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 $MgCl_2$, 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 Glu^{24}, Lys^{26} and Ser^{29} 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 Lys^{24}, Glu^{26} and Lys^{29}. RSA isolates GFMS 35, 389-3 and 389-4 had Arg^{24}, Glu^{26} and Lys^{29}, as with the mild reference isolate T346. RSA isolate 390-5 had Ala^{24}, Arg^{26} and Ser^{29}, 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 Val^{50}, Thr^{53} and Asn^{54}, 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 (Ile^{50}, Tyr^{53} and Asn^{54});

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	Atypical control			Atypical control		Atypical control
+ Control: T385 clone	Mild control		Mild control	Mild control		
+ Control: T305 clone	Severe control	Severe control		Severe control	Severe control	
- Control: Virus free						
- Control: Buffer						

\+ represents a positive control
\- represents a negative control

isolate K (Ile^{50}, Ala^{53} and Ser^{54}); and isolate c268-2 (Val^{50}, Thr^{53} and Ser^{54}). 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 Ile^{50}, Asn^{53} and Ser^{54}, as well as the atypical reference isolate T36.

In the region 78-80 isolates GFMS 35, 389-3 and 389-4 had Ala^{78}, Leu^{79} and Lys^{80}, 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 Ala^{78}, Ser^{79} and Arg^{80}, 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 Gly^{78}, Leu^{79} and Lys^{80}, 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 Gly^{78}, Leu^{79} and Arg^{80}.

Region 125-129 showed high variability amongst isolates. Isolates GFMS 35, 389-3 and 389-4 had Glu^{125}, Leu^{128} and Tyr^{129}; 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 Asp^{125}, Met^{128} and Tyr^{129} 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 Asp^{125}, Ser^{128} and Tyr^{129}. RSA isolates GFMS 12, 12-5, 12-7 and 12-9 had Asp^{125}, Met^{128} and His^{129}, 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.

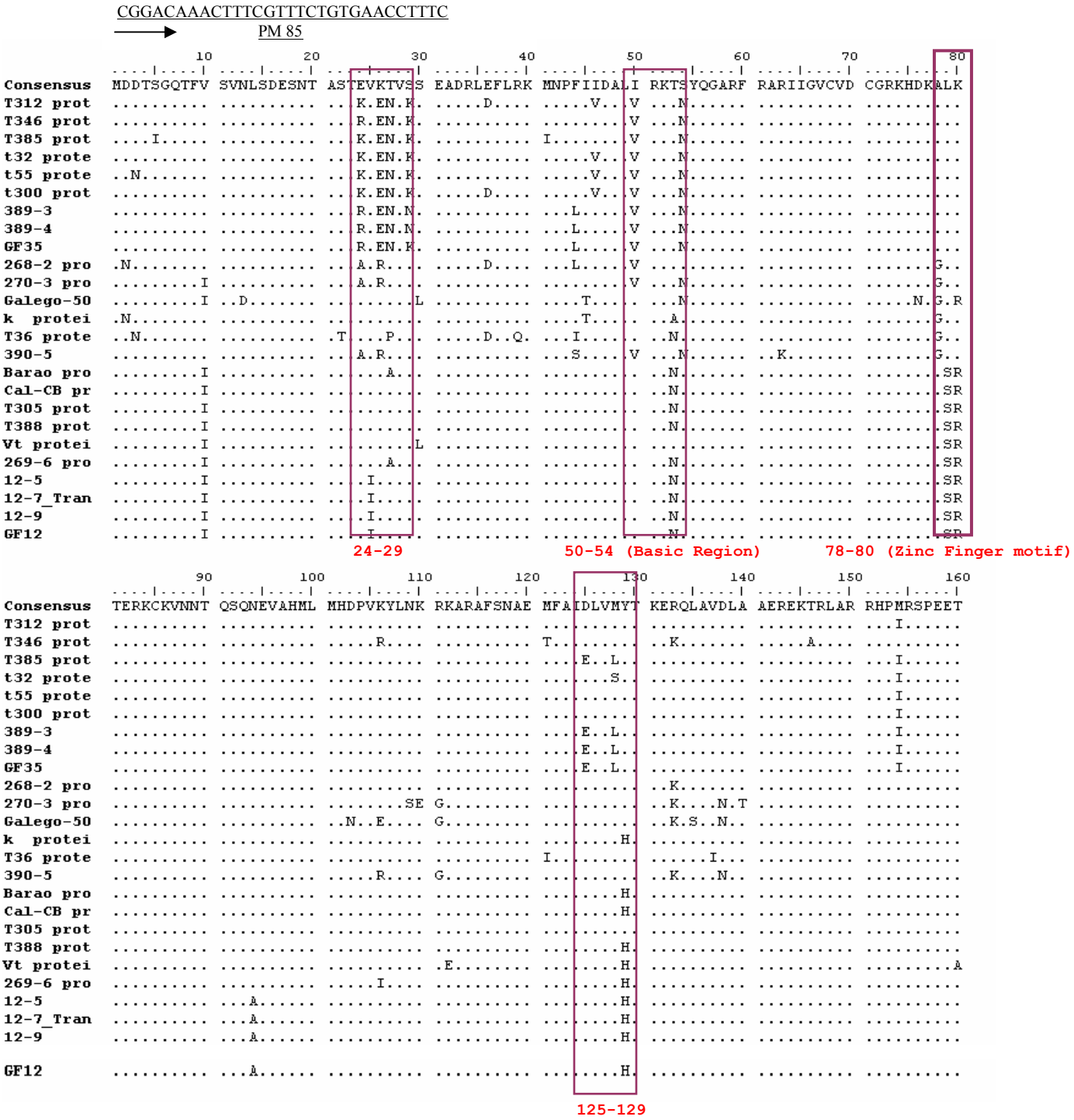

Figure 1 continued

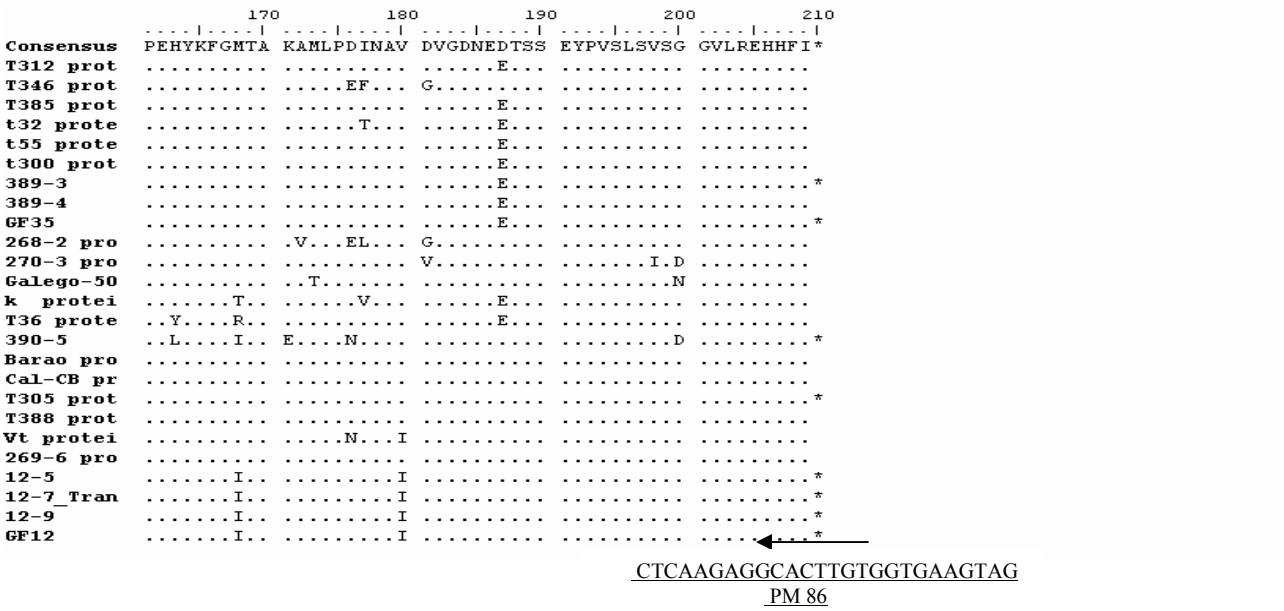

Table 2:

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	VTS	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).

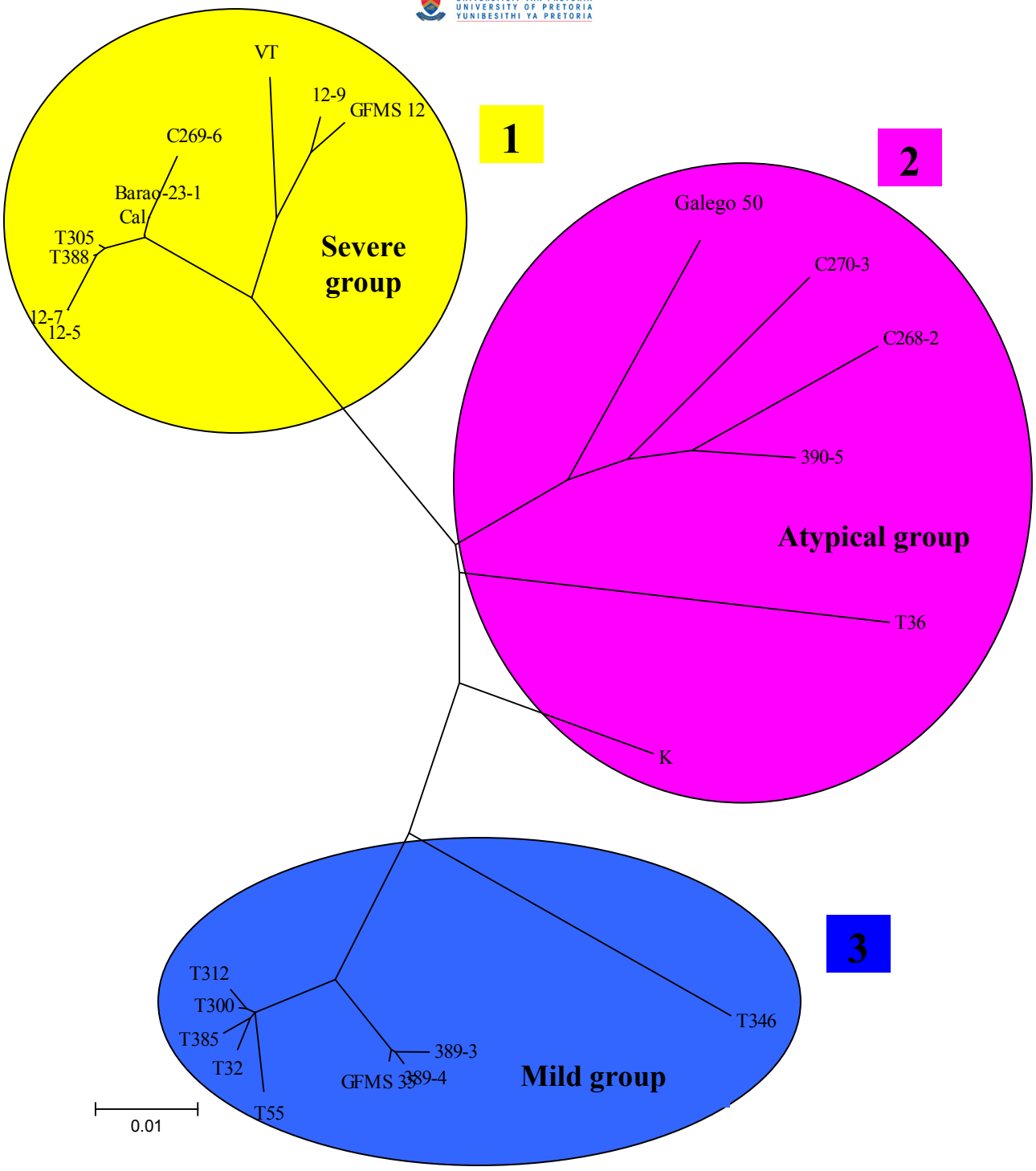

Figure 2: An unrooted phylogenetic tree obtained by the neighbour joining method with the nucleotide sequences of the p23 gene of RSA and other reference CTV isolates.

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 Glu^{24}, Lys^{26}, Ser^{29} and Lys^{24}, Glu^{26}, Lys^{29} 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 Glu^{24}, Lys^{26} and Ser^{29} 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 Lys^{24}, Glu^{26} and Lys^{29}. RSA isolates GFMS 35, 389-3 and 389-4 had Arg^{24}, Glu^{26} and Lys^{29}, 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 Ala^{24}, Arg^{26} and Ser^{29}, 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 Glu^{24}, Lys^{26} and Ser^{29} 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 (Glu^{24}, Lys^{26}, Ser^{29}) and (2) isolates c270-3, c268-2 and 390-5 had a completely unique and divergent composition (Ala^{24}, Arg^{26} and Ser^{29}). 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 Val^{50}, Thr^{53} and Asn^{54}, 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 Ile^{50}, Asn^{53} and Ser^{54}, 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 Ala^{78}, Leu^{79} and Lys^{80}, expected of the mild group. RSA isolates GFMS 12, 12-5, 12-7

and 12-9 had Ala^{78}, Ser^{79} and Arg^{80}, typical of the severe group of isolates (Sambade *et al.*, 2003). RSA isolate 390-5 had Gly^{78}, Leu^{79} and Lys^{80} characteristic of the atypical group (Sambade *et al.*, 2003). The only exception was atypical isolate Galego-50 with Gly^{78}, Leu^{79} and Arg^{80}. 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 Asp^{125}, Met^{128} and Tyr^{129}. 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 Glu^{125}, Leu^{128} and Tyr^{129}; the same as the reference isolate T385 from the mild group. Reference isolate T32 from the mild group had a unique composition of Asp^{125}, Ser^{128} and Tyr^{129}. RSA isolates GFMS 12, 12-5, 12-7 and 12-9 had Asp^{125}, Met^{128} and His^{129}, 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 Asp^{125}, Met^{128} and Tyr^{129}; and severe Asp^{125}, Met^{128} and His^{129} 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 (group2). 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.

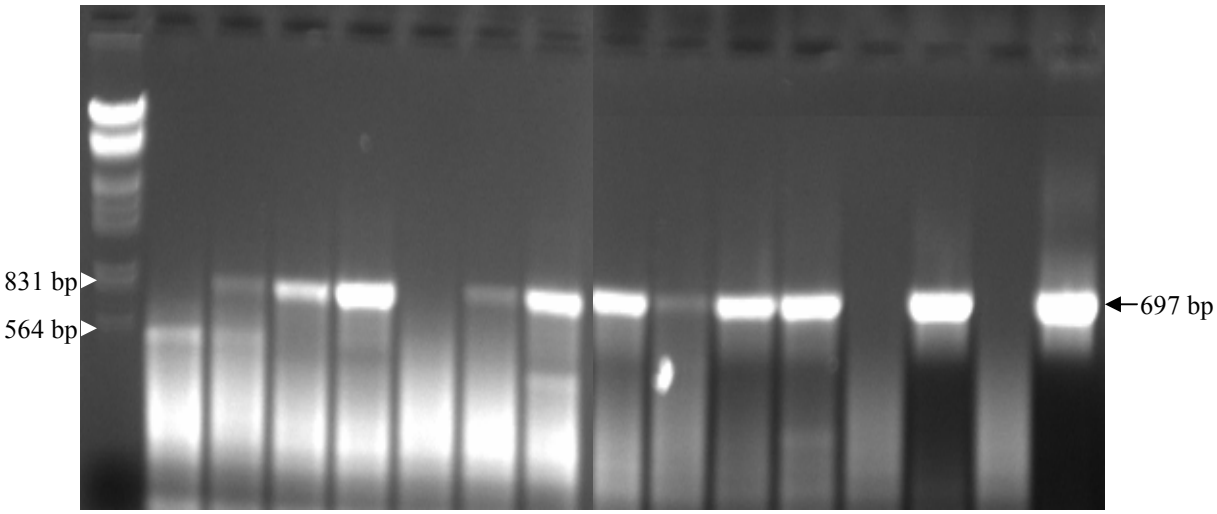

Figure 3: 1 % Gel electrophoresis photo of the conserved PCR of p23 gene from 11 CTV isolates using primers PM50 and PM51. The expected product is 697 bp. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA), (1) T30 strain plant; (2) GFMS 35; (3) 390-3; (4) 389-4; (5) 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.

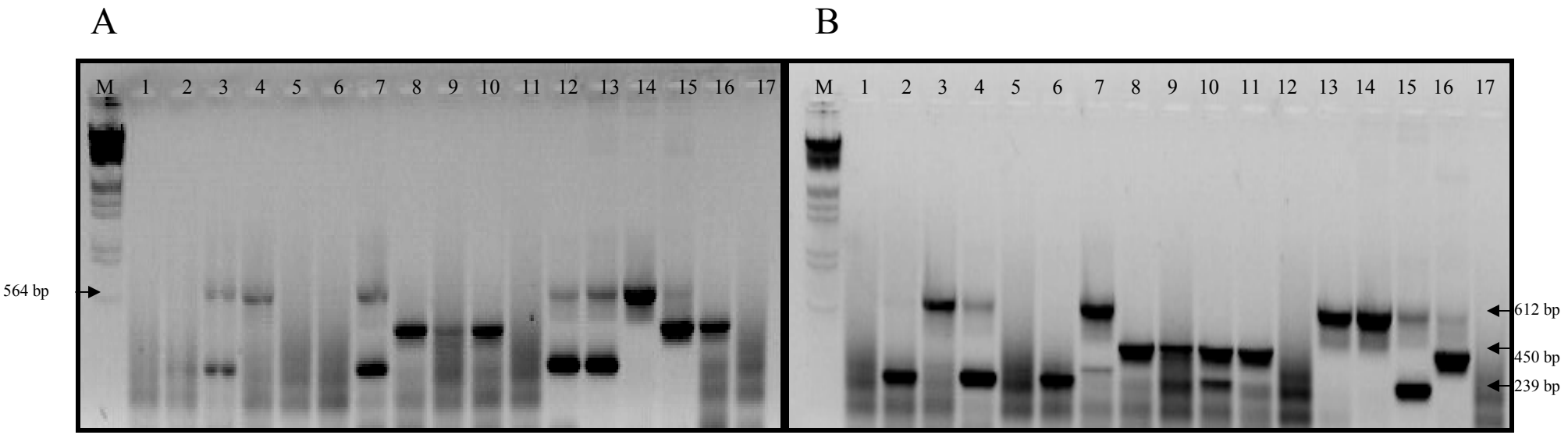

Figure 4 : 1 % Gel electrophoresis photo of the bi-directional PCRs of p23 gene from 11 CTV isolates. **(A)** Using primers PM 85, 86, 82 and 84 to distinguish between severe and atypical strains. Lanes: (M) Molecular marker (Promega Lambda DNA marker – *Eco* R I + *Hind* III, USA); (1) T30 strain plant; (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.