	56%	8.4±0.4	4%
228-T36	0.8±0.3	38%	0.9±0.5	52%	0.7±0.09	13%
230-T36	0.9±1.3	144%	0.4±0.6	150%	1.7±1.9	111%
264-T36	1.5±0.7	47%	1.6±0.8	50%	1.0±0.1	1%
268-T36	1±1.4	140%	0.24±0.01	4%	1.7±1.9	111%
Con1	1.1±0.8	72%	1.7±0.8	47%	0.5±0.02	3%
Con2	3.6±3.2	89%	2.8±3.4	121%	4.5±3.4	75%
WNV1	0.7±0.8	114%	1.0±1.1	111%	0.4±0.6	150%
WNV-prM	0.6±0.5	83%	0.4±0.2	50%	2±0.7	35%
RabMF	0.95±0.9	95%	0.5±0.6	120%	1.4±1.1	79%
40LNA-T36	2.1±1.9	90%	1.6±2.1	131%	2.6±1.9	73%
230LNA-T36	1.4±1	71%	1.4±1.3	95%	1.4±0.65	45%
264LNA-T36	1.4±0.3	21%	1.1±0.01	0.5%	1.8±0.6	33%
T3 2+	5.9±4.7	80%	5.3±7.4	139%	6.5±0.9	14%
T3 3+	6.2±3.2	52%	4.7±3.4	72%	8.4±1.2	14%
T3 5+	0.1±0.16	160%	0.1±0.3	300%	0.1±0.001	10%
T3 6+	6.7±2.3	34%	7.6±2.5	30%	5.9±2.1	40%
T36 1+	1.1±0.63	60%	0.8±0.7	88%	1.3±0.6	46%
T36 2+	1.6±1.4	88%	0.7±0.1	14%	2.5±1.6	64%
T36 3+	0.9±0.9	100%	0.9±1	111%	0.9±1	111%
T36 5+	1.8±0.9	50%	2.4±0.7	29%	1.1±0.7	64%
T36 6+	1.7±1.4	82%	2.5±1.6	64%	1±0.7	70%
T36 7+	0.6±0.2	33%	0.6±0.1	17%	0.6±0.2	33%
T30 1+	1.4±1.3	93%	0.3±0.6	200%	2.4±0.6	26%
T30 2+	5.2±4.7	90%	4.7±7.1	151%	5.6±1.7	30%
T30 3+	1.6±1.4	88%	1.7±2	121%	1.5±0.6	40%
T30 4+	2.4±2.6	108%	3±3	100%	1.7±2.5	147%
T30 5+	0.4±0.2	50%	0.3±0.3	100%	0.4±0.07	18%
T30 6+	5.8±2.2	38%	6.7±2.5	38%	4.4±1	23%
T30 7+^a	2.65±2.7	100%	4.9±1.5	30%	0.4±0.1	25%
VT 1+^a	3±2.9	97%	5.5±1.3	24%	0.5±0.6	120%
VT 2+	1.7±0.5	29%	1.8±0.4	22%	1.6±0.6	38%
VT 3+	5.5±2.3	42%	6.8±1.7	25%	4.3±2.3	53%
VT 4+	4.4±3.5	80%	3±4.9	163%	5.7±1.3	22%
VT 5+^a	3±2.9	97%	5.1±2.9	57%	1±0.2	20%
VT 6+	4±1.4	35%	5±1	25%	3.1±1	32%
Water	1±0.3	30%	N/A		N/A	
Averages		75%		78%		48%

^a Oligos highlighted as bold indicate high variance between left and right sides of the slide

3. With the hybridization of strain T36 at 42 °C (slide code 13257420) there were a few oligos *64-T36*, *Con2*, *T3-3+* and *T30-1+* which showed unacceptably high interspot variance of above 70 % CV between replicates (Table 15).

Oligos 59-T36, Con2, T3 3+ and 64-T36 had large spatial variance between the two sides, with the left side having a much higher average for the replicates than the right side. Oligo T30-1+ had a much higher average for the replicates on the right side compared to the left side. All the rest of the oligos on this slide had low %CV and therefore low variability between replicates and so can be reproducibly used in further analysis in this study. The slide had low background noise and represents a model example of acceptable levels of background.

4. With the hybridization of strain T36 at 52 °C (slide code 13257421) there were quite a few oligos which showed unacceptably high interspot variance between replicates (Table 16). This was attributed to the large amount of background noise on the slide which increased the signal intensity. The effect of high background on each spot was assessed by ensuring that the mean signal intensity (F532) was twofold higher than the background means (B532).

Oligos *230-T36* and *VT-1+* had large spatial variance between the two sides of the slide, with the left side having a much higher average for the replicates than the right side. Oligos *6-T36*, *57-T36*, *59-T36*, *T36-6+*, *T36-7+* and *40LNA-T36* had a much higher average for the replicates on the right side compared to the left side. Even though many of the spots had large variances each oligo will be assessed for further use in this study.

Table 15:

The SNR 532 averages and standard deviation results for each oligonucleotide after spot filtering hybridized against strain T36 at 42 °C.

Oligonucleotide	Ave±SD	%CV	LeftAve±SD	%CV left	Right Ave±SD	%CV right
2-T36	6.2±0.9	14%	6.4±1.2	19%	5.9±0.6	10%
6-T36	41±4.5	11%	42±4.9	12%	39.5±4.9	12%
40-T36	2.7±1.4	51%	3.5±2	57%	2.2±0.4	18%
57-T36	3.4±0.65	19%	3.7±0.2	5%	3.2±0.7	21%
59-T36^a	137±83	60%	211.4±21.7	10%	62±18	29%
64-T36^a	4.9±3.2	65%	7.6±2	26%	2.3±1.1	47%
82-T36	3.1±1.5	48%	3.8±2.2	58%	2.5±0.1	4%
123-T36	25.1±2	8%	32.9±2.1	16%	17.3±0.1	0.5%
228-T36	24±8.6	36%	29.3±8.1	27%	18.3±5.2	28%
230-T36	5.6±2.1	37%	9.5±1.3	14%	6±0.5	8%
264-T36	7.5±1.9	25%	9.4±0.1	1%	5.6±0.3	5%
268-T36	6.4±2.2	34%	7.9±0.7	9%	4.2±1.4	33%
Con1	3.1±1.1	35%	2.4±0.5	21%	3.8±1.2	31%
Con2^a	1.6±2.1	131%	2.7±2.6	96%	0.5±0.9	18%
WNVI	3.1±0.9	29%	2.9±1.3	45%	3.3±0.8	24%
WNV-prM	2.6±0.6	23%	2.9±0.1	3%	2.3±0.8	34%
RabMF	2.4±0.7	29%	2.9±0.5	17%	1.9±0.6	32%
40LNA-T36	3.6±1.1	31%	3.3±1.3	39%	3.9±1	26%
230LNA-T36	6.6±3.3	50%	5.6±2.6	46%	7.6±2.9	40%
264LNA-T36	7.5±2.2	29%	7.4±1.2	16%	7.7±3.2	42%
T3 2+	4.5±1.9	42%	5±2.6	52%	4.1±1.6	40%
T3 3+^a	3.7±3.2	86%	6.3±2.3	37%	1.2±0.6	50%
T3 5+	2.1±0.4	19%	2.4±0.2	7%	1.7±0.3	18%
T3 6+	9.4±4.5	48%	12.2±4	33%	6.6±3.4	52%
T36 1+	1.9±1	52%	2.6±0.7	27%	1.6±0.5	31%
T36 2+	0.9±0.5	56%	0.9±0.6	67%	0.9±0.5	55%
T36 3+	2.4±0.6	25%	2.4±0.3	13%	2.4±0.9	4%
T36 5+	2.4±0.8	33%	2.5±1.2	48%	2.4±0.2	8%
T36 6+	1±0.7	70%	1.4±0.8	57%	0.7±0.3	43%
T36 7+	1.4±1.2	86%	2.3±0.8	35%	0.5±0.02	4%
T30 1+^a	3.8±3	79%	1.5±1.9	13%	6.2±1.5	24%
T30 2+	4.1±1.5	37%	2.9±0.4	14%	5.5±0.8	15%
T30 3+	22.4±5.3	24%	18.8±5.5	29%	26±1.3	5%
T30 4+	3.4±1.3	38%	3.5±1.8	51%	3.3±1.1	33%
T30 5+	3.1±0.7	22%	2.8±0.3	11%	3.3±0.9	27%
T30 6+	3.3±1.3	39%	3.5±1.9	56%	3.2±0.9	29%
T30 7+	2.6±0.8	31%	2.6±0.8	30%	2.7±1	37%
VT 1+	2.1±0.6	29%	1.9±0.4	20%	2.3±0.8	35%
VT 2+	3.5±1.5	42%	3.3±1	30%	3.6±2.6	72%
VT 3+	4.2±0.4	9%	4.1±0.5	12%	4.4±0.4	9%
VT 4+	3.1±1.4	45%	3±1.8	60%	3.3±1.3	39%
VT 5+	4.4±1.8	41%	3.1±0.3	9%	6.3±0.7	11%
VT 6+	12.6±7.8	62%	9.6±3.5	36%	15.7±10.6	68%
Water	1.3±0.8	60%	N/A		N/A	
Averages		42%		30%		27%

^a Oligos highlighted as bold indicate high variance between left and right sides of the slide

Table 16:

The SNR 532 averages and standard deviation results for each oligonucleotide after spot filtering hybridized against strain T36 at 52 °C.

Oligonucleotide	Ave±SD	%CV	Left Ave±SD	%CV Left	Right Ave±SD	%CV Right
2-T36	0.5±0.6	120%	0.3±0.1	33%	0.6±0.9	150%
6-T36^a	11.4±10.2	89%	1.8±0.5	28%	21.1±8.8	41%
40-T36	1.8±1.9	105%	0.9±0.6	67%	2.8±2.4	86%
57-T36^a	2.8±1.6	57%	1.6±1.1	67%	4.1±0.9	22%
59-T36^a	2.3±3.4	150%	0±0.1	-	4.6±3.6	78%
64-T36	0.5±0.5	100%	0.7±0.6	86%	0.4±0.3	75%
82-T36	0.8±0.6	75%	1.2±0.5	42%	0.4±0.15	38%
123-T36	1.2±0.9	75%	1.3±0.6	46%	1.1±1.4	130%
228-T36	7.5±1.3	17%	6.7±0.6	9%	9.2±0.1	1%
230-T36^a	2.5±1.8	72%	4.2±0.02	0.4%	0.9±0.7	78%
264-T36	0.7±0.9	128%	2.6±1.4	53%	2.3±0.4	17%
268-T36	5.9±2.8	47%	7.2±2	27%	4.5±3.2	71%
Con1	1.6±1	62%	0.9±0.6	67%	2.4±0.8	33%
Con2	0.9±0.95	105%	1.6±0.7	43%	0.2±0.4	200%
WNV1	1.9±1.1	58%	1.2±0.6	50%	2.9±1.3	45%
WNV-prM	0.5±0.7	140%	1±0.7	70%	0.07±0.3	400%
RabMF	2.9±1.4	48%	2.2±1.5	68%	3.5±1.3	37%
40LNA-T36^a	1.7±1.5	88%	0.7±0.6	86%	2.7±1.5	56%
230LNA-T36	4.2±2.5	59%	4±1.5	38%	4.4±3.3	75%
264LNA-T36	0.7±1.1	157%	1.1±1.2	110%	0±0.02	-
T3 2+^a	3±2.8	93%	5.6±0.4	7%	0.6±0.4	67%
T3 3+^a	3±2.4	80%	5±1.6	32%	1.1±0.7	64%
T3 5+	1±0.9	90%	1.5±1.2	80%	0.5±0.05	10%
T3 6+^a	3.9±4.2	107%	7.2±3.4	47%	0.6±0.6	100%
T36 1+	3.1±1.6	51%	1.6±1.6	100%	4.2±0.4	9%
T36 2+	2.4±0.4	17%	1.8±0.1	5%	2.6±0.2	7%
T36 3+	2.2±1.0	45%	1.5±0.2	13%	3.3±0.3	9%
T36 5+	2.8±1.3	46%	1.7±0.3	17%	3.9±0.5	13%
T36 6+^a	2±1.3	65%	0.9±0.5	56%	3.2±0.3	9%
T36 7+^a	1.9±1.3	68%	0.9±0.4	44%	3±0.4	13%
T30 1+	0.5±0.5	100%	0.6±0.7	116%	0.5±0.4	80%
T30 2+	0.2±0.3	150%	0.3±0.4	133%	0±0.07	-
T30 3+	0.1±0.3	300%	0.1±0.3	300%	0±0.09	-
T30 4+	0.6±1.1	183%	1±1.3	130%	0±0.23	-
T30 5+	1.1±0.6	54%	0.7±0.4	57%	1.6±0.2	12%
T30 6+	1.7±1.4	82%	0.7±0.8	114%	2.5±1.2	48%
T30 7+	0.7±0.5	71%	0.8±0.7	88%	0.6±0.4	67%
VT 1+^a	2.9±2.5	86%	4.6±2.3	50%	1.3±1.5	115%
VT 2+	6.3±1.8	28%	6.2±2.7	43%	6.3±0.6	9%
VT 3+	0.1±0.1	100%	0.1±0.05	50%	0.1±0.2	200%
VT 4+	0.5±0.3	60%	0.3±0.6	200%	0.6±0.1	16%
VT 5+	0.9±1.7	188%	2.1±1.8	86%	0±0.1	-
VT 6+	0.4±0.5	125%	0.8±0.3	37%	0±0.08	-
Water	1.6±0.6	38%	N/A	-	N/A	-

^a Oligos highlighted as bold indicate high variance between left and right sides of the slide

The results for each oligo will include: 1) the SNR 532 average of all replicates for two different hybridization temperatures (42 °C and 52 °C); and 2) comparison of the two different strains T30 and T36 used as the hybridizing templates (Figures 19-25). The results also will include for each oligo the: (a) position of the mismatch/es; (b) the percentage of mismatches against the non-target strain; and (c) which labelled amplicon region was used as the target for hybridization (Table 17). Three LNA modified oligos (40LNA-T36, 230LNA-T36 and 264LNA-T36) were compared with oligos not containing LNA modified bases. Of the 20 T36 specific oligonucleotides, 14 had increased intensity at a hybridisation temperature of 42 °C compared with 52 °C for hybridization against strain T36. At a temperature of 60 °C no oligonucleotides hybridized and no signal was found for any spots (results not shown). However it must be noted that a few oligonucleotides performed better at 52 °C and it was clear that during development of the technique, each oligonucleotide should be reviewed individually for potential use.

Two student t-tests were performed to identify if hybridization of the T36 target to each T36-specific oligo was more significant than the hybridization of T30 target to same oligos at (a) 42 °C and (b) 52 °C.

The Null hypothesis: There is no significant difference between hybridization of T36 and T30 targets to T36 specific oligos.

- a) At 42°C the t distribution for 28 degrees of freedom, the probability that $t > 1.19$ is 0.26. Also the probability that $t < -1.19$ is 0.26. At 95 % ($\alpha = 0.025$) confidence interval with $t \pm 1.19$: $-6.945 < \mu < 21.53$. Therefore $t\text{-value} > \alpha$ and the Null hypothesis is rejected in favour of the alternative hypothesis, where the difference is significant.
- b) At 52°C the t distribution for 28 degrees of freedom, the probability that $t > 0.98$ is 0.37. Also the probability that $t < -0.98$ is 0.37. At 95 % ($\alpha = 0.025$) confidence interval with $t \pm 0.98$: $0.96 < \mu < 3.5$. Therefore $t\text{-value} > \alpha$ and the Null hypothesis is rejected in favour of the alternative hypothesis, where the difference is significant.

Table 17:

A description and function of the oligos used in the array and their targets and non-targets.

Oligonucleotide	Target strain/s	^a Amplicon region for hybridization	^b Position of mismatches	^c Amount of mismatches (bp)	Total Length (bp)
2-T36	T36 strain	1	scattered	6	19
6-T36	T36 strain	4	5' end	7	20
40-T36	T36 strain	1	5' end	4	21
57-T36	T36 strain	3	3' end	7	20
59-T36	T36 strain	1	scattered	6	18
64-T36	T36 strain	1	centre	3	19
82-T36	T36 strain	2	centre	4	19
123-T36	T36 strain	5	3' end	7	20
228-T36	T36 strain	4	scattered	10	20
230-T36	T36 strain	3	3' end	3	22
264-T36	T36 strain	5	scattered	4	22
268-T36	T36 strain	4	5' end	12	20
Con1	T36 & T30	2	none	0	18
Con2	T36 & T30	2	none	0	20
WNVI	None	none	whole oligo	18	18
WNV-prM	None	none	whole oligo	27	27
RabMF	None	none	whole oligo	24	24
40LNA-T36	T36 strain	1	5' end	4	21
230LNA-T36	T36 strain	3	3' end	3	22
264LNA-T36	T36 strain	5	scattered	4	22
T3 2+	T3 strain	2	scattered	6-7	20
T3 3+	^a None	none	scattered	2-4	21
T3 5+	T3 strain	4	scattered	7-9	23
T3 6+	^a None	none	scattered	4-5	19
T36 1+	T36 strain	1	scattered	14	20
T36 2+	T36 strain	2	scattered	12	20
T36 3+	T36 strain	3	scattered	18	21
T36 5+	T36 strain	4	scattered	11	21
T36 6+	T36 strain	5	scattered	12	20
T36 7+	^a None	none	scattered	11	21
T30 1+	T30 strain	1	whole oligo	19	19
T30 2+	T30 strain	2	scattered	7	22
T30 3+	T30 strain	3	scattered	11	21
T30 4+	^a None	none	scattered	3	23

T30 5+	^aNone	none	scattered	5	21
T30 6+	T30 strain	4	scattered	10	20
T30 7+	T30 strain	5	scattered	11	20
VT 1+	VT strain	1	whole oligo	19	19
VT 2+	VT strain	2	scattered	7	22
VT 3+	VT strain	3	scattered	6-8	21
VT 4+	^aNone	none	scattered	6	23
VT 5+	^aNone	none	scattered	3-4	18
VT 6+	VT strain	4	scattered	8-9	20

^a Refers to all the amplicons in the same sequence location on the CTV genome. There are 5 regions ranging from the 5' end to the central area of genome.

^b The amplicons available do not target the sequence region of the oligo, and serves here as a negative control.

^c The amount of mismatches to the non-target strain

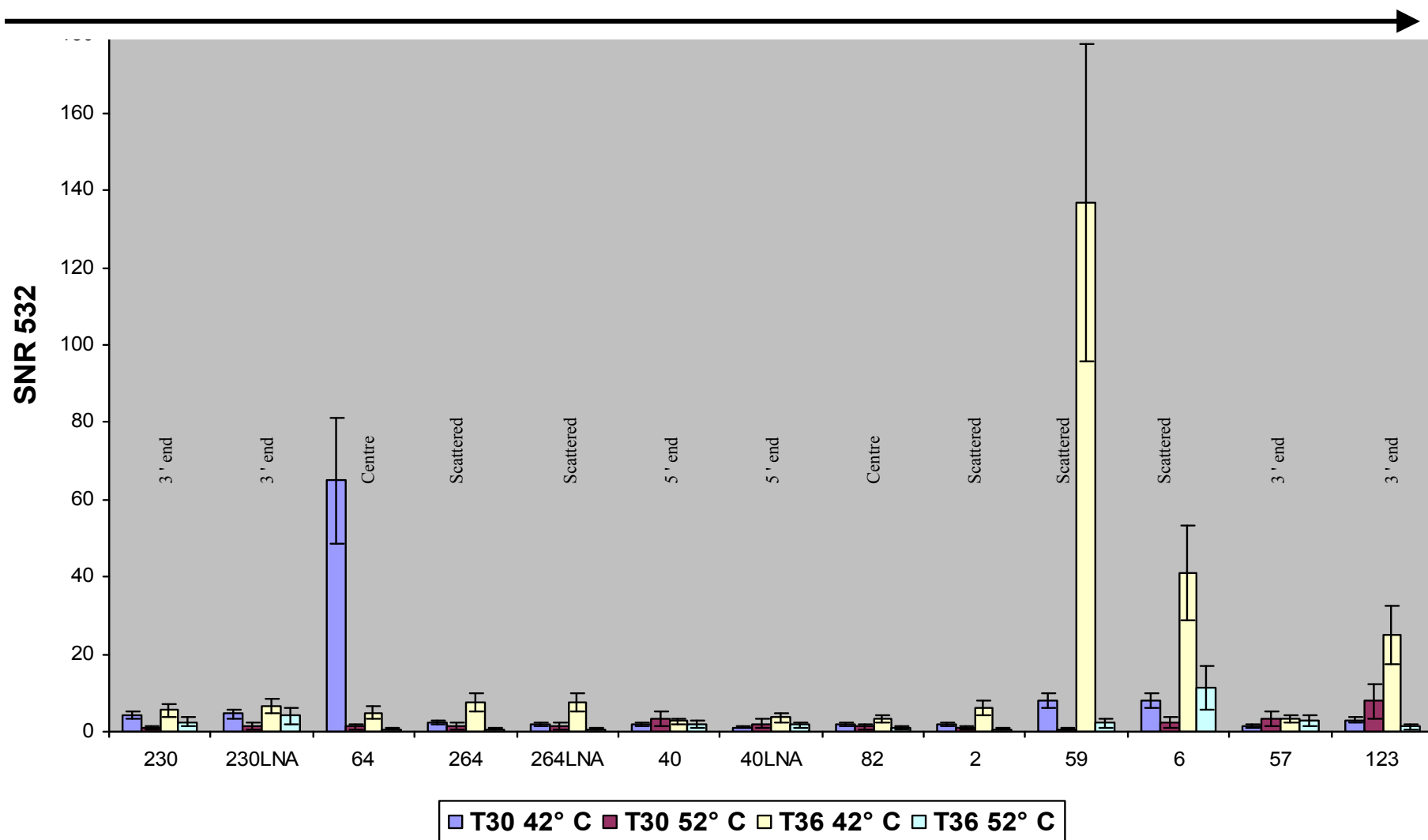


Figure 19: The SNR 532 Averages of the T36 specific oligonucleotides at 42 °C and 52 °C hybridized against strains T30 and T36. The X-axis represents the oligonucleotide name and the Y-axis represents the Average SNR 532. The arrow shows an ascending percentage of mismatches from 14-35 %. The approximate location of the mismatches are shown above each oligo.

Standard error bars are represented above each bar depicted in the figure.

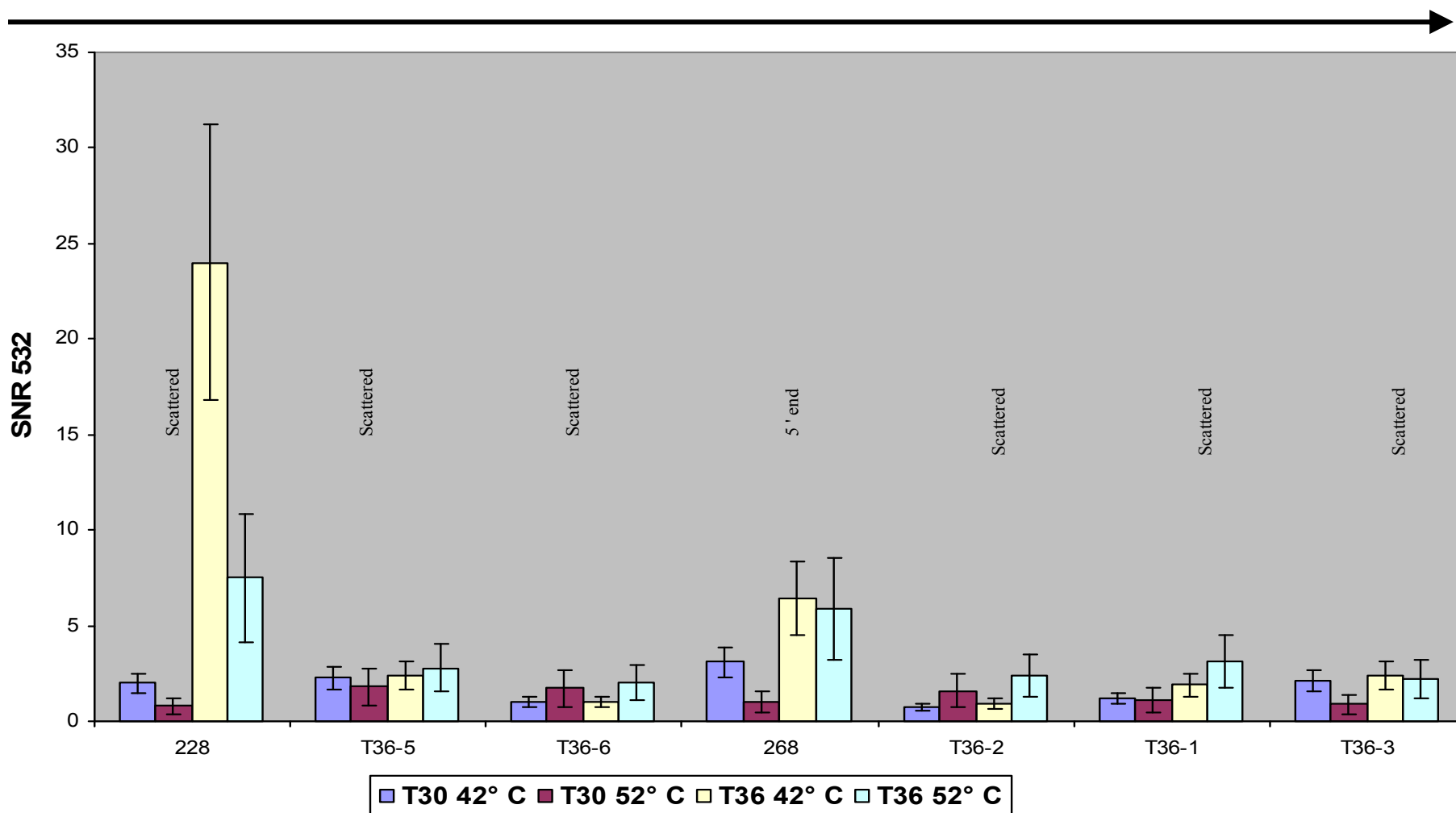


Figure 20: The SNR 532 Averages of the T36 specific oligonucleotides at 42 °C and 52 °C hybridized against strains T30 and T36. The X-axis represents the oligonucleotide name and the Y-axis represents the Average SNR 532. The arrow shows an ascending percentage of mismatches from 50-86 %. The approximate locations of the mismatches are shown above each oligo.

Standard error bars are represented above each bar depicted in the figure.

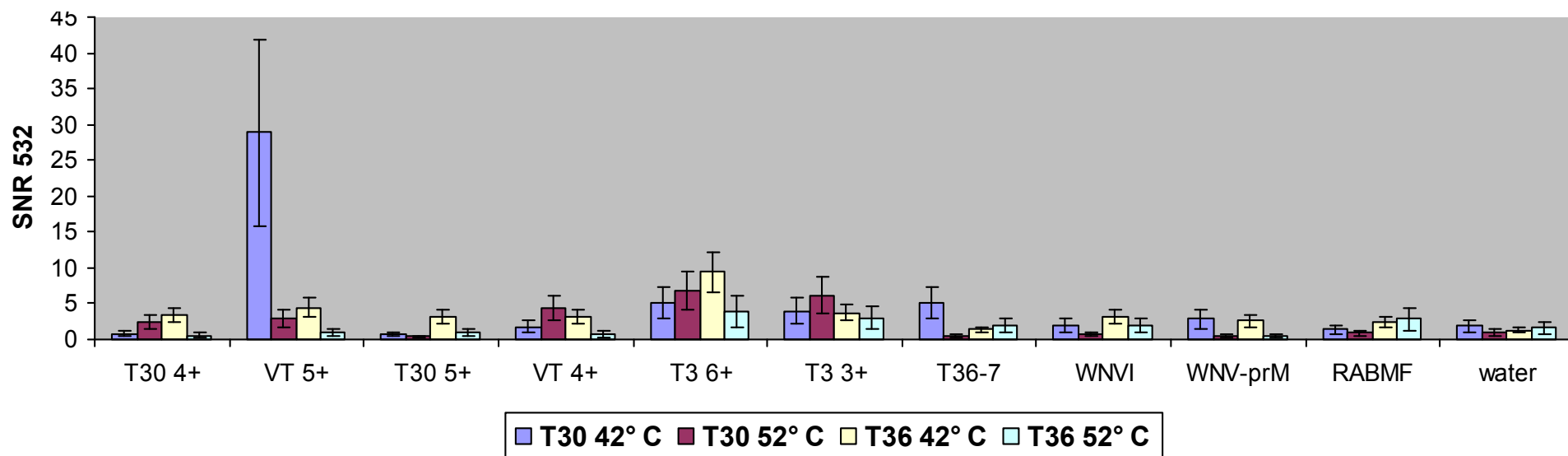


Figure 21: The SNR 532 Averages of the negative control oligonucleotides at 42 °C and 52 °C hybridized against strains T30 and T36. The X-axis represents the oligonucleotide name and the Y-axis represents the Average SNR 532. Standard error bars are represented above each bar depicted in the figure.

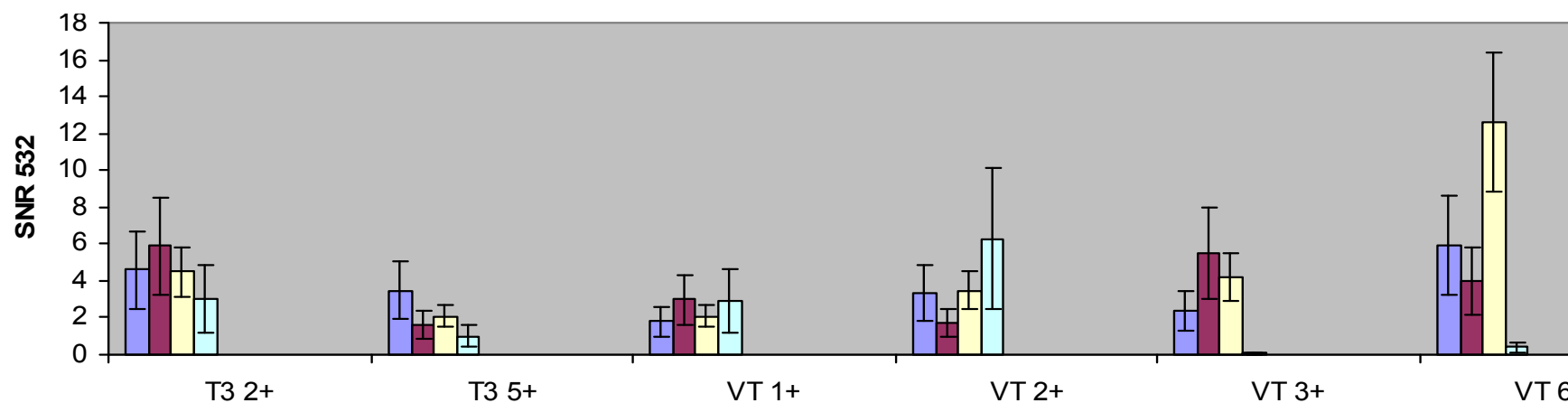


Figure 22: The SNR 532 Averages of the CTV strain negative control oligonucleotides of strains T30 and T36 at 42 °C and 52 °C hybridized against strains T30 and T36. The X-axis represents the oligonucleotide name and the Y-axis represents the Average SNR 532. Standard error bars are represented above each bar depicted in the figure.

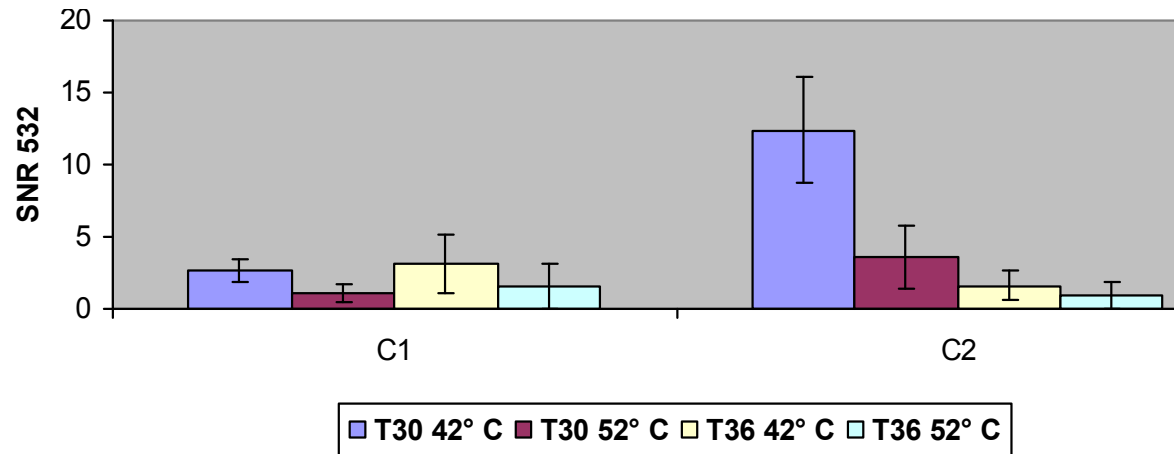


Figure 23: The SNR 532 Averages of the CTV conserved oligonucleotides at 42 °C and 52 °C hybridized against strains T30 and T36. The X-axis represents the oligonucleotide name and the Y-axis represents the Average SNR 532.

Standard error bars are represented above each bar depicted in the figure.

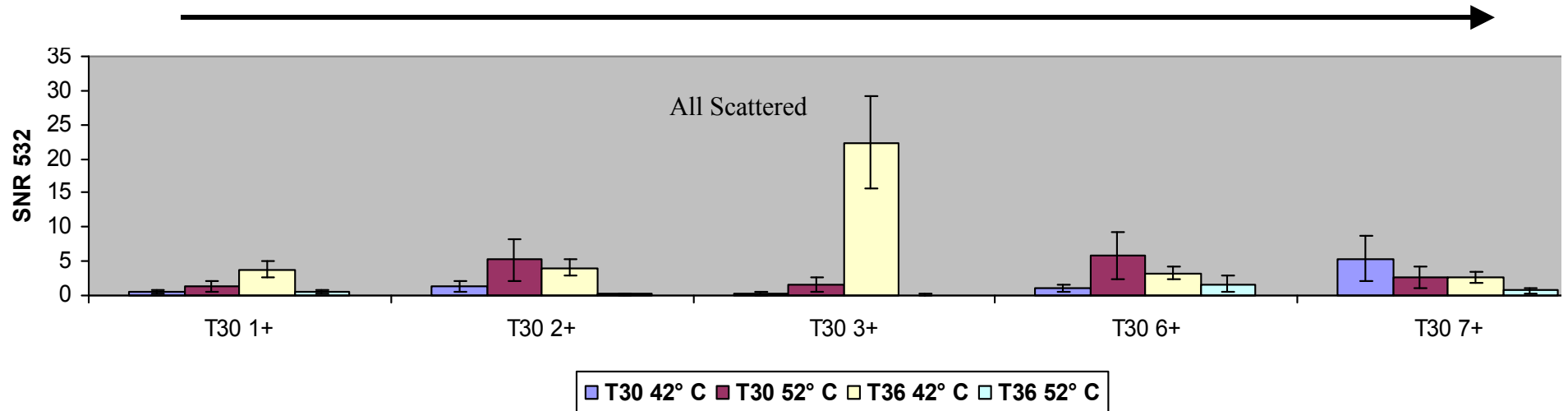


Figure 24: The SNR 532 Averages of the T30 specific oligonucleotides at 42 °C and 52 °C hybridized against strains T30 and T36. The X-axis represents the oligonucleotide name and the Y-axis represents the Average SNR 532. The arrow shows an ascending percentage of mismatches from 32-100 %. The mismatches are scattered.

Standard error bars are represented above each bar depicted in the figure.

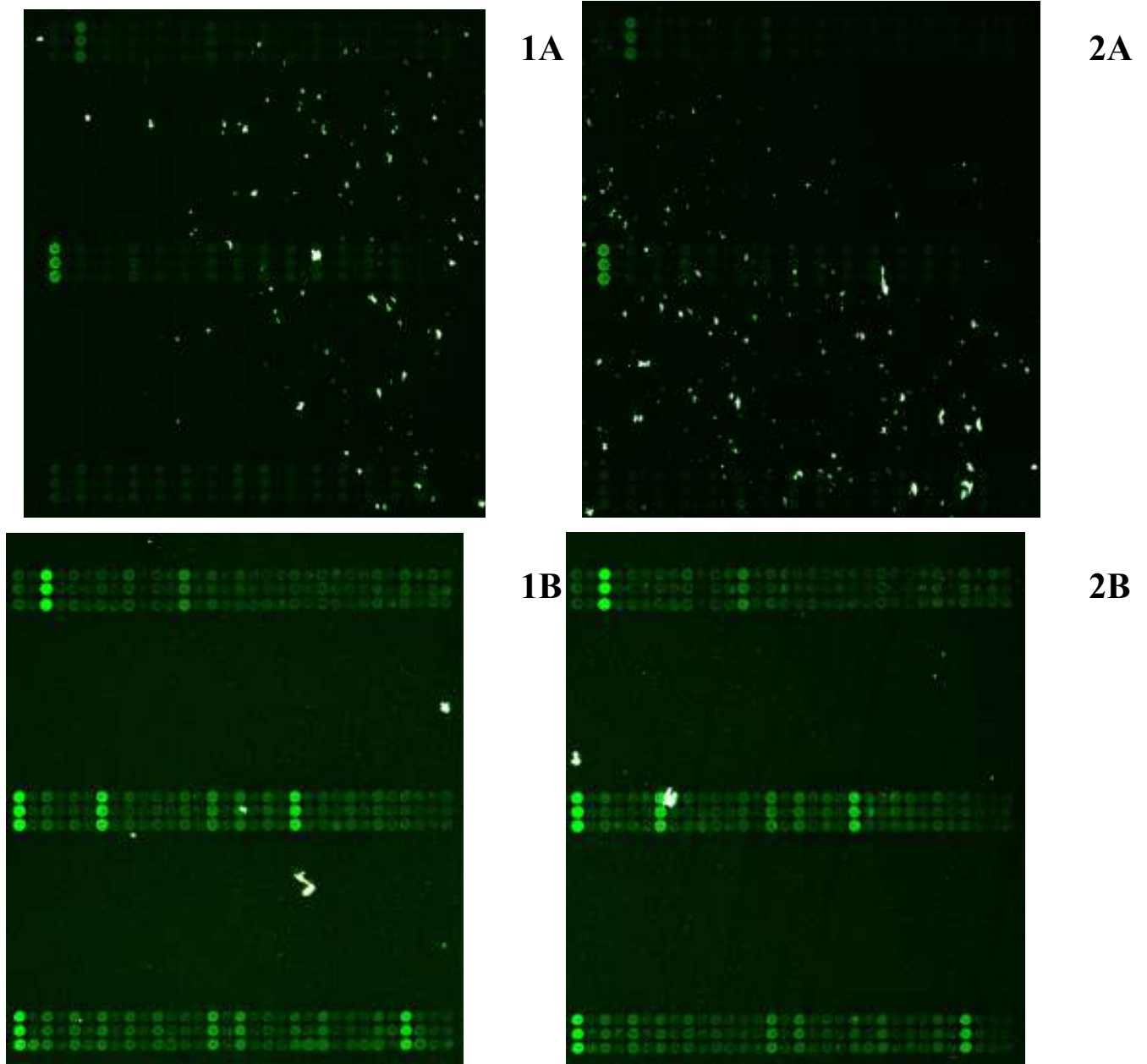


Figure 25: Hybridization results of the T36 strain at 52 °C (Images 1A-B) and at 42 °C (Images 2A-B) on a two-dimensional array. Cy3-labelled DNA amplicons were hybridized to bound oligonucleotides on glass slides. Fluorescent patterns were recorded with a Genepix 400B microarray scanner (Molecular Devices, USA) at a wavelength of 532 nm. Images 1A-B represents the left side and images 2A-B represent the right side of the s

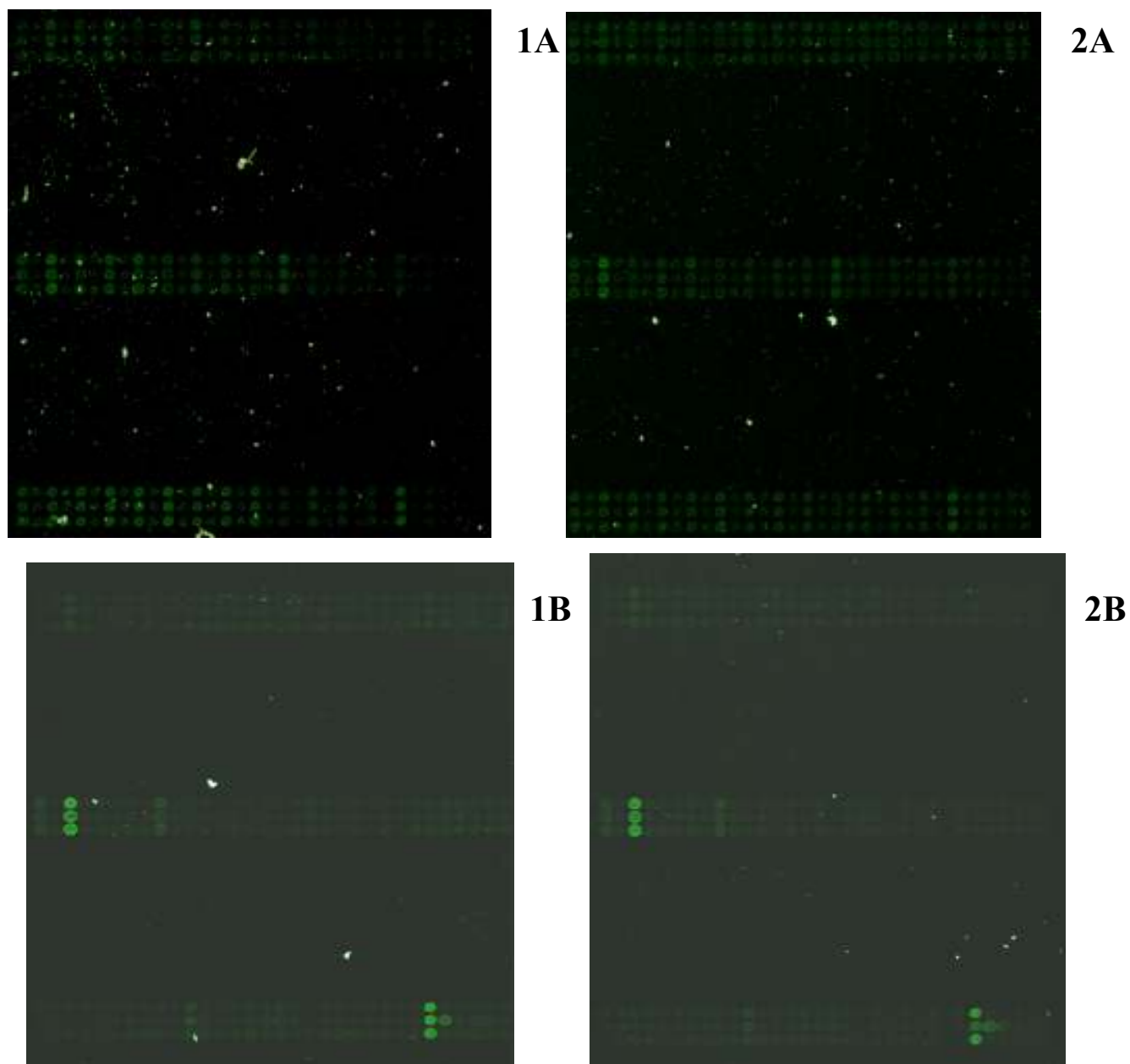


Figure 26: Hybridization results of the T30 strain at 52 °C (Images 1A and 2A) and at 42 °C (Images 1B and 2B) on a two-dimensional array. Cy3-labelled DNA amplicons were hybridized to bound oligonucleotides on glass slides. Fluorescent patterns were recorded with a Genepix 400B microarray scanner (Molecular Devices, USA) at a wavelength of 532 nm. Images 1A-B represents the left side and images 2A-B represent the right side of the slide.

Table 18:
Final results of expected results versus obtained results for oligos

Oligo Name	Expected Result	Result Obtained		Notes
		T36	T30	
02-T36	T36-Specific	√	×	42 °C best temperature
06-T36	T36-Specific	√	×	52 °C has less cross-hybridization
40-T36	T36-Specific	√	√	cross-hybridization to strain T30
57-T36	T36-Specific	√	×	42 °C best temperature, 52°C too much cross-hyb
59-T36	T36-Specific	√	×	42 °C best temperature, increase SNR threshold to 60
64-T36	T36-Specific	√	√	Cross-hybridization at 42 °C to strain T30, weak SNR at 52°C
82-T36	T36-Specific	√	×	42 °C best temperature
123-T36	T36-Specific	√	×	42 °C best temperature, increase SNR threshold to 20
228-T36	T36-Specific	√	×	52 °C best temperature, increase SNR threshold to 5
230-T36	T36-Specific	√	×	52 °C best temperature, decrease SNR threshold to 2
264-T36	T36-Specific	√	×	42 °C best temperature, increase SNR threshold to 5
268-T36	T36-Specific	√	×	52 °C best temperature, 42°C too much cross-hyb
40LNA-T36, 264LNA-T36	T36-Specific	√	×	42 °C best temperature
230LNA-T36	T36-Specific	√	×	52 °C best temperature, 42°C too much cross-hyb
Con1	Conserved	√	√	42 °C best temperature
Con2	Conserved	×	√	Poor hybridization to T36 strain at both temperatures
WNV1	Negative	×	×	52 °C best temperature, 42°C too much cross-hyb
WNV-prM	Negative	×	×	52 °C best temperature, 42°C too much cross-hyb
RabMF	Negative	×	√	Cross hybridization for T36 strain at both temperatures
T3 2+,3+,6+	Negative	√	√	Lots Cross hybridization at both temperatures
T3 5+	Negative	×	×	52 °C best temperature
VT 3+,4+,5+, 6+	Negative	×	√	52 °C best temperature, cross hyb with strain T30
VT 2+	Negative	√	×	52 °C best temperature, cross hyb with strain T36
VT 1+	Negative	×	×	42 °C best temperature
T36 1+, 5+	T36-Specific	√	×	52 °C best temperature
T36 2+, 3+, 6+	T36-Specific	×	×	Weak hybridization to T36 strain at both temperatures
T30 1+, 3+	T30 Specific	×	×	Poor hybridization to T30 strain at both temperatures, Cross-hybridization to strain T36 at 42 °C
T30 2+, 6+	T30 Specific	×	√	52 °C best temperature
T30 7+	T30 Specific	×	√	42 °C best temperature, increase SNR threshold to 4

√ Represents a positive hybridization result with the respective strain

× Represents a negative hybridization result with the respective strain

4.3.4 STRAIN DIFFERENTIATION

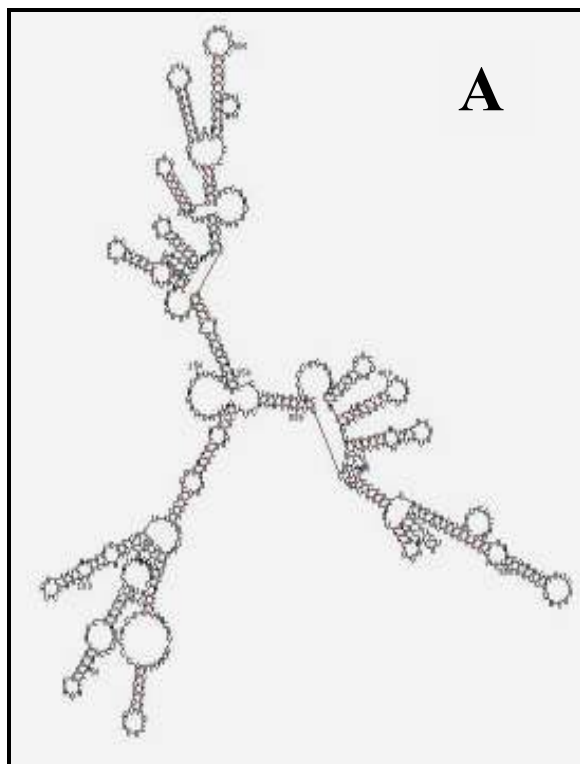
In this study, two CTV strains T30 and T36 were differentiated by oligonucleotides developed to target the T36 strain and have mismatches to the T30 strain preventing hybridization. The genotypes of these clones were determined and verified in the previous chapter and are confirmed as being their respective strains T30 and T36. Most of the negative control oligonucleotides displayed no or very low signals when hybridized against the two strains.

Of the oligos tested which were specific to the T36 strain (Table 18) the following functioned successfully in differentiating the strains: 2-T36, 40LNA-T36, 57-T36, 59-T36, 82-T36, 123-T36, 264-T36 and 264LNA-T36 at 42 °C and 6-T36, 228-T36, 268-T36, T36-1+ and 230LNA-T36 at 52 °C. Of the oligos tested which were specific to the T30 strain (Table 18) the following functioned successfully: T30-2+, T30-6+ and T30-7+ at 52 °C. Of the negative non-CTV oligos tested the following functioned successfully: WNVI and WNV-prM at 52 °C. Of the oligos tested as negative controls for the T30 and T36 CTV strains, the following functioned successfully: VT-1+ at 42 °C and T3-5+ at 52 °C. The only conserved oligonucleotide to successfully hybridize to both strains was Con1 at 42 °C.

4.3.5 SECONDARY STRUCTURES

The mfold web server is available for the prediction of the secondary structures of single stranded nucleic acids. The core algorithm predicts a minimum free energy, ΔG , as well as free energies for folding that must contain any particular base pair (Zuker., 2003). It is important to determine the secondary structures of the DNA amplicons used as the target sequence in the array that could play a role in preventing binding to their complementary oligonucleotides. The predicted secondary structures of amplicons of T30 (22-592 bp) and T36 (7370-7967 bp) are depicted in Figures 27 A-B for the T30 and T36 strains respectively. The effects of secondary structures of DNA on hybridization with oligonucleotides are not well understood. The images below represent the conditions at which the highest degree of complexity of secondary structure is found. The predicted secondary structures depicted for the T30 and T36 amplicons showed a lot of complex stem and loop structures and a lot of these stem and

loop structures were folded back on each other adding to the complexity of the structure. There were also a lot of areas on the structure where there could be steric constraints for the hybridization to an oligonucleotide because of complex folding.



B

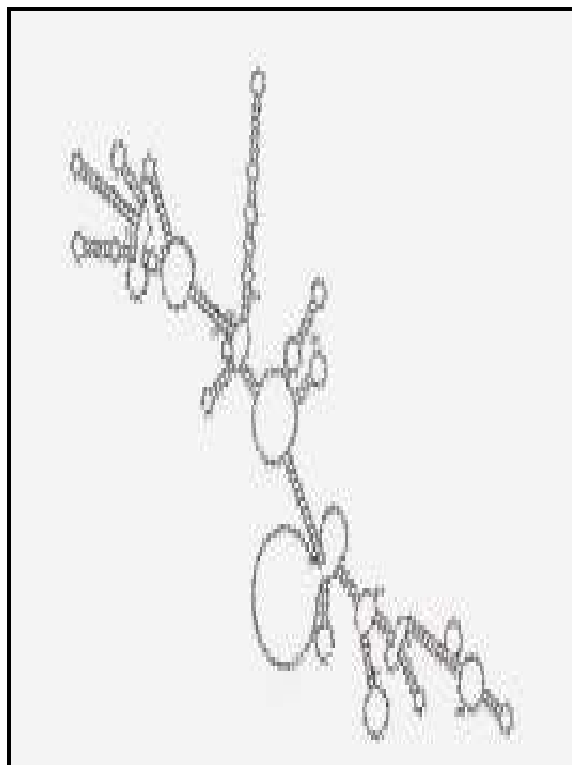


Figure 27: Representative secondary structures of CTV strains DNA PCR products constructed by the mfold programme (Zuker., 2003), (A) T30 strain, 22-592 bp, amplicon region 1 (dG= -215.64), (B) T36 strain, 7370-7967 bp, amplicon region 5 (dG= -162.73).

4.4 DISCUSSION

The oligos design was standardized to have a T_m between 55-65 ° C and the GC content preferably above 60 %. It was not always possible to design an oligonucleotide to have a 60 % GC content in the selected CTV region of the genome and therefore some oligos are below a 60 % GC content as seen with oligo Con 2. There were limitations designing oligos using the Array Designer programme. The CTV genome had to be separated into 70 mer sequences and this prevented oligo design of the overlap area of two 70 mer sequences located next to each other on the genome. There was also no way of aligning two sequences and designing an oligo based on differences between them or pre-selecting the position on the oligo of the intended mismatch.

Different steps of the microarray experiment were tested to develop a standard operating procedure (SOP) including labelling with different methods to ensure a high yield of target DNA and Cy3 dye incorporation. Different labelling techniques were tested for their efficiency and initially in the development of this array this was a major challenge. Post-PCR labelling with Klenow Fragment exonuclease using sequence specific primers or random hexamers produced fragments with a low Cy3 dye reading on the Nanodrop spectrophotometer. This could have been attributed to a high amount of secondary structures on the DNA fragments leading to steric constraints preventing Cy3 dye incorporation. These secondary structures could have caused the Klenow enzyme to stall and break off the template strand therefore decreasing the yield and Cy3 dye incorporation of full-length fragments. Attempts to label and amplify the target strains with F3-dye and Cy3-dUTP were unsuccessful. However labelling the template with a double-stranded target DNA using a Cy3-5' end labelled reverse primer was successful in producing a high DNA yield and a good signal on the array. This method also proved to be easier to control and replicate since it was an existing optimised PCR protocol. Gel electrophoresis confirmed the correct sized product was amplified and Spectrophotometer readings confirmed the concentration and Cy3 dye incorporation was sufficient. A previous study comparing different labelling methods for the production of labelled target DNA for microarray hybridization showed that the highest signals were obtained using a post-PCR labelling method like with Klenow (Franke-

Whittle *et al.*, 2005). This method however had more non-specific hybridizations (Franke-Whittle *et al.*, 2005). Hybridization with a double-stranded PCR product labelled with a Cy3-labelled primer resulted in acceptable signal to noise ratios (SNR) except for probes towards the 3' end of the gene (Franke-Whittle *et al.*, 2005). Their study concluded that labelling via the Cy3-primer approach was the most appropriate of the methods for their purposes (Franke-Whittle *et al.*, 2005). And in this study this was also the case with labelling via a Cy3-primer, even though labelling with Klenow did not show similar labelling efficiency results to Franke-Whittle *et al.* (2005), but as discussed this could have been influenced by factors such as secondary structures etc and steric constraints.

Two different hybridization mixes and two different washing step strategies were tested to select the most effective methods. The most optimal methods were with the Deyong *et al.* (2005) strategy of Denhardt's derived buffer and low stringency washing at 42 °C (6 minutes in 2X SSC and 0.2 % SDS; 2 minutes in 0.2X SSC and 0.2 % SDS, 2 minutes in 0.075X SSC). The less stringent washing step was selected. However, in the future, as the chip becomes more selective and specific a more stringent washing step might be needed to prevent high levels of background noise.

The testing of the microarray system consisted of hybridizing the slides at different temperatures to strains T30 and T36. As expected 60 °C was too stringent and no signal was obtained (data not shown). This was expected since most oligonucleotides designed had a melting temperature (T_m) of between 53-58 °C. However at 42 °C and 52 °C there were varying results for the T36 specific oligos. For some oligos 52 °C was too stringent and a lowered hybridization was seen and therefore 42 °C was more optimal and specific. However with the increased stringency at 52 °C some oligos were more specific compared to 42 °C.

Data was carefully analysed by filtering spots with abnormalities. Measuring and quantifying microarray variability was essential to prevent unwanted systemic variability. Estimates of these variabilities were essential to gaining an understanding of how well the microarray chip was performing and also important for determining the number of replicates required for a microarray experiment.

Variability exists in various different forms: (1) between individual strains; (2) between separately labelled samples; (3) between sample preparations; (4) between microarray slides and hybridisations; (5) between different areas on a slide; and (6) between replicate features. Most of the aforementioned sources of variability are introduced by the experimental process. In contrast, the variation between individual strains was independent of the microarray process.

To test the level of variability induced by the experimental process the %CV of each oligo's spots were determined. Each individual slide was assessed for oligos which had unacceptably high variability between interspot replicates of each oligo. To accurately determine variability it is important to statistically represent variability of replicate features by determining the %CV. Comparisons were made of all the CV's between replicates on the slide as well as between the left and right sides of the slide. This was possible since the entire microarray layout was duplicated on both sides of the slide. Such variance between replicates could lead to false positives or false negatives and it is therefore important to filter the data before interpreting the true results. Oligonucleotides which showed a high degree of variance between different locations were noted for further interpretation in the final conclusion about each oligo. Both the T30 and T36 slides at 52 °C had unacceptable levels of background noise and the final data for these slides reflects the subtraction of this high level of non-specific signal. For the four slides there was no clear difference in variance between the left and right sides for all oligos of the slides. There were, however, some oligos which had unexplained but distinct differences between left and right sides.

By using the signal to noise ratios (SNR) as the data values it was possible to determine the confidence with which one can quantify a signal peak of a given value, especially a signal near background. The SNR is defined as the median for a feature subtracting the median background signal and dividing by the variation in background (standard deviation of background). The confidence in quantifying the peak increases as the variation in background signal (the "noise") decreases regardless of the absolute value of the average background. An average SNR of above 3 for an oligo after filtering was regarded as the threshold for a positive hybridization based on a previous study where cucumber mosaic virus (CMV) isolates were differentiated by hybridization to oligonucleotides in a similar microarray format (Deyong *et al.*, 2005).

4.4.1 SPECIFICITY OF OLIGONUCLEOTIDES

The oligonucleotides (oligos) were designed based on the alignment of the T36 and T30 strain sequences of CTV from GenBank. The oligos targeted regions on the 5' end towards the central region of the CTV genome. This area displays the highest amount of variability (Ayllón *et al.*, 2001) and is optimal for strain differentiation of strain T30, with mild symptoms and T36, with severe decline symptoms. Since these two genotypes/strains are more common (Hilf *et al.*, 2005) than other genotypes but induce different symptom severity it would be a very useful test to determine which strain type/s are present in RSA sources. This array was developed to have an on and off signal for the two strains tested. The performance of the oligos will be discussed by categorizing them as good, weak and bad oligos.

1. T36 specific oligos

Most of the oligos were designed to be T36 specific and therefore have increased intensities against the T36 target. It was shown that the hybridization of the T36 target to the T36 specific oligos was more significant than the hybridization of T30 target to the same oligos for both temperatures. At 42 °C the t-test probability was 0.26, indicating that the T30 and T36 targets hybridization had a probability of being 26% similar and therefore 74% different from each other. At 52 °C the t-test probability was 0.37, indicating that the T30 and T36 targets hybridization had a probability of being 37% similar and therefore 63 % different from each other.

• Useful oligos

For oligos 2-T36, 40LNA-T36, 57-T36, 59-T36, 82-T36, 123-T36, 264-T36 and 264LNA-T36: 42 °C was the most optimal temperature for hybridization to the target T36 strain with an average SNR of above 3. There was a low amount of cross-hybridization at 42 °C for strain T30 with a value below a SNR of 3. At 52 °C the conditions were too stringent for hybridization to the T36 strain and there was low cross-hybridization at 52 °C to the T30 strain.

With oligo *59-T36* at 42 °C hybridized against the T36 strain there was an average SNR 17 times greater than for the T30 strain. This was significantly different and warrants this oligonucleotide getting a different threshold SNR average of above 40 to reflect a positive hybridization. With oligo *123-T36* at 42 °C hybridized against the T36 strain there was an average SNR 10 times greater than for the T30 strain. This was significantly different and warrants this oligonucleotide getting a different threshold SNR average of above 40 to reflect a positive hybridization.

For oligos *6-T36*, *228-T36*, *230LNA-T36*, *268-T36* and *T36-1+*: 52 °C was the most specific and optimal temperature for hybridization to the target T36 strain with an average SNR of above 3 and low amounts of cross-hybridization to strain T30 due to increased stringency. These oligos are definitely suitable for use in future work on strain differentiation.

These oligos ranged from 18-35% mismatches to the T30 strain. The majority of the oligos had scattered mismatches ranging from 18-33%. Oligos *57-T36* and *123-T36* had a high amount of mismatches around 35% at the 3' end. Generally the positioning of the mismatches was irrelevant if the amount of mismatches was above 25%. With low amounts of mismatches it was favourable to position the mismatches as scattered or in the centre with the only exception being oligo *230LNA-T36* which had 14% mismatches on the 3' end. However this oligo had increased specificity with the LNA modified bases. There were only two oligos that functioned successfully in differentiating the strains with mismatches below 20%.

- **Weak oligos**

For oligos *230-T36*, *T36-3+* and *T36-5+*, the average SNR at 52 °C for strain T36 was just below a SNR of 3, the threshold for a positive hybridization with very little cross-hybridization to the T30 strain. At 42 °C there was a lot of cross-hybridization to the T30 strain. These oligos do not seem to be suitable for use in future work on strain differentiation unless the threshold is dropped to between 2.3-2.5. This however would not be ideal for such an array system. Oligo *230-T36* had a low amount of mismatches around 14% in a sub-optimal position on the 3' end and oligos *T36-3+* and *T36-5+* were not designed according to the parameters specified and therefore had low GC content and were too AT rich.

- **Unacceptable oligos**

Oligos *40-T36*, *T36-2+* and *T36-6+* had low SNR averages below 3 for the T36 strain, with no significant difference between the T36 and T30 strains at both temperatures. For oligo *64-T36* there was a considerable amount of cross-hybridization to strain T30 at 42 °C. At 52 °C the average SNR were well below 3 for both strains. Oligos *40-T36* and *64-T36* had a low amount of mismatches below 20% and this could have been the cause of their poor performance. Oligos *T36-2+* and *T36-6+* were not designed according to the parameters specified and therefore had low GC content and were too AT rich. It is not known what makes these oligos unsuccessful, but they should not be included in future development of an array.

The possible reason for the high amount of cross-hybridization at 42 °C could be because there was a smaller amount of mismatches (below 20%) to the T30 strain at less stringent conditions. It is possible that conditions were too stringent at both these temperatures, but it is unlikely as all oligonucleotides were designed to have similar properties. Possibly secondary structures on the T36 strain DNA fragment prevented the oligonucleotide from accessing the target area.

It was expected that more non-specific hybridization would occur at 42 °C since this was a less stringent condition than at 52 °C. This however was not the case for a few oligos. It could be that at 42 °C the temperature was too low for a stable hybridization complex to form and perhaps 52 °C was a more optimal temperature. It could also be possible that there were more complex secondary structures at 42 °C than at 52 °C, preventing the target from accessing the oligo sequence and hybridizing to it.

2. T30 specific oligos

These oligos were expected to function as T30 strain-specific oligos with a SNR above 3 for hybridization to the T30 strain.

- **Useful oligos**

Oligos *T30-2+* and *T30-6+* had SNR above 3 for the T30 strain at 52 °C with insignificant cross-hybridization signals to the T36 strain. At 42 °C the conditions were not optimal. Oligo *T30-7+* had a SNR of 3 for strain T30 at 42 °C with a small degree of

cross-hybridization to the T36 strain. At 52 °C there was also a good SNR for strain T30 but closer to 3. It seems this oligo could be used at both temperatures. Therefore these oligos can be used as T30 strain-specific oligos in future array development.

- **Unacceptable oligos**

Oligos T30-1+ and T30-3+ had poor SNR well below 3 for hybridisation to the T30 strain at both temperatures and high amounts of cross-hybridization to the T36 strain at 42 °C. It is unclear why these oligos are not successful since they have 100 % homology to the T30 strain and have a high percentage of mismatches (50-100%) to the T36 strain. These oligos should not be used in future development of a CTV array.

3. Conserved control oligos

These oligos were expected to function as positive controls for both strains as their sequences are conserved.

Oligo *Con1* had average SNR around 3 for both strains at 42 °C. However at 52 °C there was a decrease in the average SNR as the conditions became more stringent and poor hybridization took place. It appears that 42 °C was the more specific and optimal temperature to ensure the best hybridization signal for both strains around a SNR of 3. The level of signal was unexpectedly poor as both strains have a 100 % homology to the oligo and should be producing a very strong signal. This oligo can be used as a positive control in this array for the two strains at 42 °C, however it must be noted that the threshold SNR of 3 will probably have to be lowered to 2.5 which was not an ideal situation for a positive control.

Oligo *Con2* had average SNR values above 3 for strain T30 at the different temperatures. However the average SNR for the T36 strain were well below 3 at both temperatures. At a hybridization of 52 °C for strains T30 and T36 the conditions were too stringent. At 42 °C strain T30 was well above a SNR of 3, whereas T36 was well below a SNR of 3. This was unexpected as both strains have a 100 % homology to the oligo and should be producing a very strong signal. A possibility for this could be that the design of the oligo having an AT rich sequence was not optimal for hybridization. This oligo can not be used as a positive control in this array. More oligos must be developed which are reliable and good oligos to function correctly as positive and

conserved markers for these two strains. A possibility for both oligos not functioning as expected could be that there were complex secondary structures preventing the oligo from accessing the target sequence.

4. Negative control oligos

These oligos were expected to not hybridize to any CTV sequences or strains and serves as a negative control in the array and therefore was expected to display a low signal.

- **Good oligos**

Oligos *WNVI*, *WNV-prM* and *T3-5+* had very low SNR signals being below 3 at 52 °C for both strains and the specificity of this was increased. At 42 °C there was cross-hybridisation to both strains. Oligo *VT-1+* had below SNR of 3 for both strains at 42 °C with insignificant cross-hybridization. Therefore these oligos can be used as negative controls in this array for the two strains at their respective optimal temperature.

It was unexpected to have signals above 1.5 at 42 °C for oligos which represents a 100 % mismatch to any of the CTV strains sequences represented in this array. There were no other regions on the CTV genome which represented a potential cross-hybridization site. A possibility could be that secondary structures with complex folding could have made certain non-target sequences recognisable and accessible as a target for the oligonucleotide.

- **Bad oligos**

Oligo *RabMF* had very low SNR signals for the T30 strain at both 42 °C and 52 °C. However for the T36 strain there were high amounts of cross-hybridization at both temperatures. This was unexpected since there were no regions of homology on the CTV genome with this oligo sequence. This could be due to secondary structures allowing certain non-target sequences to be recognisable and accessible as a target for the oligo.

Oligos *T3-2+*, *T3-3+* and *T3-6+* had very high amounts of cross-hybridisation with SNR well above 3 for both strains at both temperatures. Oligos *VT-2+*, *VT-3+*, *VT-4+*, *VT-5+* and *VT-6+* had varying degrees of cross-hybridization to either the T30 or T36 strains or both at 42 °C and 52 °C. Therefore these oligos can not be used as a negative CTV controls.

General Summary

There does seem to be any significant amount of cross-hybridization with the water/blank spots on all the slides. Therefore the SNR threshold can be supported as being above 3 to be regarded as a positive hybridization result. It also shows that either background noise or artefacts did not have a significant effect on the overall functioning of the slides.

As shown in Figures 19-20, the T36-specific oligos 2-T36, 6-T36, 40LNA-T36, 52-T36, 57-T36, 59-T36, 82-T36, 123-T36, 228-T36, 268-T36, 264LNA-T36, 264-T36, T36-1+ and 230LNA-T36 worked perfectly and detected the respective T36 strain without any significant cross-hybridization. The faint to high cross-hybridization signals of oligos 40-T36, 64-T36 and 230-T36 in hybridization with the T30 strain originates from the fact that they differ from the T36 strain sequences by 14-20 % as mismatches, which represents only a few base pairs. Oligos 40-T36, T36-2+, T36-3+, T36-5+ and T36-6+ did not successfully differentiate strains because the hybridization to the target T36 strain was poor. The failures could be due to experimental design and/or from complex secondary structures preventing recognition of the target.

There were 13 T36-specific oligos that were successful in differentiating the T36 and T30 strains. All these oligonucleotides fell into the medium (21-40 %) and high (above 40 %) of categories of mismatches. These oligos had their mismatches predominantly scattered, with only a few on the 3' or 5' ends and in the centre. Two successful oligos had a low (below 20 %) amount of mismatches but the mismatches were positioned at the 5' end or scattered. It was previously reported that mismatches distributed over the entire length of the oligonucleotide were most specific, while those with mismatches grouped at either the 3' or 5' end were the least specific (Letowski *et al.*, 2004). While the current data does not allow a clear answer as to which position was the most optimal, it is definitely clear that the oligos with a scattered mismatch pattern were more successful than the other positions tested irrespective of their amount of mismatches.

The T36 specific oligonucleotides which had cross-hybridization to the T30 strain all had a low amount of mismatches and position of mismatches appeared irrelevant in accounting for the cross-hybridization being so high in these oligonucleotides.

Brown *et al.* (2003) showed that hybridization success with a low amount of mismatches was dependent on position and type of mismatch and probably also on pH, and hybridization was successful with one mismatch between the oligo and the template if the mismatch was located near 3' end of the probe (Brown *et al.*, 2003). If a mismatch occurred in the centre region of the probe hybridization was strongly dependent on the mismatch type, for example as G-T mismatch is among the most stable mismatches observed in DNA (Brown *et al.*, 2003). In this study this is an important consideration to prevent hybridization to a closely related yet different strain (in this study referred to as cross-hybridization). This type of effect might explain possible cross-hybridization found in this study.

The result of this study on differential hybridization against strain differentiating oligos clearly demonstrates the great potential of oligo based microarray technology for virus strain differentiation. Previously it was shown that short oligos are suitable for discriminative hybridization of strains differing by less than 8 % in an amplified PCR product larger than 700 bp (Deyong *et al.*, 2005). An array such as the one developed here as a pilot study could be advanced to include more markers for the T30 and T36 strains; for other CTV strains as well as generic oligos for other citrus viruses and could become a faster and more powerful technique in viral diagnostics than performing multiple amplification steps and sequencing. This test has the advantage of acquiring genetic information on many different sequence regions of a particular strain in just one test which would require a large number of different PCRs equating to the same result, which would be impossible to routinely perform.

4.4.2 LOCKED NUCLEIC ACIDS (LNA)

Experimental success in molecular biology is dependent upon specific and discriminating hybridization events involving synthetic oligonucleotides and their complementary target sequences. While unmodified oligodeoxynucleotides will routinely form desired DNA:DNA and DNA:RNA duplexes, synthesis of various

modifications that confer enhanced high-affinity recognition of DNA and RNA targets have been an ongoing endeavour. The Locked Nucleic Acid (LNA) was first described by Wengel and co-workers in 1998 as a novel class of conformationally restricted oligonucleotide analogues (You *et al.*, 2006). A locked nucleic acid (LNA), often referred to as an inaccessible RNA, is a modified RNA nucleotide (You *et al.*, 2006). The ribose moiety of a LNA nucleotide is modified with an extra bridge connecting 2' and 4' carbons (You *et al.*, 2006). The bridge “locks” the ribose in 3'-endo structural conformation, which is often found in A-form of DNA or RNA (You *et al.*, 2006). LNA nucleotides can be mixed with DNA or RNA bases in the oligonucleotide. The locked base conformation enhances base stacking and backbone pre-organization (You *et al.*, 2006).

McTigue *et al.* (2004) showed that LNA pyrimidines contribute more to stability than do LNA purines with average ΔT_m values of 4.4 ± 1.5 °C for LNA-C, 3.2 ± 1.4 °C for LNA-T, 2.8 ± 1.7 °C for LNA-G and 2.1 ± 1.3 °C for LNA-A. The synthesis and incorporation of LNA bases can be achieved by using standard DNA synthesis chemistry (You *et al.*, 2006). Due to the high affinity and thermal stability of the LNA: DNA duplex it is not advisable to have more than 15 LNA bases in an oligo as this induces strong self-hybridization (You *et al.*, 2006).

Three of the DNA oligos in this study were modified to contain LNA-G bases. The LNA modified oligo referred to as 40LNA-T36, had a higher average SNR 532 at 42 °C compared to the normal DNA based oligonucleotide 40-T36, making it more specific at the same temperature. There was also increased specificity with the non-target T30 strain with less cross-hybridization with the LNA modified oligo compared to the unmodified oligo. The mismatches were on the 5' end and had a low (below 20 %) amount of mismatches to strain T30. The LNA modified G was bordered by a G on the 5' side and a mismatch A on the 3' side. The addition of LNA in this case definitely increased specificity.

The LNA modified oligo referred to as 230LNA-T36, had a higher average SNR of 532 at 42 °C and 52 °C compared to the unmodified DNA oligo 230-T36, making it more specific at the same temperatures. Cross-hybridization levels were similar with the LNA modified oligo compared to the unmodified DNA oligo. Therefore this LNA modified

oligo increased the specificity of the T36 strain especially at 52 °C. The mismatches of the oligo were on the 3' end and were classified as having a 14 % of mismatches to strain T30. The first LNA modified G was bordered by a T on the 5' side and a G on the 3' side; and the second LNA modified G was bordered by a mismatch A on the 5' side and a G on the 3' side.

The LNA modified oligonucleotide referred to as 264LNA-T36, had very similar SNR for both strains at the same temperatures. The specificity remains the same for the T30 strain with a low degree of cross-hybridization with the LNA modified oligo and the unmodified DNA oligo.

A triplet of LNA nucleotides surrounding a single-base mismatch site maximises LNA probe specificity unless the oligo contains a G-T mismatch (You *et al.*, 2006). LNA increases significantly the thermal stability of the oligos hybridized to complementary DNA/RNA, which allows excellent mismatch discrimination and therefore increases the sensitivity and specificity of microarrays (You *et al.*, 2006). The high binding affinity of LNA oligos allows for the use of short probes in antisense protocols and LNA is recommended for use in any hybridization assay that requires high specificity and/or reproducibility e.g. microarray oligos (You *et al.*, 2006). Furthermore, LNA offers the possibility to adjust T_m values of oligos in complex assays such as this (You *et al.*, 2006). It is clear that this modification of an oligo with LNA can be very useful especially when the unmodified oligo is a differentiating marker but does not have good discriminatory power.

4.4.3 SECONDARY STRUCTURES

The sensitivity of an oligonucleotide can be determined by two factors: (1) The oligonucleotide should not have internal secondary structures or bind to other identical oligonucleotides on the array; and (2) The oligonucleotide is able to access its complementary target, which could potentially be unavailable as a result of secondary structure. The latter being much more difficult to predict and quantify (Stekel., 2003). It is important to note that the success of an oligonucleotide is not determined primarily by base composition or sequence but also by secondary structures (Southern *et al.*, 2004). The effects of thermodynamic and kinetic factors on folding and stabilities of the folded structures are also not well understood.

The predicted secondary structures of the DNA amplicons used in this study as the target DNA were depicted by mfold software (Zuker *et al.*, 2003). Many amplicons showed very complex secondary stem and loop structures particularly with regions 1, 2 and 5. The best example of this was illustrated in the group 5 amplicon for strain T36 (as represented in Figure 27). It is important though when analysing secondary structure effects to locate the oligo sequence on the structure and establish the possible effect on hybridization based on the complexity of the sequence area of the target. In addition to the possible effects caused by secondary structures there are also negative effects from prominent steric constraints due to complex folding in the surrounding sequence area, which prevents the oligo from accessing the target sequence even if it is not found on a particularly complex site. These sites may be masked by the three-dimensional folding of the molecules and also by additional tertiary interactions.

4.5 ACKNOWLEDGEMENTS

I'd like to acknowledge Prof G. Pietersen for his novel idea for this study. My sincere thanks to Danie Theron, Sanushka Naidoo, Luke Solomon, Nicholas Olivier, Prof. D. Berger and Prof L.H. Nel for their advice and technical help throughout the stages of the development of this chip.

4.6 REFERENCES

- Boonham, N., Walsh, K., Preston, S., North, J., Smith, P., Barker, I. 2002. The detection of tuber necrotic isolates of Potato virus Y, and the accurate discrimination of PVYO, PVYn and PVYc strains using RT-PCR. *J. of Virol. Meth.* 102: 103-112.
- Brown, J., Brown, T., Fow, K.R. 2003. Cleavage of fragments containing DNA mismatches by enzymatic and chemical probes. *Biochem.* 371: 697-708.
- Bystricka, D., Lenz, O., Mraz, I., Piherova, L., Kmoch, S., Sip, M. 2005. Oligonucleotide-based microarray: A new improvement in microarray detection of plant viruses. *J. of Virol. Meth.* 128: 176-182.

- Deyong, Z., Willingmann, P., Heinze, C., Adam, G., Monica, P., Frey, B., Frey, J.E. 2005. Differentiation of Cucumber mosaic virus isolates by hybridization to oligonucleotides in a microarray format. *J. of Virol. Meth.* 123: 101-108.
- Franke-Whittle, I. H., Klammer, S.H., Mayrhofer, S., Insam, H. 2005. Comparison of different labelling methods for the production of labelled target DNA for microarray hybridisation. *J. of Micro. Meth.* 65: 117-126.
- Lee, G.P., Min, B.E., Kim, C.S., Choi, S.H., Harn, C.H., Kim, S.U., Ryu, K.H. 2003. Plant virus cDNA chip hybridization for the detection and differentiation of four cucurbit-infecting Tobamoviruses. *J. of Virol. Meth.* 110: 19-24.
- Letowski, J., Brousseau, R., Masson, L. 2004. Designing better probes: effect of probe size, mismatch position and number on hybridization in DNA oligo microarrays. *J. of Virol. Meth.* 57: 269-278.
- Letschert, B., Adam, G., Lesemann, D. E., Willingmann, P., Heinze, C. 2002. Detection and differentiation of serologically cross-reacting tobamoviruses of economical importance by RT-PCR and RT-PCR-RFLP. *J of Virol. Meth.* 106: 1-10.
- McTigue, P.M., Peterson, R.J., Kahn, J.D. 2004. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochem.* 43: 5388-5405.
- Song, Y., Dai, E., Wang, J., Liu, H., Zhai, J., Chen, C., Du, Z., Guo, Z., Yang, R. 2006. Genotyping of hepatitis B virus (HBV) by oligonucleotide microarray. *J. of Mol. and Cell. Probes* 20: 121-127.
- Sohail, M., Akhtar, A., Southern, E.M. 1999. The folding of large RNAs studied by hybridization to arrays of complementary oligonucleotides. *J. of RNA* 5: 646-655.
- Southern, E.M., Case-Green, S.C., Elder, J.K., Johnson, M., Mir, K.U., Wang, L., Williams. J.C. 1994. Arrays of complementary oligonucleotides for analyzing the hybridization behaviour of nucleic acids. *Nucleic Acids Res.* 22: 1368-1373.

Stekel., D. 2003. Microarray Bioinformatics. Cambridge University Press, Cambridge, UK.

Thompson, J. D., Higgins, D.J., Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673-4680.

Xiong, Z., Barthelson, R., Weng, Z., Galbraith, D.W. 2005. Analysis of the *Citrus tristeza virus* (CTV) genome using resequencing microarrays. *Phytopathology* 95: 114.

You, Y., Moreira, B.G., Behlke, M.A., Owczarzy, R. 2006. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* 34: 60.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13): 3406-3415.

CONCLUDING REMARKS

The continuous selection of superior protective mild CTV isolates is the most critical and important step to maintain a mild isolate cross protection programme for the Southern African grapefruit industry. Biological characterisation is time consuming, inaccurate and too costly for strain differentiation on its own. Molecular characterisation of isolates is rapid and useful for gaining insight into severity molecular markers. It is essential that molecular based tests are set up to differentiate strains and to ensure that mild isolates are tested thoroughly before being approved to be used in the industry. These tests together with current biological indexing will

ensure more reliable mild strains are selected. These molecular tests can also be used to prevent severe strains from being introduced into the country as well as to assess the status and dynamics of the strains in citrus producing areas of South Africa.

The aim of this study was to establish PCR based methods to differentiate CTV strains and to use these methods to characterize isolates either already in use in the cross-protection scheme or ones that have been identified as potential candidates for cross-protection. Another part of this study was to also develop an oligonucleotide microarray system to differentiate the T30 and T36 strains.

Strain differentiation based on the differences in the p23 gene developed by Sambade *et al.* (2003) was established to differentiate strains into three groups: mild represented by strain T385; severe represented by strain T305; and atypical represented by strain T36. The test consisted of two bi-directional RT-PCRs per isolate of which each PCR consisted of five primers. The first PCR differentiated between mild and severe strains and the other between severe and atypical strains. The area of target was within a conserved gene with very little variability between strains. However previous work showed that this gene was responsible in part for symptom expression. The domain of interest was within the RNA-binding region and showed clear differences in three amino acids (78-80) that differed according to the strain group. Each group was representative of a certain amino acid composition for that area. The results concluded that certain isolates in the cross-protection scheme did not have only mild sequence types but also surprisingly severe and atypical sequence types. This is of great concern since the GFMS 12 cross-protecting source consisting of sub-isolates 12-5, 12-7 and 12-9 are currently being used in cross-protection. This could indicate why cross-protection breakdown has been observed. RSA isolates 390-3 and 390-5 were atypical; 390-4, 389-4 and 389-3 were mild; GFMS 35 had mild and atypical isolates; GFMS12, 12-7 and 12-9 had mild and severe isolates and; 12-5 was severe. In this study it was also essential to fully sequence the p23 gene of these isolates and to not only compare them to other sequences from around the world but also to validate the PCR results. Interestingly phylogenetic grouping of the isolates based on the full p23 gene represented the same grouping as the PCR results showed. This PCR system was dynamic, reproducible and easy to establish. This PCR had the power to amplify sequences of more than one group in one PCR; this will prove to be valuable in detecting mixed strains in isolates as well as

their sequence types. This method also showed that strain differentiation was possible with differences in only a few nucleotides. The disadvantages of this test include: (a) such small differences in a conserved gene do not reflect other more complex variability of other genome regions; (b) there is uncertainty over the atypical group and where it fits since isolates in this group vary from symptomless to severe quick decline; (c) once an isolate has been classified into a group, it is still not possible to know what type of symptom severity it causes without sequencing it and comparing it to other known sequences; and (d) classifying the isolates into three wide groups also does not indicate the more specific strain/genotype, for example many isolates in the severe group displayed seedling yellows and stem-pitting symptoms and belonged to very different sequence types. This test however should be considered in the South African Citrus industry as a first step strain differentiating screening tool in molecularly selecting mild strains for cross-protection.

Many different CTV differentiation methods have been developed in the last few years. There is however no method as comprehensive as the RT-PCR system developed by Hilf *et al.* (2000) which targets four different genotypes (T30, T36, T3 and VT). The four genotypes selected for differentiation all display different symptoms and have distinct sequence types. The VT isolate originates from Israel whereas the T30, T36 and T3 were isolated from Florida, USA but it is thought that T3 originates from South America or Asia. These genotypes were differentiated by 23 different primer sets based on the first 10 kb of the 21 kb CTV genome, which represents the highest variability of all the regions. In this study the same isolates were tested as in the previous PCR based system to compare the results and establish which method was more informative. Each isolate was tested with RT-PCR using the 23 individually optimised genotype specific primer sets. The most common genotype detected was T30 and the least common was T3, which is consistent with other studies around the world.

The advantages of this method are that there are many different primers acting as genotype molecular markers across different regions of variability even though the test only represents approximately 1-1.5 % of the entire genome. Some disadvantages of this method are the uncertainties surrounding possible non-specific amplifications of genotypes as seen with the T3 and T30 genotypes as well as regions that fail to amplify any genotype specific amplicons. These PCRs were very difficult to optimise and are time-consuming to perform on a lot of samples. Both PCR based methods also run the

risk of selecting certain dominant strains that are better and more abundant templates for amplification, neglecting other strains that could be present in the sample. This method however is very valuable for in depth analysis of a genetically homogenous strain that is in the final stages of testing for cross-protective abilities before being approved for use in the South African mild strain cross-protection scheme, as well as shedding light on any isolate of interest and possibly the dynamics and geographic significance of the strains. Also the tests may be very useful in determining possible areas where recombination has occurred. This test however requires a lot of experience to ensure results are reproducible and to prevent false negatives occurring due to sub-optimal PCR and extraction conditions.

In comparing of the two PCR based tests it was found that the p23 gene showed that isolate 390-3 was atypical but the 5' end PCR tests only showed a predominantly mild T30 genotype. Isolate GFMS 35 showed mixed mild and atypical strains in the p23 gene PCR test but in the 5' end PCR tests test a homogenous T30, mild genotype. Isolate 12-7 showed mild and severe strains in the p23 gene PCR test but only a severe VT genotype in the 5' end PCR tests. Isolate 12-9 showed mild and severe strains in the p23 gene PCR test but only mixed genotypes VT, T30 and T36 with a predominant VT genotype in the 5' end PCR tests. These differences could reflect the different degrees of variability that were targeted in the two tests (highly conserved versus highly variable) as well as the isolates possibly containing mixed strains or recombinants. The test of choice would definitely be related to the depth of the study as to the composition of the strain/s in an isolate and the time needed to get an answer. All the other isolates showed similar results in both PCR based tests.

The question still remains, do we really know if certain symptoms are directly related to a specific strain or isolate? Or are the dynamics of strain dominance, strain interactions, strain multiplication efficiencies, environmental influence and inherent host/cultivar properties etc. more of an indication as to the real effect of a particular strain on symptom expression?

The recent advances in microarray technology have opened doors for studies on differentiation of virus strains with oligonucleotides. Microarrays have the ability to detect many hundreds of molecular markers distinctive of particular strains in one test,

which decreases the time and variability in testing compared to PCR based tests. In this study an oligonucleotide microarray was developed to differentiate DNA clones of strains T30 and T36, as a pilot study to determine the feasibility of future expansion and use in the South African citrus industry. The development was difficult due to labelling problems; oligonucleotide (oligo) design; optimisation of conditions and parameters surrounding scoring a positive result. However once these problems were sorted out the method became very simple to execute and in return mass amounts of data were retrieved. The array was designed to produce an on and off signal where oligos specifically hybridize to the T36 and do not hybridise to the T30 strain. Controls were included as negatives for all CTV strains, negatives for only T30 and T36 CTV strains, positives for both strains and conserved for both strains to eliminate possible problems.

Eleven T36-specific oligos were selected for future use in differentiating T30 and T36 strains. It was found that oligos with mismatches relative to the heterologous strain distributed over the length of the probe were most reliable followed unexpectedly by 5' and 3' positioned mismatches, while no successful oligos have centrally positioned mismatches. All oligos that successfully differentiated the two strains had mostly above 21 % mismatches with the heterologous strain. A few oligos showed cross-hybridization to the non-target strain and were removed from further analysis. These oligos had below 18 % mismatches, irrespective of their position. The conserved oligos did not perform as expected and in future work other more reliable ones would have to be designed. The advantages of this method include it being very quick to perform; having fewer variables than PCR. And producing a more informative result as to the molecular markers and therefore the sequence composition of strain/s, found within an isolate in one single test. This single array covered 1.88 % of the total T36 genome and cumulatively differentiated about 2.6 % of the T36 or T30 genomes. The equivalent test in PCR would require multiple tubes and different optimisation, therefore increasing the amount of variables and possible contamination sources. The main disadvantage of the developed array would be that variability depends on the stringency of the hybridization conditions. Thus the test might be highly subjective if carried out by different laboratories. The future prospects of this developed array could be expanded to include more molecular markers, designed to be optimal based on these results and to increase the amount of different strains targeted. This test has huge a potential in the

citrus industry to trace strains; identify certain strain markers relating to certain cultivars; to understand strain dynamics and dominance and also to identify strains relating to cross-protection breakdown.

APPENDIX 1

ISOLATE 12-5

ATGGACGATACTAGCGGACAAACTTTTATTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCTAGTACTGAAATCAAAACCGTAAGTTCGGAAGCGGATCGCTTGGAATTTTACG

GAAAATGAATCCCTTCATTATCGACGCTTTGATACGGAAAAATAGTTATCAAGGCGCT
CGCTTTCGCGCGAGAATAATAGGAGTGTGCGTGGATTGTGGTAGAAAAACACGATAAG
GCATCGAGAACTGAACGTAAGTGTAAAGGTCAACAATACGCAGTCTCAGGCCGAGGTG
GCGCATATGTTAATGCACGATCCCGTAAAATATTTAAATAAAAAGAAAGCTAGAGCCT
TTTCTAACGCAGAGATGTTTGCGATCGATTTGGTTATGCACACCAAAGAAAGGCAATT
AGCGGTTGATTTGGCCGCTGAAAGGGAGAAGACGAGATTGGCTCGTAGACACCCGAT
GCGTTCTCCGGAAGAGACTCCGGAACATTATAAATTCGGTATCACTGCCAAGGCAATG
TTACCGGACATCAACGCTATAGACGTTGGCGATAACGAAGACACCTCGTCGGAGTACC
CAGTAAGTCTGAGTGTTCGCGCGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 28: The p23 gene DNA sequence of South African Isolate 12-5.

ISOLATE 12-7

ATGGACGATACTAGCGGACAACTTTTATTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCTAGTACTGAAATCAAAACCGTAAGTTCGGAAGCGGATCGCTTGGAATTTTTACG
GAAAATGAATCCCTTCATTATCGACGCTTTGATACGGAAAAATAGTTATCAAGGCGCT
CGCTTTCGCGCGAGAATAATAGGAGTGTGCGTGGATTGTGGTAGAAAAACACGATAAG
GCATCGAGAACTGAACGTAAGTGTAAAGGTCAACAATACGCAGTCTCAGGCCGAGGTG
GCGCATATGTTAATGCACGATCCCGTAAAATATTTAAATAAAAAGAAAGCTAGAGCCT
TTTCTAACGCAGAGATGTTTGCGATCGATTTGGTTATGCACACCAAAGAAAGGCAATT
AGCGGTTGATTTGGCCGCTGAAAGGGAGAAGACGAGATTGGCTCGTAGACACCCGAT
GCGTTCTCCGGAAGAGACTCCGGAACATTATAAATTCGGTATCACTGCCAAGGCAATG
TTACCGGACATCAACGCTATAGACGTTGGCGATAACGAAGACACCTCGTCGGAGTACC
CAGTAAGTCTGAGTGTTCGCGCGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 29: The p23 gene DNA sequence of South African Isolate 12-7.

ISOLATE 12-9

ATGGACGATACTAGCGGACAACTTTTATTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCTAGCACTGAAATCAAAACCGTAAGTTCGGAAGCGGATCGCTTGGAATTTTTACG
GAAAATGAATCCCTTCATTATCGACGCTTTGATACGGAAAAATAGTTATCAAGGCGCT
CGCTTTCGCGCGAGAATAATAGGAGTGTGCGTGGATTGTGGTAGAAAAACACGATAAG
GCATCGAGGACTGAACGTAATGTAAGGTCAACAATACACAATCTCAGGCCGAGGTG
GCGCATATGTTAATGCACGATCCCGTAAAATATTTAAATAAAAAGAAAGCTAGAGCTT
TTTCTAACGCAGAGATGTTTGCGATCGATTTGGTTATGCATACCAAAGAAAGGCAATT
AGCGGTTGATTTGGCCGCTGAAAGGGAGAAGACGAGATTGGCTCGTAGACACCCGAT
GCGTTCTCCGAGGAACTCCGGAACATTATAAGTTCGGTATAACTGCTAAGGCAATG
TTACCGGACATCAACGCTATAGACGTTGGTGATAACGAAGACACTTCGTCGGAATACC
CAGTGAGTCTGAGTGTTCAGGCGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 30: The p23 gene DNA sequence of South African Isolate 12-9.

ISOLATE GFMS 12

ATGGACGATACTAGCGGACAACTTTTATTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCTAGTACTGAAATCAAAACCGTAAGTTCGGAAGCGGATCGCTTGGAATTTTTACG
GAAAATGAATCCCTTCATTATCGACGCTTTGATACGGAAAAATAGTTATCAAGGCGCT
CGCTTTCGCGCGAGAATAATAGGAGTGTGCGTGGATTGTGGTAGAAAAACACGATAAG

GCATCGAGGACTGAACGTAAGTGTAAAGTCAACAATACACAATCTCAGGCCGAGGTG
GCGCATATGTTAATGCACGATCCCGTGAAATATTTAAATAAAAGAAAAGCTAGAGCTT
TTTCTAACGCAGAGATGTTTGCGATCGATTTGGTTATGCATACCAAAGAAAGGCAATT
AGCGGTTGATTTGGCCGCTGAAAGGGAGAAGACGAGATTGGCTCGTAGACACCCGAT
GCGTTCTCCGGAGGAACTCCGGAACATTATAAGTTCGGGTATAACTGCTAAGGCAATG
TTACCGGACATCAACGCTATAGACGTTGGTGATAACGAGGACACTTCGTCGGAATACC
CAGTGAGTCTGAGTGTTCAGGCCGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 31: The p23 gene DNA sequence of South African Isolate GFMS 12.

ISOLATE 389-3

ATGGATGATACTCAGCGGACAACTTTCTGTTTCTGTGAACCTTTCTGACGAAAGCAAC
ACAGCGAGCACTAGAGTTGAAAACGTAAATTCGGAAGCGGATCGCTTGGAGTTTTTAC
GTAAAATGAATCCCCTCATTATTGACGCTCTGGTGCGGAAAACCAATTATCAGGGCGC
TCGCTTTCTGTGCGAGAATAATAGGAGTATGCGTGGATTGTGGTAGAAAACACGACAA
GGCGCTCAAGACTGAACGTAAGTGTAAAGGTCAACAATACGCAATCTCAGAACGAGGT
GGCGCATATGTTGATGCACGATCCCGTTAAGTATTTGAACAAAAGAAAGGCTAGAGCC
TTTTCTAACGCAGAGATGTTTGCGATTGAATTGGTTTTGTACACCAAGGAAAGGCAAT
TGGCGGTCGATTTAGCCGCTGAAAGGGAGAAGACGAGACTGGCTCGTAGACACCCAA
TACGTTCTCCGGAAGAACTCCGGAACATTATAAATTCGGTATGACTGCTAAGGCAAT
GTTACCGGACATCAACGCCGTAGACGTTGGTGATAACGAGGAACTTCGTCGGAGTAC
CCAGTGAGTCTGAGTGTTCCTGGCGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 32: The p23 gene DNA sequence of South African Isolate 389-3.

ISOLATE 389-4

ATGGATGATACTATCGGACAACTTTCTGTTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCGAGCACTAGAGTTGAAAACGTAAATTCGGAAGCGGATCGCTTGGAGTTTTTACG
TAAAATGAATCCCCTCATTATTGACGCTCTGGTGCGGAAAACCAATTATCAGGGCGCT
CGCTTTCTGTGCGAGAATAATAGGAGTATGCGTGGATTGTGGTAGAAAACACGACAAG
GCGCTCAAGACTGAACGTAAGTGTAAAGGTCAACAATACGCAATCTCAGAACGAGGTG
GCGCATATGTTGATGCACGATCCCGTTAAGTATTTGAACAAAAGAAAGGCTAGAGCCT
TTTCTAACGCAGAGATGTTTGCGATTGAATTGGTTTTGTACACCAAGGAAAGGCAATT
GGCGGTCGATTTAGCCGCTGAAAGGGAGAAGACGAGACTGGCTCGTAGACACCCAAT
ACGTTCTCCGGAAGAACTCCGGAACATTATAAATTCGGTATGACTGCTAAGGCAATG
TTACCGGACATCAACGCCGTAGACGTTGGTGATAACGAGGAACTTCGTCGGAGTACC
CAGTGAGTCTGAGTGTTCCTGGCGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 33: The p23 gene DNA sequence of South African Isolate 389-4.

ISOLATE 390-5

ATGGACGATACTAGCGGACAACTTTCTGTTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCAAGTACTGCAGTTAGAACCGTAAGTTCGGAAGCGGATCGCTTGGAAATTTTTACG
AAAAATGAATCCCCTTATCATCGACGCTTTGGTACGGAAAACCAATTATCAGGGTGCT
CGCTTTCTCGCGCAAGAATAATAGGAGTGTGCGTAGATTGTGGTAGAAAACACGATAAG

GGGTTGAAGACCGAACGTAATGTAAGGTCACAATACACAGTCTCAGAACGAGGTG
GCGCATATGTTAATGCACGATCCCGTTAGGTATTTAAATAAAGGAAAGGCTAGAGCCT
TTTCTAACGCAGAGATGTTTGCGATCGATTTGGTTATGTACACCAAGGAAAAGCAGTT
GGCGGTTAATTTGGCCGCTGAAAGGGAGAAGACGAGACTGGCTCGTAGACACCCGAT
GCGTTCTCCGGAAGAACTCCGGAACACTATAAATTCGGTATAACTGCTAAAGCAATG
TTACCGAACATCAACGCTGTGGACGTTGGTGATAACGAGGACACTTCGTCGGAGTACC
CAGTGAGTCTGAGTGTTTCTGACGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 34: The p23 gene DNA sequence of South African Isolate 390-5.

ISOLATE GFMS 35

ATGGATGATACTAGCGGACAACTTTTCGTTTCTGTGAACCTTTCTGACGAAAGCAACA
CAGCGAGCACTAGAGTTGAAAACGTAAAATCGGAAGCGGATCGCTTGGAGTTTTTAC
GTAAAATGAATCCCCTCATTATTGACGCTCTGGTGCGGAAAACCAATTATCAGGGCGC
TCGCTTTTCGTGCGAGAATAATAGGAGTATGCGTGGATTGTGGTAGAAAACACGACAA
GGCGCTCAAGACTGAACGTAAGTGTAAGGTCAACAATACGCAATCTCAGAACGAGGT
GGCGCATATGTTGATGCACGATCCCGTTAAGTATTTGAACAAAAGAAAGGCTAGAGCC
TTTTCTAACGCAGAGATGTTTGCGATTGAATTGGTTTTGTACACCAAGGAAAGGCAAT
TGGCGGTCGATTTAGCCGCTGAAAGGGAGAAGACGAGACTGGCTCGTAGACACCCAA
TACGTTCTCCGGAAGAACTCCGGAACATTATAAATTCGGTATGACTGCTAAGGCAAT
GTTACCGGACATCAACGCCGTAGACGTTGGTGATAACGAGGAACTTCGTCGGAGTAC
CCAGTGAGTCTGAGTGTTTCTGGCGGAGTTCTCCGTGAACACCACTTCATCTGATT

Figure 35: The p23 gene DNA sequence of South African Isolate GFMS 35.

Figure 36

Multiple nucleotide sequence alignment of the p23 gene of CTV isolates from RSA and other geographic areas using Clustal W. Nucleotides that differ from the consensus sequence are shown; dots indicate where sequence identity occurs.

```

GFMS 12 ATGGACGATACTAGCGGACAAACACCTTTCTGACGAAAAGCAACACAGCTAGTACTG
GFMS 35 .....T.....CG.....G..C...A
12-5 .....
12-7 .....
12-9 .....C.....C...
389-3 .....T...TC...CG.....G..C...A
389-4 .....T...T...CG.....G..C...A
390-5 .....T...CG.....A...
VT-23-1 .....C.....C...
C268-2 ...A.....G...CG.....G..C...A
C270-3 .....C.....C...T.....G.....
Barao-2 .....
C269-6 .....C.....
Cal-23- .....
50-23-1 .....C.....G.....C...
clone K ...A.T.....CG.....C.....G...C.
32-23-1 .....T.....CG.....G..A..C...A
36-23-1 .....TA.....CG.....G..CC...
T55-23- .....TA.....CG.....G..A..C...A
T300-23 .....T.....CG.....G..A..C...A
T305-23 .....
T312-23 .....T.....CG.....G..A..C...A
T346-23 .....CG.....G.....A...
T385-23 .....T.....T.....CG.....G..A..C...A
T388-23 .....

GFMS 12 AAATCAAAACCGTAAGTTCGGAAGCGGATCGCTTGGAATTTTACGGAAAATGAATCCCTTCATTATCGA
GFMS 35 G.G.TG...A...AA.....G.....T.....C.....T..
12-5 .....
12-7 .....
12-9 .....
389-3 G.G.TG...A...A.....G.....T.....C.....T..
389-4 G.G.TG...A...A.....G.....T.....C.....T..
390-5 C.G.T.G.....A.....T..C...
VT-23-1 ..G.....T.....T.....T.....
C268-2 G.G.TG...A...AA.....T.....T.....A.....T..
C270-3 ..G.T.....T.....T..C...
Barao-2 .....G.....
C269-6 .....G.....
Cal-23- .....
50-23-1 ..G.....T.....T..C...
clone K ..G.T.....A.....T..C...
32-23-1 ..G.TG...A...AA.....T.....T.....G.T..
36-23-1 ..CG...C...G.....T.....A.....A.T...
T55-23- ..G.TG...A...AA.....G.....T.....T.....G.T..
T300-23 ..G.TG...A...AA.....T.....T.....T.....G.T..
T305-23 .....
T312-23 ..G.TG...A...AA.....T.....T.....T.....G.T..
T346-23 G.G.TG...A...AA.....G.....T.....T.....T..
T385-23 ..G.TG...A...AA.....T.....T.....T.....G.T..
T388-23 .....

GFMS 12 CGCTTTGATACGGAAAAATAGTTATCAAGGCGCTCGCTTTTCGCGCGAGAATAATAGGAGTGTGCGTGGAT
GFMS 35 ...C..G.G.....CC.A...G.....T.....A.....
12-5 .....
12-7 .....
12-9 .....
389-3 ...C..G.G.....CC.A...G.....T.....A.....
389-4 ...C..G.G.....CC.A...G.....T.....A.....
390-5 ...G.....CC.A...G..T.....A.....A...
VT-23-1 ...C.....C.....G..T.....
C268-2 ...C..G.G.....CC.A...G.....T.....A.....
C270-3 ...C.....CC.....G..T.....G.....A...
Barao-2 .....
C269-6 .....C.....
Cal-23- .....
50-23-1 ...C.....CC.A...G..T.....
clone K .....GCC.....G..T.....C.....
32-23-1 ...C..G.G.....CC.A...G..T.....A.....
36-23-1 T.....G.....G.....
T55-23- ...C..G.G.....CC.A...G..T.....A.....
T300-23 ...C..G.G.....CC.A...G..T.....A.....
T305-23 .....
T312-23 ...C..G.G.....CC.A...G..T.....A.....
T346-23 ...C..G.G.....C..A...G.....T.....
T385-23 ...C..G.G.....CC.A...G..T.....A.....
T388-23 .....

```

Figure 36 continued

GFMS 12	TGTGGTAGAAAAACACGATAAGGCATCGAGGACTGAACGTAAGTGTAAGTCAACAATACACAATCTCAGG
GFMS 35	C.....GCTC.A.....G.....G.....G.....A
12-5	A.....G.....G.....G.....
12-7	A.....G.....G.....
12-9	A.....G.....G.....G.....
389-3	C.....GCTC.A.....G.....G.....A
389-4	C.....GCTC.A.....G.....G.....A
390-5	GG.T.A..C.....A.....G.....G.....A
VT-23-1	C.....GCTC.A.....G.....G.....A
C268-2	C.....GCTC.A.....G.....G.....A
C270-3	GG.T.A..C.....A.....G.....G.....A
Barao-2	G.....G.....G.....A
C269-6	G.....G.....G.....A
Cal-23-	G.....G.....G.....A
50-23-1	A.....GG.T.....C.....A.....G.....G.....A
clone K	GG.T.A..C.....G.....G.....A
32-23-1	C.....GCTC.A.....G.....G.....A
36-23-1	..C.....GG.T.A.....G.....G.....A
T55-23-	C.....GCTC.A.....G.....C.G.....A
T300-23	C.....GCTC.A.....G.....G.....A
T305-23	A.....G.....G.....A
T312-23	C.....GCTC.A.....G.....G.....A
T346-23	C.....GCTC.A.....G.....G.....A
T385-23	C.....GCTC.A.....G.....G.....A
T388-23	A.....G.....G.....A
GFMS 12	CCGAGGTGGCGCATATGTTAATGCACGATCCCGTGAAATATTTAAATAAAAAGAAAAGCTAGAGCTTTTTC
GFMS 35	A.....G.....T..G.....G.....C.....G.....C.....
12-5	GCTC.A.....A.....C.....
12-7	A.....C.....
12-9	G.....C.....
389-3	A.....G.....T..G.....G.....C.....G.....C.....
389-4	A.....G.....T..G.....G.....C.....G.....C.....
390-5	A.....T.GG.....G.....G.....C.....
VT-23-1	A.....G.....
C268-2	A.....G.....T.....C.....G.....C.....
C270-3	A.....T..G.....G.G..G.....C.....
Barao-2	A.....A.....C.....
C269-6	A.....A.T.....C.....
Cal-23-	A.....A.....C.....
50-23-1	A.....A.....TG.G.....G.....G.....C.....
clone K	A.....T..G.....C.....G.....C.....
32-23-1	A.....G.....T..G.....G.....C.....G.....C.....
36-23-1	A.....C.....T..G.....C.....C.....
T55-23-	A.....G.....T..G.....G.....C.....G.....C.....
T300-23	A.....G.....T.....G.....C.....G.....C.....
T305-23	A.....A.....C.....
T312-23	A.....G.....T..G.....G.....C.....G.....C.....
T346-23	A.....G.....T.GG.....C.....G.....C.....
T385-23	A.....G.....T..G.....G.....C.....G.....C.....
T388-23	A.....A.....C.....
GFMS 12	TAACGCAGAGATGTTTGGCGATCGATTTGGTTATGCATACCAAAGAAAGGCAATTAGCGGTTGATTTGGCC
GFMS 35	T..A.....T..T.C.....G.....G.....C.....A...
12-5	C.....
12-7	C.....
12-9	T..A.....T..T.C.....G.....G.....C.....
389-3	T..A.....T..T.C.....G.....G.....C.....A...
389-4	T..A.....T..T.C.....G.....G.....C.....A...
390-5	T.C.....G.....A..G..G.....A.....
VT-23-1	T..A.....T..T.C.....G.....A.....G.....
C268-2	T.C.....G.....A..G..G.....A.....AA..
C270-3	C.....
Barao-2	C.....
C269-6	C.....
Cal-23-	C.....
50-23-1	T.C.....G.....A..G.C.....A.....A...
clone K	T..A.....T..C.....G.....G.....C.....A...
32-23-1	T..A.....T..T.C.....G.....G.....C.....A...
36-23-1	..T..G.....A.....T.....T.C.....G.....G..TA.....
T55-23-	T..AC.....TC.T.C.....G.....G.....C.....A...
T300-23	T..A.....T..T.C.....G.....G.....C.....A...
T305-23	T.C.....
T312-23	T..A.....T..T.C.....G.....G.....C.....A...
T346-23	C.....T..A.....T..T.C.....G.....A.....G.....
T385-23	T..A.....T..T.C.....G.....G.....C.....A...
T388-23	C.....

Figure 36 continued

GFMS 12	GCTGAAAAGGGAGAAGACGAGATTGGCTCGTAGACACCCGATGCGTTCTCCGGAGGAAACTCCGGAACATT
GFMS 35	C.....A..A.....A.....
12-5	A..G.....
12-7	A..G.....
12-9	
389-3	C.....A..A.....A.....
389-4	C.....A..A.....A.....
390-5	C.....A.....C.
VT-23-1	G.....
C268-2	..C.....G.....A.....A.....
C270-3	C.....A.....C.
Barao-2	A..G.....
C269-6	A..G.....
Cal-23-	A..G.....
50-23-1	A.....C.....C.
clone K	C.....A.....A.....
32-23-1	C.....A..A.....A.....
36-23-1	A.....C.....A.....T.
T55-23-	C.....A..A.....A.....
T300-23	C.....A..A.....A.....
T305-23	A..G.....
T312-23	A.....C.....A..A.....A.....
T346-23	..C.....G..G.....A.....A.....
T385-23	C.....A..A.....A.....
T388-23	A..G.....
GFMS 12	ATAAGTTCGGTATAAAGTCTAAGGCAATGTTACCGGACATCAACGCTATAGACGTTGGTGATAACGAGGA
GFMS 35	...A.....G.....CG.....
12-5	...A.....C....C.....C.....A..
12-7	...A.....C....C.....C.....A..
12-9	A.....
389-3	...A.....G.....CG.....
389-4	...A.....G.....CG.....
390-5	...A.....A.....A.....G.G.....
VT-23-1	G.....A.....A.....
C268-2	...A.....G..C.....A.TG.....AC.....G..G.....A..
C270-3	...A.....G.....A.....G.G.T.....
Barao-2	...A.....G.....G.....A..
C269-6	...A.....G..C.....G.....A..
Cal-23-	...A.....G.....G.....A..
50-23-1	...A.....G.....A.....C.....T.....G.....
clone K	...A.....CG.....G.....CG.....
32-23-1	...A.....G.....C.....CG.....
36-23-1	...A.....GG.....A.....CG.....
T55-23-	...A.....G.....CG.....
T300-23	...A.....G.....CG.....
T305-23	...A.....G.....C.....G.....C.....A..
T312-23	...A.....G.....AT.....CG.....
T346-23	.C..A.....G.....A..G.....AT.....G..G.....A..
T385-23	...A.....G.....AT.....CG.....
T388-23	...A.....G.....C.....G.....C.....A..

Figure 36 continued

GFMS 12 CACTTCGTCGGAATACCCAGTGAGTCTGAGTGTTCAGGCGGAGTTCTCCGTGAACACCACTTCATCTGA
GFMS 35 A.....G.....T.....
12-5 ...C.....G.....A.....C.....
12-7 ...C.....G.....A.....C.....
12-9
389-3 A.....G.....T.....
389-4 A.....G.....T.....
390-5 G.....T.A.....
VT-23-1
C268-2 C..T.....
C270-3 G.....A.C..T.A.....
Barao-2 ...C.....G.....A.....C.....
C269-6 ...C.....G.....A.....C.....
Cal-23- ...C.....G.....A.....C.....
50-23-1 G.....TAA.....
clone K A.....G.....T.....
32-23-1 A.....G.....T.....T.....
36-23-1 A.....G..T.....T.....
T55-23- A.....G.....T.....
T300-23 A.....G.....T.....
T305-23 ...C.....G.....A.....C.....
T312-23 A.....G.....T.....
T346-23 A.....C..T.....
T385-23 A.....G.....T.....
T388-23 ...C.....G.....A.....C.....

GFMS 12 TT-
GFMS 35 ..-
12-5 ..-
12-7 ..-
12-9 ..-
389-3 ..-
389-4 ..-
390-5 ..-
VT-23-1 ..-
C268-2 ..G
C270-3 ..G
Barao-2 ..-
C269-6 ..G
Cal-23- ..-
50-23-1 ..-
clone K ..-
32-23-1 ..-
36-23-1 ..-
T55-23- ..-
T300-23 ..-
T305-23 ..-
T312-23 ..-
T346-23 ..-
T385-23 ..-
T388-23 ..-

Table 19:

Summary of optimised PCR conditions for each of the 23 primer sets

Primer set	Immunocapture step ^a	[MgCl ₂ , mM]	PCR Cycles	PCR Annealing temperature (° C)
T30 1+/1-	No	2.0	35	55
T30 2+/2-	No	2.0	35	55
T30 3+/3-	No	2.0	35	55
T30 4+/4-	No	2.0	30	45
T30 5+/5-	No	2.0	30	45
T30 6+/6-	Yes	1.5	35	55
T30 7+/7-	No	2.0	35	55
T36 1+/1-	Yes	2.0	35	55
T36 2+/2-	Yes	2.0	35	55
T36 3+/3-	Yes	2.0	35	55
T36 5+/5-	Yes	2.0	35	55
T36 6+/6-	No	2.0	35	55
T36 7+/7-	No	2.0	35	55
VT 1+/1-	Yes	1.5	35	55
VT 2+/2-	Yes	1.5	35	38
VT 3+/3-	No	1.5	35	55
VT 4+/4-	No	1.5	35	45
VT 5+/5-	Yes	1.5	35	45
VT 6+/6-	Yes	1.5	35	55
T3 2+/2-	No	1.5	25	55
T3 3+/3-	No	1.5	30	55
T3 5+/5-	No	1.5	30	55
T3 6+/6-	No	1.5	25	55

^a “No” indicates that Immunocapture step was not performed and “Yes” indicates it was used.

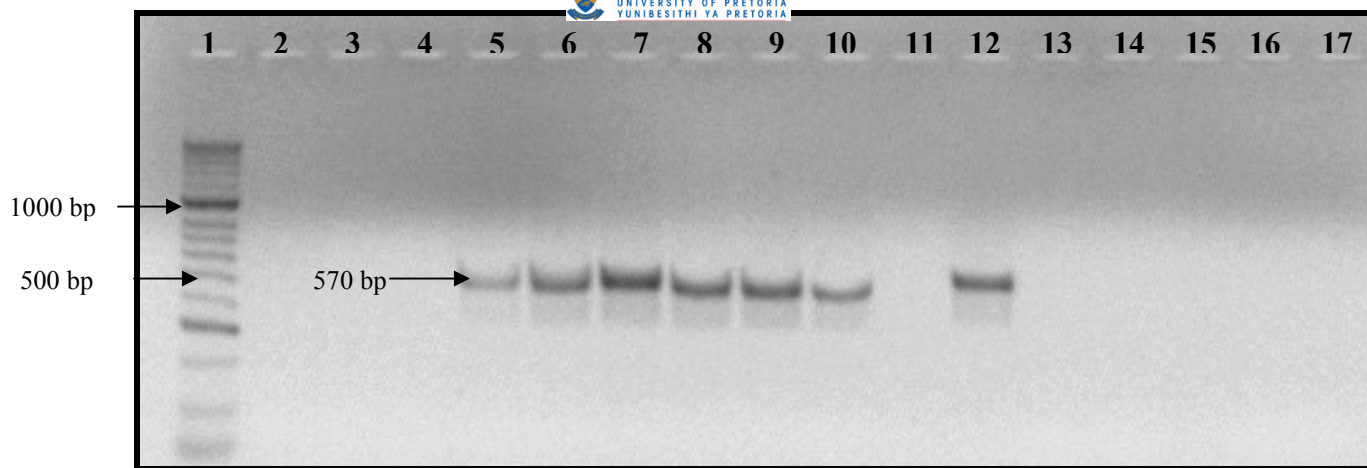


Figure 37: 1% Agarose gel electrophoresis of the T30 1+/- molecular marker PCR primer set. Expected PCR product is 570 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) T30 clone; (13) T36 clone; (14) T3 clone; (15) VT clone; (16) 390-5 and (17) Negative control.

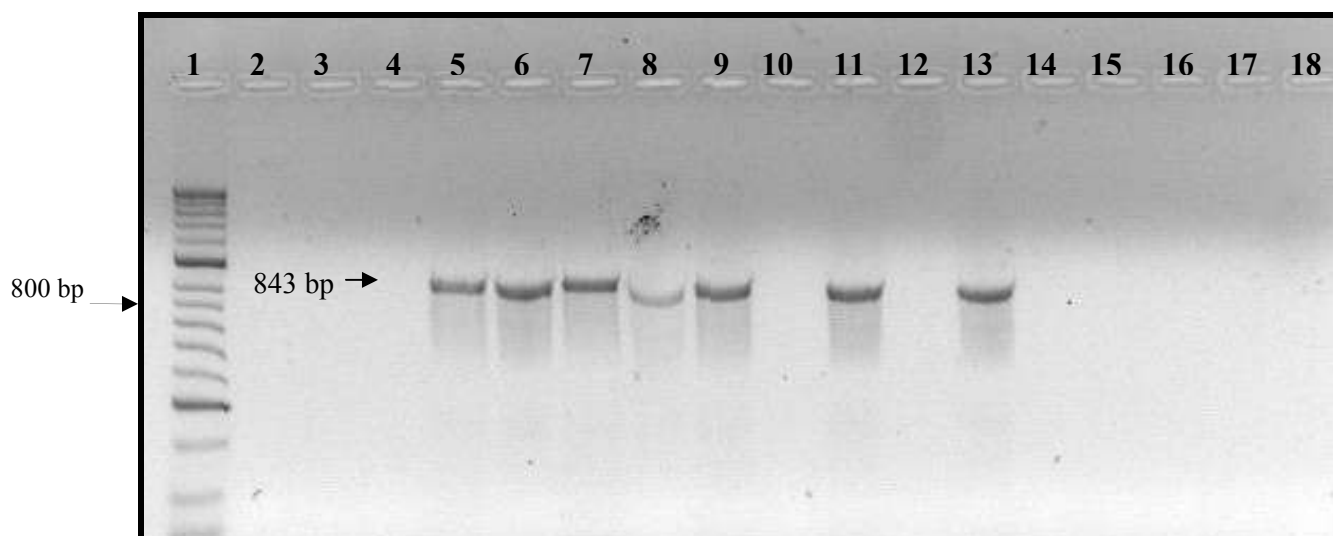


Figure 38: 1% Agarose gel electrophoresis of the T30 2+/- molecular marker PCR primer set. Expected PCR product is 843 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone (16) VT clone; (17) Virus Free Control and (18) Negative control.

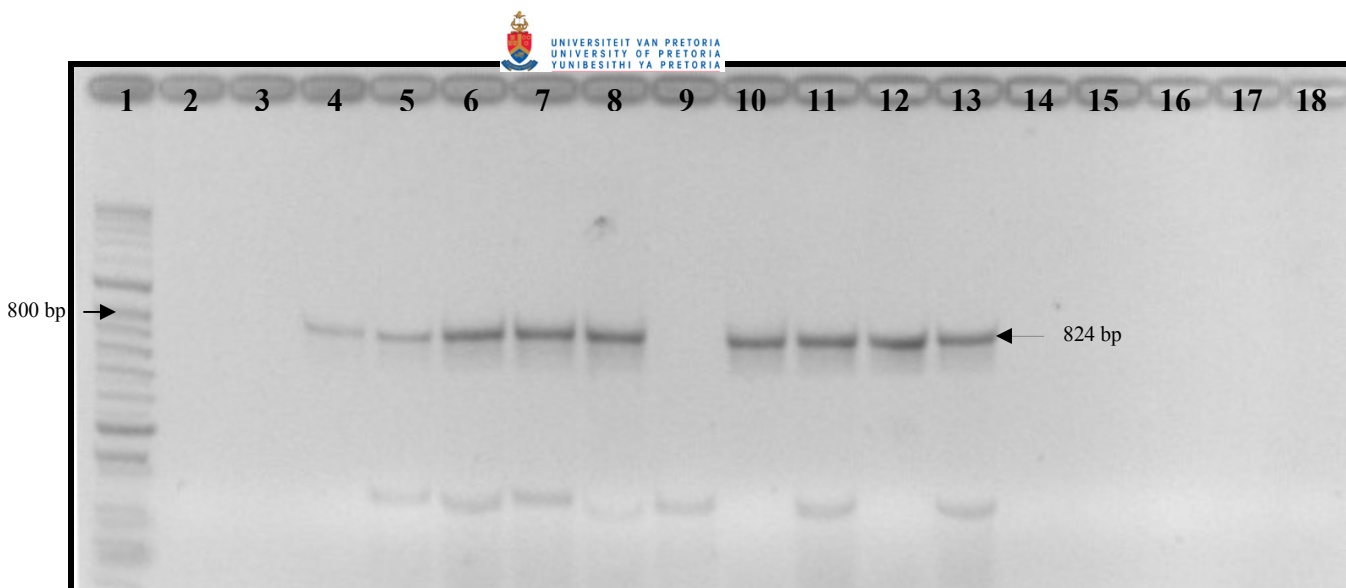


Figure 39: 1% Agarose gel electrophoresis of the T30 3+/- molecular marker PCR primer set. Expected PCR product is 824 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

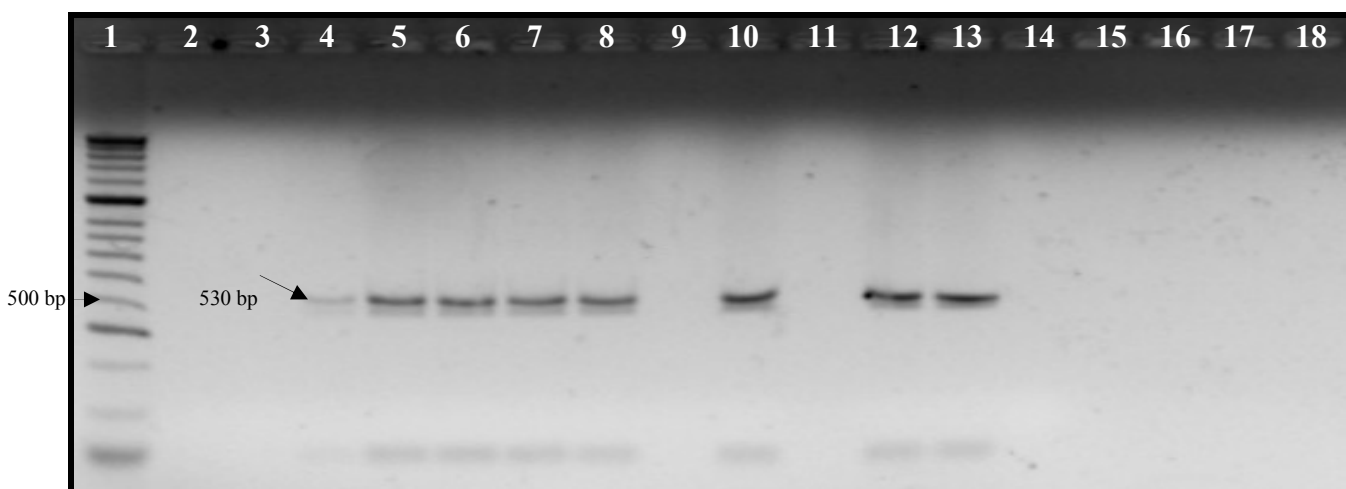


Figure 40: 1% Agarose gel electrophoresis of the T30 4+/- molecular marker PCR primer set. Expected PCR product is 530 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

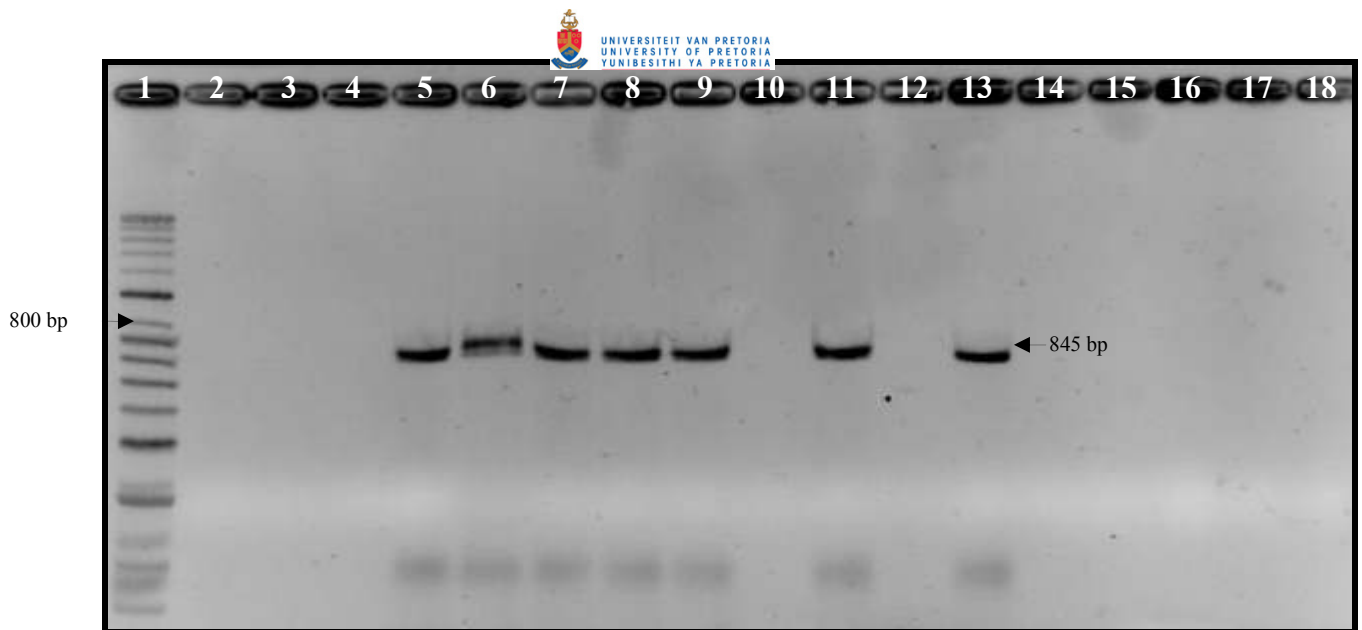


Figure 41: 1% Agarose gel electrophoresis of the T30 5+/- molecular marker PCR primer set. Expected PCR product is 845 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

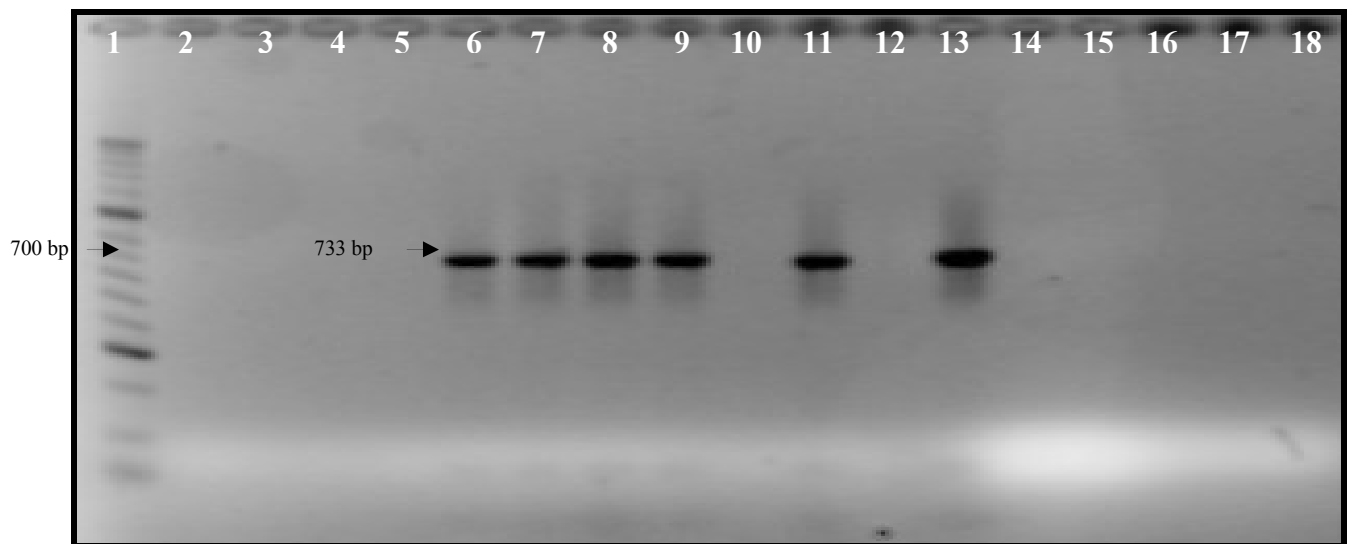


Figure 42: 1% Agarose gel electrophoresis of the T30 6+/- molecular marker PCR primer set. Expected PCR product is 733 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

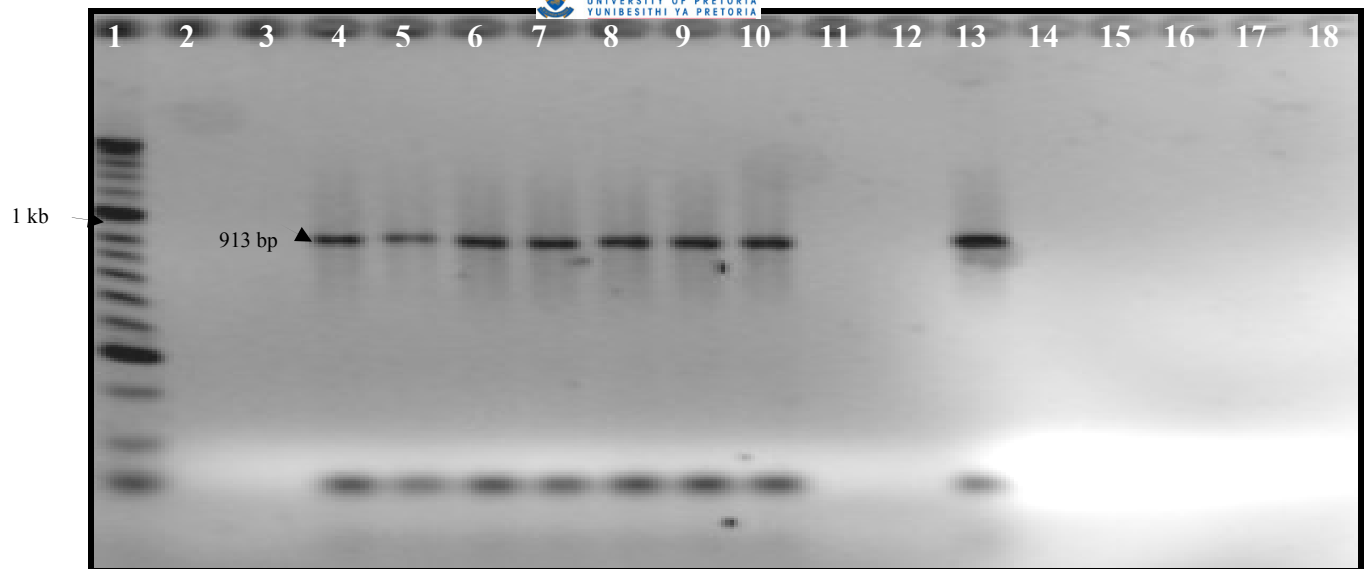


Figure 43: 1% Agarose gel electrophoresis of the T30 7+/- molecular marker PCR primer set. Expected PCR product is 913 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

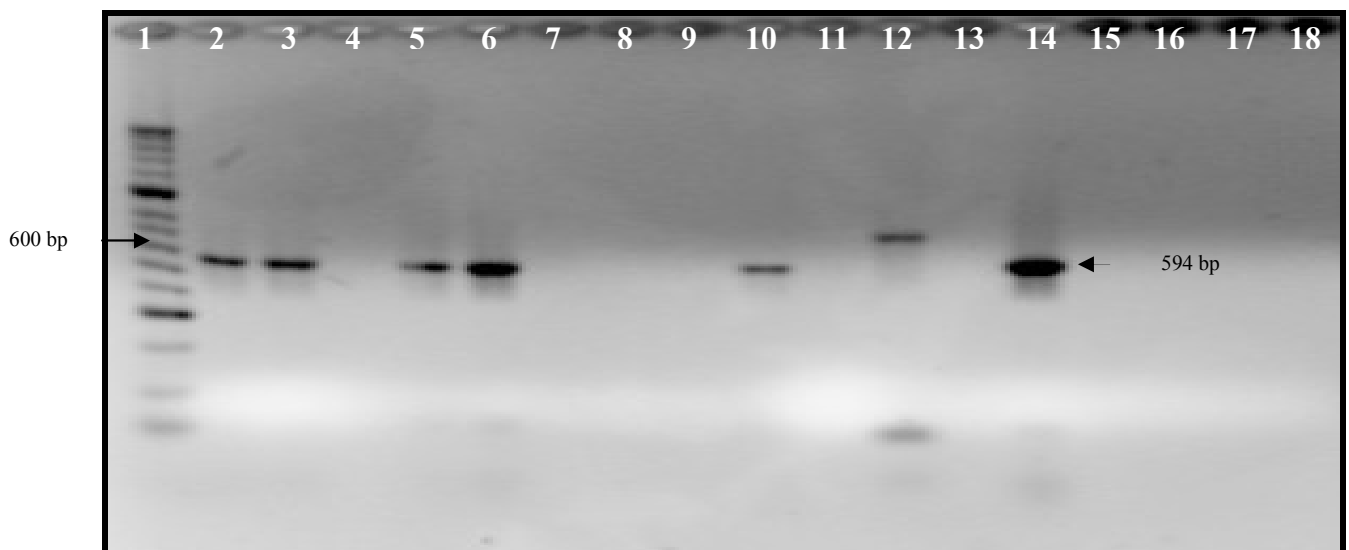


Figure 44: 1% Agarose gel electrophoresis of the T36 1+/- molecular marker PCR primer set. Expected PCR product is 594 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-9; (3) 12-7; (4) 12-5; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-5; (11) 390-4; (12) 390-3; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

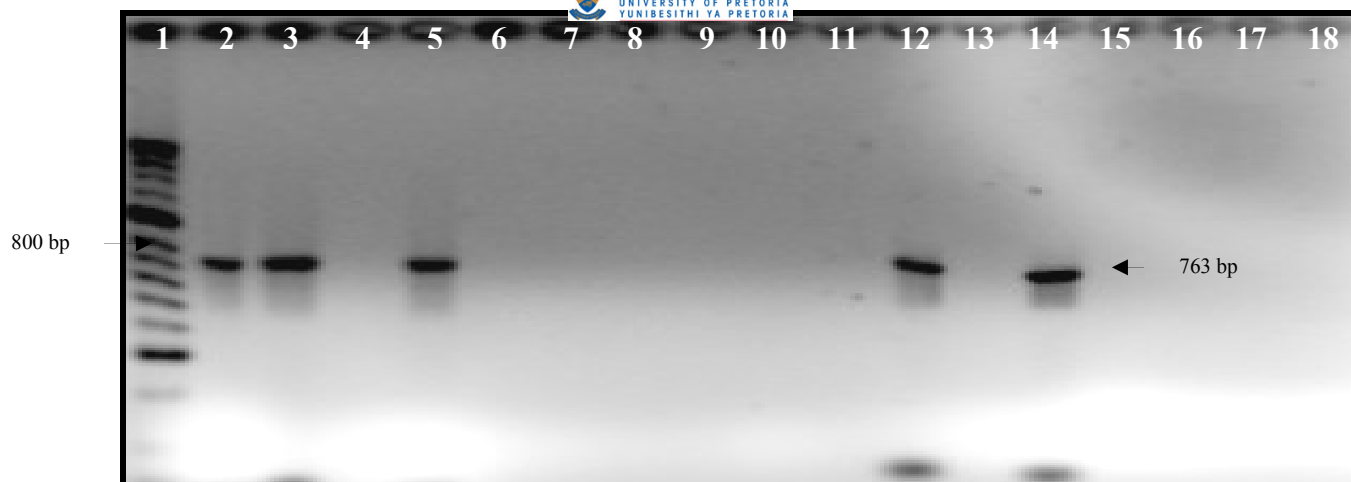


Figure 45: 1% Agarose gel electrophoresis of the T36 2+/- molecular marker PCR primer set. Expected PCR product is 763 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-9; (3) 12-7; (4) 12-5; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

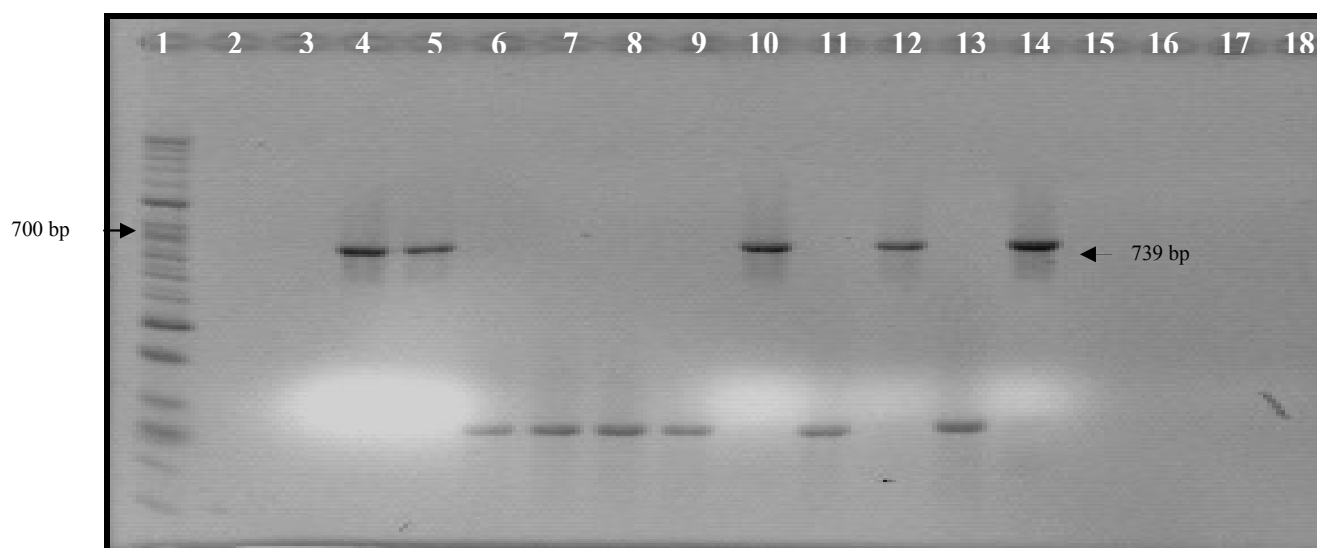


Figure 46: 1% Agarose gel electrophoresis of the T36 3+/- molecular marker PCR primer set. Expected PCR product is 739 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

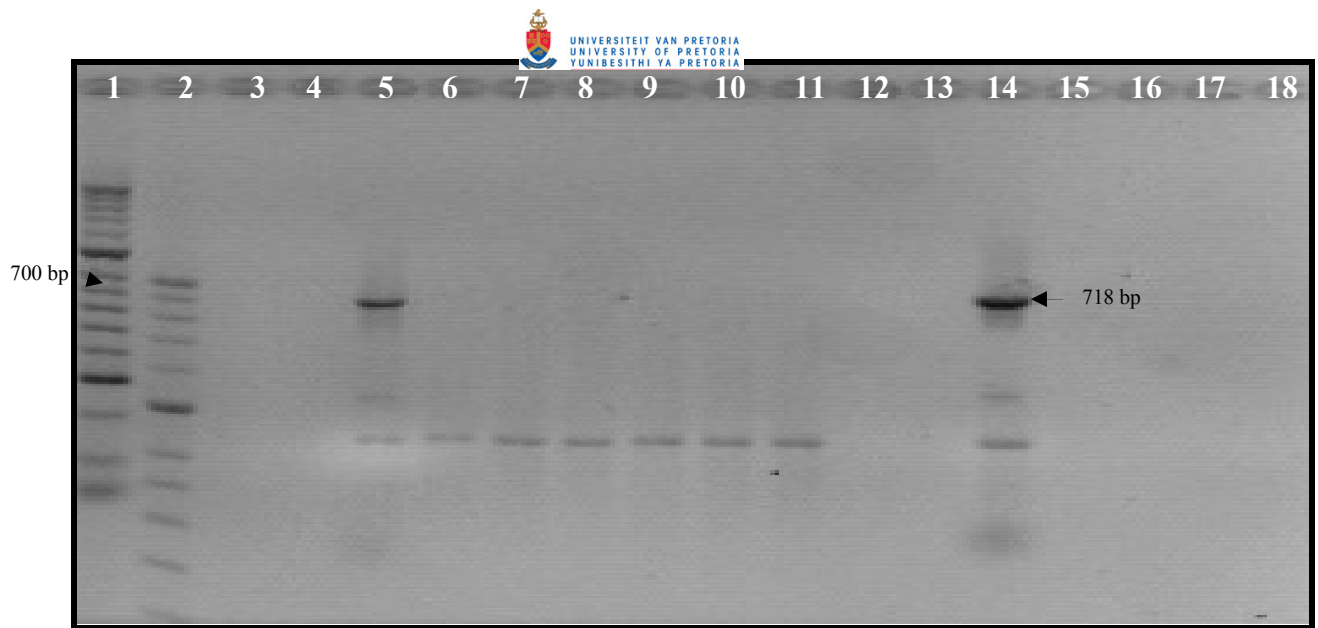


Figure 47: 1% Agarose gel electrophoresis of the T36 5+/- molecular marker PCR primer set. Expected PCR product is 718 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) Marker; (3) 12-5; (4) 12-7; (5) GFMS12; (6) 12-9; (7) GFMS35; (8) T30; (9) 389-3; (10) 389-4; (11) 390-3; (12) 390-4; (13) 390-5; (14) T36 clone; (15) T30 clone; (16) T3 clone; (17) VT clone; (18) Virus Free Control and (19) Negative control.

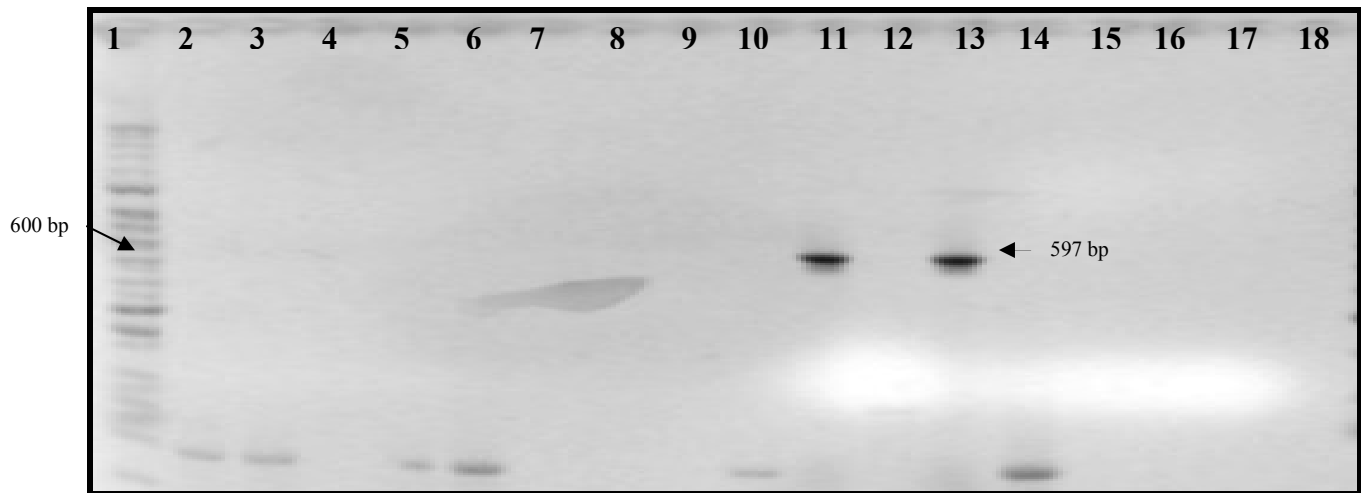


Figure 48: 1% Agarose gel electrophoresis of the T36 6+/- molecular marker PCR primer set. Expected PCR product is 597 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-5; (12) 390-4; (13) T36 clone; (14) T30 clone; (15) T3 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

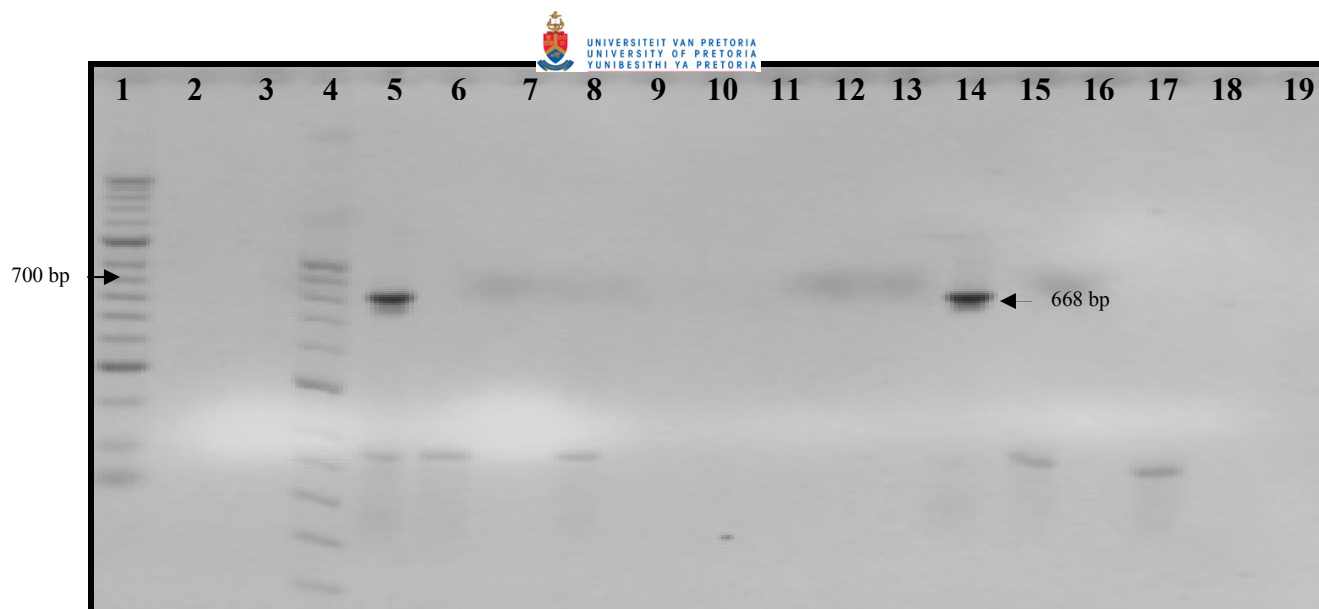


Figure 49: 1% Agarose gel electrophoresis of the T36 7+/- molecular marker PCR primer set. Expected PCR product is 668 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) Marker; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) T3 clone; (16) VT clone; (17) 12-9; (18) Virus Free Control and (19) Negative control.



Figure 50: 1% Agarose gel electrophoresis of the VT 1+/- molecular marker PCR primer set. Expected PCR product is 564 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-5; (11) 390-4; (12) 390-3; (13) T30 clone; (14) VT clone; (15) T36 clone; (16) T3 clone; (17) Virus Free Control and (18) Negative control.



Figure 51: 1% Agarose gel electrophoresis of the VT 2+/- molecular marker PCR primer set. Expected PCR product is 834 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-5; (11) 390-4; (12) 390-3; (13) T30 clone; (14) VT clone; (15) T36 clone; (16) T3 clone; (17) Virus Free Control and (18) Negative control.



Figure 52: 1% Agarose gel electrophoresis of the VT 3+/- molecular marker PCR primer set. Expected PCR product is 824 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) VT clone; (15) T36 clone; (16) T3 clone; (17) Virus Free Control and (18) Negative control.

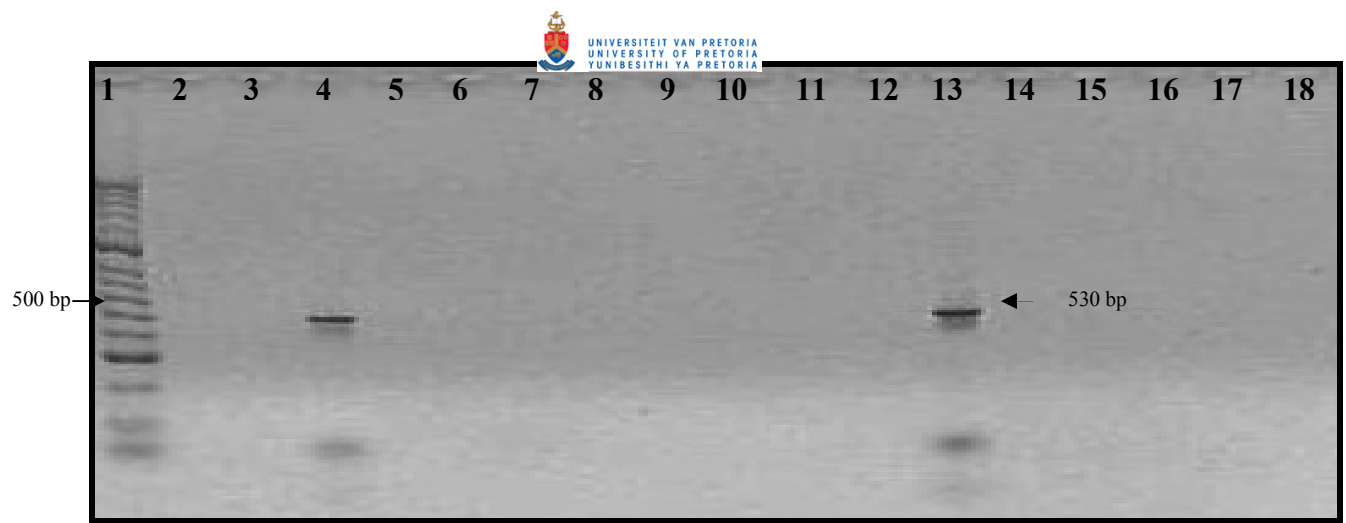


Figure 53: 1% Agarose gel electrophoresis of the VT 4+/- molecular marker PCR primer set. Expected PCR product is 530 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) T30; (6) 389-3; (7) 389-4; (8) 390-3; (9) 390-4; (10) 390-5; (11) T30 clone; (12) T36 clone; (13) VT clone; (14) T3 clone; (15) GFMS 12; (16) GFMS 35; (17) Virus Free Control and (18) Negative control.

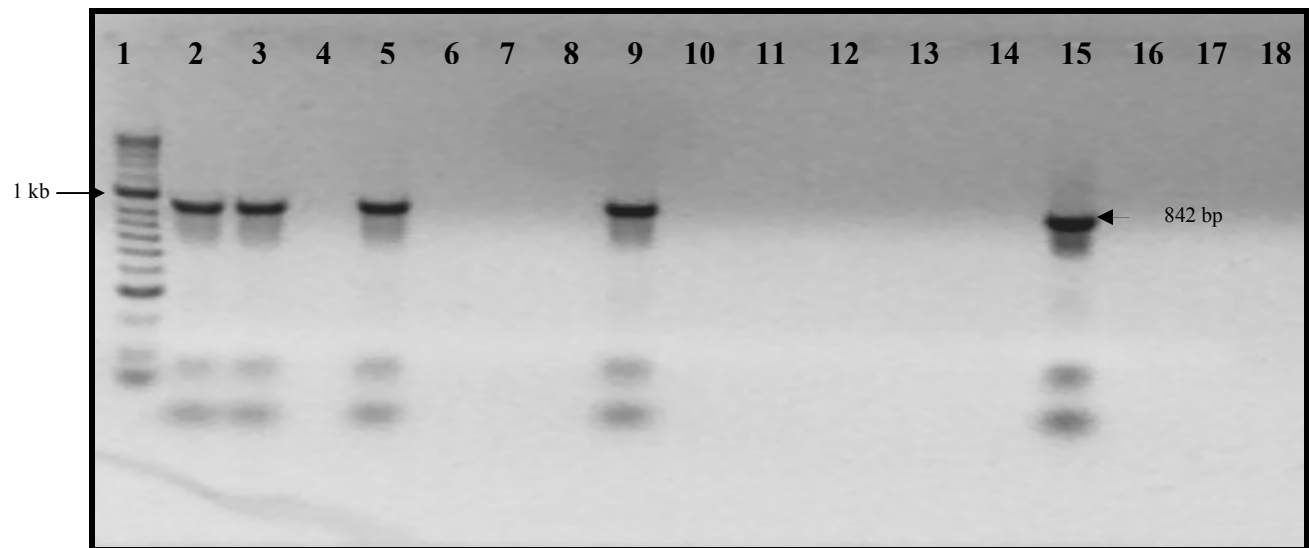


Figure 54: 1% Agarose gel electrophoresis of the VT 5+/- molecular marker PCR primer set. Expected PCR product is 842 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) GFMS 35; (5) 12-9; (6) T30; (7) 389-3; (8) 389-4; (9) GFMS 12; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) VT clone; (16) T3 clone; (17) Virus Free Control and (18) Negative control.



Figure 55: 1% Agarose gel electrophoresis of the VT 6+/- molecular marker PCR primer set. Expected PCR product is 733 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS 12; (6) GFMS 35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T36 clone; (15) VT clone; (16) T3 clone; (17) Virus Free Control and (18) Negative control.

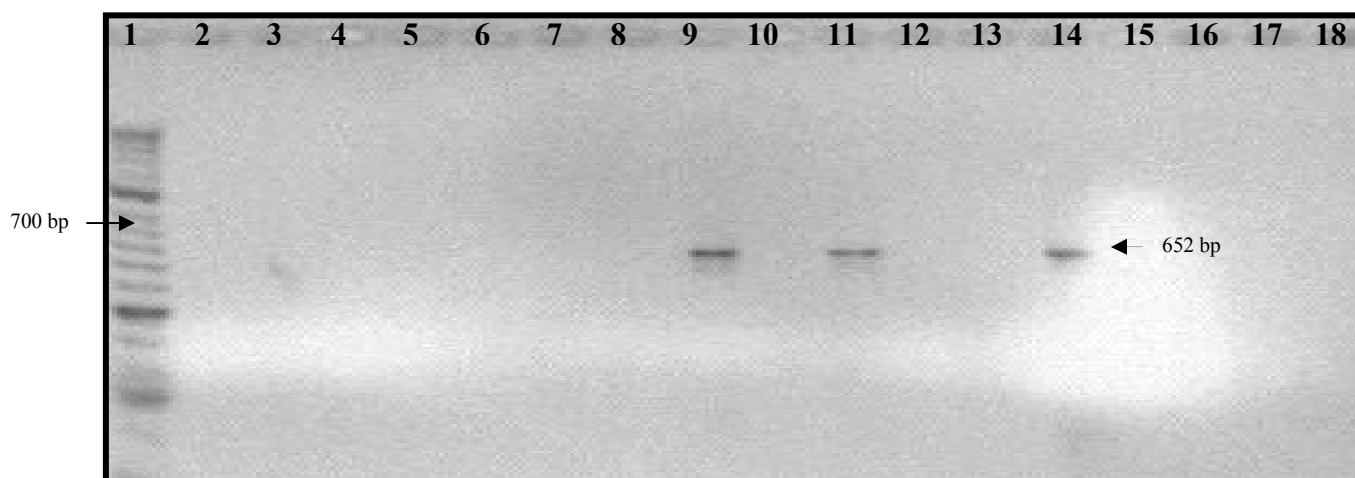


Figure 56: 1% Agarose gel electrophoresis of the T3 2+/- molecular marker PCR primer set. Expected PCR product is 652 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T3 clone; (15) T36 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

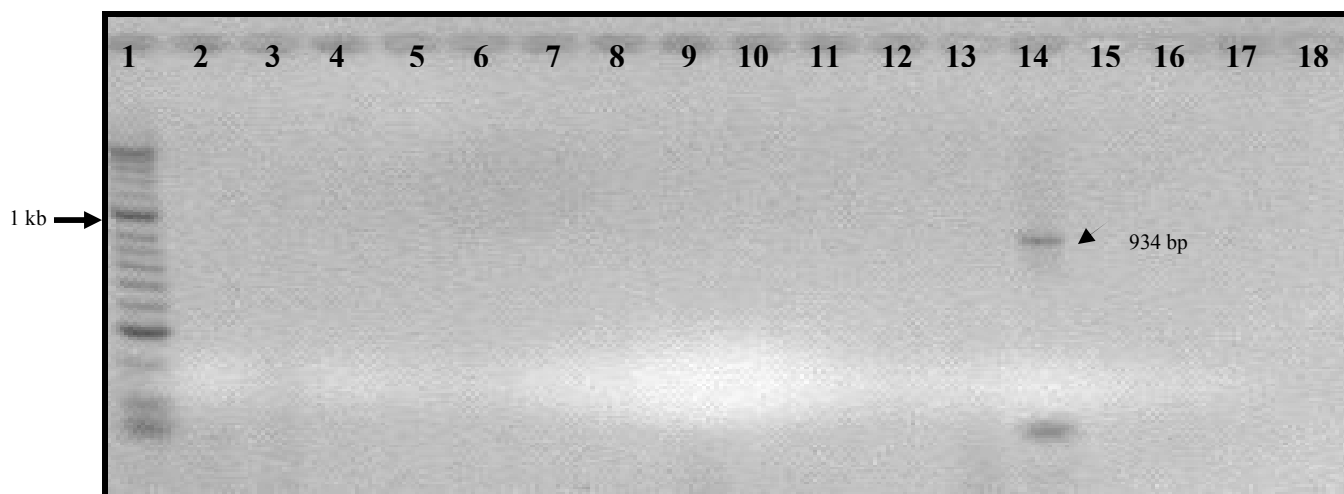


Figure 57: 1% Agarose gel electrophoresis of the T3 3+/- molecular marker PCR primer set. Expected PCR product is 934 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T3 clone; (15) T36 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.



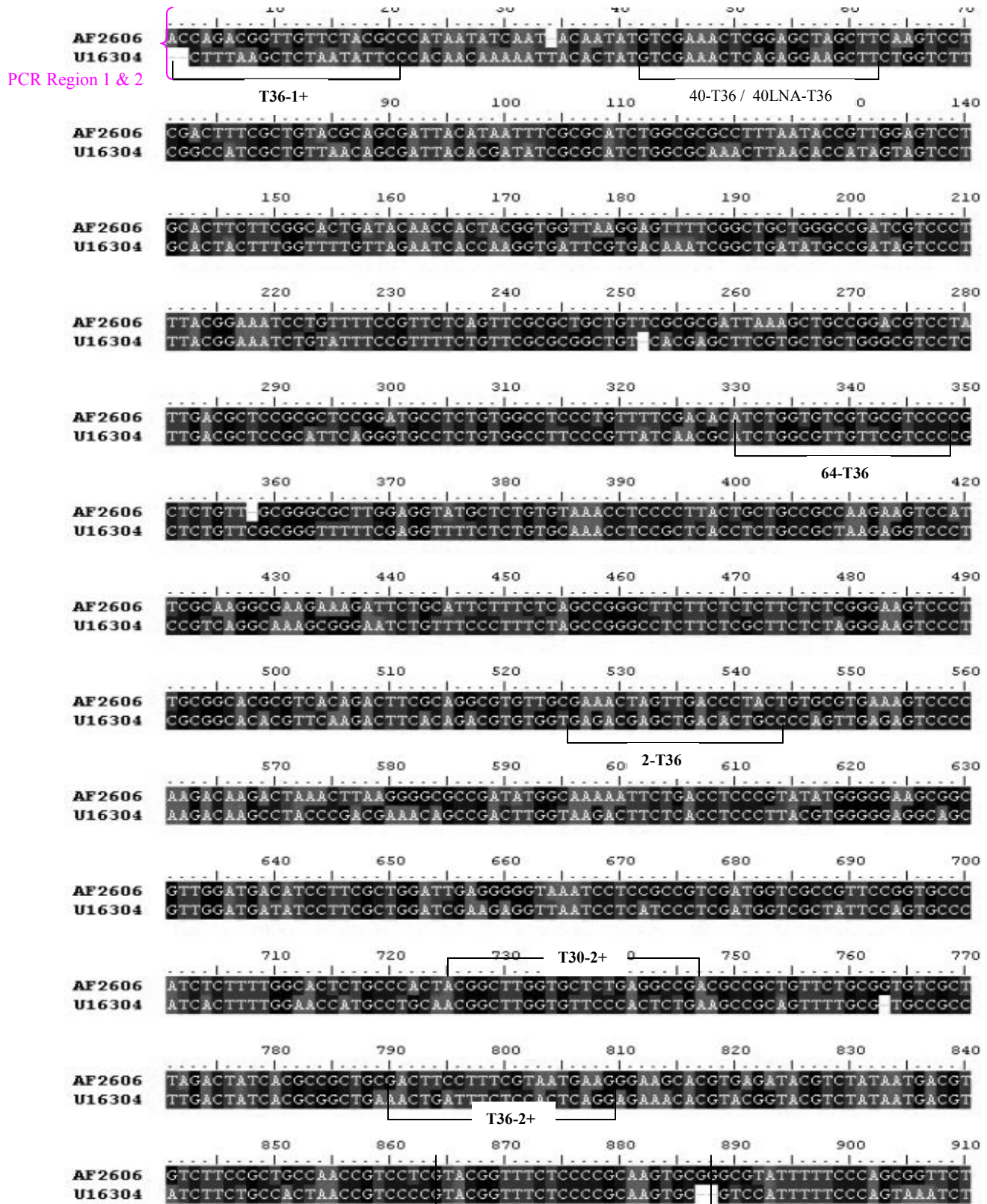
Figure 58: 1% Agarose gel electrophoresis of the T3 5+/- molecular marker PCR primer set. Expected PCR product is 397 bp. Lanes: (1) 100 bp Molecular marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T3 clone; (15) T36 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.



Figure 59: 1% Agarose gel electrophoresis of the T3 6+/- molecular marker PCR primer set. Expected PCR product is 649 bp. Lanes: (1) 100 bp Molecular Marker Hyperladder II (Bioline); (2) 12-5; (3) 12-7; (4) 12-9; (5) GFMS12; (6) GFMS35; (7) T30; (8) 389-3; (9) 389-4; (10) 390-3; (11) 390-4; (12) 390-5; (13) T30 clone; (14) T3 clone; (15) T36 clone; (16) VT clone; (17) Virus Free Control and (18) Negative control.

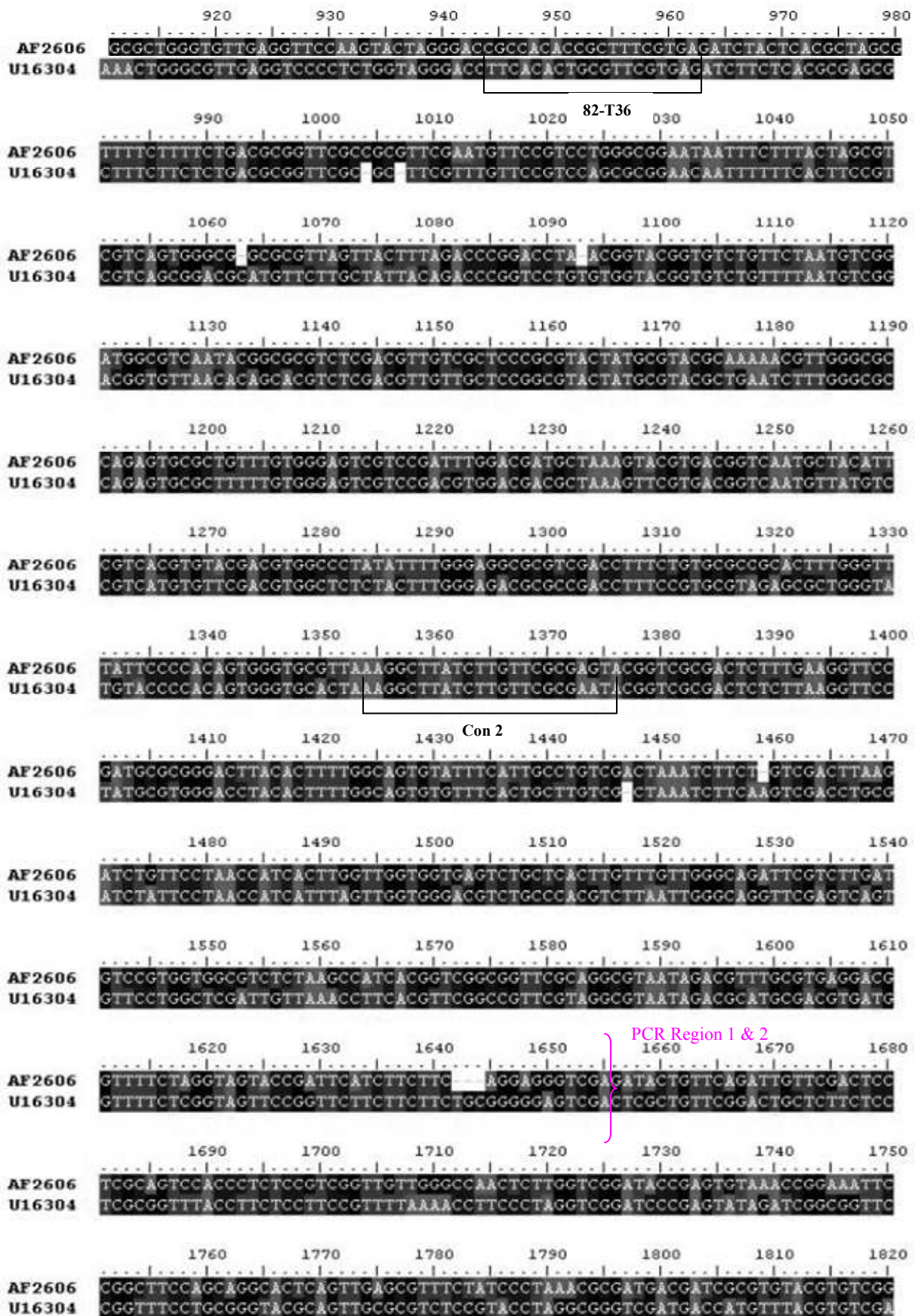


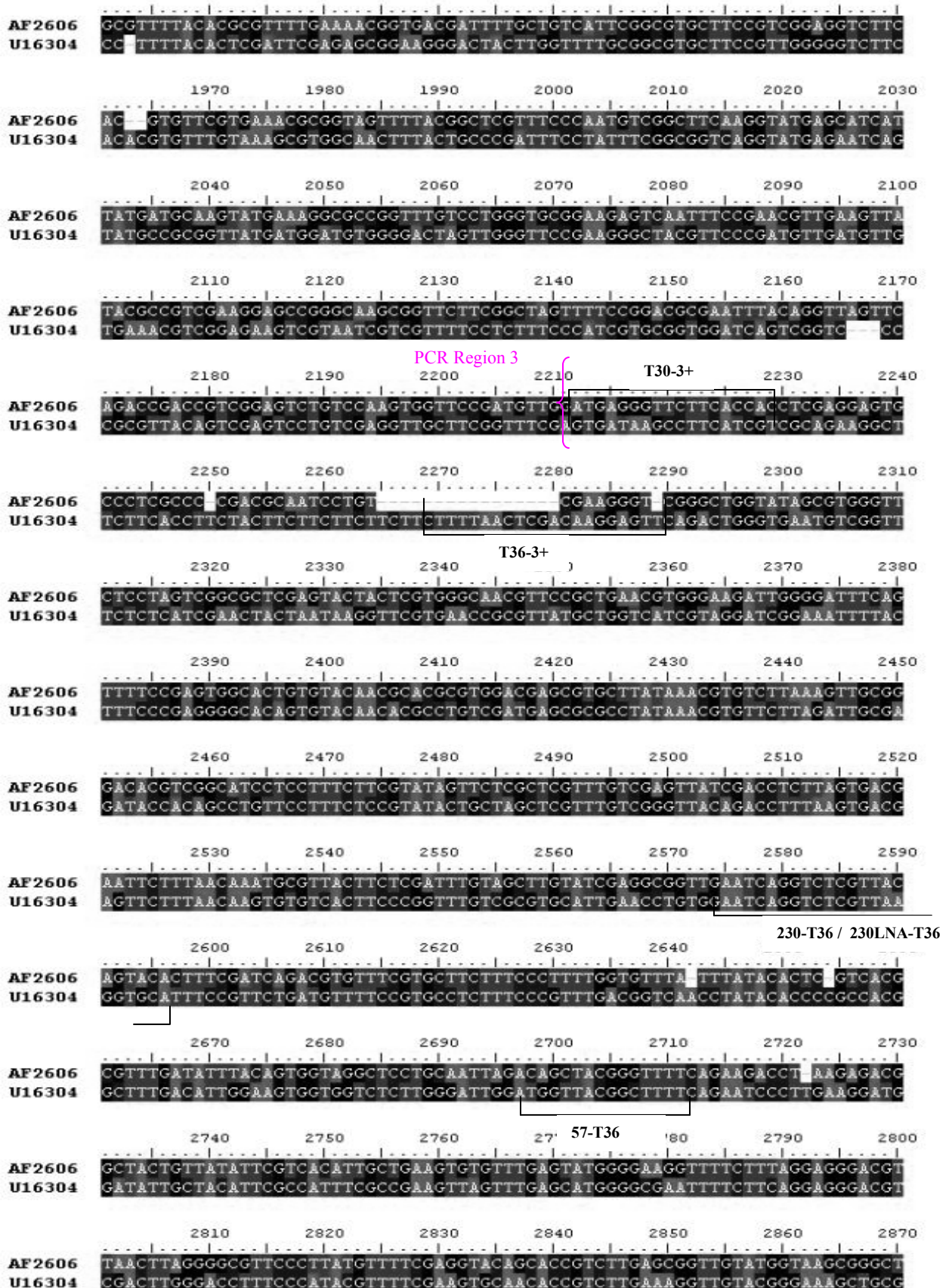
Figure 60: Pairwise alignment of T36 (U16304) and T30 (AF260651) strains of CTV from 26-8345 nt. Oligos designed for microarray use and the PCR regions targeted are annotated below.

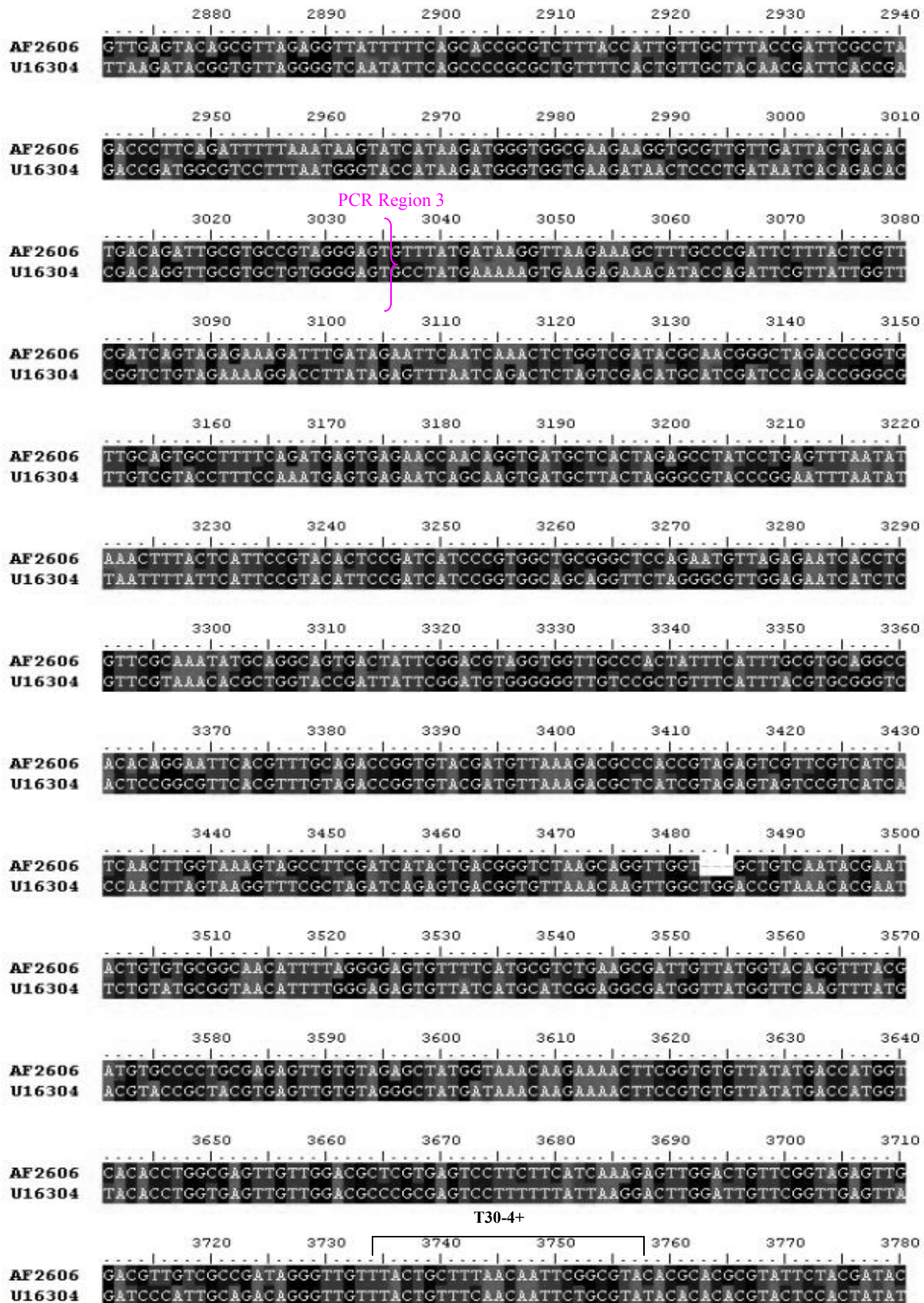


Con 1

Figure 60 continued







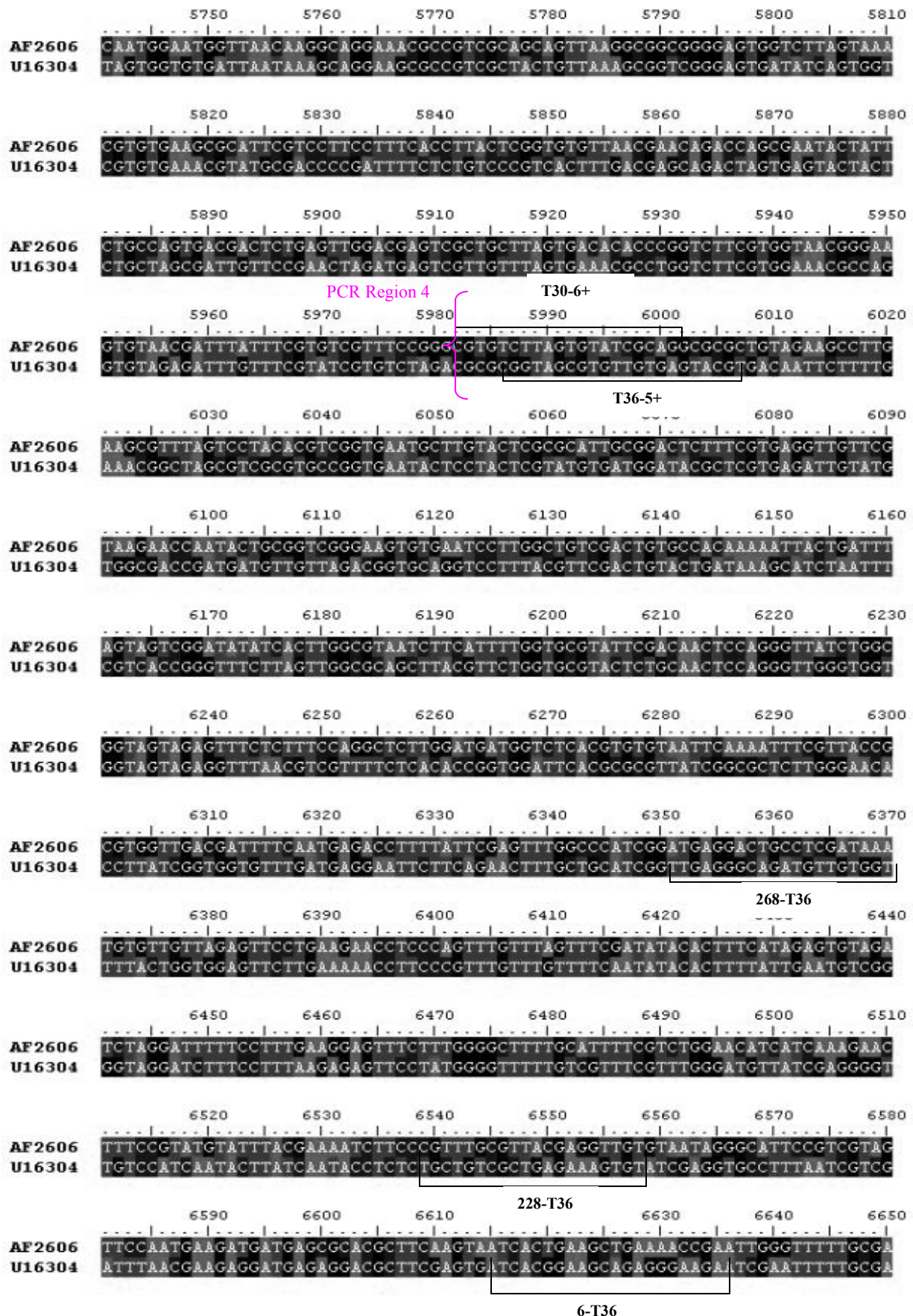
T30-4+

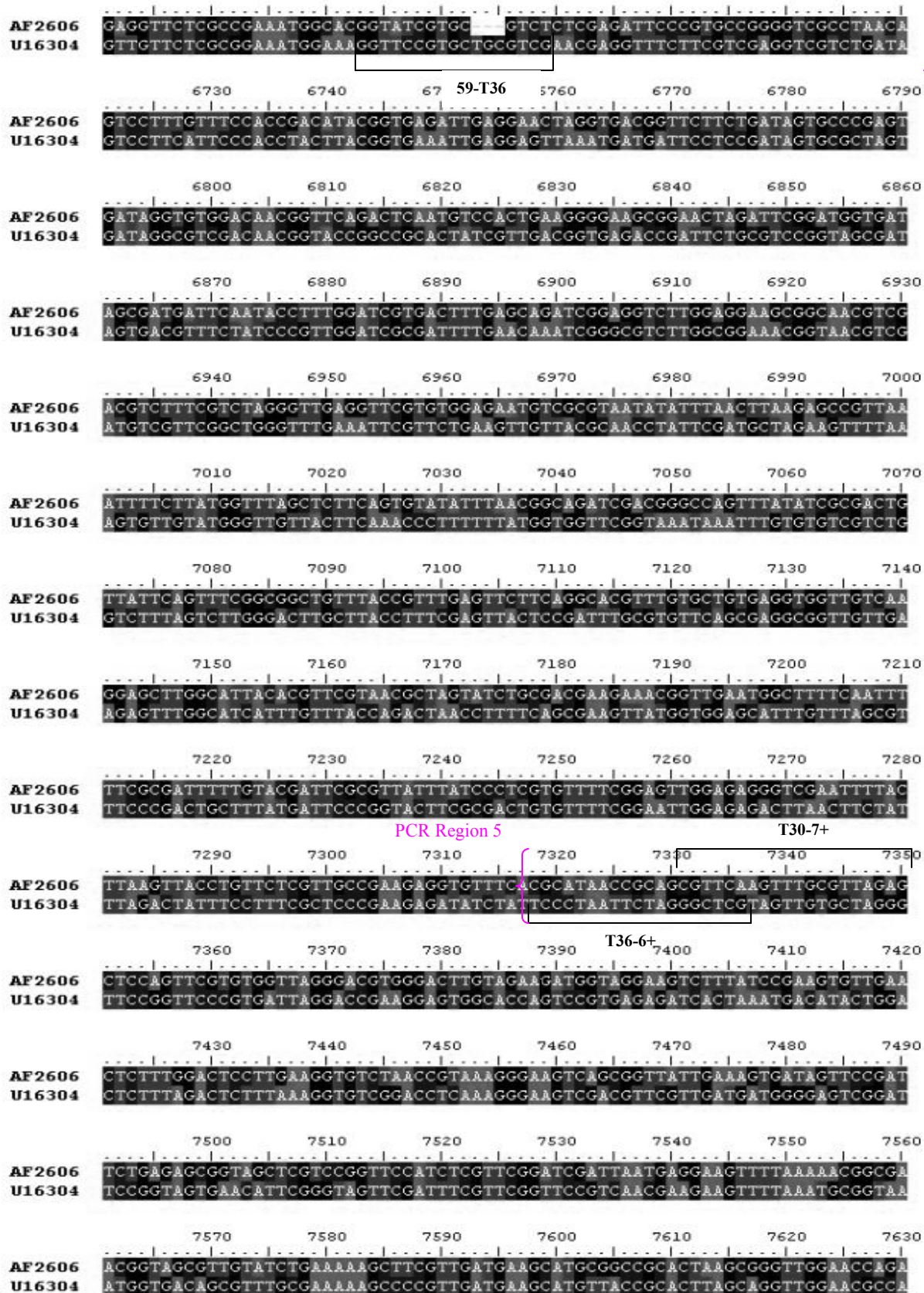
Figure 60 continued

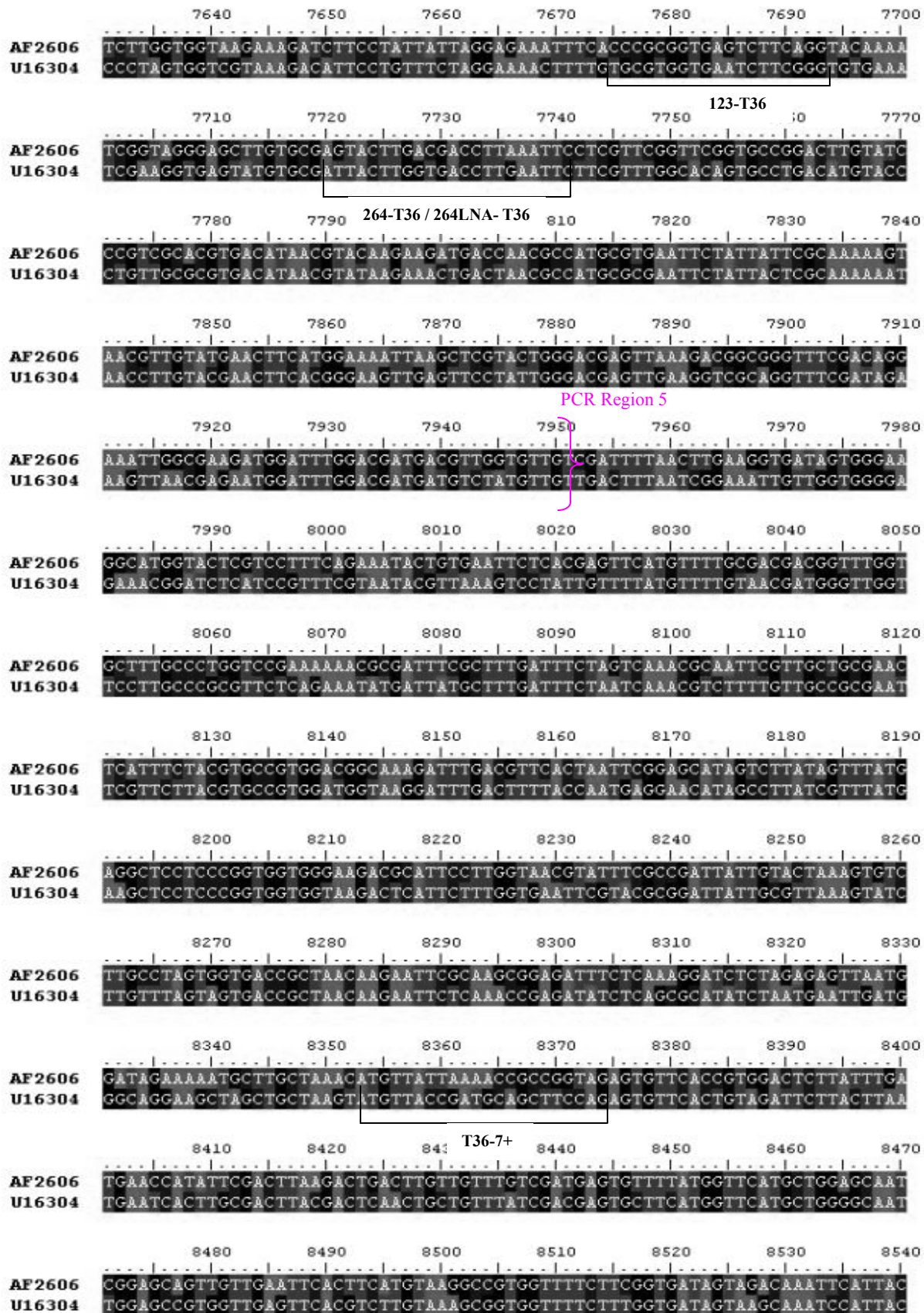
AF2606	GCGAATACATGCGTACTCCTTGTGTGGTGGATGGTTTTCTATTCACTATAGAGATGGTTTCGCTTAG
U16304	GTGAGTGTATGCCAACTCCTTGGTGTAGTGGACGGGTTTTTGTCTTACTATCGAAATGGTATCGCTCAG
	3860 3870 3880 3890 3900 3910 3920
AF2606	ATGCAGCGTTAATTATTATTGCATCACCAAAATCGCGGGTCTGTCCGAGAATTAGCGAGGTTAAACGACTA
U16304	GTGTAGCGTTAATTACTATTGTATCACCAAGTCGTCGGTCTGTCCGAGAATTAGTGAGACTAAGCGGTTG
	3930 3940 3950 3960 3970 3980 3990
AF2606	CGCTATCGTCGGTGTGATAGCGACTTGATTAGGATTAAGATTCCCTCGGTACTCCAGCAAGACTCGTAGTT
U16304	CGCTACCGTCGGTGTGATAGCGACTTGATTAGAATCAAAATTCCCGGTACTCCAGTAAGACTCGTAGTT
	4000 4010 4020 4030 4040 4050 4060
AF2606	GTTTACCAGGGGTGTTATTATTTGTACTTGGACGCTAAGTTCGTTAGTCGCGTGTACGAATACGTAGTTAA
U16304	GTTTACCTGGGTGTTACTATTTGTACTTGGACGCCAAGTTTGTAAAGTCGCGTGTACGAGTACGTGGTCAA
	4070 4080 4090 4100 4110 4120 4130
AF2606	TAATTGTGTGTGGTGAATGCTAAGACTTTTGAATGGACTTGGAACTACATCAAGTCCTGTAAGTCGAGA
U16304	TAAGTGGTGTGGTGAACGCGAAAACTTTTGAAGTGGACTTGGAAATTACATCAAACTTTGCAAGTCGAGA
	4140 4150 4160 4170 4180 4190 4200
AF2606	GTGGTCATTAGTGGGAAGGTCATCCATAGAGATGTTCCATAGCGTTGGAGTATTTGGATGGGTTTTCCG
U16304	GTGGTGATCAGCGGGAAGATTATACACCGAGACGAACCAATCGCTTTAGAGTACTTAGATGGGTTTTCTG
	4210 4220 4230 4240 4250 4260 4270
AF2606	CGGTTATGTTGTCCGCGGGCGTGAAAGGTAGGCAGAACGCAGAACGTTTTCTAGGCGTCTGGCTGCCTT
U16304	CGGTTATGCTGTCCGCTGGCGTGAAAGGTAGGCAAAACGCGGACGTTTCTAGACGCTTAGCCGCTTT
	4280 4290 4300 4310 4320 4330 4340
AF2606	TTTAGGAGATACTTCTCTTTTCGAATTGGTGAGGTTTGCCGTTTCAGAGAAAGTGTGCGGATTTATTGTG
U16304	TTCCGGGTGAAACTTTCGATTTTCGACTTAGTGGGGTTTGCTATATCCGAGAAAGTGCCGCGACTTGTTCTTG
	4350 4360 4370 4380 4390 4400 4410
AF2606	GGCATTACAGGAGGCTGTGGAGAGGTGCGTGAGATCATACCTTAGGAACTCGTTTAATATGAGTTTCGTAG
U16304	GGTATTACAGGACGCTGTGGAAAGATGCGTGAAAGTCGTACCTCAGGAATTCGTTTAATATGAGTTTCGTGG
	4420 4430 4440 4450 4460 4470 4480
AF2606	ACTTGTCCAGATCCTTTGCTTACTATTAGTGAGTACTCGGAGTTAGATGTTCCCTATCGATTTGCCGGGTTT
U16304	ACTTGTCCGACCCCTTACTTACGATTAGCGAATATTTCGAATTGGACGTTCCCTATCGACTTACCGGGTTT
	4490 4500 4510 4520 4530 4540 4550
AF2606	CGGCTGCATAACAGAAAGGGGATGAGGCTAAACTCATAGAAAGATGGGGTTAAGTCGGCTTTTGCTACGAAAA
U16304	CGGTGGAATAGTGGACGGTAACGAAACTAGACTTATGGAAGACGGTGTTAAATCGGCTTTATTACGTAAA
	4560 4570 4580 4590 4600 4610 4620
AF2606	GCGGTTTCGTGACGAGGCGCAACTTATTCTCGAAAGGGGTGGTACGTCCGAGATCGTTCCCTTGTGGGGATG
U16304	GCGGTTCAAGCCGAGGCTCAGTTGAGTCTTAAGAGAGGCAATACTTCCGAAATCGTTCCGTGTGTGACG
	4630 4640 4650 4660 4670 4680 4690
AF2606	ATGATGGTTTGGGCAGCGATGGTTCTGACAGTTCACTCGGGTCAGG TAGTGATTGTCCTCTAAAGC
U16304	ATGACGACTCGAACGGTGATGATTCCGCGTGTTCGCTCGATACGGACTGGTGTTTCGTCCTTCGGGGTGG
	4700 4710 4720 4730 4740 4750 4760
AF2606	TAGTTCTGGTGGTTTACGTGCCGGAGCCGCCCCITTCATGTTTGAAGCTTACTGAAGAAAGTTGGGCGGG
U16304	TAGTTCTGGCGGGTTGCGAGCTGGAGCCGCCCCITCTTTGTTTGGTGTCTTACTCAGGAAATTTGGGTGGG

Figure 60 continued

	4770	4780	4790	4800	4810	4820	4830
AF2606	GGTTTTTGTGTGGCCAGAAGGTTATTGGACTTGTGTGCGAGCGATTAGTTGCTTCGTTGTGCGCGCTAA						
U16304	GGTTTATGTTTGGCAAAGAAAGTGTGGAACTGTTAGTCAAGCGGGTGGTTGAATCGTTTTCTCGTTTGA						
	4840	4850	4860	4870	4880	4890	4900
AF2606	AGTTTCGGCACCACCTCATTTAACTCATTCTTTAGCATGTTACGCGCTAAGCTTTCGTGTAATTTCTTCGA						
U16304	AGCGGAAAAACCGCTCGTTTAACTCCTTCGCTAATTTGTTGCTTGCGAGGCTCTCTTGCGGGTTTTTCGA						
	4910	4920	4930	4940	4950	4960	4970
AF2606	CGACTCGATTTTTACCGCCGTTTGGGACACGGTAATGAACCTTAATTAGAATCGTTAACGGGGTGACCAAG						
U16304	TGACACTGTTTTTCTGCGGTCTGGGAAATAGCGATCAACCTGACCAGGATCGTCGATGGAGTAGTCAAG						
	4980	4990	5000	5010	5020	5030	5040
AF2606	ACGAAGTTGGAGTGTTCGCAACGTAATTCGGTTAATACCTAAGTTGTGTTCTGGTTGTAAAGACTTTG						
U16304	ACAAAACTGGGGTGCTTTTCGTAATTTATTTTCGATTGATTCCCAAGTTATACGCTGGTTGTAAAGGACGTTG						
	5050	5060	5070	5080	5090	5100	5110
AF2606	CGAAGGACTTTGCGGACATGTGCAAGTCAGTTTTGCTATTTTCTACGGAATTTCTCAGTAAAAGCTATAC						
U16304	TGAAGGACGTTAGTAACGCTTACAGGTTCGGTTGTCTCGTTTTCCGCTGAGGTTTTGGGTAAGAGTTACTC						
	5120	5130	5140	5150	5160	5170	5180
AF2606	TAAACTGTGTCGGTTTTAGTTTGGAAATATTTCGATGTTGGACTTATTGAAATTTTCAGCCCCGTAATATGGGA						
U16304	TAAACTTTGTCAGTTTTGCTTGGAAATATTTGACATCGGGCTTATTGAGGTTTCTGCTCGTAATTTGGGA						
	5190	5200	5210	5220	5230	5240	5250
AF2606	CTGGAGACTTGTGTTTCGATGGCTTCAGCTATTTTTGATGTCATACTTACCTATCCTCTGGGAAGTATTA						
U16304	CTGGAACATGCGTTTCAATAGCACTAGCTATTTTTGATGTTGATTACTTTTCCTTTGGGGAGCGTTA						
	5260	5270	5280	5290	5300	5310	5320
AF2606	GCTTTCCCGTCGTTTGTGCGAGAGTTTTTGGCGCCCTCGCTATGGAAGTGGGATCGAACTATTGCTTTAA						
U16304	GTTTTCCCGTCGTTTGTGCGCTAGGGTTTTTCGGTCGCTCTCGCTATGGAAGTGGGATCGAACTTTTTGTTAA						
	5330	5340	5350	5360	5370	5380	5390
AF2606	GAAAACTCTCGGCCAAACCCGAAATTTTAGGAGAGGATGCTTTTCGTAGAACCGTAAGTATGGTGTGAGC						
U16304	GAAAGTTTTCCGTAACCCGAAACTCTTGGTGAAGATACTTTCCGTAGGACTGTGAGTATGGTTTTGAGC						
	5400	5410	5420	5430	5440	5450	5460
AF2606	GCTAAGGCTTTTCGACAAACTTTTCAGTCGATGCTAAGGGTTTAGTTAGATGTTCCGGTATTATTCTCGAG						
U16304	CCCAAGCGTTTGACAAGCTTTCCGTGGATGCGAAAGGTTTAGTTAGGTGCTCTGGTGTTATTCTCGAG						
	5470	5480	5490	5500	5510	5520	5530
AF2606	TATTGAGGATAGTGTTTTATCTCTGTTTTCCGATGCTGACGCATGGGTGGGTATAGCAAGCACGAAGT						
U16304	TGTTAAGGATAATACTATCTTCTGTTTTCCGAAGATGACGCTTGGGTGGGTATAGTAAACATGAAGT						
	5540	5550	5560	5570	5580	5590	5600
AF2606	TTCTCTACTTCCCATGATCAGTATAGTGCGAAAAAGTTATCTTTTCGTGTTGGACGAGACTAAGGAAGCG						
U16304	TTCTTCCCTTCTATTGTTACCTTTTGTGTGAAGAGATTGCTTATTCTCTGGAGCGGTGTAAGGAAGCG						
	5610	5620	5630	5640	5650	5660	5670
AF2606	TGCATCGCTTTTCATAGTCGATTTCGTCTAAGAGAATCTTAGCGTCTTTGACCTCGGAGATTACAGAGTCTT						
U16304	TGTGTGCAATTTATTGTTGACTCTGCTAAGAAAGATTTTGAAATCTTTAACTTCAGAGTTCAACGAATCCA						
	5680	5690	5700	5710	5720	5730	5740
AF2606	TGCAAACTCGGAACCTTACTAGTGGTGCTAAATCCTTGTATGGACGTGTGAGTTCTATACTCTCAGCATC						
U16304	TTCAAACTCCGACTTGACCAAGTGGTGCTAAGTCCTTGTACGGGCACGTGAGCTCGATACCTTTACACCTC						







Characterization and strain differentiation of the p23 gene of South African CTV isolates using a bi-directional RT-PCR system and Phylogenetic analysis.

Stewart, K.A.¹, Pietersen, G.².

¹ Room 9-11, Department of Microbiology, New Agricultural building, University of Pretoria, Lunnon Road, Hatfield, South Africa, 0002

² Room 9-8, Department of Microbiology, New Agricultural building, University of Pretoria/CRI, Lunnon Road, Hatfield, South Africa, 0002

Corresponding author: K.A.Stewart
Email address: katherines@cri.co.za

Summary:

A bi-directional PCR was developed by Sambade *et al.* (2003) that differentiated CTV strains into mild, severe and atypical groups based on differences at amino acid residues 78-80 of the p23 gene. Here we describe the establishment of this method to fully characterise the p23 gene of isolates selected for possible mild-strain cross-protection in South Africa and phylogenetically compare these RSA isolates and reference isolates biologically indexed and sequenced of Sambade *et al.* (2003) from around the world. RSA isolates were amplified with the strain differentiating bi-directional RT-PCRs. The predicted amino acid sequences were compared for areas of possible variability for further strain differentiation. The PCR results and phylogenetic groupings showed that RSA isolates 390-3 and 390-5 were atypical; 390-4, 389-4 and 389-3 were mild; 12-5 was severe; GFMS 35 had mixed mild and atypical strains; and GFMS12, 12-7 and 12-9 had mixed mild and severe strains. A Phylogenetic tree showed that there were three distinct groups of isolates corresponding to previously identified mild, severe and atypical symptom isolates of Sambade *et al.* (2003). Isolates in the atypical group were more diverse than ones in the mild or severe groups. There were 53 polymorphic amino acid sites of p23 gene of the RSA and reference isolates, with 4 distinct regions of variability. However the amino acid region 78-80 was confirmed as being very useful in grouping these isolates as mild, severe or atypical. This PCR system was dynamic, reproducible and capable of detecting mixed infections, it should therefore be considered in the South African Citrus industry as a screening tool in molecularly selecting mild strains for cross-protection.

INTRODUCTION

Citrus tristeza virus (CTV) has a single-stranded RNA genome and is a member of the genus *Closterovirus*, family *Closteroviridae*. CTV is the causal agent of an economically devastating disease of citrus and is spread by aphid vectors and infected budwood. There are many strains of CTV causing different symptoms yet how these strains; vectors; host plants and the environment interact is still unknown and adds complexity to managing the disease and detecting dangerous strains. Essentially there are a few typical symptom patterns known: (1) symptomless CTV isolates (Albertini *et al.*, 1988); (2) mild inducing vein clearing; (3) decline and death of most citrus species propagated on sour orange; (4) stem-pitting, stunting, poor quality fruit and reduced yield; and (5) stunting and leaf yellowing (known as seedling yellows).

Current methods of strain differentiation include bio-indexing, serological and PCR based methods. The 3' half of the genome is relatively conserved with 90 % sequence identity, compared to the 5' half with less than 70 % sequence identity (Ayllón *et al.*, 2001). Strategies to differentiate strains in the 3' half of the genome aim to target possible pathogenicity functional differences. The pathogenicity determinants of CTV are currently unknown however transgenic Mexican lime over-expressing p23, display typical CTV symptoms which is associated with p23 accumulation (Ghorbel *et al.*, 2001). The conserved p23 gene is an RNA-binding protein involved in regulating the asymmetrical accumulation of viral RNA strands (Satyanarayana *et al.*, 2002) and a predicted ribosome binding capacity (Dolja *et al.*, 1994); is a suppressor of post-transcriptional gene silencing (Lu *et al.*, 2003); is suspected to play a regulatory role in the expression of other CTV genes (López *et al.*, 2000); and may serve as an indicator of disease severity (López *et al.*, 2000). This has significance in respect to finding markers for strain differentiation. Its corresponding sub-genomic RNA (p23-sgRNA) is the second most abundant in infected plants (Hilf *et al.*, 1995) and the protein accumulates very early in cell infection (Navas-Castillo *et al.*, 1997) down-regulating negative-stranded RNA accumulation which indirectly increases expression of the 3' genes (Ghorbel *et al.*, 2001).

The sequence analysis of the deduced amino acid sequence of the p23 gene from several CTV isolates of different biological properties and geographical origins have shown a remarkable conservation of both the basic region; and the cysteine and histidine residues in the core of the proposed zinc-finger domain, suggesting their functional importance (Pappu *et al.*, 1997). A cluster dendrogram correlated the biological properties of the isolates forming distinct groups of mild, quick decline or stem-pitting isolates (Pappu *et al.*, 1997) suggested a possible role for p23 in pathogenicity. All mild and stem-pitting isolates formed distinctive groups (Pappu *et al.*, 1997). Associated relationships between sequence data and biological properties were also found with the CP (Pappu *et al.*, 1993) and p27 (Febres *et al.*, 1995) proteins. The involvement of p23 in pathogenicity of a particular isolate may occur by p23 acting in concert with or by regulating the expression of the other disease severity determinants (Pappu *et al.*, 1997). Sambade *et al.* (2003) reported that polymorphisms located in the amino acid region 78-80 of the p23 gene allowed discrimination between mild and severe isolates.

18 CTV isolates of different geographic origin and pathogenicity characteristics were phylogenetically classified into 3 groups: 1) mild (vein clearing), 2) severe causing stem pitting and 3) an atypical group showing variable symptoms.

In this study a bi-directional PCR, sequencing and phylogenetic analysis of the p23 gene differentiating mild, severe and atypical isolates developed by Sambade *et al.* (2003) was used to test 10 South African CTV infected sources and compared to reference sequences of Sambade *et al.* (2003).

MATERIALS & METHODS

CTV isolates: 10 South African CTV sources 12-5, 12-7, 12-9, 389-3, 389-4, 390-3, 390-4 and 390-5 were tested (kindly supplied by S.P. van Vuuren, South Africa) from single aphid transmissions. Mild strain sources GFMS 12 and 35 used for cross-protection were included. Sub-isolates 12-5, 12-7 and 12-9 were made from the mild-strain cross-protecting source GFMS 12. These isolates were previously biologically indexed as mild. The clones T36, T305, and T385 were used as positive controls (kindly supplied by Pedro Moreno, Spain). The atypical group control T36 isolate induced variable seedling yellows reactions on sour orange or grapefruit. The severe group control T305 isolate induced seedling yellows on Mexican lime, sweet orange/sour orange and also variable stem-pitting on sweet orange. The mild group control T385 isolate belongs to a collection kept at the Instituto Valenciano de Investigaciones Agrarias, Valencia, Spain. Virus free citrus plants served as a healthy control.

Reference isolates characterised by Sambade *et al.* (2003) for use in phylogenetic comparisons were divided into 3 groups based on symptoms and sequence data. Isolates T32, T55, T300, T312, T385 and T346 have mild to moderate symptoms; T305, Barao B, C269-6, Cald-CB, T388 and VT have severe symptoms of seedling yellows and stem-pitting; and K, C270-3, Galego-50, C268-2, T36 have atypical symptoms ranging from stem-pitting, seedling yellows, decline and asymptomatic.

RNA extraction and cDNA synthesis: Total RNA was extracted from Mexican lime plants infected with CTV sources using the SV Total RNA Isolation System (Promega, USA) according to the manufacturer's protocol. The conserved reverse primer PM51 (IDT, USA) (Sambade *et al.*, 2003) was used to prime cDNA synthesis and is specific for

the p23 gene. Reverse transcription was performed as follows: 3 µg of total RNA extract was heated at 65 °C for 15 minutes, 55 °C for 10 minutes and held at room temperature for 5 minutes. The reaction mixture was added to a volume of 25 µl containing 50 pmol reverse primer PM51, 5U of RNAsin (Promega, USA), 10 U of AMV reverse transcriptase (Roche, Germany), 1x AMV RT-Buffer (50 mM Tris-HCl, 8mM MgCl₂, 30 mM KCl, 1 mM DTT; pH 8.5) (Roche, Germany) , 0.2 mM each of dATP, dCTP, dGTP, and dTTP. Incubated at 47 °C for 1 hour on an Eppendorf Gradient Mastercycler (Eppendorf, Germany), after which 12 µl of nuclease-free water was added.

Conserved PCR, sequencing and phylogenetic analysis: The PCR reaction contained 5 µl of cDNA and was performed according to Sambade et al. (2003) in an Eppendorf Gradient Mastercycler (Eppendorf, Germany). PCR products (5-25 µl) were analyzed by gel electrophoresis through 1 % agarose (Whitehead Scientific) in Sodium boric acid electrophoresis buffer (5 mM disodium borate decahydrate, adjusted to pH 8.5 with boric acid), stained with ethidium bromide (0.5 µg/ml), and photographed under UV light. The nucleotide sequences of the complete p23 gene were determined in both directions using primers PM 50 and 51 and an ABI PRISM DNA 377 sequencer (Perkin-Elmer). The following programs were used: Vector-NTI (Invitrogen, USA) for overlapping and compiling complete sequences; BLAST for seeking homologous sequences in databases; ClustalW for sequence alignment (Thompson *et al.*, 1994); DNAMAN for predicted amino acid sequences; EMBL www Gateway for isoelectric point determination; MEGA 2 (Kumar *et al.*, 1993) for estimating nucleotide distances using Jukes and Cantor method and for phylogenetic, bootstrap analyses using a Neighbourhood joining tree and intra/inter-group nucleotide distances.

Bi-directional PCRs: Primers PM85 and PM86 designed by Sambade *et al.* (2003) based on two p23 regions conserved in most CTV isolates, whereas primers PM82 (VT isolate, severe group), PM83 (T385 isolate, mild group) and PM84 (T36 isolate, atypical group) were based on a short internal region which allows discrimination into the three groups (mild, severe and atypical). Sources were tested using primers PM85, 86, 82 and 83 and with primers PM85,86, 82 and 84.

RESULTS

Classification of isolates by Bi-directional PCR

Primers PM50 and 51 (Sambade *et al.*, 2003) were used as a conserved CTV first step screening PCR and to amplify sequences. All RSA isolates yielded the expected 697 bp DNA fragment except isolate 390-4. The bi-directional PCRs were created using two external primers (PM 85 and 86), and two group-specific internal primers of opposite polarity (PM82 and 83, or PM82 and 84) (Table 1). PCR amplification with primers PM85/86 and PM82/86 yielded DNA fragments of 612 bp and 450 bp respectively, whereas amplification with primers PM85/83 or PM85/84 yielded a fragment of 239 bp. Therefore, RT-PCR amplification with primers PM85, 82, 83 and 86 permits detection of sequences characteristic of the mild (products of 612 and 239 bp), the severe (products of 612 and 450 bp) or both groups (products 612, 239 and 450 bp). Likewise, PCR amplification with primers PM85, 82, 84 and 86 permitted detection of sequences characteristic of the severe (products of 612 and 450 bp), the atypical (products of 612 and 239 bp), or both groups (products 612, 239 and 450 bp). In both bi-directional PCRs the following were used as group controls: T36 (atypical), T305 (severe) and T385 (mild).

PCR amplification of the 10 RSA isolates using primers PM85, 82, 83 and 86 to differentiate between mild and severe isolates (Figure 4 and Table 3) yielded: (i) isolates GFMS 35, 389-4, and the mild group control T385 clone had 612 and 239 bp DNA amplicons; (ii) the severe group DNA control clone T305 had 612 and 450 bp DNA amplicons and; (iii) RSA isolates 390-3, 390-5 and the atypical group control T36 had a 612 bp DNA fragment. Isolates that did not group with any group classification system were: (iv) isolates GFMS12, 390-4 and 389-3 which amplified one of the two mild group fragments of 230 bp and not the 612 bp fragment; (v) isolates 12-7 and 12-9 amplified both the 450 and 239 bp fragments of the mild and severe groups but not the conserved 612 bp fragment and; (vi) isolate 12-5 amplified one of the two expected fragments of the severe group of 450 bp but not the 612 bp fragment.

PCR amplification using PM85, 82, 84 and 86 to differentiate between atypical and severe isolates generated: (i) DNA fragments of 612 and 239 bp with isolates 390-3, 390-5 and the T36 atypical group control; (ii) isolate 389-4 and T385 mild group control had a 612 bp DNA amplicon; (iii) the T305 severe group control had the 612 and

450 bp fragments as expected. Isolates that did not group with the expected group classification system were: (iv) isolates GFMS 12, 12-5, 12-7, 12-9 amplified one of the two expected fragments of the severe group of 450 bp but not the 612 bp fragment; (v) isolate GFMS 35 only produced one of the two atypical fragments of 239 bp and not the 612 bp fragment and (vi) isolates 389-3 and 390-4 had no amplification. There were no mixed groupings in this PCR. In both bi-directional PCRs the virus free plant and the buffer control showed no amplification. The gel photographs depicting these amplifications and the summarized results are shown in figures 4-5 and table 1 respectively.

Predicted amino acid sequence analysis

Figure 1 depicts the multiple sequence alignment of predicted amino acid sequences of the complete p23 gene of the 8 RSA isolates and 18 reference sequences from Sambade *et al.* (2003) of different geographic origins. There were 53 polymorphic amino acid positions (approximately 25 % of the complete gene). There was high variability at positions 24-29, 50-54 and 78-80, as was also found by Sambade *et al.* (2003) and at position 125-129. The summarised results of these positions are depicted in Table 3.

RSA isolates 12-9, 12-7, 12-5 and GFMS 12 had Glu²⁴, Lys²⁶ and Ser²⁹ as with previously characterized severe isolates (VT, Cal-CB, T388, c269-6, Barao and T305). However atypical isolates K, T36, and Galego-50 also had this composition. Most reference mild isolates (T385, T32, T55, T300 and T312) had Lys²⁴, Glu²⁶ and Lys²⁹. RSA isolates GFMS 35, 389-3 and 389-4 had Arg²⁴, Glu²⁶ and Lys²⁹, as with the mild reference isolate T346. RSA isolate 390-5 had Ala²⁴, Arg²⁶ and Ser²⁹, as with other atypical reference isolates 270-3 and 268-2.

In the region 50-54 RSA isolates 390-5, GFMS 35, 389-3 and 389-4 had Val⁵⁰, Thr⁵³ and Asn⁵⁴, as with mild group reference isolates (T346, T385, T32, T55, T300 and T312) and an atypical reference isolate c270-3. Other atypical reference isolates did not fall into any defined amino acid pattern and included: Galego-50 (Ile⁵⁰, Tyr⁵³ and Asn⁵⁴);

Table 1:

A graphical representation of the amplified products for each isolate, using the primers PM 85, 86, 82 and 84 (severe versus atypical); and primers PM 85, 86, 82 and 83 (severe versus mild). The mild (T385), severe (T305) and atypical (T36) group controls are also included.

Isolate	PM 85, 86, 82, 83 (Severe vs. Mild)			PM 85, 86, 82, 84 (Severe vs. Atypical)		
	612 bp	450 bp	239 bp	612 bp	450 bp	239 bp
T30						
GFMS35						
GFMS12						
390-3						
390-4						
390-5						
389-3						
389-4						
12-5						
12-7						
12-9						
+ Control: T36 clone						
+ Control: T385 clone						
+ Control: T305 clone						
- Control: Virus free						
- Control: Buffer						

- + represents a positive control
- represents a negative control

isolate K (Ile⁵⁰, Ala⁵³ and Ser⁵⁴); and isolate c268-2 (Val⁵⁰, Thr⁵³ and Ser⁵⁴). RSA isolates GFMS12, 12-9, 12-7, 12-5 and all the severe reference sequences (Cal-CB, T305; T388, VT, c269-6 and Barao) had Ile⁵⁰, Asn⁵³ and Ser⁵⁴, as well as the atypical reference isolate T36.

In the region 78-80 isolates GFMS 35, 389-3 and 389-4 had Ala⁷⁸, Leu⁷⁹ and Lys⁸⁰, as described as the mild group defining amino acid region for mild isolates (T346, T385, T32, T55, T300 and T312) (Sambade *et al.*, 2003). Isolates GFMS 12, 12-5, 12-7 and 12-9 had Ala⁷⁸, Ser⁷⁹ and Arg⁸⁰, as described as the severe group defining amino acid region for the reference isolates (T305, Barao, Cal-CB, T388, VT and c269-6) (Sambade *et al.*, 2003). Isolate 390-5 had Gly⁷⁸, Leu⁷⁹ and Lys⁸⁰, as described as the atypical group defining amino acid region for the atypical reference isolates (T36, c270-3, K and c268-2) (Sambade *et al.*, 2003). The only exception was Galego-50, an atypical reference isolate with Gly⁷⁸, Leu⁷⁹ and Arg⁸⁰.

Region 125-129 showed high variability amongst isolates. Isolates GFMS 35, 389-3 and 389-4 had Glu¹²⁵, Leu¹²⁸ and Tyr¹²⁹; the same as the mild reference isolate T385. Many mild (T346, T55, T312 and T300) and atypical (270-3, Galego-50, T36, and c268-2) reference isolates had Asp¹²⁵, Met¹²⁸ and Tyr¹²⁹ as well as the RSA isolate 390-5. The only exception was a severe isolate, T305 which had this composition. Isolate T32 from the mild group had a unique composition of Asp¹²⁵, Ser¹²⁸ and Tyr¹²⁹. RSA isolates GFMS 12, 12-5, 12-7 and 12-9 had Asp¹²⁵, Met¹²⁸ and His¹²⁹, the same as the severe reference isolates (VT, 269-6, Barao, Cal CB and T388). Interestingly isolate K from the atypical group also had this composition.

Figure 1:

Multiple Alignment using ClustalW of predicted amino acid sequences of the p23 protein. Regions described in the text are boxed. The basic residues and zinc finger motif residues of the RNA binding domain are indicated in bold with boxes. For each position, identical residues are indicated by dots. Locations of the primers used in the PCR amplifications are indicated.

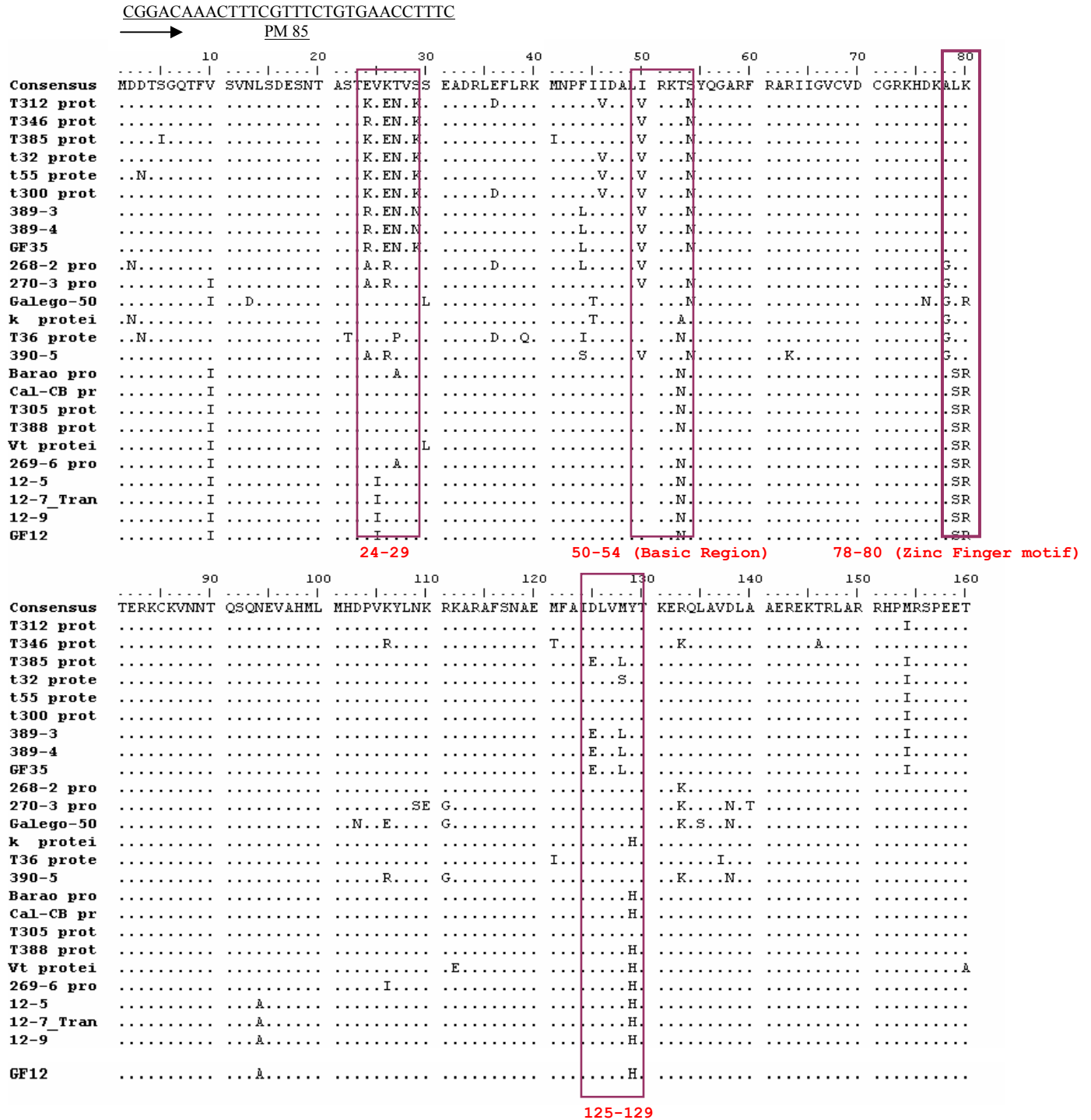


Figure 1 continued

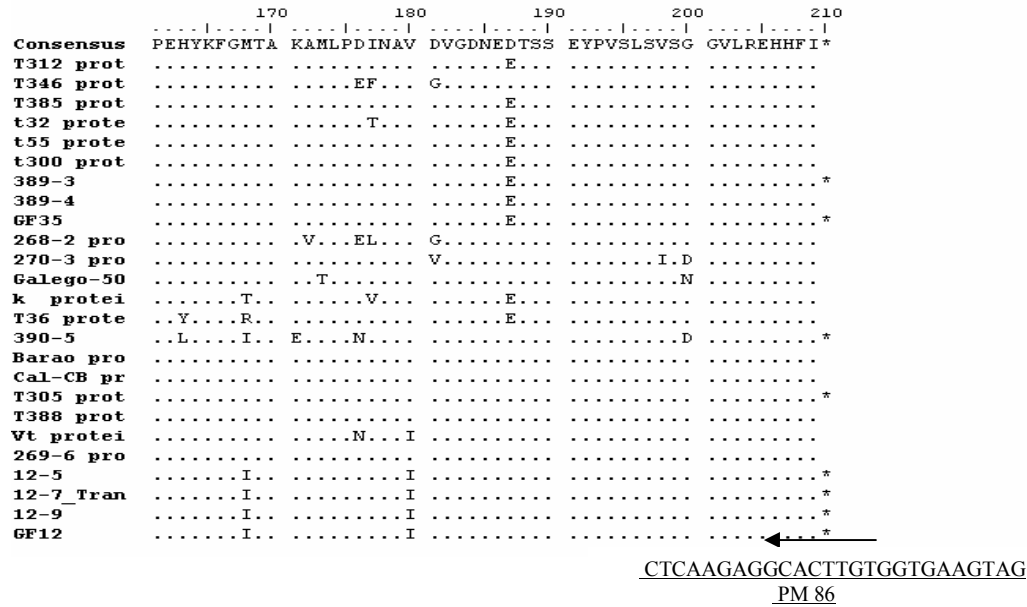


Table 2:

Intra-group and inter-group genetic diversity values estimated for mild, severe and atypical CTV isolates. Nucleotide distances were calculated with the MEGA program.

	Group 3 (Mild)	Group 1 (Severe)	Group 2 (Atypical)
Group 3 (Mild)	0.043 (4.3 %)	0.094 (9.4 %)	0.108 (10.8 %)
Group 1 (Severe)		0.019 (1.9 %)	0.094 (9.4 %)
Group 2 (Atypical)			0.106 (10.6 %)

Table 3:

Difference in Amino acid residues of 4 different regions of variability of the reference and RSA (shown in bold) isolates.

Isolate	Region 24,26,29	Region 50,53,54	Region 78,79,80	Region 125, 128, 129
Mild				
T346	REK	VTN	ALK	DMY
T385	KEK	VTN	ALK	ELY
T32	KEK	VTN	ALK	DSY
T55	KEK	VTN	ALK	DMY
T300	KEK	VTN	ALK	DMY
T312	KEK	VTN	ALK	DMY
GFMS35	REK	VTN	ALK	ELY
389-3	REK	VTN	ALK	ELY
389-4	REK	VTN	ALK	ELY
Atypical				
C270-3	ARS	VTN	GLK	DMY
Galego-50	EKS	ITN	GLR	DMY
K	EKS	IAS	GLK	DMH
T36	EKS	INS	GLK	DMY
C268-2	ARS	VTN	GLK	DMY
390-5	ARS	VTN	GLK	DMY
Severe				
Cal-CB	EKS	INS	ASR	DMH
T305	EKS	INS	ASR	DMY
T388	EKS	INS	ASR	DMH
VT	EKS	ITS	ASR	DMH
C269-6	EKS	INS	ASR	DMH
Barao	EKS	INS	ASR	DMH
GFMS12	EKS	INS	ASR	DMH
12-5	EKS	INS	ASR	DMH
12-7	EKS	INS	ASR	DMH
12-9	EKS	INS	ASR	DMH

Phylogenetic Analysis

The 697 bp DNA products were sequenced for each RSA isolate and trimmed to contain 627 bp of the complete p23 gene (209 amino acids). The p23 gene DNA sequence of isolate 12-7 was 100 % homologous to isolate 12-5. The p23 gene DNA sequences of RSA isolates 12-9 and GFMS 12 had a 96-97 % nucleotide similarity to the cognate region of many defective RNA sequences (AJ579773; U35120; AY206452). RSA isolates 389-3, 389-4 and GFMS 35 had a 98 % nucleotide similarity to the cognate regions of isolates 464-2 (AY995566); 464-1 (AY995565); 425 (AY995564); and 420-1 (AY995563) of unpublished work (Roy) of single aphid and graft transmissions of a naturally infected tree in California, USA, where no biological indexing records are available. The DNA sequence of the p23 gene of RSA isolates GFMS 35, 389-4 and 389-3 were very similar. The sequence of RSA isolate 390-5 had a 98 % nucleotide similarity to clones of isolate C315-14 (AY962359) originating from a population of strains of a grapefruit isolate selected for pre-immunisation assays in Argentina in 2005 (Iglesias, unpublished data).

The phylogenetic tree (figure 2) showed two groups of isolates that appeared clearly separated: group 1 & 3. Group 2 did not appear to be clearly separated from the other groups. Reference isolates within the first group (T305, T388, Cal-CB, Barao, 269-6 and VT) were previously characterized as inducing severe inducing seedling yellows and stem-pitting (Sambade *et al.*, 2003). RSA isolates 12-5, 12-7, 12-9 and GFMS 12 grouped with these severe isolates. Reference isolates in the second group (K, 270-3, T36, Galego-50 and 268-2) showed variable pathogenicity characteristics (Sambade *et al.*, 2003) including the symptomless isolate K and others inducing seedling yellows or even stem-pitting in grapefruit (c270-3) (Sambade *et al.*, 2003). Within the severe group, RSA isolates 12-9 and GFMS 12 clustered very closely to the severe VT isolate. RSA isolates 12-5 and 12-7 which are 100 % identical clustered closely to severe isolates T388 and T305. The group 1 sequences were distinctly separated into 2 subgroups consisting of: (A) 12-9, GFMS12 and VT; and (B) 12-5, 12-7, T388, T305, Cal-CB, Barao, and 269-6.

Isolate T36 and K were more distantly related to the other atypical isolates. RSA isolate 390-5 grouped together with the atypical isolates (Sambade *et al.*, 2003). Reference isolates in the third group (T346, T55, T32, T385, T300, and T312) showed only mild to

moderate symptoms (Sambade *et al.*, 2003) in Mexican lime and sweet orange. RSA isolates 389-3, 389-4 and GFMS35 grouped together with these mild reference isolates. Within the atypical clade (group 2) the isolate 390-5 clustered most closely with c268-2 and together with c270-3 and Galego-50 formed a separate subgroup. Isolates T36 and K formed their own branches distinctly separate from the other atypical isolates.

Within the mild clade (group 3) of isolates there were three distinct sub-branches. RSA isolates GFMS35, 389-3 and 389-4 clustered together in a subgroup on their own and isolates T312, T300, T32, T385 and T55 formed another subgroup. These two subgroups were more closely related to each other than the mild isolate T346 which formed its own subgroup.

Table 2 represents the intra-group genetic diversity (average nucleotide distance between two isolates of the same group selected randomly) and the inter-group genetic diversity (average nucleotide distance between two isolates, one of each group, selected randomly). A score of zero would indicate a 100% homology. The intra-group diversity of groups 1 and 3 were two to five folds (0.019 and 0.043 respectively) smaller than their inter-group diversity (0.094). Therefore the isolates of group 1 are similar to each other, as with group 3 isolates. The isolates of group 1 and group 3 are distinctively different from each other.

In contrast, intra-group and inter-group values of group 2 compared with groups 1 and 3 were similarly high, in agreement with the topology of the phylogenetic tree. This indicates that isolates are more distantly related within group 2 (0.106) and to groups 1 (0.094) and 3 (0.108). The inter-group diversities of groups 1 and 2 and of groups 1 and 3 were equally high (0.094), however it was higher between the group 2 and 3 (0.108).

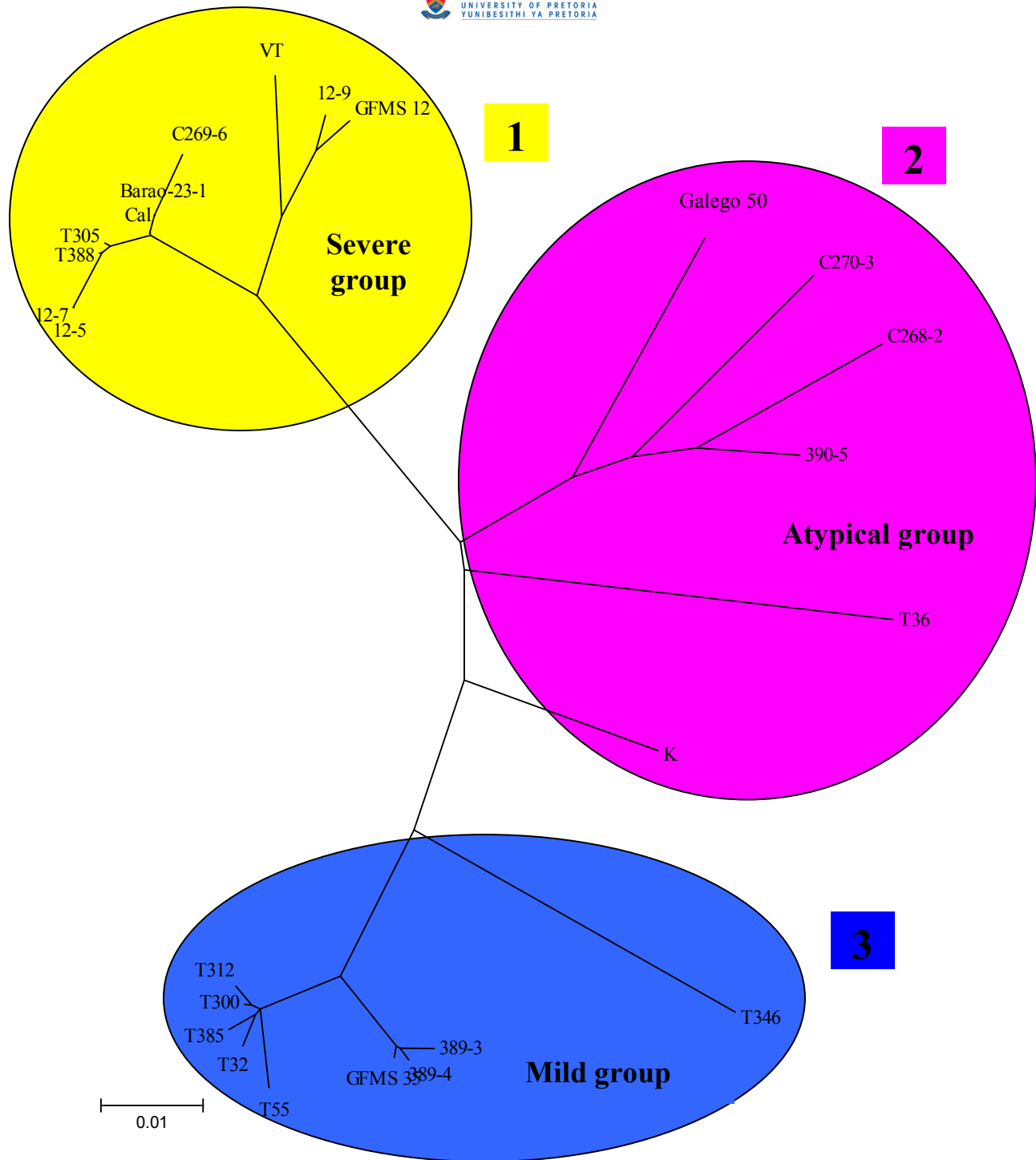


Figure 2: An unrooted phylogenetic tree obtained by the neighbour joining method with the nucleotide sequences of the p23 gene of RSA and other reference CTV isolates.

DISCUSSION

Previously these RSA isolates were biologically indexed in the greenhouse on Mexican lime (van Vuuren *et al.*, 2000; Breytenbach, unpublished) and appear to be mild. However GFMS 12 pre-immunized in Star Ruby Grapefruit was shown to perform poorly and possibly contained a severe strain (Marais *et al.*, 1996) resulting in GFMS35 being introduced to pre-immunize red grapefruit selections (van Vuuren *et al.*, 2000b).

Single amino acid differences of the p23 gene at positions 78-80 can reliably be detected in this PCR system developed by Sambade *et al.* (2003). RSA isolates 390-3 and 390-5 appear to be atypical as shown by the atypical group indicator T36. RSA isolates 390-4, 389-4 and 389-3 appear to be mild, as shown by the mild group control T385. RSA cross-protecting isolate GFMS 35, a population of different strains used in the cross-protection scheme showed a mixed pattern of mild and atypical strains. RSA isolates GFMS 12, 12-7 and 12-9 have mixed infections of severe and mild strains. Isolate 12-5 belongs to the severe group as shown by the severe control T305. Detection of mixed strains/variants and severe components in isolates 12-5, 12-7 and 12-9 is of concern since these sub-isolates from single aphid transmissions were indexed as mild (van Vuuren *et al.*, 2000). This raises concerns that a severe component could become dominant and cause cross-protection breakdown, especially since the effects of strain dominance; varying host responses; or strain interactions are presently unknown.

The predicted p23 amino acid sequences of four regions had increased variability for possible differentiation of isolates between positions: 24-29; 50-54; 78-80 and 125-129 which is similar to the findings of Sambade *et al.* (2003) where positions 24-29 and 50-54 separated mild isolates from others, whereas positions 78-80 allowed discrimination between the three groups. Positions 50-54 (basic region) and 78-80 (zinc finger) are part of a p23 domain involved in binding RNA in a non-specific manner (López *et al.*, 2000). The area between amino acid residues 98-171 does not affect RNA binding (López *et al.*, 2000) and has no known function.

At amino acid positions 24, 26 and 29, Sambade *et al.* (2003) found that isolates could be separated into severe and mild isolates with the composition Glu²⁴, Lys²⁶, Ser²⁹ and Lys²⁴, Glu²⁶, Lys²⁹ respectively. These differences affect the isoelectric point from 9.94

in mild isolates to 6.97 in severe isolates. RSA isolates 12-9, 12-7, 12-5 and GFMS 12 had Glu²⁴, Lys²⁶ and Ser²⁹ as with other characterized severe isolates and some reference atypical isolates: K, T36 and Galego-50. None of the presumed mild RSA isolates had the composition of the mild group of Sambade *et al.* (2003) reference isolates with Lys²⁴, Glu²⁶ and Lys²⁹. RSA isolates GFMS 35, 389-3 and 389-4 had Arg²⁴, Glu²⁶ and Lys²⁹, the same as the reference isolate T346 from the mild group, which clustered separately from the majority of mild reference isolates. Isolate 390-5 had Ala²⁴, Arg²⁶ and Ser²⁹, the same as other atypical reference isolates 270-3 and 268-2. It seems as though this site is not reliable for differentiation of mild versus atypical groups since there are two different compositions for the biologically mild isolates and for atypical isolates. The severe strains had Glu²⁴, Lys²⁶ and Ser²⁹ which could potentially be used as a marker for severe variants and that no mild isolates had this composition. The composition of the severe and mild isolates shows that all three amino acid positions differ and this could represent a major evolutionary difference and possibly an ancient divergence of CTV variants into distinct groups and with the strong modification of the isoelectric point, could result in functional differences. The role of atypical strains in strain variability remains unclear but in the amino acid region 24-29 the two compositions suggests two main divergent paths: (1) isolates K, Galego-50 and T36 were the same as severe isolates (Glu²⁴, Lys²⁶, Ser²⁹) and (2) isolates c270-3, c268-2 and 390-5 had a completely unique and divergent composition (Ala²⁴, Arg²⁶ and Ser²⁹). This could represent a split in divergence of isolates from the severe group to isolates which have an undefined severity and variable symptoms. Interestingly the RSA isolate 390-5 consisted of the unique atypical composition of unknown significance.

Amino acid positions 50, 53 and 54 of RSA isolates 390-5, GFMS 35, 389-3 and 389-4 were Val⁵⁰, Thr⁵³ and Asn⁵⁴, the same as the mild reference isolates and the atypical reference isolate c270-3. RSA isolates GFMS12, 12-9, 12-7, 12-5 and the all severe reference sequences had Ile⁵⁰, Asn⁵³ and Ser⁵⁴, as well as T36 from the atypical group. There were six different compositions in the atypical reference group of isolates, of which three were unique to this group. There was a lot of variability among atypical isolates which raises the question as to where these isolates actually fit in.

Amino acid positions 78, 80 and 81 of RSA isolates GFMS 35, 389-3 and 389-4 had Ala⁷⁸, Leu⁷⁹ and Lys⁸⁰, expected of the mild group. RSA isolates GFMS 12, 12-5, 12-7

and 12-9 had Ala⁷⁸, Ser⁷⁹ and Arg⁸⁰, typical of the severe group of isolates (Sambade *et al.*, 2003). RSA isolate 390-5 had Gly⁷⁸, Leu⁷⁹ and Lys⁸⁰ characteristic of the atypical group (Sambade *et al.*, 2003). The only exception was atypical isolate Galego-50 with Gly⁷⁸, Leu⁷⁹ and Arg⁸⁰. The amino acid residues in this region seem to follow a distinct pattern: (1) at position 78 it is either Ala for severe and mild or Gly for atypical isolates; (2) at position 79 it is either Leu for mild and atypical or Ser for severe isolates; and lastly (3) at position 80 it is either Lys for mild and atypical or Arg for severe (including atypical Galego-50) isolates. It appears from this region that the mild and atypical group are most similar (one amino acid difference) followed by the severe and mild groups (with two amino acid changes) and then by the severe and atypical groups having all three amino acids as different. Concurring with Sambade *et al.* (2003) this region allows discrimination between the mild, severe and atypical groups and is the area from which the primers in the bi-directional PCR were developed. The three groups can clearly be separated, however the actual involvement of each group in symptom expression can not be concluded from this. More research should be done to document the effects of the 78-80 position on symptom expression.

Amino acid region 125-129 of unknown function, showed variability among isolates. The mild group of reference isolates had three different composition types with the majority of isolates having Asp¹²⁵, Met¹²⁸ and Tyr¹²⁹. Most atypical isolates also had this composition including RSA isolate 390-5 and severe isolate T305. RSA isolates GFMS 35, 389-3 and 389-4 had Glu¹²⁵, Leu¹²⁸ and Tyr¹²⁹; the same as the reference isolate T385 from the mild group. Reference isolate T32 from the mild group had a unique composition of Asp¹²⁵, Ser¹²⁸ and Tyr¹²⁹. RSA isolates GFMS 12, 12-5, 12-7 and 12-9 had Asp¹²⁵, Met¹²⁸ and His¹²⁹, the same as the majority of severe reference isolates. Interestingly isolate K from the atypical group also had this composition. It appears from this area that there are two main sequence types: the mild Asp¹²⁵, Met¹²⁸ and Tyr¹²⁹; and severe Asp¹²⁵, Met¹²⁸ and His¹²⁹ that differ by only one amino acid. In this region it is unclear where the atypical isolates fit in but it appears that they either group with a mild or severe sequence composition.

Amino acid positions 24-29, 50-54, 125-129 can potentially be used as molecular marker sites to differentiate mild and severe strains. There have been a few exceptions though to the given pattern produced by severe and mild isolates, and therefore these

sites should not be used separately but as a whole to get a better picture of what an isolate contains and from there possibly assign it to a defined group. The area 78-80 still seems to be a more suitable area to assign isolates into mild and severe groups but more importantly to place isolates with no apparent defined symptom type into another group called atypical. However isolates in this atypical group seem to all be very different and it seems unlikely that this region alone would account for such different symptom expression. It is however plausible that the region 78-80 could interact with other regions on the genome to produce defined symptoms. However any functional differences discussed in this study of the p23 gene might not relate to symptom expression at all. This site is however a step forward to understanding possible interactions among strains and their symptoms. The p23 gene is highly conserved and differences in DNA sequence could represent possible differences in function, which would not necessarily be the case in highly variable regions.

Phylogenetic analysis of the sequences of the p23 gene of eight RSA isolates and 18 biologically indexed isolates of Sambade *et al.* (2003) supported the grouping pattern of Sambade *et al.* (2003). All CTV isolates clearly separated into three groups based on distinct symptoms observed by Sambade *et al.* (2003): mild (group 3); severe (group 1); and the atypical group (group 2). Atypical isolates showed variable pathogenicity and their sequences were less closely related to one another in their group than those sequences within the mild and severe groups. The atypical reference sequences induced seedling yellows except isolate K and T36. Isolate K has been shown to contain a mixture of variants some of which cause seedling yellows and/or stem-pitting in grapefruit (Brlansky *et al.*, 2003). There are no records of T36 causing any stem-pitting or seedling yellows but only decline.

RSA isolates: 12-5, 12-7, 12-9 and GFMS 12 grouped with the severe reference isolates; 390-5 grouped together with the atypical reference isolates; and 389-3, 389-4 and GFMS35 grouped together with the mild reference isolates. These groupings were upheld by the intra-group and inter-group diversity scores. With group 2 sequences more highly diverse compared to group 1 and group 3 sequences. The most diverse inter-group score was between the groups 2 and 3. The phylogenetic tree does not support this observation, where it appears that group 1 and 3 are the most diverse.

Within the severe group RSA isolates 12-9 and GFMS 12 clustered very closely to the severe VT reference isolate; and RSA isolates 12-5 and 12-7 clustered closely to T388 and T305 severe reference isolates. Isolates in the severe group can be separated into two sub-groups on the tree: (A) 12-9, GFMS12 and VT; and (B) 12-5, 12-7, T388, T305, Cal-CB, Barao, and 269-6. The sequences of RSA isolates 12-7 and 12-5 are identical and could be a single aphid transmission in which the exact strain was selected from the mixture of strains of GFMS12 or is a different variant with a lot of conservation within the conserved p23 gene. Isolates 12-9 and GFMS 12 also had a high homology with defective RNA sequences (D-RNAs). D-RNAs are variable between isolates and are suggested to correlate to Seedling yellows (SY) symptom expression (Garnsey *et al.*, 2000). Isolate GFMS 12 is a mixture of strains used as a mild-strain cross-protecting source. Sub-isolates made from this source have shown varying degrees of stem-pitting ((van Vuuren *et al.*, 2000). This is of concern for cross-protection since these severe variants could under certain circumstances and strain dynamics result in severe symptom expression.

Within the atypical group RSA isolate 390-5 clustered together with c268-2. These isolates together with c270-3 from Argentina and Galego-50 formed a separate subgroup in the atypical group (group 2) and induce seedling yellows (SYs) and stem-pitting (SP). The sequence of isolate 390-5 had a high homology to other Argentinean isolates C315-14 and c271-8, not included in this study because of their unknown biological symptoms. It is significant that these sequences in this sub-group are all from Argentina. Reference isolates T36 and K formed their own branches distinctly separate from the other atypical isolates and are quite different from the other atypical isolates and from one another. Isolates capable of inducing SYs and SP have clustered together and isolate T36 which causes decline and K generally known to be symptomless have all clustered separately according to possibly their different symptoms. Questions still remain as to the significance of the atypical isolates in strain differentiation.

Within the mild group (group 3) of isolates there were three distinct sub-groups. RSA isolates GFMS35, 389-3 and 389-4 clustered together in a subgroup of their own and isolates T312, T300, T32, T385 and T55 formed their own subgroup. These two subgroups were more closely related than the last mild isolate T346 which formed its own subgroup more distant to the other two. The sequences of RSA isolates 389-3, 389-

4 and GFMS35 had a high homology. It appears that the RSA isolates are unique compared to other mild isolates and have formed their own sub-group. It is possible that the mild strains in South Africa have evolved separately to mild strains from other countries, possibly due to different evolutionary pressures and levels of these pressures.

There was a good correlation between designating RSA isolates 390-4, 389-3, 389-4 as mild; 12-5 as severe; and 390-3, 390-5 as atypical in groups based on the PCR results which targeted differences at amino acids 78-80 and phylogenetic analysis of the complete p23 gene. RSA isolates 12-7, 12-9 and GFMS 12 had mixed mild and severe strains; and GFMS 35 had mixed atypical and mild strains. However for routine differentiation of strains based on the p23 gene it would only be necessary to use the PCR system since DNA sequencing and phylogenetic analysis would be too cumbersome and expensive. The p23 gene however only represents a small portion of the complete genome and therefore this system should be used in conjunction with other systems targeting other regions of the genome to get a complete picture of an isolate's sequence type before assuming any definite answer as to the strain type.

ACKNOWLEDGEMENTS

Thanks to Citrus Research International for funding for this study.

REFERENCES

1. Albertini, D., Vogel, R., Bové, C., Bové, J.M. 1988. Transmission and preliminary characterization of Citrus tristeza virus strain K. In "Proceedings 10th Conference of the International Organization of Citrus Virologists" (L.W. Timmer, S.M. Garnsey, L. Navarro, Eds), pp 17-21, IOCV. Riverside, California.
2. Ayllón, M.A., López, C., Navas-Castillo, J., Garnsey, S.M., Guerri, J., Flores, R., Moreno, P. 2001. Polymorphism of the 5' terminal region of Citrus tristeza virus (CTV) RNA: Incidence of three sequence types in isolates of different origin and pathogenicity. *Arch Virol* 146: 27-40.
3. Dolja, V. V., Karasev, A.V., Koonin, E.V. 1994. Molecular biology and evolution of closteroviruses. Sophisticated build up of large RNA genomes. *Annu. Rev. Phytopathol.* 32: 261-285.

4. Febres, V. J. 1995. Ph.D. Thesis, p 104, University of Florida, Gainesville, FL.
5. Garsney, S.M., Timmer, L. W., Graham, J.H. 2000. Compendium of Citrus Diseases 2nd ed. APS Press. USA.
6. Ghorbel, R., López, C., Fagoaga, C., Moreno, P., Navarro, L., Flores, R., Pena, L. 2001. Transgenic citrus plants expressing the citrus tristeza virus p23 protein exhibit viral-like symptoms. *Mol. Plant Path.* 2: 27-36.
7. Hilf, M.E., Karasev, A.V., Pappu, H.R., Gumpf, D.J., Niblett, C.L., Garnsey, S.M. 1995. Characterization of *Citrus tristeza virus* subgenomic RNAs in infected tissue. *Virology* 208: 576–582.
8. Kumar, S., Tamura, K., Nei, M. 1993. MEGA. Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park.
9. López, C., Navas-Castillo, J., Gowda, S., Moreno, P., Flores, R. 2000. The 23-kDa protein coded by the 3'-terminal gene of Citrus tristeza virus is an RNA-binding protein. *Virology* 269: 462-470.
10. Lu, R., Folimonov, A., Li, W. X., Choi, Y.G., Shintaku, M., Falk, B.W., Dawson, W.O., Dind, S.W. 2003. Citrus tristeza virus genome encodes two distinct suppressors of RNA silencing. American Society of Virology Meeting, July 2003, Davis, (Abstract), W19-3.
11. Pappu, S.S., Febres, V.J., Pappu, H.R., Lee, R.F., Niblett, C.L. 1997. Characterization of the 3' proximal gene of the citrus tristeza closterovirus genome. *Virus Res.* 47: 51-57.
12. Rubio, L., Ayllón, M.A., Kong, P., Fernandez, A., Polek, M., Guerri, J., Moreno, P., Falk, B. 2001. Genetic variation of citrus tristeza virus isolates from California and Spain : Evidence for mixed infections and recombination. *J. of Virol.* 75: 8054-8062.
13. Sambade, A., López, C., Rubio, L., Flores, R., Guerri, J., Moreno, P. 2003. Polymorphism of a specific region in gene p23 of Citrus tristeza virus allows discrimination between mild and severe isolates. *Arch. Virol.* 148: 2325-2340.

14. Vives, M.C., Rubio, L., López, C., Navas-Castillo, J., Albaich-Martý, M.R., Dawson, W.O., Guerri, J., Flores, R., Moreno, P. 1999. The complete genome sequence of the major component of a mild Citrus tristeza virus isolate. *J.Gen.Virol.* 80: 811-816.
15. van Vuuren, S.P., Collins, R.P. 1993. Evaluation of citrus tristeza virus isolates for cross-protection of grapefruit in South Africa. *Plant Disease* 77: 24-28.
16. van Vuuren, S.P., van der Vyver, J.B. 2000. Comparison of South African pre-immunizing citrus tristeza virus isolates with foreign isolates in three grapefruit selections. In "Proceedings 14 th Conference of the International Organization of Citrus Virologists" (J.V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp 50-56, IOCV. Riverside, California..
17. van Vuuren, S.P., van der Vyver, J.B. Luttig, M. 2000. Diversity among sub-isolates of cross-protecting citrus tristeza virus isolates in South Africa. In "Proceedings 14 th Conference of the International Organization of Citrus Virologists" (J.V. da Graca, R.F. Lee, R.K. Yokomi, Eds), pp 103-110, IOCV. Riverside, California.

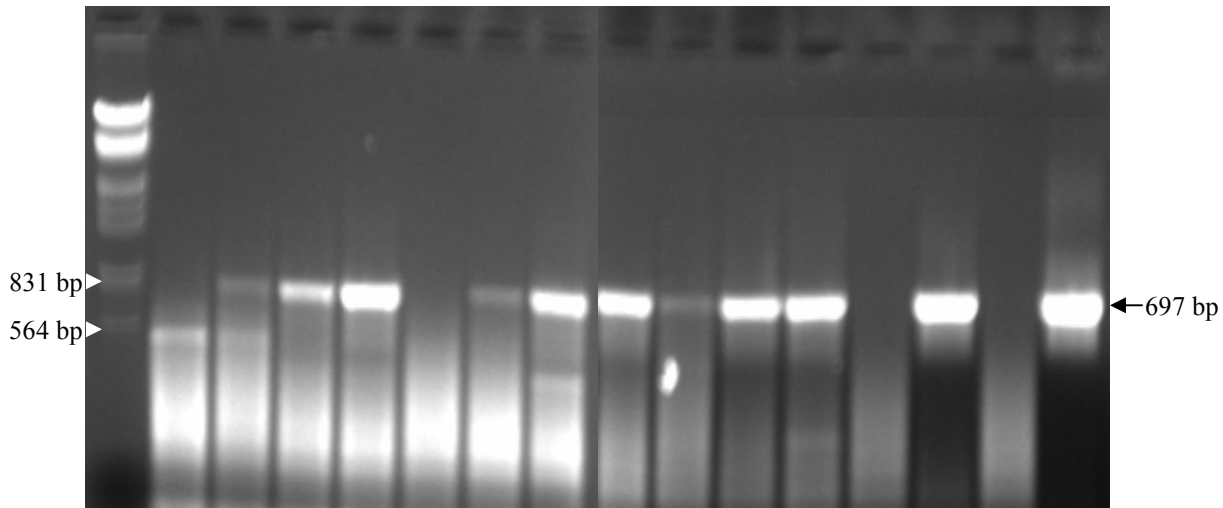


Figure 3: 1 % Gel electrophoresis photo of the conserved PCR of p23 gene from 11 CTV isolates using primers PM50 and PM51. The expected product is 697 bp. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA), (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 390-4; (6) 389-3; (7) 390-5; (8) 12-7; (9) 12-5 (10) 12-9 (11) GFMS 12; (12) Virus free; (13) T36 DNA clone; (14) Buffer control; and (15) GFMS 12 control cDNA.

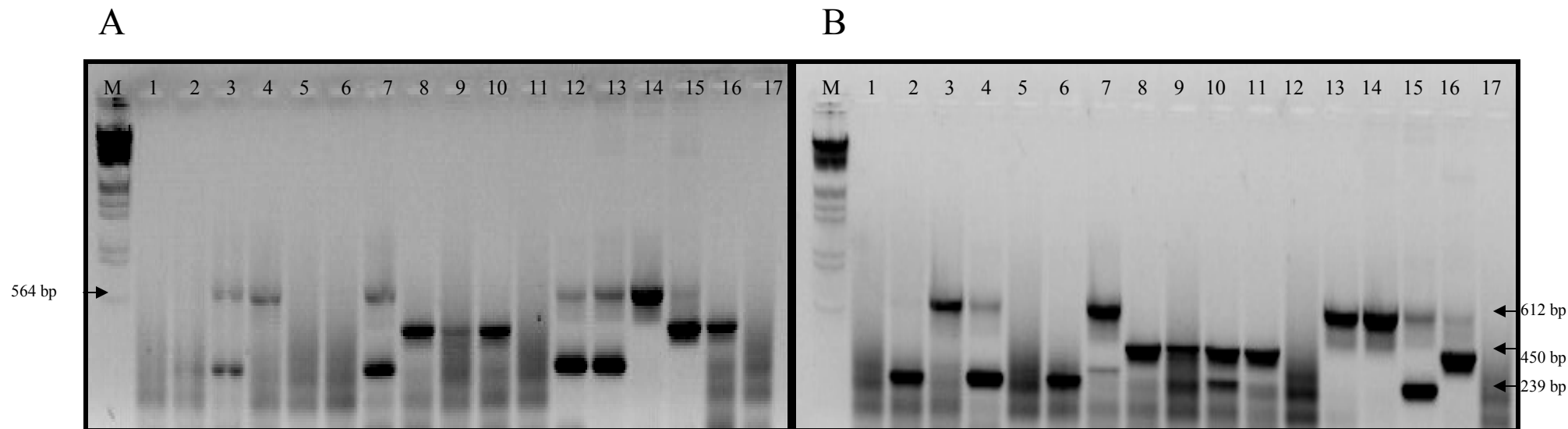


Figure 4 : 1 % Gel electrophoresis photo of the bi-directional PCRs of p23 gene from 11 CTV isolates. **(A)** Using primers PM 85, 86, 82 and 84 to distinguish between severe and atypical strains. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA); (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 390-4; (6) 389-3; (7) 390-5; (8) 12-7; (9) 12-5 (10) 12-9 (11) Virus free; (12) T36 DNA clone (13) T36 DNA clone; (14) T385 DNA clone; (15) T305 DNA clone; (16) GFMS 12; and (17) Buffer control. **(B)** Using primers PM 85, 86, 82 and 83 to distinguish between severe and mild strains. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA), (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 390-4; (6) 389-3; (7) 390-5; (8) 12-7; (9) 12-5 (10) 12-9 (11) GFMS 12; (12) Virus free; (13) T36 DNA clone; (14) T36 DNA clone; (15) T385 DNA clone; (16) T305 DNA clone; and (17) Buffer control.