# A STUDY OF SOUTH AFRICAN STRAINS OF THE SUGARCANE MOSAIC POTYVIRUS (SCMV) IDENTIFIED BY SEQUENCE ANALYSIS OF THE 5' REGION OF THE COAT PROTEIN GENE

**BERNADETTE GOODMAN**

**TECHNIKON NATAL**

# A study of South African strains of the sugarcane mosaic potyvirus (SCMV) identified by sequence analysis of the 5' region of the coat protein gene

**By**

**Bernadette Goodman**

Dissertation submitted in compliance with the requirements of the Master's Degree in Technology in the Department of Biotechnology, Technikon Natal, Durban.

APPROVED FOR FINAL SUBMISSION

21 April 1999

**Supervisor:**
**Dr Barbara Isobel Huckett**
**BSc (Hons), MSc, PhD**

**DATE**

**Durban**
**April, 1999**

# PREFACE

Experimental work described in this dissertation was conducted in the Biotechnology department of the South African Sugar Association Experiment Station under the supervision of Dr Barbara Huckett and Mrs Danile Macdonald.

Funding in the form of a bursary awarded to the author as well as running costs for the project was provided by the South African Sugar Association Experiment Station. A financial contribution towards running costs for the project was also received from Technikon Natal.

These investigations represent original work conducted by the author and have not otherwise been submitted in any form for any diploma or degree to any Technikon or University. Where use was made of previously published work, it has been duly acknowledged in the text.

# DEDICATION

"Thus when the lamp which lighted
The traveller at first goes out,
He feels awhile benighted
And looks around in fear and doubt.
But soon, the prospect clearing,
By cloudless starlight on he treads,
And thinks no lamp so cheering,
As that light which Heaven sheds."

*I'd mourn the hopes*
Thomas Moore (1779-1852)

This is dedicated to my mother, the memory of my father and in a very special way to Thomas, without whom I would not have had the inspiration to complete my journey.

# ACKNOWLEDGEMENTS

The author would like to thank and acknowledge the following for their personal, professional and financial assistance during the course of this project:

God for being my eternal guiding light.

My mother for her constant encouragement, astute advice and for just being there for me.

Thomale for all the sacrifices made and being most supportive in all that I've done.

The "Goodman Gang" (my siblings) for all that you mean to me.

Mary-Anne for all the sacrifices she has had to make for me.

Tazz, Nhla[2] and Anne for their great friendship.

Debbie Carson for her assistance with sequencing and the words of encouragement.

Derek Watt for proof-reading and valuable comments.

Pieter Cronje and Roger Bailey for their assistance and allowing me the time to complete my project.

The staff of the Biotechnology Department at SASEX for the pleasant working environment.

Keith McFarlane for collecting the sugarcane leaf samples used in this study.

The South African Sugar Association and Technikon Natal for the financial assistance.

My supervisors Barbara Huckett and Danile Macdonald for their invaluable guidance and encouragement and mostly for being more than supervisors.

# ABSTRACT

Sugarcane mosaic potyvirus (SCMV) is the causal agent of the most important viral disease of sugarcane in South Africa, mosaic. Accurate knowledge of the prevalent SCMV strain(s) in the South African sugar industry is lacking and has never been determined using molecular analysis. Identification of SCMV strains at the genomic level would provide valuable information for the development of appropriate *in vitro* diagnostic tests and in the genetic engineering of sugarcane for coat protein (CP) mediated resistance. Since most strain variations reside in the capsid structure, molecular analysis of relative homology in nucleotide and inferred amino acid sequences in the most variable (N-terminal) coat protein region is the method of choice for distinguishing strains. SCMV CP sequences were specifically amplified by PCR from infected sugarcane leaf material from several varieties and geographically distinct regions within the South African sugar industry. Using SCMV-specific primer pairs, fragment sizes of 884 and 1400bp containing the N-terminal region of the CP gene were amplified from SCMV-infected material from all regions and varieties sampled. CP fragments were cloned and the sequences of the inserted fragments determined by cycle-sequencing and automated analysis. Strain identification of the CP fragments was achieved by comparison of the nucleotide sequences with those in international sequence databases using the BLASTN search programme. Prevalent strains identified by BLASTN searches were found to be SCMV-E and SCMV-D. Pairwise comparison of the nucleotide sequences of the South African isolates with those of five SCMV strains: A, B, D, E and

SC found SCMV-SC to be the most prevalent strain amongst the isolates, followed by SCMV-E and SCMV-D. An average of 13% difference at the nucleotide level was observed for sequence relationships between several clones from geographically distinct regions and varieties. Comparison of the CP sequences at the amino acid level was more ambiguous with a general spread across all five strains. Strain identification of the South African isolates was also achieved by RFLP analysis of SCMV CP PCR products using a recent method developed by Yang & Mirkov (1997) which revealed SCMV-D and SCMV-E as the most prevalent strains. Results using this approach were obtained easily and rapidly and compared very favourably with the strain identities assigned by sequence determination. The RFLP approach was thus shown to be suitable as a diagnostic tool for the detection of SCMV infection as well as strain differentiation in the SA sugar industry. Since sequence analysis of SCMV isolated from different regions and varieties revealed the most prevalent strains to be SCMV-E and SCMV-D, which are phylogenetically very closely related, it is suggested that the CP sequence of either strain could be used for genetic engineering of local sugarcane varieties to confer CP-mediated resistance.

# TABLE OF CONTENTS


| | | |
|---|---|---|
| Preface | | i |
| Dedication | | ii |
| Acknowledgements | | iii |
| Abstract | | iv |
| Table of Contents | | vi |
| List of Figures | | ix |
| List of Tables | | xii |
| List of Abbreviations | | xiv |
| **CHAPTER 1** | **INTRODUCTION** | 1 |
| 1.1 | Mosaic | 1 |
| | 1.1.1 Symptoms of the disease | 2 |
| | 1.1.2 Transmission and host range | 3 |
| 1.2 | Economic Importance of Mosaic | 4 |
| | 1.2.1 Mosaic in South Africa | 5 |
| | 1.2.2 Disease Control | 6 |
| 1.3 | Molecular Description of the Virus | 9 |
| 1.4 | Function of the Viral Gene Products | 11 |
| 1.5 | SCMV Strain Identification | 13 |
| | 1.5.1 Symptom expression | 13 |
| | 1.5.2 Serological reactions | 14 |
| | 1.5.3 Genome based methods | 15 |
| 1.6 | Reorganisation of the SCMV Subgroup | 16 |
| 1.7 | General Information on Techniques Used | 19 |
| | 1.7.1 Polymerase chain reaction | 19 |
| | 1.7.2 Reverse transcription PCR | 24 |
| | 1.7.3 Restriction fragment length polymorphism analysis | 25 |
| | 1.7.4 DNA sequencing | 25 |
| | 1.7.5 International sequence database comparisons | 27 |
| 1.8 | Importance of Viral Strain Identification | 28 |
| | 1.8.1 SCMV diagnosis | 29 |
| | 1.8.2 Viral coat protein-mediated resistance | 29 |
| 1.9 | Aims of the Study | 30 |

**CHAPTER 2 MATERIALS AND METHODS** **32**

2.1 Materials 32
2.1.1 Source of plant material 32
2.1.2 Oligonucleotide primers 32
2.1.3 Cloning vectors 37
2.1.4 Bacterial strains 37
2.1.5 Maintenance of plasmids and bacterial strains 39
2.2 Methods 39
2.2.1 General precautions taken while working with RNA 39
2.2.2 Preparation of DEPC-treated water 40
2.2.3 RNA extraction an purification 40
2.2.4 Total RNA yield determination 41
2.2.5 First strand copy DNA (cDNA) synthesis 42
2.2.6 PCR amplification 43
a) Primary PCR amplification of viral sequences 43
b) Semi-nested PCR amplification of viral sequences 43
c) Screening of putative positive bacterial transformants 44
d) Amplification of 900bp for RFLP analysis 44
2.2.7 Restriction digests 44
2.2.8 Gel electrophoresis and visualisation of DNA 45
2.2.9 Purification of PCR products 45
a) Resolubilisation in TE buffer 46
b) QIAquick gel extraction 46
c) GenElute gel extraction 46
2.2.10 Insertion of PCR fragments into cloning vectors 47
a) Blunt-ended cloning 47
b) T-A cloning 47
2.2.11 Preparation of competent cells 48
2.2.12 Transformation of bacterial strains with plasmid vectors 48
a) *E.coli* JM83 49
b) *E.coli* INV$\alpha$F' 49
c) *E.coli* JM109 49
2.2.13 Screening for transformants 50
2.2.14 Plasmid isolation 50
2.2.15 Sequence determination 51
a) Sequencing system used 51
b) Direct sequencing of viral sequences 51
c) Sequencing after cloning 52

2.2.16 International sequence database searches 53
2.2.17 Sequence analysis 54

**CHAPTER 3 RESULTS 57**

3.1 Total RNA from Infected Leaves: Yield and Purity 57
3.2 Primary Generation of the CP fragment 59
3.3 Confirmation of SCMV CP identity 62
3.3.1 Semi-nested PCR amplification of CP fragments 62
3.3.2 Re-amplification of CP fragments 63
a) Effect of template concentration 65
b) Effect of $MgCl_2$ and KCl concentration and reaction buffer pH 65
c) Effect of annealing temperature 68
d) Effect of thermostable DNA polymerase enzyme 69
3.4 Cloning of CP Gene Fragments Into Plasmid Vectors 71
3.4.1 Cloning of the 884bp CP Gene Fragments into pCR-Script $SK^+$ 72
3.4.2 Cloning of the 1400bp CP Gene Fragments into pCR2.1 and pGEM-T Easy T-A Vectors 74
3.5 Recombinant Plasmid Analysis 77
3.5.1 Restriction enzyme analysis of recombinant plasmids 79
3.6 Direct Sequencing Compared to Sequencing After Cloning 83
3.7 Preliminary Analysis of SCMV Coat Protein Fragment Sequences 86
3.8 Sequence Comparison of CP Fragment Sequences for Strain Identification 94
3.9 Restriction Fragment Length Polymorphism (RFLP) Analysis for Strain Identification 103
3.9.1 *Hinf* I restriction analysis 104
3.9.2 *Taq* I restriction analysis 106

**CHAPTER 4 DISCUSSION 110**

**CHAPTER 5 CONCLUSIONS 118**

**LIST OF REFERENCES 119**

**APPENDIX I**
Presentations and publications based on work in this dissertation.

# LIST OF FIGURES


1.1. Examples of mosaic symptoms on the highly susceptible variety NCo376. 2

1.2. Electron micrograph of the filamentous SCMV particle (virion). 9

1.3. Diagram of the sugarcane mosaic virus genome coding region. 10

1.4. Phylogenetic relationships of coat protein gene sequences of several strains of the SCMV subgroup. 19

2.1 Geographic and varietal sources of SCMV-infected sugarcane sampled for viral genome analysis. 33

2.2 Graphic representation of the SCMV coat protein (CP) gene and a portion of the polymerase (NIb) gene. 36

2.3 Diagrammatic representation of the three plasmid cloning vectors used for insertion of SCMV CP gene fragments. 38

3.1 Qualitative analysis of total RNA from SCMV infected sugarcane leaves. 58

3.2. Effect of cDNA template concentration on PCR amplification of the 884bp SCMV CP fragment generated with primer pair BHf/GSr. 61

3.3. Effect of cDNA template concentration on PCR amplification of the1400bp SCMV CP fragment generated with primer pair AGPOT2/GSr from several sources. 61

3.4. Confirmation of SCMV CP sequence identity by semi-nested PCR amplification of a 360bp fragment from the 1400bp fragment. 64

3.5. An example of the results of decreased template concentration for re-amplification of the 884bp fragment. 66

3.6. An example of increased dilution of template in re-amplification of the 884bp fragment derived from Tongaat NCo376. 66

3.7. An example of three different annealing temperatures tested on Glendale N16 884bp fragment re-amplification using a series of template dilutions. 69

3.8. Effects of three different polymerase enzymes on the re-amplification of four different template concentrations of the 884bp fragment derived from the Ixopo N12 sample. 70

3.9. An example of blue-white selection of colonies obtained after plating *E.coli* transformed with CP gene sequences cloned in pCR-Script SK⁺ on to agar plates containing IPTG and X-gal. 73

3.10. Results of direct PCR amplification of bacterial cells from white colonies using specific primers to test for the presence of the 884bp fragment in pCR-Script SK⁺. 73

3.11. Results of direct PCR amplification of cells from white bacterial colonies using universal primers to test for the presence of the 1300bp in pCR2.1 vector. 76

3.12. Agarose gel electrophoretic analysis of plasmid DNA containing the 1400bp CP gene fragment from all regions sampled and their respective varieties. 78

3.13. Nucleotide sequence of the expected 1400bp fragment amplified by PCR using the primer pair AGPOT2 / GSr. 80

3.14. Analysis of *EcoR* I digested plasmid DNA containing the 1400bp CP gene fragment from SCMV infected source samples. 82

3.15. Analysis of *Not* I digested plasmid DNA containing the 1400bp CP gene fragment from SCMV infected source samples. 84

3.16. An example of a diagrammatic representation of the generation of consensus sequences from the primary overlap alignment of sequence repetitions of the 5' region of the cloned 1400bp CP fragment. 92

3.17. Consensus sequence of three sequences using BHf primer on clone #2 of the 1400bp CP fragment derived from Eston variety NCo376. 93

3.18. Multiple alignment of a 450bp region from the start of the coat protein gene of five strains of SCMV and plasmids containing the 1400bp CP fragment from several regions and varieties. 100

3.19. A dendrogram showing the relatedness of plasmids containing the 1400bp CP fragment from several regions and varieties with

the five SCMV strains. 101

3.20. Agarose gel electrophoretic separation of PCR products of an approximately 900bp fragment used for RFLP analysis in SCMV strain differentiation. 104

3.21. Restriction patterns after *Hinf* I digestion of the 900bp SCMV-F3/SCMV-R3 fragment from all SCMV infected source samples. 105

3.22. Restriction patterns after *Taq* I digestion of the 900bp from all SCMV infected source samples. 107

# LIST OF TABLES


1.1 Distribution of mosaic in the Natal sugarcane belt. 5

1.2. Effect of mosaic on the yield and components of yield in varieties NCo376 and N12 at Eston and Mount Edgecombe. 6

1.3. Summary of a mosaic survey of the South African sugar industry during the period of June 1995 to May 1996. 7

2.1. Viral-specific primers used in various applications. 33

2.2. Universal primers used in various applications. 34

3.1. Yield and purity of total RNA from infected leaf material from various regions and varieties as determined by spectrophotometric analysis 59

3.2. Matrix of $MgCl_2$ and KCl concentrations and pH values of the buffers of the Opti-Prime (Stratagene) kit. 67

3.3. Results of cloning the 884bp and 1300bp coat protein gene fragments into various plasmid cloning vectors. 75

3.4. Expected cleavage profiles of recombinant plasmids digested with *EcoR* I restriction endonuclease. 81

3.5. Expected cleavage profile of recombinant plasmids digested with *Not* I restriction endonuclease. 83

3.6. Comparison of sequence data obtained using direct sequencing or sequencing after cloning. 85

3.7. Results of cycle-sequencing three recombinant plasmids containing the 1400bp CP fragment from each region sampled and its respective variety. 88

3.8. Sequence results of recalcitrant plasmids using the universal primer T7 for confirmation of the presence of vector sequences. 90

3.9. Summary of results obtained from the comparison of partial 1400bp CP fragment nucleotide sequences with reference sequences in international databases. 95

3.10. Percentage sequence homology calculated from the pairwise alignment of each plasmid sequence with the nucleotide sequence of the five SCMV strains: SCMV-A (U57354); SCMV-B (U57355); SCMV-D (U57356); SCMV-E (U57357) and SCMV-SC (D00958) using the homology search function in the sequence analysis programme DNASIS v2.1. 97

3.11. Summary of results obtained from the sequence comparison of the first 150 amino acids from the start of the CP gene derived from several regions and varieties with reference sequences in international databases. 102

3.12. Expected restriction profile of *Hinf* I digested 900bp fragment of five strains in the SCMV subgroup. 105

3.13. Expected restriction profile of *Taq* I digested 900bp fragment of five strains in the SCMV subgroup 107

3.14. A summary of strain identification results restriction fragment length polymorphism analysis of the approximately 900bp SCMV-F3/SCMV-R3 fragment derived from the SCMV infected source samples. 109

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| (d)T | - | Deoxythymidine |
| $^{32}P$ | - | Phosphorus radioisotope |
| $^{35}S$ | - | Sulphur radioisotope |
| $6K_1$ | - | Putative 6Kd peptide |
| $6K_2$ | - | Second putative 6Kd peptide |
| A | - | Adenine |
| $A_{260}$ | - | Absorbance at 260nm |
| $A_{280}$ | - | Absorbance at 280nm |
| AL | - | Alignment sequence |
| AMV | - | Avian myoblastosis virus |
| ATP | - | Adenosine triphosphate |
| AzMV | - | Azuki mosaic potyvirus |
| BCMV | - | Bean common mosaic potyvirus |
| BLAST | - | Basic local alignment search tool |
| bp | - | Base pairs |
| BSA | - | Bovine serum albumin |
| C | - | Cytosine |
| $CaCl_2$ | - | Calcium chloride |
| cDNA | - | Complementary deoxyribonuleic acid |
| CI | - | Cytoplasmic inclusion protein |
| CON | - | Consensus sequence |
| CP | - | Coat protein |
| dATP | - | Deoxyadenosine triphosphate |
| dCTP | - | Deoxycytosine triphosphate |
| ddNTPs | - | Dideoxynucleotide triphosphates |
| DEPC | - | Diethyl pyrocarbonate |
| dGTP | - | Deoxyguanosine triphosphate |
| $dH_2O$ | - | Distilled water |
| DNA | - | Deoxyribonucleic acid |
| dNTPs | - | Deoxynucleotide triphosphates |
| dTTP | - | Deoxythymidine triphosphate |
| EDTA | - | Ethylene diamine trichloroacetic acid |
| ers | - | Estimated recoverable sucrose |
| G | - | Guanine |
| ha | - | Hectare |
| HC-Pro | - | Helper component protein |
| IPTG | - | Isopropyl-β-D-thiogalactoside |
| IUB | - | International Union of Biochemistry |
| JGMV | - | Johnson grass mosaic potyvirus |

| | | |
|---|---|---|
| $K^+$ | - | Potassium anion |
| Kac | - | Potassium acetate |
| kb | - | Kilobase |
| KCl | - | Potassium chloride |
| Kd | - | Kilodalton |
| LB | - | Luria Bertani |
| MDMV | - | Maize dwarf mosaic potyvirus |
| mg | - | Milligram |
| $Mg^{2+}$ | - | Magnesium anion |
| $MgCl_2$ | - | Magnesium chloride |
| µg | - | Microgram |
| µl | - | Microlitre |
| µM | - | Micromolar |
| min | - | Minute |
| ml | - | Millilitre |
| mM | - | Millimolar |
| MuLV | - | Moloney murine leukeamia virus |
| NaCl | - | Sodium chloride |
| NCBI | - | National Centre for Biotechnology Information |
| ng | - | Nanogram |
| NIa | - | Small nuclear inclusion protein |
| NIb | - | Large nuclear inclusion protein |
| nm | - | Nanometre |
| nr | - | Non-redundant |
| NSH | - | No significant homology |
| N-terminus | - | Amino terminus |
| °C | - | Degree Celsius |
| P1 | - | First protein |
| P3 | - | Third protein |
| PCR | - | Polymerase chain reaction |
| PCR-RFLP | - | Polymerase chain reaction restriction fragment length polymorphism |
| Pipes-HCl | - | Piperazine-1, 4-bis-2-ethane sulfonic acid hydrochloride |
| Poly-A | - | Poly-adenosine |
| PPV | - | Plum pox potyvirus |
| PRSV | - | Papaya ringspot potyvirus |
| PRSV-W | - | PRSV type W |
| PVY | - | Potato potyvirus Y |
| PWV | - | Passionfruit woodiness potyvirus |
| RFLP | - | Restriction fragment length polymorphism |

| | | |
|---|---|---|
| RNA | - | Ribonucleic acid |
| rpm | - | Revolutions per minute |
| RT-PCR | - | Reverse transcription polymerase chain reaction |
| SA | - | South Africa |
| SASEX | - | South African Sugar Association Experiment Station |
| SCMV | - | Sugarcane mosaic potyvirus |
| SCMV-A | - | SCMV strain A |
| SCMV-B | - | SCMV strain B |
| SCMV-D | - | SCMV strain D |
| SCMV-E | - | SCMV strain E |
| SCMV-JG | - | SCMV Johnson grass strain |
| SCMV-MDB | - | SCMV maize dwarf B strain |
| SCMV-SC | - | SCMV sugarcane strain |
| SDS | - | Sodium dodecyl sulphate |
| sec | - | Second |
| SrMV | - | Sorghum mosaic potyvirus |
| T | - | Thymine |
| TBE | - | Tris-borate-EDTA |
| TE | - | Tris-hydrochloride EDTA |
| TMV | - | Turnip mosaic potyvirus |
| Tris-HCl | - | Tris-hydrochloride |
| U | - | Unit |
| V/cm | - | Volts per centimetre |
| v/v | - | Volume per volume |
| VPg | - | Genome linked protein |
| w/v | - | Weight per volume |
| WMV-2 | - | Watermelon mosaic potyvirus 2 |
| WSMV | - | Wheat streak mosaic rymovirus |
| X-gal | - | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |
| ZYMV | - | Zucchini yellow mosaic potyvirus |

# CHAPTER 1

## INTRODUCTION AND LITERATURE REVIEW

### 1.1 Mosaic

Sugarcane mosaic potyvirus (SCMV) is the causal agent of the most important virus disease of sugarcane in South Africa, namely mosaic (Anon., 1980). It has, at one time or another, occurred in virtually every important sugarcane growing country (Edgerton, 1955; Koike & Gillaspie, 1989; Shukla *et al.*, 1994). The viral aetiology of mosaic in sugarcane was first established by Brandes (1919). Wuthrich (1921) reported that he noticed the "yellow stripe" disease in sugarcane fields in Natal in 1918 and again in 1919. The disease was probably introduced into the South African sugar industry in 1914 by the importation of Java seedlings from the Argentine (Storey, 1924). By 1922 most commercial sugarcane varieties were infected with mosaic with the exception of variety Uba which was highly resistant. All infected cane was eradicated by 1927 and no further reports on mosaic were made until 1943 when it re-emerged in the Umzinto area on the variety Co281 (King, 1952). Since 1976 a general increase in the incidence of mosaic has been evident from scattered, severe outbreaks of mosaic in susceptible varieties (Bailey, 1979). Subsequently varieties with increased resistance to mosaic have been released in an attempt to control the prevalence of mosaic in the industry (Bailey & Fox, 1980).

### 1.1.1 Symptoms of the Disease

Mosaic may be recognised by its characteristic leaf symptoms, which vary in intensity with the cane variety, growing conditions, temperature and the strain of the SCMV involved (Koike & Gillaspie, 1989). Across the natural host range, symptoms of infection range from mosaic and ringspot to necrosis and severe stunting (Brunt *et al.*, 1996). Distinct symptoms appear only on young, developing leaves approximately ten days after infection (Koike & Gillaspie, 1989). Therefore symptoms may not always be easily recognisable on mature leaves and in certain varieties (Anon., 1994a). Leaf symptoms are particularly clear in the sugarcane varieties NCo376 and NCo293 but less obvious in some other varieties (Anon., 1980). Characteristic symptoms, shown in Figure 1.1, are seen as patterns on the leaf of darker green 'islands' on a background of paler green or yellowish chlorotic areas.

**Figure 1.1**. Examples of mosaic symptoms on the highly susceptible variety NCo376. The illustration on the left shows the dark green "islands" on a paler green background, while that on the right shows the presence of yellow chlorotic areas.

Evidence of these chlorotic areas is most common in young, rapidly growing leaves and is particularly distinct at the base of the leaves (Koike & Gillaspie, 1989). Chlorotic areas may also be present on the leaf sheath, but rarely occur on the stalk. The chloroplasts in the chlorotic areas of infected leaves are smaller and less numerous than those in normal leaves (Irvine, 1972). Irvine (1972) reported that infected varieties exhibit lower photosynthetic ability as a result of the reduction in the number of their chloroplasts, ultimately leading to crop yield loss through stunting. Mosaic symptoms do not appear in leaves fully developed prior to virus infection, even though the virus may be carried to all parts of the plant. Viral concentrations are, as a result, highest in young leaves showing symptoms of the disease (Koike & Gillaspie, 1989).

### 1.1.2 Transmission and Host Range

The disease can be transmitted in several ways: naturally through insect vectors, mechanically by seed transmission or through artificial inoculation. Natural spread occurs through aphid vectors in a non-persistent manner i.e. the virus does not replicate within the insect (Brunt *et al.*, 1996). The most common aphid species involved in transmission in South African sugarcane fields are *Hysteroneura setariae* and *Rhopalosiphum maidis* (Anon., 1994a) of which the former occurs in greater numbers (Anon., 1993). Mechanical transmission of the disease is through planting infected setts. Artificial inoculation is achieved by applying plant sap obtained from infected plants to leaves which have been damaged by puncturing or by abrasives (Koike & Gillaspie, 1989). No transmission occurs through pollen (Brunt *et al.*, 1996).

SCMV has a natural host range which spans three plant families, 109 genera and 298 species (Shukla *et al.*, 1994). Depending on the strain(s) of SCMV which infect ancestral species of sugarcane (*Saccharum spp.*), *Saccharum officinarum* is the most susceptible, while *S. barberi* Jeswiet and *S. robustum* Brandes and Jeswiet are moderately susceptible and *S. sinense* and *S. spontaneum* are more resistant or immune (Koike & Gillaspie, 1989). Of the South African commercial sugarcane varieties still being grown, varieties NCo376, NCo293 and N19 are the most susceptible to mosaic infection (Anon., 1994b).

## 1.2 Economic Importance of Mosaic

Mosaic is among the most widely distributed sugarcane diseases in the world (Koike & Gillaspie, 1989) and its economic importance varies from country to country. Economic losses depend on the susceptibility of the clone, the strain of the virus, interaction with other diseases, vector populations and environmental conditions. Some countries are not greatly affected by mosaic but others have suffered substantial losses by severe outbreaks of mosaic. The most notable of these is the epidemic that struck Louisiana in the mid 1920s causing a near collapse of the industry. The detrimental effect of the mosaic infection was as a result of coinfection with *Pythium* root rot and red rot (Koike & Gillaspie, 1989). This type of synergistic interaction with unrelated pathogens to cause unusually severe diseases has also been reported for other potyviruses (Francki *et al.*, 1991).

### 1.2.1 Mosaic in South Africa

In South Africa mosaic mostly occurs in the cooler, southern areas of the sugarcane belt and the high altitude inland areas (Chambers, 1981). This is evident in the results (Table 1.1) of the survey conducted by Thomson (1962) on the occurrence of mosaic in the various regions in the industry.

**Table 1.1** Distribution of mosaic in the Natal sugarcane belt.

| **Region** | **Number of Hills Inspected** | **Percentage Mosaic Infected** |
|---|---|---|
| South Coast | 267 | 32 |
| North Coast | 556 | 29 |
| Zululand | 436 | 3 |
| Pongola | 187 | 0.5 |

(Adapted from Thomson, 1962).

In the 1978/79 season, the estimated loss due to mosaic infection was 40 000 tons of cane and 4 400 tons of sugar, since approximately 25% of the cane production is in regions of high mosaic incidence (Bailey, 1979). Reports of crop losses in susceptible South African varieties, such as NCo293 and NCo376, were as high as 50% or more (Bailey & Fox, 1980). In 1987 these susceptible varieties were reported to constitute approximately 80% of the crop grown in the high altitude and coastal hinterland areas of Natal (Bailey & Fox, 1987). Those authors found that reductions in sucrose yield were due to the reduction in stalk populations and stalk mass caused by mosaic infection (Table 1.2).

**Table1.2**. Effect of mosaic on the yield and components of yield in varieties NCo376 and N12 at Eston and Mount Edgecombe. Figures are expressed as percent differences between healthy and SCMV infected plots, where + indicates an increase and - indicates a reduction.

| Area | Variety | Stalks with symptoms | Stalk population | Stalk mass | Tons cane/ha | Estimated recoverable sugar (ers) | Tons ers/ha |
|---|---|---|---|---|---|---|---|
| Eston | NCo376 | +65 | -2 | -25 | -27 | -2 | -28 |
| | N12 | +54 | -2 | -13 | -11 | -7 | -16 |
| Mount Edgecombe | NCo376 | +85 | -13 | -18 | -29 | +4 | -25 |
| | N12 | +73 | -10 | -4 | -14 | +3 | -11 |

Taken from Bailey & Fox, 1987.

Incalculable financial losses are incurred annually by the discarding of potentially useful new sugarcane germplasm in the breeding programme as a result of mosaic susceptibility. Over 40% of new germplasm is discarded in the selection stages (Butterfield & Thomas, 1996). However, most varieties released since 1978 have increased resistance to mosaic (Bailey & Fox, 1980).

An indication of reduced percentage of susceptible cane contributing to cane crop is the reduction from 81% in the 1978-79 season to 37% in the 1993-94 season (Anon, 1994b).

Mosaic infection, however is still of serious concern in some regions such as the Midlands and the South Coast as shown in Table 1.3, where the Midlands region showed a high of 77% mosaic infection.

### 1.2.2 Disease Control

Mosaic is not very easy to control as it is a systemic disease and has the

**Table 1.3**. Summary of a mosaic survey of the South African sugar industry during the period of June 1995 to May 1996.

| Region/Mill Area | Area Surveyed (ha) | % Area with Mosaic |
|---|---|---|
| **Zululand:** | | |
| Felixton | 5622 | 9 |
| Amatikulu | 1584 | 6 |
| Entumeni | 873 | 9 |
| Mean/Total | 8079 | 8 |
| **North Coast::** | | |
| Glendale | 2160 | 83 |
| Darnall | 4146 | 44 |
| Gledhow | 2242 | 78 |
| Maidstone | 4048 | 21 |
| Mt Edgecombe | 621 | 19 |
| Mean/Total | 13217 | 49 |
| **South Coast:** | | |
| Sezela | 1215 | 69 |
| Umzimkulu | 2925 | 30 |
| Mean/Total | 4140 | 49 |
| **Midlands:** | | |
| Midlands South | 122 | 65 |
| Noodsberg/Union Co-Op | 3419 | 90 |
| Mean/Total | 3541 | 77 |

(Taken from Anon., 1996a)

potential to cause very serious losses over large areas (Anon., 1980). The presence of alternate host plants for the virus and the aphid vectors are additional factors which hinder short-term control (Anon., 1994b). Various strains of SCMV occur and they differ in their ability to infect different varieties, virulence and symptom expression (Cronje *et al.*, 1994). Disease

control is achieved mainly by varietal resistance (Anon., 1980); however certain preventative measures also need to be employed. Irons and Payn (1992) reported on the effective control of the mosaic problem on farms in the South Coast region of Dumisa. The strategy involved the implementation of the following practices:

- planting of varieties resistant to mosaic
- use of seedcane free of disease
- elimination of volunteer cane from fields to be replanted
- elimination of diseased cane stools from plant cane fields and seedcane nurseries
- avoiding planting during peak periods of mosaic aphid vector activity (mid-November to mid February)

Implementation of these measures and changing to the more resistant variety, N12, led to a remarkable drop in the mean seasonal mosaic levels from 7.5% in the 1985/86 season to 2% in the 1991/92 season (Irons & Payn, 1992). Over the past few years there has been good progress made in the breeding and selection of varieties for resistance to mosaic, based on the expression of visual symptoms (Cronje *et al.*, 1994). This has resulted in the release of more resistant varieties such as N12, N16, N17 and N21 (Anon., 1994b). However, Cronje *et al.* (1994) established that SCMV infection in sugarcane is not always indicated by symptom expression and this may affect the screening for resistance or susceptibility to SCMV in the breeding programme.

## 1.3 Molecular Description of the Virus

SCMV belongs to the *Potyvirus* genus in the Potyviridae family. The *Potyvirus* genus is the largest economically important group of plant viruses of which potato virus Y is the type member (Shukla *et al.*, 1994). A unique phenotypic characteristic of the group is the formation of pinwheel cylindrical cytoplasmic inclusions (Ward *et al.*, 1992). Electron microscopic studies show that the filamentous, flexuous particles of SCMV, shown in Figure 1.2, measure about 750nm in length (Koike & Gillaspie, 1989).

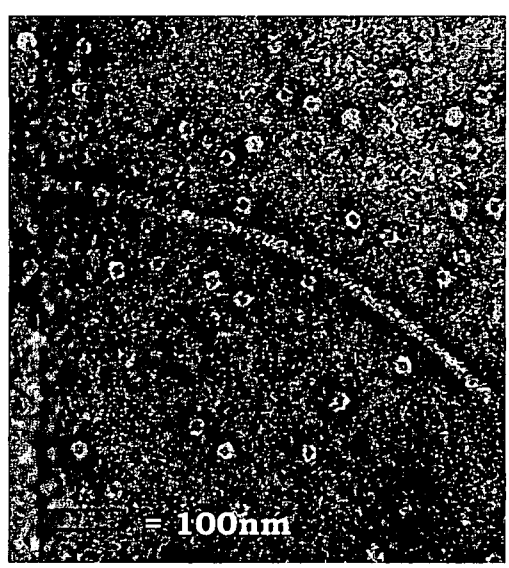



**Figure 1.2.** Electron micrograph of the filamentous SCMV particle (virion). Courtesy of CPR Cronje, SASEX.

Each SCMV particle (virion) is surrounded by a protein coat. The virion made up of a viral genome which consists of only 5% of nucleic acid while the coat protein accounts for the remaining 95% (Gough & Shukla, 1981). The genome of SCMV is approximately 9.5 kilobases (kb) in size and consists of a single-stranded, positive-sense RNA molecule which is polyadenylated at the 3' end (Brunt *et al.*, 1996). The RNA molecule encodes a polyprotein which is post-transcriptionally processed into a number of proteins. A diagrammatic

representation of the SCMV genome is shown by Figure 1.3.

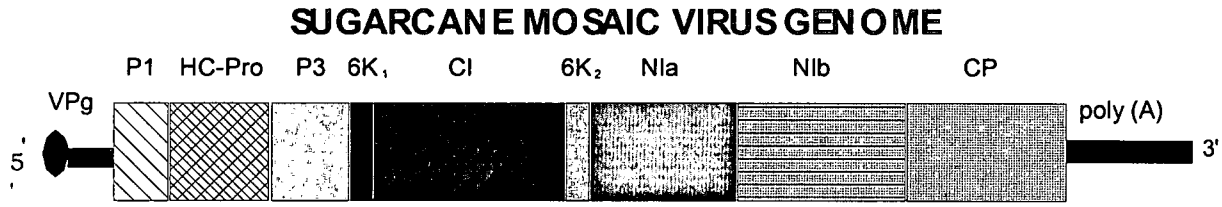



**Figure 1.3.** Diagram of the sugarcane mosaic virus genome coding region. Abbreviations: P1 = first protein, HC-Pro = helper component protease, P3 = third protein, $6K_1$ = putative 6Kd peptide (VPg), CI = cytoplasmic inclusion protein (RNA helicase), $6K_2$ = second putative 6Kd peptide, NIa = small nuclear inclusion protein (protease), NIb = large nuclear inclusion protein (RNA polymerase), CP = coat protein and poly(A) = polyadenylated 3' tail. (Adapted from Smith & Van de Velde, 1994).

The polyprotein is cleaved by proteinases into at least eight polypeptides in the following order: the first protein (P1), helper component (HC), the third protein (P3), cytoplasmic inclusion protein (CI), the small nuclear inclusion protein (NIa) (NIa + a protein covalently attached to the 5' end (VPg), the large nuclear inclusion protein (NIb) and the coat protein (CP). Cleavage site motifs are present for two small products: 6K1 and 6K2. VPg and CP are the only gene products detected in viral particles while the other gene products have been detected in infected plants (Shukla *et al.*, 1994). The major proteins are CP, VPg, HC, CI, NIa and the NIb (Shukla & Ward, 1989). The other proteins are P1 and P3 (Shukla *et al.*, 1991). The coat protein is made up of protein subunits that have a molecular weight of 28 500 - 36 500 Daltons (Koike & Gillaspie, 1989). Two thousand coat protein subunits encapsidate the RNA genome in a helical manner with seven to eight subunits in each ring, and these are considered to be important for virus infection (Shukla & Ward, 1989).

## 1.4 Functions of the Viral Gene Products

Functions have been assigned to the viral proteins but some of this work is still speculative (Zaccomer *et al.*, 1995).

*The first protein (P1 protein)*

P1 may be involved in cell-to-cell movement of the virus (Dougherty & Carrington, 1988) or specific virus-host interactions. It is also serves as the proteinase responsible for cleavage of the P1-HC junction (Shukla *et al.*, 1991).

*The third protein (P3 protein)*

P3 is considered to play a role in proteolysis of the polyprotein (Dougherty & Carrington, 1988) or other virus-specific functions (Shukla *et al.*, 1991).

*The helper component (HC protein)*

HC is responsible for insect transmission (Dougherty & Carrington, 1988) as well as the autolytic cleavage at the junction between HC and P3 (Shukla *et al.*, 1991) It exhibits a high degree of cleavage site specificity (Shukla *et al.*, 1994).

*The cytoplasmic inclusion (CI protein)*

CI protein is important in virus replication (Dougherty & Carrington, 1988), facilitating transcription and translation (Shukla *et al.*, 1991) by acting as a RNA helicase (Zaccomer *et al.*, 1995).

*The genome-linked protein (VPg)*

VPg is linked to the 5' end of the RNA molecule and is believed to perform two major functions (Shukla *et al.*, 1994). Firstly, it plays a role in replication by acting as a primer for viral nucleic acid synthesis (Dougherty & Carrington, 1988). It also protects mRNA against endonuclease attack (Shukla *et al.*,

1991).

*The small nuclear inclusion protein (NIa)*

NIa is part of the nuclear inclusion body complex formed with NIb and acts as a proteinase (Dougherty & Carrington, 1988) involved in the cleavage of the C-terminal two thirds of the polyprotein. NIa autocatalytically releases from the polyprotein by cleaving the CI-NIa and NIa-NIb junctions and catalyses the production of CI, NIb, and CP by cleaving the P3-CI and NIb-CP junctions. NIa proteinase is also involved in self-processing to liberate the genome-linked protein (VPg) (Shukla *et al.*, 1994).

*The large nuclear inclusion protein (NIb)*

NIb is the most conserved amino acid sequence of the viral gene products (Shukla *et al.*, 1991) and serves as the RNA-dependent RNA polymerase (Zaccomer *et al.*, 1995).

*The coat protein (CP)*

The CP is the most well characterised of all the gene products and consists of the highly variable, surface-exposed N-terminus, the highly conserved core region and the surface-exposed C-terminus. The N-terminus of the CP is the largest region in the virus that is unique to each virus as it contains virus-specific epitopes (Shukla *et al.*, 1991). The CP is composed of protein subunits which have a molecular weight of 28 500 - 36 500 Daltons (Koike & Gillaspie, 1989). Approximately two thousand coat protein subunits encapsidate the RNA genome in a helical manner with seven to eight subunits in each ring (Shukla & Ward, 1989). The major role of the CP is encapsidation of the genome (Dougherty & Carrington, 1988) as well as symptom production and host specificity (Shukla *et al.*, 1991).

## 1.5 Identification of SCMV Strains

Several years after mosaic was demonstrated to be a viral disease Brandes, 1919), no suggestions had been made that a number of strains might exist even though it was reported that it was possible to distinguish plants showing mild symptoms from plants that exhibited severe symptoms (Edgerton, 1955). Storey (1924) reported that in South Africa two distinct strains of mosaic were evident on members of the grass family. Summers (1934) was the first to report that he was able to recognise four strains in Louisiana and later reported on 10 strains or sub-strains of SCMV to which he assigned letter symbols. The strains A-G were earliest identified and later strains H-M (Edgerton, 1955), but Abbott and Tippet (1966) found strain G to be a variant of strain B. The identification of strains was made using the variation in the pattern of mosaic symptoms exhibited on differential sugarcane varieties (symptomatology), since strains differ with respect to their physical and chemical properties and in the symptoms they cause in sugarcane (Kondaiah & Nayudu, 1984). Strains A, B, D, F, H and M were found to cause mild mosaic symptoms on differentials, while strains C, E, I, Jg, K and L cause either severe necrosis, stunting or mosaic patterns (Chambers, 1981). Subsequently, strains of SCMV have been identified at a molecular level using serology and other genome-based methods. The merits of these various approaches are discussed below.

### 1.5.1 Symptom Expression

As mentioned above, the early identification of strains was made using the variation in the pattern of mosaic symptoms exhibited on sugarcane varieties,

sorghum or Johnson grass (symptomatology), since strains differ with respect to their physical and chemical properties and in the symptoms they cause (Kondaiah & Nayudu, 1984). Expression of symptoms is greatly influenced by environmental conditions (Gillaspie *et al.*, 1986) and as a result the use of a large number of differential hosts allows for the development of additional symptom types. Gillaspie *et al.* (1986) found that the Abbott and Tippet (1966) method of strain differentiation, in which only two varieties of sugarcane (CP 31-294 & CP 31-588) are used as differential hosts, led to ambiguities of strains that were very similar. Symptomatology, therefore cannot be used solely as a reliable means of strain identification (Shukla *et al.*, 1994).

### 1.5.2 Serological Reactions

From the late 1960s serology was used to supplement the use of differential hosts for strain identification but later became the major criterion used to characterise the different strains of SCMV. Compared to the use of differential hosts in strain identification, serology offers a means of more specifically differentiating between the strains at a molecular level based on reactions between the immunogenic viral coat proteins and the antisera produced against them. However, serological relationships between related strains of potyviruses have been found to be inconsistent (Shukla & Ward, 1989). This is because the antisera raised against the viral particles contain antibodies (polyclonal) that recognise the antigens of closely related strains of the virus and result in cross-reactions. This problem can be overcome by using monoclonal antibodies produced against the virus-specific epitopes located in the N-terminus of the coat protein (Shukla *et al.*, 1988). Monoclonal

antibodies are very expensive to produce but would be suitable for application as a large scale diagnostic technique.

### 1.5.3 Genome Based Methods

The genome-based system offers a more sensitive method of differentiating between strains since it makes use of differences that exist between the nucleotide sequences encoding viral characteristics. Therefore genome methods are a more direct indication of strain differences than serological methods. Emphasis has been placed on coat protein gene sequences as the basis of strain differentiation, reflecting the genetic source of serological distinctions.

First genome-based studies on SCMV relied on coat protein sequences being elucidated from viral cDNA library clones (Gough *et al*, 1987). The SCMV strain used in the construction of a cDNA library by those authors was the Johnson grass strain (SCMV-JG). The elucidated coat protein gene sequence is available in the GenBank sequence database under the accession number of A27631. The coat protein gene nucleotide sequence of another SCMV strain, the sugarcane strain of SCMV (SCMV- SC), was subsequently published by Frenkel *et al.* (1991).

Later, polymerase chain reaction (PCR) methodologies were used to amplify coat protein genes using primers designed from previously elucidated coat protein gene sequences (Smith & Van de Velde, 1994; Huckett & Botha, 1996; Yang & Mirkov, 1997; Gibbs & Mackenzie, 1997). The nucleotide sequences

of the PCR amplified products were then determined. Primers used for the PCR amplification of related potyviruses have also been published (Pappu *et al.*, 1993; Achon *et al.*, 1994). The principles of PCR are outlined in section 1.7.1.

Recently, sequence differences in PCR amplified regions of the coat protein gene have been analysed by restriction techniques, producing restriction fragment length polymorphisms (RFLPs), to differentiate between SCMV strains (Yang & Mirkov, 1997). This method is based on length differences of fragments produced after digestion with certain restriction endonucleases. Application of this method has facilitated the identification of unknown SCMV isolates without the need for sequence determination (Yang & Mirkov, 1997). RFLPs have also been widely used in the identification of strains of several other potyviruses such as the potato virus Y (PVY) (Glais *et al.*, 1996) and the passionfruit woodiness potyvirus (PWV) (Sokhandan *et al.*, 1997). In both cases the RFLP were based on sequences differences within the CP gene.

## 1.6 Reorganisation of the SCMV Subgroup

Until recently SCMV was considered to be a single potyvirus, consisting of a large number of strains which differed from each other in certain biological and antigenic properties (Shukla & Ward, 1994). Maize dwarf mosaic potyvirus (MDMV) was initially identified as a distinct virus of maize (Janson & Ellett, 1963) but was later considered to be a strain of SCMV because of the many common properties (Pirone, 1972). Through the use of cross-absorbed polyclonal antibodies directed towards the surface-located, virus-specific amino termini of the coat proteins, Shukla *et al.* (1989) revealed that 17

strains of SCMV (including MDMV) from Australia and the United States were not strains of one virus but in fact belonged to four distinct viruses.

It has also been shown that coat proteins of individual viruses of the potyvirus group are 38-71% identical but differ markedly in the length and sequence of their amino termini, whereas the coat proteins of strains of viruses are more than 90% identical and have amino-terminal sequences that are very similar (Frenkel *et al.*, 1991). As a result McKern *et al.* (1991) confirmed previous findings (Shukla *et al.*, 1989) that the viruses previously identified as strains of SCMV, are in fact distinct viruses. The four distinct viruses have subsequently been renamed according to the perennial host which they infect namely Johnsongrass mosaic virus (JGMV), maize dwarf mosaic virus (MDMV), sugarcane mosaic virus (SCMV )and sorghum mosaic virus (SrMV). The grouping of virus strains in the SCMV subgroup are as follows:

JGMV: SCMV-JG, MDMV-O, MDMV-KS1

MDMV: MDMV-A, D, E, F

SCMV: SCMV-A, B, D, E, SC, BC, ISIS, Brisbane, Bundaberg, MDMV-B

SrMV: SCMV-H, I, M (Shukla *et al.*, 1992).

This classification of the SCMV complex has now been confirmed by several biological, cytopathological and biochemical parameters including: reactions on differential sorghum and oat cultivars (Tosic *et al*, 1990), cytoplasmic cylindrical inclusion morphology (Lesemann *et al.*, 1992), cell-free translation of RNAs (Berger *et al.*, 1989), 3' noncoding nucleotide sequences (Frenkel *et al.*, 1989) and molecular hybridisation with probes corresponding to the 3'

noncoding regions (Frenkel *et al.*, 1992).

Recently, phylogenetic analysis using CP gene sequence data of several strains of the viruses in the SCMV subgroup (Yang & Mirkov, 1997) confirmed the reclassification of the SCMV strains to distinct viruses. Distinct clustering of the strains of each virus is shown in Figure 1.4.

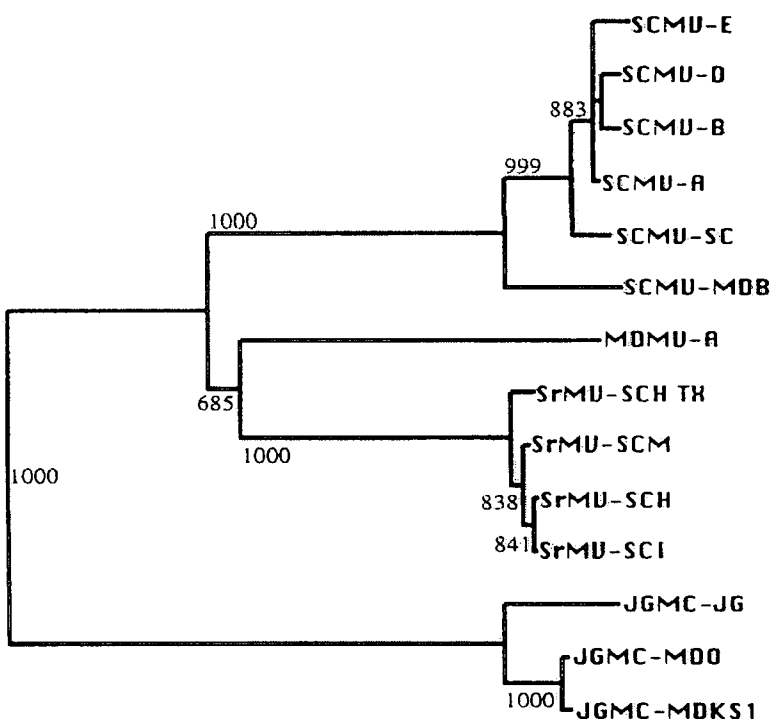



**Figure 1.4.** Phylogenetic relationships of coat protein gene sequences of several strains of the SCMV subgroup. The neighbor-joining method with 1000 bootstraps and the Clustal W parameter was used to construct the phylogenetic tree. The number alongside each node indicates the number of bootstrap replicates.
(Taken from Yang & Mirkov, 1997).

However, the reorganisation of the SCMV subgroup is based on comparisons of the former MDMV/SCMV strains only from Australia and the United States. A number of American SCMV strains namely C, F, G, K and L, have not yet

been investigated as well as MDMV/SCMV strains occurring in other countries. By using the above criteria, it would be possible to determine whether the SCMV strains not investigated belong to one of these four viruses or represent new potyviruses in the SCMV subgroup (Shukla *et al.*, 1994).

## 1.7 General Principles of Techniques Used in Strain Identification

### 1.7.1 Polymerase Chain Reaction (PCR)

PCR is a sensitive method of amplifying a segment of DNA or other double-stranded nucleic acid *in vitro* in the presence of a thermostable DNA polymerase, free deoxynucleotide triphosphates (dNTPs), a pair of oligonucleotides complementary to the segment of DNA (primers) and magnesium chloride. DNA polymerases carry out the synthesis of a complementary strand of DNA in the 5' to 3' direction using a single-stranded template, but starting from the double-stranded state. PCR employs two primers, each complementary to opposite strands of the region of DNA. The primers are arranged so that each primer extension reaction directs the synthesis of DNA towards the other. Synthesis is achieved by heating the DNA to separate the strands and cooling to allow the primers to anneal to their complementary sequences and subsequent extension. The heating and cooling cycles are repeated to accumulate DNA exponentially until one of the reaction components is depleted (Taylor, 1991).

Initially a fresh aliquot of the Klenow fragment of *E.coli* DNA polymerase I was added during each cycle because the subsequent denaturation step inactivated the enzyme. Discovery of a thermostable DNA polymerase, *Taq*

DNA polymerase, from *Thermus aquaticus* (Saiki *et al.*, 1988) eliminated the need for replenishing the enzyme and facilitated automation of the thermal cycling portion of the procedure. Automation of this process is achieved by using a thermal cycler which has programmable incubation blocks where temperatures, incubation times and cycle numbers can be manipulated as desired (Taylor, 1991). *Taq* polymerase also permitted the use of higher temperatures for annealing and extension, which improved the stringency of primer-template hybridisation and thus the specificity of the products. This has also served to increase the yield of the desired product (Kramer & Coen, 1995).

Despite its apparent simplicity, PCR amplification is influenced by several important variables. These include the magnesium concentration, dNTP concentration, primer design, template DNA, *Taq* DNA polymerase, other components and the cycling temperatures. Each factor is dealt with separately in the following sections.

a) *Magnesium Concentration*

Determining the optimum magnesium ($MgCl_2$) concentration can have an enormous influence on PCR success and is a crucial factor affecting the performance of *Taq* DNA polymerase since the enzyme requires free $Mg^{2+}$ ions for activity (Kocher & Wilson, 1991). Reaction components, including template DNA, chelating agents present in the sample, dNTPs and proteins, all affect the amount of free magnesium. In the absence of adequate magnesium, *Taq* DNA polymerase is inactive. Conversely, excess free magnesium reduces the enzyme fidelity and may increase the level of non-specific amplification. Magnesium and other salts also influence the

annealing of primers to the template DNA. For these reasons it is important to determine empirically the optimal $MgCl_2$ concentration for each reaction (Anon., 1996b).

b) *Deoxynucleotide Triphosphate (dNTP) Concentration*

Increasing the concentration of dNTPs in the reaction in an effort to increase the efficiency of PCR has deleterious effects on the amplification since dNTPs chelate magnesium and thus inhibit the activity of *Taq* DNA polymerase. Moreover, high dNTP concentrations increase the error rate of the polymerase (Gelfand, 1989).

c) *Primer Design*

This is the factor that is least predictable and most difficult to troubleshoot. An optimal primer set should hybridize efficiently to the sequence of interest with negligible hybridisation to other sequences present in the sample. PCR primers generally range in length from 15 - 30 bases and are designed to flank the region of interest. Longer primers will not help increase specificity significantly. The distance between the primers is flexible, up to 10kb. However, there is a considerable drop in synthesis efficiency with distances greater than 3kb (Jeffreys *et al.*, 1988).

Primers should ideally have a GC content of 40-60% or similar to that of the template. Primer sequences that contain stretches of polypurines or polypyrimidines should be avoided since they are likely to create internal secondary structures which affect extension. Base-pair mismatches may be intentionally or unavoidably created; therefore it is best to have mismatches at the 5' end of the primer. The closer the mismatch is to the 3' end, the more likely it is to prevent extension (Anon., 1996b).

Primer-dimers are PCR amplification artifacts which form when the 3' end of one primer anneals to the 3' end of the other primer, and polymerase then extends each primer to the end of the other. The ensuing product can compete very effectively against the PCR product of interest. Primer-dimers can best be avoided by using primers without complementarity, especially in their 3' ends (Kramer & Coen, 1995).

d) *Template DNA*

The main factors regarding template DNA are purity and quantity. Various contaminants found in DNA preparations such as sodium dodecyl sulphate (SDS) and sodium acetate, can decrease the efficiency of PCR (Gelfand, 1989). The amount of template must be sufficient to allow detection of the target sequence. However, the amount of target sequence versus nonspecific sequences can have a major effect on the yield of nonspecific products (Kramer & Coen, 1995).

e) *Taq DNA Polymerase*

Addition of more enzyme does not significantly increase product yield but may increase the yield of non-specific PCR product at the expense of the desired product. Excessive amounts of enzyme coupled with long extension times increases the likelihood of generating artefacts associated with the intrinsic 5'-3' exonuclease activity of *Taq* DNA polymerase, resulting in smearing on agarose gels (Anon., 1996b).

A very important property of *Taq* DNA polymerase is its error rate, which was initially measured at $2x10^{-4}$ nucleotides/cycle (Saiki *et al.*, 1988). The purified enzyme lacks the 3'-5' exonuclease (proof-reading) activity which leads to its high error rate. If PCR products are cloned prior to sequencing, this high

error rate will be encountered. Direct sequencing of the PCR products reduces the error rate to 0.3% since fewer amplification cycles are used (Kocher & Wilson, 1991).

Another important property of *Taq* DNA polymerase is its propensity for adding a non-template specified adenosine to the 3' end of DNA chains (Saiki *et al.*, 1988). This can be problematic if the PCR product is to be cloned by blunt-ended cloning since this would require "polishing" of the PCR product prior to cloning. However, this property of *Taq* DNA polymerase can be advantageous in cloning of the PCR products when T-tailed vectors are used (Anon., 1996b). Changing to other DNA polymerases, which have greater fidelity, may be useful to overcome the undesirable properties of *Taq* DNA polymerase.

f) *Other Components*

Additives that promote polymerase stability and processivity or increase hybridisation stringency and reduce nonspecific primer-template interactions, especially prior to the critical first cycle, can greatly improve sensitivity, specificity and yield. The addition of non-ionic detergents such as Triton X-100 or Tween-20 to PCR buffers helps to maintain full activity of *Taq* DNA polymerase during thermal cycling. β-Mercaptoethanol or dithiothreitol is added to help stabilize proteins during thermal cycling. Various carrier proteins can also be added but are not essential (Kocher & Wilson, 1991).

g) *Thermal Cycling Conditions*

Each step in the cycle requires a minimal amount of time to be effective, while too much time can be both wasteful and deleterious to the DNA polymerase (Kramer & Coen, 1995).

Denaturation

It is essential that complete strand separation occurs during the denaturation step. The standard 60 seconds used ensures that the temperature reaches 94°C. Higher denaturation temperatures may be necessary for templates with very high GC contents but this affects the activity of *Taq* DNA polymerase.

Annealing

It is critical that the primers anneal stably to the template. Primers with GC contents of <50% require low annealing temperatures, while primers with high GC contents require higher annealing temperatures. The higher temperatures minimize nonspecific primer annealing, increasing the amount specific product produced and reducing the amount of primer-dimer formation (Anon., 1996b). The time required for this step is minimal since the primers are in such excess that the annealing reaction occurs very quickly.

Extension

The extension temperature of 72°C is close to the optimal temperature for *Taq* polymerase (~75°C), yet prevents the primers from disengaging. Primer extension occurs during annealing because *Taq* is partially active at 55°C and even lower temperatures (Gelfand, 1989). The duration of this step depends on the length of the sequence to be amplified.

### 1.7.2 Reverse Transcription PCR (RT-PCR)

Polymerases used in PCR require a double-stranded template and thus limit the application of PCR. In order to apply PCR methodology to the study of RNA, the RNA sample must first be reverse transcribed to produce a double-stranded RNA-DNA hybrid or copy DNA (cDNA) to provide the necessary DNA

template for the polymerase. Reverse transcriptase enzymes such as those from Avian myeloblastosis virus (AMV) or Moloney murine leukaemia virus (MuLV) are generally used to produce cDNA from RNA directed by a oligonucleotide primer. The RNA-DNA hybrid produced is then used as a template in a PCR reaction. Recent developments have led to the use of optimised buffers for the reverse transcription and subsequent PCR to be performed in a single tube (Anon., 1996b). The same critical factors for the success of PCR apply to the RT-PCR technique.

### 1.7.3 Restriction Fragment Length Polymorphism (RFLP) Analysis

The use of PCR has made it possible to amplify a region of DNA which can then be exposed to restriction enzymes. The amplified and cleaved DNA is visualised directly on a gel after electrophoresis. The cleavage patterns derived from certain restriction enzymes can therefore be used to distinguish strains or sub-strains from one another at the nucleotide sequence level. A prerequisite of PCR-RFLP analysis is knowledge of the nucleotide sequence surrounding the restriction site (Ivinson & Taylor, 1991).

### 1.7.4 DNA Sequencing

There are two methods of DNA sequencing *in vitro*: an enzymatic or dideoxy chain termination method developed by Sanger *et al.* (1977) and a chemical degradation method developed by Maxam & Gilbert (1977). Of the two methods, the Sanger dideoxy chain termination method is the most popular. The Sanger method is based on the fact that dideoxynucleotide triphosphates (ddNTPs) are incorporated into a replicating DNA strand in the same way as

conventional dNTPs. The difference is that ddNTPs lack the 3' hydroxyl group necessary for chain elongation. As a result when ddNTPs are incorporated into the new DNA strand, the missing hydroxyl group prevents the formation of a phosphodiester bond with a successive dNTP and chain elongation is terminated at that position. Synthesis is initiated at the position where an oligonucleotide primer anneals to the template DNA and each chain is terminated at a specific base (Griffin & Griffin, 1993).

Until recently, DNA sequencing was performed using radiolabelling of extension products and autoradiography as the method of detection. $^{32}P$ was at that time the label of choice, since $^{32}P$-labelled dNTPs were readily available and are incorporated easily by polymerases (Heiner & Hunkapiller, 1989). However, the strong β particles emitted by $^{32}P$ created two problems. Firstly, scattering resulted in larger and more diffuse bands on the autoradiograph than the bands of DNA in the gel (Sambrook *et al.*, 1989). This affected the ability to read a sequence correctly and reduced the number of nucleotides that could be read from a single gel. The second problem was the radiolysis of DNA in the sequencing reaction caused by the decay of $^{32}P$ (Sambrook *et al.*, 1989). $^{32}P$-labelled sequencing reactions therefore could only be stored for a maximum of 2 days before degradation of the DNA rendered the sequencing gels indecipherable. Now $^{35}S$ is available as an alternative label and can provide significantly improved resolution due to less scattering of the weaker β particles (Biggin *et al.*, 1983). This allows unambiguous determination of several hundred nucleotides of DNA sequence from a single reaction set (Sambrook *et al.*, 1989). However, the use of $^{35}S$ requires significantly longer

exposure times than that of $^{32}P$ (Heiner & Hunkapiller, 1989).

More recently fluorescent dyes have become an alternative to radiolabels and have been used as part of automated systems using laser excitation and direct optical detection. Four dyes with different emission frequencies can be used as labels, one for each reaction. The products of all four reactions are then combined and electrophoresed on one lane of a gel. The fluorescently labelled bands are detected in real time as they migrate past a detector near the bottom of the gel. The signal spectrum is then analysed to determine which dye (i.e. which base) the bands represent. Currently, dye-labelled sequencing is available only for dideoxy reactions. These dyes may be attached to the 5' end of the primer or the terminating nucleotide analogue (Heiner & Hunkapiller, 1989).

### 1.7.5 International Sequence Database Comparisons

DNA sequence databases are linked both to the protein sequence databases which specify the translated products and to genetic maps which describe the organisation of the genetic functions in the genome (Bishop, 1989). There are three major DNA sequence databases:

- The DNA Data Bank of Japan (DDBJ)
- the European Molecular Biology Laboratory Nucleotide Sequence Data Library, (EMBL), and
- the GenBank Genetic Sequence Data Bank (GenBank)

Improvements in database accessibility have been made through the distribution at quarterly intervals of entire databases captured on optical

compact disks (CD-ROM). Using the appropriate software, this permits fast database searches on Personal Computers (Bishop, 1989). Recently, direct database searches of the most current sequence information available in sequence databases have been made possible through electronic mail and the Internet. The National Centre for Biotechnology Information (NCBI) contains a non-redundant (nr) database on its web site which combines sequence data from several sources, yielding a collection of nearly all available sequences.

In order to search these databases, a search program is required. Database searching programs can either be rigorous or can take short cuts in an attempt to speed up the alignment procedure (Bishop, 1989). Rigorous search programs employ the Smith-Waterman algorithm which compares the query sequence with each sequence in the database (target), while programs such as the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) and FASTA use heuristics to reduce the time required for alignment of the sequences (Brenner, 1998). A statistically computed score is assigned to each alignment and the best scores are reported to the user (Altschul, 1998). Variations of the BLAST and FASTA programs are available for different combinations of queries and database types (Brenner, 1998).

## 1.8 Importance of Viral Strain Identification

Viral strain identification has important implications for the development of diagnostic techniques for the detection of viral infections as well as in determining mechanisms for increased disease resistance. These are discussed in the following sections.

### 1.8.1 SCMV Diagnosis

A sensitive and reliable method for the diagnosis of SCMV infection is important for the safe movement and quarantine of sugarcane germplasm within and between countries (Smith & van der Velde, 1994). The use of RT-PCR (see section 1.7.2) has significantly enhanced the sensitivity of detection of viral infections and has been widely used in SCMV detection (Smith & van der Velde, 1994; Huckett & Botha, 1996; Yang and Mirkov, 1997; Gibbs & Mackenzie, 1997). The sensitivity offered by RT-PCR is especially useful in detecting infecting viruses since visual symptoms of viral infection can take some time to appear or may not be present at all (Cronje *et al.*, 1994). However, the application of RT-PCR in the development of a diagnostic test requires prior knowledge of the prevalent viral strains for suitable primers to be designed and applied. Knowledge of the prevalent strains in the South African sugar industry could lead to the development of diagnostic tests that make use of nucleotide sequence differences or similarities that exist between the different SCMV strains. This would facilitate the development of suitable diagnostic tests for specific or broad range strain diagnosis.

### 1.8.2 Viral Coat Protein-Mediated Resistance

Viral coat protein (CP)-mediated resistance is the expression of a viral CP gene in transgenic plants to achieve resistance to the virus from which the gene was derived (Sturtevant & Beachy, 1993). CP-mediated protection is manifested as a delay in or absence of the development of disease symptoms with a concurrent decrease in virus accumulation (Stark & Beachy, 1989). It also causes a reduction in replication and systemic spread since virus levels

are generally decreased in systemic leaves. Stark and Beachy (1989) found that levels of protection offered by CP-mediated resistance in potyviruses is greater when the CP gene used shares high sequence homology to the challenging virus. Identification of the prevalent SCMV strain(s) in the South African industry would therefore assist in determining the most applicable CP gene for use in genetically engineering resistance in local varieties of sugarcane to SCMV infection.

## 1.9 Aims of the Study

A need exists in the South African sugar industry for the development of a suitable diagnostic test to detect the presence of SCMV infection as well as the production of sugarcane containing inherent resistance through genetic engineering. However, prior knowledge of the prevalent strains of SCMV within the industry is strategically important to both these programmes. Previous work on strain identification in the industry has been based on symptomatological and serological methods, which have proven to be somewhat inconsistent and unreliable. The approach used in this study makes use of a genome-based method which offers a more sensitive means of differentiating between strains of viruses. The method makes use of differences in the nucleotide sequences which encode viral coat protein characteristics.

The primary aim of this study was to identify the strain(s) of SCMV present in current commercial sugarcane varieties within the South African sugar industry by comparative analysis of the nucleotide and inferred amino acid

sequences of the N-terminal region of the CP gene of several isolates of SCMV from various sugarcane growing regions and varieties. The following specific objectives were set in order to achieve this:

- select eight regions and two varieties (where possible) from each region representative of the industry
- isolate the N-terminal region of the CP gene from all samples
- determine the nucleotide sequence of the first 360bp of the CP gene
- establish relatedness between isolates based on sequence similarity using computer software
- use database sequence comparisons of nucleotide sequences obtained for strain identification
- extend and confirm sequence-derived strain allocations with that obtained using RFLPs.

Accomplishment of the specific objectives will therefore permit the identification of the prevalent strain(s) present in the industry.

# CHAPTER 2

## MATERIALS AND METHODS

### 2.1 Materials

#### 2.1.1 Source of Plant Material

Infected sugarcane leaf samples were collected by the South African Sugar Association Experiment Station (SASEX) Pathology department from regions of infection throughout the industry. Figure 2.1 shows the regions and varieties sampled as well as their locations within South Africa. Infected leaf material was selected from a number of plants in each region on the basis of symptom severity. Leaf samples obtained were either used directly or placed in labeled plastic bags, sealed and stored at -80°C.

#### 2.1.2 Oligonucleotide Primers

A range of primers was used in several applications including both viral specific primers and universal primers. The specific primers were AGPOT2, BHf, CPCON-F, GSf, GSfINV and GSr (Table 2.1) while universal primers were Oligo (d)T, M13Reverse, SP6 Reverse and T7 (Table 2.2). Sequences of the primers are given in Tables 2.1 and 2.2 and the relative binding positions of the primers are shown in Figure 2.2. The primers were synthesised at the University of Natal Medical School, the University of Cape Town or Boehringer Mannheim (Germany). Working stocks of the primers were diluted with sterile water to a concentration of 6μM (used for sequencing) or 10μM (all other applications) and stored at -20°C.

| | Region | Variety | Cane Age |
|---|---|---|---|
| 1 | Pongola | 86F3396 (now released as N28)<br>N19<br>N25 | 5 months<br>4 months<br>8 months |
| 2 | Melmoth | N12 | 8 months |
| 3 | Glendale | N16<br>NCo376 | 7 months<br>6 months |
| 4 | Tongaat | N12<br>NCo376 | 8 months<br>8 months |
| 5 | Mount Edgecombe | NCo376 | 8 months |
| 6 | Eston | N12<br>NCo376 | 6 months<br>6 months |
| 7 | Ixopo | N12<br>NCo376 | 7 months<br>7 months |
| 8 | Port Shepstone | NCo376 | 12 months |

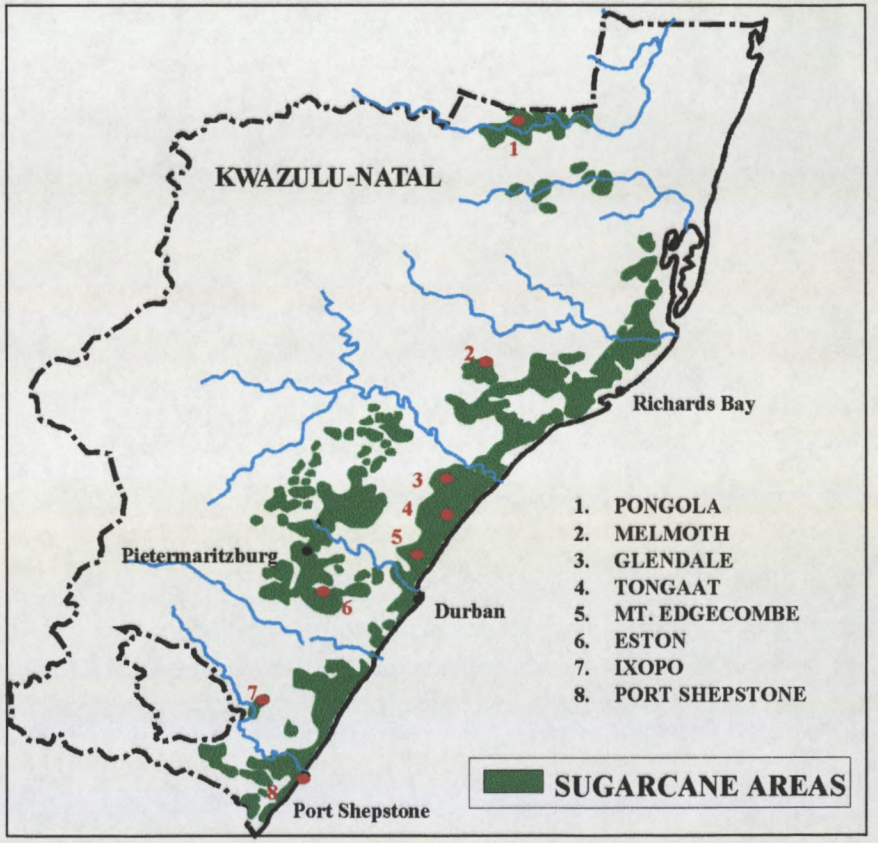



**Figure 2.1** Geographic and varietal sources of SCMV-infected sugarcane sampled for viral genome analysis.

**Table 2.1**. Viral-specific primers used in various applications.

| Primer Name | Sequence | Length (Bases) | Position of Binding* | Reference |
|---|---|---|---|---|
| AGPOT2 Forward | 5'–GGI-AA(CT)-AA(CT)-AG(CT)-GGI-C(AG)A-CC-3' | 20 | ±560bp upstream of the CP start | Gibbs & Mackenzie, 1997 |
| BHf Forward | 5'-AGC-TCC-ATA-TAT-TGC-AGA-AAC-AGC-3' | 24 | 144bp upstream of the CP start | Huckett & Botha, 1996 |
| CPCON-F Forward | 5'-TTA-AGA-TAT-CGG-ATC-CAT-GGC-TGG-AAC-AGT-CGA-TG-3' | 35 | CP start | Goodman *et al.*, 1997 |
| GSf Forward | 5'-ACA-CAA-GAG-CAA-CCA-GAG-AGG-3' | 21 | 381bp downstream of the CP start | Smith & van der Velde, 1994 |
| GSfINV Reverse | 5'–CCT-CTC-TGG-TTG-CTC-TTG-TGT-3' | 21 | 381bp downstream of the CP start | Smith & van der Velde, 1994 |
| GSr Reverse | 5'-AGT-CAA-AGG-CAT-ACC-GCG-CTA-3' | 21 | 220bp upstream of the CP termination | Smith & van der Velde, 1994 |
| SCMV F3 Forward | 5'-TTT-(CT)CA-CCA-AGC-TGG-AA-3' | 17 | CP start | Yang & Mirkov, 1997 |
| SCMV R3 Reverse | 5'-AGC-TGT-GTG-TCT-CTC-TGT-ATT-CTC-3' | 24 | 57bp upstream of the CP termination | Yang & Mirkov, 1997 |

* CP start and termination refer to the codons encoding the first and last amino acids of the CP.

I = Inosine

**Table 2.2**. Universal primers used in various applications.

| **Primer Name** | **Sequence** | **Length (Bases)** | **Position of Binding** | **Reference** |
|---|---|---|---|---|
| Oligo (d)T Reverse | T17(A, G, C) | 18 | Poly A tail | Sambrook *et al.*, 1989 |
| M13 Reverse | 5'-GTC-CTT-TGT-CGA-TAC-TG-3' | 17 | 87 downstream of the multiple cloning site | Messing, 1983 |
| SP6 Reverse | 5'–ATT-TAG-GTG-ACA-CTA-TAG-AAT-AC-3' | 20 | 99bp downstream of the multiple cloning site | Butler & Chamberlin, 1982 |
| T7 Forward | 5'–TAC-GAC-TCA-CTA-TAG-GGC-GA-3' | 18 | 73bp upstream of the multiple cloning site | Davanloo *et al.*, 1984 |

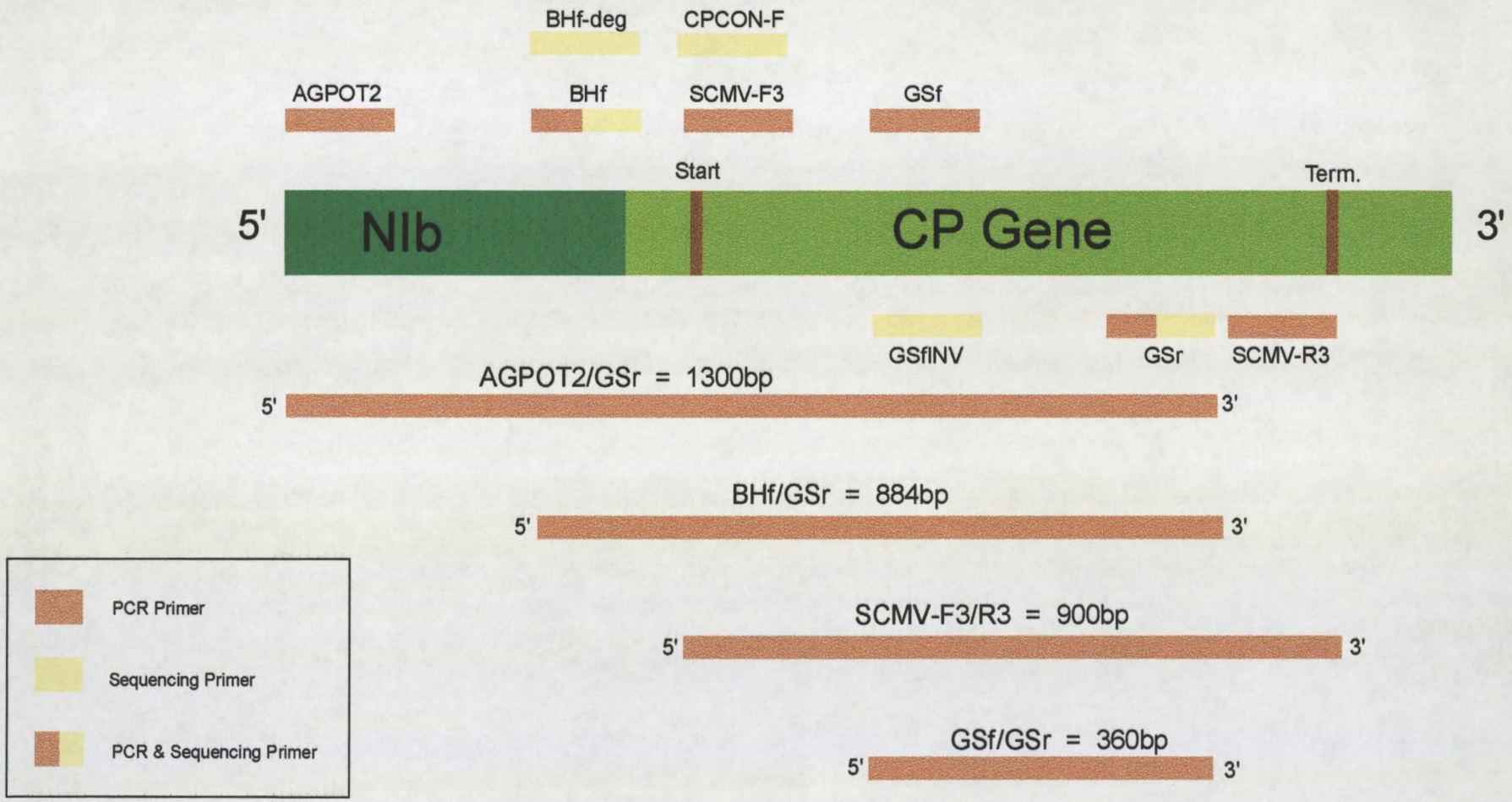



**Figure 2.2** Graphic representation of the SCMV coat protein (CP) gene and a portion of the polymerase (NIb) gene. This region is enlarged to show the position of binding of the SCMV-specific primers used in this study. Relative positions of the primers are shown in relation to the orientation of binding, where forward primers are positioned on top and reverse primers at the bottom. The estimated size of the PCR product generated using each primer-pair is also shown. Primers represented by dark blocks were used exclusively in PCR, while primers represented by the lighter shaded blocks were used exclusively in cycle-sequencing. Primers represented by two shaded blocks were used in both PCR and cycle-sequencing. Start = transcriptional start of the CP gene, Term. = transcriptional termination signal.

### 2.1.3 Cloning Vectors

The cloning vectors used in this study were the commercially available phagemids pCR-Script SK$^+$ (Stratagene), pCR2.1(Invitrogen) and pGEM-T Easy (Promega) which are 2900bp, 3890bp and 3018bp in size, respectively. The vector pCR-Script SK$^+$ is a blunt-ended cloning vector while the other two vectors are designed for the T-A method of "sticky-end" cloning. This makes use of the non-template dependent activity of *Taq* polymerase to add a single deoxyadenosine (A) to both 3' ends of amplified DNA. The DNA fragment is then ligated to a vector containing a single deoxythymidine (T) base at the 3' end of each complementary strand. The three vectors are represented in Figure 2.3.

### 2.1.4 Bacterial Strains

*Escherichia coli* strains INVαF', JM83 and JM109 were used as host cells in transformation procedures. The strains INVαF' and JM109 were commercially supplied as competent cells in conjunction with their respective vectors. Competent INVαF' cells were supplied with vector pCR2.1 and JM109 cells with pGEM-T Easy vector. JM83 cells were made competent in-house (see section 2.2.8). These strains are capable of complementation with the lacZ α–fragment encoded by the vectors to form active β-galactosidase used in blue/white selection of vector transformants with (white) and without (blue) the desired insert. All competent cells were stored at –80°C until required for transformation.

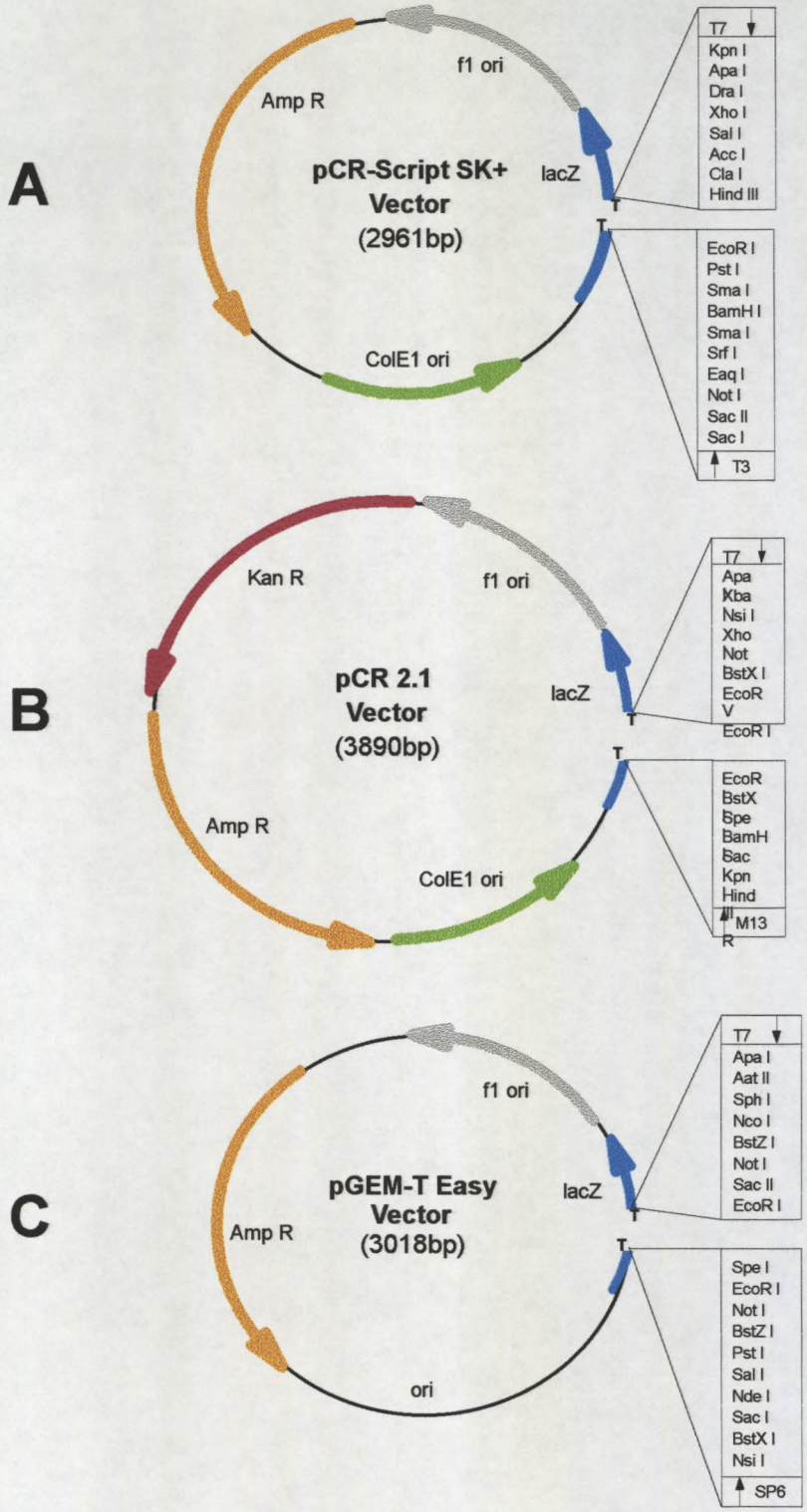



**Figure 2.3.** Diagrammatic representation of the three plasmid cloning vectors used for insertion of SCMV CP gene fragments. **A** = pCR-Script SK+ vector; **B** = pCR2.1 vector; **C** = pGEM-T Easy vector. All three vectors are capable of α-complementation with appropriate *E.coli* cells to facilitate the blue/white selection of recombinants containing the inserted CP fragment. Blue/white selection is accomplished by the insertional inactivation of the lacZ gene. In the presence of of an inducer (IPTG), expression of the β-galactosidase gene product results in the formation of blue colonies in the presence of a chromogenic substrate such as X-gal. An insertion of a DNA fragment into the lacZ gene disrupts the reading frame and prevents expression of the gene thus resulting in the formation of white colonies. White colonies are thus selected as putative recombinants. MCS is the mutliple cloning site which is flanked by vector specific priming sites, such as T7 and SP6 in the case of pGEM-T Easy vector, which allows to test by PCR for the presence of an insert. AmpR and KanR are the ampicillin and kanamycin resistance genes used for antibiotic selection of transformants.

2.1.5 Maintenance of Plasmids and Bacterial Strains

Bacterial cultures were maintained by several methods using Luria Bertani (LB) medium (1% w/v tryptone, 1% w/v NaCl, 0.5% w/v yeast extract powder, pH 7.5) as growth medium. LB-broth was used as liquid medium and 1.5% agar was added to produce the solid medium. The antibiotic ampicillin was added to the medium in a final concentration of 50μg/ml for selection when necessary. Long term bacterial cultures were preserved using glycerol (15% v/v) as cryoprotectant, frozen in liquid nitrogen and stored at -80°C. Stab cultures in LB-agar were sealed with Parafilm and stored at 4°C. Working stocks were maintained on LB-plates and stored at 4°C. These plates were re-streaked every three to four weeks and sealed with Parafilm to reduce the risk of contamination.

## 2.2 Methods

2.2.1 General Precautions Taken while Working with RNA

A number of applications in the work undertaken required the handling and use of intact RNA. RNA is easily degraded by ribonucleases (RNases) commonly found on hands, laboratory glassware, bacteria and moulds that may be present on air-borne dust particles and are extremely difficult to inactivate. Great care must be taken to avoid inadvertently introducing RNases into RNA preparations both during and after isolation. The following precautionary measures were taken to prevent the accidental contamination of RNA samples with RNases.

- Gloves were worn at all times and sterile techniques were employed

when handling any of the reagents used for RNA isolation or analysis.

- Chemicals for use in RNA isolation and analysis were reserved separately from other uses. Gloves were worn when handling these chemicals and only baked spatulas and untouched weigh boats were used.
- Whenever possible, sterile, disposable plasticware was used for handling RNA as these materials are generally RNase-free and did not require pre-treatment to inactivate RNases.
- Nondisposable glassware and plasticware were treated before use to eliminate any RNases. Glassware was baked at 200°C overnight and plasticware was thoroughly rinsed with 0.1N NaOH/1mM EDTA and then diethyl pyrocarbonate (DEPC)-treated water. Section 2.2.2 gives a description on the preparation of DEPC-treated water.
- Since autoclaving alone is not sufficient to inactivate RNases, solutions were made with DEPC-water and then autoclaved for 30min to remove any trace of DEPC.

### 2.2.2 Preparation of DEPC-Treated Water

DEPC-treated water was made by mixing 1% (v/v) DEPC with sterile deionised water in a fume cupboard. The mixture was stirred overnight using an electric magnetic stirrer. Residual DEPC was removed from the mixture by autoclaving thrice at 121°C for 20min.

### 2.2.3 RNA Extraction and Purification

Total RNA was extracted from freshly sampled infected leaves using a

commercial kit (RNeasy, Qiagen). Infected leaf tissue (minus midribs) was cut into approximately $2cm^2$ pieces and ground to a fine powder in a sterile mortar under liquid nitrogen. Two hundred milligrams of ground tissue was mixed thoroughly in a 1.5ml microcentrifuge tube containing 900µl RLT lysis buffer and 9µl of β-mercaptoethanol. Lysis of the plant cells was aided by incubating the mixture at 56°C for 5min and placing on ice immediately afterwards. The lysate was then separated from protein and cellular debris using a homogenizing spin column and centrifugation at 12000rpm for 2min. The lysate was mixed with 0.5 volumes of 95% ethanol and the nucleic acid bound to a nucleic acid-binding spin column by centrifuging for 30sec at 10000rpm. The column was washed with 700µl of RW1 wash buffer and centrifugation at 10000rpm for 30sec followed by two washes with 500µl RPE purifying buffer and the same centrifugation. Total RNA was then eluted from the column in two steps using 50µl followed by 30µl DEPC-treated, RNase free water into a sterile 1.5ml microcentrifuge tube. All the extraction steps requiring centrifugation were performed at room temperature. Once the RNA was eluted from the spin column it was kept on ice before storing at –80°C to prevent possible degradation.

### 2.2.4 Total RNA Yield Determination

The concentration of total RNA isolated was determined spectrophotometrically at a wavelength of 260nm using a Beckman DU7500 spectrophotometer. Readings were taken using aliquots of total RNA diluted fifty times in DEPC-treated water. DEPC-treated water was used as a blank to give a baseline absorbance. An absorbance of 1.0 at 260nm corresponds

to 40µg RNA per milliliter ($A_{260}$ = 1 = 40µg/ml) (Sambrook *et al.*, 1989). Concentration of RNA was calculated using the following formula:

$$\text{Concentration} = A_{260} \text{ X } 40\mu g/ml \text{ X Dilution Factor}$$

The ratio between the readings at 260nm and 280nm ($A_{260}/A_{280}$) was used to provide an estimate of the purity of the isolated RNA, where an ideal ratio is 2.0 (Sambrook *et al.*, 1989).

2.2.5 First Strand Copy DNA (cDNA) Synthesis

Viral RNA was reverse transcribed from total RNA using the Expand™ reverse transcriptase enzyme according to the supplied protocol (Boehringer Mannheim). Expand™ is an RNA-directed DNA polymerase enzyme which has reduced RNaseH activity (to prevent degradation of RNA in the DNA/RNA hybrid formed during reverse transcription). Since the extract contained total RNA (viral RNA as well as the various forms of plant RNA), hot-start reverse transcription was used. RNA and primer were first heated at 65°C for 10min before adding the rest of the reagents: 50mM Tris-HCl pH 8.3, 40mM KCl, 5mM $MgCl_2$, 0.5% Tween 20 (v/v), 2.5U Expand™ reverse transcriptase, 10mM dithiothreitol, 1mM deoxynucleoside triphosphate (dNTPs) mix and 1U of RNase inhibitor in a total volume of 20µl. Reverse transcription was then performed at 42°C for 30min. Primers GSr or Oligo (d)T were used to initiate reverse transcription of the coat protein gene region. First strand cDNA product was diluted 1:1 with sterile DEPC-treated water and was either used directly or stored at -20°C.

2.2.6 PCR Amplification

PCR was used for several applications and was performed using an Omnigene automated thermal cycler (Hybaid). Most PCR reactions were in 24µl reaction volumes containing various concentrations of template, 2µg acetylated bovine serum albumin (BSA), 10mM Tris-HCl pH 8.3, 1.5mM $MgCl_2$, 2.4µg gelatine, 100µM dNTPs, 0.1µM primers and 1U *Taq* DNA polymerase enzyme (Boehringer Mannheim). Reactions were overlaid with mineral oil and the DNA was amplified using the following thermal profile: 1 cycle of 94°C for 3min, 50°C for 5min, 72°C for 2min; 35 cycles of 94°C for 2min, 50°C for 2min, 72°C for 3min and 1 cycle of 72°C for 7min. Various modifications to the above-mentioned conditions were made for specific PCR applications and they are described below.

*a) Primary PCR amplification of viral sequences*

Several parameters were altered in the above reaction conditions to determine optimal PCR conditions using viral cDNA as a template. These included template concentration, $MgCl_2$ concentration, KCl concentration, reaction buffer pH, PCR additives, different DNA polymerase enzymes and a range of annealing temperatures. The primer-pairs used were AGPOT2/GSr and BHf/GSr.

*b) Semi-nested PCR amplification of viral sequences*

The semi-nested PCR amplification used purified PCR products as template and the GSf/GSr primer-pair. A slighter higher annealing temperature of 52°C was used with the thermal profile and reaction conditions as stated in section 2.2.6.

*c) Screening of putative positive bacterial transformants*

PCR reactions for the amplification of vector inserts from whole bacterial cells following transformation were performed in 14µl reaction volumes. The composition of the reaction was as mentioned in section 2.2.6 with the exclusion of BSA and using primer-pairs T7/M13 Reverse and T7/SP6 Reverse. Whole bacterial cells taken directly from a transformation plate, served as template DNA. Amplification was achieved using the following thermal profile: 1 cycle of 94°C for 1min, 10 cycles of 94°C for 25sec, 45°C for 30sec, 72°C for 1min, 30 cycles of 94°C for 15sec, 40°C for 30sec, 72°C for 1min and 1 cycle of 72°C for 2min.

*d) Amplification of 900bp fragment for RFLP analysis*

Amplification of an approximately 900bp fragment used for differentiation of SCMV strains by RFLP was performed in 24ul reaction volumes. The composition of the reaction was as mentioned in section 2.2.6 with the exception that the primer-pair SCMV-F3 / SCMV-R3 was used. A slighter higher annealing temperature of 52°C was used with the thermal profile and reaction conditions as stated in section 2.2.6.

2.2.7 Restriction Digests

Ten micrometer aliquots of PCR products were restricted with 5U of *Taq I* endonuclease at 65°C for 1h or with 5U of *Hinf I* endonuclease at 37°C for 1h. Reactions were performed in 15µl reaction volumes using the appropriate buffer supplied by the manufacturer (Boehringer Mannheim).

Restriction digest of plasmid DNA was carried out in 15µl reaction volumes containing 1µg of plasmid DNA, 10U of either *EcoR* I or *Not* I endonuclease and

the appropriate incubation buffer supplied by the manufacturer (Boehringer Mannheim). Reactions were incubated at 37°C for 2h and the restricted DNA visualised on 1.5% (w/v ) in TBE buffer [45mM Tris borate, 1mM EDTA, pH8.0] agarose gels.

Expected restriction profiles of PCR amplified products and plasmids was determined using the restriction enzyme search function of the DNASIS software programme and published sequence data.

### 2.2.8 Gel Electrophoresis and Visualisation of DNA

One percent agarose gels were prepared using half strength TBE buffer and ethidium bromide to a final concentration of 1μg/ml. PCR products were mixed with 2μl gel loading buffer (50% ficoll, 0.03% bromophenol blue, 0.03% xylene cyanol FF) and loaded on to the gel with a molecular weight standard (λ DNA *Eco*R I/*Hind* III digested or pBR322 DNA *Hae* III digested). The samples were electrophoresed at 6V/cm for 120min. DNA fragments were visualised on a Hoefer Mighty Bright UV transilluminator at 300nm and photographed using a Vilber Lourmat photographic system with adjustable focal length. Gel images were printed on light sensitive, high density paper using a Sony video graphic printer.

### 2.2.9 Purification of PCR Products

DNA fragments were excised from agarose gels using sterile scalpels. After removal of excess agarose, the gel pieces were treated in one of the following ways:

*a) Resolubilisation in TE buffer*

Excised gel pieces were placed in 500µl sterile TE buffer (10mM Tris-HCl, 1mM EDTA, pH 8.0) and either stored at -20°C or used directly in PCR after the DNA fragment was resolubilised by heating to 95°C for five minutes or until the agarose had melted completely.

*b) QIAquick gel extraction*

Agarose gel pieces containing the DNA fragment were weighed in sterile 1.5ml centrifuge tubes and 300µl QX1 solubilisation buffer was added for every 100mg of gel. The agarose was dissolved by incubating the mixture at 50°C for 10min or until completely dissolved. Vortexing the mixture every 2min also helped ensure that the agarose was completely dissolved. The mixture was then adding directly to a DNA-binding spin column and the DNA bound by centrifuging at 13000rpm for 60sec. Unbound agarose and buffer in the flow-through were discarded. Trace amounts of agarose were removed with an addition of 500µl QX1 buffer followed by centrifugation at 13000rpm for 60sec. Excess salt was removed from the column by adding 750µl PE wash buffer, leaving it at room temperature for 5min followed by centrifuging at 13000rpm for 60sec. Residual PE wash buffer was removed from the column after discarding the flow-through and centrifuging for an additional 60sec at 13000rpm. Membrane-bound DNA was eluted with 35µl of sterile distilled water. The column was left to stand at room temperature for 5min before centrifuging at 1300rpm for 60sec. DNA was used directly or stored at -20°C.

*c) GenElute gel extraction*

Gel pieces were placed directly in agarose-binding spin columns (GenElute, Supelco) pre-equilibrated with 100µl of TE buffer and centrifuged at

12000rpm for 10min to remove the agarose. The resultant DNA was further purified using the commercial kit, QIAquick PCR purification kit (Qiagen) as follows. Five volumes of PB binding buffer was mixed with the DNA solution and placed in a spin column. DNA was bound to the membrane in the nucleic acid-binding spin column by centrifuging at 12000rpm for 60sec and the resultant flow-through discarded. The membrane-bound DNA was washed with 750µl PE purifying buffer and centrifuging at 12000rpm for 60sec. Residual PE buffer was removed from the column by an additional centrifugation at 12000rpm for 60sec. DNA was eluted from the column using 35µl of sterile distilled water and stored at –20°C.

2.2.10 Insertion of PCR Fragments into Cloning Vectors

*a) Blunt-ended cloning*

One hundred nanograms of linearised pCR-Script $SK^+$ vector and 100ng of the coat protein gene fragment were ligated using a molar ratio of 1 : 3 (vector : insert). Ligations were carried out in 10µl reaction volumes containing 33mM Tris-HCl, 2.5mM $MgCl_2$, 0.5mM dithiothreitol and 0.5mM ATP and 10U T4 DNA ligase (Boehringer Mannheim). The reactions were incubated at 22°C for 16h.

*b) T-A cloning*

The molar ratio used in the cloning of coat protein gene fragments into the vectors pCR2.1 and pGEM-T Easy was 1: 3 (vector : insert) where the amount of vector used was 50ng. Vector and coat protein gene fragment were mixed in a 10µl reaction. The reaction using the vector pCR2.1 contained 60mM Tris-HCl (pH 7.5), 60mM $MgCl_2$, 50mM NaCl, 1mg/ml BSA, 70mM β-

mercaptoethanol, 1mM ATP 20mM dithiothreitol 10mM spermidine and 4U T4 DNA ligase (Invitrogen). Reaction components in the pGEM-T Easy ligation mixture were 300mM Tris-HCl (pH 7.8), 100mM $MgCl_2$, 100mM dithiothreitol, 10mM ATP and 3U T4 DNA ligase (Promega). Reactions were incubated overnight at 4°C (pGEM-T Easy) or 14°C (pCR2.1).

2.2.11 Preparation of Competent Cells

Five millilitres of LB broth was inoculated with *E.coli* JM 83 cells and incubated at 37°C overnight. This culture was then used to inoculate 200ml of LB-broth which was incubated at 37°C with agitation (aeration). The culture was incubated until the absorbance, measured at a wavelength of 650nm on the spectrophotometer, was 0.2-0.4 units. Cells were then cooled on ice for ten minutes and subsequently pelleted at 1000rpm at 4°C for 10min. The supernatant was discarded and the cell pellet was resuspended in 4ml transformation buffer (60mM $CaCl_2$, 10mM Pipes-HCl, pH 7.0, 15% v/v glycerol) and kept on ice. Another 36ml transformation buffer was added to the cells which were incubated on ice for a further 30min. Centrifugation at 1000rpm for 10min at 4°C precipitated the cells which were then resuspended in 6ml transformation buffer and kept on ice. The competent cell suspension was either used directly and/or divided into 200µl aliquots in sterile 1.5ml microcentrifuge tubes, flash frozen in liquid nitrogen and stored at -80°C.

2.2.12 Transformation of Bacterial Strains with Plasmid Vectors

Each *E.coli* strain was transformed according to a method that provides optimal transformation efficiency. Each strain is dealt with separately in the

following sections.

*(a) E.coli JM 83*

Five microlitres of the ligation mixture was used to transform 50µl of competent *E.coli* JM 83 cells by incubating on ice for 20min. Thereafter it was heat-shocked by incubating at 37°C for 5min and quenching on ice for 2min. Preheated (37°C) LB-broth (200µl) was added to the cells and incubated at 37°C for 30min. Between 50-100µl of the cell suspension was plated on to selective LB-plates and incubated overnight at 37°C. The plates were prepared by adding 50µg/ml ampicillin to the medium prior to setting and spreading 20µl of X-gal (10%w/v) and 20µl of IPTG (0.2M) with a sterile glass spreader over the entire surface of the plate.

*(b) E.coli INVαF'*

Fifty microlitres of INVαF' competent cells were thawed on ice and subsequently mixed with 2µl of 0.5M β-mercaptoethanol. Two microlitres of the ligation mixture was then added to the cells and incubated on ice for 30min. The cells were heat-shocked by incubating at 42°C for exactly 30sec and cooled on ice for 2min. SOC-broth (2% [w/v] tryptone, 0.5% [w/v] yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$-$6H_2O$, 20mM glucose) pre-equilibrated to room temperature (250µl) was added to the cells and the mixture incubated at 37°C for 1h with agitation. Aliquots of 50 and 200µl were plated on to selective LB-plates (as in section 2.2.12a) and incubated at 37°C overnight.

*(c) E.coli JM109*

Two microlitres of ligation mixture were mixed with 50µl of competent JM109 cells in pre-chilled polypropylene tubes (17x100mm) and incubated on ice for

20min. The cell suspension was heat-shocked by incubating at 42°C for exactly 45sec and quenching on ice for 2min. An aliquot of room temperature SOC-broth (950µl) was added to the cell suspension and incubated at 37°C for 1.5h with agitation. One hundred microlitres of the transformation mixture was plated on to duplicate selective LB-plates (as in section 2.2.12a) and incubated at 37°C overnight.

2.2.13 Screening for Transformants

Putative positive colonies (white) chosen after blue/white selection (see section 2.1.4) were tested directly for the presence of the inserted fragment by PCR analysis using primers which flank the multiple cloning site in the vector. A small amount of each bacterial colony, picked off with a sterile toothpick, served as template in a PCR reaction. Details of the PCR reaction conditions are given in section 2.2.6c.

2.2.14 Plasmid Isolation

PCR positive cells were used to inoculate LB-broth containing 50µg/ml ampicillin and grown at 37°C overnight with agitation. Cells were pelleted by centrifuging at 3000rpm for 10min. Plasmid isolation was done using the Nucleobond AX plasmid isolation kit (Macherey-Nagel). The cell pellet was resuspended in S1 suspension buffer (50mM Tris-HCl, 10mM EDTA, 100ug/ml RNase A, pH 8.0). Lysis of the cells was achieved by mixing the cell suspension with S2 lysis buffer (200mM NaOH and 1% SDS) and incubating at room temperature for 5min. S3 denaturing buffer (2.8M Kac, pH5.1) was added to the mixture and incubated on ice for 5min to denature the proteins.

Cellular debris was pelleted by centrifuging at 12000g for 40min. The supernatant was filtered before being loaded on to a DNA-binding cartridge pre-equilibrated with N2 equilibration buffer (100mM Tris, 15% ethanol, 900mM KCl, pH6.3). The cartridge was washed thrice with N3 wash buffer (100mM Tris, 15% ethanol, 1150mM KCl, pH6.3) and the bound DNA eluted twice with N5 elution buffer (100mM Tris, 15% ethanol, 1000mM KCl, pH8.5). Plasmid DNA was precipitated with room temperature isopropanol and centrifuged at 15000g for 30min. The DNA pellet was washed with 70% ethanol and centrifuged at 15000g for 10min. Residual ethanol was removed by air drying and the DNA resuspended in sterile water and stored at –20°C.

2.2.15 Sequence Determination

*a) Sequencing system used*

Nucleotide sequences were determined by cycle sequencing using multicolour fluorescent dye terminator chemistry. The four bases are represented by four different dyes and can be differentiated by their respective wavelengths. Extension products were resolved by electrophoretic separation in a single sequencing capillary filled with a pre-formulated liquid polymer, Performance Optimised Polymer-6 (POP-6) (Perkin-Elmer). POP-6 is specifically formulated for DNA sequencing and allows high resolution of large fragments in a short run time. Samples were automatically injected into the capillary and electrophoresed at 6V/cm for approximately 3h. The nucleotide sequence was then determined using an ABI Prism 310 Genetic Analyser.

*b) Direct Sequencing of Viral Sequences*

Purified PCR fragments were sequenced directly using the AmpliTaq$^R$ DNA

polymerase FS ready reaction kit (Perkin Elmer). A 15µl reaction contained: 0.8µM specific primer, 100ng PCR fragment and 6µl terminator premix (A-dye terminator, C-dye terminator, G-dye terminator, T-dye terminator, dITP, dATP, dCTP, dTTP, Tris-HCl [pH 9.0], $MgCl_2$, thermal stable pyrophosphatase and AmpliTaq DNA polymerase FS) (Perkin Elmer). The following thermal profile was used: 1 cycle of 95°C for 5min, 30 cycles of 95°C for 30sec, 40°C for 30sec, 60°C for 3min with a ramp of 2sec/°C. The extension products were purified by mixing the reactions with 55.5µl 70% ethanol containing 0.5mM $MgCl_2$. The mixture was vortexed briefly and the extension products were precipitated at room temperature for 10min. Extension products were pelleted by centrifuging at 12000rpm for 20min. Supernatant was aspirated and the pellet dried under vacuum in a vacuum drier (SpeedVac, Sorvall) at room temperature for 5min or until completely dry. Template suppression buffer (Perkin Elmer) (18µl) was added to the pellet before denaturing the extension products in a boiling waterbath for 2min followed by ice-quenching.

*c) Sequencing after cloning*

Plasmid DNA of positive clones was cycle sequenced using the BigDye Terminator kit (Perkin Elmer). A 15µl reaction contained: 3.2 pico moles specific primer, 1µg plasmid DNA and 6µl terminator premix (see section 2.2.15b). The same thermal profile was used as for direct sequencing (2.2.15b). After completion of the sequencing reaction the extension products were mixed thoroughly with 16µl of sterile deionised water and 64µl of 99% non-denatured ethanol and incubated at room temperature for 10min. Extension products were pelleted by centrifuging in a microcentrifuge at 12000rpm at room temperature for 15min. Supernatant was aspirated and

the pellet washed with 100µl of 70% and centrifuged at 10000rpm for 10min. The ethanol was decanted and the pellet dried in a SpeedVac vacuum drier at room temperature for 10min. Dried pellet was resuspended in 18µl of template suppression buffer and denatured in a boiling waterbath for 2min followed by ice-quenching. The nucleotide sequence was then determined as mentioned in section 2.2.15a.

2.2.16 Sequence Database Searches

The National Center for Biotechnology Information (NCBI) website non-redundant (nr) nucleotide and peptide databases were used to compare elucidated sequences of the cloned 1300bp CP fragment derived from all source samples (see Figure 2.1) with sequences contained within the databases. Sequence comparisons were performed using the Basic Local Alignment Search Tool (BLAST) program developed by Altschul *et al.* (1990). This algorithm uses certain parameters to compute the homology between a query sequence and target (database) sequence and assigns statistically significant scores to each alignment. The highest scores are reported to the user and any score above 80 signifies a significant homology.

Of the programs in the BLAST package, the BLASTN (Altschul *et al.*, 1990) program was used to compare nucleotide query sequences against nucleotide sequences in the NCBI nr nucleotide database. The BLASTX (Altschul *et al.*, 1997) program was used to compare the six-frame conceptual translation products of the nucleotide query sequences (both strands) against sequences in the NCBI nr protein database. An expect value of 1.0 was used for all

database searches for stringent alignments.

2.2.17 Sequence Analysis

Sequences were manually edited to resolve ambiguities using the ABI Prism Genetic Analyser software program, Sequence Navigator v1.0.1, 1991 (Perkin Elmer). For most sequences ambiguities were resolved by comparing the electropherograms and computed sequences of three separate sequence runs of the same clone. In the case of single sequence runs, ambiguities were resolved by checking the colour of the highest peak at the position of the ambiguity in the elecctropherogram and confirming the base assigned by the software. Sequence Navigator was also used to compare sequences to each other using the clustal alignment function. This alignment function uses the Clustal V algorithm which is based on the Needleman-Wunsch algorithm which assumes that an evolutionary relationship exists between the sequences being aligned. Ordered similarity scores are assigned to the sequences and this is subsequently used to perform pairwise alignments of the sequences.

Consensus sequences for each of the clones were obtained using the Contig Manager or Shotgun Fragment Matching function of the sequence analysis software program, DNASIS v2.1, 1991 (Hitachi). This function locates regions at the ends of sequences that overlap with other sequences, by checking both normal and complementary orientations. The length and minimum matching homology of the overlap determine the stringency of the alignment. For the generation of consensus sequences using this function, an overlap length of

30bp and 70% minimum matching homology was used. Consensus sequences obtained contained regions of matching bases interspersed with IUB codes representing unmatched characters. IUB codes are universally used codes for ambiguous bases, where for example, Y codes for pyrimidines cytosine (C) or thymine (T). Gaps in the sequences were represented by question marks.

Sequence relationships of the isolates were established using the Multiple Alignment function of DNASIS. This function uses the Higgins-Sharp algorithm which is based on an assumed evolutionary relationship between sequences. More closely related sequences are presumed to have fewer deletions, transitions, transversions and other mutations than more distantly related sequences. Regions of overlap between sequences are identified using a matrix of similarity scores based on the number of exactly matched residues minus a fixed gap penalty for every gap created and the sequences assigned in order in an initial phylogenetic tree. Sequences are then aligned in groups corresponding to the order of branching in the initial tree by finding the optimal midpoint of an alignment. The remainder of the alignment is found by recursively finding optimal midpoints on either side of the initial midpoint. Parameters such as gap penalty, which is the score assigned to each gap created and the K-tuple, which specifies the number of residues that must match in the alignment, determine the stringency of the alignment. Default values of 5 for the gap penalty and 4 for the K-tuple were used in all multiple alignments. The final dendrogram displayed showed the calculated matching percentages between sequences at each branch point.

Pairwise alignment of each sequence obtained with the five SCMV strains SCMV-A, -B, -D, -E and -SC was performed using the Database Homology Matching function of DNASIS. Nucleotide sequences of the five SCMV strains were obtained from GenBank under the following accession numbers:

U57354 - SCMV-A;

U57355 - SCMV-B;

U57356 - SCMV-D;

U57357 - SCMV-E;

D00948 - SCMV-SC.

This function is governed by the Lipman-Pearson FASTA algorithm in which homologous regions are matched by firstly shifting the query sequence relative to the target sequence and assigning a score to the highest number of consecutively matching residues. Secondly, the alignment is optimised to increase the number of matching residues by inserting gaps at a penalty. The scores are recalculated to reflect these modifications and displayed as a percentage.

# CHAPTER 3

## RESULTS

### 3.1 Total RNA from Infected Leaves: Yield and Purity

The SCMV genome consists of a single-stranded, positive sense RNA molecule. Since the CP gene, the gene of interest, is thus in RNA form, it was necessary to first isolate the SCMV RNA genome by extracting total RNA from infected sugarcane leaves. Total RNA i.e. plant and viral RNA, was extracted from leaf tissue sampled from all the geographical regions mentioned in Figure 2.1. After spin-column purification the RNA was analysed spectrophotometrically (see section 2.2.3). Examples of typical RNA spectra obtained are shown in Figure 3.1 with peaks at 260nm indicative of nucleic acids. These spectra represent the extremes of purity levels of RNA obtained, the lowest being 1.5 (A) and an ideal 2.1 (B). The purity ratios obtained for all the samples ranged from 1.5-2.7 (Table 3.1). Yields of RNA obtained for each region and the respective varieties are also shown in Table 3.1. These ranged between 0.086 and 1.33 μg/μl.

As the RNA yields and quality for all the varieties from each region were within acceptable limits, it was possible to proceed with preparing first strand cDNA from the total RNA obtained.

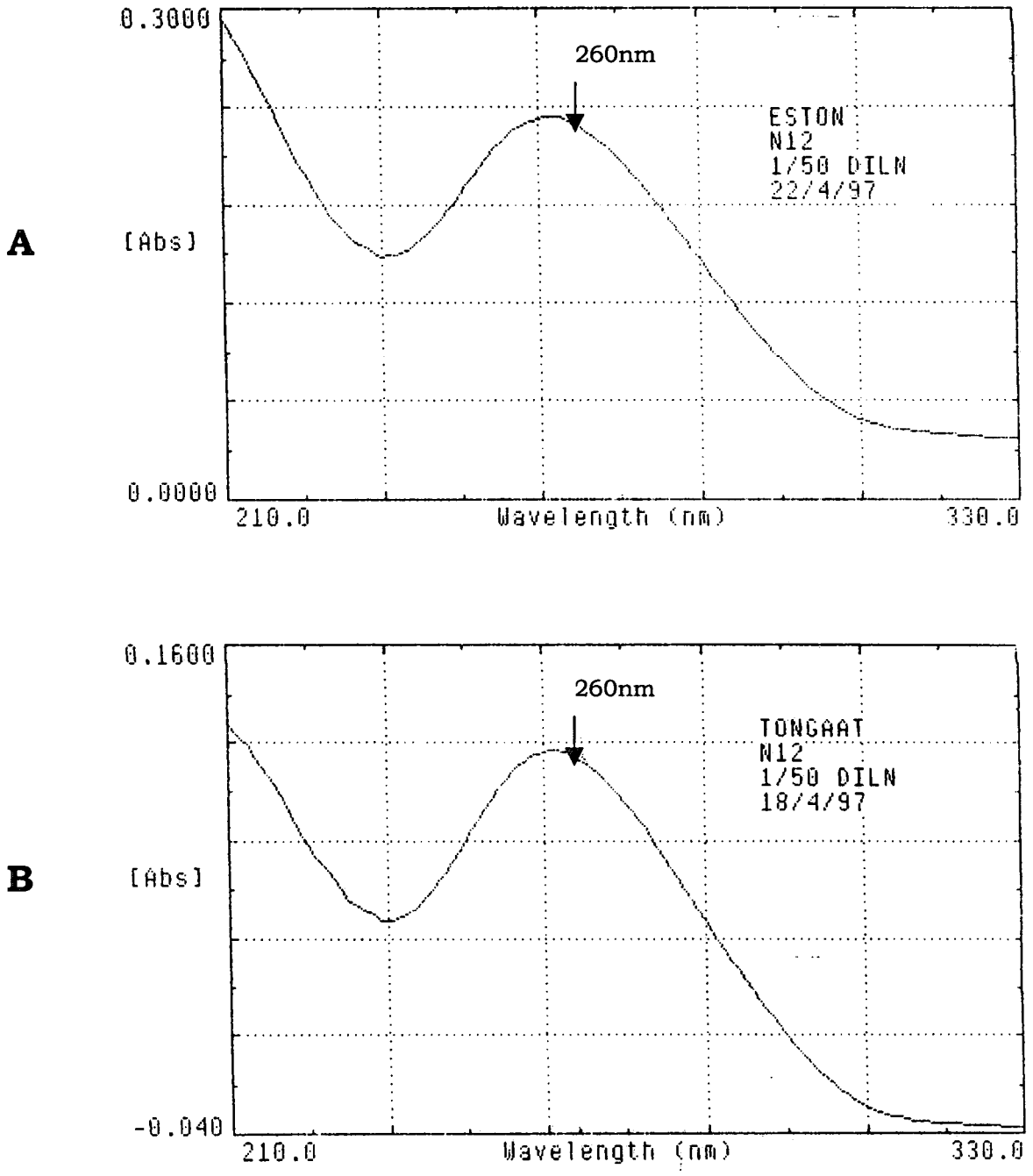



**Figure 3.1** Qualitative analysis of total RNA from SCMV infected sugarcane leaves. **A**. An example of a relatively impure RNA sample with a purity ratio of 1.5. **B**. An example of a pure RNA sample with a purity ratio of 2.1.

**Table 3.1.** Yield and purity of total RNA from infected leaf material from various regions and varieties as determined by spectrophotometric analysis.

| Region | Variety | Yield (µg/µl) | $A_{260}/A_{280}$ |
|---|---|---|---|
| Eston | N12<br>NCo376 | 0.47<br>0.28 | 1.5<br>1.8 |
| Glendale | N16<br>NCo376 | 0.68<br>0.63 | 1.9<br>1.7 |
| Ixopo | N12<br>NCo376 | 0.84<br>0.66 | 1.8<br>1.7 |
| Mount Edgecombe | NCo376 | 0.69 | 1.68 |
| Melmoth | N12 | 1.33 | 1.6 |
| Pongola | 86F3396 (N28)<br>N19<br>N25 | 0.41<br>0.37<br>0.69 | 2.1<br>1.8<br>1.65 |
| Port Shepstone | NCo376 | 0.086 | 2.7 |
| Tongaat | N12<br>NCo376 | 0.2<br>0.36 | 2.0<br>2.1 |

## 3.2 Primary Generation of the Coat Protein Gene Fragment

In order to PCR amplify the CP gene from the viral RNA genome, it was necessary to first reverse transcribe the RNA to first strand cDNA, since the process of PCR requires a double-stranded template for amplification. Purified total RNA was reverse transcribed using the Expand™ reverse transcriptase system (Boehringer Mannheim) described in section 2.2.4. First strand cDNA synthesis was initiated by either the specific primer, GSr or the universal primer, Oligo (d)T. The cDNA product was diluted in a 1:1 ratio with nuclease-free, sterile distilled water ($dH_2O$) and used as template in PCR. Two different sized PCR amplified fragments containing CP gene sequences were used in this study. Two primer pairs were used to generate the fragments: AGPOT2/GSr and BHf/GSr which amplified fragments of approximately 1400bp and 884bp respectively. A diagrammatic representation of the

position of the primers within the SCMV genome and the two fragments amplified is shown in Figure 2.2. Reaction conditions for the primary amplification of the 884bp fragment had been optimised previously (Huckett & Botha, 1996). The same conditions were used for the amplification of the 1400bp fragment with the exception that an annealing temperature of 50°C was used instead of 52°C.

Although cDNA template concentrations were not determined directly, a range of 0.5 - 8µl of cDNA product was used in PCR to assess the effect of template concentration using the reaction conditions stated in section 2.2.6. Results for amplification of the 884bp CP fragment are shown in Figure 3.2, while results for the amplification of the 1400bp CP fragment are shown in Figure 3.3. Amplification of a single fragment approximately 884bp was achieved across a wide range of cDNA concentration from all source samples while no amplification was evident in the no template control (Figure 3.2). This indicated that PCR conditions were optimal for specific amplification of the 884bp fragment regardless of the origin of the sample and confirmed the findings of Huckett & Botha (1996).

Similarly, the 1400bp CP fragment was amplified from all source samples while no amplification was achieved in the no template control. However, amplification of the 1400bp fragment appeared to be dependent on higher concentrations of cDNA than the 884bp fragment. Figure 3.3 shows that the best amplification occurs at the highest cDNA template used. This is

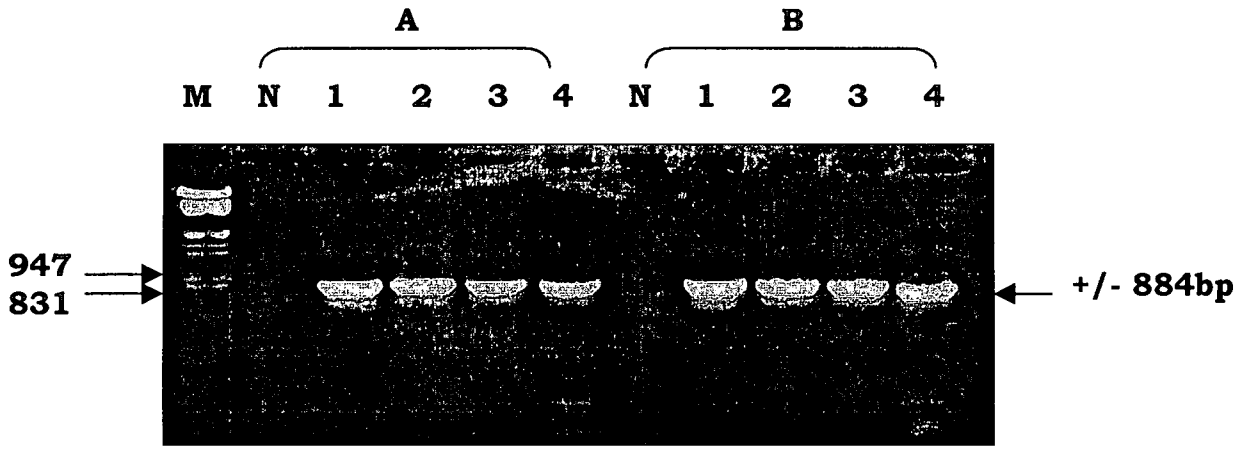



**Figure 3.2.** Effect of cDNA template concentration on PCR amplification of the 884bp SCMV CP fragment generated with primer pair BHf/GSr. M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested); N = no template control. Sources: A = Ixopo N12; B = Ixopo NCo376; Lanes 1 - 4 = 0.5; 1.0; 5.0 and 8.0µl of cDNA reaction product, diluted 1:1.

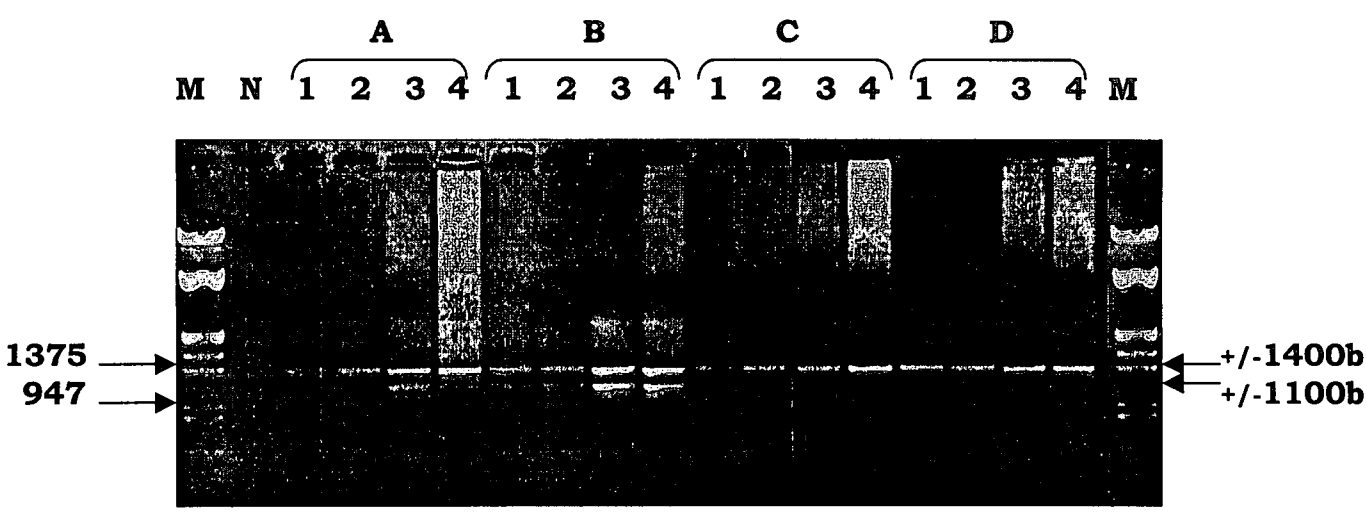



**Figure 3.3.** Effect of cDNA template concentration on PCR amplification of the1400bp SCMV CP fragment generated with primer pair AGPOT2/GSr from several sources. M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested); N = no template control; A = Glendale N16; B = Glendale NCo376; C = Ixopo N12; D = Ixopo NCo376; Lanes 1 - 4 = 0.5; 1.0; 5.0 and 8.0µl of cDNA reaction product, diluted 1:1.

especially evident in the case of the sample of variety N12 from Ixopo (C, lanes 1-4). Some of the samples (Eston N12, Eston NCo376, Glendale N16, Glendale NCo376, Tongaat N12 and Tongaat NCo376) co-amplified a second smaller fragment (approximately 1100bp) with the 1400bp CP fragment (Figure 3.3). Gibbs & Mackenzie (1997) also found that a second smaller fragment was co-amplified with the primary amplification product with other potyviruses (papaya ringspot potyvirus and watermelon mosaic 2 potyvirus) but not with SCMV. The authors found that the smaller fragment comprises the portion of the *Potyvirus* genome between that encoding motif II of the NIb gene and the N-terminus of the CP gene.

Both the 884bp and the 1400bp fragments were generated for all the samples. Each fragment band was excised from agarose gel and resolubilised in an aliquot of TE buffer or purified using a commercial kit (see section 2.2.9). This was then used as a template in a second PCR reaction to confirm the identity of the amplified fragments.

## 3.3. Confirmation of SCMV Identity of Amplified CP Fragments

Since the cDNA used in the PCR amplification of the 884bp and 1400bp CP fragments was derived from total RNA, confirmation of the viral origin of the two amplified CP fragments was a prerequisite to further characterization.

### 3.3.1 Semi-Nested PCR Amplification of CP Fragments

Verification that the primary PCR amplified product contains the desired sequence can be achieved by performing a second amplification using a set of

primers which bind internally to the target region. This approach is referred to as nested PCR. In this study, a variation of this method was used in which only one internal primer and the original downstream primer was used, resulting in a semi-nested PCR amplification. The primer pair used in the semi-nested PCR was GSf/GSr, designed to amplify a 360bp fragment of the CP gene. Relative binding positions to the SCMV genome of this primer pair are shown in Figure 2.2.

Amplified 884bp and 1400bp CP fragments from all the samples were excised from the agarose gels, resolubilised in TE buffer or gel purified and used as template in semi-nested PCR reactions. Using 0.5µl of the resolubilised or purified 884bp and 1400bp fragments and PCR reaction conditions described in section 2.2.6.b, a fragment of the expected size, 360bp, was produced. Figure 3.4 shows the confirmatory 360bp semi-nested fragment amplified from the 1400bp fragment from all the samples. The smearing in some of the lanes is probably a result of excess template in the reaction. Similar results were obtained for the 884bp fragment.

### 3.3.2. Re-amplification of CP Fragments

Another approach used to confirm the viral identity of the amplified 884bp and 1400bp fragments, was to re-amplify the PCR products for further characterization. This procedure had the additional advantage of preparatively purifying the CP gene fragment for later sequencing and cloning. Initially, re-amplification of the 884bp fragment was attempted using a template concentration range in the form of 0.5-8.0µl of resolubilised or gel

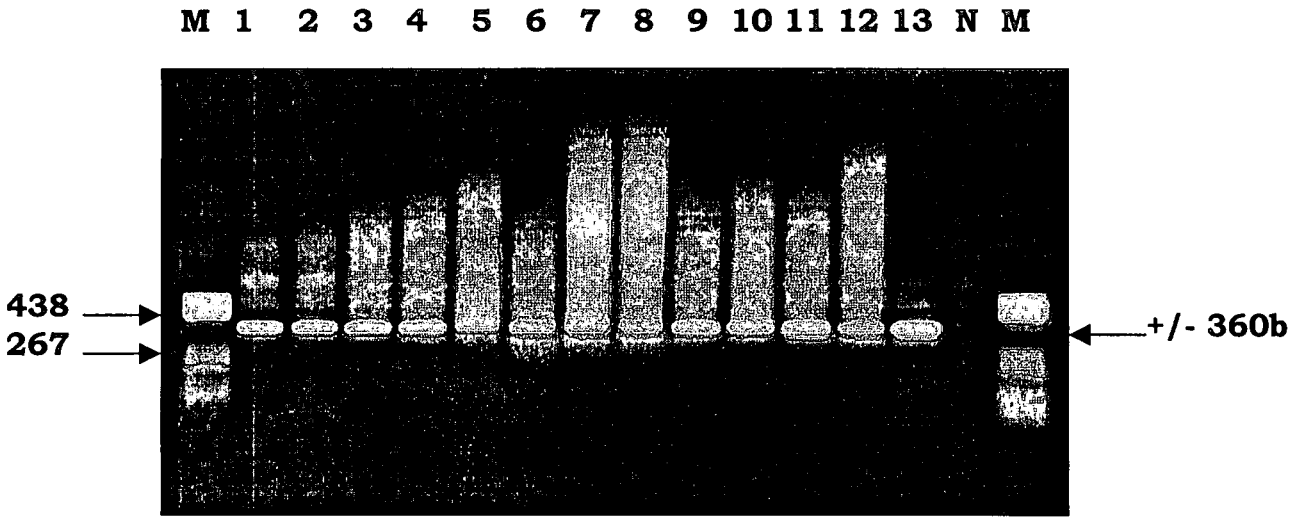



**Figure 3.4.** Confirmation of SCMV CP sequence identity by semi-nested PCR amplification of a 360bp fragment from the 1400bp fragment. Template samples represent all regions and their respective varieties using GSf/GSr primer pair. M = Molecular weight marker V (pBR322 DNA *Hae* III digested), N = No template control, Lanes 1-2 = Eston N12, NCo376; 3-4 = Glendale N16, NCo376; 5-6 = Ixopo N12, NCo376; 7 = Mount Edgecombe NCo376; 8 = Melmoth N12; 9-10 Pongola N19, N25; 11 = Port Shepstone NCo376; 12-13 = Tongaat N12, NCo376.

purified fragment solution and reactions conditions as described for primary amplification of the fragment. However, amplification was not achieved for all samples, indicating that these reaction conditions could not be applied consistently to all samples. Using the 884bp fragment as a preliminary model, various parameters were altered in an attempt to optimise the re-amplification conditions. These included: template concentration; $MgCl_2$ concentration and reaction buffer pH; annealing temperature and thermostable DNA polymerase enzymes. The effect of each of the parameters is discussed in the following sections. Standard PCR reaction conditions were: 10mM Tris-HCl; pH8.3, 1.5mM $MgCl_2$; 2.4µg gelatine; 1U *Taq* DNA polymerase (Boehringer Mannheim) in 24µl reaction volume and an annealing temperature of 50°C.

*a) Effect of template concentration*

Since the template concentrations used initially were relatively high for re-amplification, a dilution factor ranging between $10^{-1}$ and $10^{-3}$ of the resuspended 884bp fragment was tested using reaction conditions as described in section 2.2.6b. A typical example of the results of re-amplification of the 884bp fragment is shown in Figure 3.5. Dilution of the template enhanced re-amplification for some of the templates, for example variety NCo376 from Glendale and variety N12 from Ixopo (B, lanes 1 - 4 and C, lanes 1 - 4). In some cases the effect of further dilution of template was to allow amplification, but at low levels. A typical example of this is shown in Figure 3.6 where the template was diluted $10^{-5}$ fold.

However, the amount of re-amplified product was not sufficient for characterization in downstream applications such as sequencing and cloning indicating that re-amplification reaction conditions had to be optimised further.

*b) Effect of $MgCl_2$ and KCl concentrations and reaction buffer pH*

The amount of free $Mg^{2+}$ ions in a PCR reaction can affect the outcome of the amplification since the activity of *Taq* DNA polymerase is dependent on free $Mg^{2+}$ ions but is inhibited by excess $Mg^{2+}$ ions. Incorrect levels of KCl can also affect the activity of Taq DNA polymerase as well as alter the denaturation and annealing temperature of the DNA. The concentration range of $MgCl_2$ used in the standard PCR reaction (see section 2.2.6) is 1.5mM, pH is 8.3 and no KCl is present. A range of $MgCl_2$ concentrations between 1 and 3mM was tested

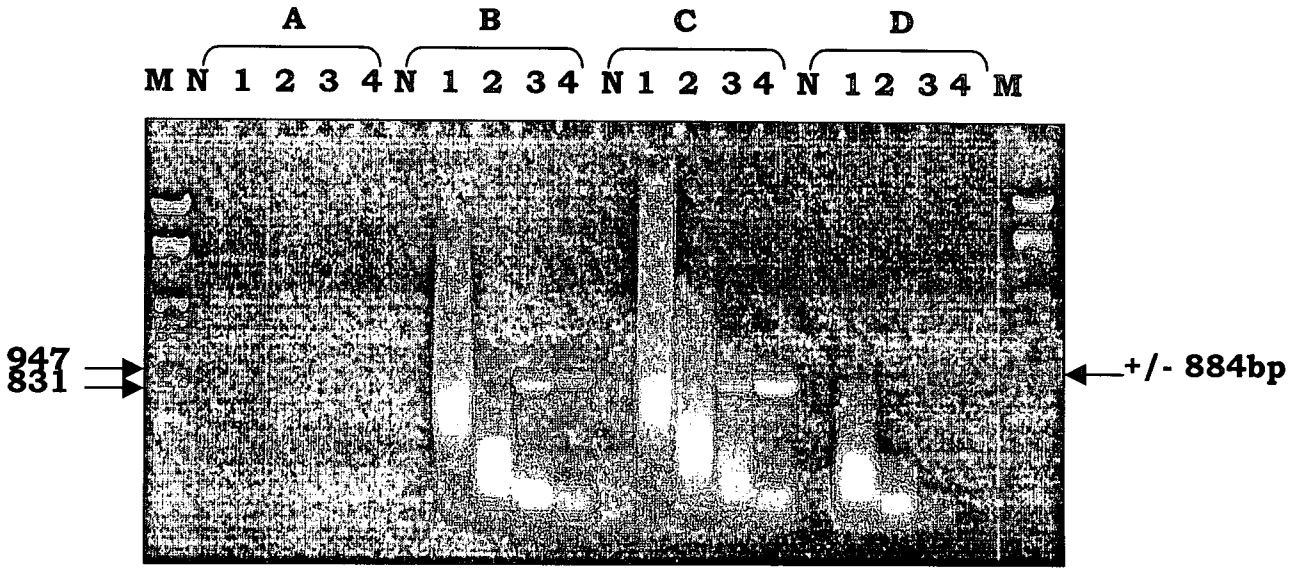



**Figure 3.5**. An example of the results of decreased template concentration for re-amplification of the 884bp fragment. A = Glendale N16; B = Glendale NCo376; C = Ixopo N12; D = Ixopo NCo376; M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested); N = no template control; 1 = undiluted; 2 = $10^{-1}$; 3 = $10^{-2}$; 4 = $10^{-3}$ dilution of resolubilised 884bp CP fragment.

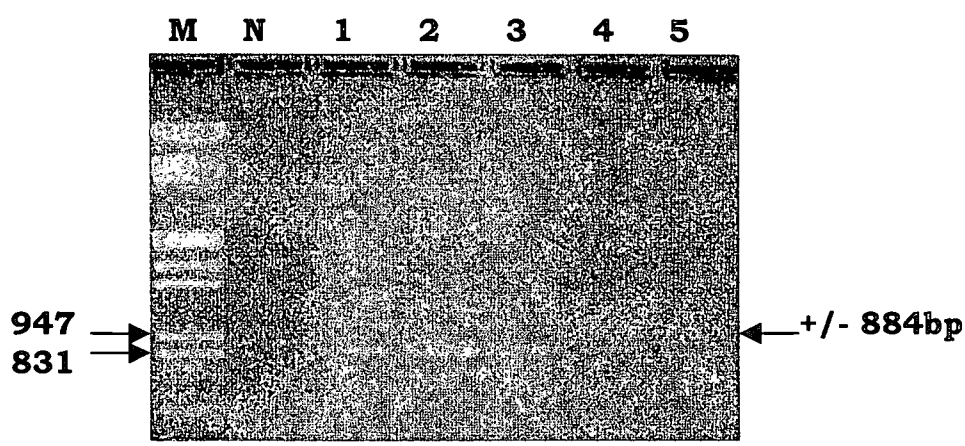



**Figure 3.6**. An example of increased dilution of template in re-amplification of the 884bp fragment derived from Tongaat NCo376. M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested); N = no template control; Lane 1 = $10^{-1}$; 2 = $10^{-2}$; 3 = $10^{-3}$; 4 = $10^{-4}$ and 5 = $10^{-5}$ dilution of resolubilised 884bp fragment.

on a few of the samples that had weak re-amplification to determine if this would improve the results. In addition, a commercial kit (OptiPrime, Stratagene) was used to determine if the pH of the reactions were optimal as well as the effect of various concentrations of KCl. A matrix of $Mg^{2+}$, $K^{+}$ and pH variables was devised to test the effect of these in one large experiment (Table 3.2).

**Table 3.2.** Matrix of $MgCl_2$ and KCl concentrations and pH values of the buffers of the OptiPrime (Stratagene) kit.

| 10mM Tris-HCl | $MgCl_2$ | KCl |
|---|---|---|
| pH 8.3 | 1.5mM | 25mM |
| | | 75mM |
| | 3.5mM | 25mM |
| | | 75mM |
| pH 8.8 | 1.5mM | 25mM |
| | | 75mM |
| | 3.5mM | 25mM |
| | | 75mM |
| pH 9.2 | 1.5mM | 25mM |
| | | 75mM |
| | 3.5mM | 25mM |
| | | 75mM |

However, none of the $MgCl_2$ and KCl concentrations or the different pH values had any positive effect on re-amplification of any of the samples tested (results not shown).

*c) Effect of annealing temperature*

Stable binding of the primers in a PCR reaction is also crucial to the success of the amplification of a fragment. If the annealing temperature of the primers is too high, no binding and extension can take place because the primers have lost attachment to the template. If the annealing temperature is too low, mispriming can occur leading to non-specific amplification. Several different annealing temperatures were tested to determine an optimal temperature for re-amplification of the weakly re-amplified 884bp fragment of some samples. Figure 3.7 shows an example of the results obtained using annealing temperatures of 40, 50 and 60°C and template dilutions of $10^{-1}$ to $10^{-3}$. As can be seen in Figure 3.7 the best re-amplification was achieved using an annealing temperature of 60°C. However, the 60°C annealing temperature appeared to result in length polymorphisms in the amplicons at different template concentrations (lanes C2 and C3). The smearing caused by the 50°C annealing temperature was the result of lengthy annealing times within the thermal profile and shorter annealing times resulted in a more specific re-amplification product (results not shown). The results of the effects of the various annealing temperatures confirmed that the standard annealing temperature used was best suited for re-amplification of the 884bp fragment. However, it was unexpected that no amplification took place at the lower annealing temperature of 40°C, since non-specific amplification is known to occur at such low temperatures. The product of re-amplification formed using the altered cycling time of the annealing temperature was not sufficient for use in downstream applications.

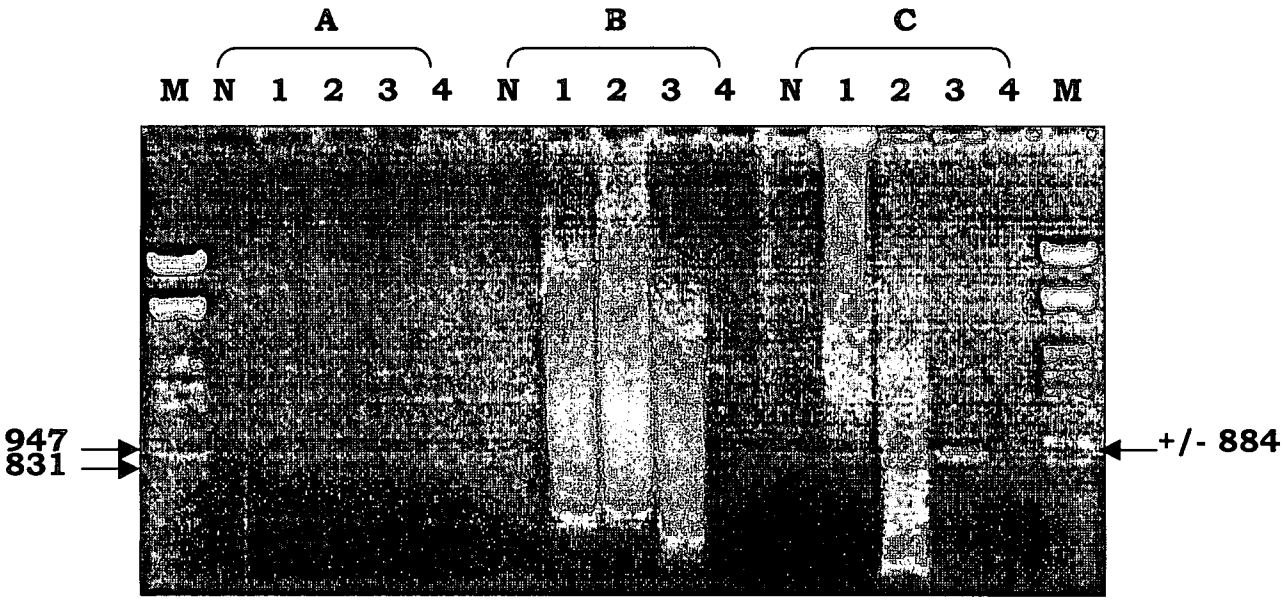



**Figure 3.7.** An example of three different annealing temperatures tested on Glendale N16 884bp fragment re-amplification using a series of template dilutions. A = 40°C; B = 50°C; C = 60°C; M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested); N = no template control; 1 = undiluted; 2 = $10^{-1}$; 3 = $10^{-2}$; 4 = $10^{-3}$ dilution of resolubilised 884bp CP fragment.

*d) Effect of thermostable DNA polymerase enzyme*

Since the first discovery of *Taq* DNA polymerase, enhancements have been made to improve the activity of *Taq* DNA polymerase. Several other thermostable DNA polymerase enzymes have since been discovered or engineered and are commercially available. The first polymerase enzyme tested, AmpliTaq$^{R}$ Gold (Perkin Elmer), did not re-amplify any of the samples, regardless of the template concentration used. Subsequently, a comparison of Boehringer Mannheim *Taq* polymerase and two other polymerase enzymes was made to determine the effects on re-amplification of one sample using several template dilutions of the 884bp fragment. The two enzymes tested were Ex-*Taq* polymerase (Takara) and Expand™ High Fidelity PCR system (Boehringer Mannheim). Expand™ High Fidelity PCR system consists of a mixture of *Taq* DNA polymerase and *Pwo* DNA polymerase enzymes. Each

polymerase enzyme was tested using the reaction conditions described by the respective manufacturers. The effects of the three polymerase enzymes on the re-amplification of the 884bp fragment of the Ixopo N12 sample is shown in Figure 3.8.

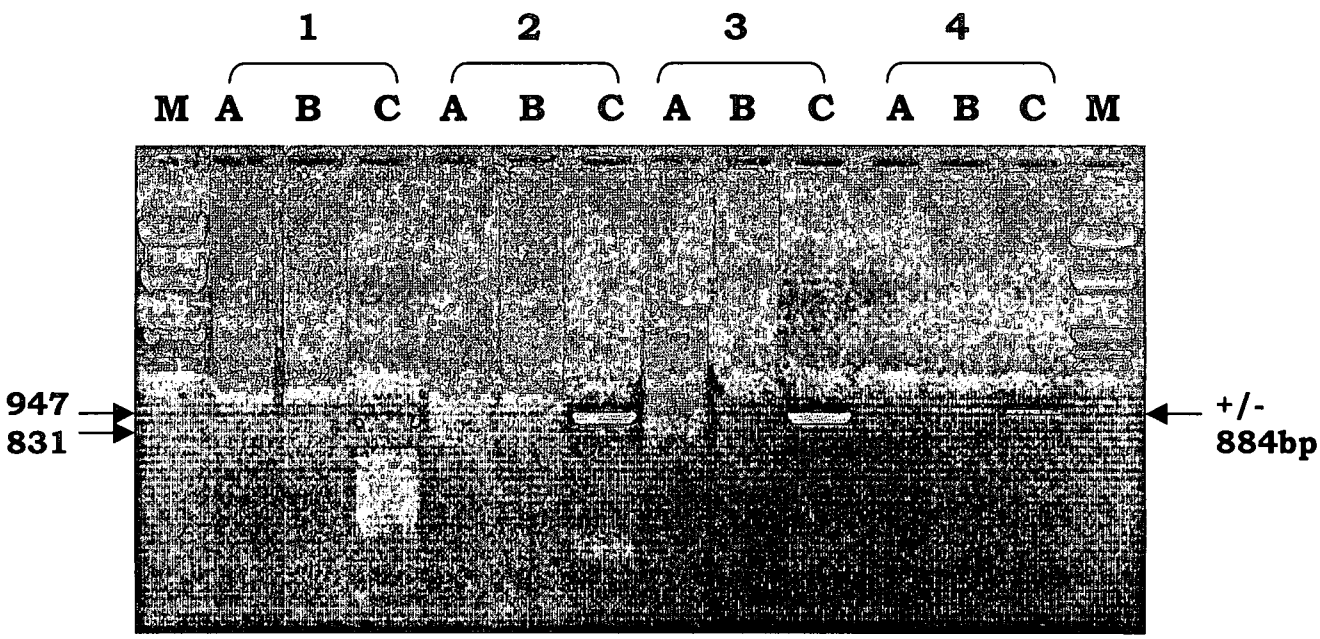



**Figure 3.8.** Effects of three different polymerase enzymes on the re-amplification of four different template concentrations of the 884bp fragment derived from the Ixopo N12 sample. M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested); A = *Taq* polymerase (Boehringer Mannheim); B = Ex-*Taq* polymerase (Takara), C = Expand™ High Fidelity PCR system (Boehringer Mannheim); 1 = undiluted; 2 = $10^{-1}$; 3 = $10^{-2}$; 4 = $10^{-3}$ dilution of resolubilised 884bp fragment.

It is clear that the Expand™ High Fidelity PCR system is the best of the three enzymes tested for the re-amplification of the 884bp fragment of the Ixopo N12 sample. Similar amplification was obtained when the Expand™ polymerase was used on all the other source samples. Using the Expand™ polymerase enzyme for re-amplification of the 884bp fragment it was possible to generate sufficient quantities for use in sequence determination and/or cloning into a plasmid vector.

All the above parameters were tested and optimised using the 884bp fragment from all the regions and their respective varieties. However, none of the parameters had any effect on the re-amplification of the 1400bp fragment from any of the regions or their respective varieties.

### 3.4 Cloning of CP Gene Fragments into Plasmid Vectors

Once the identity of the CP gene fragments had been confirmed by nested PCR, as well as re-amplification in the case of the 884bp fragment, further characterisation of the fragments was undertaken by means of sequencing.

Two methods exist to determine the nucleotide sequence of PCR amplified fragments. These are direct sequencing of a fragment population or sequencing of a fragment after it has been inserted and amplified in a plasmid cloning vector. Both approaches require the use of a purified product to prevent the contamination of the desired sequence with non-specific sequence information.

Re-amplified CP gene 884bp fragments and primary amplified 1400bp fragments were excised from agarose gels and purified from gelling agents and buffer contaminants using the methods described in section 2.2.8. Of the two purification methods used, the two step purification (GenElute followed by QIAquick PCR purification) yielded higher concentrations of purified fragment for the 884bp fragment (results not shown). This method was therefore used for the purification of 884bp and 1400bp fragments from all the regions and their respective varieties.

3.4.1 Cloning of the 884bp CP Gene Fragments into pCR-Script SK$^+$

Purified 884bp fragments of samples from Eston variety NCo376 and Pongola variety 86F3396 (N28) were inserted into the pCR-Script SK$^+$ cloning vector and used to transform competent *E.coli* HB101 cells. Putative positive transformants were selected by colour discrimination, where white colonies were considered as positive and blue colonies regarded as negative. An example of the blue/white selection of colonies obtained after transformation is shown in Figure 3.9. White colonies were sampled (a few cells) with sterile toothpicks and used as template in PCR to confirm the presence of the inserted 884bp fragment using the T7/M13 Reverse primer pair. Amplification of an approximately 1100bp fragment (containing the CP gene sequence and approximately 200bp flanking vector sequences) confirmed the insertion of the 884bp fragment in the plasmid vector. Results of the PCR amplification of several white colonies tested for the presence of the cloned 884bp fragment from Pongola variety 86F3396 (N28) in pCR-Script SK$^+$ is shown in Figure 3.10. Similar results were obtained for the 884bp fragment from Eston variety NCo376 cloned in pCR-Script SK$^+$ plasmid vector. The smaller fragments (approximately 200bp) in Figure 3.10 are the amplification products of the region between the two primers (T7/M13 Reverse) which flank the insertion site in the plasmid. These fragments indicate that the white colonies contained a plasmid that did not contain the 884bp fragment insert.

However, the plasmid in the white colonies may have contained a mutation in the *lacz* gene which resulted in the cells inability to metabolise the X-gal substrate that results in white colonies. A total of six positive clones were

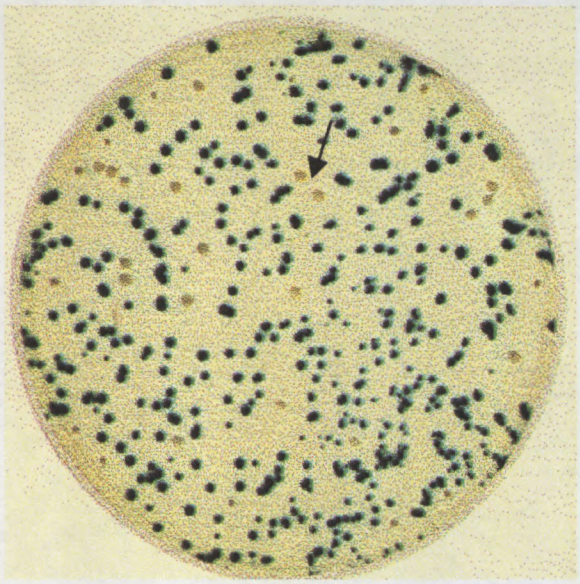


**Figure 3.9**. An example of blue-white selection of colonies obtained after plating *E.coli* transformed with CP gene sequences cloned in pCR-Script SK+ on to agar plates containing IPTG and X-gal. White colonies (indicated by an arrow) represent cells transformed with plasmid containing the inserted 884bp fragment, while blue colonies represent cells transformed with plasmid without an 884bp fragment.

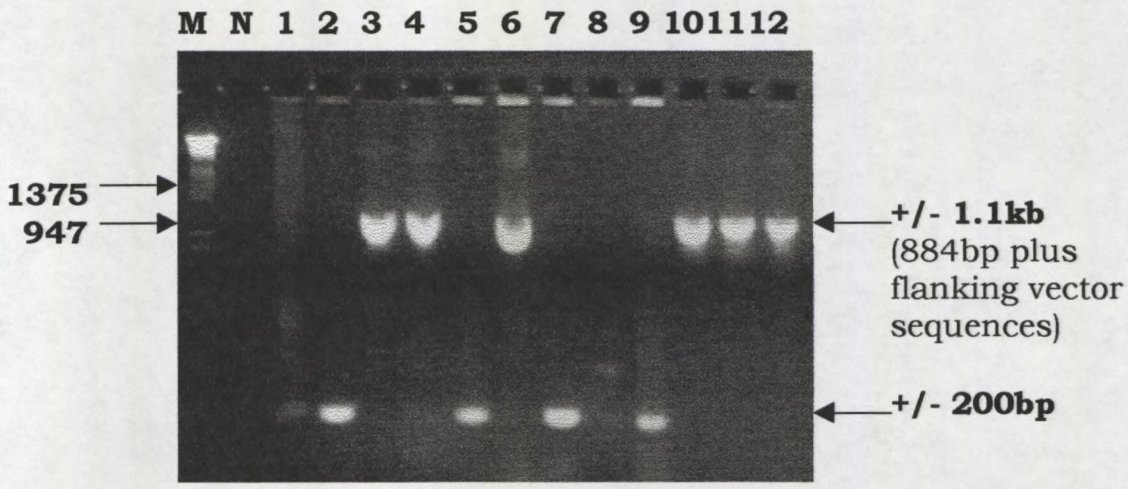



**Figure 3.10.** Results of direct PCR amplification of bacterial cells from white colonies using specific primers to test for the presence of the 884bp fragment in pCR-Script SK+. M = molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested), N = No template control. Lanes 1-12 are 12 randomly selected putative positive (white) colonies containing the 884bp fragment from Pongola variety 86F3396 (N28).

obtained for 884bp fragment of the Pongola variety 86F3396 as shown in Figure 3.10 and five positive clones of the Eston variety NCo376 (results not shown). However, the modified pCR-Script SK$^+$ plasmid vector was not very efficient as a plasmid cloning vector as no clones were obtained when the 884bp fragments of the remaining samples were attempted to be inserted into the plasmid vector. As a result, two other plasmid cloning vectors were used for further cloning experiments. Since the 1400bp fragment contained the same genic region as the 884bp fragment, all subsequent cloning was done using this fragment.

### 3.4.2 Cloning of the 1400bp CP Gene Fragments into pCR2.1 and pGEM-T Easy T-A Vectors

Initially, the cloning vector pCR2.1 was used to clone the 1400bp fragment, as it was reported to be the most efficient of the commercially available T-A cloning vectors. However, due to high cost and difficulties of availability of the product, the pGEM-T Easy vector was used as an alternative. Consequently, the 1400bp CP gene fragments of all the regions and their respective varieties were cloned into two different cloning vectors. Table 3.3 is a summary of the cloning results of the 884bp and 1400bp CP gene fragments, from all the regions and varieties used, into the various cloning vectors.

Purified 1400bp fragments were ligated into the vectors pCR2.1 and pGEM-T Easy as described in section 2.2.10 and then delivered to their respective *E.coli* cell lines by transformation as described in section 2.2.12. White colonies obtained after transformation were selected based on colour

**Table 3.3**. Results of cloning the 884bp and 1400bp coat protein gene fragments into various plasmid cloning vectors.

| Region | Variety | Plasmid Vector Used | Number of Clones Obtained | Clone Designations |
|---|---|---|---|---|
| Eston | N12<br>NCo376 | pCR2.1<br>pCR-Script SK$^+$<br>pCR2.1 | 8<br>(5)*<br>10 | $E_{12}$ #1 - 8<br>pCPN001-5<br>$E_{376}$ #1 - 10 |
| Glendale | N16<br>NCo376 | pCR2.1<br>pCR2.1 | 7<br>5 | $G_{16}$ #1 - 7<br>$G_{376}$ #1 - 5 |
| Ixopo | N12<br>NCo376 | pCR2.1<br>pGEM-T Easy | 6<br>4 | $I_{12}$ #1 - 6<br>$I_{376}$ #1 - 4 |
| Melmoth | N12 | PGEM-T Easy | 4 | $MM_{12}$ #1 - 4 |
| Mount Edgecombe | NCo376 | pGEM-T Easy | 4 | $ME_{376}$ #1 - 4 |
| Pongola | 86F3396<br>N19<br>N25 | pCR-Script SK$^+$<br>pGEM-T Easy<br>pCR2.1 | (6)*<br>5<br>10 | pCPN101 - 6<br>$P_{19}$ #1 - 5<br>$P_{25}$ #1 - 10 |
| Port Shepstone | NCo376 | pGEM-T Easy | 3 | $PS_{376}$ #1 - 3 |
| Tongaat | N12<br>NC0376 | pCR2.1<br>pGEM-T Easy | 3<br>4 | $T_{12}$ #1 - 3<br>$T_{376}$ #1 - 4 |

* Figure in parenthesis denotes the number of clones containing the 884bp CP gene fragment. All other clones contain the 1400bp CP gene fragment.

discrimination as mentioned for pCR-Script SK$^+$ (section 3.4.1). The cells were tested directly in PCR using the T7/M13 Reverse primer pair for cells transformed with the pCR2.1 vector and the T7/SP6 primer pair for the cells transformed with pGEM-T vector. Amplification of an approximately 1600bp fragment confirmed the insertion of the 1400bp fragment in either of the two plasmid vectors. Although different primers pairs were used for the confirmation of an inserted fragment, a fragment of similar size was expected for both vectors. An example of the PCR amplification of the 1600bp fragment (1400bp fragment and 200bp vector sequences) from white colonies obtained after transformation with the pCR2.1 vector is shown in Figure 3.11.

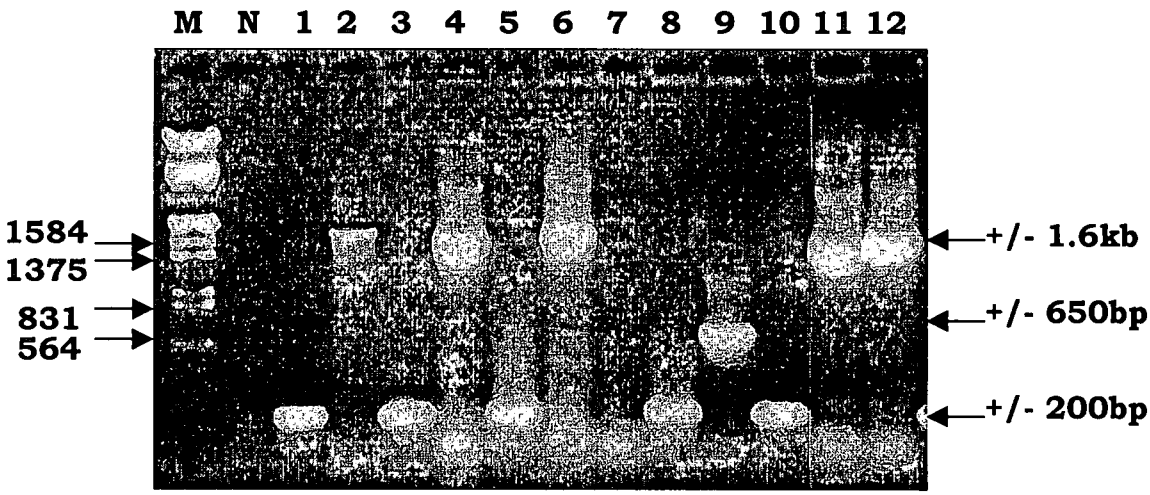



**Figure 3.11**. Results of direct PCR amplification of cells from white bacterial colonies using universal primers to test for the presence of the 1400bp in pCR2.1 vector. M = molecular weight marker (λ DNA EcoR I/Hind III digested), N = No template control. Lanes 1-12 are 12 randomly selected putative positive (white) colonies containing the 1400bp fragment from Pongola variety N25.

The smaller amplified fragments are due to the absence of an inserted 1400bp fragment in the plasmid vector which was used to transform the bacterial cell. A slightly larger fragment (approximately 650bp) was amplified from one of the colonies tested (Figure 3.11, lane 9) and this may be due to only a portion of the 1400bp fragment being inserted in the vector. Some colonies tested co-amplified a larger fragment (approximately 3000bp) which may have been of bacterial origin, since whole cells were used as template in PCR.

As several positive CP gene fragment clones were obtained for all the regions and their respective varieties (Table 3.3), it was possible to proceed in the isolation and purification of the plasmid DNA for sequence determination.

### 3.5. Recombinant Plasmid Analysis

Since poor template quality and quantity is the most common cause of sequencing failures (Griffin & Griffin, 1993), it was essential to isolate high yields of pure plasmid DNA for sequence determination.

Plasmid DNA, containing the CP gene fragment, was isolated from pure cultures of transformed bacterial cells using the method described in section 2.2.14. Three clones from each variety within each region were selected for sequence determination since three was the minimum number of clones obtained for all of the samples. The underlying reason for the choice of three clones as the minimum number was that sequence analysis of three clones would indicate the likelihood or otherwise of strain identity within each infected source sample. Concentration and purity of the isolated plasmid DNA was determined spectrophotometrically. Average plasmid DNA yield ranged between 50 and 100µg with $A_{260}/A_{280}$ purity ratios of between 1.5 - 2.1. In addition, an aliquot of the plasmid DNA was visualised by agarose gel electrophoresis and ethidium bromide staining to confirm results obtained from spectrophotometric analysis. Figure 3.12 shows an example of results of agarose gel separation of plasmid DNA isolated from three clones containing the 1400bp CP gene fragment from all regions sampled and their respective varieties. The plasmid DNA for most of the clones was of the expected size of approximately 5300bp for the pCR2.1-insert-containing vector and approximately 4400bp for the pGEM-T Easy-insert-containing vector. However, a few plasmids seen in Figure 3.12 in lanes F3, I1, K2 and K3 were larger than the expected size. This may have been due to a double-insertion

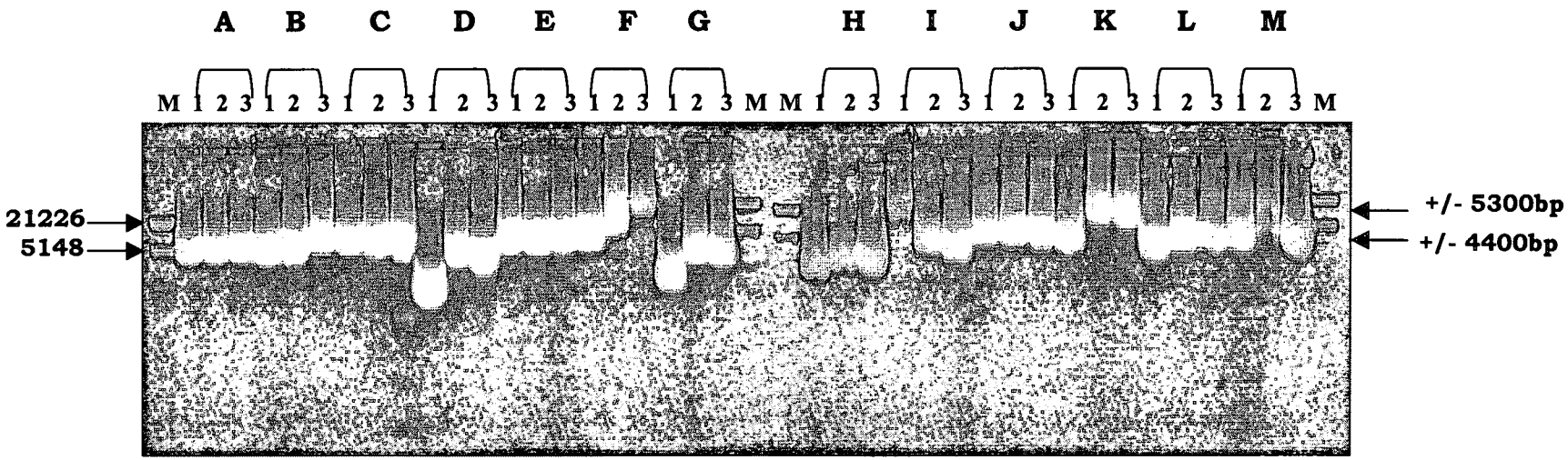



**Figure 3.12**. Agarose gel electrophoretic analysis of plasmid DNA containing the 1400bp CP gene fragment from all regions sampled and their respective varieties. M = Molecular weight marker III (λ DNA *EcoR*I/*Hind*III digested), A = Eston, N12; B = Eston, NCo376; C = Glendale, N16; D = Glendale, NCo376; E = Ixopo, N12; F = Ixopo, NCo376; G = Mount Edgecombe, NCo376; H = Melmoth, N12; I = Pongola, N19; J = Pongola, N25; K = Port Shepstone, NCo376; L = Tongaat, N12; M = Tongaat, NCo376. 1, 2, 3, = three separate clones containing the CP gene fragment.

of the fragment into the plasmid. The smaller sized plasmids seen in lanes D3, G1 - 3 and H1 - 3 appeared to be the result of clones containing smaller sized inserts; that is, not full length 1400bp CP fragment. The considerably smaller sized plasmid in lane D1 is probably due to rearrangement within the plasmid which may have resulted in a portion of the inserted 1400bp fragment or other plasmid sequences being excised. An extra band of approximately 1000bp, which is probably of bacterial origin, was present in lane C3 containing the 1400bp CP fragment from Glendale variety N16. The size discrepancies between some of the clones as seen in Figure 3.12 may also be due to the presence of contaminants within each sample such as protein and salts left over from the extraction procedure, which affect electrophoretic mobility.

### 3.5.1 Restriction Enzyme Analysis of Recombinant Plasmids

Since there appeared to be a lack of uniformity in the expected size of the extracted plasmid DNA, it was necessary to determine the conformation and structure of the recombinant plasmids by restriction enzyme analysis.

Plasmid DNA was first digested with the restriction enzyme *EcoR* I to determine whether the recombinant plasmids contained the 1400bp CP fragment (see Figure 2.3 for position of restriction enzyme sites within each vector). This enzyme would also serve to confirm the presence of the expected *EcoR* I site within the inserted 1400bp fragment shown in Figure 3.13.

```
GGAAACAACA-GTGGTCAACC-ATCAACTGTA-GTCGACAACA-CGCTAATGGT-TATCTTAGCA-      60
TTCAATTATG-CAATGTTATC-AAGTGGAATA-CAAGACAATG-AAATCGACAA-CTGTTGCAGA-     120
ATGTTCGCCA-ATGGAGACGA-TTTATTGTTG-GCAGTACATC-CCGATTATGA-ACATATACTG-     180
GATGGATTTC-AGACCCATTT-TGGAAACCTT-GGTTTAAATT-TTGAATTCAA-TTCACGAACG-     240
AGGGAAAAAT-CGAACTTGTG-GTTCATGTCA-ACACAGGGAC-TCAAGTATGA-AGGCATTTAT-     300
ATACCAAAAC-TCGAAAAGGA-AAGAATAGTC-GCCATACTCG-AGTGGGATCG-TTCAAATCTA-     360
CCCGAGCACC-GTTTAGAAGC-CATTTGTGCA-GCAATGGTCG-AAGCGTGGGG-CTATTCAGAT-     420
CTCATACATG-AAATTCGGAA-ATTTTATGCA-TGGCTTCTGG-AAATGCAACC-ATTTGCAAAT-     480
TTAGCCAAAG-AGGGTCTAGC-TCCATATATT-GCAGAAACAG-CGCTTCGTAA-CTTATATCTA-     540
GGTTCAGGGA-TTAAGGAAGA-GGAAATTGAA-AAATACTTCA-AGCAGTTTGC-AAAGGACCTC-     600
CCTGGGTATT-TAGAGGATTA-CAACGAAGAA-GTTTTTCACC-AAGCTGGAAC-AGTCGATGCA-     660
GGCGCTCAAG-GAGGAGGTGG-AAACGCCGGA-ACTCAGCCGC-CAGCCACTGG-AGCAGCAGCT-     720
CAAGGAGGAG-CTCAACCACC-AACAACTGGA-GCAGCTGCAC-AACCGCCCGC-GACTCAAGGT-     780
TCACAACCGC-CCACAGGGGG-AGCTACTGGT-GGAGGTGGTG-CACAAACAGG-AGCTGGTGGA-     840
AATAGCTCAA-TTACAGGGGG-TCAGAAAGAC-AAGGATGTAG-ATGCTGGTAC-GACAGGCAAA-     900
ATCACAGTGC-CAAAACTTAA-AGCCATGTCG-AAGAAGATGC-GCTTGCCAAA-AGCAAAAGGA-     960
AAGGACGTTT-TGCATCTGGA-CTTTCTGTTA-ACATACAAAC-CGCAACAACA-AGACATATCA-    1020
AACACAAGAG-CAACCAGAGA-GGAGTTTGAT-AGGTGGTATG-AAGCCATAAA-GAAGGAATAT-    1080
GAAATAGATG-ACACACAAAT-GACAGTTGTC-ATGAGTGGTC-TAATGGTATG-GTGTATTGAG-    1140
AATGGTTGCT-CACCAAACAT-AAACGGAAAT-TGGACAATGA-TGGATGGAGA-TGAACAAAGA-    1200
GCTTTCCCAT-TAAAACCAAT-CATTGAAAAC-GCATCTCCAA-CATTCCGGCA-AATAATGCAT-    1260
CATTTCAGTG-ATGCAGCTGA-AGCATATATC-GAGTATAGAA-ACTCTACAGA-GCGATATATG-    1320
CCACGATATG-GACTTCAGCG-AAATCTCACC-GACTATAGCT-TAGCGCGGTA-TGCATTTGACT    1380
```

**Figure 3.13**. Nucleotide sequence of the expected 1400bp fragment amplified by PCR using the primer pair AGPOT2/GSr. The sequence of the *EcoR* I restriction site within the 1400bp fragment is shaded. The nucleotide sequence is that of SCMV strain E as found in GenBank under accession number U57357.

The expected cleavage profiles of the recombinant plasmids after digestion with *EcoR* I are shown in Table 3.4. Figure 3.14 shows the restricted fragments obtained after *EcoR* I digestion of plasmid DNA. Several plasmids produced the expected cleavage pattern of approximately 3900bp and 1160bp for pCR2.1 shown in Figure 3.14 lanes A1, C2 and J2, while that for pGEM-T Easy was approximately 3000bp and 1160bp shown in lanes D2, L3, M2 and M3. The smaller 223bp fragment was not visible on the agarose gel after electrophoretic separation of the restriction products. Partial digestion of the

**Table 3.4** Expected cleavage profiles of recombinant plasmids digested with *EcoR* I restriction endonuclease.

| Fragment / Vector | Size (bp) | No. of Sites | Expected Fragment Sizes (bp) |
|---|---|---|---|
| 1400bp fragment | 1400 | 1 | 1158; 223 |
| PCR2.1 + 1400bp insert | +/- 5300 | 2 | 3875; 1158; 223 |
| pGem-T Easy + 1400bp insert | +/- 4400 | 2 | 3000; 1158; 223 |

1400bp fragment was evident in several of the plasmids resulting in a doublet of 1400bp and 1160bp as seen in lanes A2, B1 - B3, C1, C3 and J3. Similarly, the partial digestion of vector DNA resulted in doublets of approximately 5300bp and 3900bp for pCR2.1 as shown in lanes C2, C3 and J1 - J3; and doublets of approximately 4400bp and 3000bp in the case of pGEM-T Easy as shown in lanes D3, I1 - I3, K1 - K3, L1, L2, M1 and M3. A number of plasmids appeared not to possess the *EcoR* I site within the 1400bp fragment as seen in lanes D1, D3, E1, E2, I3, J3 and M1. Surprisingly, several recombinant plasmids appeared to contain no 1400bp CP fragment as seen in lanes F1, F2, G1 - G3 and H1 - H3. This may have been due to re-arrangement within the plasmid that resulted in the excision of the 1400bp fragment.

Since the *EcoR* I restriction profiles did not conclusively elucidate the conformation of the recombinant plasmids, a second restriction endonuclease, *Not* I, was used to provide more conclusive information on the condition of the plasmids. Table 3.5 shows the expected restriction fragment sizes of recombinant vectors after digestion with the *Not* I enzyme.

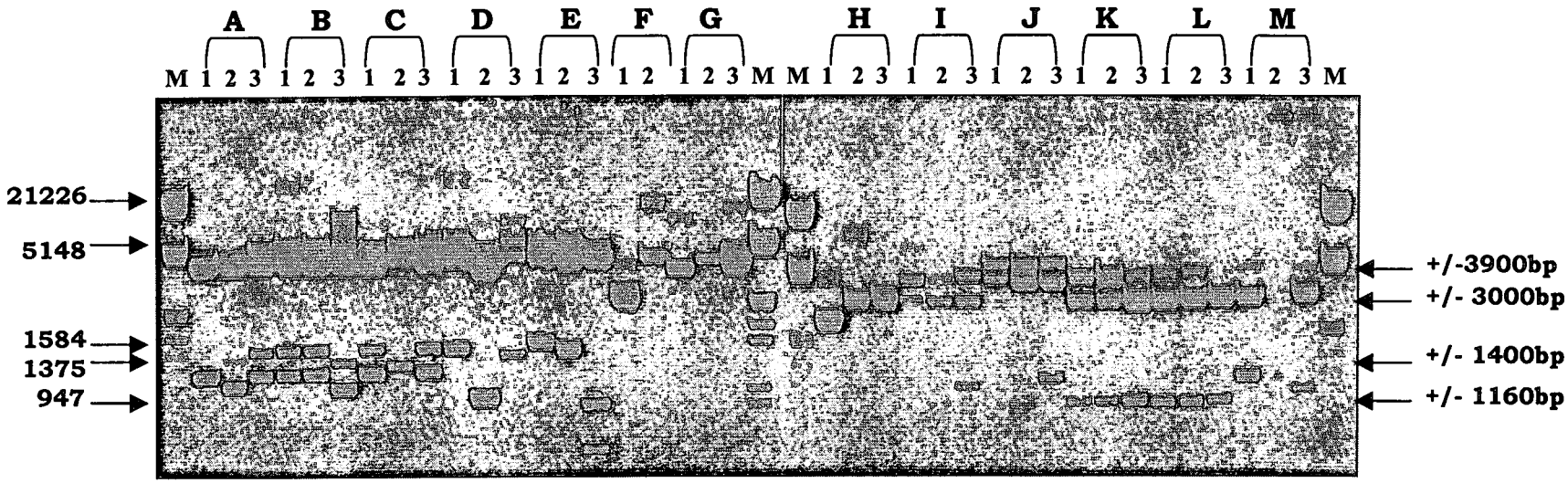



**Figure 3.14**. Analysis of *EcoR* I digested plasmid DNA containing the 1400bp CP gene fragment from SCMV infected source samples. M = Molecular weight marker III (λ DNA *EcoR*I/*Hind*III digested), A = Eston, N12; B = Eston, NCo376; C = Glendale, N16; D = Glendale, NCo376; E = Ixopo, N12; F = Ixopo, NCo376; G = Mount Edgecombe, NCo376; H = Melmoth, N12; I = Pongola, N19; J = Pongola, N25; K = Port Shepstone, NCo376; L = Tongaat, N12; M = Tongaat, NCo376. 1, 2, 3, = three separate clones containing the CP gene fragment.

**Table 3.5.** Expected cleavage profile of recombinant plasmids digested with *Not* I restriction endonuclease.

| Fragment / Vector | Size (bp) | No. of Sites | Expected Fragment Sizes (bp) |
|---|---|---|---|
| 1400bp fragment | 1400 | 0 | 1400 |
| PCR2.1 + 1400bp insert | +/- 5300 | 1 | +/- 5300 |
| pGem-T Easy + 1400bp insert | +/- 4400 | 2 | 2984; 1400; 34 |

Restriction patterns produced by *Not* I enzyme digestion of recombinant plasmids are shown in Figure 3.15. *Not* I restriction resulted in a single restriction for the pCR2.1 vector, producing a fragment of the expected size of approximately 5300bp as seen in lanes A1 - 3, B1 - 3, C1 - 3, D1 - 3, E1 - 3 and J1 - 3. Similarly, fragments of the expected sizes of approximately 3000bp and 1400bp were obtained for the pGEM-T Easy vector as seen in lanes F1, F2, I1 - 3, K1 - 3, L1 - 3 and M3. Incomplete restriction of the plasmid DNA was evident in several plasmids, but most noticeable in clone 2 of the sample from Tongaat, variety N12. *Not* I restriction also confirmed the aberrant nature of several plasmids, as seen in lanes F3, G1 - 3 and H1 - 3.

Despite these observations, it was decided to continue with sequencing of all plasmids, as it was still possible that the N-terminal region of the inserted 1400bp CP fragment would be present for sequence comparison.

## 3.6. Direct Sequencing Compared to Sequencing after Cloning

Methods for DNA sequence determination have improved over the last two decades since they were first developed (Griffin & Griffin, 1993). Since one of

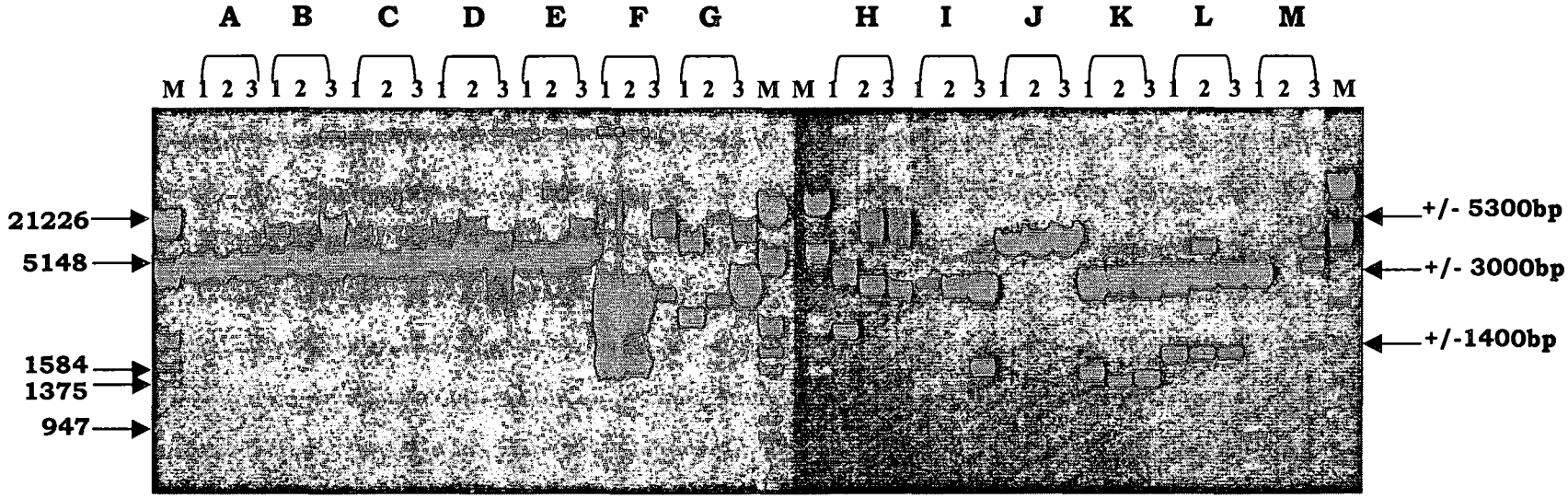



**Figure 3.15**. Analysis of *Not* I digested plasmid DNA containing the 1400bp CP gene fragment from SCMV infected source samples. M = Molecular weight marker III (λ DNA *EcoR*I/*Hind*III digested), A = Eston, N12; B = Eston, NCo376; C = Glendale, N16; D = Glendale, NCo376; E = Ixopo, N12; F = Ixopo, NCo376; G = Mount Edgecombe, NCo376; H = Melmoth, N12; I = Pongola, N19; J = Pongola, N25; K = Port Shepstone, NCo376; L = Tongaat, N12; M = Tongaat, NCo376. 1, 2, 3, = three separate clones containing the CP gene fragment.

the recent developments includes direct sequence determination of double-stranded PCR products, it was decided to investigate how this new method of sequencing compared with an older method of cloning the DNA fragment into a cloning vector prior to sequencing.

In the early stages of sequencing, samples from two regions were used to compare the difference between direct sequencing and sequencing after cloning. The two regions and their respective varieties were Eston variety NCo376 and Pongola variety 86F3396. Purified, re-amplified 884bp CP fragment from several regions and varieties were cycle sequenced directly using dye terminator chemistry as described in section 2.2.15b. Table 3.6 shows the difference in sequence lengths obtained and the percentage of failed sequences using the two methods of sequencing.

**Table 3.6.** Comparison of sequence data obtained using direct sequencing or sequencing after cloning. Comparison of Eston and Pongola 884bp fragments after cycle sequencing using the specific primer BHf.

| **Region** | **Variety** | **Direct Sequencing** | | **Sequencing after Cloning** | |
|---|---|---|---|---|---|
| | | **Longest Sequence (bp)** | **% Sequence Failure** | **Longest Sequence (bp)** | **% Sequence Failure** |
| Eston | NCo376 | 762 | 25 | 448 (pCPN001) | 25 |
| Pongola | 86F3396 | 568 | 25 | 438 (pCPN103) | 50 |

Direct sequencing produced longer sequences and a better success rate than sequencing after cloning; however the length of the sequences were not as consistent as those obtained by sequencing of cloned fragments (results not shown). Since sequencing of cloned fragments yielded more consistent

sequence lengths than direct sequencing of the re-amplified fragment, it was decided to determine the sequence of the CP gene fragments after cloning in a plasmid vector. At that time, the introduction of longer electrophoretic capillaries for sequencing permitted longer sequence lengths to be obtained for the cloned fragments, thereby adding support to the use of cloning.

### 3.7 Preliminary Analysis of SCMV Coat Protein Fragment Sequences

It was determined from the previous exercise that sequencing after cloning would be the more suitable method of sequence determination of the cloned CP fragments and thus an effective sequencing strategy had to be established to obtain the most reliable sequence data for analysis. Since the N-terminal region of the CP gene was the region of interest, it was decided that all sequencing reactions would be performed primarily with the forward primers BHf and CPCON-F for the sense strand and reverse primers GSr and GSfINV for the opposite strand. These primers would provide sufficient sequence information (at least 360bp from the start of the CP gene) of the cloned CP fragments for accurate strain identification (Pers. Comm., Mirkov, TE, Professor, Texas A & M University).

In order to obtain accurate consensus sequences of the cloned CP fragments, three repetitions of successful sequences were required for each clone. This required three separate sequencing reactions, using exactly the same reaction conditions for each clone (Pers. Comm., Groenewald, JH, Lecturer, University of Stellenbosch). In preliminary sequence determination the primers CPCON-F, GSfINV and GSr proved to be unsuitable for sequencing. CPCON-F was

found to be too long and disengaged prematurely from the template during cycle-sequencing resulting in very short sequences (results not shown). Similarly, sequence reactions directed by GSfINV mostly produced very short sequences, possibly due to premature disengaging from the template. GSr produced fairly good sequences; however, the length of most of the sequences was insufficient to provide significant overlapping region with sequences obtained using the BHf primer. It was therefore decided to proceed with the remaining sequence determination using the BHf primer exclusively. Table 3.7 shows the sequence success using the BHf primer on the cloned 1400bp CP fragment.

Several plasmids were successfully sequenced using the primer BHf and had no failed sequence reactions. However, a large number of clones did not sequence successfully with the BHf primer as seen in the shaded areas in Table 3.7. Two plasmid clones, $P_{19}$#2 and $T_{376}$#2, did produce one successful sequence with BHf therefore indicating that the failed reactions were possibly due to some inhibiting agent in the plasmid preparation.

Plasmid DNA of the clones which failed to sequence was column purified using a commercial kit (QIAquick PCR purification kit) to remove all possible contaminants that may have been inhibiting the sequencing reactions. This, however, did not have any effect on sequence success using BHf. Several other parameters were tested to improve the sequenceability of the recalcitrant plasmids and these included: increased template concentration; use of

**Table 3.7.** Results of cycle-sequencing three recombinant plasmids containing the 1400bp CP fragment from each region sampled and its respective variety. Purified plasmids were sequenced using the primer BHf. Shaded regions represent unsuccessful sequencing attempts using the BHf primer.

| Region | Variety | Clone Designation | % Sequencing Success | % Sequencing Failure |
|---|---|---|---|---|
| Eston | N12 | $E_{12}$#1 | 100% (3/3) | 0% (0/3) |
| | | $E_{12}$#2 | 100% (3/3) | 0% (0/3) |
| | | $E_{12}$#3 | 100% (3/3) | 0% (0/3) |
| | NCo376 | $E_{376}$#1 | 100% (3/3) | 0% (0/3) |
| | | $E_{376}$#2 | 100% (3/3) | 0% (0/3) |
| | | $E_{376}$#3 | 100% (3/3) | 0% (0/3) |
| Glendale | N16 | $G_{16}$#1 | 100% (3/3) | 0% (0/3) |
| | | $G_{16}$#2 | 0% (0/8) | 100% (8/8) |
| | | $G_{16}$#3 | 100% (3/3) | 0% (0/3) |
| | NCo376 | $G_{376}$#1 | 100% (3/3) | 0% (0/3) |
| | | $G_{376}$#2 | 0% (0/6) | 100%(6/6) |
| | | $G_{376}$#3 | 0% (0/5) | 100%(5/5) |
| Ixopo | N12 | $I_{12}$#1 | 0% (0/5) | 100%(5/5) |
| | | $I_{12}$#2 | 0% (0/5) | 100%(5/5) |
| | | $I_{12}$#3 | 0% (0/5) | 100%(5/5) |
| | NCo376 | $I_{376}$#1 | 0% (0/5) | 100%(5/5) |
| | | $I_{376}$#2 | 0% (0/5) | 100%(5/5) |
| | | $I_{376}$#3 | 0% (0/5) | 100%(5/5) |
| Mount Edgecombe | NCo376 | $ME_{376}$#1 | 0% (0/3) | 100%(3/3) |
| | | $ME_{376}$#2 | 0% (0/3) | 100% (3/3 |
| | | $ME_{376}$#4 | 0% (0/3) | 100% (3/3 |
| Melmoth | N12 | $MM_{12}$#1 | 0% (0/2) | 100% (2/2) |
| | | $MM_{12}$#2 | 0% (0/2) | 100% (2/2) |
| | | $MM_{12}$#3 | 0% (0/2) | 100% (2/2) |
| Pongola | N19 | $P_{19}$#1 | 100% (3/3) | 0% (0/3) |
| | | $P_{19}$#2 | 10% (1/10) | 90% (9/10) |
| | | $P_{19}$#3 | 100% (3/3) | 0% (0/3) |
| | N25 | $P_{25}$#1 | 100% (3/3) | 0% (0/3) |
| | | $P_{25}$#2 | 100% (3/3) | 0% (0/3) |
| | | $P_{25}$#3 | 0% (0/8) | 100% (8/8) |
| Port Shepstone | NCo376 | $PS_{376}$#1 | 0% (0/2) | 100% (2/2) |
| | | $PS_{376}$#2 | 0% (0/2) | 100% (2/2) |
| | | $PS_{376}$#3 | 0% (0/2) | 100% (2/2) |
| Tongaat | N12 | $T_{12}$#1 | 100% (3/3) | 0% (0/3) |
| | | $T_{12}$#2 | 100% (3/3) | 0% (0/3) |
| | | $T_{12}$#3 | 100% (3/3) | 0% (0/3) |
| | NCo376 | $T_{376}$#1 | 0% (0/2) | 100% (2/2) |
| | | $T_{376}$#2 | 34% (1/3) | 66% (2/3) |
| | | $T_{376}$#3 | 0% (0/2) | 100% (2/2) |

stringent annealing temperatures in cycle-sequencing reactions and the use of freshly prepared primer stocks. These too had little effect on improving sequencing of the plasmids.

Since none of the above mentioned modifications appeared to have had any effect on improving sequencing, the possibility that the recalcitrant plasmids may have developed altered BHf priming sites due to some mutational occurrence seemed more likely. A modification of the BHf primer, BHf-deg, containing several degeneracies in the primer sequence, was used to sequence several of the recalcitrant plasmids. None of the plasmids sequenced with the BHf-deg primer produced any successful sequences indicating that the BHf priming site, nor any variation of it, was present in those cloned CP fragments.

In a final attempt to confirm the presence of vector sequences and to elucidate the sequences of the cloned fragments in the recalcitrant plasmids, the vector-specific primer T7 was used. Table 3.8 shows the sequencing results using the T7 primer on the recalcitrant plasmid clones. Most of the recalcitrant plasmids produced successful sequences when sequenced with the T7 primer, indicating that the plasmid did possess a priming site for the T7 primer but had lost the region of the CP fragment containing the priming site for the BHf primer. This may have occurred due to mutations or as a result of rearrangement within the plasmid. The failure of several of the plasmids to produce successful sequences using the T7primer as seen in the shaded areas in Table 3.8, may be attributed to the aberrant nature of the plasmid DNA (see Figure 3.14 lanes F1-3, G1-2, H1-3, M1-3 section 3.5). These plasmids displayed unusual restriction patterns and differed in size from the other recombinant plasmids. This however does not explain the failure of clone $P_{25}$#3 to sequence with the T7 primer, since it appeared to be normal in every

**Table 3.8**. Sequence results of recalcitrant plasmids using the universal primer T7 for confirmation of the presence of vector sequences.

| Region | Variety | Clone Designation | Sequence Success using T7 primer |
|---|---|---|---|
| Glendale | N16 | $G_{16}$#2 | Yes |
| | NCo376 | $G_{376}$#2 | Yes |
| | | $G_{376}$#3 | Yes |
| | N12 | $I_{12}$#1 | Yes |
| | | $I_{12}$#2 | Yes |
| | | $I_{12}$#3 | Yes |
| Ixopo | NCo376 | $I_{376}$#1 | No |
| | | $I_{376}$#2 | No |
| | | $I_{376}$#3 | No |
| Mount Edgecombe | NCo376 | $ME_{376}$#1 | No |
| | | $ME_{376}$#2 | No |
| | | $ME_{376}$#4 | Yes |
| Melmoth | N12 | $MM_{12}$#1 | No |
| | | $MM_{12}$#2 | No |
| | | $MM_{12}$#3 | No |
| Pongola | N19 | $P_{19}$#2 | Yes |
| | N25 | $P_{25}$#3 | No |
| Port Shepstone | NCo376 | $PS_{376}$#1 | Yes |
| | | $PS_{376}$#2 | Yes |
| | | $PS_{376}$#3 | Yes |
| Tongaat | NCo376 | $T_{376}$#1 | No |
| | | $T_{376}$#2 | No |
| | | $T_{376}$#3 | No |

respect.

Sequence lengths obtained using the T7 primer where not sufficient to provide sequence information on the N-terminal region of the inserted CP fragment. No further sequence determination was therefore attempted for these plasmids.

Since the sequence determination of the normal plasmid clones was performed in triplicate, it was necessary to obtain a consensus of these sequences. The Contig Manager function in the sequence analysis computer programme DNASIS v2.1, 1991 (Hitachi) was used to generate consensus

sequences for each of the successfully sequenced plasmids. Default values of a 30bp region of a 70% minimum matching homology was used for the overlap. Consensus sequences produced using the Contig Manager function were displayed as a diagram showing the aligned sequences in one colour and the resultant consensus sequence in a different colour at the bottom as seen in Figure 3.16A and 3.16C. Several plasmid clones were sequenced in both orientations and an example of the consensus sequence obtained from these sequences is shown in Figure 3.16B. The complementary sequences are coloured black to show the reverse orientation.

The consensus sequences were also displayed as text-based residues as seen in Figure 3.17. These sequences were saved with a file name containing the name of the plasmid followed by the CON assignment to designate it as a consensus sequence. For example, E12#1CON.DNA was the file name of the consensus sequence of clone 1 of Eston variety N12. Single sequences (no consensus possible) were obtained for several clones using either the BHf or T7 primer. These sequences files were named using the clone designation followed by the name of the primer used to obtain the sequence. For example, I12#1-T7.DNA is the file name of the sequence of clone 1 of Ixopo variety N12 obtained with the T7 primer. All nucleotide sequences analysed in DNASIS required a DNA file extension for easy recognition. These consensus sequences were then used in subsequent comparison studies.

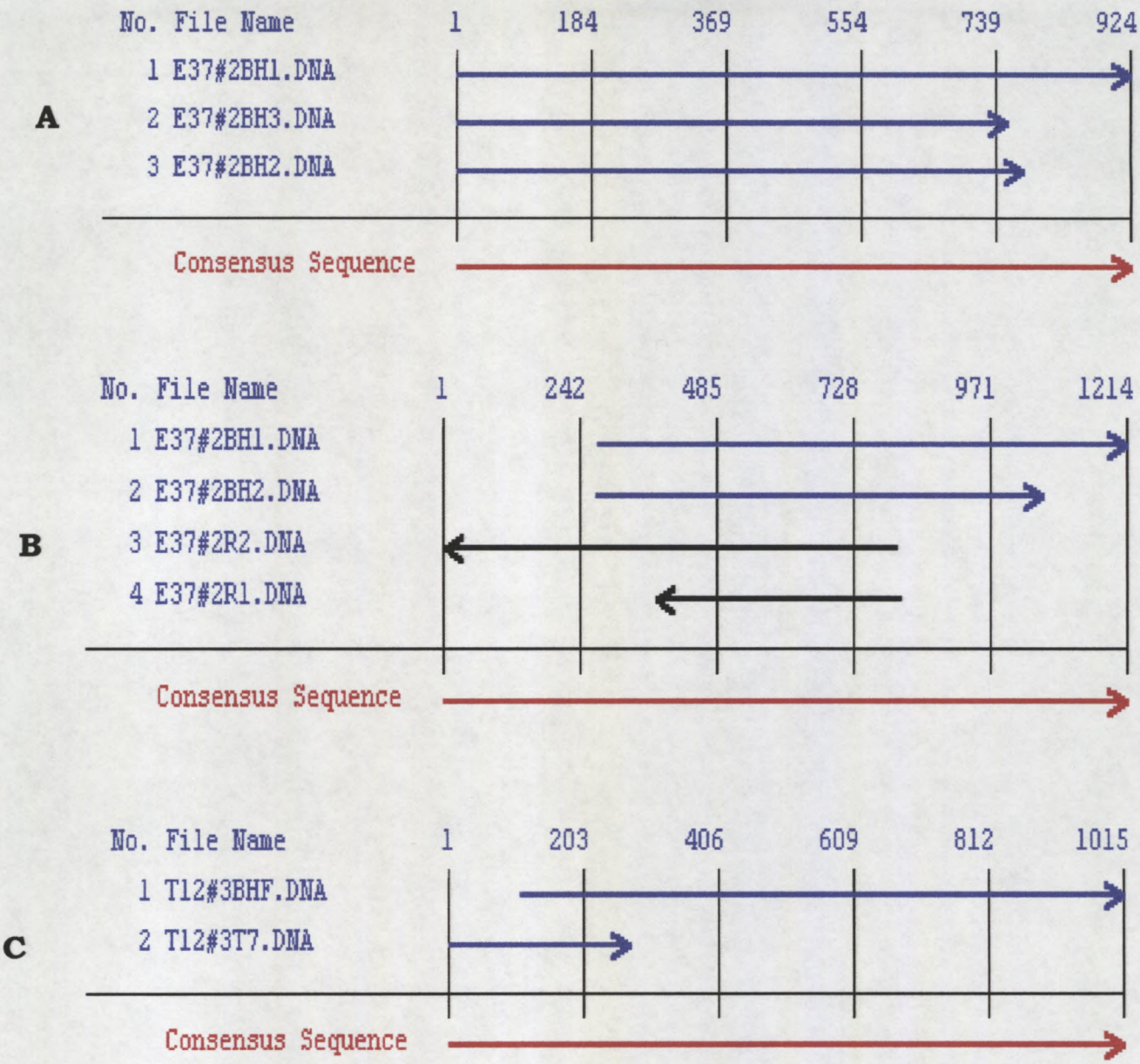



**Figure 3.16.** Diagrammatic representation of typical generation of consensus sequences from the primary overlap alignment of sequence repetitions of the 5' region of the cloned 1400bp CP fragment. The alignment was generated using a 30bp overlap region and 70% minimum homology value in the Contig Manager function of the sequence analysis programme DNASIS v2.1, 1991 (Hitachi). **A** = consensus sequence determination of three sequence repetitions using the primer BHf for clone #2 containing the 1400bp CP fragment derived from Eston variety NCo376; **B** = alignment of both forward and reverse sequences obtained using the forward primer BHf and reverse primer GSfINV of clone #2 from Eston variety NCo376; **C** = consensus sequence determination of sequences obtained using the primer BHf and T7 for clone #3 containing the 1400bp CP fragment derived from Tongaat variety N12.

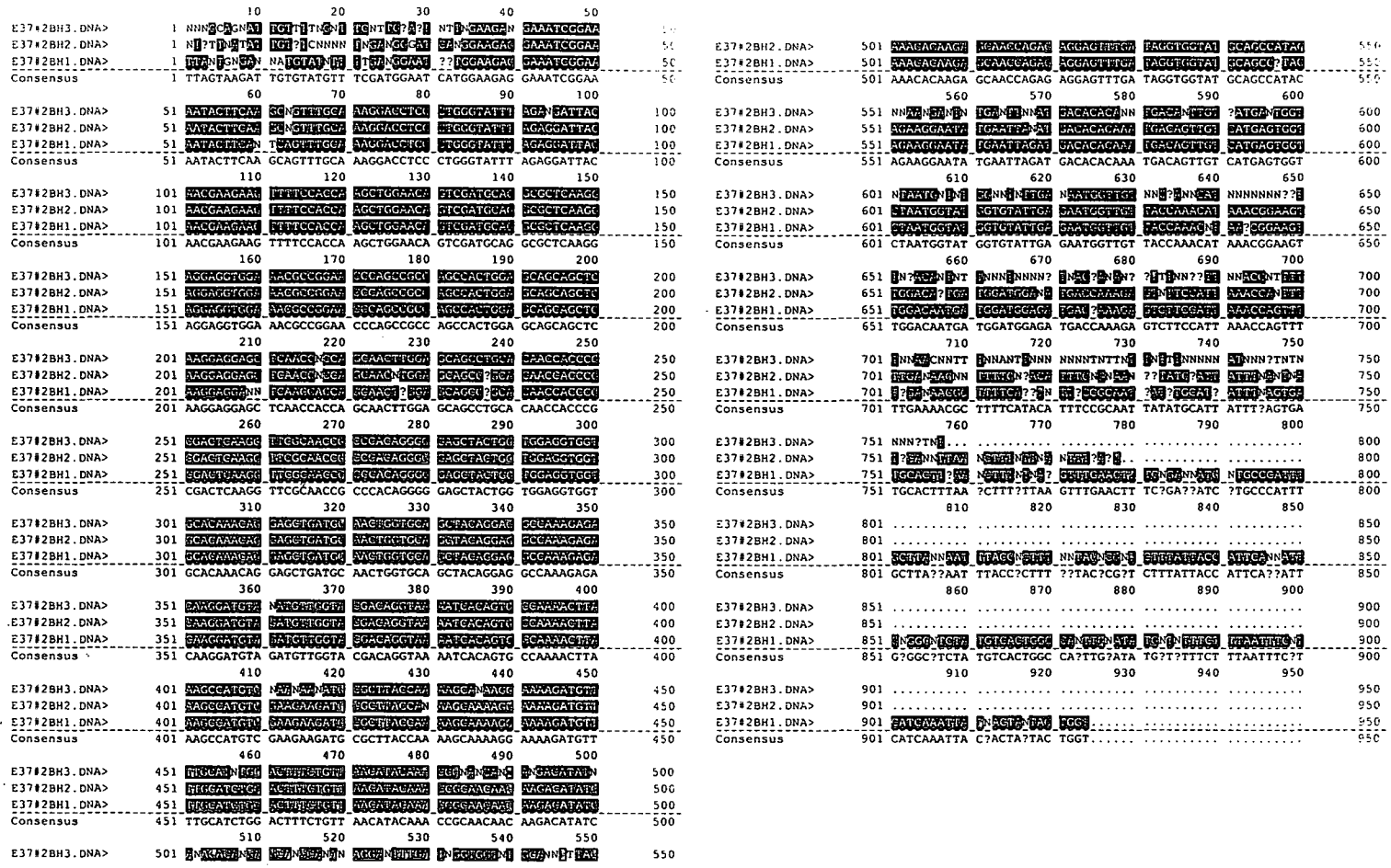


**Figure 3.17**. Consensus sequence of three sequences using BHf primer on clone #2 of the 1400bp CP fragment derived from Eston variety NCo376. Consensus sequence was generated using the Contig Manager function of DNASIS (Hitachi, 1991) using an overlap region of 30bp and a 70% minimum homology. Matching bases are highlighted and gaps are represented by question marks.

## 3.8 Sequence Comparisons of SCMV Coat Protein Fragments for Strain Identification

Since the primary aim of the study was to identify the prevalent strain(s) in the industry using coat protein gene sequences, it was necessary to compare the sequences obtained with those of reference SCMV strains in international databases. Available consensus sequences and individual sequences obtained for the clones from each region and its respective variety were submitted individually for comparison at the nucleotide level using the BLASTN (Altschul *et al.*, 1990) search programme. Searches were conducted against sequences contained in the nr database of the NCBI (see section 1.7.5) using an expect value of 1 which is used for stringent alignments. Reported alignments with bit scores above 80 were considered to have significant homologies. Sequence search results obtained using the BLASTN programme are shown in Table 3.9

Significance scores reported ranged between 82 and 1043 as seen in Table 3.9, with the average being 754. The number of identically matched bases in each of the reported alignments is shown and expressed as a percentage. Percentages of identity ranged between 91 and 98% as seen in Table 3.9, with an average of 94%. The organism most frequently identified was strain E of SCMV and strain D the next most frequent. SCMV strain B was also found to have high homology to the sequence obtained for T12#3 using the T7 primer. However, this was disregarded as the sequence obtained for the same clone using the BHf primer had a higher significance score and was identified as being most homologous to SCMV strain E. Most of the clones with scores below 80 were identified as being homologous to the SCMV CP as seen in

**Table 3.9.** Summary of results obtained from the comparison of partial 1400bp CP fragment nucleotide sequences with reference sequences in international databases. Sequences searches were conducted using the BLAST search programme (BLASTN) to search all non-redundant nucleotide databases.

| Query Sequence | | BLASTN Results | | |
|---|---|---|---|---|
| Name | Length (bp) | Organism | Score | Identity |
| $E_{12}$#1CON | 996 | SCMV - E | 1043 | 621/649 - 95% |
| $E_{12}$#2CON | 957 | SCMV - D | 585 | 447/478 - 93% |
| $E_{12}$#3CON | 936 | SCMV - E | 934 | 567/595 - 95% |
| $E_{376}$#1CON | 847 | SCMV - E | 864 | 619/667 - 92% |
| $E_{376}$#2CON | 899 | SCMV - E | 1001 | 616/649 - 94% |
| $E_{376}$#3CON | 979 | SCMV - E | 311 | 179/185 - 96% |
| $G_{16}$#1CON | 809 | SCMV - E | 985 | 641/681 - 94% |
| $G_{16}$#3CON | 945 | SCMV - E | 985 | 622/657 - 94% |
| $G_{376}$#1BHF | 739 | SCMV - E | 781 | 526/558 - 94% |
| $G_{376}$#2BHF | 1129 | NSH | NSH | NSH |
| $G_{376}$#3-T7 | 403 | SCMV CP | 42 | NSH |
| $I_{12}$#1-T7 | 656 | SCMV CP | 42 | NSH |
| $I_{12}$#2-T7 | 682 | SCMV CP | 42 | NSH |
| $I_{12}$#3-T7 | 739 | SCMV CP | 44 | NSH |
| $ME_{376}$#4-T7 | 728 | Cloning vector pKS145 MCS | 1005 | 540/549 - 98% |
| $P_{19}$#1CON | 815 | SCMV - E | 990 | 608/642 - 94% |
| $P_{19}$#2-BHF | 130 | NSH | NSH | NSH |
| $P_{19}$#2-T7 | 133 | NSH | NSH | NSH |
| $P_{19}$#3CON | 782 | SCMV - E | 1006 | 636/673 - 94% |
| $P_{25}$#1CON | 849 | SCMV - E | 996 | 630/667 - 94% |
| $P_{25}$#2CON | 826 | SCMV - E | 1041 | 643/680 - 94% |
| $P_{25}$#3-BHF | 904 | NSH | NSH | NSH |
| $PS_{376}$#1-T7 | 335 | SCMV CP | 42 | NSH |
| $PS_{376}$#2-T7 | 288 | SCMV CP | 42 | NSH |
| $PS_{376}$#3-T7 | 612 | NSH | NSH | NSH |
| $T_{12}$#1-BHF | 748 | SCMV - E | 769 | 582/638 - 91% |
| $T_{12}$#2-BHF | 206 | SCMV - D | 82 | 46/48 - 95% |
| $T_{12}$#2-T7 | 363 | SCMV - D | 549 | 292/297 - 98% |
| $T_{12}$#3-BHF | 873 | SCMV - E | 932 | 621/666 - 93% |
| $T_{12}$#3-T7 | 265 | SCMV - B | 341 | 187/193 - 96% |
| $T_{376}$#1-T7 | 388 | SCMV CP | 42 | NSH |
| $T_{376}$#2-T7 | 145 | SCMV - D | 145 | 84/87 - 96% |
| $T_{376}$#3-T7 | 432 | SCMV CP | 42 | NSH |

NSH = no significant homology

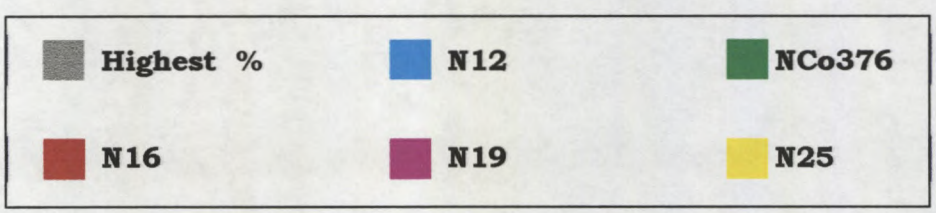
Table 3.9. Interestingly one of the sequences obtained for clone $ME_{376}$#4, showed very high homology to the cloning vector pKS145 indicating that the 1400bp CP fragment was not located in the multiple cloning site of the vector.

The CP nucleotide sequences of the plasmids were also compared to the five strains of SCMV (SCMV-A, -B, -D, -E and -SC) using the Homology Search function of the DNASIS sequence analysis programme. The nucleotide sequences of the five strains were obtained from the GenBank sequence database using the following accession numbers: U57354 (SCMV-A), U57355 (SCMV-B), U57356 (SCMV-D), U57357 (SCMV-E) and D00958 (SCMV-SC). The Homology Search function in DNASIS uses the Lipman-Pearson FASTA algorithm to calculate the similarity between two sequences. Scores obtained are expressed as a percentage indicating the degree of similarity between the two sequences. Table 3.10 shows the percent similarity obtained for the sequence of each of the cloned CP fragments compared with that of the SCMV strains. The highest percent homology obtained for each sequence is shaded in dark grey. Pairwise alignment of the sequences revealed a slightly different assignment of SCMV strains with the CP fragment sequences. Most of the CP sequences had the highest homology to SCMV-SC followed by SCMV-E as seen in Table 3.10. An equal number of sequences had the highest homology to SCMV-B and SCMV-D. Only one was found to be most homologous to SCMV-A. The high homology of a large number CP sequence of with SCMV-E is in agreement with the results obtained using the BLASTN programme.

Sequences obtained from each of varieties used in this study were colour-coded as seen in Table 3.10 to determine the possible relationship between the different varieties and the SCMV strains. No association was observed between a specific variety and a strain of SCMV. There also did not appear to be any significant correlation between a specific region and a strain of SCMV.

**Table 3.10.** Percentage sequence homology calculated from the pairwise alignment of each CP fragment sequence with the nucleotide sequence of the five SCMV strains: SCMV-A (U57354); SCMV-B (U57355); SCMV-D (U57356); SCMV-E (U57357) and SCMV-SC (D00958) using the homology search function in the sequence analysis programme DNASIS v2.1. (GenBank database accession numbers for each SCMV strain is in parenthesis). Significance of shaded areas is explained in the key at the bottom of the table.

| Region | Query Sequence | SCMV-A | SCMV-B | SCMV-D | SCMV-E | SCMV-SC |
|---|---|---|---|---|---|---|
| Eston | $E_{12}$#1CON | 69.5 | 87.6 | 80.4 | 81.7 | 81.5 |
| | $E_{12}$#2CON | 58.1 | 58.0 | 59.2 | 56.7 | 57.3 |
| | $E_{12}$#3CON | 68.8 | 66.8 | 69.4 | 85.1 | 84.3 |
| | $E_{376}$#1CON | 71.6 | 74.1 | 74.8 | 89.8 | 89.7 |
| | $E_{376}$#2CON | 68.5 | 84.2 | 85.2 | 84.3 | 84.0 |
| | $E_{376}$#3CON | 62.9 | 56.7 | 60.3 | 64.2 | 64.4 |
| Glendale | $G_{16}$#1CON | 68.7 | 72.8 | 74.1 | 89.4 | 88.6 |
| | $G_{16}$#3CON | 68.4 | 71.2 | 72.2 | 85.2 | 84.8 |
| | $G_{376}$#3-T7 | 45.5 | 46.3 | 45.5 | 45.5 | 46.8 |
| Ixopo | $I_{12}$#1-T7 | 44.2 | 44.0 | 45.4 | 45.3 | 42.8 |
| | $I_{12}$#2-T7 | 45.3 | 42.7 | 41.3 | 40.7 | 45.9 |
| | $I_{12}$#3-T7 | 43.3 | 45.8 | 40.8 | 43.9 | 43.2 |
| Mount Edgecombe | $ME_{376}$#4-T7 | 39.4 | 40.4 | 39.5 | 40.2 | 41.2 |
| Pongola | $P_{19}$#1CON | 60.2 | 80.7 | 81.0 | 82.3 | 83.0 |
| | $P_{19}$#3CON | 63.0 | 87.2 | 87.7 | 88.9 | 89.1 |
| | $P_{25}$#1CON | 62.9 | 82.4 | 82.7 | 84.0 | 84.0 |
| | $P_{25}$#2CON | 63.2 | 84.2 | 84.8 | 85.7 | 86.0 |
| | $P_{25}$#3-BHF | 27.1 | 27.9 | 26.9 | 27.9 | 28.4 |
| Port Shepstone | $PS_{376}$#1-T7 | 42.8 | 42.8 | - | 43.0 | 42.0 |
| | $PS_{376}$#2-T7 | 44.4 | 41.9 | 45.0 | 45.4 | 43.4 |
| | $PS_{376}$#3-T7 | 43.4 | 47.8 | 44.9 | 47.6 | 44.9 |
| Tongaat | $T_{12}$#1-BHF | 62.8 | 84.0 | 84.2 | 85.6 | 85.7 |
| | $T_{12}$#2-BHF | 61.5 | 62.5 | 64.1 | 63.0 | 63.0 |
| | $T_{12}$#2-T7 | 49.0 | 49.0 | 49.7 | 50.0 | 49.7 |
| | $T_{12}$#3-BHF | 66.8 | 82.3 | 82.6 | 83.9 | 84.2 |
| | $T_{12}$#3-T7 | 87.1 | 92.7 | 91.4 | 89.5 | 53.2 |
| | $T_{376}$#1-T7 | 63.6 | 51.1 | 55.4 | 46.4 | 42.4 |
| | $T_{376}$#2-T7 | 48.5 | 48.0 | 49.0 | 48.0 | 50.0 |
| | $T_{376}$#3-T7 | 44.3 | 49.2 | 44.7 | 45.5 | 44.1 |



Instead, a general spread across several strains was observed for both the varieties and regions. There were however, CP sequences from two regions that were all homologous to the same SCMV strain. These were the plasmids from Pongola and Port Shepstone that were most homologous to SCMV-SC and SCMV-E respectively.

As the above comparisons were done using consensus sequences and individual sequences which differed in length, it was difficult to observe any particular strain relationships based on these sequences. It was therefore necessary to use the same region of the CP gene for all further comparisons. As a result, a 450bp region from the start of the CP gene obtained from consensus sequences was used. This region contains the hypervariable region of the CP gene where most strain variation is known to occur and would be sufficient for the strain identification of each sequence (Mirkov, pers comm). Analysis was therefore done using the following plasmids: $E_{12}$#1-3; $E_{376}$#1-3; $G_{16}$#1,3; $P_{19}$#1,3 and $P_{25}$#1,2. These plasmids were representative of all varieties used as well as from several widely separated and geographically distinct regions. Information obtained from further comparative analysis of these sequences would be indicative of possible varietal or regional effects on strain prevalence.

The 450bp region from the consensus sequences of all the above mentioned CP fragment sequences and the five SCMV strains was saved with an AL (denoting **al**ignment sequence) at the end of the plasmid or strain name. The Multiple Alignment function of DNASIS was used to compare the 450bp region

of the selection of plasmids mentioned above and the five SCMV strains. A gap penalty of 5 and K-tuple value of 4 was used in all multiple alignments. The Multiple Alignment function uses the Higgins-Sharp algorithm which is based on an assumed evolutionary relationship between the sequences. Aligned 450bp sequences are shown in Figure 3.18.

All the sequences had a high degree of homology in the first 110bp with only a few single base substitutions occurring. However, the following 150bp appear to contain the majority of deletions and substitutions with the largest deletion occurring between bases 207 and 234 of SCMV-A. The phylogenetic relationship between these sequences is shown in Figure 3.19.

A high degree of homology is evident between most of the CP sequences and that of the SCMV strains with an average sequence identity of 87%. Most of the CP sequences cluster with SCMV-SC while the other SCMV strains form a separate cluster and the remaining sequences are separate from the two clusters. No sequence homogeneity is observed between CP sequences from the same region and variety as previously observed (Goodman *et al.*, 1998). This might be explained by the difference in sequence length used for the comparisons since the 450bp region in this comparison is almost twice the length of that used previously. Also, CP sequences did not cluster according to regional nor varietal origin indicating the possibility of the presence of several strains of SCMV in each region and variety.

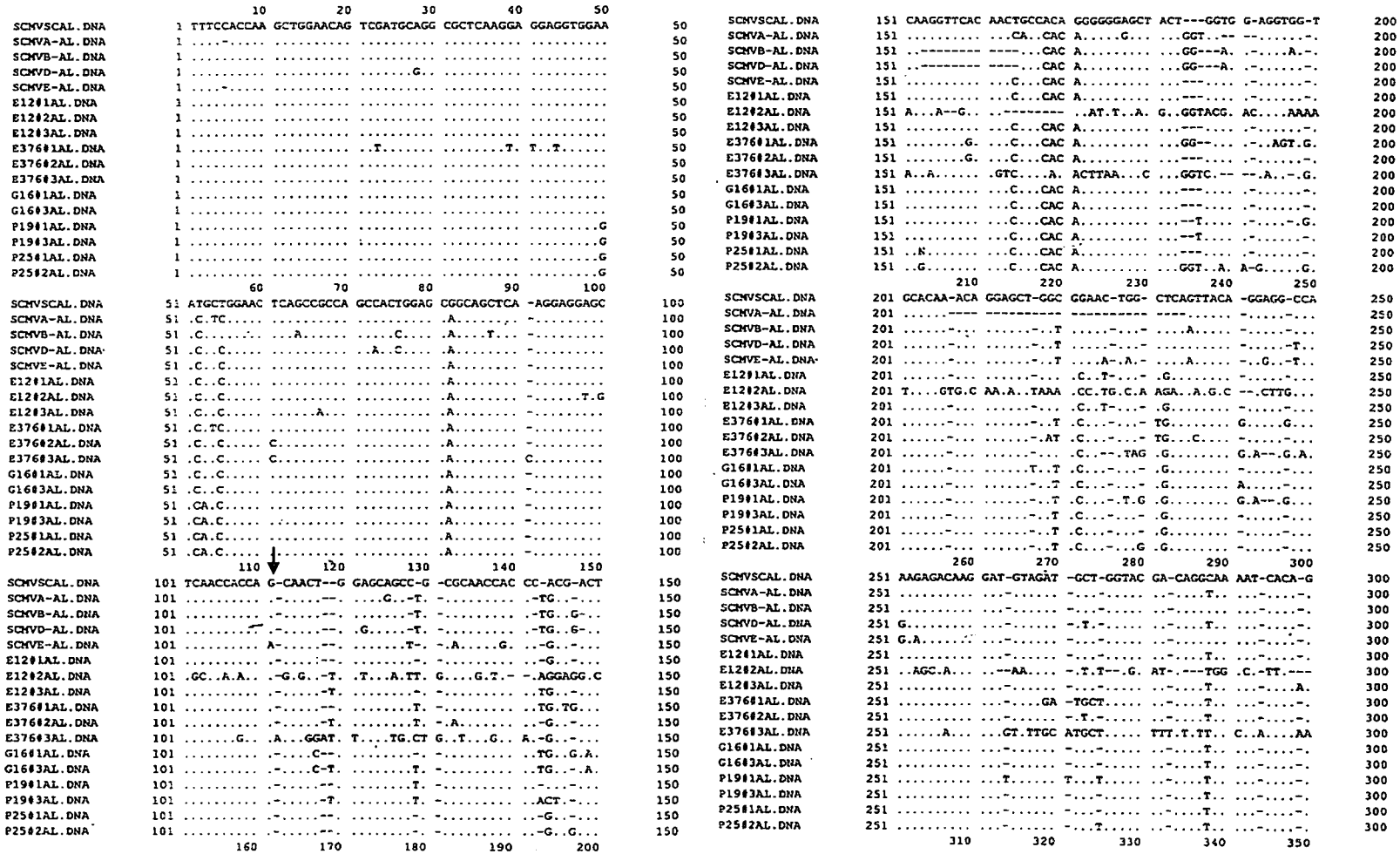


**Figure 3.18** Multiple alignment of a 450bp region from the start of the coat protein gene of five strains of SCMV and CP fragment sequences from several regions and varieties. Sequences were aligned using a gap penalty of 5 and a K-tuple value of 4 in the Multiple Alignment function of DNASIS (Hitachi, 1991). A 300bp section of the aligned sequences is represented in the figure. Matching bases are represented by dots and gaps are represented by dashes. The arrow indicates the start of the region of high variation.

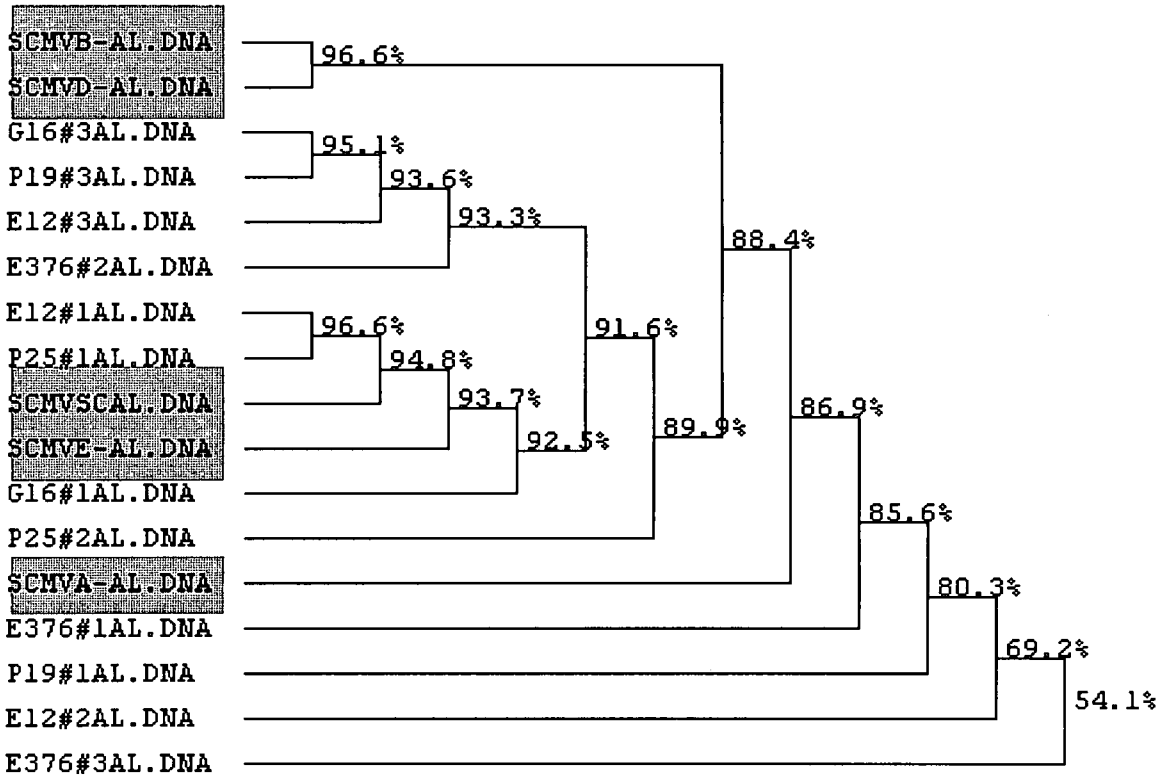



**Figure 3.19**. A dendrogram showing the relatedness of CP fragment sequences from several regions and varieties with the five SCMV strains. A 450bp nucleotide sequence from the start of the CP gene was used to generate the dendrogram. The degree of nucleotide sequence similarity between the sequences is expressed as a percentage. Reference SCMV strains used in the pairwise comparisons are highlighted.

All preliminary comparisons were done using nucleotide sequences and it was considered desirable to assess the strain associations using inferred amino acid sequences. The 450bp region of the CP sequences were similarly compared to international databases for strain identity. Sequences were submitted for comparison at the amino acid level using all six conceptual reading frames in the BLASTX (Altschul *et al.*, 1997) search programme. Searches were conducted against sequences contained in the nr peptide database of the NCBI using an expect value of 1. Reported alignments with bit scores above 80 were considered to have significant homologies. Sequence search results obtained using the BLASTX programme are shown in Table 3.11.

**Table 3.11.** Summary of results obtained from the sequence comparison of the first 150 amino acids from the start of the CP gene derived from several regions and varieties with reference sequences in international databases. Sequences searches were conducted using the BLAST search programme (BLASTX) to search all non-redundant peptide databases.

| Query Sequence | | BLASTX Results | | | |
|---|---|---|---|---|---|
| **Region** | **Name** | **Organism** | **Score** | **Identity** | **Positives** |
| Eston N12 | $E_{12}$#1AL | SCMV Polyprotein | 116 | 65/147 - 44% | 69/147 - 48% |
| | $E_{12}$#2 AL | SCMV CP | 71 | 44/83 - 53% | 50/83 - 60% |
| | $E_{12}$#3 AL | SCMV CP | 118 | 56/60 - 93% | 58/60 - 98% |
| Eston NCo376 | $E_{376}$#1 AL | SCMV CP | 131 | | |
| | $E_{376}$#2 AL | SCMV CP | 147 | 74/108 - 68% | 74/108 - 68% |
| | $E_{376}$#3 AL | SCMV CP | 50 | 33/65 - 50% | 38/65 - 57% |
| Glendale N16 | $G_{16}$#1 AL | SCMV CP | 166 | 80/81 - 98% | 81/81 - 99% |
| | $G_{16}$#3 AL | SCMV CP | 155 | 74/78 - 94% | 77/78 - 97% |
| Pongola N19 | $P_{19}$#1 AL | SCMV CP | 69 | | |
| | $P_{19}$#3 AL | SCMV CP | 146 | 70/71 - 98% | 71/71 - 99% |
| Pongola N25 | $P_{25}$#1 AL | SCMV CP | 141 | 67/70 - 95% | 69/70 - 97% |
| | $P_{25}$#2 AL | SCMV CP | 131 | | |

All the sequences were found to have high homology to SCMV CP except one ($E_{12}$#1) which was homologous to the polyprotein. Three of the sequences did not have significant bit scores (shaded areas in the Table 3.11) but were most homologous to the SCMV CP.

As no strain identification was possible using these results, further analysis of the sequences at the amino acid level was necessary. Nucleotide sequences were translated into three reading frames and aligned using the Multiple Alignment function of DNASIS. No significant alignment was obtained with the plasmid sequences and that of the SCMV strains (results not shown).

However, all the CP fragment sequences did possess the aspartic acid-alanine-glycine (DAG) triplet codon which has been found to be required for aphid transmissibility (Atreya *et al.*, 1995). Direct phylogenetic comparison of the CP sequences with the amino acid-based dendrogram published by Yang & Mirkov, 1997 was not possible as those authors used the conserved core region and the complete CP gene in their phylogenetic comparison.

## 3.9 Restriction Fragment Length Polymorphism (RFLP) Analysis for Strain Identification

During the course of this study, an article on SCMV strain differentiation using a PCR-based RFLP technique was published (Yang & Mirkov, 1997). The technique developed by those authors presented a novel approach to the differentiation of SCMV strains by specifically amplifying a polymorphic region of the CP gene. Digestion of this amplified region with specific restriction enzymes results in cleavage pattern that permit differentiation of four SCMV strains (SCMV-A, -B, -D and -E) (Yang & Mirkov, 1997). This method was applied to the SCMV samples used in this study as a means of confirming the sequence-based strain allocations observed from sequence analysis of the CP gene fragments.

Using the Yang & Mirkov (1997) primer pair SCMV-F3/SCMV-R3, specific for amplification of the CP gene of SCMV strains in the SCMV subgroup, an expected fragment of approximately 900bp was amplified from cDNA obtained from all the SCMV-infected source samples as seen in Figure 3.20.

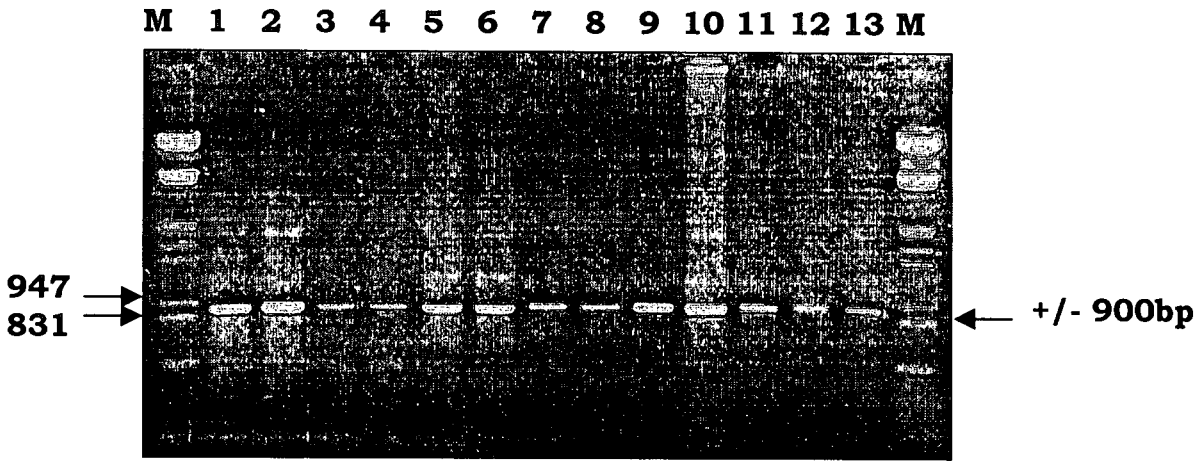



**Figure 3.20**. Agarose gel electrophoretic separation of PCR products of an approximately 900bp fragment used for RFLP analysis in SCMV strain differentiation. Template samples represent all regions and their respective varieties using the SCMV-F3/SCMV-R3 primer pair of Yang & Mirkov (1997). M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested), Lanes 1-2 = Eston N12, NCo376; 3-4 = Glendale N16, NCo376; 5-6 = Ixopo N12, NCo376; 7 = Mount Edgecombe NCo376; 8 = Melmoth N12; 9-10 Pongola N19, N25; 11 = Port Shepstone NCo376; 12-13 = Tongaat N12, NCo376.

Although some non-specific amplification was visible for some of the samples, it was in very low concentrations as seen in Figure 3.20 lanes 2, 5, 6 and 10. This however, was not of serious concern since the fragment DNA for most samples was sufficient to produce unambiguous restriction profiles.

### 3.9.1 *Hinf* I Restriction Profiles

Based on Yang & Mirkov's (1997) findings, restriction enzyme *Hinf* I was used to identify the source samples as SCMV strains A, B/D and E/SC based on the expected restriction profiles of the approximately 900bp fragment as shown in Table 3.12. Strains B/D and E/SC produce identical fragment lengths when restricted with *Hinf* I and cannot be differentiated from each other using this enzyme. It does however serve to differentiate the samples into three categories of strain A, strain B/D and strain E/SC.

**Table 3.12.** Expected restriction profile of *Hinf* I digested 900bp fragment of five strains in the SCMV subgroup as determined by the restriction enzyme search function of the DNASIS software programme.

| SCMV Strain | No. of Sites | Expected Fragment Sizes (bp) |
|---|---|---|
| A | 2 | 138; 331; 401 |
| B | 1 | 331; 555 |
| D | 1 | 331; 555 |
| E | 1 | 139; 759 |
| SC | 1 | 139; 759 |

The 900bp fragment from all the source samples were first restricted with *Hinf* I enzyme for broad strain identification and the restriction profiles obtained are shown in Figure 3.21.

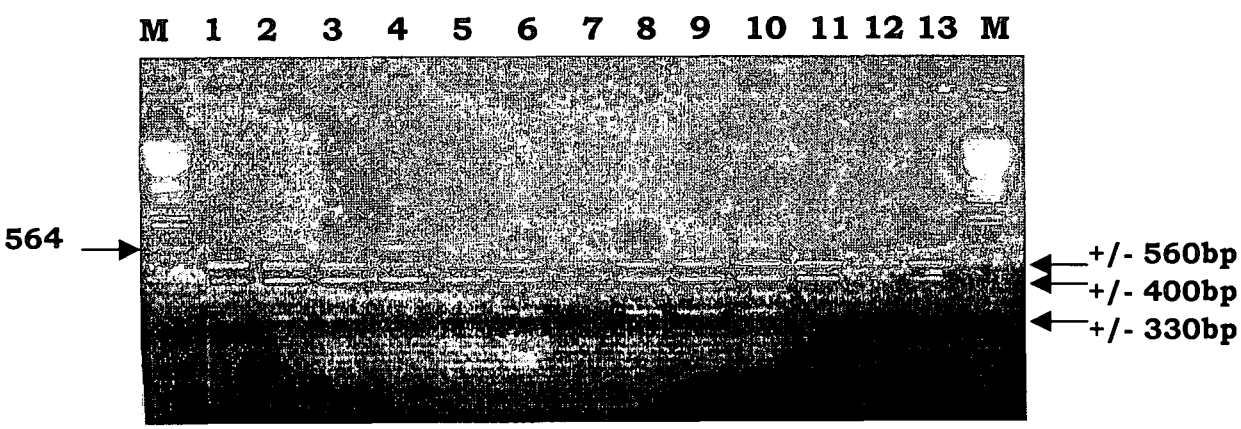



**Figure 3.21.** Restriction patterns after *Hinf* I digestion of the 900bp SCMV-F3/SCMV-R3 fragment from all SCMV infected source samples. M = Molecular weight marker III (λ DNA *EcoR* I/*Hind* III digested), Lanes 1-2 = Eston N12, NCo376; 3-4 = Glendale N16, NCo376; 5-6 = Ixopo N12, NCo376; 7 = Mount Edgecombe NCo376; 8 = Melmoth N12; 9-10 Pongola N19, N25; 11 = Port Shepstone NCo376; 12-13 = Tongaat N12, NCo376.

All the samples were susceptible to digestion with *Hinf* I restriction enzyme. Several samples produced restriction profiles similar to that of strain A, i.e., fragment sizes of approximately 400bp, 330bp and 140bp. These samples are Eston variety N12; Ixopo variety N12; Mount Edgecombe variety NCo376 and Tongaat variety N12 as seen in Figure 3.21 lanes 1, 5, 7 and 12 respectively. The remainder of the samples produced restriction profiles similar to strain A, but containing an additional fragment, i.e. fragment sizes of approximately 550bp, 400bp, 330bp and 140bp. This would indicate the presence of a mixture of strain A and strain B/D within these samples. A second restriction using a different enzyme was therefore necessary to further differentiate the samples.

3.9.2 *Taq* I Restriction Analysis

The restriction enzyme *Taq* I used by Yang & Mirkov (1997) was found to produce restriction profiles of the 900bp fragment that could be used to differentiate between SCMV strain A, strains B and D and between strains E and SC. The expected fragment sizes produced by *Taq* I restriction of the five SCMV strains are shown in Table 3.13. Restriction profiles produced by *Taq* I restriction of the samples used in this study are shown in Figure 3.22.

All the samples were susceptible to digestion with *Taq* I restriction enzyme. Several samples produced restriction profiles similar to that of strain D, i.e. fragment sizes of approximately 240bp, 260bp and 360bp. These are Eston varieties N12 and NCo376, Glendale variety N16, Ixopo variety N12, Pongola varieties N19 and N25, Port Shepstone variety NCo376 and Tongaat varieties

**Table 3.13.** Expected restriction profile of *Taq* I digested 900bp fragment of five strains in the SCMV subgroup as determined by the restriction enzyme search function of the DNASIS software programme.

| SCMV Strain | No. of Sites | Expected Fragment Sizes (bp) |
|---|---|---|
| A | 2 | 20; 241; 612 |
| B | 4 | 21; 87; 154; 263; 361 |
| D | 3 | 21; 241; 263; 361 |
| E | 3 | 21; 241; 275; 361 |
| SC | 4 | 21; 120; 121; 275; 361 |

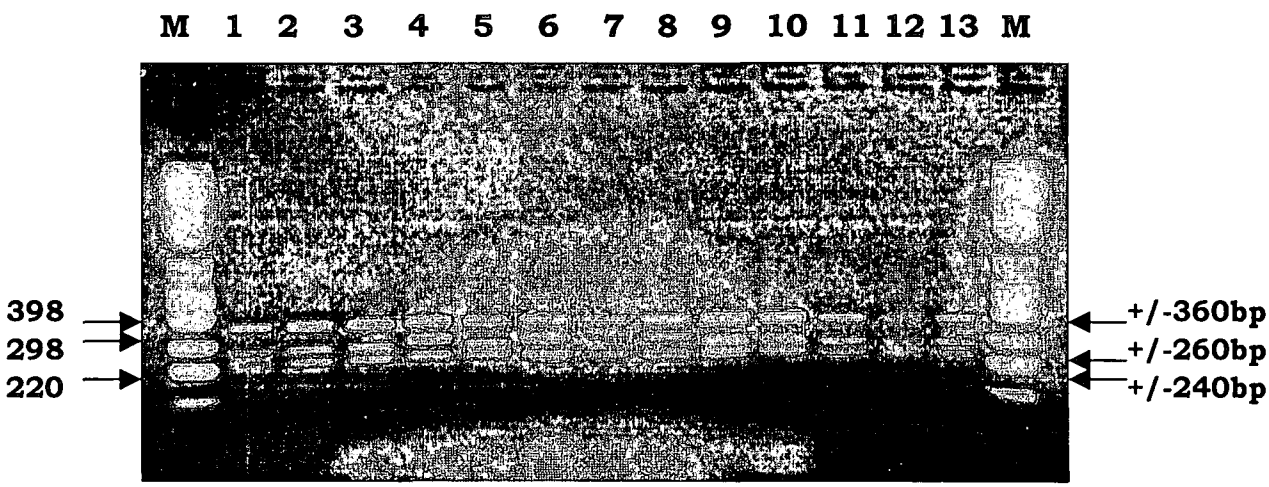



**Figure 3.22.** Restriction patterns after *Taq* I digestion of the 900bp from all SCMV infected source samples. M = Molecular weight marker VI (pBR328 DNA *Bgl* I/*Hinf* I digested), Lanes 1-2 = Eston N12, NCo376; 3-4 = Glendale N16, NCo376; 5-6 = Ixopo N12, NCo376; 7 = Mount Edgecombe NCo376; 8 = Melmoth N12; 9-10 Pongola N19, N25; 11 = Port Shepstone NCo376; 12-13 = Tongaat N12, NCo376.

N12 and NCo376 as seen in Figure 3.22 lanes 1 - 3, 5 and 9 - 13 respectively. Four of the samples produced restriction profiles similar to that of strain E, i.e. fragment sizes of approximately 240bp, 280bp and 360bp. These samples were Glendale variety NCo376, Ixopo variety NCo376, Mount Edgecombe

variety NCo376 and Melmoth N12 as seen in lanes 4, 6, 7 and 8 respectively (Figure 3.22). An extra fragment of approximately 450bp was visible in lane 6 which may have been as a result of incomplete digestion of the fragment. The restriction profiles produced by *Taq* I digestion of the source samples revealed more unambiguously the strain identification of each sample than that of *Hinf* I as shown by the summarised results in Table 3.14.

Results obtained using the RFLP technique revealed the presence of two SCMV strains, SCMV-D and SCMV-E in the various regions and varieties used in this study. These results are not in complete agreement with those obtained using sequence comparisons of CP gene fragments which revealed a strong association with SCMV-SC as well as SCMV-E and SCMV-D as shown in Tables 3.9 and 3.10. However, the RFLP strain identification appears to be simple and less ambiguous than that obtained from sequence comparisons.

**Table 3.14.** A summary of strain identification results restriction fragment length polymorphism analysis of the approximately 900bp SCMV-F3/SCMV-R3 fragment derived from the SCMV infected source samples.

| **Region** | **Variety** | **SCMV Strain ID produced by restriction enzyme cleavage patterns using** | |
|---|---|---|---|
| | | ***Hinf* I** | ***Taq* I** |
| Eston | N12 | A | D |
| | NCo376 | A + B/D | D |
| Glendale | N16 | A + B/D | D |
| | NCo376 | A + B/D | E |
| Ixopo | N12 | A | D |
| | NCo376 | A + B/D | E |
| Mount Edgecombe | NCo376 | A | E |
| Melmoth | N12 | A + B/D | E |
| Pongola | N19 | A + B/D | D |
| | N25 | A + B/D | D |
| Port Shepstone | NCo376 | A + B/D | D |
| Tongaat | N12 | A | D |
| | NCo376 | A + B/D | D |

# CHAPTER 4

## DISCUSSION

SCMV CP sequences were specifically amplified by PCR from infected sugarcane leaf material from several varieties and geographically distinct regions within the South African sugar industry. Using SCMV-specific primer pairs, fragment sizes of 884 and 1400bp containing the N-terminal (variable) region of the CP gene were amplified from SCMV-infected material from all regions and varieties sampled. Semi-nested PCR and re-amplification of the fragments confirmed the specificity of the amplification and verified the findings of Huckett & Botha (1996) for the 884bp fragment. The specificity of amplification of the 1400bp fragment was confirmed by semi-nested PCR but not by re-amplification since this proved technically recalcitrant. CP fragments were cloned and the sequences of the inserted fragments determined by cycle-sequencing and automated analysis. Strain identification of the CP fragments was achieved by comparison of the nucleotide sequences with those in international sequence databases using the BLASTN search programme. Prevalent strains identified by BLASTN searches were found to be SCMV-E and SCMV-D. Pairwise comparison of the nucleotide sequences of the South African isolates with those of five SCMV strains: A, B, D, E and SC indicated SCMV-SC to be the most prevalent strain amongst the isolates, followed by SCMV-E and SCMV-D. These strains are all closely related and cluster together, with SCMV-E and -D more closely related than -SC as shown in Figure 1.5. Comparison of the CP sequences at the amino acid level was more ambiguous with a general spread across all five strains. Strain

identification of the South African isolates was also achieved by RFLP analysis of SCMV CP PCR products using a recent method developed by Yang & Mirkov (1997) which revealed SCMV-D and SCMV-E as the most prevalent strains. Results using this approach compared very favourably with the strain identities assigned by the BLASTN search results.

Initial sequence comparisons undertaken in a preliminary study had suggested that there was little sequence variation within the CP gene among South African (SA) isolates from a small number of widely separated geographic areas and their varieties (Goodman *et al.*, 1998). However, the present study in which isolates were more widely sampled and a larger region of the CP gene was used for sequence comparison (450bp as opposed to the 248bp in the preliminary study) revealed more variation than was first recognised. Phylogenetic analysis of several clones from geographically distinct regions and varieties revealed an approximately 13% difference at the nucleotide level (Figure 3.19). This is compatible with the 12% observed by Bateson *et al.* (1994) for Australian isolates of papaya ringspot potyvirus (PRSV). Sequence difference observed by those authors, however, was based on amino acid sequences of the entire CP gene. Interestingly, Handley *et al.* (1998) observed only 3.3% sequence difference at the nucleotide level between local SCMV isolates and those from the United States and South Africa. However, the low sequence difference observed by those authors is based on comparison of the entire CP gene sequences. Thus an average of 87% sequence identity was shared amongst the various SA isolates, but this figure would probably increase if the entire CP gene was used for comparison. This

would then be compatible with the findings of Shukla & Ward (1988) that strains of the same virus share 90-99% CP sequence identities, while distinct viruses share only 38-71% identity. An additional sequence identity cluster of 75-89% identified by McKern *et al.* (1991) has also been regarded as that of distinct strains by Shukla *et al.* (1992). These figures are based on amino acid sequence comparison of the entire CP gene, which might result in higher similarities if the SA isolates are compared in this manner. However, Collmer *et al.* (1996) observed no significant change in the phylogenetic relationships between strains of bean common mosaic potyvirus (BCMV) when only the variable CP N-terminus was used for comparison as opposed to the entire CP gene. Overall, the majority of SA isolates clustered with SCMV strains SC, E and D, while two isolates, $E_{12}$#2 and $E_{376}$#3, appeared to be quite distinct with similarity percentages of only 69.2 and 54.1 respectively to the five SCMV strains and other local isolates (Figure 3.19). Interestingly, there was a significant difference of approximately 40% between two clones obtained from the same sample from variety NCo376 at Eston ($E_{376}$#2 and $E_{376}$#3), as shown in Figure 3.19, indicating the existence of multi-strain infection in sugarcane. This phenomenon has previously been shown to occur in sugarcane by Gillaspie & Koike (1973) when they artificially inoculated varieties of sugarcane with mixtures of SCMV and MDMV strains. More recently, Candresse *et al.* (1998) identified mixed infections of plum pox potyvirus (PPV) serotypes D and M in peach isolates by PCR analysis.

Using several methods in this study for strain identification, as mentioned above the most prevalent strains were found to be SCMV-E, SCMV-D and to

a lesser extent SCMV-SC. This is the first molecular analysis of strain variability within the SA sugar industry and the first time that SCMV-E has been associated with sugarcane in this country. Previous strain identification (Abbott & Stokes, 1966; Gillaspie & Mock, 1987) found SCMV-A and SCMV-D to be the prevalent strains in South Africa at that time. However, strain identification by those authors was through symptom expression on differential hosts, which has been found to be inconsistent due to climatic and environmental influences (Kondaiah & Nayudu, 1984). Studies by those authors were also based on a limited sample selection and were thus not able to demonstrate the extent of strain variation in South Africa. Differences in the strain identification as found in this study compared to previous studies is unlikely to be due to differences in the varieties tested, since the data indicates clearly that no association exists between strain prevalence and specific varieties or regions. This is in agreement with the findings of McNeil *et al.* (1996) that showed no correlation between strains of wheat streak mosaic virus (WSMV), a member of the *Potyviridae* family (*Rymovirus*), and geographic origin.

Molecular analysis of CP gene sequences has been used to differentiate between strains of other potyviruses such as PRSV (Bateson *et al.*, 1994; Jain *et al.*, 1998), zucchini yellow mosaic potyvirus (ZYMV) (Thomson *et al.*, 1995) and turnip mosaic potyvirus (TMV) (Stavolone *et al.*, 1998). ZYMV was distinguished from two closely related potyviruses (PRSV-W and watermelon mosaic potyvirus 2 [WMV-2]) using CP gene sequence comparisons and RFLP analysis (Thomson *et al.*, 1995). Azuki bean mosaic potyvirus (AzMV) was

found to be a strain of BCMV based on sequence identities between CP genes (Collmer et al., 1996). The work of those authors supports the findings of Shukla & Ward (1988) that strains of the same virus can be differentiated based on sequence identities of their CP genes and that this has proven to be a more reliable method of identifying strains of potyviruses than other conventional methods such as symptomatology.

Technical difficulties encountered using molecular analysis of CP genes of SCMV isolates in this study should in no way discredit the validity of the findings, although they do emphasise that cloning and sequencing is not a method of choice for routine diagnosis. The use of the 1400bp CP fragment without re-amplification prior to cloning is acceptable since the specificity of the amplification was confirmed by semi-nested PCR and subsequent sequencing, where this was possible. The inability to reamplify the 1400bp fragment may have been due to ineffective binding of the primers to the fragment since some primers require a certain number of bases upstream of the binding site in order to anneal stably. The aberrant nature of some recombinant clones may have been a result of rearrangement and possible partial excision of inserted viral sequences within the plasmids, since the PCR test for the presence of inserts in recombinant plasmids generally resulted in fragments of the expected sizes. Rearrangement may have occurred during the successive subculturing of the bacterial colonies containing the recombinant plasmids. Where sequence failure was experienced it was directly related to the aberrant nature of plasmids, with one exception, $P_{25}$#3. The original plasmid-encoded T7 priming site was still present for most of the

aberrant plasmid clones. Sequences obtained using the T7 primer revealed high homology to SCMV indicating the presence of partial CP fragments. The absence of the BHf priming site within the aberrant clones strongly supports partial excision of the inserted CP fragments. Sequence information obtained from normal clones provided sufficient indication of the identity of the prevalent strains in the industry. This sequence information obtained was also adequate to establish that no significant correlation could be made between strain prevalence and geographic or varietal origin of the isolates.

This study has shown that SCMV strain identification is possible using the nucleotide sequence of the N-terminal region of the CP gene. SCMV strain associations with the SA isolates established by sequence determination were confirmed with RFLP analysis. Use of the RFLP technique developed by Yang & Mirkov (1997) confirmed the findings of those authors of the specific amplification of a polymorphic region of the CP gene and the differentiation of the four SCMV strains (SCMV-A, -B, -D and -E) by unique restriction patterns. Sequence-based strain identification is not suitable as a diagnostic tool due to the high cost of sequence determination, the time required for analysis and the technical difficulties which may be encountered. However, RFLP analysis was proven in this study to be a simple and accurate method of valid strain identification which would be suitable as a diagnostic tool for the detection of SCMV infection of SA sugarcane. This method has the additional benefit of strain discrimination as opposed to simple PCR-based diagnostic tests such as those developed by Smith & van der Velde (1994) and Huckett & Botha (1996) which only identify SCMV infection.

Knowledge of the prevalent strain(s) of SCMV is also strategically important for application in achieving transgenic resistance to SCMV infection through genetic engineering. Variability amongst strain(s) / isolates of SCMV in the nucleotide or amino acid sequences of the CP gene of the challenge virus may have significant effect on the level and stability of trangenic resistance (Stark & Beachy, 1989). It is therefore important to have adequate information on the extent of the sequence variation that is found between strains of SCMV. The strains identified in this study as being most prevalent in the SA sugar industry, SCMV-D and SCMV-E, are closely related to each other as well as to the other SCMV strains with the exception of the maize dwarf B strain (SCMV-MDB) which appears to be quite distinct from the other SCMV strains as shown by Yang & Mirkov (1997). Thus, the use of any SCMV-E or SCMV-D sequence in genetically engineering sugarcane to confer CP-mediated resistance is likely to offer protection to transgenic sugarcane against all local SCMV strains.

A suitable diagnostic technique for the rapid strain-specific detection of SCMV would be most beneficial to the SA sugar industry's quarantine procedures. Using the Yang & Mirkov (1997) RFLP approach tested in this study, imported foreign sugarcane varieties as well as local varieties destined for export could then be rapidly screened for not only for the presence of SCMV but also the specific infecting strain. This technique could also be applied to investigations on the incidence of symptomless infection of SCMV in sugarcane as observed by Cronje et al. (1994) as well as for the screening of new varieties for susceptibility to mosaic.

A well established transgenic programme is in place at SASEX for the transformation of sugarcane. The findings of this study have facilitated the commencement of a study to transfer CP sequences from one of the local SCMV isolates to generate transgenic sugarcane with CP-mediated resistance within the next two years.

# CHAPTER 5

## CONCLUSIONS

- The molecular approach of sequence determination of the CP variable N-terminal region as a method of strain differentiation has proven to be an accurate and plausible method of strain identification.

- The prevalent SCMV strains in the SA sugar industry are SCMV-E and SCMV-D.

- There is no correlation between strain prevalence and variety or geographic origin of isolates in the SA sugar industry.

- Identification of strains by sequence determination is entirely compatible with a recently developed RFLP approach and the latter would serve as a quick, routine diagnostic test for SCMV detection and identification of strains in the SA sugar industry.

- The use of any SCMV-E or SCMV-D sequence would be suitable for the development of genetically engineered sugarcane with conferred CP-mediated resistance.

# LIST OF REFERENCES

Abbott, E.V. & Stokes, I.E. 1966. A world survey of sugar cane mosaic virus strains. *Sugar Azucar*, **61**: 27-29.

Abbott, E.V. & Tippett, R.L. 1966. Strains of sugarcane mosaic virus. *U.S. Department of Agriculture Technical Bulletin*, **1340**: 1-25.

Achon, M.A., Medina, V., Shanks, M., Markham, P. & Lomonossoff, G.P. 1994. Characterisation of a maize-infecting potyvirus from Spain. *European Journal of Plant Pathology*, **90**: 157-165.

Altschul, S.F. 1998. Fundamentals of Database Searching. In: *Trends Guide to Bioinformatics*, pp 7-9. Amsterdam: Elsevier.

Altschul, S.F., Gish, W., Miller, W., Meyers, E.W. & Lipman, D.J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, **215**: 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, **25**: 3389-3402.

Anonymous, 1980. Revised Sugarcane Diseases in South Africa. *South African Sugar Association Experiment Station Bulletin* **9**: 5-9.

Anonymous, 1993. Pathology department report. *South African Sugar Association Experiment Station Annual Report 1993*, pp 21-23.

Anonymous, 1994a. Mosaic. *Disease Information Sheet 2.4.* Mount Edgecombe: South African Sugar Association Experiment Station. 2p.

Anonymous, 1994b. Pathology department report. *South African Sugar Association Experiment Annual Report 1994*, pp 24-26.

Anonymous, 1996a. Mosaic survey 1985/6 to 1995/6. *South African Sugar Association Experiment Pest and Disease Report 1995/6*, p12.

Anonymous, 1996b. *Protocols and applications guide.* Third edition. USA: Promega Corporation. 404p.

Atreya, P.L., Lopez-Moya, J.J., Chu, M., Atreya, C.D. & Pirone, T.P. 1995. Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. *Journal of General Virology*, **76**: 265-270.

Bailey, R.A. 1979. An assessment of the status of sugarcane diseases in South Africa. *Proceedings of the South African Sugarcane Technologists Association*, **53**: 120-129.

Bailey, R.A, & Fox, P.H. 1980. The susceptibility of varieties to mosaic and the effect of planting date on mosaic incidence in South Africa. *Proceedings*

*of the South African Sugar Technologists Association*, **54**: 161-167.

Bailey, R.A. & Fox, P.H. 1987. A preliminary report on the effect of sugarcane mosaic virus on the yield of sugarcane varieties NCo376 and N12. *Proceedings of the South African Sugarcane Technologists Association*, **61**: 120-129.

Bateson, M.F., Henderson, J., Chaleeprom, W., Gibbs, A.J. & Dale, J.L. 1994. Papaya ringspot potyvirus: isolate variability and the origin of PRSV type P (Australia). *Journal of General Virology*, **75**: 3547-3553.

Berger, P.H., Luciano, C.S., Thornbury, D.W., Benner, H.I., Hill, J.H. & Zeyen, R.H. 1989. Properties and in vitro translation of maize dwarf mosaic virus RNA. *Journal of General Virology*, **70**: 1845-1851.

Biggin, M.D., Gibson, T.J. & Hong, G.F. 1983. Buffer gradient gels and $^{35}S$ label as an aid to rapid DNA sequence determination. *Proceedings of the National Academy of Science*, **80**: 3963.

Bishop, M.J. 1989. Computing. In: *Nucleic Acids Sequencing: A practical approach*, pp 185-219. Howe, C.J. & Ward, E.S. (Eds.). Oxford: IRL Press at Oxford University Press.

Brandes, E.W. 1919. The mosaic disease of sugarcane and other grasses. *U.S. Department of Agriculture Technical Bulletin*, **829**: 26.

Brenner, S.E. 1998. Practical Database Searching. In: *Trends Guide to Bioinformatics*, pp 9-12. Amsterdam: Elsevier.

Brunt, A.A., Crabtree, K., Dallwitz, M.J., Gibbs, A.J., Watson, L. & Zurcher, E.J. (Eds). 1996. Plant viruses online: Descriptions and lists from the VIDE database. Version: 16th January 1997. URL: http://biology.anu.edu.au/Groups/MES/vide

Butler, E.T. & Chamberlin, M.J. 1982. Bacteriophage SP6-specific RNA polymerase. I: Isolation and characterisation of the enzyme. *Journal of Biological Chemistry*, **257**: 5772.

Butterfield, K.M. & Thomas, D.W. 1996. Sucrose yield and disease resistance characteristics of sugarcane varieties under test in the SASEX breeding programme. *Proceedings of the South African Sugar Technologists Association*, **70**: 103-105.

Candresse, T., Cambra, M., Dallot, S., Lanneau, M., Asensio, M., Gorris, M.T., Revers, F., Macquaire, G., Olmos, A., Boscia, D., Quiot, J.B. and Dunez, J. 1998. Comparison of monoclonal antibodies and polymerase chain reaction assays for the typing of isolates belonging to the D and M serotypes of plum pox potyvirus. *Phytopathology*, **88**: 198-204.

Chambers, K.R. 1981. *Sugarcane mosaic virus*. Mount Edecombe, Natal: SA Sugar Experiment Station. 75p.

Collmer, C.W., Marston, M.F., Albert, S.M., Bajaj, S., Maville, H.A., Ruuska, S.E., Vesely, E.J. & Kyle, M.M. 1996. The nucleotide sequence of the coat protein gene and 3' untranslated region of azuki bean mosaic potyvirus, a member of the bean common mosaic virus subgroup. *Molecular Plant-Microbe Interactions*, **9**: 758-761.

Cronje, C.P.R., Bechet, G.R. & Bailey, R.A. 1994. Symptom expression of sugarcane mosaic virus and associated effects on sugarcane yield. *Proceedings of the South African Sugar Technologists Association*, **68**: 8-11.

Davanloo, P., Rosenberg, A.H., Dunn, J.J. & Studier, F.W. 1984. Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Science*, **81**: 2035.

Dougherty, W.G. & Carrington, J.C. 1988. Expression and function of potyviral gene products. *Annual Review of Phytopathology*, **26**: 123-143.

Edgerton, C.M. 1955. *Sugarcane and its diseases*. Baton Rouge: Louisiana University State Press. 301p.

Francki, R.I.B., Fauquet, C.M., Knudson, D.L. & Brown, F. 1991. *Classification and Nomenclature of Viruses: Fifth Report of the International Committee on Taxonomy of Viruses*. New York: Springer.

Frenkel, M.J., Jilka, J.M., McKern, N.M., Strike, P.M., Clark, J.M, Shukla, D.D. & Ward, C.W. 1991. Unexpected sequence diversity in the amino-terminal ends of the coat proteins of strains of sugarcane mosaic virus. *Journal of General Virology,* **70**: 2775-2783.

Frenkel, M.J., Jilka, J.M., Shukla, D.D. & Ward, C.W. 1992. Differentiation of potyviruses and their strains by hybridisation with the 3' non-coding region of the viral genome. *Journal of Virological Methods*, **36**: 51-62.

Frenkel, M.J., Ward, C.W. & Shukla, D.D. 1989. The use of 3' non-coding nucleotide sequences in the taxonomy of potyviruses: application to watermelon mosaic virus 2 and soybean mosaic virus. *Journal of General Virology*, **70**: 2775-2783.

Gelfand, D.H. 1989. *Taq* DNA polymerase. In: *PCR Technology: Principles and Applications for DNA Amplification.* Erlich, H.A. (Ed.). pp 17-22. New York: Stockton Press.

Gibbs, A & Mackenzie, A. 1997. A primer pair for amplifying part of the genome of all potyvirids by RT-PCR. *Journal of Virological Methods*, **63**: 9-16.

Gillaspie, A.G. Jr & Koike, H. 1973. Sugarcane mosaic virus and maize dwarf mosaic virus in mixed infections of sugarcane and other grasses. *Phytopathology*, **63**: 1300-1307.

Gillaspie, A.G. Jr & Mock, R.G. 1987. World distribution of strains of sugarcane mosaic virus. *Sugarcane*, **6**: 11-12.

Gillaspie, A.G. Jr, Gipe, M.L., Mock, R.G., Handojo, H & Irawan 1986. Sugarcane mosaic virus strains in Indonesia. *Proceedings of the International Society of Sugarcane Technologists* **19**: 352-356.

Glais, L., Kerlan, C., Tribodet, M., Tordo, V.M., Robaglia, C. & Astier-Manifacier, S. 1996. Molecular characterisation of potato virus $Y^N$ isolates by PCR-RFLP. *European Journal of Plant Pathology*, **102**: 655-662.

Goodman, B.S., Groenewald, J-H., Botha, F.C. & Huckett, B.I. 1997. Use of CP sequences in transformations. Fifth International Congress of Plant Molecular Biology, Singapore, 21-27 September 1997.

Goodman, B.S., Macdonald, D. & Huckett, B.I. 1998. A survey of South African sugarcane mosaic virus (SCMV) strains based on coat protein gene sequence analysis. *Proceedings of the South African Sugar Technologists Association,* **72**: 146-148.

Gough, K.H. & Shukla, D.D. 1981. Coat protein of potyviruses I: Comparison of the four Australian strains of sugarcane mosaic virus. *Virology*, **111**:455-462.

Gough, K.H., Azad, A.A. Hanna, P.J. & Shukla, D.D. 1987. Nucleotide sequence of the capsid and nuclear inclusion protein gene from the Johnson Grass strain of sugarcane mosaic virus RNA. *Journal of General Virology*, **68**: 297-304.

Griffin, H.G. & Griffin, A.M. 1993. Plasmid sequencing. In: *Methods in Molecular Biology, Vol. 23: DNA Sequencing Protocols*, pp 131-135. Griffin, H.G. & Griffin, A.M. (Eds.). New Jersey: Humana Press Incorporated.

Groenewald, J-H. 1998. Electronic communication to author. 2 Oct. 1998. Email address: jhgr@maties.sun.ac.za

Handley, J.A., Smith, G.R., Dale, J.L. & Harding, R.M. 1998. Sequence diversity in the coat protein coding region of twelve sugarcane mosaic potyvirus isolates from Australia, USA and South Africa. *Archives of Virology*, **143**: 1145-1153.

Heiner, C. & Hunkapiller, T. 1989. Automated DNA Sequencing. In: *Nucleic Acids Sequencing: A practical approach*, pp 221-235. Howe, C.J. & Ward, E.S. (Eds.). IRL Press at Oxford University Press, Oxford.

Huckett, B.I. & Botha, F.C. 1996. Progress towards a definitive diagnostic test for sugarcane mosaic virus infection. *Proceedings of the South African Sugarcane Technologists Association*, **70**: 11-13.

Irons, R.B. & Payn, G.C. 1992. The introduction of a new variety to farms in the Dumisa area in an effort to combat a mosaic problem. *Proceedings of the South African Sugar Technologists Association*, **66**: 41-43.

Irvine, J.E. 1972. Photosynthetic rate in sugarcane: Effect of virus diseases, genetic disorders and freezing. *Proceedings of the International Society of Sugarcane Technologists*, **14**: 1131-1138.

Ivinson, A.J. & Taylor, G.R. 1991. PCR in genetic diagnosis. In: *PCR: A practical approach.* pp 15-27. McPherson, M.J., Quirke, P. & Taylor, G.R. (Eds.). Oxford: IRL Press at Oxford University Press.

Jain, R.K., Pappu, H.R., Pappu, S.S., Varma, A. & Ram, R.D. 1998. Molecular characterisation of papaya ringspot potyvirus isolates from India. *Annals of Applied Biology*, **132**: 413-425.

Janson, B.F. & Ellett, C.W. 1963. A new corn disease in Ohio. *Plant Disease Reporter*, **47**: 1107-1108.

Jeffreys, A.J., Wilson, V., Neumann, R. & Keyte, J. 1988. Amplification of human minisatelites by the polymerase chain reaction: Towards fingerprinting of single cells. *Nucleic Acids Research*, **16**: 10953-1097.

King, N.C. 1952. Pathological department report. *South African Sugar Association Experiment Annual Report 1952*, pp18-20.

Kocher, T.D. & Wilson, A.C. 1991. DNA Amplification by the Polymerase Chain Reaction. In: *Essential Molecular Biology: A practical approach*, Vol. 2, pp 185-207. Brown, T.E. (Ed.). Oxford: IRL Press at Oxford University Press.

Koike, H. & Gillaspie, A.G.Jr. 1989. Mosaic. In: *Diseases of sugarcane: major diseases*. pp 301-322. Ricaud, C., Egan, B.T., Gillaspie, A.G.Jr & Hughes, C.G. (Eds.) Amsterdam: Elsevier.

Kondaiah, E. & Nayudu, M.V. 1984. A key to the identification of sugarcane mosaic virus strains. *Sugar Cane*, **6**: 3-8.

Kramer, M. & Coen, D.M. 1995. Enzymatic amplification of DNA by PCR: Standard procedures and optimisation. In: Current Protocols in Molecular Biology. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (Eds.) New York: John Wiley & Sons, Inc.

Lesemann, D-E., Shukla, D.D., Tosic, M & Huth, W. 1992. Differentiation of the four viruses of the sugarcane mosaic virus complex based on cytopathology. In: *Potyvirus Taxonomy*. pp353-361. Barnett, O.W. (Ed). New York: Springer.

Maxam, A.M. & Gilbert, W. 1977. A new method for DNA sequencing. *Proceedings of the National Academy of Science*, **74**: 5460.

McKern, N.M., Shukla, D.D., Toler, R.W., Jensen, S.G., Tosic, M., Ford, R.E., Leon, O. & Ward, C.W. 1991. Confirmation that the sugarcane mosaic virus subgroup consists of four distinct potyviruses by using peptide profiles of coat proteins. *Phytopathology*, **81**: 1025-1029.

McNeil, J.E., French, R., Hein, G.L., Baenziger, P.S. & Eskridge, K.M. 1996. Characterisation of genetic variability among natural populations of wheat streak mosaic virus. *Phytopathology*, **86**: 1222-1227.

Messing, J. 1983. New M13 vectors for cloning. *Methods in Enzymology*, **101**: 20.

Mirkov, T.E. 1997. Electronic communication to author. 13 March 1997. Email address: e.mirkov@tamu.edu

Pappu, S.S., Brand, R., Pappu, H.R., Rybicki, E.P., Gough, K.H., Frenkel, M.J. & Niblett, C.L. 1993. A polymerase chain reaction method adapted for selective amplification and cloning of 3' sequences of potyviral genomes: application to dasheen mosaic virus. *Journal of Virological Methods*, **41**: 9-20.

Pirone, T.P. 1972. Sugarcane mosaic virus. Number 88. In: *Descriptions of Plant Viruses*. Commonwealth Mycological Institute / Association of Applied Biology, Kew, Surrey, England.

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.

Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. *Molecular cloning: A Laboratory Manual.* Second ed. New York: Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. & Coulson, A.R. 1977. DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Science*, **74**: 5463.

Shukla, D.D., Frenkel, M.J., McKern, N.M., Ward, C.W., Jilka, J., Tosic, M. & Ford, R.E. 1992. Present status of the sugarcane mosaic subgroup of potyviruses. *Archives of Virology* [Suppl 5], 363-373.

Shukla, D.D., Frenkel, M.J. & Ward, C.W. 1991. Structure and function of the potyvirus genome with special reference to the coat protein coding region. *Canadian Journal of Plant Pathology*, **13**: 178-191.

Shukla, D.D., Strike, P.M., Tracy, S.L., Gough, K.H. & Ward, C.W. 1988. The N- and C-termini of the coat proteins of potyviruses are surface located and the N-terminus contains the major virus-specific epitopes. *Journal of General Virology*, **69**: 1497-1508.

Shukla, D.D., Tosic, M., Jilka, J., Ford, R.E., Toler, R.W. & Langham, M.A.C. 1989. Taxonomy of potyviruses infecting maize, sorghum and sugarcane in Australia and the United States as determined by reactivities of polyclonal antibodies directed towards virus-specific N-termini of coat proteins. *Phytopathology*, **79**: 223-229.

Shukla, D.D. & Ward, C.W. 1988. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *Journal of General Virology*, **69**: 2703-2710.

Shukla, D.D. & Ward, C.W. 1989. Identification and classification of potyviruses on the basis of coat protein sequence data and serology. *Archives of Virology*, **106**: 171-200.

Shukla, D.D. & Ward, C.W. 1994. Recent developments in the identification and classification of viruses comprising the sugarcane mosaic virus subgroup of potyviruses. In: *Current trends in sugarcane pathology*. pp 185-198. Gillaspie Jr, A.G.; Bergamin, P.P.; Agnihotri, V.P. & Chen, C.T. (Eds). Dehli, India: International Books and Periodicals Supply Service.

Shukla, D.D., Ward, C.W. & Brunt, A.A. 1994. *The Potyviridae*. Cambridge: Centre for Agriculture and Biosciences International, Cambridge University Press. p516.

Smith, G.R. & Van de Velde, R. 1994. Detection of sugarcane mosaic virus and fiji disease virus in diseased sugarcane using the polymerase chain reaction. *Plant Disease*, **6**: 557-561.

Sokhandan, N., Gillings, M.R. & Bowyer, J.W. 1997. Polymerase chain reaction detection of genetic variation in New South Wales isolates of passionfruit woodiness potyvirus. *Australian Plant Pathology*, **26**: 155-164.

Stark, D.M. & Beachy, R.N. 1989. Protection against potyviral infection in transgenic plants: evidence for broad spectrum resistance. *Biotechnology*, **7**: 1257-1262.

Stavolone, L., Alioto, D., Raggozzino, A. & Laliberte, J. 1998. Variability among turnip mosaic potyvirus isolates. *Phytopathology*, **88**: 1200-1204.

Storey, H.H. 1924. Diseases of sugarcane of the mosaic type in South Africa Part I. *U.S. Department of Agriculture*, **9**: 108-117.

Sturtevant, A.P. & Beachy, R.N. 1993. Virus resistance in transgenic plants: coat protein-mediated resistance. In: Transgenic plants: fundamentals and applications. pp 93-112. Hiatt, A. (Ed). New York: Marcel Dekker Inc.

Summer, E.M. 1934. Effect of mosaic caused by different strains of virus upon yields of susceptible sugarcane varieties. *Sugar Bulletin*, **21**: 181-183.

Taylor, G.R. 1991. Polymerase chain reaction: basic principles and automation. In: *PCR: A practical approach.* pp 1-14. McPherson, M.J., Quirke, P. & Taylor, G.R. (Eds.). Oxford: IRL Press at Oxford University Press.

Thomson, G.M. 1962. A report on the disease position in the Natal sugar industry. *The South African Sugar Journal*, **6**: 475-477.

Thomson, K.G., Dietzgen, R.G., Gibbs, A.J., Yamg, Y.C., Liesack, W., Teakle, D.S. & Stackebrandt, E. 1995. Identification of zucchini yellow mosaic potyvirus by RT-PCR and analysis of sequence variability. *Journal of Virological Methods*, **55**: 83-96.

Tosic, M., Ford, R.E., Shukla, D.D.& Jilka, J. 1990. Differentiation of sugarcane, maize dwarf, johnsongrass and sorghum mosaic viruses based on reaction of oat and some sorghum cultivars. *Plant Disease*, 74: 549-552.

Ward, C.W., McKern, N.M., Frenkel, M.J. & Shukla, D.D. 1992. Sequence data as the major criterion for potyvirus classification. Archives of Virology, (Suppl. 5) 283-297.

Wurthrich, E. 1921. Yellow stripe disease in Natal. *The South African Sugar Journal*, 1920-21: 157-158.

Yang, Z.N. & 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.

Zaccomer, B., Haenni, A. & Macaya, G. 1995. The remarkable variety of plant RNA virus genomes. *Journal of General Virology*, **76**: 231-247.

**APPENDIX I**

Presentations and publications based on work in this dissertation.

1. Goodman, B.S., Botha, F.C. and Huckett, B.I.
   Identification of South African isolates of sugarcane mosaic potyvirus (SCMV) by sequence analysis of the 5' end of the coat protein gene. Poster presentation at the International Society of Sugar Cane Technologists' Pathology and Molecular Biology Workshop, 11-21 May 1997 at Umhlanga Rocks, South Africa.

2. Goodman, B.S., Groenewald, J-H., Botha, F.C. and Huckett, B.I.
   South African strains of sugarcane mosaic potyvirus: identification and application in engineering coat protein-mediated resistance. Presentation at the International Society of Sugar Cane Technologists' Pathology and Molecular Biology Workshop, 11-21 May 1997 at Umhlanga Rocks, South Africa.

3. Goodman, B.S., Groenewald, J-H., Botha, F.C. and Huckett, B.I.
   South African strains of sugarcane mosaic potyvirus: identification and application in engineering coat protein-mediated resistance. Poster presentation at the 5th International Congress of Molecular Biology, 21-27 September 1997 in Singapore.

4. Goodman, B.S., Macdonald, D. and Huckett, B.I.
   Use of coat protein gene sequence analysis to survey South African

sugarcane mosaic virus strains. Poster presentation at the Joint KwaZulu-Natal Biochemistry and Microbiology Symposium, 15-17 October 1997 in Durban, South Africa.

5. Goodman, B.S., Macdonald, D. and Huckett, B.I.

A survey of South African sugarcane mosaic virus (SCMV) strains based on coat protein gene sequence analysis. *Proceedings of the South African Sugar Technologists' Association,* **72**: 146-148.

RECEIVED
29 APR 1999
FACULTY OFFICE:
APPLIED SCIENCE