DETECTION, COMPLETE NUCLEOTIDE SEQUENCE AND CONSTRUCTION OF FULL cDNA CLONES OF A GREEK ISOLATE OF *TOMATO CHLOROSIS VIRUS*

BY

AMR KATAYA

CHANIA, GREECE

2007

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor Dr. Ioannis Livieratos, for the guidance and support he has provided throughout the course of this work, and specially his valuable comments and constructive criticism.

I also acknowledge Dr. Antonios Makris, Coordinator of the Department of Natural Products, and Mr. A. Nikolaidis, Director of the Mediterranean Agronomic Institute of Chania, for giving me the opportunity to continue my M.Sc. postgraduate studies at the institute.

I am grateful to my collaborators and friends in the lab: Eleftheria, Rita, Ahou, Nika, Nikos and Theodore, for their continuous help and encouragement.

Special thanks also to Mrs. Monica Lagoudaki, Secretary of the Sustainable Agriculture Department at the Mediterranean Agronomic Institute of Chania for her help and kindness. I would also like to thank Mrs. Maria Verivaki for the English correction.

It would be a long list to mention all of the friends and colleagues I am indebted to. It is my great pleasure to thank all of them for their support and friendship, specially Jawad, Laureta, Ikram, Aly, Hadi and Reyad.

Finally, I would like to express my deepest love and recognition to my father, my mother and my brothers, for their love and continuous support. I owe them everything I have today and I dedicate this thesis to them.

ABSTRACT

Tomato chlorosis virus (ToCV; genus *Crinivirus*; family *Closteroviridae*) is a whitefly-transmitted crinivirus with a bi-partite RNA genome inducing yellowing diseases in greenhouse and open-field tomato cultivations. Diagnostic reverse transcription-polymerase chain reaction (RT-PCR) assays showed that similarly with dot-blot hybridization experiments, the virus can be well detected in 20ng of total RNA extracts from infected plants. The complete sequence of both genomic RNA molecules (RNA1 and RNA2) of a Greek ToCV isolate (Gr-535) has been determined and compared with the American and Spanish isolates as well as other fully sequenced members of the genus *Crinivirus*. Overall, it was observed that the Greek ToCV isolate is most closely related to the American at a protein amino acid level, while when nucleotide sequences of the respective untranslated regions (UTRs) are compared, the American isolate contains nucleotide additions or deletions which are not found in the Greek and Spanish isolates. Specifically, the two genomic RNA molecules of a Greek isolate comprise 8594 and 8242 nucleotides (nt) and both share high identity with the Spanish (97%) and American (99%) isolates. Phylogenetic analysis using the deduced amino acid sequences of the RNA dependent RNA polymerase, p22 and coat protein showed that the Greek isolate is more closly related to the American, and that ToCV is most similar to S*weet potato chlorotic stunt virus* and C*ucurbit yellow stunting disorder virus*. Prediction of the 3`-terminal folding of ToCV RNA1, showed the presence of four stem loops and a pseudo-knot, which has also been predicted for other sequenced criniviruses. Similar secondary structures could be generated for both American and Spanish RNA2 3`-UTR but not for the Greek isolate. The complete genomic ToCV RNAs 1 and 2 were reverse-transcribed, PCR-amplified and inserted immediately downstream the sequence of the bacteriophage T7 RNA polymerase promoter into a pUC19 vector to construct a full length cDNA clone of the virus. RNA transcripts can now be transcribed from both these clones and tested for infectivity in a protoplast system using the T7 DNA dependent RNA polymerase. These experiments will facilitate future studies of ToCV replication in a single-cell level and provide important details about a diverse in terms of genome organization.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **ii**
ABSTRACT ..... **iii**
TABLE OF CONTENTS ..... **iv**
LIST OF FIGURES ..... **vii**
LIST OF TABLES ..... **ix**
LIST OF ABBREVATIONS ..... **x**

CHAPTER 1 Literature review **1**
1.1 Economical importance of closteroviruses 1
1.2 An overview of the family *Closteroviridae* 1
1.2.1 Taxonomy 1
1.2.2 Virus morphology and biology 4
1.2.2.1 Morphology 4
1.2.2.2 Cytopathology 5
1.2.2.3 Genome Organization 7
1.2.2.4 Genome expression strategy. 9
1.2.2.5 Cell-to-cell movement and assembly 11
1.2.2.6 Transmission 12
1.2.3 Genus *Crinivirus* 12
1.2.4 ToCV 15
1.2.4.1 Introduction- Symptomatology 15
1.2.4.2 Virus transmission 15
1.2.4.3 Genome organization and virus replication 15
1.3 Detection and diagnosis 17
1.3.1 Detection based on virus biology 17
1.3.2 Virus detection based on properties of the virus proteins 18
1.3.3 Virus detection based on properties of the viral nucleic acid 18
1.4 Infectious viral cDNA clones 19
1.5 The scope of the investigation 21

CHAPTER 2 Materials and Methods 22
2.1 Sterilization and precautions 22
2.2 Bacterial strains 22
2.3 Bacterial media and antibiotics 23
2.4 Cloning vectors 24
2.5 Oligonucleotides 26
2.6 Molecular biology techniques 28
2.6.1 Virus source and purification of total RNA 28
2.6.2 Reverse transcription (RT) from total RNA extracts 29
2.6.3 Polymerase Chain Reaction (PCR) 30
2.6.4 Agarose gel electrophoresis 31
2.6.5 Purification of the PCR product following gel electrophoresis 31
2.6.6 Cloning of PCR products 32
2.6.6.1 Cloning in plasmid vector 32
2.6.6.2 Plasmid DNA purification 32
2.6.6.3 Restriction endonuclease digestion of DNA 33
2.7 Bacterial transformation 33
2.7.1 Preparation of bacterial competent cells 33
2.7.2 Transformation 34
2.8 Phenol-chloroform extraction 34
2.9 Ethanol precipitation 34
2.10 Nucleotide sequencing analysis 35

CHAPTER 3 Results 36
3.1 Diagnosis of the virus using RT-PCR 36
3.2 Strategies for ToCV RNA1 & 2 full cDNA clones construction 38
3.2.1 RT-PCR amplifications of ToCV RNA1 42
3.2.2 Cloning of ToCV RNA1 fragments into pGEM-T Easy Vector 44
3.2.3 RT-PCR amplification of ToCV RNA2 46
3.2.4 Cloning of ToCV RNA2 fragments into pGEM-T Easy Vector 48

3.3 Sequence analysis 49
3.4 Construction of ToCV RNA1 full length cDNA clones 56
3.5 Construction of ToCV RNA2 full length cDNA clones 62

CHAPTER 4 Discussion 65

REFERENCES 67

APPENDICES 77

LIST OF FIGURES

Figure 1.1 Morphology of the BYV virions 4

Figure 1.2 Microscopic analysis of ToCV-infected *Nicotiana clevelandii* tissue 6

Figure 1.3 Genome maps of the selected representatives of the family *Closteroviridae* ... 8

Figure 1.4 Map of the BYV genome..... 10

Figure 1.5 Hypothetical model of BYV assembly and cell-to-cell movement 11

Figure 1.6 Genome organizations for the members of the genus *Crinivirus* 14

Figure 1.7 ToCV genome organizations 16

Figure 1.8 Leaf symptoms of TICV and ToCV 17

Figure 2.1 Schematic representation of pGEM-T Easy vector 24

Figure 2.2 pUC19 vector circle map..... 25

Figure 3.1 Detection of ToCV in extracts of infected tomato plants by RT-PCR 37

Figure 3.2 Dot-blot hybridization by DIG-labelled probe corresponding to ToCV 37

Figure 3.3 Strategy followed to construct full length cDNA version of ToCV RNA1 39

Figure 3.4 Oligonucleotides used at 5`- and 3`-termini of the ToCV RNA1 full cDNA . 39

Figure 3.5 Strategy followed to construct full length cDNA version of ToCV RNA2 41

Figure 3.6 Oligonucleotides used at 5`- and 3`-termini of ToCV RNA2 full cDNA 41

Figure 3.7 Gel electrophoresis of RT-PCR products representing ToCV RNA1 43

Figure 3.8 Digestions of the RNA1 cloned genes into pGEM-T Easy by *EcoR*I 44

Figure 3.9 Digestions of pG-PCR9 45

Figure 3.10 Gel electrophoresis of RT-PCR products representing ToCV RNA2 47

Figure 3.11 Electrophoresis analysis for the digestions of pGEM-T Easy cloned fragments by *Not*I 48

Figure 3.12a Multiple alignments of selected genomic regions of ToCV isolates 50

Figure 3.12b Hydrophobic profile for putative proteins of 3` end of RNA1 & 2..... 51

Figure 3.13 Phylogenetic analyses of criniviruses from CP, RdRp and p22 genes. 53

Figure 3.14 Secondary structure models for the 3`-UTRs of ToCV-RNA1. 54

Figure 3.15 Secondary structure models for the 3`-UTRs of ToCV-RNA2 55

Figure 3.16 Agarose gel electrophoresis of steps were used in cloning a full-length cDNA version of ToCV RNA1 59

Figure 3.17 Electrophoresis analysis of: (A) the restriction digestion of pU-PCR1/T7+2+3+4..... 59

Figure 3.18 Gel electrophoresis for (A) digestion of pG-PCR8b+9..... 60

Figure 3.19 Gel electrophoresis of fragments were used in cloning a full-length cDNA version of ToCV RNA1 60

Figure 3.20 Gel electrophoresis of the final step in cloning of ToCV RNA1 cDNA..... 61

Figure 3.21Gel electrophoresis of control digestions of the full-length cDNA version of ToCV RNA1. 61

Figure 3.22 Electrophoresis analysis for steps used in cloning a full-length cDNA version of ToCV RNA2..... 63

Figure 3.23 Electrophoresis analysis for A: double digested pG-PCR7+8+9 63

Figure 3.24 Control digestions of the full-length cDNA version of ToCV RNA2..... 64

LIST OF TABLES

Table 1-1 Taxonomy of the family *Closteroviridae* 2

Table 1-2 List of virus members of the family *Closteroviridae* 2

Table 2-1 RNA1 specific oligonucleotide primers 26

Table 2-2 RNA2 specific oligonucleotide primers 27

Table 2-3 Diagnostic RT-PCR oligonuclotide primers 28

Table 3-1 Percentage of identity between the nucleotide and amino acid sequences..... 52

LIST OF ABBREVATIONS

%	Percentage
ATP	Adenosine triphosphate
AYV	*Abutilon yellows virus*
B. tabaci	*Bemicia tabaci*
bp	Base pairs
BPYV	*Beet pseudoyellows virus*
BuYV	*Burdock yellows virus*
BYSV	*Beet yellow stunt virus*
BYV	*Beet yellows virus*
BYV	*Beet yellows virus*
cDNA	Complimentary DNA
CNFV	*Carnation necrotic fleck virus*
CP	Coat protein
CPm	Minor coat protein
CTV	*Citrus tristeza virus*
CTV	*Citrus tristeza virus*
CYLV	*Carrot yellow leaf virus*
CYSDV	*Cucurbit yellow stunting virus*
CYV	*Clover yellows virus*
CYV	*Cucumber yellows virus*
ddH_2O	Deionized distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxythymidine
dNTPs	Deoxy nucleotide triphosphate
DTT	Dithiotreito
DVCV	*Diodea vein chlorosis virus*
DVNV	*Dendrobium vein necrosis virus*
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetra acetic acid
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
FNV	*Festuca necrosis virus*
g	Gram
g	Gravity
GLRaV-1	*Grapevine leafroll-associated virus 1*
GLRaV-2	*Grapevine leafroll-associated virus 2*
GLRaV-3	*Grapevine leafroll-associated virus 3*
GLRaV-3	*Grapevine leafroll-associated virus 3*
GLRaV-4	*Grapevine leafroll-associated virus 4*
GLRaV-5	*Grapevine leafroll-associated virus 5*

GLRaV-6	*Grapevine leafroll-associated virus 6*
GLRaV-8	*Grapevine leafroll-associated virus 8*
h	Hour
HEL	Helicase
HSP70	Heat shock protein
IBs	Inclusion bodies
ICTV	International Committee on Taxonomy of Viruses
IPTG	Isopropyl-beta-D-thiogalactopyranoside
kb	Kilobase pairs (1000bp)
KCl	Potassium Chloride
kDa	Kilodalton(s)
L	Liter
LB	Lauria-Bertani broth
LChV-2	*Little cherry virus 2*
LCV	*Lettuce chlorosis virus*
LIYV	*Lettuce infectious yellows virus*
L-Pro	Papain-like cysteine proteinase
M	Molar
M	Molecular weight marker
mA	Milliampere
MET	Methyltransferase
mg	Milligram
mg/ml	Milligram per milliliter
$MgCl_2$	Magnisium Chloride
min	Minute
ml	Milliliter
mM	Milimolar
NaCl	Sodium chloride
NaOH	Sodium hydroxide
nm	Nanometer
nt	Nucleotide (s)
^oC	Degree Celsius
OD	Optical density
ORF	Open reading frame
p	Protein
PBNSPaV	*Plum bark necrosis and stem pitting-associated virus*
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pH	Potential hydrogen
PLDS	Plasmalemma deposits
pmol	Picomole
pmol/µl	Picomole per microliter
PMSF	Phenylmethylsulfonyl fluoride
PMWa-1	*Pineapple mealybug wilt-associated virus 1*

PMWa-2	*Pineapple mealybug wilt-associated virus 2*
PTGS	Post-transcriptional gene silencing
PYVV	*Potato yellow vein virus*
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RNase A	Ribonuclease
rpm	Revolutions per minute
RT	Reverse transcriptase
RT-PCR	Reverse transcription polymerase chain reaction
s	Second
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sgRNAs	Subgenomic RNAs
SMMV	*Sugarcane mild mosaic virus*
SPCSV	*Sweet potato chlorotic stunting virus*
ssRNA	Single-stranded ribonucleic acid
TAE	Tris-acetate-EDTA
ToCV	*Tomato chlorosis virus*
TICV	*Tomato infectious chlorosis virus*
U	Units
U/µl	Units per microliter
UTRs	Untranslated regions
V	Volts
vol/vol	Volume/volume
wt/vol	Weight per volume
WYLV	*Wheat yellow leaf virus*
µg	Microgram
µg/ml	Microgram per milliliter
µl	Microliter

CHAPTER ONE

LITERATURE REVIEW

1.1 Economic importance of closteroviruses

Members of the family *Closteroviridae* include viruses with long, thread-like particles and messenger-sense single-stranded RNA genomes. Closteroviruses infect plants of agriculturally important families such as *Cucurbitaceae*, *Rosaceae*, *Solanaceae*, *Compositae*, *Vitaceae* and *Rutaceae*, causing significant economic losses all over the world (Tzanatekis & Martin, 2004). *Citrus tristeza virus* (CTV) is the most devastating citrus virus. In Spain, about 10 million citrus trees have been lost since 1956 due to CTV infections (Bar-Joseph *et al.*, 1989). In 1998, economic losses due to *Lettuce infectious yellows virus* (LIYV) infections on lettuce, sugar beets and melons reached $20 million (Duffus *et al.*, 1996). In one season, growers in the Orange County (California) suffered $2 million in losses caused by *Tomato infectious chlorosis virus* (TICV) (Wisler *et al.*, 1998).

1.2 An overview of the family *Closteroviridae*

1.2.1 Taxonomy

The family *Closteroviridae* comprises more than 35 filamentous plant viruses (Table 1.2), containing the largest and the most complex positive-stranded plant RNA viruses with genomes up to 20 kb in size (Koonin & Dolja, 1993; Karasev, 2000). Members of the family *Closteroviridae* possess highly flexuous, thread-like long particles (1,200 to 2,000 nm) (Fig.1.1) and induce distinct cytopathological effects (membranous vesicles) exclusively in the phloem of infected plants (Karasev, 2000). Bearing these two characteristic features in mind, closteroviruses were initially divided into two subgroups according to particle length: long types with particles ranging from 1,200 to 2,000 nm, and short types with particles ranging from 700 to 800 nm. As some important properties

have been added, and new virus members have been reported over the last 20 years, the taxonomy within the family *Closteroviridae* has evolved and now includes three genera: the genus *Closterovirus* with aphid-transmitted monopartite members, the genus *Crinivirus* with whitefly-transmitted bipartite members and the genus *Ampelovirus* with mealybug-transmitted monopartite members (Table 1.1) (Karasev, 2000; Martelli *et al.,* 2002).

Table 1.1 Taxonomy of the family *Closteroviridae*

Genus	Type species	Number of particles	Number of virus species	Tentative Species	Vectors
Closterovirus	*Beet yellows virus* (BYV)	1	8	4	aphids
Crinivirus	*Lettuce infectious yellows virus* (LIYV)	2	7	3	mealybugs
Ampelovirus	*Grapevine leafroll-associated virus 3* (GLRaV-3)	1	6	5	whiteflies

Table 1.2 List of virus members of the family *Closteroviridae*

Virus	Abbreviation	Genome size (kbp)	References
Species in the genus *Closterovirus*			
Beet yellows virus	(BYV)	~15,5	Agranovsky *et al.*, 1994
Beet yellow stunt virus	(BYSV)	~10,5	Duffus, 1972
Burdock yellows virus	(BuYV)		Martelli *et al*., 2002
Carnation necrotic fleck virus	(CNFV)		Inouye *et al*., 1973
Citrus tristeza virus	(CTV)	19,3	Shepherd *et al.*, 1976
Wheat yellow leaf virus	(WYLV)		Inouye et al.,1973
Carrot yellow leaf virus	(CYLV)		Agranovsky *et al*., 1994
Clover yellows virus	(CYV)		Pringle, 1996
Dendrobium vein necrosis virus	(DVNV)		Fauquet & Martelli, 1995
Heracleum virus 6	(HV-6)		Bern *et al*.,1979
Festuca necrosis virus	(FNV)		Schmidt *et al*., 1963
Festuca necrosis virus	(FNV)		Schmidt *et al*., 1963

Virus	Abbreviation	Genome size (kbp)	References
Species in the genus *Crinivirus*			
Lettuce infectious yellows virus	(LIYV)	~8,1 + ~7,2	Duffus *et al.*, 1986
Abutilon yellows virus	(AYV)		Duffus *et al.*, 1987
Cucurbit yellow stunting disorder virus	(CYSDV)	~9,1 + ~7,9	Aguilar *at al*., 2003
Tomato chlorosis virus	(ToCV)	~8,6 + ~8,2	Wisler *et al.*, 1998
Strawberry pallidosis-associated virus	(SPaV)	8+ 7.9	Tzanetakis *et al*., 2004a
Tomato infectious chlorosis virus	(TICV)		Duffus *et al*., 1996
Sweet potato chlorotic stunting virus	(SPCSV)	~ 9,4 + ~ 8,2	Cohen *et al*., 1992
Lettuce chlorosis virus	(LCV)		Duffus *et al.*, 1996
Potato yellow vein virus	(PYVV)	~8 + ~5,3 + ~3,9	Salazar *et al.*, 2000;
Cucumber yellows virus	(CYV)	~7,9 + ~7,6	Hartono *et al.*, 2003
Diodea vein chlorosis virus	(DVCV)		
Species in the genus *Ampelovirus*			
Grapevine leafroll-associated virus 1	(GLRaV-1)		Zee *et al*., 1987
Grapevine leafroll-associated virus 2	(GLRaV-2)	~16,5	Zee *et al*., 1987
Grapevine leafroll-associated virus 3	(GLRaV-3)	~17,9	Zee *et al*., 1987
Grapevine leafroll-associated virus 5	(GLRaV-5)		Zee *et al*., 1987
Pineapple mealybug wilt-associated virus 1	(PMWaV-1)	*~10,7*	
Pineapple mealybug wilt-associated virus 2	(PMWaV-2)	*~14,9*	
Little cherry virus 2	(LChV-2)	~15	
Sugarcane mild mosaic virus	(S MMV)		
Tulip severe mosaic virus			
Grapevine leafroll-associated virus 4	(GLRaV-4)		
Grapevine leafroll-associated virus 6	(GLRaV-6)		
Grapevine leafroll-associated virus 8	(GLRaV-8)		
Plum bark necrosis and stem pitting-associated virus	(PBNSPaV)		
Unclassified *Closteroviridae*			
Grapevine leafroll-associated virus 7	(GLRaV7)		Morales & Monis., 2007
Little cherry virus 1			Jelkmann *et al*., 1997
Mint vein banding virus			Tzanetakis *et al*., 2005

1.2.2 Virus morphology and biology

1.2.2.1 Morphology

Members of the family *Closteroviridae* possess long (up to 2000nm), thread-like, flexuous filamentous virions with a chacteristic "rattlesnake" structure morphology due to the presence of a cluster of structural proteins at one tip of the virion (Fig. 1.1) (Agranovsky *et al*., 1995; Karasev, 2000; *Dolja et al*., 2006). Specifically for BYV, the major coat protein (CP) forms a virion body of helical symmetry that constitutes approximately 95% of the virion length and the short virion tail is assembled by five virus-encoded proteins (CP, minor capsid protein [CPm], heat shock protein 70 homologue [HSP70h], 64-kDa protein, and 20-kDa protein) and specifically covers the 5`-terminus of the genomic RNA (Zinovkin *et al*., 1999; Dolja *et al*., 2006).

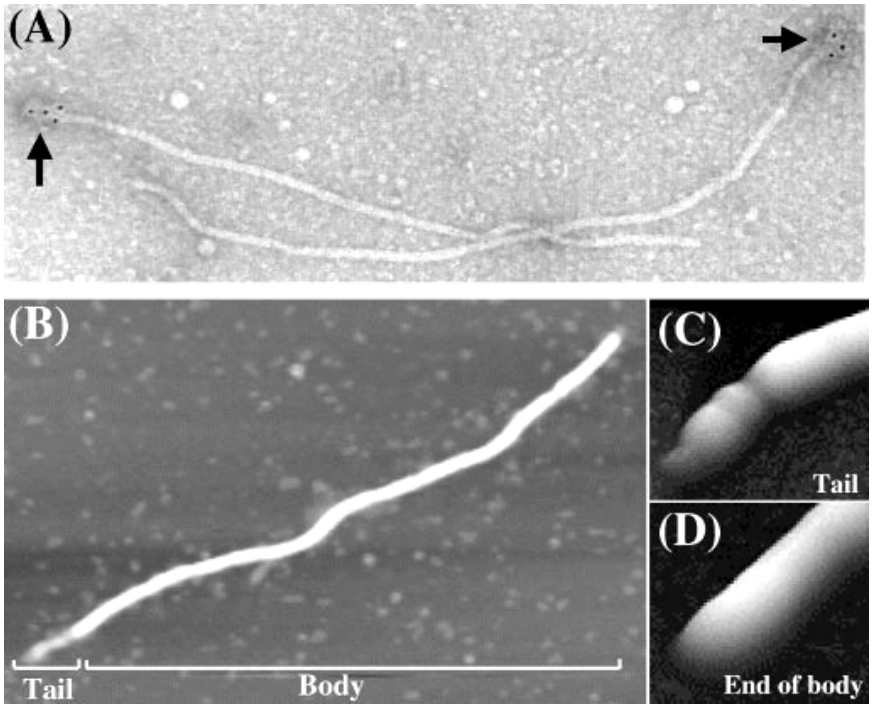


Figure 1.1 Morphology of the BYV virions. (A) Electron micrograph of two BYV virions. Virion tails marked by arrows were immunogold labeled using anti-CPm serum (courtesy of Alberto J. Napuli). (B) Atomic force microscopy image of a single BYV virion. (C and D) 3D reconstruction of the virion ends obtained using atomic force microscopy (Dolja, 2003).

1.2.2.2 Cytopathology

In infected plants, closteroviruses are located primarily in the phloem, where they form massive particle aggregates. Virions have been reported in the phloem parenchyma, mesophyll and epidermal cells depending on the host plant (Dolja *et al*., 1994). The association of closteroviruses with the phloem tissue, and in many cases their restriction in this compartment, results in low titers in infected plants and low virion yields during purification (Karasev, 2000).

Medina *et al*. (2003) has reported that the main cytopathological effects induced by criniviruses include the presence of virus-like particles in the sieve tubes and vascular parenchyma cells as scattered particles, or in companion cells as large masses forming cross-banded inclusions. Furthermore, Wisler and co-workers reported that in ToCV-induced infections, inclusion bodies and cytoplasmic vesicles are consistently observed and accumulated in the phloem tissue (Fig. 1.2 A, B & C). In later stages of infection, degeneration of the cytoplasm and organelles has been observed in companion and phloem parenchyma cells with viral particles aggregating into large bundles (Fig. 1.2 D) (Wisler *et al*., 1998). Interestingly, studies on LIYV have shown that infections of protoplasts with LIYV RNA1 alone, induce the formation of vesiculated inclusion bodies (Medina *et al*., 1998), while co-infection of LIYV RNA1 and 2 induces the production of conical plasmalemma deposits (PLDs) (Medina *et al*., 1998) distinguishing the presence of LIYV p26 and virions themselves (Medina *et al*., 2005).

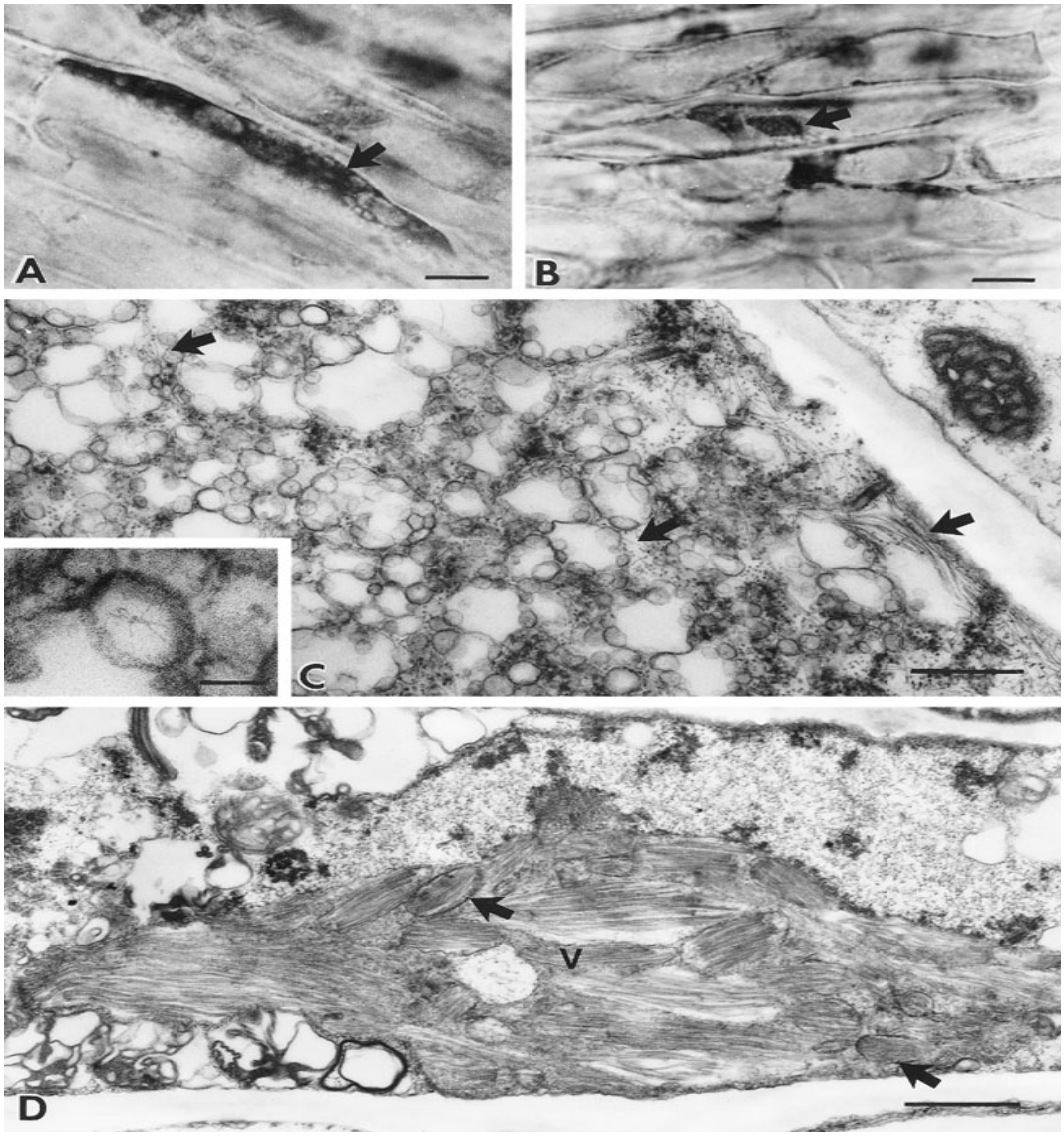


Figure 1.2 Light and electron microscopic analysis of ToCV-infected *Nicotiana clevelandii* tissue. A and B, Freehand sections through vascular tissues were stained with azure A. Arrows indicate A, vacuolate and B, pluglike viral inclusion bodies in phloem tissue. C, Cytoplasmic vesicles in a phloem companion cell sampled at 14 days post-inoculation. Virus particles (arrows) are interspersed with vesicles. Many vesicles contain a reticulate arrangement of fine fibrils. D, Phloem companion cell was sampled at 24 days postinoculation. A large aggregate of virus particles (V) is observed in the cytoplasm. Note various orientations of particles in discrete organized bundles within the large aggregate. Groups of particles appear to be surrounded by membranes (arrows). Cellular contents exhibit extreme degradation (Wisler *et al*., 1998).

1.2.2.3 Genome organization

The size of closterovirus genomes varies from ~15.5 to ~19.5 kb with a coding capacity of 10-14 proteins (Karasev, 2000). The RNA genomes of closteroviruses contain two conserved gene blocks. The first gene block consists of ORFs 1a and 1b and is involved in replication. The product of the 5`-terminal ORF 1a contains the papain-like leader proteinase (L-Pro), RNA methyltransferase (MET) and RNA helicase (HEL) domains, while ORF 1b encodes for the RNA-dependent RNA polymerase (RdRp) (Karasev, 2000). Some closteroviruses (e.g. CTV) possess a tandem of the leader proteinases that probably evolved via gene duplication. One of these proteinases (L1) can functionally substitute for a single L-Pro by enhancing virus RNA amplification. When L1 and L2 are expressed together, they act synergistically to provide for even higher levels of viral RNA (Fig. 1.3) (Dolja *et al*., 2006).

The second characteristic gene block of closterovirus genomes encodes products involved in encapsidation and cell to cell movement: a small hydrophobic protein (p6), the HSP70h, a ~ 60 kDa protein (p60), and two structural proteins CP and CPm (Dolja *et al*., 2006). Even though the second gene block contains universally conserved ORFs within the family, the order and size of the ORFs encoding for the capsid proteins does not. For example, in the genera *Ampelovirus* and *Crinivirus*, the order of CPm and CP is reversed and the CPm is much larger than in the genus *Closterovirus* (Klassen *et al*., 1995; Kreuze *et al*., 2002). Other examples of second gene block variation include duplication and divergence of the CPm gene in the *Grapevine leafroll-associated virus-1*, reshuffling of the ORFs within the second gene block in *Little cherry virus-1* and occurrence of a few enigmatic additional ORFs in criniviruses (e.g. p9 and p26) and certain ampeloviruses (Fig. 1.3) (Dolja *et al*., 2006).

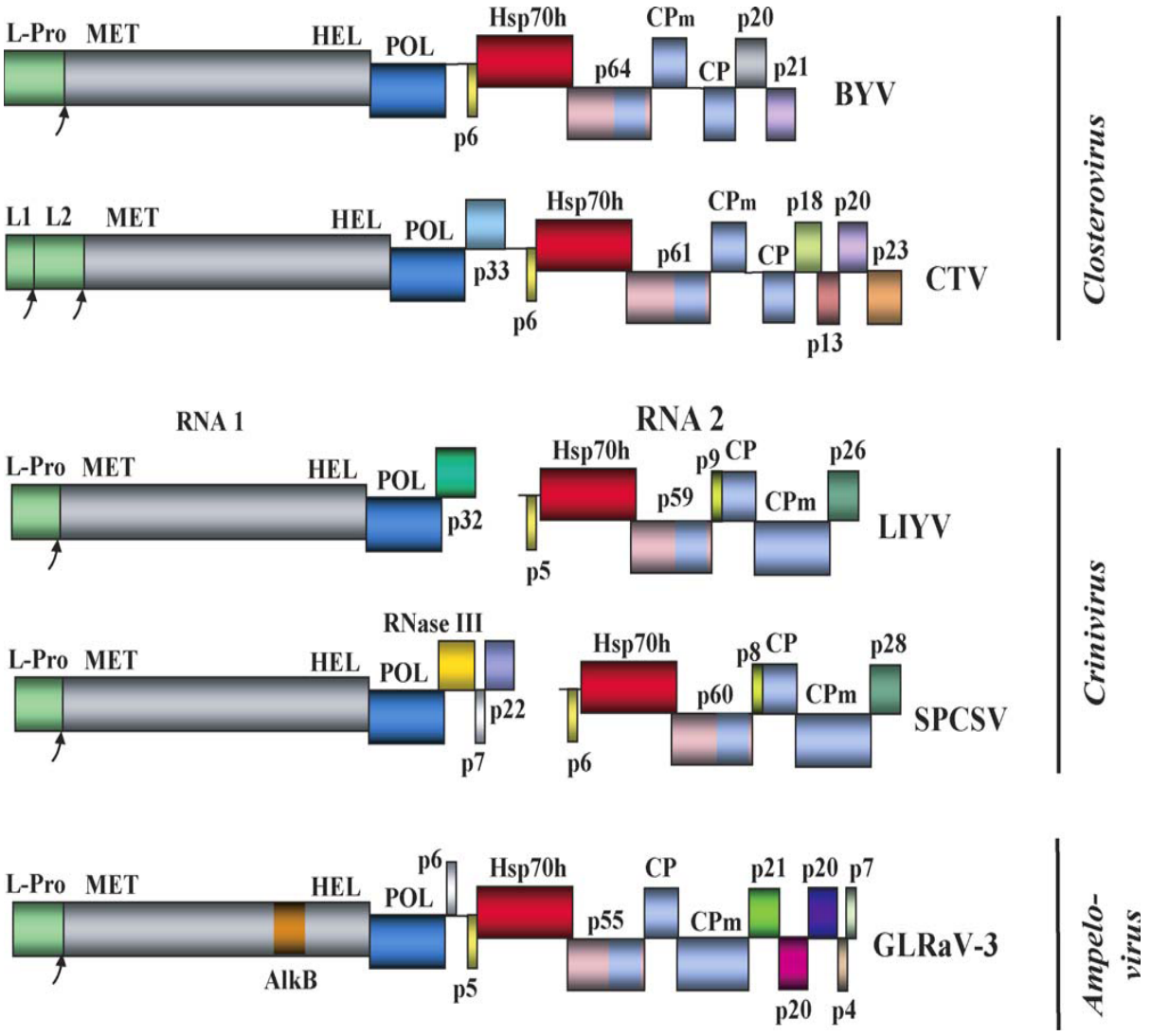


Figure 1.3 Genome maps of the selected representatives of the family Closteroviridae. The genera names are shown on the right. BYV (Agranovsky *et al*., 1994); CTV (Karasev *et al*., 1995); LIYV (Klaassen *et al*., 1995); SPCSV (Kreuze *et al*., 2002); GLRaV-3 (Ling *et al*., 2004). L1 and L2: tandem of the leader proteinases in CTV. The unique proteins in each genome are shown in colors as dissimilar as possible with their approximate moleular weight (Dolja *et al*., 2006).

1.2.2.4 Genome expression strategy

Genome amplification of RNA viruses requires two fundamental processes, viral RNA translation to produce virus-specific proteins (including the viral replicase) and RNA replication itself (Revers *et al*., 1999). Since viruses replicate in cellular environments that are dedicated to translation of monocistronic mRNAs, all ORFs except the most 5' proximal ORF in a viral positive-strand ssRNA will have a disadvantage to initiate translation. Therefore, closteroviruses utilize several mechanisms to regulate their gene expression: i) translational frameshifting ii) polyprotein processing, and iii) 3`-co-terminal subgenomic (sg) RNAs (Fig. 1.4) (Karasev, 2000; Dolja *et al*., 2006).

Closterovirus RdRp is encoded by a downstream ORF 1b that is expressed via a +1 translational frameshift (Fig. 1.4) (Agranovsky *et al*., 1994). Thus, translation of the genomic RNA results in two large polyproteins, one spanning MET-HEL, and the other encompassing MET-HEL-RdRp domains. This second, larger polyprotein is produced in much smaller quantities due to the low frequency of frameshifting (Dolja *et al*., 2006).

Another feature of closterovirus expression strategy is the production of a multiple (up to ten) nested set of up to ten 3`-coterminal subgenomic (sg) RNAs (Fig. 1.3). Usually, these sgRNAs are functionally monocistronic, each expressing only one protein from the 5`-proximal ORF. For example, the direct expression of the BYV ORFs 2-8 is via synthesis of a 3`-coterminal nested set of subgenomic messenger RNAs (Fig. 1.4) (Reed *et al.*, 2003). An analogous set of sgRNAs corresponding to ORFs 2 to 10 has been observed in plants infected with CTV (Gowda *et al*., 2006).

The proteins resulting from translation of ORFs 1a and 1b are involved in virus replication (Karasev, 2000). Nevertheless, efficient BYV and LIYV RNA amplification requires expression of p21 and p32 proteins respectively, which are both expressed to high levels early in infection (Yeh *et al*., 2000). BYV p21 (Reed *et al*., 2003) and the CTV-encoded proteins (p23, p20, and CP) have been found to suppress post-

transcriptional gene silencing (PTGS). The p23 and CP function intracellularly and intercellularly, respectively, whereas p20 acts both ways (Lu *et al*., 2004). Remarkably for SPCSV, it has been reported that the p22 and an RNase type III protein synergistically suppress post-transcriptional gene silencing (Kreuze *et al*., 2005).

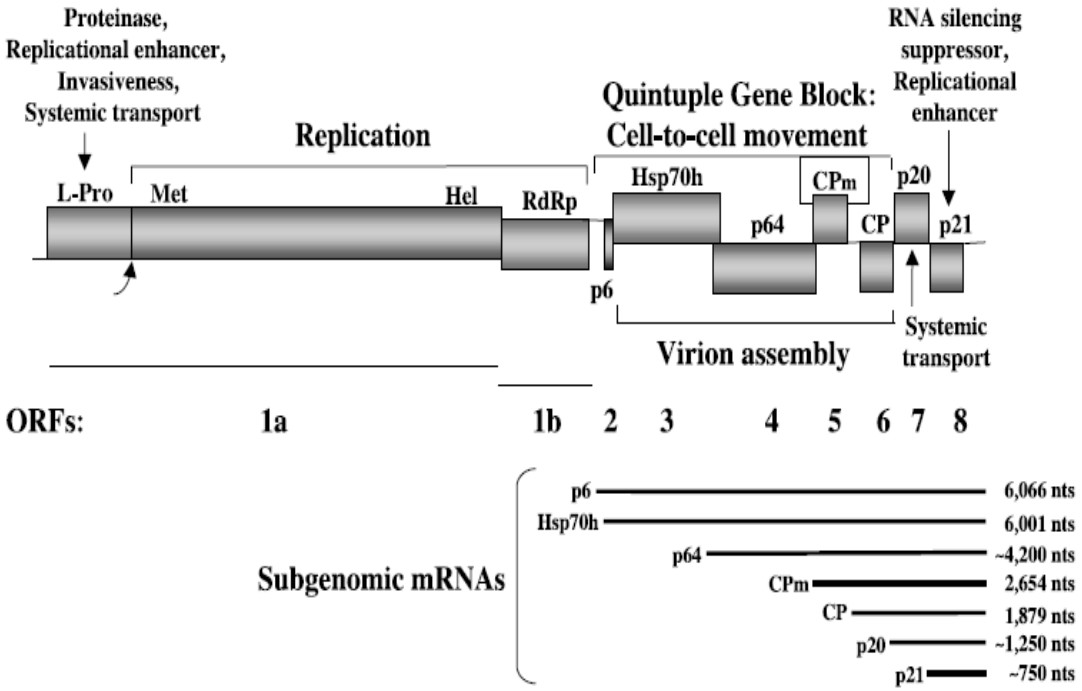


Figure 1.4 Map of the BYV genome. The ORFs are shown as boxes with the numbering shown below the diagram. The lines correspond to non-coding regions. L-Pro, leader proteinase; Met, Hel, and RdRp, methyltransferase, RNA helicase, and RNA-dependent RNA polymerase domains of the replicase, respectively; p6, a 6-kDa protein; Hsp70h, a Hsp70-homologue; p64, a 64-kDa protein; CPm and CP, the minor and major capsid proteins, respectively; p20 and p21, the 20-kDa and 21-kDa proteins, respectively. The functions of the encoded proteins are indicated. The subgenomic mRNAs are shown at the bottom. The proteins encoded by each sgRNA are indicated at the left, whereas the sizes of these RNAs are shown at the right. Because the exact 5`-termini of three sgRNAs are not known, their lengths were determined arbitrarily (Dolja, 2003).

1.2.2.5 Cell-to-cell movement and assembly

Systemic infection in infected plants occurs when the virus is able to move, after a genome amplification step, from the primary infection focus to invade distal regions of the plant (Lucas and Gilbertson, 1994; Revers *et al*., 1999). This requires that the infectious unit should move locally from cell-to-cell through the plasmodesmata, and then over longer distances through the phloem. The virion body of a closterovirus is self-assembled from viral RNA and CP, and is required for genome protection, whereas the tail is considered a device for cell-to-cell movement and in the case of LIYV a transmission determinant (Tian *et al*., 1999). The tail of BYV virion is formed by CPm, HSP70h, a 64-kDa protein and a 20-kDa protein, and attached to the body with the aid of HSP70h (Alzhanova *et al*., 2001). By attaching HSP70h to cytoskeleton, virions are considered to be chaperoned towards plasmodesmata and the subsequent virion translocation through the channel supposed to involve mechanical force generated by HSP70h (Peremyslov *et al*., 2004; Alzhanova *et al*., 2001; Dolja *et al*., 2006). Relocalization of HSP70h to plasmodesmata may trigger tail disassembly and destabilize the body. As a result, a short region of virion RNA may be exposed to interact with ribosomes and start a new cycle of infection (Alzhanova *et al*., 2001).

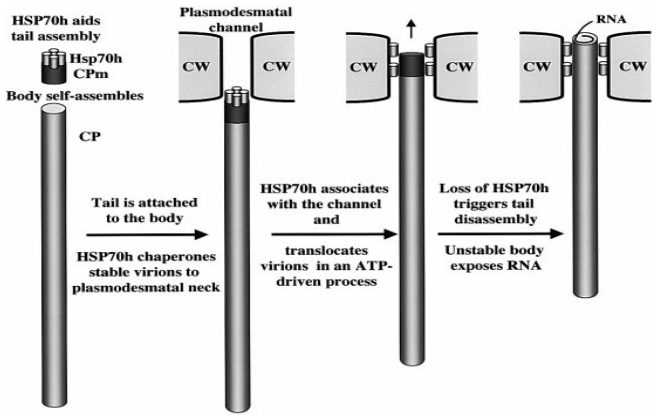


Figure 1.5 Hypothetical model of BYV assembly and cell-to-cell movement. CW, cell wall separating adjacent plant cells (Alzhanova *et al*, 2001).

1.2.2.6 Transmission

Insect vectors transmitting closteroviruses all belong to the order *Homoptera*, suborder *Sternorrhyncha*, and have been found in three families, *Aphididae* (aphids, transmitting members of the genus *Closterovirus*), *Aleyrodidae* (whiteflies, transmitting members of the genus *Crinivirus*), and *Pseudococcidae* (mealybugs, transmitting members of the genus *Ampelovirus*). These insects prefer feeding on a phloem tissue, and their piercing-sucking mouthparts generally cause little damage to the plant. Closteroviruses are transmitted in a semipersistent manner, with a minimum acquisition period of 0.5–1 h, and may retain infectivity in an insect for up to 9 days, depending on a virus-insect combination. Mechanical transmission is possible for some closteroviruses, such as BYV, CTV, and GLRaV-2, although it is difficult to achieve (Karasev, 2000). CPm and overall virion capsid integrity are required for the whitefly transmission of LIYV (Tian *et al*., 1999; Ng and Falk, 2006).

1.2.3 Genus *Crinivirus*

The complete nucleotide sequence of several crinivirus genomic RNAs have been reported over the last decade including: *Blackberry yellow vein virus* (BYVaV; Tzanetakis *et al*., 2006), *Cucumber yellows virus* (CuYV; Hartono *et al*., 2003), *Cucurbit yellow stunting disorder virus* (CYSDV; Aguilar *et al*., 2003), *Lettuce infectious yellows virus* (LIYV; Klaassen *et al*., 1995), *Strawberry psallidosis*-associated *virus* (SPaV; Tzanetakis *et al*., 2004b), *Sweet potato chlorotic stunt virus* (SPCSV; Kreuze *et al*., 2002) and *Tomato chlorosis virus* (ToCV; Wintermantel *et al*., 2005; Lozano *et al*., 2006; Lozano *et al*., 2007). Members of the genus *Crinivirus* possess 5`-capped (Kreuze *et al*., 2002) positive-sense bi-partite RNA genomes with a size ranging from 15000 to 19000 nt long (Martelli *et al*., 2002). Recent data suggests that *Potato yellow vein virus* (PYVV) from Peru and Colombia (Salazar *et al*., 2000) consists of a tri-partite genome (Livieratos *et al*., 2004).

Members of the genus *Crinivirus* have some unique features: i) the number of ORFs downstream the RdRp varies amongst the different species of the genus from zero to three (Fig. 1.6), and ii) in RNA2, all criniviruses possess two ORFs that encode proteins (p9 and p26) which remarkably share no significant homology between them or any protein in the database (Klaassen *et al.*, 1995). Furthermore, in LIYV protoplast experiments, it was shown that RNA1 accumulates earlier (24 hrs post-inoculation) in infection than those of RNA2 (48 hrs) and LIYV RNA1 is replication-competent in the absence of LIYV RNA2, while expression of p32 is essential for LIYV RNA2 replication (Yeh *et al.*, 2000). On the other hand, LIYV RNA2 and its encoded proteins do not affect LIYV RNA1 replication (Yeh *et al.*, 2000).

SPCSV is the only known RNA virus encoding a RNase III and uses two independent proteins, cooperatively for RNA silencing suppression. Data has shown that RNase III enhances the suppression of RNA silencing mediated by p22 (Kreuze *et al.*, 2004).

It has been predicted for all the fully characterized members of the genus *Crinivirus* and some members of the genera *Closterovirus* and *Ampelovirus* that the 3`-UTRs contain conserved secondary structures that include four stem-loops followed by a pseudoknot structure (Livieratos *et al.*, 2004). The 5` ends of the genomic and sgRNAs are capped in infected plants (Kreuze *et al.*, 2002).

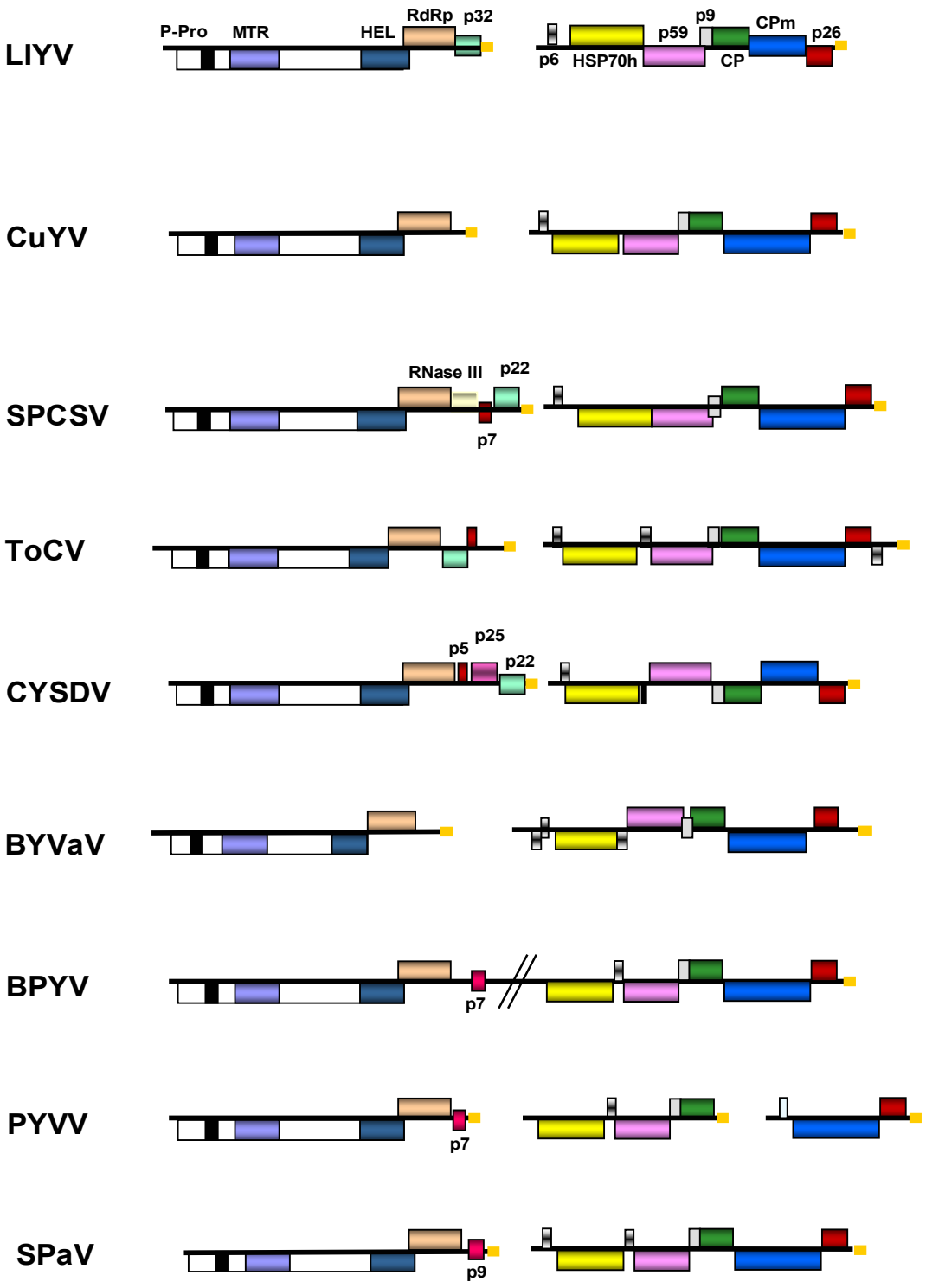


Figure 1.6 Genome organization for the members of the genus *Crinivirus*.

1.2.4 ToCV

1.2.4.1 Introduction-Symptomatology

ToCV, a member of the genus *Crinivirus* (family *Closteroviridae*), was discovered initially in diseased tomato plants and has since been identified as a serious problem for tomato crops in several parts of the world, and particularly in the United States, Europe and southeastern Asia (Wintermantel *et al*., 2005). Yellowing symptoms caused by ToCV include irregular chlorotic mottling that develops first on lower leaves, gradually advancing toward the growing point. Red or brown necrotic flecks and significant yield reduction occurs due to a loss of photosynthetic area (Fig. 1.8) (Wisler *et al*., 1998).

1.2.4.2 Virus transmission

ToCV is transmitted in a semi-persistent manner by four whitefly species belonging to two genera: *Bemisia tabaci* Gennadius, biotypes A and B (*B. argentifolii*), *Trialeurodes abutilonea* Haldeman and *T. vaporariorum* Westwood (Wisler *et al*., 1998; Wintermantel *et al*., 2005). The efficiency of ToCV transmission varies between the vectors, with *T. abutilonea* being the most efficient (Wisler *et al*., 2001). ToCV is persisting up to 6 days in its vector *T. abutilonea*, the longest among all the other whitefly vectors. There is no virus transmission 6 days after virus acquisition, while the transmission efficiency continues 2 days after acquisition and starts decreasing by day 3 (Wintermantel *et al*., 2006).

1.2.4.3 Genome organization and virus replication

The complete nucleotide sequence of American and Spanish ToCV isolates have been determined (Wintermantel *et al*., 2005; Lozano *et al*., 2006; Lozano *et al*., 2007). ToCV RNA1 of the American isolate is 8595 nucleotides long and contains four ORFs (Fig 1.7). RNA1 contains a 302- and 207-nucleotide long at 5`- and 3`- untranslated

region respectively and ToCV ORF 1 contains two overlapping ORFs (ORF1a and ORF1b). ORF 1a encodes a 221 kDa multifunctional protein, which contains PRO, MET and HEL motifs, while ORF lb encodes the 59 kDa RdRp most likely through a +1 translational frameshift. ORF 2 and ORF 3 encode two putative proteins (p22 and p5) with molecular weights of approximately 22 kDa and 5 kDa respectively, both of unknown function (Wisler *et al*., 1998; Wintermantel *et al*., 2005). ToCV RNA2 of the American isolate is 8247 nucleotides long and contains nine ORFs and a 238 and 218 nt long 5`-and 3`-untranslated region, respectively (Wintermantel *et al*., 2005). ToCV RNA2 ORFs potentially encode in a 5` to 3` direction for proteins of 4 kDa (ORF 1; hydrophobic protein), 65 kDa (ORF2; HSP70h), 8 kDa (ORF 3; protein of unknown function), 59 kDa (ORF 4; p59) that possibly associates with virion tails and is involved in virus cell-to-cell movement, 9 kDa protein of unknown function (ORF 5; p9), two coat proteins (CP and CPm), a 27 kDa protein (p27) and a 7 kDa protein (p7), which is a unique protein in terms of amino acid homology and location among all sequenced criniviruses (Wisler *et al*., 1998; Wintermantel *et al*., 2005).

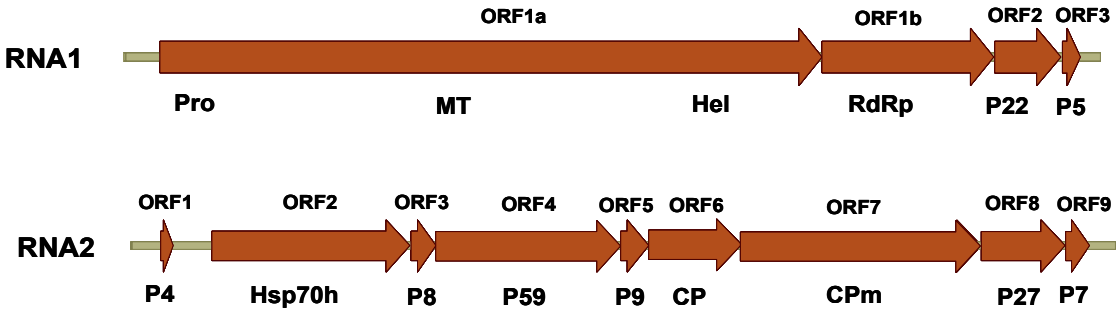


Figure 1.7 ToCV genome organization. The names of putative proteins are listed under each open reading frame (ORF). Acronyms: CP, coat protein; CPm, minor coat protein; Hsp70h, Heat shock protein 70 homolog; Hel, Helicase; MT, Methyl transferase; Pro, Proteinase, RdRp, RNA dependent RNA polymerase.

1.3 Detection and diagnosis

ToCV-induced infections in greenhouse tomato cultivations are prominent in the Mediterranean area and the United States, where reliable diagnostic tests are required to routinely and specifically detect the virus. ToCV and TICV are showing identical symptoms that cannot be easily distinguished in infected tomato plants. In principle, methods currently used for virus diagnosis are based on virus properties such as biology, viral proteins and nucleic acid. Presently, ToCV and TICV can readily be distinguished based on molecular detection as well as vector specificity (Wisler *et al*., 1998).

Figure 1.8 (A) Leaf symptoms of tomato infectious chlorosis virus (TICV) infecting field-grown tomatoes in Orange County, California. (B) TICV infecting China aster showing typical interveinal yellowing. (C) Tomato chlorosis virus (ToCV) infecting greenhouse-grown tomatoes in north central Florida (Wisler *et al*., 1998).

1.3.1 Detection based on virus biology

In principle, the use of an indicator host plant may provide routine diagnosis providing the development of clear, characteristic and consistent symptoms. Differences in the symptomatology of ToCV and TICV are seen mainly on two indicator plants: *Nicotiana benthamiana* and *N. clevelandii*. Both species show interveinal yellowing

when infected with either virus, but only TICV causes necrotic flecking (Wisler *et al*., 1998). Another diagnostic host includes *N. glutinosa*, which can be used to distinguish between these two criniviruses because it develops a severe interveinal chlorosis when infected by ToCV exclusively (Wintermantel *et al*., 2006).

1.3.2 Detection based on properties of the virus proteins

Depending on the availability of a specific antiserum, tissue-print hybridization and enzyme-linked immunosorbent assay (ELISA) provide simple, sensitive and fast diagnosis, allowing simultaneous analysis of multiple virus-infected plant samples. Polyclonal antibodies have been used to detect SPaV in petiole tissue blots (Tzanetakis *et al*., 2004) and CYSDV (Livieratos *et al*., 1999; Hourani & Abou-Jawdah., 2003). A polyclonal antiserum has been prepared against purified TICV virions but in indirect ELISA assays, it failed to distinguish between TICV and ToCV (Wisler *et al*., 1998).

1.3.3 Detection based on properties of the viral nucleic acid

The reverse transcriptase-polymerase chain reaction (RT-PCR) method has been used successfully to detect very low levels (20ng per 10µg of total plant RNA extracted) in CYSDV-infected cucurbit leaves (Marco *et al*., 2003). RT-PCR has also been used to detect *Blackberry yellow vein virus* (BYVaV) in blackberry leaves using degenerated oligonucleotide primers designed against conserved motifs of crinivirus genomes (Martin *et al*., 2003). Another useful, sensitive, and cost-saving method is the multiplex RT-PCR which has been developed to detect simultaneously TICV and ToCV in a large number of samples (Dovas *et al*., 2002). In this technique, degenerate primers are used to amplify part of the HSP70h gene of ToCV and TICV, followed by nested PCR using two pairs of specific oligonucleotide primers, one for each virus, in one step.

1.4 Infectious viral cDNA clones

Infectious cDNA clones are DNAs that when *in vitro*-transcribe, their RNA copies corresponding to the genomes of RNA viruses have the ability to initiate infection and replicate causing identical symptoms to the virus they derive from. In general, the construction of a cDNA involves reverse transcription of the viral RNA into a single stranded DNA using a primer hybridizing specifically to the 3`end of the viral genome and conversion of the single-stranded DNA into a double-stranded DNA molecule (usually by PCR) using an oligonucleotide primer corresponding to the 5` end viral RNA nucleotides. Also, for the design of a full-length cDNA clone from which infectious RNAs are expected to be produced (*in vitro*), the choice of the RNA polymerase promoter is of critical importance because it directly and highly affects the yield of transcripts and the nucleotide sequence at their extremities (Boyer & Haenni, 1994). In a procedure firstly described by Weiland and Dreher (1989) and is now widely used for its simplicity and convenience, oligonucleotides containing a promoter directly linked to the 5` end of the viral sequence. cDNA synthesis, cloning strategy, and the design of sequences bordering the viral insert can have a strong influence on the infectivity of the resulting transcripts. It is also known that the presence of nonviral nucleotides at the 5` end of viral transcripts strongly reduces infectivity (Boyer & Haenni, 1994). The nucleotide sequence of the 3` end of *in vitro* transcribed full-length viral RNAs is commonly dictated by the position of the restriction site used for "run-off" transcription (Boyer & Haenni, 1994).

Infection of host plants with infectious clones is usually carried out by mechanical inoculation but can also be achieved through expression of infectious transcripts *via* transgenic plants (Yamaya *et al*., 1988) or agroinfection (i.e., infection of plants by infiltirating leaves with an *Agrobacterium tumefaciens* suspension harbouring a plasmid containing a full-length copy of the genome of the virus under study) (Leiser *et al*., 1992; Liu & Lomonossoff, 2002).

Another problem is the high instability of full-length cDNA clones in bacteria. For instance, for two related flaviviruses, yellow fever virus (YFV; *Rice et al*., 1989) and Japanese encephalitis virus (JEV; Sumiyoshi *et al*., 1992) full-length cDNA clones were successfully constructed but none of the corresponding *in vitro* transcripts was infectious. In these cases, the only way to circumvent this problem was to transcribe directly from *in vitro*-ligated fragments. Although the reasons for this are not clearly understood, this example might reflect the potential toxicity and/or instability of some viral sequences in bacteria, it also provides a possible explanation of why no infectious transcripts have been obtained to date for some important viruses. In some cases, similar problems could also be tackled by alteration of the *Esherichia coli* strain and/or the DNA vector (Boyer & Haenni, 1994).

1.5 The scope of the investigation

ToCV is a whitefly-transmitted crinivirus inducing yellowing diseases on tomato crops of great economic importance all over the Mediterranean and the US. There is currently a significant necessity for a reliable and sensitive diagnostic methodology to identify ToCV-induced infections. Moreover, the availability of infectious cDNA clones is currently restricted to the prototype member of the genus *Crinivirus*, LIYV and there is currently a great need to produce full and infectious cDNA clones for other members of the genus *Crinivirus* too. This will provide further analysis of the replication strategies in plant or protoplast systems, particularly bearing in mind some distinguishing features of their genomic organization (variable number of ORFs downstream the polymerase, existence of conserved 3`-UTR secondary structures, etc.). Taking these in mind, the aims of this thesis will be:

a. to elucidate the complete nucleotide sequence of ToCV RNA of a Greek isolate and compare it with previously reported sequences from two isolates from the US and Spain in an attempt to position it phylogenetically in relation to the other two isolates;

b. to carry out a RT-PCR detection assay and compare results with available data from dot hybridization assays in terms of sensitivity; and

c. to construct a full cDNA clone for each of the two genomic RNA molecules of ToCV in order to be tested for infectivity in a protoplast system.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Sterilization and precautions

All plastic, glassware and solutions were autoclaved at 121^{o}C for 20 min, 151lb/sq.in. on liquid cycle. The antibiotics were filter-sterilized and stored at -20^{o}C.

2.2 Bacterial strains

The following *Escherichia coli* strains were used:

XL1-Blue Genotype *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1* lac [F′ *proAB lacI*^{q}*ZΔM15* Tn10 (Tetr)] is the strain of choice for preparation of high-quality plasmid DNA, which allows blue-white color screening and single-strand rescue of phagemid DNA. It contains an antibiotic-resistant F' episome, eliminating time-consuming selection on minimal media plates. It is also available in a wide variety of transformation efficiencies.

JM109 Genotype: *e14-(McrA-) recA1 endA1 gyrA96 thi-1 hsdR17*(rK- mK+) *supE44 relA1 Δ(lac-proAB)* [F` *traD36 proAB lacI*^{q}*ZΔM15]* cells are endonuclease (*end*A) deficient, greatly improving the quality of miniprep DNA, and are recombination (*rec*A) deficient, improving insert stability. The *hsd*R mutation prevents the cleavage of cloned DNA by the *Eco*K endonuclease system. JM109 cells contain the lacIqZΔM15 gene on the F′ episome, allowing blue-white screening for recombinant plasmids.

2.3 Bacterial media and antibiotics

***Luria-Bertani* (LB) medium:**

1% (w/v) NaCl

1% (w/v) Bacto-tryptone

0.5 (w/v) Yeast extract

For preparation of solid LB medium, the above mixture was supplemented with agar to a concentration of 1.5% (w/v) after adjusting the pH to 7. The medium was autoclaved under standard conditions (20 min, 1511b/sq.in. on liquid sycle). When the temperature of the media was 50-60^{o}C, ampicillin antibiotic was added as a selectable marker for plasmids, to a final working concentration of 100μg/ml. The media was immediately poured into plates and solidified plates were stored at 4ºC for a maximum period of two weeks.

2.4 Cloning vectors

The pGEM®-T Easy Vector System (Promega) is convenient for the cloning of PCR products. The vectors are prepared by cutting pGEM®-T Easy Vectors with a blunt-ended restriction endonuclease (*EcoR*V) and adding a 3` terminal thymidine to both ends (Figures 2.1). These single 3` T overhangs at the insertion site greatly improve the efficiency of ligation of a PCR product into the plasmids by preventing recircularization of the vector and providing a compatible overhang for PCR products (generated using a nonproofreading DNA polymerase) with 5` A overhangs (Robles and Doers, 1994). pGEM®-T Easy Vectors contain T7 and SP6 RNA Polymerase promoters flanking a multiple cloning site (MCS) within the peptide coding region of the enzyme beta galactosidase.

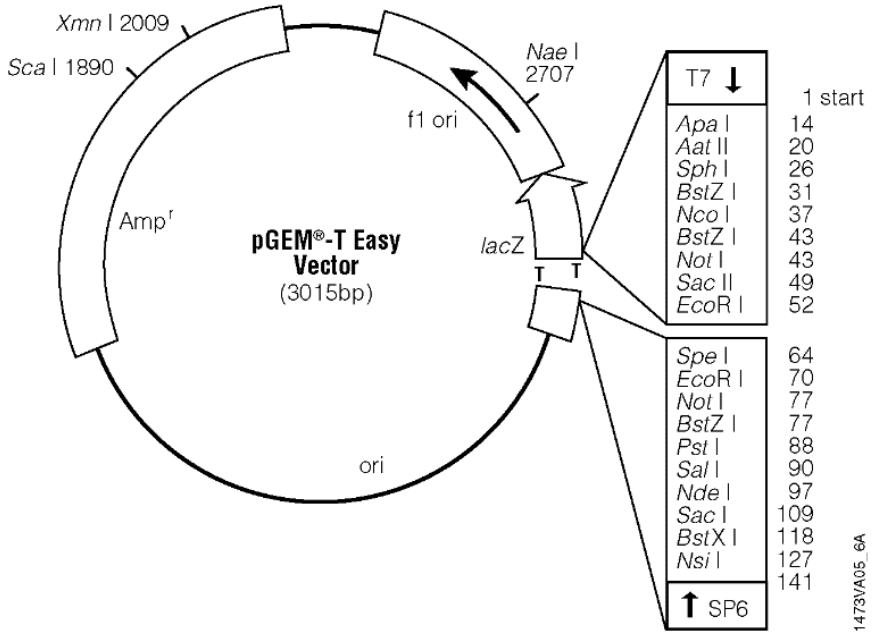


Figure 2.1 Schematic representation of pGEM-T Easy vector (Promega). The T7 and SP6 RNA polymerase promoters flank multiple cloning sites and two 3`-terminal thymidine (T) overhangs. The vector carries a single ampicillin resistance site (amp^r).

pUC19 is a small, high-copy number *E. coli* plasmid cloning vector. It contains the pMB1 origin of replication from pBR322 but lacks the *rop* gene and carries a point mutation in the RNAII transcript (G 2975 in pBR322 to A 1308 in pUC19). These changes together result in a temperature-dependent copy number of about 75 per cell at 37°C and >200 per cell at 42°C. The multiple cloning site (MCS) is in frame with the lacZα gene, allowing screening for insertions using α-complementation (Yanisch-Perron *et al*., 1985).

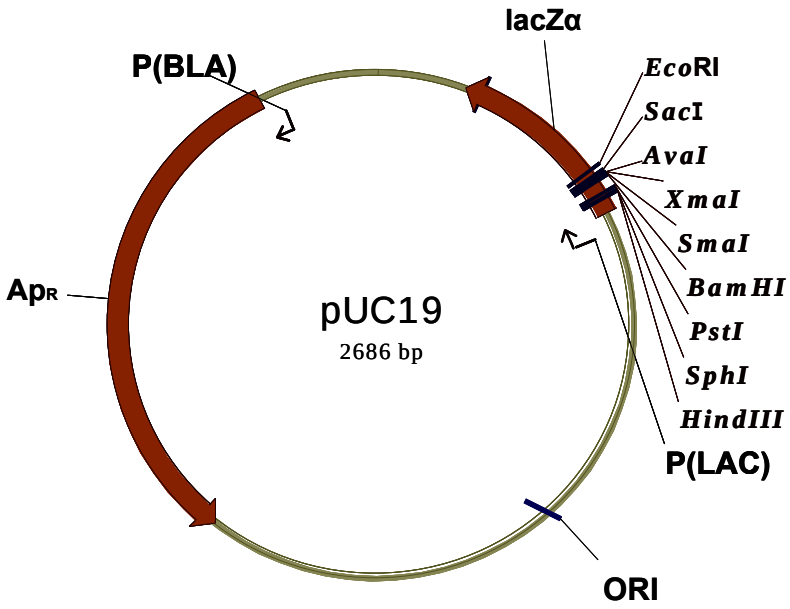


Figure 2.2 puc19 vector circle map showing the multiple cloning site, LAC repressor binding site, *BLA* gene which codes for a signal peptide and lacZ gene encoding the N-terminal fragment of beta-galactosidase. The vector carries a single ampicillin resistance site (Ap^R).

2.5 Oligonucleotides

Specific oligonucleotide primers based on the nucleotide sequences of the American isolate of ToCV reported by Wintermantel *et al*. (2005) were designed for RT-PCR. The underlined stretch of sequences represents restriction sites that were incorporated in order to facilitate subcloning into expression vectors.

Table 2.1 RNA1 specific oligonucleotide primers. The underlined nucleotides (nt) are for restriction enzymes, while the nt in italics are for the T7 promoter.

Name	Sequence (5`-3`)
1. BamHI- F/T7	**GGATCC***TAATACGACTCACTATA*GAAATA GTATTCGTGTGATTACA
1. BamHI- F	**GGATCC**GAAATAGTATTCGTGTGATTACA
2. NcoI-R	**GAGCTC**AATGTTGACGTCAGAGTAAAGT
3. NcoI-F	CTCGAGGAAATCGAGAGAATAGA
4. AvrII-R	**GAGCTC**TAATGTCCGATCAAATCATCAA
5. AvrII-F	GCGCCTGAAAAACCCATCAATCT
6. AfiII-R	**GAGCTC**CTGTGCAAGTTTCAGTTTTATA
7. AfiII-F	TGCCCATAGACACATTTCGAGGA
8. EcoRV-R	**GAGCTC**TAACAATAGATTGTACATCCGA
9. EcoRV-F	GTGCTAAAGAGATGAAGGGGCT
10. MluI-R	**GAGCTC**ATTCTAACTCTGTAGGGAGTTA
11. MluI-F	TGAGACGGTCTCTCCAATAGTTA
12. NdeI-R	**GAGCTC**ATCAGTATCCATCTTACCACCGT
13. NdeI-F	AGATCATACATAAGCTCAGGAAG
14. SmaI,XmaI-R	**GAGCTC**AAACAATATAGTTCTGTGCAAA
15. SmaI,XmaI-F	ACTTTTTCGAGACCTGCCTCAT
16. PstI-R	**GAGCTC**CAATCTCCAAAGTGTGATATGA
16b. EcoR1/R	**GAGCTC**CGGCACTGAATTCTCTATCATC
17. PstI-F	GTTATGTATTCCGGGATGAATCT
18. Sac-TCG-R	**GAGCTCTCG**CGACCTATTTATTTATATACTA

Table 2.2 RNA2 specific oligonuclotide primers. The underlined nt are for restriction enzymes, while the nt in italics are for the T7 promoter.

Name	Sequence (5`-3`)
19. Sac- F/T7	**GAGCTC***TAATACGACTCACTATA*GAAAT ACTAGTCCAGGTGTTT
19. Sac- F	**GAGCTC**GAAATACTAGTCCAGGTGTTT
20. SalI-R	**GCATGC**AGTTATGGAGAGACAAGTTGGA
21. SalI-F	GTCTTCCTCTCGCGTATTTAGA
21b.SalI-Fb	GTGCGTTTACGTGGTCAGATGG
22. AvrII-R	**GCATGC**TGTGACCTCCCTCTCGATCTTG
22b.AvrII-Rb	GCATGCAAATCTTGTGACCTCCCTCTCG
23. AvrII-F	CATCCGTTGATGGATTTGGGTT
24. BamHI-R	**GCATGC**TCAGTGACGAAAGGTAAATACT
25. BamHI-F	CCAATTGTCGAGTCCGGCCCAT
26. BstZ17I-R	**GCATGC**AAAGAAGTCGAAGCGCATATA
27. BstZ17I-F	TGACAGAGCGAAGCAAATGTT
28. Acc65I-R	CAGGTTCGAGATAAGTTGATCATCT
29. Acc65I-F	TATCAACTCCGGTCTTATTAAG
30. NsiI-R	**GCATGC**TTGTCGAGATCTTCGCTAAAGT
31. NsiI-F	ATGAAAGGAAGTTCAACGTTGG
32. KasI-R	**GCATGC**CAAGGATATTTAACCAGTGAAT
33. KasI-F	GAAGACTGATACTGGAGCGCAG
34. NruI-R	**GCATGC**GCAACATTATTGACATAAGGACA
35. MscI-F	**TGGCCA**AAGATAAGATAAGTGACGA
36. SmaI-R	**GCATGCTCG**CGACCTATTTATTTATATACTA

Table 2.3 Diagnostic RT-PCR oligonuclotide primers.

Name	Sequence (5`-3`)
TICVCPFor	AGGTCTTTCACAGTGGATTT
TICVCPRev	GTCCGAAACTGATTGAACCATCG
ToCVp22ForBamHI	GGATCCGATCTCACTGGTTGTTTGCGT
ToCVp22RevSalI	GGATCCGATCTCACTGGTTGTTTGCGT

2.6 Molecular biology techniques

2.6.1 Virus source and purification of total RNA

Tomato leaves with chlorotic symptoms were supplied from the area of Lakonia (Pelloponese) and freeze-dried to be used for RT-PCR amplification and elucidation of the complete viral genome. Symptomatic and asymptomatic tomato plants were collected from the area of Platanos (western Crete) to be used in diagnostic RT-PCR.

Freeze-dried ToCV-infected tomato samples (100mg) were chopped into fresh pieces (approximately 5mm^2), placed in liquid nitrogen and ground briefly in a pre-cooled pestle and mortar. Total plant RNA was isolated using TRIZOL reagent (Invitrogen) according to the manufacturer's instructions. Briefly, the powder-ground plant tissue was mixed with 1ml of TRIZOL reagent per 50-100 mg of tissue. The mixture was incubated for 5 min at room temperature. About 0.2 ml of chloroform per 1ml of TRIZOL reagent was added and mixed for 15 s. The tubes were incubated at room temperature for 2 to 3 min and centrifuged at 12000g for 15 min at 4^{o}C. The top aqueous phase was transferred to a new eppendorf tube and the RNA was precipitated with the addition of 0.5 ml isopropyl alcohol per 1ml of TRIZOL reagent used for the initial homogenization. The samples were incubated for 10 min at room temperature and then centrifuged at 12000g for 10 min at 4^{o}C. After centrifugation, the pellet was washed with 75% ethanol and air-dried. The pellet was then resuspended in 20μl of RNase-free water, incubated for 10 min at 60^{o}C and the eluted RNA was stored at -80ºC.

2.6.2 Reverse transcription (RT) from total RNA extracts

To generate cDNA from ToCV total RNA extraction, prepare in a sterile tube: 10µl total RNA extraction (10ng-5µg) and 1µl antisense sequence-specific primer (100pmol/µl), and incubate the mix at 70°C for 5 minutes and chill on ice. Add the following in the order indicated:

5X reaction buffer	10µl
10mM 4 dNTPs mix (1.0mM – final concentration)	2µl
Ribonuclease Inhibitor 20u	3µl
DEPC-treated water	up to 39µl
RevertAid™ H Minus M-MuLV Reverse Transcriptase (200 U/ µl, Fermentas)	1.5µl

After incubating for 60 min at 42°C, the reaction was terminated by incubation for 10 min in 70°C.

2.6.3 Polymerase Chain Reaction (PCR)

Following reverse transcription, gene fragments were amplified using 1/5 or 1/10 diluted cDNA products by polymerase chain reaction (PCR).

Materials

10x LA PCRTM Buffer II (Mg^{2+} free)	5µl
25 mM $MgC1_2$ (final 2.5 mM)	5µl
dNTP Mixture (2.5 mM each)	8µl
Sense oligonucleotide (100pmol/µl)	1µl
Antisense oligonucleotide (100pmo1/µl)	1µl
DNA Template 1/10 of cDNA template	5µl
DNA polymerase Takara LA Taq ^{TM} (5 units/µl)	0.5µl
Nuclease-free ddH_2O	up to 50µl

The DNA template was amplified according to the following program:

Initial denaturation: 95°C for 5 min	× 1 cycle
Denaturation: 95°C for 30 s	
Annealing: (52-62°C) for 40 s	× 35 cycles
Polymerization: 72°C for 2 min	
Polymerization: 72°C for 10 min	× 1 cycle

2.6.4 Agarose gel electrophoresis

Solutions

Running gel buffer: 50x TAE buffer was prepared by dissolving 24.2 % (w/v) Tris-base (2M) and 3.72% (w/v) EDTA (0.05M) in 800 ml ddH_2O. The pH was then adjusted to 8.0 with glacial acetic acid and the total volume was brought up to 1 litre with water. The TAE working solution was made from the stock solution by dilution, to a final concentration of 0.5x.

Method

Agarose gel electrophoresis was prepared by melting molecular grade agarose to a final concentration of 1% in 0.5x TAE buffer. Ethidium bromide was added to a final concentration of 0.2µg/ml, when the temperature was cooled down to 55°C. The gel was poured into the tray and left until it solidified. The tank was filled with 0.5x TAE solutions covering the top of the gel. The DNA samples were mixed with one tenth of the volume l0x loading buffer (1% SDS, 50% glycerol, and 0.05 bromophenol blue) before loading. The gel was run at 100V for at least 20 min and the DNA bands were visualized under UV light.

2.6.5 Purification of the PCR product following gel electrophoresis

In order to purify PCR amplified fragments, a 50µl PCR reaction was mixed with the appropriate volume of l0x loading buffer and loaded on a 1% agarose gel. The gel was run at 100V for 20-30 min in 0.5x TAE. At the end of the electrophoresis, the band corresponding to the desired product was excised from the gel using a clean scalpel with minimum exposure to UV light. The gel slice was weighted in a tube and 3 volumes of Buffer QX1 (*QJAEXII Gel Extraction Kit, QIAGEN*) were added to 1 volume of the excised gel. The mixture was incubated at 50°C for 10 minutes, or until the gel had dissolved completely. 12µl of QIAEX II was added and incubated for 20 min, mixing occasionally by vortexing to keep QIAEX II in suspension, following by centrifugation

for 1 min to pellet the resin. Pellets were washed with 500µl of buffer QX1, and twice with buffer PE. The pellets were then air-dried for 5-10 min and eluted with 12µl sterile ddH_2O. The elution was placed into a clean (fresh) 1.5ml tube and stored at -20°C. The presence of DNA was checked in a 1% agarose gel.

2.6.6 Cloning of PCR products

2.6.6.1 Cloning in plasmid vector

Following purification, PCR products were cloned into the pGEM-T Easy vector (Promega) according to the manufacturer's instructions. An approximate 5.5:1 molar ratio of the insert and vector was ligated together in a final reaction volume of 15 µl, which included 7.5 µl of 2×ligation buffer [60 mM Tris-HCl, pH 7.8; 20 mM $MgCl_2$, 20 mM DTT, 2 mM adenosine triphosphate (ATP), 10% polyethylene glycol], and 1 µl of T4 DNA ligase (Promega). The mixture was incubated at 4^oC overnight.

2.6.6.2 Plasmid DNA purification

A single recombinant colony was grown overnight at 37^oC in 5ml LB medium supplemented with 100µg/ml ampicillin. The isolation of plasmid DNA was done following the QIAprep Spin miniprep Kit Protocol. The overnight culture (3ml) was centrifuged at 13000rpm for 3 min at room temperature and the supernatant was removed. The pellet was resuspended in 250µl resuspension buffer (50mM Tris HCl pH 8.0, 10mM EDTA, 10µg/ml RNase A) and the cells were lysed by the addition of 250µl lysis buffer [200mM NaOH, 1% sodium dodecyl sulfate (SDS)] and inversion of the tube 4-6 times. When the suspension was viscous and clear, 350µl of neutralization buffer (buffer N3) was added and mixed. After centrifugation at 13000rpm for 10 min, the supernatant was collected and applied to the QIAprep spin column. The column was centrifuged for 60 s and the flow-through was discarded. The column was washed with 750µl of wash solution (PE buffer). The DNA was eluted with 50µl elution buffer (10mM Tris-Cl, pH 8.5) by centrifugation for 1 min.

2.6.6.3 Restriction endonuclease digestion of DNA

Restriction digestion of DNA fragments was carried out in the appropriate buffer and 5U of restriction enzyme per 500µg of DNA by incubation at 37^{o}C for 3-4 h. Digestions were checked by DNA electrophoresis on 1% agarose gel prior to any further manipulations.

Dephosphorylation of pUC19 DNA was carried out only once where the *EcoR*I-digested vector was first gel-purified and then incubated with shrimp alkaline phosphatase SAP (1 unit/µg DNA) (Promega) at 37°C for 15 minutes in 1X SAP reaction buffer (final volume of 30–50µl). The alkaline phosphatase was removed from the reaction by heating to 65°C for 15 minutes.

2.7 Bacterial transformation

2.7.1 Preparation of bacterial competent cells

A single colony of *E. coli* was grown overnight in 5ml of LB media in the absence of antibiotic at 37^{o}C with vigorous shaking (200-250 rpm). About 2ml of an *E. coli* culture was used to inoculate 200ml LB medium in a 1L sterile flask to an optical density (OD) of 0.35-0.4 for 3-4 h at 37^{o}C with constant rotation. The cells were immediately chilled on ice for 5 min, before centrifugation at 3000 rpm for 10 min at 4^{o}C. The bacterial pellet was resuspended in 10ml of ice-cold filter sterile transformation storage solution [TSS; 1 x TSS is LB broth containing 10% (wt/vol) polyethylene glycol (PEG 8000), 5% (vol/vol) dimethyl sulfoxide (DMSO), and 50mM $MgCl_2$, pH 6.5] by careful pipetting. Aliquots of 50µl were frozen in liquid nitrogen, and kept at -80^{o}C for future use.

2.7.2 Transformation

Half of the ligation mixture was added to 50-100µl of competent cells and gently mixed. The cells were placed on ice for 20 min and then heat-shocked for 50 s in a water bath at exactly 42^{o}C. The tube was returned immediately on ice for 2 min. After adding 500µl LB broth, the mixture was incubated at 37^{o}C with constant rotation (225 rpm) for 1 h.

The cells were quickly pelleted and 400µl of the supernatant was discarded. The bacterial pellet was resuspended with the rest of the supernatant (100µl) by gentle pipetting. Using a sterile spreader, the 100µl aliquot was spread onto LB/Amp (100mg/ml) plates (pre-warmed for 1 h at room temperature). The plate was incubated at 37^{o}C for 16h.

2.8 Phenol-chloroform extraction

Sterile water was added to a solution containing nucleic acids to bring its volume to 430µl. An equal volume of (pH 7.9) equilibrated phenol /chloroform was added and the mixture was vortexed for one minute. The mixture was then centrifuged for 5 min at 13000 rpm. The upper aqueous layer (approximately 400µl) was carefully removed and transferred to a new tube avoiding the phenol interface. If desired, the aqueous layer could also be subjected to two chloroform extractions to remove residual phenol.

2.9 Ethanol precipitation

Following phenol-chloroform extraction, the nucleic acids from the aqueous phase were precipitated with ethanol by adding 2.5 volumes of 99% ethanol and 1/10 of the volume 3M sodium acetate (pH 5.2). The solution was incubated at -80^{o}C for 2 h or at -20^{o}C overnight and centrifuged at maximum speed for 20 min. The supernatant was aspirated and the pellet was washed with 70% ethanol, centrifuged at maximum speed for 20 min, air-dried and resuspended in the appropriate volume of sterile water.

2.10 Nucleotide sequencing analysis

RT-amplified products were cloned into the pGEM-T Easy vector (Promega) and sequenced at least twice and on both orientations by the Sanger chain-termination method, using labelled primer cycle sequencing with Sequitherm EXCEL II DNA Sequencing Kit-LC (66cm gel) of Epicentre Biotechnologies on both orientations (Li-Cor Long Read IR2 4200). M13, T7 and SP6 promoter universal oligonucleotide (labeled with IRD700 or IRD800) primers were utilized. Sequence data were assembled and compared with databases using the BLAST server on the NCBI Web server [http://www.ncbi.nlm.nih.gov/BLAST], vector NTI program, CLUSTALX and TreeView programmes (Thompson, 1996). The secondary structure of the 3`-termini of both genomic ToCV RNA fragments were analyzed using the Mfold algorithm programme (Zuker & Mathews, 1999).

CHAPTER THREE

RESULTS

3.1 Diagnosis of ToCV using RT-PCR and dot-blot hybridization assays

A diagnostic RT-PCR was performed using serial dilutions (5µg, 2.5µg, 0.5µg, 100ng and 20ng) of total RNA preparations from symptomatic tomato plants. The complete ToCV p22 gene, which has no homology with any other gene in the database, was used as the target fragment using ToCV-p22For*BamH*I and ToCV-p22Rev*Sal*I oligonucleotide primers containing *BamH*I and *Sal*I restriction sites, respectively (Table 2.3, Materials and Methods). Specific oligonucleotide primers (TICV-CPFor and TICV-CPRev) for TICV CP were also designed (Table 2.3, Materials and Methods). Following RT-PCR amplification, DNA products were analyzed in a 1% agarose gel and a DNA fragment of approximately 650 nt in size was observed in all samples from symptomatic plants using ToCV-specific oligonucleotide primers (Fig. 3.1, lanes 1-5). No DNA products were amplified when total RNA extracts from asymptomatic plants or when TICV-specific oligonucleotide primers were used (Fig. 3.1, lanes 6-7, respectively). The highest dilution of total RNA to be used successfully for detection purposes was 20ng of total plant RNA (lane 5). The PCR-amplified DNA product was further cloned and sequenced verifying its viral origin. A smaller DNA product (~350nt) present in all dilutions of ToCV-infected plant material is of unknown origin and might represent a product of oligonucleotides mis-priming.

Comparison of RT-PCR with dot blot hybridization experiments available from a previous thesis showed that both RT-PCR and dot-blot hybridization assays seem to perform satisfactorily in terms of sensitivity. It was shown in dot blot hybridization experiments that the virus could well be detected with comparable sensitivity (20ng of total RNA extracts) using a DIG-labeled probe derived from ToCV p22 ORF and overnight film exposure (Fig. 3.2).

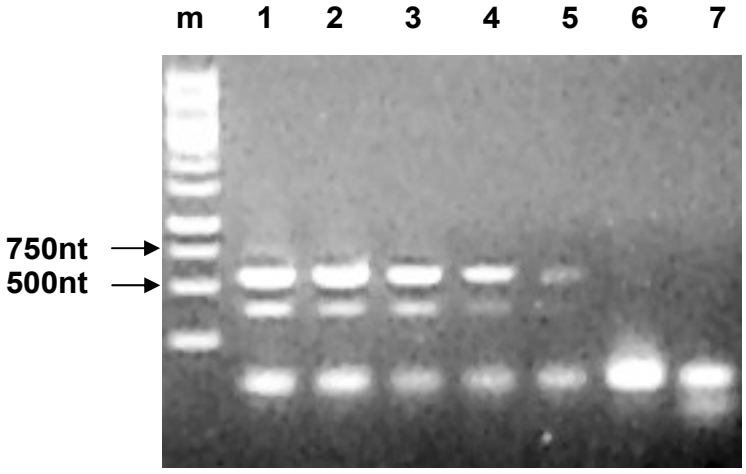


Figure 3.1 Detection of ToCV in extracts of infected tomato plants by RT-PCR using oligonucleotides ToCV-p22-F and ToCV-p22-R. 1% agarose gel electrophoresis of RT-PCR products using serial dilutions of total RNA extracts of ToCV-tomato plants 5mg (lane 1), 2.5mg (lane 2), 0.5mg (lane 3), 0.1mg (lane 4), 20ng (lane 5). In lane 6, TICV-specific oligonucleotide primers were used with 5mg of total RNA extracts from the same tomato plants. In lane 7, ToCV-specific oligonucleotides were used with total RNA extracts (5mg) from asymptomatic tomato plants. DNA molecular size marker is shown in lane m.

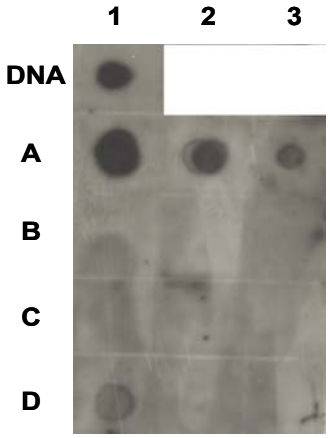


Figure 3.2 Dot blot hybridization using a DIG-labelled probe corresponding to the minus strand of the ToCV-p22 gene. Five-fold dilutions of total RNA (A & B) or plant (C & D) extracts from infected (A & D) and healthy (B & C) tomato plants were spotted on nitrocellulose membranes. For total RNA extracts, 0.5mg (lane 1), 0.1mg (lane 2), and 20ng (lane 3) were used. For plant sap, 10ml of 1/5, 1/25, 1/125 plant extract dilution (g/ml) in phosphate buffer was used. 20ng of a DNA clone of ToCV p22 was used as a positive control (top of the membrane).

3.2 Strategies for the construction of full ToCV RNA1 and RNA2 cDNA clones

In order to obtain the complete nucleotide sequence of the Greek ToCV isolate and to generate full-length cDNA clones of the virus, two strategies were devised for each one of the two (RNA1 and RNA2) molecules of the viral genome. Each strategy was modified and optimised so as to overcome difficulties and hindrances appearing during the implementation of the work.

.

The first strategy devised for ToCV RNA1 was based on a nine-fragment amplification and cloning approach (Fig. 3.3). It consisted of nine RT-PCR generated overlapping amplicons, covering ToCV RNA1 in its entirety depending on the presence of unique enzymatic restriction sites located on the ToCV genomic RNA1 molecule. Based on the location of these restriction sites, a pair of primers was designed for the amplification of each internal portion of the genome (Table 2.1, Materials and Methods). Furthermore, two oligonucleotide primers corresponding respectively to the 5`- and 3`-termini of ToCV RNA1, were designed for cloning purposes to each comprise a unique restriction site (*BamH*I and *Sac*I, respectively) which is absent from the ToCV RNA1 sequence and only found on the cloning vector (pUC19). Furthermore, the 5`-oligonucleotide primer (BamHI-F/T7) included the bacteriophage T7 RNA polymerase promoter sequence (5`-TAATACGACTCACTATA-3`) placed immediately downstream the restriction site and upstream the 5`-terminus of the RNA1 sequence (Fig. 3.4.A) allowing transcription initiation at the +1 position of the first nucleotide of ToCV RNA1. Restriction site *Nru*I was also included at the most 3`-end oligonucleotide (Sac-TCG-R) by adding three extra nucleotides (5`-CGA-3`) immediately downstream the viral genomic RNA1 sequence (5`-TCG-3`) and upstream the *Sac*I restriction site. According to this strategy, digestion of the full length construct by *Nru*I would linearize the plasmid in a position to facilitate the transcription up to the very last nucleotide of the viral genome towards the 3`-terminus of the viral genome (Fig. 3.4.B).

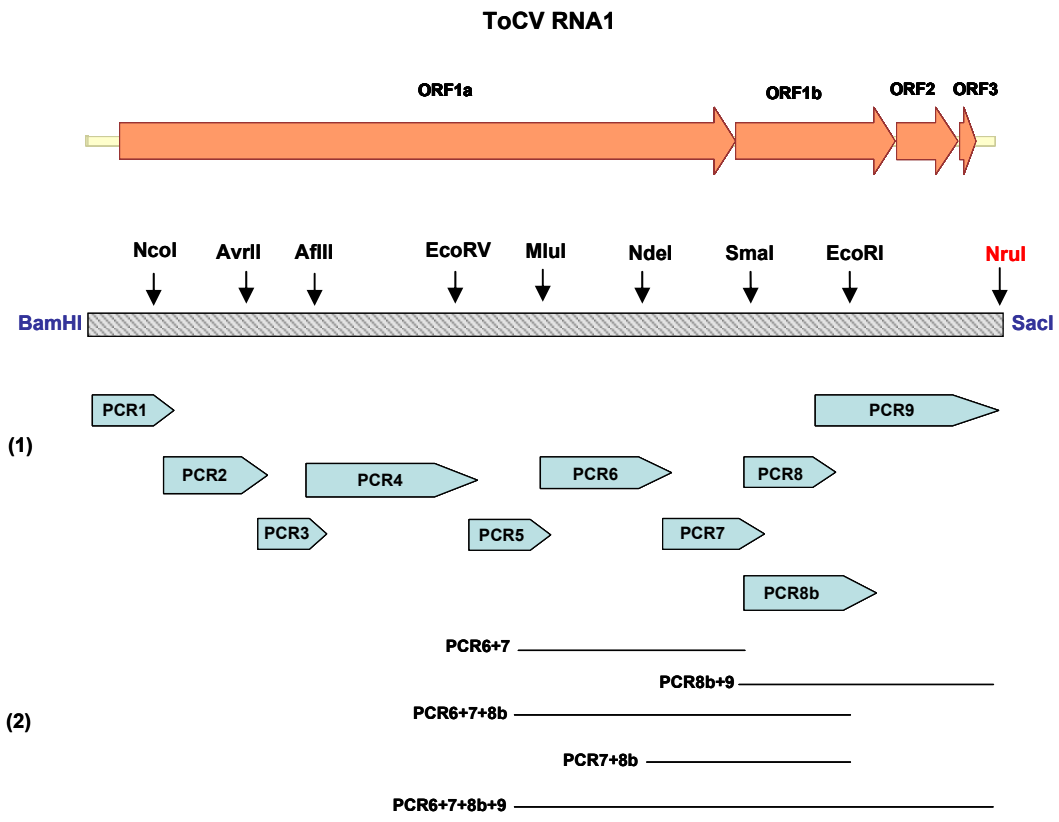


Figure 3.3 Strategy followed to construct full length cDNA version of ToCV RNA2. (1) Diagrammatic representation of the steps involved in the construction of the full length ToCV RNA1 cDNA clones flanked by the T7 RNA polymerase promoter at the 5` end. (2) Modified constructs to facilitate the formation of full cDNA of ToCV RNA1. Unique restriction sites used to assemble the clones and their approximate positions in the genome are shown.

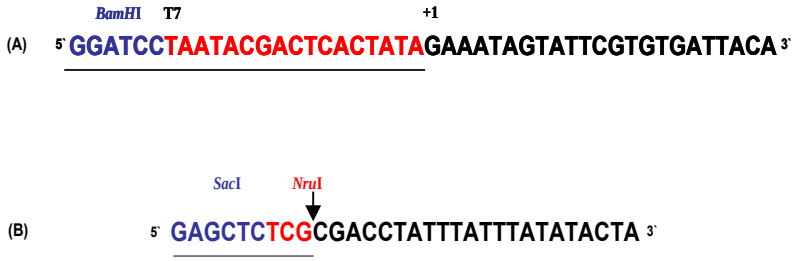


Figure 3.4 Oligonucleotides used at the 5`- and 3`-termini of ToCV RNA1 full cDNA. A, BamHI- F/T7 primer showing the added nt to generate *BamH*I and T7 promoter (in red), while the +1 base is the first base incorporated into RNA during transcription. B, Sac-TCG-R primer showing the underlined nt added to the 3` end primer to generate NruI and SacI at the 3` end of ToCV RNA1.

The second strategy devised for the construction of the full-length ToCV RNA2 cDNA clone (Figure 3.5) was analogous to the one for ToCV RNA1. The size and position of nine RT-PCR amplified fragments was selected depending on the presence of unique enzymatic restriction sites on the ToCV RNA2 molecule. For the amplification of each internal portion of the genome, oligonuclotide primers used for ToCV RNA2 are listed in Table 2.2, Materials and Methods. Two more oligonucleotide primers corresponding respectively to the 5`- and 3`-termini of ToCV RNA2 were designed to each include a restriction site (*Sac*I and *Sph*I, respectively) which is absent from the ToCV RNA2 sequence but found exclusively on the cloning vector (pUC19). Restriction site *Nru*I (5`-TCGCGA-3`), which was originally located on ToCV RNA2 (position 7929-7934 bp), was mutated (5`-TCGC**C**A-3`) by ligating the first three nucleotides of the *Nru*I site of PCR8 and the last three nucleotides of an artificial *MSc*I [5`-TGGCCA-3`] site created at the 5`-terminus of PCR9. This nucleotide mutation was incorporated in order to kill the *Nru*I site within the viral genome and at the same time to create a single *Nru*I site at the very 3`-terminus of the molecule (Fig. 3.6.B) and facilitate blunt end digestion and transcription of the precise viral molecule. This single nucleotide mutation would not alter the expression of the predicted amino acid of the p7 protein. Similarly to ToCV RNA1, the sequence of the bacteriophage T7 RNA polymerase promoter sequence was incorporated in the oligonucleotide primer immediately upstream of the 5`-terminus of the RNA2 sequence to allow transcription of the molecule at position +1 of the 5`-terminus of ToCV genomic RNA2 (Fig. 3.6.A).

The fragments were obtained by reverse transcription, while the PCR conditions initially followed the manufacturer's instructions of high-fidelity DNA polymerase (*Takara LA Taq*) tested, but were later optimised. These fragments were then cloned in the pGEM-T Easy vector, and each one of them was subcloned (from the 5`- to the 3`-end) into the pUC19 vector using in each case a unique restriction site found in the viral genome and one of the *Sac*I or *Sph*I at the 3`-terminus of the RNA1 and RNA2, respectively (Fig. 3.4.B and 3.6.B, respectively).

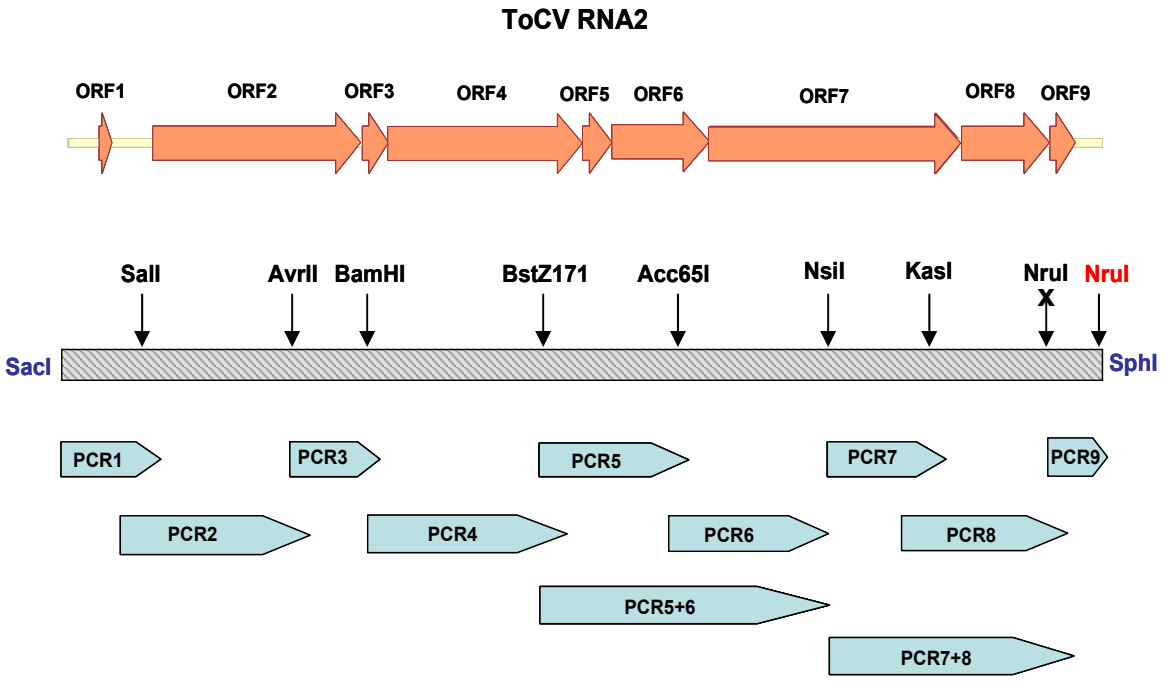


Figure 3.5 Strategy followed to construct full length cDNA version of ToCV RNA2. Diagrammatic representation of the steps involved in the construction of the full length ToCV RNA2 cDNA clones flanked by the T7 RNA polymerase promoter at the 5` end. The restriction sites used to assemble the clones and their approximate positions in the genome are shown as well.

Figure 3.6 Oligonucleotides used at the 5`- and 3`-termini of ToCV RNA2 full cDNA. A, Sac- F/T7 primer showing the underlined nt added to generate *Sac*I and T7 promoter (in red), while the +1 base is the first base incorporated into RNA during transcription. B, SmaI-R primer showing the underlined nt added to the 3` end primer to generate NruI and *Sph*I**.**

3.2.1 RT-PCR amplifications of ToCV RNA1

Total RNA extraction was isolated from ToCV-infected tomato plants, and used as a template for RT-PCR. Reverse transcription was carried out as described in 2.6.2 using M-MLV reverse transcriptase and DNA fragments were PCR-amplified using *TaKaRa LA Taq* polymerase according to the manufacturer's instructions. Standard amplification profiles consisted of initial denaturation at 94ºC for 5 min, followed by 35 cycles of denaturation at 94ºC for 30 s, annealing at 50-65ºC for 30 s (annealing temperature occasionally varied depending on primer melting point), and extension at 72ºC 1 to 2 min, with a final extension cycle of 72ºC for 10 min at the end of the reaction series.

Agarose gel electrophoresis showed that oligonucleotide primers designed on ToCV RNA1 (Table 2.1, Materials and Methods) BamHI-F and NcoI-R, NcoI-F and AvrII-R, AvrII-F and AfiII-R, AfiII-F and EcoRV-R, EcoRV-F and MluI-R, MluI-F and NdeI-R, NdeI-F and SmaI, XmaI-R, SmaI, XmaI-F and PstI-R, SmaI, XmaI-F and EcoR1/R, PstI-F and Sac-TCG-R generated PCR products of the expected size: PCR1 (758bp), PCR2 (985bp), PCR3 (568bp), PCR4 (1584bp), PCR5 (821bp), PCR6 (1347bp), PCR7 (941bp), PCR8 (768bp), PCR8b (1170bp) and PCR9 (1793bp) respectively (Fig.3.7). PCR-generated product T7/PCR1 was amplified using BamHI-F/T7 and NcoI-R oligonucleotide primers using the pGEM-T-easy-PCR1 (pG-PCR1) clone as a DNA template in order to include the T7 bacteriophage promoter (Fig. 3.7.G)

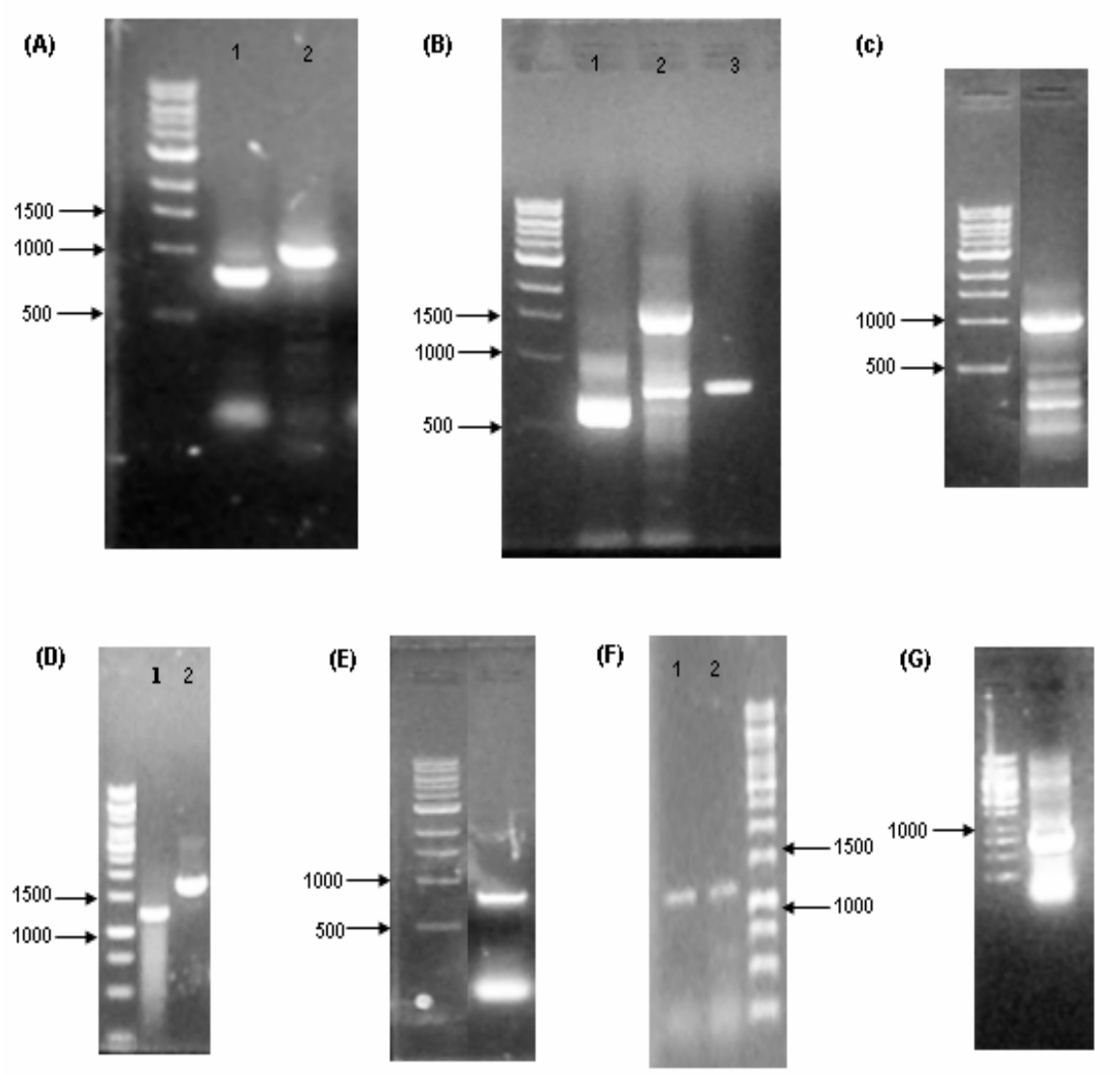


Figure 3.7 Gel electrophoresis of RT-PCR products representing ToCV RNA1, derived from total RNA extracts isolated from ToCV-infected tomato plants. (A) lane 1: PCR1, 758 bp; lane 2: PCR2, 985 bp (B) lane 1: PCR3, 586 bp; lane 2: PCR4, 1584 bp; lane 3: PCR5, 821 bp (C) PCR7, 941 bp (D) lane 1: PCR6, 1347bp; lane 2: PCR9, 1793 bp (E) lane 1:PCR8, 768 bp (F) lanes 1& 2: PCR8b, 1170 bp (G) T7/PCR1.

3.2.2 Cloning of ToCV RNA1 fragments into the pGEM-T Easy Vector

All RT-PCR generated products from RNA1 were gel-purified by the QIAEX II Gel Extraction Kit (QIAGEN) and cloned into the pGEM-T Easy vector to produce, respectively, pG-T7/PCR1, pG-PCR2, pG-PCR3, pG-PCR4, pG-PCR5, pG-PCR5, pG-PCR6, pG-PCR7, pG-PCR8, pG-PCR8b and pG-PCR9 recombinant plasmids. The insertion of the right size fragments initially checked by digestion with *EcoR*I (flanking the insertion site of the vector) (Fig. 3.8) where for each treatment, two bands were observed representing the pGEM-T Easy vector (3kbp) and the cloned insert. pG-PCR9 showed three bands (Fig. 3.8 lane 9) due to the presence of a *EcoR*I restriction site in PCR9 fragment. All recombinant plasmids were sequenced at least twice and on both orientations to allow the determination of the full sequence of the Greek ToCV RNA1.

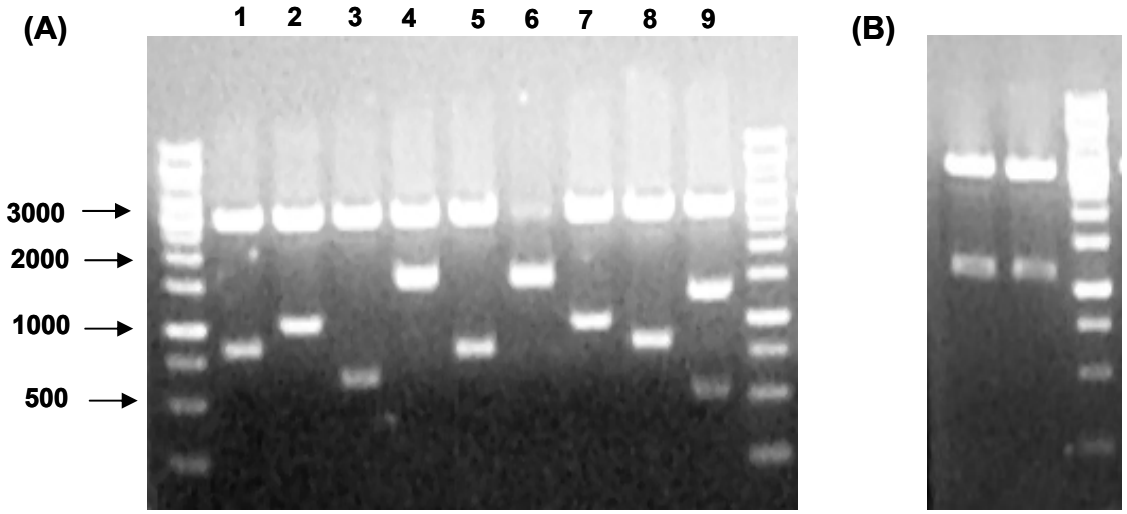


Figure 3.8 Digestions of the RNA1cloned genes into pGEM-TEasy by *EcoR*I; (A) lane 1: pG-T7/PCR1, lane 2: pG-PCR2, lane 3: pG-PCR3, lane 4: pG-PCR4, lane 5: pG- PCR5, lane 6: pG-PCR6, lane 7: pG-PCR7, lane 8: pG-PCR8, lane 9: pG-PCR9. (B) pG-PCR8b clones 1 & 2. Position of marker (base pair) is shown on the left.

In order to facilitate the nucleotide sequence of the internal part of PCR9 (1793bp), recombinant clone pG-PCR9 was digested using *Pst*I, *Hind*III and *Sac*I to result in two fragments (PCR9a: 773 bp and PCR9b: 910 bp) that were further cloned into similarly digested pUC19 plasmids to generate pU-PCR9a and pU-PCR9b clones (Fig. 3.9.C). Both these two clones were sequenced in both orientations.

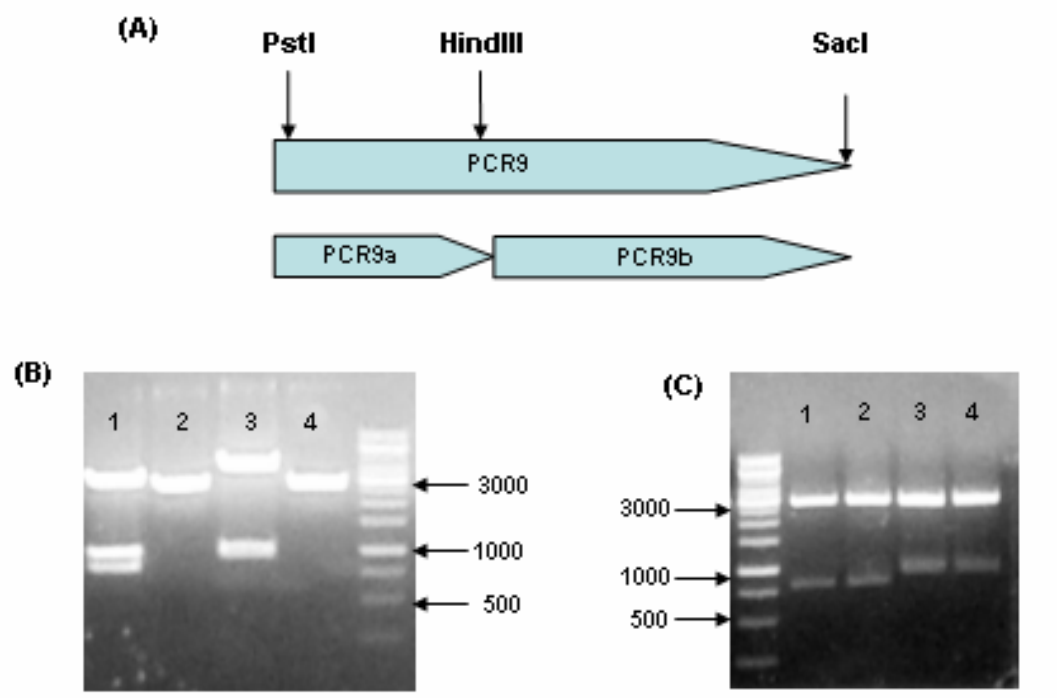


Figure 3.9 Digestions of pG-PCR9 (A) Diagrammatic representation of subcloning of excised fragments PCR9a and PCR9b. (B) Electrophoresis analysis of the restriction digestions of lanes 1 and 2: pG-PCR9 and pUC19 vector digested by *Pst*I and *Hind*III. Lanes 3 and 4: pG-PCR9 and pUC19 vector digested by *Hind*III and *Sac*I. (C) Digestions for checking the correct constructs, lanes 1 and 2: pU-PCR9a digested by *Pst*I and *Hind*III; lanes 3 and 4: pU-PCR9b digested by *Hind*III and *Sac*I.

3.2.3 RT-PCR amplification of ToCV RNA2

Agarose gel electrophoresis of the RT-PCR amplified products showed that oligonucleotide primers designed ToCV RNA2 (Table 2.2, Materials and Methods) SacI-F and SalI-R, SalI-F and AvrII-R, AvrII-F and BamHI-R, BamHI-F and BstZ17I-R, BstZ17I-F and Acc65I-R, Acc65I-F and NsiI-R, NsiI-F and KasI-R, KasI-F and NruI-R, MscI-F and SmaI-R, generated respectively PCR1 (858bp), PCR2 (1379bp), PCR3 (698bp), PCR4 (1511bp), PCR5 (1136bp), PCR6 (1225bp), PCR7 (850bp), PCR8 (1202bp), and PCR9 (313bp) (Fig. 3.10). In additional experiments, oligonucleotide primers BstZ17I-F and NsiI-R, NsiI-F and NruI-R amplified fragments PCR5+6 (2222bp) and PCR7+8 (1971bp) (Fig. 3.10.7, lanes 1 and 2, respectively). PCR-generated product T7/PCR1 was amplified using Sac-F/T7 and SalI-R oligonucleotide primers using the pGEM-T-easy-PCR1 (pG-PCR1) clone as a DNA template in order to include the T7 bacteriophage promoter (Fig. 3.10.8).

Particularly for PCR2, a number of modifications: i) increase of Mg^{2+} levels , ii) addition of 5-10% DMSO, and iii) increase of the annealing temperature to meet the requirements of the high melting temperature of the 3`-primer (67^{o}C), failed to generate a correct size DNA product. Only when oligonucleotide primer (AvrII-R) was used to reverse-transcribe a cDNA molecule followed by nested PCR using SalI-F and AvrII-Rb oligonucleotide primers, the correct size product was generated (Fig. 3.10.2).

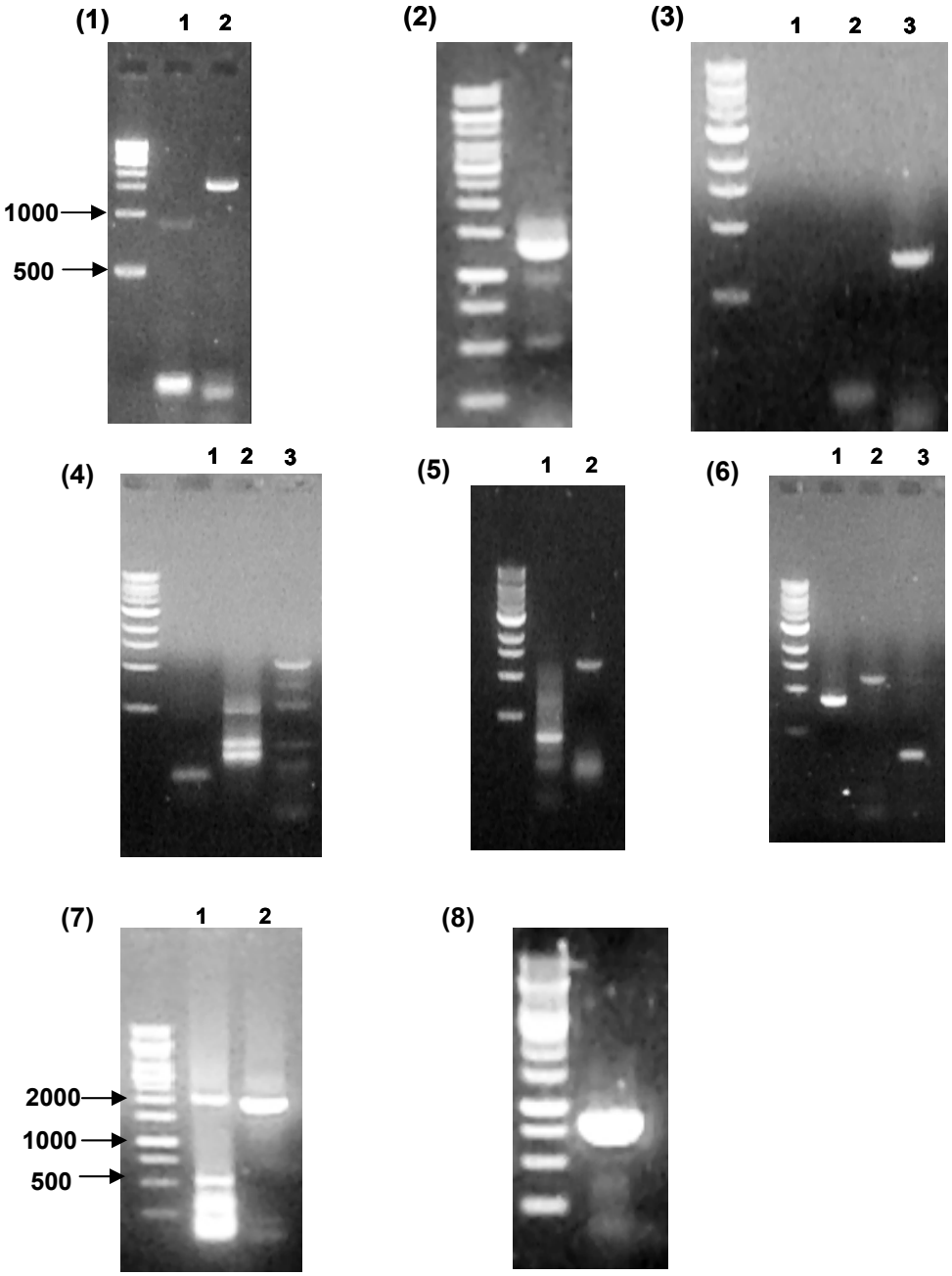


Figure 3.10 Gel electrophoresis of RT-PCR products representing ToCV TNA2. (1) lane 1: PCR1, 858 bp; lane 2: PCR4, 1511 bp (2) PCR2, 1379 bp (3) lane 3: PCR3, 689 bp (4) lane 3: PCR5, 1136 bp (5) lane 2: PCR6, 1225 bp (6) lane 1: PCR7, 850 bp; lane 2: PCR8, 1202 bp; lane 3: PCR9, 313 bp (7) lane 1: PCR5+6, 2222 bp; lane 2: PCR7+8, 1971 (8) T7/PCR1.

3.2.4 Cloning of ToCV RNA2 fragments into pGEM-T Easy Vector

Each of the ToCV RNA2 PCR-amplified products (T7/PCR1, PCR1, PCR2, PCR3, PCR4, PCR5, PCR6, PCR7, PCR8, PCR9, PCR5+6, PCR7+8) were cloned into the pGEM-T Easy vector to produce, respectively pG-T7/PCR1, pG-PCR2, pG-PCR3, pG-PCR4, pG-PCR5, pG-PCR5, pG-PCR6, pG-PCR7, pG-PCR8, pG-PCR8, pG-PCR9, pG-PCR5+6 and pG-PCR7+8. The recombinant plasmids were digested with *Not*I (flanking the insertion site of the vector), to verify the presence of the correct size fragments. Following digestion, two bands were observed per treatment representing the pGEM-T Easy vector (3kbp) and the insert (Fig. 3.11). The recombinant plasmids were sequenced at least twice on both orientations to allow the determination of the full sequence of the Greek ToCV RNA2.

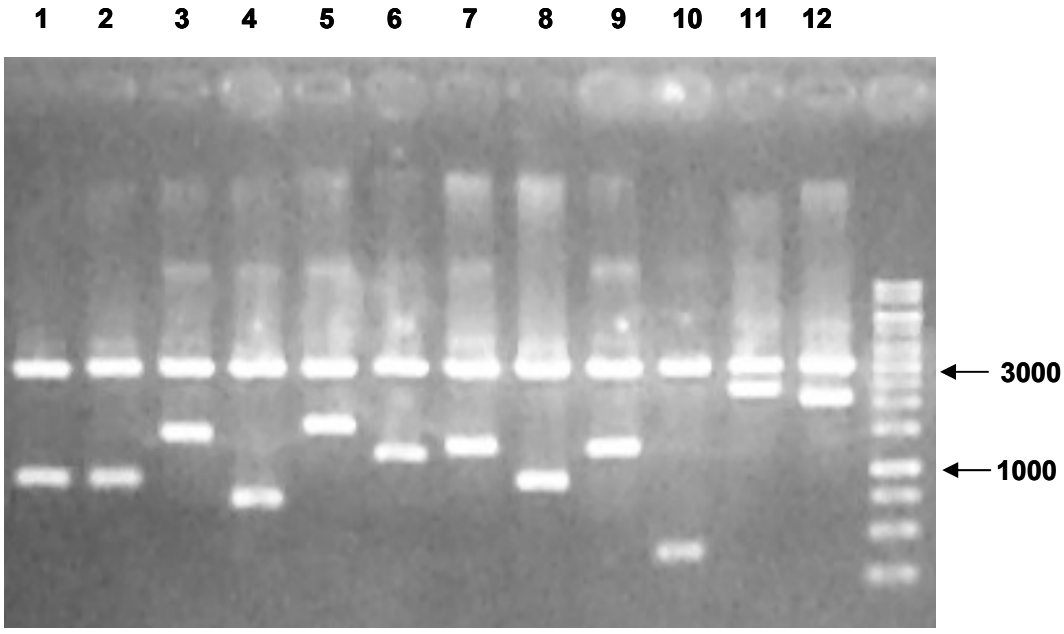


Figure 3.11 Electrophoresis analysis for the digestions of pGEM-T Easy cloned fragments by *Not*I; lane 1: pG-T7/PCR1, lane 2: pG-PCR1, lane 3: pG-PCR2, lane 4: pG-PCR3, lane 5: pG-PCR4, lane 6: pG-PCR5, lane 7: pG-PCR6, lane 8: pG-PCR7, lane 9: pG-PCR8, lane 10: pG-PCR9, lane 11: pG-PCR5+6, lane 12: pG-PCR7+8.

3.3 Sequence analysis

The complete nucleotide sequence of the complete bi-partite genome of the Greek ToCV isolate consists of 8594 and 8242 nt, for RNA1 and 2, respectively (Appendix 1). Using the Vector NTI and the AlignX program, percentages of nucleotide sequence identity were estimated and aligned between the American, Spanish and Greek isolates (Appendix 2). Greek ToCV isolate RNA1 exhibits high identity with the equivalent RNAs of the American and Spanish isolates (99.1% and 97.3%) and possesses 301 and 176 nt at the 5` and 3`-untranslated regions (UTRs) respectively. The 5`-UTR presents precisely the same number of nucleotides as the Spanish isolate and two less than the American due to a common deletion for both isolates (Fig.3.12a.A). Moreover, in another region of the 5`-UTR RNA1, both Spanish and Greek 5`-UTR have an extra nucleotide compared to the American isolate (Fig. 3.12a.B). ToCV RNA1 includes ORF1a (302-6139) and b (6141-7655), encoding the conserved closterovirus L-Pro, MTR, HEL (ORF1a) and RdRP (ORF1b) domains and ORF 2 (7662-8243) and 3 (8263-8418), putatively encoding for 22 and 5-kDa proteins respectively. As with other criniviruses, the region between the ToCV MTR and HEL domains was found to contain no significant similarities with any other known proteins and the RdRp was predicted to be expressed by a +1 translational frameshift, similar to other members of the family *Closteroviridae*. Several criniviruses have been reported to contain zero (CuYV, PYVV, SPaV, BYVaV), one (LIYV, ToCV) or two (SPCSV, CYSDV) ORFs similar (in terms of location and size) to the ToCV p22 ORF but with no significant nucleotide identity amongst each other. ToCV p5 possesses a central transmembrane domain between residues 20 and 42 (Fig. 3.12b) but is also of unknown function.

Analysis of the determined sequence of ToCV RNA2 of the Greek ToCV isolate also showed high identity with the equivalent RNAs of the American and Spanish isolates (99% and 97.6% sequence identity). ToCV RNA2 possesses 238 and 213 nt long 5` and 3`-UTRs. The 3`-UTR of both Greek and Spanish isolates has in common a 5 nt deletions in comparison with the American isolate (Fig.3.12a.C). Nevertheless, both

American and Greek isolates have 3 nt less (in the UTR region between ORF 1 and 2) than the Spanish isolate (Fig.3.12a.D). ToCV RNA2 includes nine ORFs: ORF 1 (nucleotides 239-340) encodes a small 4 kDa protein with a large hydrophobic domain between residues 10 and 32, suggesting it may function as a transmembrane protein. ORF 2 (733-2397) encodes HSP70h which, as with other closteroviruses, is likely associated with virion tails, virion assembly and cell to cell movement. This protein is conserved among criniviruses. ORF 3 (2407-2610) encodes an 8 kDa protein. ORF 4 (2562-4115) encodes a 59 kDa protein (p59) that also associates with virion tails and is likely involved in viral cell-to-cell movement. ORF 5 (4097-4333) encodes a 9 kDa protein for which no function has been identified. ORFs 6 (4333-5106) and 7 (5112-7121) encode the 29 kDa and 76 kDa CP and CPm, respectively. A 27 kDa protein (p27) is putatively encoded by ORF8 (7125-7823) downstream from the CPm, while ORF 9 (7832-8029) possesses a unique position to criniviruses putatively encoding a 7 kDa polypeptide with a transmembrane domain between residues 5 and 27 (Fig. 3.12b) with no homology found to any other proteins in the databases.

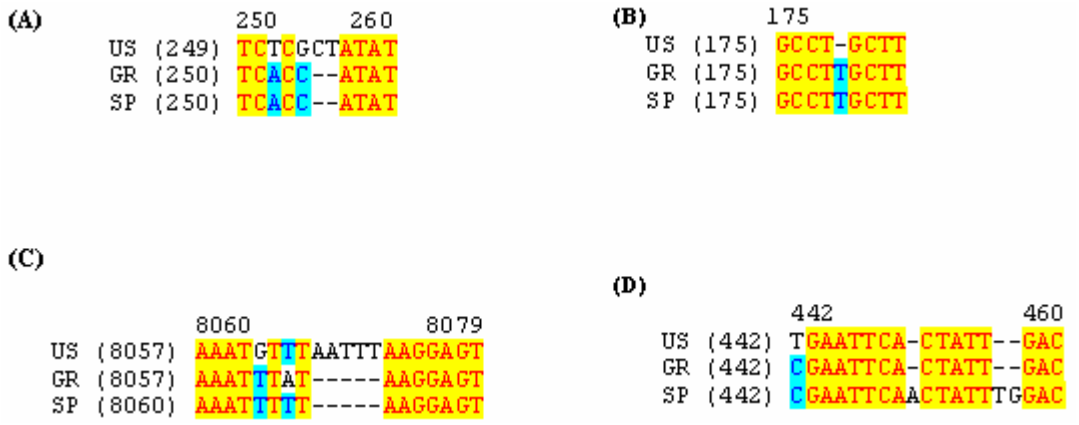


Figure 3.12a Multiple alignments of selected genomic regions (A and B: RNA1, C and D: RNA2) of the American, Greek and Spanish ToCV isolates. Numbers correspond to nucleotide positions in respective viral genomes.

RNA1 P5

RNA2 P7

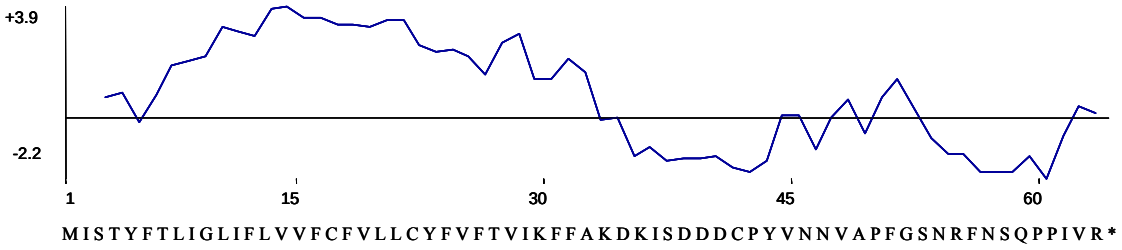


Figure 3.12b Hydrophobic profile (Kyte and Doolittle, 1982) for putative proteins encoded by the small ORFs at the 3` end of RNA 1 and RNA 2, showing transmembrane domains.

Nucleotide and amino acid sequence comparisons of the ORFs and UTRs (Appendix 2) of the genomic RNAs of the Greek, American and Spanish isolates are shown in Table 3.1. Phylogenetic analyses performed on the complete nucleotide and amino acid sequences of the CP, RdRp and p22 ORFs and proteins of all sequenced ToCV isolates and all other sequenced members of the genus *Crinivirus*, showed that the Greek isolate is mostly related to the American isolate (Fig.3.13; Appendix 2).

Table 3.1 Percentage of identity between the nucleotide (underlined) ORFs amd UTRs and their deduced amino acid (parenthesis) sequences between the Greek, American and Spanish isolates.

Greek	**American (nt/aa%)**	**Spanish (nt/aa%)**
RNA 1		
ORF1a	99 /(98.6)	97.1/(97.3)
ORF1b/RdRp	99.5/(100)	98.2/(99.8)
ORF2/P22	98.8/(99)	96.6/(95.4)
ORF3/P5	100/(100)	98.7/(100)
5`-UTR	97.7	97
3-UTR	100	100
RNA 2		
ORF1/P4	96.1/(91.2)	96.1/(91.2)
ORF2/HSP70h	99/(98.6)	98.1/(98.7)
ORF3/P8	100/(100)	98.5/(98.5)
ORF4/P59	99.2/(99)	97.8/(98.1)
ORF5/P9	99.6/(100)	100/(100)
ORF6/CP	98.7/(98.4)	96.9/(95.7)
ORF7/CPm	99.3/(99.1)	97.4/(97.8)
ORF8/P27	99.3/(99.1)	97.4/(97.4)
ORF9/P7	100/(100)	99/(98.5)
5`-UTR	98.7	98.3
3-UTR	96.8	97.7

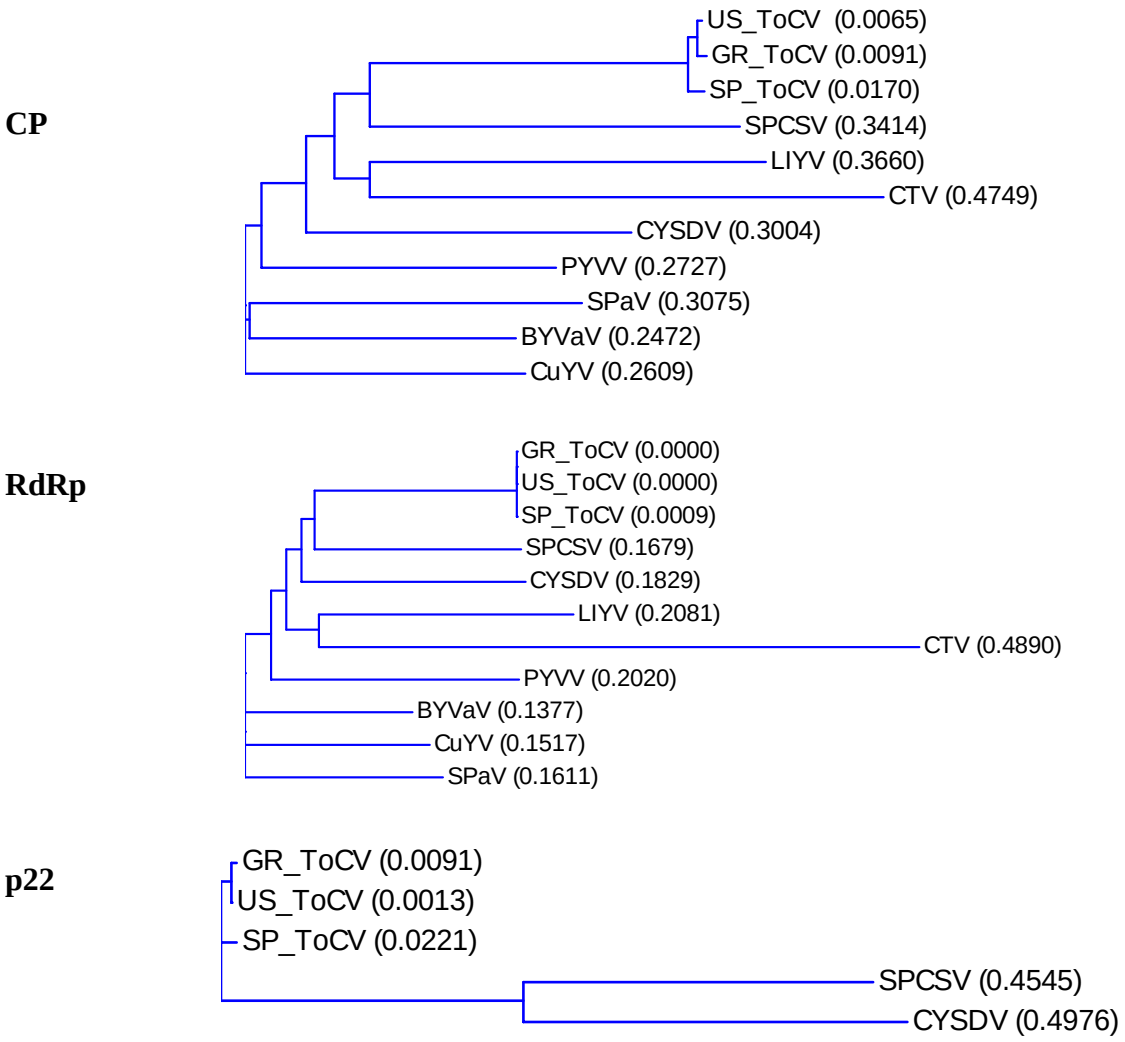


Figure 3.13 Phylogenetic analysis of criniviruses as determined from amino acid sequences of the CP, RdRp and p22 genes. The phylogram was generated by the AlignX program using the Neighbor Joining method (NJ) of Saitou and Nei. The NJ method works on a matrix of distances between all pairs of sequence to be analyzed. These distances are related to the degree of divergence between the sequences. The calculated distance values are in parentheses, following the molecule name. CTV was used as representative of the genus *Closterovirus*. The GenBank accession numbers are as follows: BYVaV, NC_006962 and NC_006963; CTV, U16304; CuYV, AB085612 and AB085613; CYSDV, AJ537493 and AJ439690; LIYV, U15440 and U15441; PYVV, AJ557128, AJ557129 and AJ508757; SPaV, NC_005895 and NC_005896; SPCSV, AJ428554 and AJ428555; ToCV-US, AY903447 and AY903448; ToCV-SP, DQ983480 and DQ136146.

Using the Mfold algorithm, four stem-loop structures (hpI-hpIV), were identified at the 3`-terminal of both ToCV RNA1 (Fig. 3.14), bearing in mind that all the isolates have 100% RNA1 3` UTR (Table 3.1). Amongst the four, stem-loop hpI was strongly conserved in both RNAs and showed the possibility of pseudoknot formation with the G/C-rich sequence near the very 3`-end (not predicted by Mfold). Similar secondary structures could be generated for both American and Spanish RNA2 3`-UTR but not the Greek isolate (Fig.3.15).

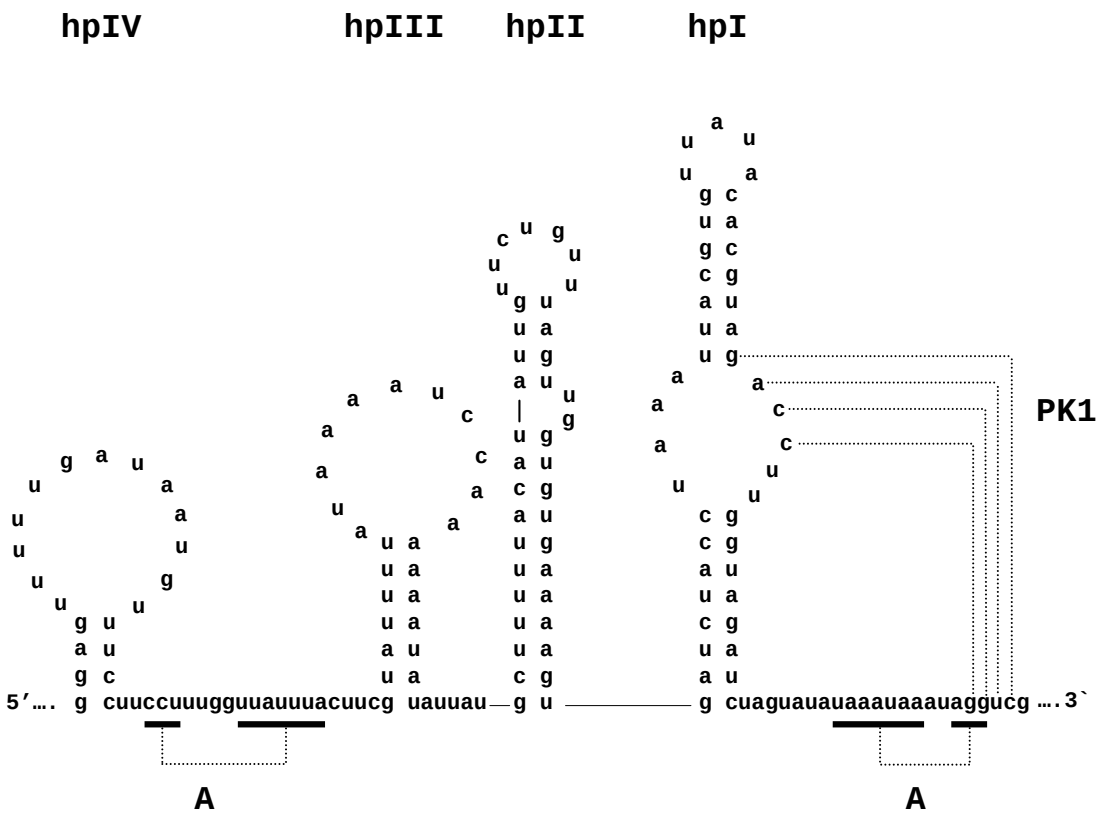


Figure 3.14 Secondary structure model for the 176 long 3`-UTR of the Greek ToCV isolate RNA1. RNA1: Stem loop structures are denoted as hpI, hpII, hpIII and hpIV. Sequences involved in the putative interactions are also underlined. Dotted lines indicate the pseudoknot interaction PK1.

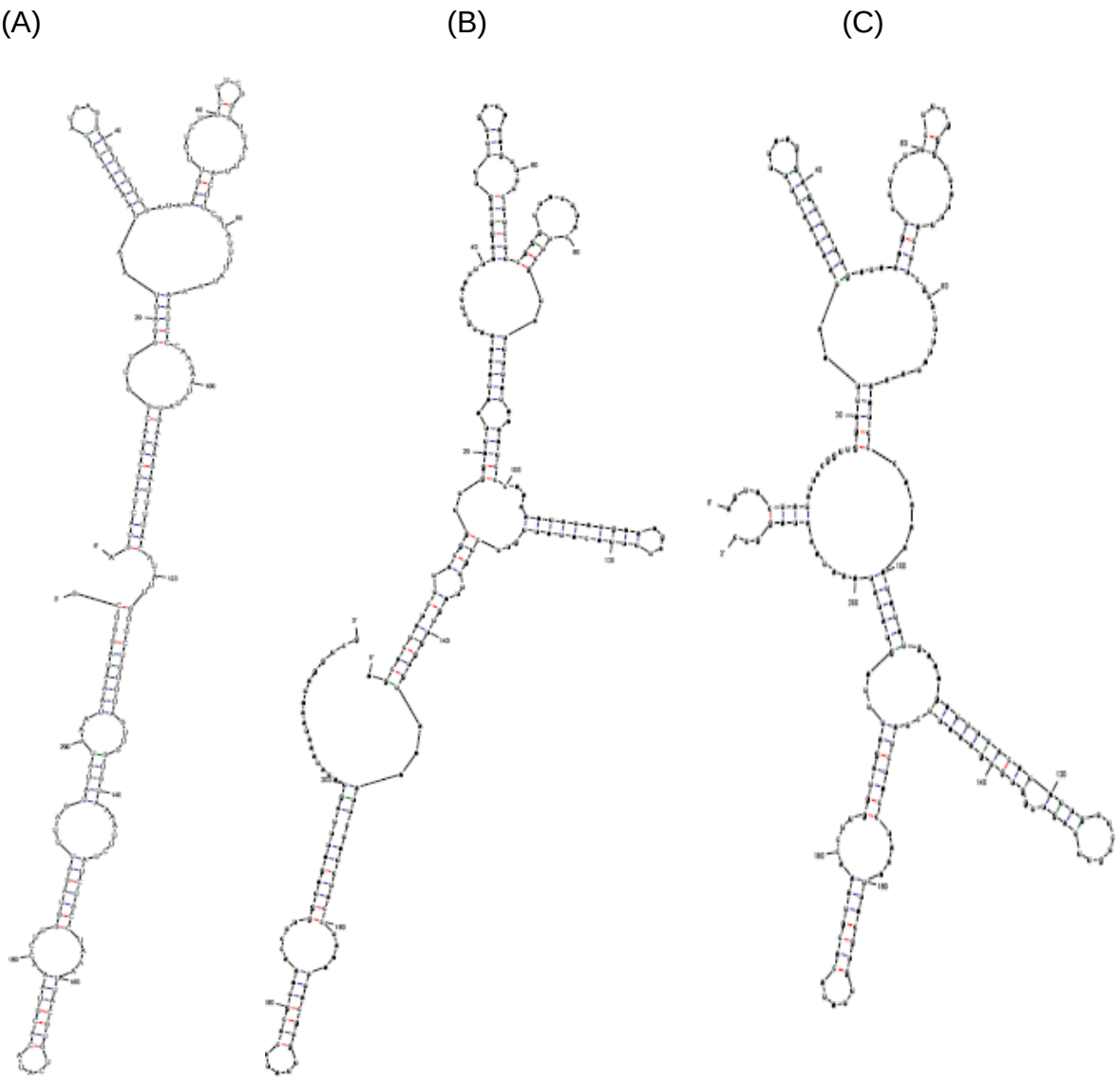


Figure 3.15 Secondary structure models predicted by the Mfold algorithm for the 212 nt long 3`-UTR of the ToCV RNA2 of the (A) Greek (B) American (C) Spanish isolates.

3.4 Construction of ToCV RNA1 full length cDNA clones

In order to construct the full length ToCV RNA1 cDNA clone, plasmid pG-T7/PCR1 was digested with *BamH*I and *Sac*I to release T7/PCR1 which was subcloned into a similarly digested pUC19 vector to generate pU-T7/PCR1. Single (*Sac*I) and double digestion (*BamH*I and *Sac*I) of pUC-T7/PCR1 followed by 1% agarose gel indicated that the generated construct included the correct size insert (PCR1) (Fig. 3.16 lanes 1 and 2, respectively). The newly assembled pU-T7/PCR1 construct was further linearised with double digestion with *Nco*I and *Sac*I and further ligated to the PCR2 insert which was excised from pG-PCR2 (*Nco*I and *Sac*I) to generate pU-T7/PCR1+2. Single (*Sac*I) and double digested (*Nco*I and *Sac*I) products of pU-T7/PCR1+2 were run on 1% agarose gel to confirm the insertion of the correct size fragment (PCR2) (Fig. 3.16 lanes 3 and 4, respectively). The pU-PCR1/T7+2 fragment was linearised following double digestion with *Avr*II and *Sac*I and ligated to the downstream *Avr*II and *Sac*I-excised PCR3 insert. Following ligation, transformation and plasmid isolation, pU-PCR1/T7+2+3 was generated and electrophoresed as single digested (*Sac*I) and double digested (*Avr*II and *Sac*I) on 1% agarose gel and were shown to include the PCR3 insert (Fig. 3.16 lanes 5 and 6, respectively). The pU-T7/PCR1+2+3 fragment was linearised following double digestion with *Afi*II and *Sac*I and ligated downstream to *Afi*II and *Sac*I-excised PCR4 insert. Following ligation, transformation and plasmid isolation, pU-T7/PCR1+2+3+4 was generated and electrophoresed as single (*Sac*I) and double digested (*Afi*II and *Sac*I) on 1% agarose gel to confirm the size of the insert (PCR4) (Fig. 3.17.A lane 2). pU-T7/PCR1+2+3+4 was linearised following double digestion with *EcoR*V and *Sac*I and further ligated to *EcoR*V and *Sac*I-excised PCR5 insert. Following ligation, transformation and plasmid isolation, pU-T7/PCR1+2+3+4+5 was generated and electrophoresed as single (*Sac*I) and double digested (*EcoR*V and *Sac*I) on 1% agarose gel and were shown to include the right size insert (PCR5) (Fig. 3.16 lanes 7 and 8, respectively). At this point and prior to further completion of the work, preliminary verification of clone pU-T7/PCR1+2+3+4+5 was carried out by digestion using selected restriction enzymes: *BamH*I and *Sac*I, *Hind*III, *BamH*I and *Avr*II, *Sac*I, *Sac*I and *Avr*II,

*Nco*I and *EcoR*V. Electrophoresis of the digested fragments confirmed the presence of DNA molecules of the expected molecular weight in all cases (Fig. 3.17. B).

PCR6 was released from pG-PCR6 recombinant plasmid by *Mlu*I and *Sac*I and was eluted from the resin (incubation at 60ºC for 20 min) and ligated into similarly digested pU-T7/PCR1+2+3+4+5. As a result, the construct pU-T7/PCR1+2+3+4+5+6 was generated. After transformation, ligation and plasmid purification single (*Sac*I) and double digested (*Mlu*I and SacI) fragments of the plasmid were electrophoresed on 1% agarose gel to show the correct size insert (PCR6) (Fig. 3.16 lanes 9 and 10, respectively). At this point, PCR7 was also subcloned into the pG-PCR6 plasmid construct following double digestion of pG-PCR7 with the *Nde*I and *Sac*I restriction enzymes and ligation to similarly digested pG-PCR6 recombinant plasmid to generate an additional recombinant construct pG-PCR6+7 (Fig. 3.17. C). In order to create the pU-T7/PCR1+2+3+4+5+6+7 construct, PCR7 was initially digested from pG-PCR7 using *Nde*I and *Sac*I and ligated into similarly digested pU-PCR1/T7+2+3+4+5+6 plasmid without success. In an alternative method, insert PCR6+7 was released from pG-PCR6+7 recombinant plasmid using *Mlu*I and *Sac*I and was ligated into the similarly digested pU-T7/PCR1+2+3+4+5. These experiments were successful and confirmed by single (*Sac*I) and double digestion (*Nsi*I and *Sac*I) followed by electrophoresis on a 1% agarose gel. As a result, the recombinant construct pU-T7/PCR1+2+3+4+5+6+7 was generated (Fig. 3.16 lanes 11 and 12, respectively).

To further complete the work towards the 3`-terminus of ToCV RNA1, four different constructs were created: pG-PCR8b+9, pG-PCR7+8b, pG-PCR6+7+8b and pG-PCR6+7+8+9. pG-PCR8b+9 recombinant plasmid in particular was created by digestion of pG-PCR9 by *Pst*I and *Sac*I and subcloning of the PCR9 fragment into pG-PCR8b (Fig 3.18. A). Secondly, PCR7 was released by digestion of pG-PCR7 by *SmaI* and *Sal*I and subcloned into pG-PCR8b to generate pG-PCR7+8b (Fig. 3.18.B. lanes 1, 2 and 3). Thirdly, pG-PCR8b was digested by *SmaI* and *Sac*I to release PCR8b to be subcloned into pG-PCR6+7 to create pG-PCR6+7+8b (Fig. 3.18.B. lanes 4 and 5). Fourthly, PCR8-

9 was digested from pG-PCR8+9 by *SmaI* and *Sac*I and subcloned into pG-PCR6+7 to create pG-PCR6+7+8+9 (Fig. 3.18.B, lane 6).

At this point, three different strategies were envisaged to proceed with the construction of a full cDNA ToCV RNA1:

i) PCR7+8b was excised by digestion (*Nde*I and *Sac*I) from pG-PCR6+7+8b and successfully ligated into pU-T7/PCR1+2+3+4+5+6 to create pU-T7/PCR1+2+3+4+5+6+7+8b (Fig.3.19.B, lanes 1-5; Fig.3.16, lanes 13 & 14).

ii) PCR6+7+8b+9 was released by digestion of pG-PCR6+7+8b+9 by (*Mlu*I and *Sac*I) and unsuccessfully ligated into pU-T7/PCR1+2+3+4+5.

iii) PCR8b+9 was released by *Sma*I and *Sac*I from pG-PCR8b+9 and unsuccessfully ligated into pU-T7/PCR1+2+3+4+5+6+7.

For the last 3`-terminal fragment of RNA1, PCR9 was excised using *EcoR*I from pG-PCR9 (*EcoR*I) and ligated into similarly digested pU-T7/PCR1+2+3+4+5+6+7+8b recombinant plasmid previously treated with alkaline phosphatase. Since *EcoR*I does also exist in the pUC19 vector immediately downstream the *Sac*I restriction site, orientation of the insert had to be confirmed (the insert could be cloned in two possible orientations). The final plasmid constructs (three clones in total) were also confirmed to include the correct size PCR8b+PCR9 inserts and in the correct orientation using *Sma*I and *Sac*I restriction enzymes digestion (Fig 3.20.C and D).

The three clones representing the complete ToCV RNA1 cDNA into the pUC19 vector were named pU-RNA1-1, pU-RNA1-2 and pU-RNA1-3 (Fig 3.20 C, lanes 2 and 10; D, lane 3). One of these clones was further tested by digestion using specific restriction enzymes in order to check whether the generated products were of the expected size depending on their position in the viral genomic sequence. These digestions were carried out using the following combinations of enzymes: *BamH*I, *BamH*I and *Nco*I, *BamH*I and *Avr*II, *BamH*I and *EcoR*V, *EcoR*V, *BamH*I and *Mlu*I, *BamH*I and *Nsi*I, *BamH*I and *Sac*I, *Sac*I, *Nru*I. All digestions produced DNA fragments of the expected molecular weights (Fig. 3.21).

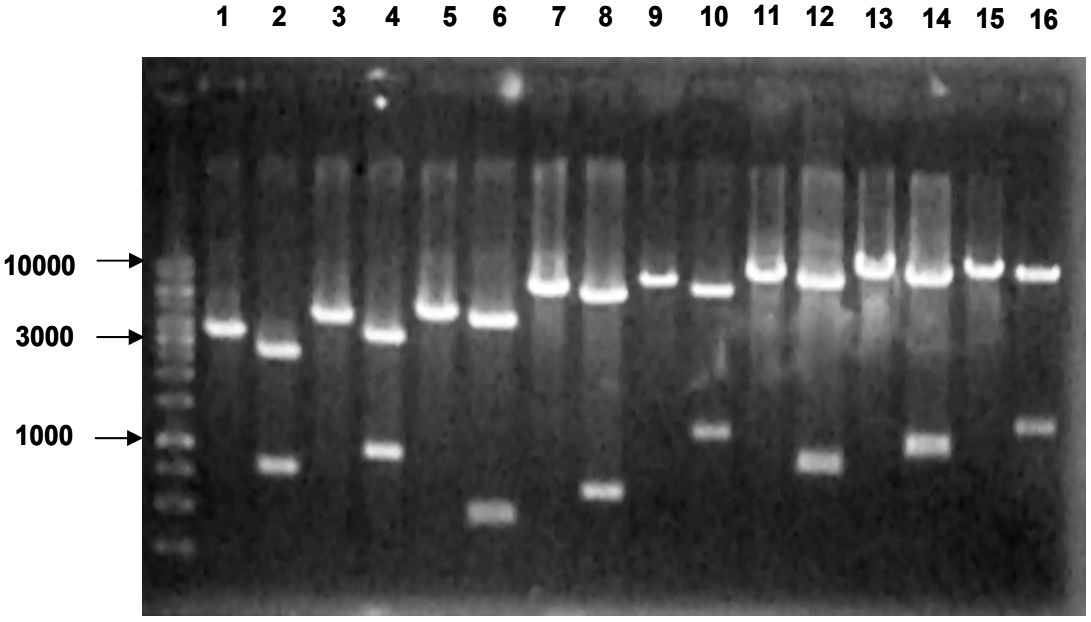


Figure 3.16 Agarose gel electrophoresis of steps used in cloning a full-length cDNA version of ToCV RNA1. Lanes 1 and 2: digestions of pU-PCR1/T7, lanes 3 and 4: digestions of pU-PCR1/T7+2, lanes 5 and 6: digestions of pU-PCR1/T7+2+3, lanes 7 and 8: digestions of pU-PCR1/T7+2+3+4+5, lanes 9 and 10: digestions of pU-PCR1/T7+2+3+4+5+6, lanes 11 and 12: digestions of pU-PCR1/T7+2+3+4+5+6+7, lanes 13 and 14: digestions of pU-PCR1/T7+2+3+4+5+6+7+8, lanes 15 and 16: digestions of pU-RNA1.

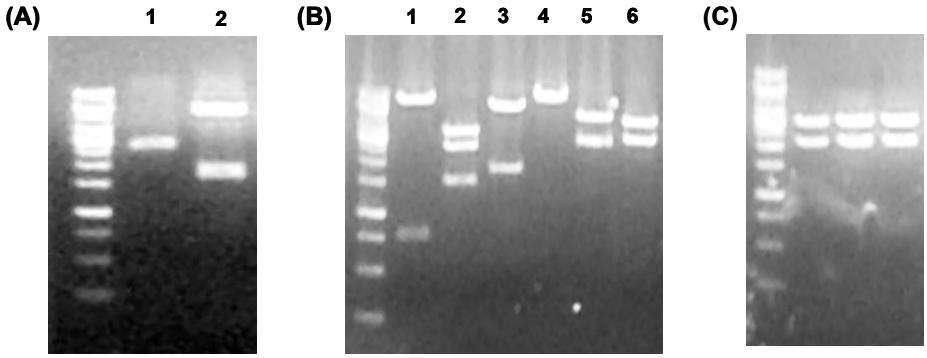


Figure 3.17 Electrophoresis analysis of: (A) the restriction digestion of pU-PCR1/T7+2+3+4 clones by *Afi*II and *Sac*I (lane 2); (B) the restriction digestions of the pU-PCR1/T7+2+3+4+5 clone. Lane 1: *BamH*I and *Sac*I, lane 2: *Hind*III, lane 3: *BamH*I and *Avr*II, lane 4: *Sac*I, lane 5: *Sac*I and *Avr*II, lane 6: *Nco*I and *EcoR*V; (C) the restriction digestion of pG-PCR6+7 clones.

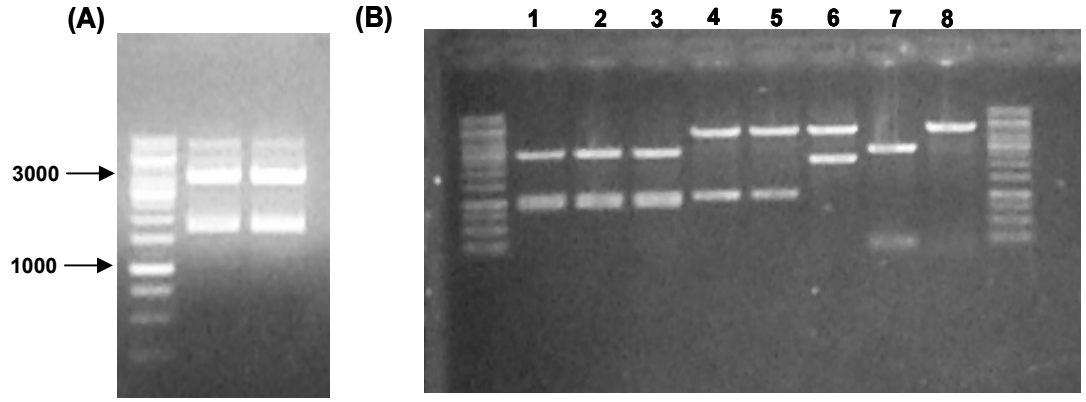


Figure 3.18 Gel electrophoresis for (A) digestion of pG-PCR8b+9 for 2 colonies by *Pst*I and *Sac*I, (B) digestion of pG-PCR7+8b clones 1, 2and 3 (lanes 1, 2, and 3, respectively), digestion of pG-PCR6+7+8b *SmaI* and *Sac*I (lanes 4 and 5), digestion of pG-PCR6+7+8+9 by *SmaI* and *Sac*I (lane 6). Position of 1 kbp DNA molecular marker is shown on the left.

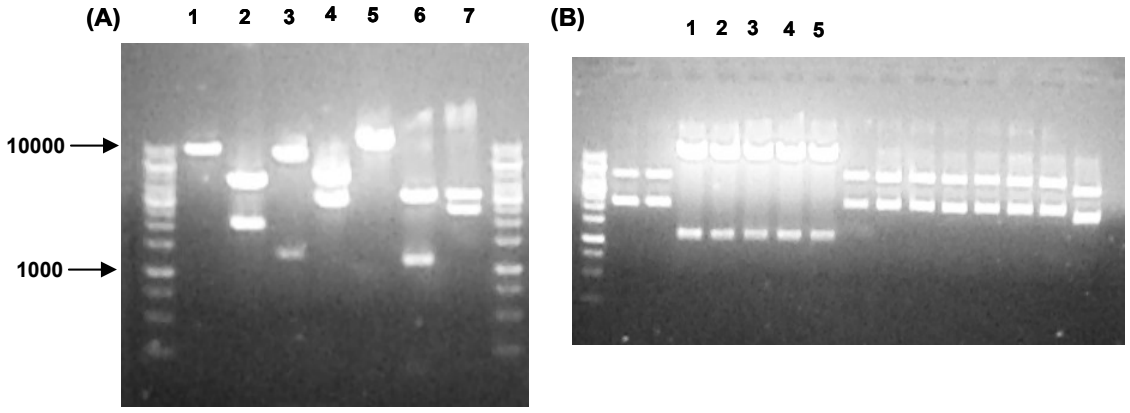


Figure 3.19 Gel electrophoresis of fragments were used in cloning a full-length cDNA version of ToCV RNA1. A: *Nde*I and *Sac*I digested pU-PCR1/T7+2+3+4+5+6 and pG-PCR6+7+8 (lanes 1 and 2, respectively), *Mlu*I and *Sac*I digested pU-PCR1/T7+2+3+4+5+6 and pG-PCR6+7+8+9 (lanes 3 and 4, respectively), *Sma*I and *Sac*I digested pU-PCR1/T7+2+3+4+5+6+7, pG-PCR8b and pG-PCR8b+9 (lanes 5, 6 and 7, respectively). B: Digestion of different successful clones of pU-PCR1/T7+2+3+4+5+6+7+8b by *Sma*I and *Sac*I (lanes 1 to 5). Position of 1 kbp DNA molecular marker is shown on the left.

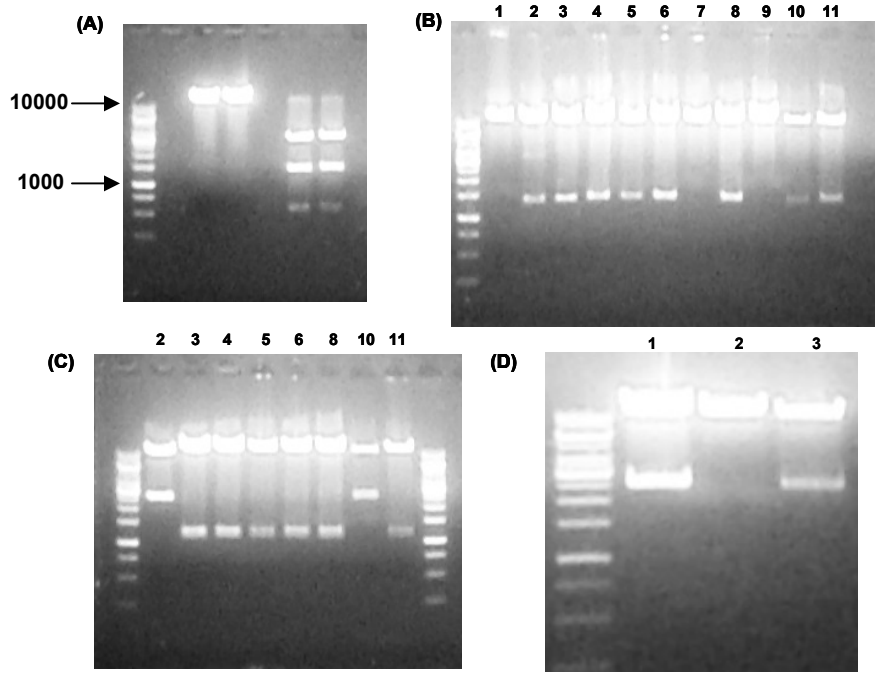


Figure 3.20 Gel electrophoresis for final step in cloning of ToCV RNA1 cDNA. A: *EcoR*I digested pU-PCR1/T7+2+3+4+5+6++7+8b and pG-PCR9, B: digestion to verify which clone has PCR9 by *EcoR*I, C: the clones which succeeded from B were re-digested by *Sma*I and *Sac*I (lanes 2 and 10 conatin successful pU-RNA1), D: Lanes 1 and 3, other clones contain the full cDNA of ToCV RNA1.

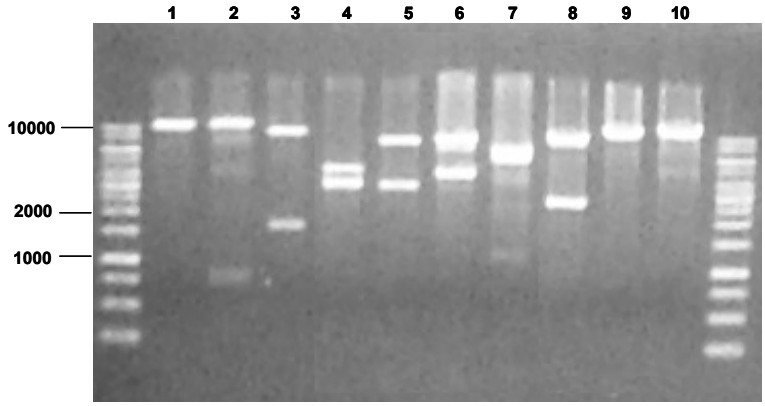


Figure 3.21 Gel electrophoresis of control digestions of the full-length cDNA version of ToCV RNA1. Lane 1: *BamH*I, lane 2: *BamH*I and *Nco*I, lane 3: *BamH*I and *Avr*II, lane 4: *BamH*I and *EcoR*V, lane 5: *EcoR*V, lane 6: *BamH*I and *Mlu*I, lane 7: *BamH*I and *Nsi*I, lane 8: *BamH*I and *Sac*I, lane 9: *Sac*I, lane10: *Nru*I. Position of 1 kbp DNA molecular marker is shown on the left.

3.5 Construction of ToCV RNA 2 full length cDNA clones

Recombinant plasmid pG-T7/PCR1 was digested by *Sac*I and *Sph*I to release insert T7/PCR1, which was further cloned into the pUC19 vector (Fig. 3.22 lanes 1 and 2, respectively). The newly assembled pU-T7/PCR1 construct was linearised following double digestion (*Sal*I and *Sph*I) and ligated with PCR2 carrying compatible ends. Following ligation, transformation and plasmid isolation, the pU-T7/PCR1+2 clone was generated and electrophoresed as single digested (*Sac*I) and double digested (*Sal*I and *Avr*II) on 1% agarose gel which was shown to include the correct size insert (PCR2) (Fig. 3.22 lanes 3 and 4, respectively). The newly assembled pU-T7/PCR1+2 clone was double digeste (*Avr*II and *Sph*I) before ligation with a *Avr*II and *Sph*I-excised PCR3 insert. Following transformation and plasmid isolation, the pU-T7/PCR1+2+3 plasmid was electrophoresed as single (*Sac*I) and double digested (*Avr*II and *BamH*I) to show that the correct size insert (PCR3) was cloned successfully (Fig. 3.22 lanes 5 and 6, respectively). Recombinant plasmid pU-T7/PCR1+2+3 was further linearised following double digestion with *BamH*I and *Sph*I and ligated to the similarly digested PCR4 insert excised from pG-PCR4 to create pU-T7/PCR1+2+3+4 (Fig. 3.22 lanes 7 and 8, respectively). Recombinant plasmid pU-T7/PCR1+2+3+4 was double digested with *BstZ17*I and *Sph*I before ligation with a similarly digested PCR5+6 insert to generate pU-T7/PCR1+2+3+4+5+6. The correct size insert was verified by single (*Sac*I) and double digestion (*BamH*I and *Nsi*I) and electrophoresis on 1% agarose gel (Fig. 3.22 lanes 9 and 10, respectively). Following the determination of PCR7+8 and PCR9 inserts orientation into constructs pG-PCR7+8 and pG-PCR9, respectively. PCR9 was excised by *Msc*I and *Sal*I and cloned into the recombinant plasmid pG-PCR7+8 which had been digested by *Nru*I and *Sal*I restriction enzymes (Fig. 3.23.A, lane 2). Finally, recombinant plasmid pU-T7/PCR1+2+3+4+5+6 was linearised following double digestion with *Nsi*I and *Sph*I and ligated to the downstream *Nsi*I and *Sph*I-excised PCR7+8+9 insert which represents the 3`-terminus of the ToCV RNA2. Single (*Sac*I) and double (*Nsi*I and *Sph*I) digestion products were electrophoresed on 1% agarose gel to show that insert PCR7+8+9 had been successfully cloned to generate pU-RNA2 (Fig. 3.22 lanes 11 and 12, respectively).

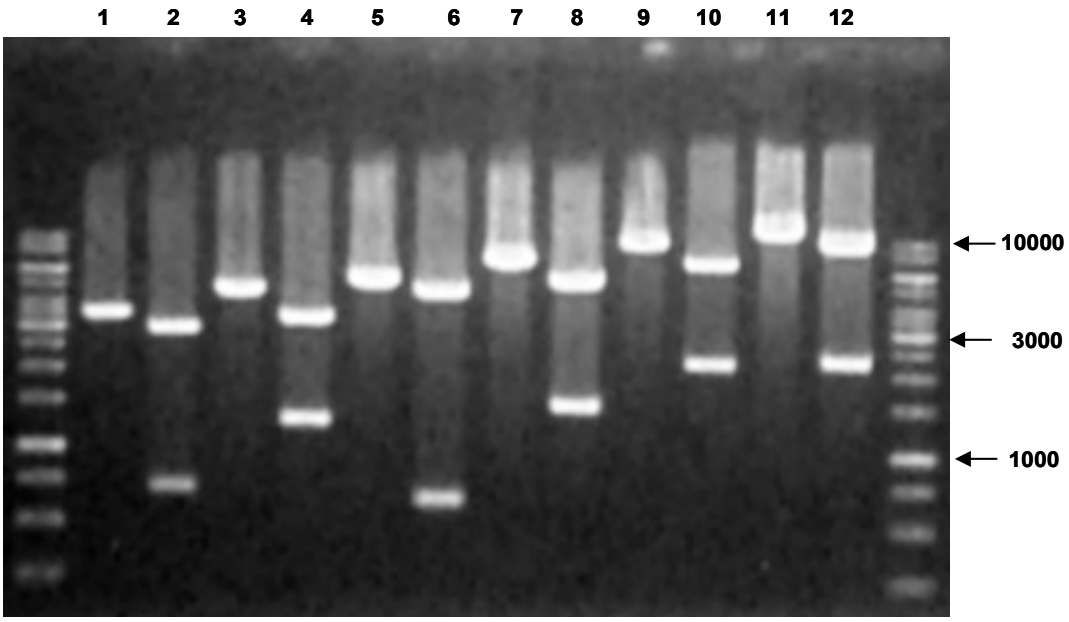


Figure 3.22 Electrophoresis analysis for steps used in cloning a full-length cDNA version of ToCV RNA2. Lanes 1 and 2: digestions of pU-PCR1/T7, lanes 3 and 4: digestions of pU-PCR1/T7+2, lanes 5 and 6: digestions of pU-PCR1/T7+2+3, lanes 7 and 8: digestions of pU-PCR1/T7+2+3+4, lanes 9 and 10: digestions of pU-PCR1/T7+2+3+4+5+6, lanes 11 and 12: digestions of pU-RNA2. Position of 1 kbp DNA molecular marker is shown on the right.

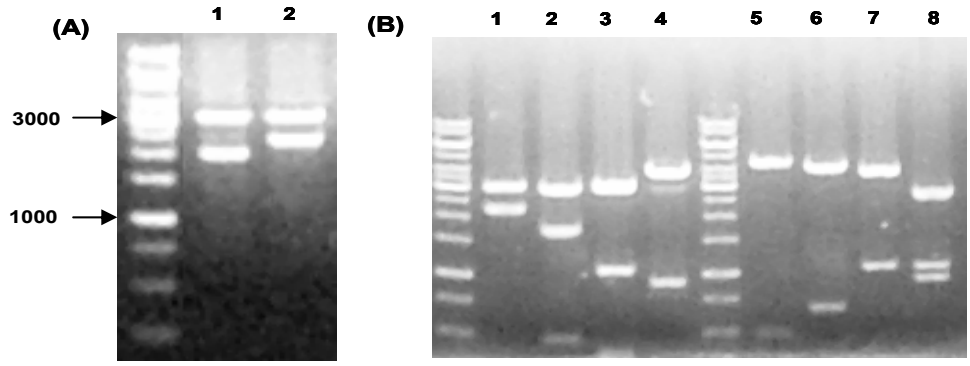


Figure 3.23 Electrophoresis analysis for A: double digested pG-PCR7+8+9 by *Nsi*I and *Sph*I (lane 2); B: *EcoR*I, *Sph*I and *Hind*III digested pG-PCR5+6 (lanes 1 and 2 respectively), *EcoR*I, *BamH*I and *Nco*I digested pG-PCR7+8 (lanes 3 and 4 respectively), *EcoR*I, *Sal*I and *Sac*I, *Hind*III, *Hind*III and *Sac*I of pU-PCR1/T7+2 (lanes 5 to 8, respectively). Position of 1 kbp DNA molecular marker is shown on the left.

Many verification steps throughout the work were carried out in order to confirm the correct clones (pG-PCR5+6, pG-PCR7+8 and pU-T7/PCR1+2) (Fig.3.23.B). Finally, further verification of the final construct pU-RNA2 was carried out by digesting one of the three clones obtained using specific restriction enzymes to generate fragments of predictable molecular weight for comparison purposes: *Sac*I, *Sac*I and *Sal*I, *Sac*I and *Avr*II, *Sac*I and *BamH*I, *BamH*I, *Sac*I and *BstZ17*I, *Sac*I and *Nsi*I, *Sac*I and *Sph*I, *Sph*I, *Nru*I. All digestions produced fragments of the expected molecular weight (Fig. 3.24).

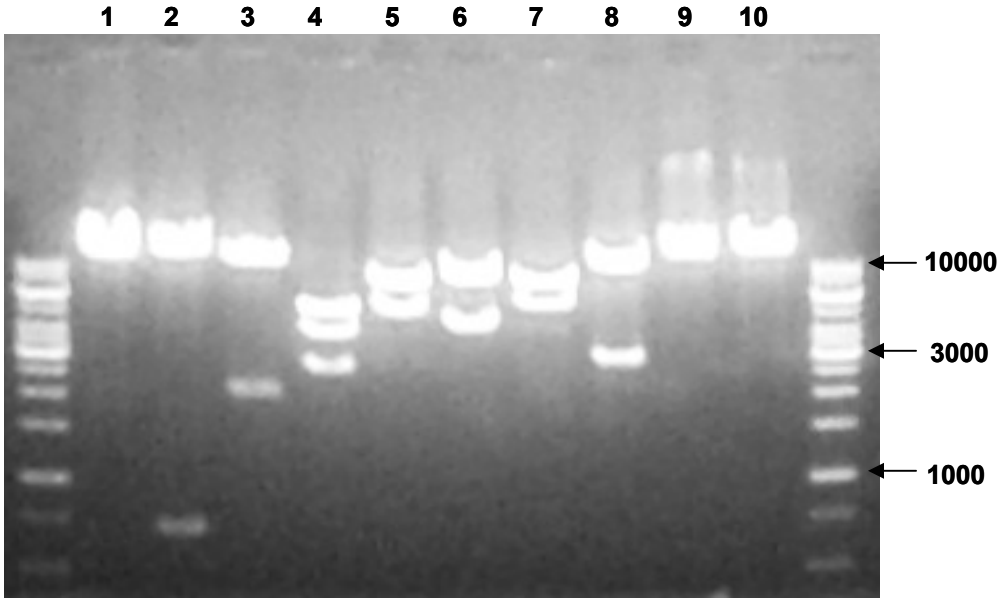


Figure 3.24 Control digestions of the full-length cDNA version of ToCV RNA2 (pU-RNA2). Lane 1: *Sac*I, lane 2: *Sac*I and *Sal*I, lane 3: *Sac*I and *Avr*II, lane 4: *Sac*I and *BamH*I, lane 5: *BamH*I, lane 6: *Sac*I and *BstZ17*I, lane 7: *Sac*I and *Nsi*I, lane 8: *Sac*I and *Sph*I, lane 9: *Sph*I , lane 10: *Nru*I. Position of 1 kbp DNA molecular marker is shown on the right.

DISCUSSION

RT-PCR and dot-blot hybridization have proved to be reliable and sensitive techniques in ToCV diagnosis. Until recently, the probes used in dot-blot hybridization were radioactively labeled and their use has been restricted to specialized laboratories due to safety reasons. Non-isotopic dot-blot hybridization becomes more accessible to non-specialized or less equipped laboratories, since tomato samples can be homogenized even in the field, and sap extract immediately can be applied to nylon membranes. The necessity to fix the total RNA extracts onto the membrane by a trans-illuminator is a disadvantage, but the technique has several advantages: it is faster (no RNA extraction kit/methodology required), more economical (the probe can be re-used up to 3 times, can be used for hundreds of samples simultaneously) and the samples can be stored until more samples accumulate. On the other hand, in terms of sensitivity, both RT-PCR and dot-blot hybridization perform satisfactorily and this becomes essential when early detection is required, bearing in mind low virus concentrations and phloem tropism of criniviruses.

Both nucleotide and deduced amino acid sequences were analyzed and compared with the American and Spanish isolates. The nucleotide sequence of the Greek ToCV isolate enabled us to perform phylogenetic analysis to show that the Greek isolate clusters together with the American isolate rather than a European isolate (Spanish) as would be expected. However, the RNA1 5`-UTR presents precisely the same number of nucleotides as the Spanish isolate and two less than the American isolate due to a common deletion for both isolates. Moreover, both Spanish and Greek 5`-UTR have an extra nucleotide than the American isolate. Also, the RNA2 3`-UTR of both Greek and Spanish isolates has common 5 nt deletions when compared with the American isolate. Generally it could be mentioned that the Greek isolate of ToCV is more closely related to the US isolate when ORFs and their deduced amino acid sequences are compared, while this is reversed when UTR sequences are compared. In terms of evolution, this may suggest recombination of two pre-existing ToCV sequences. Despite previous reports on

restricted genetic diversity within each species in the genus *Crinivirus*, a thorough investigation to reveal the genetic diversity of ToCV isolates would be intriguing in the future. Four stem-loop structures (hpI-hpIV) and the possibility of pseudoknot formation have been identified at the 3`-terminal of ToCV RNA1. Similar secondary structures have been predicted for all other sequenced criniviruses and several closteroviruses (Livieratos *et al*., 2004) with a role in virus replication that needs to be determined. For several positive single strand RNA plant viruses (*Barley mosaic virus*, *Turnip yellow mosaic virus*, *Tobacco mosaic virus*), the involvement of 3`-termini secondary structures in the initiation of negative-strand synthesis is well documented (Olsthoorn & Bol, 2002; Bringloe *et al*., 1999; Deiman & Pleij, 1997; Osman *et al*.,2000; Sivakumaran *et al*., 2000).

The availability of infectious *in vitro* transcripts from the cloned cDNA of several viral RNAs has facilitated gene function studies by using a mutagenesis and reverse genetics approaches (Boyer and Haenni, 1994). In the case of criniviruses, an infectious cDNA *in vitro* system is available solely for LIYV (Klaassen *et al*., 1996), which enforces the necessity for the development of infectious cDNA clones and replication studies for another member of this diverge group of plant viruses. Full-length cloned cDNAs of ToCV RNAs 1 and 2 were constructed in a pUC19 vector immediately downstream the sequence of the bacteriophage T7 RNA polymerase promoter in order to be tested in the future for infectivity in a protoplast system. Protoplast experiments will shed light on the understanding of how criniviruses utilize their extensive and diverse genetic information to complete their phloem-limited life cycle and initiate plant disease.

REFERENCES

Abou-Jawdah, Y. and Hourani, H. (2004). Detection of the crinivirus, *Cucurbit yellow stunting disorder virus*: using antibodies produced against its recombinant coat protein. *Phytopathology* **94**.

Agranovsky, A. A., Koonin, E. V., Boyko, V. P., Maiss, E., Frotschl, R., Lunina, N. A. and Atabekov, J. G. (1994). Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Virology* **198:** 311-24.

Agranovsky, A. A., Lesemann D. E., Maiss, E., Hull, R. and Atabekov, J. G. (1995). "Rattlesnake" structure of a filamentous plant RNA virus built of two capsid proteins. *National Acad Sciences* **92:** 2470-2473.

Aguilar, J. M. and Abad, J. (2006). Resistance to *Cucurbit yellow stunting disorder virus* in cucumber. *General Virology* **90:** 583-586.

Aguilar, J. M., Franco, M., Marco, C. F., Berdiales, B., Rodriguez-Cerezo, E., Truniger, V. and Aranda, M. A. (2003). Further variability within the genus *Crinivirus*, as revealed by determination of the complete RNA genome sequence of *Cucurbit yellow stunting disorder virus. General Virology* **84:** 2555-2564.

Alicai, T., Fenby, N. S., Gibson, R. W., Adipala, E., Vetten, H. J., Foster, G. D. and Seal, S. E. (1999). Occurrence of two serotypes of *Sweet potato chlorotic stunt virus* in East Africa and their associated differences in coat protein and HSP70 homologue gene sequences. *Phytopathology* **48:** 718.

Alzhanova, D. V., Napuli, A. J., Creamerl, R. and Dolja, V. V. (2001). Cell-to-cell movement and assembly of a plant closterovirus: roles for the capsid proteins and Hsp70 homolog. *EMBO* **20:** 6997-7007.

Bar-Joseph, M., Marcus, R. and Lee, R. F. (1989). The continuous challenge of *Citrus Tristeza Virus* control. *Annual Review of Phytopathology* **27:** 291-316.

Boyer, J. C. and A. L. Haenni (1994). Infectious transcripts and cDNA clones of RNA viruses. *Virology* **198:** 415-26.

Bringloe, D. H., Pleij, C. W., and Coutts, R. H. (1999). Mutation analysis of cis-elements in the 39- and 59-untranslated regions of *Satellite tobacco necrosis virus* strain C RNA. *Virology* **264:** 76-84.

Cohen, J., Franck, A., Vetten, H. J., Leseman, D. E. and Loebenstein, G. (1992). Purification and properties of closterovirus-like particles associated with a whitefly-transmitted disease of sweet potato. *Phytopathology* **121:** 257-268.

Deiman, B. and Pleij, C. W. A. (1997). Pseudoknots: a vital feature in viral RNA. *Academic Press* **8:** 166-175.

Dolja, V. V. (2003). *Beet yellows virus*: the importance of being different. *Molecular Plant Pathology* **4:** 91-98.

Dolja, V. V., Kreuze, J. F. and Valkonen, J. P. (2006). Comparative and functional genomics of closteroviruses. *Virus Research* **117:** 38-51.

Dolja, V. V., Karasev, A. V. and Koonin, E. V. (1994). Molecular biology and evolution of closteroviruses: sophisticated build-up of large RNA genomes. *Annual Review of Phytopathology* **32:** 261-285.

Dolja, V. V., Kreuze, J. F. and Valkonen, J. P. (2006). Comparative and functional genomics of closteroviruses. *Virus Research* **117:** 38-51.

Dovas, C. I., Katis, N. I. and Avgelis, A. D. (2002). Multiplex detection of criniviruses associated with epidemics of a yellowing disease of tomato in Greece. *Plant Disease* **86:**1345-1349.

Duffus, J., Larsen, R. and Liu, H. (1986). Lettuce infectious yellows virus- a new type of whitefly-transmitted virus. *Phytopathology* **76:** 97-100.

Duffus, J. E., Liu, H. Y., Wisler, G. C. and Li, R. (1996). *Lettuce chlorosis virus*-A new whitefly-transmitted closterovirus. *European Journal of Plant Pathology* **102:** 591-596.

Francki, R. I. B., Fauquet, C. M., Knudson, D. and Brown, F. (1991). Classification and nomenclature of viruses. Fifth report of the international committee on taxonomy of viruses. *Archives of Virology* **2:** 345-347.

Gowda, S., Satyanarayana, T., Ayllon, M.A., Albiach-Marti, M. R., Mawassi, M., Rabindran, S., Garnsey, S. M., Dawson, W. O., (2001). Characterization of the cis-acting elements controlling subgenomic mRNAs of *Citrus tristeza virus*: production of positive- and negative-stranded 3`-terminal and positive-stranded 5_-terminal RNAs. *Virology* **286:** 134–151.

Hartono, S., Natsuaki, T., Genda, Y. and Okuda, S. (2003). Nucleotide sequence and genome organization of *Cucumber yellows virus*, a member of the genus *Crinivirus*. *General Virology* **84:** 1007–1012.

Hourani, H. and Abou-Jawdah, Y. (2003). Immunodiagnosis of *Cucurbit yellow stunting disorder virus* using polyclonal antibodies developed against recombinant coat protein. *Plant Pathology* **85:** 197-204.

Jelkmann, W., Fechtner, B. and Agranovsky. A. A. (1997). Complete genome structure and phylogenetic analysis of *little cherry virus*, a mealybug-transmissible closterovirus. *General Virology* **78:** 2067-2071.

Karasev, A. V. (2000). Genetic diversity and evolution of closteroviruses. *Annual Reviews* **38:** 293-324.

Klaassen, V. A., Boeshore, M. L., Koonin, E. V., Tian, T. and Falk, B. W. (1995). Genome structure and phylogenetic analysis of *Lettuce infectious yellows virus*, a whitefly-transmitted, bipartite closterovirus. *Virology* **208:** 99–110.

Klaassen, V. A., Mayhew, D., Fisher, D. and Falk, B. W. (1996). *In vitro* transcripts from cloned cDNAs of the *Lettuce infectious yellows closterovirus* bipartite genomic RNAs are competent for replication in *Nicotiana benthamiana* protoplasts. *Virology* **222:** 169-175.

Koonin, E. V. and Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Critical Reviews in Biochemistry and Molecular Biology* **28:** 375-430.

Kreuze, J. F., Savenkov, E. I., Cuellar, W., Li, X. and Valkonen, J. P. (2005). Viral class 1 RNase III involved in suppression of RNA silencing. *Virology* **79**: 7227-7238.

Kreuze, J. F., Savenkov, E. I. and Valkonen, J. P. T. (2002). Complete genome sequence and analyses of the subgenomic RNAs of *Sweet potato chlorotic stunt virus* reveal several new features for the genus *Crinivirus*. *Virology* **76:** 9260-9270.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Molecular Biology* **157:** 105-132.

Leiser, R., Ziegler-Graff, V., Reutenauer, A., Herrbach, E., Lemaire, O., Guilley, H., Richards, K. and Jonard, G. (1992). Agroinfection as an alternative to insects for infecting plants with *Beet western yellows luteovirus*. *National Acad Sciences* **89:** 9136-9140.

Liu, H. Y., Wisler, G. C. and Duffus, J. E. (2000). Particle lengths of whitefly-transmitted criniviruses. *Plant Disease* **84**, 803-805.

Liu, L. and G. Lomonossoff (2002). Agroinfection as a rapid method for propagating *Cowpea mosaic virus*-based constructs. *Virological Methods* **105:** 343-8.

Livieratos, I. C., Avgelis, A. D. and Coutts, R. H. A. (1999). Molecular characterization of the *cucurbit yellow stunting disorder virus* coat protein gene. *Virology* **89:** 1050-1055.

Livieratos, I. C., Eliasco, E., Muller, G., Olsthoorn, R. C. L., Salazar, L. F., Pleij, C. W. A. and Coutts, R. H. A. (2004). Analysis of the RNA of *Potato yellow vein virus*: evidence for a tripartite genome and conserved 3'-terminal structures among members of the genus *Crinivirus*. *General Virology* **85:** 2065-2075.

Lozano, G., Moriones, E. and Navas-Castillo, J. (2006). Complete nucleotide sequence of the RNA2 of the crinivirus *Tomato chlorosis virus*. *Archives of Virology* **151:** 581-587.

Lozano, G., Moriones, E. and Navas-Castillo, J. (2007). Complete sequence of the RNA1 of a European isolate of tomato chlorosis virus. *Archives of Virology* **152:** 839-841.

Lu, R., Folimonov, A., Shintaku, M., Li, W. X., Falk, B. W., Dawson, W. O. and Ding, S. W. (2004). Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. *Proceedings of the National Academy of Sciences* **101:** 15742-15747.

Lucas, W. J. and Gilbertson, R. L. (1994). Plasmodesmata in relation to viral movement within leaf tissues. *Annual Reviews* **32:** 387-411.

Marco, C. F., Aguilar, J. M., Abad, J., Gómez-Guillamón, M. L. and Aranda, M. A. (2003). Melon resistance to *Cucurbit yellow stunting disorder virus* is characterized by reduced virus accumulation. *Virology* **93:** 844-852.

Martelli, G. P., Agranovsky, A. A., Bar-Joseph, M., Boscia, D., Candresse, T., Coutts, R. H. A., Dolja, V. V., Falk, B. W., Gonsalves, D. and Jelkmann, W. (2002). The family *Closteroviridae* revised. *Archives of Virology* **147:** 2039-2044.

Martin, R. R., Tzanetakis, I. E., Gererich, R., Fernandez, G. and Pesic, Z. (2003). *Blackberry yellow vein associated virus*: a new crinivirus found in blackberry. ISHS Acta Horticulture **656:** 137-142

Medina, V., Rodrigo, G., Tian, T., Juarez, M., Dolja, V. V., Achon, M. A. and Falk, B. W. (2003). Comparative cytopathology of *Crinivirus* infections in different plant hosts. *Annual Applied Biology* **143:** 99–110.

Medina, V., Sudarshana, M. R., Tian, T., Ralston, K. S., Yeh, H. H. and Falk, B. W. (2005). The *Lettuce infectious yellows virus* (LIYV)-encoded P26 is associated with plasmalemma deposits within LIYV-infected cells. *Virology* **333:** 367-373.

Medina, V., Tian, T., Wierzchos, J. and Falk, B. W. (1998). Specific inclusion bodies are associated with replication of *Lettuce infectious yellows virus* RNAs in *Nicotiana benthamiana* protoplasts. *General virology* **79:** 2325-2329.

Ng, J. C. K. and Falk, B. W. (2006). *Bemisia tabaci* transmission of specific *Lettuce infectious yellows virus* genotypes derived from in vitro synthesized transcript-inoculated protoplasts. *Virology* 352: 209-215.

Offei, S. K., Arciniegas, M., Muller G., Guzman, M., Salazar, L. F. and Coutts, R. H. A. (2004). Molecular variation of *Potato yellow vein virus* isolates. *Virology* **149:** 821-827.

Olsthoorn, R. C. L. and Bol, J. F. (2002). Role of an essential triloop hairpin and flanking structures in the 3' untranslated region of *Alfalfa mosaic virus* RNA in *in vitro* transcription. *Virology* **76:** 8747-8756.

Osman, T. A. M., Hemenway, C. L. and Buck, K. W. (2000). Role of the 3 ' tRNA-like structure in *tobacco mosaic virus* minus-strand RNA synthesis by the viral RNA-dependent RNA polymerase *in vitro. Virology* **74:** 11671-11680.

Peremyslov, V. V., Andreev, I. A., Prokhnevsky, A. I., Duncan, G. H., Taliansky, M. E. and Dolja, V. V. (2004). Complex molecular architecture of *Beet yellows virus* particles. *National Acad Sciences* **101:** 5030-5035.

Pringle, C. R. (1996). Virus taxonomy 1996-a bulletin from the Xth international congress of virology in Jerusalem. *Virology* **141:** 2251-2256.

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

Revers, F., Gall, O. L., Candresse, R. and Maule, J. (1999). New advances in understanding the molecular biology of plant/potyvirus interactions. *Am. Phytopath. Society* **12:** 367-376.

Robles, J. and Doers, M. (1994). pGEM®-T Vector Systems troubleshooting guide. Promega Notes **45:** 19.

Rubio, L., Soong, J., Kao, J. and Falk, B. W. (1999). Geographic distribution and molecular variation of isolates of three whitefly-borne closteroviruses of cucurbits: *Lettuce infectious yellows virus*, *Cucurbit yellow stunting disorder virus*, and *Beet pseudo-yellows virus*. *Virolog* **89:** 707-711.

Salazar, L. F., Muller, G., Querci, M., Zapata, J. L. and Owen, R. A. (2000). *Potato yellow vein virus*: its host range, distribution in South America and identification as a crinivirus transmitted by *Trialeurodes vaporariorum*. *Annals Applied Biology* **37:** 7-19.

Sivakumaran, S., Hariharaputran, S., Mishra, J. and Bhalla, U. S. (2003). The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Oxford Univ Press* **19:** 408-415.

Sumiyoshi, H., Hoke, C. H. and Trent, D. W. (1992). Infectious Japanese encephalitis virus RNA can be synthesized from in vitro-ligated cDNA templates. *Virology.* **66:** 5425-5431.

Tian, T., Rubio, L., Yeh, H. H., Crawford, B. & Falk, B. W. (1999). *Lettuce infectious yellows virus*: in vitro acquisition analysis using partially purified virions and the whitefly *Bemisia tabaci*. *General virology* **80:** 1111-1117.

Tzanetakis, I. E. and Martin, R. R. (2004). Complete Nucleotide Sequence of a Strawberry Isolate of *Beet Pseudo yellows Virus*. *Virology* **28:** 239-246.

Tzanetakis, I. E., Postman, J. D. and Martin, R. R. **(2005).** A member of the *Closteroviridae* from mint with similarities to all three genera of the family. *Plant Disease* **89:** 654-658.

Tzanetakis IE, Susaimuthu J, Gergerich RC, Martin RR. (2006). Nucleotide sequence of *Blackberry yellow vein associated virus*, a novel member of the *Closteroviridae. Virus Research* **116:** 196-200.

Weiland, J. J. and T. W. Dreher (1989). Infectious TYMV RNA from cloned cDNA: effects *in vitro* and *in vivo* of point substitutions in the initiation codons of two extensively overlapping ORFs. *Nucleic Acids Res.* **17:** 4675-87.

Wintermantel, W. M., Wisler, G. C., Anchieta, A. G., Liu, H. Y., Karasev, A. V. and Tzanetakis, I. E. (2005). The complete nucleotide sequence and genome organization of *Tomato chlorosis virus*. *Archives of Virology* **150:** 2287-2298.

Wintermantel, W. M. and Wisler, G. C. (2006). Vector specificity, host range, and genetic diversity of *Tomato chlorosis virus*. *Plant Disease* **90:** 814-819.

Wisler, G. C., Duffus, J. E., Liu, H. Y. and Li, R. H. (1998). Ecology and epidemiology of whitefly-transmitted closteroviruses. *Plant Disease* **82:** 270-280.

Wisler, G. C., Li, R. H., Liu, H. Y., Lowry, D. S. and Duffus, J. E. (1998). *Tomato chlorosis virus*: A new whitefly-transmitted, phloem limited, bipartite closterovirus of tomato. *Phytopathology* **88:** 402-409.

Yamaya, J., Yoshioka, M., Meshi, T., Okada, Y. and Ohno, T. (1988). Expression of *Tobacco mosaic virus* RNA in transgenic plants. *Molecular and General Genetics* **211:** 520-525.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33:** 103-119.

Yeh, H. H., Tian, T., Rubio, L., Crawford, B. and Falk, B. W. (2000). Asynchonous accumulation of *Lettuce infectious yellows virus* RNAs 1 and 2 and identification of an RNA 1 Trans enhancer of RNA 2 accumulation. *Virology* **74:**5762-5768.

Zee, F., Gonsalves, D., Goheen, A., Kim, K. S., Pool, R. and Lee, R. F. (1987). Cytopathology of leafroll-diseased grapevines and the purification and serology of associated closterovirus like particles. Phytopathology 77: 1427-1434.

Zinovkin, R. A., Jelkmann, W. and Agranovsky, A. A. (1999). The minor coat protein of *Beet yellows closterovirus* encapsidates the 5' terminus of RNA in virions. *General Virology* **80:** 269-272.

APPENDICES

Appendix 1

Nucleotide Sequences

ToCV-RNA1 Sequence

```
gaaatagtat tcgtgtgatt acacaaagta ctaattaact tagctttaag gctttctggt        60
gtgttgccag tttgcctgcc cggttgccac tgtgttcagt gtcgacctta ttataccagc       120
attcaccgcc atagcattac tagaaacagc acgacgagta cactaaacta atctgccttg       180
ctttactttg cgttctcctt ttcgtctcgg agagcgcgcc ccttgcccgc cgttgtgggt       240
tgttagtgtt caccatataa ccccctcttg ttgtgattta gtagtttgta ttttagctat       300
tatggattct cagcaaaacc tagtttcgtt taacgtagat tttactgaaa agaaaataaa       360
agatacgttt agagtagtta agaggcatat tagtaataaa tataataaat cctttaagaa       420
gagactgttt ttgtgcagtt gcgatttaaa tgtctcaatt gccgcaaact cagtgtcgac       480
tgctcaggga tgcagtgtga gggctcggat aaaaactcgc atgaacgtac ttagacggtt       540
atgtgggata ccacattgca atttcaacaa attaccagtt tctgtataca gaaaattcgg       600
tcacgatttc catagaataa attctgcaat agaccggtat ttggaaagtt ctgtggggtc       660
aacaggtgga aaatctctcg aggaaatcga gagaatagat ggttacacgg accatggcga       720
tggttatcgt gtaagacttt actctgacgt caacatttgt gacgtgtttg taaggttcaa       780
tgctgatgtt gccactggtc acgatttgaa gatcaggctc cagagacaga agaatagaat       840
taccggccaa gttaggacgg tcgttcatgc tgcgaagaat ggttttgggc ttgatttttc       900
tctgtggtgt gatagtttcc ttctcacaga cagtaaaact tccggccaga agatggttat       960
ggatatgata gcggctgttg ctatgaaggt tccggatgtt ttgcctggat ttgaccgttt      1020
gtactccaga gtttttagta agaaccttga gcagatcaga tccgctttta tcagagagtg      1080
caagaagtac atgcaaactt cttgtgataa ggctaatgca agtcaccgcg agaaagtgac      1140
actcaaagtg aagccgtcag tcgcggtgaa gaacgtgcca gatgtgaccc ataaggtcgg      1200
acccgacggt gcagaatcgt tcgtggtcac ttattctgat gggaaacaac gcataattgt      1260
gaatgatgac ggtgccgtta ggaatttgtt taacgctact ctcaccaacg gtaaatattt      1320
cattcacccg aaggctatga tcccagataa aagctttttc tcgtccaaaa caactgagta      1380
ttgttggttg aatgccctcg ctgccgtcaa taagaaaata ccagatttcg ttgtacctta      1440
cccatgtctt aggatgcggg tgttgtacaa ctgtggtttg ggcagtgtag ttgagaagca      1500
ttgtaagttt gtaaaagccg gtttatacca ttttgatctt aggtactgtg cgcctggaaa      1560
acccatcaat ctcaatggtt atgttggttc caaggttgat actgatatac cgtccctagg      1620
taagaacata aatgtcatct ttgatgattt gatcggacat tatgttcaag gtactaattt      1680
gagatctgac aatctgttat ctagcaacat cgttaataga ttgtctgaca gaataaacac      1740
aatgtttagc aaaccgaaag atccctcgat agcgacttcg ctcaccgcat cagagaaacg      1800
taaagtgatc gatatgtttc ctgagttgtg tttgaatttc accgacacat cttactcctc      1860
tcatcctatt gccactgcaa ttaggtgttg tgagaacttc atcatggcca aacgatgtgg      1920
taatgaggat tttattgatg ctggtggtga tgttgtgcac tatatgttag agagtgtgaa      1980
aaacgttcat gtttgcacac caatagttga taccaaggat gcccatagac acatttcgag      2040
gagtgctatg ctcgatacca tgtggggact taaggacaag gtgtctttct gtgagcataa      2100
aactgaaact tgcacagtgg agaaaactaa tattgtagcg gtggaggttt atgatatgac      2160
tctgaaacaa atggcacaag ctcttctgtc acataaggcc aaaaggtttg acttcagcct      2220
tataatacca cctgaggtgt gcgacccggt gtgtgatgtg taccagctta acaatagcct      2280
acacgttacc aataacggtg acaagattga gtacgcatac ggcgactttg gtgagagtta      2340
ttttcatgat agagagaact tgagagacat tctacgaact caaatgttcg tttataatgg      2400
agttgttttc aagaagtcat tagagtgttc tcgtgacaat ctgcatttct tctctgtggt      2460
tccttgcttg ggcattaaac cgggtgtcta cactttgtca actcattata aaaaatctga      2520
gtctgataag attgagatga caatacccgt tcgggagaaa ctcggacaag tcgtcgacaa      2580
gcggacgagt gttgataggg cacttactta cagcttgata gaatatgtca tgaacacagc      2640
ggctagggtg gatgagaaat cctctgacta tttagtttct cagtacaggg ccaagaagag      2700
tatgaccata aaaggtaata aagtggtgcc taatgactgc gatttacctg tggagttggt      2760
gccaggttac ctagctatta ttttggccga gggattgagg cttagggaga agattcaata      2820
ctttgcaaaa attagttatc atcgtcacta cagcccgtcg attttgaaga tattttccct      2880
cttgatacaa gagtttgtaa cgttcgctaa ggctaagtgt tacgatgggt ttgtttggtt      2940
ccttcgcaag tgtctgtctg ataaggtact tgataaggtg atatttggtg agaggcgtat      3000
```

```
ccatgactgc cagttcataa tgcggttcaa gcaagaggtc actgtggtcg gcgagagtgg      3060
taagcattcg atcctcggtg attctatcca cgccttcact gaagcaagtg agagagctga      3120
aactggactg gccgattttg tggacggtag accggatttg ttcaagccag atgactatga      3180
gaagttgtat gatctagttc attctggtgg tggaaacgca gggtttttta gtaagacggc      3240
taattattac tctccgtttt ctttgtatta ttctatttac aatttttct cgaattttat      3300
ttcttcagct tctcgaattt ccttgtatac caatttcttt atttcagtta tagattggat      3360
gcgttatcac tctatctctt gttttgaata ttttaaggag gtcgtaaaag tcatagtttg      3420
tgctactatg gggaagttga aatcaggtat gcagtctctt tgggacaaac tcaagcggtg      3480
tgctaaagag atgaaggggc tcgtgaacag cgacaggctg tctttggaga aaaaacttga      3540
tgatatcttc tctgcggacg acgagattgc tgaagcattg tcctcggatg tacaatctat      3600
tgttactgaa gcagtgcttt cagagcaatt gatctcgtct ggaggcggtg gagtcatatc      3660
acgaccgatg aaatgcgttc gtgaaatgta cgatcattca cggcgatcct tgcggtggtt      3720
ttgtacttgg tttctccagc aatttaagct ttacaaaaga aatattttgt acattcctga      3780
atttttaagg gagtcctttg atgagatctt tgaggcccta aaattgaaga ttttgtctaa      3840
cgatttcatc gagatgacga tccaaggtat gtctttcttt gttgttaatt taactacatc      3900
gttgtttttt ggtaagatag gtttaggttt gagtactgtg tctactgttt tctatctttc      3960
cattaaatac actggtttgg agagaagatt tcttggtact tcatttgtca atgaacatct      4020
ctcttccgct ataacgactg gtggtttttc ccatccgtac atgattccag tgaggtcttt      4080
ggttatgaaa agttgccaga attggatgaa acagaaactg ttgaggtatg agacggtctc      4140
tccaatagtt actgacctta ttgctaagga cgtgttaaac gcgtcagttt atcagtatgt      4200
aactccctac agagttagaa tggggttgta ttttagtatt ttgttggccc tactgaaacc      4260
aacatttgct gtcgtgctcc tggtgtgtat tctgttggtg gcggaacatg caaaatttta      4320
cagaactgtt gttgttcaag ccaatgttca tttatccttt gcttctcgac tgaggagatt      4380
gaatcctact tcgaaggcta gagcgattaa gcagttactc gtttccaagt ttgataggaa      4440
gagatttagc aaggataaca gtgacgataa gttctgcgat acatttgagg actttgattg      4500
cggtccgacc gagtttggaa taaactctaa agatttggtt ggtgctgggg aggcctcgac      4560
gtctgaacag cctgtggtag agttggatta ctccggcaat gagatcaaag ggaagagttg      4620
gggagatgtg actgacgatt ctatttctga gccggatatt gtggacagac cggacggtct      4680
ggtgttttca agtttggaga taactaatca aagggagaaa ttcttgtgtg atagcagtct      4740
tcttttgagt gatgctcttt tgcaatatcc agttgtggat gtgacgacca acgtcctgac      4800
tggtgataat attgttgaca ctttttcaga gtttttcttt ttagagaaaa agaaattaca      4860
tgtggagtta ggtaagatta ataatgttgt tgatttttac aaatcctcat tgaccgggaa      4920
gaatgctttt tataataagg tttggagctt gcgaaacaaa ttcgacgatt cgtcgttata      4980
tgtttcggaa aactcaaagg tctggtacaa gttgaaacag ggtgagaagg gtcatgttca      5040
acttgaaggt gtgtgcaaat acacgcttga caacaagctg gttcctttca cttactttta      5100
tgatgacttt caggtcacca gtgatgaatt gatgggcatg ttttccaaca ggagatgttt      5160
ggcattgcag tcaatcacac cgcgtgagtc cgggtttgac cttaatagca tgttggagaa      5220
tgtcacgttt ttcaataaac ctcctggtgc tggtaagacg accaccattg tgaggaatat      5280
ggtgagagac atcaaaagta atgtgaggtg tttggccctg acctgcacaa atgctgggaa      5340
gaaagagatc atacataagc tcaggaagga aggagtcaac aatgcattca gcttagtgat      5400
gacatatgac tcttttctca tcaatggtgg taagatggat actgatatgg tctattgtga      5460
tgagatcttt atgattcatg ctggactgtg ggtggcccta ttatcaatgt tgcaattcaa      5520
gaaaatggaa tgttacggtg acaaaaacca gattccattt atcaatagag tgccgaatac      5580
tctttgtcag tactcccaga agatattttt cttgtttagg atgatccacg acaatgtttc      5640
atacagatgc ccgccagatg tttgctacat tttgtctaat ttgagagatg ctgccggtaa      5700
tctgttgtat ccaaatggtg tcaaggctgt agggccgaac agcaatcttc tccgttctat      5760
gttcgtggtg ccactgagat ctgccgagga agttccatac agtccggatg tgaagatgat      5820
tgctttcacc aaacctgaga aggacgacat aatgcggcat ggtagaacag cggatggtaa      5880
gaccaattca gctcagactg ttaatgaggt tcaaggtggg acttttccca aagtcgagtt      5940
gtacagattg aggcaatatg ataatcccat ctacaatgat gtcaaccaat tcgtggttag      6000
tatatccaga cacacggagg ttatgaaata tagagtcttg tctacgaaga tgcacgacac      6060
tgttggacaa catatatcat ctttggataa ggttgctgat catatcataa gggagtgtgc      6120
atttaaacag caggtttaac acttatcggt tgaccattga aggttgttat attcctgaca      6180
ctttttcgag acctgcctca tctcatttga tggcagtcaa cgattttatg tcagtggtca      6240
acccgggatt agcttggatg caatttttgc acagaactat attgtttgag tatggtgatt      6300
ttgacatgcc acctgttgag aagatggtcc ttgatttttc aaaatacaag ccttatgtcg      6360
caggggagtt tgttgtctcg aaaattcttg gcaaaggtga gaggacgaga ccggacagta      6420
tgaaacaggg gataatctca ttgtcacata gaaatttttc tgcaccgaga ataaatgaac      6480
gtctggacgt ttataagact gctgaacgtt tatgtcagaa tctcgttaga tctttcgact      6540
tctcaaggtt gtatgagaac tatgatgtga ttcttcctga catgttcaaa attgacgatt      6600
```

```
ggttgcagga tagagatggt tcgaagtttg gtcggataaa aagggatatg gaccacaaat      6660
tgttggtcga acagtttgag agcttaaaat tcatgatcaa aggggagatg aaaccgaaga      6720
tggatatgtc gtcctataca gcttataatc caccggcgaa tatcatctat tataaccatc      6780
tggtaagtat gtattattct ccgttgtttc tggaggtctt tgataggata tcatactgtc      6840
ttagcaagaa gatagttatg tattccggga tgaatctaga aactctcggc accctgattg      6900
gttctaaact gcagaagccg ttgacatcat atcacacttt ggagattgat ttctcaaagt      6960
ttgataagtc ccaaggtatc ctatttaaag tttatgaggg gatgatttac cggtttttca      7020
agttttccga ggattactat accaacatag aggccactga atacttcata aagtatcgtg      7080
gtaggtgtgg aatcagcggg gagttgggtg cacaaaggag aacggggtca ccgaacactt      7140
ggttgtcaaa cacattggtt actatgggta tcatactcag tgtttacgac ctggatgata      7200
ttgatttatt cttagtaagt ggtgacgaca gtttgatctt ttcgagtaaa cccttgaaga      7260
ataaaactga tgagataaac agagatttcg gttttgaggc taagatgata gagaattcag      7320
tgccgtattt ttgctccaaa tatatcatca gtgatagagg aaaaatcaga gtcgttcctg      7380
atcctgtgag gttttttgag aagttgtctg tcccaattcg agttcaagat tttatgagtg      7440
acactctcat gcgggaaaaa tttaggtctt ataaggactt gatgaaggac tttgattacg      7500
atacaacgtg cgttttggtg gatgctttgg tgtgttatag gtacaattta ccaccgatgt      7560
gttcatatgc agcgttgtgt tatattcatt gtctgtgtgc aaattttact actttcagaa      7620
gagtctatga gagcgatttg actgttgtta tttaggtcag tatggatctc actggttgct      7680
tgcgtaagct tcgacagtgt gatcgacttc ttgagaggcg gggaaatgcc gtttcagaag      7740
ttcatttaag agcgatctta attgatcttg atgagtgttc tgagtgtttg atgctttgtg      7800
agcaggagta tatcagagac acggactgcc ttatgtcatt tctgttggcg ctgaaacact      7860
atgagattaa atttcacatg gatatgttga atatgattta tgactttaaa ctgaaaacgt      7920
cccagttgat tcaggatgtt tttagaatta aagtaatcat cagagtgtat cttgagttgt      7980
gtgaaattga tccacttttg gctatgactg aggcttgtca agacattctt gagagtggta      8040
ttttgaacat cggtttcatc tcttcggccc tcgggcatga accgaacata cttatcacaa      8100
tattgtcgat ggtcgatttt atagtcgtca ttgatgatcg accactggtc tttatccctt      8160
caaaaataag gtttgttggc gacaagttgg ggtcaggtca ttttaggtgg tttgataagt      8220
ttttctttgg gagtgacata taatctggtt aatattcaaa ggatgggtga tcgttttgct      8280
tgcgattctt tggatagtat agccagagat atacactctt tgtatatttt gtttttctat      8340
acgtttctag taggtgtttt gttgacattc ataatgtctt gtgttagagg tgttatagaa      8400
atttatcgtg ttaattaggg agtttttgat aatgtttcct tcctttggtt atttacttcg      8460
tattttataa aatccaaaaa atatattatg ctttttacat attgttctgt ttagttggtg      8520
tgaaaagtga tctacctaaa ttacgtgtta tacacgtaga ccttggtaga tctagtatat      8580
aaataaatag gtcg                                                       8594
```

ToCV-RNA2 Sequence

```
gaaatactag tccaggtgtt tcctgtgggt acgcgatgag cctcccccac gttaattacc        60
ccaccgtcac taggtggacg tgattgtggg cgctgccggc tttggccgtc ggtgtctgct       120
ttcattttct agtgacgcta tgttaattaa tattaacata taacaaaaca aaacaaaaag       180
aaaataaaat agcaggccga gtaccataaa cttaactgaa ctgctcgagt ttttgcacat       240
gcctacggct ggtatgttgc agcccctctt gcactttgtt atagtgtatg ttatggtctc       300
gcctataccc ccttattctt tccttgtgaa atttaaatga ccaaatagtg agtttgtacg       360
atcaaattac tataaaccgg aaagttcaga gttactctgc actttcgaag tcggcaagtt       420
ggtaccgaca gttgatccat acgaattcac tattgacgtt atttagaaac gtgttttcgt       480
gctttactat ttgcaataat aaatgttaaa ggttgagtgg ttgtgcgttt acgtggtcag       540
atggaaactt tcttattaac ttacacatgt taatggttta gtcttcctct cgcgtactta       600
gaatcattca gtctagtaca acagtttcat tgtattgtcg acaataatta gcggtccagt       660
taactcagtt taacttgatt ccgtttgttt ttccgattct tttatttgaa ctgtattctt       720
gtctttagta ttatgagtat taaagctggt ttggattttg gtactacatt cagtactatt       780
agttgtttct ataataacaa attgttttca ttaaaactca atgggaccga gtacattcca       840
acttgtctct ccataactcc aaataatgag gtgatagtcg ggggcccttc tcaagtttta       900
gaagcttccg aaactccgtc ttgttatttc tatgatttga agagatgggt tggtgtcact       960
tcggtcaatt atgaggtcgt gaaagcgaag ataaacccaa tgtataanac gcgtttatct      1020
aataataagg tgtatataac tggtatcaat aaaggtttct cgaccgagtt ttcggttgag      1080
caacttatat tacattatgt taatacttta gttcgattat tctcaaaaac agaaaactta      1140
aaaataaccg atctcaatgt gtctgttccg gctgattaca agtctgggca gagacttttc      1200
atgcaggcag tttgttcctc tttgggtttc aatttacgtc gcatagtcaa tgaaccgtcg      1260
gctgccgcta tttactgcgt ttctaaatat ccgcagtatg cttatttcta tatttacgat      1320
```

```
tttggtggcg gtactttcga cacttcttta atagtgcgat acggtaagtt tgtcactgtt      1380
gctgataccc agggagattc gtttcttggt gggcgagata tagataaagc catatcgaaa      1440
ttcataatgg ataaaaatgc tttgagcgcc ccactgtcgg cagatatgtt agcgtctata      1500
aaggaagaga caaattctac cgggcgcagt tcatacaata taataagtga tgatgggagt      1560
ataatcaata ttcagtttac gtttgacgat ttggtcaagt gcgttgaacc attcactaga      1620
cgcagttttt caatacttcg aagtctcgtt tctcgtgaca aaacttcgaa tggagcgctg      1680
tttcttgtag gtggttcctc attgcttaga ccgattcaga atagagcaga tggttttgcg      1740
cgtaatcatg ggttagctct cattatagac ccagatctca gagctgctgt gtcatttggt      1800
tgttcaatgc tccatgcaca agaggattct gggaatatga catatataga ctgcaattca      1860
catccgttga tggatttggg tttatattgt catcctagga ttatcatcag aaaacccatg      1920
tctgttccgt acacacacaa gatcgagagg gaggtcacaa gatttatcac cactgctctc      1980
aatgtttatg aagggtctga ccttttcgta cttaataacg attggttaat cagcgcagat      2040
gttgactatt caaagtacgc taaaatgggt gagacattag tcagtgttta taagtacacc      2100
attgacggca ttttggaact gtcgatggct aacaaaacca ctggcaagtc gtgggtcctt      2160
ccgaacacct ttgccaggtc cgaaaagata gtcattagtg atttaacctt gactcaattg      2220
tcgaatgtcg atgaattggc cactatagtg tcaattttga gttattttga tgcaacattc      2280
aactatctta tttccatgtt caacactccg tcgatttttg agcgagaagt tggaaagata      2340
tctgacgcca aaggtcttta taaccgactc gttgaacaaa acagaaattt ctcttgaagg      2400
tttgttatgt tagtttgttt agtactgaga actaactgtc ttgttaaagt cttgtatttc      2460
tctgatactc caattgtcga gtccggccca ttagtaataa gaagctcgga tccgaccatt      2520
atagaagacc ttctagagta tttacctttc gtcactgagt catggaaccc gttgatactt      2580
ccgaaagagt caaaagactt ttctcagtag tgttcaagaa atcgaacaac gatgagaaga      2640
ttcataaatt ggctgattac ctcttgaaat attattcaac ggaaaacaga aatttgtata      2700
ggacgaccat taacaataaa gctttctctt ttacttctac atattcggtc tctggtggta      2760
aagtttatct cgacaccaag gaaccttggc aggttgtgaa actgatcatc atatatttgt      2820
acaaggtaga gcctggttat ctcaagaaaa ctaactactc tcccgaaaat cttttcgcta      2880
gattgaggtt cgatgattat tacgatgagt ggaataagta ttttgacaag gatgtcaacg      2940
attatcttgc cgaccatcct gaagagggat gtttgtacac catgaacgac attatgaagg      3000
aatatccggg tgaagaaccg actgcacaac tgactctgta cagggtttgt aactcacttg      3060
ggaaaaagat atcggttcga gaacttaagg agggaaaaat tagtgctttt aaaattgagt      3120
caaaaactga taacgctgaa attggagaag gcgttggtgg taatgctctg ttcaaagagt      3180
gtgttgagat tttgcagagt tatttgctct tgaattcttc caaagcgggg cgggagaaga      3240
tccgcgctaa tgctaagatt tttgagtgct atttgtcgag tctggttcca aaaggtctag      3300
ataagaaatt agcggcgaat ccattggttg tggctaaatt cgtcaatgcg tttacggtcc      3360
ggactgtgaa cagcaaggga tttggtgaca attttaaggc cgtgaaggaa ctgtctcctg      3420
aacttttgag tttcatcaag agagtgtttt tggttgacgc caggcttaat gaagatgtgt      3480
tgtttatagc acttccgaag aactctgtag ttgagattct tggtgacaaa tttgctgtcg      3540
gtgaatattt aaaagtgcaa aatgttttgc ctgcttcgag caactcgagt agcctcccac      3600
cagatatcga taagtgcgtg tctgatgcct tggttacttt tatgcggacg ttcggtaatt      3660
ttcaaccggc cttcatacta gacatttggt tattcgtgtt tgggaagatg accaccaatt      3720
ccaaactttg gagagaggat aatgagattc tagtgacggt gggggacgta gttgtgaaat      3780
cgactaccag taggttattg tcacatgtga aaaactgtgt cagacgggac tttccccagt      3840
tctcgaccga caatataatc agacagtggg ctaatttgag gggtgacaga gcgaagcaaa      3900
tgtttcaatt gatgaacttc agacccggtt tgttttcgag tataccgggt atcaaatcgt      3960
atatgcgctt cgacttcttt aagatgttag atttgtcgaa atgtactcgt gaagaaattg      4020
aaagttatca aacattacgt cgggtgacgg aaagtaggtc caataagact gcttgtgacg      4080
ataggtgttt ggagtcatgg atcttgagga aatgatcaaa gagttaggtc tggctaaagt      4140
tgagagattt ctcactgtct ataatcaagg taggtttgta gctttcggaa atatagaaac      4200
tctactctgc ctgattaatc aacattttgt ggagtttaat cctcaaagag ctaaactgga      4260
cattgaattg tctgaagtga gagatttctt gaggtgtttt gaatctttta gaagctttgg      4320
tttaaggaaa taatggagaa cagtgctgtt gcaaacactg gtgataacgg tggtggccgc      4380
aatcctctgg ttagaccgtt agatgatggc gtagatgacg aggtgcagaa cttaggcagg      4440
agggacgatt caacatctat cattccggct aatcctaatc gatcttccag ttgggctttg      4500
ttgaacccgg atactattaa ttataacgag ttaaggaaat tgaaggtaca ctccactagg      4560
ggtgatactc ttaccttgac tcaggaagag gagttcgaga agatactcga atccttttgc      4620
aggcgaataa tcggtgagac ccagatgacg gataagattt tcgctggttt ctacatgtcc      4680
atgtgtcagg ccattgtaaa ccaagggacc tcagttaaag cagccggtaa taacagtctt      4740
gaaaactact ttgaggtaga tggtgcgaaa tttaagtgga aaactccgga tttgataaat      4800
gaggttagac ccaaaatggc cgatgttcca aacgctatac gtcggtacgc cagaagtcat      4860
gaaaagatta ttcaggactt tatcaactcc ggtcttatta agcctgatta tcatttacaa      4920
```

```
ttcaaacatg gcgtattacc aagccatgtg tttggtaccg gcgattatat aaatggttcg      4980
ttgatgaata tctcagatga tcaacttatc tcgaacctgc ttatgaaaag aaacgctttg      5040
tgcaagggta acgagggcaa ggaactgtac aacgttaacc aacttgcatc gataactggt      5100
tgctaaatta tatggatgaa aatgaaatct atgaggatca agaggatctc tctgctcgtg      5160
gcggtggggg tttctattac cagactgtga ctttgggttc cggtgatgtg tttcccgttg      5220
atttagccct aacgagatcg gctgaatttg attcgacaat tttctcctta tatattaggt      5280
ttgtaattaa ggaggggaat gtgcgtttaa agatcgattt tggaaacaat tgggatgtga      5340
ctatgcaaca ggtgagactt tctggatggt ttgcggcgtt tggtaagatt gaaaaaccga      5400
gaaccgccag gtctggatgg tcatacccaa taaaattgtt taaagaggct ggagaagtca      5460
tagtgtccat tagtggttgg aggtgttata aaatttataa tgggtatccc gtagatcgcg      5520
ttgatttggt tctggcagta cccgttcgtg aagtaacagc cgatttaaaa cgaccattgg      5580
ttggggatta cgtcaatttt catgatgtat ttactcttat aaaaagtaaa aattctgaca      5640
tcactttacc taacccgagt ctgatattca acgattcaac aagtaaggtt aatttagatg      5700
tgtctccagg tgcgcgaaaa caaattgctc aagtcaaggc tgagaaagat ctgaacatta      5760
agaatcctga agactcaaag cctgacgttc ctaatgattc gctgagtgaa gtcgaatatc      5820
ataatcactc tgatgtttcc agtgttttca gattgtatta cacatggagg gttgaaagag      5880
attttgagag atcagttgag tcgagaattt tctttccgaa tatatttccg accgatttca      5940
caatacttca acaaatgtgg tatgggacga ctgccggtaa cgttgagact tttgtggaga      6000
taggtaaaaa tgaaaggaag ttcaacgttg gggtcgccgc ttggaaggac aatgcatttg      6060
gacatttcaa attagatggc ccgactttag cgaagatctc aacaattcca ggcaggttcg      6120
tggaccataa aattgaaaaa gaccctaaag gacatttgat tgtgagtgtt gataatactg      6180
ttcttgtgcg tactaacaaa ctgatagtca aaccgagtat tcagatcggt tgggaatttc      6240
atttaccatg ggacgcgatt ggaaagtatg gagttggtaa tttggccagg ttcacagaca      6300
tcatcaaacc caatcacatc aagtatgatg gttctgaagt tcctttggtg caaactaata      6360
ccatagaaag tgatcgttcc aagtctggtc ataaactatc tctagtcaac ctgaaaagtt      6420
tcagacgcat tagttctact gcagatttct tcttcgaacc accaccaccc tctgagtccg      6480
atgacaaaac ttgggaagat aagacccaaa ccgaagtgga tataaagaaa gaagagacta      6540
ttccgactaa tgagggtacc tcttcctctg acttgccgag tgagaagtca caatttgttg      6600
cagctaatca ttatctcttg tcgatagctg aagacaggaa cattttcaaa gcggctgtag      6660
atcggtacac cggattgggt ttctcaaagg atcaagctgt gttgataata tatcaattgg      6720
gggtaacatt cggcacttcc agaaattgtt gcagtgataa ttcatcgttt ctagtctgga      6780
agactgatac tggagcgcag gttataatca gaaaaggcgc ccactccagg tttctcaatt      6840
cactggttaa atatccttgc aacgtggaga gattgatact acgaagacgt agtgcggaga      6900
tattggcgtt gttgaggaac aagaaattgg cttacccaga cagattggcc aaaaagaaag      6960
gggtaagtca gggattcaca tatatggcat gtgattttct cgattacact gcggtaacgt      7020
taactcaaga agagcagttg actatgaatt ctgttgtgca gtacgtgaga ctccataata      7080
aacatcgaag aagcattgtg agcacgagtc agcttttctg atcgatggag gtcgtgtaca      7140
attcagacga tgttaacagt ggagttggat ccggtgaaga tgtaaatacg accgtggcca      7200
agaactttta ttctataact cacgttatga gtaattaccg taattacaca ccagacgaaa      7260
ttaaggatgc tgtgaatata ggttatgggt tactgaattt gtgtgagagg ttggatagag      7320
atgtaatact tgtgtctccg aattcaccag tttacaacaa ttaccgagat gccggaattc      7380
cacacaatct gcttatggaa aatactgcac ggtatttccc agtagttaat ccgagcgaat      7440
tgggaaaagt tctattgggt catatcagtg ttttaaagtt tttggagtat ttcacaagat      7500
acggggttga cgatatgcta atcacaagac tgttctcaaa ttttgtcttg tggtccactg      7560
gtgacgtgaa cgcagcattg tattcaattt atcaacagga ttttcacttt cctgttgagg      7620
taagagcgaa tttcaatttc ttatttttga attcaagtga aattgacaga aggttgagta      7680
acattagaag gaaaggttat ccaaactctg agaatttcaa ttggttcaaa aatatgataa      7740
gtaattactt atattttgat tttgtgttca gatactctgg tacaaaaatc aatatagaaa      7800
gaatcacaaa ttattatatt tgattttcaa tatgatttcc acttatttta ccttaatagg      7860
tttaattttc ttggtggtat tctgttttgt tttattgtgt tatttcgtct tcactgtcat      7920
taaattcttc gcgaaagata agataagtga cgatgattgt ccttatgtca ataatgttgc      7980
tccattcggg agtaacaggt ttaactcaca acctccaata gttcgttaaa gtactatctt      8040
acggttggat taataaaaat ttataaggag tttttgataa agttttcttc ttcgggttat      8100
ttacttcgta ttttataaaa tcccaaaaat atatgaaaga ttttacatat tgttctgttt      8160
agttggtgta aaattctatc tacctaaatt acgtgtcata cacgtaaacc ttggtagatt      8220
tagtatataa ataaataggt cg
```

Greek RNA1

ORF1a (L-Pro, MTR, HEL) Translation

MDSQQNLVSFNVDFTEKKIKDTFRVVKRHISNKYNKSFKKRLFLCSCDLNVSIAANSVSTAQGCSVRARIKTRMNVLRR
LCGIPHCNFNKLPVSVYRKFGHDFHRINSAIDRYLESSVGSTGGKSLEEIERIDGYTDHGDGYRVRLYSDVNICDVFVR
FNADVATGHDLKIRLQRQKNRITGQVRTVVHAAKNGFGLDFSLWCDSFLLTDSKTSGQKMVMDMIAAVAMKVPDVLPGF
DRLYSRVFSKNLEQIRSAFIRECKKYMQTSCDKANASHREKVTLKVKPSVAVKNVPDVTHKVGPDGAESFVVTYSDGKQ
RIIVNDDGAVRNLFNATLTNGKYFIHPKAMIPDKSFFSSKTTEYCWLNALAAVNKKIPDFVVPYPCLRMRVLYNCGLGS
VVEKHCKFVKAGLYHFDLRYCAPGKPINLNGYVGSKVDTDIPSLGKNINVIFDDLIGHYVQGTNLRSDNLLSSNIVNRL
SDRINTMFSKPKDPSIATSLTASEKRKVIDMFPELCLNFTDTSYSSHPIATAIRCCENFIMAKRCGNEDFIDAGGDVVH
YMLESVKNVHVCTPIVDTKDAHRHISRSAMLDTMWGLKDKVSFCEHKTETCTVEKTNIVAVEVYDMTLKQMAQALLSHK
AKRFDFSLIIPPEVCDPVCDVYQLNNSLHVTNNGDKIEYAYGDFGESYFHDRENLRDILRTQMFVYNGVVFKKSLECSR
DNLHFFSVVPCLGIKPGVYTLSTHYKKSESDKIEMTIPVREKLGQVVDKRTSVDRALTYSLIEYVMNTAARVDEKSSDY
LVSQYRAKKSMTIKGNKVVPNDCDLPVELVPGYLAIILAEGLRLREKIQYFAKISYHRHYSPSILKIFSLLIQEFVTFA
KAKCYDGFVWFLRKCLSDKVLDKVIFGERRIHDCQFIMRFKQEVTVVGESGKHSILGDSIHAFTEASERAETGLADFVD
GRPDLFKPDDYEKLYDLVHSGGGNAGFFSKTANYYSPFSLYYSIYNFFSNFISSASRISLYTNFFISVIDWMRYHSISC
FEYFKEVVKVIVCATMGKLKSGMQSLWDKLKRCAKEMKGLVNSDRLSLEKKLDDIFSADDEIAEALSSDVQSIVTEAVL
SEQLISSGGGGVISRPMKCVREMYDHSRRSLRWFCTWFLQQFKLYKRNILYIPEFLRESFDEIFEALKLKILSNDFIEM
TIQGMSFFVVNLTTSLFFGKIGLGLSTVSTVFYLSIKYTGLERRFLGTSFVNEHLSSAITTGGFSHPYMIPVRSLVMKS
CQNWMKQKLLRYETVSPIVTDLIAKDVLNASVYQYVTPYRVRMGLYFSILLALLKPTFAVVLLVCILLVAEHAKFYRTV
VVQANVHLSFASRLRRLNPTSKARAIKQLLVSKFDRKRFSKDNSDDKFCDTFEDFDCGPTEFGINSKDLVGAGEASTSE
QPVVELDYSGNEIKGKSWGDVTDDSISEPDIVDRPDGLVFSSLEITNQREKFLCDSSLLLSDALLQYPVVDVTTNVLTG
DNIVDTFSEFFFLEKKKLHVELGKINNVVDFYKSSLTGKNAFYNKVWSLRNKFDDSSLYVSENSKVWYKLKQGEKGHVQ
LEGVCKYTLDNKLVPFTYFYDDFQVTSDELMGMFSNRRCLALQSITPRESGFDLNSMLENVTFFNKPPGAGKTTTIVRN
MVRDIKSNVRCLALTCTNAGKKEIIHKLRKEGVNNAFSLVMTYDSFLINGGKMDTDMVYCDEIFMIHAGLWVALLSMLQ
FKKMECYGDKNQIPFINRVPNTLCQYSQKIFFLFRMIHDNVSYRCPPDVCYILSNLRDAAGNLLYPNGVKAVGPNSNLL
RSMFVVPLRSAEEVPYSPDVKMIAFTKPEKDDIMRHGRTADGKTNSAQTVNEVQGGTFPKVELYRLRQYDNPIYNDVNQ
FVVSISRHTEVMKYRVLSTKMHDTVGQHISSLDKVADHIIRECAFKQQV

ORF1b (RdRp) Translation

VHLNSRFNTYRLTIEGCYIPDTFSRPASSHLMAVNDFMSVVNPGLAWMQFLHRTILFEYGDFDMPPVEKMVLDFSKYKP
YVAGEFVVSKILGKGERTRPDSMKQGIISLSHRNFSAPRINERLDVYKTAERLCQNLVRSFDFSRLYENYDVILPDMFK
IDDWLQDRDGSKFGRIKRDMDHKLLVEQFESLKFMIKGEMKPKMDMSSYTAYNPPANIIYYNHLVSMYYSPLFLEVFDR
ISYCLSKKIVMYSGMNLETLGTLIGSKLQKPLTSYHTLEIDFSKFDKSQGILFKVYEGMIYRFFKFSEDYYTNIEATEY
FIKYRGRCGISGELGAQRRTGSPNTWLSNTLVTMGIILSVYDLDDIDLFLVSGDDSLIFSSKPLKNKTDEINRDFGFEA
KMIENSVPYFCSKYIISDRGKIRVVPDPVRFFEKLSVPIRVQDFMSDTLMREKFRSYKDLMKDFDYDTTCVLVDALVCY
RYNLPPMCSYAALCYIHCLCANFTTFRRVYESDLTVVI

ORF2 (P22) Translation

MDLTGCLRKLRQCDRLLERRGNAVSEVHLRAILIDLDECSECLMLCEQEYIRDTDCLMSFLLALKHYEIKFHMDMLNMI
YDFKLKTSQLIQDVFRIKVIIRVYLELCEIDPLLAMTEACQDILESGILNIGFISSALGHEPNILITILSMVDFIVVID
DRPLVFIPSKIRFVGDKLGSGHFRWFDKFFFGSDI

ORF3 (P5) Translation

MGDRFACDSLDSIARDIHSLYILFFYTFLVGVLLTFIMSCVRGVIEIYRVN

Greek RNA2

ORF1 (P4) Translation

MPTAGMLQPLLHFVIVYVMVSPIPPYSFLVKFK

ORF2 (HSP70h) Translation

MSIKAGLDFGTTFSTISCFYNNKLFSLKLNGTEYIPTCLSITPNNEVIVGGPSQVLEASETPSCYFYDLKRWVGVTSVN
YEVVKAKINPMYXTRLSNNKVYITGINKGFSTEFSVEQLILHYVNTLVRLFSKTENLKITDLNVSVPADYKSGQRLFMQ
AVCSSLGFNLRRIVNEPSAAAIYCVSKYPQYAYFYIYDFGGGTFDTSLIVRYGKFVTVADTQGDSFLGGRDIDKAISKF
IMDKNALSAPLSADMLASIKEETNSTGRSSYNIISDDGSIINIQFTFDDLVKCVEPFTRRSFSILRSLVSRDKTSNGAL
FLVGGSSLLRPIQNRADGFARNHGLALIIDPDLRAAVSFGCSMLHAQEDSGNMTYIDCNSHPLMDLGLYCHPRIIIRKP
MSVPYTHKIEREVTRFITTALNVYEGSDLFVLNNDWLISADVDYSKYAKMGETLVSVYKYTIDGILELSMANKTTGKSW
VLPNTFARSEKIVISDLTLTQLSNVDELATIVSILSYFDATFNYLISMFNTPSIFEREVGKISDAKGLYNRLVEQNRNF
S

ORF3 (P8) Translation

MLVCLVLRTNCLVKVLYFSDTPIVESGPLVIRSSDPTIIEDLLEYLPFVTESWNPLILPKESKDFSQ

ORF4 (P59) Translation

MEPVDTSERVKRLFSVVFKKSNNDEKIHKLADYLLKYYSTENRNLYRTTINNKAFSFTSTYSVSGGKVYLDTKEPWQVV
KLIIIYLYKVEPGYLKKTNYSPENLFARLRFDDYYDEWNKYFDKDVNDYLADHPEEGCLYTMNDIMKEYPGEEPTAQLT
LYRVCNSLGKKISVRELKEGKISAFKIESKTDNAEIGEGVGGNALFKECVEILQSYLLLNSSKAGREKIRANAKIFECY
LSSLVPKGLDKKLAANPLVVAKFVNAFTVRTVNSKGFGDNFKAVKELSPELLSFIKRVFLVDARLNEDVLFIALPKNSV
VEILGDKFAVGEYLKVQNVLPASSNSSSLPPDIDKCVSDALVTFMRTFGNFQPAFILDIWLFVFGKMTTNSKLWREDNE
ILVTVGDVVVKSTTSRLLSHVKNCVRRDFPQFSTDNIIRQWANLRGDRAKQMFQLMNFRPGLFSSIPGIKSYMRFDFFK
MLDLSKCTREEIESYQTLRRVTESRSNKTACDDRCLESWILRK

ORF5 (P9) Translation

MDLEEMIKELGLAKVERFLTVYNQGRFVAFGNIETLLCLINQHFVEFNPQRAKLDIELSEVRDFLRCFESFRSFGLRK*

ORF (CP) Translation

MENSAVANTGDNGGGRNPLVRPLDDGVDDEVQNLGRRDDSTSIIPANPNRSSSWALLNPDTINYNELRKLKVHSTRGDT
LTLTQEEEFEKILESFCRRIIGETQMTDKIFAGFYMSMCQAIVNQGTSVKAAGNNSLENYFEVDGAKFKWKTPDLINEV
RPKMADVPNAIRRYARSHEKIIQDFINSGLIKPDYHLQFKHGVLPSHVFGTGDYINGSLMNISDDQLISNLLMKRNALC
KGNEGKELYNVNQLASITGC

ORF7 (CPm) Translation

MDENEIYEDQEDLSARGGGGFYYQTVTLGSGDVFPVDLALTRSAEFDSTIFSLYIRFVIKEGNVRLKIDFGNNWDVTMQ
QVRLSGWFAAFGKIEKPRTARSGWSYPIKLFKEAGEVIVSISGWRCYKIYNGYPVDRVDLVLAVPVREVTADLKRPLVG
DYVNFHDVFTLIKSKNSDITLPNPSLIFNDSTSKVNLDVSPGARKQIAQVKAEKDLNIKNPEDSKPDVPNDSLSEVEYH
NHSDVSSVFRLYYTWRVERDFERSVESRIFFPNIFPTDFTILQQMWYGTTAGNVETFVEIGKNERKFNVGVAAWKDNAF
GHFKLDGPTLAKISTIPGRFVDHKIEKDPKGHLIVSVDNTVLVRTNKLIVKPSIQIGWEFHLPWDAIGKYGVGNLARFT
DIIKPNHIKYDGSEVPLVQTNTIESDRSKSGHKLSLVNLKSFRRISSTADFFFEPPPPSESDDKTWEDKTQTEVDIKKE
ETIPTNEGTSSSDLPSEKSQFVAANHYLLSIAEDRNIFKAAVDRYTGLGFSKDQAVLIIYQLGVTFGTSRNCCSDNSSF
LVWKTDTGAQVIIRKGAHSRFLNSLVKYPCNVERLILRRRSAEILALLRNKKLAYPDRLAKKKGVSQGFTYMACDFLDY
TAVTLTQEEQLTMNSVVQYVRLHNKHRRSIVSTSQLF

ORF8 (P27) Translation

MEVVYNSDDVNSGVGSGEDVNTTVAKNFYSITHVMSNYRNYTPDEIKDAVNIGYGLLNLCERLDRDVILVSPNSPVYNN
YRDAGIPHNLLMENTARYFPVVNPSELGKVLLGHISVLKFLEYFTRYGVDDMLITRLFSNFVLWSTGDVNAALYSIYQQ
DFHFPVEVRANFNFLFLNSSEIDRRLSNIRRKGYPNSENFNWFKNMISNYLYFDFVFRYSGTKINIERITNYYI

ORF9 (P7) Translation

MISTYFTLIGLIFLVVFCFVLLCYFVFTVIKFFAKDKISDDDCPYVNNVAPFGSNRFNSQPPIVR

Appendix 2

Nucleotide sequence alignments

Greek ToCV RNA1 and the American and Spanish isolates

```
            1                                                          60
US      (1) GAAATAGTATTCGTGTGATTACACAAAGTACTAATTAACTTAGCTTTAAGGCTTTCTGGT
GR      (1) GAAATAGTATTCGTGTGATTACACAAAGTACTAATTAACTTAGCTTTAAGGCTTTCTGGT
SP      (1) GAAATAGTATTCGTGTGATTACACAAAGTACTAATTAATTAGCTTTAAGGCTTTCTGGT

            61                                                         120
US     (61) GTGTTGCCAGTTTGCCTGCCCGGTTGCCACTGTGTTCAGTGTCGACCTTATTATACCAGC
GR     (61) GTGTTGCCAGTTTGCCTGCCCGGTTGCCACTGTGTTCAGTGTCGACCTTATTATACCAGC
SP     (61) GTGTTGCCAGTTTGCCTGCCCGGTTGCCACTGTGTTCAGTGTCGACCTTATCATACCAGC

            121                                                        180
US    (121) ATTCACCGCCATAGCATTACTAGAAACAGCACGACGAGTACACTAAACTAATCTGCCT-G
GR    (121) ATTCACCGCCATAGCATTACTAGAAACAGCACGACGAGTACACTAAACTAATCTGCCTTG
SP    (121) ATTCACCGCCATAGCATTACTAGAAACAGCACGACGAGTACACTAAACTAATCTGCCTTG

            181                                                        240
US    (180) CTTTACTTTGCGTTCTCCTTTTCGTCCCGGAGAGCGCGCCCCTTGCCCGCCGTTGTGGGT
GR    (181) CTTTACTTTGCGTTCTCCTTTTCGTCTCGGAGAGCGCGCCCCTTGCCCGCCGTTGTGGGT
SP    (181) CTTTACTTTGCGTTCTCCTTTTCGTCTCGGAGAGCGCGCTCCTTGCCCGCCGTTGTGGGT

            241                                                        300
US    (240) TGTTAGTGCTCTCGCTATATAACCCCCTCTTGTTGTGATTTAGTAGTTTGTATTTTAGCT
GR    (241) TGTTAGTGTTCACC--ATATAACCCCCTCTTGTTGTGATTTAGTAGTTTGTATTTTAGCT
SP    (241) TGTTAGCGTTCACC--ATATAACCCCCTTTTGTTGTGATTTAGTAGTTTGTATTTAAGTC

            301                                                        360
US    (300) ATTATGGATTCTCAGCAAAACCTAGTTTCGTTTAACGTAGATTTTACTGAAAAGAAAATA
GR    (299) ATTATGGATTCTCAGCAAAACCTAGTTTCGTTTAACGTAGATTTTACTGAAAAGAAAATA
SP    (299) ATCATGGATTCTCAGCAAAACCTTGTTTCGTTTAACGTAGATTTTACTGAAAAGAAAATA

            361                                                        420
US    (360) AAAGATACGTTTAGAGTAGTTAAGAGGCATATTAGTAATAAATATAATAAATCCTTTAAG
GR    (359) AAAGATACGTTTAGAGTAGTTAAGAGGCATATTAGTAATAAATATAATAAATCCTTTAAG
SP    (359) AAAAATACGTTTAGAGTAGTTAAGAGGCATATCAGTAGTAGATATAATAAATCCTTTAAG

            421                                                        480
US    (420) AAGAGACTGTTTTTGTGCAGTTGCGATTTAAATGTCTCAATTGCCGCAAACTCAGTGTCG
GR    (419) AAGAGACTGTTTTTGTGCAGTTGCGATTTAAATGTCTCAATTGCCGCAAACTCAGTGTCG
SP    (419) AAGAGACTGTTTTTGTGCAGTTGTGATTTAAATGTCTCAATTGTCACAAACTCAGTGTCC

            481                                                        540
US    (480) ACTGCTCAGGGATGTGGTGTGAGGGCTCGGATAAAAACTCGCATGAACGTACTTAGACGG
GR    (479) ACTGCTCAGGGATGCAGTGTGAGGGCTCGGATAAAAACTCGCATGAACGTACTTAGACGG
SP    (479) ACGGTTCAGGGATGCAGTGTGAGGGCTCGGATAAAAACTCGCATGAACGTACTTAGACGG

            541                                                        600
US    (540) TTATGTGGGATACCGCATTGCAATTTCAACAAACTACCAGTTTCTGTATACAGAAAATTC
GR    (539) TTATGTGGGATACCACATTGCAATTTCAACAAATTACCAGTTTCTGTATACAGAAAATTC
SP    (539) TTATGTGGGATACCGCATTGCAATTTCAATAAATTACCAGTTTCTGTTTATAGAAAATTC
```

```
            601                                                       660
US    (600) GGTCACGATTTCCATAGAATAAATTCTGCAATAGACCGGTATTTGGAAAGTTCTGTGGGG
GR    (599) GGTCACGATTTCCATAGAATAAATTCTGCAATAGACCGGTATTTGGAAAGTTCTGTGGGG
SP    (599) GGTCATGATTTCTATAGAATAAATTCTGCAATTGACCGATATTTGGAAAGTTCTGTGGGG

            661                                                       720
US    (660) TCAACAGGTGAAAAATCTCTCGAGGAAATCGAGAGAATAGATGGTTACACGGACCATGGC
GR    (659) TCAACAGGTGGAAAATCTCTCGAGGAAATCGAGAGAATAGATGGTTACACGGACCATGGC
SP    (659) TCAACAGGTGGAAAATCTCTCGAGGAAATCGAGAGAATAGATGGTTACACGGACCATGGC

            721                                                       780
US    (720) GATGGTTATCGTGTAAAACTTTACTCTGACGTCAACATTTGTGACGTGTTTGTAAAGTTC
GR    (719) GATGGTTATCGTGTAAGACTTTACTCTGACGTCAACATTTGTGACGTGTTTGTAAGGTTC
SP    (719) GATGGTTATCGCGTAAAACTTTATTCTGACGTCAACATTTGTGACGTGTTTGTAAAGTTC

            781                                                       840
US    (780) AATGCTGATGTTGCCACTGGTCACGATTTGAAGATCAGGCTCCAGAGACAGAAGAATAGA
GR    (779) AATGCTGATGTTGCCACTGGTCACGATTTGAAGATCAGGCTCCAGAGACAGAAGAATAGA
SP    (779) AATGCTGATGTTGCCACTGGTCACGATTTGAAGATCAGACTCCAGAGACAGAAGAATAGG

            841                                                       900
US    (840) ATTACCGGCCAAGTTAGGACGGTCGTTCATACTGCGAAGAATGGTTTTGGGCTTGATTTT
GR    (839) ATTACCGGCCAAGTTAGGACGGTCGTTCATGCTGCGAAGAATGGTTTTGGGCTTGATTTT
SP    (839) ATTACCGACCAAGTTAGGACGGTCGTTCATGCTGCGAAGAATGGTTTTGGGCTTGATTTT

            901                                                       960
US    (900) TCTCTGTGGTGTGATAGTTTCCTTCTCACAGACAGTAAAACTTCCGGCCAGAAGATGGTT
GR    (899) TCTCTGTGGTGTGATAGTTTCCTTCTCACAGACAGTAAAACTTCCGGCCAGAAGATGGTT
SP    (899) TCTCTGTGGTGTGATAGTTTCCTTCTCACAGACAGTAAAACTTCCGGCCAGAAGATGGTT

            961                                                      1020
US    (960) ATGGATATGATAGCGGCTGTTGCTATGAAGGTTCCGGATGTTTTACCTGGATTTGGCCGT
GR    (959) ATGGATATGATAGCGGCTGTTGCTATGAAGGTTCCGGATGTTTTGCCTGGATTTGACCGT
SP    (959) ATGGATATGATAGCGGCTGTTGCTATGAAGGTTCCGGATGTTTTGCCTGGATTCGGCCGT

            1021                                                     1080
US   (1020) TTGTACTCCAGAGTTTTTAGTAAGAACTTTAAGCAGATCAGATCCGCTTTTATCAGAGAG
GR   (1019) TTGTACTCCAGAGTTTTTAGTAAGAACCTTGAGCAGATCAGATCCGCTTTTATCAGAGAG
SP   (1019) TTGTACTCCAGAGTTTTTAGTAAGAACTTTAAGCAGATCAGATCCGCTTTTATCAGAGAG

            1081                                                     1140
US   (1080) TGCAAGAAGTACATGCAAACTTCTTGTGATAAGGCTAATGCAAGTCACCGCGAGAAAGTG
GR   (1079) TGCAAGAAGTACATGCAAACTTCTTGTGATAAGGCTAATGCAAGTCACCGCGAGAAAGTG
SP   (1079) TGCAAGAAGTATATGCGAACTTCTTGTGATAAGGCTAATGCGAGTCACCGCGAGAAAGTG

            1141                                                     1200
US   (1140) ACACTCAAAGTGAAGCCGTCAGTCGCGGTGAAGAACGTGCCAGATGTGACCCATAAGGTC
GR   (1139) ACACTCAAAGTGAAGCCGTCAGTCGCGGTGAAGAACGTGCCAGATGTGACCCATAAGGTC
SP   (1139) ACACTCAAAGTGAAACCGTCAGTCGCGGTGAAGAACGTGCCTGATGTGACCCATAAGGTC

            1201                                                     1260
US   (1200) GGACCCGATGGTGCAGAATCGTTCGTGGTCACTTATTCTGATGGGAAACAACGCATAATT
GR   (1199) GGACCCGACGGTGCAGAATCGTTCGTGGTCACTTATTCTGATGGGAAACAACGCATAATT
SP   (1199) GGACCCGACGGTGCGGAATCGTTTGTGGTCACCTATTCTGATGGGAAACAGCGCATAGTT

            1261                                                     1320
US   (1260) GTGAATGATGACGGTGCCGTTAGGAATTTGTTTAACGCTACTCTCACCAACGGTAAATAT
GR   (1259) GTGAATGATGACGGTGCCGTTAGGAATTTGTTTAACGCTACTCTCACCAACGGTAAATAT
SP   (1259) GTGAATGATGATGGTGCCGTTAGGAACTTGTTTAACGCTACTCTCACCAACGGTAAATAT
```

```
            1321                                                        1380
US   (1320) TTCATTCACCCGAAGGCTATGATCCCAGATAAAAGCTTTTTCTCGTCCAAAACAACTGAG
GR   (1319) TTCATTCACCCGAAGGCTATGATCCCAGATAAAAGCTTTTTCTCGTCCAAAACAACTGAG
SP   (1319) TTCATTCACCCGAAGGCCATGATTCCTGATAAAAGCTTTTTCTCGTCCAAAACAACTGAA

            1381                                                        1440
US   (1380) TATTGTTGGTTGAATGCCTTCGCTGCCGTCAATAAGAAAATACCTGATTTCGTTGTACCT
GR   (1379) TATTGTTGGTTGAATGCCCTCGCTGCCGTCAATAAGAAAATACCAGATTTCGTTGTACCT
SP   (1379) TATTGTTGGTTGAATGCCTTCGCTGCCGTCAATAAGAAAATACCAGATTTCGTTGTACCT

            1441                                                        1500
US   (1440) TACCCATGTCTTAGGATGCGGGTGTTGTACAACTGTGGTTTGGGCAGTGTAGTTGAGAAG
GR   (1439) TACCCATGTCTTAGGATGCGGGTGTTGTACAACTGTGGTTTGGGCAGTGTAGTTGAGAAG
SP   (1439) TACCCATGTCTTAGGATGCGGGTGTTGTACAACTGTGGTTTGGGCAGTGTAGTTGAGAGG

            1501                                                        1560
US   (1500) CATTGTAAGTTTGTAAAAGCCGGTTTATACCATTTTGATCTTAGGTACTGTGCGCCTGAA
GR   (1499) CATTGTAAGTTTGTAAAAGCCGGTTTATACCATTTTGATCTTAGGTACTGTGCGCCTGGA
SP   (1499) CATTGTAAGTTCGTAAGAGCCGGTTTATACCATTTTGATCTTAGGTACCGTACGCCTGGA

            1561                                                        1620
US   (1560) AAACCCATCAATCTCAATGGTTATGTTGGTTCCAAGGTTGATACTGATATACCGTCCCTA
GR   (1559) AAACCCATCAATCTCAATGGTTATGTTGGTTCCAAGGTTGATACTGATATACCGTCCCTA
SP   (1559) AAACCGATCGATCTCAATGGTTATGTTGGTTCAAAGGTTGACACTGATATACCGTCCCTA

            1621                                                        1680
US   (1620) GGTAAGAACATAAATGTCATCTTTGATGATTTGATCGGACATTATGTTCAAGGTACTAAT
GR   (1619) GGTAAGAACATAAATGTCATCTTTGATGATTTGATCGGACATTATGTTCAAGGTACTAAT
SP   (1619) GGTAAGAACATAAATGTCATTTTTGATGATTTGATCGGACATTATGTTCAGGGTACTAAT

            1681                                                        1740
US   (1680) TTGAGATCTGACAATCTGTTATCTAGCAACATCGTTAATAGATTGTCTGACAGAATAAAC
GR   (1679) TTGAGATCTGACAATCTGTTATCTAGCAACATCGTTAATAGATTGTCTGACAGAATAAAC
SP   (1679) TTGAGATCTGACAATCTGTTATCTAGCAACATCGTTAATAGGTTGTCTGACAGAATAAAC

            1741                                                        1800
US   (1740) ACAATGTTTAGCAAACCGAAAGATCTCTCGATAGCGACTTCGCTCACCGCATCAGAGAAA
GR   (1739) ACAATGTTTAGCAAACCGAAAGATCCCTCGATAGCGACTTCGCTCACCGCATCAGAGAAA
SP   (1739) ACAATGTTTAGCAAACCGAAAGATCTCACAATAGCGACTTCGCTCACCGCATCAGAGAAA

            1801                                                        1860
US   (1800) CGTAAAGTGATCGATATGTTTCCTGAGTTGTGTTTGAATTTCACCGACACATCTTACTCC
GR   (1799) CGTAAAGTGATCGATATGTTTCCTGAGTTGTGTTTGAATTTCACCGACACATCTTACTCC
SP   (1799) CGTAAAGTGATCGATATGTTTCCTGAGTTGTGTTTGAACTTCACAGACACATCTTATTCC

            1861                                                        1920
US   (1860) TCTCATCCTATTGCCACTGCAATTAGGTGTTGTGAGAACTTCATCATGGCCAAACGATGT
GR   (1859) TCTCATCCTATTGCCACTGCAATTAGGTGTTGTGAGAACTTCATCATGGCCAAACGATGT
SP   (1859) TCTCATCCTATTGCCACTGCAATTAGGTGTTGTGAGAACTTCATCATGGCCAGACGATGT

            1921                                                        1980
US   (1920) GGTAATGAGGATTTTATTGATGCTGGTGGTGATGTTGTGCACTATATGTTAGAGAGTGTG
GR   (1919) GGTAATGAGGATTTTATTGATGCTGGTGGTGATGTTGTGCACTATATGTTAGAGAGTGTG
SP   (1919) GGTTATGAGGACTTTATTGATGCTGGTGGCGATGTTGTGCACTACATGTCAGAGAGTGTG
```

```
             1981                                                        2040
US  (1980) AAAAACGTTCACGTTTGCACACCAATAGTTGATACCAAGGATGCCCATAGACACATTTCG
GR  (1979) AAAAACGTTCATGTTTGCACACCAATAGTTGATACCAAGGATGCCCATAGACACATTTCG
SP  (1979) AAAAACGTTCACGTTTGCACACCAATAGTTGATTCTAAAGATGCCCATAGACACATTTCG

             2041                                                        2100
US  (2040) AGGAGTGCTATGCTCGATACCATGTGGGGACTTAAGGACAAGGTGTCTTTCTGTGAGTAT
GR  (2039) AGGAGTGCTATGCTCGATACCATGTGGGGACTTAAGGACAAGGTGTCTTTCTGTGAGCAT
SP  (2039) AGGAGTGCTATGCTCGATACCATGTGGGGACTTAAGGACAAGGTGTCTTTCTGTGAGCAT

             2101                                                        2160
US  (2100) AAAACTGAAACTTGCACAGTGGAGAAAACTAATATTGTAGCGGTGGAGGTTTATGATATG
GR  (2099) AAAACTGAAACTTGCACAGTGGAGAAAACTAATATTGTAGCGGTGGAGGTTTATGATATG
SP  (2099) AAAACTGAAACTTGCACGGTGGAGAAAACTAATATCGTAGCGGTGGAGGTTTATGATATG

             2161                                                        2220
US  (2160) ACTCTGAAACAAATGGCACAAGCTCTTCTGTCACATAAGGCCAAGAGGTTTGACTTCAGC
GR  (2159) ACTCTGAAACAAATGGCACAAGCTCTTCTGTCACATAAGGCCAAAAGGTTTGACTTCAGC
SP  (2159) ACTCTGAAACAGATGGCACGAGCTCTTCTGTCACATAAGGCCAAAAGGTTTGACTTCAGC

             2221                                                        2280
US  (2220) CTTATAATACCACCTGAGGTGTGCGACCCGGTGTGTGATGTGTACCTGCTTAACAATAGC
GR  (2219) CTTATAATACCACCTGAGGTGTGCGACCCGGTGTGTGATGTGTACCAGCTTAACAATAGC
SP  (2219) CTTATAATACCACCTGAGGTGTGTGACTCGGTGTGTGATGTGTACCTGCTTAACAATAGC

             2281                                                        2340
US  (2280) CTACACGTTACCAATAACGGTGACAAGATTGAGTACGCATACGGCGACTTTGGTGAGAGT
GR  (2279) CTACACGTTACCAATAACGGTGACAAGATTGAGTACGCATACGGCGACTTTGGTGAGAGT
SP  (2279) CTACACGTTACCAATAACGGTGACAAGATTGAGTACGCATACGGCGACTTTGGTGAGAGT

             2341                                                        2400
US  (2340) TATTTTCATGATAGAGAGAACTTGAGAGACATTCTACGAACTCAAATGTTCGTTTATAAT
GR  (2339) TATTTTCATGATAGAGAGAACTTGAGAGACATTCTACGAACTCAAATGTTCGTTTATAAT
SP  (2339) TATTTTCATGATAGAGAGAACTTGAGAGACATTTTACGAACCCAAATGTTCGTTTATAAT

             2401                                                        2460
US  (2400) GGAGTTGTTTTCAAGAAGTCATTAGAGTGTTCTCGTGACAATCTGCATTTCTTCTCTGTG
GR  (2399) GGAGTTGTTTTCAAGAAGTCATTAGAGTGTTCTCGTGACAATCTGCATTTCTTCTCTGTG
SP  (2399) GGAGTTGTTTTCAAGAAGTCATTAGAGTGTTCTCGTGACAATCTGCATTTCTTCTCTGTG

             2461                                                        2520
US  (2460) GTTCCTTGCTTGGGCATTAAACCGGGTGTCTACACTTTGTCAACTCATTATAAAAAATCT
GR  (2459) GTTCCTTGCTTGGGCATTAAACCGGGTGTCTACACTTTGTCAACTCATTATAAAAAATCT
SP  (2459) GTTCCTTGCTTGGGCATTAAACCGGGTGTCTACACTTTGTCAACTCATTATAAAAAATCT

             2521                                                        2580
US  (2520) GAGTCTGATAAGATTGAGATGACAATACCCGTTCGGGAGAAACTCGGACAAGTCGTCGAC
GR  (2519) GAGTCTGATAAGATTGAGATGACAATACCCGTTCGGGAGAAACTCGGACAAGTCGTCGAC
SP  (2519) GAGTCTGACAAGATTGAGATGACAATACCTGTTCGGGAGAAACTCGGACAAGTCGTCGAC

             2581                                                        2640
US  (2580) AAGCGGACGAGTGTTGATAGGGCACTTACTTACAGCTTGATAGAATATGTCATGAACACA
GR  (2579) AAGCGGACGAGTGTTGATAGGGCACTTACTTACAGCTTGATAGAATATGTCATGAACACA
SP  (2579) AAGCGGACGAGTGTTGATAGGGCACTTACTTACAGCTTGATAGAATACGTCATGAACACA

             2641                                                        2700
US  (2640) GCGGCTAGGGTGGATGAGAAATCCTCTGACTATTTAGTTTCTCAGTACAGGGCCAAGAAG
GR  (2639) GCGGCTAGGGTGGATGAGAAATCCTCTGACTATTTAGTTTCTCAGTACAGGGCCAAGAAG
SP  (2639) GCGGCTAGGGTGGATGAGAAATCCTCTGACTATTTAGTTTCTCAGTACAGGGCCAAGAAG
```

```
            2701                                                        2760
US   (2700) AGTATGACTATAAAAGGTAATAAAGTGGTGCCTAATGACTGCGATTTACCCGTGGAGTTG
GR   (2699) AGTATGACCATAAAAGGTAATAAAGTGGTGCCTAATGACTGCGATTTACCTGTGGAGTTG
SP   (2699) AGTATGACCATAAAAGGTAATAAAGTGGTGCCTAATGACTGCGATTTACCCGTGGAGTTG

            2761                                                        2820
US   (2760) GTGCCAGGTTACCTTGCTATTATTTTGGCCGAGGGATTGAGGCTTAGGGAGAAGATTCAA
GR   (2759) GTGCCAGGTTACCTAGCTATTATTTTGGCCGAGGGATTGAGGCTTAGGGAGAAGATTCAA
SP   (2759) GTGCCAGGTTACCTAGCTATTATTTTGGCCGAGGGATTGAGGCTTAGGGAGAAGATTCAA

            2821                                                        2880
US   (2820) TACTTTGCAAAAATTAGTTATCATCGTCACTACAGCCCGTCGATTTTGAAGATATTCTCC
GR   (2819) TACTTTGCAAAAATTAGTTATCATCGTCACTACAGCCCGTCGATTTTGAAGATATTTTCC
SP   (2819) TACTTTGCAAAAATTAGTTATCATCGTCACTACAGCCCGTCAATTTTGAAGATATTCTCC

            2881                                                        2940
US   (2880) CTCTTGATACAAGAGTTTGTAACGTTCGCTAAGGCTAAGTGTTACGATGGGTTTGTTTGG
GR   (2879) CTCTTGATACAAGAGTTTGTAACGTTCGCTAAGGCTAAGTGTTACGATGGGTTTGTTTGG
SP   (2879) CTCTTGATACAAGAGTTTGTAACGTTCGCTAAGGCTAAGTGTTACGATGGGTTTGTTTGG

            2941                                                        3000
US   (2940) TTCCTTCGCAAGTGTCTGTCTGATAAGGTGCTTGATAAGGTGATATTTGGTGAGAGGCGT
GR   (2939) TTCCTTCGCAAGTGTCTGTCTGATAAGGTACTTGATAAGGTGATATTTGGTGAGAGGCGT
SP   (2939) TTCCTTCGCAAGTGTCTGTCTGATAAGGTGCCTGATAAGGTGATATTTGGTGAGAGGCGT

            3001                                                        3060
US   (3000) ATCCATGACTGCCAGTTCATAATGCGGTTCAAGCAAGAGGTCACTGTGGTCGGCGAGAGT
GR   (2999) ATCCATGACTGCCAGTTCATAATGCGGTTCAAGCAAGAGGTCACTGTGGTCGGCGAGAGT
SP   (2999) ATCCATGACTGCCAGTTCATAATGCGGTTCAAGCAAGAGGTCACTGTGGTCGGCGAGAGT

            3061                                                        3120
US   (3060) GGTAAGTATTCGATCCTCGGTGATTCTATCCACGCCTTCACTGAAGCAAGTGAGAGAGCT
GR   (3059) GGTAAGCATTCGATCCTCGGTGATTCTATCCACGCCTTCACTGAAGCAAGTGAGAGAGCT
SP   (3059) GGCAAGCACTCGATCCTCGGTGATTCCATCCATGCCTTCACTGAAGCTAGTGAGAGAGCT

            3121                                                        3180
US   (3120) GAAACTGGACTGGCCGATTTTGTGGACGGTAGACCGGATTTGTTCAAGCCAGATGACTAT
GR   (3119) GAAACTGGACTGGCCGATTTTGTGGACGGTAGACCGGATTTGTTCAAGCCAGATGACTAT
SP   (3119) GAAACTGGACTGGCCGATTTTGTGGACGGCAGACCAGATTTGTTCAAGCCAGATGACTAT

            3181                                                        3240
US   (3180) GAGAAGTTGTATGATCTAGTTCATTCTGGTGGTGGAAACGCAGGGTTTTTTAGTAAGACG
GR   (3179) GAGAAGTTGTATGATCTAGTTCATTCTGGTGGTGGAAACGCAGGGTTTTTTAGTAAGACG
SP   (3179) GAAAAGTTGTATGATCTAGTTCATTCTGGTGGTGGAAACGCAGGGTTTTTTAGTAAGACG

            3241                                                        3300
US   (3240) GCTAATTATTACTCTCCGTTTTCTTTGTATTATTCTATTTACAATTTTTTCTCGAATTTT
GR   (3239) GCTAATTATTACTCTCCGTTTTCTTTGTATTATTCTATTTACAATTTTTTCTCGAATTTT
SP   (3239) GCTAATTATTACTCTCCGTTTTCTTTGTATTATTCTATTTACAATTTTTTCTCAAATTTT

            3301                                                        3360
US   (3300) ACTTCTTCAGCTTCTCGAATTTCCTTGTATACCAATTTCTTTATTTCAGTTATAGATTGG
GR   (3299) ATTTCTTCAGCTTCTCGAATTTCCTTGTATACCAATTTCTTTATTTCAGTTATAGATTGG
SP   (3299) ACTTCTTCAGCTTCTCGAATTTCCTTGTATACCAATTTCTTTATTTCAGTTATGGATTGG

            3361                                                        3420
US   (3360) ATGCGTTATCACTCTATCTCTTGTTTTGAATATTTTAAGGAGGTCGTAAAAGTCATAGTT
GR   (3359) ATGCGTTATCACTCTATCTCTTGTTTTGAATATTTTAAGGAGGTCGTAAAAGTCATAGTT
SP   (3359) ATGCGTTATCACCCTATCTCTTGTTTTGAATATTTTAAGGAGGTCGTAAAAGTCATAGTT
```

```
            3421                                                        3480
US   (3420) TGTGCTGCTATGGGGAAGTTGAAATCAGGTATGCAGTCTCTTTGGGACAAACTCAAGCGG
GR   (3419) TGTGCTACTATGGGGAAGTTGAAATCAGGTATGCAGTCTCTTTGGGACAAACTCAAGCGG
SP   (3419) TGTGCTGCTATGGGGAAGTTGAAATCAGGTATGCAGTCTCTTTGGGGCAAACTCAAGCGG

            3481                                                        3540
US   (3480) TGTGCTAAAGAGATGAAGGGGCTCGTGAACAGCGACAGGCTGTCTTTGGAGAAGAAACTT
GR   (3479) TGTGCTAAAGAGATGAAGGGGCTCGTGAACAGCGACAGGCTGTCTTTGGAGAAAAAACTT
SP   (3479) TGTGCTAAAGAGATGAAGGGGCTCGTGAACAGCGACAGGCTGTCTTTGGAGAAGAAACTT

            3541                                                        3600
US   (3540) GATGATATCTTCTCTGCGGACGACGAGATTGCTGAAGCATTGTCCTCGGATGTACAATCT
GR   (3539) GATGATATCTTCTCTGCGGACGACGAGATTGCTGAAGCATTGTCCTCGGATGTACAATCT
SP   (3539) GATGATATCTTCTCTGCGGACGACGAGATTGCTGAAGTGTTGTCCTCGGAAGTACAATCT

            3601                                                        3660
US   (3600) ATTGTTACTGAAGCAGTGCTTTCAGAGCAATTGATCTCGTCTGGAGGCGGTGGAGTCATA
GR   (3599) ATTGTTACTGAAGCAGTGCTTTCAGAGCAATTGATCTCGTCTGGAGGCGGTGGAGTCATA
SP   (3599) ATTGTTACTGAAGCAGTACTTTCAGAGCAATTGATCTCGTCTGGAGGCGGTGGAGTCATA

            3661                                                        3720
US   (3660) TCACGACCGATGAAATGCGTTCGTGAAATGTACGATCATTCACGGCGATCCTTGCGGTGG
GR   (3659) TCACGACCGATGAAATGCGTTCGTGAAATGTACGATCATTCACGGCGATCCTTGCGGTGG
SP   (3659) TCACGACCGATGAAATGCGTTCGTGAAATGTACGACCATTCACGGCGATCCTTGCGGTGG

            3721                                                        3780
US   (3720) TTTTGTACTTGGTTTCTCCAGAAATTTAAGCTTTACAAAAGAAATATTTTGTACATTCCT
GR   (3719) TTTTGTACTTGGTTTCTCCAGCAATTTAAGCTTTACAAAAGAAATATTTTGTACATTCCT
SP   (3719) TTCTGTACTTGGTTTCTCCAGCAATTTAAGCTTTACAAAAGAAATATTTTGTACATTCCT

            3781                                                        3840
US   (3780) GAATTTTTAAGGGAGTCCTTTGATGAGATCTTTGAGGCCCTAAAAATGAAGATTTTGTCT
GR   (3779) GAATTTTTAAGGGAGTCCTTTGATGAGATCTTTGAGGCCCTAAAATTGAAGATTTTGTCT
SP   (3779) GAATTTTTAAGGGAGTCCTTTGATGAGATCTTTGAGGCCCTAAAAATGAAGATTTTGTCT

            3841                                                        3900
US   (3840) AACGATTTCATCGAGATGACGATCCAAGGTATGTCTTTCTTTGTGGTTAATTTAACTACA
GR   (3839) AACGATTTCATCGAGATGACGATCCAAGGTATGTCTTTCTTTGTTGTTAATTTAACTACA
SP   (3839) AACGATTTCATCGAGATGACGATCCAAGGTGTGTCTTTCTTTGTGGTTAATTTAACTACA

            3901                                                        3960
US   (3900) TCGTTGTTTTTTGGTAAGATAGGTTTAGGTTTGAGTGCTGTATCTACTGTTTTCTATCTT
GR   (3899) TCGTTGTTTTTTGGTAAGATAGGTTTAGGTTTGAGTACTGTGTCTACTGTTTTCTATCTT
SP   (3899) TCGGTGTTTTTTGGTAAGATAGGTTTAGGTTTGAGTGCTGTGTCTACTGTTTTCTATCTT

            3961                                                        4020
US   (3960) TCCATTAAATACACTGGTTTAGAGAGAAGATTTCTTGGTACTTCATTTGTCAATGAACAT
GR   (3959) TCCATTAAATACACTGGTTTGGAGAGAAGATTTCTTGGTACTTCATTTGTCAATGAACAT
SP   (3959) TCCATTAAATACACTGGTTTGGAGAGGAGATTTCTTGGTACTTCATTTGTTAATGAACAT

            4021                                                        4080
US   (4020) CTCTCTTCCGCTATAACGACTGGTGGTTTTTCCCATCCGTACATGATTCCAGTGAGGTCT
GR   (4019) CTCTCTTCCGCTATAACGACTGGTGGTTTTTCCCATCCGTACATGATTCCAGTGAGGTCT
SP   (4019) CTCTCTTCTGCCATAACGACTGGTGGTTTTTCCCATCCGTACATGATTCCAGTGAGGTCT

            4081                                                        4140
US   (4080) TTGGTTATGAAAAGTTGCCAGAATTGGATGAAACAGAAACTGTTGAGGTATGAGACGGTC
GR   (4079) TTGGTTATGAAAAGTTGCCAGAATTGGATGAAACAGAAACTGTTGAGGTATGAGACGGTC
SP   (4079) TTGGTTATGAAAAGTTGCCAGAATTGGATGAAACAGAAACTGTTGAGGTATGAAACGGTC
```

```
            4141                                                        4200
US   (4140) TCTCCAATAGTTACTGACCTTATTGCTAAGGACGTGTTAAACGCGTCAGTTTATCAGTAT
GR   (4139) TCTCCAATAGTTACTGACCTTATTGCTAAGGACGTGTTAAACGCGTCAGTTTATCAGTAT
SP   (4139) TCTCCAATAGTTACTGACCTTATTGCTAAGGATGTGTTAAACGCGTCGGTTTATCAGTAT

            4201                                                        4260
US   (4200) GTAACTCCCTACAGAGTTAGAATGGGGTTGTATTTTAGTATTTTGTTGGCCCTACTGAAA
GR   (4199) GTAACTCCCTACAGAGTTAGAATGGGGTTGTATTTTAGTATTTTGTTGGCCCTACTGAAA
SP   (4199) GTAACTCCTTACAGAGTTAGAATGGGGTTGTATTTTAGTATTTTGTTGGCCCTACTGAAA

            4261                                                        4320
US   (4260) CCAACATTTGCTGGCGTGCTCCTGGTGTGTATTCTGTTGGTGGCGGAACATGCAAAATTT
GR   (4259) CCAACATTTGCTGTCGTGCTCCTGGTGTGTATTCTGTTGGTGGCGGAACATGCAAAATTT
SP   (4259) CCGACATTTGCTGGCGTGCTCCTGGTGTGTATTCTGTTGGTAGCGGAACATGCAAAATTT

            4321                                                        4380
US   (4320) TACAGAACTGTTGTTGTTCAAGCCAATGTTCATTTATCCTTTGCTTCTCGACTGAGGAGA
GR   (4319) TACAGAACTGTTGTTGTTCAAGCCAATGTTCATTTATCCTTTGCTTCTCGACTGAGGAGA
SP   (4319) TACAGAACTGTTGTTGTTCAAGCCAATGTTCATTTATCCTTTGCTTCTCGACTGAGGAGA

            4381                                                        4440
US   (4380) TTGAATCCTACTTCGAAGGCTAGAGCGATTAAGCAGTTACTCGTTTCCAAGTTCGATAGG
GR   (4379) TTGAATCCTACTTCGAAGGCTAGAGCGATTAAGCAGTTACTCGTTTCCAAGTTTGATAGG
SP   (4379) TTGAATCCTACTTCGAAGGCTAGAGCGATTAAGCAGTTACTTGTTTCCAAGTTTGATAGG

            4441                                                        4500
US   (4440) AAGAGATTTAGCAAGGATAACAGTGACGATAAGTTCTGCGATACATTTGAGGACTTTGAT
GR   (4439) AAGAGATTTAGCAAGGATAACAGTGACGATAAGTTCTGCGATACATTTGAGGACTTTGAT
SP   (4439) AAGAGATTTAACAAGGATAACAGTGACGATAAGTTCTGCGATACATTTGAAGACTTTGAT

            4501                                                        4560
US   (4500) TGCGGTCCGACCGAGTTTGGAATAAACTCTAAAGATTTGGTTGGTGCTGGGGAGGCCTCG
GR   (4499) TGCGGTCCGACCGAGTTTGGAATAAACTCTAAAGATTTGGTTGGTGCTGGGGAGGCCTCG
SP   (4499) TGCGGTCCGACCGAGTTTGGAATAAACTCTAAAGATATGGTTGGTGCTGGAGAGGCCTCG

            4561                                                        4620
US   (4560) ACGTCTGAACTGCCTGTGGTAGAGTTGGATTACTCCGGCAATGAGATCAAAGGGAAGTGT
GR   (4559) ACGTCTGAACAGCCTGTGGTAGAGTTGGATTACTCCGGCAATGAGATCAAAGGGAAGAGT
SP   (4559) ACGTCTGAACTGCCTGTAGTAGAGTTGGATTACTCCGGCAAGGAGATCAAAGGAAGAGAT

            4621                                                        4680
US   (4620) TGGGGAGATGTGACTGACGATTCTATTTCTGAGCCGGATATTGTGGACAGACCGGACGGT
GR   (4619) TGGGGAGATGTGACTGACGATTCTATTTCTGAGCCGGATATTGTGGACAGACCGGACGGT
SP   (4619) TGGGGAGATGTGACTGACGATACTATTTCTGAGCCGGATGTTGTGGACAGACCGGACGGT

            4681                                                        4740
US   (4680) CTGGTGTTTTCAAGTTTGGAAATAACTAATCAAAGGGAGAGATTTTGTGTGATAGCAGT
GR   (4679) CTGGTGTTTTCAAGTTTGGAGATAACTAATCAAAGGGAGAAATTCTTGTGTGATAGCAGT
SP   (4679) CTGGTGTTTTCAAGTTTGGAGATAACTAATCAAAGGGAGAAATTCTTGTGTGATAGCAGT

            4741                                                        4800
US   (4740) CTTCTTTTGAGTGATGCCCTTTTGCAATATCCAGTTGTGGATGTGACGACCAACGTCCTG
GR   (4739) CTTCTTTTGAGTGATGCTCTTTTGCAATATCCAGTTGTGGATGTGACGACCAACGTCCTG
SP   (4739) CTTCTTTTGAGTGATGCCCTTTTGCAATATCCAGTTGTGGATGTGACGACCAACGTCCTG

            4801                                                        4860
US   (4800) ACTGGTGATAATATTGTTGATACTTTTTCAGAGTTTTTCTTTTTAGAGAAAAAGAAATTA
GR   (4799) ACTGGTGATAATATTGTTGACACTTTTTCAGAGTTTTTCTTTTTAGAGAAAAAGAAATTA
SP   (4799) ACTGGTGACAATATTGTTGACACTTTTTCAGAGTTTTTCTTTTTAGAGAAAAAGAAATTA
```

```
            4861                                                    4920
US   (4860) CATGTGGAGTTAGGTAAGATTAATAATGTTGTTGATTTTTACAAATCCTCATTGACCGGG
GR   (4859) CATGTGGAGTTAGGTAAGATTAATAATGTTGTTGATTTTTACAAATCCTCATTGACCGGG
SP   (4859) CATGTGGAGTTAGGTAAGATTAATAATGTTGTTGATTTTTACAAATCTTCATTGACCGGG

            4921                                                    4980
US   (4920) AAGAATGCTTTTTATAATAAGGTTTGGAGCTTGCGAAACAAATTCGACGATTCGTCGTTA
GR   (4919) AAGAATGCTTTTTATAATAAGGTTTGGAGCTTGCGAAACAAATTCGACGATTCGTCGTTA
SP   (4919) AAGAATGCTTTTTATAATAAGGTTTGGAGCTTGCGAAACAAATTCGACGATTCGTCGTTA

            4981                                                    5040
US   (4980) TATGTTTCGGAAAACTCAAAGGTCTGGTACAAGTTGAAACAGGGTGAGAAGGGTCATGTT
GR   (4979) TATGTTTCGGAAAACTCAAAGGTCTGGTACAAGTTGAAACAGGGTGAGAAGGGTCATGTT
SP   (4979) TATGTTTCGGAAAACTCAAAGGTCTGGTACAAGTTGAAACAGGGTGAGAAAGGTCACGTT

            5041                                                    5100
US   (5040) CAACTTGAAGGTGTGTGCAAATACACGCTTGACAACAAGCTGGTTCCTTTCACTTACTTT
GR   (5039) CAACTTGAAGGTGTGTGCAAATACACGCTTGACAACAAGCTGGTTCCTTTCACTTACTTT
SP   (5039) CAACTTGAAGGTGTGTGCAAATACACGCTTGACAATAAGCTGGTTCCTTTCACTTACTTT

            5101                                                    5160
US   (5100) TATGATGACTTTCAGGTCACCAGTGATGAATTGATGGGCATGTTTTCCAACAGGAGATGT
GR   (5099) TATGATGACTTTCAGGTCACCAGTGATGAATTGATGGGCATGTTTTCCAACAGGAGATGT
SP   (5099) TATGATGACTTTCAGGTCACCAGTGATGAATTGATGGGCATGTTTTCCAACAGGAGATGT

            5161                                                    5220
US   (5160) TTGGCATTGCAGTCAATCACACCGCGTGAGTCCGGGTTTGACCTTAATAGCATGTTGGAG
GR   (5159) TTGGCATTGCAGTCAATCACACCGCGTGAGTCCGGGTTTGACCTTAATAGCATGTTGGAG
SP   (5159) TTGGCATTACAGTCAATCACACCGCGTGAGTCCGGGTTTGACCTTAATAGCATGTTGGAG

            5221                                                    5280
US   (5220) AATGTCACGTTTTTCAATAAACCTCCTGGTGCCGGTAAGACGACCACCATTGTGAGGAAT
GR   (5219) AATGTCACGTTTTTCAATAAACCTCCTGGTGCTGGTAAGACGACCACCATTGTGAGGAAT
SP   (5219) AATGTCACGTTTTTCAATAAACCTCCTGGCGCCGGTAAGACGACCACCATTGTGAGGAAT

            5281                                                    5340
US   (5280) ATGGTGAGAGACATCAAAGGTAATGTGAGGTGTTTGGCCCTGACCTGCRCAAATGCTGGG
GR   (5279) ATGGTGAGAGACATCAAAAGTAATGTGAGGTGTTTGGCCCTGACCTGCACAAATGCTGGG
SP   (5279) ATGGTGAGAGACATCAAAGGTAATGTGAGGTGTTTGGCCCTGACCTGCACAAACGCTGGG

            5341                                                    5400
US   (5340) AAGAAAGAGATCATACATAAGCTCAGGAAGGAAGGAGTCAACAATGCATTCAGCTTAGTG
GR   (5339) AAGAAAGAGATCATACATAAGCTCAGGAAGGAAGGAGTCAACAATGCATTCAGCTTAGTG
SP   (5339) AAGAAAGAGATCATACATAAGCTCAGGAAGGAAGGAGTCAACAATGCATTCAGCTTAGTG

            5401                                                    5460
US   (5400) ATGACATATGACTCTTTTCTCATCAACGGTGGTAAGATGGATACTGATATGGTCTATTGT
GR   (5399) ATGACATATGACTCTTTTCTCATCAATGGTGGTAAGATGGATACTGATATGGTCTATTGT
SP   (5399) ATGACATATGACTCTTTTCTTATCAATGGTGGTAAGATGGATACTGACATGGTATTATTGT

            5461                                                    5520
US   (5460) GATGAGATCTTTATGATTCATGCTGGACTGTGGGTGGCCCTATTATCAATGTTGCAATTC
GR   (5459) GATGAGATCTTTATGATTCATGCTGGACTGTGGGTGGCCCTATTATCAATGTTGCAATTC
SP   (5459) GATGAGATCTTTATGATTCATGCTGGACTGTGGGTGGCCCTATTATCAATGTTGCAATTC

            5521                                                    5580
US   (5520) AAGAAAATGGAGGTGTTACGGTGACAAAAACCAGATTCCATTTATCAATAGAGTGCCGAAT
GR   (5519) AAGAAAATGGAATGTTACGGTGACAAAAACCAGATTCCATTTATCAATAGAGTGCCGAAT
SP   (5519) AAGAAAATGGAGTGTTACGGTGACAAAAACCAGATTCCATTTATCAATAGAGTGCCGAAT
```

```
                5581                                                5640
US   (5580) ACTCTTTGTCAGTACTCCCAGAAGATATTTTTCTTGTTTAGGATGATCCACGACAACGTT
GR   (5579) ACTCTTTGTCAGTACTCCCAGAAGATATTTTTCTTGTTTAGGATGATCCACGACAATGTT
SP   (5579) ACTCTTTGTCAGTACTCCCAGAAGATATTTTTCTTGTTTCGGATGATCCACGACAATGTT

                5641                                                5700
US   (5640) TCATACAGATGCCCGCCAGATGTTTGCTACATTTTGTCTAATTTGAGAGATGCTGCCGGT
GR   (5639) TCATACAGATGCCCGCCAGATGTTTGCTACATTTTGTCTAATTTGAGAGATGCTGCCGGT
SP   (5639) TCATACAGGTGCCCGCCAGATGTTTGCTACATTTTGTCTAATTTGAGAGATGCTGCCGGT

                5701                                                5760
US   (5700) AATCTGTTGTATTCAAATGGTGTCAAGGCTATGGGGCCGAACAGCAATCTTCTCCGTTCT
GR   (5699) AATCTGTTGTATCCAAATGGTGTCAAGGCTGTAGGGCCGAACAGCAATCTTCTCCGTTCT
SP   (5699) AATCTGTTGTATCCAAATGGTGTCAAGGCTGTGGGACCAAACAGCAATCTTCTCCGTTCT

                5761                                                5820
US   (5760) ATGTTCGTGGTGCCACTGAGATCTGCCRAGGAAGTTCCATACAGTCCGGATGTGAAGATG
GR   (5759) ATGTTCGTGGTGCCACTGAGATCTGCCGAGGAAGTTCCATACAGTCCGGATGTGAAGATG
SP   (5759) ATGTTCGTGGTGCCACTGAGATCTGCCGAGGAAGTTCCATACAGTCCGGATGTGAAGATG

                5821                                                5880
US   (5820) ATTGCTTTCACCAAACCTGAGAAGGACGACATAATGCGGCATGGTAGAACAGCGGATGGT
GR   (5819) ATTGCTTTCACCAAACCTGAGAAGGACGACATAATGCGGCATGGTAGAACAGCGGATGGT
SP   (5819) ATTGCTTTCACTAAACCTGAGAAGGACGACATAATGAGGCATGGTAGAACAGCGGATGGT

                5881                                                5940
US   (5880) AAGACCAATTCAGCTCAGACTGTTAATGAGGTTCAAGGTGGGACTTTTCCCAAAGTCGAG
GR   (5879) AAGACCAATTCAGCTCAGACTGTTAATGAGGTTCAAGGTGGGACTTTTCCCAAAGTCGAG
SP   (5879) AAGACCAATTCAGCTCAGACTGTTAATGAGGTTCAAGGTGGGACTTTTCCCAAAGTGGAG

                5941                                                6000
US   (5940) TTGTACAGATTGAGGCAATATGATAATCCCATCTACAATGATGTCAACCAATTCGTGGTT
GR   (5939) TTGTACAGATTGAGGCAATATGATAATCCCATCTACAATGATGTCAACCAATTCGTGGTT
SP   (5939) TTGTACAGATTGAGGCAGTATGATAATCCCATCTACAATGATGTCAACCAATTCGTAGTT

                6001                                                6060
US   (6000) AGTATATCCAGACACACGGAGGTTATGAAATATAGAGTCTTGTCTACGAAAATGCACGAC
GR   (5999) AGTATATCCAGACACACGGAGGTTATGAAATATAGAGTCTTGTCTACGAAGATGCACGAC
SP   (5999) AGTATATCCAGACACACGGAGGTTATGAAATATAGAGTCTTGTCTACGAAAATGCACGAC

                6061                                                6120
US   (6060) ACTGTTGGACAACATATATCATCTTTGGATAAGGTTGCTGATCATATCATAAGGGAGTGT
GR   (6059) ACTGTTGGACAACATATATCATCTTTGGATAAGGTTGCTGATCATATCATAAGGGAGTGT
SP   (6059) ACTGTTGGACAACATATATCATCTTTGGATAAGGTAGCTGATCATATCATAAGGGAGTGT

                6121                                                6180
US   (6120) GCATTTAAACAGCAGGTTTAACACTTATCGGTTGACCATTGAAGGTTGTTATATTCCTGA
GR   (6119) GCATTTAAACAGCAGGTTTAACACTTATCGGTTGACCATTGAAGGTTGTTATATTCCTGA
SP   (6119) GCATTTAAACAGCAGGTTTAACACTTATCGGTTGACCATTGAAGGTTGTTATATTCCTGA

                6181                                                6240
US   (6180) CACTTTTTCGAGACCTGCCTCATCTCATTTGATGGCAGTCAACGATTTTATGTCAGTGGT
GR   (6179) CACTTTTTCGAGACCTGCCTCATCTCATTTGATGGCAGTCAACGATTTTATGTCAGTGGT
SP   (6179) CACTTTTTCGAGACCTGCCTCATCTCATTTGATGGCAGTCAACGATTTTATGTCAGTGGT

                6241                                                6300
US   (6240) CAACCCGGGATTAGCTTGGATGCAATTTTTGCACAGAACTATATTGTTTGAGTATGGTGA
GR   (6239) CAACCCGGGATTAGCTTGGATGCAATTTTTGCACAGAACTATATTGTTTGAGTATGGTGA
SP   (6239) CAACCCGGGATTAGCTTGGATGCAATTTTTGCACAGAACTATATTGTTTGAGTATGGTGA
```

```
            6301                                                    6360
US   (6300) TTTTGACATGCCACCTGTTGAGAAGATGGTCCTTGATTTTTCAAAATACAAGCCTTATGT
GR   (6299) TTTTGACATGCCACCTGTTGAGAAGATGGTCCTTGATTTTTCAAAATACAAGCCTTATGT
SP   (6299) TTTTGACATGCCACCTGTTGAGAAGATGGTCCTTGATTTTTCAAAATACAAGCCTTATGT

            6361                                                    6420
US   (6360) CGCAGGGGAGTTTGTTGTCTCGAAAATTCTTGGCAAAGGTGAGAGGACGAGACCGGACAG
GR   (6359) CGCAGGGGAGTTTGTTGTCTCGAAAATTCTTGGCAAAGGTGAGAGGACGAGACCGGACAG
SP   (6359) CGCAGGGGAGTTTGTTGTCTCGAAAATTCTTGGCAAAGGTGAGAGGACCAGACCGGACAG

            6421                                                    6480
US   (6420) TATGAAACAGGGGATAATCTCATTGTCACATAGAAATTTTTCTGCACCGAGAATAAATGA
GR   (6419) TATGAAACAGGGGATAATCTCATTGTCACATAGAAATTTTTCTGCACCGAGAATAAATGA
SP   (6419) TATGAAACAGGGGATAATCTCATTGTCACATAGAAATTTTTCTGCACCGAGAATAAATGA

            6481                                                    6540
US   (6480) ACGTCTGGACGTTTATAAGACTGCTGAACGTTTATGTCAGAATCTCGTTAGATCTTTCGA
GR   (6479) ACGTCTGGACGTTTATAAGACTGCTGAACGTTTATGTCAGAATCTCGTTAGATCTTTCGA
SP   (6479) ACGTCTGGACGTTTATAAAACTGCTGAACGTTTATGTCAGAATCTCATCAGATCTTTCGA

            6541                                                    6600
US   (6540) CTTTTCGAGGTTGTATGAGAACTATGATGTGATTCTTCCTGACATGTTTAAAATTGACGA
GR   (6539) CTTCTCAAGGTTGTATGAGAACTATGATGTGATTCTTCCTGACATGTTCAAAATTGACGA
SP   (6539) CTTTTCGAGGTTGTATGAGAACTATGATGTGATTCTTCCTGACATGTTCAAAATTGACGA

            6601                                                    6660
US   (6600) TTGGTTGCAGGATAGAGATGGTTCGAAGTTTGGTCGGATAAAAAGGGATATGGACCACAA
GR   (6599) TTGGTTGCAGGATAGAGATGGTTCGAAGTTTGGTCGGATAAAAAGGGATATGGACCACAA
SP   (6599) TTGGTTGCAGGATAGAGATGGTTCCAAGTTTGGTCGGATAAAAAGGGATATGGATCACAA

            6661                                                    6720
US   (6660) ATTGTTGGTCGAACAGTTTGAGAGCTTAAAATTCATGATCAAAGGGGAGATGAAACCGAA
GR   (6659) ATTGTTGGTCGAACAGTTTGAGAGCTTAAAATTCATGATCAAAGGGGAGATGAAACCGAA
SP   (6659) ATTGTTGGTCGAACAGTTTGAGAGCTTAAAATTCATGATCAAAGGGGAGATGAAACCGAA

            6721                                                    6780
US   (6720) GATGGATATGTCGTCCTATACAGCTTATAATCCACCGGCGAATATCATCTATTATAACCA
GR   (6719) GATGGATATGTCGTCCTATACAGCTTATAATCCACCGGCGAATATCATCTATTATAACCA
SP   (6719) GATGGATATGTCGTCCTATACAGCTTATAATCCACCGGCGAATATCATATTATAACCA

            6781                                                    6840
US   (6780) TCTGGTAAGTATGTATTATTCTCCGTTGTTTCTGGAGGTCTTTGATAGGATATCATACTG
GR   (6779) TCTGGTAAGTATGTATTATTCTCCGTTGTTTCTGGAGGTCTTTGATAGGATATCATACTG
SP   (6779) TCTGGTGAGTATGTATTATTCTCCGTTGTTTCTGGAGGTCTTTGATAGGATATCATACTG

            6841                                                    6900
US   (6840) TCTTAGCAAGAAGATAGTTATGTATTCCGGGATGAATCTAGAAACTCTCGGCACCCTGAT
GR   (6839) TCTTAGCAAGAAGATAGTTATGTATTCCGGGATGAATCTAGAAACTCTCGGCACCCTGAT
SP   (6839) TCTCAGCAAGAAGATAGTCATGTATTCCGGGATGAATCTAGAAACTCTCGGCACCCTGAT

            6901                                                    6960
US   (6900) TGGTTCTAAACTGCAGAAGCCATTGACATCATATCACACTTTGGAGATTGATTTCTCAAA
GR   (6899) TGGTTCTAAACTGCAGAAGCCGTTGACATCATATCACACTTTGGAGATTGATTTCTCAAA
SP   (6899) TGGTTCTAAACTGCAGAAGCCATTGACATCATATCACACTTTGGAGATCGATTTCTCAAA

            6961                                                    7020
US   (6960) GTTTGATAAGTCTCAAGGTATCCTATTTAAAGTTTATGAGGGGATGATTTACCGGTTTTT
GR   (6959) GTTTGATAAGTCCCAAGGTATCCTATTTAAAGTTTATGAGGGGATGATTTACCGGTTTTT
SP   (6959) GTTTGATAAGTCTCAAGGTATCCTATTTAAAGTTTATGAGGGGATGATTTACCGGTTTTT
```

```
            7021                                                    7080
US   (7020) CAAGTTTTCCGAGGATTACTATACCAACATAGAGGCCACTGAATACTTCATAAAGTATCG
GR   (7019) CAAGTTTTCCGAGGATTACTATACCAACATAGAGGCCACTGAATACTTCATAAAGTATCG
SP   (7019) CAAGTTTTCCGAGGATTACTATACTAATATAGAGGCTACTGAATACTTCATAAAGTATCG

            7081                                                    7140
US   (7080) TGGTAGGTGTGGAATCAGCGGGGAGTTGGGTGCACAAAGGAGAACGGGGTCACCGAACAC
GR   (7079) TGGTAGGTGTGGAATCAGCGGGGAGTTGGGTGCACAAAGGAGAACGGGGTCACCGAACAC
SP   (7079) TGGAAGGTGTGGAATCAGCGGGGAGTTGGGTGCACAAAGGAGAACGGGGTCACCGAACAC

            7141                                                    7200
US   (7140) TTGGTTGTCGAACACATTGGTTACTATGGGTATCATACTCAGTGTTTACGACCTGGATGA
GR   (7139) TTGGTTGTCAAACACATTGGTTACTATGGGTATCATACTCAGTGTTTACGACCTGGATGA
SP   (7139) TTGGTTGTCAAACACATTGGTTACTATGGGTATCATACTTAGTGTTTACGACCTTGATGA

            7201                                                    7260
US   (7200) TATTGATTTATTCTTAGTAAGTGGTGACGACAGTTTGATCTTTTCGAGTAAACCCTTGAA
GR   (7199) TATTGATTTATTCTTAGTAAGTGGTGACGACAGTTTGATCTTTTCGAGTAAACCCTTGAA
SP   (7199) TATTGATTTATTCTTAGTGAGTGGTGACGACAGTTTGATCTTTTCGAGTAAACCCTTGAA

            7261                                                    7320
US   (7260) GAATAAAACTGATGAGATAAACAGAGATTTCGGTTTTGAGGCTAAGATGATAGAGAATTC
GR   (7259) GAATAAAACTGATGAGATAAACAGAGATTTCGGTTTTGAGGCTAAGATGATAGAGAATTC
SP   (7259) GAATAAAACTGATGAGATAAACAGAGATTTCGGTTTTGAGGCTAAGATGATAGAGAATTC

            7321                                                    7380
US   (7320) AGTGCCGTATTTTTGCTCCAAATATATCATCAGTGATAGAGGAAAAATCAGAGTCGTTCC
GR   (7319) AGTGCCGTATTTTTGCTCCAAATATATCATCAGTGATAGAGGAAAAATCAGAGTCGTTCC
SP   (7319) AGTGCCGTATTTTTGTTCCAAATACATCATCAGTGATAGAGGAAAAATCAGAGTCGTTCC

            7381                                                    7440
US   (7380) TGATCCTGTGAGGTTTTTTGAGAAGTTGTCTGTCCCAATTCGAGTTCAAGATTTTATGAG
GR   (7379) TGATCCTGTGAGGTTTTTTGAGAAGTTGTCTGTCCCAATTCGAGTTCAAGATTTTATGAG
SP   (7379) TGATCCTGTGAGGTTTTTTGAGAAGTTGTCTGTCCCAATTCGAGTTCAAGATTTTATGAG

            7441                                                    7500
US   (7440) TGACACTCTCATGCGGGAGAAATTTAGGTCTTATAAGGACTTGATGAAGGACTTTGATTA
GR   (7439) TGACACTCTCATGCGGGAAAAATTTAGGTCTTATAAGGACTTGATGAAGGACTTTGATTA
SP   (7439) TGACACTCTCATGCGGGAAAAATTTAGGTCTTATAAGGACTTGATGAAGGATTTTGATTA

            7501                                                    7560
US   (7500) CGATACAACGTGCGTTTTGGTGGATGCTTTGGTGTGTTATAGGTACAATTTACCACCGAT
GR   (7499) CGATACAACGTGCGTTTTGGTGGATGCTTTGGTGTGTTATAGGTACAATTTACCACCGAT
SP   (7499) CGACACAACGTGCGTTTTGGTGGATGCTTTGGTGTGTTATAGGTACAATTTACCACCGAT

            7561                                                    7620
US   (7560) GTGTTCATATGCAGCGTTGTGTTATATTCATTGTCTGTGTGCAAATTTTACAACTTTCAG
GR   (7559) GTGTTCATATGCAGCGTTGTGTTATATTCATTGTCTGTGTGCAAATTTTACTACTTTCAG
SP   (7559) GTGTTCATATGCAGCGTTGTGTTATATTCATTGTTTGTGTGCAAATTTTACTACTTTCAG

            7621                                                    7680
US   (7620) AAGAGTCTATGAGAGCGATTTGACTGTTGTTATTTAGGTCAGTATGGATCTCACTGGTTG
GR   (7619) AAGAGTCTATGAGAGCGATTTGACTGTTGTTATTTAGGTCAGTATGGATCTCACTGGTTG
SP   (7619) AAGAGTCTATGAGAGCGATTTGACTGTTGTTATTTAGGTCAGTATGGATCTCACTGGTTG

            7681                                                    7740
US   (7680) TTTGCGTAAGCTTCGACAGTGTGATCGACTTCTTGAGAGGCTGGGTAATGACGTTTCAGA
GR   (7679) CTTGCGTAAGCTTCGACAGTGTGATCGACTTCTTGAGAGGCGGGGAAATGCCGTTTCAGA
SP   (7679) CTTGCGTAAGCTTCGACAGTGTGATCGACTTCTTGAGAGGCTGGGTAATGATGTTTCAGA
```

```
            7741                                                    7800
US   (7740) AGTTCATTTAAGAGCGATCTTAATTGATCTTGATGAGTGTTCTGAGTGTTTGATGCTTTG
GR   (7739) AGTTCATTTAAGAGCGATCTTAATTGATCTTGATGAGTGTTCTGAGTGTTTGATGCTTTG
SP   (7739) AGTTCATTTAAGAGCGATCTTAATTGATCTTGATGAGTGTTCTGAGTGTCTGATGCTTTG

            7801                                                    7860
US   (7800) TGAGCAGGAGTATATCAGAGACACGGACTGCCTTATGTCATTTCTGTTGGCGCTGAAACA
GR   (7799) TGAGCAGGAGTATATCAGAGACACGGACTGCCTTATGTCATTTCTGTTGGCGCTGAAACA
SP   (7799) TGAGCAGGAGTATATCAGAGACACGGACTGCCTTTTGTCATTTTTGTTGGTACTGAAACA

            7861                                                    7920
US   (7860) CTATGAGATTAAATTTCACATGGATATGTTGAATATGATTTATGACTTTAAACTGAAGAC
GR   (7859) CTATGAGATTAAATTTCACATGGATATGTTGAATATGATTTATGACTTTAAACTGAAAAC
SP   (7859) CTATGAGATCAAATTTCACATGGATATGTTGAATATGATTTATGACTTTAAACTGAAAAC

            7921                                                    7980
US   (7920) GTCCCAGTTGATTCAGGATGTTTTTAGAATTAAAGTAATCATCAGAGTGTATCTTGAGTT
GR   (7919) GTCCCAGTTGATTCAGGATGTTTTTAGAATTAAAGTAATCATCAGAGTGTATCTTGAGTT
SP   (7919) GTCCCAGTTGATTCAGGATGTTTTTAGAATTAAAGTAATCATCAGAGTGTATCTTGAGTT

            7981                                                    8040
US   (7980) GTGTGAAATTGATCCGCTTTTGGCTATGACTGAGGCTTGTCAAGACATTCTTGAGAGTGG
GR   (7979) GTGTGAAATTGATCCACTTTTGGCTATGACTGAGGCTTGTCAAGACATTCTTGAGAGTGG
SP   (7979) GTGTGAAATCGATCCACTTTTAGCTATGACTGAGGCTTGTCAAGACATTCTTGAGGGTGG

            8041                                                    8100
US   (8040) TATTTTGAACATCGGTTTCATCTCTTCGGCCCTCGGGCATGAACCGAACATACTTATCAC
GR   (8039) TATTTTGAACATCGGTTTCATCTCTTCGGCCCTCGGGCATGAACCGAACATACTTATCAC
SP   (8039) TATTTTGAACATCGGTTTCGTCTCTTCGGCCCTCGGGCACGAACCGAACATACTTATCAC

            8101                                                    8160
US   (8100) AATATTGTCGATGGTCGATTTTATAGTCGTCATTGATGATCGACCACTGGTCTTTATCCC
GR   (8099) AATATTGTCGATGGTCGATTTTATAGTCGTCATTGATGATCGACCACTGGTCTTTATCCC
SP   (8099) AATATTGTCGATGGTCGATTTTATAGTCGTCATTGATGATCGACCACTGGTCTTTATCCC

            8161                                                    8220
US   (8160) TTCAAAAATAAGGTTTGTTGGCGACAAGTTGGGGTCAGGTCATTTTAGGTGGTTTGATAA
GR   (8159) TTCAAAAATAAGGTTTGTTGGCGACAAGTTGGGGTCAGGTCATTTTAGGTGGTTTGATAA
SP   (8159) TTTAAAAATGAGGTTTGTTGGCGACAAGTTGGGGTCGGGTTATTTTAGGTGGTTTGATAA

            8221                                                    8280
US   (8220) GTTTTTCTTTGGGAGTGATATATAATCTGGTTAATATTCAAAGGATGGGTGATCGTTTTG
GR   (8219) GTTTTTCTTTGGGAGTGACATATAATCTGGTTAATATTCAAAGGATGGGTGATCGTTTTG
SP   (8219) GTTTTTCTTTGGGAGTGATATATAATCTGGTTAATATTCAAAGGATGGGTGATCGTTTTG

            8281                                                    8340
US   (8280) CTTGCGATTCTTTGGATAGTATAGCCAGAGATATACACTCTTTGTATATTTTGTTTTTCT
GR   (8279) CTTGCGATTCTTTGGATAGTATAGCCAGAGATATACACTCTTTGTATATTTTGTTTTTCT
SP   (8279) CTTGCGATTCTTTGGATAGTATAGCCAGAGATATACACTCTTTGTATATTTTGTTTTTCT

            8341                                                    8400
US   (8340) ATACGTTTCTAGTAGGTGTTTTGTTGACATTCATAATGTCTTGTGTTAGAGGTGTTATAG
GR   (8339) ATACGTTTCTAGTAGGTGTTTTGTTGACATTCATAATGTCTTGTGTTAGAGGTGTTATAG
SP   (8339) ATACGTTTCTAGTGGGTGTTTTGTTGACATTTATAATGTCTTGTGTTAGAGGTGTTATAG

            8401                                                    8460
US   (8400) AAATTTATCGTGTTAATTAGGGAGTTTTTGATAATGTTTCCTTCCTTTGGTTATTTACTT
GR   (8399) AAATTTATCGTGTTAATTAGGGAGTTTTTGATAATGTTTCCTTCCTTTGGTTATTTACTT
SP   (8399) AAATTTATCGTGTTAATTAGGGAGTTTTTGATAATGTTTCCTTCCTTTGGTTATTTACTT
```

```
            8461                                                    8520
US   (8460) CGTATTTTATAAAATCCAAAAAAATATATTATGCTTTTTACATATTGTTCTGTTTAGTTGG
GR   (8459) CGTATTTTATAAAATCCAAAAAAATATATTATGCTTTTTACATATTGTTCTGTTTAGTTGG
SP   (8459) CGTATTTTATAAAATCCAAAAAAATATATTATGCTTTTTACATATTGTTCTGTTTAGTTGG

            8521                                                    8580
US   (8520) TGTGAAAAGTGATCTACCTAAATTACGTGTTATACACGTAGACCTTGGTAGATCTAGTAT
GR   (8519) TGTGAAAAGTGATCTACCTAAATTACGTGTTATACACGTAGACCTTGGTAGATCTAGTAT
SP   (8519) TGTGAAAAGTGATCTACCTAAATTACGTGTTATACACGTAGACCTTGGTAGATCTAGTAT

            8581        8596
US   (8580) ATAAATAAATAGGTCG
GR   (8579) ATAAATAAATAGGTCG
SP   (8579) ATAAATAAATAGGTCG
```

Greek ToCV RNA2 and the American and Spanish isolates

```
            1                                                         60
US     (1) GAAATACTAGTCCAGGTGTTTCCTGTGGGTACGCGATGAGCCTCCCCCACGTTAATTACC
GR     (1) GAAATACTAGTCCAGGTGTTTCCTGTGGGTACGCGATGAGCCTCCCCCACGTTAATTACC
SP     (1) GAAATACAAGTCCAGGTGTTTCCTGTGGGTACGCGATGAGCCTCCCCCACGTTAATTACC

            61                                                       120
US    (61) CCACCGTCACTAGGTGGACGTGATTGTGGGCGCTGCCGGCTTCGGTCGTCGGTGTCTGCT
GR    (61) CCACCGTCACTAGGTGGACGTGATTGTGGGCGCTGCCGGCTTTGGCCGTCGGTGTCTGCT
SP    (61) CCACCGTCACTAGGTGGACGTGATTGTGGGCGCTGCCGGCTTCGGTCGTCGGTGTCTGCT

            121                                                      180
US   (121) TTCATTTTCTAGTGACGCTATGTTAATTAATATTAACATATAACAAAACAAAACAAGAAG
GR   (121) TTCATTTTCTAGTGACGCTATGTTAATTAATATTAACATATAACAAAACAAAACAAAAAG
SP   (121) TTCATTTTCTAGTGACGCTATGTTGATTAATATTAACATATAACAAAACAAAACAAAAAG

            181                                                      240
US   (181) AAAATAAAATAGCAGGCCGAGTACCATAAACTTAACTGAACTGCTCGAGTTTTTGCACAT
GR   (181) AAAATAAAATAGCAGGCCGAGTACCATAAACTTAACTGAACTGCTCGAGTTTTTGCACAT
SP   (181) AAAATAAAATAGCAGGCCGAGTACCATAAACTTAACTGAACTGCTCGAGTTTTTGCACAT

            241                                                      300
US   (241) GCCTACGGCTGGTATGTTGCAGCCCCTCTTGCACTTTGTTATAGTGTATGTTATGTTCTC
GR   (241) GCCTACGGCTGGTATGTTGCAGCCCCTCTTGCACTTTGTTATAGTGTATGTTATGGTCTC
SP   (241) GCCTACGGCTGGTATGTTGCAGCCCCTCTTGCACTTTGTTATAGTGTATGTTATGGTCTC

            301                                                      360
US   (301) GCCTATACCCCCTTACTCTTTCTTTGTGAAACTTAAATGACCAAATAGTGAGTTTGTACG
GR   (301) GCCTATACCCCCTTATTCTTTCCTTGTGAAATTTAAATGACCAAATAGTGAGTTTGTACG
SP   (301) GCCTATATCCCCTTACTTTTTCTTTGTGAAATTTAAATGACCAAACAGTGAGTTTGCACG

            361                                                      420
US   (361) ATCAAATTACTATAAACCGGAAAGTTCAGAGTTACTCTGCACTTTCGAAGTCGGCAAGTT
GR   (361) ATCAAATTACTATAAACCGGAAAGTTCAGAGTTACTCTGCACTTTCGAAGTCGGCAAGTT
SP   (361) ATCAAATTACTATAGACCGGAAAGTTCAGAGTTACTCTGCACTTTCGAAGTCGGCTAGTT

            421                                                      480
US   (421) GGTACCGACAGTTGATCCATATGAATTCA-CTATT--GACGTTATTTAGAAACGTGTTGT
GR   (421) GGTACCGACAGTTGATCCATACGAATTCA-CTATT--GACGTTATTTAGAAACGTGTTTT
SP   (421) GGTACCGACAGTTGATCTATACGAATTCAACTATTTGGACGTTATTTAGAAACGTGTTTT
```

```
                481                                                     540
US     (478)    CGTGCTTTACTATTTGTAATAATAAATGTTAGAGGTTGAGTGGTTGTGCGTTTACGTGGT
GR     (478)    CGTGCTTTACTATTTGCAATAATAAATGTTAAAGGTTGAGTGGTTGTGCGTTTACGTGGT
SP     (481)    TGTGAATTGTTATTTGTAATAATAAATGTTAAAGGTTGAGTGGTTGAGCATTTACGTCGT

                541                                                     600
US     (538)    CAGATGGAAACTTTCTTATTAACTTACACATGTTAATGGTTTAGTCTTCCTCTCGCGTAT
GR     (538)    CAGATGGAAACTTTCTTATTAACTTACACATGTTAATGGTTTAGTCTTCCTCTCGCGTAC
SP     (541)    CAGATGGAAACTTTCTTATTAACTCATACATGTTAATGGTTTAGTCTTCCTCTCGCGTAT

                601                                                     660
US     (598)    TTAGAATCATTCAGTCTAGTACAACAGTTTCATTGTATTGTCGACAATAATTAGCGGTCC
GR     (598)    TTAGAATCATTCAGTCTAGTACAACAGTTTCATTGTATTGTCGACAATAATTAGCGGTCC
SP     (601)    TTAGAATCATTCAGTCTAGTACAACAGTCTCATTGTATTATCGACAATAATTAGCGGTCC

                661                                                     720
US     (658)    AGTTAACTCAGTTTAACTTGATTCCGTTTGTTTTTCCGATTCTTTCATTTGAACTGTATT
GR     (658)    AGTTAACTCAGTTTAACTTGATTCCGTTTGTTTTTCCGATTCTTTTATTTGAACTGTATT
SP     (661)    AGTTAACTCAGTTTAACTTGATTCCGTTTGTTTTTCCCATTCTTTTATTTGAACTGTATT

                721                                                     780
US     (718)    CTTCTCTTTTGTATTATGAGTATTAAAGCTGGTTTGGATTTTGGTACTACATTCAGTACT
GR     (718)    CTTGTCTTTAGTATTATGAGTATTAAAGCTGGTTTGGATTTTGGTACTACATTCAGTACT
SP     (721)    CTTGTCTTTTGTATTATGAGTATTAAAGCTGGTTTGGATTTTGGTACTACGTTCAGTACT

                781                                                     840
US     (778)    ATTAGTTGTTTCTATAATAACAAATTGTTTTCATTAAAACTCAATGGGACCGAGTACATT
GR     (778)    ATTAGTTGTTTCTATAATAACAAATTGTTTTCATTAAAACTCAATGGGACCGAGTACATT
SP     (781)    ATTAGTTGTTTCTATAATAACAAATTGTTTTCATTAAAACTCAATGGGACCGAGTATATT

                841                                                     900
US     (838)    CCAACTTGTCTCTCCATAACTCCAAATAATGAGGTGATAGTCGGGGGCCCCTTCTCAAGTT
GR     (838)    CCAACTTGTCTCTCCATAACTCCAAATAATGAGGTGATAGTCGGGGGCCCCTTCTCAAGTT
SP     (841)    CCAACTTGTCTCTCCATAACTCCAAATAATGAGGTGATAGTCGGAGGCCCCTTCTCAAGTT

                901                                                     960
US     (898)    TTAGAAGCTTCCGAAACTCCGTCTTGTTATTTCTATGATTTGAAAAGATGGGTTGGTGTC
GR     (898)    TTAGAAGCTTCCGAAACTCCGTCTTGTTATTTCTATGATTTGAAGAGATGGGTTGGTGTC
SP     (901)    TTAGAAGCCTCCGAAACTCCGTCTTGTTATTTCTATGATTTGAAAAGATGGGTTGGTGTC

                961                                                    1020
US     (958)    ACTTCGGTCAATTATGAGGTCGTGAAAGCGAAGATAAACCCAACGTATAAAACGCGTTTA
GR     (958)    ACTTCGGTCAATTATGAGGTCGTGAAAGCGAAGATAAACCCAATGTATAANACGCGTTTA
SP     (961)    ACTTCGGTCAATTATGAGGTAGTGAAAGCGAAGATAAACCCAATGTATAAAACGCGTTTA

                1021                                                   1080
US   (1018)     TCTAATAATAAAGTGTATATAACTGGTATCAATAAAGGTTTCTCGACCGAGTTTTCGGTT
GR   (1018)     TCTAATAATAAGGTGTATATAACTGGTATCAATAAAGGTTTCTCGACCGAGTTTTCGGTT
SP   (1021)     TCTAATAATAAAGTGTATATAACTGGTATCAATAAAGGTTTCTCGACCGAGTTTTCGGTT

                1081                                                   1140
US   (1078)     GAGCAACTTATATTACATTATGTTAATACTTTAGTTCGATTATTCTCAAAAACAGAAAAC
GR   (1078)     GAGCAACTTATATTACATTATGTTAATACTTTAGTTCGATTATTCTCAAAAACAGAAAAC
SP   (1081)     GAGCAACTTATATTACATTATGTTAACACTTTAGTTCGATTGTTCTCAAAAACAGAAAAC

                1141                                                   1200
US   (1138)     TTAAAAATAACCGATCTCAATGTGTCTGTTCCGGCTGATTACAAGTCTGGGCAGAGACTT
GR   (1138)     TTAAAAATAACCGATCTCAATGTGTCTGTTCCGGCTGATTACAAGTCTGGGCAGAGACTT
SP   (1141)     TTAAAAATAACCGATCTCAACGTGTCTGTTCCGGCTGATTACAAGTCTGGGCAGAGACTT
```

```
            1201                                                        1260
US   (1198) TTCATGCAGGCAGTTTGTTCCTCTTTGGGTTTCAATTTACGTCGCATAGTCAATGAACCG
GR   (1198) TTCATGCAGGCAGTTTGTTCCTCTTTGGGTTTCAATTTACGTCGCATAGTCAATGAACCG
SP   (1201) TTCATGCAGGCAGTTTGTTCCTCTTTGGGTTTCAATTTACGTCGCATAGTCAATGAACCG

            1261                                                        1320
US   (1258) TCGGCTGCCGCTATTTACTGCGTTTCTAAATATCCGCAGTATGCTTATTTCTATATTTAC
GR   (1258) TCGGCTGCCGCTATTTACTGCGTTTCTAAATATCCGCAGTATGCTTATTTCTATATTTAC
SP   (1261) TCGGCTGCCGCTATTTACTGCGTTTCTAAATATCCGCAGTATGCTTATTTCTATATTTAC

            1321                                                        1380
US   (1318) GATTTTGGTGGCGGTACTTTCGACACTTCTTTAATAGTGCGATATGGTAAGTTTGTCACT
GR   (1318) GATTTTGGTGGCGGTACTTTCGACACTTCTTTAATAGTGCGATACGGTAAGTTTGTCACT
SP   (1321) GATTTTGGTGGCGGTACTTTCGACACTTCTTTAATAGTGAGATATGGTAAGTTTGTCACT

            1381                                                        1440
US   (1378) GTTGCTGATACCCAGGGAGATTCGTTTCTAGGTGGGCGAGATATAGATAAAACCATATCG
GR   (1378) GTTGCTGATACCCAGGGAGATTCGTTTCTTGGTGGGCGAGATATAGATAAAGCCATATCG
SP   (1381) GTTGCTGATACCCAGGGAGATTCGTTTCTTGGTGGGCGAGATATAGATAAAGCCATATCG

            1441                                                        1500
US   (1438) AAATTCATAATGGACAAAAATGCTTTGAACGCCCCACTGTCGGCAGATATGTTAGCGTCT
GR   (1438) AAATTCATAATGGATAAAAATGCTTTGAGCGCCCCACTGTCGGCAGATATGTTAGCGTCT
SP   (1441) AAGTTCATAATGGACAAAAATGCTTTGAACACCCCACTGTCGGCAGATATGTTAGCGTCT

            1501                                                        1560
US   (1498) ATAAAGGAAGAGACAAATTCTACCGGGCGCAGTTCATACAATATAATAAGTGATGATGGGG
GR   (1498) ATAAAGGAAGAGACAAATTCTACCGGGCGCAGTTCATACAATATAATAAGTGATGATGGGG
SP   (1501) ATAAAGGAAGAGACAAATTCTACCGGGCGCAGTTCATACAATATAATAAGTGATGATGGGG

            1561                                                        1620
US   (1558) AGTATAATCAATATTCAGTTTACGTTTGACGATTTGGTCAAGTGCGTTGAACCATTCGCT
GR   (1558) AGTATAATCAATATTCAGTTTACGTTTGACGATTTGGTCAAGTGCGTTGAACCATTCACT
SP   (1561) AGTATAATCAATATTCAGTTCACGTTTGACGATTTGGTCAAGTGCGTTGAACCATTCGCT

            1621                                                        1680
US   (1618) AGACGCAGTTTTTCAATACTTCGAAGTCTCGTTTCTCGTAACAAAACTTCGAATGGAGCG
GR   (1618) AGACGCAGTTTTTCAATACTTCGAAGTCTCGTTTCTCGTGACAAAACTTCGAATGGAGCG
SP   (1621) AGACGCAGTTTTTCAATACTTCGAAGTCTCGTTTCTCGTAACAAAACTTCGAATGGAGCG

            1681                                                        1740
US   (1678) CTGTTTCTTGTCGGTGGTTCCTCATTGCTTAGACCGATTCAGAATAGAGCAGATGGTTTT
GR   (1678) CTGTTTCTTGTAGGTGGTTCCTCATTGCTTAGACCGATTCAGAATAGAGCAGATGGTTTT
SP   (1681) CTGTTTCTTGTCGGTGGTTCTTCATTGCTTAGACCGATTCAGAATAGAGCAGATGGTTTT

            1741                                                        1800
US   (1738) GCGCGTAATCATGGGTTAGCTCTCATTATAGACCCAGATCTCAGAGCTGCTGTGTCATTT
GR   (1738) GCGCGTAATCATGGGTTAGCTCTCATTATAGACCCAGATCTCAGAGCTGCTGTGTCATTT
SP   (1741) GCGCGTAATCATGGGTTGGCTCTCATTATAGACCCAGATCTCAGAGCTGCCGTGTCATTT

            1801                                                        1860
US   (1798) GGTTGTTCAATGCTCCATGCACAAGAGGATTCTGGGAATATGACATATATAGACTGCAAT
GR   (1798) GGTTGTTCAATGCTCCATGCACAAGAGGATTCTGGGAATATGACATATATAGACTGCAAT
SP   (1801) GGTTGTTCAATGCTCCATGCACAAGAGGATTCTGGGAATATGACATATATAGACTGCAAT

            1861                                                        1920
US   (1858) TCACATCCGTTGATGGATTTGGGTTTATATTGTCATCCTAGGATTATCATCAGAAAACCC
GR   (1858) TCACATCCGTTGATGGATTTGGGTTTATATTGTCATCCTAGGATTATCATCAGAAAACCC
SP   (1861) TCACATCCGTTGATGGATTTGGGTTTATATTGTCATCCTAGGATTATCATCAGAAAACCC
```

```
            1921                                                        1980
US   (1918) ATGTCTGTTCCGTACACACACAAGATCGAGAGGGAGGTCACAAGATTTATTACCACTGCT
GR   (1918) ATGTCTGTTCCGTACACACACAAGATCGAGAGGGAGGTCACAAGATTTATCACCACTGCT
SP   (1921) ATGTCTGTTCCGTACACACACAAGATTGAGAGGGAGGTCACAAGATTTATCACCACTGCT

            1981                                                        2040
US   (1978) CTCAATGTTTATGAAGGGTCTGACCTTTTCGTACTTAACAACGATTGGTTAATCAGCGCA
GR   (1978) CTCAATGTTTATGAAGGGTCTGACCTTTTCGTACTTAATAACGATTGGTTAATCAGCGCA
SP   (1981) CTCAATGTTTATGGAGGGTCTGACCTTTTCGTACTTAACAACGACTGGTTAATCAGCGCA

            2041                                                        2100
US   (2038) GATGTTGACTATTCAAAGTACGCTAAAATGGGTGAGACATTAGTTAGTGTTTATAAGTAC
GR   (2038) GATGTTGACTATTCAAAGTACGCTAAAATGGGTGAGACATTAGTCAGTGTTTATAAGTAC
SP   (2041) GATGTTGACTATTCAAAGTATGCTAAAATGGGTGAGACATTAGTTAGTGTTTATAAGTAC

            2101                                                        2160
US   (2098) ACCATTGACGGCATTTTGGAACTGTCGATGGCTAATAAAACCACTGGCAAGTCGTGGGTC
GR   (2098) ACCATTGACGGCATTTTGGAACTGTCGATGGCTAACAAAACCACTGGCAAGTCGTGGGTC
SP   (2101) ACCATTGACGGCATTTTGGAACTGTCGATGGCTAATAAAACCACTGGCAAGTCGTGGGTC

            2161                                                        2220
US   (2158) CTTCCGAACACCTTTGCCAGGTCCGAAAAGATAGTCATTAGTGATTTAACCTTGACTCAA
GR   (2158) CTTCCGAACACCTTTGCCAGGTCCGAAAAGATAGTCATTAGTGATTTAACCTTGACTCAA
SP   (2161) CTTCCGAACACCTTTGCCAGGTCCGAAAAGATAGTCATTAGTGATTTAACCTTGACTCAA

            2221                                                        2280
US   (2218) TTGTCGAATGTCGATGAATTGGCCACTATAGTGTCAATTTTGAGTTATTTTGATACAACA
GR   (2218) TTGTCGAATGTCGATGAATTGGCCACTATAGTGTCAATTTTGAGTTATTTTGATGCAACA
SP   (2221) TTGTCGAATGTCGATGAATTGGCCACTATAGTGTCAATTTTGAGTTATTTTGATACAACA

            2281                                                        2340
US   (2278) TTCAACTATCTTATTTCCATGTTCAACACTCCGTCGATTTTTGAGCGAGAAGTTGGAAAG
GR   (2278) TTCAACTATCTTATTTCCATGTTCAACACTCCGTCGATTTTTGAGCGAGAAGTTGGAAAG
SP   (2281) TTCAACTATCTTATTTCCATGTTCAACACTCCGTCGATTTTTGAGCGAGAAGTTGGAAAG

            2341                                                        2400
US   (2338) ATATCTGACGCCAAAGGTCTTTATAACCGACTCGTTGAACAAAACAGAAATTTCTCTTGA
GR   (2338) ATATCTGACGCCAAAGGTCTTTATAACCGACTCGTTGAACAAAACAGAAATTTCTCTTGA
SP   (2341) ATATCTGACGCCAAAGGTCTTTATAACCGACTCGTTGAACAAAACAGAAATTTCTCTTGA

            2401                                                        2460
US   (2398) AGGTTTGTTATGTTAGTTTGTTTAGTACTGAGAACTAACTGTCTTGTTAAAGTCTTGTAT
GR   (2398) AGGTTTGTTATGTTAGTTTGTTTAGTACTGAGAACTAACTGTCTTGTTAAAGTCTTGTAT
SP   (2401) AGGTTTGTTATGTTAGTTTGTTTAGTACTGAGAACTAACTGTCTTGTTAAAGTCTTGTAT

            2461                                                        2520
US   (2458) TTCTCTGATACTCCAATTGTCGAGTCCGGCCCATTAGTAATAAGAAGCTCGGATCCGACC
GR   (2458) TTCTCTGATACTCCAATTGTCGAGTCCGGCCCATTAGTAATAAGAAGCTCGGATCCGACC
SP   (2461) TTCTCTGATACTCCAATTGTCGAGTCCGGCCCATTAGTAATAAGAAGCTCAGATCCGACC

            2521                                                        2580
US   (2518) ATTATAGAAGACCTTCTAGAGTATTTACCTTTCGTCACTGAGTCATGGAACCCGTTGATA
GR   (2518) ATTATAGAAGACCTTCTAGAGTATTTACCTTTCGTCACTGAGTCATGGAACCCGTTGATA
SP   (2521) ATTATAGAAGACCTTCTAGAGTACTTACCTTTCGTCACTGAGTCATGGAGCCCGTTGATA

            2581                                                        2640
US   (2578) CTTCCGAAAGAGTCAAAAGACTTTTCTCAGTAGTGTTCAAGAAATCGAACAACGATGAGA
GR   (2578) CTTCCGAAAGAGTCAAAAGACTTTTCTCAGTAGTGTTCAAGAAATCGAACAACGATGAGA
SP   (2581) CTTCCGAAAGAGTCAAAAGACTTTTCTCAGTAGTATTCAAGAAATCGAACAACGATGAGA
```

```
                2641                                                2700
US   (2638) AGATTCATAAATTGGCTGATTACCTCTTGAAATATTATTCAACGGAAAACAGAAATTTGT
GR   (2638) AGATTCATAAATTGGCTGATTACCTCTTGAAATATTATTCAACGGAAAACAGAAATTTGT
SP   (2641) AGATTCATAAATTGGCTGATTATCTCTTGAAATATTATTCAACGGAAAACAGAAATTTGT

                2701                                                2760
US   (2698) ATAGGACGACCATTAACAATAAAGCTTTCTCTTTTACTTCTACATATTCGGTMTCTGGTG
GR   (2698) ATAGGACGACCATTAACAATAAAGCTTTCTCTTTTACTTCTACATATTCGGTCTCTGGTG
SP   (2701) ATAGGACGACCATTAATAATAGAGCTTTCTCTTTTACTTCTACATATTCGGTCTCTGGTG

                2761                                                2820
US   (2758) GTAAAGTTTRTCTCGACACCAAGGAACCTTGGCAGGTTGTGAAAACTGATCATCATATATT
GR   (2758) GTAAAGTTTATCTCGACACCAAGGAACCTTGGCAGGTTGTGAAAACTGATCATCATATATT
SP   (2761) GTAAAGTTTATCTCGACACCAAGGAACCTTGGCAGGTTGTGAAGCTGATCATCATATATT

                2821                                                2880
US   (2818) TGTAYAAGGTAGAGCCTGGTTATCTCAAGAAAACTAACTACTCTCCCGAAAATCTTTTCG
GR   (2818) TGTACAAGGTAGAGCCTGGTTATCTCAAGAAAACTAACTACTCTCCCGAAAATCTTTTCG
SP   (2821) TGTACAAGGTAGAGCCTGGTTATCTTAAGAAAACTAACTACTCTCCCGAAAATCTTTTCG

                2881                                                2940
US   (2878) CTAGATTGAGGTTCGATGATTATTACGATGAGTGGAATAAGTATTTTGACAAGGATGTCA
GR   (2878) CTAGATTGAGGTTCGATGATTATTACGATGAGTGGAATAAGTATTTTGACAAGGATGTCA
SP   (2881) CTAGATTGAGGTTCGATGATTATTACGATGAGTGGAATAAGTATTTTGACAAGGATGTCA

                2941                                                3000
US   (2938) ACGATTATCTTGCCGACCATCCTGAAGAGGGATGTTTGTACACCATGAACGACATTATGA
GR   (2938) ACGATTATCTTGCCGACCATCCTGAAGAGGGATGTTTGTACACCATGAACGACATTATGA
SP   (2941) ACGATTATCTTGCCGACCATCCTGAAGAGGGATGTTTGTACACCATGAACGACATTGTGA

                3001                                                3060
US   (2998) AGGAATATCCGGGTGAAGAACCGACTGCACAACTGACTCTGTACAGGGTTTGTAACTCAC
GR   (2998) AGGAATATCCGGGTGAAGAACCGACTGCACAACTGACTCTGTACAGGGTTTGTAACTCAC
SP   (3001) AGGAATATCCGGGTGAAGAACCGACTGCACAACTAACTCTGTACAGGGTTTGTAACTCGC

                3061                                                3120
US   (3058) TTGGAAAAAAGATATCGGTTCGAGAACTTAAGGAGGGAAAAATTAGTGCTTTTAAAATTG
GR   (3058) TTGGGAAAAAGATATCGGTTCGAGAACTTAAGGAGGGAAAAATTAGTGCTTTTAAAATTG
SP   (3061) TTGGAAAAAAGATATCGGTTCGAGAACTTAAGGAGGGAAAAATTAGTGCTTTTAAAATTG

                3121                                                3180
US   (3118) AGTCAAAAACTGATAACGCTGAAATTGGAGAAGGCGTTGGTGGTAATGCTCTGTTCAAAG
GR   (3118) AGTCAAAAACTGATAACGCTGAAATTGGAGAAGGCGTTGGTGGTAATGCTCTGTTCAAAG
SP   (3121) AGTCAAAAACTGATAACGCTGAAATTGGAGAAGGTGTTGGTGGTAATGCTCTGTTCAAAG

                3181                                                3240
US   (3178) AGTGTGTTGAGACTTTGCAGAGTTATTTGCTCTTGAATTCTTCCAAAGCGGGGCGGGAGA
GR   (3178) AGTGTGTTGAGATTTTGCAGAGTTATTTGCTCTTGAATTCTTCCAAAGCGGGGCGGGAGA
SP   (3181) AGTGTGTTGAGACTTTGCAGAGTTATTTGCTCTTGAATTCTTCCAAAGCGGGGCGGGAGA

                3241                                                3300
US   (3238) AGATCCGCGCTAATGCTAAGATTTTTGAGTGCTATTTGTCGAGTCTGGTTCCAAAAGGTC
GR   (3238) AGATCCGCGCTAATGCTAAGATTTTTGAGTGCTATTTGTCGAGTCTGGTTCCAAAAGGTC
SP   (3241) AGATCCGCGCTAATGCTAAGATTTTTGAGTGCTATTTGTCGAGTCTGGTTCCAAAAGGTC

                3301                                                3360
US   (3298) TAGATAAGAAATTAGCGGCGAATCCATTGGTTGTGGCTAAATTCGTCAATGCGTTTACGG
GR   (3298) TAGATAAGAAATTAGCGGCGAATCCATTGGTTGTGGCTAAATTCGTCAATGCGTTTACGG
SP   (3301) TAGATAAGAAATTAGCGGCGAATCCATTGGTTGTGGCTAAATTCGTTAATGCGTTTACGG
```

```
                3361                                                    3420
US   (3358) TCCGGACTGTGAACAGCAAGGGATTTGGTGACAATTTTAAGGCCGTGAAGGAACTGTCTC
GR   (3358) TCCGGACTGTGAACAGCAAGGGATTTGGTGACAATTTTAAGGCCGTGAAGGAACTGTCTC
SP   (3361) TCCGGACTGTGAACAGCAAGGGATTTGGTGACAATTTTAAGGCCGTGAAGGAATTGTCTC

                3421                                                    3480
US   (3418) CCGAACTTTTGAGTTTCATCAAAAGAGTGTTTTTGGTTGACGCCAGGCTTAATGAAGATG
GR   (3418) CTGAACTTTTGAGTTTCATCAAGAGAGTGTTTTTGGTTGACGCCAGGCTTAATGAAGATG
SP   (3421) CTGAGATTTTGAGTTTCATCAAGAGAGTGTTTTTGGTCGACGCCAGGCTTAATGAAGATG

                3481                                                    3540
US   (3478) TGTTGTTTATAGCACTTCCGAAGAACTCTGTAGTTGAGATTCTTGGTGACAAATTTGCTG
GR   (3478) TGTTGTTTATAGCACTTCCGAAGAACTCTGTAGTTGAGATTCTTGGTGACAAATTTGCTG
SP   (3481) TGTTGTTTATAGCACTCCCGAAGAACTCTGTAGTTGAGATTCTTGGTGACAAATTTGCTG

                3541                                                    3600
US   (3538) TCGGTGAATATTTAAAAGTGCAAAATGTTTTGCCTGCTTCGAGCAACTCGAGTAGCCTCC
GR   (3538) TCGGTGAATATTTAAAAGTGCAAAATGTTTTGCCTGCTTCGAGCAACTCGAGTAGCCTCC
SP   (3541) TCGGTGAATATTTAAGAGTGCAAAATGTTTTGCCAGCTTCGAGCAACTCAAGTAGCCTCC

                3601                                                    3660
US   (3598) CACCAGATATCGATAAGTGCGTGTCTGATGCCTTGGTTACTTTTATGCGGACGTTCGGTA
GR   (3598) CACCAGATATCGATAAGTGCGTGTCTGATGCCTTGGTTACTTTTATGCGGACGTTCGGTA
SP   (3601) CACCAGATATCGATAAGTGCGTGTCTGATGCCTTGGTTACTTTCATGCAGTCGTTCGGTA

                3661                                                    3720
US   (3658) ATTTTCAACCGGCCTTCATACTAGACATTTGGTTATTCGTGTTTGGGAAGATGACCACCA
GR   (3658) ATTTTCAACCGGCCTTCATACTAGACATTTGGTTATTCGTGTTTGGGAAGATGACCACCA
SP   (3661) ATTTTCAACCAGCCTTCATACTAGATATTTGGTTATTTGTGTTTGGGAAGATGACCACCA

                3721                                                    3780
US   (3718) ATTCCAAACTTTGGAGAGAGGATAATGAGATTCTAGTGACGGTGGGGGACGTAGTTGTGA
GR   (3718) ATTCCAAACTTTGGAGAGAGGATAATGAGATTCTAGTGACGGTGGGGGACGTAGTTGTGA
SP   (3721) ATTCCAAACTTTGGAGAGAGGATAATGAGATCCTAGTGACGGTGGGGGACGTAGTTGTGA

                3781                                                    3840
US   (3778) AATCGACTACCAGTAGGTTATTGTCACATGTGAAAAACTGTGTTAGACGGGACTTTCCCC
GR   (3778) AATCGACTACCAGTAGGTTATTGTCACATGTGAAAAACTGTGTCAGACGGGACTTTCCCC
SP   (3781) AGCCAACTACCAGTAGGTTATTGTCACATGTGAAAAACTGTGTCAAACGGGACTTTCCCC

                3841                                                    3900
US   (3838) AGTTCTCGACCGACAATATAATCAGACAGAGGGCTAATTTGAGGGGTGACAGAGCGAAGC
GR   (3838) AGTTCTCGACCGACAATATAATCAGACAGTGGGCTAATTTGAGGGGTGACAGAGCGAAGC
SP   (3841) AGTTCTCGACCGACAATATAATCAGACAGTGGGCTAATTTGAGGGGTGACAGAGCGAAGC

                3901                                                    3960
US   (3898) AAATGTTTCAATTGATGAACTTTAGACCCGGTTTGTTTTCGAGTATACCGGGTATCAAAC
GR   (3898) AAATGTTTCAATTGATGAACTTCAGACCCGGTTTGTTTTCGAGTATACCGGGTATCAAAT
SP   (3901) AAATGTTTCAATTGATGAACTTCAGACCCGGTTTGTTTTCGAGTATACCGGGTATCAAAC

                3961                                                    4020
US   (3958) CGTATATGCGCTTCGACTTCTTTAAGATGTTAGATTTGTCGAAATGTACTCGTGAAGAAA
GR   (3958) CGTATATGCGCTTCGACTTCTTTAAGATGTTAGATTTGTCGAAATGTACTCGTGAAGAAA
SP   (3961) CGTATATGCGCTTCGACTTCTTTAAGATGTTAGATTTGTCGAAATGTACTCGTGAAGAAA

                4021                                                    4080
US   (4018) TTGAAAGTTATCAAACATTACGTCGGGTGACGGAAAGTAGGTCCAATAAGACTGCTTGTG
GR   (4018) TTGAAAGTTATCAAACATTACGTCGGGTGACGGAAAGTAGGTCCAATAAGACTGCTTGTG
SP   (4021) TTGAAAGTTATCAAACATTACGTCGGGTGACGGAAAGTAGGTCCAATAAGACTGCTTGTG
```

```
                4081                                                    4140
US   (4078) ACGATAGGCGTTTGGAGTCATGGATCTTGAGGAAATGATCAAAGAGTTAGGTCTGGCTAA
GR   (4078) ACGATAGGTGTTTGGAGTCATGGATCTTGAGGAAATGATCAAAGAGTTAGGTCTGGCTAA
SP   (4081) ACGATAGGTGTTTGGAGTCATGGATCTTGAGGAAATGATCAAAGAGTTAGGTCTGGCTAA

                4141                                                    4200
US   (4138) AGTTGAGAGATTTCTCACTGTCTATAATCAAGGTAGGTTTGTTGCTTTCGGAAATATAGA
GR   (4138) AGTTGAGAGATTTCTCACTGTCTATAATCAAGGTAGGTTTGTAGCTTTCGGAAATATAGA
SP   (4141) AGTTGAGAGATTTCTCACTGTCTATAATCAAGGTAGGTTTGTAGCTTTCGGAAATATAGA

                4201                                                    4260
US   (4198) AACTCTACTCTGCCTGATTAATCAACATTTTGTGGAGTTTAATCCTCAAAGAGCTAAACT
GR   (4198) AACTCTACTCTGCCTGATTAATCAACATTTTGTGGAGTTTAATCCTCAAAGAGCTAAACT
SP   (4201) AACTCTACTCTGCCTGATTAATCAACATTTTGTGGAGTTTAATCCTCAAAGAGCTAAACT

                4261                                                    4320
US   (4258) GGACATTGAATTGTCTGAAGTGAGAGATTTCTTGAGGTGTTTTGAATCTTTTAGAAGCTT
GR   (4258) GGACATTGAATTGTCTGAAGTGAGAGATTTCTTGAGGTGTTTTGAATCTTTTAGAAGCTT
SP   (4261) GGACATTGAATTGTCTGAAGTGAGAGATTTCTTGAGGTGTTTTGAATCTTTTAGAAGCTT

                4321                                                    4380
US   (4318) TGGTTTAAGGAAATAATGGAGAACAGTGCCGTTGCAAACACTGGTGATAACGGTGGTGGC
GR   (4318) TGGTTTAAGGAAATAATGGAGAACAGTGCTGTTGCAAACACTGGTGATAACGGTGGTGGC
SP   (4321) TGGTTTAAGGAAATAATGGAGAACGATGCTGTTACAAACACTGGTGATAATGGTGGCAGC

                4381                                                    4440
US   (4378) CGCAATCCTCTGGTTAGACCGTTAGATGATGGCGTAGATGACGAGGTGCAGAACTTGGGC
GR   (4378) CGCAATCCTCTGGTTAGACCGTTAGATGATGGCGTAGATGACGAGGTGCAGAACTTAGGC
SP   (4381) CGCAATCCTCTGGTTAGACCGTTAGATGATAGCGTAGATGACGAGGTGCAGAACTTAGGC

                4441                                                    4500
US   (4438) AGGAGGGACGATTCGACATCTCTCATTCCGGCTAATCCTAATCGATCTTCCAGTTGGGCT
GR   (4438) AGGAGGGACGATTCAACATCTATCATTCCGGCTAATCCTAATCGATCTTCCAGTTGGGCT
SP   (4441) AGGAGGGACGATCCGACATCTCTCATTCCGGCTAATCCTAATCGATCTTCCAGTTGGGCT

                4501                                                    4560
US   (4498) TTGTTGAACCCGGATACTATTAATTATAACGAGTTAAGGAAATTGAAGGTACACTCCACT
GR   (4498) TTGTTGAACCCGGATACTATTAATTATAACGAGTTAAGGAAATTGAAGGTACACTCCACT
SP   (4501) TTGTTGAACCCGGATACTATTAATTATAACGAGTTAAGGAAATTGAAGGTACGCTCCACT

                4561                                                    4620
US   (4558) AGGGGTGATACTCTTACCTTGACTCAGGAAGAGGAGTTCGAGAAGATACTCGAATCCTTT
GR   (4558) AGGGGTGATACTCTTACCTTGACTCAGGAAGAGGAGTTCGAGAAGATACTCGAATCCTTT
SP   (4561) AGGGGTGATACTCTTACCTTGACTCAGGAAGAGGAGTTCGAGAAGATACTCGAATCCTTC

                4621                                                    4680
US   (4618) TGCAGGCGAATAATCGGTGAAACCCCGATGACGGATAAGATTTTCGCTGGTTTCTACATG
GR   (4618) TGCAGGCGAATAATCGGTGAGACCCAGATGACGGATAAGATTTTCGCTGGTTTCTACATG
SP   (4621) TGCAGGCGAATAATCGGTGAGACCCCGATGACAGATAAGATTTTCGCTGGTTTCTACATG

                4681                                                    4740
US   (4678) TCTATGTGTCAGGCCATTGTAAACCAAGGGACCTCAGTTAAAGCAGCCGGTAATAACAGT
GR   (4678) TCCATGTGTCAGGCCATTGTAAACCAAGGGACCTCAGTTAAAGCAGCCGGTAATAACAGT
SP   (4681) TCTATGTGTCAGGCCATTGTAAACCAAGGGACCTCAGTTAAAGCAGCCGGTAATAACAGT

                4741                                                    4800
US   (4738) CTTGAAAACTACTTTGAGGTAGATGGTGCGAGATTTAAGTGGAAAACTCCGGATTTGATA
GR   (4738) CTTGAAAACTACTTTGAGGTAGATGGTGCGAAATTTAAGTGGAAAACTCCGGATTTGATA
SP   (4741) CTTGAAAATTACTTTGAGGTAGATGGCGCGAGATTTAAGTGGAAAACTCCGGATTTGATA
```

```
            4801                                                        4860
US   (4798) AATGAGGTTAGACCCAAAATGTCCGATGTTCCAAACGCTATACGTCGGTACGCCAGAAGT
GR   (4798) AATGAGGTTAGACCCAAAATGGCCGATGTTCCAAACGCTATACGTCGGTACGCCAGAAGT
SP   (4801) AATGAGGTTAGACCCAAAATGTCCGATGTTCCAAACGCTATACGTCGGTACGCCAGGAGT

            4861                                                        4920
US   (4858) CATGAAAAGATTATTCAGGACTTTATCAACTCCGGTCTTATTAAGCCTGATTATCATTTA
GR   (4858) CATGAAAAGATTATTCAGGACTTTATCAACTCCGGTCTTATTAAGCCTGATTATCATTTA
SP   (4861) CATGAAAAGATTATTCAGGATTTTATTAACTCCGGTCTTATTAAGCCTGATTATCATTTA

            4921                                                        4980
US   (4918) CAATTCAAACATGGCGTATTGCCAAGCCATGTGTTTGGTACCGGCGATTATATAAATGGT
GR   (4918) CAATTCAAACATGGCGTATTACCAAGCCATGTGTTTGGTACCGGCGATTATATAAATGGT
SP   (4921) CAATTCAAACATGGCGTATTACCAAGCCATGTGTTTGGTACCGGCGATTATATAAATGGT

            4981                                                        5040
US   (4978) TCGTTGATGAATATCTCAGATGATCAACTTATCTCGAACCTGCTTATGAAAAGAAACGCT
GR   (4978) TCGTTGATGAATATCTCAGATGATCAACTTATCTCGAACCTGCTTATGAAAAGAAACGCT
SP   (4981) TCGTTGATGAATATCTCAGATGATCGACTTATCTCGAACCTGCTTATGAAAAGAAACGCT

            5041                                                        5100
US   (5038) TTGTGCAAGGGTAACGAGGGCAAGGAACTGTACAACGTTAACCAACTTGCATCGATAACT
GR   (5038) TTGTGCAAGGGTAACGAGGGCAAGGAACTGTACAACGTTAACCAACTTGCATCGATAACT
SP   (5041) TTGTGCAAGGGTAACGAGGGCAAGGAACTGTACAACGTTAACCAGCTTGCATCGATAACT

            5101                                                        5160
US   (5098) GGTTGCTAAATTATATGGATGAAAATGAAATCTATGAGGATCAAGAGGATCTCTCTGCTC
GR   (5098) GGTTGCTAAATTATATGGATGAAAATGAAATCTATGAGGATCAAGAGGATCTCTCTGCTC
SP   (5101) GGTTGCTAAATTATATGGATGAAAATGAAATCTATGAGGATCAAGAGGATCTCTCTGCTC

            5161                                                        5220
US   (5158) GTGGCGGTGGGGGTTTCTATTACCAGACTGTGACTTTGGGTTCCGGTGATGTGTTTCCCG
GR   (5158) GTGGCGGTGGGGGTTTCTATTACCAGACTGTGACTTTGGGTTCCGGTGATGTGTTTCCCG
SP   (5161) GTGGCGGTGGGGGTTTCTATTACCAGACCGTGACTTTGGGTTCCGGTGATGTGTTTCCCG

            5221                                                        5280
US   (5218) TTGATTTAGCCCTAACGAGATCGGCTGAATTTGATTCGACAATTTTCTCCTTATATATTA
GR   (5218) TTGATTTAGCCCTAACGAGATCGGCTGAATTTGATTCGACAATTTTCTCCTTATATATTA
SP   (5221) TTGATTTAGCCCTAACGAGATCGGCCGAATTTGATTCGACAATTTTCTCCTTATATATTA

            5281                                                        5340
US   (5278) GGTTTGTAATTAAGGAGGGGAATGTGCGTTTAAAGATCGATTTTGGAAATAATTGGGATG
GR   (5278) GGTTTGTAATTAAGGAGGGGAATGTGCGTTTAAAGATCGATTTTGGAAACAATTGGGATG
SP   (5281) GGTTCGTAATTAAGGAGGGGAATGTGCGTTTAAAGATCGATTTTGGAAATAATTGGGATG

            5341                                                        5400
US   (5338) TGACTATGCAACAGGTGAGACTTTCTGGATGGTTTGCGGCGTTTGGTAAGATTGAAAAAC
GR   (5338) TGACTATGCAACAGGTGAGACTTTCTGGATGGTTTGCGGCGTTTGGTAAGATTGAAAAAC
SP   (5341) TGTCTATGCAGCAGGTGAGACTTTCTGGATGGTTTGCGGCGTTTGGTAAAATTGAAAAAC

            5401                                                        5460
US   (5398) CGAGAACCGCCAGGTCTGGATGGTCATATCCAATAAAATTGTTTAAAGAGGCTGGAGAAG
GR   (5398) CGAGAACCGCCAGGTCTGGATGGTCATACCCAATAAAATTGTTTAAAGAGGCTGGAGAAG
SP   (5401) CGAGAACCGCCAGGTCTGGATGGTCATACCCAATAAAATTGTTTAAAGAGGCTGGAGAAG

            5461                                                        5520
US   (5458) TCATAGTGTCCATTAGTGGTTGGAGGTGTTATAAAATTTATAATGGGTATCCCGTAGATC
GR   (5458) TCATAGTGTCCATTAGTGGTTGGAGGTGTTATAAAATTTATAATGGGTATCCCGTAGATC
SP   (5461) TCATAGTGTCCATTAGTGGTTGGAGGTGTTATAAAATTTATAATGGATATCCTGTAGATC
```

```
            5521                                                    5580
US   (5518) GCGTTGATTTGGTTCTGGCAGTACCCGTTCGTGAAGTAACAGCCGATTTAAAACGACCAT
GR   (5518) GCGTTGATTTGGTTCTGGCAGTACCCGTTCGTGAAGTAACAGCCGATTTAAAACGACCAT
SP   (5521) GCGTTGATTTGGTTCTGGCAGTACCCGTTCGTGAAGTAACAGCCGATTTAAAACGACCGT

            5581                                                    5640
US   (5578) TGGTTGGGGATTACGTCAACTTTCATGATGTATTTACTCTTATAAAAAGTAAAAATTCTG
GR   (5578) TGGTTGGGGATTACGTCAATTTTCATGATGTATTTACTCTTATAAAAAGTAAAAATTCTG
SP   (5581) TGGTTGGGGATTACGTCAACTTTCATGATGTATTTACTCTTATAAAGAGTAAAAATATTG

            5641                                                    5700
US   (5638) ACATCACTTTACCTAACCCGAGTCTGATATTCAACGATTCAACAAGTAAGGTTAATTTAG
GR   (5638) ACATCACTTTACCTAACCCGAGTCTGATATTCAACGATTCAACAAGTAAGGTTAATTTAG
SP   (5641) ACATCACTTTACCTAACCCGAGTCTGATATTCAACGATTCGACAAGTAAGGTTAATTTAG

            5701                                                    5760
US   (5698) ATGTGTCTCCAGGTGCGCGAAAACAAATTGCTCAAGTCAAGGCTGAGAAAGATCTGAACA
GR   (5698) ATGTGTCTCCAGGTGCGCGAAAACAAATTGCTCAAGTCAAGGCTGAGAAAGATCTGAACA
SP   (5701) ATGTGTCTCCAGGTGCGCGAAAACAAATTGCTCAAGTTAAGGCTGAGAAAGACTTGAACA

            5761                                                    5820
US   (5758) TTAAGAATCCTGAAGACTCAAAGCCTGACGTTCCTAATGATTCGCTGAGTGAAGTCGAAT
GR   (5758) TTAAGAATCCTGAAGACTCAAAGCCTGACGTTCCTAATGATTCGCTGAGTGAAGTCGAAT
SP   (5761) TTAAGAATCCTGAAGACTTAAAGCCTGATGTTCCTAATGATTCGCTGAGTGAAGTCGAAT

            5821                                                    5880
US   (5818) ATCATAATCACTCTGATGTTTCCAGTGTTTTCAGATTGTATTACACATGGAGGGTTGAAA
GR   (5818) ATCATAATCACTCTGATGTTTCCAGTGTTTTCAGATTGTATTACACATGGAGGGTTGAAA
SP   (5821) ATCATAATCACTCTGATGTTTCCAGTGTTTTCAGATTGTATTACACATGGAGGGTTGAAA

            5881                                                    5940
US   (5878) GAGATTTTGAGAGATCAGTTGAGTCGAGAATTTTCTTTCCGAATATATTTCCGACCGATT
GR   (5878) GAGATTTTGAGAGATCAGTTGAGTCGAGAATTTTCTTTCCGAATATATTTCCGACCGATT
SP   (5881) GAGATTTTGAAAGATCAGTTGAGTCGAGAATTTTCTTTCCGAATATATTTCCGACCGATT

            5941                                                    6000
US   (5938) TCACAATACTTCAACAAATGTGGTATGGGACGACTGCCGGTAACGTTGAGACTTTTGTGG
GR   (5938) TCACAATACTTCAACAAATGTGGTATGGGACGACTGCCGGTAACGTTGAGACTTTTGTGG
SP   (5941) TCACAATACTTCAACAAATGTGGTACGGGACGACTGCCGGTAACGTTGAGACTTTTGTGG

            6001                                                    6060
US   (5998) AGATAGGTAAAAATGAAAGGAAGTTCAACGTTGGGGTCGCCGCTTGGAAGGACAATGCAT
GR   (5998) AGATAGGTAAAAATGAAAGGAAGTTCAACGTTGGGGTCGCCGCTTGGAAGGACAATGCAT
SP   (6001) AGATAGGTAAAAATGATAGGAAGTTCAACGTTGGGGTCGCCGCTTGGAAGGACAATTCAT

            6061                                                    6120
US   (6058) TTGGACATTTCAAATTGGACGGCCCGACTTTAGCGAAGATCTCGACAATTCCAGGCAGGT
GR   (6058) TTGGACATTTCAAATTAGATGGCCCGACTTTAGCGAAGATCTCAACAATTCCAGGCAGGT
SP   (6061) TTGGACATTTCAAATTAGATGGCCCGACTTTAGCGAAGATCTCAACAATTCCAGGCAGGT

            6121                                                    6180
US   (6118) TCGTGGACCATAAAATTGAAAAAGACTCTAAAGGACATTTGATTGTGAGTATTGATAATA
GR   (6118) TCGTGGACCATAAAATTGAAAAAGACCCTAAAGGACATTTGATTGTGAGTGTTGATAATA
SP   (6121) TCGTGGACCATAAAATTGAAAAAGACTCTAAAGGACATTTGATTGTGAGTATTGATAATA

            6181                                                    6240
US   (6178) CTGTTCTTGTGCGTACTAACAAACTGATAGTCAAACCGAGTATTCAGATCGGTTGGGAAT
GR   (6178) CTGTTCTTGTGCGTACTAACAAACTGATAGTCAAACCGAGTATTCAGATCGGTTGGGAAT
SP   (6181) CTGTTCTTGTGCGTACTAACAAACTGATAGTCAAACCGAGTATTCAGATCGGCTGGGAAT
```

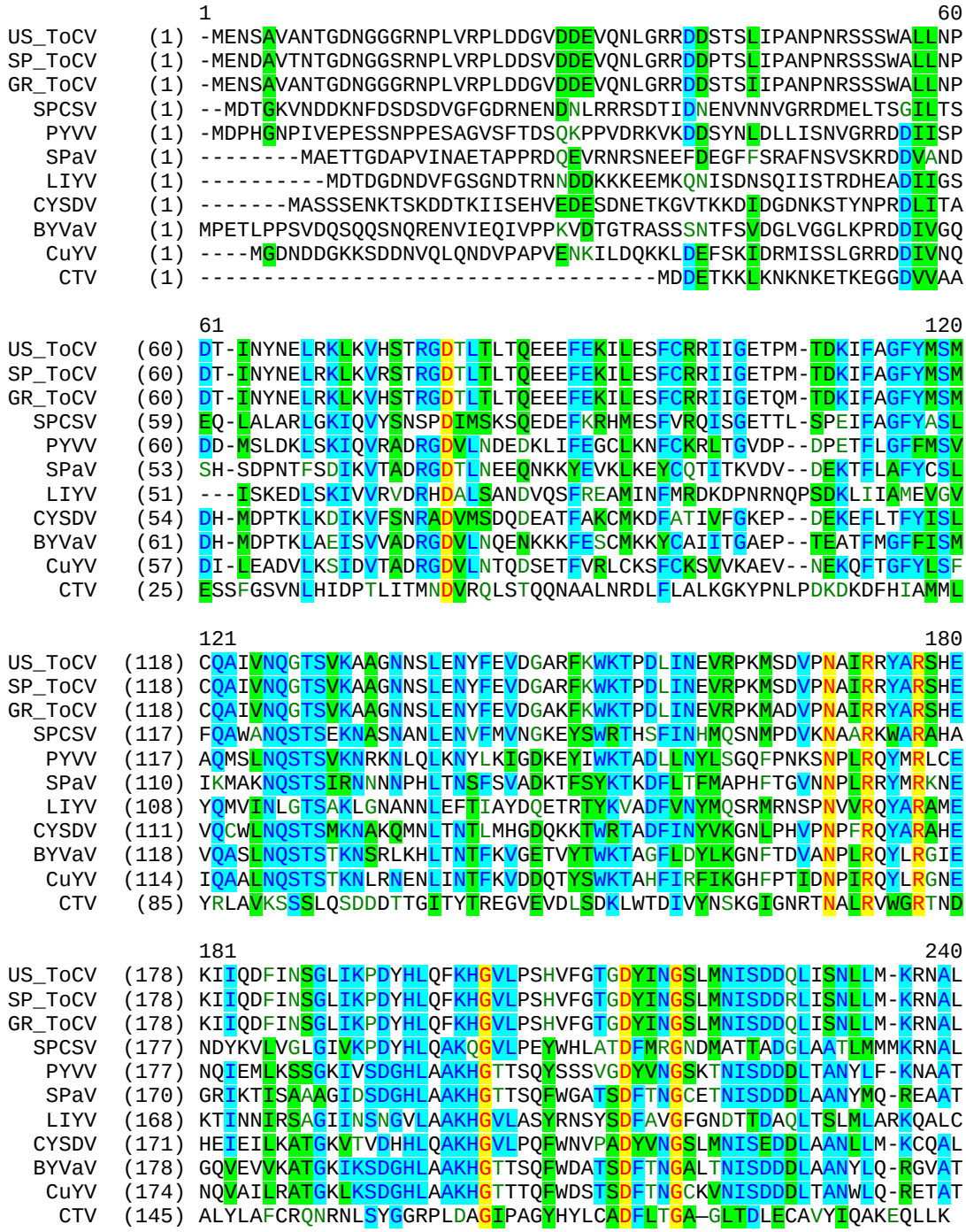
```
            6241                                                    6300
US   (6238) TTCATTTACCATGGGACGCGATTGGAAAGTATGGAGTTGGTAATTTGACCAGGTTCACAG
GR   (6238) TTCATTTACCATGGGACGCGATTGGAAAGTATGGAGTTGGTAATTTGGCCAGGTTCACAG
SP   (6241) TTCATTTACCATGGGACGCGATCAGAAAGTATGGAGTTGGTAATTTGGCCAGGTTCACAG

            6301                                                    6360
US   (6298) ACATCATCAAACCCAATTACATCAAGTATGATGGTTCTGAAGTTCCTTTGGTGCAAACTA
GR   (6298) ACATCATCAAACCCAATCACATCAAGTATGATGGTTCTGAAGTTCCTTTGGTGCAAACTA
SP   (6301) ACATCATCAAACCCAATTACATCAAGTTTGATAGTTCTGAAGTTCCTTTGGTACAAACTA

            6361                                                    6420
US   (6358) ATACCATAGAAAGTGATCGTTCCAAGTCTGGTCATAAACTATCACTAGTCAACCTGAAAA
GR   (6358) ATACCATAGAAAGTGATCGTTCCAAGTCTGGTCATAAACTATCTCTAGTCAACCTGAAAA
SP   (6361) ATACTATAGAAAGTGATCGTTCCAAGTCTGGTCATAAACTATCTCTAGTCAACCTGAAAA

            6421                                                    6480
US   (6418) GTTTCAGACGCATTAGTTCTACTGCAGATTTCTTCTTCGAACCACCACCACCCTCTGAGT
GR   (6418) GTTTCAGACGCATTAGTTCTACTGCAGATTTCTTCTTCGAACCACCACCACCCTCTGAGT
SP   (6421) GTTTCAGACGTATTAGTTCTACTGCAGATTTCTTCTTCGAACCACCACCACCCTCTGAGT

            6481                                                    6540
US   (6478) CCGATGACAAAACTTGGGAAGATAAGATCCAAACCGAAGTGGATATAAAGAAAGAAGAGA
GR   (6478) CCGATGACAAAACTTGGGAAGATAAGACCCAAACCGAAGTGGATATAAAGAAAGAAGAGA
SP   (6481) CCGATGACAAAAGTTGGGAAGATAAGGTCCAAACCGAAGTGGATATAAAGAAAGAAGAGA

            6541                                                    6600
US   (6538) CTATTCCGACTAATGAGGGTATCTCTTCCTCTGACTTGCCGAGTGAGAAGTCACAATTTG
GR   (6538) CTATTCCGACTAATGAGGGTACCTCTTCCTCTGACTTGCCGAGTGAGAAGTCACAATTTG
SP   (6541) CTATTCCGTCTAATGAAGGCACCTCTTCCTCTGATTTGCCGAGTGAGAAGTCACAATTTG

            6601                                                    6660
US   (6598) TTGCAGCTAATCATTATCTCTTGTCGATAGCTGAAGACAGGAACATTTTCAAAGCGGCTG
GR   (6598) TTGCAGCTAATCATTATCTCTTGTCGATAGCTGAAGACAGGAACATTTTCAAAGCGGCTG
SP   (6601) TTGCAGCTAATCATTACCTCTTGTCGATAGCTGAAGACAGGAATATTTTTAAAGCGGCTG

            6661                                                    6720
US   (6658) TAGATCGGTACACCGGATTGGGTTTCTCAAAGGATCAAGCTGTGTTGATAATATATCAAT
GR   (6658) TAGATCGGTACACCGGATTGGGTTTCTCAAAGGATCAAGCTGTGTTGATAATATATCAAT
SP   (6661) TAGATCGGTATACCGGCTTGGGTTTCTCAAAGGATCAAGCTGTGTTGATAATATATCAAT

            6721                                                    6780
US   (6718) TGGGGGTAACATTCGGCACTTCCAGAAATTGTTGCAGTGATAATTCATCGTTTCTAGTCT
GR   (6718) TGGGGGTAACATTCGGCACTTCCAGAAATTGTTGCAGTGATAATTCATCGTTTCTAGTCT
SP   (6721) TGGGGGTAACATTCGGCACTTCCAGAAATTGTTGCAGTGATAATTCATCGTTTCTAGTCT

            6781                                                    6840
US   (6778) GGAAGACTGATACTGGAGCGCAGGTTATAATCAGAAAGGGCGCCCACTCCAGGTTTCTCA
GR   (6778) GGAAGACTGATACTGGAGCGCAGGTTATAATCAGAAAAGGCGCCCACTCCAGGTTTCTCA
SP   (6781) GGAAGACTGATACCGGAACGCAGGTTATAATCAGAAAGGGCGCACACTCCAGGTTTCTCA

            6841                                                    6900
US   (6838) ATTCACTGGTTAAATATCCTTGCAACGTGGAGAGATTGATACTACGAAGACGTAGTGCGG
GR   (6838) ATTCACTGGTTAAATATCCTTGCAACGTGGAGAGATTGATACTACGAAGACGTAGTGCGG
SP   (6841) ATTCACTGGTTAAATATCCTTGTAACGTGGAGAGATTGATACTACGAAGACGTAGTGCGG

            6901                                                    6960
US   (6898) AGATATTGGCGTTGTTGAGGAACAAGAAATTGGCTTACCCAGACAGATTGGCCAAAAAGA
GR   (6898) AGATATTGGCGTTGTTGAGGAACAAGAAATTGGCTTACCCAGACAGATTGGCCAAAAAGA
SP   (6901) AGATATTGGCGTTGTTGAGGAACAAGAAATTGGCTTACCCAGATAGATTGGCCAAAAAGA
```

```
            6961                                                        7020
US   (6958) AAGGGGTAAGTCAGGGATTTACATATATGGCATGTGATTTTCTCGATTACACTGCGGTAA
GR   (6958) AAGGGGTAAGTCAGGGATTCACATATATGGCATGTGATTTTCTCGATTACACTGCGGTAA
SP   (6961) AAGGGGTAAGTCAGGGATTCACATATATGGCATGTGATTTTCTCGATTACACTGCGGTAA

            7021                                                        7080
US   (7018) CGTTAACTCAAGAAGAGCAGTTGACTATGAATTCTGTTGTGCAGTACGTGAGACTCCATA
GR   (7018) CGTTAACTCAAGAAGAGCAGTTGACTATGAATTCTGTTGTGCAGTACGTGAGACTCCATA
SP   (7021) CGTTAACTCAAGAAGAGCAGTTGACTATGAATTCTGTTGTGCAGTACGTGAGACTCCATA

            7081                                                        7140
US   (7078) ATAAACATCGAAGAAGCATTGTGAGCACGAGTCAGCTTTTCTGATCGATGGGGGTCGTGT
GR   (7078) ATAAACATCGAAGAAGCATTGTGAGCACGAGTCAGCTTTTCTGATCGATGGAGGTCGTGT
SP   (7081) ATAAACATCGAAGAAGCATTGTGAGCACGAGTCAGCTTTTCTGATCGATGGAGGTCGTGT

            7141                                                        7200
US   (7138) ACAATTCAGACGATGTTAACAGTGGAGTTGGATCCGGTGAAGATGTAAATACGACCGTGG
GR   (7138) ACAATTCAGACGATGTTAACAGTGGAGTTGGATCCGGTGAAGATGTAAATACGACCGTGG
SP   (7141) ACAATTCAGACGATGTTAACAGTGGAACTGGATCCGGTGAAGATGTAAATACGACCGTGG

            7201                                                        7260
US   (7198) CCAAGAACTTTTATTCTATAACTCACGTTATGAGTAACTACCGTAATTACACACCAGACG
GR   (7198) CCAAGAACTTTTATTCTATAACTCACGTTATGAGTAATTACCGTAATTACACACCAGACG
SP   (7201) CCAAGAACTTTTATACTATAACTCACGTTATGAGTAACTACCGTAATTACACACCAGACG

            7261                                                        7320
US   (7258) AAATTAAGGATGCTGTGAATATAGGTTATGGGTTACTGAATTTGTGTGAGAGGTTGGATA
GR   (7258) AAATTAAGGATGCTGTGAATATAGGTTATGGGTTACTGAATTTGTGTGAGAGGTTGGATA
SP   (7261) AAATTAAGGATGCTGTGAATGTAGGTTATGGGTTACTGAATTTGTGTGAGAGGTTGGATA

            7321                                                        7380
US   (7318) GAGATGTAATACTTGTGTCTCCGAATTCACCAGTTTACAACAATTACCGAGATGCCGGAA
GR   (7318) GAGATGTAATACTTGTGTCTCCGAATTCACCAGTTTACAACAATTACCGAGATGCCGGAA
SP   (7321) GAGATGTAATACTTGTGTCTCCGAATTCACCAGTTTACAACAATTACCGAGATGCCGGAA

            7381                                                        7440
US   (7378) TTCCACACAATCTACTTATGGAAAATACTGCACGGTATTTCCCAGTAGTTAATCCGAGCG
GR   (7378) TTCCACACAATCTGCTTATGGAAAATACTGCACGGTATTTCCCAGTAGTTAATCCGAGCG
SP   (7381) TTCCACACAATTTACTTATGGAAAATACTGCACGGTATTTCCCAGTAGTTAATCCGAGTG

            7441                                                        7500
US   (7438) AATTGGGAAAAGTTCTATTGGGTCATATCAGTGTTTTAAAGTTTTTGGAGTATTTCACAA
GR   (7438) AATTGGGAAAAGTTCTATTGGGTCATATCAGTGTTTTAAAGTTTTTGGAGTATTTCACAA
SP   (7441) AATTAGGAAAAATTCTATTGGGTCATATCAGTGTTTTAAAGTTTTTGGAGTGTTTCACAA

            7501                                                        7560
US   (7498) GATACGGGGTTGACGATATGCTAATCACAAGACTGTTCTCAAATTTTGTCTTGTGGTCCA
GR   (7498) GATACGGGGTTGACGATATGCTAATCACAAGACTGTTCTCAAATTTTGTCTTGTGGTCCA
SP   (7501) GATACGGGGTTGACGATATGCTAATTACAAGATTGTTCTCAAATTTTGTCTTGTGGTCCA

            7561                                                        7620
US   (7558) CTGGTGACGTGAACGCAGCATTGTATTCTATTTATCAACAGGATTTTCACTTTCCTGTTG
GR   (7558) CTGGTGACGTGAACGCAGCATTGTATTCAATTTATCAACAGGATTTTCACTTTCCTGTTG
SP   (7561) CTGGTGACGTGAACGCAGCATTATATTCTATTTATCAACAGGATTTTCACTTTCCTGTTG

            7621                                                        7680
US   (7618) AGGTAAGAGCGAATTTCAATTTCTTATTTTTGAATTCAAGTGAAATTGACAGAAGGTTGA
GR   (7618) AGGTAAGAGCGAATTTCAATTTCTTATTTTTGAATTCAAGTGAAATTGACAGAAGGTTGA
SP   (7621) AGGTAAGAGCGAATTTTAATTTCTTATTTTTGAATTCAAGTGAAATTGACAGAAGGTTGA
```

```
            7681                                                    7740
US   (7678) GTAACATTAGAAGGAAAGGTTATCCAAACTCTGAGAATTTCAATTGGTTCAAAAATATGA
GR   (7678) GTAACATTAGAAGGAAAGGTTATCCAAACTCTGAGAATTTCAATTGGTTCAAAAATATGA
SP   (7681) GTAACATTAGAAGGAAAGGTTATCCAAACTCTGAGAATTTCAATTGGTTCAAAAATATGA

            7741                                                    7800
US   (7738) TAAGTAATTACTTATATTTTGATTTTGTGTTCAGATACTCTGGTACAAAAATCAATATAG
GR   (7738) TAAGTAATTACTTATATTTTGATTTTGTGTTCAGATACTCTGGTACAAAAATCAATATAG
SP   (7741) TAAGTAATTACTTATATTTTGATTTTGTGTTCAGATACTCTGGTACAAAAATCAATATAG

            7801                                                    7860
US   (7798) AAAGAATCTCAAATTATTATATTTGATTTTCAATATGATTTCCACTTATTTTACCTTAAT
GR   (7798) AAAGAATCACAAATTATTATATTTGATTTTCAATATGATTTCCACTTATTTTACCTTAAT
SP   (7801) AAAGAATCTCAAATTACTATATTTGATTTTCAATATGATTTCCACTTATTTTACCTTAAT

            7861                                                    7920
US   (7858) AGGTTTAATTTTCTTGGTGGTATTCTGTTTTGTTTTATTGTGTTATTTCGTCTTCACTGT
GR   (7858) AGGTTTAATTTTCTTGGTGGTATTCTGTTTTGTTTTATTGTGTTATTTCGTCTTCACTGT
SP   (7861) AGGTTTAATTTTCTTGGTGGTATTCTGTTTTGTTTTATTGTGTTATTTCGTCTTCACTGT

            7921                                                    7980
US   (7918) CATTAAATTCTTCGCGAAAGATAAGATAAGTGACGATGATTGTCCTTATGTCAATAATGT
GR   (7918) CATTAAATTCTTCGCGAAAGATAAGATAAGTGACGATGATTGTCCTTATGTCAATAATGT
SP   (7921) CATTAAATTCTTCGCGAAAGATAAGATGAGTGACGATGATTGTCCTTATGTCAATAATGT

            7981                                                    8040
US   (7978) TGCTCCATTCGGGAGTAACAGGTTTAACTCACAACCTCCAATAGTTCGTTAAAGCACTAT
GR   (7978) TGCTCCATTCGGGAGTAACAGGTTTAACTCACAACCTCCAATAGTTCGTTAAAGTACTAT
SP   (7981) TGCTCCATTCGGGAGTAACAGGTTTAACTCACAACCTCCAATCGTTCGTTAAAGTACTAT

            8041                                                    8100
US   (8038) CTTACGGTTGGATTAATAAAAATGTTTAATTTAAGGAGTTTTTGATAAAGTTTTCTTCTT
GR   (8038) CTTACGGTTGGATTAATAAAAATTTAT-----AAGGAGTTTTTGATAAAGTTTTCTTCTT
SP   (8041) ATTACGGTTGGATTAATAAAAATTTT-----AAGGAGTTTTTGATAAAGTTTTCTTCTT

            8101                                                    8160
US   (8098) CGGGTTATTTACTTCGTATTTTATAAAATCCCAAAAATATATGAAAGATTTTACATATTG
GR   (8093) CGGGTTATTTACTTCGTATTTTATAAAATCCCAAAAATATATGAAAGATTTTACATATTG
SP   (8096) CGGGTTATTTACTTCGTATTTTATAAAATCCCAAAAATATATGAAAGATTTTACATATTG

            8161                                                    8220
US   (8158) TTCTGTTTAGTTGGTGTAAAATTCTATCTACCTAAATTACGTGTTATACACGTAAACCTT
GR   (8153) TTCTGTTTAGTTGGTGTAAAATTCTATCTACCTAAATTACGTGTCATACACGTAAACCTT
SP   (8156) TTCTGTTTAGTTGGTGTGAAATTCTATCTACCTAAATTACGTGTTATACACGTAAACCTT

            8221                        8250
US   (8218) GGTAGATTTAGTATATAAATAAATAGGTCG
GR   (8213) GGTAGATTTAGTATATAAATAAATAGGTCG
SP   (8216) GGTAGATTTAGTATATAAATAAATAGGTC-
```

Amino acid alignments

CP



```
                241               262
US_ToCV  (237) CKGNEGKELYNVNQLASITGC-
SP_ToCV  (237) CKGNEGKELYNVNQLASITGC-
GR_ToCV  (237) CKGNEGKELYNVNQLASITGC-
  SPCSV  (237) CNKDNKNSIYNVTQLTGTGLHC
   PYVV  (236) KSTKKTKSIYNVAQLAGSIE--
   SPaV  (229) KNKARSRTIFNVSQLAGNVQ--
   LIYV  (228) KGEGGSVEHYNTMQLANLKHPC
  CYSDV  (230) KRNEKEKKYYNVSQLAPGGCGN
  BYVaV  (237) KGKQQRKQIYNVSQIAGNV---
   CuYV  (233) KGKNKKNTIYNVSQLASYGN--
    CTV  (203) KRGADEVVVTNVRQLGKFNTR-
```

RdRp

```
                1                                                          60
GR_ToCV    (1) -TYRLTIEGCYIPDTFSRPASSHLMAVNDFMSVVNPGLAWMQFLHRTILFEYGDFDMPPV
SP_ToCV    (1) -TYRLTIEGCYIPDTFSRPASSHLMAVNDFMSVVNPGLAWMQFLHRTILFEYGDFDMPPV
US_ToCV    (1) -TYRLTIEGCYIPDTFSRPASSHLMAVNDFMSVVNPGLAWMQFLHRTILFEYGDFDMPPV
  BYVaV    (1) DSYILEVDSLKTALTYSRPSASHHRAIGEFMMLINTNLSAYDYIHRTLLYEFEDYELPVV
   CuYV    (1) DTYDLKIDCLRAPSTFSRPSASHHRAINEFMTLINPGLSAYNFIHRTLLFEYEQFELPPV
  CYSDV    (1) ------------------------------------------------------YHDFEMPYL
   LIYV    (1) ----------------------------MDIISPGVAFYNYLHRTLIFEYSDYYLPPC
   PYVV    (1) DEYRIEVDVCRFGNTLSRAPSSHFVAVNEFMELINPGLSAHDYLYRTLLYEYEQYELPFV
   SPaV    (1) DVYEMVVEDLCLPATFSRPPSSHHRSINEFMCFINPGLSAYNYIHRTLIFEYEQYELPVV
  SPCSV    (1) DVYNLQVEDFFLPQTFSRPPASTLVAVNEFMSLINPGIIDHDFLHRTMISEYSMFELPPV
    CTV    (1) ------------------------------------------------------------

                61                                                         120
GR_ToCV   (60) EKMVLDFSKYKPYVAGEFVVSKILGKGERTRPDSMKQGIISLSHRNFSAPRINERLDVYK
SP_ToCV   (60) EKMVLDFSKYKPYVAGEFVVSKILGKGERTRPDSMKQGIISLSHRNFSAPRINERLDVYK
US_ToCV   (60) EKMVLDFSKYKPYVAGEFVVSKILGKGERTRPDSMKQGIISLSHRNFSAPRINERLDVYK
  BYVaV   (61) EDLELSISRSKAYQPGEYIIPDLLGKGERSRPNTWKQVILSLSHRNFSAPRINERLDVTS
   CuYV   (61) ENAKLVLSHSKPYAGAEFIVPDIMGKGERSRPNTWKQVMISLSHRNFSAPRINENLDTTK
  CYSDV   (10) EDVDIKLNKTKIYQPGEYIVSNLLGKGERSRPDTWKQALISLSKRNFSAPRVNEKLDVLK
   LIYV   (31) EDLRITLSKSKPYHPGAYVVSKILGKGERNRPNTWKQVIQSLSHRNFNAPIINHKLDVKR
   PYVV   (61) DDLTLDITRSAPYTSGDFLVPKILGKGERTRPDTWKQVLISLSHRNFSAPRINERFDNVK
   SPaV   (61) GEVDLVLSRSKPYNPGLYVIPDLLGKGERSRPDTWRQVLLSLSHRNFSAPRVNENCDTLA
  SPCSV   (61) GDLTIDLSKSKPYISGDFVTSGVLGKGERSRPDTWRQAIASLSHRNYSAPRVNERLDTFK
    CTV    (1) ---ETPPLLTRVTYTNRLAFGVVRSQAIPPRKASLQENLLSYESRNYNFIKTERFVGPSE

                121                                                        180
GR_ToCV  (120) TAERLCQNLVRSFDFSRLYENY--DVILPDMFKIDDWLQDRDGSKFGRIKRDMDHKLLVE
SP_ToCV  (120) TAERLCQNLIRSFDFSRLYENY--DVILPDMFKIDDWLQDRDGSKFGRIKRDMDHKLLVE
US_ToCV  (120) TAERLCQNLVRSFDFSRLYENY--DVILPDMFKIDDWLQDRDGSKFGRIKRDMDHKLLVE
  BYVaV  (121) SAERLVQSLMKCFELSKLAENF--DLIVPDLYRIDKWLETRDGEKIRKLKRSFSHSLMIS
   CuYV  (121) TAERLMSGLMTGLNPLKLSENY--DTILPDMYSIDKWLNTRDGDKYRKLKRSFSHTLMIQ
  CYSDV   (70) TAERLSHSLFKAFDFTKLFENY--DPVLPDLNKLGEWLTTRDGMRYGKLKRSMNHRLVVE
   LIYV   (91) SAQILYDSVVKSLRQDRLTEWY--EPILPDLFKIGKWLDDRDGSKYRMLNRRLDFASLAD
   PYVV  (121) TAEILSKNLIGCMKLERLYENF--DPVLPDLRRVDKWLISRDPNKLRRLYRSFSTDLLVN
   SPaV  (121) SAEILAQSLMKAFDFLKLSENF--DTVLPDIWSITKWIEDREPNKVRKLKRSFGHDLMNS
  SPCSV  (121) TAELLCESLLRAFDMSKLFENY--DYIIPDIYRIDQWLSDRDVSKFNRIKRDMNHNLMYE
    CTV   (58) FGRAMAAAVIERCFKMEEMAKIRCDIISLTEANILKWLDKRTPCQIKAVHGELKLPFSVE
```

```
                  181                                                   240
GR_ToCV   (178) -QFESLKFMIKGEMKPKMDMSSYTAYNPPANIIYYNHLVSMYYSPLFLEVFDRISYCLSK
SP_ToCV   (178) -QFESLKFMIKGEMKPKMDMSSYTAYNPPANIIYYNHLVSMYYSPLFLEVFDRISYCLSK
US_ToCV   (178) -QFESLKFMIKGEMKPKMDMSSYTAYNPPANIIYYNHLVSMYYSPLFLEVFDRISYCLSK
  BYVaV   (179) -QFESMKLMIKGEMKPKMDTSSYSTYNPPANIIYYEHAINMFYSPMFLEVFDRITYCLSN
   CuYV   (179) -QFENLRLMVKGDMKPKMDMSSYSTYSPPSNIIYYEHVINMFFSPMFLEVFDRLTYCLAG
  CYSDV   (128) -QFQPLNFMIKGDMKPKMDMSSYSQYDPPSNIIYYKNCINLFYSPLFLEIFDRIVYCLKG
   LIYV   (149) -KFKTLNLMVKGETKPKMDLSTYDSYNAPANIVYYQQIVNLYFSPIFLECFARLTYCLSD
   PYVV   (179) -KFNDLKLMVKGSMKPKLDTSSYSMYTPPSNIVYYEQVVNMFYSPIFLEVFERIRYCMND
   SPaV   (179) -QFSRMKLMIKGEMKPKMDMSSYGTYGPSSNIIYYEQIVNMFYSPMFLRIFDRIGYCLNS
  SPCSV   (179) -QFSTMKMMIKGDLKPKMDLSCYTTYAPPANIIYYKHIVSMFFSPLFLEVFDRITYCLSN
    CTV   (118) EQISNFKLMVKRDAKVKLDDSSLSKHPAAQNIMFHKKFINAIFSPCFDEFKNRVLSSLND

                  241                                                   300
GR_ToCV   (237) KIVMYSGMNLETLGTLIGSKLQKPLTSYHTLEIDFSKFDKSQGILFKVYEGMIYRFFKFS
SP_ToCV   (237) KIVMYSGMNLETLGTLIGSKLQKPLTSYHTLEIDFSKFDKSQGILFKVYEGMIYRFFKFS
US_ToCV   (237) KIVMYSGMNLETLGTLIGSKLQKPLTSYHTLEIDFSKFDKSQGILFKVYEGMIYRFFKFS
  BYVaV   (238) KIILYSGMNLETLGSLIRSKLPLPLDEYKTVEIDFSKFDKSQGVLFKVYEEIVYKFFKFS
   CuYV   (238) NIVLYSGMNLDTLGELISAKLQYPISEYKTIEIDFSKFDKSQGVVFKLYEEIVYKFFRFR
  CYSDV   (187) KIIMYSGMNLMTLADLIGSTLPMPVEAYHTTEIDFRMFDKSQGVLFKVYEEIIYKVFKFS
   LIYV   (208) KIVLYSGMNTDVLAELIESKLPLGLNAYHTLEIDFSKFDKSQGTCFKLYEEMMYKMFGFS
   PYVV   (238) KIVMYSGMNLETLSEIISCKLTMPLNEYYTTEIDFSKFDKSQGVVFKLYEEMVYKFFKVS
   SPaV   (238) KVVLYSGMNLETLGKLIRSKLDFPIQEYRTTEIDFSKFDKSQGVIFKLYEEIIYKFFKFD
  SPCSV   (238) KVVMYSGMNLETLSRVVKSKLPLPLDSYYTLEIDFSKFDKSQGCVFKLYEGMIYRFLKFS
    CTV   (178) NIVFFTEMTNAGLAEIIR-RIIGDDDNLFVGEVDFSKFDKSQDLFIKEYERTLYSEFGFD

                  301                                                   360
GR_ToCV   (297) EDYYT-NIEATEYFIKYRGRCGISGELGAQRRTGSPNTWLSNTLVTMGIILSVYDLDDID
SP_ToCV   (297) EDYYT-NIEATEYFIKYRGRCGISGELGAQRRTGSPNTWLSNTLVTMGIILSVYDLDDID
US_ToCV   (297) EDYYT-NIEATEYFIKYRGRCGISGELGAQRRTGSPNTWLSNTLVTMGIILSVYDLDDID
  BYVaV   (298) EELYE-NIKMTEYFCRAKSRSGVSLELGAQRRTGSPNTWLSNTLVTLGIILSTYKIDDID
   CuYV   (298) KIYME-NIKLTEYFCKAKSRSGVSVELGAQRRTGSPNTWLSNTLVTLGIVLSTYDLRDID
  CYSDV   (247) EEMYD-NIKLTEYFTRYTGTCGVSGELGAQRRTGSPNTWLSNTLVTMGILLSQYNLDDID
   LIYV   (268) PELYDRDFKYTEYFCRAKATCGVDLELGTQRRTGSPNTWLSNTLVTLGMMLSSYDIDDID
   PYVV   (298) EDVYE-NIKFSEYFCRVRSRSGIQTELGAQRRTGSPNTWLSNTLTTMAVVLSSYNLDDVE
   SPaV   (298) SKTYE-AIKFSEYFCRAKSSCGISVELGAQRRTGSPNTWLSNTLVTLAMILTHYDLDDID
  SPCSV   (298) EEAYL-NIETTEYFCRFKSASGLTGELGAQRRTGSPNTWLSNTLVTMGMLLNVYDLDDID
    CTV   (237) TELLDVWMEGEYRARATTLDGQLSFSVDGQRRSGGSNTWIGNSLVTLGILSLYYDVSKFD

                  361                                                   420
GR_ToCV   (356) LFLVSGDDSLIFSSKPLKNKTDEINRDFGFEAKMIENSVPYFCSKYIISDRGKIRVVPDP
SP_ToCV   (356) LFLVSGDDSLIFSSKPLKNKTDEINRDFGFEAKMIENSVPYFCSKYIISDRGKIRVVPDP
US_ToCV   (356) LFLVSGDDSLIFSSKPLKNKTDEINRDFGFEAKMIENSVPYFCSKYIISDRGKIRVVPDP
  BYVaV   (357) LLLVSGDDSLIFSKEDLPNLANQINQDFGMEAKFIVNSVPYFCSKFIIEDRGEIKVVPDP
   CuYV   (357) LLLVSGDDSLIFSRKSLKNQVNEINRDFGMEAKFIENSVPYFCSKFIIEDRGSIRVLPDP
  CYSDV   (306) LMLVSGDDSLIFSKKPLPNVTAEINKDFGFEAKFLMNSVPYFCSKYIFTDDGKVRVVPDA
   LIYV   (328) LLLVSGDDSLIFSRKHLPNKTQEINKNFGMEAKYIEKSSPYFCSKFIVELNGKLKVIPDP
   PYVV   (357) LFLVSGDDSLIFSKYPLDNRTQRMNVDFGMEAKFIENSVPYFCSKFIIQDRGRIKVIPDP
   SPaV   (357) LLLVSGDDSLIFSRKDLGNKANEINRDFGMEAKFIMNSVPYFCSKFIIEDRGEIKVVPDP
  SPCSV   (357) LMLVSGDDSLIFSKKQLENKTNELNINFGFEAKFIENSVPYFCSKFIVDDRGSIKVVPDP
    CTV   (297) LLLVSGDDSLIYSSEKISNFSSEICLETGFETKFMSPSVPYFCSKFVVQTGNKTCFVPDP
```

```
                421                                                                 480
GR_ToCV  (416) VRFFEKLSVPIRVQDFMSDTLMREKFRSYKDLMKDFDYDTTCVLVDALVCYRYNLPPMCS
SP_ToCV  (416) VRFFEKLSVPIRVQDFMSDTLMREKFRSYKDLMKDFDYDTTCVLVDALVCYRYNLPPMCS
US_ToCV  (416) VRFFEKLSVPIRVQDFMSDTLMREKFRSYKDLMKDFDYDTTCVLVDALVCYRYNLPPMCS
  BYVaV  (417) VRFFEKLSVPIRLEDFLSETLLKEKYTSFKDLMVFFDSDTVCVLVDRLICIRYGIPEMSS
   CuYV  (417) VRFFEKLSVPMSLSDFESGEMLRERYTSYKDLMVGYNLDTNIILVDSLISVRYSIPLMSS
  CYSDV  (366) QRMFEKLSNPIRRSDFEEGTILKERFISYKDLMYYYRFDTTCLAVDRLICKRHGLPEMSS
   LIYV  (388) IRFFEKLSIPIRQEDFVNGSVVKERFISFKDLMKEYDNDVAVIRIDEAVCYRYSIPVGCS
   PYVV  (417) VRFFEKLSVPVSYQDYENWNMIRERFISYKDLMVEFDYDTSCMLVDVMVSKRYSLPNMAS
   SPaV  (417) VRFFEKLSVPISLQDFQVGDLLRERFVSFKDLMIGYDSDAVIILVDNLISIRYDIPRMTS
  SPCSV  (417) VRFFEKLSVPVRLSDFLAETTLRERFTSYKDLMSEYDNDSVCILVDSLISKRYSLPLMTS
    CTV  (357) YKLLVKLGAP---QNKLTDVELFELFTSFKDMTQDFGDQVVLEKLKLLVEAKYGFASGTT

                481                                                                 540
GR_ToCV  (476) YAALCYIHCLCANFTTFRRVYESDLTVVI-------------------------------
SP_ToCV  (476) YAALCYIHCLCANFTTFRRVYESDLTVVI-------------------------------
US_ToCV  (476) YAALCYIHCLCANFTTFRRVYESDLTVVI-------------------------------
  BYVaV  (477) YSALCYIHCLLANVLSFKRLYTDAMTVVI-------------------------------
   CuYV  (477) YAALCYIHCMCSNITAYKRIYHDSFVVVI-------------------------------
  CYSDV  (426) YAALCYIHCMFANVVAFRKL-------------------------------------------
   LIYV  (448) YAALCYIHCCMSNFVSFRRIYDNCEIVWI-------------------------------
   PYVV  (477) YAALCYIHCLLANAQSFKRIYLDSILVSI-------------------------------
   SPaV  (477) YAALCYIHCLTSNVLAYTKLFIEGFTVAI-------------------------------
  SPCSV  (477) YAALCFIHCLCANFSSFRKIFDEFFVVDI-------------------------------
    CTV  (414) MPALCAIHCVRSNFLSFERLFPFIRGWYVVDALKLRQLRKLTNLICERVVYDNRVSYFSY

                541                     567
GR_ToCV  (505) --------------------------
SP_ToCV  (505) --------------------------
US_ToCV  (505) --------------------------
  BYVaV  (506) --------------------------
   CuYV  (506) --------------------------
  CYSDV  (446) --------------------------
   LIYV  (477) --------------------------
   PYVV  (506) --------------------------
   SPaV  (506) --------------------------
  SPCSV  (506) --------------------------
    CTV  (474) FDNPFTKPDANDDNVDDLGQAGELATG
```

p22

```
                1                                                                   60
GR_ToCV    (1) MDLTGCLRKLRQCDRLLERRGNAVSEVHLRAILIDLDECSECLMLCEQEYIRDTDCLMSF
SP_ToCV    (1) MDLTGCLRKLRQCDRLLERLGNDVSEVHLRAILIDLDECSECLMLCEQEYIRDTDCLLSF
US_ToCV    (1) MDLTGCLRKLRQCDRLLERLGNDVSEVHLRAILIDLDECSECLMLCEQEYIRDTDCLMSF
  SPCSV    (1) MSSGAIYSELNRFLKYLRELDFSNFCFSDFLFRFSSLKSLIDEYSSHWLVNTNELVWY
  CYSDV    (1) MQSVGVGIPTTRVERHPDQDLMGVYRCYHVADFSLGLPCNLRLKVLTNGRIDDLKRIFTI

                61                                                                  120
GR_ToCV   (61) LLALKHYEIKFHMDMLNMIYDFKLKTSQLIQDVFRIKVIIRVYLELCEIDPLLAMTEACQ
SP_ToCV   (61) LLVLKHYEIKFHMDMLNMIYDFKLKTSQLIQDVFRIKVIIRVYLELCEIDPLLAMTEACQ
US_ToCV   (61) LLALKHYEIKFHMDMLNMIYDFKLKTSQLIQDVFRIKVIIRVYLELCEIDPLLAMTEACQ
  SPCSV   (59) RTICEQRLHPHLFDANRLL--TRINVSDLKHDLIMIRDHFSSIDDFLNKYPAIPILVSFT
  CYSDV   (61) QFWFTVDEKTKCEFGLQQIENSEKLTPFVFTTHDGYRSACENGNIVEWEEKDLKNFKDCR
```

```
                121                                                       180
GR_ToCV  (121) DILESGILNIGFISSALGHEPNILITILSMVDFIVVIDDRPLVFIPSKIRFVGDKLGSGH
SP_ToCV  (121) DILEGGILNIGFVSSALGHEPNILITILSMVDFIVVIDDRPLVFIPLKMRFVGDKLGSGY
US_ToCV  (121) DILESGILNIGFISSALGHEPNILITILSMVDFIVVIDDRPLVFIPSKIRFVGDKLGSGH
  SPCSV  (117) KNI-SAFLEDNEWESVIHHTDSLNRLVFRESWSDGYGYVNHHRLSASSHDFQHMKLCDIF
  CYSDV  (121) LVTSESIDIISEFEIEYFWIDDVKSVFINRECVGFIFTKDVMDVKMIRFGVESMQFLETS

                181           197
GR_ToCV  (181) FRWFDKFFFGSDI----
SP_ToCV  (181) FRWFDKFFFGSDI----
US_ToCV  (181) FRWFDKFFFGSDI----
  SPCSV  (176) RLSKDKLNSGVFTFLL-
  CYSDV  (181) PIKIYRLMGTIP-----
```