

**SUGARCANE STRIATE MOSAIC ASSOCIATED VIRUS:
RNA SEQUENCE AND GENOME ORGANISATION,
TAXONOMY AND DETECTION**

Nicole Thompson

B. Sc. (Hons) Adelaide University, Australia

Department of Applied and Molecular Ecology
Waite Campus, Adelaide University
Australia

Thesis submitted for the degree of
Doctor of Philosophy
at Adelaide University

March, 2001

“Be charming to my virus”

-Dave Lister,

Red Dwarf

TABLE OF CONTENTS

Table of Contents	i
Abstract	viii
Statement.....	x
Acknowledgments	xi
Abbreviations and virus acronyms	xii

CHAPTER 1: GENERAL INTRODUCTION

1.1 THE SUGAR INDUSTRY	1
1.2 SUGARCANE VIRUS DISEASES	2
1.3 SUGARCANE STRIATE MOSAIC DISEASE (SCSMD).....	2
1.3.1 Introduction	2
1.3.2 Symptoms of SCSMD	3
1.3.2.1 Plant symptoms	3
1.3.2.2 Leaf and cell structural changes	3
1.3.3 Identification of a putative causal agent	4
1.3.3.2 Virions.....	4
1.3.3.3 Preliminary sequence analysis.....	5
1.4 VIRUS TAXONOMY AND PHYLOGENETICS	5
1.4.1 Origin of viruses	5
1.4.2 Recombination.....	6
1.4.3 Virus taxonomy	7
1.4.4 Phylogenetic studies	7
1.5 GENERA REPRESENTING VIRUSES RESEMBLING SCSMaV	8
1.5.1 <i>Foveavirus</i>	8
1.5.2 <i>Carlavirus</i>	8
1.5.3 <i>Potexvirus</i>	10
1.6 DETECTION METHODS.....	11
1.6.1 Serological methods	11
1.6.2 Nucleic acid methods	11

1.6.2.1 Hybridisation.....	11
1.6.2.2 RT-PCR.....	12
1.7 VECTORS.....	13
1.8 SCOPE OF THIS THESIS.....	14

CHAPTER 2: MATERIALS AND METHODS

2.1 MATERIALS	15
2.1.1 Sugarcane plants.....	15
2.1.2 Chemical reagents	15
2.1.3 Bacterial strains	15
2.1.4 Kits	16
2.1.5 Oligonucleotide primers	16
2.1.6 Solutions, buffers and media	17
2.2 METHODS	17
2.2.1 Gel electrophoresis	17
2.2.1.1 Agarose gel electrophoresis (AGE).....	17
2.2.1.2 Polyacrylamide gel electrophoresis (PAGE).....	17
2.2.1.3 Sodiumdodecylsulphate PAGE (SDS-PAGE)	18
2.2.1.4 Staining of polyacrylamide gels (PAG)	18
2.2.1.4a Silver staining for visualisation of nucleic acids	18
2.2.1.4b Coomassie blue staining for protein	19
2.2.1.5 Maintenance of stained PAG	19
2.2.2 Extraction of total nucleic acid (TNA)	19
2.2.2.1 From sugarcane leaf tissue	19
2.2.2.2 From insect samples.....	20
2.2.3 Double stranded RNA (dsRNA) extraction.....	20
2.2.4 Plasmid DNA preparation	22
2.2.5 Restriction enzyme analysis	22
2.2.6 Reverse transcriptase-polymerase chain reaction (RT-PCR)	23
2.2.6.1 Designing of oligonucleotide primers	23
2.2.6.2 Reverse transcription.....	23
2.2.6.2a dsRNA as a template.....	23
2.2.6.2b ssRNA as a template	24

2.2.6.3 PCR reactions.....	24
2.2.6.3a DyNAzyme DNA polymerase reaction.....	24
2.2.6.3b <i>Taq</i> DNA polymerase reactions.....	24
2.2.6.3c Duplex RT-PCR using TMV- RNA as an internal control.....	25
2.2.6.4 Rapid amplification of cDNA ends (RACE)	25
2.2.6.4a 5' RACE.....	25
2.2.6.4b 3' RACE.....	26
2.2.7 Cloning of PCR products and DNA sequencing	26
2.2.7.1 Purification of PCR products	26
2.2.7.2 Ligation of PCR products into plasmid vectors	27
2.2.7.2a Ligation into pGEM [®] -Teasy vector	27
2.2.7.2b Ligation into pQE-31 bacterial expression vector	27
2.2.7.3 Preparation of competent <i>E. coli</i> cells.....	27
2.2.7.3a <i>E. coli</i> JM109 cells.....	27
2.2.7.3b <i>E. coli</i> M15[pREP4] cells.....	28
2.2.7.4 Transformation of competent cells and selection of positive clones	29
2.2.7.4a <i>E. coli</i> DH5 α cells.....	29
2.2.7.4b <i>E. coli</i> JM109 cells	29
2.2.7.4c <i>E. coli</i> M15[pREP4] cells	30
2.2.7.5 Screening of bacterial colonies containing plasmids	30
2.2.7.5a Amplification of plasmid inserts.....	30
2.2.7.5b Colony screening	30
2.2.7.6 Automated sequencing.....	31
2.2.7.7 Sequence analysis.....	31
2.2.8 Detection of homologous nucleic acids with radioactively labelled probes.....	32
2.2.8.1 Capillary blot.....	32
2.2.8.2 Nucleic acid dot-blot.....	32
2.2.8.3 Tissue squash	33
2.2.8.4 Radiolabelling of DNA probe with ³² P	33
2.2.8.5 Hybridisation of nylon filters	33
2.2.8.6 Autoradiography.....	34
2.2.8.7 Stripping of nylon membranes	34
2.2.9 Virus preparation	34

2.2.10 Preparation of electron microscope grids for transmission electron microscopy (TEM)	36
--	----

CHAPTER 3: SEQUENCING THE SCSMaV-RNA

3.1 INTRODUCTION	37
3.2 METHODS	37
3.2.1 Isolation of dsRNA and TNA	37
3.2.2 Sequencing of existing clones	38
3.2.3 Designing of oligonucleotide primers	38
3.2.4 RT-PCR reactions.....	38
3.2.4.1 Bridging of gaps using known sequence.....	38
3.2.4.2 Extension of the RNA sequence using anchored RT-PCR	39
3.2.4.3 Extension into the coat protein region	40
3.2.4.4 Rapid amplification of cDNA ends (RACE)	40
3.2.5 Production and selection of clones for sequence analysis	40
3.3 RESULTS.....	41
3.3.1 Sequencing of existing clones	41
3.3.2 Amplification by RT-PCR.....	41
3.3.2.1 Bridging of gaps	41
3.3.2.2 Sequence extension using anchored RT-PCR.....	41
3.3.2.3 Extension using PXCP primer	42
3.3.2.4 5' and 3' RACE	42
3.3.3 Sequence assembly	42
3.4 DISCUSSION.....	43

CHAPTER 4: GENOME ORGANISATION OF THE SCSMaV-RNA

4.1 INTRODUCTION.....	44
4.2 MATERIALS AND METHODS	44
4.2.1 Identification of open reading frames.....	44
4.2.2 Nucleotide sequence analysis	45
4.2.3 Peptide analysis	45
4.3 RESULTS.....	45
4.3.1 Genome organisation.....	45

4.4 DISCUSSION.....	46
4.4.1 ORF1	46
4.4.2 ORF2, ORF3 and ORF4.....	47
4.4.2.1 ORF2.....	47
4.4.2.2 ORF3	48
4.4.2.3 ORF4.....	48
4.4.3 ORFs 2, 3 and 4 comprise the Triple Gene Block (TGB)	49
4.4.5 ORF5	50
4.4.6 The 5' and 3' non-coding regions	51
4.4.7 Other potential ORFs.....	52

CHAPTER 5: PHYLOGENETIC ANALYSIS OF SCSMaV-RNA AND PUTATIVE ORF PRODUCTS

5.1 INTRODUCTION.....	53
5.2 MATERIALS AND METHODS	54
5.2.1 Phylogenetic analyses.....	54
5.3 RESULTS.....	55
5.3.1 Phylogenetic analysis using the full nucleotide sequence	55
5.3.2 Phylogenetic analysis using peptide sequences	55
5.4 DISCUSSION.....	55
5.4.1 Relationships based on nucleotide sequence	55
5.4.2 Relationships based on RdRP sequence	56
5.4.3 Relationships based on TGB protein sequences.....	56
5.4.4 Relationships based on CP sequence.....	57
5.4.5 Summation of phylogenetic analyses	57

CHAPTER 6: CLONING OF THE SCSMaV-CP AND PRODUCTION OF ANTISERA

6.1 INTRODUCTION.....	58
6.2 MATERIALS AND METHODS	59
6.2.1 SCSMaV preparation.....	59
6.2.2 Cloning of the CP of SCSMaV	59
6.2.3 Expression and purification of SCSMaV-CP from <i>E. coli</i>	60

6.2.3.1 Expression of the recombinant protein	60
6.2.3.2 Purification of induced protein under native conditions	61
6.2.3.3 Purification of induced protein under denaturing conditions.....	61
6.2.4 Preparation of antigen.....	62
6.2.4.1 Antigen A	62
6.2.4.2 Antigen B	63
6.2.5 Preparation of antiserum.....	63
6.2.6 Dot immuno binding assay (DIBA).....	63
6.2.6.1 Preparation of membrane	64
6.2.6.2 Detection of antigen	64
6.2.7 Western blot assay	65
6.2.7.1 Preparation of membrane	65
6.2.7.2 Detection of antigen	65
6.3 RESULTS.....	65
6.3.1 Virus preparation	65
6.3.2 Cloning of the SCSMaV-CP	66
6.3.3 Expression of the SCSMaV-CP in a bacterial expression system.....	66
6.3.4 DIBA	66
6.3.5 Western blot	67
6.4 DISCUSSION.....	67

CHAPTER 7: DEVELOPMENT OF NUCLEIC ACID BASED DIAGNOSTIC TESTS FOR SCSMaV

7.1 INTRODUCTION.....	70
7.2 MATERIALS AND METHODS	70
7.2.1 NA sample preparation.....	70
7.2.1.1 Dot blot	70
7.2.1.2 Tissue blot.....	71
7.2.1.3 Extraction of insect nucleic acid	71
7.2.2 Hybridisation of membranes	71
7.2.3 RT-PCR	71
7.2.3.1 SCSMaV- <i>RdRP</i> specific RT-PCR	71
7.2.3.2 SCSMaV- <i>CP</i> specific RT-PCR	72

7.2.3.3 Incorporation of internal control	72
7.3 RESULTS.....	72
7.3.1 Nucleic acid dot blot hybridisation.....	72
7.3.2 Tissue blot hybridisation	72
7.3.3 RT-PCR	73
7.3.3.1 SCSMaV- <i>RdRP</i> specific RT-PCR	73
7.3.3.2 Duplex RT-PCR.....	73
7.3.3.3 SCSMaV- <i>CP</i> specific RT-PCR	74
7.3.3.4 Insect RT-PCR	74
7.4 DISCUSSION.....	74
7.4.1 Hybridisation assays	74
7.4.2 RT-PCR assays.....	75
7.4.3 Recommendations	76

CHAPTER 8: GENERAL DISCUSSION

8.1 INTRODUCTION	77
8.2 ANALYSIS OF SCSMaV-RNA NUCLEOTIDE SEQUENCE.....	78
8.3 ANALYSIS OF ORF1 - RdRP	78
8.4 ANALYSIS OF ORF2, ORF3 AND ORF4 - TGB	79
8.5 ANALYSIS OF ORF5 – CP	81
8.6 NON-CODING REGIONS	82
8.7 TAXONOMIC CLASSIFICATION OF SCSMaV	82
8.8 DETECTION OF SCSMaV	84
8.9 CONCLUSIONS	87

APPENDICES

APPENDIX A: BIOCHEMICALS AND MISCELLANEOUS CHEMICALS	89
APPENDIX B: OLIGONUCLEOTIDE PRIMERS.....	93
APPENDIX C: POLYACRYLAMIDE AND AGAROSE GELS, SOLUTIONS, BUFFERS AND MEDIA ..	101
APPENDIX D: HOMOLOGIES	108
LITERATURE CITED.....	114

**“Sugarcane striate mosaic associated virus: RNA sequence and genome organisation,
taxonomy and detection”**

Abstract

Sugarcane striate mosaic associated virus (SCSMaV) is associated with the sugarcane striate mosaic disease (SCSMD) of sugarcane in the lower Burdekin region of northern Queensland, Australia. SCSMaV has flexuous 900 x 15 nm particles with a ssRNA genome, and a previous study suggested it did not classify into established virus genera.

This thesis describes the complete nucleotide sequence and genome organisation of SCSMaV-RNA, examines the taxonomy of SCSMaV by phylogenetic analysis, and describes the development of diagnostic tests for application to field study.

SCSMaV-RNA comprises 8146 nucleotides, is polyadenylated at the 3' terminus and encodes five potential open reading frames. ORF1 encodes the putative RNA dependent RNA polymerase (RdRP), ORFs 2, 3, and 4 comprise a triple gene block (TGB) encoding putative movement function proteins and ORF5 encodes the putative coat protein (CP).

SCSMaV-RNA has the highest overall nucleotide sequence identity (38.3%) with Sour cherry green ring mottle virus, a tentative member of the genus *Foveavirus*. SCSMaV-ORF1 has the highest amino acid sequence identity (32%) with *Blueberry scorch virus* (genus *Carlavirus*). SCSMaV-ORF5 has the highest amino acid sequence identity (22.9%) with *Potato virus X* (genus *Potexvirus*). Phylogenetic analyses were consistent with these findings, showing that SCSMaV did not classify within any established virus genera. Furthermore, the particle morphology, natural host range, and mode of spread of SCSMaV differs from established virus genera.

The SCSMaV-CP was cloned, expressed in *E. coli* and the expressed protein purified and injected into rabbits to produce antisera. The resulting antisera detected the homologous antigen by Western blot assay, but showed no specificity for SCSMD extracts by Western blot or dot-immunobinding assays (DIBA).

Nucleic acid dot-blot and leaf squash hybridisations detect SCSMaV-RNA in total nucleic acid (TNA) and leaf sap from SCSMD affected sugarcane. RT-PCR directed towards part of the RdRP or the full CP detected SCSMaV-RNA in TNA extracts from SCSMD affected sugarcane, and a control to identify false negative reactions was adapted and tested. Routine CP RT-PCR gave a distinct product that identified SCSMaV-RNA in glasshouse grown SCSMD affected sugarcane and in four hemipteran insect species collected from the roots of SCSMD affected sugarcane in the Burdekin region.

It was concluded that SCSMaV has a unique RNA sequence, and molecular and biological properties indicate that it is not a member of any described plant virus genus. The description of the complete SCSMaV-RNA sequence led to the development of diagnostic RT-PCRs for the specific detection of SCSMaV with potential application to determining epidemiology and disease control strategies.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this these contains no material previously published or written by another person, except where due reference is given in the text.

I consent to this thesis being made available for photocopying or loan when deposited in the University Library.

Nicole Thompson

March 2001.

ACKNOWLEDGMENTS

Firstly, I would like to thank my supervisors Dr. John W. Randles and Dr. M. Ali Rezaian for their guidance, encouragement, support and discussions throughout the course of this study.

Special thanks to the members of the Plant Virology Laboratory, past and present including: Dr. Yoongi Choi, Dr. Valeria Torok, Mr. Anthony Mowles, Ms. Kylie Rogers, Dr. Dagmar Hanold, Dr. Akhtar Ali, Ms Georgina Smith, and Mr. Darren Graetz.

Thanks also to the members of: **CSIRO Horticulture** unit (past and present) including Dr. Dean Pountney, Dr. Claudia Fazeli, Mr Alan Little, Mr Jamus Stoner and Mr Mark Seemanpillai; **BSES** including Mr Barry Croft, Dr Peter Whittle, Dr Rob Magarey, Mr Lionel Jensen, Mr Keith Chandler and Ms Laurelea Pickering; **Dept. of Primary Industries, Qld** including Dr. John Thomas, Ms Cherie Gambley and Mr Murray Sharman; and Dr Murray Fletcher at **NSW Dept. Agriculture** for insect identification

I would also like to take this opportunity to thank members of the Department of Applied and Molecular Ecology staff: Dr Gary Taylor, Mr. Terry Feckner, Dr. Andy Austin, Prof. Otto Schmidt, Ms Anke Johnson, Ms Sue Eaton.

And many thanks to all Postgraduate students and Postdocs for their support. Special thanks to Lisa DeCandia (FUSA) for suggestions and proofing and Kylie Rogers who helped in the printing of this thesis.

I would also like to acknowledge the support of the Bureau of Sugar Experiment Stations (BSES) for the provision of a postgraduate award for this project.

Finally, I dedicate this thesis to my family and friends. Thankyou for your love and support during this time – I couldn't have done it without you!

ABBREVIATIONS

%	Percent
α	Alpha
A	Absorbance
ABS	Australian Bureau of Statistics
AGE	Agarose gel electrophoresis
AMV-RT	Avian myeloblastosis virus reverse transcriptase
ANGIS	Australian National Genome Information Service
Anon.	Anonymous
APS	Ammonium persulphate
ATP	Adenosine 5' - triphosphate
BCIP	5-bromo-4-chloro-3-indolyl phosphate
Bis	N,N'-methylene bis-acrylamide
bp	Base pairs
BSA	Bovine serum albumin
BSES	The Bureau of Sugar Experiment Stations
C	Concentration
°C	Degrees Celsius
ca.	Circa (around)
cDNA	Complementary DNA
cm	Centimetre
CP	Coat protein
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CTP	Cytidine 5' - triphosphate
dATP	2'-Deoxyadenosine, 5'-Triphosphate
dCTP	2'-Deoxycytidine, 5'-Triphosphate
DEPC	Diethyl pyrocarbonate
dGTP	2'-Deoxyguanosine, 5'-Triphosphate
DIBA	Dot immuno binding assay
DNA	Deoxyribonucleic acid
dNTP	Mixture of deoxynucleotide-triphosphates in equimolar amounts
ds	Double stranded

DTT	Dithiothreitol
dTTP	2'-Deoxythimidine, 5'-Triphosphate
E	Specific extinction coefficient
EDTA	Ethylenediaminetetra-acetic acid disodium salt
ELISA	Enzyme-linked immunosorbent assay
EtBr	Ethidium bromide
FUSA	Flinders University of South Australia
g	Gram
g	Gravity
GCG	Genetics Computer Group, Wisconsin USA
GTP	Guanosine 5' - triphosphate
H ₂ O	18 mΩ water (Milli-Q ultrapure water system)
I	Deoxyinosine
IPA	Isopropyl alcohol
IPTG	<i>Iso</i> -Propyl-β-D-thiogalactopyranoside
kb	Kilobase
kbp	Kilobase pairs
kDa	Kilodalton
L	Litre
μ-	Micro- (10 ⁻⁶)
M	Molar
m-	Milli- (10 ⁻³)
m	Metre
mA	Milli Ampere
M _r	Molecular ratio
mRNA	Messenger ribonucleic acid
MW	Molecular weight
n-	Nano- (10 ⁻⁹)
NA	Nucleic acid
NBT	Nitro blue tetrazolium
nt	nucleotide
nts	nucleotides
OD	Optical density
ORF	Open reading frame

PAG	Polyacrylamide gel
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PNK	Polynucleotide kinase
PVP	Polyvinylpyrrolidone
RACE	Rapid amplification of cDNA ends
RdRP	RNA dependent RNA polymerase
RNA	Ribonucleic acid
RNAse	Ribonuclease
rPCR	Randomly primed PCR
rpm	Revolution per minute
RT	Reverse transcriptase
SCSMD	Sugarcane striate mosaic disease
SDDW	Sterile 18 mΩ water
SDS	Sodium dodecyl sulphate
ss	Single stranded
TAE	Tris-acetate EDTA
TBE	Tris-borate EDTA
TEMED	N N N' N'-Tetramethylethylenediamine
TGB	Triple gene block
TGBp	Triple gene block protein
TNA	Total nucleic acid
Tris	Tris(hydroxymethyl)aminomethane
U	Unit
UTP	Uridine 5' - triphosphate
UTR	Untranslated region
UV	Ultra violet
V	Voltage
v/v	Volume:volume ratio
vol	Volumes
w/v	Weight:volume ratio

X-Gal 5-Bromo-4-chloro-3-indolyl- β -D-galactosidase

VIRUS ACRONYMS

Acronym	Virus (Genus)
ACLSV	<i>Apple chlorotic leaf spot virus (Trichovirus)</i>
AltV	<i>Alternanthera virus (Potexvirus)</i>
ALV	<i>Apricot latent virus (Foveavirus)</i>
ASGV	<i>Apple stem grooving virus (Capillovirus)</i>
ASPV	<i>Apple stem pitting virus (Foveavirus)</i>
BBScV	<i>Blueberry scorch virus (Carlavirus)</i>
CaCMV	<i>Cassava common mosaic virus (Potexvirus)</i>
CaLV	<i>Carnation latent virus (Carlavirus)</i>
CCA	<i>Cherry capillovirus A (Capillovirus)</i>
CIYMV	<i>Clover yellow mosaic virus (Potexvirus)</i>
CPMMV	<i>Cowpea mild mottle virus (Carlavirus)</i>
CTLV	<i>Citrus tatter leaf virus (Capillovirus)</i>
CVB	<i>Chrysanthemum virus B (Carlavirus)</i>
CyMV	<i>Cymbidium mosaic virus (Potexvirus)</i>
FDV	<i>Fiji disease virus (Fijivirus)</i>
GarLV	<i>Garlic latent virus (Carlavirus)</i>
Gar-VA	<i>Garlic virus A (Allexivirus)</i>
Gar-VB	<i>Garlic virus B (Allexivirus)</i>
Gar-VC	<i>Garlic virus C (Allexivirus)</i>
GRSPaV	<i>Grapevine rupestris stem pitting associated virus (Foveavirus)</i>
GVA	<i>Grapevine virus A (Vitivirus)</i>
GVB	<i>Grapevine virus B (Vitivirus)</i>
HelVS	<i>Helenium virus S (Carlavirus)</i>
LVX	<i>Lily virus X (Potexvirus)</i>
PAMV	<i>Potato aucuba mosaic virus (Potexvirus)</i>
PapMV	<i>Papaya mosaic virus (Potexvirus)</i>
PasiaMV	<i>Plantago asiatica mosaic virus (Potexvirus)</i>
PCV	<i>Peanut clump virus (Pecluvirus)</i>

PVM	<i>Potato virus M (Carlavirus)</i>
PVS	<i>Potato virus S (Carlavirus)</i>
PVT	<i>Potato virus T (Trichovirus)</i>
PVY	<i>Potato virus Y (Potyvirus)</i>
PVX	<i>Potato virus X (Potexvirus)</i>
PVYV	<i>Pear vein yellows virus (Foveavirus)</i>
RSPaV	<i>Rupestris stem pitting associated virus (Foveavirus)</i>
SCBV	<i>Sugarcane bacilliform virus (Badnavirus)</i>
SCGRMV	<i>Sour cherry green ring mottle virus (tentative Foveavirus)</i>
SCMMV	<i>Sugarcane mild mosaic virus (Tentative Closterovirus)</i>
SCMV	<i>Sugarcane mosaic virus (Potyvirus)</i>
SCSMaV	<i>Sugarcane striate mosaic associated virus</i>
SCYLV	<i>Sugarcane yellow leaf virus (Luteovirus)</i>
ShV-X	<i>Shallot virus X (Alexivirus)</i>
SLV	<i>Shallot latent virus (Carlavirus)</i>
SMYEaV	<i>Strawberry mild yellow edge associated virus (Potexvirus)</i>
SSEV	<i>Sugarcane streak Egypt virus (Mastrevirus)</i>
SSMV	<i>Sugarcane streak Mauritius virus (Mastrevirus)</i>
SSV	<i>Sugarcane streak virus (Mastrevirus)</i>
TMV	<i>Tobacco mosaic virus (Tobamovirus)</i>
WCIMV	<i>White clover mosaic virus (Potexvirus)</i>

CHAPTER 1

GENERAL INTRODUCTION



CHAPTER 1: GENERAL INTRODUCTION

1.1 The sugar industry

Sugar is produced from two sources, sugarbeet and sugarcane, with the latter producing the majority of sugar traded worldwide. Australia is one of the largest exporters of raw sugar, exporting 3.46 million tons (85%) of a record 4.3 million tons produced in 1993/94 (Australian Bureau of Statistics (ABS), <http://www.abs.gov.au>). In Australia sugarcane is a crop of major economic importance having the fourth largest gross value in 1999-2000 totalling \$886 million, after wheat (\$4500 million), seed cotton (\$1500 million) and grapes (\$1100 million) (ABS, Value of Principal Agricultural Commodities Produced, 7501.0). The sugar industry in Australia had grown steadily to exports of \$1.2 million with yields of 39,531,000 tonnes with crops covering over 415,000 hectares in 1997-98 (ABS, AgStats (7117.0); Value of Agricultural Commodities Produced, Australia (7503.0); Agriculture, Australia (7113.0)), but a recent fall in world sugar prices has seen the gross value of sugar cane cut for crushing down by 15% in 1999-2000 (ABS, Value of Principal Agricultural Commodities Produced, 7501.0).

Sugarcane is grown along the 2100km coastal strip from Mossman in northern Queensland to Grafton in northern New South Wales (Figure 1.1), and at a small site near the Ord river in Western Australia. Queensland produces 95% of raw sugar with the remaining 5% produced in northern NSW. In Queensland the greatest production is in the Burdekin region, just south of Ayr (Figure 1.1).

The varieties of sugarcane grown in Australia are tightly controlled to optimise the sugar yield whilst taking into account factors such as climate and disease resistance.

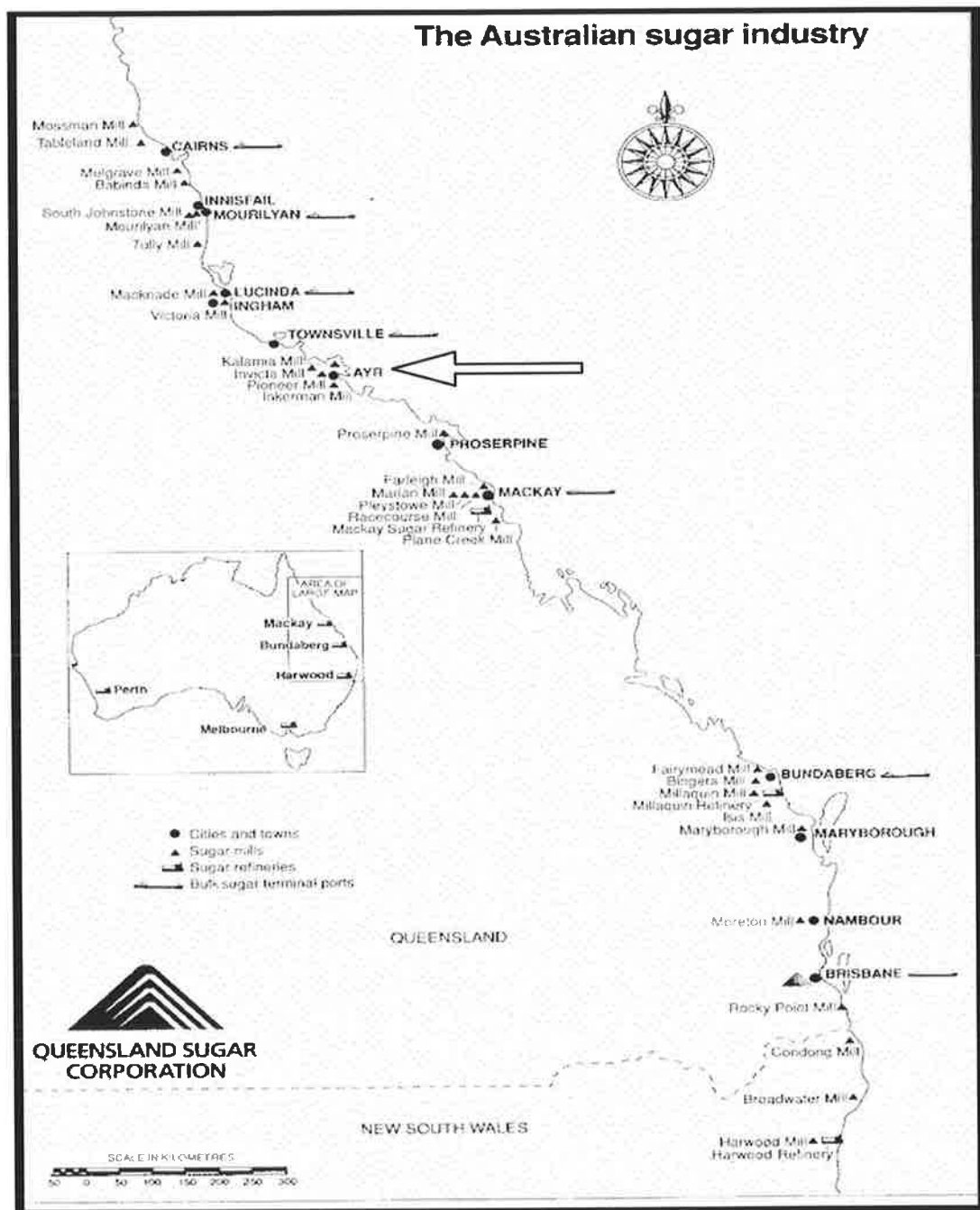


Figure 1.1

The sugar growing regions in Australia.

The arrow shows Ayr in the Burdekin region of Queensland.

Sugarcane is a vegetatively propagated crop, with cane being grown from setts/cuttings with two or more buds. Seed is only used for breeding to produce new varieties. Transgenic sugarcane is successfully being developed to introduce new desirable characteristics into high yielding varieties, however this is still experimental (Birch, 1996).

1.2 Sugarcane virus diseases

An estimated yield loss of approximately 20% per annum worldwide is caused by diseases of sugarcane (Food and Agriculture Organisation (FAO), 1982). Table 1.1 lists diseases that have a major impact on the world production of raw sugar. The presence of the disease in Australia is indicated and detection and control methods, if available, are shown. Sugarcane striate mosaic disease is found only in Australia, and is discussed in the following sections.

1.3 Sugarcane striate mosaic disease (SCSMD)

1.3.1 Introduction

SCSMD was first reported in 1961 in the sugarcane hybrid clones Pindar and Q56 in the lower Burdekin district, near Ayr, Queensland (Hughes, 1961a; Hughes 1961b). The distribution of this disease is limited, but it causes complete crop loss in affected areas due to premature plough-out of fields (B. Croft, pers.comm.).

Nematicidal and fungicidal treatment have been reported to reduce the incidence of the disease and it has been shown that the disease recurs on the same farm after ploughing out. These findings suggest that the disease is soil-borne (Anon., 1965). In the field, SCSMD is spread by infected planting material (Hughes, 1961a; Hughes, 1961b) and

Table 1.1: Examples of virus diseases of sugarcane with detection and control methods
(Adapted from Croft and Smith, 1996; Rott *et al.*, 2000).

'A' indicates the virus disease is present in Australia

	Disease, virus and acronym	Virus genus	Detection and Control
A	Fiji disease <i>Fiji disease virus</i> (FDV)	<i>Fijivirus</i>	RT-PCR (Smith <i>et al.</i> , 1992; Smith and Vandavelde, 1994), NA hybridisation (Skotnicki <i>et al.</i> , 1986) and ELISA (Wagih and Adkins, 1996). Disease-free planting and resistant varieties
A	Sugarcane mosaic disease <i>Sugarcane mosaic virus</i> (SCMV)	<i>Potyvirus</i>	Tissue print immunoblotting (Hohmann <i>et al.</i> , 1996), PCR (Smith and Vandavelde, 1994). Heat treatment and resistant varieties
A	Striate mosaic disease Sugarcane striate mosaic associated virus (SCSMaV)	Unclassified	RT-PCR (Chapter 7, this thesis). Disease-free planting material and resistant varieties
A	Bacilliform virus disease <i>Sugarcane bacilliform virus</i> (SCBV)	<i>Badnavirus</i>	PCR (Braithwaite <i>et al.</i> , 1995). Disease-free planting
A	Mild mosaic disease Sugarcane mild mosaic virus (SCMMV)	Tentative <i>Closterovirus</i>	Immuno electron microscopy (Lockhart <i>et al.</i> , 1996). Disease-free planting.
A	Yellow leaf syndrome (YLS) <i>Sugarcane yellow leaf virus</i> (SCYLV) NB. YLS has also been associated with a phytoplasma	<i>Luteovirus</i>	ELISA (Scagliusi and Lockhart, 2000); RT-PCR (Comstock <i>et al.</i> , 1998). Disease-free planting. [Detection of phytoplasma by PCR (Cronje <i>et al.</i> , 1998a and b)]
	Red leaf mottle disease <i>Peanut clump virus</i> (PCV)	<i>Pecluvirus</i>	Electron microscopy and serological methods (Rott, 1995; Rott and Chatenet, 2000). Disease-free planting.
	Streak disease <i>Sugarcane streak virus</i> , <i>Sugarcane streak Mauritius virus</i> , <i>Sugarcane streak Egypt virus</i> (SSV, SSMV, SSEV)	<i>Mastrevirus</i>	Serological methods and electron microscopy. Resistant varieties. (Rott and Peterschmitt, 2000)

can be transmitted vegetatively, however cuttings can not be cured, and the only means of control is the use of resistant cultivars (Hughes, 1961a; Hughes, 1961b).

Transmission of SCSMD from diseased to healthy sugarcane by mechanical inoculation was attempted using pinpricks and was reported to be successful (Hughes, 1964). A repeat of this experiment was attempted recently using other susceptible sugarcane hybrids and indicator species, but this was not successful (Choi, 1997). Transmission of this disease may thus be dependent on a vector, but this has not been identified.

1.3.2 Symptoms of SCSMD

1.3.2.1 Plant symptoms

The characteristic symptoms of this disease are short chlorotic striations found on most leaves and a general yellowing of the sugarcane. Chlorotic striations on the leaf sheath and stems of the cane are also reported in some cases. The stems of SCSMD cane have a shorter internode length, less thickening of the diameter, and the symptoms worsen for subsequent ratoons (Choi, 1997). SCSMD plants had a reduced growth rate compared to healthy plants from the time of germination onwards, with SCSMD cane growing at approximately 20 cm/week and healthy cane growing at approximately 50 cm/week under glasshouse conditions (Choi, 1997).

1.3.2.2 Leaf and cell structural changes

Differences between SCSMD and healthy plant leaf surface were observed by scanning electron microscopy, with the affected plants having a smoother texture due to fewer leaf spikes, and a thicker wax layer, particularly on the lower surface. Chloroplasts of SCSMD plant cells had a deformed thylakoid membrane, and electron dense material

was present between the membranes when viewed by transmission electron microscopy. Membrane bound, single walled, apparently empty vesicles of varying size were found in the cytoplasm, especially near the deformed chloroplasts. A greater number of mitochondria were observed in diseased plant cells, as well as more and thicker plasmodesmata (Choi, 1997).

1.3.3 Identification of a putative causal agent

Three dsRNAs were found to be associated with striate mosaic disease suggesting that a virus may be associated with the disease (Anon., 1985). dsRNA was isolated from SCSMD sugarcane (Choi and Randles, 1997), and random PCR (rPCR) clones constructed from the largest dsRNA fragment to use as a hybridisation probe for fractions collected during virus preparation (Choi, 1997; Choi *et al.*, 1999). Virions were observed in a fraction that hybridised to the dsRNA rPCR probe (Choi, 1997).

1.3.3.2 Virions

Virions observed by transmission electron microscopy (TEM) were 15nm wide, with the distribution of length bimodal comprising two main components ca. 1000 ± 100 nm and ca. 2000 ± 100 nm, probably representing monomers and dimers (Choi, 1997). It was concluded that the size of the virions was 15 nm x 950 nm (Choi *et al.*, 1999). A protein component associated with the virus preparations was 51 kDa (Choi, 1997; Choi *et al.*, 1999). The nucleic acid was single stranded RNA (ssRNA) polyadenylated at the 3' terminus (Choi, 1997). The SCSMD associated virions were named *Sugarcane striate mosaic associated virus* (SCSMaV). It was suggested that SCSMaV may be a member of the genera *Foveavirus* or *Carlavirus*, as the virions resemble *Apple stem pitting virus* (ASPV) and *Blueberry scorch virus* (BBScV) (Choi, 1997; Choi *et al.*, 1999).

1.3.3.3 Preliminary sequence analysis

Five rPCR clones generated from the disease associated dsRNA were sequenced and gave three non-overlapping segments representing ca. 28% of the SCSMaV-RNA genome (Choi, 1997; Choi *et al.*, 1999). Genome database comparison showed nucleotide sequence similarity with rod-shaped viruses in the genera *Potexvirus*, *Foveavirus*, *Carlavirus*, *Capillovirus*, *Trichovirus*, and *Potyvirus*. The greatest similarity was found between SCSMaV-RNA and the open reading frame 1 (ORF1) of ASPV, which encodes the putative RNA dependent RNA polymerase (RdRP) (Choi, 1997; Choi *et al.*, 1999). It was concluded that SCSMaV was most closely related to species in the genus *Foveavirus*, but does not classify within any established plant virus genus (Choi *et al.*, 1999).

1.4 Virus taxonomy and phylogenetics

1.4.1 Origin of viruses

The origin of viruses is believed to be very ancient, with viruses possibly present before the divergence of the three major kingdoms (Gibbs and Keese, 1995) or emerging soon after that in a polyphyletic manner (Ward, 1993). It is also possible that the expression and replication mechanisms of viruses evolved more than once (Koonin and Dolja, 1993). Popular theories as to how viruses arose include: (i) viruses evolved from autonomous self replicating host cell molecules (eg plasmids, transposons) by acquiring genes that encoded for packaging molecules; (ii) viruses arose from degeneration of primitive cells by losing their cellular membranes and utilising the host-cell machinery for replication; (iii) RNA viruses may be descendants from pre-biotic RNA polymers; or

(iv) that viruses evolved from viroids (Ward, 1993). It is possible that any or all of these theories may be true, or that they arose by a different mechanism (Roossinck, 1997).

1.4.2 Recombination

Recombination is involved in the evolution of viruses in a horizontal manner (Lai, 1995; Mayo and Pringle, 1998; Meyers *et al.*, 1995; Roossinck, 1997; Simon and Bujarski, 1994; Ward, 1993). Recombination can be of three distinct types: (i) homologous recombination - recombination between two viruses of near identical sequence at the same place of the genome; (ii) aberrant homologous recombination - recombination between two viruses of near identical sequence at different parts of the genome, and; (iii) non-homologous recombination - recombination between any two sequences independent of the original position, be it viral or host nucleic acid (Meyers *et al.*, 1995; Simon and Bujarski, 1994). Homologous recombination between viruses is believed to play a role in the maintenance of the RNA genome sequence as the RdRP does not have a proof-reading ability (Lai, 1995). This could also lead to rapid evolution of RNA viruses because the lack of proof-reading ability means that mutations are retained, which could lead to the alteration of biological properties and therefore speciation (Lai, 1995). Aberrant and non-homologous recombination may lead to blocks of nucleic acid encoding a gene, or genes, being transferred from one species to another. Horizontal or 'modular' evolution (Ward, 1993) can be illustrated by phylogenetic analysis of different regions of a virus genome. Different trees and relationships can be observed for each genome region analysed, indicating recombination has taken place (Mayo and Pringle, 1998; Ward, 1993). A recent example of large genome rearrangements by recombination has been found in the family *Luteoviridae*, in which it is proposed that

Sugarcane yellow leaf virus evolved by recombination between Luteoviral and Poleroviral ancestors (Moonan *et al.*, 2000).

1.4.3 Virus taxonomy

The International Committee for the Taxonomy of Viruses (ICTV) lists criteria for the classification of viruses into species (van Regenmortel *et al.*, 2000; Table 1.2). In most cases, the order and family are not developed as a concept and classification of plant viruses is generally limited to the genus and species level (Mayo and Pringle, 1998). The following discussion is restricted to positive sense ssRNA viruses with flexuous filamentous particles.

1.4.4 Phylogenetic studies

The taxonomy of a virus should accurately reflect its phylogeny (Mayo and Pringle, 1998). For ssRNA viruses, generally phylogenetic analysis of the RdRP was found to separate virus species into their genera (Mayo and Pringle, 1998; Ward, 1993). Phylogenetic classification based on complete nucleotide genome sequence is possible if major recombination has never occurred in the ancestry of groups of viruses (Mayo and Pringle, 1998). However, recombination amongst distantly related viruses is likely to be a major evolutionary trend (Dolja *et al.*, 1991) which would make it impossible to determine the evolutionary history of positive sense ssRNA viruses using nucleotide sequence data. An evolutionary model for viruses must incorporate both evolution from one generation to the next (vertical evolution) and evolution within generations through recombination (horizontal evolution, see 1.4.2; Koonin and Dolja, 1993). Therefore in order to be accurate in the classification of a new virus, it is necessary to take into account all available data including phylogenetic analysis using different algorithms (Weiller *et*

Table 1.2: List of criteria demarcating different virus taxa.

(From ICTV web site <http://www.ncbi.nlm.nih.gov/ICTV/Criteria.html>)

I. Order	<u>Common properties between several families including:</u> • Biochemical composition • Virus replication strategy • Particle structure (to some extent) • General genome organisation
II. Family	<u>Common properties between several genera including:</u> • Biochemical composition • Virus replication strategy • Nature of the particle structure • Genome organisation
III. Genus	<u>Common properties within a genus including:</u> • Virus replication strategy • Genome size, organisation and/or number of segments • Sequence homologies (hybridization properties) • Vector transmission
IV. Species	<u>Common properties within a species including:</u> • Genome rearrangement • Sequence homologies (hybridization properties) • Serological relationships • Vector transmission • Host range • Pathogenicity • Tissue tropism • Geographical distribution

al., 1995) and different regions of the genome, the viral RNA sequence, genome, biological and morphological characteristics (van Regenmortel *et al.*, 2000).

1.5 Genera representing viruses resembling SCSMaV

The most similar viruses to SCSMaV belong to three genera: *Foveavirus*, *Carlavirus* and *Potexvirus* (Choi, 1997; Choi *et al.*, 1999).

1.5.1 *Foveavirus*

The type member of the genus *Foveavirus* is *Apple stem pitting virus* (ASPV). Members of this genus have flexuous filamentous virions of ca. 800 nm x 15 nm which show a tendency for end-to-end aggregation (Koganezawa and Yanese, 1990), SCGRMV has slightly larger virions ca. 1000 nm (Zagula *et al.*, 1989). The genome is ca. 8.4-9.3 kb and has 5 ORFs, with ORF 1 encoding the viral replicase, ORF 2-4 the triple gene block (TGB) encoding putative movement proteins, and ORF 5 encoding the coat protein (Jelkmann, 1994; Meng *et al.*, 1998; Zhang *et al.*, 1998). The genome replication is likely to occur in the cytoplasm with the direct expression of ORF1 from the RNA, and the expression of downstream ORFs through subgenomic RNAs (van Regenmortel *et al.*, 2000). The natural host range is restricted to one or a few dicotyledonous hosts. No vector is known for any members of this group, however they are transmitted by grafting and mechanical transmission has been reported (Van Dijk *et al.*, 1987).

1.5.2 *Carlavirus*

Diseases which are associated with viruses of the genus *Carlavirus* generally have mild symptoms or are symptomless (Brunt *et al.*, 1996 onwards; van Regenmortel *et al.*, 2000). Mildness of the symptoms may cause problems in the control of the disease, as

there are no means of visually monitoring the spread of the virus. Some members of this genus are mechanically transmitted (van Regenmortel *et al.*, 2000) and the RNA alone has also been shown to be infectious (Hillman and Lawrence, 1994). Aphids can also transmit some members of this genus in a non-persistent manner (Cavileer *et al.*, 1994), and this route is suggested to be the most common mode of transmission in nature (Brunt, 1995; Hillman and Lawrence, 1994). Whiteflies can transmit some members of this genus, but the efficiency of transmission is often low (Badge *et al.*, 1996). The variety of hosts infected by members of this genus is wide, but the range of an individual virus is restricted (Brunt, 1995; Hillman and Lawrence, 1994).

Carlavirus virions are slightly flexuous filaments ca. 610-700 nm x 12-15 nm (van Regenmortel *et al.*, 2000). The genome is ca. 7.4-8.5 kb and has six ORFs, ORF1 encodes the putative RdRP, ORFs 2, 3 and 4 comprise the TGB encoding putative movement proteins, ORF5 encodes the putative CP, and ORF 6 encodes a 11 kDa putative cysteine rich protein of unknown function (van Regenmortel *et al.*, 2000). The four cysteine residues of ORF6 are predicted to be within a nucleic acid binding zinc finger (Haylor *et al.*, 1990).

The five 3' ORFs are suggested to be expressed by at least 2 subgenomic RNAs (sgRNAs): one for the expression of the TGB and one for the CP (Hillman and Lawrence, 1994). Potential ribosome binding sites have been identified upstream of ORF2 and ORF5 (Foster and Mills, 1991) and the resultant subgenomic RNAs have sequence similarity at the 3' end and a poly-A tail (Hillman and Lawrence, 1994). The expression of ORF6 may be either an internal initiation of a second start codon or ribosomal frameshifting producing a CP/12K transframe protein (Gramstat *et al.*, 1994).

Members of the genus *Carlavirus* and genus *Potexvirus* have some nucleic acid and amino acid sequence similarity (Mackenzie *et al.*, 1989; Memelink *et al.*, 1990; Morozov *et al.*, 1990; Rupasov *et al.*, 1989) and display some cross reaction in serological tests, and are therefore assumed to be closely related (Hillman and Lawrence, 1994).

1.5.3 Potexvirus

The type member for the genus *Potexvirus* is *Potato virus X* (PVX). Virions are flexuous helical rods ca. 470-580 nm x 13 nm. The genome is ca. 6.5 kb and contains five putative ORFs: ORF1 encoding the RdRP; ORFs 2, 3 and 4 comprise the TGB which are involved in cell-to-cell movement; and ORF5 encodes the CP. The 5' UTR of PVX is 83 nt and is reported to have a stem-loop structure required for plus sense strand RNA accumulation during replication (Miller *et al.*, 1998). ORF1 is translated directly from the viral RNA and ORFs 2-5 are expressed via either two or three sgRNAs (van Regenmortel *et al.*, 2000), with the TGB requiring two subgenomic RNAs for translation (Verchot *et al.*, 1998). Members of the *Potexvirus* genus are usually moderately pathogenic causing mosaic or ringspot symptoms in a wide range of monocotyledonous and dicotyledonous plants, however the host range of individual viruses is limited (van Regenmortel *et al.*, 2000). Transmission is generally mechanical, however Potato aucuba mosaic virus (PAMV) is transmitted in conjunction with a helper virus (Potato virus Y *Potyvirus*) in a non-persistent manner by an aphid vector (Matthews, 1991).

1.6 Detection methods

1.6.1 Serological methods

The detection of plant viruses using antisera generated against the virion includes dot immuno-binding assay (DIBA), enzyme-linked immunosorbent assay (ELISA) and Western blot analysis (Cеровска and Filigarova, 1995; Clark and Adams, 1977; Filhart *et al.*, 1998; Garcia *et al.*, 1997; Hill *et al.*, 1996; Hsu *et al.*, 1995; Makkouk *et al.*, 1993; Schonborn *et al.*, 1991; Wagih and Adkins, 1996; Zagula *et al.*, 1989; Zhang *et al.*, 1998). A high level of purity of antigen is required to immunise animals to minimise non-specific reactions (Randles *et al.*, 1996). This can be achieved either by using purified virus particles (eg. Zagula *et al.*, 1989) or cloning the *CP* and expressing it in an expression system (eg. Minafra *et al.*, 2000). This versatile technique allows *CP* expression for antisera production for detection systems (eg. Ling *et al.*, 2000), virus capture for immuno-electron microscopy (Jelkmann and Keim-Konrad, 1997) as well as for structural studies of the *CP* (Yusibov *et al.*, 1996). The sensitivity of ELISA is less than that of RT-PCR (see below), however it has been reported that ELISA can have a higher disease detection rate than RT-PCR due to the highly specific RT-PCR being unable to detect some virus variants (Garcia *et al.*, 1997).

1.6.2 Nucleic acid methods

1.6.2.1 Hybridisation

The detection of viruses by the hybridisation of the viral nucleic acid (NA) to a radioactively labelled probe is an easy and fast method of detection. NA dot blot hybridisation onto nylon membrane followed by either radioactive or non-radioactive

detection provides a rapid screening tool for virus RNA (Ali *et al.*, 1998; Caciagli and Bosco, 1996; James *et al.*, 1998; Skotniki *et al.*, 1986). Tissue printing has a wide range of uses including the localisation of protein, enzymes and RNA whilst maintaining cellular structure (Varner and Ye, 1994), and has been adapted for the study of long-distance movement of viruses (Mas and Pallas, 1995). Detection of viruses by hybridisation in squash blots onto nylon membrane has been done for both DNA (*Tomato yellow leaf curl virus*: TYLCV) and RNA (*Tobacco mosaic virus*: TMV) viruses. This method provides rapid tissue processing with little labour cost.

1.6.2.2 RT-PCR

The advantages of using RT-PCR for the detection of plant RNA viruses are: no culturing of organisms is necessary; it is very sensitive, rapid and versatile allowing the detection of a few target molecules (Saiki *et al.*, 1988); and once it has been set up the cost per reaction is less than serological based detection methods such as ELISA (Henson and French, 1993).

RT-PCR can be used for detection of either a single virus species (Halpern and Hillman, 1996; Smith *et al.*, 1992; Thomson *et al.*, 1996), taxonomically similar viruses (Braithwaite *et al.*, 1995; Robertson *et al.*, 1991), taxonomically different viruses (Seoh *et al.*, 1998), or suspected viral components such as dsRNA (Haber *et al.*, 1995). Detection of RT-PCR products can be done by several methods including agarose or polyacrylamide gel electrophoresis (Sambrook *et al.*, 1989) and molecular beacons (Eun and Wong, 2000).

Immuno-capture RT-PCR or PCR uses antibodies to capture virus particles from infected sap, and RT-PCR or PCR to amplify the virus nucleic acid (Harper *et al.*, 1999;

Jelkmann and Keim-Konrad, 1997). The products can be colorimetrically detected by PCR-ELISA (Rowhani *et al.*, 1998). These methods are useful when PCR reactions on TNA extracts fail due to inhibitors in the reaction from the plant, and provide enrichment of the starting material for amplification (Jelkmann and Keim-Konrad, 1997).

RT-PCR is more sensitive than serological methods such as ELISA (Haber *et al.*, 1995; Hadidi *et al.*, 1993; Halpern and Hillman, 1996). NA dot blots are restricted in their sensitivity by the amount of viral nucleic acid in the sample, and are predicted to have a sensitivity equal to or slightly higher than ELISA (Fenby, 1995). The increased sensitivity of RT-PCR is useful for the early detection of infection, and for detection of viruses that are suspected to be of low titre (Fenby, 1995).

1.7 Vectors

SCSMD has been reported to be retained in the soil for many years with little spread (Croft and Randles, 2000) and also after fumigation treatments (Anon., 1965; section 1.3.1). This suggests that there may be no biotic vector or any possible vector is soil-based. The species in the genera that are thought to be most closely related to SCSMaV (section 1.5) have either no known vector or are transmitted by whiteflies or aphids (Matthews, 1991; van Regenmortel *et al.*, 2000). Some hemipteran insect species are known to live underground at some stage of their life cycle, and this group of vectors cannot be precluded as potential vectors of SCSMaV. The soil borne nature of SCSMD suggests three other possible modes of transmission. SCSMaV could be transmitted on or within fungal spores (Campbell, 1996), by a nematode vector or the virus could persist in the soil with no vector (Matthews, 1991).

1.8 Scope of this thesis

This thesis describes the complete nucleotide sequence of the SCSMaV-RNA molecule and its genome organisation. Taxonomic classification of SCSMaV by phylogenetic analysis is described, and the data gained from this thesis was collated with previously gained data to conclude that SCSMaV is unique and does not readily classify into any known taxa. The SCSMaV-CP was cloned and expressed in a bacterial expression system, however the attempt to produce antisera specific to SCSMD was not successful. Nucleic acid based detection methods including NA dot blot and tissue squash hybridisation and RT-PCR assays directed towards the SCSMaV-RdRP and SCSMaV-CP detect SCSMaV-RNA in SCSMD affected sugarcane. The SCSMaV-CP RT-PCR was used for the preliminary screening of possible insect vectors of SCSMaV.

CHAPTER 2

MATERIALS AND METHODS

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

2.1.1 Sugarcane plants

SCSMD affected sugarcane was obtained from known striate mosaic affected patches on farms in the Burdekin region, Queensland, and grown under glasshouse conditions (22°C) at the Waite Campus, University of Adelaide. Healthy cane of variety Q96 from the Bureau of Sugar Experiment Stations (BSES), Ayr was germinated and grown separately from affected cane.

2.1.2 Chemical reagents

Chemical and biochemical reagents and their suppliers are listed Appendix A.

2.1.3 Bacterial strains

Escherichia coli DH5 α : F' ϕ 80dlacZ Δ M15 Δ (lacZYA-argF) U169 *deoR* *recA1* *endA1* *hsdR17* (*r_k*⁻, *m_k*⁺) *phoA* *supE44*, λ *thi-1* *gyrA96* *relA1* cells were obtained competent from Gibco BRL Life Technologies.

E. coli M15[pREP4]: *Nal*^s, *Str*^s, *Rif*^s, *Thi*⁻, *Lac*⁻, *Ara*⁺, *Gal*⁺, *Mtl*⁻, F⁻, *RecA*⁺, *Uvr*⁺, *Lon*⁺, [pREP4] cells were obtained from Qiagen.

E. coli JM109: *recA1*, *supE44*, *endA1*, *hsdR17*, *gyrA96*, *relA1*, *thi*, Δ (*lac-proAB*), [F['], *traD36* *proAB*, *lacI*^qZ Δ M15] cells were obtained from Promega.

Previously prepared SCSMaV random PCR (rPCR) clones (see 1.3.3; rPCR products in pGEM[®]-T plasmid vector) in *E. coli* DH5 α cells were stored as glycerol stocks at -70°C (Choi, 1997).

2.1.4 Kits

JetQuick Plasmid Mini-Prep Spin kit was obtained from Astral Scientific; the QIAquick PCR and Gel purification kits were obtained from Qiagen; the pGEM[®]-Teasy vector kit was obtained from Promega; the 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 was obtained from Gibco BRL Life Technologies; and the QIAexpressionist kit components were obtained from Qiagen.

2.1.5 Oligonucleotide primers

Oligonucleotides were obtained from Gibco BRL Life Technologies (Melbourne, Australia), GENSET (Singapore) or Geneworks (Adelaide, Australia). A list of all oligonucleotides and their predicted annealing sites to SCSMaV (where appropriate) are shown in Appendix B. Oligonucleotides obtained from GENSET were resuspended in SDDW. Oligonucleotides from Gibco BRL Life Technologies and Geneworks were initially resuspended in 100 μl SDDW, and the concentration calculated using the following formula:

$$C \text{ (mM)} = \frac{A_{260}}{E_{260} \times \text{path length (1 cm)}} \times \text{Dilution Factor} \quad (\text{Anon., 1999a})$$

All primers were diluted to 10 μM and 5 μM and stored at -20°C .

2.1.6 Solutions, buffers and media

Solutions, buffers, media, and polyacrylamide and agarose gel components are listed in Appendix C.

2.2 Methods

2.2.1 Gel electrophoresis

2.2.1.1 Agarose gel electrophoresis (AGE)

Nucleic acids (NA) were fractionated by electrophoresis on a 1%-1.5% (w/v) agarose gel in TAE (AGE) (Appendix C) containing 0.5 µg/ml ethidium bromide. Samples were loaded with 1 x Loading Buffer I (Appendix C). The 1kb Plus DNA ladder was loaded and the sizes of the bands on each gel are shown in base pairs/nucleotides (bp/nt). After electrophoresis at 1-5 volts per cm of agarose gel, the components were identified by UV fluorescence.

2.2.1.2 Polyacrylamide gel electrophoresis (PAGE)

Double stranded RNA (dsRNA) was fractionated on a 6% polyacrylamide (87:1 acrylamide:bisacrylamide) gels (Appendix C). Samples were loaded with 1 x Loading Buffer I and electrophoresed in 1 x TBE (Appendix C) at 100 V for 15-17 hours for 150 x 180 x 0.5 mm gels, and at 150 V for 2-3 hours on Bio-rad[®] minigels (100 x 100 x 0.5 mm). The dsRNA was visualised by silver staining (2.2.1.4a).

2.2.1.3 Sodium dodecyl sulphate PAGE (SDS-PAGE)

Discontinuous SDS-PAGE (Laemmli, 1970; Sambrook *et al.*, 1989) was used for the analysis of protein. Gels were prepared (Appendix C) with a 5% polyacrylamide stacking layer (buffered in 125 mM Tris-HCl pH 6.8 and 0.1% SDS) and a 12-15% polyacrylamide resolving layer (buffered in 375 mM Tris-HCl pH 8.8 and 0.1% SDS). Protein samples were heated at 95°C for 5 minutes in 1 x SDS-PAGE loading buffer (Appendix C) and electrophoresed in 1 x Tris-glycine buffer (pH 8.3; Appendix C) at 8 V/cm in the stacking gel increasing to 15 V/cm for the resolving gel. Gels were run until the buffer front had reached the bottom of the gel. Precision Prestained Protein Standards (BioRad®) were loaded and the size of the bands shown in kDa. The proteins were visualised by Coomassie blue staining (2.2.1.4b).

2.2.1.4 Staining of polyacrylamide gels (PAG)

2.2.1.4a Silver staining for visualisation of nucleic acids

Gels were washed for 5-10 minutes in 10% ethanol/ 5% acetic acid, and then for 5-10 minutes in 10% ethanol/ 0.5% glacial acetic acid, and stained for 1 hour in 0.2% silver nitrate with gentle agitation. The gel was washed briefly in degassed water once for 10 seconds, and twice for 30-60 seconds. The developing solution (0.375 M NaOH, 0.15% formaldehyde, 3 mM NaBH₄) was added and the gel developed with constant agitation until the bands were visible. The reaction was stopped with 5% acetic acid.

2.2.1.4b Coomassie blue staining for protein

Gels were stained for at least 4 hours in 0.25% Coomassie brilliant blue stain (Appendix C). Gels were de-stained with several changes of destaining solution I (Appendix C) for 1-2 hours and 8-16 hours in destaining solution II (Appendix C).

2.2.1.5 Maintenance of stained PAG

Stained PAG between 5% and 12% were air-dried between cellophane sheets using the Bio-Rad[®] gel-air-drying apparatus. Stained SDS-PAG of >12% were stored in a sealed plastic bag in destaining solution II (Appendix C).

2.2.2 Extraction of total nucleic acid (TNA)

2.2.2.1 From sugarcane leaf tissue

TNA was extracted from sugarcane using a modification of the method of Robertson *et al.* (1991). A small amount of sugarcane leaf material (~50 mg) was crushed in liquid nitrogen in a 1.5 ml microcentrifuge tube with a pellet pestle (Sigma). Buffer G (500 µl; Appendix C) was added and mixed by vortexing briefly. An equal volume of phenol (pH 4.5) was added and the solution mixed again by vortexing. The phenol phase was removed by centrifugation at 13000 g for 15 minutes at room temperature, and the supernatant recovered. Nucleic acids were precipitated by the addition of 0.3 M ammonium acetate and 2.5 x volume ice-cold 100% ethanol, pelleted by centrifugation at 13000 g at 4°C for 30-60 minutes, and the pellet was washed once in 70% ethanol, dried *in vacuo* and resuspended in 20 - 50 µl TE buffer. The quality of the TNA was determined by AGE (2.2.1.1).

2.2.2.2 From insect samples

Insects received in 70% ethanol were re-hydrated by washing in 100 mM NaCl, 10 mM Tris-HCl, pH 8.0, 1 mM MgCl₂, three times for 10 minutes each. Total RNA was extracted from whole insect samples (Table 7.1) consisting of 1 - 10 insects (7.2.1.3) with a modified acid guanidium thiocyanate-phenol-chloroform extraction method (Chomczynski and Sacchi, 1987) using 30 µg glycogen carrier per extraction added before the precipitation step to aid in the precipitation of insect total RNA. Insect total RNA was resuspended in 10-15 µl TE.

2.2.3 Double stranded RNA (dsRNA) extraction

dsRNA was extracted by a modification of the method of Choi and Randles (1997). Frozen sugarcane leaf tissue was powdered by blending in liquid nitrogen in a VirTis 45 homogenizer (VirTis, New York, USA). The frozen tissue was then transferred to a large beaker containing 4 vol (w/v) of STE with 1% SDS and 1% monothioglycerol. This slurry was stirred at room temperature for 30 – 60 minutes. One half volume (v/v) of 0.1M Tris-HCl saturated phenol, pH 8.0 (Appendix C) was added and stirred for an additional 30 minutes and the phases allowed to separate. The aqueous phase was mixed with 0.5 volumes (v/v) of chloroform for 30 minutes. The chloroform phase was removed by centrifugation at 10000 g for 10 minutes, and the aqueous phase collected. Nucleic acids were precipitated by the addition of 0.9 vol (v/v) isopropylalcohol (IPA) and incubation at -20°C overnight then pelleted by centrifugation at 12000 g for 30 – 60 minutes, air-dried and resuspended overnight at 4°C in 1.3 x STE (Appendix C) using 0.8 ml per 3 g of original leaf tissue. Insoluble material was removed by centrifugation at 10000 g for 20 minutes. Microgranular cellulose (Appendix A) was added to the

supernatant using 10 mg / 3 g original leaf tissue, and the slurry mixed vigorously for 30 minutes. Ethanol was added dropwise to a final concentration of 20% (v/v) with constant vigorous stirring to bind the nucleic acid to the cellulose. The slurry was mixed vigorously for 30 minutes, then transferred into a series of 2 ml microcentrifuge tubes. The cellulose was collected by centrifugation at 13000 g for 5 minutes, and the supernatant carefully decanted. The cellulose was washed with STE containing 16.5% ethanol by vortexing for 30 seconds and collected again by centrifugation. The supernatant was removed and the cellulose washed at least four more times with STE containing 16.5% ethanol. After the final wash, the cellulose pellet was drained and air dried. The dsRNA was eluted from the cellulose by the addition of 200 µl STE per 10 mg cellulose, vigorous vortexing and centrifugation as above, and collection of the dsRNA solution. The dsRNA was precipitated with 0.9 x volume IPA in 0.3 M ammonium acetate pH 5.2, for 16 hours at -20°C. The dsRNA pellet was collected by centrifugation at 13000 g, 30 min, 4°C, washed with 70% ethanol, dried *in vacuo* and resuspended in TE. The quality of the dsRNA was determined by PAGE (2.2.1.2) followed by silver staining (2.2.1.4a).

Contaminating DNA and ssRNA in microgranular cellulose purified dsRNA was removed by successive DNase, RNase and Proteinase K treatment. DNase treatment was done by adding 12.5 µl RQ1 RNase-free DNase and 12.5 µl 10 x RQ1 Buffer (Promega) to 100 µl of the dsRNA sample and incubating at 37°C for 60 minutes. The reaction was stopped by the addition of 12.5 µl of RQ1 stop solution (Promega) and incubation at 65°C for 10 minutes. The solution was adjusted to 0.3M NaCl, 50 mM Tris-HCl pH 8.0 and RNase A was added to 1 ng/µl in a total volume of 250 µl, and incubated at 37°C for 60 minutes. The reaction was stopped by the addition of 1 µg/µl

Proteinase K, 1 x Proteinase K buffer (Appendix C) in a total volume of 500 µl and incubation at 37°C for at least 60 minutes. The enzyme treatment reactions were extracted with phenol-chloroform-IPA (25:24:1), and the dsRNA precipitated as described above. The preparations were analysed by PAGE (2.2.1.2) with dsDNA size markers (1kb Plus DNA Ladder) and stained with silver (2.2.1.4a).

2.2.4 Plasmid DNA preparation

Single colonies were inoculated into 5 ml of LB containing 50 µg/ml ampicillin and grown overnight at 37°C with constant vigorous shaking. Plasmid DNA was isolated from the overnight cultures using the JetQuick plasmid DNA isolation. Generally, the plasmid DNA was eluted in 30-50 µl TE buffer and analysed by 1-1.5% AGE (2.2.1.1).

2.2.5 Restriction enzyme analysis

EcoRI digestions of plasmid preparations were done in a 20 µl reaction with 10-20 U of enzyme, 2-5 µl (1-2 µg) of purified plasmid DNA and the reaction buffer H (Promega) and incubated at 37°C for 1-2 hours. Digestion of PCR products with *Bam*HI and *Hind*III was done in a 20 µl reaction volume containing 10 U of each enzyme, 5-10 µl (1 µg) DNA and reaction buffer E (Promega), and incubated at 37°C for 1 hour. Digestions were analysed by 1.5% AGE (2.2.1.1)

2.2.6 Reverse transcriptase-polymerase chain reaction (RT-PCR)

2.2.6.1 Designing of oligonucleotide primers

Oligonucleotide primers for the extension of the SCSMaV sequence were designed to anneal to a position at least 100 nt within the known sequence (Choi and Randles, 1999), and internal to new sequence as it was obtained. Oligonucleotide primers of ca. 30 nucleotides were designed and tested using the computer programs Oligo[®] and Amplify[®]. 5' and 3' Rapid Amplification of cDNA Ends (RACE) primers were designed (Gibco BRL Life Technologies) and an anchored oligo-dT primer was used for 3' poly-A priming. The potexvirus-coat protein degenerate primer was designed by C. Gambley and M. Sherman, DPI, Qld. Oligonucleotide primers are listed in Appendix B.

2.2.6.2 Reverse transcription

2.2.6.2a dsRNA as a template

dsRNA was reverse transcribed with Superscript II Reverse Transcriptase using a modification of the manufacturer instructions (Gibco BRL Life Technologies). SCSMaV specific primer (0.4 μ M) and 8 μ L SCSMaV dsRNA were initially denatured in water at 100°C for 10 minutes, then chilled on ice. The reverse transcription was done in a 20 μ L reaction containing the denatured primer-dsRNA mix, 1 x Superscript II Reverse Transcriptase first strand buffer, 10 mM DTT and 2 mM of each dNTP (dATP, dTTP, dCTP, dGTP). This mix was equilibrated at 42°C for 2 minutes and 200 Units of Superscript II Reverse Transcriptase added and the reaction incubated at 42°C for 50 minutes. The enzyme was heat inactivated at 70°C for 15 minutes.

2.2.6.2b ssRNA as a template

TNA extracts were reverse transcribed with AMV Reverse Transcriptase with using a modification of the manufacturer instructions (Geneworks). SCSMaV specific primer (0.6 μ M) and 1 μ L SCSMaV TNA (diluted 1:10 or undiluted) were combined with water to a total volume of 7.5 μ L and denatured at 70°C for 10 minutes, then chilled on ice. AMV reaction buffer, 0.2 mM of each dNTP, 7.5 U AMV Reverse Transcriptase and 15 U RNAsin® RNase Inhibitor per reaction was then added and the volume made to 15 μ L with water. The reaction was incubated at 42°C for 40-60 minutes.

2.2.6.3 PCR reactions

PCR reactions used either DyNAzyme DNA polymerase for the production of clones for sequence analysis, or *Taq* DNA polymerase for diagnostic reactions.

2.2.6.3a DyNAzyme DNA polymerase reaction

DyNAzyme DNA polymerase was used in a 20 μ L reaction volume containing 2 μ L of the RT reaction product, forward and reverse SCSMaV primers at 0.4 μ M each (Appendix B), 1.5 mM MgCl₂, 0.2 mM each dNTP, DyNAzyme DNA polymerase reaction buffer and 0.2 U DyNAzyme DNA polymerase. Unless otherwise stated, the thermal cycling was done using an initial denaturation at 94°C for 5 minutes, followed by 30 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 60°C, and 1-3 minutes extension at 72°C, with a final extension at 72°C for 10 minutes.

2.2.6.3b *Taq* DNA polymerase reactions

Taq DNA polymerase was used in a 20 μ L reaction volume containing 2 μ L of the RT reaction product, forward and reverse SCSMaV primers at 0.4 μ M each (Appendix B),

1.5 mM MgCl₂, 0.2 mM of each dNTP, *Taq* DNA polymerase reaction buffer and 0.2 U *Taq* DNA polymerase. Unless otherwise stated, the thermal cycling was done using an initial denaturation at 94°C for 3 minutes, followed by 30-35 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at 60°C, and 1 minute extension at 72°C, with a final extension at 72°C for 7 minutes.

2.2.6.3c Duplex RT-PCR using TMV- RNA as an internal control

Tobacco mosaic virus (TMV) RNA and specific primers (Torok and Randles, 2001) were incorporated into the SCSMaV diagnostic test to test for false negative reactions. This was done by adding 10-100 pg TMV RNA and 0.3 µM TMVR3 primer to the AMV RT reaction (2.2.6.2b) and 0.2 µM of both TMVF and TMVR to the PCR reaction (2.2.6.3a or 2.2.6.3b).

2.2.6.4 Rapid amplification of cDNA ends (RACE)

5' and 3' ends were amplified using RACE of SCSMaV-RNA in TNA extracts using the Gibco BRL Life Technologies 5' RACE kit. 5' and 3' RACE primers are shown in Appendix B.

2.2.6.4a 5' RACE

5' RACE was done using the 5' RACE kit (Gibco BRL) with SCSMaV specific primers SCSMaV-L19 and SCSMaV-5GSP1 (Appendix B). Reverse transcription was done in a 25 µl reaction as suggested by the manufacturer with 0.1 µM L19 primer and 1 µl TNA extract. The cDNA was purified using a Qiagen PCR purification kit, eluted in 30 µl 10 mM Tris-HCl, pH 8.0 and poly-dC tailed with terminal deoxynucleotidyl transferase (TdT). PCR was done with DyNAzyme DNA polymerase in a 50 µl reaction containing

the reaction buffer supplied, 1.5 mM MgCl₂, 0.2 mM each dNTP, 0.4 μM Anchored Abridged Primer (Gibco BRL Life Technologies), 0.4 μM 5GSP1 primer and 0.5 U DyNAzyme DNA polymerase. The thermal cycling was done with an initial denaturation step of 94°C for 5 minutes, followed by 30 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 55°C, and 2 minute extension at 72°C, with a final extension at 72°C for 10 minutes. 5' RACE products were analysed by 1.5% AGE (2.2.1.1).

2.2.6.4b 3' RACE

3' RACE was done using an anchored oligo-dT primer (PCC-dT; Appendix B) to prime reverse transcription using Superscript II Reverse Transcriptase of SCSMaV-RNA in SCSMD TNA sample as described in 2.2.6.2a except that the initial denaturation was at 70°C. PCR was done with primers SCSMaV-3RACE and PCC (Appendix B) and DyNAzyme DNA polymerase (2.2.6.3a).

2.2.7 Cloning of PCR products and DNA sequencing

2.2.7.1 Purification of PCR products

PCR products were purified using a QIAquick PCR Purification Kit (Qiagen). NA fractionated by AGE were recovered using the QIAquick Gel Extraction Kit (Qiagen), and eluted in 30 μl 10 mM Tris-HCl, pH 8.0. For both purification methods, if concentration of the DNA was required, the DNA was eluted in 200 μl EB and precipitated by adding 0.3 M sodium acetate pH 5.2 and 2.5 vol (v/v) ethanol and incubation for 4-16 hours at -20°C. The DNA was recovered by centrifugation at 13000

g, 30 minutes, at 4°C, washed once with 70% ethanol, dried *in vacuo* and resuspended in a minimal amount (5-10 µl) of TE.

2.2.7.2 Ligation of PCR products into plasmid vectors

2.2.7.2a Ligation into pGEM[®]-Teasy vector

PCR products were cloned using the pGEM[®]-Teasy vector cloning system (Promega). Ligations were done at 25°C for 1-4 hours, and an aliquot used for transformation (2.2.7.4a or 2.2.7.4b). The remainder of the ligation mix was incubated at 4°C overnight to complete the ligation reaction and was used for transformation if the number of positive transformants was low.

2.2.7.2b Ligation into pQE-31 bacterial expression vector

Ligation of the cohesive termini of the restricted PCR product and restricted plasmid vector (pQE-31) was done in a 10 µl reaction as described (Sambrook *et al.*, 1989). Vector DNA (1 µl of 600 ng/µl) was added to 6.5 µl of purified, restricted PCR product and the reaction incubated at 45°C for 5 minutes to melt any cohesive termini. To this mixture, T4 DNA ligase buffer, bacteriophage T4 DNA ligase (0.1 Weiss unit) and 1 µl of 5 mM ATP was added and the reaction incubated at 16°C for 16 hours. An aliquot was used for transformation of competent cells (2.2.7.4b).

2.2.7.3 Preparation of competent *E. coli* cells

2.2.7.3a *E. coli* JM109 cells

A single colony of *E. coli* JM109 cells was isolated and grown in 10 ml LB for 16 hours at 37°C with constant vigorous shaking. A 2 ml aliquot was removed to a flask

containing 50 ml of pre-warmed LB, and the *E. coli* grown at 37°C with constant vigorous shaking until an OD₆₀₀ of 0.6 was reached. The cells were cooled and pelleted by centrifugation at 4000 g for 10 minutes at 4°C, then resuspended in 25 ml cold 0.1 M MgCl₂. The cells were pelleted as above and resuspended in 12.5 ml cold 0.1 M CaCl₂, then incubated on ice for 30 minutes. The cells were pelleted again as above and resuspended in 1.6 ml cold 0.1 M CaCl₂ containing 15% glycerol. The cells were distributed in 200 µl aliquots in pre-cooled 1.5 ml tubes, snap frozen in liquid nitrogen and stored at -70°C.

2.2.7.3b *E. coli* M15[pREP4] cells

Competent *E. coli* M15[pREP4] cells were prepared as recommended by the manufacturer (Qiagen). A single colony was inoculated into 10 ml LB containing 25 µg/ml kanamycin and incubated at 37°C with constant vigorous shaking. A 1 ml aliquot from the overnight culture was added to 100 ml pre-warmed LB containing 25 µg/ml kanamycin in a 250 ml flask, and incubated at 37°C with constant vigorous shaking until an OD₆₀₀ of 0.5 was reached. The culture was cooled on ice for 5 minutes then centrifuged at 4000 g for 5 minutes at 4°C to pellet the cells. The cells were resuspended in 30 ml cold TFB1 buffer (Appendix C) and incubated on ice for 90 minutes. The cells were pelleted by centrifugation as above and resuspended in 4 ml cold TFB2 buffer (Appendix C). The cells were distributed in 100 µl aliquots in pre-cooled 1.5 ml tubes, snap frozen in liquid nitrogen and stored at -70°C.

2.2.7.4 Transformation of competent cells and selection of positive clones

2.2.7.4a *E. coli* DH5 α cells

Transformation of competent *E. coli* DH5 α cells with 50ng of ligated vector was done by heat shock (Gibco BRL Life Technologies). A 100 μ l aliquot of the *E. coli* DH5 α cells was thawed on ice and 2 μ l of ligation reaction added, and the mixture incubated on ice for 30 minutes. The tube was transferred to a 37°C waterbath for 90 seconds, then returned to ice and 900 μ l LB was added. The cells were allowed to recover at 37°C for 60 minutes with shaking, and a 50 – 200 μ l aliquot of the cells was plated onto LB-agar plates supplemented with 100 μ g/ml ampicillin, 200 μ g/ml X-gal and 160 μ g/ml IPTG allowing blue-white colour selection of recombinant colonies after growth for 16 hours at 37°C.

2.2.7.4b *E. coli* JM109 cells

Transformation of competent *E. coli* JM109 cells was done by heat shock (Sambrook *et al.*, 1989). An aliquot of the competent *E. coli* JM109 cells was thawed on ice, 2 μ l of ligation reaction was added, and the mixture incubated on ice for 30 minutes. The tube was transferred to a 42°C waterbath for 90 seconds, then returned to ice and 1 ml LB was added. The cells were allowed to recover at 37°C for 60 minutes with shaking, then pelleted gently by centrifugation at 2000 g. The cells were resuspended in 200 μ l LB and plated on LB-agar plates containing 100 μ g/ml ampicillin.

2.2.7.4c *E. coli* M15[pREP4] cells

Transformation of competent *E. coli* M15[pREP4] cells was done as described above for JM109 cells (Sambrook *et al.*, 1989), except that Psi broth was used instead of LB (Appendix C), and the cells were not pelleted by centrifugation prior to plating on LB-agar plates containing 25 µg/ml kanamycin and 100 µg/ml ampicillin.

2.2.7.5 Screening of bacterial colonies containing plasmids

Bacterial colonies containing plasmids were screened either by PCR of purified plasmid (2.2.4) or PCR of single colonies.

2.2.7.5a Amplification of plasmid inserts

Plasmid inserts were screened by PCR with the flanking primers T7 and SP6, or with specific primers within the insert. PCR reactions with a total volume of 20 µL contained 2 µL of purified plasmid DNA (2.2.4; diluted 1 in 1000 with water), *Taq* DNA polymerase reaction buffer, 1.5 mM MgCl₂, 0.2 mM of each dNTP, 0.4 µM of each T7 and SP6 primer (or specific primers) and 0.2 U *Taq* DNA polymerase. The thermal cycling conditions for T7 and SP6 primers were an initial denaturation at 94°C for 3 minutes, followed by 25-30 cycles of 94°C denaturation for 30 seconds, 48°C annealing for 30 seconds, and a 72°C extension for 1 minute per kb to be amplified, with a final extension at 72°C for 5 minutes. PCR products were analysed by AGE (2.2.1.1).

2.2.7.5b Colony screening

A 10 µL PCR reaction containing *Taq* DNA polymerase buffer, 2.5 mM MgCl₂, 0.4 µM each primer (T7 and SP6 or two specific primers), 250 µM of each dNTP and 0.2 U *Taq*

DNA polymerase was set up in a 0.2 ml PCR tube. A single colony was picked using a sterile toothpick and a small amount added to the reaction mix. The thermal cycling conditions were an initial denaturation at 94°C for 7 minutes, followed by 35 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at 60°C, and 1 minute extension per kb to be amplified at 72°C, with a final extension of 10 minutes at 72°C. PCR products were analysed by AGE (2.2.1.1).

2.2.7.6 Automated sequencing

Each clone was sequenced at least twice in both directions (forward and reverse) to confirm the sequence. Automated DNA sequencing of purified plasmid clones and PCR products was done either by Flinders University Sequencing Service, Flinders University (FUSA), or the Nucleic Acid Sequencing Unit, Department of Plant Science, Adelaide University. Primers T7 and SP6 used for sequencing of plasmid clones (Appendix B) were supplied by the sequencing service. SCSMaV primers for sequencing of PCR product were supplied to the sequencing service (Appendix B).

2.2.7.7 Sequence analysis

Initial sequence editing was done using the program SeqEd[®] Version 1.0.3. This program was used to remove the vector and primer sequences and to do simple overlap analysis of clones. Most subsequent sequence analysis was done using programs of the GCG Sequence Analysis Software Package version 8 (Genetics Computer Group, Madison, WI, 1984; Devereux *et al.*, 1984) through the webANGIS interface (<http://www.angis.org.au>).

2.2.8 Detection of homologous nucleic acids with radioactively labelled probes

2.2.8.1 Capillary blot

Agarose gels were transferred into a container covered with Southern Solution I – Denaturing Solution (Appendix C) and incubated with gentle shaking for 30 minutes at 25°C. The solution was removed and replaced with Southern Solution II – Neutralising Solution (Appendix C) and incubated at 25°C with gentle shaking for 30 minutes, replacing the solution once after 15 minutes. A transfer stack was made by placing (in order) in a tray 2-3 layers of 10 x SSC soaked sponge, 2 layers of 10 x SSC soaked Whatman[®] No 1 filter paper, the denatured and neutralised gel and water soaked Zetaprobe[®] membrane cut to the exact size of the gel. A shield to ensure that capillary action was only through the gel/membrane, was made from a plastic sheet by cutting a hole in it that was slightly smaller than the gel/membrane. This was placed onto the membrane and the stack completed by adding 3 layers of 10 x SSC soaked Whatman[®] No 1 filter paper and ca. 15 cm paper stack, with a weight on top. Bubbles between each wet layer were removed by rolling with a glass rod. The tray was then filled with 10 x SSC. The transfer was done for 16 hrs at 25°C, after which the stack was disassembled and the nucleic acid fixed on the membrane by crosslinking in a Bio-rad[®] UV crosslinker at 125 mJoule.

2.2.8.2 Nucleic acid dot-blot

TNA extracts were dotted directly onto ZetaProbe[®] nylon membranes, dried and crosslinked as above.

2.2.8.3 Tissue squash

Sugarcane leaf tissue was squashed directly onto dry Magna[®] nylon membrane (Geneworks) as described in Chapter 7 and crosslinked as described above.

2.2.8.4 Radiolabelling of DNA probe with ³²P

The "cons7" probe was prepared by RT-PCR amplification of an 850 nt fragment of the SCSMaV RNA using primer L6 to prime reverse transcription with AMV Reverse Transcriptase (2.2.6.2b) and PCR with primers U1 and L3 using DyNAzyme DNA polymerase with reaction conditions as described (2.2.6.3a). Routinely, an aliquot (5 µl) of the PCR product was analysed by AGE (2.2.1.1) and its concentration estimated by comparison with the 1 kb Plus DNA ladder. The remainder of the PCR product was purified using the QIAquick PCR purification kit. 50-200 ng of purified DNA was labelled by primer extension of random oligonucleotides (Hodgson and Fisk, 1987). Unincorporated nucleotides were removed with the QIAquick PCR purification kit, and the labelled probe was eluted in 200 µl 10 mM Tris-HCl, pH 7.2.

2.2.8.5 Hybridisation of nylon filters

Hybridisation was done using a modification of the ZetaProbe[®] membrane protocol. The membranes were placed into a hybridisation bottle and covered with prehybridisation solution containing denatured herring or salmon sperm DNA (Appendix C). This was allowed to pre-hybridise at the hybridisation temperature (65°C for hybridisation without formamide, 42°C for hybridisation with formamide) for a minimum of 1 hour to overnight. The probe was prepared (2.2.8.4) and denatured by boiling for 5 minutes, then chilled and added to the hybridisation bottle. Hybridisation of the membrane was

done for 16 hours at 65°C for hybridisation not containing formamide, and 42°C for hybridisation containing 50% formamide.

After hybridisation the membranes were washed sequentially in ZetaProbe[®] Wash Solutions 1, 2 and 3 (Appendix C) which have increasing stringency. Washes were done at the hybridisation temperature for 30 minutes each, checking the membrane for radioactivity using a Geiger counter after each wash. When the signal appeared to be no longer non-specific the washes were ceased.

2.2.8.6 Autoradiography

Radioactive nylon membranes were covered in plastic wrap and exposed to X-ray film (Kodak) from 2 hours to 6 days with an intensifying screen at -70°C and developed with an X-ray film developer CP1000 (AGFA, Gevaert, Germany).

2.2.8.7 Stripping of nylon membranes

Zetaprobe[®] membranes were stripped of radioactively labelled probe by the addition of boiling 0.1 x SSC, 1% (w/v) SDS, and mixing until cooled to room temperature. They were assessed for residual radioactivity using a Geiger counter and exposure to film for 16 hours (2.2.8.6) and stripping was repeated if signals were observed on film. The stripped filters were covered in plastic wrap and stored at 4°C prior to rehybridisation.

2.2.9 Virus preparation

SCSMaV was prepared using a modification of Foster (1998). SCSMD affected sugarcane leaf tissue (1 kg) was homogenised in 2 volumes (w/v) Virus Extraction Buffer (Appendix C) in a 5 L Waring blender at 4°C. The extract was filtered through cheesecloth then stirred with 1 volume (v/v) chloroform for 90 minutes at 4°C. The

chloroform phase was separated by allowing the mixture to stand, and the aqueous phase transferred to a new beaker. PEG 8000 was added to 6% w/v and the slurry stirred overnight (16 hours) at 4°C. The precipitate was collected by centrifugation at 10000 g for 30 minutes, and resuspended for 16 hours in 200 ml 0.5 M borate buffer (pH 7.8) containing 0.5% Triton X-100 at 4°C. The extract was clarified by centrifugation at 10000 g for 30 minutes, and the supernatant retained. Virus particles were sedimented by high-speed centrifugation at 30000 g for 90 minutes in a Beckman 60-Ti rotor, and the supernatant discarded. The pellet was resuspended in 100 ml buffer (50 mM sodium borate buffer with 1% Triton X-100) using a glass homogeniser. The low speed clarification and high speed sedimentation was repeated, and the resulting pellet resuspended in 10 ml 0.05 M borate buffer (pH 7.8) using a glass homogeniser as above. The virus was further purified by isopycnic centrifugation in caesium sulphate at 1.326 g/ml and centrifuged at 110000 g in a Beckman SW-65Ti rotor at 4°C for 20 hours. The virus band was observed by light scattering and fractions collected using a 18 gauge needle. The virus was pelleted by centrifugation at 45000 g in a Beckman 65Ti fixed angle rotor at 4°C for 90 minutes, and resuspended in 0.5 ml sterile water. Electron microscope grids were prepared (2.2.10), and the remainder of the pellet stored in 50% glycerol at -20°C.

SCSMaV particles were further fractionated by Nycodenz[®] isopycnic density gradient centrifugation using the method of Randles and Hanold (1989). A Nycodenz[®] gradient was prepared by layering 0.6 ml of 30, 40, 50, and 60% Nycodenz[®] in a 3ml polycarbonate Ti-100 Beckman centrifuge tube. Glycerol in stored virus preparations was diluted 1:3 (v:v) with water and layered on top of the gradient. The preparation was fractionated by centrifugation in the Beckman TLA100 rotor at 103000 g (45000 rpm)

for 15 hours at 5°C. The virus was visualised by light scattering and fractions were collected using an 18 gauge needle as described above. Electron microscope grids were prepared (2.2.10) and the remaining preparation stored in 50% glycerol at -20°C.

2.2.10 Preparation of electron microscope grids for transmission electron microscopy (TEM)

400 mesh copper grids were coated with formvar and carbon (Hayat, 1989), and were made hydrophilic using a spark tester *in vacuo* (Edwards, England). Droplets (ca 10 µl) of the virus preparation samples and 1% PTA (Appendix C) were placed on Parafilm®. The electrostatic grids were placed onto the sample for 1-5 minutes, and excess sample removed by draining carefully with filter paper placed against the edge of the grid. The grid was then placed onto the 1% PTA stain for 10 seconds, and the excess drained. The grids were air-dried and observed using a Philips CM100 Transmission Electron Microscope at the Centre for Electron Microscopy and Microstructure Analysis (CEMMSA, Adelaide University, Australia).

CHAPTER 3

**SEQUENCING THE SCSMaV
RNA GENOME**

CHAPTER 3: SEQUENCING THE SCSMaV-RNA GENOME

3.1 Introduction

Determining the complete nucleotide sequence of the SCSMaV-RNA is essential to describe its genome organisation, taxonomy and to develop nucleic acid based diagnostic tests for SCSMD. A random PCR cDNA library from purified disease associated dsRNA was previously generated and approximately 28% of the genome was sequenced and gave three segments (Choi, 1997; Choi *et al.*, 1999). Nucleotide sequence analysis of the segments suggested that SCSMaV was most closely related to *Apple stem pitting virus* (ASPV), type member of the genus *Foveavirus* (Choi *et al.*, 1999). Alignment of the SCSMaV-RNA segments to the nucleotide sequence of ASPV gave the relative positions of the segments (Figure 3.1). This chapter describes the determination and nucleotide sequence of SCSMaV-RNA.

3.2 Methods

3.2.1 Isolation of dsRNA and TNA

The dsRNA used as a template for cDNA synthesis was isolated from SCSMD sugarcane leaf tissue (2.2.3). TNA for 5' and 3' RACE was isolated from the same source of SCSMD sugarcane leaf tissue (2.2.2.1).

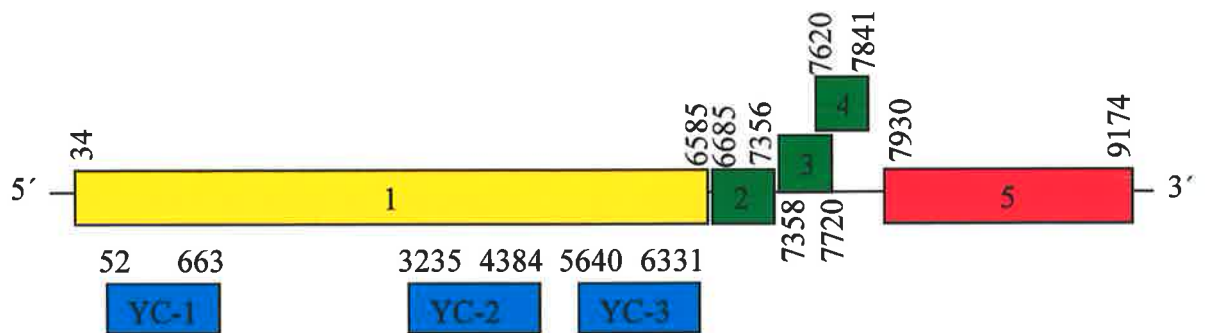


Figure 3.1

Alignment of the previously sequenced regions of SCSMaV (Choi, 1997; Choi *et al.*, 1999) to the ASPV genome. The nucleotide positions of the ASPV ORFs are shown on the diagram. ORF 1 (yellow) encodes the RdRP, ORF 2, 3, 4 (green) encode the TGB, and ORF 5 (red) encodes the CP. The three SCSMaV regions YC-1, YC-2 and YC-3 (blue) are shown beneath the ASPV genome at the positions of the greatest sequence similarity.

3.2.2 Sequencing of existing clones

Glycerol stocks of the existing rPCR clones in *E.coli* DH5 α cells (Choi, 1997; Choi *et al.*, 1999; 2.1.3) were streaked on LBA plates containing 50 μ g/ml ampicillin and single colonies grown overnight in 5 ml LB containing 50 μ g/ml ampicillin at 37°C. Plasmids were prepared (2.2.4), restricted with *Eco*RI (2.2.5) and the restriction fragments analysed by 1.5 % AGE (2.2.1.1) to check that the size of the inserts was the same as previously described (Choi, 1997). Plasmid clones were sequenced by the Nucleic Acid sequencing service (2.2.7.6) and the sequence analysed (2.2.7.7).

3.2.3 Designing of oligonucleotide primers

Oligonucleotide primers were designed (2.2.6.1) and a list of primers used for sequencing SCSMaV RNA can be found in Appendix B.

3.2.4 RT-PCR reactions

3.2.4.1 Bridging of gaps using known sequence

Bridging the gaps between previously sequenced regions (Choi, 1997) was done using RT-PCR with two SCSMaV specific primers to amplify the region in between (Figure 3.1). Reverse transcription was done with Superscript II Reverse Transcriptase (2.2.6.2a) and PCR was done using DyNAzyme DNA polymerase (2.2.6.3a). The primers used for RT-PCR amplification of the YC1-YC2 gap were U2 and L1 and for amplification of the YC2-YC3 gap were U1 and L3. Either primer could be used to prime reverse transcription because the template was dsRNA. RT-PCR products were analysed by AGE (2.2.1.1).

3.2.4.2 Extension of the RNA sequence using anchored

RT-PCR

Sequential clones were generated using a modification of the methods of Froussard (1992) and Fazeli and Rezaian (1999). Primers internal to the known sequence were used to synthesise the first strand cDNA by reverse transcription with Superscript II Reverse Transcriptase (2.2.6.2a).

The second strand reaction was initiated at random sites within the genome by the PCC-RH primer (0.15 µg). The primer and first strand reaction were denatured at 100°C for 3 minutes and the second strand synthesised using Klenow DNA polymerase (Promega). The reaction conditions were 50 mM Tris-HCl pH 7.4, 10 mM MgSO₄, 0.1 mM DTT, 0.3 mM each dNTP and 5-10 U Klenow fragment DNA polymerase, and the reaction was incubated at 37°C for 30-60 minutes. Excess PCC-RH primer was removed using Qiagen PCR purification kit (2.2.7.1).

PCR was done with a 2 µl aliquot of the second strand reaction and DyNAzyme DNA polymerase (2.2.6.3a) using 0.4 µM PCC primer and 0.4 µM of the primer used to initiate reverse transcription. The thermal cycling was done using an initial denaturation at 94°C for 5 minutes, followed by 30 cycles of 1 minute denaturation at 94°C, annealing at 60°C for 1 minute, and extension at 72°C for 3 minutes, with a final extension for 10 minutes at 72°C. PCR products were analysed by 1% AGE (2.2.1.1), and fragments greater than 500 nt were excised from the gel with a scalpel blade, and extracted from the gel (2.2.7.1).

3.2.4.3 Extension into the coat protein region

A degenerate potexvirus-coat protein primer (PXCP) was used to amplify regions towards the 3' end of the SCSMaV-RNA. dsRNA was reverse transcribed using SuperScript II Reverse Transcriptase and primer SCSMaV-U11 (2.2.6.2a). PCR was done using DyNAzyme DNA polymerase with 0.4 μ M PXCP and 0.4 μ M SCSMaV-U11 primers (2.2.6.3a). RT-PCR products were analysed by AGE (2.2.1.1).

3.2.4.4 Rapid amplification of cDNA ends (RACE)

The 5' and 3' ends of SCSMaV were amplified using RACE on TNA extracts as described in 2.2.6.4. RACE products were analysed by AGE (2.2.1.1).

3.2.5 Production and selection of clones for sequence analysis

RT-PCR products were purified (2.2.7.1), cloned into pGEM[®]-Teasy vector (2.2.7.2a), transformed into *E. coli* DH5 α (2.2.7.4a). Positive colonies had their plasmids purified (2.2.4), screened (2.2.7.5a), sequenced by FUSA (2.2.7.6) and the sequence edited and analysed (2.2.7.7).

Existing rPCR clones (Choi, 1997; Choi *et al.*, 1999) were plasmid purified (2.2.4) and restricted to assess the size of the insert (2.2.5). The plasmids with insert sizes that were the same as Choi (1997) were sequenced (2.2.7.6) and their nucleotide sequence edited (2.2.7.7). Overlapping regions were identified and the sequence compiled sequentially.

RT-PCR clones were generated step-wise with the sequence extended as more clones were generated. The sequence of the new clone was compared to that of known sequence using SeqEd[®] (2.2.7.7). The 100 nucleotide overlap produced by the nesting of the reverse transcription primer, was used to confirm that clones obtained were

extensions of the known sequence and each nucleotide was sequenced at least twice in both directions.

3.3 Results

3.3.1 Sequencing of existing clones

The majority of the existing rPCR clones confirmed the known sequence but did not extend it. One clone extended the known sequence by approximately 200 nt, but it did not overlap with other regions. This gave three regions (YC-1, YC-2 and YC-3) which were estimated to comprise 25-30% of the genome sequence. These regions were aligned by maximising regions of homology with the complete ASPV nucleotide sequence (Figure 3.1).

3.3.2 Amplification by RT-PCR

3.3.2.1 Bridging of gaps

RT-PCR was done to complete the sequencing between the known regions (Figure 3.1). The gap between YC-1 and YC-2 was amplified by PCR and gave a fragment of 2.1 kb, and the gap between YC-2 and YC-3 gave a fragment of 0.85 kb (Figure 3.2).

3.3.2.2 Sequence extension using anchored RT-PCR

The extension of the known sequence using an anchored RT-PCR strategy (3.2.4.2) generally gave a ladder of bands (Figure 3.3). The arrows indicate the range of bands that were removed from the agarose and cloned. The largest fragment cloned using this method was ca. 600 bp.

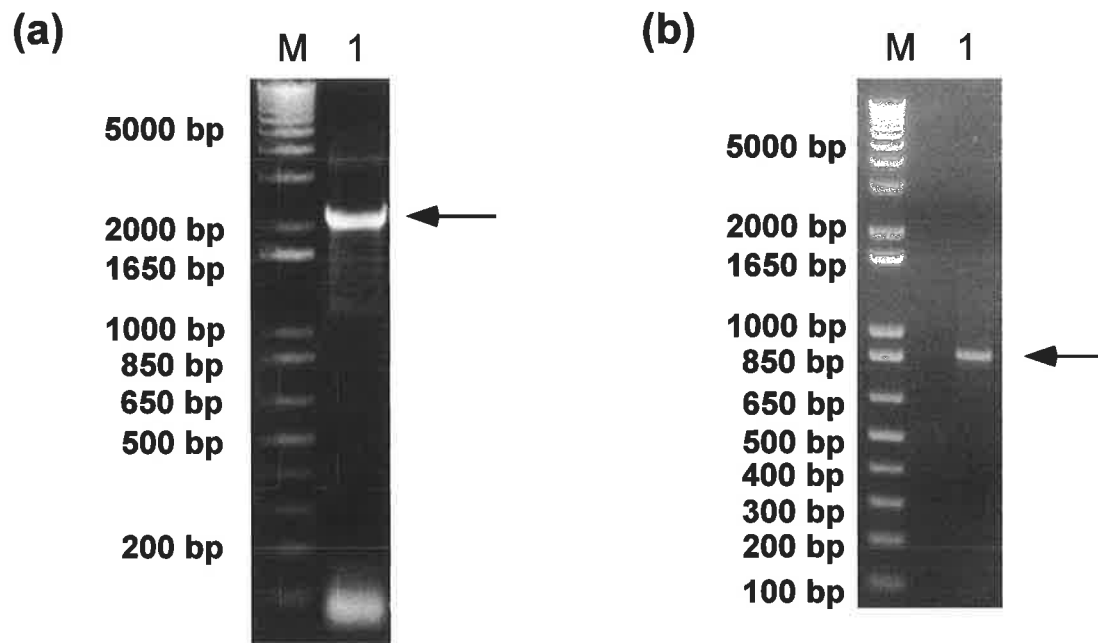


Figure 3.2

Analysis by 1.5% AGE of RT-PCR products spanning the gaps between existing regions of known sequence.

(a) The gap between YC-1 and YC-2 was amplified using primers U2 and L1 to generate a fragment of ca. 2.1 kb.(shown by the arrow)

(b) The gap between YC-3 and YC-2 was amplified using primers U1 and L3 to generate a fragment of ca. 0.8 kb.(shown by the arrow)

M - 1kb Plus DNA ladder with sizes shown adjacent

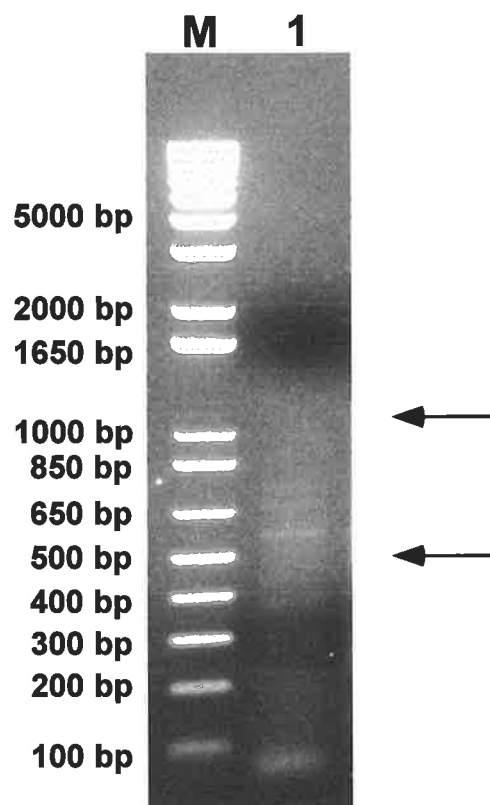


Figure 3.3

Analysis by 1% AGE of anchored RT-PCR amplification products from SCSMaV dsRNA. The arrows indicate the region of the gel that was excised from lane 1 for cloning reactions.

M-1kb Plus DNA ladder used as a standard DNA marker.

1- Ladder of bands resulting from the anchored RT-PCR amplification.

3.3.2.3 Extension using PXCP primer

RT-PCR with the PXCP primer and SCSMaV-U11 primer produced an ca. 1100 nt fragment (3.2.4.3; Figure 3.4), that had its sequence determined through 2-3 rounds of sequencing in forward and reverse directions.

3.3.2.4 5' and 3' RACE

The results from 5' RACE gave a product of ca. 150 nt (Figure 3.5a). The results from 3' RACE gave a PCR fragment of either 1650 nt for the U14 primer or ca. 1200 nt for the U20 primer (Figure 3.5b). The products were cloned, and clones containing the predicted insert size were sequenced.

3.3.3 Sequence assembly

The sequence was assembled sequentially with the program SeqEd[®] (3.2.5) using the 100 nt overlap between clones to orient and align the sequences. A schematic diagram showing the relative amounts of sequence gained from each method is shown in Figure 3.6. Towards the completion of the sequencing, Frames 2.3 (<http://www.nih.go.jp/jun/cgi-bin/frameplot.pl>) was used identify potential ORFs. Regions that showed a potential frame shift within an ORF were re-amplified by RT-PCR, cloned, and sequenced to confirm the accuracy of the original sequencing.

The sequence of the SCSMaV-RNA is shown in Figure 3.7. The SCSMaV-RNA genome comprised 8146 nucleotides and was polyadenylated at the 3' end. The nucleotide sequence of SCSMaV has the GenBank accession number AF315308.

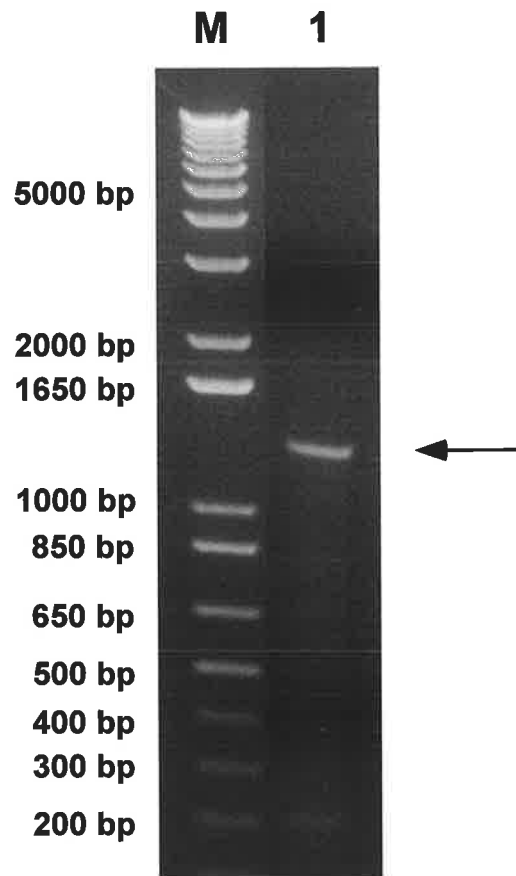


Figure 3.4

Analysis by 1% AGE of RT-PCR using the potexvirus degenerate primer (PXCP) and a SCSMaV specific primer (U11). The arrow indicates the ca 1100 bp fragment that was cloned.

M- 1kb Plus DNA ladder.

1- RT-PCR fragment from the PXCP-U11 reaction.

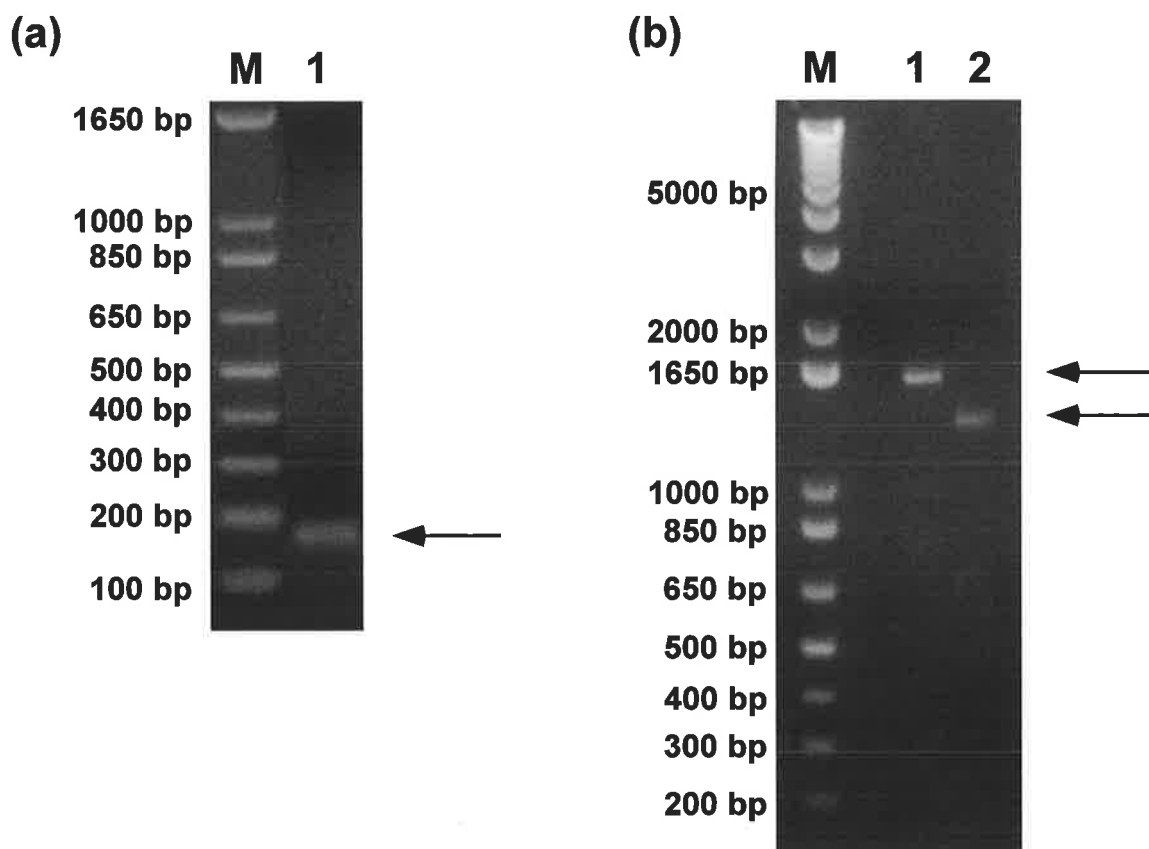


Figure 3.5

(a) Analysis by 1.5% AGE of the 5' RACE product from TNA extracts of diseased sugarcane. The arrow indicates the ca 150 bp fragment that was cloned.

M- 1kb Plus DNA ladder.

1- 5' RACE product that used SCSMaV-L19 and SCSMaV-5GSP1 primers.

(b) Analysis by 1% AGE of 3' RACE products from TNA extracts of diseased sugarcane.

The arrows indicate the ca 1500 bp and 1200 bp fragments that were clones.

M- 1kb Plus DNA ladder.

1- 3' RACE product using SCSMaV-U14 and PCC primers.

2- 3' RACE product using SCSMaV-U20 and PCC primers.

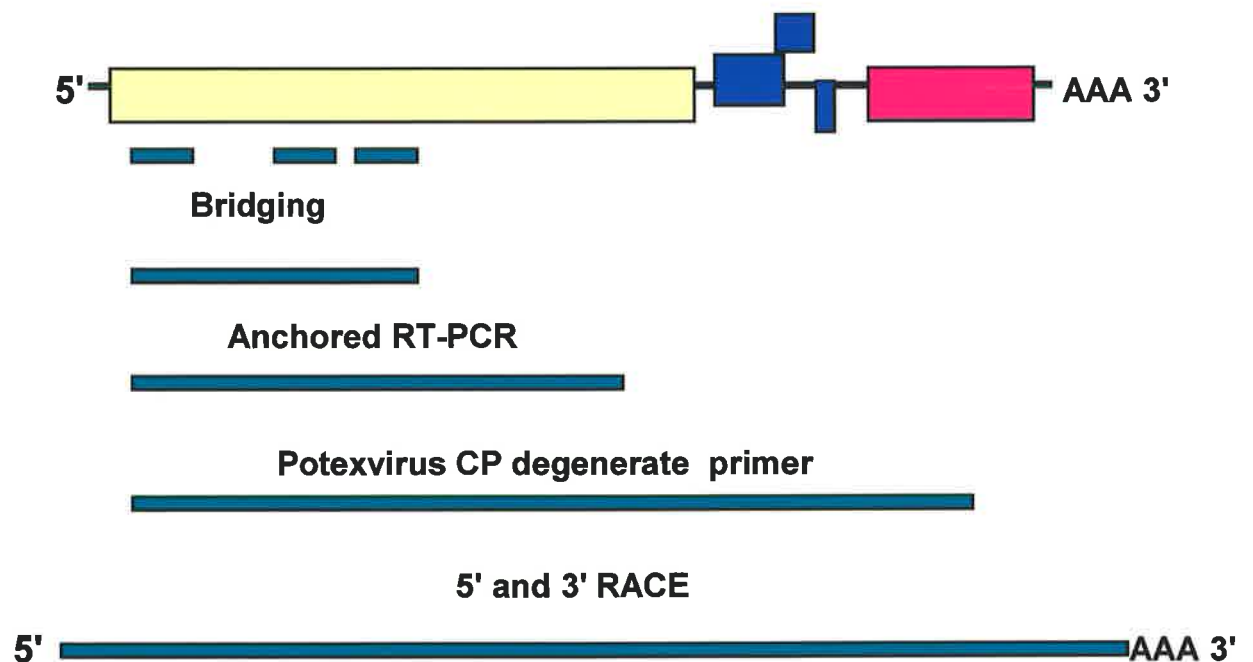


Figure 3.6

Schematic representation of the methods used for the sequencing of SCSMaV RNA and the relative amounts of sequence obtained by each method.

Figure 3.7

The nucleotide sequence of SCSMaV-RNA comprising 8146 nt.

(A)_n indicates the 3' terminus polyadenylation.

1	UGAAAACAAA	ACUACACAAU	UGAGCUCAAA	CUGCUACAAU	AUCCACUCAA
51	UAUACCUUAU	UGAUUCAUAU	UGAUGGCAAU	UCUCACCUUC	AAACCCCUA
101	AGCAGCUCUU	AAUCGCUCAA	UUGUCCUCUG	AAGAGCAGGC	CAGAAUUAGG
151	AGCUCUGUGA	CACGAAAGUU	AGCGCAGGAG	GAAGAGAAUG	AUCAAGGUUU
201	AUUCAGUUAU	CAUGUCUCUC	CAUACUGUGC	AGAGAAGUUG	AUCAAGUCUG
251	GUGUUCAACU	CUGUGCCUAC	UCCUCUGAUG	UACAUAGUCA	UCCUUGUUGU
301	AAAAUGCUAG	AAAAUCAUAU	UUUAUACAGA	GUUCUCCUU	CUUAUGUAGA
351	UUCAGAUUUU	AUCCUUGUAG	GUAUUAAGCA	UUCUAAGUUA	GCCUUUCUUA
401	AACCAAGAGG	CGCGAAAAAG	AAUGUUGAUC	UUUCAACUGU	UAAGGUGCUC
451	AAUCGUUUUG	UCACAAGUCU	UGACAGAUCC	AGGUAUGGUA	AGAGUUUAA
501	UACCUCUAAU	AUCAUGUCA	CAUUCACUCC	UGAACCUAAG	UGAAGCAGUU
551	CCUAUACCCC	UGAACUUAAA	AGCCUAAUUC	CAAAUCUGGA	GCGAUUUAA
601	GGCCACAAAC	UUUUUUUCCA	UGAUGAGUUG	CAUUACUGGA	AGGAUGAAUC
651	UAUUUUUAAA	AUGCUAUCAU	CUUUAGAGCC	UUCAGUAGUG	UACGCAACUG
701	UUGUAUGCCC	ACCCGAAUUG	CUUAUUGGUG	CUAAGCAGAG	UUUAUAUGAU
751	UGGCUAUACA	CUUUUGAAAU	CAUUGGGGAU	CAAAUCUUAU	UUGCUCUAGA
801	CGGAAAGAUG	AAUAGUGCCU	ACUUCCAACC	UGUCAAUUGU	GCCCGCCUUC
851	UAAGAGCAAA	AAGGAUAUGG	CUGGAUAAGC	AUAACUAUUA	UGAUUAUGAU
901	CUGGUCUAUU	CAAAUUUCUC	GCAUCAUUG	CUCUGCUUCU	CAAAGGGCUC
951	AAAUUUUAA	AAUCAAGAUC	UUAGGGCUUU	CUCUGACUUU	GACGCAAUUA
1001	AAUUAAAUUA	CAUAAGUGAG	CUCAUGGUGA	AUGCUGAGAG	UUGC UUCCA
1051	AUUAGUUGGA	CUGUUGUAAA	AGGGCUUUUA	GAUUUUCUUG	AGAGCUUGAC
1101	AAAAGAGGAU	UACAGAUCAU	GUAAGGCUAA	GCAAACUCAU	UAUAUGCCCA
1151	AUUGCUCAGG	GAAACAACUU	AAAUUUUGCA	CUGACCUUAU	CAAACUUUAU
1201	CUUGAAGCUC	CACAUUUUAA	AACUAUGAUA	GGUUCUAGCC	CAGUUCACAA
1251	CUUUAAAGGUC	UUGUUUACAC	GGAUGUUCU	UGC UUUGCAU	AAGAAAAAUU
1301	UCUCUUUUUA	UAAAGGAAAA	UUGUUGAAAA	GUUUUAUUUA	ACAUCUAGCU
1351	CCUUUGGCUU	UCCAUGUCA	AACUUAUGAU	CUUCAUUUAG	GUUUUGUUUU
1401	CAAAUUCAG	GAGCUGAAAA	AUGUGACGGA	UCUAAUGGAG	GAUGUGCAG
1451	AAUUUGAAGU	UGACUUUGUC	AGAGAAUAUG	ACCCAUUCCA	GAUGCGCAGC
1501	ACUCCUUAUG	UUGGUCUUGC	CCCUUUGCGU	GAUGAAUCAU	AUUUAUAGAG
1551	GAUGAAUUGU	GGUGAGGCUU	UCUUUGACAA	CAUGAUUGGG	CAAAUUAUGA
1601	GAUCCCAUUU	AAGCGGGUAU	GGCACAAAUA	CUAAUCCUGC	UGAACUUUUC
1651	CUAAGAAUCA	AUUUGACCAU	UGUUGGAAGG	GUUGAGUUUC	UUGUGAAUGG
1701	CCAUCAUGAU	UUUGUAAUUU	UUUAUAUCAA	CAAACAUUUU	AAGGUAAAGA
1751	AGGACCCCAG	GUACAAAAAC	UUGUACUUUG	CAAAUAUGAC	CCUCGAAUUG
1801	AAAUUGAGGU	GGCUUUCGUG	CAGACGUUUU	AGAAAGUAUC	AGAAGUUCUU
1851	ACCCUGCGGU	CUGGUGAGUC	UUGCUGAAGU	ACAAAUUCUU	UGGAGAAUA
1901	UUGCGAGAAG	CAUCCAGAGU	GGAUCACAUG	AGGCUCAAGG	ACAAAUGAA
1951	CAAGUUGAUU	CCCAUGGGGA	UGGAGUAGCA	AUUGAGGAGC	CAGAAAGGAA
2001	GGGGAAUACU	UUGGAAAAAU	UAAAUGAAGC	UACUUUAGUA	ACCCACCUCC
2051	AGCACUACCU	CAAGAAUCCU	GUUGUUGAUG	AUUUUUCAAG	AUGCGCCUUU
2101	GAUGUCGACU	CAAAGACUCU	AUCUAAUCUA	AUUGAUUGUC	UUAAACAGAU
2151	UUGUGAUACU	AGCAUAAAUG	UUGCCAUGG	AUGUGACUGU	UCAAUUUCAG
2201	AUCUAAAUUU	ACUCAAGAUU	AGUAAAAUGG	CCUCCUCUGU	GGCAUCAUAU
2251	GUUAAUCUGG	ACAAUGUCA	AUCUGUGAAA	AUCAAUUGGA	CCGAAGCUGG
2301	UGAGAUCUUA	GUCGGUUUAG	AGGAUCAGCC	UCUUUUUGUC	GUGGUUCUUG
2351	GGAAGGAUUA	UAAUUCAAUU	ACAUAAGCAA	GGAUCAUGU	UAUGCAAUUU
2401	CCUGAGCUUG	AAGAAAAUUC	AGCAAUUUUA	AUUGAAAGGA	AUAACAUCUC
2451	UAAAAUUUCU	GUAUCUGGCC	CAUCUGCUCU	CAUGAUCUUU	CCAAAAGAUU
2501	AGAAUUUGCU	GGAUUUGCAU	UUAAUUGGCU	GCGAUCAAUU	CCCCCUGCA
2551	AAGAUUGAUU	UCAUAGAUGG	GUACAUCAAA	ACUACGGUGG	AUCCUAGUCU
2601	UGAUUCUCGG	UUAACAUGCA	UCAAUGUACC	CGCAAUUGGU	GAUUGUUUUU

Figure 3.7: Nucleotide sequence of SCSMaV-RNA

2651	GGCAUAGUGU	AUCUCUUUAU	CUUAGUGUUG	AGGCCAAAUA	GAUCAAAGAU
2701	GCACCGUUC	UAAGAAACAG	GAAUCCCCAA	AAUGCCAGAU	UAGUUGAACC
2751	AAUGGGUGAU	AAAGUUUGGG	CUGAAAAUGA	GGCUAUACUA	AACUGUGCAU
2801	GUUUUUACAA	CUUGAAUAUA	CGUAUAGUCU	UGAACAUGA	AUUGCAUACC
2851	UCCGUCUCCC	AUGAUUUGAA	UGAAGAUUCA	ACUGAAGUUU	GGCUCAAAAA
2901	UUUGGAUCAU	AUGCAUUUUU	GUCCAGCAAA	GUUUUAUAAU	UCUUGUGUCA
2951	UGGAGCCAAU	UUCUAAAGCA	GUUGACAUUC	CGCCUUUGAA	UGUACAAAGA
3001	AUGAUUCUU	CACAGGAGAG	UCUUCACUCC	GAGUAUGAGA	GAUUGAGCAG
3051	UGGGAAAUCC	UUUGAUCUCC	UUGGACUGGA	GAGUUUGAUG	ACUUUCUUCA
3101	ACAUAGGGGC	UGAAGUCCAG	GAAAAUGGUC	ACAAGUUUAU	UUUGAACGAG
3151	AAUGGCAGUA	GGGCUCGAAG	CUUCAUUCUG	GAGAAUAAUC	AUAUUGAAUU
3201	UGUUCGUGGA	GAUAAAAAGG	AUCUAAGAAA	GGAGGGAAUU	GGGUUUGUUA
3251	AGCUAAUGAA	UGCAAAUCAA	AGUAAGUCUG	UGCUGUCUCU	CGCCUGCUCA
3301	AAGGUUAAAU	ACACCCCAUC	UUUAGACAGA	GCCUUUAAAU	UGUCAAAGC
3351	UUUUGCUAUU	GGUUUAUCAG	GCAUAAUGCU	UAGCAGUGAA	AAAUUUGGCG
3401	AGAUUGAAAU	CAACGGUAAG	AAUGACGAGA	GGGAAAUCAA	UGUUAUGAUG
3451	GGAACAUUUG	GGGCUGGAAA	GACUAGAGUA	UUUGUGGAUU	AUAUUCAGGC
3501	UUACUCUGGU	AGAGGUUUUAU	UUUAUGUUAG	CCCAAGAAAG	AACCUUAAGG
3551	AGCUUUUCGA	UAACUCUGCC	AUCCAGGCAG	UUCAAGGCUU	UAAAACCAAG
3601	UCAAGUGACA	AGAAAGGCAG	UGUAAAAGAGU	UUUCAUAGCU	CUUUCACCUU
3651	UGAGACUGCU	AUAAUGAAAAG	CCGGCCAGAU	UGGUGAUGAU	UCAAUUCUUA
3701	UUAUAGACGA	GAUUCAGCUC	UAUCCGCCUG	GCUACCUGA	UUUAAUCCUU
3751	CUAUUGAUAC	CUUCAAGUUG	CAUGGUAAUU	UGUGCUGGUG	ACCCUGUUCA
3801	GUCUGACUAU	GAUUCAGAUG	GUGACCGUUU	GUAUUUUCUU	AAUGAUAAAC
3851	CAGAUUUUGA	GCGGUUGUUA	GAGGGUCAAU	CUUAUAAGUA	UGCAAUCAA
3901	UCAAGAAGGU	UUAAAUCCUC	CUUAUUCAUU	GGACGACUAC	CUUGUGAUUU
3951	GCCCGGUCUU	CAAGGGAUGG	GCAUGGACAC	UUUUGGGAUU	UAUGGAGACA
4001	UUAGCAGUGU	GCCUCAGGAU	UUAAUUAAGA	GAUCCUUCUU	UUUGGUGAGU
4051	AGUUUUUGAUG	AGAAGAAAAU	UAUUAAAAGC	AAUUUUAAAAG	AGACUCAUAA
4101	CAAGAUUUUG	ACUUUUUGGGG	AAUCCACAGG	AUCAACUUUU	CAAGAAAUUG
4151	UCAUUCUCAU	AACUACUAGU	UCAGAAAAGG	CAAGUGAGAG	GAGAUGGGUA
4201	ACAGCAUUA	GCAGAGCAUC	GACAAGCAUA	UCUUUUUCUA	ACCUGACUGG
4251	CAUUGGAAUA	CAAGAGCUUA	UAUCGACUAC	UUUUUUUGGC	AGAUUCUUGC
4301	AAAAAUUUUU	AAGUGCUGAA	UGCAAGAGGG	AGGAUAUUA	GCCAUAUUUA
4351	UACGGAGAGC	CAAAAUUUGU	UGAUGACUUC	AGGAGCAUAG	GAAGACGAAG
4401	UGCAGAGGAG	AAAGAAGUCA	AAGUUGAUGG	UGAUCCGUGG	UUAAAGUGCA
4451	UCUUGCAUUU	CUUUGAAACU	CCAAUCCAG	AUGAAGUUGA	GUGUGUUAAA
4501	GUUGAUUUGG	UAGAGCAAAA	AGCAAAAACU	CACUUACCGG	UUUCGUCCUU
4551	UGAGGUUUUA	UGGGCCCAUU	UUAAUGAAAA	ACUUUUGCUU	AAGGAGAUCA
4601	GGGAGCGGAG	GAUAGAUGAU	CUUGUUUCUG	ACCAGUUUGC	AUGGGAUCAU
4651	AGGUCACAAG	GACUAGCAAU	UAAUAAUGCU	GCCACAAGGU	AUGAAUCAAU
4701	AUUCCCUAGG	CAUACCACA	AUGAUAAAGU	GACCUUUUUG	AUGGCAGUUA
4751	AGAAAAGGUU	GACCUUUGGA	UCGCCCGCAA	AGAAUGGGAG	GAUGCUUAGA
4801	GAGGCUAUGC	CCUAUGGUGA	AAUAUACUC	AAGGAGUUUC	UCAAGAGGGU
4851	CCCUAUUAAG	AAUGAUAGAA	ACGAUCAUUU	AUUUUUAGAA	UGUCUAAUGG
4901	AUUUUGAAGA	AAAAAAAACU	UCCAAGGACU	CAGCCACAAU	CGCAAAUAC
4951	UGUAUACGCU	CACAGAAUUC	AUGGCCAAUU	GAUAUUGCUG	AGUAUAUCAU
5001	GAAGUCACAA	CUCUGCACAA	AAAUGGAGAG	UAUGUUCUGC	GAGGCAAAGG
5051	CUGGUCAGAC	UUUAGCUUGU	UCCAUCAUG	UAGUUCUUGC	ACGAUUUUCU
5101	CCAUGGGUCC	GUUACAUAGA	AAAGAAGCUG	GUCAGUUCAC	UGCCAGAUAA
5151	UUACUAUAUA	CACUCUGGAA	AGAACUUUGA	UCAACUCAAU	GCUUGGGUUA
5201	AGGCAAACAA	UUUUGUGGGU	GAAUGCACUG	AAUCGGAUUA	UGAGGCCUUU
5251	GAUGCAAGUC	AAGAUCACUA	UAUACUGGCC	UUUGAAUUAG	CAUUAUGAA
5301	GUACCUUGGU	UUACCCAAUG	AGUUGAUAGA	GGACUACAAG	UACAUUAAGA
5351	UUCAUUUGGG	UUCAAAAUUG	GGUGAUUUUG	CCAUAAUGAG	AUUCACCGGU
5401	GAAGCCAGCA	CUUUCUUAUU	UAACACAAUG	GCAAUAUUGC	UUUUCACUUU
5451	UCUGGUCUAU	GAUAUCAAAAG	GCAUUGAGUC	AAUUUGCUUU	GCUGGCGAUG
5501	ACAUGUAUGC	AAAUACAAGU	CUUAAGACAU	GCAGCAAGAA	CAAAAGUUUG
5551	CUUAAAAAGU	UUACACUCAA	AGCUAAAGUU	CAGGUAACUG	AAAGCCCAAC
5601	AUUUUGCGGA	UGGCACUUGA	CAGAGGAUGG	AAUUUAUAAG	GACCCAAUAC
5651	UUGUGUCUUU	CAGAUUCAUG	AUUGCACUUG	AGAAAGGGAU	CUUACACGAA

Figure 3.7: Nucleotide sequence of SCSMaV-RNA

5701	UGCCUUGAGU	CUUACUCGAU	UGAGUGUUCA	UAUGCAUAUA	ACAUGGGAAG
5751	UGCAGUUUCA	AGACAUUAUGG	ACCAGCUUAG	CCUUGCUUAU	CAUUACUGCU
5801	GCGUGAGGUU	UAUUCUAAAA	AAUUCAAAAAC	UUCUUCGCUC	UGAUGUGAAG
5851	AAGUUCUUUA	AGGACGGGCU	CCAGCUUACG	CAGUAGCCAC	GUUUAGGCUG
5901	CGACUUGUAU	AUAUUGAUUA	GGAAAUAGUC	UGUCGUUACU	UAGACGACUU
5951	UAAGUUUGAG	AGAACUAGCU	UAGAGCUUUC	UUCACCUAUA	GUGGUUCACG
6001	CAGUCCAGG	UUCAGGGAAG	UCUACCCUCA	UAAGAAAGUG	CAUAACUGAA
6051	AAUCCUUCUU	UGACAGCCUG	CACCUUUGGA	AAGCCAGACU	UGCCCAUUCU
6101	UACUGGUACC	UAUAUAAAAG	GGUAUAAAAG	AGGUGAGAAG	UAUGACAUAA
6151	UUGAUGAGUA	CAUCGGCAAC	AGAUAUUGAGU	GCGUUCUUGU	UGCACUCUUU
6201	UCUGAUCCUU	UUCAGAAUUG	UGAACAUAAC	AUUAUAACAG	CACAUUAUAU
6251	UGGACGAACU	UCUCACCGUU	UCGGCGUUGU	ACUGCUGAGC	UAUUGC GCGG
6301	GCUUGGUUAU	CAAGUUUCCU	CAUCUCGCAA	AGACAAGGUC	ACUCAGGGUU
6351	CAGGUUACUC	UAUUGAAUCU	ACAGGUGCUA	UCAUAUGC UU	UGAGCCUGAC
6401	ACUGAGGACU	AUUUACUUGC	ACAUAAACCUC	AAGCCUCUUC	AUCCCUUGGA
6451	CAUCCGCGGA	GAGACUUUUG	ACAAAGUUAC	AUUCUAUCAU	UCUAGUGAUC
6501	UCCGCGAUCU	UGUUGGUCCU	GAGCUAUUAU	UUUCGCUGAC	UCGCCAUAGA
6551	AAUGAAAUCA	AAUAUUGCAA	GCUUAAUGUC	AGCUCUUAGG	CAGCCAUCUG
6601	AUAAUUCUGA	GUCAAUUAUA	AUUUGCGCUG	CUCUAGGGGG	GACUGCAUUA
6651	CUUCUUUAUA	UUUGGUGUCA	GCCUAAUAAU	CCAAAAUCAG	GGGACCUGGG
6701	CCAUUGGUUA	CCUUUUGGGG	GAGUUUACAA	AGACGGUAAC	AAAUUCGUCA
6751	GGUAUAAUUC	ACCUCUUCAC	ACCGCUCCUC	CUUUCUGGCU	UUUAAUUGCA
6801	AUACCUUUGG	CCUUCUUGC	UCUCACUCAC	UUUAGCUACU	UCAGUAACAC
6851	ACCCAGAAGA	AGAGUGACCA	UAACACAUAU	UCCAUUGUGU	GGCCACUUUG
6901	UAUGCAGGCG	CUAGAGAUUA	AUAUUCUGGU	GGUUUUUCUA	CUGCUGAUAG
6951	CCUCCUUACU	CUUCAACAA	GAGAGUCAGC	AGUGCAUAAU	UCACAUUGAC
7001	GGGACCAAAG	CUUUUAUCAC	CGGCUGCACU	GAUCAAAGU	GGUUGGAUAC
7051	UGAAGUACAA	CACCUAGAC	CAGACAAUCA	AAGAUCAGGU	UGGUUUAGAA
7101	GAUGAGUGGG	CCACCCCUU	CAACUGGUGC	CAAUGCUUCC	AGUGGUGCUC
7151	AACCAACAAC	UCAGCAGGAA	ACGAGCACAG	UUUUUAAGAA	AAGCCGCAGA
7201	GGAUCAAAAU	CUAAAUUUGU	UUCUGCUGUU	ACAAAGGAGC	AGGUUAGGAG
7251	GGUGGUCGAG	CUUGACAUUG	AGACAGACAA	AAUAGCCCGG	GAUGAGGACC
7301	UUGAGAAAGU	GAAAGCUGCU	UGGCUUGCCA	UUAAUGUGCC	AUCCGACAAG
7351	AUAUUUGAAG	UUUCCCUUGA	AAUAACACUU	UCCUGUGCUC	ACACUUCGAC
7401	UUCAAAUAAG	CAGAAAUUA	AAGGCAAGUC	CAAAUUCUGC	GACAUAUUGC
7451	UUGCUGAAAU	AGCUGGCUGU	AUCACUAAGG	AGGAGGUAAAC	AAUACGUUCC
7501	UUUUGUGCUU	AUUGGGCAAA	AUACGUCUGG	AACGCAAUGC	AUGAACGUAA
7551	UGAGCCUCCU	CUCAAUUGGG	UUGCUCUUGG	GUUAACAGAA	CAAAACCGUU
7601	UUGCAGCUUU	UGACUUUUUC	UAUGGUGUUG	ACUCUCCUGC	UUCAGCAGAA
7651	CCAGAAGGUG	GGUUAUUGC	UCCUACCACU	GAGUCUAUGA	GAGCUGCAAA
7701	CAUGGCUAGG	AAAGAGAUAG	CCAUCAAAAG	AUGAGAGAAA	GGGAGGGUAU
7751	GUCAUAUUUG	AAUCUGUCAG	CAGUCACAGG	UGCCAGGAU	AGUAUAAGGC
7801	CAAGGUUAGA	UUUCUCACCU	GCAGAAAUAA	AUUAGCUUUU	CUUUUAUUCU
7851	ACCCCCUGUU	CAAGCAGGUC	GGACCUUCGG	CAAGUCCAAC	AAAAAGAUUU
7901	CUUUCUAUUU	GUAAAUAUACU	GUCGAGGUAC	AGAGUGUUGC	UUUGCACGAC
7951	ACCUUUGGGU	GAACCUAGU	CUUAGAACUA	GUGCUUUGUC	AAUCAUUUAA
8001	UGGCCAAGGC	AAGGCAUUGU	CUUGGGCACU	AUUUGAUUGA	CAUUGUGGUC
8051	UUUCUAAGAC	CCUCGGCUUC	AAUAAACAGC	AUUGACAAGC	AUGUACGUAU
8101	CUCAUGCACG	CUGUAUAAAA	AUAGCUUUGG	UUUUACAUAU	UUUCCC (A) _n

3.4 Discussion

The existing rPCR clones did not give a representation of the whole SCSMaV-RNA genome (Choi, 1997; Choi *et al.*, 1999). It was therefore necessary to extend the known sequence using a variety of methods (Figure 3.6). The alignment with the ASPV genome showed three regions that could be linked together by bridging RT-PCR. This gave a continuous sequence of approximately 3kb. Extension towards the 5' and 3' ends was done to gain more sequence information. A number of methods were employed for this extension, including an anchored RT-PCR approach and the use of a degenerate oligonucleotide designed to anneal to the *CP* from a number of species from the genus *Potexvirus*. The 5' and 3' termini sequence were obtained using the RACE technique. Sequence assembly was done sequentially using the overlapping regions to ensure that the extended sequence was accurate.

CHAPTER 4

**GENOME ORGANISATION OF
THE SCSMaV RNA**

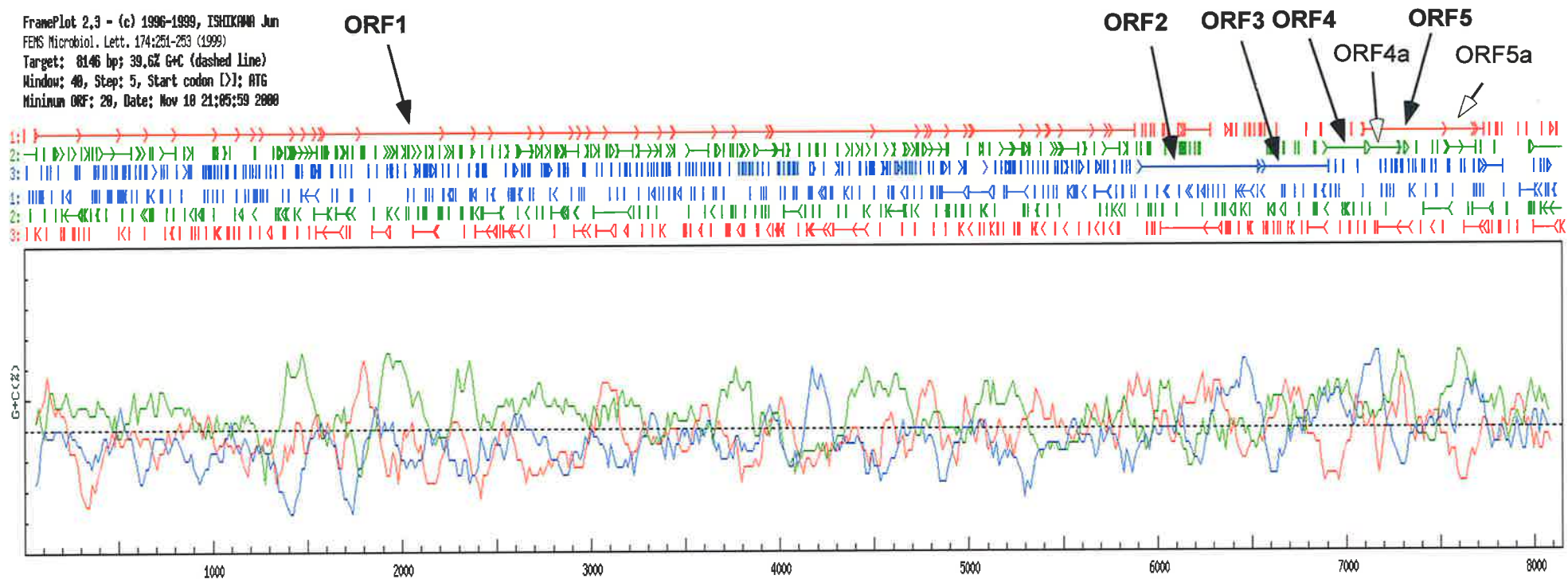


Figure 4.1

Data output from Frames 2.3 of the nt sequence of SCSMaV RNA. The six reading frames are represented as three forward and three reverse, the top three representing the forward reading frames. The arrow heads show the potential start codons (AUG), the vertical lines show the stop codons (UAA, UAG and UGA) and the horizontal lines show the potential ORFs greater than 20 amino acids for each frame. The percentage G+C is shown for the areas of the nucleotide sequence and the nucleotide numbers are shown below.

The SCSMaV ORFs 1 to 5 are indicated by the solid head arrows and ORFs 4a and 5a are indicated by the open head arrows.

Table 4.1: Potential ORFs identified by Frames 2.3 analysis of the complete nucleotide sequence of SCSMaV RNA showing the predicted position, size (Frames 2.3) and molecular ratio (PepStats).

ORF	Nucleotide position	Potential peptide size	M_r
1	73-5886	1937	221526
2	5919-6539	206	23311
3	6552/6576-6914	120/112	13511/12667
4	6902-7105	67	7834
5	7102-7734	210	23220
4a	7133-7285	50	5838
5a	7541-7687	49	5428

CHAPTER 4: GENOME ORGANISATION OF THE SCSMaV-RNA

4.1 Introduction

The sequence of SCSMaV was determined in Chapter 3 (Figure 3.7). Analysis of the nucleotide sequence for start and stop codons is used to identify potential open reading frames (ORFs). The program Frames 2.3 identifies potential ORFs of a candidate nucleotide sequence and calculates the third letter G + C content (Ishikawa and Hotta, 1999). The program uses internet three web pages for analysis of a sequence: a sequence submission page; a results page showing potential ORFs, their nucleotide and amino acid sequences; and a feature page allowing Blast searches of the genome database to identify nucleotide or amino acid sequence similarities.

The genome sequence and organisation together with the virion morphology and host range can be used to determine the taxonomic position of a virus. This chapter describes the genome organisation of the SCSMaV-RNA and characteristics of specific SCSMaV ORFs to determine its relationship to similar viruses using available sequence data.

4.2 Materials and methods

4.2.1 Identification of open reading frames

The Frames 2.3 program and the linked Blast genome database searching facility were used to identify ORFs and for preliminary genome database searches for amino acid and

nucleotide sequence similarities. The program is available for academic purposes at <http://www.nih.go.jp/~jun/cgi-bin/frameplot.pl>.

4.2.2 Nucleotide sequence analysis

Programs for nucleotide sequence analysis were accessed through the webANGIS interface GCG programs (<http://www.angis.org.au>). Nucleotide database comparisons were done using the programs FastA and Blast. The percentage sequence identity between sequences that showed similarity was calculated using the GCG program HOMOLOGIES.

4.2.3 Peptide analysis

Programs for peptide analysis were accessed through the webANGIS interface GCG programs (<http://www.angis.org.au>). Peptide composition was analysed using Pepstats. Motif searches were done using the PROSITE protein motif dictionary accessed through the Motif program. Hydropathy plots were done using the method of Kyte and Doolittle (1982) with the program Pepwindow. Genome database searches were done using FastA and Blast. The percentage amino acid identity was calculated using HOMOLOGIES. Multiple sequence alignments of amino acid data were done using eClustalW (Thompson *et al.*, 1994) and interpreted using the GeneDoc program.

4.3 Results

4.3.1 Genome organisation

Analysis of the complete nucleotide sequence using Frames 2.3 identified seven potential ORFs (Table 4.1, Figure 4.1). Blast genome database searches using the

putative peptide sequence showed that five of these potential ORFs had similarity to viral peptide sequences in the GenPept database, and were therefore included in the genomic map. The two remaining peptides had no similarity to peptide sequences in the genome database and therefore could not be assigned a potential function. The ORFs are shown on the nucleotide sequence (Figure 4.2a) and in a genomic map of SCSMaV (Figure 4.2b).

4.4 Discussion

4.4.1 ORF1

ORF1 spans nt 73-5886 and encodes a protein of 1937 amino acids with a predicted M_r of 221000 (Table 4.1). The predicted amino acid sequence of the ORF1 peptide is shown in Figure 4.3a. Blast and FastA genome database searches showed that the ORF1 peptide had similarity to RNA polymerases of viruses from a range of genera (Table 4.2) with amino acid sequence identity of up to 32% with BBScV (Appendix D). A high degree of similarity between the C-terminal portion of the ORF1 putative product and the RdRP of viruses from a large number of genera, including *Potexvirus*, *Foveavirus*, *Carlavirus* and *Trichovirus* was observed by multiple sequence alignment (Figure 4.3b). A search for protein motifs in the ORF1 product found a predicted ATP/GTP binding site spanning amino acids 1121-1139 (Figure 4.3a) and alignment of the RdRPs and the ORF1 product identified eight amino acid motifs specific to RdRPs (Figure 4.3b; Koonin, 1991; O'Reilly and Kao, 1998). The two major domains of the RdRP are the helicase and polymerase domains, and there is evidence that these are autocatalytically cleaved (Lawrence *et al.*, 1995). The most conserved motifs were IV-VIII including the GDD domain at amino acid position 1808-1810 (Figure 4.3a, motif VI). The similarity

Figure 4.2a: Nucleotide sequence of SCSMaV-RNA showing ORFs

Figure 4.2a

Nucleotide sequence of SCSMaV-RNA showing the positions of ORFs 1 to 5

1	UGAAAACAAA	ACUACACAAU	UGAGCUCAAA	CUGCUACAAU	AUCCACUCAA
51	UAUACCUUAAU	UGAUUCAUAAU	UGAUGGCAAU	UCUCACCUUC	AAACCCCUAA
101	AGCAGCUCUU	AAUCGCUCAA	UUGUCCUCUG	AAGAGCAGGC	CAGAAUUAGG
151	AGCUCUGUGA	CACGAAAGUU	AGCGCAGGAG	GAAGAGAAUG	AUCAAGGUUU
201	AUUCAGUUAU	CAUGUCUCUC	CAUACUGUGC	AGAGAAAGUUG	AUCAAGUCUG
251	GUGUUAACU	CUGUGCCUAC	UCCUCUGAUG	UACAUAGUCA	UCCUUGUUGU
301	AAAUGCUAG	AAAUCAUAAU	UUUAUACAGA	GUUCUCCUU	CUUAUGUAGA
351	UUCAGAUUUU	AUCCUUGUAG	GUAAUAAGCA	UUCUAAGUUA	GCCUUUCUUA
401	AACCAAGAGG	CGCGAAAAAG	AAUGUUGAUC	UUUCAACUGU	UAAGGUGCUC
451	AAUCGUUUUG	UCACAAGUCU	UGACAGAUCC	AGGUAUGGUA	AAGAGUUUAA
501	UCACCUCAAU	AUCAUGUCAA	CAUUCACUCC	UGAACCUAAG	GGUAGCAGUU
551	CCUAUACCCC	UGAACUUAAA	AGCCUAAUUC	CAAAUCUGGA	GCGAUUUUAG
601	GGCCACAAAC	UUUUUUUCCA	UGAUGAGUUG	CAUUACUGGA	AGGAUGAAUC
651	UAUUAAUAAA	AUGCUAUCAU	CUUUAGAGCC	UUCAGUAGUG	UACGCAACUG
701	UUGUAUGCCC	ACCCGAAUUG	CUUAUUGGUG	CUAAGCAGAG	UUAAUAUGAU
751	UGGCUAUACA	CUUUUGAAAU	CAUUGGGGAU	CAAAUCUUUAU	UUGCUCUAGA
801	CGGAAAGAUG	AAUAGUGCCU	ACUCCAACCC	UGUCAAUUGU	GCCCGCCUUC
851	UAAGAGCAAA	AAGGAUAUGG	CUGGAUAAGC	AUAACUAUUA	UGAUUAUGAU
901	CUGGUCUAUU	CAAAUUUCUC	GCAUCAUUUG	CUCUGCUUCU	CAAAGGGCUC
951	AAAUUUUAA	AAUCAAGAUC	UUAGGGCUUU	CUCUGACUUU	GACGCAUUUA
1001	AAUUAAAUAU	CAUAAGUGAG	CUCAUGGUGA	AUGCUGAGAG	UUGCUUUCCA
1051	AUUAGUUGGA	CUGUUGUAAA	AGGGCUUUUA	GAUUUAUCUUG	AGAGCUUGAC
1101	AAAAGAGGAU	UACAGAUCAU	GUAAGGCUAA	GCAAACUCAU	UAUAUGCCCA
1151	AUUGCUCAGG	GAAACAACUU	AAAUUUUUGA	CUGACCUUUA	CAAACUUUAU
1201	CUUGAAGCUC	CACAUUUUAA	AACUAUGAUA	GGUUCUAGCC	CAGUUCACAA
1251	CUUUAAAGGUC	UUGUUUACAC	GGAUGUUCUU	UGCUUGCAUU	AAGAAAAAUU
1301	UCUCUUUUUA	UAAAGGAAAA	UUGUUGAAAA	GUUUUAUUUGA	ACAUCUAGCU
1351	CCUUUGGCUU	UCCAUGUCAA	AACUUAUGAU	CUUCAUUUAG	GUUUUGUUUU
1401	CAAAUUCAG	GAGCUGAAAA	AUGUGACGGA	UCUAAUGGAG	GAUGUUAUG
1451	AAUUUGAAGU	UGACUUUGUC	AGAGAAUAUG	ACCCAUUCCA	GAUGCGCAGC
1501	ACUCCUUAUG	UUGGUCUUGC	CCCUUUGCGU	GAUGAAUCAU	AUUUAUAGAG
1551	GAUGAAUUGU	GGUGAGGCUU	UCUUUGACAA	CAUGAUUGGG	CAAAUAAUGA
1601	GAUCCCAUUU	AAGCGGGUAU	GGCACAAAUA	CUAAUCCUGC	UGAACUUUUC
1651	CUAAGAAUCA	AUUUGACCAU	UGUUGGAAGG	GUUGAGUUUC	UUGUGAAUGG
1701	CCAUCAUGAU	UUUGUAAUUU	UUUAUAUCAA	CAAAUUAUUG	AAGGUAAAGA
1751	AGGACCCCAG	GUACAAAAAC	UUGUACUUUG	CAAAUAUGAC	CCUCGAAUUG
1801	AAAUUGAGGU	GGCUUUCGUG	CAGACGUUUU	AGAAAGUAUC	AGAAGUUCUU
1851	ACCCUGCGGU	CUGGUGAGUC	UUGCUGAAGU	ACAAAUUCUU	UGGAGAAUAU
1901	UUGCGAGAAG	CAUCCAGAGU	GGAUCACAUG	AGGCUCAAGG	ACAAAUGAA
1951	CAAGUUGAUU	CCCAUGGGGA	UGGAGUAGCA	AUUGAGGAGC	CAGAAAGGAA
2001	GGGGAUUACU	UUGGAAAAAU	UAAAUGAAGC	UACUUUAGUA	ACCCACCUCC
2051	AGCACUACCU	CAAGAAUCCU	GUUGUUGAUG	AUUUUUCAAG	AUGCGCCUUU
2101	GAUGUCGACU	CAAAGACUCU	AUCUAAUCUA	AUUGAUUGUC	UUAAACAGAU
2151	UUGUGAUACU	AGCAUAAAUG	UUGCCAUGG	AUGUGACUGU	UCAAAUUCAG
2201	AUCUAAAUUU	ACUCAAGAUU	AGUAAAAUGG	CCUCCUCUGU	GGCAUCAUAU
2251	GUUAAUCUGG	ACAAUGUCAA	AUCUGUGAAA	AUCAAUUGGA	CCGAAGCUGG
2301	UGAGAUCUUA	GUCGGUUUAG	AGGAUCAGCC	UCUUUUUUGUC	GUGGUUCUUG
2351	GGAAGGAUUA	UAAUUCAAUU	ACAUUAGCAA	GGGAUCAUGU	UAUGCAUUUU
2401	CCUGAGCUUG	AAGAAAAUUC	AGCAAUUUUA	AUUGAAAGGA	AUAACAUCUC
2451	UAAAAUUUCU	GUUUCUGGCC	CAUCUGCUCU	CAUGAUCUUU	CCAAAAGAUC
2501	AGAAUUUGCU	GGAUUGUCAU	UUAAUUGGCU	GCGAUCAAUU	CCCCCUGCA
2551	AAGAUUGAUU	UCAUAGAUGG	GUACAUCAAA	ACUACGGUGG	AUCCUAGUCU
2601	UGAUUCUCGG	UUAACAUGCA	UCAAUUGUAC	CGCAAUUGGU	GAUUGUUUUU
2651	GGCAUAGUGU	AUCUCUUUAU	CUUAGUGUUG	AGGCCAAAUA	GAUCAAAGAU
2701	GCACCUGUUC	UAAGAAACAG	GAAUCCCCAA	AAUGCCAGAU	UAGUUGAACC

Figure 4.2a: Nucleotide sequence of SCSMaV-RNA showing ORFs

2751	AAUGGGUGAU	AAAGUUUGGG	CUGAAAUGA	GGCUAUACUA	AACUGUGCAU
2801	GUUUUUACAA	CUUGAAUAUA	CGUAUAGUCU	UUGAACAUGA	AUUGCAUACC
2851	UUCCGUCCCC	AUGAUUUGAA	UGAAGAUUCA	ACUGAAGUUU	GGCUCAAAAA
2901	UUUGGAUCAU	AUGCAUUUUU	GUCCAGCAAA	GUUUUAUAAU	UCUUGUGUCA
2951	UGGAGCCAAU	UUCUAAAGCA	GUUGACAUUC	CGCCUUUGAA	UGUACAAAGA
3001	AUGAUAUCUU	CACAGGAGAG	UCUUCACUCC	GAGUAUGAGA	GAUUGAGCAG
3051	UGGGAAAUCC	UUUGAUCUCC	UUGGACUGGA	GAGUUUGAUG	ACUUUCUUCA
3101	ACAUAGGGGC	UGAAGUCCAG	GAAAAUGGUC	ACAAGUUUAU	UUUGAACGAG
3151	AAUGGCAGUA	GGGCUCGAAG	CUUCAUUCUG	GAGAAUAAUC	AUAUUGAAUU
3201	UGUUCGUGGA	GAUAAAAAGG	AUCUAAGAAA	GGAGGGAAUU	GGGUUUUGUUA
3251	AGCUAAUGAA	UGCAAAUCAA	AGUAAGUCUG	UGCUGUCUCU	CGCCUGUCUA
3301	AAGGUUAAAU	ACACCCCAUC	UUAUGACAGA	GCCUUUAAAU	UGUCAAAGC
3351	UUUUGCUAAU	GGUUAUACAG	GCAUAAUGCU	UAGCAGUGAA	AAAUUUGGCG
3401	AGAUUGAAAU	CAACGGUAAG	AAUGACGAGA	GGGAAUCAA	UGUUAUGAUG
3451	GGAACAUUUG	GGGCUGGAAA	GACUAGAGUA	UUUGUGGAUU	AUAUUCAGGC
3501	UUACUCUGGU	AGAGGUUUUA	UUUAUGUUAG	CCCAAGAAAG	AACCUUAAGG
3551	AGCUUUUCGA	UAACUCUGCC	AUCCAGGCAG	UUCAAGGCUU	UAAAACCAAG
3601	UCAAGUGACA	AGAAAGGCAG	UGUAAAGAGU	UUUCAUAGCU	CUUUCACCUU
3651	UGAGACUGCU	AUAAUGAAAG	CCGGCCAGAU	UGGUGAUGAU	UCAAUUCUUA
3701	UUAUAGACGA	GAUUCAGCUC	UAUCCGCCUG	GCUACCUUGA	UUUAAUCCUU
3751	CUAUUGAUAC	CUUCAAGUUG	CAUGGUAAUU	UGUGCUGGUG	ACCCUGUUCA
3801	GUCUGACUAA	GAUUCAGAUG	GUGACCGUUU	GUUUUUUCUU	AAUGAUAAAC
3851	CAGAUUUUGA	GCGGUUGUUA	GAGGGUCAAU	CUUAUAAGUA	UGCAAUCAA
3901	UCAAGAAGGU	UUAAAUCCUC	CUUAUUCAUU	GGACGACUAC	CUUGUGAUUU
3951	GCCCGGUCUU	CAAGGGGAUGG	GCAUGGACAC	UUUUGGGAUU	UAUGAGAGCA
4001	UUAGCAGUGU	GCCUCAGGAU	UUAAUUAAGA	GAUCCUUCUU	UUUGGUGAGU
4051	AGUUUUUGAUG	AGAAGAAAAU	UAUUAAAGCC	AAUUUUAAAG	AGACUCAUAA
4101	CAAGAUUUUG	ACUUUUGGGG	AAUCCACAGG	AUCAACUUUU	CAAGAAAUUG
4151	UCAUUCUCAU	AACUACUAGU	UCAGAAAAAGG	CAAGUGAGAG	GAGAUGGGUA
4201	ACAGCAUUA	GCAGAGCAUC	GACAAGCAUA	UCUUUUUCUUA	ACCUGACUGG
4251	CAUUGGAAUA	CAAGAGCUUA	UAUCGACUAC	UUUUUUUGGC	AGAUUCUUGC
4301	AAAAAUUUUU	AAGUGCUGAA	UGCAAGAGGG	AGGAUAUUAA	GCCAUUUUA
4351	UACGGAGAGC	CAAAAUUUGU	UGAUGACUUC	AGGAGCAUAG	GAAGACGAAG
4401	UGCAGAGGAG	AAAGAAGUCA	AAGUUGAUGG	UGAUCCGUGG	UUAAAGUGCA
4451	UCUUGCAUUU	CUUUGAAACU	CCAAUUCAG	AUGAAGUUGA	GUGUGUUAAA
4501	GUUGAUUUGG	UAGAGCAAAA	AGCAAAAACU	CACUUACCGG	UUUCGUCCUU
4551	UGAGGUUUUA	UGGGCCCAUU	UUAAUGAAAA	ACUUUUGCUU	AAGGAGAUCA
4601	GGGAGCGGAG	GAUAGAUGAU	CUUGUUUCUG	ACCAGUUUGC	AUGGGAUCAU
4651	AGGUCACAAG	GACUAGCAAU	UAAUAAUGCU	GCCACAAGGU	AUGAAUCAAU
4701	AUUCCCUAGG	CAUACCACA	AUGAUAAAGU	GACCUUUUUG	AUGGCAGUUA
4751	AGAAAAGGUU	GACCUUUGGA	UCGCCCAGAA	AGAAUGGGAG	GAUGCUUAGA
4801	GAGGCUAUGC	CCUAUGGUGA	AAUAUUACUC	AAGGAGUUUC	UCAAGAGGGU
4851	CCCUAUUAAG	AAUGAUAGAA	ACGAUCAUUU	AUUUUUAGAA	UGUCUAAUGG
4901	AUUUUGAAGA	AAAAAAAACU	UCCAAGGACU	CAGCCACAAU	CGCAAUACAC
4951	UGCAUACGCU	CACAGAAUUC	AUGGCCAAUU	GAUAUUGCUG	AUAUAUUCAU
5001	GAAGUCACAA	CUCUGCACAA	AAAUGGAGAG	UAUGUUCUGC	GAGGCAAAGG
5051	CUGGUCAGAC	UUUAGCUUGU	UUCCAUCAUG	UAGUUCUUGC	ACGAUUUUUCU
5101	CCAUGGGUCC	GUUACAUAGA	AAAGAAGCUG	GUCAGUUCAC	UGCCAGAUAA
5151	UUACUAUUA	CACUCUGGAA	AGAACUUUGA	UCAACUCAAU	GCUUGGGUUA
5201	AGGCAAACAA	UUUUGUGGGU	GAAUGCACUG	AAUCGGAUUA	UGAGGCCUUU
5251	GAUGCAAGUC	AAGAUCACUA	UAUACUGGCC	UUUGAAUUAG	CAUUAUAGAA
5301	GUACCUUGGU	UUACCCAAUG	AGUUGAUAGA	GGACUACAAG	UACAUUAAGA
5351	UUCAUUUGGG	UUCAAAAUUG	GGUGAUUUUG	CCAUAUAGAG	AUUCACCGGU
5401	GAAGCCAGCA	CUUUCUUAUU	UAACACAAUG	GCAAUAUUGC	UUUUCACUUU
5451	UCUGGUCUAA	GAUAUCAAAG	GCAAUGAGUC	AAUUUGCUUU	GCUGGCGAUG
5501	ACAUGUAUGC	AAAUACAAGU	CUUAAGACAU	GCAGCAAGAA	CAAAGUUUG
5551	CUUAAAAAGU	UUACACUCAA	AGCUAAAGUU	CAGGUAACUG	AAAGCCCAAC
5601	AUUUUGCGGA	UGGCACUUGA	CAGAGGAUGG	AAUUUAUAA	GACCCAAUAC
5651	UUGUGUCUUU	CAGAUUCAUG	AUUGCACUUG	AGAAAGGGAU	CUUACACGAA
5701	UGCCUUGAGU	CUUACUCGAU	UGAGUGUUCA	UAUGCAUUA	ACAUGGGAAG
5751	UGCAGUUUCA	AGACAUUAUGG	ACCAGCUUAG	CCUUGCUUAU	CAUUAUCUGU

Figure 4.2a: Nucleotide sequence of SCSMaV-RNA showing ORFs

5801	GCGUGAGGUU	UAUUCUAAA	AAUUCAAAAC	UUCUUCGCUC	UGAUGUGAAG
5851	AAGUUCUUUA	AGGACGGGCU	CCAGCUUACG	CAGUAGCCAC	GUUUAGGCUG
5901	CGACUUGUUA	AUAUUGAUUA	GGAAAUAGUC	UGUCGUUACU	UAGACGACUU
5951	UAAGUUUGAG	AGAACUAGCU	UAGAGCUUUC	UUCACCUAUA	GUGGUUCACG
6001	CAGUUCAGG	UUCAGGGAAG	UCUACCCUCA	UAAGAAAGUG	CAUAAACUGAA
6051	AAUCCUUCUU	UGACAGCCUG	CACCUUUGGA	AAGCCAGACU	UGCCCAAUCU
6101	UACUGGUACC	UAUAUAAAAG	GGUAUAAAAG	AGGUGAGAAG	UAUGACAUAA
6151	UGAUGAGUA	CAUCGGCAAC	AGAUUUGAGU	GCGUUCUGU	UGCACUCUUU
6201	UCUGAUCCUU	UUCAGAAUUG	UGAACAAUAC	AUUAUACAG	CACAUUAUUA
6251	UGGACGAACU	UCUCACCGUU	UCGGCGUUGU	ACUGCUGAGC	UAUUGCGCGG
6301	GCUUGGUUAU	CAAGUUUCCU	CAUCUCGCAA	AGACAAGGUC	ACUCAGGGUU
6351	CAGGUUACUC	UAUUGAAUCU	ACAGGUGCUA	UCAUAUGCUU	UGAGCCUGAC
6401	ACUGAGGACU	AUUUACUUGC	ACAUAAACCUC	AAGCCUCUUC	AUCCCUUGGA
6451	CAUCCGCGGA	GAGACUUUUG	ACAAAGUUAC	AUUCUAUCAU	UCUAGUGAUC
6501	UCCGCGAUCU	UGUUGGUCCU	GAGCUAUUA	UUUCGCUGAC	UCGCCAUAGA
6551	AAUGAAAUCA	AAUAUUGCAA	GCUUAAUGUC	AGCUCUUAAG	CAGCCAUCUG
6601	AUAAUUCUGA	GUCAAUAAA	AUUUGCGCUG	CUCUAGGGGG	GACUGCAUUA
6651	CUUCUUUAUA	UUUGGUGUCA	GCCUAAUAAU	CCAAAUCAG	GGGACCUGGG
6701	CCAUUGGUUA	CCUUUUGGGG	GAGUUUACAA	AGACGGUAAC	AAAUUCGUCA
6751	GGUAUAAUUC	ACCUCUUCAC	ACCGCUCCUC	CUUUCUGGCU	UUUAAUUGCA
6801	AUACCUUUGG	CCUUCUUGC	UCUCACUCAC	UUUAGCUACU	UCAGUAACAC
6851	ACCCAGAAGA	AGAGUGACCA	UAACACAUUA	UCCAUGUGU	GGCCACUUUG
6901	UAUGCAGGCG	CUAAGAUUA	AUAUUCUGGU	GGUUUUUCUA	CUGCUGAUAG
6951	CCUCCUUAUCU	CUUCAACAA	GAGAGUCAGC	AGUGCAUAAU	UCACAUUGAC
7001	GGGACCAAAG	CUUUUAUCAC	CGGCUGCACU	GAUCAAAGU	GGUUGGAUAC
7051	UGCAAUCAA	CACCUGAGAC	CAGACAAUCA	AAGAUCAGGU	UGGUUUAGAA
7101	GAUGAGUGGG	CCACCCCUU	CAACUGGUGC	CAAUGCUUCC	AGUGGUGCUC
7151	AACCAACAAC	UCAGCAGGAA	ACGAGCACAG	UUUUUAAGAA	AAGCCGCAGA
7201	GGAUCAAAAU	CUAAAUUUGU	UUCUGCUGUU	ACAAAGGAGC	AGGUUAGGAG
7251	GGUGGUCGAG	CUUGACAUUG	AGACAGACAA	AAUAGCCCGG	GAUGAGGACC
7301	UUGAGAAAGU	GAAAGCUGCU	UGGCUUGCCA	UUAAUGUGCC	AUCCGACAAG
7351	AUAUUUGAAG	UUUCCCUUGA	AAUAACACUU	UCCUGUGCUC	ACACUUCGAC
7401	UUCAAAUAAG	CAGAAAUUUA	AAGGCAAGUC	CAAAUUCUGC	GACAUAUUGC
7451	UUGCUGAAAU	AGCUGGCUGU	AUCACUAAGG	AGGAGGUAAC	AAUACGUUCC
7501	UUUUGUGCUU	AUUGGGCAA	AUACGUCUGG	AACGCAAUGC	AUGAACGUAA
7551	UGAGCCUCCU	CUCAAUUGGG	UUGCUCUUGG	GUUAACAGAA	CAAAACCGUU
7601	UUGCAGCUUU	UGACUUUUUC	UAUGGUGUUG	ACUCUCCUGC	UUCAGCAGAA
7651	CCAGAAGGUG	GGUUAUUGC	UCCUACCACU	GAGUCUAUGA	GAGCUGCAAA
7701	CAUGGCUAGG	AAAGAGAUAG	CCAUCAAAAG	AUGAGAGAAA	GGGAGGGUAU
7751	GUCAUAUUUG	AAUCUGUCAG	CAGUCACAGG	UGCCAGGAAU	AGUAUAAGGC
7801	CAAGGUUAGA	UUUCUCACCU	GCAGAAAUAA	AUUAGCUUUU	CUUUUAUUCU
7851	ACCCCCUGUU	CAAGCAGGUC	GGACCUUCGG	CAAGUCCAAC	AAAAAGAUUU
7901	CUUUCUAUUU	GUAAAUAUCU	GUCGAGGUAC	AGAGUGUUGC	UUUGCACGAC
7951	ACCUUUGGGU	GAACCUCAGU	CUUAGAACUA	GUGCUUUGUC	AAUCAUUUAA
8001	UGGCCAAGGC	AAGGCAUUGU	CUUGGGCACU	AUUUGAUUGA	CAUUGUGGUC
8051	UUUCUAAGAC	CCUCGGCUUC	AAUAAACAGC	AUUGACAAGC	AUGUACGUUA
8101	CUCAUGCACG	CUGUAUAAAA	AUAGCUUUGG	UUUUACAUUA	UUUCCC (A) _n

ORF1: Blue; ORF2: Pink; ORF3: Purple; ORF4: Apricot; ORF5: Yellow

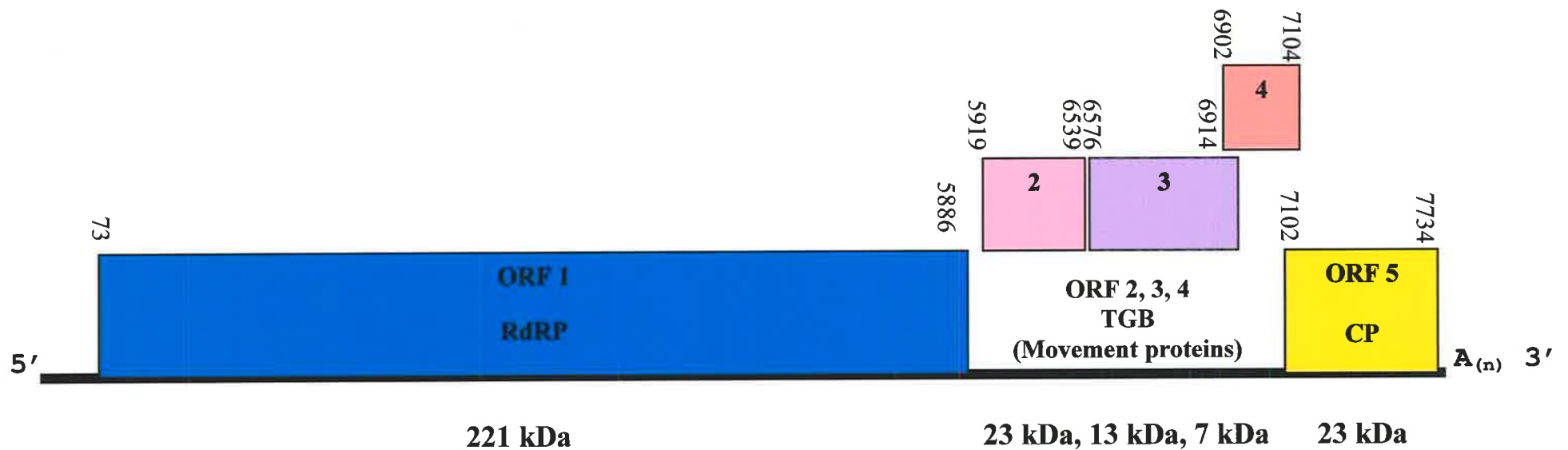


Figure 4.2b

A genomic map of SCSMaV showing the size and potential function of ORFs and their products. ORF 1 spans nt 73 – 5886 (Blue), ORF 2 spans nt 5919 – 6539 (Pink), ORF 3 spans nt 6576 – 6914 (Purple), ORF 4 spans nt 6902 – 7104 (Apricot). ORF 5 spans nt 7102 – 7734 (Yellow).

Table 4.2 FastA database search results for ORF 1 peptide sequence

The best scores are:		length	initn	initl	opt	z-sc	E(542617)
X	BAA04853 Apple stem pitting virus (ASPV)	(2183)	2627	1620	2829	3193.9	2.3e-170
X	AAA68984 Blueberry scorch carlavirus mRN	(1967)	2560	1769	2819	3183.2	9.1e-170
X	CAA92815 Garlic latent virus genes for r	(1924)	2798	1863	2813	3176.6	2.1e-169
X	AAC35433 Sour cherry green ring mottle v	(2023)	3075	1658	2647	2988.1	6.7e-159
	AAD10481 Sugarcane striate mosaic virus	(416)	2632	2632	2632	2981.7	1.5e-158
X	AAC62910 Grapevine Rupestris stem pittin	(2161)	3088	1520	2615	2951.4	7.4e-157
	AAC35498 Rupestris stem pitting associat	(2161)	3092	1523	2601	2935.6	5.7e-156
X	CAA37232 Potato virus M complete genome	(1968)	2513	1234	2564	2894.3	1.1e-153
	BAA03339 Potato virus M genomic RNA, com	(1968)	2513	1234	2564	2894.3	1.1e-153
X	CAA68080 Apple spot chlorotic leaf spot	(1887)	1769	1279	1594	1795.3	1.9e-92
	BAA03641 Apple chlorotic leaf spot virus	(1885)	1793	1283	1584	1784.0	7.9e-92
	AAA42587 Apple chlorotic leaf spot virus	(1884)	1780	1274	1563	1760.2	1.7e-90
	CAB46073 Apple chlorotic leaf spot virus	(1883)	1714	1265	1550	1745.5	1.1e-89
X	CAA57896 Cherry capillovirus A RNA for O	(2342)	1810	1419	1521	1711.2	9.1e-88
	AAD10482 Sugarcane striate mosaic virus	(223)	1451	1451	1456	1653.2	1.5e-84
X	BAA00350 Potato virus S genomic RNA, 3'-	(353)	1346	1346	1369	1551.5	7e-79
	CAA75709 Potato virus S gene encoding re	(355)	1342	1342	1368	1550.4	8.2e-79
	CAA75710 Potato virus S gene encoding re	(355)	1344	1344	1364	1545.8	1.5e-78
	CAA75714 Potato virus S gene encoding re	(352)	1343	1343	1363	1544.7	1.7e-78
	CAA75712 Potato virus S gene encoding re	(355)	1339	1339	1359	1540.2	3e-78
	CAA75711 Potato virus S gene encoding re	(355)	1339	1339	1359	1540.2	3e-78
	CAA75716 Potato virus S and genes encodi	(355)	1339	1339	1356	1536.8	4.7e-78
	CAA75715 Potato virus S gene encoding re	(355)	1336	1336	1356	1536.8	4.7e-78
	CAA75713 Potato virus S gene encoding re	(355)	1328	1328	1348	1527.7	1.5e-77
	BAA01931 Garlic latent virus genomic RNA	(312)	1207	1207	1209	1371.0	8e-69
X	BAA03639 Apple stem grooving virus genom	(2105)	1294	438	1180	1325.4	2.8e-66
X	CAA53196 Grapevine virus B genomic RNA f	(1707)	1405	868	1083	1216.9	3.1e-60
X	CAA53182 Grapevine virus A genomic RNA,	(1707)	1351	908	1061	1192.0	7.5e-59
X	BAA03869 Citrus tatter leaf virus (CTLV)	(937)	983	439	1034	1165.4	2.3e-57
X	AAB20076 ORF1...ORF6 {CVB 3' terminal	(260)	987	987	1013	1150.1	1.6e-56
	BAA03351 Citrus tatter leaf virus RNA fo	(870)	892	348	940	1059.4	1.8e-51
X	BAA61806 Garlic virus A genomic RNA, com	(1625)	586	402	686	767.3	3.4e-35
X	BAA61818 Garlic virus C genomic RNA, com	(1550)	600	400	675	755.2	1.6e-34
X	BAA61812 Garlic virus B genomic RNA, 3'	(474)	544	424	664	750.6	2.9e-34
	AAC58809 Garlic virus B complete genome	(1543)	617	418	662	740.5	1.1e-33
X	AAA47787 Shallot virus X DNA sequence, o	(1718)	512	373	643	718.2	1.8e-32
X	BAA01042 Potato virus T genomic RNA, 3'-	(248)	563	563	618	702.8	1.3e-31
X	BAA02082 Strawberry mild yellow edge-ass	(1323)	578	199	624	698.5	2.3e-31
X	CAA39324 Potato Virus X complete genomic	(1456)	502	231	620	693.3	4.5e-31
	AAA47172 Potato virus X complete genome	(1456)	502	231	620	693.3	4.5e-31
	CAA80774 Potato virus X coat protein RNA	(1456)	507	231	620	693.3	4.5e-31
	AAA47178 Potato virus X 166K RNA, comple	(1456)	502	231	620	693.3	4.5e-31
	AAA03489 Potato virus X mRNA, complete c	(1456)	531	225	619	692.2	5.2e-31
	CAA51012 Potato Virus X strain HB RNA	(1456)	507	231	616	688.8	8e-31
	CAB57958 Lily latent virus mRNA for part	(171)	592	592	602	687.2	9.8e-31
	AAA47167 Potato virus X complete genome	(1456)	533	225	614	686.5	1.1e-30
	CAA28826 Potato virus X genome	(1456)	533	225	614	686.5	1.1e-30
	CAB70969 Poinsettia mosaic virus complet	(1987)	498	288	612	682.1	1.9e-30
	AAB32330 putative polymerase...capsid pr	(1646)	482	240	610	681.1	2.1e-30
	BAA00249 Potato virus X genomic RNA	(1456)	531	225	608	679.7	2.6e-30
	BAA03050 Papaya mosaic virus genomic RNA	(1547)	506	213	571	637.4	5.9e-28
	CAA34628 White clover mosaic virus total	(1294)	469	256	562	628.3	1.9e-27
	CAA29904 Potexvirus white clover mosaic	(1294)	476	257	560	626.1	2.5e-27
	AAC80555 Erysimum latent virus, complete	(1748)	547	312	551	613.9	1.2e-26
	CAA34415 Turnip yellow mosaic virus comp	(1844)	395	169	551	613.5	1.2e-26
	CAA30322 Turnip yellow mosaic virus geno	(1844)	395	169	549	611.3	1.7e-26

AAA46592	Turnip yellow mosaic virus Club	(1844)	385	161	548	610.1	1.9e-26
AAB92649	Turnip yellow mosaic Blue Lake	(1844)	388	164	547	609.0	2.2e-26
AAA46796	Ononis yellow mosaic virus comp	(1776)	603	289	544	605.8	3.3e-26
BAA02891	Narcissus mosaic virus genomic	(1643)	483	215	538	599.6	7.5e-26
CAA76071	Physalis mottle tymovirus OP, R	(1932)	514	279	529	588.3	3.2e-25
AAA43826	Foxtail mosaic virus, complete	(1335)	359	195	526	587.3	3.6e-25
CAA74273	Strawberry mild yellow edge-ass	(371)	493	196	518	586.8	3.8e-25
AAB71853	Cymbidium mosaic virus Korean t	(1417)	426	205	512	571.1	2.9e-24
AAF09240	Chayote mosaic tymovirus, compl	(1835)	616	272	512	569.4	3.6e-24
BAA00532	Kennedya yellow mosaic virus ge	(1874)	569	269	511	568.1	4.2e-24
AAB57879	Cymbidium mosaic virus, complet	(1417)	420	205	497	554.1	2.5e-23
BAA05033	Bamboo mosaic virus genomic RNA	(1365)	387	217	494	550.9	3.8e-23
AAB70562	Bamboo mosaic virus strain S co	(1365)	384	217	484	539.6	1.6e-22
CAA75810	Apple chlorotic leaf spot trich	(121)	457	457	457	525.2	1e-21
AAA43039	Eggplant mosaic virus genome	(1839)	496	274	461	511.5	6e-21
AAC54775	Cassava common mosaic virus Bra	(1449)	450	185	456	507.5	1e-20
AAC57874	Oat blue dwarf virus complete g	(2066)	522	281	456	505.1	1.4e-20
CAA79761	P.asiatica mosaic virus genomic	(1385)	423	210	452	503.2	1.7e-20
CAA75920	Potato trichovirus T gene encod	(121)	423	423	427	491.2	8.1e-20
CAA75814	Heracleum latent virus RdRpol g	(121)	390	390	402	462.8	3.1e-18
BAA74944	Garlic virus D genomic RNA, 3'	(285)	375	375	403	458.3	5.5e-18
CAA75813	Grapevine virus D RdRpol gene,	(121)	370	370	381	439.0	6.5e-17
AAC25015	Andean potato latent virus repl	(341)	284	193	360	408.3	3.3e-15
AAD51844	Aconitum latent virus replicati	(183)	318	278	349	400.0	9.7e-15
AAC40599	Hepatitis E virus strain US-1 n	(452)	145	85	270	304.5	2e-09
AAD15812	Hepatitis E virus strain HEV-US	(1698)	177	85	270	295.6	6.3e-09
AAC97208	Swine hepatitis E virus nonstru	(1708)	179	85	267	292.2	9.9e-09
AAA45733	Hepatitis E virus RNA-directed	(290)	151	83	256	291.6	1.1e-08
AAA03206	Hepatitis E virus polyprotein (	(727)	151	83	260	290.0	1.3e-08
AAC27934	Hepatitis E virus from Hyderaba	(1693)	216	83	264	288.9	1.5e-08
AAB82002	Hepatitis E virus nonstructural	(1693)	216	83	264	288.9	1.5e-08
CAA67802	Hepatitis E virus ORF1, ORF2 &	(1693)	210	82	263	287.7	1.7e-08
AAA03188	Hepatitis E virus ORF1, ORF2, O	(487)	151	83	255	287.0	1.9e-08
AAD15815	Hepatitis E virus strain HEV-US	(1708)	177	85	261	285.4	2.4e-08
AAA45725	Hepatitis E Virus, complete cds	(1692)	216	83	260	284.3	2.7e-08
BAA01865	Hepatitis E virus complete geno	(1693)	216	83	260	284.3	2.7e-08
AAA65488	Hepatitis E virus ORFs 1-3	(1693)	216	83	260	284.3	2.7e-08
AAA91078	Hepatitis E virus genes, comple	(1693)	216	83	260	284.3	2.7e-08
BAA01172	Hepatitis E virus genomic RNA	(1693)	216	83	260	284.3	2.7e-08
AAA45734	Hepatitis E structural and nons	(1693)	216	83	260	284.3	2.7e-08
CAA66936	Hepatitis E virus, complete gen	(1693)	216	83	260	284.3	2.7e-08
AAC97186	Hepatitis E virus strain TK15/9	(1693)	216	83	259	283.2	3.1e-08
AAA03189	Hepatitis E virus ORF1, ORF2, O	(1693)	216	83	255	278.7	5.6e-08
AAA96139	Hepataitis E virus complete gen	(1693)	183	83	255	278.7	5.6e-08

X: Used for phylogenetic analyses in Chapter 5

length: The size of the sequence in amino acids.

init1: The alignment score derived by FastA from the best initial alignment found between the query sequence and a database sequence in the first phase of a search.

initn: The alignment score calculated in the second phase of a FastA search, after the shorter alignments found in the first phase of the search are joined together.

opt: The alignment score derived from the 'optimised' sensitive alignment derived in the third phase of the FastA search

z sc: Normalised score derived from the opt score.

E (expectation value): Conversion of the z sc into a statistical expectation value which approximates the probability that a given match is a chance occurrence.

Figure 4.3a

Predicted peptide sequence for SCSMaV-ORF1 showing a nucleotide binding site motif (1121-1139) and eight motifs that are common to RdRPs (1641-1890) including the GDD at position 1808-1810.

Figure 4.3b

Alignment of SCSMaV, ASPV, GRSPaV, SCGRMV and BBScV RdRP amino acid sequences. Conserved regions are shown in black and non-conserved regions are shown grading black to white for no conserved motifs. The N and C terminal regions are the most conserved. Dashes represent the gaps inserted to maximise the alignment.

Figure 4.3a: SCSMaV-ORF1 peptide sequence

1	MAILTFKPPK	QLLIAQLSSE	EQARIRSSVT	RKLAQEEEND	QGLFSYHVSP
51	YCAEKLIKSG	VQLCAYSSDV	HSHPCKKMLE	NHILYRVLPS	YVDSDFILVG
101	IKHSKLAFLK	PRGAKKNVDL	STVKVLNRFV	TSLDRSRYGK	EFNHLNIMST
151	FTPEPKGSSS	YTPCLKSLIP	NLERFKGHKL	FFHDELHYWK	DESIKMLSS
201	LEPSVVYATV	VCPELLIGA	KQSYNDWLYT	FEIIGDQILF	APDGKMNSAY
251	FQPVNCARLL	RAKRIWLDKH	NYDIDLVS	NFSHLLCFS	KGSNLKNQDL
301	RAFSDFDAIK	LNYISELMVN	AESCFPISWT	VVKGLYDYLE	SLTKEDYRSG
351	KAKQTHYPN	CSGKQLKFT	DLIKLYLEAP	HIKTMIGSSP	VHNFKVLFTR
401	MFPACIKKNF	SFYKGKLLKS	LFEHLAPLAF	HVKTYDLHLG	FVFKFQELKN
451	VTDLMEDVHE	FEVDFVREYD	PFQMRSTPYV	GLAPLRDESY	YKRMKCGEAF
501	FDNMIGQIMR	SHLSGYGTNT	NPAELFLRIN	LTIVGRVEFL	VNGHHDFVIF
551	YINKHFVKVK	DPRYKNLYFA	NMTLELKLW	LSCRRFRKYQ	KFLPCGLVSL
601	AEVQILWRNI	ARSIQSGSHE	AQGQNEQVDS	HGDGVAIEEP	ERKGNTLEKL
651	NEATLVTHLQ	HYLKNPVVDD	FSRCAFVDVS	KTLSNLIDCL	KQICDTSINV
701	AIGCDCSNSD	LNLLKISKMA	SSVASVYNLD	NVKSVMKINWT	EAGEILVGLE
751	DQPLFVVVLG	KDINSITLAR	DHVMQFPELE	ENSAILIERN	NISKISVSGP
801	SALMIFPKDQ	NLLDCHLIGC	DQIPPAKIDF	IDGYIKTTVD	PSLDSRLTCI
851	NVPANGDCFW	HSVSLYLSVE	AKMIKDAPVL	RNRNPQNARL	VEPMGDKVWA
901	ENEAILNCAC	FYNLNIRIVF	EHELHTFRPH	DLNEDSTEVW	LKNLDHMHFC
951	PAKFINSVCVM	EPISKAVDIP	PLNVQRMIS	QESLHSEYER	LSSGKSFDDL
1001	GLESLMTFFN	IGAENVQENGH	KFILNENGSR	ARSFILENNH	IEFVRGDKKD
1051	LRKEGIGFVK	LMNANQSKSV	LSLACSKVKY	TPSYDRAFKL	SKAFANGYTG
1101	IMLSSEKFGE	IEINGKNDER	EINVMGTFG	AGKTRVFVDY	IQAYSGRGLF
1151	YVSPRKNLKE	LFDNSAIQAV	QGFKTKSSDK	KGSVKSFHSS	FTFETAIMKA
1201	GQIGDSDILI	IDEIQLYPPG	YLDLILLIP	SSCMVICAGD	PVQSDYDSGD
1251	DRLYFLNDKP	DIERLLEGQS	YKYAIKSRRF	KSSLFIGRLP	CDMPGLQGMG
1301	MDTYGIYESI	SSVPQDLIKR	SFFLVSSFDE	KKIIKANFKE	THNKILTFGE
1351	STGSTFQEI	ILITTSSEKA	SERRWVTALS	RASTSISFLN	LTGIGIQELI
1401	STTYFGRFLQ	KFLSAECKRE	DIKPYLYGEP	KFVDDFRSIG	RRSAEEKEVK
1451	VDGDPWLKCI	LHFFETPIPD	EVECVKQDMV	EQKAKTHLPV	SSFEVLWAHF
1501	NEKLLLKEIR	ERRIDDLVSD	QFAWDHRSQG	LAINNAATRY	ESIFPRHHHN
1551	DKVTFLMAVK	KRLTFGSPAK	NGRMLREAMP	YGEILLKEFL	KRVPIKNDRN
1601	DHLFLECLMD	FEEKKTSKDS	ATIANHCIRS	QNSWPIDIAD	IFMKSQ ^I LCTK
1651	MESMFCEAKA	GQTLACFHHV	VLARFSPWVR	YIEKKLVSSL	PDNYYIHSGK ^{III}
1701	NFDQLNAWVK	ANNFVGECTE	SDYEA ^{II} F ^{IV} DAQ	DHYILAFELA	LMKYLGLPNE
1751	LIEDYKYIKI	HLGSKLGDA	IMRFTGEAST	FLFNTMANML	FTFLVYDIKG
1801	NESICFAGDD	MYANTSLKTC	SKNKSLLKKE	TLKAKVQVTE	SPTFCGWHLT ^{VII}
1851	EDGIYKDPIL	VSFRFMIALE	KGILHECLES	YSIECSYAYN	MGS ^{VIII} AVSRHMD
1901	QLSLAYHYCC	VRFILKNSKL	LRSDVKKFFK	DGLQLTQ*	

Figure 4.3b: Alignment of RdRP amino acid sequences

SCSMaV	:	MAITTPPPKQLTQLQEEAATRSSVTRKLAQEEEDOGSSVHVSPYCAEILKSGOLCAISDVHSHFCRMLEHILRLDPYVES	:	94
ASPV	:	MLLSATAAGVIVSPEEASTOSTOVLALTNVKPKHDLNVALEELAMFNSTLTHYRKHSHVCKTLENLILNLPYLDN	:	93
GRSPaV	:	LSYPAYSEVILKQDAGVSATKALVDLEESQHNHSAALDRSELISSCVLSEYFRHSHFCKTLEHILNLPYLDN	:	93
SCGRMV	:	HTITPASEVLOSEEAAGASSNFSKLEAHNHSRHNLYAYASLSERCFLLPYETHSHVCKTLEHILNLPYLDN	:	93
BL5V	:	HTTYSPEVSEVITLLEPTALLASAAAFORHKKNFEMRPSVLAFAEHSKAGTILSEAGFHSHPCKTLEHILNLPYLDN	:	93
SCSMaV	:	DLILVCIKHSKALPRGAFNVDISTYKVNKRVTSLSRYGKEENHLNIMSTFTPEPK--GSSSYIPSTKSLINERFKGH----	:	182
ASPV	:	EFVLSTIKKAVDLELR---SHPLQMYETINRISSIDKTRYGCFHVSPSKISAKFKCDRTGFEDDASIDIGFCMGEARK----	:	179
GRSPaV	:	EFVLSTIKKAVDLELR---SHPLQMYETINRISSIDKTRYGCFHVSPSKISAKFKCDRTGFEDDASIDIGFCMGEARK----	:	177
SCGRMV	:	DLILVCIKHSKALPRGAFNVDISTYKVNKRVTSLSRYGKEENHLNIMSTFTPEPK--GSSSYIPSTKSLINERFKGH----	:	179
BL5V	:	EFVLSTIKKAVDLELR---SHPLQMYETINRISSIDKTRYGCFHVSPSKISAKFKCDRTGFEDDASIDIGFCMGEARK----	:	175
SCSMaV	:	HDELHYMKDESIKMSLELSVYATVVCHELLGAKOSYNDWLTBEELDQILFAPDCMNSAYFOPNCARLLAKRWLDKHNYS	:	276
ASPV	:	HDELHYMKDESIKMSLELSVYATVVCHELLGAKOSYNDWLTBEELDQILFAPDCMNSAYFOPNCARLLAKRWLDKHNYS	:	273
GRSPaV	:	HDELHYMKDESIKMSLELSVYATVVCHELLGAKOSYNDWLTBEELDQILFAPDCMNSAYFOPNCARLLAKRWLDKHNYS	:	271
SCGRMV	:	HDELHYMKDESIKMSLELSVYATVVCHELLGAKOSYNDWLTBEELDQILFAPDCMNSAYFOPNCARLLAKRWLDKHNYS	:	272
BL5V	:	HDELHYMKDESIKMSLELSVYATVVCHELLGAKOSYNDWLTBEELDQILFAPDCMNSAYFOPNCARLLAKRWLDKHNYS	:	269
SCSMaV	:	LYENSHHLCPCKGSNKNQDLRASSDPAKKNYISEMVNAESCPSWITVKGLYDYLSLIRBYRSGHAKOTHYPCNCSGLKFKCT	:	370
ASPV	:	LYENSHHLCPCKGSNKNQDLRASSDPAKKNYISEMVNAESCPSWITVKGLYDYLSLIRBYRSGHAKOTHYPCNCSGLKFKCT	:	366
GRSPaV	:	LYENSHHLCPCKGSNKNQDLRASSDPAKKNYISEMVNAESCPSWITVKGLYDYLSLIRBYRSGHAKOTHYPCNCSGLKFKCT	:	365
SCGRMV	:	LYENSHHLCPCKGSNKNQDLRASSDPAKKNYISEMVNAESCPSWITVKGLYDYLSLIRBYRSGHAKOTHYPCNCSGLKFKCT	:	365
BL5V	:	LYENSHHLCPCKGSNKNQDLRASSDPAKKNYISEMVNAESCPSWITVKGLYDYLSLIRBYRSGHAKOTHYPCNCSGLKFKCT	:	362
SCSMaV	:	DLIKYVYAPHIKHIGSSPVHNFVLTMPFACKNFSPYGLKLSLFEHAPLAHRYTYDHLG-----PVFKFQELKNVTDMEDVH	:	459
ASPV	:	DLIKYVYAPHIKHIGSSPVHNFVLTMPFACKNFSPYGLKLSLFEHAPLAHRYTYDHLG-----PVFKFQELKNVTDMEDVH	:	460
GRSPaV	:	DLIKYVYAPHIKHIGSSPVHNFVLTMPFACKNFSPYGLKLSLFEHAPLAHRYTYDHLG-----PVFKFQELKNVTDMEDVH	:	459
SCGRMV	:	DLIKYVYAPHIKHIGSSPVHNFVLTMPFACKNFSPYGLKLSLFEHAPLAHRYTYDHLG-----PVFKFQELKNVTDMEDVH	:	459
BL5V	:	DLIKYVYAPHIKHIGSSPVHNFVLTMPFACKNFSPYGLKLSLFEHAPLAHRYTYDHLG-----PVFKFQELKNVTDMEDVH	:	456
SCSMaV	:	EFEVDFVREYD-----FQMRSTPLGLAPLRDESYYKPKCEAFFDNMGOIMRS---HLSGYGTNTNPAELFLRNTIVGVVEFLVNG	:	543
ASPV	:	EANGTK-KITVNV---ITTVHHSPLYSKFESYDHOFSHISVKSISAIRAKIVLSLYDPCVVEAFSESRTVNLAVNIIAANLR---ACFA	:	546
GRSPaV	:	NSWFGL-GDQFNQRAPFFLGSYSNSKFSVEHKFSGTNSCIMQVLSIPFSD---DPTFRPSSTEVNLALSEVKAALZATGQSKLPRL	:	549
SCGRMV	:	KKSSGLQLNLT-----VATPCLVGGCCCVFTGIHVNLTGHFVKSTNANWVGCNGLGYEALIALAASHQRG---KRTFLHDEFMKLAIL	:	545
BL5V	:	ERWEGK-QSADR-----TEAP-----GLAPFVDYKIOWRFQFNIPKFNQIAELFVN---SCSVNGG---VRMSIPAYRRLATCS---CVGR	:	533
SCSMaV	:	HHDFVIFYINKHFKVKDPRYKNLYFANMTLELRLSCSR-RFKYQPCGLVSLAEVQILWRNIARSQSQS---HEAQGQNEQDSHG--	:	632
ASPV	:	VFDLWRIFFGILLCECK---RAQCKMRKRFFHFLGRLLFV---DVSNQWPPCRDGLIARSVFXQFIKGCQDNLSHNGRMSLRQ/LKG--	:	633
GRSPaV	:	VDDCVHREVRSYSVGL---FKHIALTHCFNSCGQLLLR---QSNLPLKNRASSFADLCEVIRKAYQFVISQVILPEALLSLTKFVKDSD	:	639
SCGRMV	:	ANLVNSLAFKMALVEV---TRRLRRRN---AIRGLRDRLESEPED-KRERAVTLNFSRVVSEVLELATECPVTSLEATHYITKLGFPR---	:	630
BL5V	:	AMLCCLETVDIASLRVV---VRNRYPYTEDFYRCRRRLRGAQRPPSYEDAKHLERLGQFEEQFORPMRSRSLYTLASVSMNGTDDP---	:	622
SCSMaV	:	-----DCVAIEEPERKGNTEKLNEA-----TLVTHLQHYLKNPVVDDFSRCAF---DVDSKTSLNLDCLKQCDTSIIVAGCD	:	705
ASPV	:	-----PKLQALFDVSELSIHNVENENAP-----EAGSTLDAGIKPTSSPLEVVP-IE---NARCNLAPCKCDLNCFOPADVSLHGNL	:	709
GRSPaV	:	SKGVSIPRLASRNHGLAHFANSALDEPQSVDCNAGRVQASDSSQQLADTHSLSSVKSIE---TASKAFNLEEIRIMIRLPEDFVWVKNI	:	731
SCGRMV	:	-----SNGMVAIGASEEGSVQELKIPFHPEN-----LQTEICVDEKVKGERAPEVTEKEPPIGDENLHRSFSSIKNGEFKDALRKA	:	713
BL5V	:	-----PCSDCF-YDVPVVARAKIVPTET-----VIVERALEPLAIDGTGTSADPCDAP---GATCLRGAQAQVCACGSMAVSAPVYAE	:	698
SCSMaV	:	CNSNDLNLKISKMASSVSY---VNLNVKSVKIMTEAGEILGLEDO---PLVVVVGKDIN---SITLAPDHVMGEP-----ELEE	:	781
ASPV	:	VFLDFIGGS---KGGS---DLKGSTVGFHVSRVAFPLDKPSDNKIPLNFYQCVLESTGHGQSMKDSISYDINHO/LNYS---	:	800
GRSPaV	:	GFKDLRIG---GSS---PGISCSGCGHTSLKAFMDQISSTGG-RNYVSCVLIENSKALKDSQVEIGHKL/NITC---	:	819
SCGRMV	:	SGDTRVYKHKGT/L/S---GSFSG---SVYPSQQAFFEEAYG---SR---SCVLIENSKALKDSQVEIGHKL/NITC---	:	800
BL5V	:	KMDFYPDAL---KGDIA---EDREKLGCGHLCTAKWQLQVQANGV-DETCMLARGAQKNGFALNIEEMRGAP/HLS/D---	:	788
SCSMaV	:	NSAILIER-NNISKISVS-----S-----ALMIFKDONLLDCLIGCDOIPPAKLLSIDGYIKT-----	:	837
ASPV	:	DLFCIECLGSGFEIPLS-----QMLLGFYKEHRECKSPKGLHLLSLTEKEDSQVPIQEVVITICDHGSDDDRA-----	:	877
GRSPaV	:	STFTISSRNLVEGNHCSLTI---N-----EPFEPRMCMNYFSGSNCPGRVSLTFRQKLED-----DD-----	:	882
SCGRMV	:	EACIAFRVONVEVLEQVNETSQENYVEVKLSDGFFLLRFPQGFQPSRYANNNVITLFLQSRDLN-----	:	874
BL5V	:	NDFGTCEAAGROYTTLR---N-----VQFTLSESTETHRAVRNTAGVSYLFRALKKD-----	:	844
SCSMaV	:	-----	:	971
ASPV	:	ALKALERRSHQSGRPAVELEGHEREKVNSSDSDSAPVQEFLIQIDSSLLLEYALKSLSGLSKNVNVNDCMLCNSPWLKNEELRFSEALRDLAGA	:	951
GRSPaV	:	-LIFIN-----FOVPIELN-----HEKLDKSMQMLGHLGKKSISMNGTSFTSDLCSCFSCHNFHFKFDLINNRALGA	:	951
SCGRMV	:	-----E-----	:	845
SCSMaV	:	-----	:	845
ASPV	:	QGLIQIDFLCLKVLR-CAEVNRIHSELPTHVFLRGTMHIVDLDDRESIRGDVKEGFSFGFR-RWKVMSCSTDLMILAFKPKMTLGQELRSHE	:	1063
GRSPaV	:	QGLGQCDRVVFATTPGLSKVLEMPRSKQKSLVLEB-----ALSTETDYGPVKLVGSEVFKGDFQTKMMEGSIFFVITYKPIRSTGRVRVHS	:	1040
SCGRMV	:	-----G-----KPMQCGPIQDSAKKK-----	:	900
BL5V	:	-----S-RVIEEVEVETKDMGFSS-----	:	874
SCSMaV	:	-----	:	905
ASPV	:	DETELSDLTEKLHGCSVLSRKFPDLPHSPDVEACG-NCPFWHVGPLTGDEYLRRIHQDAKKD-----GVKCPRSKLGNTWARAV	:	1151
GRSPaV	:	SESPSGSKVLLGCGIEACADYDIDDPNTPSVPGCG-NCPFWHVGPLTGDEYLRRIHQDAKKD-----GVKCPRSKLGNTWARAV	:	1133
SCGRMV	:	NGCSISKSKGP-----DICLSLVFPKADG-DCPFWHVGPLTGDEYLRRIHQDAKKD-----GVKCPRSKLGNTWARAV	:	980
BL5V	:	EPDGVET-----ENVOCVPGDG-NCPFWHVGPLTGDEYLRRIHQDAKKD-----GVKCPRSKLGNTWARAV	:	943

Figure 4.3b: Alignment of RdRP amino acid sequences

SCSMaV	: LNCACFYNNTRIVFE---HELHTRPHDLNED---STEVMKNDLHHEAKFINSEVVEHISKVDIPPLVORSSQESLHSEYORSS	: 993
ASPV	: AYFCSHYGRRNVLVT---RESCTWIKPHEVLK---AATICQDNHKKCMVNGCVIRHESLNRRETVVAUGKP-AHEDLFIKVAE	: 1236
GRSPaV	: ALFIRHHVPLIVITP---EYFVSWKFGEGEWPL---CGTCLKSNHFCALNCGMITYISFLGRRETVVNYCRP-STNHIFLQCO	: 1218
SCGRMV	: TATFLMNRHLLIKLVGADDEGFTVEPLTESDKISISGYL---NORCHHDLAVKEGCVIRVVEFLKQNPTKLSAN-CSKELLALMS	: 1073
BISV	: CAASAHLGVDIVHDE---DQGFKNLYRYPGNKR---EAPRLKGSHEPLLEKEMCVVKTAQVKKRSPVIRVAAKK-MGEDFKKPCR	: 1028
SCSMaV	: GKSFDLGEESMTFETIGTQEQENGHKTILNENSRARSITENNHEFVRGDKDLRKEGIGFVKLIANQGG---SVSLACSKVRYTPSY	: 1084
ASPV	: GRFSPFDTRFPEFSICGSVDITGGELIMVRENRIAPESPEKEHIAHPTLSRRKFSPIVSDLNRSNSAMR---FLAKNGAE---VDYRPSI	: 1326
GRSPaV	: GGLNMYIAEAFEARDFCKDINGIEVINPHKISALDDITNEHVRHVEKIGNGPSIKVDELRRKRSALD---FLSSNGSK---ITFPPSP	: 1308
SCGRMV	: GLTQEPHEEFPRVPEHKECDGSLRLINKKCSREARSAVENDIFSYCPGVKAATNLGAFKQPHGNSIIPEDRYNDPDKSNANIVPEPSL	: 1167
BISV	: SKVMVDVFMVAKIFVSCVLOGTEQIMINPKRIKIGLRITTTIESYDGVDPKVKHSEVNVYKHDALQIED---LIEIRELSSIIVETPSF	: 1120
SCSMaV	: DIAFKSKAFANGYTGIMLSSEKFGSEIEING---KNDEBENVMGMFTGAGKTRVDYIQ---AYSFGFVSPRNLKELFNSAIQAVQ	: 1171
ASPV	: DIASTLUSFEICATGVLCOGIKEAQDLASKLIPELVHEKIMILGTFCGKSSKKKPIE---KSPAAITVSPRSLAESINHDGLARV	: 1418
GRSPaV	: ELAEKLGCLLGLAGVISDEKFSDAKFWLSGISTTDTKPPRTVVVLTGFCAGKSPKSPKK---RSETPVTVSPRNLANSIKNDIMDD	: 1400
SCGRMV	: SIAKILANSFLSSQGVINSRIVAGQFWLADTNKLCFDEKIGVMVGTFCGSKSHSVVELIK---ANLNYONILSPRSLKDOIGMDLVSK	: 1259
BISV	: SIAKILADCLHDSGVMSSELVNDKGLCP---EGRETRVTIGVLLTFCGCKSKKEIF---KLCFSSCISPRALCDSFDEIRKARG	: 1210
SCSMaV	: GPKTSSDKGSVKSFSHSTFETATMKAGTGDDSIIDEOLPPGYDLILIPSCMIIAGDFVOSYDDEGDFLYLNDKPDIPIL	: 1265
ASPV	: GGRVTKQKSD---LKNVRVETPELFIHLHDSKEGHTVVIDEOLPPGYDLILGLKPNVMIIAGDFVOSYDDEGDFLYLNDKPDIPIL	: 1509
GRSPaV	: CKVARAGRS---KEGWDVVTFFVPRKAVAGKAGHGVVIDEOLPPGYDLILIRSDAFISLAGDFVOSYDDEGDFLYLNDKPDIPIL	: 1491
SCGRMV	: R---KKG---ISTEVATFEVALKKNGLVKKTRIFNDTOLPPGYDLILVAGRDSSELVMDPAOSSYDDEGDFLYLNDKPDIPIL	: 1344
BISV	: N---MGE---RG---IKHYKSTFEKAILQASKHKGSLVIDEOLPPGYDLILVAGPTMKYFALGDFVOSYDDEGDFLYLNDKPDIPIL	: 1297
SCSMaV	: LQTSYVAKSKIRKSSLEIRLPQDMPGLOGMGMDTYCTESSSVPQDLIKR---SFFLVSSFEKTIIRNFKETHNKLTFGESTGTF	: 1356
ASPV	: LQTSYVAKSKIRKSSLEIRLPQDMPGLOGMGMDTYCTESSSVPQDLIKR---SFFLVSSFEKTIIRNFKETHNKLTFGESTGTF	: 1601
GRSPaV	: LQTSYVAKSKIRKSSLEIRLPQDMPGLOGMGMDTYCTESSSVPQDLIKR---SFFLVSSFEKTIIRNFKETHNKLTFGESTGTF	: 1584
SCGRMV	: LQTSYVAKSKIRKSSLEIRLPQDMPGLOGMGMDTYCTESSSVPQDLIKR---SFFLVSSFEKTIIRNFKETHNKLTFGESTGTF	: 1437
BISV	: LQTSYVAKSKIRKSSLEIRLPQDMPGLOGMGMDTYCTESSSVPQDLIKR---SFFLVSSFEKTIIRNFKETHNKLTFGESTGTF	: 1387
SCSMaV	: QEIVLLITTEKASERWVALSRASTSHSLNLTIGIQLISTTYFQFHOELAECKREITPYLYCEKVVDEERSIGR-SRAEEKEV	: 1449
ASPV	: QKGAILLVYESALTSDFRMVALSRFESHKINGMVTWNAI---HGVNPPHFFKRAKNDLIDILGRELLEGOSQVG-AEGVREA	: 1693
GRSPaV	: KNGGILSHSEFHTDDRMVALSRFESHLDLVNITPRVESFL---HGAQIPYHPLAKSGENVITDILGEBNFGSNVSG-KAGVREE	: 1676
SCGRMV	: DVVCVLITQESMLVDERMVALSRARITSEVNLSSLSQEFCCQMVGGIHFPTSTANYNGLAILGEBNFGSNVSG-KAGVREE	: 1528
BISV	: DVGCVLITSTSAHTEORVALSRPERKIVVNAASAWDTIQSVANMGGELNRSARQESIRLILGTLVBEGBQKNLGLAEGREK	: 1480
SCSMaV	: KVGDPWLKCIHFETPI-PDEVCKVDMVEQKAKHETSSPVLMAHNEKLUKEIRERRDDASDQAWDHR-SGLAINAATPYE	: 1541
ASPV	: KVGDPWLKCIHFETPI-PDEVCKVDMVEQKAKHETSSPVLMAHNEKLUKEIRERRDDASDQAWDHR-SGLAINAATPYE	: 1786
GRSPaV	: KVGDPWLKCIHFETPI-PDEVCKVDMVEQKAKHETSSPVLMAHNEKLUKEIRERRDDASDQAWDHR-SGLAINAATPYE	: 1769
SCGRMV	: KVGDPWLKCIHFETPI-PDEVCKVDMVEQKAKHETSSPVLMAHNEKLUKEIRERRDDASDQAWDHR-SGLAINAATPYE	: 1621
BISV	: KVGDPWLKCIHFETPI-PDEVCKVDMVEQKAKHETSSPVLMAHNEKLUKEIRERRDDASDQAWDHR-SGLAINAATPYE	: 1573
SCSMaV	: STPEPRHHNRVYFLMAKKRLTFGSLAKNGRM---BAPYCEILKEFLRVIRKNDNDHFLCUMDFEKKTSRSDATIANECIRSONSP	: 1635
ASPV	: AITYPRHGTDAFLMAKKRLTFGSLAKNGRM---BAPYCEILKEFLRVIRKNDNDHFLCUMDFEKKTSRSDATIANECIRSONSP	: 1880
GRSPaV	: AITYPRHGTDAFLMAKKRLTFGSLAKNGRM---BAPYCEILKEFLRVIRKNDNDHFLCUMDFEKKTSRSDATIANECIRSONSP	: 1863
SCGRMV	: AITYPRHGTDAFLMAKKRLTFGSLAKNGRM---BAPYCEILKEFLRVIRKNDNDHFLCUMDFEKKTSRSDATIANECIRSONSP	: 1715
BISV	: AITYPRHGTDAFLMAKKRLTFGSLAKNGRM---BAPYCEILKEFLRVIRKNDNDHFLCUMDFEKKTSRSDATIANECIRSONSP	: 1667
SCSMaV	: LILIFKSCQCTKMDLTAAKAOTLCEHVLARSPWRYIEKNEVSLDLYIHSGKNFLINAWKANNVCGCTESDYEAFLDS	: 1729
ASPV	: LILIFKSCQCTKMDLTAAKAOTLCEHVLARSPWRYIEKNEVSLDLYIHSGKNFLINAWKANNVCGCTESDYEAFLDS	: 1974
GRSPaV	: LILIFKSCQCTKMDLTAAKAOTLCEHVLARSPWRYIEKNEVSLDLYIHSGKNFLINAWKANNVCGCTESDYEAFLDS	: 1957
SCGRMV	: LILIFKSCQCTKMDLTAAKAOTLCEHVLARSPWRYIEKNEVSLDLYIHSGKNFLINAWKANNVCGCTESDYEAFLDS	: 1809
BISV	: LILIFKSCQCTKMDLTAAKAOTLCEHVLARSPWRYIEKNEVSLDLYIHSGKNFLINAWKANNVCGCTESDYEAFLDS	: 1761
SCSMaV	: ODPIILAFELALMKYLLPFLNELSDYKIRILASKLGFIMRFTGEASTFLFNTHAMMLETFLVYINQESIFAGDDMYANTSITCCKN	: 1823
ASPV	: ODPIILAFELALMKYLLPFLNELSDYKIRILASKLGFIMRFTGEASTFLFNTHAMMLETFLVYINQESIFAGDDMYANTSITCCKN	: 2068
GRSPaV	: ODPIILAFELALMKYLLPFLNELSDYKIRILASKLGFIMRFTGEASTFLFNTHAMMLETFLVYINQESIFAGDDMYANTSITCCKN	: 2051
SCGRMV	: ODPIILAFELALMKYLLPFLNELSDYKIRILASKLGFIMRFTGEASTFLFNTHAMMLETFLVYINQESIFAGDDMYANTSITCCKN	: 1903
BISV	: ODPIILAFELALMKYLLPFLNELSDYKIRILASKLGFIMRFTGEASTFLFNTHAMMLETFLVYINQESIFAGDDMYANTSITCCKN	: 1855
SCSMaV	: KSLRNFLKAKVYVIESPTFCGWHITEDGIRLHISFPMIALKSGIHCELESYSTECAYMMCSAVSRHMDLSLAYVCCVRFILKN	: 1917
ASPV	: SNRLDILKAKVYVIESPTFCGWHITEDGIRLHISFPMIALKSGIHCELESYSTECAYMMCSAVSRHMDLSLAYVCCVRFILKN	: 2162
GRSPaV	: EGRANMLKAKVYVIESPTFCGWHITEDGIRLHISFPMIALKSGIHCELESYSTECAYMMCSAVSRHMDLSLAYVCCVRFILKN	: 2145
SCGRMV	: EHILDILKAKVYVIESPTFCGWHITEDGIRLHISFPMIALKSGIHCELESYSTECAYMMCSAVSRHMDLSLAYVCCVRFILKN	
BISV	: KNPMSILKAKVYVIESPTFCGWHITEDGIRLHISFPMIALKSGIHCELESYSTECAYMMCSAVSRHMDLSLAYVCCVRFILKN	
SCSMaV	: SKLLRSPKKFKD---GLQUTQ---	: 1937
ASPV	: NHLLVIRDLAK---GESLPASS---	: 2183
GRSPaV	: KHKRCSISGLGE---AID---	: 2161
SCGRMV	: KQTFHAIKFKKEGEGCISPRNFS---	: 2023
BISV	: KHLLEFATVQ---TARVD---	: 1967

between known RdRP amino acid sequences and the presence of motifs consistent with RdRPs led to the conclusion that ORF1 encodes the RdRP.

It is predicted that species in the genera *Potexvirus* and *Foveavirus* express the ORF1 directly from the genomic RNA (van Regenmortel *et al.*, 2000). Therefore the context of the ORF1 initiation codon is important for determining the potential translatability of the protein. The context of the SCSMaV ORF1 initiation codon is TTGAUGGC (Figure 4.2a). This is identical in the +1 nt and +2 nt to the plant mRNA consensus sequence of AACAUGGC (Lütcke *et al.*, 1987), and has been reported in some species in the genus *Potexvirus* (Bancroft *et al.*, 1991) and species in the genus *Capillovirus* (Yoshikawa *et al.*, 1992). The -3 nt, -2 nt, and -1 nt of the SCSMaV ORF1 initiation codon region are complementary to the consensus sequence: i.e. the positions of pyrimidines and purines are conserved. This conservation may be important in the translation of the SCSMaV RdRP.

4.4.2 ORF2, ORF3 and ORF4

4.4.2.1 ORF2

ORF2 spans nt 5919-6539 and encodes a peptide of 206 amino acids with a predicted M_r of 23000 (Table 4.1). The predicted amino acid sequence of the ORF2 peptide is shown in Figure 4.4. Blast and FastA genome database comparisons show up to 29.7% amino acid sequence similarity with the ORF2 protein (TGBp1) of species in the genera *Carlavirus*, *Foveavirus* and *Potexvirus*, which comprise part of the TGB encoding movement proteins (Table 4.3; Appendix D). A search using the PROSITE dictionary found an ATP/GTP binding site motif at amino acid positions 23-40 (see Figure 4.4).

Table 4.3 FastA database search results for ORF 2 peptide sequence (see Table 4.2 for legend)

The best scores are:		length	initn	init1	opt	z-sc	E(467554)
BAA04854	Apple stem pitting virus (ASPV)	(223)	354	238	373	457.8	5.1e-18
BAA00351	Potato virus S genomic RNA, 3'-	(227)	298	210	327	402.6	6e-15
CAB57959	Lily latent virus mRNA for part	(228)	312	214	324	399.0	9.6e-15
CAA33398	Lily symptomless virus RNA 3-te	(228)	286	193	306	377.4	1.5e-13
CAA75717	Potato virus S and genes encodi	(227)	272	196	304	375.0	2.1e-13
AAC62911	Grapevine Rupestris stem pittin	(221)	232	158	302	372.8	2.7e-13
AAC35499	Rupestris stem pitting associat	(221)	232	158	301	371.6	3.2e-13
AAA85900	Lily symptomless virus 25 kDa p	(228)	281	193	296	365.4	7.1e-13
CAA37233	Potato virus M complete genome	(229)	235	171	270	334.3	3.8e-11
BAA03340	Potato virus M genomic RNA, com	(229)	235	171	270	334.3	3.8e-11
AAC35434	Sour cherry green ring mottle v	(222)	200	151	262	324.9	1.3e-10
BAA03051	Papaya mosaic virus genomic RNA	(233)	181	142	259	321.0	2.1e-10
CAA33394	Lily virus X RNA 3' terminal re	(216)	202	169	256	317.9	3.1e-10
CAA34629	White clover mosaic virus total	(236)	243	178	255	316.1	3.9e-10
AAB20077	ORF1...ORF6 {3' terminal reigon	(231)	224	162	249	309.1	9.7e-10
CAA29905	Potexvirus white clover mosaic	(236)	248	177	248	307.7	1.2e-09
CAA79762	P.asiatica mosaic virus genomic	(230)	215	190	245	304.3	1.8e-09
AAB81274	Potato virus M strain Idaho 25K	(217)	201	132	243	302.3	2.3e-09
AAB26348	triple block genes: putative he	(230)	202	177	232	288.8	1.3e-08
AAC54776	Cassava common mosaic virus Bra	(231)	169	169	232	288.7	1.3e-08

1	MEIVCRYLDD	FKFERTSLEL	SSPIVVHAVP	GSGKSTLIRK	CITENPSLTA
51	CTFGKPDLPN	LTGTIYIKGYK	EGEKYDIIDE	YIGNRFECVP	VALFSDPFQN
101	CEQYINTAHY	IGRTSHRFGV	VLLSYCAGLV	IKFPHLAKTR	SLRVQVTLLN
151	LQVLSYALSL	TLRTIYLHIT	SSLFIPWTS	ERLLTKLHSI	ILVISAILLV
201	LSYIFR				

Figure 4.4

Predicted peptide sequence for SCSMaV-ORF2 showing the predicted nucleotide binding site motif at position 23-40 represented by a horizontal line above the sequence.

4.4.2.2 ORF3

ORF3 was predicted to span nt 6552-6914 and to encode a peptide of 120 amino acids with a predicted M_r of 13000 (Table 4.1). The predicted amino acid sequence of the ORF3 peptide is shown in Figure 4.5a. Blast and FastA genome database comparisons show amino acid sequence similarity of up to 27.5% to the ORF3 protein (TGBp2) of species in the genera *Foveavirus*, *Potexvirus* and *Carlavirus* that comprise part of the TGB encoding movement proteins, and the putative movement protein encoding ORF of species in the genus *Trichovirus* (Table 4.4; Appendix D). Protein motif searches using the PROSITE dictionary did not find any consensus motifs in the peptide sequence. Hydropathy plots were compared for ORF3 peptide sequences of SCSMaV, *Bamboo mosaic potexvirus* (BaMV), *Cassava common mosaic potexvirus* (CsCMV) and ASPV. Comparison of these plots showed that they had similar trends, however SCSMaV did not appear to have a hydrophilic leader sequence (Figure 4.5b). Analysis of the nucleotide sequence revealed a second potential start codon at position 6576 that gave a peptide sequence of 112 amino acids. A hydropathy plot showed that this peptide had a hydrophilic leader sequence, which is predicted for membrane spanning proteins. ORF3 is therefore predicted to span nt 6576-6912 and to encode TGBp2, a peptide of 112 amino acids with a predicted M_r of 12600 (Table 4.1).

4.4.2.3 ORF4

ORF4 spans nt 6902-7105 and encodes a peptide of 67 amino acids with a predicted M_r of 7000 (Table 4.1). The predicted amino acid sequence of ORF4 peptide is shown in Figure 4.6a. FastA and Blast genome database comparisons show amino acid sequence similarity of up to 22.5% with the ORF4 proteins (TGBp3) of species in the genera *Potexvirus*, *Carlavirus* and *Foveavirus* (Table 4.5; Appendix D) which comprise part of

Table 4.4 FastA database search results for ORF 3 peptide sequence (see Table 4.2 for legend)

The best scores are:		length	initn	init1	opt	z-sc	E(467541)
AAB70564	Bamboo mosaic virus strain S co	(119)	137	137	204	271.5	1.2e-07
BAA05036	Bamboo mosaic virus genomic RNA	(119)	137	137	202	269.0	1.7e-07
AAC54777	Cassava common mosaic virus Bra	(112)	116	116	181	243.1	4.6e-06
BAA61814	Garlic virus B genomic RNA, 3'	(108)	145	145	175	235.9	1.2e-05
AAC58811	Garlic virus B complete genome	(106)	128	128	166	224.8	4.8e-05
BAA03052	Papaya mosaic virus genomic RNA	(111)	113	113	166	224.5	5e-05
BAA61808	Garlic virus A genomic RNA, com	(103)	133	96	164	222.5	6.5e-05
AAA47789	Shallot virus X DNA sequence, o	(103)	129	99	159	216.2	0.00014
CAA33395	Lily virus X RNA 3' terminal re	(108)	87	87	159	215.9	0.00015
AAA43828	Foxtail mosaic virus, complete	(105)	147	147	157	213.6	0.0002
AAB26349	triple block genes: putative he	(110)	135	111	155	210.8	0.00029
CAA79763	P.asiatica mosaic virus genomic	(110)	135	111	155	210.8	0.00029
CAA29906	Potexvirus white clover mosaic	(116)	108	108	151	205.4	0.00058
AAA69636	White clover mosaic potexvirus	(116)	108	108	151	205.4	0.00058
CAA44531	Cymbidium mosaic virus RNA for	(112)	66	66	150	204.4	0.00066
AAB57881	Cymbidium mosaic virus, complet	(112)	66	66	150	204.4	0.00066
AAB71855	Cymbidium mosaic virus Korean t	(112)	66	66	147	200.7	0.0011
CAA34630	White clover mosaic virus total	(117)	120	120	147	200.4	0.0011
AAB81270	Potato virus M strain Idaho 25K	(109)	108	108	146	199.6	0.0012
BAA01933	Garlic latent virus genomic RNA	(110)	115	115	143	195.8	0.002

Figure 4.5a

Predicted peptide sequence for SCSMaV-ORF3 showing the two potential start codons at amino acid position 1 (nt position 6552; blue) and amino acid position 9 (nt position 6576; red).

Figure 4.5b

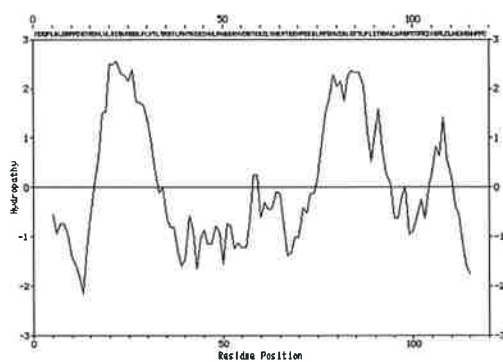
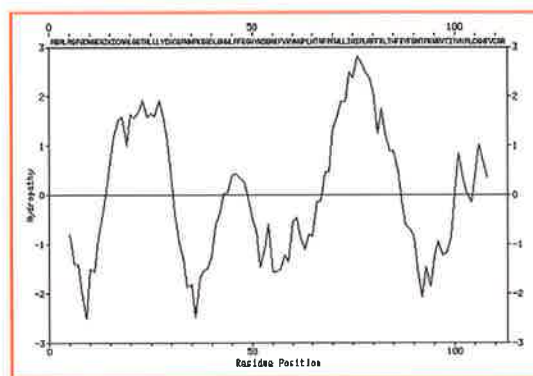
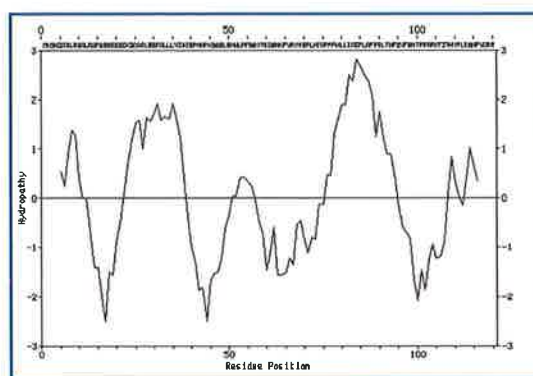
Hydropathy plots of SCSMaV-ORF3 using peptide translated from position 1 (blue) and position 9 (red) compared to the hydropathy plots of similar viruses with TGBs (BaMV, CsCMV and ASPV). A hydrophilic leader sequence is present for BaMV, CsCMV, ASPV and the SCSMaV peptide translated from the second (red) start codon.

(a)

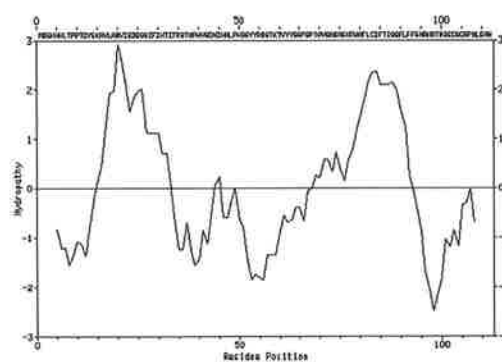
```
1  MKSNIASLMS ALRQPSDNSE SIKICAALGG TALLYIWCQ PNNPKSGDLG
51  HWLPFGGVYK DGNKFVRYNS PLHTAPFFWL LIAIPLAFFA LTHFSYFSNT
101 PRRRVTITHY PLCGHFVCRR
```

(b)

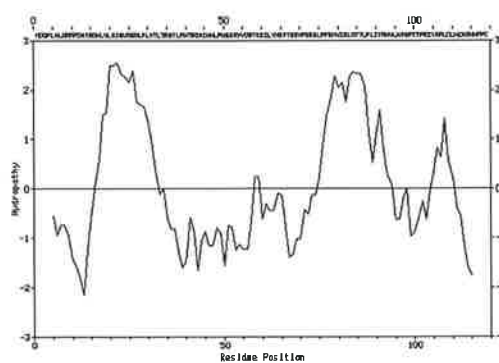
SCSMaV-ORF3



BaMV-ORF3



CsCMV-ORF3



ASPV-ORF3

Table 4.5 FastA database search results for ORF 4 peptide sequence (see Table 4.2 for legend)

The best scores are:		length	initn	init1	opt	z-sc	E(467568)
BAA04856	Apple stem pitting virus (ASPV)	(70)	80	80	103	166.9	0.081
AAB81271	Potato virus M strain Idaho 25K	(63)	52	52	82	137.8	3.4
AAB31047	12K, 7K=coat protein [Helenium	(61)	25	25	81	136.6	3.9
BAA02627	Poplar mosaic virus (PMV) genes	(71)	43	43	78	131.4	7.7
AAB20079	ORF1...ORF6 {3' terminal reigon	(64)	26	26	77	130.6	8.5
AAB71856	Cymbidium mosaic virus Korean t	(91)	43	43	77	128.4	11
AAB57882	Cymbidium mosaic virus, complet	(91)	43	43	77	128.4	11
CAA44530	Cymbidium mosaic virus RNA for	(91)	43	43	77	128.4	11
CAB40580	Bacillus cereus bc297a, alkD ge	(203)	37	37	80	127.4	13
AAA47170	Potato virus X complete genome	(70)	27	27	75	127.2	13
AAA03492	Potato virus X mRNA, complete c	(70)	27	27	75	127.2	13
BAA00252	Potato virus X genomic RNA	(70)	27	27	75	127.2	13
AAF13452	Potato virus X 12 kDa movement	(70)	27	27	75	127.2	13
CAA28829	Potato virus X genome	(70)	27	27	75	127.2	13
CAA20982	Caenorhabditis elegans cosmid Y	(158)	37	37	78	126.2	15
AAC54778	Cassava common mosaic virus Bra	(97)	32	32	74	123.7	21
AAA83294	Caenorhabditis elegans cosmid Z	(188)	69	69	77	123.7	21
BAA11122	Garlic latent virus genes for n	(64)	55	55	71	122.1	25
CAA91413	Caenorhabditis elegans cosmid M	(429)	35	35	79	121.2	28
AAD37496	Corynebacterium diphtheriae hem	(119)	30	30	73	121.0	29

the TGB encoding movement proteins. Searches for protein motifs using the PROSITE dictionary were not informative, with no motifs found. Hydropathy plots were done for the ORF4 peptides of SCSMaV, ASPV, *Garlic latent virus* (GarLV) and *Helenium virus S* (HelVS) and comparison of the profiles showed that they had similar trends (Figure 4.6b).

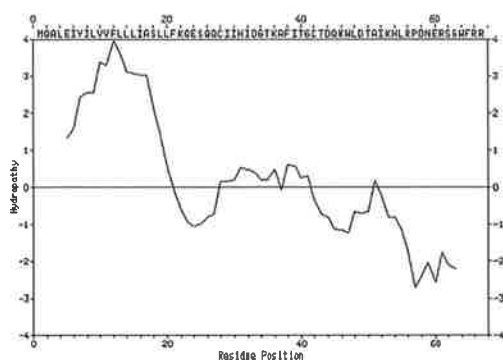
4.4.3 ORFs 2, 3 and 4 comprise the Triple Gene Block (TGB)

Genome database comparisons of peptide sequences suggest that SCSMaV ORFs 2, 3 and 4 comprise the TGB encoding movement proteins. The TGB is found in species in the genera *Potexvirus*, *Carlavirus*, *Foveavirus*, *Benyvirus* and *Hordeivirus* (van Regenmortel *et al.*, 2000) and some similarity is shared with the TGB of the *Furovirus* which has a segmented RNA genome (Koenig *et al.*, 1996). The TGB of these genera encodes three proteins. TGBp1 encodes a helicase that is a member of a distinct group within superfamily I of DNA and RNA helicases (Mushegian and Koonin, 1993). TGBp1 has two reported functional domains. The N terminus has probable enzymatic function and the C terminus mediates interactions with other virus products (Morozov *et al.*, 1999). The reported RNA binding activity (Wung *et al.*, 1999) is therefore likely to be mediated by the N terminus, whilst the C terminus mediates the formation of a ribonucleoprotein complex between the CP and TGBp1 (Lough *et al.*, 2000). During cell to cell movement the TGBp1 has been reported to form a ribonucleoprotein complex (Lough *et al.*, 2000) and localises to the plasmodesmata (Erhardt *et al.*, 1999). TGBp2 and TGBp3 are predicted to encode two small membrane spanning proteins (Mushegian and Koonin, 1993) which may interact with TGBp1 or the ribonucleoprotein complex to assist cell to cell movement (Lough *et al.*, 1999; Morozov *et al.*, 1999).

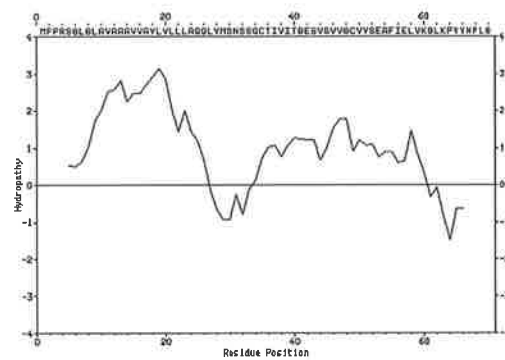
(a)

```
1  MQALEIYILV VFLLLIASLL FKQESQQCII HIDGTKAFIT GCTDQKWLDI
51 AIKHLRPDNE RSGWFRR
```

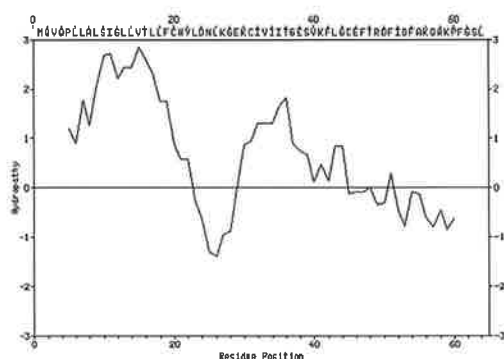
(b)



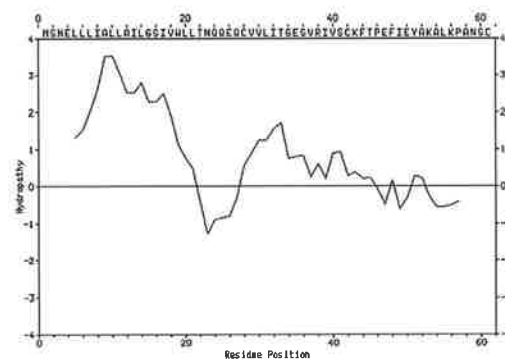
SCSMaV-ORF4



ASPV-ORF4



GarLV-ORF4



HeIVS-ORF4

Figure 4.6

(a) Predicted peptide sequence for SCSMaV-ORF 4.

(b) Comparison of hydropathy plots for the ORF 4 peptide sequences of SCSMaV and ASPV, GarLV and HeIVS, showing similar profiles.

SCSMaV ORF2 putative protein has a potential nucleotide binding domain and is therefore predicted to encode the helicase protein (TGBp1). ORF3 and ORF4 putative proteins have hydrophobicity profiles that are similar to the predicted small membrane spanning proteins of viruses that have TGB proteins and are predicted to encode the membrane spanning proteins of the triple gene block (TGBp2 and TGBp3 respectively).

4.4.5 ORF5

ORF5 spans nt 7102-7734 and encodes a peptide of 210 amino acids with a predicted M_r of 23000 (Table 4.1). The predicted amino acid sequence of ORF5 peptide is shown in Figure 4.7. FastA and Blast genome database comparisons showed amino acid sequence similarity of up to 22.9% with the coat protein (CP) of species within the genera *Foveavirus* and *Potexvirus* (Table 4.6). A search for protein motifs found a *Potexvirus* and *Carlavirus* CP signature in the ORF5 peptide sequence (Figure 4.7). This motif was recognised in a 70 amino acid central portion of the CP peptide sequence conserved between species in the genera *Potexvirus* (AbouHaidar and Lai, 1989) and *Carlavirus* (Mackenzie *et al.*, 1989). The most conserved region is located in the central part and is used as the signature pattern. It is suggested that this region is found within the native molecule and is important in maintaining correct tertiary structure through interactions between the viral RNA and CP (Mackenzie *et al.*, 1989), and thus maintains structurally similar arrangements of the CP (AbouHaidar and Lai, 1989). The sequence similarities to CP amino acid sequences of a range of species in the genera *Foveavirus*, *Potexvirus* and *Carlavirus* and the presence of a CP specific motif led to the conclusion that SCSMaV ORF5 encodes the CP.

Table 4.6 FastA database search results for ORF 5 peptide sequence (see Table 4.2 for legend)

The best scores are:			length	initn	init1	opt	z-sc	E(542720)
X	BAA00373	Clover yellow mosaic virus geno	(257)	297	244	401	498.5	3.2e-20
X	BAA04857	Apple stem pitting virus (ASPV)	(414)	285	253	371	458.7	5.2e-18
X	BAA04852	Pear vein yellows-associated vi	(414)	280	253	371	458.7	5.2e-18
X	AAC35502	Rupestris stem pitting associat	(259)	279	223	367	457.0	6.5e-18
	BAA00375	Clover yellow mosaic virus geno	(212)	244	244	365	455.8	7.6e-18
	AAC62913	Grapevine Rupestris stem pittin	(259)	279	223	366	455.7	7.7e-18
X	CAA61261	Potato virus X coat protein gen	(248)	303	303	361	449.9	1.6e-17
	CAA61264	Potato virus X coat protein gen	(248)	303	303	361	449.9	1.6e-17
	CAA82533	Potato virus X (CP2) mRNA for 8	(236)	309	309	360	449.0	1.8e-17
	CAA28830	Potato virus X genome	(237)	300	300	356	444.1	3.4e-17
	AAA47177	Potato X virus coat protein gen	(237)	300	300	356	444.1	3.4e-17
	AAA47171	Potato virus X complete genome	(237)	300	300	356	444.1	3.4e-17
	CAA61260	Potato virus X coat protein gen	(248)	298	298	356	443.8	3.5e-17
	CAA31294	Potato Virus X (PVX) mRNA for v	(236)	304	304	355	442.9	4e-17
	CAA39328	Potato Virus X complete genomic	(236)	304	304	355	442.9	4e-17
	CAA61267	Potato virus X coat protein gen	(237)	305	305	355	442.9	4e-17
X	AAC16234	Apricot latent virus putative c	(421)	289	257	358	442.7	4.1e-17
	AAF13450	Potato virus X capsid protein m	(237)	304	304	354	441.7	4.7e-17
	CAA61262	Potato virus X coat protein gen	(237)	304	304	354	441.7	4.7e-17
	CAA61263	Potato virus X coat protein gen	(237)	304	304	354	441.7	4.7e-17
	CAA61266	Potato virus X coat protein gen	(237)	304	304	354	441.7	4.7e-17
	BAA00253	Potato virus X genomic RNA	(237)	304	304	354	441.7	4.7e-17
	CAA84016	Potato virus X (MS) coat protei	(237)	304	304	354	441.7	4.7e-17
	CAA46149	Potato Virus X genomic RNA for	(237)	304	304	354	441.7	4.7e-17
	AAA03493	Potato virus X mRNA, complete c	(237)	304	304	354	441.7	4.7e-17
	BAA13515	Potato virus X mRNA for 12kDa n	(237)	303	303	353	440.5	5.4e-17
	CAA80778	Potato virus X coat protein RNA	(236)	302	302	352	439.3	6.3e-17
	CAA82535	Potato virus X (S6111) mRNA for	(237)	300	300	350	436.8	8.7e-17
	CAA51016	Potato Virus X strain HB RNA	(236)	302	302	349	435.6	1e-16
	CAA61265	Potato virus X coat protein gen	(236)	297	297	348	434.4	1.2e-16
	AAA47181	Potato virus X 166K RNA, comple	(236)	264	264	347	433.2	1.4e-16
	AAA47176	Potato virus X complete genome	(236)	264	264	347	433.2	1.4e-16
	CAA82531	Potato virus X (P551) mRNA for	(237)	294	294	346	431.9	1.6e-16
X	BAA03054	Papaya mosaic virus genomic RNA	(215)	283	159	342	427.7	2.8e-16
	AAA73480	Potato virus X coat protein mRN	(237)	290	290	342	427.0	3e-16
	BAA00169	Papaya mosaic potexvirus gene f	(214)	234	159	336	420.4	7.1e-16
X	AAB20080	ORF1...ORF6 {CVB 3' terminal	(315)	257	159	336	417.8	9.9e-16
X	CAA33396	Lily virus X RNA 3' terminal re	(201)	211	211	323	404.9	5.2e-15
X	AAB94083	Cowpea mild mottle virus CPMMV-	(288)	241	180	323	402.5	7.1e-15
X	AAB32335	PAMV putative polymerase...caps	(249)	246	246	322	402.3	7.3e-15
X	AAC35435	Sour cherry green ring mottle v	(268)	184	184	322	401.8	7.8e-15
X	AAC54779	Cassava common mosaic virus Bra	(229)	167	167	320	400.4	9.3e-15
X	AAD12241	Alternanthera potexvirus capsid	(207)	178	178	314	393.7	2.2e-14
	AAB94082	Cowpea mild mottle virus CPMMV-	(288)	209	155	313	390.3	3.4e-14
X	AAB71857	Cymbidium mosaic virus Korean t	(223)	242	242	310	388.3	4.3e-14
	AAB71858	Cymbidium mosaic virus Korean t	(223)	242	242	310	388.3	4.3e-14
	AAB57883	Cymbidium mosaic virus, complet	(223)	242	242	309	387.1	5.1e-14
	CAA44532	Cymbidium mosaic virus RNA for	(223)	241	241	306	383.5	8.1e-14
	CAA44064	Cymbidium mosaic virus gene for	(220)	239	239	305	382.3	9.4e-14
	AAF18460	Cymbidium mosaic virus triple g	(220)	241	241	302	378.7	1.5e-13
X	BAA01248	Helenium virus S (HelVS) gene f	(299)	297	187	303	377.8	1.7e-13
X	CAA34632	White clover mosaic virus total	(208)	232	232	300	376.6	2e-13
X	BAA25764	Shallot latent virus RNA for co	(294)	224	197	298	371.9	3.6e-13
	BAA25766	Shallot latent virus RNA for co	(294)	224	197	298	371.9	3.6e-13
	BAA25757	Shallot latent virus RNA for vi	(294)	224	197	298	371.9	3.6e-13
	CAA29908	Potexvirus white clover mosaic	(190)	232	232	295	371.1	4e-13

AAA69638	White clover mosaic potexvirus	(190)	232	232	295	371.1	4e-13
X CAA79765	P.asiatica mosaic virus genomic	(207)	186	186	295	370.5	4.3e-13
X CAA09306	Carnation latent virus RNA, 3'	(306)	227	181	290	361.8	1.3e-12
BAA25775	Garlic latent virus RNA for coa	(297)	249	186	288	359.6	1.7e-12
BAA11123	Garlic latent virus genes for n	(295)	190	190	285	356.0	2.8e-12
AAD20295	Lily symptomless virus strain K	(291)	169	135	284	354.8	3.2e-12
BAA25779	Shallot latent virus RNA for co	(297)	186	186	282	352.2	4.5e-12
BAA25774	Garlic latent virus RNA for coa	(297)	186	186	282	352.2	4.5e-12
BAA25776	Garlic latent virus RNA for coa	(297)	186	186	282	352.2	4.5e-12
BAA01935	Garlic latent virus genomic RNA	(297)	182	182	282	352.2	4.5e-12
BAA25777	Garlic latent virus RNA for coa	(297)	186	186	282	352.2	4.5e-12
BAA25778	Shallot latent virus RNA for co	(297)	186	186	282	352.2	4.5e-12
CAA33401	Lily symptomless virus RNA 3-te	(291)	169	135	281	351.2	5.1e-12
CAA57047	Garlic common latent carlavirus	(308)	269	205	281	350.8	5.4e-12
BAA25782	Garlic common latent virus RNA	(319)	269	206	281	350.5	5.5e-12
BAA25780	Garlic common latent virus RNA	(319)	269	205	281	350.5	5.5e-12
BAA25772	Garlic common latent virus RNA	(319)	266	202	281	350.5	5.5e-12
BAA25768	Garlic latent virus RNA for coa	(297)	186	186	280	349.8	6.1e-12
CAA37236	Potato virus M complete genome	(304)	180	180	279	348.4	7.3e-12
BAA03343	Potato virus M genomic RNA, com	(304)	180	180	279	348.4	7.3e-12
BAA07859	Lily symptomless virus mRNA for	(291)	169	135	278	347.5	8.2e-12
BAA25770	Garlic latent virus RNA for coa	(297)	181	181	278	347.4	8.3e-12
AAB70212	Lily symptomless virus 32kDa vi	(291)	169	135	277	346.3	9.6e-12
BAA05926	Garlic virus 1 genes for coat p	(297)	182	182	277	346.1	9.7e-12
AAF06328	Potato virus X coat protein gen	(120)	235	235	272	346.1	9.8e-12
CAB57961	Lily latent virus mRNA for part	(291)	169	135	276	345.1	1.1e-11
AAB81272	Potato virus M strain Idaho 25K	(304)	180	180	276	344.8	1.2e-11
AAF01313	Indian citrus ringspot virus co	(325)	232	205	276	344.3	1.2e-11
CAA56941	Cymbidium mosaic potexvirus gen	(220)	231	231	273	343.3	1.4e-11
CAA59433	Potato M carlavirus RNA for ORF	(304)	190	190	274	342.3	1.6e-11
CAA75693	Potato virus S gene encoding co	(294)	207	170	271	338.9	2.5e-11
AAA85903	Lily symptomless virus 25 kDa p	(291)	135	135	270	337.7	2.9e-11
CAA40688	Potato Virus M genes for capsid	(304)	172	172	269	336.2	3.5e-11
CAA75720	Potato virus S and genes encodi	(294)	201	168	268	335.2	4e-11
CAA75701	Potato virus S gene encoding co	(294)	203	170	267	334.0	4.6e-11
CAA75705	Potato virus S gene encoding co	(294)	203	170	267	334.0	4.6e-11
CAA75699	Potato virus S gene encoding co	(294)	203	170	267	334.0	4.6e-11
CAA75695	Potato virus S gene encoding co	(294)	203	170	267	334.0	4.6e-11
BAA00355	Potato virus S genomic RNA, 3'-	(293)	225	169	266	332.8	5.4e-11
BAA00353	Potato virus S genomic RNA, 3'-	(375)	225	169	266	331.2	6.7e-11
CAA75697	Potato virus S gene encoding co	(294)	203	170	262	327.9	1e-10
CAA92819	Garlic latent virus genes for r	(294)	186	186	261	326.7	1.2e-10
AAB65084	Potato virus S coat protein and	(294)	192	159	256	320.6	2.6e-10
AAC58813	Garlic virus B complete genome	(243)	184	184	252	316.9	4.1e-10

1	MKDQVGLEDE	WPPPSTGANA	SSGAQPTTQQ	ETSTVFKKSR	RGSKSKFVSA
51	VTKEQVRRVV	ELDIETDKIA	RDEDLEKVKA	AWLAINVPSD	KIFEVSLEIT
101	LSCAHTSTSN	KQKFKGKSKF	CDILLAEIAG	CITKEEVTIR	SFCAYWAKYV
151	WNAMHERNEP	PLNWVALGYT	<u>EQNRFAAFDF</u>	FYGVDPASA	EPEGGLLAPT
201	TESMRAANMA	RKEIAIKE			

Figure 4.7

Predicted peptide sequence for SCSMaV-ORF5 showing the motif conserved between species in the genera *Potexvirus* and *Carlavirus* at amino acid position 169-194 and represented by a horizontal line above the sequence.



4.4.6 The 5' and 3' non-coding regions

The SCSMaV 5' untranslated region (UTR) comprises 72 nucleotides and does not have nucleotide sequence similarity to any reported plant viruses by FastA or Blast genome database searches, and contains no recognised nucleotide sequence motifs. The size of the 5' UTR is between that of ASPV-RNA (52nt; Jelkmann, 1994) and SCGRMV-RNA (102nt; Zhang *et al.*, 1998). Sequence similarity in the 5' UTR is reported between species in the genera *Carlavirus* and *Potexvirus* (Huisman *et al.*, 1988; Zavriev *et al.*, 1991, Zhang *et al.*, 1998) but the SCSMaV-RNA 5' UTR does not share this.

The SCSMaV 3' UTR comprises 312 nucleotides and has ca. 31 % nucleotide sequence similarity to *Cowpea mosaic virus* and *Red clover mottle virus* (both genus *Comovirus*), but no similarity to rod-shaped plus sense ssRNA viruses by FastA and Blast genome database searches. A polyadenylation signal (AAUAAA; Nevins, 1983) was found at nucleotide position 8071-8077, 76-70 nt upstream from the polyA tail. The presence of a polyadenylation signal in the 3' UTR is not a consistent feature of viruses in the same genus, as it is present in some species of the genus *Potexvirus* (eg. CYMV; AbouHaidar, 1983) but absent from others (eg PVX: Huisman *et al.*, 1988; Morozov *et al.*, 1983). However, when the polyadenylation signal is present, it has been noted that it is further from the polyA tail than for eukaryotic mRNAs (AbouHaidar, 1988; AbouHaidar and Lai, 1989; Nevins, 1983). This is consistent with the findings for SCSMaV-RNA where the polyadenylation signal is located in a position spanning 76-70 nt upstream from the polyA tail, whilst the polyadenylation signal for eukaryotic mRNAs 10-30 nt from the polyA tail (Nevins, 1983). A hexanucleotide motif (ACUUAA) in the 3' UTR that is conserved between species in the *Potexvirus* and *Carlavirus* genera (Bancroft *et al.*, 1991) and has a suggested role in the synthesis of viral RNA (White *et al.*, 1992) was

not found in SCSMaV-RNA. These differences provide further evidence that SCSMaV is not closely related to viruses in either of these genera. An analysis of the 3' UTR of species in the genus *Foveavirus* showed that SCGRMV has a polyadenylation signal at nucleotide positions 8292-8298 and a hexanucleotide motif located immediately adjacent at nucleotide positions 8299-8304 (Zhang *et al.*, 1998), whereas the 3' UTR of ASPV and RSPaV had neither the polyadenylation signal nor the hexanucleotide motif (Jelkmann, 1994; Meng *et al.*, 1998). Therefore the presence or absence of these motifs cannot be used to distinguish species in the *Foveavirus* genus from viruses in other genera.

4.4.7 Other potential ORFs

Two additional potential ORFs were identified (Fig 4.1). ORF4a spans nt 7133-7285 and encodes a potential peptide of 50 amino acids and ORF5a spans nt 7541-7690 and encodes a potential peptide of 49 amino acids (Table 4.1; Figure 4.8). FastA and Blast searches showed that neither of these putative peptides have amino acid sequence similarity with any reported peptide sequences. Two additional potential ORFs (ORF2a and ORF5a) were reported in SCGRMV and like the potential proteins encoded by SCSMaV ORF4a and ORF5a, were found to have no significant similarity to any reported peptide sequence (Zhang *et al.*, 1998). Amino acid and nucleotide sequence comparisons between the protein products and nucleotide sequences of the additional potential ORFs showed no similarity between SCSMaV and SCGRMV (data not shown). Therefore the potential proteins encoded by ORF4a and ORF5a could not be assigned a function.

ORF4a

1	MLPVVLNQQL	SRKRAQFLRK	AAEDQNLNLF	LLLQRSRLGG	WSSLTLRQTK
---	------------	------------	------------	------------	------------

ORF5a

1	MNVMSLLSIG	LLLGIQNKTV	LQLLTFSMVL	TLLLQQNQKV	GYLLLPLSL
---	------------	------------	------------	------------	-----------

Figure 4.8

Predicted peptide sequence for two additional potential SCSMaV-ORFs: SCSMaV-ORF4a and SCSMaV-ORF5a.

CHAPTER 5

**PHYLOGENETIC ANALYSIS
OF SCSMaV RNA AND
PUTATIVE ORF PRODUCTS**

CHAPTER 5: PHYLOGENETIC ANALYSIS OF SCSMaV-RNA AND PUTATIVE ORF PRODUCTS

5.1 Introduction

Chapter 4 compared the amino acid sequences of the SCSMaV ORF products to assign putative functions. Comparison of these nucleotide and amino acid sequences to similar viruses can be used to determine regions of sequence identity (Koonin, 1991) and for phylogenetic analysis. Phylogenetic trees can be used to examine taxonomic relationships (Koonin and Dolja, 1993; Mayo and Pringle, 1998; van Regenmortel *et al.*, 2000; Ward, 1993) and to determine if a new virus species can be classified within a known genus. The RdRP generally gives a phylogeny that is accurate for taxonomic purposes because it is a highly conserved region (Breunn, 1991; Koonin, 1991; Koonin and Dolja, 1993; Poch *et al.*, 1989). However, it has been reported that every gene retains phylogenetic relationships that are independent of adjacent genomic elements (Morozov *et al.*, 1989). There is also no evidence to suggest that any one algorithm is better than any other (Weiller *et al.*, 1995), therefore several algorithms for phylogenetic analysis of both the nucleotide sequence and predicted protein products of SCSMaV-RNA are compared in this chapter in an attempt to establish the taxonomic position of SCSMaV.

5.2 Materials and methods

5.2.1 Phylogenetic analyses

Phylogenetic trees were constructed using the data described in Chapters 3 and 4. For the nucleotide sequence analysis, 16 full length virus RNA sequences were chosen (Table 5.1), for the RdRP amino acid sequences of 21 viruses were chosen (Table 4.2), and for the CP amino acid sequences of 20 viruses were chosen (Table 4.6). The accession numbers for these virus sequences are shown in Table 5.2. All programs used for phylogenetic analyses were accessed through the webANGIS interface (adapted from Felsenstein, 1989, unless otherwise noted) using the default parameters unless otherwise noted. Multiple sequence alignments produced by EclustalW (Thompson, 1994) Chapter 4) were resampled by bootstrapping using the eSeqboot program with 1000 bootstraps. The bootstrapped alignment was then used in four algorithms to produce phylogenetic trees. The Neighbor-Joining tree was constructed by firstly producing a distance matrix using Ednadist (Kimura 2-Parameter with a Transition/Transversion ratio of 2.0), which was interpreted using the program Eneighbor. The distance matrix was also used in a Fitch algorithm with global rearrangements, and a Kitsch algorithm. Maximum Parsimony trees were constructed using the programs Ednapars for nucleotide data and Eprotpars for peptide data. A consensus tree was constructed for each of the four methods by interpreting the output files using Econsense and viewed using TreeView (Page, 1998).

Table 5.1 FastA database search results using the full nucleotide sequence of SCSMaV-RNA

The best scores are:		length	initn	init1	opt	z-sc	E(944650)
†	SSU57923 Sugarcane striate mosaic virus	(1248)	5952	5952	5952	7410.7	0
†	SSU57924 Sugarcane striate mosaic virus	(671)	3221	3221	3230	4012.8	2.1e-216
†	SSU57922 Sugarcane striate mosaic virus	(632)	2219	1918	2359	2924.6	9.1e-156
X	AITASPV Apple stem pitting virus (ASPV)	(9306)	2264	1528	2323	2861.8	2e-153
X	BLUORF Blueberry scorch carlavirus mRNA	(8512)	1246	1246	2217	2729.9	4.8e-146
X	AF017780 Sour cherry green ring mottle v	(8372)	1854	1477	2155	2652.5	9.9e-142
	AF026278 Grapevine Rupestris stem pittin	(8725)	1168	753	1984	2438.5	7.9e-130
X	AF057136 Rupestris stem pitting associat	(8744)	1402	814	1978	2431.0	2.1e-129
X	GLVGNS Garlic latent virus genes for rep	(8353)	760	656	1956	2403.8	7e-128
X	PTMPVVMCG Potato virus M genomic RNA, com	(8535)	1407	808	1636	2003.7	1.3e-105
	POPVMCG Potato virus M complete genome	(8535)	1407	808	1636	2003.7	1.3e-105
	PVS Potato virus S genomic RNA, 3'-termi	(3552)	1575	1421	1607	1973.3	1.6e-103
	PVSASRE14 Potato virus S gene encoding r	(1072)	1451	1451	1572	1937.5	5.2e-101
	PVSASRE4 Potato virus S gene encoding re	(1072)	1030	916	1544	1902.5	4.6e-99
	PVSVLRE2 Potato virus S gene encoding re	(1072)	1424	1424	1509	1858.8	1.3e-96
	PVSVLRE5 Potato virus S gene encoding re	(1072)	1406	1406	1500	1847.5	5.3e-96
	PVSKBRE11 Potato virus S gene encoding r	(1062)	1291	958	1491	1836.3	2.2e-95
	PVSKARE9 Potato virus S gene encoding re	(1072)	1282	949	1482	1825.0	9.5e-95
	PVSKOBRA Potato virus S and genes encodi	(3553)	1164	949	1464	1794.6	1.4e-93
	PVSKBRE1 Potato virus S gene encoding re	(1066)	1164	949	1464	1802.6	1.7e-93
	PVSKBRE2 Potato virus S gene encoding re	(1069)	1155	940	1455	1791.3	7.2e-93
X	CCAORF Cherry capillovirus A RNA for ORF	(7383)	653	337	1379	1683.5	1.1e-87
	D11161 Garlic latent virus genomic RNA f	(3468)	790	790	1225	1496.0	6.2e-77
	S60150 ORF1...ORF6 {3' terminal reigon}	(3426)	1087	1008	1206	1472.4	1.3e-75
X	ACLSVCOMP Apple spot chlorotic leaf spot	(7549)	899	899	1201	1460.9	2.6e-75
	ACH243438 Apple chlorotic leaf spot viru	(7545)	664	664	1153	1400.9	5.7e-72
	APCCOMS Apple chlorotic leaf spot virus,	(7552)	780	780	1105	1340.9	1.2e-68
	APCCG Apple chlorotic leaf spot virus, c	(7555)	790	790	977	1180.9	1e-59
X	ASG241KP Apple stem grooving virus genom	(6496)	596	324	767	919.5	4.3e-45
X	GVMPCP Grapevine virus A genomic RNA, MP	(7349)	162	162	766	917.4	5e-45
	LSY131812 Lily latent virus mRNA for par	(3000)	645	645	737	887.1	5.9e-43
	PVTGR Potato virus T genomic RNA, 3'-ter	(2392)	372	372	700	842.4	2.3e-40
X	GVBMPCP Grapevine virus B genomic RNA fo	(7599)	220	129	688	819.7	1.3e-39
	CTLA Citrus tatter leaf virus (CTLV) gen	(2956)	204	204	669	802.2	3.2e-38
X	CTLVPP Citrus tatter leaf virus gene for	(6496)	204	204	668	795.7	3.4e-38
	CTLVCC Citrus tatter leaf virus RNA for	(2755)	204	204	634	758.9	8.9e-36
	ACL15889 Apple chlorotic leaf spot trich	(363)	176	176	428	514.9	2.6e-21
	GVY15892 Grapevine virus D RdRpol gene,	(363)	261	261	411	493.7	4e-20
	PTT988 Potato trichovirus T gene encodin	(363)	148	148	396	474.9	4.5e-19
	PAVYV Pear vein yellows-associated virus	(1598)	320	320	332	385.1	1e-14
	HLY15893 Heracleum latent virus RdRpol g	(363)	207	207	315	373.7	1.9e-13
	AF177225 Aconitum latent virus replicati	(620)	195	195	258	298.9	1.7e-09
	S73580 putative polymerase...capsid prot	(7059)	118	118	249	271.5	4.9e-09
X	POWCLMV4 White clover mosaic virus total	(5846)	104	104	236	256.5	4.1e-08
	AB010303 Garlic virus D genomic RNA, 3'	(4497)	81	81	236	258.3	4.2e-08
	CMVIRCP Cymbidium mosaic virus gene for	(1290)	110	110	236	266.6	5.1e-08
	CMFVCP Cymbidium mosaic potexvirus genom	(748)	105	105	234	267.7	7.6e-08
	POLX3TERM Lily virus X RNA 3' terminal r	(2343)	102	102	231	256.3	1e-07
X	AF016914 Cymbidium mosaic virus Korean t	(6227)	100	100	228	246.1	1.4e-07
	I66494 Sequence 14 from patent US 567036	(7218)	85	85	219	233.9	6e-07
	POLS3TER Lily symptomless virus RNA 3-te	(2613)	199	199	218	239.4	8.2e-07
	COCPMVB Cowpea mosaic virus bottom compo	(5889)	120	120	201	212.7	1.1e-05
	LSU43905 Lily symptomless virus 25 kDa p	(2511)	181	181	200	217.1	1.5e-05
	HENSG Helenium virus S (HelVS) gene for	(1389)	149	149	183	199.8	0.00025
X	SMY13938 Strawberry mild yellow edge-ass	(3037)	280	157	181	192.1	0.0003
	BJU68009 Borrelia japonica strain IKA2 p	(1351)	135	135	177	192.5	0.00065

MCPRNA3B Cowpea mosaic virus B RNA 3' te	(80)	120	120	178	212.5	0.00084
RCMMIDCA Red clover mottle virus (RCMV)	(3543)	97	97	173	181.1	0.0011
CELT21D12 Caenorhabditis elegans cosmid	(35092)	120	120	170	162.1	0.0012
AC005827 Homo sapiens clone 14503, compl	(94376)	94	94	169	154.3	0.0012
HS35M08 Homo sapiens chromosome 17 clone	(104520)	107	107	168	152.4	0.0014
AB004456 Shallot latent virus RNA for co	(1292)	119	119	172	186.6	0.0014
AC005159 Homo sapiens BAC clone GS1-262D	(79914)	134	134	168	154.2	0.0015
AC004386 Homo Sapiens Chromosome X clone	(172657)	127	127	167	147.8	0.0016
HSA004956 Homo sapiens Immunoglobulin ka	(711)	81	81	172	190.5	0.0016
CNS01DW7 Human chromosome 14 DNA sequenc	(169466)	108	108	165	145.5	0.0022
SYVCPE Strawberry mild yellow edge-assoc	(5966)	152	152	168	171.4	0.0022
CEF59F4 Caenorhabditis elegans cosmid F5	(20712)	104	104	166	160.6	0.0025
AC023301 Homo sapiens chromosome 18 clon	(134859)	206	124	164	145.7	0.0026
AB023149 Homo sapiens mRNA for KIAA0932	(6261)	96	96	166	168.6	0.003
AC018414 Homo sapiens clone RP11-17M19,	(138711)	102	102	163	144.3	0.0031
AF024629 Cowpea mild mottle virus CPMMV-	(2511)	85	85	165	173.4	0.004
AL354975 Homo sapiens chromosome 10 clon	(205528)	94	94	160	137.9	0.0047
AJ397593 dkfz426 Gallus gallus cDNA clo	(678)	89	89	164	180.8	0.0058
AL135938 Homo sapiens chromosome 20 clon	(83290)	111	111	159	142.7	0.0063
CELT24H7 Caenorhabditis elegans cosmid T	(20823)	108	108	160	153.1	0.0066
AL353754 Homo sapiens chromosome 10 clon	(173322)	112	112	156	134.1	0.0091
AC010376 Homo sapiens chromosome 5 clone	(146839)	101	101	156	135.2	0.0093
AC000058 Homo sapiens BAC clone CTB-111F	(46419)	84	84	157	144.0	0.0094
HSA004955 Homo sapiens Immunoglobulin ka	(618)	81	81	161	177.7	0.0094
AF190647 Tetrahymena thermophila SerH3 i	(2562)	90	90	159	165.8	0.011
TETSATR Tetrahymena thermophila surface	(2560)	90	90	159	165.8	0.011
AC006820 Caenorhabditis elegans clone Y7	(20507)	82	82	157	149.5	0.011
CELY73C8B Caenorhabditis elegans cosmid	(20507)	82	82	157	149.5	0.011
AC003106 Homo sapiens Xp22 GS-524I1 (Gen	(128888)	105	105	155	134.8	0.011
TLFORPKPF T.litoralis FOR, pki and pflX	(4516)	90	90	158	160.8	0.011
AL136228 Homo sapiens chromosome 6 clone	(93003)	137	137	155	136.9	0.012
AL161433 Homo sapiens chromosome 6 clone	(165942)	96	96	154	131.8	0.013

X: used for phylogenetic analysis

†: partial sequences of SCSMaV-RNA (Choi, 1997; Choi *et al.*, 1999)

length: The size of the sequence in nucleotides.

init1: The alignment score derived by FastA from the best initial alignment found between the query sequence and a database sequence in the first phase of a search.

initn: The alignment score calculated in the second phase of a FastA search, after the shorter alignments found in the first phase of the search are joined together.

opt: The alignment score derived from the 'optimised' sensitive alignment derived in the third phase of the FastA search

z sc: Normalised score derived from the opt score.

E (expectation value): Conversion of the z sc into a statistical expectation value which approximates the probability that a given match is a chance occurrence.

Table 5.2: Sequences used for the construction of phylogenetic trees

Nucleotide: Accession numbers for GenBank or *EMBL			
Acronym	Virus	Accession No.	Reference (DS: Direct submission/unpublished)
SCSMaV	Sugarcane striate mosaic associated virus	AF315308	Thompson This thesis
GarLV	<i>Garlic latent virus (Carlavirus)</i>	Z68502	DS
BBScV	<i>Blueberry scorch virus (Carlavirus)</i>	L25658	Cavileer <i>et al.</i> , 1994
PVM	<i>Potato virus M (Carlavirus)</i>	D14449* D00515* X53062*	Zavriev <i>et al.</i> , 1991
SCGRMV	Sour cherry green ring mottle virus (tentative <i>Foveavirus</i>)	AF017780	Zhang <i>et al.</i> , 1998
ASPV	<i>Apple stem pitting virus (Foveavirus)</i>	D21829	Jelkmann, 1994
RSPaV	<i>Rupestis stem pitting associated virus (Foveavirus)</i>	AF057136	Meng <i>et al.</i> , 1998
SMYEaV	<i>Strawberry mild yellow edged associated virus (Potexvirus)</i>	D12517* D01227*	Jelkmann <i>et al.</i> , 1992b
WCIMV	<i>White clover mosaic virus (Potexvirus)</i>	X16636*	Beck <i>et al.</i> , 1990
CyMV	<i>Cymbidium mosaic virus (Potexvirus)</i>	AF016914	DS
PAMV	<i>Potato aucuba mosaic virus (Potexvirus)</i>	S73580*	Xu <i>et al.</i> , 1994
CCA	<i>Cherry capillovirus A (Capillovirus)</i>	X82547	Jelkmann, 1995
GVA	<i>Grapevine virus A (Vitivirus)</i>	X75433	DS
GVB	<i>Grapevine virus B (Vitivirus)</i>	X75448 X75896	Saldarelli <i>et al.</i> , 1996
ACLSV	<i>Apple chlorotic leaf spot virus (Trichovirus)</i>	M58152 M31714	German <i>et al.</i> , 1990
ASGV	<i>Apple stem grooving virus (Capillovirus)</i>	D14995 S47260	Yoshikawa <i>et al.</i> , 1992
CTLV	<i>Citrus tatter leaf virus (Capillovirus)</i>	D16681	Ohira <i>et al.</i> , 1995

ORF1: Peptide sequence accession numbers for GenPept database

BBScV	<i>Blueberry scorch virus (Carlavirus)</i>	AAA68984	Cavileer <i>et al.</i> , 1994
PVS	<i>Potato virus S (Carlavirus)</i>	CAA75709	DS
GarLV	<i>Garlic latent virus (Carlavirus)</i>	CAA92815	DS
CVB	<i>Chrysanthemum virus B (Carlavirus)</i>	AAB20076	Levay and Zavriev, 1991
PVM	<i>Potato virus M (Carlavirus)</i>	CAA37232	Zavriev <i>et al.</i> , 1991
ASPV	<i>Apple stem pitting virus (Foveavirus)</i>	BAA04853	Jelkmann, 1994
GRSPaV	<i>Grapevine rupestris stem pitting associated virus (Foveavirus)</i>	AAC62910	DS
SCSMaV	Sugarcane striate mosaic associated virus	AF315308	Thompson This thesis
SCGRMV	Sour cherry green ring mottle virus (tentative <i>Foveavirus</i>)	AAC35433	Zhang <i>et al.</i> , 1998
ACLSV	<i>Apple chlorotic leaf spot virus (Trichovirus)</i>	CAA68080	German-Ratana <i>et al.</i> , 1997
CCA	<i>Cherry capillovirus A (Capillovirus)</i>	CAA57896	Jelkmann, 1995
CTLV	<i>Citrus tatter leaf virus (Capillovirus)</i>	BAA03869	Yoshikawa <i>et al.</i> , 1993
ASGV	<i>Apple stem grooving virus (Capillovirus)</i>	BAA03639	Yoshikawa <i>et al.</i> , 1992

PVT	<i>Potato virus T (Trichovirus)</i>	BAA01042	Ochi <i>et al.</i> , 1992
GVB	<i>Grapevine virus B (Vitivirus)</i>	CAA53196	Minafra <i>et al.</i> , 1994
GVA	<i>Grapevine virus A (Vitivirus)</i>	CAA53182	Minafra <i>et al.</i> , 1994
SMYEaV	<i>Strawberry mild yellow edge associated virus (Potexvirus)</i>	BAA02082	Jelkmann <i>et al.</i> , 1992b
PVX	<i>Potato virus X (Potexvirus)</i>	CAA39324	Orman <i>et al.</i> , 1990
ShV-X	<i>Shallot virus X (Allexivirus)</i>	AAA47787	Kanyuka <i>et al.</i> , 1992
Gar-VA	<i>Garlic virus A (Allexivirus)</i>	BAA61806	Sumi <i>et al.</i> , 1999
Gar-VB	<i>Garlic virus B (Allexivirus)</i>	AAC58809	DS
Gar-VC	<i>Garlic virus C (Allexivirus)</i>	BAA61818	Sumi <i>et al.</i> , 1999

ORF5: Peptide sequence accession numbers for GenPept database

PasiaMV	<i>Plantago asiatica mosaic virus (Potexvirus)</i>	CAA79765	Solovyev <i>et al.</i> , 1994
CaCMV	<i>Cassava common mosaic virus (Potexvirus)</i>	AAC54779	Calvert <i>et al.</i> , 1996
AltV	<i>Alternanthera virus (Potexvirus)</i>	AAD12241	Geering and Thomas, 1999
PapMV	<i>Papaya mosaic virus (Potexvirus)</i>	BAA03054	Sit <i>et al.</i> , 1989
CIYMV	<i>Clover yellow mosaic virus (Potexvirus)</i>	BAA00373	Abouhaidar and Lai, 1989
LVX	<i>Lily virus X (Potexvirus)</i>	CAA33396	Memelink <i>et al.</i> , 1990
PVX	<i>Potato virus X (Potexvirus)</i>	CAA61261	Santa Cruz and Baulcombe, 1995
WCIMV	<i>White clover mosaic virus (Potexvirus)</i>	CAA34632	Beck <i>et al.</i> , 1990
PAMV	<i>Potato aucuba mosaic virus (Potexvirus)</i>	AAB32335	Xu <i>et al.</i> , 1994
CyMV	<i>Cymbidium mosaic virus (Potexvirus)</i>	AAB71857	DS
ALV	<i>Apricot latent virus (Foveavirus)</i>	AAC16234	Nemchinov <i>et al.</i> , 2000
ASPV	<i>Apple stem pitting virus (Foveavirus)</i>	BAA04857	Jelkmann, 1994
PVYV	<i>Pear vein yellows virus (Foveavirus)</i>	BAA04852	Jelkmann, 1994
RSPaV	<i>Rupestris stem pitting associated virus (Foveavirus)</i>	AAC35502	Meng <i>et al.</i> , 1998
SCGRMV	<i>Sour cherry green ring mottle virus (tentative Foveavirus)</i>	AAC35435	Zhang <i>et al.</i> , 1998
SCSMaV	<i>Sugarcane striate mosaic associated virus</i>	AF315308	Thompson and Randles this thesis
CaLV	<i>Carnation latent virus (Carlavirus)</i>	CAA09306	Meehan and Mills, 1991
SLV	<i>Shallot latent virus (Carlavirus)</i>	BAA25764	Tsuneyoshi <i>et al.</i> , 1998
CPMMV	<i>Cowpea mild mottle virus (Carlavirus)</i>	AAB94083	DS
CVB	<i>Chrysanthemum virus B (Carlavirus)</i>	AAB20080	Levay and Zavriev, 1991
HelVS	<i>Helenium virus S (Carlavirus)</i>	BAA01248	Foster <i>et al.</i> , 1990

5.3 Results

5.3.1 Phylogenetic analysis using the full nucleotide sequence

Neighbor-Joining and Maximum Parsimony phylogenetic trees are shown in Figures 5.1a and 5.1b respectively. The phylogenetic data are represented by a cladogram and an unrooted tree.

5.3.2 Phylogenetic analysis using peptide sequences

Neighbor-Joining and Kitsch phylogenetic trees of the RdRP sequences are shown in Figure 5.2 a and 5.2b, respectively. Kitsch and Maximum Parsimony phylogenetic trees of the CP sequences are shown in Figure 5.3a and 5.3b, respectively. The phylogenetic data are represented by a cladogram and an unrooted tree.

5.4 Discussion

5.4.1 Relationships based on nucleotide sequence

Neighbor-Joining showed clustering of viruses into genera, except for species in the genus *Capillovirus* which were dispersed into two positions. Maximum Parsimony showed a similar arrangement, however species in the genus *Potexvirus* were more distant from species in the genera *Carlavirus* and *Foveavirus* than in the Neighbor-Joining analysis. SCSMaV was found between the members of the genera *Foveavirus* and *Carlavirus* for both algorithms.

Figure 5.1

(a) Neighbor joining phylogenetic trees for the complete nucleotide sequence of 17 viruses, shown as a cladogram (left) and unrooted tree (right). The percentages on the cladogram represent the number of times the branch fell to the right of the node out of 1000 bootstraps. The distances are arbitrary.

(b) Maximum parsimony trees for the complete nucleotide sequences of the same 17 viruses as for (a).

The colours on the unrooted trees represent the following genera:

Blue – *Carlavirus*

Red – *Foveavirus*

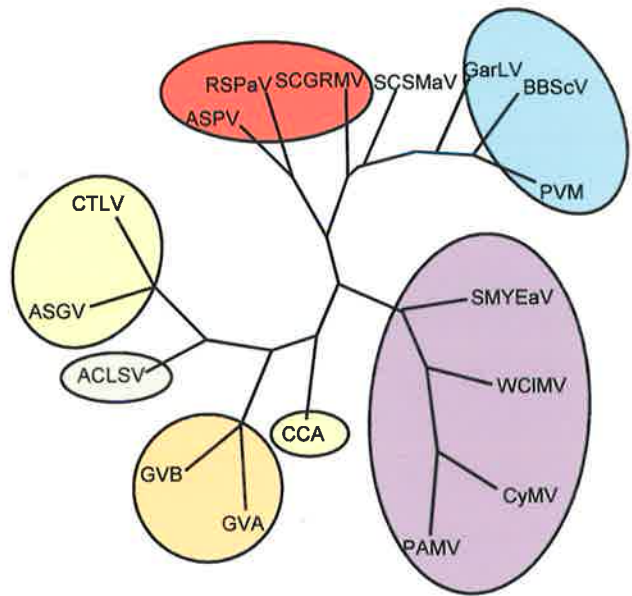
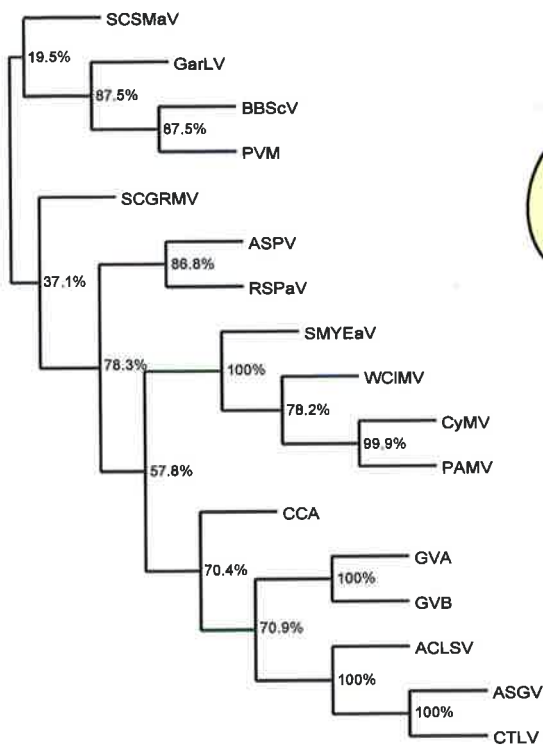
Purple – *Potexvirus*

Yellow – *Capillovirus*

Green – *Trichovirus*

Orange - *Vitivirus*

(a)



(b)

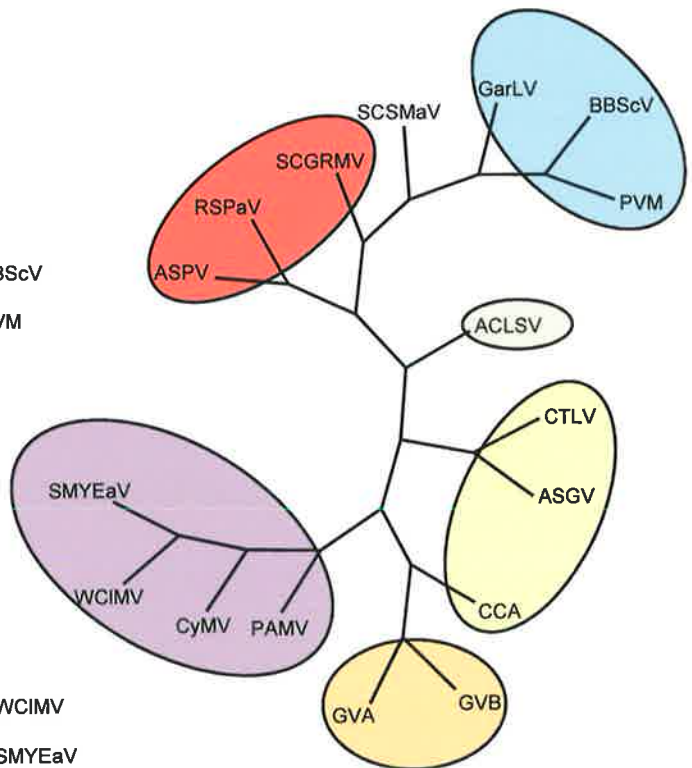
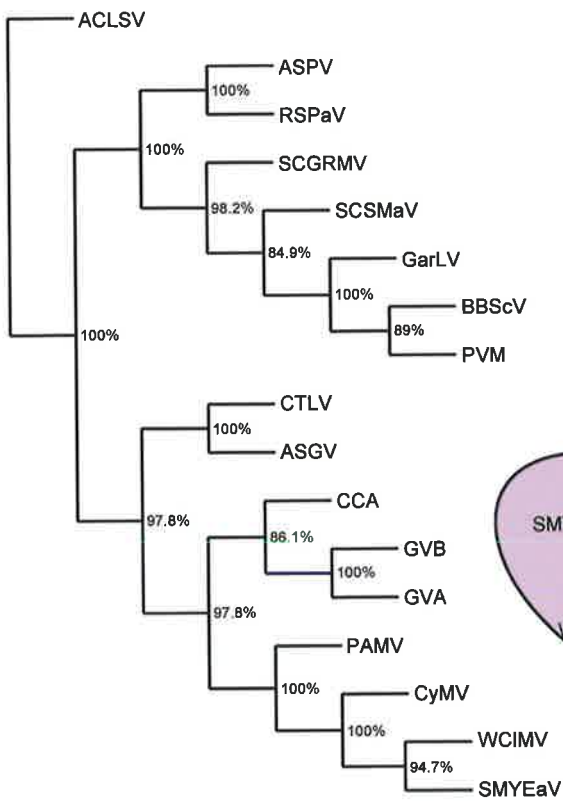


Figure 5.2

(a) Neighbor joining phylogenetic trees for the complete amino acid sequence of the RdRP of 22 viruses, shown as a cladogram (left) and unrooted tree (right). The percentages on the cladogram represent the number of times the branch fell to the right of the node out of 1000 bootstraps. The distances are arbitrary.

(b) Kitsch phylogenetic trees for the complete RdRP amino acid sequences of the same 22 viruses as for (a).

The colours on the unrooted trees represent the following genera:

Blue – *Carlavirus*

Red – *Foveavirus*

Purple – *Potexvirus*

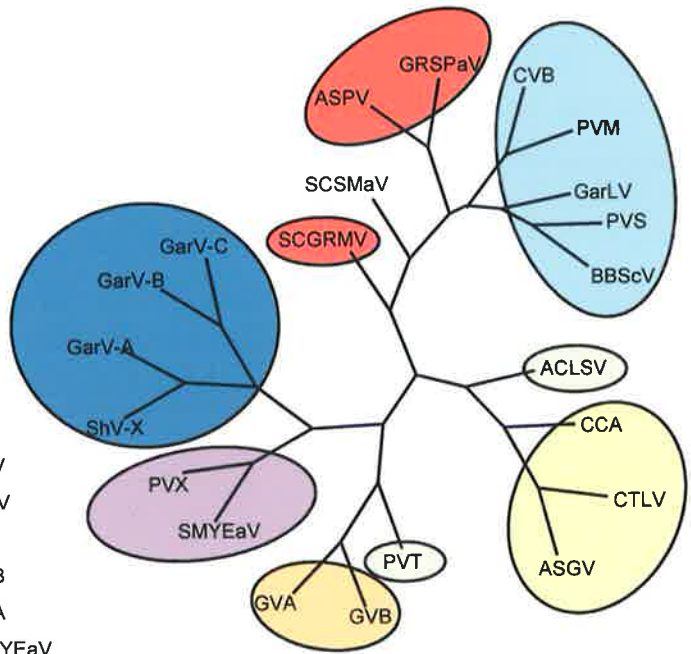
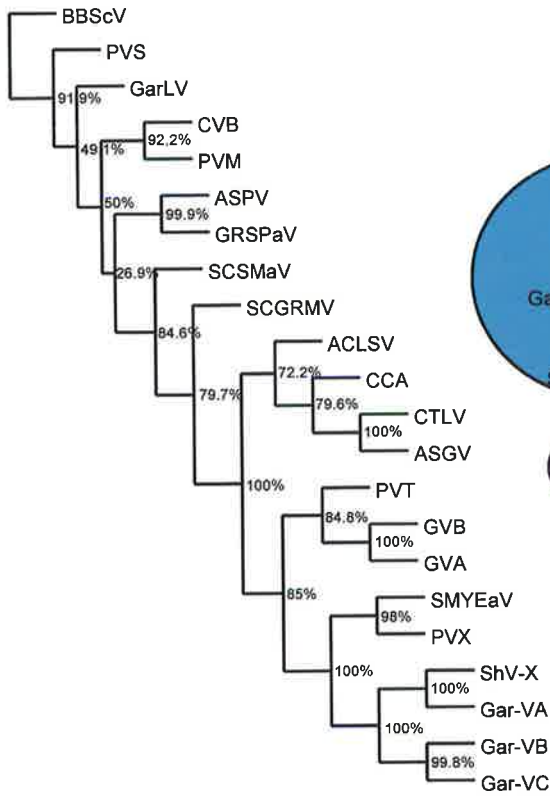
Yellow – *Capillovirus*

Green – *Trichovirus*

Orange – *Vitivirus*

Deep blue – *Allexiviruses*

(a)



(b)

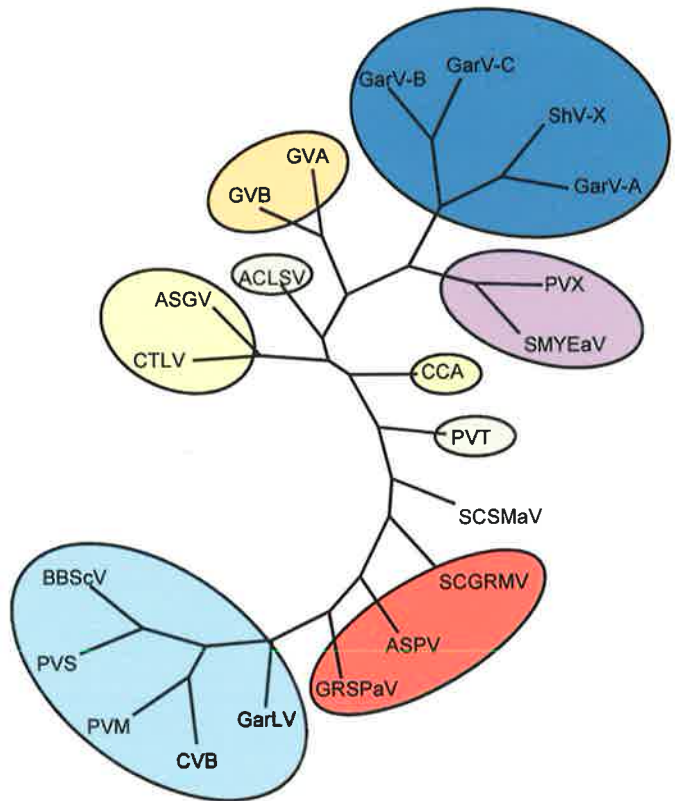
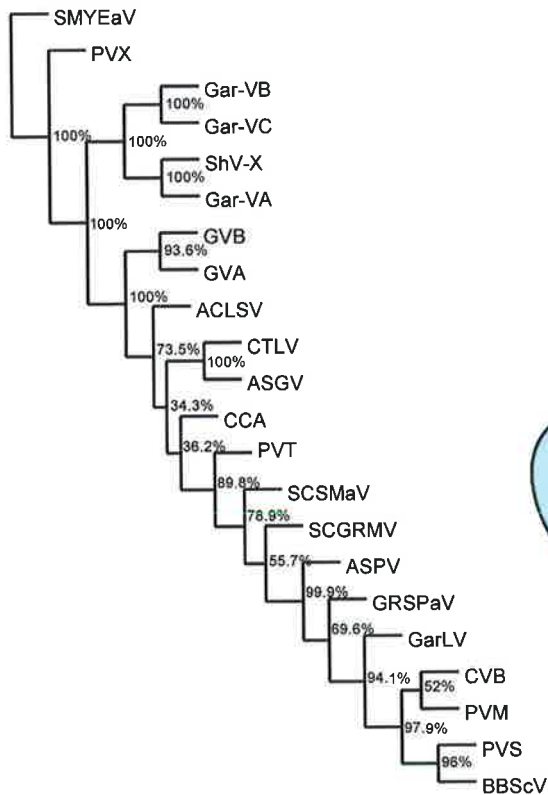


Figure 5.3

(a) Kitchin phylogenetic trees for the complete CP amino acid sequence of 21 viruses, shown as a cladogram (left) and unrooted tree (right). The percentages on the cladogram represent the number of times the branch fell to the right of the node out of 1000 bootstraps. The distances are arbitrary..

(b) Maximum parsimony trees for the complete CP amino acid sequences of the same 21 viruses as for (a).

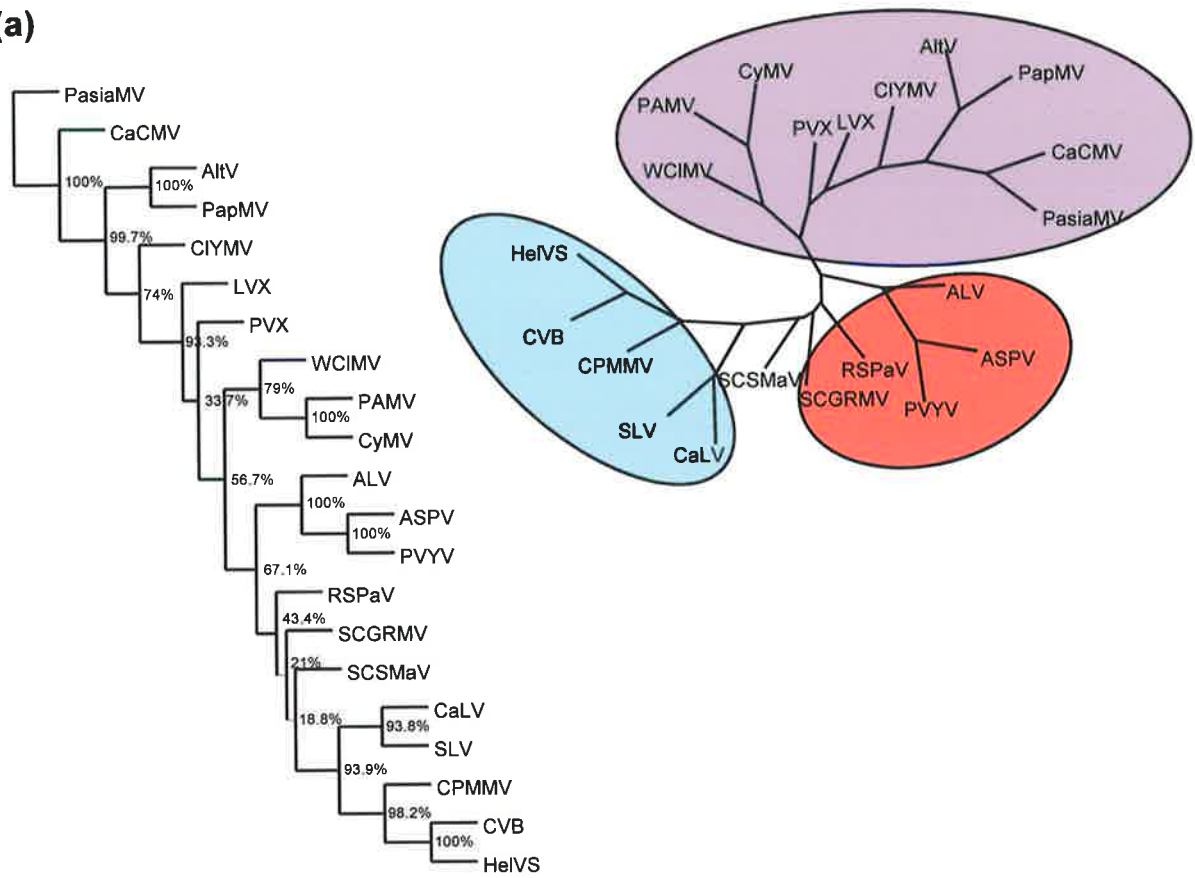
The colours on the unrooted trees represent the following genera:

Blue – *Carlavirus*

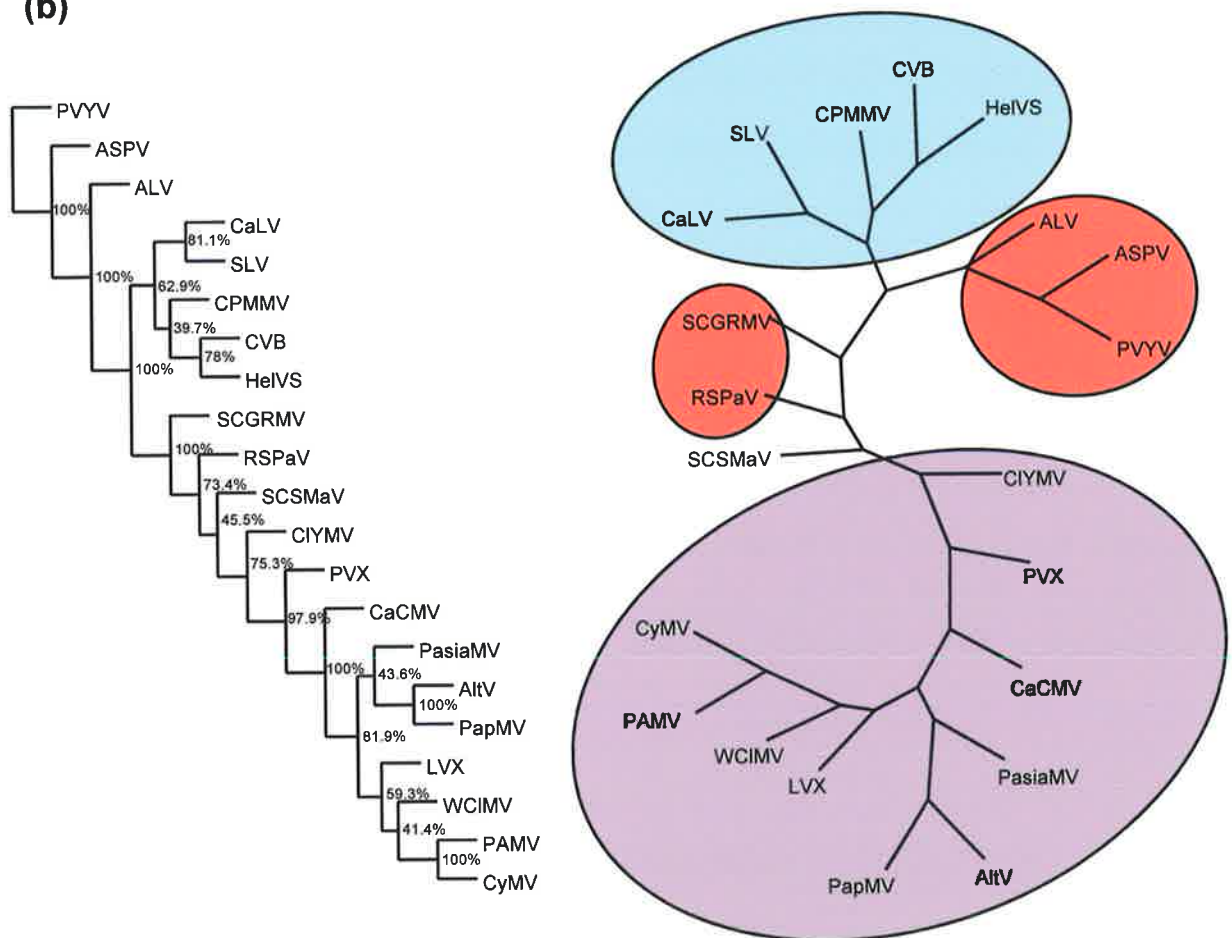
Red – *Foveavirus*

Purple – *Potexvirus*

(a)



(b)



5.4.2 Relationships based on RdRP sequence

Neighbor-Joining analysis of the RdRP amino acid sequence separated the viruses into clusters representing established genera, except for the genus *Trichovirus* (Figure 5.2a). The species in the genus *Foveavirus* were diverse with SCGRMV separated from ASPV and RSPaV. SCSMaV fell between SCGRMV and the other members of the genus *Foveavirus*. Kitsch analysis showed clustering of species from all genera except members of the genera *Capillovirus* and *Trichovirus*. Species of the genus *Foveavirus* species were located between members of the genus *Carlavirus* and members of the genera *Trichovirus/Capillovirus*, but not in a distinct clade. SCSMaV was found between members of the genus *Foveavirus* and members of the genera *Capillovirus/Trichovirus*. Analyses of these peptide trees show relationships similar to those obtained for the complete nucleotide sequence tree, with SCSMaV most closely allied to species of the genus *Foveavirus*, but still distinct.

5.4.3 Relationships based on TGB protein sequences

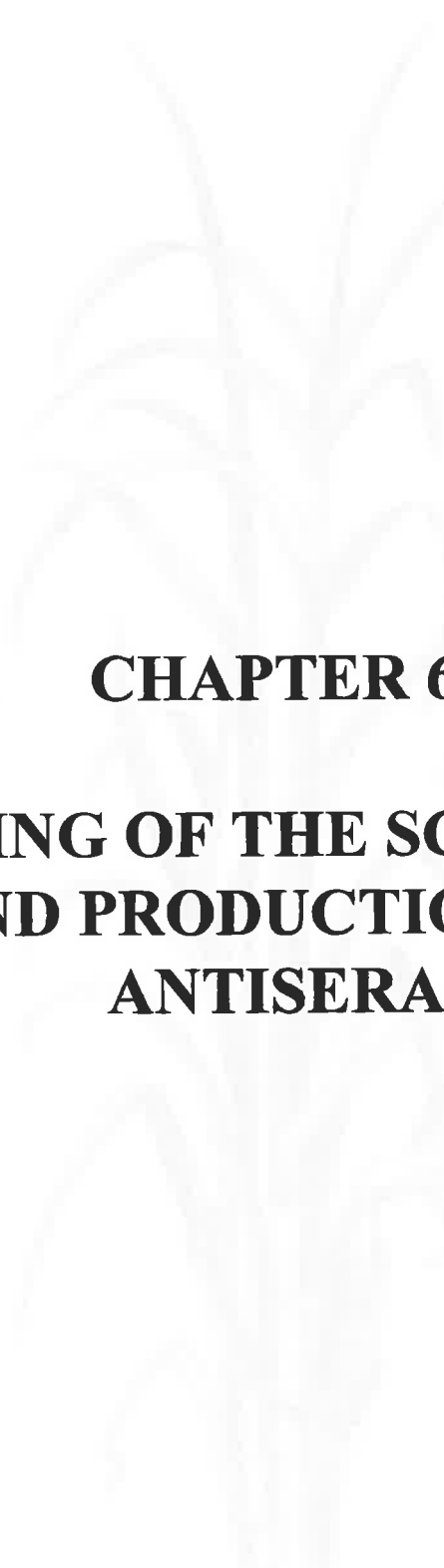
The phylogenetic trees produced from the TGBp1, TGBp2 and TGBp3 amino acid sequences using the viruses shown in Tables 4.3, 4.4, 4.5 and Appendix D, did not cluster viruses into established genera. Therefore these phylogenies were not used to attempt taxonomic classification of SCSMaV. Mushegian and Koonin (1993) also noted that the classification of movement proteins based on their amino acid sequence does not correlate with the type of genomic nucleic acid, or with groupings based on either the phylogenies of the nucleic acid replicase or host range.

5.4.4 Relationships based on CP sequence

Kitsch analysis showed clustering of the species in the genera *Potexvirus* and *Carlavirus* with the species in the genus *Foveavirus* located between the two (Figure 5.3a). ASPV, PVYV and ApLV formed a clade with RSPaV and SCGRMV on separate branches near this clade. SCSMaV was located between members of the genera *Foveavirus* and *Carlavirus*, showing the same phylogenetic positioning as for the nucleotide sequence tree (Figure 5.1) and for the RdRP tree (Figure 5.2). Analysis by Maximum Parsimony showed clustering of the species of the genera *Carlavirus* and *Potexvirus* into separate clades. The members of the genus *Foveavirus* were a diverse group, with ASPV, PVYV and ApLV more closely related to species in the genus *Carlavirus*, and SCGRMV and RSPaV located between species in the genera *Potexvirus* and *Carlavirus*. SCSMaV was located between members of the genera *Foveavirus* and *Potexvirus*.

5.4.5 Summation of phylogenetic analyses

The phylogenies described in this chapter were not used to infer relationships between virus species, but to assess if any algorithm and nucleotide or amino acid sequence combination separated viruses into their established genera. The position of SCSMaV could then be assessed to infer taxonomic classification. The trees shown were the best at clustering viruses into their genus. None of the algorithms was more accurate than any other, and only the complete nucleotide sequence, RdRP and CP protein sequences reflected established genera. These findings are supported by previous findings by Weiller *et al.* (1995) and Morozov *et al.* (1989), respectively. From these phylogenies it was concluded that SCSMaV was most closely related to viruses in the genus *Foveavirus*, and most closely related to SCGRMV, a tentative member of the this genus.



CHAPTER 6

CLONING OF THE SCSMaV-CP AND PRODUCTION OF ANTISERA

CHAPTER 6: CLONING OF THE SCSMaV-CP AND PRODUCTION OF ANTISERA

6.1 Introduction

Detection of plant viruses using antibodies developed against the coat protein or virion has wide applicability, high sensitivity and specificity, and allows simultaneous processing of many samples at a relatively low cost (Randles *et al.*, 1996). Pure antigen used to immunise animals will give polyclonal antibodies, which has an advantage over monoclonal antibodies for a detection test as they are not specific for one epitope. The antigenicity and antigenic specificity of a protein is determined by features such as its size, the number and nature of epitopes and the structure of the protein (Randles *et al.*, 1996). Antigens used for development of antibodies to plant viruses for use in diagnostic assays are usually either virions isolated from plant tissue or recombinant CP. Expression of CP in a bacterial expression vector system, such as the M15[pREP4] [pQE-31] system from Qiagen is one method for producing recombinant CP (Anon., 1999b). In this case the CP is amplified by RT-PCR with primers to allow cloning into the pQE-31 expression vector downstream of a histidine tag and under the control of an IPTG inducible promoter. The CP-vector construct is transformed into *E. coli* M15[pREP4] cells and the histidine-tagged CP can be overproduced by induction with IPTG and the histidine-tagged CP purified using a nickel-nitrilotriacetic acid (Ni-NTA) column (Anon., 1999b). This chapter describes the cloning, expression and purification of the SCSMaV-CP and an attempt to produce SCSMaV-CP specific antibodies for detection of SCSMD by DIBA and Western blot analysis.

6.2 Materials and methods

6.2.1 SCSMaV preparation

SCSMaV particles were prepared, separated by Cs₂SO₄ density centrifugation, and fractions collected (2.2.9). Each fraction was further separated by Nycodenz[®] gradient centrifugation and the visible bands were collected (2.2.9). Electron micrograph grids were prepared from these fractions and viewed by transmission electron microscopy (2.2.10).

6.2.2 Cloning of the CP of SCSMaV

Primers were designed to amplify the full SCSMaV-CP from SCSMaV-RNA using RT-PCR. These had *Bam*H1 and *Hind*III restriction enzyme sites incorporated into the sequence to allow cloning into pQE-31 (Figure 6.1). The SCSMaV-CP was amplified from dsRNA (2.2.3) by RT-PCR using Superscript II Reverse Transcriptase and CPLnew to prime the RT (2.2.6.2a) and DyNAzyme DNA polymerase with the CPUnew and CPLnew primers for the PCR (2.2.6.3a). The RT-PCR product was analysed by AGE (2.2.1.1), purified from the gel (2.2.7.1), restricted with *Bam*HI and *Hind*III in the supplied reaction buffer at 37°C for 1 hour (2.2.5), and further purified (2.2.7.1). The plasmid vector pQE-31 was also restricted and purified as described above. The restricted and purified RT-PCR product was ligated into the restricted and purified pQE-31 plasmid vector using T4 DNA ligase (2.2.7.2b). A 2 µl aliquot of the ligation reaction was used for transformation by heat shock into *E. coli* JM109 cells (2.2.7.3a), and positive colonies identified by colony screening PCR using the CPUnew and CPLnew primers (2.2.7.5b). The positive colonies were selected, plasmids prepared (2.2.4) and a 2µl aliquot of the purified plasmid transformed into competent *E. coli* M15

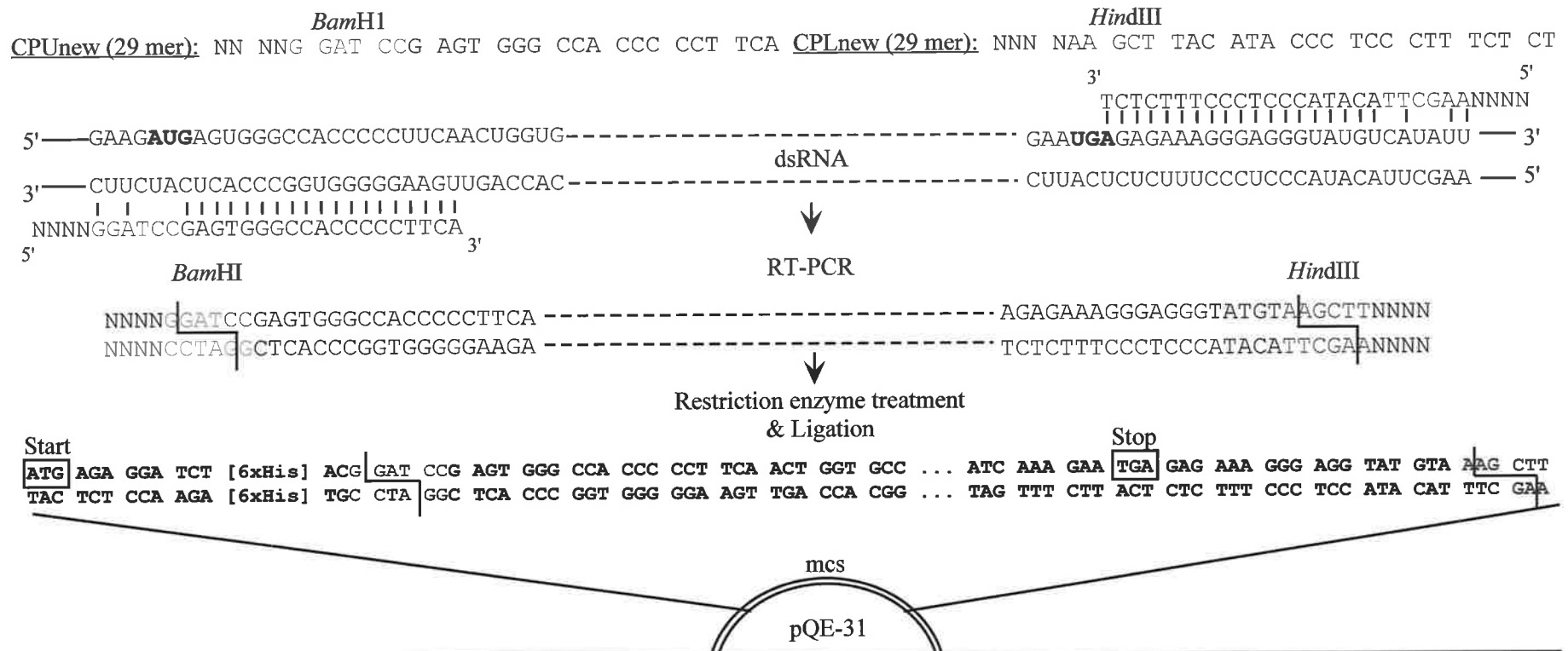


Figure 6.1: RT-PCR and cloning of SCSMaV-CP into pQE-31.

Primers CPUnew incorporating a *Bam*HI (red) restriction site and CPLnew incorporating a *Hind*III (blue) restriction site are shown as they are predicted to anneal to the SCSMaV dsRNA. The start and stop codons for SCSMaV-CP are in bold. The RT-PCR product of the amplified SCSMaV-CP with modified ends containing the restriction enzyme sites and indicating the sticky ends generated by restriction enzyme treatment are shown. The modified SCSMaV-CP is incorporated into the pQE-31 multiple cloning site (mcs) by ligation of the vector and RT-PCR products cohesive termini. The pQE-31-SCSMaV-CP construct contains the start codon for translation of the fusion protein at the 5' end of the sequence, the [6 x His] tag upstream of the SCSMaV-CP and the SCSMaV-CP stop codon is included in frame.

[pREP4] cells (2.2.7.3b and 2.2.7.4c). The colonies were screened by PCR with CPUnew and CPLnew (2.2.7.5b), and positive clones selected for expression and purification of the SCSMaV-CP. The purified plasmid also had the insert and flanking regions sequenced (2.2.7.6) using primers QE-31F and QE-31R (Appendix B; supplied by Dr. D. Pountney, CSIRO).

6.2.3 Expression and purification of SCSMaV-CP from *E. coli*

Figure 6.2 is a flow diagram describing the purification of proteins from the transformed *E. coli* M15[pREP4] and the fractions collected for SDS-PAGE analysis.

6.2.3.1 Expression of the recombinant protein

Selected colonies containing the pQE-31-SCSMaV-CP construct were grown for preparative purification using a scaled down version of Protocol 7 'Culture growth for preparative purification' (Anon., 1999b). A single colony was inoculated into 10 ml LB containing 100 µg/ml ampicillin and 25 µg/ml kanamycin and incubated for 16 hours at 37°C with constant vigorous shaking. This culture was inoculated into 500 ml LB containing 100 µg/ml ampicillin and 25 µg/ml kanamycin and grown at 37°C with constant vigorous shaking until an OD₆₀₀ of 0.6 was reached. A 1 ml sample (non-induced control) was removed, the cells pelleted and resuspended in 50 µl 2 x SDS-PAGE loading buffer (Appendix C) and stored at -20°C. Expression of the SCSMaV-CP was induced by the addition of IPTG to 1 mM and incubation of the culture for a further 4-5 hours, after which a second 1 ml sample (induced control) was collected and stored as above. The cells were then harvested by centrifugation at 4000 g for 20 minutes and the pellet stored at -20°C overnight.

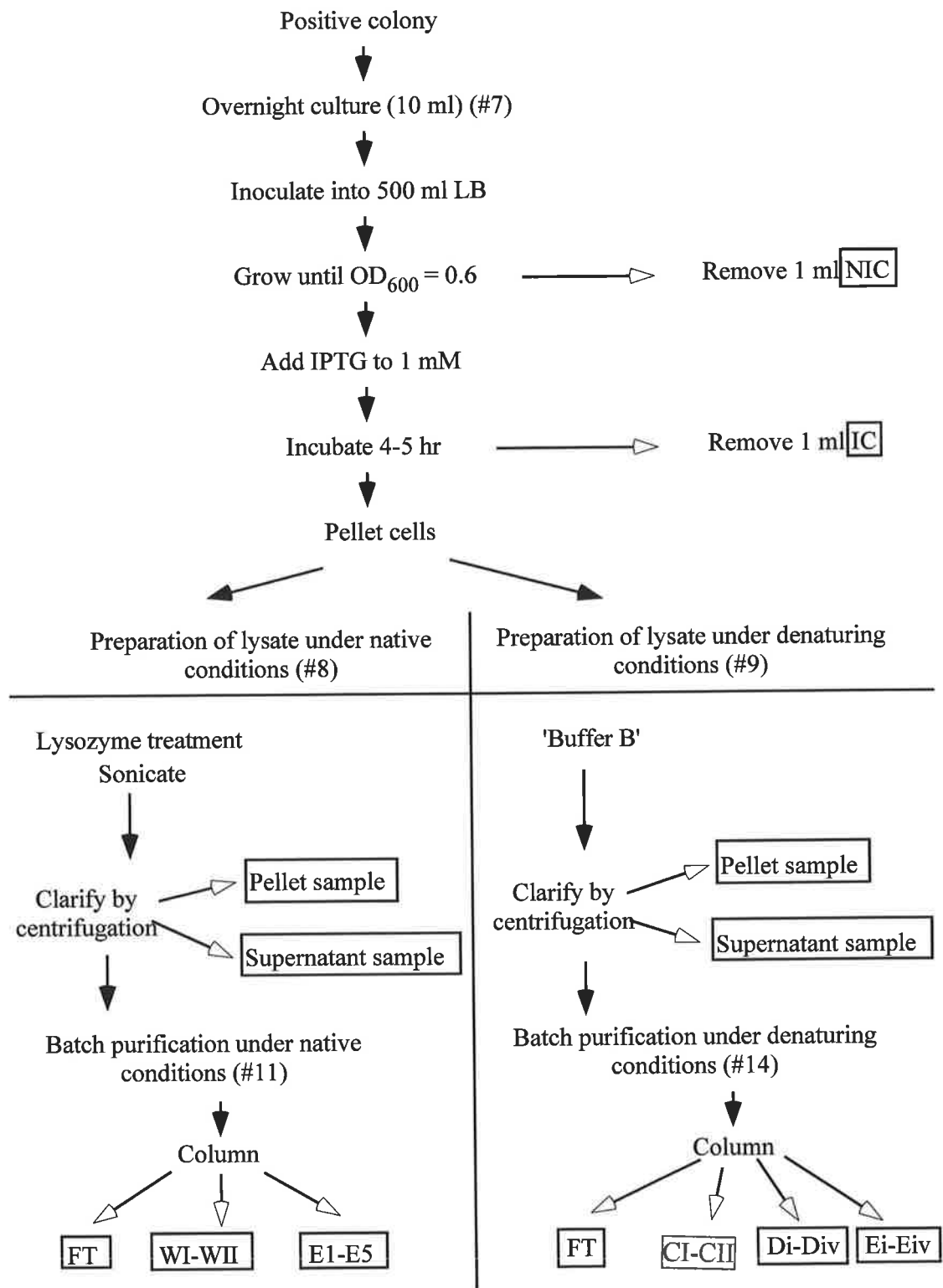
Figure 6.2

Flow diagram showing the samples collected during protein purification for SDS-PAGE analysis.

The numbers in parentheses show the protocol number from Anon (1999).

The solid head arrows show steps in the purification procedures and the open head arrows indicate the steps at which samples were removed. The samples removed for SDS-PAGE analysis are boxed. The composition of the buffers is in Appendix C.

NIC	Non-induced control
IC	Induced control
FT	Flow through of column before wash steps
WI-WII	Fractions collected after washing with native wash buffer
E1-E5	Fractions collected after elution with native elution buffer
CI-CII	Fractions collected after washing with denaturing wash buffer C
Di-Div	Fractions collected after elution with denaturing elution buffer D
Ei-Eiv	Fractions collected after elution with denaturing wash buffer E



6.2.3.2 Purification of induced protein under native conditions

The 'Preparation of cleared lysates under native conditions' was done according to Protocol 8 (Anon., 1999b). The cell pellet was thawed for 15 minutes on ice and the cells resuspended in native lysis buffer (Appendix C) at 2-5 ml per gram wet weight. Lysozyme was added to 1 mg/ml, the cells were incubated on ice for 30 minutes then sonicated on ice as recommended (Anon., 1999b). The cellular debris was removed by centrifugation at 10000 g for 30 minutes at 4°C. A sample of the pellet was resuspended in 200 µl 2 x SDS-PAGE loading buffer (Appendix C) for further analysis (pellet sample). The supernatant was used for 'Batch purification under native conditions' (Protocol 11; Anon., 1999b). The 50% Ni-NTA slurry (Qiagen) was equilibrated with native lysis buffer (Appendix C), and 1 ml added to 4 ml of cleared lysate and mixed gently on ice for 60 minutes. The lysate-Ni-NTA mix was loaded onto a column, and the flow-through (FT) collected. The column was washed twice with 4 ml native wash buffer (Appendix C), the wash fractions collected then the protein eluted by application of 0.5 ml native elution buffer (Appendix C), collecting each fraction in a separate tube. The samples, including non-induced control (NIC), induced control (IC), pellet, flow-through (FT), washes (W) and eluates (E) were analysed by 12% or 15% resolving SDS-PAGE (2.2.1.3).

6.2.3.3 Purification of induced protein under denaturing conditions

The 'Preparation of cleared lysates under denaturing conditions' was done according to Protocol 9 (Anon., 1999b). The cell pellet was thawed as described and resuspended in

denaturing buffer B (Appendix C) at 5 ml per gram wet weight. The cells were lysed by stirring for 60 minutes at 25°C, and the lysate centrifuged at 10000 g for 30 minutes to pellet the cellular debris. Samples of the pellet and supernatant were collected as described above (Figure 6.2), and the supernatant used for 'Batch purification under denaturing conditions' (Protocol 14; Anon., 1999b). The Ni-NTA slurry was equilibrated in denaturing buffer B (Appendix C), and 1 ml was added to 4 ml cleared lysate and mixed gently by shaking for 60 minutes at 25°C. The lysate-Ni-NTA mix was loaded onto a column, and the flow-through (FT) collected. The column was washed twice with 4 ml denaturing buffer C (Appendix C) and the wash fractions collected. The protein was eluted from the column by the application of 0.5 ml denaturing buffer D (Appendix C) four times, followed by four times of 0.5 ml denaturing buffer E (Appendix C) collecting all fractions in separate tubes. The samples, including non-induced control (NIC), induced control (IC), pellet (P), flow-through (FT), washes (W) and eluates from both buffers (D and E) were analysed by 12% or 15% resolving SDS-PAGE (2.2.1.3).

6.2.4 Preparation of antigen

6.2.4.1 Antigen A

The fractions containing protein that was purified under denaturing conditions using the Ni-NTA column (fractions Di to Eii, Figure 6.5b) were pooled and concentrated by drying *in vacuo*. The protein was resuspended in 3 ml sterile saline solution (0.14 M NaCl), and stored in 0.5 ml aliquots at -20°C. The antigen was emulsified with one volume of Freund's complete adjuvant (FCA) for the first injection, and with one volume of Freund's incomplete adjuvant (FIA) for each subsequent injection.

6.2.4.2 Antigen B

An aliquot of the induced pellet fraction was run on a 150 x 150 x 0.1 mm 15% resolving SDS-PAGE gel with one lane 100 mm wide (2.2.1.3), stained with Coomassie blue (2.2.1.4b) and the induced band identified and excised from the gel using a scalpel blade. The gel containing the protein was macerated in a glass homogenizer in saline solution (0.14 M NaCl; 1:1 v/v) until fine enough to pass through 19 gauge and 23 gauge needles. The total volume was made up to 6 ml with saline solution and stored in 1 ml aliquots at -20°C. No adjuvant was added to the aliquots before injection as polyacrylamide has been reported to have some adjuvant properties which can enhance the immune response ('Guidelines for the production of polyclonal and monoclonal antibodies in rodents and rabbits' <http://www.upstate.edu/dlar/antibody.htm>).

6.2.5 Preparation of antiserum

The antigens were injected subcutaneously into two New Zealand White Rabbits. The first bleed was done after 3 weeks. Four more injections were done two-weekly with test bleeds being taken prior to each injection. The final bleed was done by heart puncture three weeks after the final injection. The red blood cells were removed by allowing a clot to form overnight at 4°C, and the serum clarified by centrifugation at 3000 rpm for 5 minutes in a Beckman TJ-6 centrifuge with a swing-out rotor. Sodium azide was added to a concentration of 0.02% as a preservative and the serum stored at 4°C. The final bleed antiserum was used for most assays described below.

6.2.6 Dot immuno binding assay (DIBA)

DIBA was done using a modification of Hibi and Saito (1985) and Graddon and Randles (1986).

6.2.6.1 Preparation of membrane

A 0.5 g sample of leaf was crushed in PBS (1:1 w/v) in a plastic bag and the slurry transferred to a 1.5 ml centrifuge tube and clarified by centrifugation at 13000 g for 2 minutes. 1 µl dots of the supernatant were applied to nitrocellulose membrane (Schleicher and Schuell, BA85, 0.45 µm pore diameter) in duplicate and allowed to air dry. The samples applied to the membrane were SCSMD affected leaf, healthy sugarcane leaf, antigen A and antigen B.

6.2.6.2 Detection of antigen

Detection of the antigen was done using a modification of Ali and Randles (1997). Protein binding sites were saturated with blocking buffer B (Appendix C) for 1 hour at 25°C with constant agitation. The blocking buffer B was replaced with antiserum diluted 1/1000 in blocking buffer B, and incubated for 1 hour at 25°C with constant agitation. The membrane was washed three times for 3 minutes each with blocking buffer A, and incubated with goat-anti-rabbit phosphatase conjugate diluted to 1/1000 in PBS containing 1% BSA, for 1 hour at 25°C with constant agitation. The membrane was washed twice for three minutes each in AP 7.5 (Appendix C), and twice for three minutes each in AP 9.5 (Appendix C). The substrate was prepared by mixing 44 µl NBT stock (Appendix C) and 33 µl BCIP stock (Appendix C) with 10 ml AP 9.5, and added to the membrane. The membrane was incubated in the dark until dots appeared, to a maximum of 16 hours, and the reaction terminated with DIBA stop buffer (Appendix C). The membrane was air dried and stored sealed in a plastic bag.

Testing for endogenous phosphatase activity in the samples was done as described above, except that goat-anti-rabbit phosphatase conjugate was omitted.

An alternative blocking agent was tested which contained healthy plant sap and 600 mM mannose and 600 mM glucose (Ligat *et al.*, 1991; Ali and Randles, 1997).

6.2.7 Western blot assay

6.2.7.1 Preparation of membrane

Proteins fractionated on the resolving SDS-PAGE gel were transferred to nitrocellulose membrane (Schleicher and Schuell, BA85, 0.45 µm pore diameter) using a Bio-Rad Trans-Blot® SD Semi-Dry Electrophoretic Transfer Cell with Towbin transfer buffer at 10 V for 30 minutes (Towbin *et al.*, 1979). After transfer, the membrane was air-dried and stored at 4°C in an air-tight container.

6.2.7.2 Detection of antigen

Antigen was detected on nitrocellulose membranes as described for the DIBA (6.2.6.2), except that membranes were blocked with blocking buffer A.

6.3 Results

6.3.1 Virus preparation

The densities of fractions from the Cs₂SO₄ gradients that were refractionated on the Nycodenz® gradient were: Tube 1 = 1.27 g/l; Tube 2 = 1.328 g/l; Tube 3 = 1.337 g/l; Tube 4 = 1.516 g/l (Figure 6.3). Nycodenz® gradients and the corresponding transmission electron micrographs are shown in Figure 6.3. Electron micrographs a-e showed particles characteristic of SCSMaV (Figure 6.3).

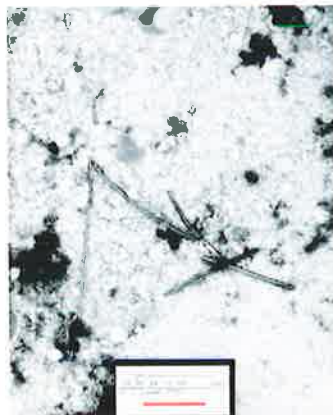
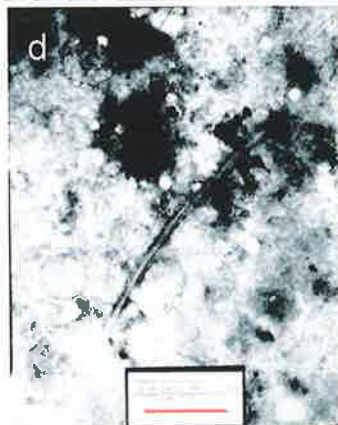
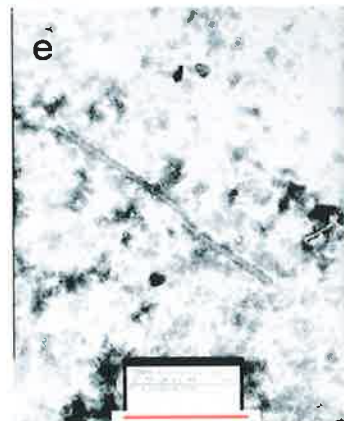
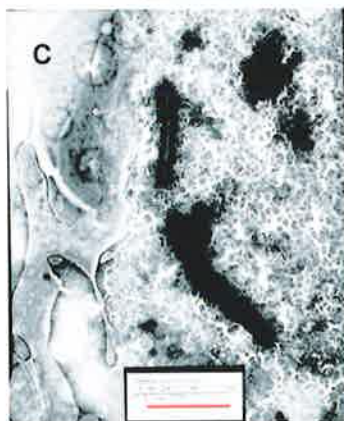
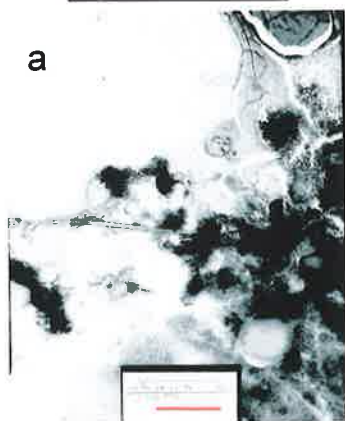
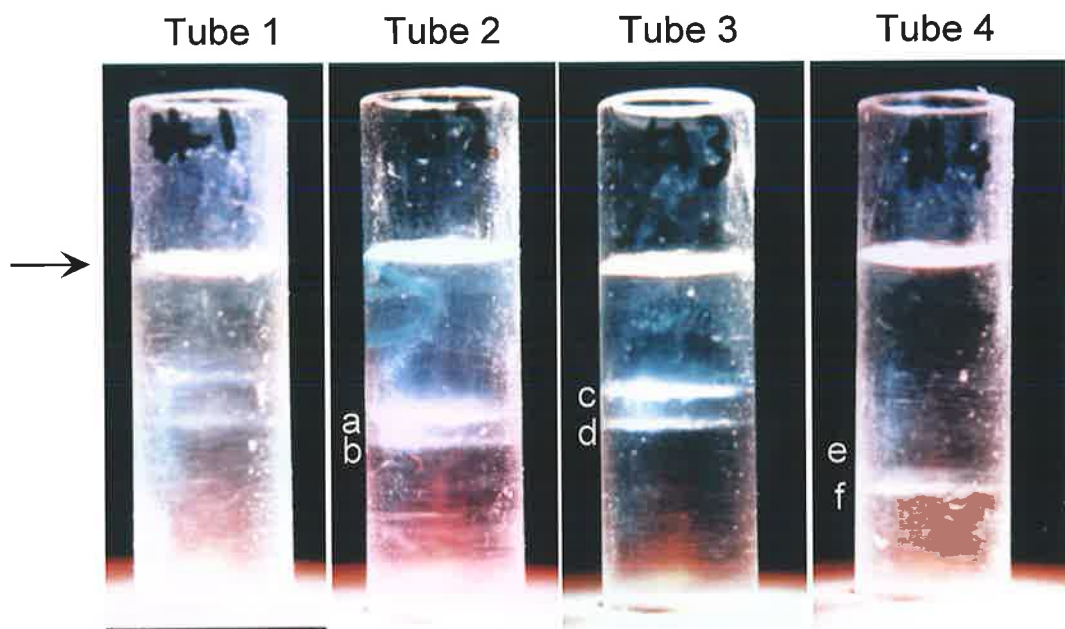
Figure 6.3

Nycodenz[®] gradients after centrifugation, the arrow indicates the meniscus (Tubes 1-4).

Transmission electron micrographs of virus preparations stained with 1% PTA.

The visible bands in the Nycodenz[®] gradients that showed particles by electron microscopy analysis are labelled (a) to (f) and the corresponding electron micrographs shown below.

The red scale bars represent 1 μm .



6.3.2 Cloning of the SCSMaV-CP

RT-PCR of the SCSMaV CP gave a 650 nt RT-PCR fragment (Figure 6.4), which was cloned complete and in frame into pQE-31 as described above (Figure 6.1). RT-PCR using CPL-new as the primer for reverse transcription gave a more strongly staining product than when the CPU-new primer was used (Figure 6.4).

6.3.3 Expression of the SCSMaV-CP in a bacterial expression system

SDS-PAGE analysis of the protein fractions collected from the protein purification protocol under native conditions showed multiple protein bands in the eluates (Figure 6.5a). The induced protein was most highly concentrated in the pellet fraction, and this band was excised (antigen B, 6.2.4.2). Purification under denaturing conditions identified a single protein product at low yield in eluates Di to Eii (antigen A, 6.2.4.1, Figure 6.5b). Antigen A was ca. 25 kDa in size and antigen B was ca. 28 kDa in size (Figure 6.5c).

6.3.4 DIBA

Antiserum B did not produce a detectable reaction with the homologous antigen by DIBA (antigen B; Figure 6.6b). This antiserum was therefore not used. Antiserum A showed a reaction to the homologous antigen (antigen A) and the SDS-PAGE gel purified antigen B (Figure 6.6a). Antiserum A also showed a reaction to both healthy and SCSMD affected sugarcane leaf sap after prolonged development (Figure 6.6a). The reactions of the antisera collected from the four test bleeds were compared, and no differences between reactions were observed by DIBA (not shown). Testing for

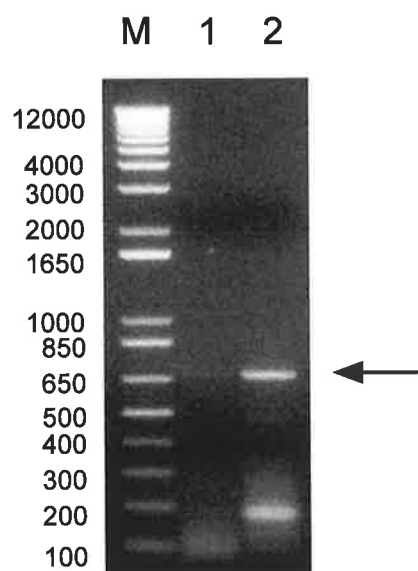


Figure 6.4

Analysis by 1.5% AGE of the RT-PCR amplification product of the SCSMaV-*CP* gene produced from dsRNA using the primers CPUnew and CPLnew.

M = 1 kb Plus DNA ladder, with sizes shown adjacent

1 = RT with CPUnew primer and PCR with CPUnew and CPLnew

2 = RT with CPLnew primer and PCR with CPUnew and CPLnew

The arrow indicates the ca. 650 nt RT-PCR product.

The size of the marker is shown in bp.

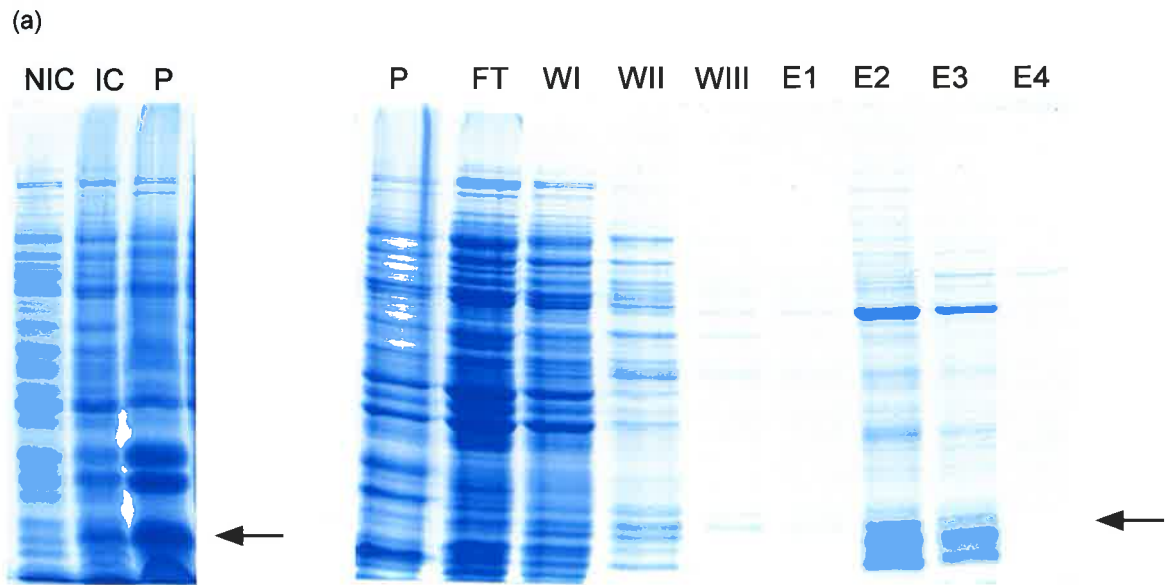


Figure 6.5a

Analysis by 12% SDS-PAGE of samples taken during the Ni-NTA fractionation of the induced protein under native conditions (Figure 6.2). The buffer compositions are in Appendix C.

NIC = Non-induced control sample of M15[pREP4] cells taken prior to IPTG induction

IC = Induced control sample of M15[pREP4] cells taken after IPTG induction

P = Sample of the pellet representative of insoluble proteins

FT = Flow-through of the column prior to washes

WI-WIII = Wash fractions with native wash buffer (contains 50 mM imidazole)

E1-E4 = Eluted fractions with native elution buffer (contains 250 mM imidazole)

The arrow indicates a protein band induced by IPTG ca 28 kDa

(b)

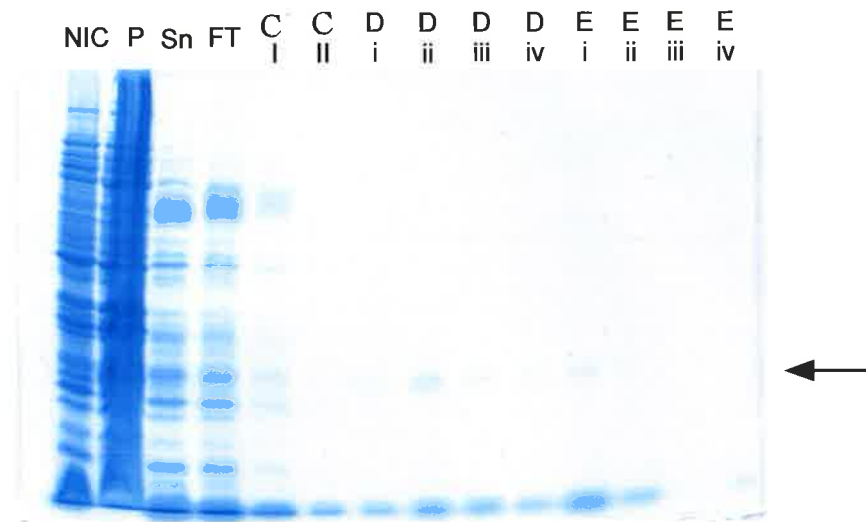


Figure 6.5b

Analysis by 15% SDS-PAGE of samples taken from steps during the Ni-NTA fractionation of induced proteins under denaturing conditions (Figure 6.2). The buffer compositions are in Appendix C.

NIC, P, FT samples as for Figure 6.5a

Sn = supernatant fraction taken before Ni-NTA column fractionation

C[-II = Fractions collected using Buffer C, pH 6.3

Di-Div = Fractions collected using Buffer D, pH 5.9

Ei-Eiv = Fractions collected using Buffer E, pH 4.5

The arrow shows protein eluted from the column in fractions Di - Eii.

(c)

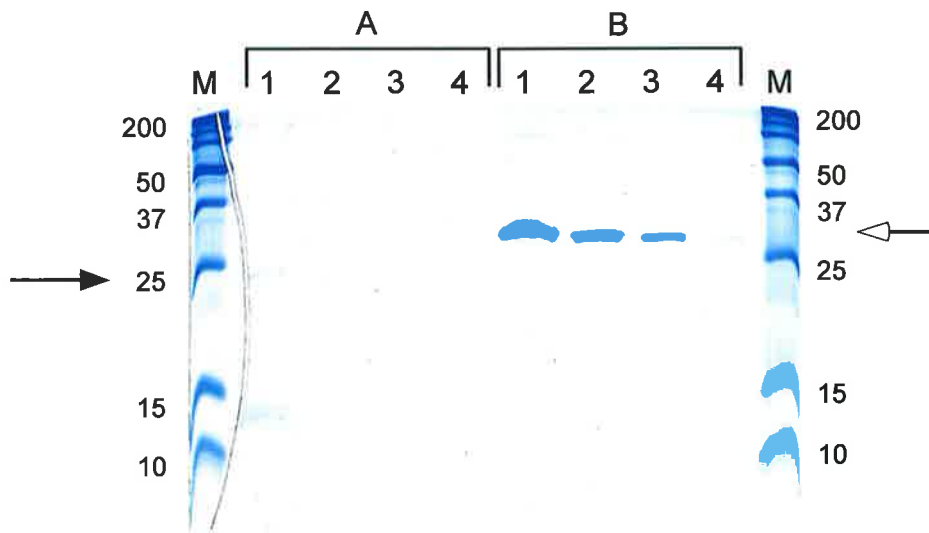


Figure 6.5c

Analysis by 15% SDS-PAGE of a dilution series of (A) Ni-NTA column purified antigen A and (B) SDS-PAGE gel purified antigen B that were injected into rabbits.

M = Prestained SDS-PAGE protein marker with sizes shown adjacent

A1 = 1/25th of amount at each injection

A2 = 1/50th

A3 = 1/100th

A4 = 1/500th

B1 = 1/50th of amount at each injection

B2 = 1/100th

B3 = 1/200th

B4 = 1/1000th

The closed head arrow shows the Ni-NTA column purified antigen A of ca. 25 kDa, and the open head arrow shows the SDS-PAGE gel purified Ag B of ca. 28 kDa.

The size of the marker is shown in kDa

Antigen A				SCSMD				
1/1	1/5	1/25	1/125	1/1	1/5	1/25	1/125	
1/1	1/5	1/25	1/125	1/1	1/5	1/25	1/125	
1/1	1/5	1/25	1/125	1/1	1/5	1/25	1/125	
1/1	1/5	1/25	1/125	1/1	1/5	1/25	1/125	
Antigen B				Healthy				PBS

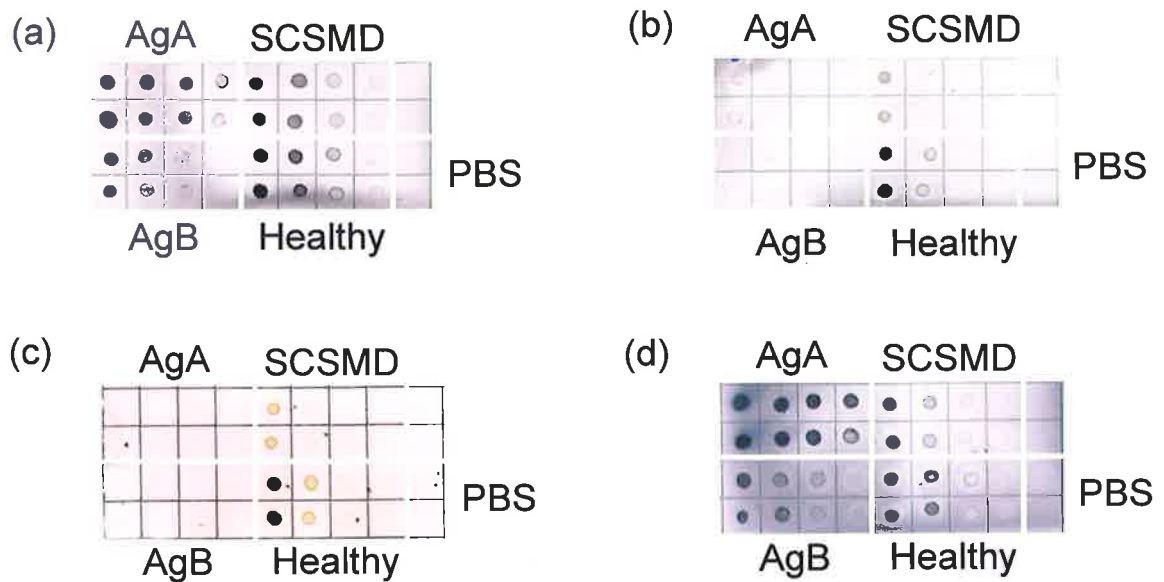


Figure 6.6

DIBA with antisera A and B against dilutions of antigen A (AgA), antigen B (AgB), SCSMD affected sugarcane leaf extract and healthy sugarcane leaf extract. Each sample was dotted in duplicate and at four concentrations. PBS was used as a negative control.

(a) Antiserum A

(b) Antiserum B

(c) DIBA excluding goat-anti-rabbit phosphatase conjugate using antiserum A

(d) DIBA with alternate blocking buffer containing 600 mM glucose and 600 mM mannose using antiserum A

endogenous phosphatase activity of the samples showed no reaction (Figure 6.6c; Note that the dots visible on the membrane were green without evidence of blue, indicating a negative serological reaction). An alternate blocking buffer containing 600 mM mannose and 600 mM glucose did not increase the sensitivity of the antisera for SCSMD affected sugarcane (Figure 6.6d).

6.3.5 Western blot

The specificity of both antisera was assessed by Western blot analysis (Figure 6.7). An identical SDS-PAGE gel to the gels used for Western blots is shown stained for total protein (Figure 6.7a). Antiserum A showed a reaction to the homologous antigen, antigen B, and to multiple protein bands from non-induced and induced *E. coli*. There was little or no reaction to either SCSMD affected or healthy sugarcane extracts (Figure 6.7b). Antiserum B showed a reaction to the homologous antigen and to multiple protein bands from non-induced and induced *E. coli*. There was little or no reaction to the healthy or SCSMD affected sugarcane extracts or antigen A (Figure 6.7c).

6.4 Discussion

Fractionation of the virus preparation on the Cs_2SO_4 gradient showed a light scattering zone at 1.32 g/l. Isolated virus particles were observed from around this zone after further fractionation by Nycodenz[®] gradient centrifugation (Figure 6.3). The density of SCSMaV particles in Cs_2SO_4 gradients was reported to be 1.27 g/l (Choi, 1997), and the yield was expected to be low (Choi *et al.*, 1999). It is possible that the bands collected and viewed by electron microscopy did not contain the majority of particles. The low yield and possible contamination made these virus preparations unsuitable for the production of antibodies. An alternate strategy of cloning and expressing the SCSMaV-

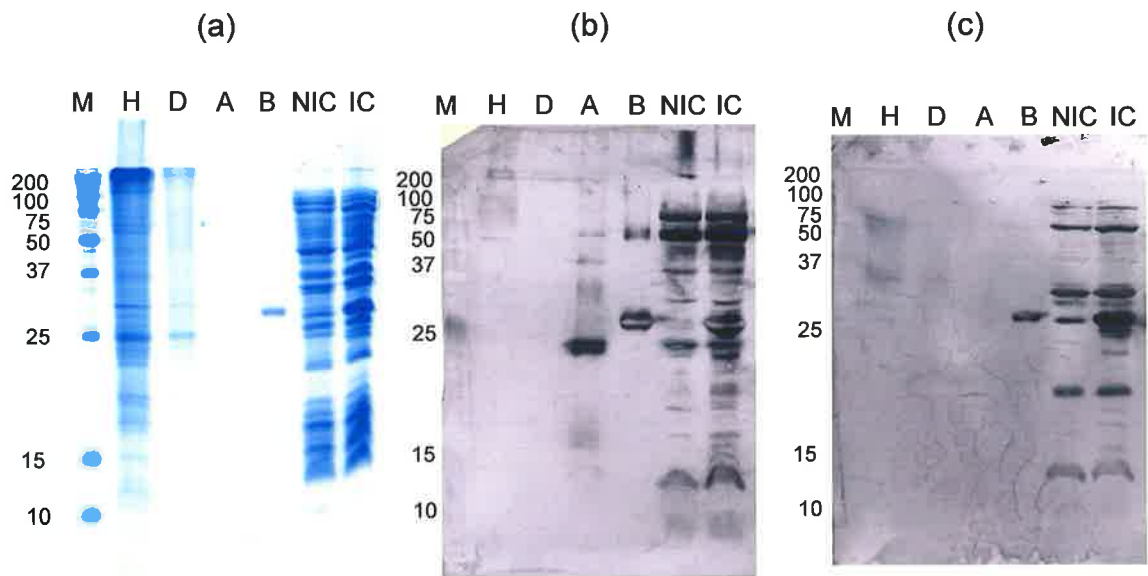


Figure 6.7

Analysis by 15% SDS-PAGE and Western blot assay testing the specificity of antisera A and B

(a) Stained gel showing total protein identical to gels used in the Western blot assays

(b) Western blot assay of SDS-PAGE identical to (a) probed with antisera A

(c) Western blot assay of SDS-PAGE identical to (a) probed with antisera B

M = Prestained SDS-PAGE protein markers with sizes shown adjacent (kDa)

H = 20 μ l healthy sugarcane extract

D = 20 μ l SCSMD affected sugarcane extract

A = 20 μ l aliquot of antigen A

B = 20 μ l aliquot of antigen B

NIC = Non-induced control sample of cells before induction with IPTG

IC = Induced control sample of cells after induction with IPTG

CP using the QIAexpressionist system (Anon., 1999b) was therefore used to produce antigen.

The complete SCSMaV-CP was cloned into the bacterial expression vector pQE-31 with the start codon removed and a 6xHis tag inserted upstream. This was initially transformed into *E. coli* JM109 cells for propagation and then transformed into *E. coli* M15[pREP4] for expression.

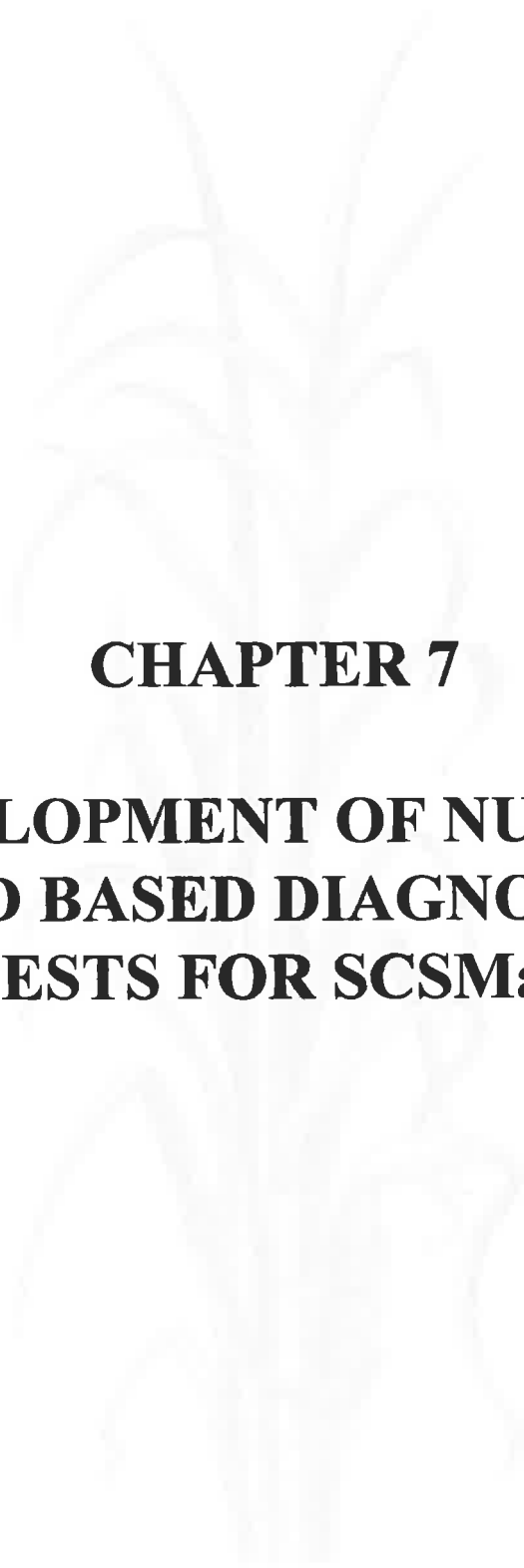
Multiple protein bands were observed by SDS-PAGE analysis after the expression and fractionation of the protein under native conditions. The induced protein appeared to be in the pellet, and the induced protein band of ca. 27 kDa (antigen B) was excised from the gel and used to produce antibodies (antiserum B). Expression and purification of the protein under denaturing conditions yielded ca. 50 µg protein of ca. 25 kDa (antigen A) of which ca. 8 µg was used per injection to produce antibodies (antiserum A).

Neither of the antisera produced were specific for SCSMD affected sugarcane by DIBA. No reaction was observed when goat-anti-rabbit phosphatase conjugate was excluded from the DIBA, indicating that the reaction with healthy sap was not due to endogenous phosphatase activity of the samples. The use of an alternate membrane blocking solution containing 600 mM mannose and 600 mM glucose (Ligat *et al.*, 1991; Ali and Randles, 1997) did not increase the specificity of the reaction for SCSMD samples over healthy samples. When the antisera from the earlier test-bleeds were compared by DIBA, no difference was found between the specificity of the reactions.

Although neither antiserum was specific for SCSMD affected sugarcane by Western blot assay, both reacted with their homologous antigens and with multiple protein bands from non-induced and induced *E. coli*, thus neither were suitable for the detection of SCSMaV by DIBA or Western blot analysis.

Both antisera had strong reactions to many proteins in the bacteria used for the expression of the SCSMaV-CP and both antisera reacted with their homologous antigens and to proteins from *E. coli*. It is therefore possible that the protein used for the production of antisera was contaminated with antigen from *E. coli*. To see if there are SCSMD specific antibodies in either antiserum, it could be extensively absorbed with untransformed *E. coli* extracts and used for Western blot analysis.

Alternatively, the antisera produced could be directed towards SCSMaV-CP epitopes that are not exposed on the SCSMaV particle. This may be a result of the system used for the expression of the protein, as it has recently been shown that different levels of expression and products are observed for the expression of the same product in the Baculovirus and *E. coli* systems (Merits *et al.*, 2000). It may be necessary to attempt the expression of the SCSMaV-CP in an alternate system to get the correct tertiary structure of the protein so that antibodies can be generated to specifically detect SCSMaV-CP in SCSMD affected sugarcane extracts by serological detection methods. It is also possible that SCSMaV is not present in the extracts in a high enough concentration for the SCSMaV-CP to be detected by either DIBA or Western blot analysis. Therefore more sensitive diagnostic assays, such as RT-PCR and hybridisation assays, were developed (Chapter 7).



CHAPTER 7

DEVELOPMENT OF NUCLEIC ACID BASED DIAGNOSTIC TESTS FOR SCSMaV

CHAPTER 7: DEVELOPMENT OF NUCLEIC ACID BASED DIAGNOSTIC TESTS FOR SCSMaV

7.1 Introduction

Nucleic acid sequence information allows the development of nucleic acid (NA) diagnostic methods by identifying viral nucleic acids in extracts of diseased plants. For example: oligonucleotide primers can be developed for RT-PCR amplification of part of the virus genome; clones or amplified fragments can be radioactively or non-radioactively labelled to be used as a probe for tissue or nucleic acid blot hybridisation assays; the RT-PCR amplicon can be detected by AGE and used as a diagnostic test. This chapter describes the development of hybridisation and RT-PCR based assays for the detection of SCSMaV-RNA.

7.2 Materials and methods

7.2.1 NA sample preparation

7.2.1.1 Dot blot

TNA was extracted (2.2.2.1) and 5 µl dotted on Zetaprobe[®] nylon membrane and fixed by crosslinking (2.2.8.2).

7.2.1.2 Tissue blot

A leaf piece ca. 1 cm long was cut from diseased and healthy sugarcane and either frozen briefly in liquid nitrogen or placed directly on Magna[®] nylon membrane. It was covered with a plastic sheet and squashed onto the membrane using a roller (Figure 7.1). Cane pieces were cut and immediately imprinted on Magna[®] nylon membrane. The exudate was fixed by crosslinking (2.2.8.3).

7.2.1.3 Extraction of insect nucleic acid

Table 7.1 lists the insect specimens used for RT-PCR analysis, and the number of insects used for TNA extractions (2.2.2.2).

7.2.2 Hybridisation of membranes

Hybridisation was done at 42°C for 16 hours (2.2.8.5) with ³²P labelled cons7 probe produced by amplification of part of the RdRP (2.2.8.4, Figure 7.2). The membranes were washed at low stringency using Magna[®] Wash Buffer 1 (Appendix C) twice for 15 minutes each at 25°C, and once at high stringency using Magna[®] Wash Buffer 2 (Appendix C) for 15 minutes at 25°C.

7.2.3 RT-PCR

7.2.3.1 SCSMaV-RdRP specific RT-PCR

TNA was reverse transcribed with primer L6 using AMV Reverse Transcriptase (2.2.6.2b) and PCR was done with primers U1 and L3 and *Taq* DNA polymerase (2.2.6.3b; Figure 7.2). PCR conditions were an initial denaturation at 94°C for 5 minutes followed by 30 cycles of 94°C denaturation for 30 seconds, 65°C annealing for 30

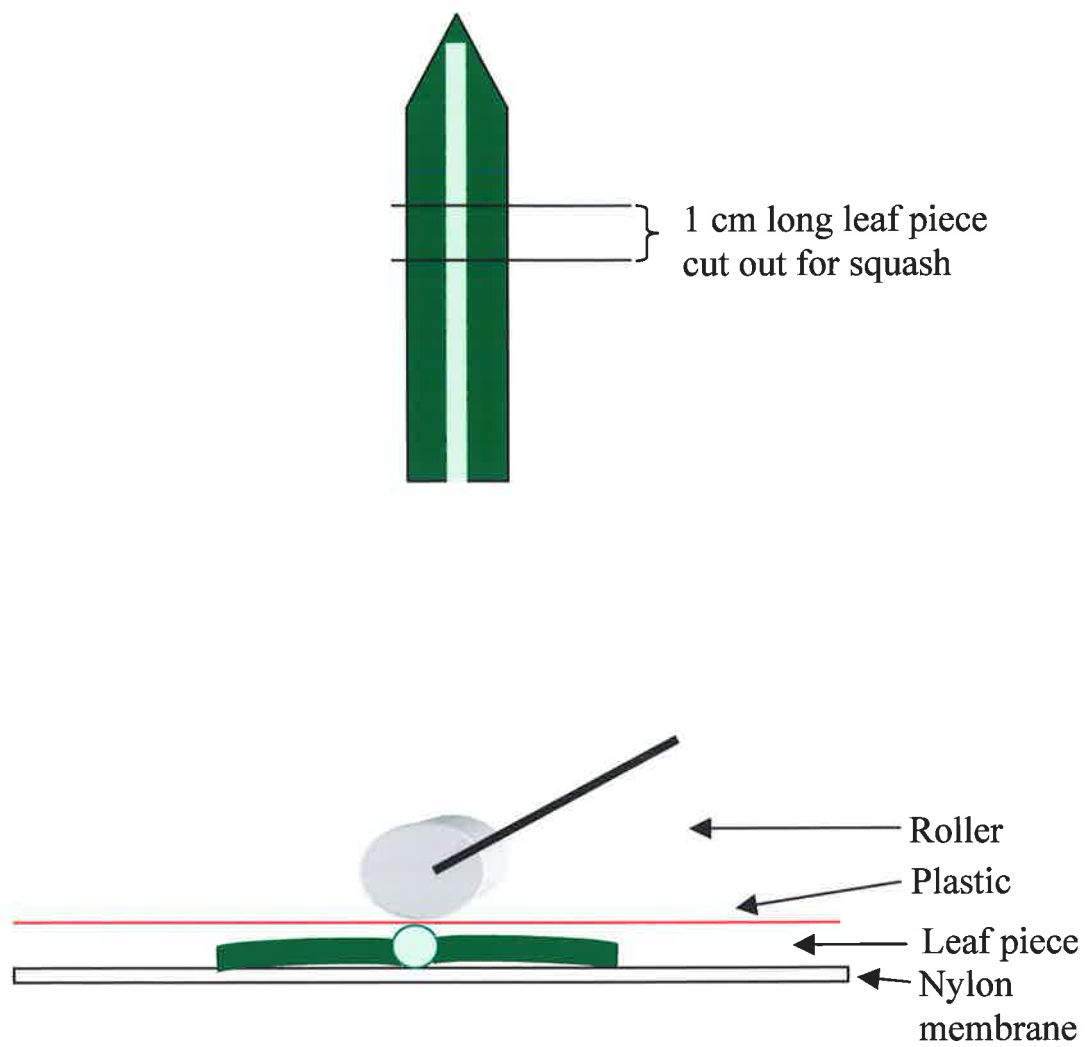


Figure 7.1

Diagram illustrating the preparation of membranes for tissue blot hybridisation.

The leaf was imprinted on the membrane by rolling along the mid-rib of the leaf piece.

Table 7.1 Insect species collected from the roots of SCSMD affected sugarcane.

Collected by K.J. Chandler at a site on Klondyke Rd., ca. 3 km from Ayr, Queensland.

No. indicates the number of insects used per extraction.

Insect	Identification	No.
Leafhoppers (found in ants nests)	Cicadellidae: Xestocephalinae <i>Myrmecophryne formiceticola</i> Kirkaldy	5
Bugs	?Lygaeidae (nymphs only)	3
Aphids	Pemphigidae: Pemphiginae <i>Geoica lucifuga</i> Zehntner	>10
Mealybugs	Pseudococcidae <i>Saccharicoccus sacchari</i> Cock II	>10

Insects identified by K. Chandler, BSES and Dr. M. Fletcher, NSW Dept of Agriculture.

Figure 7.2

Primers and probe used for SCSMD diagnostic tests.

Partial NA sequence of SCSMaV-RNA shown from nt 4151 – 5450 (part of SCSMaV-*RdRP*) and nt 7107 – 7790 (encompassing the SCSMaV-*CP*). '>>' indicates a break in the NA sequence

The region amplified to generate the 'cons7' hybridisation probe is shown in pink.

The annealing positions for:

SCSMaV-*RdRP* specific RT-PCR primers L6 to prime reverse transcription [red], and U1 and L3 for PCR [blue]

SCSMaV-*CP* specific RT-PCR primers CP-L and CP-U [green]

5' >>

4151	UCAUUCUCAU	AACUACUAGU	UCAGAAAAAGG	CAAGUGAGAG	GAGAUGGGUA	U1
4201	ACAGCAUUA	GCAGAGCAUC	GACAAGCAUA	UCUUUUCUUA	ACCUGACUGG	
4251	CAUUGGAAUA	CAAGAGCUUA	UAUCGACUAC	UUUUUUUGGC	AGAUUCUUGC	
4301	AAAAUUUUU	AAGUGCUGAA	UGCAAGAGGG	AGGAUAUUAA	GCCAUAUUUA	
4351	UACGGAGAGC	CAAAAUUUGU	UGAUGACUUC	AGGAGCAUAG	GAAAGACGAAG	
4401	UGCAGAGGAG	AAAGAAGUCA	AAGUUGAUGG	UGAUCCGUGG	UUAAAGUGCA	
4451	UCUUGCAUUU	CUUUGAAACU	CCAAUCCAG	AUGAAGUUGA	GUGUGUUAAA	
4501	GUUGAUUUGG	UAGAGCAAAA	AGCAAAAACU	CACUUACCGG	UUUCGUCCUU	
4551	UGAGGUUUUA	UGGGCCCAUU	UUAAUGAAAA	ACUUUUGCUU	AAGGAGAUCA	
4601	GGGAGCGGAG	GAUAGAUGAU	CUUGUUUCUG	ACCAGUUUGC	AUGGGAUCAU	
4651	AGGUCACAAG	GACUAGCAAU	UAAUAAUGCU	GCCACAAGGU	AUGAAUCAAU	
4701	AUUCCCUAGG	CAUCACCACA	AUGAUAAAGU	GACCUUUUUG	AUGGCAGUUA	
4751	AGAAAAGGUU	GACCUUUGGA	UCGCCCAGAA	AGAAUGGGAG	GAUGCUUAGA	
4801	GAGGCUAUGC	CCUAUGGUGA	AAUUAUACUC	AAGGAGUUUC	UCAAGAGGGU	
4851	CCCUAUUAAG	AAUGAUAGAA	ACGAUCAUUU	AUUUUUAGAA	UGUCUAAUGG	
4901	AUUUUGAAGA	AAAAAAAACU	UCCAAGGACU	CAGCCACAAU	CGCAAAUCAC	
4951	UGCAUACGCU	CACAGAAUUC	AUGGCCAAUU	GAUAUUGCUG	AUAUAUUCAU	
5001	GAAGUCACAA	CUCUGCACAA	AAAUGGAGAG	UAUGUUCUGC	GAGGCAAAGG	L3
5051	CUGGUCAGAC	UUUAGCUUGU	UUCCAUCAUG	UAGUUCUUGC	ACGAUUUUCU	
5101	CCAUGGGUCC	GUUACAUAGA	AAAGAAGCUG	GUCAGUUCAC	UGCCAGAUAA	
5151	UUACUAUUA	CACUCUGGAA	AGAACUUUGA	UCAACUCAAU	GCUUGGGUUA	
5201	AGGCAAACAA	UUUUGUGGGU	GAAUGCACUG	AAUCGGAUUA	UGAGGCCUUU	
5251	GAUGCAAGUC	AAGAUCACUA	UAUACUGGCC	UUUGAAUUAG	CAUUAUAUGAA	
5301	GUACCUUGGU	UUACCCAAUG	AGUUGAUAGA	GGACUACAAG	UACAUUAAGA	
5351	UUCAUUUGGG	UUCAAAAUUG	GGUGAUUUUG	CCAUAUUGAG	AUUCACCGGU	
5401	GAAGCCAGCA	CUUUCUUAUU	UACACAAUUG	GCAAAUAUGC	UUUUCACUUU	
5451	UCUGGUCUUA	GAUAUCAAAAG	GCAAUGAGUC	AAUUUGCUUU	GCUGGCGAUG	
5501	ACAUGUAUGC	AAAUACAAGU	CUUAAGACAU	GCAGCAAGAA	CAAAAGUUUG	
5551	CUUAAAAAGU	UUACACUCAA	AGCUAAAAGU	CAGGUAACUG	AAAGCCCAAC	L6
5601	AUUUUGCGGA	UGGCACUUGA	CAGAGGAUGG	AAUUUAUAAG	GACCCAAUAC	
>>	>>	>>	CP-U	>>	>>	
7101	GAUGAGUGGG	CCACCCCUU	CAACUGGUGC	CAAUGCUUCC	AGUGGUGCUC	
7151	AACCAACAAC	UCAGCAGGAA	ACGAGCACAG	UUUUUAAGAA	AAGCCGCAGA	
7201	GGAUCAAAAU	CUAAAUUUGU	UUCUGCUGUU	ACAAAGGAGC	AGGUUAGGAG	
7251	GGUGGUCGAG	CUUGACAUUG	AGACAGACAA	AAUAGCCCGG	GAUGAGGACC	
7301	UUGAGAAAGU	GAAAGCUGCU	UGGCUUGCCA	UUAAUGUGCC	AUCCGACAAG	
7351	AUAUUUGAAG	UUUCCCUUGA	AAUAACACUU	UCCUGUGCUC	ACACUUCGAC	
7401	UUCAAAUUAG	CAGAAAUUUA	AAGGCAAGUC	CAAAUUCUGC	GACAUAUUGC	
7451	UUGCUGAAAU	AGCUGGCUGU	AUCACUAAGG	AGGAGGUAAC	AAUACGUUCC	
7501	UUUUGUGCUU	AUUGGGCAAA	AUACGUCUGG	AACGCAAUGC	AUGAACGUAA	
7551	UGAGCCUCCU	CUCAAUUGGG	UUGCUCUUGG	GUAUACAGAA	CAAAACCGUU	
7601	UUGCAGCUUU	UGACUUUUUC	UAUGGUGUUG	ACUCUCCUGC	UUCAGCAGAA	
7651	CCAGAAGGUG	GGUUACUUGC	UCCUACCACU	GAGUCUAUGA	GAGCUGCAAA	
7701	CAUGGCUAGG	AAAGAGAUAG	CCAUCAAGA	AUGAGAGAAA	GGGAGGGUAU	CP-L
7751	GUCAUAUUUG	AAUCUGUCAG	CAGUCACAGG	UGCCAGGAAU	>> 3'	

NA
hybridisation
probe
"cons7"

seconds and 72°C extension for 1 minute, with a final extension of 10 minutes at 72°C.

RT-PCR products were analysed by 1.5 % AGE (2.2.1.1).

7.2.3.2 SCSMaV-CP specific RT-PCR

TNA was reverse transcribed with primer CP-L using AMV Reverse Transcriptase (2.2.6.2b) and PCR was done with CP-L and CP-U and *Taq* DNA polymerase (2.2.6.3b; Figure 7.2) with PCR conditions as described above, except that annealing was done at 60°C and the extension was for 30 seconds at 72°C. RT-PCR products were analysed by 1.5% AGE (2.2.1.1).

7.2.3.3 Incorporation of internal control

TMV-RNA and primers were incorporated into the RT-PCRs above as an internal control for false negative RT-PCRs (2.2.6.3c; Torok and Randles, 2001).

7.3 Results

7.3.1 Nucleic acid dot blot hybridisation

Results of the NA dot blot hybridisation assay of TNA extracts from six SCSMD and four healthy sugarcane plants are shown in Figure 7.3. The SCSMD dots showed strong hybridisation to the 'cons7' probe, and the healthy extracts showed no hybridisation.

7.3.2 Tissue blot hybridisation

The tissue blot hybridisation assay result is shown in Figure 7.4. The fresh SCSMD leaf tissue showed hybridisation to the probe, and a weaker reaction was observed for frozen leaf tissue and the cane cutting. No hybridisation of the probe was observed for the

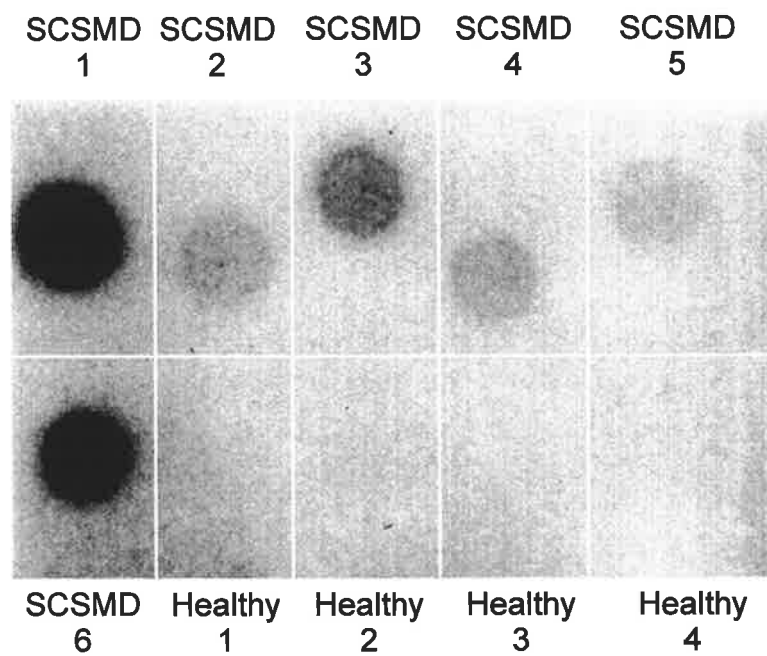


Figure 7.3

Autoradiograph of the NA dot blot assay showing hybridisation of the cons7 probe.

SCSMD affected sugarcane TNA samples 1-6

Healthy sugarcane TNA samples 1-4

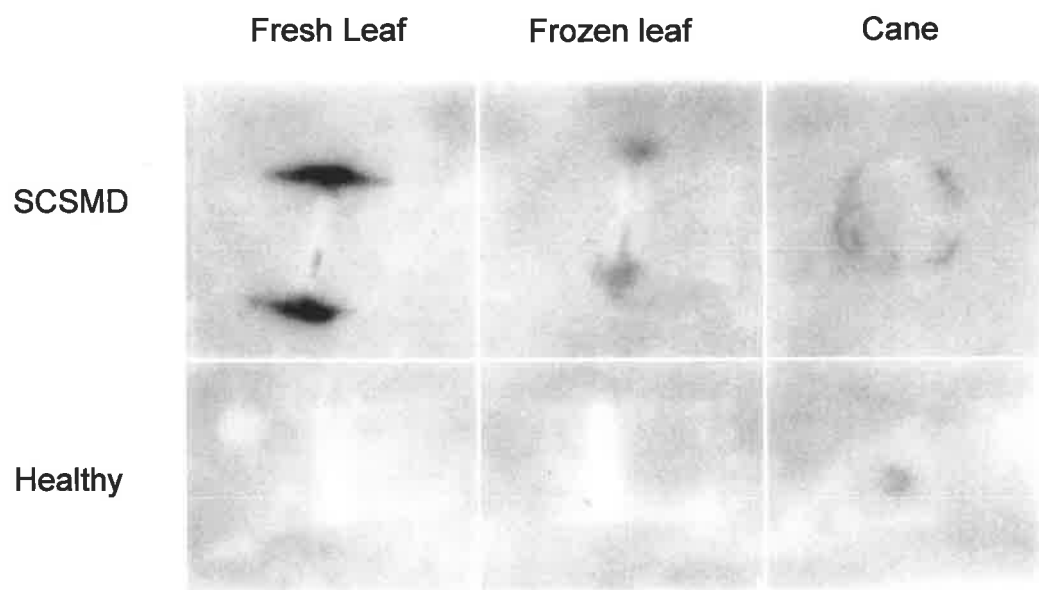


Figure 7.4

Autoradiograph of tissue blot assay showing the hybridisation of the ^{32}P labelled 'cons7' probe.

The samples "Fresh leaf" "Frozen leaf" and "Cane" are shown for both SCSMD affected (upper) and healthy (lower) sugarcane tissue.

The leaf samples had the mid-ribs ca. vertically aligned on the membrane.

healthy leaf sample, either fresh or frozen, but a weak signal was observed for the tissue blot from the healthy cane cutting (Figure 7.4). When this was repeated on a nitrocellulose filter no signal was observed for healthy cane samples (results not shown).

7.3.3 RT-PCR

7.3.3.1 SCSMaV-RdRP specific RT-PCR

The SCSMaV-RdRP specific RT-PCR amplification products from 28 TNA extracts are shown in Figure 7.5. No RT-PCR products were observed for any healthy (H) plants. All SCSMD affected sugarcane TNA extracts assayed, except D4 and D9, showed an 850 nt RT-PCR product

7.3.3.2 Duplex RT-PCR

A control for false negative reactions (2.2.6.3; Torok and Randles, 2001) was incorporated in the tests developed previously (7.3.3.1) to test the reaction conditions for each RT-PCR. The TMV primer pairs were shown not to amplify healthy sugarcane TNA and did not interfere with the amplification of the SCSMaV-RdRP by U1 and L3 (Figure 7.6). No interference of the SCSMaV-RdRP RT-PCR was observed when the TMV positive control components were added to the reaction. There was no RT-PCR amplification of healthy sugarcane TNA samples using the TMV primers, with only the TMV-RNA being amplified, and there was no amplification of the TMV-RNA using the SCSMaV specific primers (Figure 7.6c). Figure 7.7 shows the amplification products from RT-PCR with incorporated TMV-RNA control for 6 diseased and 4 healthy samples analysed by 1.5% AGE (2.2.1.1). TMV-RNA specific 400 nt products and SCSMaV-RdRP specific 850 nt products were present in samples D1, D2, D3, D5 and



Figure 7.5

Analysis by 1.5% AGE of the SCSMaV-*RdRP* specific RT-PCR on SCSMD affected and healthy sugarcane TNA extracts.

M - 1kb Plus DNA ladder with sizes shown adjacent

D1 - D14: SCSMD affected sugarcane

H1 - H14: Healthy sugarcane

The arrow indicates the 850 nt product amplified by the SCSMaV-*RdRP* RT-PCR.

An asterisk indicates samples D4 and D9 which are SCSMD affected sugarcane TNA extracts that show no RT-PCR amplification of the SCSMaV-*RdRP*.

The size of the marker is shown in bp.

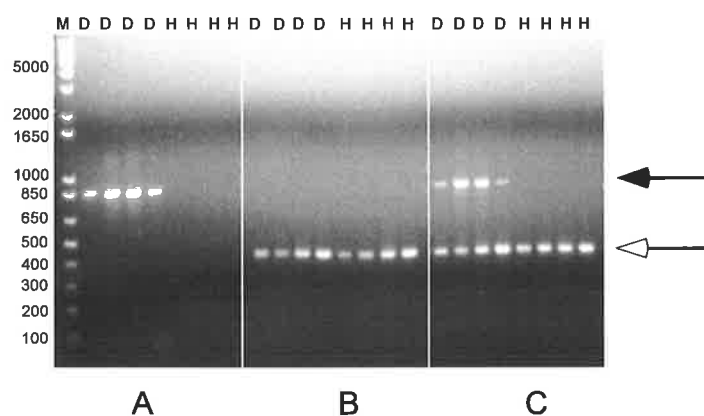


Figure 7.6

Analysis by 1.5% AGE of RT-PCR amplification products.

TMV-RNA was added as the internal control template to a reaction containing either TNA extracts from SCSMD affected (D) or healthy (H) sugarcane. The primers used in the RT-PCR were varied:

- (a) Only SCSMaV-*RdRP* specific primers present
- (b) Only TMV specific primers present
- (c) Both SCSMaV-*RdRP* and TMV primers present in a duplex RT-PCR

M - 1kb Plus DNA ladder with sizes shown adjacent

The block arrow shows the 850 nt amplified region of the SCSMaV-*RdRP*.

The open arrow shows the 400 nt amplified region of the TMV RNA.

The size of the marker is shown in bp.

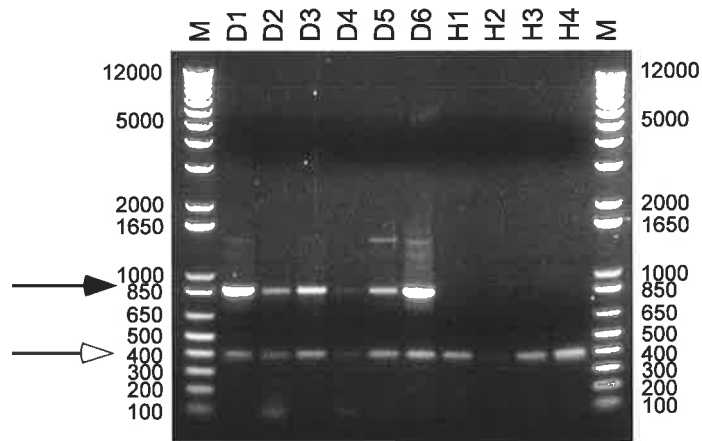


Figure 7.7

Analysis by 1.5% AGE of duplex RT-PCR amplification from TNA extracts from SCSMD affected and healthy sugarcane.

M - 1kb Plus DNA ladder with sizes shown adjacent

D1 - D6: SCSMD affected sugarcane

H1 - H4: Healthy sugarcane

The block arrow shows the 850 nt amplified region of the SCSMaV-*RdRP*.

The open arrow shows the 400 nt amplified region of the TMV-RNA.

The size of the marker is shown in bp.

D6 while only the TMV-RNA product was present in samples H1, H3 and H4. Samples D4 and H2 had no products, showing that the RT-PCR had failed and that the tests were invalid.

7.3.3.3 SCSMaV-CP specific RT-PCR

The SCSMaV-CP specific RT-PCR was done in duplex with TMV-RNA and primers as above, and the amplification products for four healthy and two SCSMD affected sugarcane TNA samples were analysed by 1.5% AGE (Figure 7.8). The TMV-RNA RT-PCR product was present in all samples, and the 650 nt SCSMaV-CP specific RT-PCR product was found in the SCSMD affected samples only.

7.3.3.4 Insect RT-PCR

The SCSMaV-CP specific RT-PCR amplification product from TNA extracts of the aphid, mealybug, lygaeid bug and leafhopper specimens are shown in Figure 7.9. The RT-PCR was done without (Figure 7.9a) or with (Figure 7.9b) the TMV internal control. Reactions that had suitable conditions showed amplification of the internal control. Amplification without the incorporation of the internal control appears to give stronger staining products by AGE, and it was only possible to amplify both the internal control and the SCSMaV-CP when the TNA was diluted.

7.4 Discussion

7.4.1 Hybridisation assays

The nucleic acid based hybridisation tests described detect SCSMaV-RNA from either TNA extracts or leaf tissue blots. The leaf tissue blot showed that the SCSMaV-RNA was present in the exudate that was squashed out of the midrib onto the nylon. This

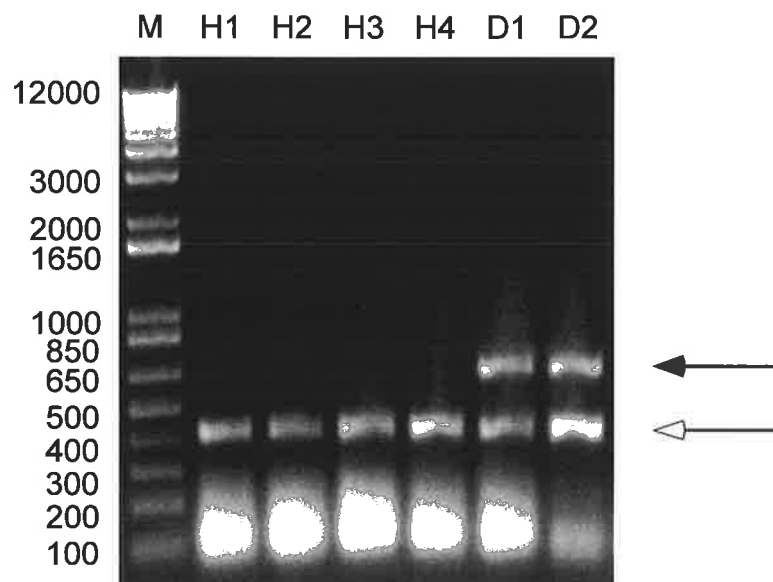


Figure 7.8

Analysis by 1.5% AGE of duplex RT-PCR amplification from TNA extracts from SCSMD affected sugarcane and healthy samples.

M - 1kb Plus DNA ladder with sizes shown adjacent

D1 - D2 : SCSMD affected sugarcane

H1 - H4: Healthy sugarcane

The block arrow shows the 650 nt amplified region of the SCSMaV-CP

The open arrow shows the 400 nt amplified region of the TMV-RNA.

The size of the marker is shown in bp

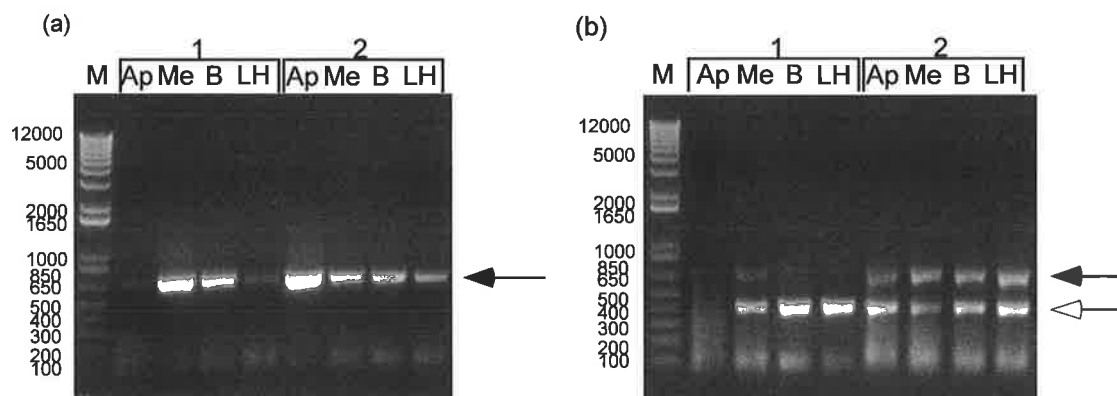


Figure 7.9

Analysis by 1.5% AGE of SCSMaV-CP specific RT-PCR products from insect TNA samples (see Table 7.1)

Ap - *Geioica lucifuga*

Me - *Saccharicoccus sacchari*

B - Lygaeid bug

LH - *Myrmecophryne formiceticola*

1 - undiluted TNA used for reverse transcription

2 - TNA diluted 1:10 before reverse transcription

(a) without TMV internal control

(b) with TMV internal control

Block arrow indicates 650 nt amplified fragment from SCSMaV-CP

Open arrow indicates 400 nt amplified fragment from TMV-RNA control

M - 1kb Plus DNA ladder with sizes shown adjacent in nt

indicates that SCSMaV was present in the midribs of the sugarcane. The single weak reaction to the healthy cane sample may be non-specific, as a repeat of this experiment using nitrocellulose membrane showed no hybridisation of the probe to healthy cane.

7.4.2 RT-PCR assays

RT-PCR amplification of part of the SCSMaV-*RdRP* gave a single diagnostic product of 850 nt in SCSMD affected sugarcane TNA samples, and no product in healthy sugarcane TNA samples. A second diagnostic RT-PCR was developed to amplify the SCSMaV-*CP* and gave a product of 650 nt. Some SCSMD affected sugarcane TNA samples gave no RT-PCR product when analysed by AGE, indicating that the reaction conditions were not suitable for amplification. An exogenous internal control developed by Torok and Randles (2001) was used to test the reaction conditions and was successfully incorporated into most RT-PCR reactions, however some interference of the amplification of SCSMaV-*CP* was observed (Figure 7.9a and 7.9b). Dilutions of the TNA were required to achieve amplification of both products. This implies that although the internal control duplex RT-PCR is a good indicator of suitable RT-PCR amplification conditions all template and primer concentrations should be thoroughly tested before incorporating this control into routine procedures.

When identical samples were used for RT-PCR analysis a more strongly staining product was observed for the SCSMaV-*CP* RT-PCR than for the SCSMaV-*RdRP* RT-PCR (results not shown). This may be because the smaller product (650 vs 850) was more easily amplified under the conditions specified (Fenby, 1995). Another possibility could be the presence of SCSMaV sub-genomic RNAs containing the SCSMaV-*CP* in the TNA sample, that increased the amount of target RNA to be amplified, and therefore the amount of RT-PCR amplification product.

Detection of SCSMaV from four hemipteran insect samples (Table 7.1) showed that the SCSMaV-*CP* RT-PCR amplification has a potential application for the detection of SCSMaV in vectors. This preliminary result could implicate any of these insects as a possible vector for SCSMaV. The insects were collected from the roots of diseased sugarcane so SCSMaV-RNA may have been present in the gut or tissue of the insect. These preliminary experiments should be followed up with transmission and localisation studies using the SCSMaV-*CP* specific RT-PCR.

7.4.3 Recommendations

The hybridisation assays developed are less labour intensive than RT-PCR, but differences are observed in the intensity of positives. RT-PCR has been shown to be more sensitive than hybridisation assays (Fenby, 1995) and the SCSMaV-*CP* specific RT-PCR duplexed with the TMV internal control provides for negative reaction control. Interference between the SCSMaV-*RdRP* and the TMV-RNA reactions in the duplex RT-PCR were overcome by using at least two dilutions of the TNA extract for detection of SCSMaV in this study. All of the tests described were developed using glasshouse maintained SCSMD affected plants. The possibility remains that there may be strains of SCSMaV that will not be able to be detected by all or any of the tests described, and more than one type of assay for diagnosis of SCSMD is recommended.

CHAPTER 8

GENERAL DISCUSSION

CHAPTER 8: GENERAL DISCUSSION

8.1 Introduction

The primary aim of this project was to complete the sequencing of the SCSMaV-RNA and to determine the genome organisation.

Previous studies had associated dsRNA with SCSMD (Anon., 1985; Choi 1997; Choi and Randles, 1997). Random PCR (rPCR) clones were generated from the associated dsRNA, and one was selected as a probe to screen a virus preparation. The probe hybridised to a fraction enriched for viral particles, and the associated virus was observed by electron microscopy and named SCSMaV (Choi *et al.*, 1999). Transmission of SCSMD to sugarcane or other plants by inoculation of SCSMaV or SCSMaV-RNA has not been done, and thus Kochs postulates to implicate SCSMaV as the sole agent of SCSMD have not been demonstrated.

The nucleic acid sequence of five rPCR clones was determined (Choi, 1997; Choi *et al.*, 1999), and this gave a discontinuous sequence of ca. 3 kb that showed sequence similarity to viral RNAs from the genera *Foveavirus*, *Carlavirus*, *Potexvirus*, *Capillovirus*, *Trichovirus* and *Potyvirus* (Choi, 1997; Choi *et al.*, 1999). The sequence with the highest similarity to SCSMaV was ASPV, type member of the genus *Foveavirus* (Choi *et al.*, 1999).

This thesis now reports the determination of the remainder of the SCSMaV-RNA sequence by RT-PCR amplification of dsRNA and ssRNA using primers designed from the known sequence. The SCSMaV-RNA sequence has been used to develop nucleic acid based detection methods.

8.2 Analysis of SCSMaV-RNA nucleotide sequence

The SCSMaV-RNA was 8146 nt, polyadenylated at the 3' end and five potential ORFs were identified encoding a putative RdRP (ORF1), TGB encoding putative movement function proteins (ORF2, 3 and 4) and a putative CP (ORF5).

The complete nucleotide sequence of SCSMaV-RNA was found to be most similar to SCGRMV (38%), BBScV (37.9%), GarLV (37.5%), PVM (37.5%) and RSPaV (33.7%). ASPV, which was found to be most similar by previous analysis of the partial sequence (Choi, 1997; Choi *et al.*, 1999), was the sixth most similar (31.6%) when the full nucleotide sequence was analysed. Nucleotide sequence identities between species in the same genus range from 32-41% for the genus *Foveavirus*, 43-46% for the genus *Carlavirus* and 30-36% for the genus *Potexvirus* (Appendix D). SCSMaV appears to be as closely related to species from the genera *Foveavirus* and *Carlavirus* as species in these two genera are to each other (PVM and ASPV have nucleotide sequence identity of 33.8%). The phylogenetic analysis of the full nucleotide sequences of a range of genera support these nucleotide identities and show that SCSMaV does not cluster within established genera and is located between members of the genera *Foveavirus* and *Carlavirus* and closest to SCGRMV, a tentative member of the genus *Foveavirus*.

8.3 Analysis of ORF1 - RdRP

Analysis of the ORF1 peptide showed that it had characteristics of a RdRP, with eight motifs characteristic of RdRPs, a NTP binding domain and amino acid sequence identity of 32% with the RdRP of BBScV (Chapter 4). The importance of RNA template recognition and polymerase activity of the RdRP is reflected in the high degree of amino acid sequence conservation between all positive sense ssRNA viruses for this protein

(Breunn, 1991; Koonin; 1991; Koonin and Dolja, 1993; Poch *et al.*, 1989). Phylogenies of the RdRP correlate to some extent with other means of classification (Koonin, 1991; Koonin and Dolja, 1993; Poch *et al.*, 1989). The amino acid sequence identity for the ORF1 of the species of the genus *Foveavirus* was >31%, and for the genus *Carlavirus* was >40%. The amino acid sequence identity between members (ASPV and PVM) of these two genera is 36%. SCSMaV had 32% identity with species in the genus *Carlavirus* and 30.2% identity to species in the genus *Foveavirus*. Therefore species in the genera *Carlavirus* and *Foveavirus* were as closely related to each other as they were to SCSMaV, and the identity of SCSMaV was closer to the range of species in the genus *Foveavirus* than of species in the genus *Carlavirus*. This was consistent with the finding for the complete nucleotide sequence. Phylogenetic analysis using the ORF1 putative peptide places SCSMaV near members of both the genera *Foveavirus* and *Carlavirus* and closest to SCGRMV. By Neighbor-Joining analysis, SCSMaV was placed between SCGRMV and the members of the *Foveavirus* genus, however SCGRMV was as closely related to the other species in the genus *Foveavirus* as they were to the members of the genus *Carlavirus* (Figure 5.2a; Appendix D). The Kitsch analysis showed that SCGRMV was more closely related to the other members of the genus *Foveavirus* than members of the genus *Carlavirus* and SCSMaV was found to occupy a taxonomic position between SCGRMV and PVT (5.2b).

8.4 Analysis of ORF2, ORF3 and ORF4 - TGB

ORFs 2, 3 and 4 encode proteins with amino acid sequence similarity to a range of TGB proteins (TGBp). The ORF2 product has a NTP binding site motif, and had the greatest amino acid similarity with RSPaV TGBp1 (29.5%) which suggests that it may encode SCSMaV-TGBp1 and have a helicase function, similar to other TGBp1s.

The ORF3 product showed the greatest amino acid sequence identity to GVA-TGBp2 (27.5%) and BaMV-TGBp2 (27%), which are small membrane spanning proteins. ORF3 is therefore suggested to encode SCSMaV-TGBp2.

The ORF4 product showed the greatest amino acid sequence identity to PVM-TGBp3 (22.5%), which encodes the second small membrane spanning protein of the PVM-TGB. ORF4 is therefore suggested to encode SCSMaV-TGBp3.

Analysis of these three ORF products did not give phylogenetic trees that separated species into their respective genera. This is consistent with the amino acid percentage identities, which did not distinguish species of different genera from each other. It has been suggested that TGBp1 may be a useful tool in the classification of plant viruses that contain a TGB encoding movement proteins (Wong *et al.*, 1998), however none of the four methods used in this study (Neighbor-Joining, Maximum Parsimony, Kitsch and Fitch) separated the species into established genera. Other evidence suggesting that the TGB proteins are not useful for taxonomic determination include a comparative study of the amino acid sequences of movement proteins from a large number of species from a range of genera. This study concluded that the classification of movement proteins based on their amino acid sequence does not correlate with the type of genomic nucleic acid, or with grouping based on phylogenies of replicative proteins, or with virus host range (Mushigian and Koonin, 1993). Thus, SCSMaV-TGB protein sequences were found not to be useful for genus or species demarcation.

The SCSMaV-TGB encodes proteins that are approximately the same size as TGB proteins for the genera *Carlavirus*, *Potexvirus*, *Foveavirus*, and *Allexivirus* (Table 8.1), but genome database comparisons showed that each of the SCSMaV-TGB proteins had similarity to viruses of different genera (compare Tables 4.3, 4.4, 4.5 and Appendix D).

Table 8.1: A comparison of the sizes of open reading frame translation products (kDa), virion characteristics and selected biological properties of SCSMaV with viruses in the most closely related genera, and the tentative *Foveavirus* CGRMV.

Virus genus (member representing ORF product)	ORF 1 (RdRP)	ORF 2 (TGB)	ORF 3 (TGB)	ORF 4 (TGB)	ORF 5 (CP)	ORF 6 (Ala- rich)	Virion characteristics	Host Range: Mono/Dicot	Mode of spread
Unclassified (SCSMaV)	221 ☆	23 ☆	13 ☆	7 ☆	23 ☆	-	Slightly flexuous rod-shaped 950 nm x 15 nm*	1 known host* +/-	Soil-borne, propagation †
<i>Potexvirus</i> ‡ (PVX)	147	26	13	4	21	-	Straight or flexuous rod- shaped 445-775 nm x 11- 15 nm	Wide +/+	Mechanical, one with a Hemipteran vector (PAMV)
<i>Carlavirus</i> ‡ (PVM)	223	25	12	7	34	11	Straight or flexuous rod- shaped 509-720 nm x 11- 18 nm	Wide +/+	Mechanical, Hemipteran vector, seed
<i>Foveavirus</i> § (ASPV)	247	23	13	7	44	-	Very flexuous rod-shaped 800 nm x 15 nm	Very narrow -/+	Propagation
Tentative <i>Foveavirus</i> (CGRMV[▼])	230	25	12	7	30	-	Thin, flexuous rod-shaped 1000-2000 nm x 5-6 nm	Very Narrow -/+	Root grafting

☆ Chapter 4, This Thesis; *Choi (1997), Choi *et al.* (1999); † Anon (1985); ‡ Brunt *et al.* (1996 onwards), Dallwitz, M.J. (1980), Dallwitz, M. J. *et al.* (1993), van Regenmortel *et al.* (2000); § Martelli and Jelkmann (1998), van Regenmortel *et al.* (2000); ▼ Zhang *et al.* (1998), Zagula *et al.* (1989).

This suggests that the SCSMaV-TGB is not closely related to the TGBs of any other known species or established genera.

8.5 Analysis of ORF5 – CP

Analysis of the ORF5 product showed that it had a *Carlavirus* and *Potexvirus* CP signature, and had the highest amino acid sequence identity with PVX-CP (22.9%), CIYMV-CP (22.7%) and RSPaV-CP (22.6%). The SCSMaV-ORF5 was predicted to be encode the ca. 23 kDa SCSMaV-CP. A protein component associated with the SCSMaV virus preparation was ca. 51 kDa (Choi, 1997; Choi *et al.*, 1999), and it is possible that this represents a CP dimer. The amino acid sequence identity for the CP of species in the genus *Foveavirus* appears to fall into two or three distinct groupings, with ASPV, *Pear vein yellows virus* (PVYV) and ALV forming one group with 55-99% identity. RSPaV and SCGRMV have 24% amino acid identity, and therefore may form one or two additional groups, which have 15-20% amino acid identity with the first group. For the CP of species in the genus *Carlavirus*, the amino acid sequence identity is >32% and for species in the genus *Potexvirus* amino acid sequence identity ranges from 20-75%. The amino acid sequence identity between the type members of the genera *Foveavirus* and *Carlavirus* is 16%, between the type members of the genera *Potexvirus* and *Carlavirus* is 25% and between the type members of the genera *Foveavirus* and *Potexvirus* is 18%. Species in the genera *Potexvirus* and *Carlavirus* are more closely related to each other than to species in the genus *Foveavirus* (<20%). This is consistent with the phylogenetic analyses of the CP which showed that SCSMaV-CP falls between members of the genera *Potexvirus* and *Foveavirus* by Maximum Parsimony analysis, and between members of the genera *Carlavirus* and *Foveavirus* by Kitsch analysis (Figure 5.3). The differences between the phylogenetic trees may be indicative of the

different analyses or of the relatively low but similar amino acid sequence identities between members of the *Carlavirus*, *Foveavirus* and *Potexvirus* genera for the CP (Appendix D). In both phylogenetic trees, SCSMaV was closest to SCGRMV.

8.6 Non-coding regions

The 5' UTR of SCSMaV-RNA does not show nucleotide sequence similarity to any sequence in the genome database using FastA or Blast searches, and does not contain conserved sequence elements that are present in members of the *Carlavirus* and *Potexvirus* genera (Foster and Mills, 1991).

The 3' UTR has a polyadenylation signal but not a hexanucleotide motif that is conserved in the genera *Carlavirus* and *Potexvirus* (White *et al.*, 1992). The presence of the polyadenylation signal without the hexanucleotide motif is unique to SCSMaV-RNA.

8.7 Taxonomic classification of SCSMaV

The 'Key to placement of viruses in taxa' (van Regenmortel *et al.*, 2000; and http://www.ncbi.nlm.nih.gov/ICTV/new_virkey.html) uses criteria including the genome type, virion morphology, replication strategy, host, vector and the number and size of structural and non-structural proteins to place a virus into a genus. To discriminate between virus species in the same genus the characteristics analysed are: genome sequence relatedness, natural host range, cell and tissue tropism, pathogenicity and cytopathology, the mode of transmission, physicochemical properties and antigenic properties of virions (Table 1.2; van Regenmortel *et al.*, 2000).

The classification of SCSMaV using the 'Key' described above suggests that SCSMaV has characteristics similar to members of the genus *Foveavirus*. Several algorithms used for phylogenetic analysis of the full nucleotide sequence, and the SCSMaV-RdRP and SCSMaV-CP amino acid sequences showed SCSMaV was most closely related to members of the genus *Foveavirus*, and in almost all analyses showed the greatest similarity to SCGRMV. The characteristics of the genera *Potexvirus*, *Carlavirus*, *Foveavirus* and the tentative *Foveavirus* SCGRMV are compared in Table 8.1. SCGRMV has been included separately because it has some characteristics that are not consistent with the other members of the genus *Foveavirus* including a long and thin flexuous virion, a CP with a size approximately two thirds of that of the type member ASPV, and a slightly smaller RdRP (Table 8.1).

SCSMaV is excluded from the genus *Carlavirus* as it does not have a sixth ORF that is characteristic of this genus. The SCSMaV-RdRP is approximately 1.5 times larger than that of the type member for the genus *Potexvirus* (PVX), and the virion characteristics of SCSMaV are also distinct from viruses in this genus. A comparison of characteristics of members of the genus *Foveavirus* and SCSMaV show that the SCSMaV-CP is about half the size of the ASPV-CP, and is about two-thirds the size of the SCGRMV-CP. Particle morphology distinguishes SCSMaV from SCGRMV and ASPV, and the host range of SCSMaV is limited to the monocotyledon sugarcane, whereas members of the *Foveavirus* genus infect woody dicotyledonous hosts such as apple (ASPV), pear (PYVY), cherry (SCGRMV) and grapevine (RSPaV). Propagation is the only known means of spread for members of the *Foveavirus* genus and the symptoms are the characteristic pitting of the stems, whereas SCSMaV is apparently soil-borne and causes a striate mosaic on the leaves (Anon., 1985; Choi, 1997; Hughes, 1961a; Hughes, 1961b).

The characteristics of all genera considered in this study including the data from Chapter 3 and Chapter 4, and the virus genera included in phylogenetic analysis from Chapter 5 are compared in Table 8.2.

SCSMaV is most similar to members of the described *Foveavirus* genus yet differs from them on genomic, morphological and biological criteria. It therefore may represent a new plant virus genus.

8.8 Detection of SCSMaV

The accurate detection of SCSMaV is important for the development of a disease management strategy. The attempt to produce antibodies for specifically detecting SCSMD in leaf tissue extracts was not successful. However, a range of nucleic acid based detection methods was developed including nucleic acid dot-blot hybridisation, leaf squash hybridisation and RT-PCR amplification of part of the SCSMaV-*RdRP* or SCSMaV-*CP* of SCSMaV-RNA. Leaf squash hybridisation was the least labour intensive of these assays, and could be further developed to be used with a non-radioactive probe. Detection of SCSMaV-RNA using the SCSMaV-*CP* specific RT-PCR gave strong positive amplification products from SCSMD affected sugarcane leaf TNA extracts and the incorporation of an internal control allowed the detection of false negative assays (Torok and Randles, 2001), thus increasing confidence in the results.

It has been suggested that for the detection of plant virus NA, more than one set of primers should be used for accurate diagnosis of a disease (Nolasco *et al.*, 2000). Two sets of primers (SCSMaV-*RdRP* and SCSMaV-*CP* directed) were designed for the detection of SCSMaV-RNA and many more primers were developed for the sequencing

Table 8.2 Selected characteristics of genera that are similar to SCSMaV.

Taxa	Virion morphology	NA size, features	No. ORFs	Host Range	Transmission	Geographic spread
SCSMaV	900 x 12 nm	8.1, polyA	5	Sugarcane only	?	Limited
<i>Foveavirus</i>	800 x 12 nm	8.4-9.3, polyA, cap?	5	Restricted natural and experimental	No known vector, grafting and mechanical	All have wide geographic distribution
<i>Carlavirus</i>	610-700 x 12-15 nm	7.4-8.5, polyA, cap?	6	Restricted natural, may have wide experimental	Aphids, whitefly, seedborne and mechanical	Varies depending on species
<i>Potexvirus</i>	470-580 x 13 nm	5.9-7, polyA, cap	5	Wide for genus, but individual viruses have limited natural	Mechanical. Only PAMV known to have insect vector	-
<i>Allexivirus</i>	800 x 12 nm	9, polyA	6	Restricted natural, experimental also somewhat restricted	Mite-borne and mechanical	Varies depending on species
<i>Vitivirus</i>	725-825 x 12 nm	7.6, polyA, cap	5	Restricted natural, Experimental can be wide	Mechanical, grafting, mealybugs, scale insect. HLV is transmitted by aphids with a helper virus	Varies depending on species
<i>Trichovirus</i>	640-760 x 12 nm	7.5, polyA	3	Restricted natural, Experimental limited to herbaceous hosts	Mechanical, grafting, seed, no vector known	Varies depending on species
<i>Capillovirus</i>	640-700 x 12 nm	6.5-7.4, polyA	2	Pome fruits, citrus, lily, cherry	No vector, seed, grafting	Varies depending on virus
<i>Tymovirus</i>	Icosahedral, d = 30 nm	6.3, polyA, cap	3	Restricted to dicots	Mechanical, beetles, seed	-

Tabulated from van Regenmortel *et al.*, 2000.

of SCSMaV-RNA (Appendix B). Any pair of these primers that were specific for SCSMaV-RNA could theoretically be developed for the detection of SCSMaV-RNA. In addition, the two RT-PCRs described for the detection of SCSMaV-RNA are directed towards amplification of genes on opposite ends of the SCSMaV-RNA. This may allow the detection of SCSMaV variants with mutations of either the SCSMaV-*RdRP* or SCSMaV-*CP*, or viruses that have undergone recombination. The variability of SCSMaV in the field is unknown, so the use of more than one set of primers may give a greater chance of detecting any SCSMaV-RNA that has nucleic acid sequence variation. The most variable region of positive sense ssRNA viruses is the TGB (Chapter 5, Appendix D). Therefore primers developed for RT-PCR amplification of all or part of this region of the SCSMaV genome may be useful for distinguishing between strains or variants in the field. Other methods of studying SCSMaV-RNA variability in the field could be developed using RT-PCR followed by restriction fragment length polymorphism (RFLP) as has been done for SCMV and Sorghum mosaic virus (Yang and Mirkov, 1997), amplification fragment length polymorphism (AFLP), RNase protection assays (RPA) and direct sequence analysis.

The multiplexing of RT-PCR primer pairs has been reported to be successful for the detection of up to five viruses simultaneously (Nie and Singh, 2000), and both the SCSMaV-*CP* and SCSMaV-*RdRP* specific RT-PCRs were able to be duplexed with TMV-RNA as an internal control. It may be possible to multiplex one of these SCSMaV specific RT-PCRs with established RT-PCRs directed towards other sugarcane viruses (Table 1.1) and still incorporate the internal control. Nevertheless, loss of sensitivity has

been observed in duplex RT-PCR (Torok and Randles, 2001) when the target or primer concentrations are not optimal, and this needs to be considered for such assays.

The sensitivity of RT-PCR makes it a useful tool for the detection of viral NA in a range of tissues, including possible vectors (Singh, 1998). It is suggested that SCSMaV is soil-borne, and four hemipteran insect species were collected from sugarcane roots and tested for the presence of SCSMaV. All four insect species tested were positive. Whole insects were used for these preliminary tests therefore the location of SCSMaV-RNA in the insect was unknown and negative controls were not available. Further study to determine the location of SCSMaV in the insects and transmission studies would be required to determine if any of the insects can vector SCSMaV.

All of the insect specimens collected are related to plant virus vectors. Aphids are prolific vectors of plant viruses with ca. 66% of plant viruses with invertebrate vectors being transmitted by aphids (Matthews, 1991). There are no known vectors in the subspecies of leafhopper studied, however there are over 130 known vector species in the Cicadellidae (Harris, 1990). There are at least two known viruses transmitted by members of the Lygaeidae family: *Beet leaf curl virus* (tentative *Rhabdovirus*) and *Centrosema mosaic virus* (*Potexvirus*) (Brunt *et al.*, 1996 onwards; van Regenmortel *et al.*, 2000), and the pink sugarcane mealybug (*Saccharicoccus sacchari*) can transmit *Sugarcane bacilliform virus* (Comstock and Lockhart, 1996). Therefore it is possible that one of the insects studied could vector SCSMaV.

Of the closely related viruses (Table 8.1 and 8.2), invertebrate vectors are known for species in the genera *Carlavirus* (aphids and whitefly), *Vitivirus* (mealybugs, and scale insects/aphids in association with a helper virus), *Tymovirus* (beetles), *Allexivirus* (mite)

and one *Potexvirus* (aphid transmission of PAMV in association with PVY). No vector is known for species in the genera *Capillovirus*, *Trichovirus* or *Foveavirus*. A study of the possible hemipteran vectors may aid in the taxonomic classification of SCSMaV.

The possibility remains that SCSMaV may have a vector that was not collected in this study such as another insect species, a nematode or fungal vector, or that SCSMaV is retained in the soil without a vector. These potential vectors and soil samples could be tested for the presence of SCSMaV-RNA using the SCSMaV-CP specific RT-PCR and transmission studies could then be done to determine the vector of SCSMaV.

For the accurate and sensitive detection of SCSMaV-RNA, it is suggested that the less labour intensive hybridisation assays be used in conjunction with both the SCSMaV-RdRP and SCSMaV-CP specific RT-PCRs with the internal control.

8.9 Conclusions

This study has made the following contributions to the understanding of SCSMaV:

- (1) The SCSMaV-RNA comprises 8146 nt and is polyadenylated at the 3' terminus.
- (2) SCSMaV-RNA has 5 potential ORFs: ORF1 encodes a putative RdRP; ORF2, ORF3 and ORF4 comprise a TGB encoding putative movement function proteins; and ORF5 encodes a putative CP.
- (3) Comparison of nucleotide sequence and genome organisation indicates that SCSMaV is most closely related to viruses in the *Foveavirus*, *Carlavirus*, and *Potexvirus* genera, but it does not easily classify into any of these genera.
- (4) Phylogenetic analysis of the complete nucleotide sequence, and the amino acid sequences of the RdRP and CP, shows that SCSMaV is most consistently allied with

SCGRMV (a tentative member of the genus *Foveavirus*), yet differences including the size of the putative protein products and morphological and biological criteria suggest that SCSMaV may represent a new plant virus genus.

- (5) Detection of SCSMaV-RNA in sugarcane sap and nucleic acid extracts by hybridisation and RT-PCR assays is consistent with SCSMD affected sugarcane. The SCSMaV-CP specific RT-PCR assay used with an internal control gives diagnostic products by AGE and the ability to detect false negative reactions.
- (6) Preliminary studies for insect vectors of SCSMaV have shown that four hemipteran insect species collected from the roots of SCSMD affected sugarcane plants test positive for SCSMaV-RNA by SCSMaV-CP specific RT-PCR. Further study is required to determine if these insects can actively vector SCSMaV.

APPENDIX A: BIOCHEMICALS AND MISCELLANEOUS CHEMICALS

<u>Biochemical</u>	<u>Source</u>
α - ³² P dCTP	Geneworks (Australia)
1kb Plus DNA ladder	Gibco BRL Life Technologies (USA)
Ampicillin	Progen (Australia)
AMV Reverse Transcriptase	Geneworks (Australia)
<i>Bam</i> HI	Promega (USA)
dATP	Promega (USA)
dCTP	Promega (USA)
dGTP	Promega (USA)
dTTP	Promega (USA)
<i>Eco</i> RI	Promega (USA)
Herring Testes DNA (Type XIV)	Sigma (USA)
<i>Hind</i> III	Promega (USA)
Kanamycin sulphate	Progen (USA)
MLLV Reverse Transcriptase	Epicentre Technologies (USA)
pGEM [®] -T Vector system	Promega (USA)
Protein Marker - Precision Prestained	BioRad (USA)
Protein Standards	
pQE-31 Vector	Qiagen (Australia)
Proteinase K	Amresco (USA)
RNasin [®]	Promega (USA)
RQ1 RNase-free DNase	Promega (USA)
SP6 RNA Polymerase	Promega (USA)
Superscript II Reverse Transcriptase	Gibco BRL Life Technologies (USA)
T4 PNK	GeneWorks (Australia)
<i>Taq</i> DNA Polymerase in storage buffer A	Promega (USA)
TdT	Gibco BRL Life Technologies (USA)

<u>Chemical</u>	<u>Source</u>
2-mercaptoethanol	Riedel-deHaën (Germany)
8-hydroxyquinoline	M&B Pronalys (Australia)
Acetic acid	BDH (Australia)
Acetone	BDH (Australia)
Acrylamide	BioRad (USA)
Agar Bacteriological (Agar No.1)	Oxoid (England)
Agarose 1	Amresco (USA)
Ammonium acetate	UNIVAR (Australia)
APS	Amresco (USA)
Bis	BioRad (USA)
Boric acid	UNIVAR (Australia)
Bromphenol blue (sodium salt)	Probing & Structure (USA)
BSA (Fraction V)	Sigma (USA)
CaCl ₂ (dihydrate)	BDH (Australia)
CaCl ₂ dried	Unilab (Australia)
Cellulose powder MN 300 (Microgranular cellulose)	Macherey, Nagel & Co. (Germany)
Chloroform	BDH (Australia)
Cs ₂ SO ₄	BDH (England)
CTAB	Merck (Germany)
D(+) Mannose	Sigma (USA)
DEPC	Sigma (USA)
D-Glucose	BDH (Australia)
EDTA	BDH (Australia)
EtBr	Amresco (USA)
Ethanol	BDH (Australia)
Ficoll (Type 400)	Pharmacia (Sweden)
Formaldehyde	M&B Pronalys (Australia)
Formamide	Boehringer Mannheim (Germany)

Glycerol	BDH (Australia)
Imidazol	Sigma (USA)
IPTG	Sigma (USA)
Isoamyl alcohol	UNIVAR (Australia)
Isopropanol	UNIVAR (Australia)
KCl	BDH (Australia)
KH_2PO_4	BDH (Australia)
Maleic acid	BDH (Australia)
MgCl_2	UNIVAR (Australia)
$\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$	UNIVAR (Australia)
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$	UNIVAR (Australia)
MOPS	Sigma (USA)
NN-dimethylformamide	Sigma (USA)
$\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$	BDH (Australia)
Na_2SO_3	BDH (Australia)
NaCl	BDH (Australia)
NaN_3	Sigma (USA)
NaOH	BDH (Australia)
NBT	Sigma (USA)
N-lauroyl sarcosine	Sigma (USA)
PEG ₆₀₀₀	Sigma (USA)
Phenol	UNIVAR (Australia)
Potassium acetate	UNILAB (Australia)
PVP	Sigma (USA)
SDS	BDH (England)
Sephadex G-25	Pharmacia (Sweden)
Sodium acetate	UNIVAR (Australia)
Sucrose	UNIVAR (Australia)
TEMED	BioRad (USA)

Appendix A: Biochemicals and miscellaneous chemicals

Tri sodium citrate	BDH (Australia)
Tris	Amresco (USA)
Triton [®] X-100	Amresco (USA)
Tryptone	Oxoid (England)
Tween [®] 20	BioRad (USA)
Urea	BDH (Australia)
X-Gal	Amresco (USA)
Xylene cyanol	LABCHEM (Australia)
Yeast Extract	Oxoid (England)

APPENDIX B: OLIGONUCLEOTIDE PRIMERS

All oligonucleotide primers and their application are shown in the table below.

Sequencing Primers	
NicUp1	5' GGC AAG TGA GAG GAG ATG GGT AAC AGC ATT 3'
NicLo1	5' ATC TGC GGG TAC ATT GAT GCA TGT TTA ACC 3'
NicLo93	5' ACT ATG CCA AAA ACA ATC GCC ATC TGC GGG 3'
NicUp2	5' CCA TGA ATG AAG TTT GCA TTA CTG GAA GGG 3'
NicLo2	5' ACT TTC GTG TCA CAG AGC TGG TAA TTC TGG 3'
NicUp3	5' CAA AGC TAA AGT TCA GGT AAC TGA AAG CCC 3'
U3new	5' AAG ACA TGC AGC AAG AAC AAA AG 3'
NicLo3	5' CCA GCC TTT GCC TCG CAG AAC ATA CTC TCC 3'
NicUp6	5' CCT GCA AAG ATT GAT TTC ATA GGG ATG GGT 3'
NicLo6	5' CAG TCA AGT GCC ATC CGC AAA ATG TTG GGA 3'
NicUp7	5' TGC TTA AGG AGA TCA GGG AGC GGA GGA TAG 3'
NicLo7	5' CTT GAC TTG CAT CAA AGG CCT CAT AAT CCG 3'
NicLo8	5' CCA TTT TCC TGT ACT TCA GCC CCT ATG TTG 3'
NicUp8	5' GAG CAG TGG GAA ATC CTT TGA TCT CCT TGG 3'
NicLo9	5' CCT GGA TGG CAG AGT TAG TCG AAA ACG TCC 3'
NicUp9	5' AGG AGC TTT TCG ACT AAC TCT GCC ATC CAG 3'
NicLo10	5' AAT GCT GTT ACC CAT CTC CCT CTC TCA CTT 3'
NicLo11	5' GTC TTT GCG AGA TGA GGA AAC TTG ATA ACC 3'
NicUp11	5' GTC CTG AGC TAT ATA TTT CGC TGA CTC GCC 3'
NicUp12	5' CTT GAC AGA GGA TGG AAT TTA TAA GGA CCC 3'
NicUp13	5' CTG ATG TGA AGA AGT TCT TTA AGG ACG GGC 3'
NicLo13	5' CAA GTC TGG CTT TCC AAA GGT GCA GGC TGT 3'
NicUp14	5' GGT GTC AGC CTA ATA ATC CAA AAT CAG GGG 3'
NicLo14	5' GAG AGG AGG CTC ATT ACG TTC ATG CAT TGC 3'
NicUp15	5' TTT CTA TGG TGT TGA CTC TCC TGC TTC AGC 3'
NicUp16	5' AGT GCC TAC TTC CAA CCT GTC AAT TGT GCC 3'

NicLo16	5' AGA TCA TGA GAG CAG ATG GGC CAG ATA CAG 3'
NicUp17	5' AGG GCT TTA TGA TTA TCT TGA GAG CTT GAC 3'
NicLo17	5' AAT TAG ATT AGA TAG AGT CTT TGA GTC GAC 3'
NicLo18	5' GAC CGC AGG GTA AGA ACT TCT GAT ACT TTA 3'
NicUp19	5' GTT AGC GCA GGA GGA AGA GAA TGA TCA AGG 3'
NicLo19	5' CTT AGA AGG CGG GCA CAA TTG ACA GGT TGG 3'
NicUp20	5' CTG GTG GTT TTT CTA CTG CTG ATA GCC TCC 3'
NicLo21	5' GAA TTA GGC TTT TAA GTT CAG GGG TAT AGG 3'
Primers for amplification and cloning of CP	
CPUNew	5' NNN NGG ATC CGA GTG GGC CAC CCC CTT CA 3'
CPLNew	5' NNN NAA GCT TAC ATA CCC TCC CTT TCT CT 3'
Degenerate potexvirus coat protein primer – supplied by Dr. J. Thomas, DPI Qld.	
PXCP	5' TTC GCN GCN TTY GAY TTC TTN GA 3'
General Extension Primers	
PCC-RH	5' GCC GGA GCT CTG CAG AAT TCN NNN NN 3'
PCC	5' GCC GGA GCT CTG CAG AAT TC 3'
PCC-dT ₍₁₅₎	5' GCC GGA GCT CTG CAG AAT TCT TTT TTT TTT TTT TT 3'
Oligo-dT	5' TTT TTT TTT TTT TTT 3'
dT-RH	5' TTT TTT TTT TTT TTT TNN NNN N 3'
T7	5' TAA TAC GAC TCA CTA TAG GG 3'
SP6	5' TAT TTA GGT GAC ACT ATA G 3'
3' and 5' RACE Primers	
PCC-dT ₁₅	5' GGC GGA GCT CTG CAG AAT TCT TTT TTT TTT TTT 3'
5RACE	5' GGG CCC TTA AAT TCG CTC CAG ATT GG 3'
3RP2	5' AGC TTT TCT TTA TAT CTA CCC CCT GTT CAA 3'
Sequencing Primers for pQE-31 vector (supplied by D. Pountney, CSIRO)	
QE31F	5' CGG ATA ACA ATT TCA CAC A 3'
QE31R	5' GGT CAT TAC TGG AGT CTT G 3'

N = either A T G or C

Y = either C or T

1 UGAAAACAAA ACUACACAAU UGAGCUCAAA CUGCUACAAU AUCCACUCAA
51 UAUACCUUUAU UGAUUCUAUUAU UGAUGGCAAU UCUCACCUUC AAACCCCUA
101 AGCAGCUCUU AAUCGCUCAA UUGUCCUCUG AAGAGCAGGC CAGAAUUAGG
151 AGCUCUGUGA CACGAAAGUU AGCGCAGGAG GAAGAGAAUG AUCAAGGUUU
201 AUUCAGUUUAU CAUGUCUCUC CAUACUGUGC AGAGAAGUUG AUCAAGUCUG
251 GUGUUAACU CUGUGCCUAC UCCUCUGAUG UACAUAGUCA UCCUUGUUGU
301 AAAAUGCUAG AAAAUCAUUAU UUUUAUACAGA GUUCUCCUU CUUAUGUAGA
351 UUCAGAUUUU AUCCUUGUAG GUAAUUAAGCA UUCUAAGUUA GCCUUUCUUA
401 AACCAAGAGG CGCGAAAAAG AAUGUUGAUC UUCAACUGU UAAGGUGCUC
451 AAUCGUUUUG UCACAAGUCU UGACAGAUCC AGGUAUGGUA AAGAGUUUAA
501 UCACCUCAAU AUCAUGUCA CAUUCACUCC UGAACCUAAG GGUAGCAGUU
551 CCUAUACCCC UGAACUUAAA AGCCUAAUUC CAAAUUCUGGA GCGAUUUUAG
601 GGCCACAAAC UUUUUUUCCA UGAUGAGUUG CAUUCACUGGA AGGAUGAAUC
651 UAUUAUUAAA AUGCUAUCAU CUUUAGAGCC UUCAGUAGUG UACGCAACUG
701 UUGUAUGCCC ACCCGAAUUG CUUAUUGGUG CUAAGCAGAG UUAUAUAGAU
751 UGGCUAUACA CUUUUGAAA AUUUGGGGAU CAAAUUCUUAU UUGCUCACAA
801 CGGAAAGAUG AAUAGUGCCU ACUUCCAACC UGUCAAUUGU GCCCGCCUUC
851 UAAGAGCAAA AAGGAUAUGG CUGGAUAAGC AUAACUAUUA UGAUAUAGAU
901 CUGGUCUAUU CAAAUUUCUC GCAUCAUUG CUCUGCUUCU CAAAGGGCUC
951 AAAUCUUAAG AAUCAAGAUC UUAGGGCUUU CUCUGACUUU GACGCAUUUA
1001 AAUUAUUUA CAUAAGUGAG CUCAUGGUGA AUGCUGAGAG UUGCUUUCCA
1051 AUUAGUUGGA CUGUUGUAAA AGGGCUUUUAU GAUUAUCUUG AGAGCUUGAC
1101 AAAAGAGGAU UACAGAUCAU GUAAGGCUAA GCAAACUCAU UUAUGCCCA
1151 AUUGCUCAGG GAAACAACUU AAAUUUUGCA CUGACCUUAU CAAACUUUAU
1201 CUUGAAGCUC CACAUAUUA AACUAUGAUA GGUUCUAGCC CAGUUCACAA
1251 CUUUAAGGUC UUGUUUACAC GGAUGUUCUCC UGCUUGCAUU AAGAAAAAUU
1301 UCUCUUUUUA UAAAGGAAAA UUGUUGAAAA GUUUUUUGA ACAUCUAGCU
1351 CCUUGGCCUU UCCAUGUCA AACUUAUGAU CUUCAUUUAG GUUUUGUUUU
1401 CAAAUUCCAG GAGCUGAAAA AUGUGACGGA UCUAUUGGAG GAUGUUCAGU
1451 AAUUUGAAGU UGACUUUGUC AGAGAAUAUG ACCCAUCCA GAUGCGCAGC
1501 ACUCCUUAUG UUGGUCUUGC CCCUUUGCGU GAUGAAUCAU AUUAUAAGAG

1551 GAUGAAAUGU GGUGAGGCUU UCUUUGACAA CAUGAUUGGG CAAAUAAUGA
1601 GAUCCCAUUU AAGCGGGUUAU GGCACAAAUA CUAUCCUGC UGAACUUUUC
1651 CUAAGAAUCA AUUUGACCAU UGUUGGAAGG GUUGAGUUUC UUGUGAAUGG
1701 CCAUCAUGAU UUUGUAAUUU UUUUAUAUCAA CAAACAUUUU AAGGUAAAGA
1751 AGGACCCCAG GUACAAAAAC UUGUACUUUG CAAAUUUGAC CCUCGAAUUG
1801 AAUUGAGGU GGCUUUCGUG CAGACGUUUU AGAAGUAUC AGAAGUUCUU
1851 ACCCUGCGGU CUGGUGAGUC UUGCUGAAGU ACAAUUCUU UGGAGAAUA
1901 UUGCGAGAAG CAUCCAGAGU GGAUCACAUG AGGCUCAAGG ACAAUUGAA
1951 CAAGUUGAUU CCCAUGGGGA UGGAGUAGCA AUUGAGGAGC CAGAAAGGAA
2001 GGGGAUACU UUGGAAAAAU UAAUUGAAGC UACUUUAGUA ACCCACCUC
2051 AGCACUACCU CAAGAAUCCU GUUGUUGAUG AUUUUUAAG AUGCGCCUUU
2101 GAUGUCGACU CAAAGACUCU AUCUAAUCUA AUUGAUUGUC UUAACAGAU
2151 UUGUGAUACU AGCAUAAAUG UUGCCAUGG AUGUGACUGU UCAAUUCAG
2201 AUCUAAUUU ACUCAAGAUU AGUAAAUGG CCUCCUCUGU GGCAUCAU
2251 GUUAAUCUGG ACAUUGCAA AUCUGUGAAA AUCAAUUGGA CCGAAGCUGG
2301 UGAGAUUUU GUCGGUUUAG AGGAUCAGCC UCUUUUUGUC GUGGUUCUUG
2351 GGAAGGAUUA UAAUUCAAU ACAUUGCAA GGAUCAUGU UAUGCAAUUU
2401 CCUGAGCUUG AAGAAAUUC AGCAUUUUUA AUUGAAAGGA AUAACAUCUC
2451 UAAAAUUUCU GUAUCUGGCC CAUCUGCUCU CAUGAUCUU CCAAAGAU
2501 AGAAUUUGCU GGAUUGUCAU UAAUUGGCU GCGAUCAAU CCCCCUGCA
2551 AAGAUUGAUU UCAUAGAUGG GUACAUCAAA ACUACGGUGG AUCCUAGUCU
2601 UGAUUCUCGG UUAACAUGCA UCAUUGUACC CGCAAUUGGU GAUUGUUUUU
2651 GGCAUAGU GU AUCUCUUUAU CUUAGUGUUG AGGCCAAAUA GAUCAAAGAU
2701 GCACCUGUUC UAAGAAACAG GAAUCCCCAA AAUGCCAGAU UAGUUGAACC
2751 AAUGGGUGAU AAAGUUUGG CUGAAAAUGA GGCUAUACUA AACUGUGCAU
2801 GUUUUUACAA CUUGAAUUAU CGUAUAGUCU UUGAACAUUA AUUGCAUACC
2851 UUCCGUCCCC AUGAUUUGAA UGAAGAUUCA ACUGAAGUUU GGCUCAAAAA
2901 UUUGGAUCAU AUGCAUUUUU GUCCAGCAA GUUUUAUAAU UCUUGUGUCA
2951 UGGAGCCAAU UUCUAAAGCA GUUGACAUUC CGCCUUUGAA UGUACAAAGA
3001 AUGAUUUCU CACAGGAGAG UCUUCACUCC GAGUAUGAGA GAUUGAGCAG
3051 UGGGAAAUCC UUUGAUCUCC UUGGACUGGA GAGUUUGAUG ACUUUCUUA

U6 L1 L93
U8

3101	← L8 ACAAGGGGC UGAAGUCCAG GAAA AUGGUC	ACAAGUUUAU UUUGAACGAG
3151	AAUGGCAGUA GGGCUCGAAG CUUCAUUCUG	GAGAAUAAUC AUAUUGAAUU
3201	UGUUCGUGGA GAUAAAAAGG AUCUAAGAAA	GGAGGGAAUU GGGUUUGUUA
3251	AGCUAAUGAA UGCAAAUCAA AGUAAGUCUG	UGCUGUCUCU CGCCUGCUCA
3301	AAGGUUAAAU ACACCCCAUC UUAUGACAGA	GCCUUUAAAU UGUCAAAAGC
3351	UUUUGCUAU GGUUAUACAG GCAUAAUGCU	UAGCAGUGAA AAAUUUGGCG
3401	AGAUUGAAU CAACGGUAAG AAUGACGAGA	GGGAAUCAA UGUUAUGAUG
3451	GGAACAUUUG GGGCUGGAAA GACUAGAGUA	UUUGUGGAUU AUAUUCAGGC
3501	UUACUCUGGU AGAGGUUUUAU UUUAUGUUAG	CCCAAGAAAG AACCUUAAGG ←
3551	← L9 AGCUUUUCGA UAACUCUGCC AUCCAGG	CAG UUCAAGGCUU UAAAACCAAG
3601	U9 UCAAGUGACA AGAAAGGCAG UGUAAAGAGU	UUUCAUAGCU CUUUCACCUU
3651	UGAGACUGCU AUAUGAAAG CCGGCCAGAU	UGGUGAUGAU UCAAUUCUUA
3701	UUAUAGACGA GAUUCAGCUC UAUCCGCCUG	GCUACCUUGA UUUAUCCU
3751	CUAUUGAUAC CUUCAAGUUG CAUGGUAAU	UGUGCUGGUG ACCCUGUUA
3801	GUCUGACUUA GAUUCAGAU GUGACCGUU	GUAUUUUCU AAUGAUAAAC
3851	CAGAUUUGA GCGGUUGUUA GAGGGUCAA	CUUAUAAGUA UGCAAUCAA
3901	UCAAGAAGGU UUAUUAUCCUC CUUAUUCU	GGACGACUAC CUUGUGAUU
3951	GCCCGGUCU CAAGGGAUGG GCAUGGACAC	UUAUGGGAUU UAUGAGAGCA
4001	UUAGCAGUGU GCCUCAGGAU UUAUUAAGA	GAUCCUUCU UUUGGUGAGU
4051	AGUUUUGAUG AGAAGAAAAU UAUUAAGCC	AAUUUUAAG AGACUCAUA
4101	CAAGAUUUUG ACUUUUGGG AAUCCACAGG	AUCAACUUU CAAGAAUUG
4151	UCAUUCUUA AACUACUAGU UCAGAAAAG	← L10 CG CAAGUGAGAG GAGAUGGGUA
4201	← U1 ACAGCAUUA GCAGAGCAUC GACAAGCAUA	UCUUUUCUUA ACCUGACUGG
4251	CAUUGGAAUA CAAGAGCUUA UAUCGACUAC	UUAUUUUGGC AGAUUCUUGC
4301	AAAAAUUUU AAGUGCUGAA UGCAAGAGGG	AGGAUAUUAA GCCAUUUUA
4351	UACGGAGAGC CAAAUUUGU UGAUGACUUC	AGGAGCAUAG GAAGACGAAG
4401	UGCAGAGGAG AAAGAAGUCA AAGUUGAUGG	UGAUCCGUGG UUAAGUGCA
4451	UCUUGCAUUU CUUUGAAACU CCAAUCCAG	AUGAAGUUGA GUGUGUAAA
4501	GUUGAUUUGG UAGAGCAAAA AGCAAAAACU	CACUUACCGG UUUCGUCCU
4551	UGAGGUUUUA UGGGCCCAU UUAUUGAAAA	ACUUUUGCUU AAGGAGAUCA
4601	← U7 GGGAGCGGAG GAUAGAUGAU	CUUGUUUCUG ACCAGUUUGC AUGGGAUCAU

4651 AGGUCACAAG GACUAGCAAU UAAUAAUGCU GCCACAAGGU AUGAAUCAAU
4701 AUUCCCUAGG CAUCACCACA AUGAUAAAGU GACCUUUUUG AUGGCAGUUA
4751 AGAAAAGGUU GACCUUUGGA UCGCCCGCAA AGAAUGGGAG GAUGCUUAGA
4801 GAGGCUAUGC CCUAUGGUGA AAUAUUACUC AAGGAGUUUC UCAAGAGGGU
4851 CCCUAUUAAG AAUGAUAGAA ACGAUCAUUU AUUUUUAGAA UGUCUAAUGG
4901 AUUUUGAAGA AAAAAAACU UCCAAGGACU CAGCCACAAU CGCAAUCAC
4951 UGCAUACGCU CACAGAAUUC AUGGCCAAU GAUAUUGCUG AUUAUUCAU
5001 GAAGUCACAA CUCUGCACAA AAAU^{L3}GGAGAG UAUGUUCUGC GAGGCCAAAGG
5051 CUGGUCAGAC UUUAGCUUGU UUCAUCAUG UAGUUCUUGC ACGAUUUUCU
5101 CCAUGGGUCC GUUACAUAGA AAAGAAGCUG GUCAGUUCAC UGCCAGAUAA
5151 UUACUAUAUA CACUCUGGAA AGAACUUUGA UCAACUCAAU GCUUGGGUUA
5201 AGGCAAACAA UUUUGUGGGU GAAUGCACUG AAU^{L7}CGGAUUA UGAGGCCUUU
5251 GAUGCAAGUC AAGAUCACUA UAUACUGGCC UUUGAAUUAG CAUUA AUGAA
5301 GUACCUUGGU UUACCCAAUG AGUUGAUAGA GGACUACAAG UACAUUAAGA
5351 UUCAUUUGGG UUCAAAAUGG GGUGAUUUUG CCAUAAUGAG AUUCACCGGU
5401 GAAGCCAGCA CUUUCUUAUU UAACACAAUG GCAAAUAUGC UUUUCACUUU
5451 UCUGGUCUUA GAUAUCAAAG GCAUUGAGUC AAUUUGCUUU GCUGGCGAUG
5501 ACAUGUAUGC AAAUACAAGU CUUAAGACAU GCAGCAAGAA CAAAAGUUUG
5551 CUUAAAAAGU UUACACUCAA AGCUAAAGUU CAGGUAACUG AAAG^{U3new}GCCAAC
5601 AUUUUGCGGA UGGCA^{L6}CUUGA CAGAGGAUGG AAUUUAUAAG GACCC^{U3}AAUAC
5651 UUGUGUCUUU CAGAUUCAUG AUUGCACUUG AGAAAGGGAU CUUACACGAA
5701 UGCCUUGAGU CUUACUCGAU UGAGUGUUCA UAUGCAUAUA ACAUGGGAAG
5751 UGCAGUUUCA AGACAU AUGG ACCAGCUUAG CCUUGCUUUA CAUACUGCU
5801 GCGUGAGGUU UAUUCUUAUA AAUUCAAAAC UUCUUCGCU UGAUGUGAAG
5851 AAGUUCUUUA AGGACGGGU CCAGCUUACG CAGUAGCCAC GUUUAGGCUG
5901 CGACUUGUAU AUUAUGAUUA GGAAAUAGUC UGUCGUUACU UAGACGACU
5951 UAAGUUUGAG AGAACUAGCU UAGAGCUUUC UUCACCUAUA GUGGUUCACG
6001 CAGUCCAGG UUCAGGGAAG UCUACCCUCA UAAGAAAGUG CAUAACUGAA
6051 AAUCCUUCU ^{L13}UGACAGCCUG CACCUUUGGA AAGCCAGACU UGCCCAAUCU
6101 UACUGGUACC UAUAUAAAAG GGUAUAAAAG AGGUGAGAAG UAUGACAUAA
6151 UUGAUGAGUA CAUCGGCAAC AGAUUUGAGU GCGUUCUGU UGCACUCUUU

6201 UCUGAUCCUU UUCAGAAUUG UGAACAAUAC AUUAAUACAG CACAUAUAU
6251 UGGACGAACU UCUCACCGUU UCGGCGUUGU ACUGCUGAGC UAUUGCGCGG
6301 GCUUGG^{L11}UUAU CAAGUUUCCU CAUCUCGCAA AGACAAGGUC ACUCAGGGUU
6351 CAGGUUACUC UAUUGAAUCU ACAGGUGCUA UCAUAUGCUU UGAGCCUGAC
6401 ACUGAGGACU AUUUACUUGC ACAUAACCUC AAGCCUCUUC AUCCCUUGGA
6451 CAUCCGCGGA GAGACUUUUG ACAAGUUAC AUUCUAUCAU UCUAGUGAUC
6501 UCCGCGAUCU UGUUGGUCCU GAGCUAUAUA UUUCGCUGAC UCGCCAUAGA
6551 AAUGAAAUCA AAUAUUGCAA GCUUAAUGUC AGCUCUAGG CAGCCAUCUG
6601 AUAUUUCUGA GUCAAUAAA AUUUGCGCUG CUCUAGGGGG GACUGCAUUA
6651 CUUCUUUAUA UUU^{U11}GGUGUA GCCUAAUAAU CCAAAUACAG GGGACCUGGG
6701 CCAUUGGUUA CCUUUUGGGG GAGUUUACAA AGACGGUAAAC AAAUUCGUCA
6751 GGUAAUUAUC ACCUCUUCAC ACCGCUCCUC CUUUCUGGCU UUUAAUUGCA
6801 AUACCUUUGG CCUUCUUUGC UCUCACUCAC UUUAGCUACU UCAGUACAC
6851 ACCCAGAAGA AGAGUGACCA UAACACAUUA UCCAUGUGU GGCACUUUG
6901 UAUGCAGGCG CUAGAGAUAU AUAU^{U14}CUGGU GGUUUUUCUA CUGCUGAUAG
6951 CCUCCUUACU CUUCAACAA GAGAGUCAGC AGUGCAUAAU^{U20} UCACAUUGAC
7001 GGGACCAAAG CUUUUAUCAC CGGCUGCACU GAUCAAAGU GGUUGGAUAC
7051 UGCAAUCAA CACCUGAGAC CAGACAAUCA AAGAUCAGGU UGGUUUAGAA
7101 GAU^{CP-U}GAGUGGG CCACCCCUU CAACUGGUGC CAAUGCUICC AGUGGUGCUC
7151 AACCAACAAC UCAGCAGGAA ACGAGCACAG UUUUUAGAA AAGCCGCAGA
7201 GGAUCAAU CUAUUUUGU UUCUGCUGUU ACAAAGGAGC AGGUUAGGAG
7251 GGUGGUCGAG CUUGACAUUG AGACAGACAA AAUAGCCCGG GAUGAGGACC
7301 UUGAGAAAGU GAAAGCUGCU UGGCUUGCCA UUAUUGGCC AUCCGACAAG
7351 AUAUUUGAAG UUUCCCUUGA AAUAACACUU UCCUGUGCUC ACACUUCGAC
7401 UUCAAAUAG CAGAAAUUA AAGGCAAGUC CAAUUCUGC GACAUUUGC
7451 UUGCUGAAU AGCUGGCUGU AUCACUAAGG AGGAGGUAAC AAUACGUUCC
7501 UUUUGUGCUU AUUGGGCAA AUACGUCUG AAC^{L14}GCAAUGC AUGAACGUAA
7551 UGAGCCUCCU CUCAAUUGG UUGCUCUUG GUAUACAGAA CAAACCGUU
7601 UUGCAGCUU UGACUUUUUC^{U15} UAUGGUGUUG ACUCUCCUGC UUCAGCAGAA
7651 CCAGAAGGUG GGUUACUUGC UCCUACCACU GAGUCUAUGA GAGCUGCAA
7701 CAUGGCUAGG AAAGAGAUAG^{PXCP} CCAUCAAGA^{CP-L} AUGAGAGAAA GGGAGGGUAI

7751 GUCAUAUUUG AAUCUGUCAG CAGUCACAGG UGCCAGGAAU AGUAUAAGGC
7801 CAAGGUUAGA UUUCUCACCU GCAGAAUAA AUUAGCUUUU CUUUAUAUCU
7851 ACCCCCUGUU CAAGCAGGUC GGACCUUCGG CAAGUCCAAC AAAAAGAUUU 3RP2
7901 CUUUCUAUUU GUAAAUUACU GUCGAGGUAC AGAGUGUUGC UUUGCACGAC
7951 ACCUUUGGGU GAACCUCAGU CUUAGAACUA GUGCUUUGUC AAUCAUUUAA
8001 UGGCCAAGGC AAGGCAUUGU CUUGGGCACU AUUUGAUUGA CAUUGUGGUC
8051 UUUCUAAGAC CCUCGGCUUC AAUAAACAGC AUUGACAAGC AUGUACGUAU
8101 CUCAUGCACG CUGUAUAAAA AUAGCUUUGG UUUUACAUAU UUUCCC (A)_n

APPENDIX C: POLYACRYLAMIDE AND AGAROSE GELS, SOLUTIONS, BUFFERS AND MEDIA

1. Acrylamide solutions

40% solution: 39.5 g of acrylamide was combined with 0.5 g bisacrylamide in a total volume of 100 ml, heating to 37 °C to dissolve. The solution was filtered through a 0.45 µm syringe filter and stored in a light protected bottle at 4 °C.

30% solution: 29 g of acrylamide was combined with 1 g bisacrylamide in a total volume of 100 ml. The solution was filtered and stored at 4 °C.

2. Polyacrylamide gels

6% (87 acryl-:1 bis-acrylamide) PAG for NA analysis

40% (87:1) acrylamide mix	7.5 ml
10 x TBE	5 ml
SDDW	37.5 ml
10% APS	400 µl
TEMED	50 µl

TOTAL

50 ml

5% - 12% stacking-resolving SDS-PAG for protein analysis (Laemmli, 1970)

(Prepared as described in Sambrook *et al.*, 1989).

12% Resolving gel

30% (29:1) acrylamide mix	4 ml
1 M Tris-HCl pH 8.8	2.5 ml
10% SDS	100 µl
SDDW	3.3 ml
10% APS	100 µl
TEMED	5 µl

TOTAL

10 ml

5% Stacking gel

30% (29:1) acrylamide mix	330 µl
1 M Tris-HCl pH 6.8	250 µl
10% SDS	20 µl
SDDW	1.4 ml
10% APS	20 µl
TEMED	2 µl

TOTAL

2 ml

3. Agarose gel

1.0 – 1.5 g of agarose was dissolved by heating in a microwave oven in 100 ml of TAE buffer to make 1 – 1.5% (w/v) agarose. Ethidium bromide was added to 0.5 µg/ml.

4. Solutions

1 x Denhardts: 0.02% (w/v) Ficoll[®] 400, 0.02% (w/v) polyvinylpyrrolidone, 0.02% (w/v) gelatine

100 x Modified Denhardts: 2% (w/v) polyvinylpyrrolidone, 2% (w/v) Ficoll[®] 400, 2% (w/v) polyethylene glycol 8000

1 x PBS: 120 mM NaCl, 3 mM KCl, 2 mM KH₂PO₄, 3 mM Na₂HPO₄

20 x SSC: 3 M NaCl, 0.3 M Na₃-citrate, adjusted to pH 7.2 with NaOH

1 x SSPE: 0.18 M NaCl, 10 mM NaH₂PO₄, 1 mM EDTA pH 7.2

1 x TE: 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0

Tris-HCl saturated phenol, pH 8.0 : 500 g phenol was melted in a 65°C waterbath and had 8-hydroxyquinoline added to 0.1% (w/v) and an equal volume (w/v) of 0.5 M Tris-HCl, pH 8.0 added, stirring well to mix the phases for 30-60 min. The phases were allowed to separate, the aqueous phase removed and replaced with an equal volume (v/v) of 0.1 M Tris-HCl, pH 8.0 which was mixed well as above. The phases were allowed to separate, the aqueous phase removed and the pH of the phenol checked using pH indicator paper. If the pH was not 8.0 the equilibration with 0.1 M Tris-HCl, pH 8.0 was repeated. When the pH reached 8.0 the aqueous phase was replaced with 0.1 x volume of 0.1 M Tris-HCl containing 0.2% β -mercaptoethanol and stored in a light sensitive bottle at 4°C.

5. Staining solutions

Coomassie blue stain solution: 0.25 g Coomassie Brilliant Blue R250 dissolved in 45:45:10 (v/v) methanol: SDDW: glacial acetic acid, filtered through Whatman No. 1 filter to remove particulate matter.

Coomassie blue destaining solution I: 45:45:10 (v/v) methanol: SDDW: glacial acetic acid solution.

Coomassie blue destaining solution II: 5:7:88 (v/v/v) glacial acetic acid: methanol: SDDW

1% PTA (Phosphotungstic acid): 1 g $\text{H}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$ + x H_2O dissolved in 100 ml H_2O , pH to 6.8 with NaOH.

6. Transfer solutions

Southern Solution I – Denaturing Solution: 0.5 N NaOH, 1 M NaCl

Southern Solution II – Neutralising Solution: 0.5 M Tris-HCl pH 7.4, 1.5 M NaCl

Towbin Western Transfer solution: 25 mM Tris, 192 mM glycine, 20% methanol, pH 8.3 (no additional acid or base to adjust pH)

7. Hybridisation solutions

Magna[®] membrane Northern hybridisation solution: 5 x SSPE, 50% formamide, 5 x Denhardt's, 0.5% SDS, 100 µg/ml denatured salmon sperm DNA, 10% dextran sulphate

ZetaProbe[®] Southern Prehybridisation solution: 5 x SSC, 20 mM NaH₂PO₄ pH 7.0, 2.5% SDS, 5 x Modified Denhardt's, 100 µg/ml denatured salmon sperm DNA

ZetaProbe[®] Northern Hybridisation solution: 50% formamide, 120 mM Na₂HPO₄, 250 mM NaCl, 7% (w/v) SDS, 1 mM EDTA, 100 µg/ml denatured salmon sperm DNA

Preparation of denatured salmon/herring DNA: 1 gm Herring testes type XIV DNA dissolved in 100 mL 0.4 M NaOH by stirring for 16 hrs at 25°C, boiled for 45 minutes, chilled on ice and neutralised with acetic acid to pH 4.7. Particulate matter was removed by centrifugation at 5000 g for 10 minutes, and the DNA precipitated by addition of 2 volumes of 100% ethanol and incubation at -20°C for 16 hrs. DNA was pelleted by centrifugation at 8000 g for 30 minutes, washed with 70% ethanol, dried *in vacuo* and resuspended to a concentration of 10 mg/ml in TE and stored at -20°C.

8. Membrane washing solutions

Magna[®] Wash Buffer 1 – Low stringency: 5 x SSC, 1% SDS

Magna[®] Wash Buffer 2 – High stringency: 5 x SSC, 0.5% SDS

ZetaProbe[®] Wash Buffer 1: 3 x SSC, 2.5% SDS

ZetaProbe[®] Wash Buffer 2: 0.5 x SSC, 2.5% SDS

ZetaProbe[®] Wash Buffer 3: 0.2 x SSC, 1% SDS

9. Miscellaneous buffers

6 x AGE Loading Buffer I: 0.25% (w/v) bromophenol blue, 40% (w/v) sucrose in water (Sambrook *et al.*, 1989)

5-10 x AMV reaction buffer: 250 mM Tris-HCl (pH 8.3), 50 mM MgCl₂, 500 mM KCl, 20 mM DTT

1 x DyNAzyme DNA polymerase reaction buffer: 50 mM Tris-HCl (pH 9.0), 1.5 mM MgCl₂, 15 mM (NH₄)₂SO₄, 0.1% Triton[®]X-100

GTE Buffer: 50 mM Glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0

10 x Klenow (DNA polymerase I) buffer: 500 mM Tris-HCl (pH 7.2), 100 mM MgSO₄, 1 mM DTT

1 x Proteinase K buffer: 10 mM Tris-HCl (pH 7.8), 5 mM EDTA, 0.5% SDS

RQ1 RNase-free DNase buffer: 400 mM Tris-HCl (pH 8.0), 100 mM MgSO₄, 10 mM CaCl₂

RQ1 RNase-free DNase stop buffer: 20 mM EGTA (pH 8.0)

2 x SDS-PAGE Loading Buffer: 0.75 M Tris-HCl pH 8.8, 10% (v/v) glycerol, 2% (v/v) SDS, 2% (v/v) β-mercaptoethanol, 0.25% (w/v) bromophenol blue

10 x STE: 100 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM EDTA

SuperScript II 5 x First Strand Reaction Buffer: 250 mM Tris-HCl (pH 8.3), 375 mM KCl, 15 mM MgCl₂

10 x T4 DNA Ligase Reaction Buffer: 200 mM Tris-HCl (pH 7.6), 50 mM MgCl₂, 50 mM DTT, 500 µg/ml bovine serum albumin

1 x TAE: 40 mM Tris-HCl, 20 mM sodium acetate, 2 mM EDTA, adjusted to pH 7.8 with glacial acetic acid

1 x Taq DNA polymerase reaction buffer: 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.1% Triton[®] X-100

1 x TBE: 900 mM Tris-Borate, 2 mM EDTA, pH 8.0

1 x TE: 10 mM Tris-HCl pH 8.0, 1 mM EDTA

TNA Extraction Buffer G: 100 mM Glycine pH 9.5, 100 mM NaCl, 10 mM EDTA

1 x Tris-glycine electrophoresis buffer for SDS-PAGE: 25 mM Tris-HCl pH 8.3, 250 mM glycine pH 8.3, 0.1% (v/v) SDS

10. Virus preparation buffers

Borate Buffer: 0.5 M H₃BO₃, 5 mM EDTA, pH 7.8

Virus Extraction Buffer: 100 mM Phosphate buffer pH 7.0, 10 mM EDTA, 1% MTG

Virus Storage Buffer: 10 mM Tris-HCl pH 7.4, 0.02% (w/v) NaN₃

11. Buffers for recombinant protein preparation

TFB 1 Buffer: 100 mM RbCl₂, 50 mM MnCl₂, 30 mM potassium acetate, 10 mM CaCl₂, 15% glycerol, pH to 5.8 with KOH, filter sterilize (0.2 µm)

TFB 2 Buffer: 10 mM MOPS, 10 mM RbCl₂, 75 mM CaCl₂, 15% Glycerol, pH to 6.8 with KOH, filter sterilize (0.2 µm)

Native Lysis Buffer: 50 mM NaH₂PO₄, pH 8.0, 300 mM NaCl, 10 mM imidazole

Native Wash Buffer: 50 mM NaH₂PO₄, pH 8.0, 300 mM NaCl, 20 mM imidazole

Native Elution Buffer: 50 mM NaH₂PO₄, pH 8.0, 300 mM NaCl, 250 mM imidazole

Denaturing Lysis Buffer B: 8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris-HCl, pH 8.0

Denaturing Wash Buffer C: 8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris-HCl, pH 6.3

Denaturing Elution Buffer D: 8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris-HCl, pH 5.9

Denaturing Elution Buffer E: 8 M urea, 0.1 M NaH₂PO₄, 0.01 M Tris-HCl, pH 4.5

12. Buffers for DIBA and Western blot analysis

PBS: 8 g/l NaCl, 0.2 g/l KH₂PO₄, 1.15 g/l Na₂HPO₄, 0.2 g/l KCl, pH to 7.4

Blocking buffer A (50 ml): Add skim milk powder to PBS to a final concentration of 2.6 % (gives a final casein concentration of ~ 1%)

Blocking buffer B (20 ml): Crush 2 gm healthy leaf material in 2 ml PBS. Transfer liquid to centrifuge tube and spin at ca. 1000 g for 5 minutes. Add 2 ml of supernatant to 18 ml of Blocking buffer A, store at RT

Washing buffer AP 7.5: 0.1 M Tris-HCl pH 7.5, 0.1 M NaCl, 2 mM MgCl₂, 0.05% Triton X-100

Washing buffer AP 9.5: 0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 2 mM MgCl₂

DIBA stop buffer: 10 mM Tris-HCl pH 7.5, 5 mM EDTA

NBT (Nitroblue tetrazolium) stock: 75 mg/ml in 70% dimethylformamide

BCIP (Bromo-4-chloro-3-indolyl phosphate) stock: 50 mg/ml in dimethylformamide

13. Media

Luria Broth (LB): 1% (w/v) tryptone, 1% (w/v) NaCl, 0.5% (w/v) yeast extract, adjusted to pH 7.2 with 1M NaOH

Luria Broth Agar (LBA): LB with the addition of 1% (w/v) bacteriological agar No. 1.

Psi Broth: LB media with the addition of 4 mM MgSO₄, 10 mM KCl

APPENDIX D: HOMOLOGIES RESULTS

HOMOLOGIES within ORF1 (RdRP)

Number of sequences: 22
 First residue number: 1
 Last residue number: 3088
 Threshold of comparison: 1.00
 Symbol Comparison Table: genrundata:pepunity.cmp
 Denominator: "Number of residues compared"
 Gap handling, General: "Including gaps"
 Gap handling, Termini: "Including terminal gaps"
 Gap penalty: 0.00

UNITY comparison table used by HOMOLOGIES for the comparison of protein sequences.
 Created: 09-Dec-1991, by Jack A.M. Leunissen

Key for column and row indices:

1	orflpep.msf{ACLSV}	Length: 3088	Length without gaps: 1887
2	orflpep.msf{ASGV}	Length: 3088	Length without gaps: 2105
3	orflpep.msf{ASPV}	Length: 3088	Length without gaps: 2183
4	orflpep.msf{BBSv}	Length: 3088	Length without gaps: 1967
5	orflpep.msf{CCA}	Length: 3088	Length without gaps: 2342
6	orflpep.msf{CTLV}	Length: 3088	Length without gaps: 937
7	orflpep.msf{CVB}	Length: 3088	Length without gaps: 260
8	orflpep.msf{GLV}	Length: 3088	Length without gaps: 1924
9	orflpep.msf{RSPaV}	Length: 3088	Length without gaps: 2161
10	orflpep.msf{GarV-A}	Length: 3088	Length without gaps: 1625
11	orflpep.msf{GarV-B}	Length: 3088	Length without gaps: 1543
12	orflpep.msf{GarV-C}	Length: 3088	Length without gaps: 1550
13	orflpep.msf{GVA}	Length: 3088	Length without gaps: 1707
14	orflpep.msf{GVB}	Length: 3088	Length without gaps: 1707
15	orflpep.msf{PVM}	Length: 3088	Length without gaps: 1968
16	orflpep.msf{PVS}	Length: 3088	Length without gaps: 355
17	orflpep.msf{PVT}	Length: 3088	Length without gaps: 248
18	orflpep.msf{PVX}	Length: 3088	Length without gaps: 1456
19	orflpep.msf{SCGRMV}	Length: 3088	Length without gaps: 2023
20	orflpep.msf{SMYEaV}	Length: 3088	Length without gaps: 1323
21	orflpep.msf{SCSMaVORF1.pep}	Length: 3088	Length without gaps: 1937
22	orflpep.msf{ShvX}	Length: 3088	Length without gaps: 1718

Similarity Matrix Part: 1

	1	2	3	4	5	6	7	8	9	10	11	12
1	100.0%	16.0%	22.6%	23.3%	16.5%	7.7%	6.4%	24.4%	22.8%	14.6%	14.1%	14.1%
2		100.0%	13.6%	14.6%	24.8%	38.3%	5.2%	15.0%	13.8%	11.2%	10.9%	11.0%
3			100.0%	36.6%	14.6%	6.3%	7.5%	35.1%	42.6%	11.5%	11.5%	11.7%
4				100.0%	14.8%	6.5%	9.8%	41.7%	37.0%	13.0%	12.6%	12.9%
5					100.0%	13.2%	5.4%	15.3%	14.5%	11.2%	11.3%	11.4%
6						100.0%	11.5%	6.5%	6.2%	5.4%	5.4%	5.6%
7							100.0%	9.2%	7.9%	4.9%	5.1%	5.0%
8								100.0%	34.8%	12.8%	12.6%	12.6%
9									100.0%	12.9%	12.5%	12.6%
10										100.0%	63.6%	64.1%
11											100.0%	78.6%
12												100.0%
13												

Similarity Matrix Part: 2

	13	14	15	16	17	18	19	20	21	22
1	21.6%	21.1%	23.0%	9.3%	4.9%	14.6%	25.5%	13.6%	22.1%	13.9%
2	14.0%	14.3%	14.6%	7.0%	4.3%	11.4%	14.4%	10.9%	14.3%	10.5%
3	19.6%	19.5%	33.5%	10.9%	4.5%	12.3%	31.8%	11.8%	28.6%	11.4%
4	21.3%	22.3%	41.1%	14.4%	4.8%	12.8%	32.0%	12.2%	32.0%	12.9%
5	13.7%	13.7%	14.6%	7.5%	3.7%	10.5%	15.4%	10.7%	15.5%	10.7%
6	6.5%	6.9%	6.6%	15.6%	9.7%	5.1%	6.8%	5.3%	7.2%	5.0%
7	6.1%	6.0%	10.0%	55.2%	36.7%	4.9%	7.6%	5.0%	8.0%	4.5%
8	20.6%	21.2%	36.3%	13.3%	5.0%	13.4%	31.3%	12.3%	31.1%	11.8%
9	19.9%	20.6%	34.3%	10.7%	4.6%	12.6%	32.1%	11.4%	29.2%	12.2%
10	15.0%	15.4%	13.0%	6.4%	4.3%	29.8%	12.8%	30.3%	13.4%	68.8%
11	14.9%	15.5%	12.5%	6.6%	4.7%	32.3%	12.8%	31.7%	13.0%	57.7%
12	15.3%	16.1%	12.5%	6.6%	4.6%	32.4%	12.9%	31.6%	13.4%	59.4%
13	100.0%	47.9%	19.7%	8.3%	6.7%	14.9%	21.1%	14.6%	18.6%	14.3%
14		100.0%	19.8%	8.4%	6.4%	15.3%	21.7%	15.3%	19.1%	14.4%
15			100.0%	13.0%	5.0%	13.5%	30.8%	12.8%	30.7%	12.5%
16				100.0%	27.0%	6.6%	10.4%	6.6%	10.6%	6.0%
17					100.0%	4.5%	4.9%	4.4%	5.0%	4.4%
18						100.0%	12.8%	42.9%	13.7%	27.9%
19							100.0%	12.1%	30.2%	12.3%
20								100.0%	12.6%	28.5%
21									100.0%	12.7%
22										100.0%

HOMOLOGIES within ORF2 (TGBp1)

Number of sequences: 14
 First residue number: 1
 Last residue number: 442
 Threshold of comparison: 1.00
 Symbol Comparison Table: genrundata:pepunity.cmp
 Denominator: "Number of residues compared"
 Gap handling, General: "Including gaps"
 Gap handling, Termini: "Including terminal gaps"
 Gap penalty: 0.00

UNITY comparison table used by HOMOLOGIES for the comparison of
 protein sequences.

Created: 09-Dec-1991, by Jack A.M. Leunissen

Key for column and row indices:

1	orf2pep.msf{ASPV}	Length: 442	Length without gaps: 414
2	orf2pep.msf{CVB}	Length: 442	Length without gaps: 315
3	orf2pep.msf{CsCMV}	Length: 442	Length without gaps: 97
4	orf2pep.msf{RSPaV}	Length: 442	Length without gaps: 221
5	orf2pep.msf{LLV}	Length: 442	Length without gaps: 228
6	orf2pep.msf{LSV}	Length: 442	Length without gaps: 228
7	orf2pep.msf{LVX}	Length: 442	Length without gaps: 201
8	orf2pep.msf{PVM}	Length: 442	Length without gaps: 63
9	orf2pep.msf{PVS}	Length: 442	Length without gaps: 227
10	orf2pep.msf{PapMV}	Length: 442	Length without gaps: 215
11	orf2pep.msf{PasiamV}	Length: 442	Length without gaps: 207
12	orf2pep.msf{SCGRMV}	Length: 442	Length without gaps: 268
13	orf2pep.msf{ScSmaVORF2.pap}	Length: 442	Length without gaps: 206
14	orf2pep.msf{WCLMV}	Length: 442	Length without gaps: 208

Similarity Matrix Part: 1

	1	2	3	4	5	6	7	8	9	10	11	12	
1	100.0%	11.7%	2.9%	5.6%	5.5%	5.8%	14.0%	1.0%	5.5%	13.1%	13.3%	13.9%	**
2		100.0%	1.9%	6.1%	6.4%	5.8%	14.0%	1.0%	7.3%	15.5%	13.2%	14.3%	
3			100.0%	5.8%	8.7%	7.8%	5.9%	8.9%	7.8%	7.9%	7.2%	3.0%	
4				100.0%	33.9%	33.9%	7.9%	5.4%	35.2%	6.2%	5.6%	6.1%	
5					100.0%	76.3%	11.5%	4.8%	54.6%	9.4%	8.5%	7.4%	
6						100.0%	10.3%	4.8%	54.6%	7.0%	6.8%	5.6%	
7							100.0%	3.0%	10.3%	29.7%	34.6%	19.7%	
8								100.0%	5.3%	3.3%	1.9%	1.9%	
9									100.0%	7.9%	7.3%	6.7%	
10										100.0%	38.4%	16.2%	
11											100.0%	17.6%	
12												100.0%	
13													
14													

Similarity Matrix Part: 2

	13	14	
1	5.5%	12.7%	**
2	5.3%	14.2%	
3	10.0%	5.6%	
4	29.7%	10.2%	
5	24.9%	10.4%	
6	26.2%	10.0%	
7	3.3%	24.7%	
8	5.3%	1.9%	
9	25.6%	10.0%	
10	7.4%	23.2%	
11	7.8%	21.3%	
12	5.9%	17.4%	
13	100.0%	7.9%	
14		100.0%	

HOMOLOGIES within ORF3 (TGBp2)

Number of sequences: 14
 First residue number: 1
 Last residue number: 222
 Threshold of comparison: 1.00
 Symbol Comparison Table: genrundata:pepunity.cmp
 Denominator: "Number of residues compared"
 Gap handling, General: "Including gaps"
 Gap handling, Terminal: "Including terminal gaps"
 Gap penalty: 0.00

UNITY comparison table used by HOMOLOGIES for the comparison of
 protein sequences.

Created: 09-Dec-1991, by Jack A.M. Leunissen

Key for column and row indices:

1	orf3pep.msf{BaMV}	Length: 222	Length without gaps: 119
2	orf3pep.msf{CsCMV}	Length: 222	Length without gaps: 97
3	orf3pep.msf{CymMV}	Length: 222	Length without gaps: 91
4	orf3pep.msf{FxmMV}	Length: 222	Length without gaps: 105
5	orf3pep.msf{GLV}	Length: 222	Length without gaps: 64
6	orf3pep.msf{GVA}	Length: 222	Length without gaps: 103
7	orf3pep.msf{GVB}	Length: 222	Length without gaps: 108
8	orf3pep.msf{LVX}	Length: 222	Length without gaps: 201
9	orf3pep.msf{PVM}	Length: 222	Length without gaps: 63
10	orf3pep.msf{PapMV}	Length: 222	Length without gaps: 215
11	orf3pep.msf{PasiamV}	Length: 222	Length without gaps: 207
12	orf3pep.msf{SCSMaVORF3.pep}	Length: 222	Length without gaps: 120
13	orf3pep.msf{ShVX}	Length: 222	Length without gaps: 103
14	orf3pep.msf{WCLMV}	Length: 222	Length without gaps: 208

Similarity Matrix Part: 1

	1	2	3	4	5	6	7	8	9	10	11	12	
1	100.0%	12.1%	11.8%	37.7%	6.7%	37.0%	28.1%	9.8%	5.0%	6.0%	10.1%	27.0%	
2		100.0%	11.8%	8.0%	10.1%	8.9%	11.1%	3.0%	5.0%	4.2%	6.8%	5.8%	
3			100.0%	17.1%	6.2%	11.3%	11.7%	7.0%	4.0%	5.6%	5.3%	10.6%	
4				100.0%	9.2%	29.7%	25.2%	7.5%	4.4%	7.4%	10.1%	20.6%	
5					100.0%	8.7%	13.8%	4.0%	26.4%	2.3%	3.4%	7.4%	
6						100.0%	50.0%	7.0%	3.8%	7.4%	10.1%	27.5%	
7							100.0%	4.9%	8.3%	6.0%	9.2%	25.8%	
8								100.0%	1.5%	29.7%	33.2%	8.8%	
9									100.0%	0.9%	1.9%	3.3%	
10										100.0%	38.6%	7.4%	
11											100.0%	9.7%	
12												100.0%	
13													100.0%
14													

Similarity Matrix Part: 2

	13	14	
1	31.1%	7.5%	
2	8.0%	3.4%	
3	11.3%	6.2%	
4	25.4%	8.2%	
5	14.4%	4.3%	
6	62.1%	8.7%	
7	50.0%	6.6%	
8	6.0%	31.1%	
9	9.5%	2.4%	
10	6.0%	29.8%	
11	9.2%	27.1%	
12	25.0%	9.0%	
13	100.0%	7.2%	
14		100.0%	

HOMOLOGIES within ORF4 (TGBp3)

Number of sequences: 10
 First residue number: 1
 Last residue number: 422
 Threshold of comparison: 1.00
 Symbol Comparison Table: genrundata:pepunity.cmp
 Denominator: "Number of residues compared"
 Gap handling, General: "Including gaps"
 Gap handling, Termini: "Including terminal gaps"
 Gap penalty: 0.00

UNITY comparison table used by HOMOLOGIES for the comparison of
 protein sequences.

Created: 09-Dec-1991, by Jack A.M. Leunissen

Key for column and row indices:

1	orf4pep.msf{ASPV}	Length: 422	Length without gaps: 414
2	orf4pep.msf{CVB}	Length: 422	Length without gaps: 315
3	orf4pep.msf{CsCMV}	Length: 422	Length without gaps: 97
4	orf4pep.msf{CymMV}	Length: 422	Length without gaps: 91
5	orf4pep.msf{GLV}	Length: 422	Length without gaps: 64
6	orf4pep.msf{HelVS}	Length: 422	Length without gaps: 299
7	orf4pep.msf{PMV}	Length: 422	Length without gaps: 71
8	orf4pep.msf{PVM}	Length: 422	Length without gaps: 63
9	orf4pep.msf{PVX}	Length: 422	Length without gaps: 248
10	orf4pep.msf{SCSMaVORF4.pep}	Length: 422	Length without gaps: 67

Similarity Matrix Part: 1

	1	2	3	4	5	6	7	8	9	10
1	100.0%	19.3%	4.3%	4.3%	1.9%	19.5%	2.7%	2.2%	19.3%	2.7%
2		100.0%	4.4%	6.3%	2.5%	56.2%	2.5%	2.5%	23.7%	4.8%
3			100.0%	13.8%	12.0%	4.7%	5.6%	9.7%	6.0%	13.2%
4				100.0%	11.5%	7.0%	10.0%	11.5%	4.4%	14.7%
5					100.0%	2.7%	25.4%	27.9%	2.0%	15.5%
6						100.0%	2.0%	3.3%	22.3%	4.7%
7							100.0%	17.6%	2.8%	21.3%
8								100.0%	3.2%	22.5%
9									100.0%	4.4%
10										100.0%

HOMOLOGIES within ORF5 (CP)

Number of sequences: 21
 First residue number: 1
 Last residue number: 442
 Threshold of comparison: 1.00
 Symbol Comparison Table: genrundata:pepunity.cmp
 Denominator: "Number of residues compared"
 Gap handling, General: "Including gaps"
 Gap handling, Termini: "Including terminal gaps"
 Gap penalty: 0.00

UNITY comparison table used by HOMOLOGIES for the comparison of protein sequences.

Created: 09-Dec-1991, by Jack A.M. Leunissen

Key for column and row indices:

1	orf5pep.msf{ASPV}	Length: 442	Length without gaps: 414
2	orf5pep.msf{AltV}	Length: 442	Length without gaps: 207
3	orf5pep.msf{ApLV}	Length: 442	Length without gaps: 421
4	orf5pep.msf{CPMMV}	Length: 442	Length without gaps: 288
5	orf5pep.msf{CVB}	Length: 442	Length without gaps: 315
6	orf5pep.msf{CaCMV}	Length: 442	Length without gaps: 229
7	orf5pep.msf{CaLV}	Length: 442	Length without gaps: 306
8	orf5pep.msf{CLYMV}	Length: 442	Length without gaps: 257
9	orf5pep.msf{CyMV}	Length: 442	Length without gaps: 223
10	orf5pep.msf{HelVS}	Length: 442	Length without gaps: 299
11	orf5pep.msf{LVX}	Length: 442	Length without gaps: 201
12	orf5pep.msf{PAMV}	Length: 442	Length without gaps: 249
13	orf5pep.msf{PVX}	Length: 442	Length without gaps: 248
14	orf5pep.msf{PVYV}	Length: 442	Length without gaps: 414
15	orf5pep.msf{PapMV}	Length: 442	Length without gaps: 215
16	orf5pep.msf{PasiamV}	Length: 442	Length without gaps: 207
17	orf5pep.msf{RSPaV}	Length: 442	Length without gaps: 259
18	orf5pep.msf{SCGRMV}	Length: 442	Length without gaps: 268
19	orf5pep.msf{SCSMAVORF5.pep}	Length: 442	Length without gaps: 218
20	orf5pep.msf{ShLV}	Length: 442	Length without gaps: 294
21	orf5pep.msf{WCLMV}	Length: 442	Length without gaps: 208

Similarity Matrix Part: 1

	1	2	3	4	5	6	7	8	9	10	11	12	..
1	100.0%	13.6%	55.7%	16.9%	15.4%	13.8%	16.1%	15.0%	19.8%	13.7%	15.4%	20.0%	
2		100.0%	12.4%	20.0%	20.4%	37.8%	20.1%	33.3%	27.4%	19.5%	34.6%	26.6%	
3			100.0%	18.0%	16.7%	15.0%	16.1%	15.7%	19.6%	15.3%	16.1%	18.7%	
4				100.0%	39.6%	21.0%	38.8%	22.0%	21.2%	43.0%	22.1%	17.8%	
5					100.0%	18.6%	32.9%	18.6%	17.2%	56.2%	17.4%	14.8%	
6						100.0%	21.4%	38.1%	31.9%	19.2%	31.7%	23.7%	
7							100.0%	22.3%	20.3%	35.7%	20.8%	16.7%	
8								100.0%	27.0%	20.2%	27.9%	21.8%	
9									100.0%	17.2%	34.4%	50.4%	
10										100.0%	19.3%	15.1%	
11											100.0%	28.4%	
12												100.0%	
13													
14													
15													
16													
17													
18													
19													
20													
21													

Similarity Matrix Part: 2

	13	14	15	16	17	18	19	20	21	..
1	17.9%	99.5%	14.6%	14.1%	19.1%	16.0%	14.5%	16.7%	17.2%	
2	31.7%	13.6%	75.3%	42.5%	20.8%	21.6%	18.8%	22.8%	32.1%	
3	16.9%	55.5%	14.3%	14.3%	18.3%	15.7%	13.7%	17.3%	17.6%	
4	25.3%	16.9%	21.7%	19.9%	19.8%	22.4%	19.7%	32.8%	19.2%	
5	23.8%	15.4%	18.9%	17.9%	18.2%	19.9%	17.6%	33.3%	16.0%	
6	33.3%	13.8%	39.1%	51.1%	23.4%	21.9%	19.4%	21.4%	29.9%	
7	25.4%	16.1%	21.4%	21.7%	17.8%	21.1%	17.5%	44.4%	19.0%	
8	25.7%	15.0%	32.9%	32.7%	25.3%	24.1%	22.7%	23.8%	27.4%	
9	30.4%	19.8%	28.3%	31.3%	25.0%	26.7%	21.5%	24.1%	39.1%	
10	22.7%	13.7%	19.5%	18.9%	17.9%	20.0%	17.2%	34.2%	15.8%	
11	30.2%	15.4%	34.6%	35.0%	23.3%	23.3%	19.0%	23.6%	35.7%	
12	24.5%	19.8%	25.2%	22.4%	22.1%	24.9%	19.1%	20.2%	33.5%	
13	100.0%	17.9%	31.7%	29.8%	25.4%	23.5%	22.9%	29.5%	33.7%	
14		100.0%	14.6%	14.1%	19.1%	15.7%	14.2%	16.7%	17.0%	
15			100.0%	39.6%	20.4%	22.6%	21.4%	22.8%	30.3%	
16				100.0%	21.5%	17.2%	18.3%	21.8%	29.3%	
17					100.0%	24.0%	22.6%	22.3%	24.6%	
18						100.0%	19.9%	24.4%	25.5%	
19							100.0%	18.3%	19.4%	
20								100.0%	21.4%	
21									100.0%	

HOMOLOGIES within nucleotide sequence

Number of sequences: 17
 First residue number: 1
 Last residue number: 9796
 Threshold of comparison: 1.00
 Symbol Comparison Table: genrundata:pepunity.cmp
 Denominator: "Number of residues compared"
 Gap handling, General: "Including gaps"
 Gap handling, Termini: "Including terminal gaps"
 Gap penalty: 0.00

UNITY comparison table used by HOMOLOGIES for the comparison of
 protein sequences.

Created: 09-Dec-1991, by Jack A.M. Leunissen

Key for column and row indices:

1	seqs.msf{ACLSV.dna}	Length: 9796	Length without gaps: 7555
2	seqs.msf{ASGV.dna}	Length: 9796	Length without gaps: 6496
3	seqs.msf{ASPV.dna}	Length: 9796	Length without gaps: 9306
4	seqs.msf{BBSv.dna}	Length: 9796	Length without gaps: 8512
5	seqs.msf{CCA.dna}	Length: 9796	Length without gaps: 7383
6	seqs.msf{CTLV.dna}	Length: 9796	Length without gaps: 6496
7	seqs.msf{CyMV.dna}	Length: 9796	Length without gaps: 6227
8	seqs.msf{GLV.dna}	Length: 9796	Length without gaps: 8535
9	seqs.msf{GrVA.dna}	Length: 9796	Length without gaps: 7349
10	seqs.msf{GrVB.dna}	Length: 9796	Length without gaps: 7599
11	seqs.msf{PAMV.dna}	Length: 9796	Length without gaps: 7059
12	seqs.msf{PVM.dna}	Length: 9796	Length without gaps: 8535
13	seqs.msf{RSPaV.dna}	Length: 9796	Length without gaps: 8744
14	seqs.msf{SCGRMV.dna}	Length: 9796	Length without gaps: 8372
15	seqs.msf{SMYEaV.dna}	Length: 9796	Length without gaps: 5966
16	seqs.msf{SCSMaV.seq}	Length: 9796	Length without gaps: 8146
17	seqs.msf{WCLMV.dna}	Length: 9796	Length without gaps: 5846

Similarity Matrix Part: 1

	1	2	3	4	5	6	7	8	9	10	11	12
1	100.0%	37.9%	31.6%	29.8%	28.6%	37.9%	22.1%	29.6%	27.4%	26.2%	25.0%	29.7%
2		100.0%	27.5%	26.3%	31.1%	82.9%	23.9%	25.6%	29.7%	28.3%	26.5%	25.9%
3			100.0%	34.6%	27.1%	27.6%	19.0%	33.2%	24.1%	25.1%	22.1%	33.8%
4				100.0%	28.4%	25.8%	21.5%	43.5%	26.8%	26.8%	23.0%	46.0%
5					100.0%	30.8%	24.9%	28.0%	30.1%	30.8%	27.8%	28.1%
6						100.0%	23.6%	25.5%	30.1%	28.3%	26.2%	26.0%
7							100.0%	20.8%	24.3%	24.3%	32.8%	20.9%
8								100.0%	26.1%	26.8%	22.9%	43.1%
9									100.0%	40.2%	28.1%	27.1%
10										100.0%	27.6%	27.2%
11											100.0%	23.8%
12												100.0%
13												
14												
15												
16												
17												

Similarity Matrix Part: 2

	13	14	15	16	17
1	30.6%	30.9%	22.8%	30.5%	22.8%
2	27.0%	26.9%	24.7%	27.2%	25.2%
3	41.0%	32.2%	20.3%	31.6%	19.2%
4	35.2%	37.5%	22.6%	37.9%	21.2%
5	26.8%	28.9%	24.8%	20.7%	24.0%
6	27.5%	26.7%	24.7%	27.0%	24.8%
7	20.8%	23.2%	35.5%	22.5%	34.9%
8	33.4%	36.3%	22.1%	37.5%	21.7%
9	26.6%	27.9%	25.6%	27.4%	23.0%
10	25.4%	27.2%	23.9%	26.9%	23.2%
11	22.9%	25.1%	30.5%	24.3%	30.6%
12	34.9%	36.3%	22.2%	37.5%	20.8%
13	100.0%	35.2%	22.3%	33.7%	21.5%
14		100.0%	23.0%	38.3%	22.4%
15			100.0%	23.3%	35.7%
16				100.0%	23.0%
17					100.0%

LITERATURE CITED

- AbouHaidar, M. G. (1983): The structure of the 5' and 3' ends of clover yellow mosaic virus RNA. *Canadian Journal of Microbiology* **29**, 151-156.
- AbouHaidar, M. G. (1988): Nucleotide sequence of the capsid protein gene and 3' non-coding region of papaya mosaic virus RNA. *Journal of General Virology* **69**, 219-226.
- AbouHaidar, M. G., and Lai, R. (1989): Nucleotide sequence of the 3'-terminal region of clover yellow mosaic virus RNA. *Journal of General Virology* **70**, 1871-1875.
- Ali, A., and Randles, J.W. (1997): Early season survey of pea viruses in Pakistan and the detection of two new pathotypes of Pea seedborne mosaic potyvirus. *Plant Disease* **81**, 343-347.
- Ali, A., Randles, J. W., and Hodgson, R. A. J. (1998): Sensitive detection of pea seed borne mosaic potyvirus by dot blot and tissue print hybridisation assay. *Australian Journal of Agricultural Research* **49**, 191-197.
- Anonymous (1965): Annual Report, pp70, Bureau of Sugar Experiment Stations, Indooroopilly, Australia.
- Anonymous (1985): Annual Report, pp. 16, Bureau of Sugar Experiment Stations, Indooroopilly, Australia.
- Anonymous (1999a): *Custom oligonucleotide service handbook*. GeneWorks. Adelaide.
- Anonymous (1999b): *The QIAexpressionist: A handbook for high-level expression and purification of 6xHis-tagged proteins*. QIAGEN. Melbourne.
- Badge, J., Brunt, A., Carson, R., Dagless, E., Karamagioli, M., Phillips, S., Seal, S., Turner, R., and Foster, G. D. (1996): A carlavirus-specific PCR primer and partial nucleotide sequence provides further evidence for the recognition of cowpea mild mottle virus as a whitefly-transmitted carlavirus. *European Journal of Plant Pathology* **102**, 305-310.

- Bancroft, J. B., Rouleau, M., Johnston, R., Prims, L., and Mackie, G. A. (1991): The entire nucleotide sequence of foxtail mosaic virus RNA. *Journal of General Virology* **72**, 2173-2181.
- Beck, D. L., Forster, R. L. S., Bevan, M. W., Boxen, K. A., and Lowe, S. C. (1990): Infectious transcripts and nucleotide sequence of cloned cDNA of the potexvirus White clover mosaic virus. *Virology* **177**, 152-158.
- Birch, R. J. (1996): New gene technologies and their potential value for sugarcane. *Outlook on Agriculture* **25**, 219-226.
- Braithwaite, K. S., Egeskov, N. M., and Smith, G. R. (1995): Detection of sugarcane bacilliform virus using the polymerase chain reaction. *Plant Disease* **79**, 792-796.
- Bruenn, J. A. (1991): Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Research* **19**, 217-226.
- Brunt, A. A. (1995): Genus *Carlavirus*, pp. 475-478. In F. A. Murphy, C. M. Fauquet, D. H. L. Bishop, S. A. Ghabrial, A. W. Jarvis, G. P. Martelli, M. A. Mayo, and M. D. Summers (Eds): *Virus Taxonomy: Sixth Report of the International Committee on Taxonomy of Viruses*, Springer-Verlag, New York.
- Brunt, A. A., Crabtree, K., Dallwitz, M. J., Gibbs, A. J., Watson, L., and Zurcher, E. J. (1996 onwards): Plant Viruses Online: Descriptions and Lists from the VIDE Database.
- Caciagli, P., and Bosco, D. (1996): Quantitative determination of tomato yellow leaf curl geminivirus DNA by chemiluminescent assay using digoxigenin-labelled probes. *Journal of Virological Methods* **57**, 19-29.
- Cahill, D. M. (1999): Detection, identification and disease diagnosis of soil-borne pathogens. *Australasian Plant Pathology* **28**, 34-44.
- Calvert, L. A., Cuervo, M. I., Ospina, M. D., Fauquet, C. M., and Ramirez, B. C. (1996): Characterization of cassava common mosaic virus and a defective RNA species. *Journal of General Virology* **77**, 525-530.

- Campbell, R. N. (1996): Fungal transmission of plant viruses. *Annual Review of Phytopathology* **34**, 87-108.
- Cavileer, T. D., Halpern, B. T., Lawrence, D. M., Podleckis, E. V., Martin, R. R., and Hillman, B. I. (1994): Nucleotide sequence of the carlavirus associated with blueberry scorch and similar diseases. *Journal of General Virology* **75**, 711-720.
- Cerovska, N., and Filigarova, M. (1995): Specific detection of the Andean strain of potato virus S by monoclonal antibodies. *Annals of Applied Biology* **127**, 87-93.
- Choi, Y. G. (1997): The etiology of sugarcane striate mosaic disease, pp. 106: *Department of Crop Protection*, University of Adelaide, Waite Campus, Adelaide.
- Choi, Y. G., and Randles, J. W. (1997): Microgranular cellulose improves dsRNA recovery from plant nucleic acid extracts. *BioTechniques* **23**, 610-611.
- Choi, Y. G., Croft, B. J., and Randles, J. W. (1999): Identification of sugarcane striate mosaic-associated virus by partial characterization of its double-stranded RNA. *Phytopathology* **89**, 877-883.
- Chomczynski, P., and Sacchi, N. (1987): Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**, 156-159.
- Clark, M. F., and Adams, A. N. (1977): Characteristics of the microplate method of enzyme-linked immunosorbent assay for the detection of plant viruses. *Journal of General Virology* **34**, 475-483.
- Comstock, J. C., and Lockhart, B. E. L. (1996): Effect of Sugarcane bacilliform virus on biomass production of three sugarcane cultivars. *Sugar Cane* **4**, 12-15.
- Comstock, J. C., Ireby, M. S., Lockhart, B. E. L., and Wang, Z. K. (1998): Incidence of yellow leaf syndrome in CP cultivars based on polymerase chain reaction and serological techniques. *Sugar Cane* **4**, 21-28.

- Croft, B. J., and Randles, J. W. (2000): Striate Mosaic, pp. 265-267. In P. Rott, R. A. Bailey, J. C. Comstock, B. J. Croft, and A. S. Saumtally (Eds): *A guide to sugarcane diseases*, Cirad, Montpellier.
- Croft, B. J., and Smith, G. R. (1996): Major diseases affecting sugarcane production in Australia and recent experiences with sugarcane diseases in quarantine, pp. 55-58. In B. J. Croft, C. M. Piggin, E. S. Wallis, and D. M. Hogarth (Eds): *Sugarcane Germplasm Conservation and Exchange*, ACIAR, Brisbane.
- Croft, B. J., Piggin, C. M., E.S., W., and Hogarth, D. M. (1996): *Sugarcane Germplasm Conservation and Exchange*. ACIAR, Brisbane.
- Cronje, C. P. R., Bailey, R. A., and McFarlane, K. (1998a): Update on the occurrence of yellow leaf syndrome in the South African Sugar Industry and evidence for the causal organism. *Proceedings of the South African Sugar Technology Association* **72**, 81-84.
- Cronje, C. P. R., Tymon, A. M., Jones, P., and Bailey, R. A. (1998b): Association of a phytoplasma with a yellow leaf syndrome of sugarcane in Africa. *Annals of Applied Biology* **133**, 177-186.
- Dallwitz, M. J. (1980): A general system for coding taxonomic descriptions. *Taxon* **29**, 41-46.
- Dallwitz, M. J., Paine, T. A., and Zurcher, E. J. (1993): *User's Guide to the DELTA System: a general system for processing taxonomic descriptions*. CSIRO Division of Entomology. Canberra.
- Devereux, J., Haeberli, P., and Smithies, O. (1984): A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387-395.
- Dolja, V. V., Boyko, V. P., Agranovski, A. A., and Koonin, E. V. (1991): Phylogeny of capsid proteins of rod-shaped and filamentous plant RNA viruses: Two families with distinct patterns of sequence and probable structure conservation. *Virology* **184**, 79-86.
- Erhardt, M., Stussi-Garaud, C., Guilley, H., Richards, K. E., Jonard, G., and Bouzoubaa, S. (1999): The first triple gene block protein of peanut clump virus localizes to the plasmodesmata during virus infection. *Virology* **264**, 220-229.

- Eun, A. J.-C., and Wong, S.-M. (2000): Molecular beacons: A new approach to plant virus detection. *Phytopathology* **90**, 269-275.
- Fazeli, C. F., and Rezaian, M. A. (2000): Nucleotide sequence and organisation of ten open reading frames in the genome of Grapevine leafroll-associated virus 1 and identification of three subgenomic RNAs. *Journal of General Virology* **81**, 605-615.
- Fegan, M., Croft, B. J., Teakle, D. S., Hayward, A. C., and Smith, G. R. (1998): Sensitive and specific detection of *Clavibacter xyli* subsp. *xyli*, causal agent of ratoon stunting disease of sugarcane, with a polymerase chain reaction-based assay. *Plant Pathology* **47**, 495-504.
- Felsenstein, J. (1989): Phylogeny inference package (Version 3.2). *Cladistics* **5**, 164-166.
- Fenby, N. S., Scott, N. W., Slater, A., and Elliott, M. C. (1995): PCR and non-isotopic labelling techniques for plant virus detection. *Cellular and Molecular Biology* **41**, 639-652.
- Fillhart, R. C., Bachand, G. D., and Castello, J. D. (1998): Detection of infectious tobamoviruses in forest soils. *Applied and Environmental Microbiology* **64**, 1430-1435.
- Food and Agriculture Organisation (FAO) (1982): Production Yearbook. FAO, Rome.
- Foster, G. D. (1998): Carlavirus isolation and RNA extraction, pp. 145-150. In G. D. Foster, and S. C. Taylor (Eds): *Plant Virology Protocols: From virus isolation to transgenic resistance*, Humana Press, Totowa.
- Foster, G. D., and Mills, P. R. (1991): Occurrence of chloroplast ribosome recognition sites within conserved elements of the RNA genomes of carlaviruses. *FEBS Letters* **280**, 341-343.
- Foster, G. D., Millar, A. W., Meehan, B. M., and Mills, P. R. (1990): Nucleotide sequence of the 3' terminal region of Helenium virus S RNA. *Journal of General Virology* **71**, 1877-1880.

- Froussard, P. (1992): A random-PCR method (rPCR) to construct whole cDNA library from low amounts of RNA. *Nucleic Acids Research* **20**, 2900.
- Garcia, M. L., Sanchez de la Torre, M. E., Dal Bo, E., Djelouah, K., Rouag, N., Luisoni, E., Milne, R. G., and Grau, O. (1997): Detection of citrus psorosis-ringspot virus using RT-PCR and DAS-ELISA. *Plant Pathology* **46**, 830-836.
- Geering, A. D., and Thomas, J. E. (1999): Characterisation of a virus from Australia that is closely related to papaya mosaic potexvirus. *Archives of Virology* **144**, 577-592.
- German, S., Candresse, T., Lanneau, M., Huet, J. C., Pernollet, J. C., and Dunez, J. (1990): Nucleotide sequence and genomic organisation of Apple chlorotic leaf spot closterovirus. *Virology* **179**, 104-112.
- German-Ratana, S., Bergey, B., Delbos, R. P., Candresse, T., and Dunez, J. (1997): Complete nucleotide sequence of the genome of a severe cherry isolate of apple chlorotic leaf spot trichovirus (ACLSV). *Archives of Virology* **142**, 833-841.
- Gibbs, A. J., and Keese, P. K. (1995): In search of the origins of viral genes, pp. 76-90. In A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal (Eds): *Molecular basis of virus evolution*, Cambridge University Press, Cambridge.
- Graddon, D. J., and Randles, J. W. (1986): Single antibody dot immunoassay - a simple technique for rapid detection of a plant virus. *Journal of Virological Methods* **13**, 63-69.
- Gramstat, A., Prufer, D., and Rohde, W. (1994): The nucleic acid-binding zinc finger protein of potato virus M is translated by internal initiation as well as by ribosomal frameshifting involving a shifty stop codon and a novel mechanism of P-site slippage. *Nucleic Acids Research* **22**, 3911-3917.
- Haber, S., Rymerson, R. T., Procunier, J. D., Murray, G., and Cvitkovitch, S. E. (1995): Diagnosis of flame chlorosis by reverse transcription-polymerase chain reaction (RT-PCR). *Plant Disease* **79**, 626-630.
- Hadidi, A., Montasser, M. S., Levy, L., Goth, R. W., Converse, R. H., Madkour, M. A., and Skrzeczkowski, L. J. (1993): Detection of potato leafroll and strawberry mild yellow

- edge luteoviruses by reverse transcription-polymerase chain reaction amplification. *Plant Disease* **77**, 595-601.
- Halpern, B. T., and Hillman, B. I. (1996): Detection of blueberry scorch virus strain NJ2 by reverse transcriptase-polymerase chain reaction amplification. *Plant Disease* **80**, 219-222.
- Harper, G., Dahal, G., Thottappilly, G., and Hull, R. (1999): Detection of episomal banana streak badnavirus by IC-PCR. *Journal of Virological Methods* **79**, 1-8.
- Harris, K. F. (1980): Aphids, Leafhoppers and Planthoppers, pp. 1-14. In K. F. Harris, and K. Maramorosch (Eds): *Vectors of Plant Pathogens*, Academic Press, New York.
- Hayat, M. A. (1989): *Principles and techniques of electron microscopy*. CRC Press. Boca Raton.
- Haylor, M. T. M., Brunt, A. A., and Coutts, R. H. A. (1990): Conservation of the 3' terminal nucleotide sequence in five carlaviruses. *Nucleic Acids Research* **18**, 6127.
- Henson, J. M., and French, R. (1993): The polymerase chain reaction and plant disease diagnosis. *Annual Review of Phytopathology* **31**, 81-109.
- Hibi, T., and Saito, Y. (1985): A dot immunobinding assay for the detection of tobacco mosaic virus in infected tissues. *Journal of General Virology* **66**, 1191-1194.
- Hill, M. F., Giles, R. J., Moran, J. R., and Hepworth, G. (1996): The incidence of chrysanthemum stunt viroid, chrysanthemum B carlavirus, tomato aspermy cucumovirus and tomato spotted wilt tospovirus in Australian chrysanthemum crops. *Australasian Plant Pathology* **25**, 174-178.
- Hillman, B. I., and Lawrence, D. M. (1994): Carlaviruses, pp. 35-50. In K. Kohmoto, R. P. Singh, U. S. Singh, and R. Zeigler (Eds): *Pathogenesis and Host-Parasite Specificity in Plant Diseases*, Permagon, London.
- Hodgson, C. P., and Fisk, R. Z. (1987): Hybridization probe size control: optimized 'oligolabelling'. *Nucleic Acids Research* **15**, 6295.

- Hohmann, F., Fuchs, E., Gruntzig, M., and Kuntz, L. (1996): Development of a routine test for the detection of sugarcane mosaic virus (SCMV) and the maize dwarf mosaic virus (MDMV) using tissue print-immunoblotting (TPIB). *Journal of Plant Diseases and Protection* **103**, 464-471.
- Hsu, H. T., Kim, J. Y., and Lawson, R. H. (1995): Purification of lily symptomless carlavirus and detection of the virus in lilies. *Plant Disease* **79**, 912-916.
- Hughes, C. G. (1961a): Striate Mosaic: a new disease of sugarcane. *Nature* **190**, 366-367.
- Hughes, C. G. (1961b): Striate Mosaic: a new disease of sugarcane. *International Sugar Journal* **63**, 298-299.
- Hughes, C. G. (1964): Striate Mosaic, pp. 192-195. In C. G. Hughes, E. V. Abbott, and C. A. Wismer (Eds): *Sugar-Cane Diseases of the World*, Elsevier, Amsterdam.
- Huisman, M. J., Linthorst, H. J. M., Bol, J. F., and Cornelissen, B. J. C. (1988): The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus-stranded RNA viruses. *Journal of General Virology* **69**, 1789-1798.
- Ishikawa, J., and Hotta, K. (1999): FramePlot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G+C content. *FEMS Microbiology Letters* **174**, 251-253.
- James, D., Jelkmann, W., and Upton, C. (1999): Specific detection of cherry mottle leaf virus using digoxigenin-labeled cDNA probes and RT-PCR. *Plant Disease* **83**, 235-239.
- Jelkmann, W. (1994): Nucleotide sequences of apple stem pitting virus and of the coat protein gene of a similar virus from pear associated with vein yellows disease and their relationship with potex- and carlaviruses. *Journal of General Virology* **75**, 1535-1542.
- Jelkmann, W. (1995): Cherry virus A: cDNA cloning of dsRNA, nucleotide sequence analysis and serology reveal a new plant capillovirus in sweet cherry. *Journal of General Virology* **76**, 2015-2024.

- Jelkmann, W., and Keim-Konrad, R. (1997): Immuno-capture polymerase chain reaction and plate-trapped ELISA for the detection of apple stem pitting virus. *Journal of Phytopathology* **145**, 499-503.
- Jelkmann, W., Kunze, L., Vetten, H. J., and Lesemann, D. E. (1992a): cDNA cloning of dsRNA associated with apple stem pitting disease and evidence for the relationship of the virus-like agents associated with apple stem pitting and pear vein yellows. *Acta Horticulturae* **309**, 55-62.
- Jelkmann, W., Maiss, E., and Martin, R. R. (1992b): The nucleotide sequence and genome organisation of strawberry mild yellow edge-associated potexvirus. *Journal of General Virology* **73**, 475-479.
- Kanyuka, K. V., Vishnichenko, V. K., Levay, K. E., Kondrikov, D. Y., Ryabov, E. V., and Zavriev, S. K. (1992): Nucleotide sequence of shallot virus X RNA reveals a 5'-proximal cistron closely related to those of potexviruses and a unique arrangement of the 3'-proximal cistrons. *Journal of General Virology* **73**, 2553-2560.
- Koenig, R., Beier, C., Commandeur, U., Luth, U., Kaufmann, A., and Luddecke, P. (1996): Beet soil-borne virus RNA 3 - A further example of the heterogeneity of the gene content of furovirus genomes and of triple gene block-carrying RNAs. *Virology* **216**, 202-207.
- Koganezawa, H., and Yanase, H. (1990): A new type of elongated virus isolated from apple trees containing the stem pitting agent. *Plant Disease* **74**, 610-614.
- Koonin, E. V. (1991): The phylogeny of RNA-dependent RNA polymerases of positive strand RNA viruses. *Journal of General Virology* **72**, 2196-2206.
- 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.
- Kyte, J., and Doolittle, R. F. (1982): A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology* **157**, 105-132.

- Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of the bacteriophage T4. *Nature* **227**, 680-685.
- Lai, M. M. C. (1995): Recombination and its evolutionary effect on viruses with RNA genomes, pp. 119-132. In A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal (Eds): *Molecular basis of virus evolution*, Cambridge University Press, Cambridge.
- Lawrence, D. M., Rozanov, M. N., and Hillman, B. I. (1995): Autocatalytic processing of the 223-kDa protein of blueberry scorch carlavirus by a papain-like proteinase. *Virology* **207**, 127-135.
- Levay, K., and Zavriev, S. (1991): Nucleotide sequence and gene organisation of the 3' terminal region of chrysanthemum virus B genomic RNA. *Journal of General Virology* **72**, 2333-2337.
- Ligat, J. S., Cartwright, D., and Randles, J. W. (1991). Comparison of some pea seed-borne mosaic virus isolates and their detection by dot-immunobinding assay. *Australian Journal of Agricultural Research* **42**, 441-451.
- Ling, K.-S., Zhu, H.-Y., Jiang, Z.-Y., and Gonsalves, D. (2000): Effective application of DAS-ELISA for detection of grapevine leafroll associated closterovirus-3 using a polyclonal antiserum developed from recombinant coat protein. *European Journal of Plant Pathology* **106**, 301-309.
- Lockhart, B. E. L., Autrey, L. J. C., and Comstock, J. C. (1996a): Purification and serology of sugarcane mild mosaic virus, a mealybug-transmitted closterolike virus. *Phytopathology* **82**, 691-695.
- Lockhart, B. E. L., Ireby, M. J., and Comstock, J. C. (1996b): Sugarcane Bacilliform Virus, Sugarcane Mild Mosaic Virus and Sugarcane Yellow Leaf Syndrome, pp. 108-112. In B. J. Croft, C. M. Piggin, E. S. Wallis, and D. M. Hogarth (Eds): *Sugarcane Germplasm Conservation and Exchange*, ACIAR, Brisbane.
- Lough, T. J., Netzler, N. E., Emerson, S. J., Sutherland, P., Carr, F., Beck, D. L., Lucas, W. J., and Forster, R. L. S. (2000): Cell-to-cell movement of potexviruses: evidence for

- a ribonucleoprotein complex involving the coat protein and first triple gene block protein. *Molecular Plant-Microbe Interactions* **13**, 962-974.
- Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., and Scheele, G. A. (1987): Selection of AUG initiation codons differs in plants and animals. *The EMBO Journal* **6**, 43-48.
- Mackenzie, D. J., Tremaine, J. H., and Stace-Smith, R. (1989): Organization and interviral homologies of the 3'-terminal portion of potato virus S RNA. *Journal of General Virology* **70**, 1053-1063.
- Makkouk, K. M., Hsu, H. T., and Kumari, S. G. (1993): Detection of three plant viruses by dot-blot and tissue-blot immunoassays using chemiluminescent and chromogenic substrates. *Journal of Phytopathology* **139**, 97-102.
- Martelli, G. P., and Jelkmann, W. (1998): *Foveavirus*, a new plant virus genus. *Archives of Virology* **143**, 1245-1249.
- Mas, P., and Pallas, V. (1995): Non-isotopic tissue-printing hybridization: a new technique to study long-distance plant virus movement. *Journal of Virological Methods* **52**, 317-326.
- Matthews, R. E. F. (1991): *Plant Virology*. Academic Press. San Diego.
- Mayo, M. A., and Pringle, C. R. (1998): Virus Taxonomy - 1997. *Journal of General Virology* **79**, 649-657.
- Meehan, B. M., and Mills, P. R. (1991): Nucleotide sequence of the 3' terminal region of carnation latent virus. *Intervirology* **32**, 262-267.
- Memelink, J., van der Vlugt, C. I., Linthorst, H. J., Derks, A. F., Asjes, C. J., and Bol, J. F. (1990): Homologies between the genomes of a carlavirus (lily symptomless virus) and a potexvirus (lily virus X) from lily plants. *Journal of General Virology* **71**, 917-924.
- Meng, B., Pang, S., Forsline, P. L., McFerson, J. R., and Gonsalves, D. (1998): Nucleotide sequence and genome structure of grapevine rupestris stem pitting

associated virus-1 reveal similarities to apple stem pitting virus. *Journal of General Virology* **79**, 2059-2069.

Merits, A., Fedorkin, O. N., Guo, D., Kalinina, N. O., and Morozov, S. Y. (2000): Activities associated with the putative replication initiation protein of Coconut foliar decay virus, a tentative member of the genus *Nanovirus*. *Journal of General Virology* **81**, 3099-3106.

Meyers, G., Tautz, N., and Theil, H.-J. (1995): Cellular sequences in viral genomes, pp. 91-102. In A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal (Eds): *Molecular basis of virus evolution*, Cambridge University Press, Cambridge.

Miller, E. D., Plante, C. A., Kim, K.-H., Brown, J. W., and Hemenway, C. (1998): Stem-loop structure in the 5' region of potato virus X genome required for plus-strand RNA accumulation. *Journal of Molecular Biology* **284**, 591-608.

Minafra, A., Casati, P., Elicio, V., Rowhani, A., Saldarelli, P., Savino, V., and Martelli, G. P. (2000): Serological detection of grapevine rupestris stem pitting-associated virus (GRSPaV) by a polyclonal antiserum to recombinant virus coat protein. *Vitis* **39**, 115-118.

Minafra, A., Saldarelli, P., Grieco, F., and Martelli, G. P. (1994): Nucleotide sequence of the 3' terminal region of the RNA of two filamentous grapevine viruses. *Archives of Virology* **137**, 249-261.

Moonan, F., Molina, J., and Mirkov, T. E. (2000): Sugarcane yellow leaf virus: An emerging virus that has evolved by recombination between luteoviral and poleroviral ancestors. *Virology* **296**, 156-171.

Morozov, S. Y., Dolja, V. V., and Atbekov, J. G. (1989): Probable reassortment of genomic elements among elongated RNA-containing plant viruses. *Journal of Molecular Evolution* **29**, 52-62.

Morozov, S. Y., Kanyuka, K. V., Levay, K. E., and Zavriev, S. K. (1990): The putative RNA replicase of potato virus M: Obvious sequence similarity with potex- and tymoviruses. *Virology* **179**, 911-914.

- Morozov, S. Y., Lukasheva, L. I., Chernow, B. K., Prasalov, V. S., Kozlov, Y. V., Atabekov, J. G., and Skryabin, K. G. (1983): The analysis of the primary structure and localization of the coat protein gene on the genomic RNA of potato virus X. *Doklady Akademii nauk SSSR* **271**, 211-215.
- Morozov, S. Y., Solovyev, A. G., Kalinina, N. O., Fedorkin, O. N., Samuilova, O. V., Schiemann, J., and Atabekov, J. G. (1999): Evidence for two nonoverlapping functional domains in the potato virus X 25K movement protein. *Virology* **260**, 55-63.
- Mushegian, A. R., and Koonin, E. V. (1993): Cell-to-cell movement of plant viruses: Insights from amino acid sequence comparisons of movement proteins and from analogies with cellular transport systems. *Archives of Virology* **133**, 239-257.
- Navot, N., Ber, R., and Czosnek, H. (1989): Rapid detection of tomato yellow leaf curl virus in squashes of plants and insect vectors. *Phytopathology* **79**, 562-568.
- Nemchinov, L. G., Shamloul, A. M., Zemtchik, E. Z., Verderevskaya, T. D., and Hadidi, A. (2000): Apricot latent virus: a new species in the genus Foveavirus. *Archives of Virology* **145**, 1801-1813.
- Nevins, J. R. (1983): The pathway of eukaryotic mRNA formation. *Annual Review of Biochemistry* **52**, 441-466.
- Nie, X., and Singh, R. P. (2000): Detection of multiple potato viruses using an oligo(dT) as a common cDNA primer in multiplex RT-PCR. *Journal of Virological Methods* **86**, 179-185.
- Nolasco, G., Mansinho, A., Teixeira Santos, M., Soares, C., Sequeira, Z., Sequeira, C., Correia, P. K., and Sequeira, O. A. (2000): Large scale evaluation of primers for diagnosis of rupestris stem pitting associated virus-1. *European Journal of Plant Pathology* **106**, 311-318.
- Ochi, M., Kashiwazaki, S., Hiratsuka, K., Namba, S., and Tsuchizaki, T. (1992): Nucleotide sequence of the 3' terminal region of potato virus T RNA. *Annals of the Phytopathology Society of Japan* **58**, 416-425.

- Ohira, K., Namba, S., Rozanov, M., Kusumi, T., and Tsuchizaki, T. (1995): Complete sequence of an infectious full-length cDNA clone of citrus tatter leaf capillovirus: comparative sequence analysis of capillovirus genome. *Journal of General Virology* **76**, 2305-2309.
- O'Reilly, E. K., and Kao, C. C. (1998): Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology* **252**, 287-303.
- Orman, B. E., Celnik, R. M., Mandel, A. M., Torres, H. N., and Mentaberry, A. N. (1990): Complete cDNA sequence of a South American isolate of potato virus X. *Virus Research* **16**, 293-305.
- Page, R. D. M. (1998): TreeView (Win32) Version 1.5.0. <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>.
- Pan, Y.-B., Grisham, M. P., and Burner, D. M. (1998): A polymerase chain reaction protocol for the detection of *Clavibacter xyli* subsp. *xyli*, the causal bacterium of sugarcane ratoon stunting disease. *Plant Disease* **82**, 285-290.
- Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989): Identification of four conserved motifs among the RNA- dependent polymerase encoding elements. *EMBO Journal* **8**, 3867-3874.
- Randles, J. W., and Hanold, D. (1989): Coconut Foliar Decay Virus Particles Are 20-nm Icosahedra. *Intervirology* **30**, 177-180.
- Randles, J. W., Hodgson, R. A. J., and Wefels, E. (1996): The rapid and sensitive detection of plant pathogens by molecular methods. *Australasian Plant Pathology* **25**, 71-85.
- Robertson, N. L., French, R., and Gray, S. M. (1991): Use of group-specific primers and the polymerase chain reaction for the detection and identification of luteoviruses. *Journal of General Virology* **72**, 1473-1477.
- Roossinck, M. J. (1997): Mechanisms of plant virus evolution. *Annual Review of Phytopathology* **35**, 191-209.

- Rott, P. (1995): Red leaf mottle, pp. 125. In B. J. Croft, C. M. Piggin, W. E.S., and D. M. Hogarth (Eds): *Sugarcane Germplasm Conservation and Exchange*, ACIAR, Brisbane, Queensland, Australia.
- Rott, P., and Chatenet, M. (2000): Red leaf mottle, pp. 255-258. In P. Rott, R. A. Bailey, J. C. Comstock, B. J. Croft, and A. S. Saumtelly (Eds): *A guide to sugarcane diseases*, CIRAD, Montpellier.
- Rott, P., and Peterschmitt, M. (2000): Streak, pp. 259-264. In P. Rott, R. A. Bailey, J. C. Comstock, B. J. Croft, and A. S. Saumtelly (Eds): *A guide to sugarcane diseases*, CIRAD, Montpellier.
- Rott, P., Bailey, R. A., Comstock, J. C., Croft, B. J., and Saumtally, A. S. (2000): *A guide to sugarcane diseases*. CIRAD. Montpellier.
- Rowhani, A., Biardi, L., Routh, G., Daubert, S. D., and Golino, D. A. (1998): Development of a sensitive colorimetric-PCR assay for detection of viruses in woody plants. *Plant Disease* **82**, 880-884.
- Rupasov, V. V., Morozov, S. Y., Kanyuka, K. V., and Zavriev, S. K. (1989): Partial nucleotide sequence of potato virus M RNA shows similarities to potexviruses in gene arrangement and the encoded amino acid sequences. *Journal of General Virology* **70**, 1861-1869.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988). Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.
- Saldarelli, P., Minafra, A., and Martelli, G. P. (1996): The nucleotide sequence and genomic organisation of grapevine virus B. *Journal of General Virology* **77**, 2645-2652.
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989): *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Press. New York.
- Santa Cruz, S., and Baulcombe, D. (1995): Analysis of potato virus X coat protein genes in relation to resistance conferred by the genes Nx, Nb and Rx1 of potato. *Journal of General Virology* **76**, 2057-2061.

- Scagliusi, S. M., and Lockhart, B. E. L. (2000): Transmission, characterization and serology of a luteovirus associated with yellow leaf syndrome of sugarcane. *Phytopathology* **90**, 120-124.
- Schonborn, J., Oberstrab, J., Breyel, E., Tittgen, J., Schumacher, J., and Lukacs, N. (1991): Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. *Nucleic Acids Research* **19**, 2993-3000.
- Seoh, M.-L., Wong, S.-M., and Zhang, L. (1998): Simultaneous TD/RT-PCR detection of cymbidium mosaic potexvirus and odontoglossum ringspot tobamovirus with a single pair of primers. *Journal of Virological Methods* **72**, 197-204.
- Simon, A. E., and Bujarski, J. J. (1994): RNA-RNA recombination and evolution in virus-infected plants. *Annual Review of Phytopathology* **32**, 337-362.
- Singh, R. P. (1998): Reverse-transcription polymerase chain reaction for the detection of viruses from plants and aphids. *Journal of Virological Methods* **74**, 125-138.
- Sit, T. L., AbouHaidar, M. G., and Holy, S. (1989): Nucleotide sequence of papaya mosaic virus RNA. *Journal of General Virology* **70**, 2325-2331.
- Skotnicki, A. H., Dale, J. L., and Skotnicki, M. L. (1986): Detection of Fiji disease virus in infected sugarcane by nucleic acid hybridization. *Journal of Virological Methods* **13**, 71-77.
- Skryabin, K. G., Morozov, S. Y., Kraev, A. S., Rozanov, M. N., Chernov, B. K., Lukasheva, L. I., and Atabekov, J. G. (1988): Conserved and variable elements in RNA genomes of potexviruses. *FEBS Letters* **240**, 33-40.
- Smith, G. R., and Van de Velde, R. (1994): Detection of sugarcane mosaic virus and Fiji disease virus in diseased sugarcane using the polymerase chain reaction. *Plant Disease* **78**, 557-561.
- Smith, G. R., Van de Velde, R., and Dale, J. L. (1992): PCR amplification of a double-stranded RNA region of Fiji disease virus from diseased sugarcane. *Journal of Virological Methods* **39**, 237-246.

- Solovyev, A. G., Novikov, V. K., Merits, A., Savenkov, E. I., Zelenina, D. A., Tyulkina, L. G., and Morozov, S. Y. (1994): Genome characterization and taxonomy of *Plantago asiatica* mosaic potexvirus. *Journal of General Virology* **75**, 259-267.
- Song, Y.-R., Ye, Z.-H., and Varner, J. E. (1993): Tissue print hybridization on membrane for localization of mRNA in plant tissue. *Methods in Enzymology* **218**, 671-681.
- Sumi, S., Matsumi, T., and Tsuneyoshi, T. (1999): Complete nucleotide sequences of garlic viruses A and C, members of the newly ratified genus *Allexivirus*. *Archives of Virology* **144**, 1819-1826.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Research* **22**, 4673-4680.
- Thomson, K. G., Dietzgen, R. G., Thomas, J. E., and Teakle, D. S. (1996): Detection of pineapple bacilliform virus using the polymerase chain reaction. *Annals of Applied Biology* **129**, 57-69.
- Torok, V.A. and Randles, J.W. (2001): Using *Tobacco mosaic virus* RNA as an internal control for duplex RT-PCR assay of pea germplasm. *Australasian Plant Pathology* (in Press).
- Towbin, H., Staehelin, T., and Gordon, J. (1979): Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences of the USA* **76**, 4350.
- Tsuneyoshi, T., Matsumi, T., Deng, T., Sako, I., and Sumi, S. (1998): Differentiation of *Allium* carlaviruses isolated from different parts of the world based on the viral coat protein sequence. *Archives of Virology* **143**, 1093-1107.
- Van Dijk, P., Van Der Meer, F. A., and Piron, P. G. M. (1987): Accessions of Australian *Nicotiana* species suitable as indicator hosts in the diagnosis of plant virus diseases. *Netherlands Journal of Plant Pathology* **93**, 73-85.

Van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carstens, E. B., Estes, M. K., Lemon, S. M., Maniloff, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R., and Wickner, R. B. (2000): *Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses*. Academic Press. San Diego.

Varner, J. E., and Ye, Z. (1994): Tissue printing. *The FASEB Journal* **8**, 378-384.

Verchot, J., Angell, S. M., and Baulcombe, D. C. (1998): In vivo translation of the triple gene block of potato virus X requires two subgenomic mRNAs. *Journal of Virology* **72**, 8316-8320.

Wagih, M. E., and Adkins, S. W. (1996): Detection of Fiji disease virus in sugarcane by ELISA. *Australian Journal of Experimental Agriculture* **36**, 105-110.

Ward, C. W. (1993): Progress towards a higher taxonomy of viruses. *Research in Virology* **144**, 419-453.

Weiller, G. F., McClure, M. A., and Gibbs, A. J. (1995): Molecular phylogenetic analysis. In A. J. Gibbs, C. H. Calisher, and F. Garcia-Arenal (Eds): *Molecular basis of virus evolution*, Cambridge University Press, Cambridge.

White, K. A., Bancroft, J. B., and Mackie, G. A. (1992): Mutagenesis of a hexanucleotide sequence conserved in potexvirus RNAs. *Virology* **189**, 817-820.

Wong, S.-M., Lee, K.-C., Yu, H.-H., and Leong, W.-F. (1998): Phylogenetic analysis of triple gene block viruses based on the TGB 1 homolog gene indicates a convergent evolution. *Virus Genes* **16**, 295-302.

Wung, C.-H., Hsu, Y.-H., Liou, D.-Y., Huang, W.-C., Lin, N.-S., and Chang, B.-Y. (1999): Identification of the RNA-binding sites of the triple gene block protein 1 of bamboo mosaic potexvirus. *Journal of General Virology* **80**, 1119-1126.

Xu, H., Leclerc, D., Leung, B., and AbouHaidar, M. G. (1994): The entire nucleotide sequence and genomic organisation of potato aucuba mosaic potexvirus. *Archives of Virology* **135**, 461-469.

- Yang, Z. N., and Mirkov, T. E. (1997): Sequence and relationships of sugarcane mosaic and sorghum mosaic virus strains and development of RT-PCR-based RFLPs for strain discrimination. *Phytopathology* **87**, 932-939.
- Yoshikawa, N., Imaizumi, M., Takahashi, T., and Inouye, N. (1993): Striking similarities between the nucleotide sequence and genome organisation of citrus tatter leaf and apple stem grooving capilloviruses. *Journal of General Virology* **74**, 2743-2747.
- Yoshikawa, N., Sasaki, E., Kato, M., and Takahashi, T. (1992): The nucleotide sequence of apple stem grooving capillovirus genome. *Virology* **191**, 98-105.
- Yusibov, V., Kumar, A., North, A., Johnson, J. E., and Loesch-Fries, L. S. (1996): Purification, characterization, assembly and crystallization of assembled alfalfa mosaic virus coat protein expressed in *Escherichia coli*. *Journal of General Virology* **77**, 567-573.
- Zagula, K. R., Aref, N. M., and Ramsdell, D. C. (1989): Purification, serology and some properties of a mechanically transmissible virus associated with green ring mottle disease in peach and cherry. *Phytopathology* **79**, 451-456.
- Zavriev, S. K., Kanyuka, K. V., and Levay, K. E. (1991): The genome organisation of potato virus M RNA. *Journal of General Virology* **72**, 9-14.
- Zhang, Y.-P., Kirkpatrick, B. C., Smart, C. D., and Uyemoto, J. K. (1998): cDNA cloning and molecular characterization of cherry green ring mottle virus. *Journal of General Virology* **79**, 2275-2281.

Corrigendum

Table 5.2: Pear vein yellows virus is a strain of *Apple stem pitting ideoavirus*.

Table 8.2: The *potexvirus* SMYEaV is transmitted by the aphid *Chaetosiphon fragaefolii* in a persistent manner

Throughout the thesis, where nt has been used to describe PCR products or markers, this should be replaced with bp.