THE USE OF AGROINFECTIOUS CLONES TO INVESTIGATE RECOMBINATION BETWEEN DISTINCT MAIZE STREAK VIRUS STRAINS

Wendelin Heribert Schnippenkoetter

Thesis presented for the degree of Doctor of Philosophy
in the Department of Microbiology, University of Cape Town

December 1997

The University of Cape Town has been given the right to reproduce this thesis in whole or in part. Copyright is held by the author.

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

CERTIFICATION BY SUPERVISOR

In terms of Paragraph GP9 of the regulations for the degree of Doctor of Philosophy at the University of Cape Town, I certify that I approve of the inclusion in this thesis of material already published, or submitted for publication, by candidate Wendelin Schnippenkoetter.

Signed by candidate

Signature Removed

Edward P. Rybicki, PhD
Associate Professor in Virology
Department of Microbiology
University of Cape Town

December 1997

CONTENTS

ABSTRACT

The infectivity of the replicative form (RF) DNAs of MSV-Kom, MSV-Set and PanSV-Kar contained in the plasmids pKom500, pSet100 and pPS100 was established by agroinoculating susceptible Jubilee sweetcorn with partial homodimeric *Agrobacterium tumefaciens* (C58C1) clones of RF-DNAs. Biological characteristics typical of Mastreviruses; such as, the appearance and leafhopper transmissibility of streak symptoms on infected plants, the presence of 18x30nm geminate particles in electron micrographs of leaf-dip preparations, and the presence of single-stranded and double-stranded DNA in Southern blot tests of infected plant DNA extracts, indicated that the RF-DNAs in pKom500, pSet100 and PanSV-Kar represent the entire genomes of MSV-Kom, MSV-Set and PanSV-Kar respectively. The complete nucleotide (nt) sequence of the genome of MSV-Set was determined and characterised, and compared with those of MSV-Kom and PanSV-Kar. The genome sizes of MSV-Kom, MSV-Set and PanSV-Kar are 2701, 2690 and 2705 nt respectively, and all share Mastreviral genomic features. Phylogenetic analyses on the nt sequences and the putative amino acid sequences of the movement, coat and replication-associated proteins (MP, CP and Rep respectively) indicate that MSV-Set is grouped with, yet distinct from the MSV group of viruses isolated from maize. MSV-Set shares a 78% nt sequence identity with MSV-Kom which shares a >96% nt sequence identity with other MSVs. The PanSV-Kar genome shares a 60% nt sequence identity with the MSV group and 89% with the Kenyan PanSV-Ken. PanSV-Kar causes mild non-persistent streak in Jubilee sweetcorn. MSV-Kom (previously isolated from maize in Komatipoort, Mpumalanga) and MSV-Set (previously isolated from a *Setaria* species in Mt. Edgecombe, Kwazulu/Natal) have different pathogenicities, and have overlapping, but non-identical, host ranges. Leafhopper transmission tests determined that MSV-Kom and MSV-Set generally cause severe and moderate streak in maize cultivars, or mild and severe streak in wheat cultivars respectively.

Despite their genomic and biological differences, agroinoculation of MSV-Kom/MSV-Set heterodimers were shown to be infectious as long as the tandem clones contained at least two long intergenic regions (LIR) delimiting a set of four open reading frames (ORFs). Partial sequence analyses of recombinant progeny genomes isolated from heterodimer-agroinoculated plants indicated that progeny recombinants consisted mainly of wild-type MSV-Kom sequence with LIR or V1 sequence replacements of 2-176bp to the left or right of the *Bam* HI cloning junction between the different genomes. Preferred recombinational crossover sites were identified to have anything from 2 bp to 28 bp complete homology

between the genomes of MSV-Kom and MSV-Set, demonstrating the likelyhood that MSV-Kom and MSV-Set genomes can recombine in mixed infections to form new MSV strains in nature. The rationale for this investigation was twofold: first to shed some light on the mechanism of release of the genome of MSV agroinfection, which would have implications in resolving the mechanism of MSV DNA replication in the host; and second, to investigate the possibility of using recombinant progeny in agroinoculation tests in elucidating the genomic domains that may determine biological features, such as host range and symptom severity of MSV.

The infectivity and biological characteristics of MSV-Kom/MSV-Set recombinant RF-DNAs pSeK142 and pKoS005, generated by heterodimer agroinoculations, were determined in agroinoculation tests. pSeK142 contained MSV-Kom sequence with 105 bp replaced by MSV-Set cognate sequence in the LIR right of the hairpin loop and pKoS005, with a 176 bp replaced by MSV-Set cognate sequence in the V1 ORF. Although the host ranges and leafhopper transmissibility of these recombinants were not altered with respect to wild-type MSV-Kom, their agroinfectivities and streak symptoms were attenuated. The generation of recombinant progeny using the heterodimer agroinoculation technique of relatively distinct MSVs, and the subsequent determination of infectivity of recombinant progeny, demonstrated that viable recombinants can be generated by agroinoculation of heterodimers and appear to be "naturally selected" in agroinfected plants.

pKeSeK2, a heterologous 1.1mer containing two MSV-Kom LIRs flanking the four ORFs of MSV-Set, was found to be only weakly agroinfectious (10% agroinfectivity) in slow-growing Jubilee sweetcorn plants under low light intensity. PCR detection of replicating KeSeK chimaeras indicated that the MSV-Set Rep protein recognises and initiates trans-replication in the MSV-Kom LIR sequence. The implications of the results of the KeSeK, pSeK142 and pKoS005 chimaeras with respect to their biological features and taxonomic rankings are discussed.

ACKNOWLEDGEMENTS

I am grateful to Prof. Ed Rybicki for his supervision, encouragement and support throughout the duration of this study. The time I have spent in his laboratory has been pleasant and been a scientifically rewarding experience, largely due to his easy-going, yet effective style of supervision. I am indebted to Ed for his unending patience, especially when it came to proofreading this thesis and offering advice on it.

I am grateful to the following people or organisations for their contributions towards this PhD:

- Prof. Barbara von Wechmar for her advice and supervision of the leafhopper transmission work as well as encouragement and support during this study.
- Mr Donald Solomons and Mr Derrick Rossouw for planting and managing hundreds of plants and managing thousands of leafhoppers, as well as for their advice.
- Ms Di James for patiently teaching me how to sequence and for her assistance in sequencing many of my recombinants.
- Dr Fiona Hughes for her advice and kick-starting this project by making many of her clones as well as her agroinfectious clone available.
- Dr Mohammed Jaffer for the electron microscopy.
- The Electron Microscope Unit at UCT for their help in preparing the glossy pictures.
- My laboratory mates Kenneth, Janet, Darren and Eric - all of whom have been cited in this thesis - for their constructive criticism, discussion, advice and ideas.
- My colleagues at the University of the Western Cape for affording me the space from work for me to be able to write-up this thesis.
- The FRD for financial assistance.
- Permenthri, Fran, friends and family, for providing me with the capacity to retain my sense of humour and for all their moral support.

ABBREVIATIONS

Viruses

AbMV - *Abutilon* mosaic virus
ACMV - African cassava mosaic virus (previously cassava latent virus; CLV)
BCaMV - bean calico mosaic virus
BCTV - beet curly top virus
BDMV - bean dwarf mosaic virus
BGMV - bean golden mosaic virus
CSMV - *Chloris* striate mosaic virus
DSV - *Digitaria* streak virus
HrCTV - horse radish mosaic virus
MiSV - *Miscantus* streak virus
MSV - maize streak virus
MSV-Ken - MSV from Kenya (Howell, 1985)
MSV-Nig - MSV from Nigeria (Mullineaux *et al*, 1984)
MSV-Reu - MSV from Reunion (Peterschmidt *et al*, 1996)
MSV-SA - MSV from South Africa (Lazarowitz, 1988)
MSV-Set - MSV from *Setaria* sp. Kwazulu/Natal
MYMV - mungbean yellow mosaic virus
PanSV - *Panicum* streak virus
PanSV-Ken - PanSV from Kenya (Briddon *et al*, 1992)
PanSV-Kar - PanSV from Karino, South Africa
PHV - pepper huasteco virus
SLCV - squash leaf curl virus
SSV - sugarcane streak virus
SSV-N - SSV from Kwazulu/Natal, South Africa (Hughes *et al*, 1992)
TGMV - tomato golden mosaic virus
TLCV - tomato leaf curlvirus
TomGV - tomato golden virus
ToMoV - tomato mottle virus
TPCTV - tomatoe pseudo-curly top virus
TYDV - tobacco yellow dwarf virus
TYLCV - tomato yellow leaf curl virus
WDV - wheat dwarf virus

Genomic features and protein abbreviations

AC1 - 1st ORF on complementary-sense geminivirus genome component A
AC2 - 2nd ORF on complementary-sense geminivirus genome component A
AC3 - 3rd ORF on complementary-sense geminivirus genome component A
BC1 - 1st ORF on complementary-sense geminivirus genome component B
BV1 - 1st ORF on virion-sense geminivirus genome component B
C1 - 1st ORF on complementary-sense geminivirus genome
C2 - 2nd ORF on complementary-sense geminivirus genome
C3 - 3rd ORF on complementary-sense geminivirus genome
C4 - 4th ORF on complementary-sense geminivirus genome
CP - coat protein
IR - intergenic region
LIR - long intergenic region
MP - movement protein
NTP - nucleotide triphosphate binding motive involved in Rep B helicase activity
ORF - open reading frame
PBS - primer binding site

REn - replication enhancer
Rep - replication-associated protein
rpe - rightward promoter element
SIR - short intergenic region
TrAP - transcription activator protein
UAS - upstream activating sequence
V1 - 1st ORF on virion sense geminivirus genome
V2 - 2nd ORF on virion-sense geminivirus genome
V3 - 3rd ORF on virion-sense geminivirus genome
vir - *Agrobacterium* virulence gene
Vir - *Agrobacterium* virulence protein

Nucleic Acid terms

A, C, G, T, U - nucleotides: adenine, cytosine, guanosine, thymidine, uridine
b - bases
bp - base pair(s)
ccc - covalently closed circular
DNA - deoxyribonucleic acid
ds - double-stranded
GUS - β-glucoronidase protein, gene product of the *gus* gene
kb - kilobases
LB - left border sequence flanking T-DNA
MCS - multiple cloning site of plasmid vector
npt II - neomycin phosphotransferase II gene
nt - nucleotide(s)
nt pos. - nucleotide position(s)
RB - right border sequence flanking T-DNA
RCR - rolling circle replication
RF-DNA - replicative form DNA
RNA - ribonucleic acid
ss - single-stranded
Ti - tumour inducing
pTi - tumour inducing plasmid
T-DNA - tumour DNA transferred as ss DNA by *Agrobacterium*

Units of measure

Abs_{260} - absorbance reading at wavelength 260nm
k, m, μ, n and p - kilo, milli, micro, nano and pico
μl - microlitres
μg - micrograms
ng - nanograms
°C - degrees Celsius
Da - Daltons
g - grams
h - hours
kcal/mol - kilo calories per mole
kDa - kiloDaltons
M - Molar (moles per litre)
mM - millimolar
ml - millilitres
mg - milligrams
min - minutes
mm - millimetres
ng - nanograms
nm - nanometres

Myr	- million years
OD_{600}	- optical density at 600nm wavelength
rpm	- revolutions per minute
U	- units
v/v	- volume per volume
w/v	- weight per volume

Chemicals and biochemicals

ATP	- adenosine triphosphate
$CaCl_2$	- calcium chloride
DIG	- digoxygenin
DMSO	- dimethylsulphoxide
ddATP	- dideoxyadenosine trophosphate
ddCTP	- dideoxycytosine triphosphate
ddGTP	- dideoxyguanisine triphosphate
ddTTP	- dideoxythymidine triphosphate
dATP	- deoxyadenosine triphosphate
dCTP	- deoxycytosine triphosphate
dGTP	- deoxy guanosine triphosphate
dTTP	- deoxythymidine triphosphate
dNTPs	- deoxynucleoside triphosphate
dUTP	- deoxyuracil triphosphate
EDTA	- ethylene diamine tetra acetic acid
HCl	- hydrochloric acid
H_3PO_4	- hydrogen pyrophosphate
IPTG	- isopropyl-β-D-thio-galactopyranoside
KCl	- pottasium chloride
KOAc	- pottasium acetate
$MgCl_2$	- magnesium chloride
NaCl	- sodium chloride
NaI	- sodium iodide
NaOAc	- sodium acetate
NaOH	- sodium hydroxide
Na_2SO_3	- sodium sulphite
NBT-phosphate	- nitroblue tetrazolium salt in 70% (v/v) dimethylformamide
X-phosphate	- 5-bromo-4-chloro-3-indolyl phosphate toluidium salt in 100% dimethylformamide
$RbCl_2$	- rubidium chloride
^{35}S-dATP	- deoxyadenosine triphosphate labelled with radioactive ^{35}S isotope
SDS	- sodium dodecyl sulphate
TAE	- Tris-acetate-EDTA electrophoretic buffer
TBE	- Tris-borate-EDTA electrophoretic buffer
TE	- Tris-EDTA buffer
Tris	- 2-amino-2-(hydroxymethyl)-1,3-propandiol
X-gal	- 5-bromo-4-chloro-3-indoyl-β-D-galactosylpyranoside

Antibiotics

Amp	- ampicillin
Gm	- gentamycin
Km	- kanamycin
Rif	- rifampicin

Miscellaneous

AGE	- agarose gel electrophoresis
ΔG	- free energy value in kcal/mol
cv.	- cultivar
DAS-ELISA	- double-antibody sandwich ELISA
ELISA	- enzyme-linked immunosorbant assay
LA	- Luria-Bertani Agar
LB	- Luria-Bertani broth
Mt.	- Mount
N2	- equilibration buffer (Nucleobond)
N3	- wash buffer (Nucleobond)
N5	- elution buffer (Nucleobond)
na	- not applicable
P1	- resuspension buffer (Qiagen)
P2	- lysis buffer (Qiagen)
P3	- neutralising buffer (Qiagen)
PCR	- polymerase chain reaction
pH	- hydrogen potential
GBT	- eguilibration buffer (Qiagen)
QC	- wash buffer (Qiagen)
QF	- elution buffer (Qiagen)
S1	- solution 1; resuspension buffer (Nucleobond or plasmid minipreparation)
S2	- solution 2; lysis buffer (Nucleobond or plasmid minipreparation)
S3	- solution 3; neutralising buffer (Nucleobond or plasmid minipraparation)
sp. (spp.)	- species (plural)
SSC	- sodium saline citrate buffer
TFB2	- transformation buffer 2
TSB	- transformation storage buffer
UV	- ultraviolet
v/v	- volume per volume
w/v	- weight per volume

CHAPTER 1

INTRODUCTION

---

1. INTRODUCTION

Geminiviruses are a unique group of infectious agents that cause disease in a variety of plant species including many agronomically important crops, particularly in tropical and sub-tropical regions. They are found throughout Africa, Latin America, the Mediterranean Basin, South East Asia, Australia and Western parts of North America (Lazarowitz, 1992). The diseases of these viruses were recognized well before the infectious agents were identified (Storey, 1928). They were only properly characterised as late as 1974 (Bock *et al.*, 1974; Matthews, 1979) largely due to their phloem restriction, lack of ease of mechanical transmission and physically fragile nature of the virions. However, their genomes are relatively easy to clone and thus have been well characterised in the last decade. Compared to plant RNA viruses, their small DNA genomes with extensive reliance on plant host systems are useful as models for the study of host DNA replication and transcription. Since they also replicate autonomously in the nuclei of the host, they have received much attention as vectors for the expression of foreign genes in plants (Howarth and Goodman, 1982; Hayes *et al.*, 1988; Davies and Stanley,1989; Coutts *et al.*, 1990; Meyer *et al.*, 1992; Stanley, 1993; Shen and Hohn, 1994; 1995). Considerable progress has been made toward understanding geminivirus molecular biology and pathology and reviews are available (Stanley, 1985; Lazarowitz, 1992).

1.1 Distinguishing features of geminiviruses

1.1.1 *Physical and biological features*

Members of the family *Geminiviridae* are characterised by their twinned quasi-icosahedral particles ("geminate") of approximately 18nm in diameter and 30nm in length (Figure 1.1), containing single-stranded (ss) circular DNA molecules of approximately 2.5-3 kb in length (for reviews see Stanley, 1985, and Lazarowitz, 1992). Virions consist of two incomplete icosahedra (T=1) with a total of 22 capsomers. Geminiviruses are subdivided into three subgroups or genera (Briddon and Markham, 1995) which are distinguished based on differences in insect vector specificity, genome organisation and host range. Recently the International Committee on the Taxonomy of Viruses (ICTV) have designated geminivirus subgroups I, II and III as Mastreviruses, Curtoviruses and Begomoviruses respectively (E.P. Rybicki, pers. comm.).

Mastreviruses are transmitted by leafhoppers (Homoptera: *Cicadellidae*), and have monopartite genomes. Members of this group include maize streak virus (MSV), *Chloris* striate mosaic virus (CSMV), *Digitaria* streak virus (DSV), *Panicum* streak virus (PanSV), sugarcane streak virus (SSV), and wheat dwarf virus (WDV).

Mastreviruses have relatively narrow host ranges; all the known members are hosted by monocotyledonous plants except tobacco yellow dwarf virus (TYDV; Morris, *et al.*, 1992) and bean yellow dwarf virus (BeYDV; Liu *et al.*, 1997).

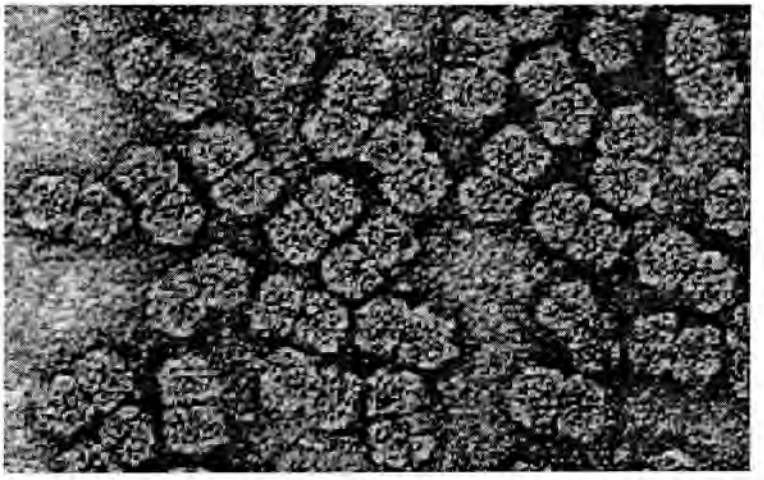

Figure 1.1 Scan of electron micrograph of geminate particles stained with uranyl acetate (particles18x30nm). Image scanned from photographic prints (Rothamsted Experimental station, 1994).

Curtoviruses also have monopartite genomes and are transmitted by leafhoppers with the exception of tomato pseudo-curly top virus (TPCTV; Briddon and Markham, 1995) which is transmitted by a treehopper (Homoptera: *Membracidae*). These viruses have a typically wide host range; for example, beet curly top virus (BCTV; Stanley *et al.*, 1986) has a very wide host range of over 300 species in 44 plant families.

Begomoviruses, on the other hand, have strikingly narrower host ranges than the Mastreviruses and the Curtoviruses; however, restrictions in the host ranges of Begomoviruses are not imposed by the feeding habits of the insect vector (Goodman, 1981a and 1981b). Begomoviruses are whitefly-transmitted (*Bemisia tabaci*; Homoptera: *Aleyrodidae*) and usually consist of bipartite genomes comprising two similar sized DNA molecules (A and a dependent B), although Begomovirus members with only the A DNA component have been identified. It has been suggested that the A component of these viruses may have independently lost their B components or do have a B component that has simply not yet been found (Rybicki, 1994). Individual Begomoviruses generally have narrow host ranges amongst dicotyledenous plants. Members include squash leaf curl virus (SLCV), African cassava mosaic virus (ACMV) and *Abutilon* mosaic virus (AbMV). Some types of tomato yellow leaf curl virus (TYLCV) and tomato leaf curl virus (TLCV) are examples of single-component Begomoviruses (Briddon and Markham, 1995).

1.1.2 *Replication*

Geminivirus DNA replication occurs through double-stranded replicative intermediates (RF-DNA) via a rolling circle mechanism (RCR) and is largely dependent on the mechanisms of the host nucleus. RF-DNA is assembled into nucleosomes in the host nucleus (Pilartz and Jeske, 1992; Abouzid *et al.*, 1988). Geminivirus DNA replication can be thought of as two-step process: first the conversion of a single stranded virion DNA to a dsDNA that will serve as a template for transcription of the viral genes; and secondly, the production of a single stranded virion DNA from the double stranded intermediate (Laufs *et al.*, 1995). Geminivirus replication has been extensively reviewed by Laufs *et al.* (1995), Bisaro (1996) and Hanley-Bowdoin *et al.* (1996). Little is known about the synthesis of viral minus strand to form the ds and supercoiled replicative DNA, however a virion-associated primer is encapsidated in Mastrevirus particles (Donson *et al.*, 1984; 1987; Hayes *et al.*, 1988a). No such primer has been found in Curto- or Begomoviruses and the origin of replication of the minus-strand is unknown (Hanley-Bowdoin *et al.*, 1996).

Accumulating evidence suggests that the dsDNA intermediates or RF-DNAs serve as a template for RCR which involves a multifunctional protein of about 41kDa designated Rep (Lazarowitz, 1992; Rogers *et al*, 1986; Saunders *et al.*, 1991). Rep shares no homology to any known DNA polymerase; however it is related to proteins catalysing the initiation of replication of ssDNA plasmids (Koonin and Ilyina, 1992). Rep is the sole virus-encoded protein required for replication in geminiviruses (Laufs *et al.*, 1995).

The plus-strand origin of replication constitutes a sequence capable of forming a hairpin loop (or stem-loop; Figure 1.2) along with flanking upstream activating sequences (UAS) found in the intergenic region (IR) (Revington *et al.*, 1989; Stenger *et al.*, 1991; Lazarowitz *et al.*, 1992). An invariant TAATATTAC loop sequence, resembling the gene A protein cleavage site of the ssDNA phage ΦX174 and related viruses, is flanked by inverted repeats (Sunter *et al.,* 1985). Rep binds with high affinity to UASs upstream of the hairpin loop either as dsDNA (Fontes *et al.*, 1992) or as ssDNA (Thömmes *et al.*, 1993b) and has been shown to initiate replication by acting as a site- and strand-specific endonuclease (Koonin and Ilyina 1992; Heyraud-Nitschke *et al.*, 1995) by nicking the TAATATT↓AC (as indicated by ↓), covalently binding the free 5' end and allowing 3' end elongation by repair polymerases (Elmer *et al.*, 1988a; Hayes and Buck, 1989; Hanley-Bowdoin *et al.*, 1990).

The UAS high affinity binding sites are apparently themselves not sufficient to direct specific origin recognition *in vivo* (Laufs *et al.*, 1995; Fontes *et al.*, 1994a; 1994b). The existence of

other specificity determinants of replication located in the IR right of the UAS and in the inverted repeat flanking the nonanucleotide TAATATTAC were identified in sequence comparison studies of related geminiviruses (Figure 1.2; Argüello-Astorga *et al.*, 1996). These iterative sequence motifs are arranged in a similar fashion among phylogenetically related geminiviruses, which suggests their involvement as a specific binding site for geminivirus replication-associated proteins (Rep).

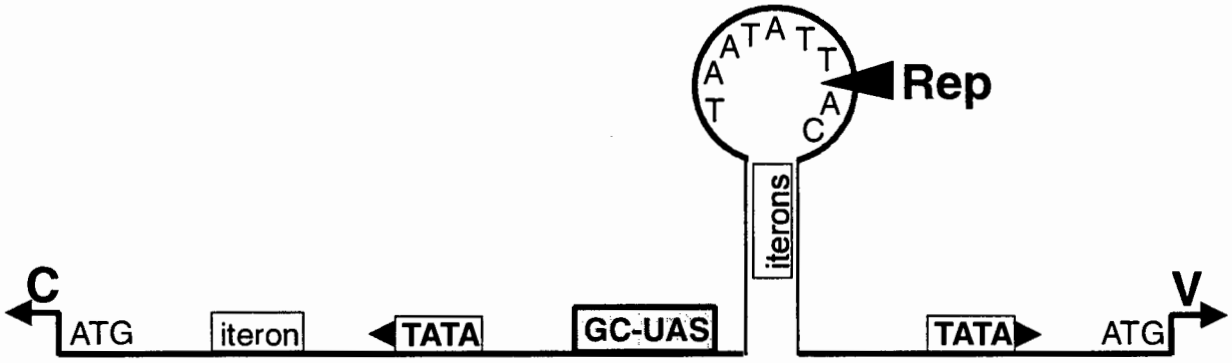

Figure 1.2: Schematic diagram of the intergenic region of geminiviruses

Diagram is not drawn to scale. C = direction of complementary sense ORF(s) starting at an ATG start codon. V = direction of virion sense ORF(s) starting at an ATG start codon. An iteron or iterative element is indicated to the left and within the stem-loop or hairpin (two complementary to each other forming part of the stem; Argüello-Astorga *et al.*, 1996). TATA boxes are indicated and their directions of promotion of transcription to the left and right of the hairpin structure. The absolutely conserved nonanucleotide TAATATT↓AC is indicated in the loop with the Rep nick site where replication of the viral genome is initiated (Koonin and Ilyina 1992; Heyraud-Nitschke *et al.*, 1995). The GC-rich high affinity Rep binding site is indicated just left of the hairpin structure in a shaded box (Laufs *et al.*, 1995; Fontes *et al.*, 1994a; 1994b).

Sequence comparison of Rep proteins (Figure 1.3) from different geminiviruses shows significant conservation of four amino acid sequence motifs (Ilyina and Koonin, 1992; Koonin and Ilyina, 1992; Gorbalenya *et al.*, 1990). The function of the FLTY motif is unknown, whereas the HLH motif is speculated to be involved in metal ion coordination (Koonin and Ilyina, 1992). The third motif contains a conserved tyrosine implicated in DNA cleavage at the TAATATTAC origin. The fifth motif represents the NTP-binding motif binding site, $EGX_4GKTX_{32}DD$; mutations of which abolish replication (Gorbalenya *et al*, 1990; Desbiez *et al.*, 1995; Hanson *et al.*, 1995). Rep encoded by the C1/C2 ORFs of Mastreviruses, C1 ORF of Curtoviruses or AC1 ORF of Begomoviruses, is thought to have helicase activity (Gorbalenya and Koonin, 1989; Gorbalenya *et al.*, 1990). Rep is also responsible for regulating viral transcription by repressing its own expression (Sunter *et al.*,1993) and

stimulates expression from the coat protein (CP) promoter, at least in Begomoviruses (Hofer *et al.*, 1992). Transcription is bidirectional and is initiated within the IR.

In additon to its cleavage and supposed helicase ability, Rep has been shown to have DNA joining activity that permits the transfer of the 5' terminal phosphate of an oligonucleotide linked to Rep to the 3' OH of a concomitantly cleaved origin sequence, or even to a preformed free 3' OH end of an oligonucleotide recognized by Rep (Laufs *et al.*, 1995). Geminivirus Rep proteins contains the conserved motif Gly-X-X-X-X-Gly-Lys-Thr/Ser (contained in the $EGX_4GKTX_{32}DD$ NTP-binding motif) specifying the P-loop of the phosphate binding pouch of a wide variety of nucleoside triphophate (NTP) binding or hydrolysing proteins (Laufs *et al.*, 1995; Gorbalenya and Koonin, 1989). Although Rep has been shown to exhibit DNA-dependent ATPase activity *in vitro* and *in vivo,* the *in vitro* joining activity does not require ATP (Laufs *et al.*, 1995).

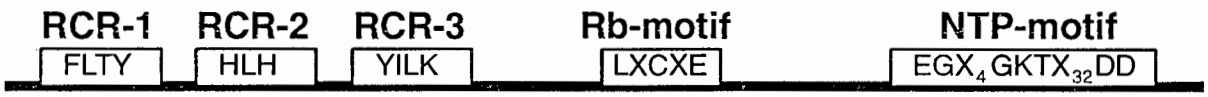

Figure 1.3: Schematic geminivirus Rep protein sequence

The N-terminus (left) contains the RCR-1, 2 and 3 conserved rolling circle replication motifs FLTY, HLH and YILK respectively (Koonin and Ilyina, 1992). The function of the RCR-1 motif' is unknown. The RCR-2 motif is thought to be involved in ion co-ordination. RCR-3 includes a conserved tyrosine thought to be involved in cleaving TAATATTAC. The Rb-motif represents the retinoblastoma binding motive found in Mastreviruses but not in Curto- or Begomoviruses (Xie *et al.*, 1995; Chapter 3). The C-terminus (right) contains the NTP-binding motif. (Figure 1.2; Gorbalenya *et al*, 1990; Desbiez *et al.*, 1995; Hanson *et al.*, 1995).

1.1.3 *Genome organisation*

The virion and complementary strands of geminiviral RF-DNA encode genes which diverge from an IR (Figure 1.4). The genome organisation of Curtoviruses resembles that of the A component of the Begomovirus bipartite genome, while that of the Mastreviruses is unique (Figure 1.4). Mastreviral genomes are approximately 2.6-2.8 kb in size and encode four significant ORFs; two on the virion-sense strand (V1 and V2) encoding the movement and coat proteins (MP and CP), respectively and two on the complementary-sense strand (C1 and C2 or *Rep* A and *Rep* B) encoding the replication-associated protein Rep (Figure 1.4; Chapter 3).

Curtoviral genomes are approximately 2.7-3.0 kb in length and potentially encode seven genes; three on the virion strand (V1, V2 and V3) and four on the complementary strand (C1-C4). C1 encodes Rep while C3 and V2 are also involved in DNA replication (Stanley *et al.,* 1992). The protein product of C3 enhances replication but it is unclear what the function of C2 is since mutations of C2 do not significantly reduce infectivity of BCTV (Stanley, 1992). The BTCV C4 coding region occurs entirely within the *Rep* coding region but in a different reading frame suggesting that expression of both genes is highly co-ordinated (Stanley and Latham, 1992; Stanley *et al.*, 1992). In the case of TGMV and ACMV Rep proteins were found to down regulate both Rep and C4 expression (Haley *et al.*, 1992; Hong and Stanley, 1995; Sunter *et al.*, 1993). Tobacco and sugarbeet plants infected with BCTV C4 mutants failed to induce wild-type-like enation and hyperplasia of phloem parenchyma cells suggesting that C4 protein may affect cell division (Stanley and Latham, 1992; Stanley *et al.*, 1992). Recent experiments with transgenic plants expressing the C4 gene resulted in abnormal growth of the plants confirming that C4 alone is sufficient to initiate cell division in permissive cells (Latham *et al.*, 1997). Homologues of the C3 protein are not found in Mastrevirus genomes and V2 is unique to BCTV (Curtovirus). V2 was shown to be involved in viral single-stranded DNA accumulation (Frischmuth *et al.*, 1993). Although V1 encodes the CP, it is also implicated in movement of the virus in the host along with the V3 product of BCTV (Briddon *et al.*, 1989; Stanley *et al.*, 1992; Hormuzdi and Bisaro, 1993; Frischmuth *et al.*, 1993).

The two genome components of Begomoviruses are 2.5-2.8kb in size, with the larger of the two components (DNA A) usually encoding four genes; one on the viral (AV1) and three on the complementary strand (AC1, AC2 and AC3, analogous to C1, C2 and C3 of Curtoviruses). Both DNAs A and B have a common intergenic region (IR) of approximately 230 nt with nearly complete sequence identity encompassing the conserved TAATATT↓AC sequence. DNA A encodes the CP and all the functions required for replication and the products of the DNA B are involved in spread within plant hosts (Brough *et al.*, 1988; Etessami, *et al.*, 1988; Gardiner *et al.*, 1988). AC1 encodes Rep, AC2 codes for a protein that transactivates the expression of virion sense promoters (TrAP), and AC3 specifies a protein that enhances viral DNA replication (REn) (Sunter *et al.*, 1990; Etessami *et al.*, 1991; Sunter and Bisaro, 1991 & 1992; Haley *et al.*, 1992). The TrAP and REn proteins have been demonstrated to complex and form oligomers and thought to be functional as protein complexes in the initiation of replication of Begomoviruses (Settlage *et al.*, 1996; Hanley-Bowdoin *et al.*, 1996). Homologues of TrAP and REn are not encoded in Mastrevirus genomes. The CP is encoded by the AV1 ORF but bipartite viruses are still infectious in its absence (Stanley and Townsend, 1986).

In monopartite Begomoviruses such as TYLCV, the genome encodes five genes as opposed to four on the A component of their bipartite counterparts (Lazarowitz, 1992). AV1, AC1, AC2 and AC3 of the monopartite Begomoviruses are homologues of AV1, AC1, AC2 and AC3 of their bipartite counterparts. The additional fifth ORF AV2 of monopartite begomoviruses potentially encodes a protein of 13kDa involved in spread. AV2 (which is normally absent in bipartite Begomoviruses) of TYLCV is homologous to an ORF similarly located in the bipartite Begomovirus ACMV (Lazarowitz, 1992). The A component of ACMV moves systemically independent of its B component 10-40% of the time, depending on the method of agroinoculation of tobacco (Klinkenberg and Stanley, 1990).

The B component encodes two genes; BC1 (MP) and BV1 (NS, nuclear shuttle) are required for viral infectivity and systemic infection and have essential but distinct roles in movement (von Arnim and Stanley, 1992a; Smith and Maxwell 1994; Haley *et al.*, 1995; Ingham *et al.*, 1995). BC1 has been shown to increase the size exclusion limit of plasmodesmata (Noueiry *et al.*, 1994) whereas the BV1 traffics ss (Pascal *et al.*, 1994) or ds (Noueiry *et al.*, 1994) viral DNA in and out of the host nucleus. Both BC1 and BV1 may act in combination with the CP (encoded by DNA A, AV1) in systemic movement (Ingham *et al.*, 1995; Jeffrey *et al.*, 1996; Pooma *et al.*, 1996). With the exception of TYLCV-Thailand (Rochester *et al.*, 1990), both genome components are required for infectivity and symptom development (Hamilton *et al.*, 1983; Stanley, 1983); however, the A genome encodes all viral functions necessary for the replication and encapsidation of viral DNA (Rogers *et al.*, 1986; Townsend *et al.*, 1986; Sunter *et al.*, 1987).

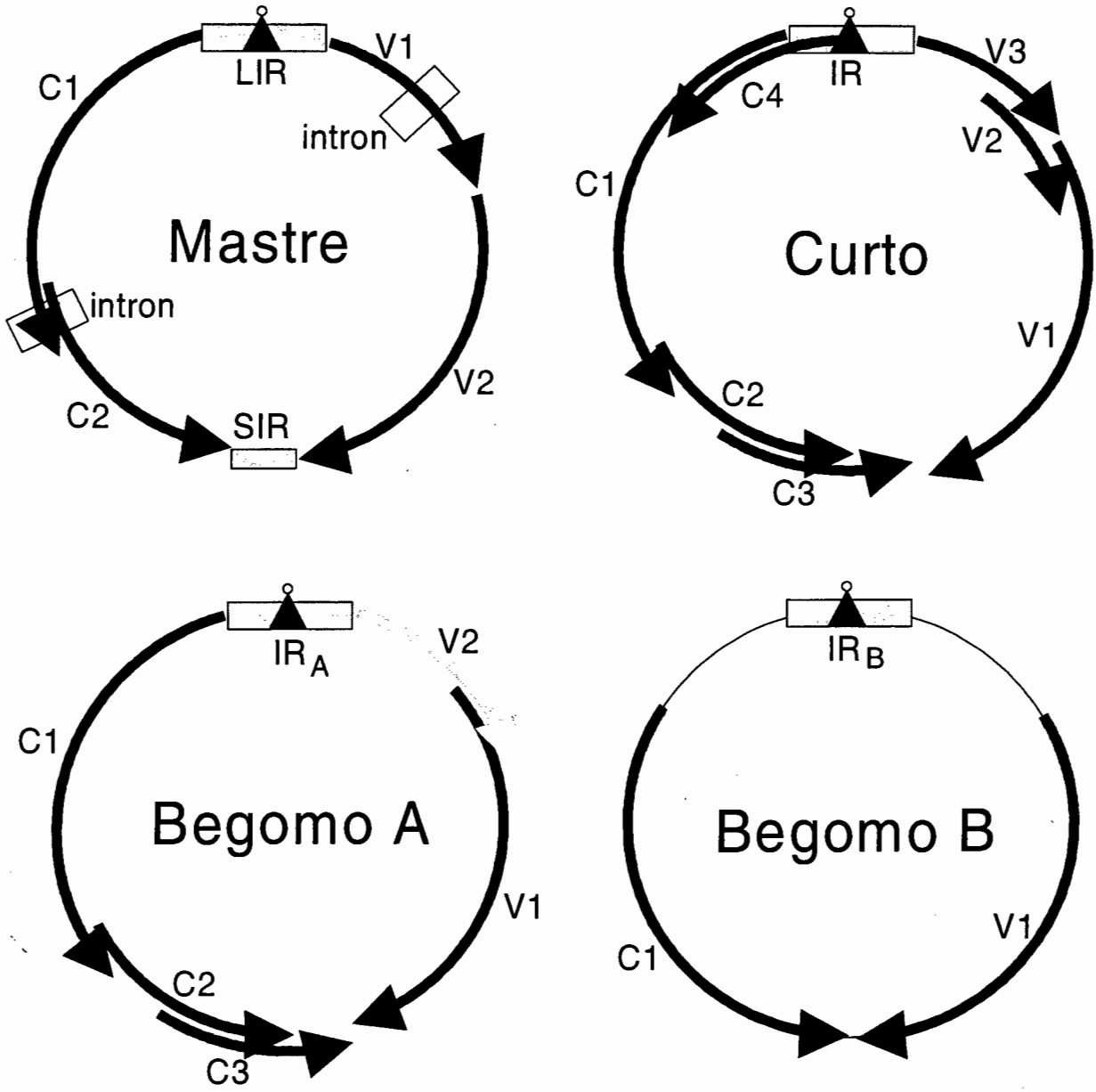

Figure 1.4 Typical genome organisations of geminiviruses

Mastrevirus, Curtovirus and Begomovirus genomes respectively. Genes are denoted V or C for virion or complementary strands respectively and their directions of transcription are shown by their arrows. **Mastreviruses**: V1 and V2 encode the MP and CP respectively and C1 and C2 encode Rep A and Rep B respectively. Introns are found in both C1/C2 and V1 transcripts. LIR and SIR = long and short intergenic regions respectively. **Curtoviruses**: V1 and C1 encode CP and Rep respectively. IR = intergenic region. V3, C2 and C3 likely to encode MP, TrAP and REn respectively. C4 affects host cell division. Most **Begomoviruses** have two components (A and B) and some a single A component. IR = common intergenic region. **Two-component Begomoviruses:** AC1, AC2, AC3, AV1, BC1 and BV1 encode Rep, TrAP, REn, CP, MP, and NS respectively. **Single-component Begomoviruses:** AC1, AC2, AC3, and V1 encode Rep, TrAP, REn and CP respectively. AV2 is only present in Old World geminiviruses, in which case it probably encodes the MP.

1.2 Phylogenetics of geminiviruses and evolutionary implications

Previous classifications of viruses were based on phenotypic characters such as symptomatology and host range, which depended on the genotypes of both the host and virus (Matthews, 1983). More refined classification of viruses depended on serological properties of capsid proteins, their amino acid composition, and their coat protein sequences (Shukla and Ward, 1988; Fauquet *et al.,* 1986). Serological studies of geminiviruses have shown that whitefly-transmitted geminiviruses are sufficiently related to cross-react with polyclonal and monoclonal antisera. Leafhopper-transmitted geminiviruses (Roberts *et al,* 1984) with the exceptions of MSV, DSV and SSV (Dollet *et al.,* 1986; Pinner *et al.*, 1992) have shown much less serological cross-reactivity, so that some serological tests, such as the double-antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) are limited in detecting the relatedness of distant viruses (Rybicki and von Wechmar, 1985; Clarke *et al.*, 1989). More recently however, CSMV, DDSMV (*Digitaria didactyla* striate mosaic virus) and PaSMV (*Paspalum* striate mosaic virus) strains from Australia; MSV, PanSV, and SSV strains from Africa; DSV from Vanuatu; and WDV from Europe were all found to be related in ELISA tests and placed into distinct Australian, African and European groups by phylogenetic analysis (Pinner *et al.*, 1992). The only geminivirus in the study conducted by Pinner and co-workers (1992) that did not cross-react in ELISA tests was MiSV from Japan. Interestingly, serological cross-reactivity was apparently detected between BCTV and TYDV, that is, between a Curto- and a Mastrevirus (Thomas and Bowyer, 1980). No antigenic cross-reactivity has been substantiated between whitefly- and leafhopper-transmitted geminiviruses (Thomas *et al.,* 1986).

Since capsid proteins, which have been shown to have good antigenic properties (Bock, 1982), are the almost sole antigenic factors for these observations, the effectiveness for comparative studies is limited. These limitations have been overcome with increasing geminivirus sequence data being made available for phylogenetic analyses of both DNA and protein sequence comparisons (Howarth and Vandemark, 1989; Rybicki, 1994; Padidam *et al.*, 1995). Phylogenetic relationships of geminiviruses are being more refined by the alignments and analyses of amino acid sequences of both gene products which all geminiviruses contain; namely, the CP and the Rep proteins (MacDowell *et al.*, 1985; Stanley *et al.*, 1986), and partial or complete nucleotide sequences. Thus even the most disparate geminiviruses can be and have been compared (Howarth and Vandemark, 1989; Rybicki, 1994; Padidam *et al.*, 1995).

1.2.1 *Sequence comparisons*

The division of Mastreviruses and Begomoviruses is supported by the phylogenetic analyses of the CP, MP and Rep protein sequences as well as the sequences of intergenic *cis*-acting sequences (Rybicki, 1994; Padidam *et al.*, 1995). Homologies detected between the V1 genes of DNA A and the V1 genes of their respective DNA B genomes of African cassava mosaic virus (ACMV), bean golden mosaic virus (BGMV) and tomato golden mosaic virus (TGMV) suggested that the bipartite genomes evolved from a monopartite ancestor (Kikuno *et al*, 1984; Rao, 1985). In addition, phylogenetic studies suggested that this evolution occurred by divergence from a common progenitor rather than segmentation of a larger sequence (Howarth and Goodman, 1986). A further distinct dichotomy within the Begomoviruses between "Old World and New World" viruses was confirmed in extensive phylogenetic analyses of virus sequence data (Rybicki, 1994; Padidam *et al*, 1995). It has been proposed to classify the Old World and New World groups as "*Archeogeminivirus* and *Neogeminivirus*" respectively; in line with the results of the phylogenetic analyses on the available sequences of geminiviruses (Padidam *et al.*, 1995); however, Rybicki (1994) determined that the sequence differences did not justify such a taxonomic separation.

Mastreviruses have been shown to be distinct as a group in all sequence comparisons done to date (Howarth and Vandemark, 1989; Rybicki, 1994; Padidam *et al.*, 1995). To justify their generic status of having a single-component genome with its unique organisation, it was proposed to designate the genus "*Monogeminivirus*" (Rybicki, 1994; Padidam *et al.*, 1995). In sequence comparisons (Howarth and Vandemark, 1989; Rybicki and Hughes, 1990; Hughes *et al.*, 1992; Rybicki, 1994) and in serological studies (Pinner, *et al.*, 1988; 1992), Mastreviruses were shown to cluster into distinct groups; the so-called "African streak group" and an "Australasian striate mosaic virus group" as well as very distinct viruses such as WDV (wheat dwarf virus), TYDV (tobacco yellow dwarf virus) and MiSV (*Miscanthus* streak virus). Mastreviruses were found to cluster largely in line with their geographical distributions (Rybicki, 1994). With the exclusion of DSV, the type of divergence apparent between Mastreviruses from Asia, Africa, Australasia and Europe/Middle East and the complete absence of reports of these viruses from the Americas, is reminiscent of a "Gondwanaland distribution" suggesting that Mastreviruses could have evolved from a common ancestor (Howarth and Goodman, 1986) some time after the separation of the Americas from Gondwana bloc (130 Myr) and then speciated along with continental drift (Rybicki, 1994). *Digitaria* streak virus is an anomaly, however, since it has closer relationships with African viruses even though its origin is near Australasia.

1.2.2 *Sequence evidence for recombination*

While Mastreviruses and Begomoviruses appear to have evolved from a common ancestor (Howarth and Goodman, 1986), Curtoviruses such as BCTV and tomato pseudo-curly top virus (TPCTV; Briddon *et al.*, 1996) have CPs which have affinities with that of Mastreviruses, and Rep proteins which have affinities with those of Begomoviruses. Thus, the genus appears to be the result of an ancient recombination event between a Mastrevirus-like virus and a Begomovirus-like virus (Rybicki, 1994; Briddon *et al.*, 1996). It was thus earlier proposed that the Curtovirus genus be designated "*Intergeminivirus*" (Rybicki, 1994; Padidam *et al.*, 1995). The prerequisites for this type of recombinational event is co-infection of a plant cell with two different geminivirus lineages followed by a recombination event, successful transmission by an insect vector and long-term survival of susceptible plants and vectors. If the B-component was present, it could be lost for similar reasons to those already invoked for "Old World" single-component Begomoviruses (Rybicki, 1994). A deliberately-engineered ACMV with a BCTV CP was found to still be infectious and could be leafhopper-transmitted rather than by whiteflies (Briddon *et al.,* 1990). Although the engineered ACMV was still two-component, this result demonstrated that these types of recombinational events were not unlikely events in nature (Rybicki, 1994).

Other viruses with genome structures indicating old recombinational events between different geminiviruses include squash leaf curl virus (SLCV), pepper huasteco virus (PHV; Torres-Pacheco *et al*, 1993), mungbean yellow mosaic virus (MYMV) and horse radish curly top virus (HrCTV; Klute *et al.*, 1996): SLCV appears to derive its IR and Rep sequences from BCaMV/PhVCMV/TomGV/PepGV-like viruses and its B-component from a virus in the New World cluster (Rybicki, 1994); PHV (pepper huasteco virus) appears to derive its CP AC2, AC3 and whole B-component from an Old World Begomovirus (Rybicki, 1994); MYMVs Rep protein groups with Old World viruses, the IRs with New World Begomoviruses and the presumptive MP with Old World viruses suggesting that MYMV is a representative of an ancestral lineage (Rybicki, 1994). The Rep and C4 proteins of the HrCTV genome share a closer relationship to the corresponding proteins of SLCV, and the rest of the genome bears a closer resemblance to the corresponding sequence of BCTV, indicating that this virus is the result of a recombinational event between a SLCV-like (Begomovirus) and a BCTV-like (Curtovirus) virus (Klute *et al*, 1996). A Ugandan variant of ACMV (AgV) was found to have a similar complete nucleotide sequence to that of a Tanzanian isolate of EACMV (East African cassava mosaic virus) except for the CP gene (Zhou *et al.*, 1997). The CP gene was found to have three distinct regions: the 5' 219nt are 99% identical to EACMV (79% to ACMV); the following 459nt are 99% identical to ACMV (75% to EACMV); and the 3' 93 nt are 98% identical to EACMV (76% to ACMV); thus UgV DNA A was

considered to have arisen by an interspecific recombination event between EACMV and ACMV (Zhou *et al.*, 1997).

Deliberately constructed recombinant genomes derived from the Logan and CFH strains of BCTV and their respective reciprocal clones were shown to be infectious as well as to exhibit novel pathogenic properties not common to their wild-type counterparts (Stenger *et al.*, 1994). This result is further indication that recombination between geminiviruses and their subsequent survival as new strains is indeed not an unlikely event in nature. More recently, the genotypic diversity of BTCV populations in the Western United States was examined by the analysis of 58 field isolates and eight laboratory and nursery isolates of BCTV using restriction endonuclease mapping data (Stenger and Mahon, 1997). The results indicated that BCTV genotypes vary within and among strains. Three distinct genotypes were identified with the suggestion of interstrain recombination occuring between the Worland and CFH strains.

1.3 Geminivirus recombination

Although there are limits to what can be said about the evolution of geminiviruses, comparative sequence analyses along with geographical distributions and biological comparisons can yield realistic conclusions. From the point of view of gene function and organisation, the *Geminiviridae* are a collection of viruses with similar architecture, genome structure and replication mechanisms with other important features specific to the genus or subgroup. It seems indisputable that recombination has and is playing a role in this curious mix of sequence conservation and divergence among the *Geminiviridae*. The types and mechanisms of recombination in the geminivirus group observed in experimental work has been reviewed by Bisaro (1994).

Although there is phylogenetic evidence for ancient recombination events between geminiviruses, recombination between wild-type components does not appear to be a frequent occurrence since most recombinational events are not likely to be advantageous to the recombined product (Bisaro, 1994). In addition, small regions of the IRs of independent bipartite viruses are conserved in their respective A and B components, suggesting that recombinational homogenisation does not occur frequently (Bisaro, 1994). The sequence differences that do exist presumably do not confer a difference in the rate of replication between the two components, since A and B components appear in equal amounts in infected plants. It is possible however, that the conserved variances between the IRs of Begomoviruses are required for differential expression of A and B component genes, as both leftward and rightward promoters are located in the IRs (Bisaro, 1994). If recombination

between different geminiviruses is relatively frequent however, most recombinants (the majority of which would be deleterious) would be "sifted out" during the growth of the plant host, presumably due to the higher rates of replication, assembly and spread of the wild-types as opposed to that of the less compatible recombinant viruses.

1.3.1 *Illegitimate recombination*

Sequence rearrangements such as deletions and insertions along with other rearrangements have been detected in geminivirus mutant experiments. These rearrangements result from homologous recombination, or alternatively, recombination mediated by little or no sequence homology (non-homologous or illegitimate recombination; Stanley 1985; Bisaro, 1994). Illegitimate recombination mechanisms operating on geminivirus DNA would include aberrant breakage-fusion events mediated by topoisomerases as well as errors of DNA replication. The types of illegitimate recombinational events that have been observed are; i) the release of infectious viral DNA from recombinant plasmids containing viral DNA monomers (Stanley and Townsend, 1986; Morris *et al.*, 1988), ii) the deletion of foreign sequences from geminivirus expression vectors (Hayes *et al.*, 1989; Elmer and Rogers, 1990; von Arnim and Stanley, 1992), iii) the reversion of viral deletion mutants to wild-type size (Stanley and Townsend, 1986; Ward, 1988; Etessami *et al.*, 1989; Klinkenberg *et al.*, 1989; Gardiner *et al.*, 1988; Hayes *et al.*, 1988a, 1988c) and iv) the production of subgenomic DNAs (Hamilton *et al.*, 1982; Stanley and Townsend, 1985; MacDonald *et al.*, 1988; Frischmuth and Stanley, 1992; Stenger *et al.*, 1992).

Mechanical inoculation of recombinant plasmids containing a single copy of A or B genomes of ACMV with their respective complementary genome components onto tobacco plants was shown to elicit infection (Stanley and Townsend, 1986). These plasmids can also be viewed as viral genomes containing a plasmid vector. This clearly demonstrated the release of infectious viral DNA from the plasmid vector or - the corollary - the loss of the interfering plasmid DNA from the viral genome. This was surprising since there was little or no homology between the two combined DNAs, thus indicating that infectious replicons were generated from non-homologous recombinational events. On closer investigation of resulting infectious replicons, small deletions of viral DNA were observed at the cloning sites and within non-essential sequences of the viral genome. These observations indicate selective pressures on oversized viral replicons, or more specifically, on foreign DNA contained in viral replicons. This selective pressure is further corroborated by observations of intermolecular homologous recombinational events between DNA A and DNA B containing plasmid DNA (Stanley and Townsend, 1986). In some progeny DNA B, deleted sequences extending from the cloning site just upstream of the BC1 ORF and the plasmid

DNA including some sequence of the IR was replaced by corresponding sequence of DNA A. This recombinational event appeared to have been mediated by homologous sequences (5-11bp) at the crossover points. Non-homologous recombinational events were also observed by Morris, *et al.* (1988) when they agroinoculated (see section 1.4) plants with Ti-plasmids containing a single copy of ACMV. It is interesting that monomers of ACMV cloned into plasmid vectors are shown to be infectious while similar constructs of related geminiviruses are not (Bisaro, 1994).

The tendency of geminiviral genomes constructed with foreign DNA to revert to wild-type size in host plants is a problem that needs to be taken into account when constructing geminivirus-derived expression vectors. However, geminivirus vectors are generally stable in transient expression systems such as protoplasts and leaf discs and have been shown to tolerate and express foreign genes as large as 1.5-2kb to give an approximate total unit-length vector size of 4kb (Brough *et al.*, 1992). Foreign genes in replicating extrachromosomal vectors in plants transformed with tandem repeats of a unit-length vector can however be unstable, especially when the insertion length is larger than the size of the wild-type (Hayes *et al.*, 1988a and 1988c; Hayes *et al.*, 1989; Elmer and Rogers, 1990). An instance was demonstrated by Elmer and Rogers (1990) where replicons isolated from plants agroinoculated with TGMV vector containing a β-glucoronidase (GUS) replacement of the CP gene, reverted from the 3.8kb unit-length vector size to 2.5kb by rearrangements that fused different sequences in the T-DNA and deleted GUS sequences. Although the mechanisms of these rearrangements are likely to be mediated by non-homologous recombination, exactly how these events occurred is a mystery, since sequences contained in deleted replicons were not from the TGMV-GUS vector. It appears evident that the selective pressure that favours wild-type sized replicons is dependent on an aspect at the level of systemic movement since: i) 3.8kb TGMV-GUS vectors stably replicate in leaf discs and transformed plants thus ruling out the level of replication as the selection factor; and ii) factors at the levels of encapsidation and assembly are also ruled out since the CP gene was replaced by GUS (Bisaro, 1994; Elmer and Rogers, 1990). Selection on the basis of replicon size, but also on the basis of the deletion of BC1 sequences of TGMV, was observed by von Arnim and Stanley (1992) in experiments revealing similar types of rearrangements.

An ACMV CP mutant with a deletion of 76 bp retained the ability to spread systemically and elicit wild-type symptoms (Stanley and Townsend, 1986). Subsequent to this observation ACMV and closely related TGMV CP replacement vectors were constructed for the purpose of expressing foreign genes with the control of the CP promoter in plants (Ward *et al.*, 1988;

Etessami *et al.*, 1988 and Hayes *et al.*, 1988c). However extensive CP deletions (727 bp) rendered the the genomic component non-infectious when mechanically inoculated onto host plants, but infectivity was restored when DNA of approximately the size of the deleted CP DNA (i.e chloamphenicol acetyl transferase [CAT] gene) rendering the mutant 60bp larger than wild-type size, replaced the deletion (Ward *et al*, 1988). Surprisingly, some ACMV CP deletion mutants, although replication competent in a protoplast system, are associated with reversion of DNA A to wild-type size (Etessami *et al.*, 1989). Analyses of these illegitimate recombinants revealed that the original and newly deleted sequences were replaced by a tandem repeat of DNA A sequence from elsewhere in the genome (Etessami *et al.*, 1989). The 5' terminus of the repeat was located in the invariant nonanucleotide sequence TAATATT↓AC suggesting that illegitimate recombination following nicking of the plus strand, preceded replication. The mutants were able to replicate in protoplasts and leaf discs suggesting that their deletion in plants was due to a selection imposed by the viral movement system for molecules of wild-type size (Etessami *et al*, 1989; Klinkenberg *et al*, 1989). However different inoculation procedures suggested that the more stringent size requirement for cell-to-cell movement is relaxed for long distance movement through vascular systems.

Genome size selection and/or size reversion is not universal among geminiviruses as was demonstrated with stably replicating CP deletion mutants of TGMV eliciting delayed and attenuated symptoms (Gardiner *et al.*, 1988; Hayes *et al.*, 1988b). This is corroborated by the well-known phenomenon of the production, replication and spread of small circular subgenomic DNAs in host plants infected with geminiviruses (Hamilton *et al.*, 1982; Stanley and Townsend, 1985; MacDonald *et al.*, 1988; Frischmuth and Stanley, 1992; Stenger *et al.*, 1992). The role of subgenomic DNA in geminivirus infections is not clear, however they could function as interfering defective circular DNAs that seem to limit the high levels of replication of wild-types thereby ameliorating disease, presumably to ensure the survival of the host plant and thus ensure the long-term survival of the virus (Frischmuth and Stanley, 1991; 1993 and 1994). Because subgenomics are so prevalent in geminivirus infections, most subgenomic DNA molecules have single deletions, however two deletions can occur and more deletions are rare. Although BCTV produces subgenomic DNA molecules approximately 600 bp to 1800 bp in size, most geminiviruses generate half-sized subgenomics that are normally from populations of molecules with similar deletions (Frischmuth and Stanley, 1992; Stenger *et al.*, 1992; Bisaro, 1994). Yet subgenomic molecules all retain their *cis*-acting sequences necessary for replication and are thus invariably dependent on their respective wild-type genomes for their amplification, spread and encapsidation. Deletions are either flanked or not by short 2-7 bp repeats that are

deletions are. Subgenomics may result from errors that occur during DNA replication such as polymerase "jumping" or from some other illegitimate recombination mechanism (Frischmuth and Stanley, 1992; Stenger *et al.*, 1992).

1.3.2 *Intermolecular homologous recombination*

Intermolecular homologous recombination between geminiviruses has been observed in complementation experiments with TGMV AC1 and AC2/AC3 mutants where Brough *et al.* (1988) detected wild-type recombinants. Respective wild-types were also detected in similar complementation experiments with ACMV BC1 and BV1 mutants (Ettessami *et al.*, 1988), MSV V1 and V2 mutants (Lazarowitz *et al.*, 1989) and BCTV V1, V2 and V3 mutants (Hormuzdi and Bisaro, 1993). Wild-type recombinant genomes were found in most inoculated plants, but constituted a small fraction of the total viral DNA. This indicated that the expected competitive edge of the wild-type genomes were not dominant as expected in infected plants, however subsequent transmissions of these wild-types were found to supercede the domination of the mutants (Ettesami *et al.*,1988).

AC1 (which is essential for replication) and AC2/3 (non-essential for replication) AbMV mutants were incapable of infecting tobacco plants following agroinoculations, but AC1 mutants and not AC2/3 mutants prevented DNA replication in leaf discs (Evans and Jeske, 1993). Co-agroinoculation of both mutants and wild-type DNA B resulted in complementation preceding recombination. Plants in which complementation occurred were symptomless and contained low amounts of virus DNA whereas recombinational events established wild-type infections (Evans and Jeske, 1993)

Roberts and Stanley (1994) showed that ACMV DNA B mutants containing nucleotide substitutions within the loop sequences of the hairpin loop structure were as infectious as their wild-type counterparts. Analysis of progeny showed that in most cases wild-type sequences were restored by homologous recombination with DNA A. Stem-loop deletion mutants of both genomic components were not infectious when co-inoculated, although they were once again efficiently rescued by recombination when inoculated with the wild-type components (Roberts and Stanley, 1994). These corrections appear to be achieved by homologous recombination between common region sequences (CIRs) which is in contrast to a previous report in which a TGMV DNA B loop mutant containing an 8 bp insert was unable to infect host plants when co-agroinoculated with wild-type DNA A (Revington *et al.*, 1989).

1.3.3 *Pseudorecombination*

Pseudorecombinants are made by exchanging cloned Begomovirus A or B components with those of a related geminivirus (Bisaro, 1994; Stanley, 1985; Gilbertson *et al.*, 1993). Pseudorecombinants are potentially useful in elucidating which genomic components are responsible for replication and/or pathogenesis of Begomoviruses (Bisaro, 1994), however viable pseudorecombinants are normally restricted to those between closely related geminiviruses. Pseudorecombination between components of geminiviruses has been reported for ACMV (Stanley *et al.*, 1985; Morris *et al.*, 1990), BGMV (Faria *et al.*, 1990; Faria *et al.*, 1994), SLCV (Lazarowitz, 1991) and TGMV (von Arnim and Stanley, 1992b). However these pseudorecombinants were combinations of closely related geminivirus isolates or strains, isolated from the same plant species but from different geographical locations (ACMV, BGMV and TGMV) or isolated from a single plant (SLCV). The inability to form viable pseudorecombinants between distinct geminiviruses appears to be due to incompatibility in the IR sequences (Lazarowitz *et al*, 1992; Gilbertson *et al*, 1993).

Interestingly, Gilbertson *et al.* (1993) reported viable pseudorecombinants of tomato mottle and bean dwarf mosaic geminiviruses (ToMoV and BDMV); the first case of viable pseudorecombination between distinct viruses which were isolated from different plant species. The IR sequences of the A and B components of most bipartite geminiviruses have >95% identities due to the presence of important *cis*-acting regulatory sequences that interact with *trans*-acting factors (Fontes *et al.*, 1992), such as the AC1 product (Rep) which is necessary for replication (Elmer *et al.*, 1988; Revington *et al.*, 1989). Mixtures of the components of geminiviruses with different common regions of <75% nucleotide identities are apparently not infectious (Gilbertson *et al.*, 1991; Lazarowitz *et al.*, 1992; Stanley and Gay, 1983). It seems clear therefore that the relatively high sequence identity of 87% of the IRs of ToMoV and BDMV was sufficient for the reduced infectivity; presumably owing to less efficient interaction between heterologous combinations (Gilbertson *et al.*, 1993). The fact that infectious pseudorecombinants can be made between distinct bipartite geminiviruses has evolutionary implications: pseudorecombination may be a mechanism by which new geminiviruses evolve in nature.

In pseudorecombination experiments the movement proteins BV1 and BC1 of bipartite Begomoviruses were shown to function interchangeably and could support the movement of a heterologous viral genome from the same virus group, but with some discrimination between Old World and New World viruses (Frischmuth *et al.*, 1993). Further complementation studies provided the evidence that AC2 (TrAP) and AC3 (REn) proteins of dicot-infecting geminiviruses can complement mutations in a heterologous viral genome

(Sunter *et al.*, 1994). C3 or AC3 was found to be, to a certain extent, interchangeable between Curtoviruses and Begomoviruses respectively whereas AC2 is more specific in its function but does not distinguish between Old World (ACMV) and New World (TGMV, SqLCV, TPGV) Begomoviruses. Begomovirus AC2 amino acid sequences have a greater than 80% sequence homology whereas that of C2 BCTV has only a 65% homology with those of Begomoviruses. Thus inability of BCTV (Curtovirus) to activate TGMV (Begomovirus) virion sense promoters is reflected in the AC2 and C2 sequence divergence (Sunter *et al.*, 1994). The variation in MPs suggests evolutionary divergences of Old World and New World geminiviruses and the interchangeability of AC2 and AC3 proteins suggests that these viruses share an evolutionary progenitor.

A 682-nt circular DNA satellite associated with a TLCV isolated in northern Australia is the first satellite reported for DNA plant viruses (Dry *et al*, 1997). This molecule is dependent on the the TLCV Rep for its replication and is encapsidated by TLCV coat protein. It has no significant ORFs and shows no sequence identity to TLCV except for a TAATATTAC invariant geminivirus nonanucleotide and a sequence identical to the putative TLCV Rep binding motif (AATCGGTGTC), both of which are situated in separate putative hairpin structures. In pseudorecombination-type experiments Dry and co-workers have shown that the replication of TLVC sat-DNA can be supported by other taxonomically distinct geminiviruses such as TYLCV, ACMV and BCTV. The Rep amino acid sequence identities of TYLCV, ACMV and BCTV with TLCV are 79, 78 and 61% respectively indicating that replication of TLCV sat-DNAs is apparently sequence non-specific and not in conformity with requirements dictating the specificity of interaction of Rep with their cognate origins as predicted by the current model of replication. The sat-DNA did not appear to attenuate symptoms of any of the viruses tested with it; however satellite infection was reduced with the other geminiviruses. Where this sat-DNA originated is unknown and is unclear whether it originated as a result of an ancient recombination event since the molecule does not appear to have significant homology with any sequences in the nonredundant nucleotide database.

1.4 Agroinfection and intramolecular homologous recombination

Although inoculation of plants with purified plasmid DNA containing tandem geminivirus genome repeats by mechanical abrasion (Hayes *et al.*, 1988; Morinaga *et al.*, 1988; Stenger *et al.*, 1990a and 1990b; Louie 1995) and by biolistic particle delivery (Gilbertson *et al.*, 1991) has been reported, many geminiviruses are not easily mechanically transmissible (Louie, 1995). Geminivirus phloem limitation is presumably a factor contributing to their lack of mechanical transmissibility, but *Agrobacterium tumefaciens*-mediated transfer of these viruses or "agroinfection" has shown to be a relatively efficient mode of transmission

(Rogers *et al*, 1986; Grimsley, *et al.*, 1986 and 1987; Grimsley and Bisaro, 1987; Sunter *et al.*, 1987; Hohn, *et al.*, 1987; Elmer *et al.*, 1988b).

1.4.1 Agrobacterium-*mediated DNA transfer*

The agroinfection technique takes advantage of the natural transfer of T-DNA mechanism of *A. tumefaciens*. *Agrobacterium*-mediated DNA transfer relies on its large tumour-inducing plasmid (pTi) encoding virulence genes (*vir* genes) located outside the T-DNA region which is delimited by two almost perfect 25 bp repeat borders left and right (LB and RB respectively) (Stachel *et al.*, 1985; Bolton *et al.*, 1986; Tinland, 1996). The *vir* region of the pTi consists of six main operons *vir*A, *vir*B, *vir*C, *vir*D, *vir*E and *vir*G. *vir*A encodes an enviromental sensor protein VirA which is constitutively produced and is sensitive to low molecular mass phenolic compounds (e.g. acetosyringone) normally excreted from wounded plant tissues. Upon VirA activation of a basal amount of VirG protein by phosphorylation, VirG stimulates the transcription of all the other *vir* operons including itself (Leroux *et al.*, 1987; Winans *et al.*, 1987). VirC1 and VirC2 proteins recognise an overdrive sequence located outside the T-DNA region at the RB improving the recognition efficiency of VirD1 and VirD2 for the RB (Toro *et al.*,1989). VirD2 functions as a site-specific endonuclease which recognises and nicks a single strand of the T-DNA at the RB and also attaches itself by a phosphotyrosine bond and remains covalently bound to the 5'-end of the released single-stranded T-DNA (Howard *et al.*, 1989). VirD1 protein has topoisomerase activity and is thought to assist in removing the corresponding T-DNA strand (Ghai and Das, 1989). The ss T-DNA-VirD2 complex is exported into the host cell through a channel associated with VirB and VirE proteins (Tinland, 1996). VirE2 is a ssDNA specific binding protein which complexes with T-DNA-VirD2, protecting the ssT-DNA against nucleases in the plant cell (Rossi *et al.*, 1996). VirD2 and VirE2 are believed to target the T-DNA to the plant cell nucleus (Citovsky, V. and Zambryski, 1993), where it is integrated into the host genome. T-DNA encodes plant growth factors (oncogenes) and genes for the synthesis of opines (bacterial nutrients) and when transferred, falls under the control of the host expression signals (Tinland, 1996). Resulting neoplastic growth of infected plants create ecological niches for *Agrobacterium* in natural infections.

Any DNA inserted between the LB and RB of pTi is readily transferred as a T-complex into host plant cell nuclei, as demonstrated by A*grobacterium*-mediated transformation of plants. If the T-DNA contains partial or complete head-to-tail repeats of DNA virus genomes, unit-length genomes are released and - at least for competent host-virus combinations - replicate and spread systemically (Gardner *et al.*, 1985; Grimsley *et al.*, 1986; Rogers *et al.*, 1986).

Agroinfection has proved to be an important technical advance in the field of plant virology. Over the last decade this technique has contributed significantly to investigations into geminivirus-derived expression vectors, recombination, gene function and studies of mutant geminiviruses, enabling the elucidation of the functions of the ORFs. "Agroinfected" plants that displayed typical streak symptoms of the Mastrevirus MSV provided the evidence that *A.tumefaciens* specifically interacts with monocotyledonous plant tissue at the molecular level at a time when *Agrobacterium* was not regarded as having a graminaceous plant host (Grimsley *et al.*, 1987). Agroinfection is a potentially convenient procedure in controlled and specific infectivity tests on new crop cultivars or prospective geminivirus-resistant transgenic plants (as opposed to insect transmission tests) and can be useful in determining host range of geminiviruses (Boulton *et al*, 1989). In addition, the agroinfection technique provides a rapid and sensitive marker for the transfer of DNA to plants and therefore can be used in determining compatibility between *Agrobacterium* strains and host plants before attempting to make *Agrobacterium*-mediated transgenic plants. The agroinfection technique has exploitative potential in the study of plant molecular genetic systems; for example, agroinfectious MSV constructs have been used to introduce transposable elements in maize plants, thereby opening up investigations into the use of this system as a tool to investigate the excisions and insertions of transposable elements in plants (Shen and Hohn, 1992; Shen *et al.*, 1992).

1.4.2 *Mechanisms of geminivirus DNA release*

Tandem repeats of the genome (A or B) of dicot-infecting geminiviruses (TGMV) cloned into *Agrobacterium* have been shown to be incorporated into the genome of host cells by the establishment of A or B component transgenic plants. (Rogers *et al.*, 1986). Presumably unit-length RF-DNAs are released from infectious constructs after the integration with the host genome; however in the case of tandem repeats where plasmid DNA is physically transferred by mechanical inoculation or biolistic bombardment, this is not the case. Further, recent work conducted by Rigden and co-workers (1996) on TLCV showed that ss or ds RF-DNA forms accumulate in the *Agrobacterium* cell. This raises the question whether unit-length single-stranded viral genomes are also transferred by the *Agrobacterium* transfer mechanism. Whatever the situation, infectious constructs need not be complete tandem repeats and as little as 1.2 unit-length copies of the viral genome are sufficient to release an infectious unit-length genome (Elmer *et al.*, 1988b). However, constructs containing two IRs have higher (up to 100%) agroinfectivity rates than those with only one (Elmer *et al.*, 1988b). This may be a consequence of the mechanism by which unit-length genomes are released from the repeats (Bisaro, 1994).

Two mechanisms have been put forward for the release of covalently closed unit-length viral DNA (in effect, RF-DNA) from tandem genomic repeats (Figure 1.5; Rogers *et al.*, 1986; Elmer *et al.*, 1988b; Stenger *et al.*, 1991). Release could occur by intramolecular homologous recombination to give free covalently closed circular DNAs which then function as replicative forms (RF-DNA) for the production of genomic viral ssDNA. Alternatively, ssDNA may be released by RCR from the tandem repeats, which formally have the same structure as RF-DNA (Figure 1.5; Elmer, 1988b; Stenger *et al.*, 1991). These two mechanisms are however not mutually exclusive.

Intramolecular homologous recombination resulting from a single cross-over can occur at random locations within the tandem viral genome repeats. Thus a number of recombinant progeny genotypes would be expected if the parental repeat genomes were derived from different distinguishable strains. The mechanism of release was addressed in studies by Stenger *et al.*, (1991), using tandem repeats of different BCTV strains. When two putative plus-strand origins of replication (i.e. hairpin loops with invariant TAATATT↓AC in IRs) were present in inocula, the replicative release mechanism was favoured that resulted in progeny genomes of a single predominant genotype containing sequences that mapped between the hairpin loops. However, when inocula contained tandem repeats with only one origin of replication, unit-length viral genomes were generated by random intramolecular recombination events. This was evident since a number of different progeny genotypes were produced and simultaneously replicated in infected plants. Along with other considerations this was an indication that geminiviral replication occurs by RCR (Lazarowitz, 1992; Rogers *et al.*, 1986; Saunders *et al.*, 1991). In addition, these experiments indicated that replicative release is a highly efficient mechanism of release between two hairpin loops in tandem repeats, predominating over random homologous recombination in agroinfections. However recombinational "hotspots" or preferred recombinational sites between different strains of geminiviruses could not be ruled out (Bisaro, 1994).

In similar studies on WDV complete heterodimers containing head-to-tail genomes with different coat protein replacements (V1 replaced by luxA or luxB genes), the genome unit flanked by two LIRs was preferentially released in barley suspension culture cells rather than the one only flanked by one LIR (Heyraud *et al.*, 1993). A third copy of the LIR on the inoculum, so that both genome units were flanked by two identical LIRs, resulted in the release of both genomes with similar efficiency, indicating that the release results primarily from RCR but intramolecular homologous recombination was shown to occur at low levels simultaneously (Heyraud *et al.*, 1993).

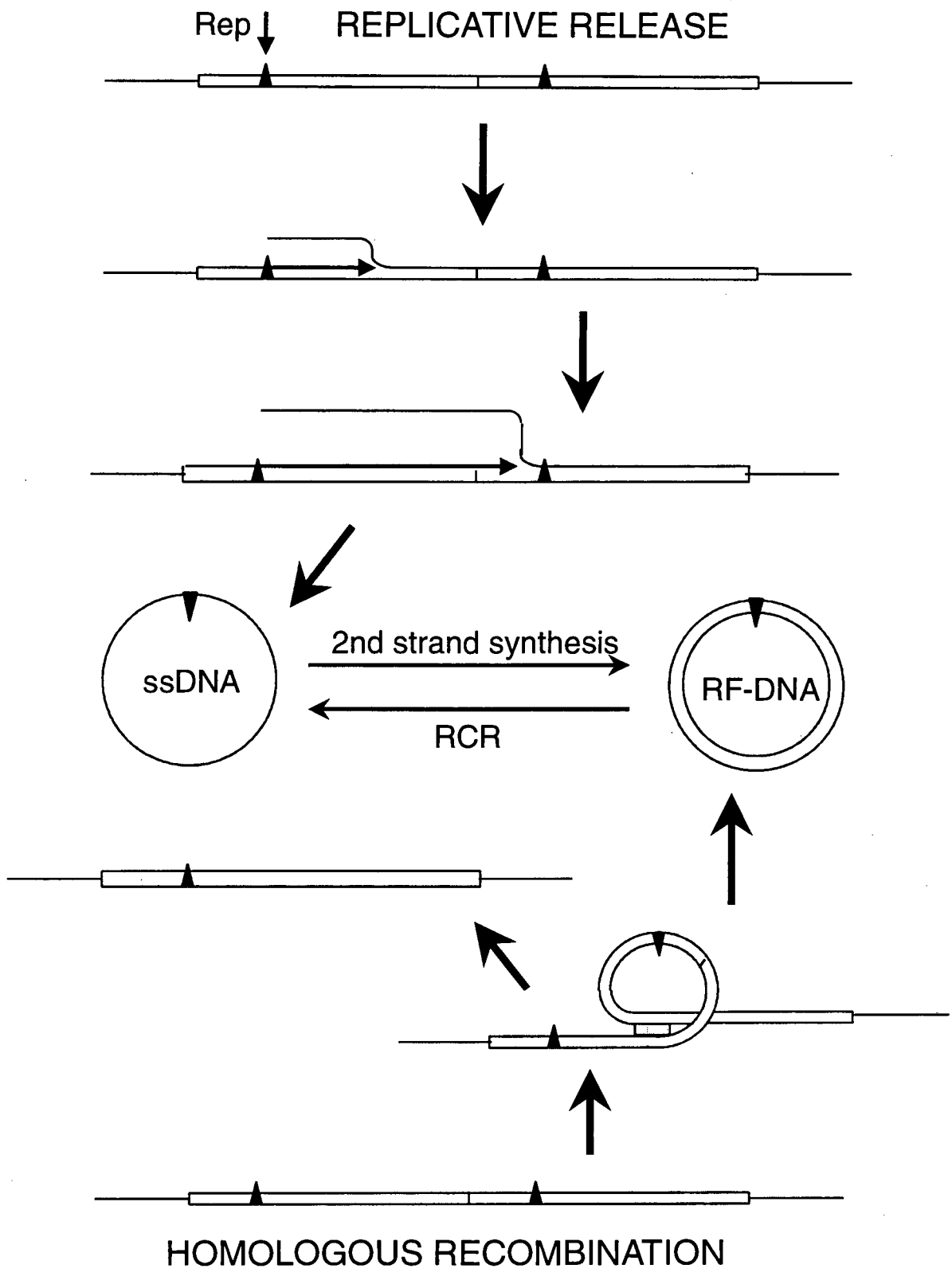

Figure 1.5 Release of unit-length geminivirus DNA from tandem repeats

Two mechanisms are depicted: homologous recombination and replicative release. Black triangles indicate the positions of the invariant TAATATTAC nonanucleotide of hairpin loop structure in the intergenic regions of geminiviruses. For replicative release Rep initiates replication and ssDNA is released which is subjected to second strand synthesis and thus rolling circle replication (RCR) follows. Homologous recombination occurs where double-stranded DNA crosslinks with homologous sequences as indicated by shaded region. (Adapted from Stenger *et al.*, 1991)

Another factor that had not been addressed before 1996 was the possibility that recombination occurs in the *Agrobacterium*. Rigden *et al.* (1996) showed for TLCV that tandem repeats of viral genomes cloned into an *Agrobacterium* plasmid are able to generate RF-DNAs in *Agrobacterium* indistinguishable from those produced in infected plants. This accumulation depended on the presence of two LIRs with origins of replication but accumulation was substantially reduced when mutations were introduced in Rep or the C1 ORF, indicating that active viral-induced replication is supported by the bacterial cellular machinery. Replication was not affected when mutations were introduced in the other viral ORFs. This evidence suggests that single-stranded DNA viruses may have evolved from prokaryotic episomal replicons (Rigden *et al.*, 1996).

1.5 Geminivirus tissue invasion, cytopathology and symptomatology

1.5.1 *Tissue specificity*

It has been suggested that host cell division may be necessary for geminivirus replication (Townsend, 1986). Some transient replication experiments have shown that cell division seems to be a prerequisite for the appearance of RF-DNA of geminiviruses (Briddon *et al.*, 1989; Matzeit *et al.*, 1991; Boulton *et al.*, 1993). A small proportion of a population of MSV-infected protoplasts - correlating with the onset of cell division - were found to have replicating MSV (Boulton *et al.*, 1993). Accotto *et al.* (1993) demonstrated that host cellular DNA replication occurring in the S-phase of mitosis - rather than cell division - was important for DSV DNA replication. In contrast to these observations however, experiments have detected geminivirus replication in the absence of cell division (Brough *et al.*, 1992; Timmermans *et al.*, 1992). For example, Timmermans and co-workers (1992) demonstrated that a vector based on a replicon from WDV initiated replication independently of cell division. In primary callus or suspension cultures of AbMV-infected leaf discs where cells are actively dividing AbMV was not detected in Southern blot analysis (Song and Jeske, 1994).

These conflicting results have been reconciled in immunolocalisation, immunohistochemical and *in situ* hybridisation studies on AbMV, MSV and TGMV infected plants (Horns and Jeske, 1991; Lucy *et al.*, 1996; Nagar *et al.*, 1995). No AbMV DNA was detected in the meristimatic tissue of infected plants (Horns and Jeske, 1991). Lucy and co-workers (1996) demonstrated in a comparison of the distribution of MSV RF-DNA with the expression of the S-phase specific host gene (H2b) that host DNA synthesis is not essential for MSV replication. Nagar and co-workers (1995) corroborate this finding for Begomoviruses in similar immunolocalisation of the Rep and REn proteins and of a host factor associated with DNA synthesis, PCNA (host DNA synthesis protein proliferating cell nuclear antigen and an

accessory factor for DNA polymerase δ) to identify and characterise cell types that supported geminivirus replication in TGMV-infected and transgenic plants. They found that Rep and REn of TGMV are located in terminally differentiated cells rather than actively dividing meristematic cells and found that the presence of these proteins induced accumulation of PCNA (normally only associated with actively dividing cells). Rep and REn were detected in epidermal, mesophyll and phloem-associated cells of the leaf as well as the differentiated cells of the outer cortex, xylem, parenchyma, phloem and pith of the infected tobacco stem (Nagar *et al.*, 1995).

It was once accepted that geminiviruses were generally limited to the phloem of infected plants (Harrison, 1985). Ultrastructural and immunological studies of AbMV-infected plants revealed geminivirus particles and coat protein only in phloem tissue (Horns and Jeske, 1991). A reinvestigation involving *in situ* hybridisation techniques revealed that AbMV DNA was not detected in the tissue which show the predominant cytopathological effects, palisade and spongy parenchyma. Hence the AbMV accumulated in the phloem while symptoms are observed in other tissues, and was undetected in meristematic tissue (Horns and Jeske, 1991). In MSV infections, viral coat protein and both positive and negative strands of MSV DNA were detected in mesophyll, vascular-associated parenchyma, and bundle sheath cells of the leaf (Lucy *et al.*, 1996). However, in the apical meristem, MSV was only present in the already differentiated vascular tissues and not detectable in actively dividing meristematic tissue. Although Lucy *et al.*, 1996 did not detect MSV in nonphotosynthetic tissue outside of the vasculature, previous workers reported that MSV was found to infect all tissues including the epidermis whereas CSMV did not infect epidermal cells (Bock *et al.*, 1974; Hatta and Francki, 1979; Francki *et al.*, 1979).

1.5.2 *Cytopathology*

Ultrastructural studies of geminiviruses have revealed viral particles in nuclei and chloroplasts of infected cells, as well as in differentiated sieve tube elements which lack nuclei at maturity (Nagar *et al.*, 1995; Rushing *et al.*, 1987; Groning *et al.*, 1990; Horns and Jeske, 1991). Distorted nuclei (swollen) with fibrillar bodies and viral aggregates have been observed in electron micrographs of TGMV-infected cells (Rushing *et al.*, 1987; Nagar *et al.*, 1995). Infected necrotic cells lack chloroplasts and can be identified by the detachment of the plasma membrane from the cell wall (Nagar *et al.*, 1995). Ultrastructural examination of leaf tissue of tobacco infected with Indian cassava mosaic virus (ICMV) revealed abnormalities in the phloem and occasionally, xylem cells (Roberts, 1989).

Ultrastructural studies of Mastrevirus-infected cells revealed aggregates of viruses in the nuclei of a variety of cell types, including bundle sheath and mesophyll cells (Pinner *et al.*, 1993); linear aggregates or large crystalline arrays of geminate particles in the nuclei of DSV-infected companion and phloem parenchyma cells (Dollet *et al.*, 1986); fibrillar rings in CSMV-infected cell nuclei and segregation of the nucleolus (Francki *et al.*, 1980); aggregates of virus-like particles in the cytoplasm of intact cells; crystalline arrays of DSV (Dollet *et al.*, 1986); extensive sheets of PSMV (Greber, 1989) and random aggregates of CSMV and geminiviruses isolated from *Bromus* and *Digitaria* grass species (Francki *et al.*, 1980; Greber, 1989). For Begomoviruses, accumulation of amorphous masses or paracrystalline AbMV inclusions in phloem was shown by Horns and Jeske, 1991; granular inclusion bodies composed of virus-like particles were detected by immunogold labelling techniques in the nuclei of ICMV-infected cells (Roberts, 1989). In progressed ICMV infections hollow spheres made up of fribillar material were produced. On the other hand, CSMV and geminiviruses causing streak in sugarcane and *Panicum* exhibited non-crystalline inclusions (Pinner *et al*, 1990). More recently however, cytopathology and ultrastructural studies of MSV and SSV infected cells revealed crystalline and non-crystalline aggregates (Ammar, 1994). Cytopathological changes in these cells were confined to the nuclei - like in the TGMV-infected cells - which were usually larger than normal, with peripheral chromatin and nucleoli. The nuclear envelope of some inclusion-containing nuclei were ruptured with an occasional crystalline aggregate in the cytoplasm of cells in which no intact nuclei were detected (Ammar, 1994).

1.5.3 *Mastrevirus symptomatology*

The chlorosis on leaves of infected plants as a result of infection is due to the failure of chloroplasts to develop in the tissue surrounding the vascular bundles. The symptoms of MSV on maize plants develop as chlorotic thick continuous streaks for severe strains or thin discontinuous stippled streaks for milder varieties. The development and extent of these symptoms can be correlated with the age of the leaf at the time of inoculation (Peterschmitt *et al.*, 1992). Lesion colour can also vary from white to yellow (Pinner *et al.*, 1988). Maize streak disease symptoms develop as small spots on lowest portion of the first exposed leaf. Only newly formed leaf tissue develops symptoms whereas below the point of infection the leaf remains healthy. Spots develop into streaks and streaks can join to form thicker streaks and eventually can extend to the entire length and breadth of the leaf (Hughes, 1991). Reduced photosynthesis and increased respiration leads to a reduction in leaf length and breadth and plant height (Hughes, 1991; McClean, 1947; Rose, 1978). Thus maize plants infected at an early stage become severely stunted, producing undersized, misshappen cobs or give no yield at all (McClean, 1947; Rose, 1978).

1.6 Transmission and host ranges of Mastreviruses

It has been reported that six different species of *Cicadulina* transmit MSV with *C.mbila* Naude being the most common and ubiquitous vector (Damsteegt, 1983). None of the recognised natural *Cicadulina* species have been reported from the Western Hemisphere, however other Mastreviruses such as DSV, WDV, CSMV and MiSV are transmitted by other insect genera of leafhoppers (Damsteegt 1983). For example DSV and CSMV are transmitted by two different *Nesoclutha* spp. (Julia and Dollet, 1989; Pinner *et al.*, 1988; Grylls, 1963). WDV is transmitted by a *Psammotettix* sp. (Lindsten *et al.*, 1980) and the vector for MiSV is unknown (Ikegami *et al.*, 1990; Chatani *et al.*, 1991).

In general the frequency of transmission of MSV by an insect vector increases with an increase in virus concentration in the plant, the duration of the acquisition access period, or the duration of the inoculation (Okoth, *et al.*, 1987; Grylls, 1963). Leafhoppers are able to acquire MSV after an acquisition period of a minimum of 15 seconds, however they may require up to 1 hour from infected mesophyll plant tissue (Storey, 1938). Before transmission can occur, there is a latent period in which there are two barriers to infection; the insect's gut lining and virus entry into the salivary glands (Storey, 1938). This length of this period in which the virus passes from the insect's gut lumen through the hoemocoel and to the salivary glands - where the virus is available for inoculation - is temperature dependent (Storey, 1928). At 30°C the minimum latent period is 6-12 hours whereas at 16°C it is at 85 hours. All 5 nymph instars are able to acquire and transmit, however the virus cannot be transmitted to progeny (Storey, 1928). Mastreviruses also have not been shown to be seed-borne in host plants (M. B. von Wechmar, pers. comm.).

The wide host ranges of Mastreviruses in part reflect the feeding range of the insect vectors. In an extensive MSV host range and vulnerability of maize germ plasm study, several grass species in the genera *Aegilops, Andropogon, Avena, Bothrtochloa, Digitaria, Echinochloa, Eleusine, Hyparrhenia, Panicum, Schizachrium, Sorghastrum, Sorghum, Trichachne, Trichloris and Zea* supported abundant oviposition and nymphal development of the vector *C. mbila,* (Damsteegt, 1983). However the feeding range of leafhoppers can be limited to preferential or more palatable host plant species, thereby allowing susceptible plants to escape infection in leafhopper transmission tests (Damsteegt, 1983).

Distinct geminiviruses from different parts of the world that are quite distinct from each other can have overlapping host ranges. For instance CSMV, PSMV (*Paspalum* striate mosaic virus), MSV and WDV share a number of cereal crop hosts and/or several wild grass hosts. More specifically, WDV, MSV, and CSMV infect *Hordeum vulgare*; CSMV, PSMV and

MSV infect *Z. mays*; and MSV and CSMV infect *Chloris, Eleusine, Lolium, Paspalum* and *Setaria* spp. (Greber, 1989; McClean, 1947; Damsteegt, 1983; Pinner *et al.*, 1988).

MSV has been reported to be a major production constraint in African lowlands and mid-altitude growing areas becoming progressively less severe as the altitude increases (Damsteegt, 1983). Damsteegt (1983) found that a severe "A-type" MSV obtained from Potchefstroom, South Africa, was able to infect 54 of 138 grass species from Andropogonoideae, Chlorodoideae, Festucoideae, and Panicoideae subfamilies in *C. mbila* transmission tests. The 54 susceptible grasses were from 33 different genera. Symptoms varied greatly probably due to varied *C. mbila* feeding preferences and/or plant heterogeneity. Although all the 529 maize hybrids, inbreds, exotic lines and sweetcorn cultivars were susceptible (except for two resistant hybrids, Revolution and J-2705), two *Z. mays* subsp. collections from Guatemala (*parviglumis* and *huehuetenangensis*) were resistant. MSV also infected several cereal crop cultivars such as barley, wheat, rye, oats and rice (Damsteegt, 1983). CSMV was found to infect 15 grass cultivated cereals genera including oats, barley, wheat, maize and the *Dacteyloctenium*, *Leptochloa* spp. (Greber, 1989).

1.7 An historical account of the African Mastreviruses

MSV was first observed as a crop disease by Fuller in 1901 as "mealie variegation" which was reported to be responsible for losses in the maize crop in South Africa at the time. By the late 1920s Storey had established the disease as "maize streak", and it was found to be obligately transmitted by leafhoppers (Rose, 1978). The disease caused by MSV is endemic in Africa and neighbouring Indian Ocean islands such as Mauritius and La Reunion and has also reportedly been found in India (Damsteegt, 1983). Breeding programmes in African countries had determined the mode of inheritance of resistance to these diseases by 1967 (Storey and Howland, 1967). In 1974 it was established by Bock and co-workers in Kenya that disease was due to "geminate" particles which are now known as the geminiviruses. Four years later (1978) Harrison and co-workers showed that the genomes of these plant viruses - previously thought to be ssRNA - were in fact, single stranded circular (ssc) DNA molecules.

These were established as being the smallest known autonomously replicating viruses when the first sequence of a geminivirus genome (ACMV) in its entirety was published in 1983 by Stanley and co-workers and those of the first Mastreviruses were published in 1984 by Mullineaux and co-workers (MSV-Nig) and - separately - by Howell (MSV-Ken). In 1985 the entire sequence of the European WDV was published by MacDowell and co-workers. By

1986 the first Curtovirus sequence was reported by Stanley *et al.* (1986). The advent of the agroinfection technique (Grimsley *et al.*, 1987) established that the MSV 2.7kb single-component genome was indeed viable and sufficient for infection - when there were doubts about the competency of this very small single component genome (Rybicki, 1988).

Mastreviruses were also found to infect indigenous African grasses (Storey and MacClean, 1930; Rose, 1978), where "maize streak" probably originated, since crops such as maize, sugarcane and barley are introduced plants in Africa (Bock, 1982; Briddon *et al.*, 1992). Depending on host source, one virus isolate may show different symptoms and host range to another. (Bock *et al.*, 1974). MSV can cause major crop losses, with the yield loss depending on the strain of the virus. Early methods of MSV typing by means of vector transmission and assessment of plant symptoms (Storey and Maclean, 1930; McClean 1947; Bock *et al.*, 1974) indicated that distinctly different viruses existed, with differing host ranges, and which did not cross-protect against one another. Different viruses such as one causing streak in *Panicum maximum* were designated as "host-adapted" strains of MSV (Bock *et al*, 1974). Earlier symptom, transmission, and serological studies indicated that all African grass Mastreviruses could be classified as strains of MSV. More recently however, serological analyses identified three distinct groups of strains: the maize, *Panicum maximum* and the sugarcane strains (Dekker *et al.*, 1988; Pinner *et al*, 1988; Pinner and Markham, 1990). Recently, the first African dicot-infecting Mastrevirus - BeYDV - was reported from South Africa (Liu *et al.*, 1997): this is the second dicot-infecting Mastrevirus discovered, the first being TYDV (Morris *et al.*, 1992).

Serological, DNA cross-hybridisation, restriction endonuclease mapping, PCR DNA amplification and sequencing data accumulating in this laboratory have established that the Natal and Mauritius SSVs are distantly related strains, whereas maize MSV isolates from widely separated geographical regions have remarkably similar genomes (Clarke *et al.*, 1989; Rybicki, *et al.*, 1989, Rybicki and Hughes, 1990; Hughes *et al.*, 1992). Comparative studies of this nature outside of this laboratory revealed the existence of an SSV/MSV/DSV group of viruses distinct from CSMV, WDV or MiSV (Howarth and Vandemark, 1989; Rybicki and Hughes, 1990; Hughes, 1991; Chatani *et al.*, 1991; Pinner *et al.*, 1992). Comparative studies in this laboratory on southern African graminaceous geminiviruses corroborating Pinner and co-workers results (1988; 1990; 1992) revealed three distinct groups; the MSVs, the PanSVs isolated from *Panicum maximum* and the SSVs isolated from sugarcane (Hughes *et al.*, 1992).

1.8 Rationale for project

Three Mastreviruses were previously isolated in this laboratory from different geographical regions in South Africa and from different graminaceous hosts; MSV-Kom was isolated from maize in Karino, Mpumalanga, MSV-Set was isolated from a *Setaria* grass species in Mt. Edgecombe, Kwazulu/Natal and PanSV-Kar was isolated from *Panicum maximum* in Komatipoort, Mpumalanga (Hughes *et al.,* 1992). The RF-DNAs of these were previously cloned into pUC plasmid vectors; pKom500, pSet100 and pPS100 respectively (Hughes 1991).

One of the aims of this thesis was to investigate the biological and genomic characteristics (Chapters 2 and 3 respectively) of these three distinct South African Mastreviruses from different geographical regions. Agroinfectious clones of the RF-DNAs of MSV-Kom, MSV-Set and PanSV-Kar contained in plasmids pKom500, pSet100 and pPS100 respectively would firstly confirm that the RF-DNAs contained in pPS100, pSet100 and pKom500 respectively were infectious, and secondly, the host range and symptomatology of these geminiviruses could be determined on an available range of differentials consisting of several maize, wheat and barley cultivars as well as *Digitaria* and *Panicum* grass species (Chapter 2).

With the establishment that the genomes contained in pPS100, pSet100 and pKom500 were infectious (Chapter 2), the infectious genome of MSV-Set was characterised by determining the entire nucleotide sequence of the genome contained in pSet100. The genomic sequencing of MSV-Set and characterisation of the genome along with those of MSV-Kom and PanSV-Kar (sequences previously determined by Dr Fiona Hughes, Ms Di James and Mr Mark Fyvie in this laboratory) is presented in Chapter 3. The phylogenetic relationships of MSV-Set, MSV-Kom and PanSV-Kar with respect to their nucleotide and protein amino acid sequences were also determined (Chapter 3).

Before the onset of this project, two mechanisms of release (intramolecular homologous recombination or replicative release) of unit-length geminiviral RF-DNA from tandem genomic repeats in BCTV agroinfection tests had been proposed (Rogers *et al*, 1986; Elmer *et al*, 1988b; Stenger *et al*, 1991). Stenger and co-workers had reported the agroinfection of heterodimers of the BCTV Logan and Worland isolates which shared a complete nucleotide sequence identity of approximately 80%. The analyses of the recombinants revealed that the majority unit-length genomes released during agroinoculations occurred from between two LIRs; however a few recombinants presumed to be released by intramolecular recombination were also present. Stenger *et al.*, (1991) mapped the origin of replication to a

20 bp BCTV sequence within the conserved hairpin indicated that geminivirus replication occurred by a rolling circle mechanism.

The present project was aimed at constructing heterodimers of two biologically and genomically distinct South African Mastreviruses - in this case, MSV-Kom and MSV-Set - in an effort to generate chimaeric Mastrevirus recombinants (Chapter 4). The rationale for this was to use generated MSV-Kom/MSV-Set recombinants in agroinfection and leafhopper transmission tests in an attempt to determine the genomic domains responsible for host range and symptom severity of Mastreviruses. During the course of this work Heyraud *et al.* (1993) had reported that the TAATATTAC conserved nonanucleotide in the LIR was the region within which the release of unit-length molecules occurred in similar experiments using WDV heterodimers consisting of a Swedish and a barley adapted strain which shared a nucleotide sequence identity of 82.5%.

Since geminiviruses are a collection of viruses sharing similar genome structure as well as a similar replication mechanism, but with genotypic and phenotypic functions that may be unique to a subgroup or even a single virus, it is not surprising that chimaeric recombinant genomes can be constructed in the laboratory between closely related viruses or even between different subgroups. For example, Briddon *et al.*, (1990) - as mentioned above - have exchanged the ACMV CP with that of BCTV and found that the engineered ACMV recombinant was specific to leafhopper transmission as opposed to wild-type ACMV's white-fly vector. This result elegantly indicated that the specificity of leafhopper transmission from insect to plant resided with the CP. In an attempt to identify regions of the BCTV genome that determine strain-specific pathogenic properties Stenger *et al.*, (1994) constructed chimaeric viruses containing portions of the Logan and CHF BCTV genomes. Similarly, the use of infectious MSV-Kom/MSV-Set recombinants in agroinfection tests and virus-free leafhopper transmission tests can be potentially used in determining genomic domains responsible for Mastrevirus pathogenicity and host range. Chapter 5 reports on the pathogenicities and host ranges of selected MSV-Kom/MSV-Set LIR or CP recombinants generated from agroinfections of MSV-Kom/MSV-Set heterodimers and the implications of these results are discussed.

The present work also reports on an agroinfectious construct containing two MSV-Kom LIRs flanking the four ORFs of MSV-Set (pKeSeK2; Chapter 5) which suggests that the Rep protein of MSV-Set recognises and initiates replication of MSV-Kom despite their biological and genomic differences. The implications of this result are discussed and the implications with respect to the phylogeny of Mastreviruses are also discussed.

CHAPTER 2

BIOLOGICAL CHARACTERISATION OF THREE MASTREVIRUSES

2.1 SUMMARY

The RF-DNAs of MSV-Kom, MSV-Set and PanSV-Kar contained in the plasmids pKom500, pSet100 and pPS100 respectively were infectious as determined by the agroinoculations of partial homodimeric *Agrobacterium* clones on 3-day-old Jubilee sweetcorn seedlings. The biological characters such as the streak symptoms on graminaceous hosts, the observation of geminate particles in electron micrographs of leafdip preparations, the leafhopper transmission of the symptoms from infected plants to healthy plants and the presence of viral linear and circular ds and ss DNA in infected plants, indicated that the infectious RF-DNAs comprise the entire genomes of MSV-Kom, MSV-Set and PanSV-Kar respectively. While the agroinfectivities of the constructs pKom504, pSet102 and pPS103 (1.4, 1.8 and 1.7mers respectively) with only one LIR, were 52, 33 and 26% respectively, those of the constructs pKom603 and pSet107 (1.1mers with two LIRs each) were 100 and 90% respectively in an agroinoculation experiment, indicating that replicational release of unit-length genomes from homodimeric constructs is favoured over that of homologous recombination in agroinfections. In leafhopper transmission tests on an available range of maize, wheat, barley and grass lines, cv.s and species, MSV-Kom caused severe streak in maize and more moderate streak in wheat whereas MSV-Set causes moderate to mild streak or no streak in "resistant" maize lines and severe streak in wheat cultivars. MSV-Kom causes streak, stunting and leaf curl symptoms on Clipper barley more severely than MSV-Set. PanSV-Kar causes distinctly mild streak in sweetcorn and "susceptible" maize cv.s. and does not produce streak symptoms in the available maize, wheat or barley cv.s. This chapter establishes the biological differences of MSV-Kom, MSV-Set and PanSV-Kar and the implications are discussed.

2.2 INTRODUCTION AND RATIONALE

Mastreviruses have attracted interest as autonomously replicating vectors since; i) they are some of the only two known classes of ssDNA plant viruses that replicate via a dsDNA intermediate in the plant cell nucleus (Hayes *et al,* 1988), ii) they have a broad host range amongst graminaceous plants which includes the world's most important cultivated crops such as wheat, barley and maize (Damsteegt, 1983), and iii) they are able to replicate at very high levels per nucleus of susceptible hosts (Timmermans, *et al.* 1992; Shen and Hohn, 1995; Palmer, 1997).

Mastreviruses are obligately leafhopper-transmitted, however the mechanical transmission for MSV into seedlings has been reported by vascular puncture of maize kernels using an engraving tool with mounted insect pins (<5% infection rate; Louie, 1995). Naked genomic RF-DNA cloned as tandem copies into a plasmid vector can be introduced into protoplasts via PEG-mediated DNA uptake or into callus, cell suspensions and protoplasts by microprojectile bombardment (Heyraud *et al*, 1993b; K. Palmer, pers. comm.). Although the *Agrobacterium*-mediated method of Mastrevirus transmission or "agroinfection" is a more convenient mode of transmission, leafhoppers - the natural vectors of most Mastreviruses - are generally more effective in transmitting these viruses (Storey, 1928; M. B. von Wechmar, pers. comm.). However the agroinfection technique is particularly suited when virus-free leafhoppers are not available.

Leafhoppers are adapted to feed from a wide range of monocotyledous plants (as well as dicotyledonous plants) by their ability to penetrate phloem tissue situated in the vasculature of leaves using their stylets, thereby releasing or acquiring Mastreviruses (Storey, 1938). Thus leafhoppers can be useful in determining host ranges for specific Mastreviruses, but the maintenance and containment of virus-free leafhoppers are inconvenient.

The agroinfection technique is dependent on the mechanical injection of *Agrobacterium tumefaciens* containing viral genome clones at or close to the apical meristem of the maize seedling in the coleoptilar node of three to ten day-old seedlings (Grimsley *et al.*, 1988). Undifferentiated meristematic tissue is not competent for agroinfection; however meristematic tissue becomes competent at developmental stages that correlate with differentiation of the first one or two leaf initials (Schläppi and Hohn, 1992). Single cells of the apical meristem of maize are susceptible to agroinfection when preinduced *Agrobacterium* clones are microinjected, however this response was found to be dependent on the maize plant genotype (Escudero *et al.* 1996). Furthermore agroinfection of maize was also found to be *Agrobacterium* strain-specific (Boulton *et al.* 1989; Schläppi and Hohn, 1992). Despite these complications, agroinfection is a potentially convenient procedure in controlled and specific infectivity/susceptibility tests on new crop cultivars or prospective Mastrevirus-resistant transgenic plants (as opposed to leafhopper transmission tests).

There is a group of Mastreviruses that have become host-adapted in several graminaceous species and share maize as a common susceptible host (McClean, 1947; Damsteegt, 1983; Hughes, 1991). Grass Mastreviruses such as PanSV were isolated from African grass species, although the related DSV was isolated from a *Digitaria* species in Vanuatu (Briddon *et al.* 1992; Dollet *et al.* 1986). It is thought that the Mastreviruses such as MSV and SSV

infecting introduced Graminaceous crops in Africa and its neighbouring islands originated from African grasses such as *Panicum maximum* which is widely distributed throughout Africa (Briddon *et al.* 1992). In serological, DNA cross-hybridisation, restriction endonuclease mapping or PCR DNA amplification and sequencing tests SSV, CSMV, WDV, and MiSV were shown to be distinct from MSV isolates (Howarth and Vandemark, 1989; Rybicki and Hughes; 1990; Hughes, 1991; Hughes *et al.* 1991; Chatani *et al.* 1991). However MSVs isolated from maize from widely separated geographical regions were found to have remarkably similar genomes (Clarke *et al.* 1989; Rybicki *et al.* 1989; Rybicki and Hughes 1990; Hughes *et al.* 1991).

MSV-Set, a virus isolated from a *Setaria* sp. was found to be a possible strain of MSV since the DNA cross-hybridised with MSV DNA in Southern blot tests, and their restriction endonuclease maps featured conserved sites (Hughes, 1992). PanSV-Kar was isolated from a *Panicum* sp. and the partial sequence data and restriction endonuclease map resembled that of PanSV-Ken isolated from a *Panicum* sp. in (Hughes *et al*, 1992, Briddon *et al*, 1992).

PanSV-Kar and MSV-Set were isolated from African grass species which displayed mild Mastrevirus-like streak symptoms (Hughes *et al.* 1992). In leafhopper transmission tests PanSV-Kar and MSV-Set produced milder symptoms in a differential panel of MSV-susceptible maize cv.s. than the maize isolates of MSV. In contrast, MSV-Kom was isolated from maize with severe streak symptoms and produced severe streak in a panel of differential hosts in leafhopper transmission tests (Hughes *et al.* 1992). MSV-Kom, MSV-Set and PanSV-Kar are from Komatipoort in Mpumalanga, Mt. Edgecombe in Kwazulu/Natal and Karino in Mpumalanga respectively (different geographical regions in South Africa). Genomic DNAs of MSV-Kom, MSV-Set and PanSV-Kar (RF-DNAs) were previously cloned into pKom500, pSet100 and pPS100 respectively (Hughes *et al.*, 1992).

This chapter reports the distinct biological features of MSV-Kom, MSV-Set and PanSV-Kar. pKom504, a 1.4mer tandem dimer of the MSV-Kom genome, was previously found to be infectious in agroinfection tests thereby ascertaining that the genome contained in pKom500 is infectious (Hughes and Rybicki, unpublished). To determine whether the cloned RF-DNAs contained in pSet100 and pPS100 respectively were viable and infectious, tandem partial dimers of these genomes were constructed and cloned into *A.tumefaciens* and agroinoculated into susceptible maize, sweetcorn or popcorn 3-day-old seedlings. This chapter reports the construction of agroinfectious clones of MSV-Kom, MSV-Set and PanSV-Kar and their relative agroinfectivities or agroinfection rates. Streak symptoms and host

ranges were established for each Mastrevirus on an available range of maize, wheat, and barley cultivars in this laboratory in leafhopper transmission tests.

2.3 MATERIALS AND METHODS

2.3.1 Origins of MSV-Kom, MSV-Set and PanSV-Kar

MSV-Kom, MSV-Set and PanSV-Kar RF-DNAs were originally isolated by Dr. Fiona Hughes in this laboratory from plants showing typical streak symptoms (Hughes 1991; Hughes *et al*, 1992; Table 2.1 for the origins of streak infected plants). RF-DNAs were cloned into the *Bam* HI restriction site of the multiple cloning sites (MCSs) of the pUC18 vector: *Bam* HI is a unique site in all three genomes. The clones were obtained for this project from Dr F. Hughes who designated them pKom500, pSet100 and pPS100 respectively (Hughes, 1991).

Table 2.1: Origins of viruses in this study

Isolate	Plant Source	RF-DNA Clone	Geographical Origin	Obtained From
MSV-Kom	*Zea mays*	pKom500	Komatipoort, Mpumalanga	M. B. von Wechmar[a]
MSV-Set	*Setaria* sp.	pSet100	Mt. Edgecombe, Kwazulu-Natal	K. Harborne[b]
PanSV-Kar	*Panicum maximum*	pPS100	Karino, Mpumalanga	M. B. von Wechmar[a]

a = University of Cape Town, South Africa.

b = South African Sugar Association Experimental Station, Mt. Edgecombe, Kwazulu/Natal.

Infectious clones of all three RF-DNAs have been entirely sequenced in this laboratory. The genomic sequence of MSV-Kom was obtained by Fiona Hughes and Di James (GENBANK accession number AF003952). The genomic sequence of PanSV-Kar was obtained by Mark Fyvie and Janet Willment (GENBANK accession number L39638) and MSV-Set was sequenced in this study (see Chapter 3, GENBANK accession number AF007881). The diagrams of the restriction endonuclease maps of the three RF-DNAs are presented in Figure 2.1. Clones pKom500, pSet100 and pPS100 had been mapped by Fiona Hughes (Hughes *et al.* 1992) using the following restriction endonucleases from Boehringer-Mannheim; *Apa* I, *Bam* HI; *Bgl* I, *Bgl* II, *Cla* I, *Eco* RI, *Hind* III, *Kpn* I, *Pvu* II, *Pst* I, *Sal* I, *Sac* I and *Xho* I (Figure 2.1).

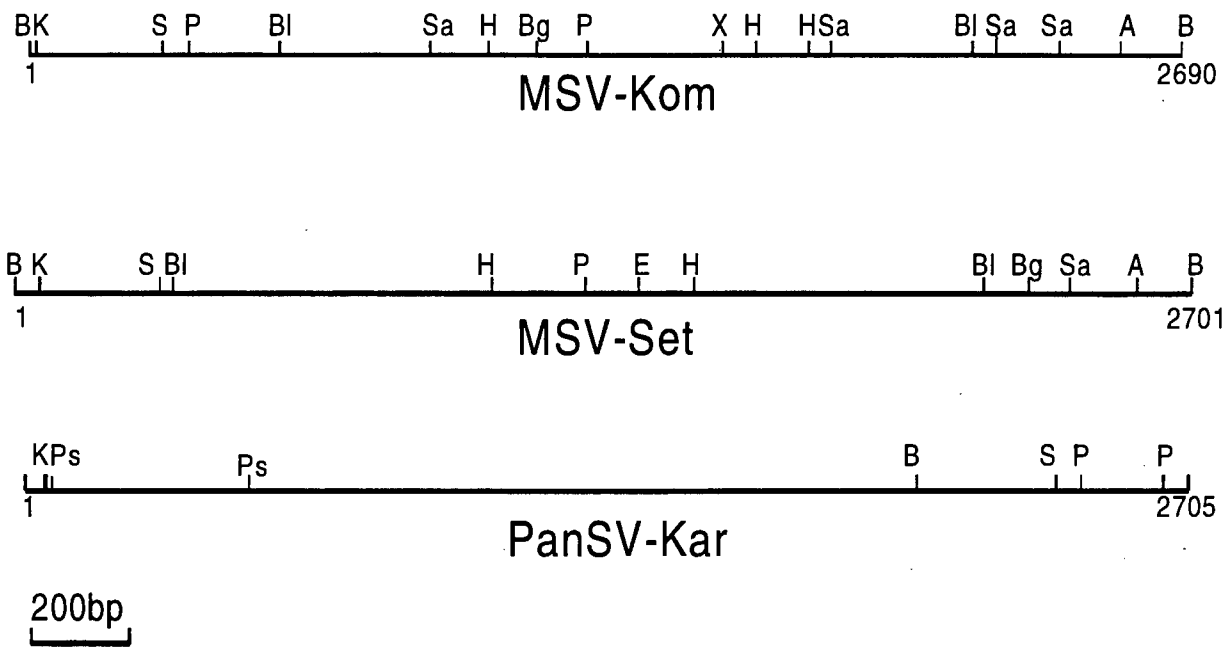

Figure 2.1: Restriction maps of viral genomes

Restriction maps aligned where the first base is the A of the ATG start codon of the V1 ORF. Abbreviations for the restriction endonuclease sites are as follows: A=*Apa* I; B=*Bam* HI; Bl=*Bgl* I; Bg=*Bgl* II; E=*Eco* RI; H=*Hind* III; K=*Kpn* I; P=*Pvu* II; Ps=*Pst* I; S=*Sal* I; Sa=*Sac* I; X=*Xho* I. MSV-Kom, MSV-Set, and PanSV-Kar genomes consist of 2690, 2704 and 2705 nucleotides respectively. These maps were derived from the sequences of MSV-Kom, MSV-Set and PanSV-Kar (accession numbers AF003952, AF007881 and L39638 respectively).

2.3.2 Construction of partially dimeric agroinfectious clones

Using standard cloning techniques (Sambrook *et al.* 1989), partial tandem homodimeric genomes were constructed in pUC18 and inserted into the broad host range vector plasmid, pBI121 (Figure 2.2). pKom504, an agroinfectious construct consisting of 1.4 MSV-Kom genomes in tandem inserted in pBI121 (Figure 2.2), and pSet101, a 1.8mer of the MSV-Set genome constructed in pUC18, were obtained from Dr. Fiona Hughes. The MSV-Set 1.8mer was excised from pSet101 using the restriction endonuclease *Xba* I (Appendix A1) and subsequently partially digested with *Sac* I (Appendix A2). Partial digests were required since a *Sac* I site is situated in the MCS of the pUC18 vector and in the MSV-Set genome. *Sac* I and *Xba* I double-digested pBI121 and pSet101, partially digested with *Sac* I and completely digested with *Xba* I, were ligated as described in Appendix A4. Competent *E.coli* cells were transformed (Appendix A5) using this ligation mix. The aim of this "shotgun" ligation was to replace the β-glucoronidase (GUS) ORF situated between the *Xba* I and the *Sac* I site in the MCS with the 1.8mer MSV-Set construct. Colony blots (Appendix A8) of LA + kanamycin plates and hybridisation tests using DIG-labelled MSV-Set genomic DNA (Appendices A7

and A9) as a DNA probe, were performed to distinguish colonies harbouring pBI121 with sequences homologous to MSV-Set. Plasmid minipreparations of selected colonies were prepared (Appendix A6) and digested with *Pvu* II. Since there is one *Pvu* II site in the genome of MSV-Set and the size of this genome is approximately 2.7kb, a *Pvu* II-digested minipreparation of the desired construct in pBI121 which yielded a 2.7kb fragment was selected for maxipreparation (Appendix A11). This plasmid was designated pSet102 (Figure 2.2).

A PanSV-Kar genomic 1.7mer was constructed in a similar fashion as decribed for pSet102 above. An entire PanSV-Kar genome, excised using *Bam* HI from pPS100, was inserted into the *Kpn* I deletion clone of pPS100 (pPS101; 0.7mer) resulting in a 1.7mer PanSV-Kar tandem construct, pPS102. The GUS ORF and NOS-terminator situated between the *Hind* III site in the MCS and *Sac* I restriction sites of pBI121 were replaced by the PanSV-Kar 1.6mer tandem repeat of pPS102 (pPS103; Figure 2.2). Since there are *Hind* III and *Sac* I restriction sites in the MCS on either side of the PanSV-Kar 1.7mer of pPS102, *Hind* III- and *Sac* I-digested pPS102 was ligated (Appendix A4) with *Hind* III- and *Sac* I-digested pBI121. Transformations were performed as described in Appendix A5. The desired clone was detected by colony blotting of LA + kanamycin plates (Appendix A8) and hybridisation tests using the PanSV-Kar genome as the DIG-labeled DNA probe (Appendix A7). DNA minipreparations (Appendix A6) of the colonies harbouring DNA homologous to the PanSV-Kar genome were prepared and digested with *Bam* HI. Since there is one *Bam* HI restriction site in the approximately 2.7kb PanSV-Kar genome, a minipreparation resulting in a 2.7kb fragment on agarose gels was selected as representative of the desired clone, pPS103. A maxipreparation of pPS103 was prepared (Appendix A11).

2.3.3 Construction of genomic clones containing two LIRs

Using standard cloning techniques (Appendix A), 1.1mer clones consisting of two LIRs flanking the ORFs were constructed for both MSV-Kom and MSV-Set in pUC18, and inserted into pBI121 for agroinfection (Figure 2.2). The orientation of the MSV-Kom RF-DNA in pKom500 was switched by digesting the plasmid with *Bam* HI and then religating the digest (Appendices A1, A4 and A5). This was required to aid in the construction of the 1.1mer. A *Sal* I-digested plasmid minipreparation revealing opposite orientation of MSV-Kom RF-DNA to that of pKom500 was selected for maxipreparation (Appendix A7) and designated pKom600. pKom600 was digested with *Sac* I and religated (Appendices A1, A4, A5, A6 and A11) to excise the ORFs and the *Bam* HI restriction site between the left side of the LIR and the MCS in the pUC18 vector similar to the construction of the deletion clone pSet100-10 (3.2.1 and Figure 3.1). This plasmid was designated pKom601.

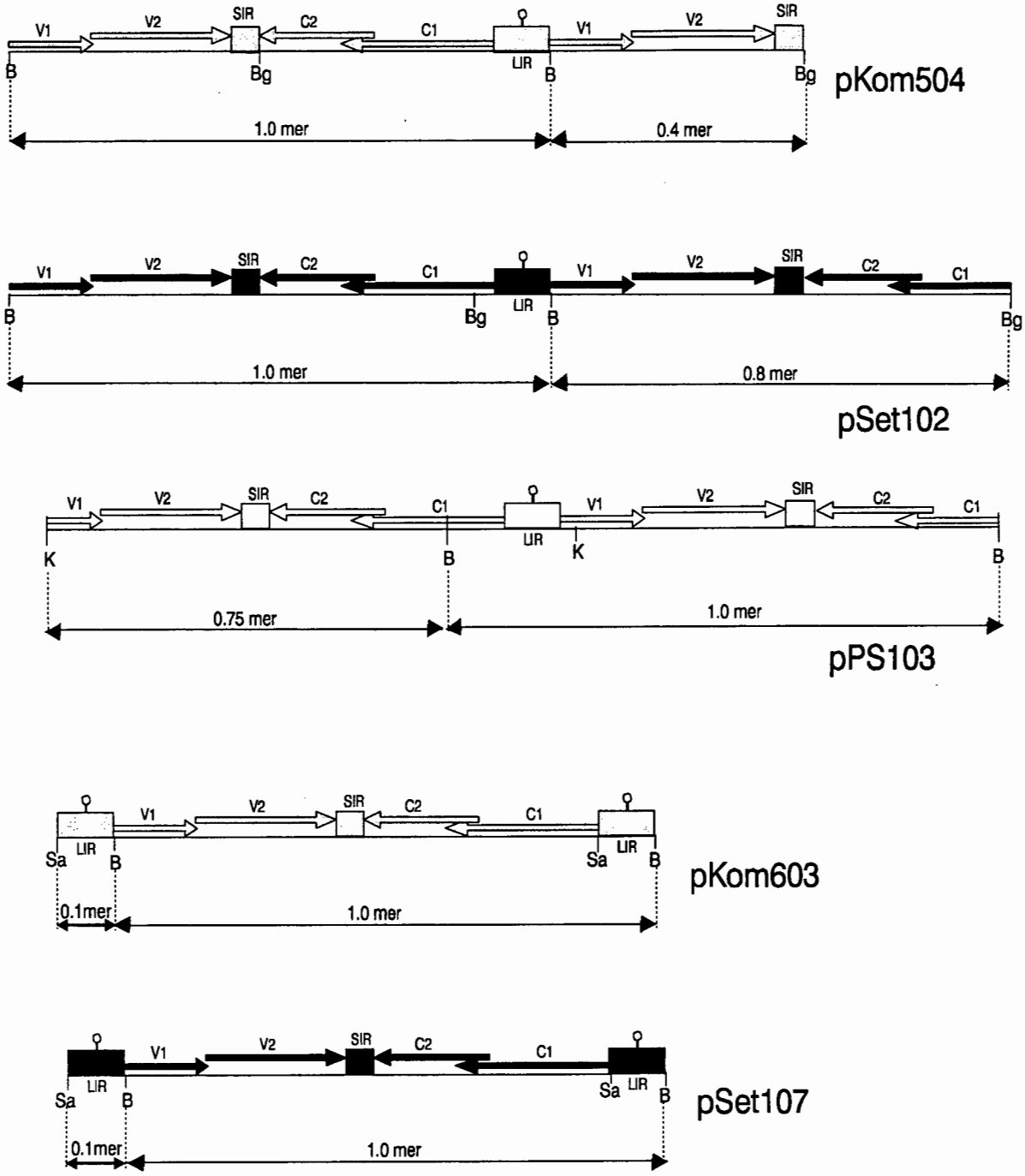

Figure 2.2: Partial homodimeric tandem dimers introduced into pBI121

pKom504, pSet102 and pPS103, tandem partial dimers of the genomes of MSV-Kom, MSV-Set and PanSV-Kar respectively. V1 = virion-sense ORF encoding MP; V2 = virion-sense ORF encoding CP; C1 and C2 = complementary-sense ORFs encoding Rep; SIR and LIR = small and long intergenic regions respectively; B = *Bam* HI; Bg = *Bgl* II, K = *Kpn* I; Sa = *Sac* I. Shaded, black and open constructs are MSV-Kom, MSV-Set and PanSV-Kar sequences respectively.

Bam HI/-digested pKom601 and pSet100-10 were ligated with *Bam* HI/*Pst* I-digested pKom500 and *Bam* HI/*Eco* RI-digested pSet100 respectively (Appendices A1, A4, A5, A6 and A11) to obtain the clones consisting of two LIRs flanking the ORFs, designated pKom602 and pSet105 respectively. The *Pst* I and *Eco* RI cuts pUC18 in the MCS thereby reducing the probability of this linear vector interfering in the "shotgun" ligations. An *Apa* I-digested plasmid minipreparation which showed a 2.7kb fragment on agarose gels (Appendices A1 and A3) was selected for each virus, since a unique *Apa* I restriction site is situated close to the hairpin origin of replication in the LIR for both viral RF-DNAs. pKom602 was digested with *Xba* I, *Pst* I and *Eco* RI and was ligated with *Xba* I/*Eco* RI/*Sac* I-digested pBI121 to replace the GUS reporter gene with the 1.1mer constructs (Appendices A1-A5) to give pKom603. Digesting pKom602 with *Pst* I and pBI121 with *Sac* I reduced the probability that the undesirable linear pUC18 and the GUS/NOS-terminator DNA would interfere in the ligation. Similarly, pSet105 was digested with *Xba* I and *Pst* I and partially digested with *Eco* RI, ligated with *Xba* I/*Eco* RI/*Sac* I-digested pBI121 to insert the MSV-Set 1.1mer in place of the GUS and NOS-terminator of pBI121. pSet105 had to be partially digested since the MSV-Set RF-DNA has an *Eco* RI site. Colony blots (Appendix A8) of LA + kanamycin plates and hybridisation tests (Appendix A8) using DIG-labelled MSV-Set genomic DNA (Appendices A7 and A9) as a probe, were performed to distinguish colonies harbouring pBI121 with sequences homologous to MSV-Kom and MSV-Set respectively. Plasmid minipreparations of selected colonies were prepared (Appendix A6) and digested with *Apa* I (Appendices A1 and A3). Since a unique *Apa* I restriction site is situated close to the hairpin in the LIR of both viral genomes and the sizes of the genomes are approximately 2.7kb, *Apa* I-digested minipreparations of the desired tandem 1.1mers in pBI121 yielding 2.7kb fragments on agarose gels were selected for maxipreparation (Appendices A1, A3, A6 and A11). Agroinfections were performed as described below.

2.3.4 Agroinoculation of maize seedlings

pKom504, pSet102, pKom603, pSet107 and pPS103 were introduced directly into *Agrobacterium tumefaciens* strain C58 with a pTiC58-derived plasmid, pMP90 (gentamycin resistance marker [Gm^R] and T-DNA deletion; Koncz and Schell, 1986), using the freeze-thaw transformation method of An *et al.* (1988) (Appendix A15). The *Agrobacterium* strain C58 was maintained in bacterial culture media containing 50-100µg/ml gentamycin and rifampicin. Transformed *Agrobacterium* clones were maintained in the same media containing 50µg/ml kanamycin (in addition to gentamycin and rifampicin) for selection of kanamycin resistance conferred to transformants by the NPT II gene of the pBI121 clones. The agroinfection methods described by Grimsley *et al.* (1987) were employed with some modifications. Maize seedlings (3-4 days old) were germinated in moist vermiculite at 30°C,

rinsed in distilled water and decapitated approximately 3-5mm above the coleoptilar node, using a sterile scalpel blade. Bacterial cultures were grown with aeration, under kanamycin selection, in 5mls Luria Broth (LB) at 30°C for approximately 24-36hrs. 1.5mls of cultures were spun down and resuspended in 100µl LB. Two to 3µl of the concentrated inoculum was injected down through the cut surface of the decapitated seedlings into the meristematic region. The inoculated seedlings were planted in punnets containing heat-steriled soil and grown in a contained plant growth room at 27°C/22°C day/night temperature and relative humidity of approximately 70%. Daylight was supplemented by a 14 hour photoperiod with fluorescent lights. Streak symptoms were monitored and agroinfectivity rates were recorded over a period of approximately 30 days.

Steps were taken to ensure that irrigation water used in the contained plant growth rooms was not allowed to escape by placing all the punnets in catch trays. All plant tissue, used soil and bacterial growth media were sterilised by autoclaving before they were disposed of. These regulations in the laboratories and growth rooms were established to ensure the clones were not released into the environment.

2.3.5 Leafhopper maintenance and transmissions

A virus-free leafhopper (*Cicadulina mbila*) colony was maintained as described by von Wechmar and Hughes (1992) in the Department of Microbiology, University of Cape Town. The insects were housed in gauze cages where they fed and bred on uninfected potted maize plantlets (5-8 leaf stage), wheat (2-3 week old) or barley (2-3 week old) seedlings. Cages were accommodated in an air conditioned environment placed close to windows to maintain good plant growth. Alternatively they were placed under fluorescent lights fitted to a timer allowing a 14/10 hour day/night cycle at a temperature of 27-30°C. Mixed populations of adults and nymphs consisting of all instars were used in the leafhopper transmission tests to determine the host ranges of the viruses. Non-viruliferous leafhoppers were allowed to feed on agroinoculated maize plants showing streak symptoms in cages situated in separate rooms allowing an acquisition feeding time of 48-96 hours. Two week-old (from germination) uninfected maize, wheat and barley potted plants being tested for susceptibility were placed into the cages containing viruliferous leafhoppers. Each pot contained 4-5 maize plantlets or approximately 10-20 barley or wheat plantlets; 9 pots per cage. Cages were kept under the above mentioned conditions during the inoculation feeding period where viruliferous leafhoppers were allowed to feed on the plants for 7-10 days for the first infectious cycle of transmissions and 5-7 days for the following infectious cycles. The longer first infectious cycle was necessary to allow for the latent phase following virus acquisition: this is the time required for the virus to move from the insect gut through to the

salivary glands before the virus is transmitted to the host plant (Storey, 1928). Movement of leafhoppers in the cages was directed by shifting a 60 Watt incandescent light source shining into the cage from the exterior to ensure that all the plants tested were fed from (leafhoppers follow the light source and warmth). Plants removed from the cages were sprayed with Efekto® G49 Agrihold™ insecticide and placed in plant growth rooms with the above mentioned conditions. All tests were quadrupled and repeated to limit chance of error. Streak symptoms were recorded as severe, moderate, mild or no streak (+++, ++, + or - respectively).

2.3.6 Plants tested for susceptibility

The plant varieties tested for susceptibility included *Zea mays*, *Triticum aestivum*, *Hordeum vulgare*, *Panicum maximum* and *Digitaria sanguinalis*. The cultivars tested are listed below where appropriate:

Z. mays

- Sweetcorn cultivars, Jubilee (commercial variety) and US More (parent breeding variety for developing hybrids), were chosen for their high susceptibility to cereal Mastreviruses as positive controls and as indicators for determining whether leafhoppers were viruliferous throughout the period of the leafhopper tests (von Wechmar and Hughes, 1992).
- Popcorn (commercial supermarket variety) was chosen for its susceptibility and its differentiation of symptom severity (Figure 2.5a and b).
- "Witplat" (cv. Kalahari early pearl); a susceptible cultivar.
- "Vaalhartz Wit and Geel"; resistant or tolerant variety from Potchestroom, North-West Province, South Africa, obtained from the Grain Crops Research Institute (GCRI) Potchefstroom, North-West Province, South Africa.
- Hybrid varieties donated by Pioneer Seed Co. (Pty) Ltd. and Asgrow (Pty) Ltd.; A7563 and PNR 6549 are susceptible hybrids, and A7541 and PNR 6552 are tolerant hybrids (von Wechmar and Hughes, 1992).
- Susceptible and tolerant hybrids obtained from Pannar Seed Co.: PAN 6191, 6166, 6133, 6363, 6480, 6043, 6564, 6474, 6481, 6496, 6099, 6549, 6578, 6552, 6195, 6528, 6364, 6462, 6479 and 6141.

T. aestivum

- Water irrigation cultivars obtained from Elsenberg Agricultural Development Institute, Western Cape, South Africa: SST66, SST44, SST33, SST55 and SST16.
- Varieties obtained from the Agronomy Section of the Elsenberg Agricultural Development Institute, Western Cape: cv.s Schooner, Palmiet, Dias, Sterling, Gamtoos, Vloekskoot, Adam Tas, Chokka and Nantes.

- Rust differentials obtained from the Elsenberg Agricultural Development Institute: cv.s Festiquay, Marquis and Agent.

The only *H. vulgare* tested was the susceptible cv. Clipper, obtained from the Elsenberg Agricultural Development Institute in the Western Cape, South Africa.

2.3.7 Detection of viral DNA in infected plants

One µg of *Pvu* II-cut (Appendix A1) and uncut total plant DNA extractions (isolated as described in Appendix A12) from agroinoculated samples was mixed with loading buffer and run on a 0.8% agarose Tris-borate-EDTA (TBE) gel in TBE containing 20ng/ml ethidium bromide (Appendix A3). Gels were capillary blotted overnight onto Hybond-N+® membrane unidirectionally without pretreatment, using 0.4M NaOH as the transfer buffer (Appendix A13). DNA probes were DIG®-labelled using the Boehringer-Mannheim non-radioactive DIG®-labelling and detection kit as described in Appendix A7. MSV-Kom, MSV-Set and PanSV-Kar genomes were excised with *Bam* HI (Appendix A1) from pKom500, pSet100 and pPS100 respectively and isolated from Tris-acetate-EDTA (TAE) 0.8% agarose gel and using the Biolab Geneclean® kit as described in Appendices A3 and A7. These were labelled using the DIG® DNA-labelling and detection kit. Hybridisations were done as described by Boehringer-Mannheim's detection manual at 68-70°C for 16h and blots were washed twice for 5min at room temperature and twice for 15min at 68-70°C (Appendix A9). Immunological Lumigen-PPD® detection was performed as described in the manufacturer's DIG DNA-labeling and detection kit manual (Appendix A10). Blots were exposed to Agfa Curix RP1 X-ray film for 10-60min and developed in an OKAMOTO™ Medical X-ray film automatic processor X2 as decribed in Appendix A10.

2.4 RESULTS

2.4.1 Agroinfectivities of homodimeric clones

Agroinoculations were performed as described and symptoms were monitored over a period of three to five weeks. Results of these agroinoculations are presented in Table 2.2 and Figure 2.3. Three-day-old sweetcorn seedlings were agroinoculated for each of partial dimers, pKom504, pSet102 and pPS103 cloned into *A. tumefaciens* (C58C1). Of the plantlets that survived the injection, the final percentages of infected sweetcorn seedlings were 52, 33 and 26% for pKom504, pSet102 and pPS103 respectively. PanSV-Kar symptoms appeared between three and four days later than those of MSV-Kom and MSV-Set. Six initial pilot agroinoculation experiments using pSet102 and pKom504 clones were performed on popcorn and Jubilee sweetcorn seedlings. These agroinfectivity rates are

summarised in Table 2.2 along with the above results and are in line with the above results obtained for the same clones in scaled up experiments.

Table 2.2 Agroinfectivity results of homodimeric clones pKom504, pSet102 and pPS103

Clone	Experiment I	Experiment II	Experiment III
pKom504	26/44 59%	45/91 49%	23/44 52%
pSet102	17/82 21%	35/127 28%	15/45 33%
pPS103	na	na	11/43 26%

Experiments I and II = popcorn (I) and sweetcorn(II) pilot agroinoculation results displaying the total agroinfectivity rates obtained in six experiments prior to experiment III. Values indicate the number of surviving seedlings showing symptoms by the number of surviving seedlings. The percentage infectivity rates are included. The agroinfectvity profiles of the experiment III results are depicted in Figure 2.3. na = not done.

In contrast, agroinoculations performed with *Agrobacterium* clones of pKom603 and pSet107 - both 1.1mer homodimeric clones containing two LIRs - revealed agroinfectivity rates significantly higher, at 88% and 82% as opposed to 45% and 32% for pKom504 and pSet102 *Agrobacterium* clones respectively, in two experiments (Table 2.3). In addition, streak symptoms of the agroinoculations of 1.1mers with 2 LIRs appeared between 4-5 days after agroinoculation whereas the appearance of streak symptoms appeared 8-9 days after agroinoculation with pKom504 and pSet102. This pattern was observed in a repeat experiment using larger numbers of seedlings (Table 2.3, Experiment II and Figure DI, Appendix D) and in the agroinfection tests when pKom603 and pSet107 were used as positive controls in heterodimer tests (Chapter 4). In all agroinoculation tests, pBI121 transformed into *Agrobacterium* produced no symptoms.

Another observation of interest is that agroinoculations of decapitated seedlings with partially homodimeric clones with one LIR (pKom504 and pSet102) resulted in a significantly lower mortality rate over the 3-5 week monitoring period compared those agroinoculations of partial dimers with two LIRs (pKom603 and pSet107). The mortalities of pKom603- and pSet107-inoculated plants were 20 and 23% as opposed to 3 and 4% for pKom504 and pSet102 repectively (Table 2.4).

Table 2.3 Agroinfectivities of 1.1mers and partial homodimeric clones

Clone	LIRs	Experiment I	Experiment II	Total Agroinfectivity
pKom603	2 (1.1mer)	40/40 100%	60/73 82%	100/113 88%
pSet107	2 (1.1mer)	36/40 90%	53/67 79%	89/107 83%
pKom504	1 (1.4mer)	11/29 38%	37/77 48%	48/106 45%
pSet102	1 (1.8mer)	7/25 28%	24/72 33%	31/97 32%

Both experiments were performed using Jubilee sweetcorn. Values indicated are the number of infected plants out of the number of agroinoculated plants that survived the injection. Percentage infections are also indicated. Total agroinfectivity (column indicated) pooled the results of both experiments. Agroinfectivity profiles of Experiment I and Experiment II are depicted in Figures 2.3 and Appendix D, Figure D1 respectively.

In a repeat experiment the death rates were 14 and 11% for pKom603 and pSet107 agroinoculations as opposed to 7 and 9% for pKom504 and pSet102 respectively. These mortality rates are presented in Table 2.4 where the totals are at 16% for the 1.1mers with two LIRs and 6 and 8% for the significantly larger partially homodimeric clones with only one LIR. These results were corroborated in agroinoculations of heterodimic clones (Chapter 4). *Agrobacterium* containing pBI121 as a negative control, produced a significantly lower mortality rate of 4% in one of the experiments.

Table 2.4 Mortality rates of agroinoculated seedlings

Clone	LIRs	Experiment I	Experiment II	Total mortality
pKom603	2 (1.1mer)	10/50 20%	12/85 14%	22/135 16%
pSet107	2 (1.1mer)	12/52 23%	8/75 11%	20/127 16%
pKom504	1 (1.4mer)	1/30 3%	6/83 7%	7/113 6%
pSet102	1 (1.8mer)	1/26 4%	7/79 9%	8/105 8%
pBI121	-	na	1/23 4%	1/23 4%

Results indicate number of dead seedlings by the number of seedlings agroinoculated and percentage mortalities. na = not applicable. LIRs indicates the number of LIRs contained in the homodimeric clones.

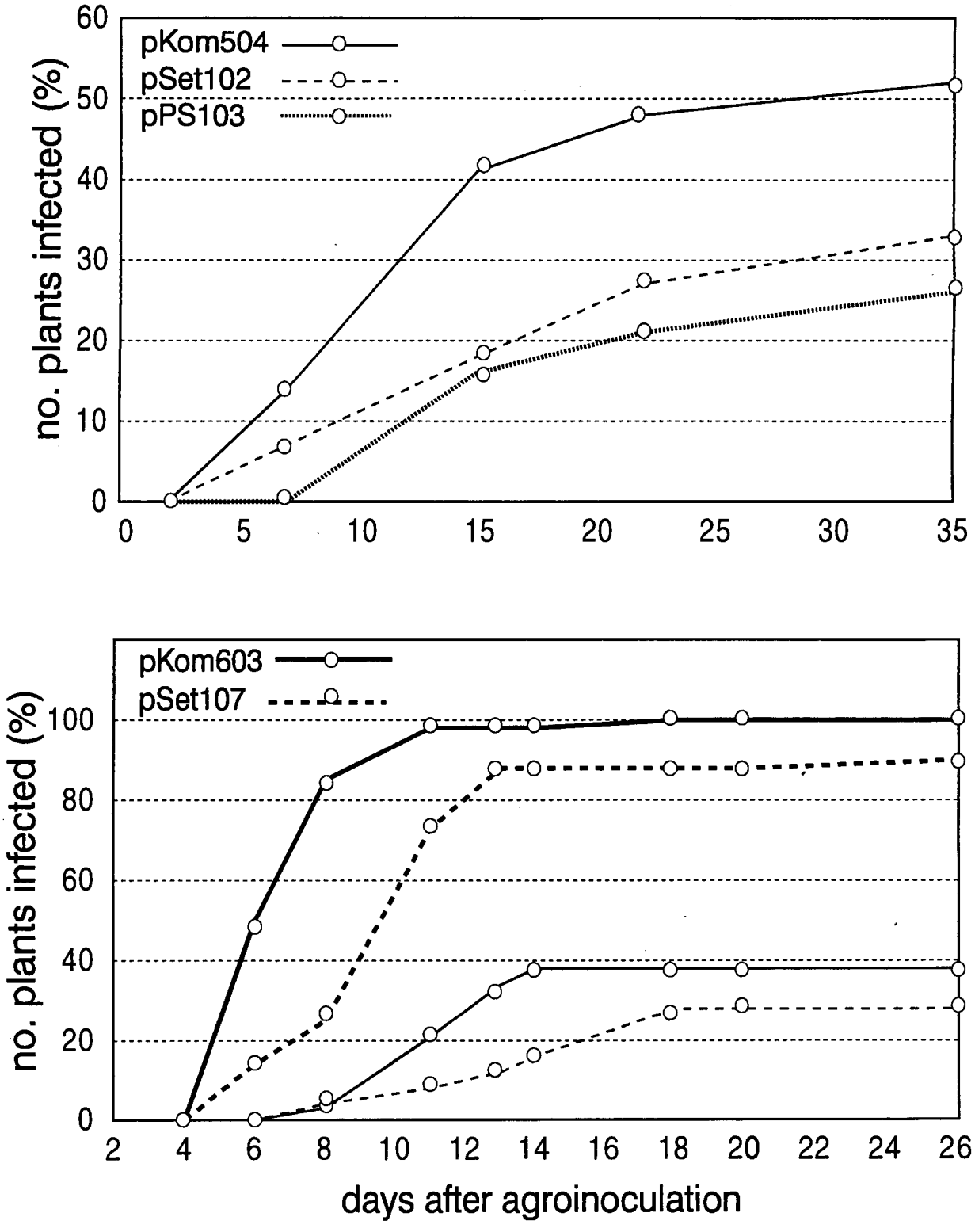

Figure 2.3 Rates of agroinfection of homodimeric dimers

Top diagram depicts the relative rates of agroinfection of the partial dimers with one LIR, pKom504, pSet102 and pPS103. Bottom diagram depicts the relative rates of agroinfection of the partial dimers (pKom504 and pSet102) and those of the 1.1mers with two LIRs (pKom603 and pSet107). See Appendix D, Figure D1 for repeat agroinfectivity rates.

2.4.2 MSV-Kom, MSV-Set and PanSV-Kar symptoms and host ranges

2.4.2.1 *Agroinfections and symptoms*

MSV-Kom dimers caused severe streak in Jubilee sweetcorn: the symptoms were continuous chlorotic yellow streaks running most of the length of the leaves (Figure 2.4). MSV-Set dimers caused more dispersed whiter chlorotic streaks which were relatively moderate in sweetcorn. In contrast, PanSV-Kar dimers were found to cause only very mild fine stippled streaks running some of the way up the leaves. PanSV does not cause a persistent infection in sweetcorn since the younger larger leaves (3rd, 4th and 5th leaves) display gradually diminishing streak symptoms to a point where the plant overcomes the infection (Figure 2.4). The severity of the streak correlated with the stunting effect on the growth of the plants which was pronounced for MSV-Kom. Stunting was not as obvious for MSV-Set, and with PanSV-Kar no stunting was observed. The symptoms produced in agroinoculated plants were generally more severe than the symptoms observed in leafhopper transmission tests.

2.4.2.2 *Leafhopper transmissions and host ranges*

The symptoms for MSV-Kom, MSV-Set and PanSV-Kar were all successfully transmitted by virus-free leafhoppers. Jubilee sweetcorn was used as an indicator for successful transmission in all cages used since it is considered our "universal host". A variety of maize plants, wheat, barley and grasses were subjected to leafhopper transmissions of these three viruses to establish their host ranges. A summary of the leafhopper transmissions was tabulated and the symptom severities scored as severe, mild or moderate (Table 2.5). These results on the maize cultivars are consistent with the findings of comparative computer image analyses being done of the streak symptoms of these viruses in this laboratory (Darren Martin, pers. comm.) The host ranges distinguish the three viruses in this study. The *Z. mays* cultivars, including the resistant varieties, were susceptible to MSV-Kom: moderately for the more resistant varieties and severely for the susceptible varieties. The resistant varieties of maize were found to be resistant to MSV-Set however the more susceptible varieties such as sweetcorn (Jubilee), popcorn and "Witplat" displayed streak symptoms typical of Mastreviruses in graminaceous plants. MSV-Kom and MSV-Set streak symptoms on popcorn leaves were very clearly (even more so than on Jubilee sweetcorn) distinctly severe with continuous chlorotic streak (MSV-Kom) or very mild with dispersed stippled streak (MSV-Set) as depicted in Figures 2.5a and b. MSV-Kom and MSV-Set infections in Clipper barley were distinguishable as severe chlorotic clearing of leaves or severe stippled streak repectively as depicted in Figures 2.5c and d. Both MSV-Kom and MSV-Set appeared to cause curling in infected leaves of Clipper barley as depicted in Figure 2.5; MSV-Kom caused more severe stunting and curling than MSV-Set.

The wheat varieties, were found to be generally susceptible to MSV-Set, whereas certain cultivars such as the rust differentials (Marquis, Festiquay, Agent and Dias) tended to be more resistant to the MSV-Kom. A good example of this result are those symptoms obtained from the resistant cultivar SST44 where MSV-Kom does not produce streak whereas MSV-Set produced mild stippled streak (Figure 2.6). In contrast however, MSV-Kom streak in susceptible wheat cultivar SST66 is able to produce streak symptoms as severe as those of MSV-Set, but MSV-Kom can also cause severe leaf distortion as depicted in Figure 2.6.

PanSV-Kar is extremely mild in sweetcorn (Jubilee) and was not transmitted to even the MSV-susceptible maize, wheat or barley varieties. Leafhoppers were only able to transmit very mild streak in Jubilee sweetcorn. PanSV-Kar infection was not persistent as for MSV-Kom and MSV-Set: the streak symptoms diminished as the plantlets grew older (Figure 2.4). No streak symptoms were observed in any of the other maize, wheat or barley cv.s. The results for MSV-Kom and MSV-Set were corroborated in similar tests when they were used as positive wild-type controls in more extensive leafhopper transmission tests in recombinant viral tests (Chapters 4 and 5, Tables 4.3 and 5.1 and 5.2).

2.4.3 Southern blot detection of viral DNA in infected plants

1μg of each total plant DNA extract was loaded on a TBE 0.8% agarose gel and run. *Pvu* II-digested total plant DNA extract was also loaded. Southern blots were performed as described above and detections were done using DIG-labelled MSV-Kom, MSV-Set and PanSV-Kar probes. The probes hybridised with their respective agroinoculated plant DNA extracts although there was slight cross-hybridisation of MSV-Set to DNA of pKom504-agroinoculated plants. Southern blots are presented in Figure 2.7. Linear and circular ss viral DNA was detected as unaltered DNA bands when DNA was cut with *Pvu* II. The positive identification of MSV-Set, MSV-Kom and PanSV-Kar in these blots by *Pvu* II-restriction and their respective hybridisations confirmed their successful transmission. MSV-Set has a single *Pvu* II site in nucleotide position 1462, whereas MSV-Kom or PanSV-Kar have two *Pvu* II sites each in nucleotide positions (nt pos.) 541 and 1464 or 98 and 2617 respectively. The analyses of the sequences of MSV-Set, MSV-Kom and PanSV-Kar are reported in Chapter 3. Thus digested total plant DNA extractions revealed a single 2701 bp fragment for pSet102 or pSet107 agroinoculations, 923 bp and 1767 bp fragments for pKom504 or pKom603 agroinoculations; and 186 bp and 2519 bp fragments for pPS103 agroinoculations in Southern blot tests (Figure 2.7). Southern blot tests on leafhopper-transmitted plants with symptoms showed similar results (results not shown). Although no

restriction digests were performed on these, the probes used specifically hybridised to their respective leafhopper-mediated infected plants.

Southern blots of pBI121-agroinoculated plant DNA did not hybridise with the MSV-Kom, MSV-Set, PanSV-Kar or pBI121 DIG-labelled probes. Likewise, DNA from homodimer-agroinoculated plants did not hybridise with the DIG-labelled pBI121 probe indicating that pBI121 vector does not interfere in the agroinoculation tests. All plants with no streak symptoms, whether from pBI121 negative control tests or homodimeric agroinoculations, showed no positive hybridisations in Southern blot tests.

2.4.4 Leafdip preparations of agroinoculated plant leaves

These were prepared by Mohammed Jaffer (Electron Microscope Unit, University of Cape Town). Leafdip preparations of MSV-Kom-, MSV-Set- and PanSV-Kar- agroinoculated Jubilee sweetcorn with streak symptoms revealed geminate particles approximately 18x30nm in size. The number of particles present as observed in electron micrographs were several, few or difficult to find for MSV-Kom, MSV-Set and PanSV-Kar respectively (Figure 2.8).

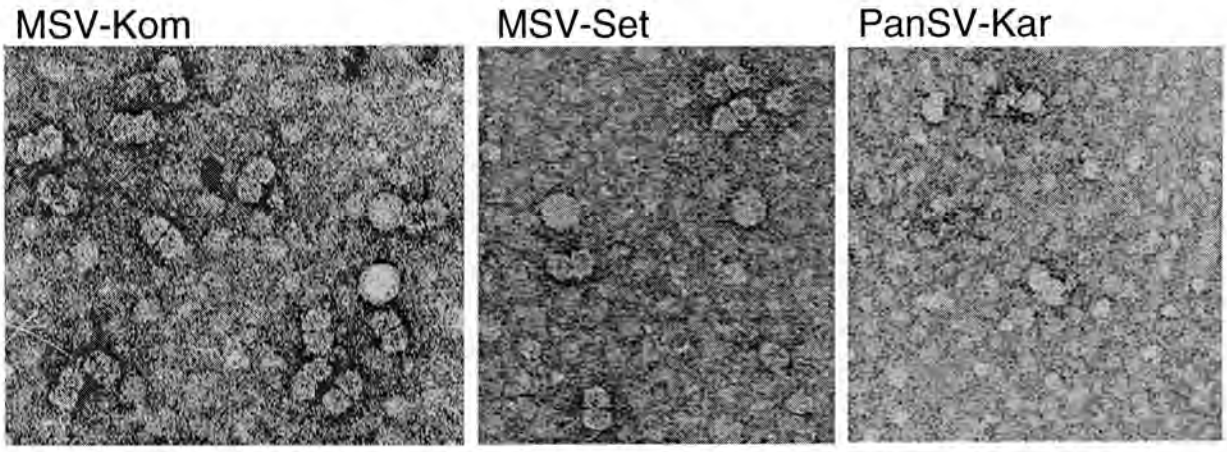

Figure 2.8 Electron micrographs of leaf-dip preparations

Leaf-dip preparations stained with uranyl acetate of - from the left - pKom504-, pSet102- and pPS103-agroinoculated Jubilee sweetcorn with streak symptoms. Geminate particles are 18x30nm.

Table 2.5: Host ranges and symptom severity as determined by leafhopper transmissions

Cultivar	PSV-Kar	MSV-Set	MSV-Kom
Z. mays			
Jubilee[a]	+	++[+]	+++[+]
US More[a]	-	+++	+++[+]
Popcorn[b]	-/+	++	+++[+]
Witplat	-/+	++[+]w	+++[+]
Vaalhartz Wit	-	-	+++
Vaalhartz Geel	-	-	+++
PNR 6549	-	-/+	+++
PNR 6552	-	-	+++
PNR 7541	-	-	+++
PNR 7563	-	-/+	+++
PAN hybrids[c]	na	-	+++
Clipper[d]	-	++	+++
T. aestivum			
SST 44	-	+	-
SST 66	-	++[+]	+++
Adam Tas	na	+++	+++
Chokka	na	+++	+
Festiquay	na	++	-
Marquis	na	-/+	-
Agent	na	-/+	-
Dias	na	+++	+
Digitaria[e] *sanguinalis*	na	+++	+++
Panicum[e] *maximum*	na	-	-

Top, *Z.mays* cultivars and bottom, *T.aestivum* cultivars and grass species; [a] Sweetcorn variety; [b] purchased at local supermarket in Cape Town; [c] see list of hybrids in section 2.2.6; [d] *H.vulgare* variety; [e] grass species; w = Wallaby ear; PAN, seed obtained from Pannah; PNR seed obtained from Pioneer Seed; na = leafhopper transmissions not performed. - = no streak symptoms; + = mild stippled streak and recovery(-/+); ++ = moderate stippled to continuous streak(++[+]); +++ = severe continuous streak and stunting and/or distortion of growth and death(+++[+]). See 5.3.2 for more host ranges of MSV-Set and MSV-Kom.

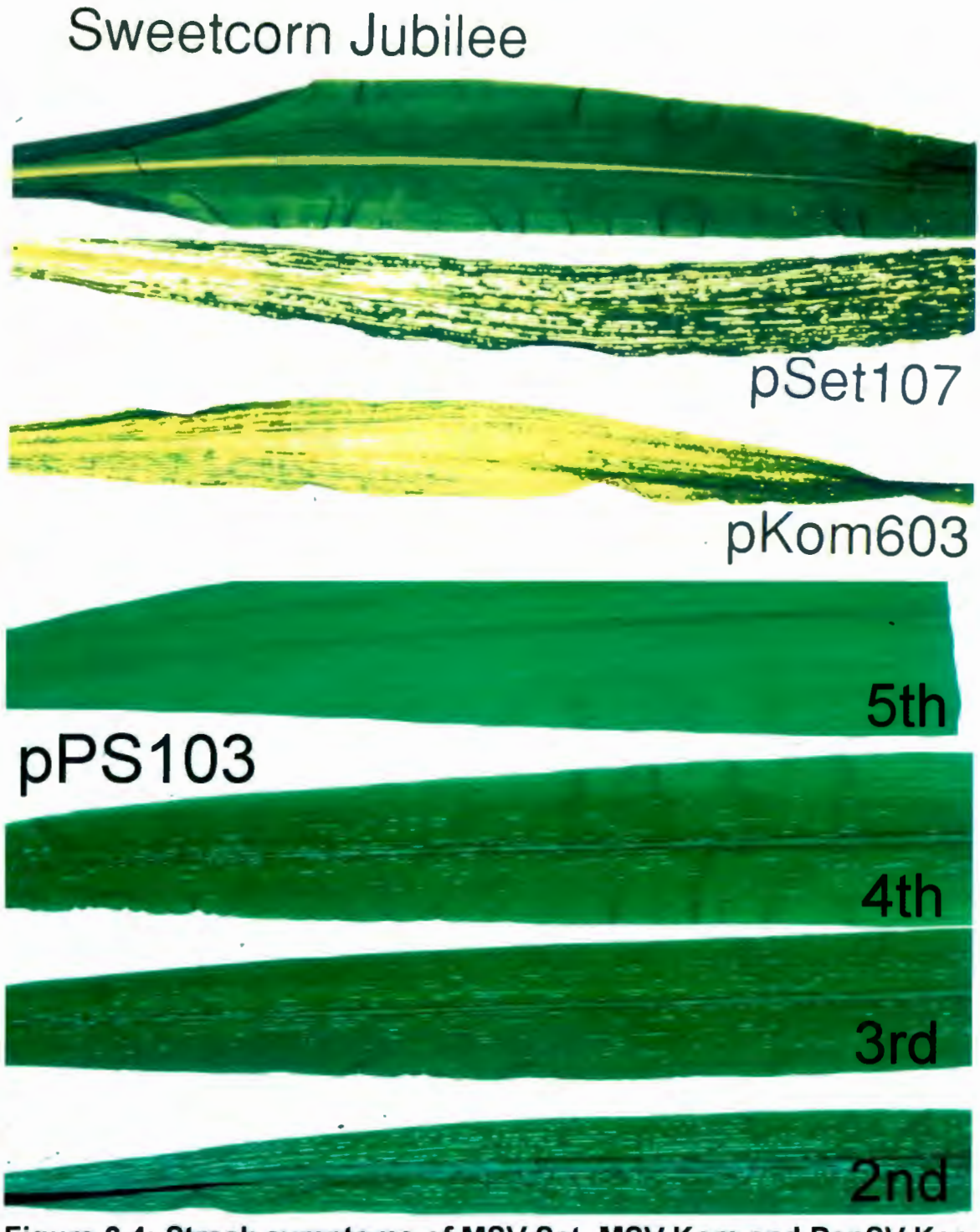

Figure 2.4: Streak symptoms of MSV-Set, MSV-Kom and PanSV-Kar on leaves of sweetcorn Jubilee

Three day old sweetcorn seedlings were agroinoculated with partially dimeric constructs of the genomes of MSV-Set, MSV-Kom and PanSV-Kar respectively (above: pSet107 or pKom603; below pPS103 as indicated). Above; the 2nd infected leaves are shown along with an healthy leaf and below; the 2nd, 3rd, 4th and 5th leaves from the same individual plant are shown.

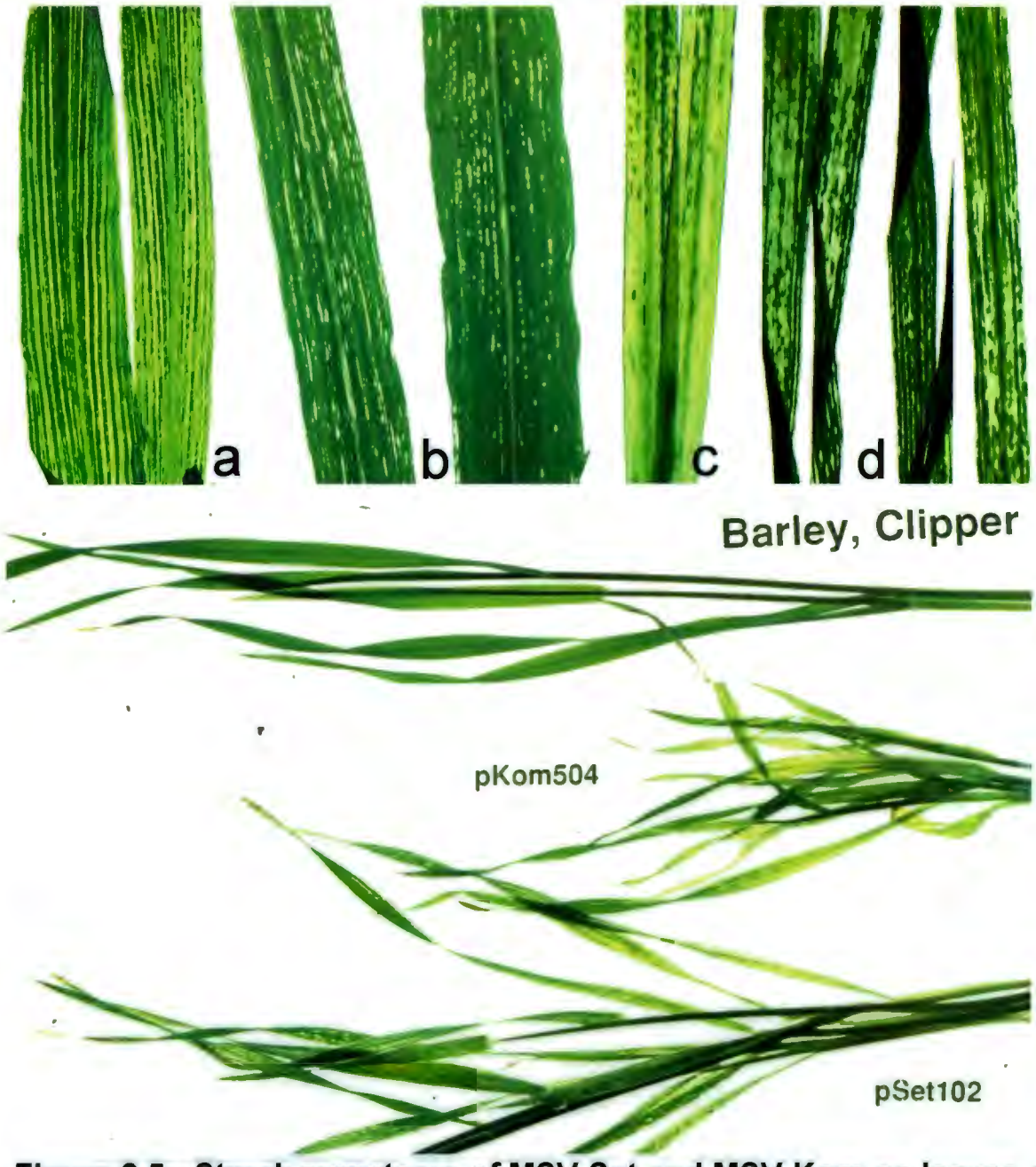

Figure 2.5: Streak symptoms of MSV-Set and MSV-Kom on leaves of barley and popcorn

Above: a, popcorn infected with MSV-Kom; b, popcorn infected with MSV-Set; c, barley leaves infected with MSV-Kom and d, barley leaves infected with MSV-Set. MSV-Kom causes continous streak or chlorotic clearing whereas MSV-Set causes milder streak in popcorn and barley (Clipper). Below: leafhopper transmitted 3-week old barley plantlets that were infected by viruliferous leafhoppers (previously allowed to feed from pKom504 or pSet102 agroinoculated maize plants).

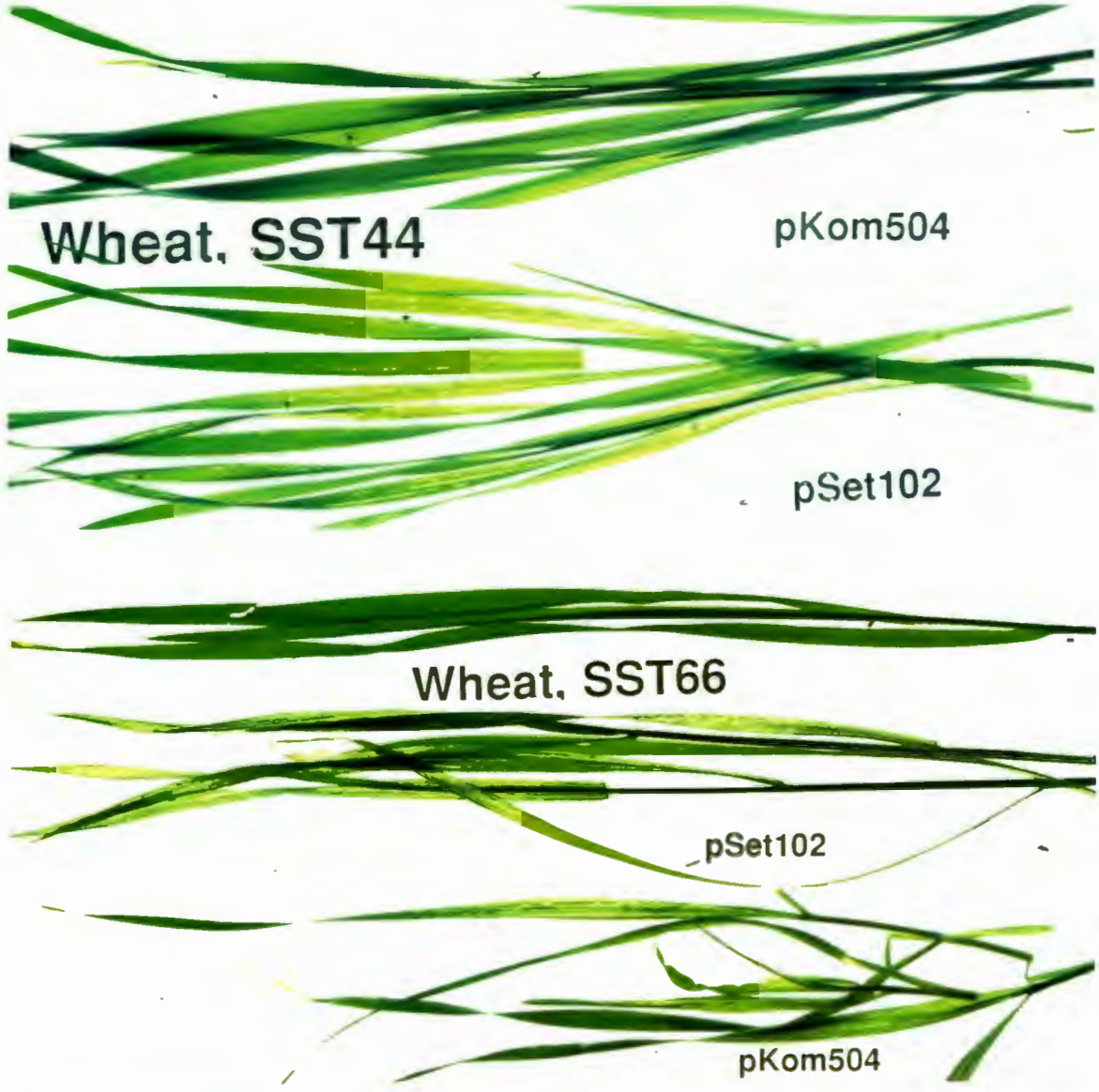

Figure 2.6: MSV-Kom and MSV-Set streak symptoms on wheat cv.s

Above: MSV-Kom is unable to produce streak symptoms on "resistant" cultivar SST44, however MSV-Set produces a mild streak on the leaves. Below: MSV-Kom and MSV-Set both produce severe streak on susceptible cultivar SST66, however MSV-Kom causes leaf curl. Three-week old seedlings were exposed to viruliferous leafhoppers which had fed from pKom504 or pSet102 agroinoculated sweetcorn plants.

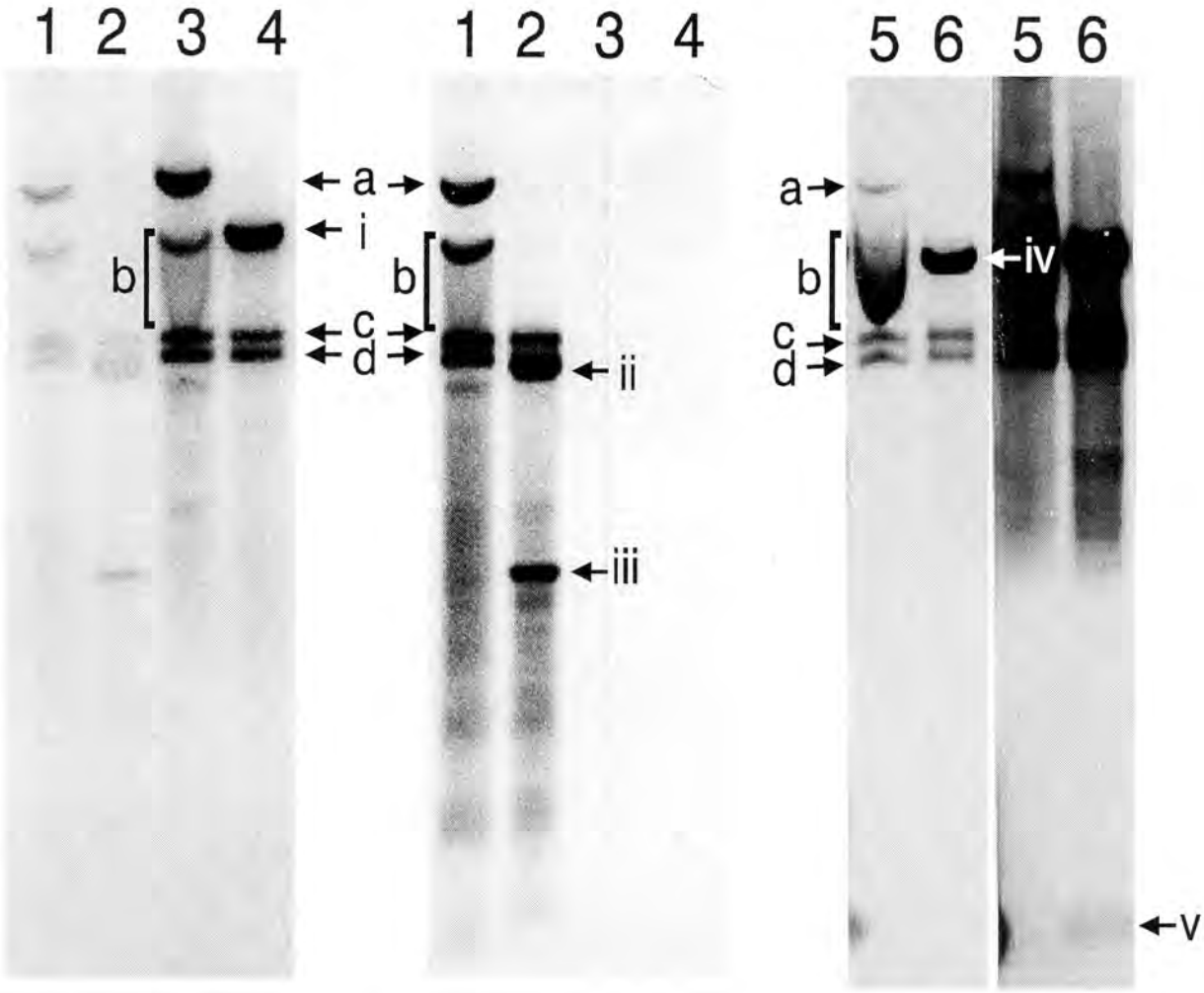

Figure 2.7 Southern blots for fhe detection of viruses in agroinfected plants

1μg total DNA isolated from agroinfected sweetcorn, Jubilee cv. was loaded in wells 1-6 on 0.8% agarose gels and run for 3hrs at 10V/cm (Appendices B3,B7, B9 and B10). DNA was capillary blotted onto Hybond-N+ nylon membranes. Left; blot hybridised with MSV-Set DNA probe; centre, same blot stripped and rehybridised with MSV-Kom DNA probe. Lanes 1 and 2 contained undigested and *Pvu* II-digested DNA respectively from an pKom504 agroinoculated plant displaying severe streak symptoms; lanes 3 and 4 contained undigested and *Pvu* II-digested DNA respectively from pSet102 agroinoculated plants displaying moderate streak. Right blot was hybridised with a PanSV-Kar DNA probe; lanes 5 and 6 contained undigested and *Pvu* II-digested DNA isolated from a pPS103 agroinfected plant respectively. Chemiluminescent detection using LumigenPPD was performed and X-ray film was exposed to the membrane for 30 min for MSV-Set detection (left exposure), and 15 min for MSV-Kom detection (centre X-ray exposure). For the PanSV-Kar detection X-ray film was exposed for 30 min (left exposure) and for 45 min to detect a 186bp fragment resulting from the *Pvu* II digestion (right). a = ds open circular DNA; b = closed circular and other conformations of viral DNA; c = linear ssDNA; d = circular ssDNA; i = MSV-Set RF-DNA contains one *Pvu* II site thus resulting in one ds 2701 bp fragment; ii + iii = MSV-Kom contains two *Pvu* II sites resulting in 1767 bp and 923 bp linear dsDNA fragments respectively; iv + v = PanSV-Kar constitutes two *Pvu* II sites resulting in 2519 and 186 bp linear dsDNA fragments respectively. Blot with lanes 1, 2, 3, and 4 was stripped and hybridised with DIG-labelled pBI121 resulting in a clear blot with no positive hybridisation.

2.5 DISCUSSION

The successful *Agrobacterium tumefaciens*-mediated delivery ("agroinfection") of cloned MSV-Kom, MSV-Set and PanSV-Kar DNA to plants allowed the determination of infectivity and pathogenicity of the viral RF-DNAs contained in plasmids pKom500, pSet100 and pPS100 respectively. This was displayed by; (i) the appearance of typical Mastrevirus-like streak symptoms on "agroinoculated" plants for each of the three viruses; (ii) the appearance of geminate particles typical of geminiviruses (approx.18x30nm) in electron micrographs of leaf-dip preparations of agroinoculated leaves (Figure 2.8; M. Jaffer, pers. comm.); (iii) the ability of leafhoppers to transmit the streak disease from agroinoculated plants to healthy uninfected susceptible plants, thereby fulfilling Koch's postulates; and (iv) the Southern blot DNA detection of ss and ds DNA forms in total DNA extracts of agroinoculated plantlets using the respective DIG-labelled RF-DNA cloned in the respective plasmids pKom500, pSet100 and pPS100. Thus the DNAs cloned in plasmids pKom500, pSet100 and pPS100 are sufficient for infection and most likely to represent the entire genomes of MSV-Kom, MSV-Set and PanSV-Kar respectively.

2.5.1 Agroinfectivities of homodimeric clones

Pilot agroinoculations using pKom504 and pSet102 were performed on a susceptible maize, and a wheat and a barley cultivar. Streak symptoms were obtained In maize but not in the wheat and barley cultivars. In agroinfection tests of WDV on wheat 30% infection rate was obtained by Woolsten *et al.*, (1988) possibly reflecting the efficiency with which *Agrobacterium* interacts with wheat cells. Therefore, due to its ability to germinate and grow to a convenient size (for easy injection using a 10µl Hamilton syringe) relatively quickly in the conditions provided, it was decided to perform agroinoculations on Jubilee sweetcorn as a standard for determining relative agroinfection rates or agroinfectivities in this study. Jubilee sweetcorn is also susceptible to all Mastreviruses previously tested in leafhopper transmissions (von Wechmar, pers. comm.; von Wechmar and Hughes, 1992) and thus is considered as the "universal Mastrevirus host" in this laboratory. It displays distinctive streak symptoms (Figure 2.4) when infected and is palatable to leafhoppers. Thus Jubilee sweetcorn is an all-round useful cultivar in a study such as this and was also used as an indicator in leafhopper transmissions tests to ensure that leafhoppers were viruliferous throughout the duration of the leafhopper transmission experiments.

The agroinfectivities of the homodimeric dimers pKom504, pSet102 and pPS103 reflect the severity of the streak symptoms on MSV-Kom-, MSV-Set- and PanSV-Kar-infected Jubilee sweetcorn. That is, the agroinfectivity rates for pKom504, pSet102 and pPS103 of 52, 33

and 28% are in line with the severe, moderate and mild streak symptoms of MSV-Kom, MSV-Set and PanSV-Kar respectively (Table 2.2 and Figures 2.3 and 2.4). This was the case whether Jubilee sweetcorn was agroinoculated or exposed to viruliferous leafhoppers (Table 2.5). In addition, streak symptoms appearing on pPS103-agroinoculated seedlings 3-5 days later than those appearing for the more severe pSet102 and pKom504 clones is an indication of the relative mildness of PanSV-Kar infection in sweetcorn. The agroinfectivities of pSet107 and pKom603 (homodimeric 1.1mers of MSV-Set and MSV-Kom respectively; Figure 2.2) of a total of 83% and 88% in two experiments (Table 2.3; Figure 2.3) appear to reflect their moderate and severe symptom severities respectively. Although these agroinfectivity rates are not significantly different in this case, those of pKom603 were consistently higher than those of pSet107 when they were used as wild-type controls in heterodimer agroinoculation tests reported in Chapter 4 and 5.

Although two mechanisms of release of unit-length genomes from dimers were proposed (Figure 1.4; Stenger *et al.*, 1991), the release of unit-length genomes of MSV-Kom, MSV-Set or PanV-Kar genomes from pKom504, pSet102 or pPS103 respectively is restricted to the homologous recombination mechanism. This is because the partial tandem dimers in pKom504, pSet102 and pPS103 (1.4, 1.8 and 1.6 unit-length partial dimers respectively), contain only one LIR in each (Figure 2.2). Thus, the agroinfectivities of these homodimeric partial dimers must have been invoked by the homologous recombination mechanism of release and the replicative release mechanism is ruled out.

The agroinfectivities of pSet107 and pKom603 (1.1mers with two LIRs; Figure 2.2) were dramatically greater at 83 and 88%) than those of pSet102 and pKom504 at 32 and 45% (Table 2.3,). In addition, 1.1mer-agroinoculated seedlings display streak symptoms 2-3 days before the symptoms appear on pKom504- or pSet102-agroinoculated seedlings (Figure 2.3). Thus, these dramatically higher agroinfectivities and shorter periods for symptoms to appear, are presumably invoked by the replicative release mechanism (Figure 1.4; Stenger *et al.*, 1991). However the homologous recombination release mechanism cannot be ruled out for the 1.1mers since homologous recombination presumably can occur between the LIRs (Figure 1.4). These results, however, strongly suggest that the replicative mechanism of release is the favoured and more efficient mechanism of release than homologous recombination for MSV-Kom and MSV-Set, corroborating the results obtained by Stenger *et al*, 1991 for BCTV and Heyraud *et al.*, 1993 for WDV.

Agroinoculated decapitated three-day-old seedlings are initially stressed and to a certain extent deformed due to the treatment. Most of the plantlets survive and grow to be normal

healthy plants, however there are a few (approximately 0-10%) that do not recover; presumably due to unrecoverable damage done to the meristem along with the severity of the disease to the host. A mortality rate of 4% was obtained in an agroinoculation experiment of non-infectious pBI121 - the binary vector used to transfer homodimeric viral genomes - presumably due to the physical manipulation and additional biological stress of the *Agrobacterium* infection the seedlings have to endure (Table 2.4). However the mortality rates of seedlings agroinoculated with homodimeric dimers containing only one LIR was twice as great at 8% suggesting that the early virus infection of plants dramatically increases the biological stress on the seedlings. Further, the mortality rate for the homodimeric 1.1mers with two LIRs is greater by a factor of two; at 16% in two experiments (Table 2.4). This is an indication that many more infectious unit-length genomes are released by replicative release than by homologous recombination release thereby significantly increasing the biological stress on the seedlings. These results corroborates the work of Stenger *et al.*, 1991 and Heyraud *et al.*, 1993) where they found that genomes between LIRs were released at greater numbers thereby increasing their rates of infection of BCTV and WDV respectively.

2.5.2 MSV-Kom, MSV-Set and PanSV-Kar host ranges and symptoms

The distinction of agroinfectivities of the three viruses that are represented by the DNAs cloned in pKom500, pSet100, and pPS103 is supported by their symptomatology and host ranges as determined by leafhopper transmission tests. MSV-Kom severely stunts and causes continuous streak up most of the length of the leaves of most sweetcorn and maize cultivars tested in this study; whether they were considered tolerant cv.s or not (Tables 2.5, 4.3 and 5.1). MSV-Set moderately infects sweetcorn and susceptible maize cv.s, however it causes very mild or no streak in the cv.s considered to be MSV-tolerant or resistant (Tables 4.3 and 5.1).

Although PanSV-Kar causes mild streak in Jubilee sweetcorn in agroinoculation tests, its ability to infect other "susceptible" cultivars such as popcorn and Witplat was not apparent in leafhopper transmissions and it was unable to produce streak symptoms in all the MSV-tolerant cv.s tested (Table 2.5). There are three possible factors that may influence PanSV-Kar's leafhopper transmission infectivity rates: First; while 3-day-old seedlings are stressed due to the nature of the agroinoculation method, it is possible that they are more vulnerable to infection, as the mortality rates suggest (mentioned above), than the 10-14 day old seedlings used in leafhopper transmission tests. An indication of this phenomenon is that streak symptoms observed on Jubilee sweetcorn were generally slightly more severe in 3-day-old agroinoculated seedlings than the 10-14 day old plantlets (already soil established)

used in leafhoppers transmission tests. Secondly; it is possible that some other species of *Cicadulina* may be more efficient vectors of PanSV than *C. mbila* (Briddon, *et al.*, 1992).

A third factor is that the efficacy of acquisition of the leafhopper depends on the extent of the streak caused by the Mastrevirus on a given host since virus distribution correlates with symptom expression in the plant (Briddon *et al.*, 1992). Thus the mild symptoms produced by PanSV-Kar in Jubilee sweetcorn hosts may be the cause of inadequate levels of virus availabity on infected leaves for leafhoppers to acquire and transmit PanSV-Kar effectively. Since von Wechmar and Hughes (1992) found that PanSV-Kar could be transmitted to sweetcorn or to two *P.maximum* subspecies with the use of leafhoppers, the use of PanSV-Kar's original host, *P.maximum*, as the acquisition host would most likely provide a more conclusive host range. However *P.maximum* was unavailable at the time these experiments were performed. Bock *et al.*, (1974) reported being unable to transmit geminiviruses from *P.maximum* to *P.maximum* using leafhoppers, and von Wechmar and Hughes, (1992) were unable to transmit PanSV-Kar to any maize lines or popcorn. In addition, Clipper barley, and SST44 and SST66 wheat cv.s did not display streak symptoms when exposed to PanSV-Kar-viruliferous leafhoppers. These results suggest that the lack of transmission is inherent in PanSV-Kar, thus as was also suggested for the Kenyan isolate of PanSV, PanSV-Kar is not likely to pose a problem in cultivated crops (Briddon *et al*, 1992).

von Wechmar and Hughes (1992) found that the number of plants infected in any one leafhopper cage was related to the actual number of leafhoppers present in the cage; the larger the colony the higher the percentage of plants that became infected. However, to ensure that there was effective transmission of the viruses in these experiments, unusually large numbers of leafhoppers were used in each cage; the leafhoppers were agitated every second day to ensure an even spread of viruliferous leafhoppers feeding off all the plants in the cage, and every leafhopper experiment was quadrupled. Thus the chances of leafhoppers not feeding off any one plant was unlikely (von Wechmar, pers. comm.).

MSV-Kom and MSV-Set are both persistent viruses in Jubilee sweetcorn with moderate to severe streak, thus the MSV-Kom- and MSV-Set-acquisition and transmission ability of leafhoppers is presumably not a limiting factor in these type of experiments. The streak symptoms of these viruses are distinctive, with MSV-Set causing moderate whiter stippled streak in Jubilee sweetcorn whereas MSV-Kom causes severe continuous yellow streak which can develop into large areas of chorotic clearing, as well as noticeably stunting the growth of the plant (Figure 2.4). The two viruses have especially distinct streak symptoms in popcorn where MSV-Set causes mild stippled streak and MSV-Kom causes severe

continuous streak (Figures 2.5 a and b). In the susceptible lines of maize such as popcorn and Witplat, MSV-Set produces moderate streak whereas MSV-Kom is able to produce severe to very severe streak symptoms (Table 2.5). MSV-Set is however unable to produce streak in all the tolerant or resistant maize lines (except for a very mild streak on PNR 6549) whereas MSV-Kom causes moderate to severe streak in all maize lines tested (Tables 2.5, 4.3 and 5.1; Figures 2.4 and 5.4). These results corroborate the findings of Hughes *et al.* (1992).

The severity of streak symptoms on wheat cv.s tends to be more severe when infected with MSV-Set and less severe for MSV-Kom. Although MSV-Kom - along with MSV-Set - can cause severe streak in susceptible wheat varieties such as the susceptible SST66, Adam Tas, Gamtoos, Nantes and Sterling cv.s (Table 5.2; Figure 2.6) it is milder than MSV-Set on varieties that MSV-Set can be moderate to severe on such as the Chokka, Festiquay and Dias cv.s. Thus the MSV-Kom and MSV-Set host ranges are distinct. Also there are wheat cv.s that are resistant (i.e do not produce streak) to MSV-Kom, but produce mild streak when infected with MSV-Set; such as supposedly resistant variety SST44 (Figure 2.6) and the Rust differentials Festiquay, Marquis and Agent cv.s. Both MSV-Set and MSV-Kom are severe on Clipper barley; however MSV-Set-infected leaves are stippled whereas MSV-Kom-infected leaves can show large areas of chlorisis (Figure 2.5c and d). Severe streak in wheat and barley cv.s tested resulted in deformation such as curling and stunting of leaves (Figures 2.5, 2.6 and 4.5). These results strongly suggest that MSV-Kom, MSV-Set and PanSV-Kar are biologically distinct viruses with MSV-Kom and MSV-Set more closely related than they are to PanSV-Kar, since the biological properties are more similar between MSV-Kom and MSV-Set.

2.5.3 Southern blot detection of viral DNA in infected plants

The Southern blot detections of MSV-Kom, MSV-Set and PanSV-Kar using the respective DNA probes also reflected their symptom severities in agroinoculated or leafhopper-transmitted Jubilee sweetcorn seedlings. While equal amounts of total plant DNA extracts from infected plants were loaded onto the agarose gels and capillary blotted onto nylon membranes, and equal amounts of the respective labelled DNA probes were used in the hybridisation tests, different exposure times were required to obtain detectable results in determining whether viral DNA was present (Figure 2.7). For the MSV-Kom, MSV-Set and PanSV-Kar hybridisations, 15, 30 and 45 minutes exposure times respectively were required to obtain approximately equal banding pattern intensities on X-ray films for the three viruses. Thus it appears that the more severe the infection, the more viral DNA is present in infected plants. This result corroborates the electron micrograph results where more geminate

particles were observed for MSV-Kom infected leaf-dip preparations than for MSV-Set, and more geminate particles were observed for MSV-Set than was found for PanSV-Kar where it was very difficult to find particles under the EM under the same conditions (M. Jaffer, pers. comm.).

Pvu II-restriction digestion of the total DNA extracts from infected plants indicated that the expected viruses - MSV-Kom, MSV-Set and PanSV-Kar - were causing the infections. In Southern blot detection tests plant DNA extracts from agroinoculated plants displaying streak symptoms always showed positive hybridisation whereas DNA extracts from symptomless agroinoculated plants or symptomless plants that were exposed to viruliferous leafhoppers, never hybridised positively. This result confirmed that DNA was not present (unless in very small amounts) unless the plant showed streak symptoms, whether the plant was agroinoculated or exposed to viruliferous leafhoppers. Further, pBI121-agroinoculated plant never displayed streak and when the blots with positive viral hybridisation results were stripped and hybridised with a pBI121-labelled probe the blots showed negative results thus indicating that pBI121, the binary vector plasmid used in all the experiments in this project, did not interfere in the agroinoculation and subsequent leafhopper transmission tests.

In hybridisation experiments cross-hybridisations were observed where the three probes cross-hybridised with MSV-Kom, MSV-Set or PanSV-Kar DNA. As depicted in Figure 2.7 the MSV-Set probe cross-hybridised with MSV-Kom-infected total DNA extract DNA (Figure 2.7, right blot lanes 1 and 2) and the MSV-Set probe hybridised to MSV-Kom-infected plant DNA extract (blot was exposed for for 30 -45 minutes; not shown). This is an indication that the nucleotide sequence of the genomes of MSV-Set and MSV-Kom are related enough to cross-hybridise and thus have a certain degree of homology. A similar result was observed in hybridisations between PanSV-Kar and MSV-Set or MSV-Kom also indicating that the MSV and PanSV-Kar sequences are related. However, the exposure times necessary to be able to detect the cross-hybridisation were approximately 60-90 mins. This is an indication that there is homology between the PanSV-Kar nucleotide sequences and the MSVs but to a lesser extent than between the MSVs. This corroborates Hughes *et al.*, (1992) and the findings in Chapter 3 finding that their sequences are related to different extents.

4 distinct DNA bands were obtained from agroinoculated plants or plants exposed to viruliferous leafhoppers, which presumably represent the different forms of DNA normally associated with Mastreviral infections (Woolston et al., 1988; Lazarowitz, 1988). These are, as indicated in Figure.7, for MSV-Kom, MSV-Set and PanSV-Kar; ds open circular DNA, ds linear DNA, ss linear and ss circular DNA. Single stranded DNA was identified as the DNA

bands that were unaltered when the total plant DNA extracts from infected plants were digested with *Pvu* II (Figure 2.7).

Hybridisation experiments using the identical MSV-Kom, MSV-Set or PanSV-Kar DNA probes on undigested and *Pvu* II-digested original total DNA extracts from the plants these viruses were isolated from, indicated identical hybridisation patterns as depicted in Figure 2.7. These results indicate that the MSV-Kom, MSV-Set and PanSV-Kar RF-DNAs cloned into pKom500, pSet100 and pPS100 respectively were indeed from their original infections and the viral progeny resulting from agroinoculations are such, thereby conventionally fulfilling Koch's postulates.

2.6 CONCLUSIONS

The symptomatology and/or host ranges for MSV-Kom and PanSV-Kar appear to corroborate those for MSV strains (Damsteegt, 1983; von Wechmar and Hughes, 1992) and PanSV-Ken (Briddon *et al.*, 1992) respectively. MSV-Set, although isolated from a *Setaria* grass species, appears to be more closely related to MSV-Kom than to PanSV-Kar as suggested in the host range results, agroinfectivities and Southern blot tests.

The availability of agroinfectious *Agrobacterium* clones of a severe, a moderate and a mild Mastrevirus could be a potential tool for determining the susceptibilty or resistance of new graminaceous crop hybrids or even prospective genetically engineered Mastrevirus-resistant crops in controlled tests. Additionally, as many geminivirologists have done, the ability to produce infection with cloned DNA provides a route for the study of mutant virus genomes produced *in vitro* and the study of marker genes in whole plants, allowing progress towards both the understanding of geminivirus molecular biology and the use of cereal geminivirus as plant transformation vectors. Indeed, mutant Mastreviruses on many occasions have been used to determine the functions of ORFs in geminiviruses; however, the genomic domains expressing biological characters - such as host range of Mastreviruses - have not been fully identified.

The availability of different strains of Mastreviruses with different biological characteristics such as those of MSV-Kom, MSV-Set and PanSV-Kar can be potentially used in genome replacement experiments in determining biological characters or functions of Mastreviruses, provided they are sufficiently related. The following Chapter (3) reports the determination of the MSV-Set nucleotide sequence and characterises it along with those of

MSV-Kom and PanSV-Kar, as well as phylogenetically determining their relatedness with respect to other Mastreviruses previously sequenced.

CHAPTER 3

GENOMIC CHARACTERISATION OF THREE MASTREVIRUSES

3.1 SUMMARY

The complete nucleotide sequence of the infectious clone of MSV-Set, pSet100, was determined. This was compared with the genomic sequences of MSV-Kom and PanSV-Kar, which were determined previously. The genome sizes of MSV-Set, MSV-Kom and PanSV-Kar are 2701, 2690 and 2705 nucleotides (nt) respectively. Two virion-sense (V1 and V2) and two complementary-sense (C1 and C2) open reading frames (ORFs), typical of Mastreviral genomes, were identified in each of the sequences. Comparative investigations and alignments were used to identify putative C1/C2 introns, Rep protein motifs, potential DNA primer binding sites and non-coding and potential control sequences typical of Mastrevirus genomes. Phylogenetic analyses on the nucleotide sequences and the putative amino acid sequences of the movement, coat and replication-associated proteins (MP, CP and Rep respectively) indicate that MSV-Set is distinct from, yet grouped with the MSV group of Mastreviruses isolated from maize. MSV-Kom shares a nt sequence identity of >96% to other sequenced MSVs isolated from maize. In contrast, MSV-Set and PanSV-Kar genomes share 78% and 60% nt sequence identity with MSV-Kom respectively. The PanSV-Kar genome shares a 89% nucleotide identity with the PanSV from Kenya (Briddon *et al.*, 1992).

3.2 INTRODUCTION AND RATIONALE

Mastreviruses are biologically and genomically distinct from the other geminivirus genera (Chapter 1). A diagrammatical representation of the the unique genome organisation of Mastreviruses is presented in linear form in Figure 3.1. Small ssDNA molecules about 80 bases in size have been isolated from Mastrevirus particles and shown to prime the synthesis of a complementary DNA strand *in vitro* (Donson *et al.*, 1984, 1987; Anderson *et al.*, 1988; Hayes *et al.*, 1988). The short intergenic region (SIR) comprises the primer binding site (PBS) for second strand synthesis to form ds RF-DNA. The genomes of geminiviruses replicate via double stranded (ds) intermediates by a rolling circle mechanism (rolling circle replication; RCR) in the nuclei of plant host cells (Saunders *et al.*, 1991; Koonin and Ilyana, 1992; Laufs *et al.*, 1995; Accotto *et al.*, 1993).

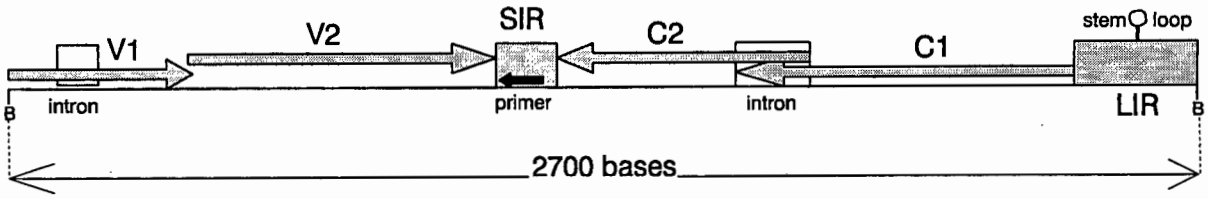

Figure 3.1: Linear representation of the genomic organisation of Mastreviruses

V1: virion sense ORF 1 encoding viral movement protein (MP) with an intron

V2: virion sense ORF 2 encoding viral coat protein (CP)

C1 and C2: complementary sense ORFs 1 and 2 respectively encoding Rep A and B respectively (replication associated proteins) where an intron exists in the overlapping region

SIR: short intergenic region where the proposed primer (black arrow) site for second strand synthesis is found

LIR: long intergenic region in which transcription occurs bidirectionally from and where the hairpin loop (stem-loop) structure containing the absolutely conserved TAATATT↓AC nanomer is found.

B: unique *Bam* HI restriction site found in all MSVs and closely related strains; diagram shows the genome linearised at this site.

Mastreviruses have circular ssDNA genomes of approximately 2700 bases in length.

Mastreviruses have four functional open reading frames (ORFs), depicted in Figure 3.1, which are bidirectionally transcribed diverging from the non-coding long intergenic region (LIR) (Mullineaux *et al.*, 1984; Morris-Krsinich *et al.*, 1985). On the virion sense (+) strand, V1 and V2 open reading frames (ORFs) encode the viral movement and coat proteins (MP and CP) respectively. The product of the V1 transcript has been shown to encode a non-structural protein of 11 kDa (Mullineaux *et al.*, 1988) which is essential for virus spread and subsequent disease development within the host but is not required for virus replication (Boulton *et al.*, 1989; Boulton *et al.*, 1993; Lazarowitz *et al.*, 1989). Recently, a 76 nucleotide intron for MSV and an 88 nucleotide intron for DSV, with plant intron features was identified within the V1 MP gene upstream of the V2 CP gene (Wright *et al.*, 1997; M. Boulton, pers. comm.). Spliced and unspliced forms of virion-sense transcript were found to be produced at different efficiencies for both MSV and DSV, suggesting that splicing is a feature in the regulation of early and late gene expression (Wright *et al.*, 1997).

The V2 transcript encodes the CP of about 30 kDa (Morris-Krsinich *et al.*, 1985). Mutations in V2 prevent the accumulation of viral ssDNA in protoplasts and around the inoculation site in plants (Boulton *et al.*, 1989, 1993). The mechanism by which the CP and MP of Mastreviruses mediate systemic infection is not clear although the requirements for movement of Begomoviruses have been studied more thoroughly (Chapter 1). ss and ds DNA was found to bind with the 104 N-terminal amino acid residues of the MSV CP in a nucleotide non-specific manner (Liu *et al.*, 1997). It is proposed that the CP of

Mastreviruses is involved in viral DNA nuclear transport as well as encapsidation and may play a role in intra- and inter-cellular movement and systemic infection.

An intron exists in the overlapping complementary (-) sense ORFs, C1 and C2 (Rep A and Rep B) which encode the replication-associated (Rep) protein. Spliced and unspliced mRNA transcripts have been detected in wheat infected with wheat dwarf virus (WDV) (Schalk *et al.*, 1989). A fused 41-kDa C1-C2 polypeptide resulting from intron excision was shown to be necessary for DNA replication of WDV (Schalk *et al.*, 1989). Transcript mapping experiments revealed that splicing occurs in the mRNA of complementary sense ORFs C1 and C2 (Accotto *et al.*, 1989; Mullineaux *et al.*, 1990). Rep initiates RCR by specifically binding to sequences in the LIR and nicking the absolutely conserved TAATATT↓AC nanonucleotide of the stem-loop structure (Saunders *et al.*, 1991; Stenger *et al.*, 1991; Heyraud, *et al.*, 1993a; 1993b; Heyraud-Nitschke *et al.*, 1995; Stanley, 1995).

The LIR also contains important *cis*-acting expression control elements such as promoters and enhancers, and the SIR contains consensus polyadenylation signals of plants, AATAAA, in both strands (Mullineaux, *et al.* 1984; Lazarowitz, 1988). Upstream activating sequences (UAS) situated upstream from the hairpin loop structure is essential for the transcription of the virion-sense ORFs and has been suggested that they determine host range by binding to host-specific factors (Fenoll *et al*, 1988, 1990; Lazarowitz, 1988). Homologous iterative elements or iterons situated in the stem of the hairpin loop and upstream of the UAS are predicted to be important in the replication of geminivirall genomes (Argüello-Astorga *et al*, 1994). DNA bending dA.dT tracts normally associated with origins of DNA replication in number of viral, prokaryotic and eukaryotic systems have been identified in WDV to the right of the hairpin loop (Suàrez-Lòpez *et al.*, 1995).

Mastreviruses have been described based on host range, symptoms (Pinner *et al.*, 1988, von Wechmar and Hughes, 1992), serology (Dekker *et al.*, 1988), restriction mapping (Clarke, *et al.*, 1989), amino acid sequence comparisons (Howarth and Vandemark, 1989) and nucleotide sequence homology (Rybicki and Hughes, 1990; Hughes *et al.*, 1993; Rybicki, 1994; Padidam, *et al.*, 1995). The complete nucleotide sequences of more than 12 Mastreviruses have already been reported - and more have most likely been sequenced (Rybicki, 1994; Padidam *et al.*, 1995, E.P. Rybicki, pers. comm.). Chapter 2 has compared the agroinfectivities and host ranges of MSV-Set, MSV-Kom and PanSV-Kar RF-DNA cloned in pSet100, pKom500 and pPS100 respectively, and thus ascertained that the clones are infectious. With the biological characterisations of MSV-Set, MSV-Kom and PanSV-Kar determined in Chapter 2 the question arises as to how do these viruses differ genomically

and how do they taxonomically fit into the Mastrevirus genus? In this chapter the complete nucleotide sequence of the infectious RF-DNAs of MSV-Set contained in pSet100 is reported and characterised, along with those of MSV-Kom and PanSV-Kar contained in pKom500 and pPS100 respectively. The addition of these sequences to the collection of Mastreviruses, and the analysis of their genomes, will contribute significantly to the field of Mastre- and geminivirology.

3.3 MATERIALS AND METHODS

3.3.1 Deletion cloning and sequencing MSV-Set

Ordered series of deletion clones and subclones derived from pPS100 and pKom500 were used to sequence the entire genomes of PanSV-Kar and MSV-Kom respectively. Selected templates cloned in pUC18 or pSK were subjected to bidirectional Sanger dideoxynucleotide chain termination plasmid sequencing using the Sequenase® II kit and protocols (United States Biochemical Corporation) or using the Taqtrack® Taq polymerase method of sequencing from Promega® (Fiona Hughes, Di James, Janet Willment, Mark Fyvie and Ed Rybicki, unpublished).

Similarly, deletion clones and subclones of the MSV-Set RF-DNA inserted in the *Bam* HI site of the MCS of pUC18 vector (pSet100) were prepared using standard cloning techniques (Sambrook *et al.* 1989), and sequenced in both directions. Restriction endonuclease recognition sites found in the MCS and in the RF-DNA were used to make a series of ordered deletion clones and subclones. pSet100 deletion clones were prepared by religating *Sal* I, *Hind* III, *Eco* RI and *Sac* I single digests (Appendices A1, A4, A5 and A6). Subclones were prepared by "shotgun" cloning the resulting RF-DNA fragments of *Sal* I, *Hind* III, *Eco* RI and *Sac* I single digests. To obtain *Bgl* II/*Bam* HI subclones, pSet100 was digested with *Bgl* II and partially digested with *Bam* HI (Appendices A1 and A2). After ammonium sulphate/isopropanol/ethanol precipitation, pSet100/*Bgl* II.p*Bam* HI double digested DNA was religated (Appendix A4). Since *Bgl* II- and *Bam* HI-digested DNA share identical single-strand 4 base cohesive ends, the *Bgl* II/*Bam* HI pSet100 deletion clones were obtained by religation (Appendix A4). Transformations were done using competent *E.coli* strain, JM105, (Appendix A5) and deletion clones and subclones were selected by digesting plasmid minipreparations (Appendix A6) of the transformed white colonies with restriction endonucleases that would result in a distinguishable banding pattern on an agarose gel (Appendices A1 and A3). Subclones and deletion clones were prepared using silica-based anion exchange columns of the commercially available plasmid isolation kit (Nucleobond®

AX100 columns Machery-Nagel) (Appendix A11) and used as templates in sequencing reactions (Appendix B). Selected template plasmids were subjected to bidirectional Sanger dideoxynucleotide chain termination plasmid sequencing using the Sequenase® II kit and protocols (United States Biochemical Corporation) or using the TaqTrack® *Taq* polymerase method of sequencing from Promega® (see Appendix B for materials and methods used). M13 derived universal reverse and forward primers (1212F and 1233R) were used in the priming reactions and ^{35}S-dATP (Amersham®) was used in the labelling reactions. Figure 3.2 depicts the deletion clones and subclones that were used to determine the entire MSV-Set RF-DNA sequence.

Unambiguous sequence data obtained off the reverse primer (1233R) from templates pSet100.2 and pSet100.5 was used to design oligonucleotides to extend and complete the entire plus-strand sequence. The Genetics Computer Group Inc. (GCG) VAX mainframe package, Version 7.1 was used to design the primers designated P1 and P2 (Table 3.1 and Figure 3.2). Similarly, unambiguous sequence data obtained off the reverse primer (1233R) from template pSet100.6, and off the forward primer (1212F) from template pSet100.3, was used to design oligonucleotides, P4 and P3 respectively. P3 and P4 were used in sequencing reactions to complete the MSV-Set sequence in the reverse direction (Table 3.1 and Figure 3.2). Oligonucleotides were purchased from the Department of Biochemistry, University of Cape Town. Each clone was sequenced twice on average in one/both directions, where required. Sequencing reactions were repeated where the sequence data showed discrepancies.

3.3.2 Sequence analyses and comparisons with other Mastreviruses

Sequence data was stored, assembled and analysed using the Genetics Computer Group, Inc. (GCG) VAX mainframe package, Version 7.1, and DNAMAN for Windows®, Version 2.2 (Lynnon Biosoft, Quebec). The sequences of BCTV (beet curly top virus; Stanley *et al.*, 1986), CSMV (chloris striate mosaic virus; Anderson *et al.*, 1988), DSV (*Digitaria* streak virus; Donson *et al*, 1987), MiSV (miscanthus streak virus; Chatani *et al.*, 1991), MSV-K (Kenyan isolate of maize streak virus; Howell, 1985), MSV-N (Nigerian isolate; Mullineaux *et al.*, 1984), MSV-SA (South African isolate; Lazarowitz, 1988), PanSV-Ken (*Panicum* streak virus, Kenyan isolate; Briddon *et al.*, 1992), SSV-N (sugarcane streak virus, Kwazulu-Natal isolate; Hughes *et al.*, 1993), TYDV (tobacco yellow dwarf virus; Morris *et al.*, 1992), MSV-Reu (Reunion isolate of maize streak virus; Peterschmitt *et al.*, 1996) and WDV (wheat dwarf virus; MacDowell *et al.*, 1985) were obtained from the GENBANK databases. Multiple sequence alignments between amino acid sequences derived from MP, CP and Rep proteins were done using PILEUP on the GCG package and the CLUSTALW package

(Version 1.5; Thompson *et al.*, 1994). Calculation of pairwise percent sequence distance values from aligned sequences was done using the neighbour-joining option in CLUSTALW. The data sets were used in the programme TreeView for Windows Version 1.2 (Page, 1996) to generate phylogenetic trees by the neighbour joining method of Saitou and Nei (1987), modified by Studier and Keppler (1988). Multiple nucleotide sequence alignments were done using the optimal alignment option on the DNAMAN package. (Feng and Doolittle 1987, Thompson *et al.*, 1994). Phylogenetic trees were generated using TreeView Version 1.2.

The genomes were screened in all six reading frames for regions with the potential to code for proteins. ORFs were identified by using the CODONPREFERENCE command in the GCG package and by comparing these ORFs with those of the published sequence of MSV-SA (Lazarowitz, 1988). ORFs were translated and predicted amino acid sequences were analysed using the GCG package. Hydrophobicity profiles of the amino acid sequences of the MP, CP and Rep proteins were compared to those of MSV-SA using the GCG and DNAMAN packages.

Table 3.1: Primers used to extend MSV-Set sequences

Primer	Size (bases)	Sequence (5'-3')	Genome location*
P1	18mer	TCGCATACCCCTCCGCAC	850-867
P2	20mer	CTCGAAGTACTGAAGTTTAG	2008-2027
P3	20mer	ACTTGTGGAAGTCAACGTTC	1016-1035
P4	18mer	CATTCCTAGGAAGAAATC	2168-2185

(*) nt pos. see MSV-Set sequence in Figure 3.2

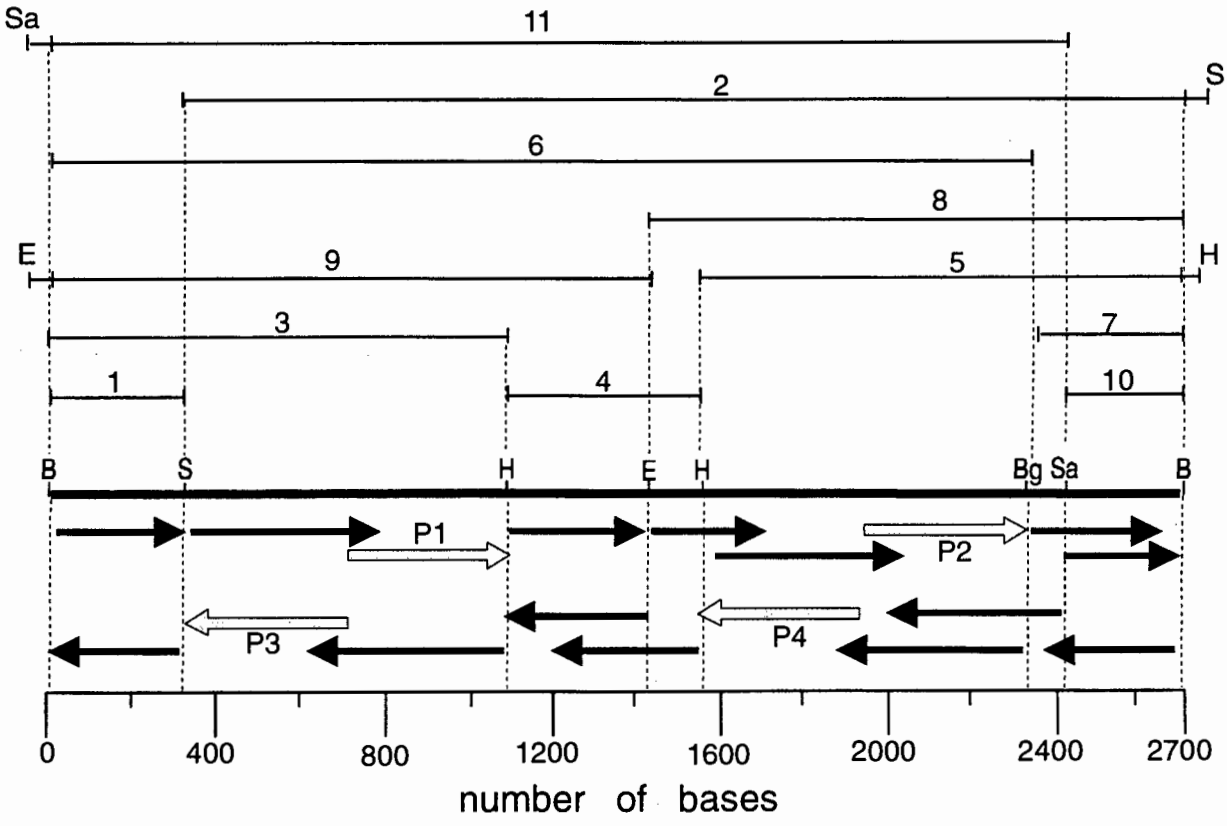

Figure 3.2 Sequencing strategy to obtain entire MSV-Set RF-DNA sequence

Subclones and deletion clones sequenced are shown above the MSV-Set RF-DNA map, numbered 1 to 11. These are designated pSet100-1, pSet100-2... and pSet100-11 in this study. Arrows below the map indicate the regions of sequence obtained which were assembled. Shaded arrows indicate the sequences obtained using primers, P1, P2, P3 and P4 to complete the sequence. Arrows pointing to the right contributed to the plus-strand sequence and those pointing the left to the complementary strand. Scale at the bottom indicates number of bases in proportion to the length of the MSV-Set genome. Sa = *Sac* I, S = *Sal* I, E = *Eco* RI, H = *Hind* III, B = *Bam* HI, Bg = *Bgl* II.

3.4 RESULTS

3.4.1 The complete nucleotide sequence of MSV-Set

The total nucleotide sequence of an infectious clone of RF-DNA of MSV-Set was determined on both strands. Each sequence reaction autoradiograph was reread until there were no discrepancies in at least three consecutive readings and the sequences of both strands were complementary. The previously determined restriction map was consistent with the genome sequence. The nucleotide sequence data has been submitted to the GenBank nucleotide sequence database, has been assigned the accession number AF007881 and is presented in Figure 3.4. Sequence orientations were deduced by comparisons with the published sequence of MSV-SA. The first nucleotide (A) after the Rep protein nick position of the absolutely conserved nonanucleotide motif TAATATT↓AC is defined as position 1 of the reported sequences in conformity with the *Geminiviridae* Study Group proposals (E.P. Rybicki, pers. comm.). The inverted repeats situated at nucleotides (nt) 2674-2691 and 3-20 flank the absolutely conserved nonanucleotide TAATATT↓AC, and the putative hairpin structure they form with the TAATATT↓AC nonanucleotide situated in the loop has a free energy value (ΔG) of -31.10 kcal/mol. Other underlined sequences are inverted repeats with free energy values [ΔG] of less than -10 that could potentially form stem-loops. These are situated in base positions 1381-1412, 2151-2177 and 2646-2667 and their ΔG values are -13.0, -10.0 and -11.9 respectively (Figure 3.4).

The sequence of MSV-Kom (Fiona Hughes, Di James and Ed Rybicki, unpublished) consists of 2690 nucleotides; this is exactly the same length as MSV-SA (Lazarowitz, 1988), and has 99% homology with MSV-SA; that is, they differ in 26 nt. PanSV-Kar was sequenced in this laboratory (Mark Fyvie, Janet Willment and Ed Rybicki, unpublished). MSV-Set and PanSV-Kar sequences were 11 and 15 nucleotides larger at 2701 and 2705 nt respectively. MSV-Set and PanSV-Kar have 78% and 60% identity with MSV-SA respectively and 59% with each other. The PanSV-Kar genome was found to have 87% homology with PanSV-Ken (Kenyan isolate of PanSV). Table 3.2 presents the CLUSTALW distance matrix comparison of the complete nucleotide sequences of all the Mastreviruses sequenced to date.

```
1    ACCTCGCCCC CTTTACCTGC GAGGGCCCGG TAGGGACCGA GCGTCTTTGA TTTAAAGCTC
61   AGATTTGCTT TTGTCGTGAA ATATCAAAGC TGCCTTGTTT AAAGAAGCCG TCCCACGCGA
121  CTATAAATTG TCTCACAAGT GCGATTCAGC CATGGATCCT CAGAGTGCTA TCTATACTCT
181  CCCACGGGTA CCCACAGCAG CTCCGACCAC CGGAGGTGTG TCGTGGAGTC ACGTCGGCGA
241  GGTAGCTATA CTGAGCTTTG TTGCTTTGAT TTGCATTTAT CTGCTTTACC TTTGGGTGTT
301  GAGAGACCTT ATCTTAGTCT TGAAGGCAAG ACGCGGGAGG TCCACGGAGG AGCTGATATT
361  TGGATCTGAA GCTGTGGATA GGAGGCACCC TATCCCTAAT ACTTTGGAAC CTACAGCTCC
421  GGTGCATCCC GGACCGTTCG TTCCAGGTCA GGGATAAGCA GTAAGCCATG TCGACGTCCA
481  AGAGGAAGCG TGCCGATGAG GCGCAATGGA ATAAGCGGTC CACCAAGAAG AAAGGTTCTG
541  CGCCGCAGGC GAAGAAGCCT GGGGGTAAGG TTGAGAAGCC TTCCCTCCAG ATACAGACTT
601  TACTCCACTC AGGTGACACG ATGATCACCG TACCTTCAGG TGGGGTCTGT GACCTCATCA
661  ATACATATGC CCGAGGGTCC GATGAGGGCA ACCGTCACAC CAGCGAAACC CTAACGTACA
721  AAGTTGGGGT CGATTACCAC TTCGTTGCTG ACGCCGCATC CTGCAAGTAC TCCAACCGCG
781  GAACAGGTGT GATGTGGCTG GTGTACGACA CGACTCCCGG CGGAAACGCA CCCACCACCC
841  AAGATATTTT CGCATACCCC TCCGCACTCA AGGCCTGGCC GACTACTTGG AAAGTTAGTC
901  GGGAGTTGTG TCATCGCTTC GTGGTGAAGC GGCGATGGCT CTTCACAATG GAGACTGACG
961  GTCGGATAGG CTCGGATACA CCACCGAGCA ACCAGAGTTG GCCGCCATGT AAGCGGAACG
1021 TTGACTTCCA CAAGTTCACT AGTGGGTTGG GAGTGAGGAC GCAGTGGAAG AACGTCACAG
1081 ACGGTGGAGT TGGTGCGATA CAGAGAGGTG CTCTGTATCT AGTCATTGCC CCCGGCAATG
1141 GTATTACGTT TACTGCCCAT GGGCAGACAC GTCTGTACTT TAAGAGTGTT GGCAACCAGT
1201 GATGAATAAA ACTGCGTTTT ATTATATCTG ATGAATGCTG AAAGCTTACA TTGATATGTC
1261 GTGGGATGAC ACGAAAAACA CACAAACAAT ACAGGGCGGT CAGGACAAAG CAGGCGGCGT
1321 ATGGGCGCGC CCGGGGCAAC ATCTTAAAAA AAAACCGATA ATAGAATTAA ACTTCCTCCG
1381 TAGGAGGAAG CTCAGGGGGA GAATACCATT TCTCCCCCGG CTGCATTACA TATATCATGC
1441 AGTTTGCTTC GAAATACTCC AGCTGCCCTG GAGTCATTTC CTTCATCCAA TCTTCATCCG
1501 AGTTGGCGAG AATGATAGTA GGTTTAGACT TCATCTGAAC CTTCTTCTTC TTCCCGTATT
1561 TAGGATTTAC AACGAATTCT TTTTGACAGC CAACCAACTG TTTCCAGCAA GGACAATATT
1621 TAAAGGGGAT ATCATCTACG ATGTTGTAGA TTGTGTCTTC GTTGTAGGAA GACCAATCAA
1681 CATTATTTTG CCAGTAATTA TGACGCCCTA AGCTTCTGGC CCATGTAGAT TTTCCTGTCC
1741 TTGTTGGACC GACGATGTAG AGGCTGCGCT TCCTTGTTCC ATAATCTGCT TGCTGGTTAA
1801 ATCATCCAGC CATTCGAGAT CAGAAATAGC GTCGTCGATG GAATAACAGT TGGGTTGAAG
1861 AAGCAGGTAC GCTTGTGGAC TGACCTGGTA TATGTTAGGT TGCAACCAAT CTTTGATTGA
1921 TTCGTTGCAT AGCAGGTCTG GTTCAGAGGT AGGATGGGGA TTGATGAACT CTTCCTGAAT
1981 ATCAGGAAAT AGTTTGTTGG CCGAGTACTC GAAGTACTGA AGTTTAGTGG ACCAATCATA
2041 AGGGAGGGAT TTCTGAACCA TGGAGAGGTA CTCCTGCTTG GAAGTAGCAT GAGAAATAAT
2101 GTCTCGCATT ATTTCATCTT TGGATGGTTT CTTATCCGAA TTTCCCTTTG AGGAGTTTCC
2161 TAGGAATGAT TTCTTCCTAG GAATGAATGT ACCTCTCTCA AACAAGGCCA GGGGTTCCTT
2221 GGTGATGTAA TCTCGTACCT TGTTGGGGCT GATTGCACTC TGAATGTTGG GGTGGAAACC
2281 ATCAATGTCG AAGAACCTTG AGTCGGTAGT CCTAACTGGT TTCTCCGTCT GTATTAATGC
2341 ATGTAAACAT TGATTTCCAT CTCTATGCGC CTCCCGGGCG CATATGATGT ACAGAGGAGT
2401 CCAATGACTA CAGAGATCCC AGATCCTCTG ACTGATGATT TCTGGCTGTT CGGGACACTG
2461 AGGATATGTC AGGAAGGTGT TAGGAGATCT ATGTGAGAAC GGACGGTTGG ATGAGGAGGA
2521 GGTCATATCC GACGACGGAG GCTGCCGCTA GCGAAGGCAG GGATGGGAGC TCTAACAAAC
2581 TCTCCTACAG CTCGCCAAGT TTGTTGCCGC CGCGGAGAGA GAAACTCCAT CGGCTTATAT
2641 AGTTGTTCTA ATGGGCCGGA CCGGGCCGGC CCA GCAGGTA AAGGGGGCGAGCAATAATAT
2701 T
```

Figure 3.4 Complete nucleotide sequence of the genome of MSV-Set

The first base position (**A**) after the Rep protein nick site in the absolutely conserved nanonucleotide **TAATATT↓AC** (bold type) is defined as position 1. The stem (18 nucleotide complementary reverse sequence) of the stem-loop is situated at nucleotides 2674-2691 and 3-20 (bold and underlined flanking the conserved nonanucleotide). The free energy value of the stem-loop is ΔG = -31.10 kcal/mol. Other underlined sequences are reverse complementary repeats with ΔG values of less than -10.

3.4.2 Analyses of MSV-Set, MSV-Kom and PanSV-Kar nucleotide sequences

3.4.2.1 *Potential coding regions*

The genome sequences of MSV-Set, MSV-Kom and PanSV-Kar were screened in all six reading frames for regions with the potential to code for proteins. ORFs were identified by using the CODONPREFERENCE command in the GCG package and by comparing these ORFs with those of the published sequence of MSV-SA. The genomic features are summarised in Table 3.4. ORFs were translated and predicted amino acid sequences were analysed using the GCG package. These putative protein sequences were also compared with those of other Mastreviruses. Hydrophobicity profiles of the putative amino acid sequences of the MP, CP and Rep proteins were found to be similar when these viruses were compared to MSV-SA or PanSV-Ken using the GCG and the DNAMAN packages, further confirming the identified ORFs congruity with those of other Mastreviruses (Figures 3.5, 3.6 and 3.7 respectively). In the virion sense strands of the genomes of MSV-Kom, MSV-Set and PanSV-Kar, two ORFs similar to the MP and the CP (products of V1 and V2 respectively), were identified. The V1 (MP) ORF of MSV-Kom potentially encodes a 10.92 kDa polypeptide - as would be obtained from the MSV-SA sequence - whereas MSV-Set and PanSV-Kar potentially encode slightly larger putative MPs at 11.02 kDa and 11.44 kDa respectively. In line with the nucleotide sequence homologies, the MP amino acid sequences of MSV-Kom and MSV-Set have identities of 99% and 75% respectively with that of MSV-SA (Table 3.3). The MP amino acid sequence of PanSV-Kar has identities of 46% and 45% to that of MSV-SA and MSV-Set respectively, which are significantly lower than the nucleotide sequence homologies of the same viruses; however it has an 81% identity with the MP of the Kenyan isolate of PanSV (Briddon *et al.*, 1992). The V1 protein has been shown to be essential to systemic spread of Mastreviruses and the conserved hydrophobic region between 32 and 52 amino acid residues have been identified in MSV-Set, MSV-Kom and PanSV-Kar and depicted in Figure 3.5 (Boulton *et al.*, 1989a; Lazarowitz *et al.*, 1989). The MP amino acid sequence distance matrices are presented in Table 3.3.

Recently, Wright *et al.* (1997) identified a plant-like intron in the V1 ORF or MP gene of MSV and DSV. These AT-rich introns, along with the consensus splice site dinucleotides (5'-GT and AG-3' respectively), were identified for MSV-Kom, MSV-Set and PanSV-Kar in nucleotide postions 241-316, 242-316 and 254-341 respectively. The size of the putative V1 intron for MSV-Kom is 76 b identical to that of MSV-Nig (Wright *et al.*, 1997). The V1 intron of PanSV-Kar is 89 b as for that of PanSV-Ken, and the MSV-Set V1 putative intron is 75 b in size.

The V2 (CP) ORFs of MSV-Kom, MSV-Set and PanSV-Kar potentially encode putative CP products of 27.01, 26.90 and 27.29 kDa respectively (Morris-Krsinich *et al.*, 1985). The CP sequence identities for MSV-Kom, MSV-Set and PanSV-Kar with MSV-SA are 99%, 84% and 69% respectively. PanSV-Kar CP sequence have identities of 89% and 67% with that of Kenyan isolate of PanSV and MSV-Set respectively. The CP sequences have been shown to be significantly more conserved amongst these viruses than those of the MPs (distance matrix presented in Table 3.3; Rybicki, 1994).

3.4.2.2 *C1/C2 introns and Rep protein motifs*

Two partly overlapping ORFs (C1 and C2) were found in the complementary sense sequences of MSV-Kom, MSV-Set and PanSV-Kar, as for all other Mastreviruses. Congruity was ascertained as described above for the virion sense ORFs (Figure 3.7). It has been shown for WDV, DSV and TYDV (Accotto *et al*, 1989; Morris, *et al.* 1992 and Schalk *et al*, 1989 respectively) and deduced for MSV and CSMV (Lazarowitz, 1988; Mullineaux *et al.*, 1984; and Schalk *et al.*, 1989 respectively) that an intron in the overlapping region of the C1/C2 transcript is spliced out to yield a fusion protein product of about 40 kDa. Similar putative introns, 92 bases in size, were situated in positions 1795-1886, 1793-1884 and 1810-1901 in MSV-Kom, MSV-Set and PanSV-Kar respectively. The intron alignments for these three distinct viruses and some selected viruses are presented in Figure 3.8. These three distinct South African viruses contain the splice acceptor and donor dinucleotides, GT and AG respectively, characteristic of mastreviruses (Schalk *et al.*, 1989). The characteristic "lariat" sequence (CTGAC) is found positioned equidistantly ahead of the acceptor dinucleotide in the three distinct South African virus genomes (Figure 3.8). However, as previously shown for SSV-N (Hughes *et al.*, 1993), the lariat sequence of MSV-Set was unambiguously CTGAT rather than the characteristic CTGAC.

The spliced C1/C2 transcript potentially codes for 41.36, 41.46 and 41.28 kDa Rep proteins for MSV-Kom, MSV-Set and PanSV-Kar respectively. Rep amino acid sequences were aligned as described above. The Rep proteins of MSV-Kom, MSV-Set and PanSV-Kar were found to have similarities of 99%, 82% and 64% with MSV-SA respectively (Table 3.2). In addition the Rep protein of PanSV-Kar was found to have identities of 86% and 65% with the Kenyan isolate PanSV-Ken and MSV-Set respectively. The distance matrix of the Rep amino acid sequences obtained from the CLUSTALW computer package is presented in Table 3.2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. MSV-SA	100	**99**	98	98	96	**78**	62	60	**60**	58	49	42	39	41	34
2. MSV-Kom	**99**	100	**98**	**98**	**95**	**78**	**62**	**59**	**60**	**58**	**49**	**42**	**39**	**41**	**34**
3. MSV-Nig	99	**98**	100	98	96	**78**	62	60	**60**	59	49	42	39	41	34
4. MSV-Ken	98	**98**	99	100	96	**79**	62	60	**60**	59	49	42	39	41	34
5. MSV-Reu	97	**96**	97	97	100	**78**	62	60	**61**	59	49	42	40	40	37
6. MSV-Set	**82**	**82**	**81**	**81**	**81**	100	**59**	**59**	**59**	**57**	**47**	**42**	**38**	**41**	**33**
7. DSV	66	**66**	66	66	66	**66**	100	57	**58**	56	46	40	38	40	32
8. PanSV-Ken	61	**62**	62	62	62	**62**	62	100	**87**	60	48	39	39	41	31
9. PanSV-Kar	**64**	**63**	**63**	**63**	**64**	**65**	**65**	**86**	100	**63**	**47**	**40**	**39**	**41**	**32**
10. SSV-N	63	**63**	63	63	63	**64**	61	61	**63**	100	46	41	38	39	32
11. CSMV	43	**44**	43	43	43	**45**	44	43	**43**	43	100	41	40	40	31
12. TYDV	48	**48**	48	48	46	**47**	46	45	**48**	45	46	100	46	61	35
13. MiSV	44	**44**	44	44	28	**43**	44	44	**44**	42	45	49	100	45	31
14. WDV	42	**42**	42	42	43	**41**	43	44	**43**	42	45	47	49	100	33
15. BCTV	31	**31**	31	30	30	**30**	30	31	**32**	31	31	38	34	34	100

Table 3.2 Pairwise sequence identity matrices for the genomes and Rep proteins of Mastreviruses

Upper right-hand side diagonal shows the entire genome nucleotide sequence percentage identity matrix. Lower left-hand side diagonal shows the amino acid sequence identity matrix of the Rep (C1/C2) proteins. Values for Rep protein sequences were obtained using CLUSTALW multiple alignment and those of the nucleotide sequence were obtained using the PILEUP alignment option on the GCG package. **Smaller** unshaded region indicates the MSV group of viruses; the larger lightly shaded box indicates the "African streak viruses" (Hughes *et al.*, 1992) and the largest darkly shaded box indicates other sequenced Mastreviruses.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. MSV-SA	100	**99**	99	99	96	**75**	40	48	**46**	46	18	25	10	30	11
2. MSV-Kom	**99**	**100**	**98**	**98**	**95**	**75**	**40**	**48**	**46**	**47**	**18**	**25**	**10**	**30**	**11**
3. MSV-Nig	98	**98**	100	100	97	**75**	40	48	**46**	46	18	25	10	30	11
4. MSV-Ken	99	**98**	99	100	97	**75**	40	48	**46**	47	18	25	10	30	11
5. MSV-Reu	98	**99**	99	98	**100**	**76**	42	46	**45**	47	18	27	69	27	13
6. MSV-Set	**84**	**84**	**83**	**84**	**83**	**100**	**44**	**45**	**43**	**45**	**12**	**23**	**9**	**26**	**11**
7. DSV	63	**62**	63	63	**62**	**60**	100	35	**35**	48	17	20	9	22	13
8. PanSV-Ken	68	**68**	68	68	**69**	**69**	68	100	**81**	74	12	25	10	21	12
9. PanSV-Kar	**69**	**69**	**68**	**68**	**69**	**67**	**69**	**89**	**100**	**76**	**15**	**26**	**11**	**20**	**14**
10. SSV-N	66	66	66	66	67	62	69	53	55	100	12	25	7	22	12
11. CSMV	44	**44**	45	45	**45**	**43**	45	46	**45**	44	100	11	10	12	7
12. TYDV	37	**37**	37	37	**37**	**34**	37	35	**37**	36	40	100	8	43	6
13. MiSV	44	**44**	44	44	**28**	**43**	44	44	**44**	42	45	49	100	7	21
14. WDV	31	**31**	32	31	**31**	**34**	33	33	**33**	33	34	33	35	100	3
15. BCTV	22	**22**	22	22	**24**	**23**	20	19	**20**	21	21	22	21	20	100

Table 3.3 Pairwise sequence identity matrices for the movement and coat proteins of Mastreviruses

Upper right-hand side diagonal shows the MP sequence (product of the V1 ORF) identity matrix shown as a percentage. Lower left-hand side diagonal shows the CP sequence (product of the V2 ORF) identity matrix. Protein sequences were aligned using the PILEUP command in GCG computer package. Values were obtained using CLUSTALW. Smaller unshaded region indicates the MSV group; larger lightly shaded region indicates the "African streak virus group" (Hughes *et al.*, 1992); and the largest darkly shaded region indicates other sequenced Mastreviruses.

Table 3.4 Genomic features of MSV-Kom, MSV-Set and PanSV-Kar

ORF	Reading Frame	TATA[a]	Start (AUG)	Stop	Poly A[b]	Product (daltons)	No. of aa residues[c]
MSV-Kom V1	3+	120	150	453[d]	1203	10 922	101
MSV-Set V1	2+	122	152	455[d]	1205	11 015	101
PanSV-Kar V1	1+	114	145	466[e]	1240	11 440	107
MSV-Kom V2	1+	120	466	1198[d]	1203	27 008	244
MSV-Set V2	3+	122	468	1199[f]	1205	26 903	244
PanSV-Kar V2	2+	114	479	1223[f]	1240	27 289	248
MSV-Kom C1	1-	2629	2528	1712[e]	1685[g]	31 397	272
MSV-Set C1	2-	2641	2526	1710[e]	1687	31 178	272
PanSV-Kar C1	1-	2654	2564	1705[d]	1257	33 022	287
MSV-Kom C2	3-	2629[h]	1815[i]	1371[d]	1223	17 200	148
MSV-Set C2	3-	2641	1813[j]	1369[d]	1223	17 223	147
PanSV-Kar C2	3-	2654	1830[k]	1413[d]	1257	16 101	138
MSV-Kom Rep		2629	2528	1371[d]	1223	41 357	355
MSV-Set Rep		2641	2526	1369[d]	1223	41 458	355
PanSV-Kar Rep		2654	2564	1413[d]	1257	41 276	353

[a] CTATAA = most likely TATA box matching the consensus promoter sequence $T^{C}/_{G}TATA^{T}/_{A}A$ for plant genes; [b] AATAA = polyadenylation signal matching the consensus sequence $^{A}/_{G}ATAA$ for plant genes; [c] number of amino acid residues making up the protein product; [d] UAA termination codon; [e] UAG termination codon; [f] UGA termination codon; [g] another possible polyadenylation site at 1519; [h] another possible TATA box; [i, j] and [k] possible initiation codons situated within 92 base introns at genomic sequence positions 1795-1886, 1793-1884 and 1810-1901 respectively. V1 = ORF encoding MP; V2 = ORF encoding CP; C1 and C2 = overlapping ORFs encoding Rep protein.

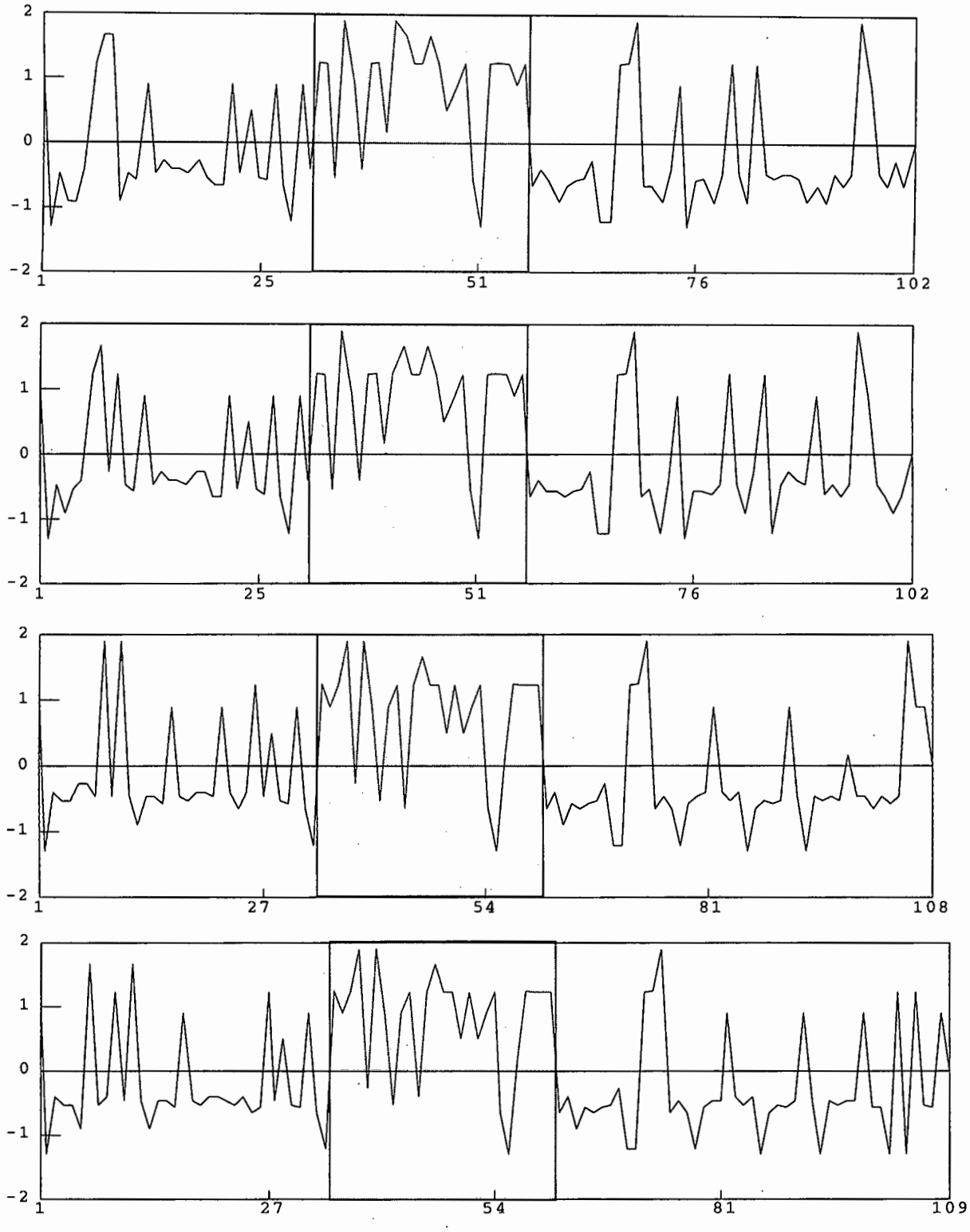

Figure 3.5 Hydrophobicity profiles of the MPs

From the top, MSV-Kom (almost identical to MSV-SA), MSV-Set , PanSV-Kar and PanSV-Ken (Briddon *et al,* 1992). Hydrophobicity values are above zero and hydrophilicity values below zero on the y-axis. X-axis indicates the amino acid count. Plots were generated using DNAMAN.

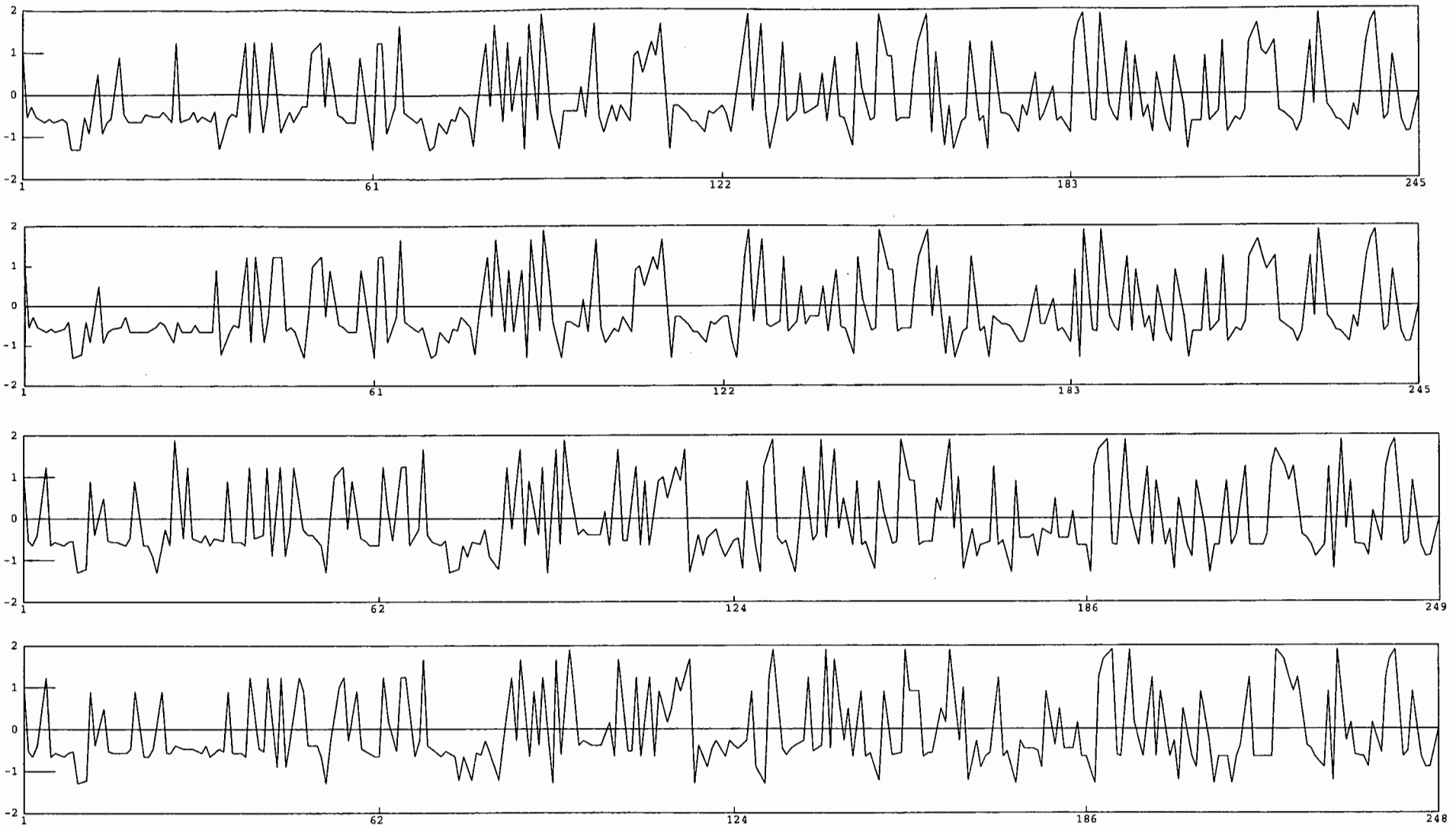

Figure 3.6 Hydrophobicity profiles of the CPs

From the top; MSV-Kom (similar to that of MSV-SA), MSV-Set, PanSV-Kar and PanSV-Ken. Hydrophobicity values are above zero on the Y-axis and the X-axis indicates the amino acid count. Plots were generated using DNAMAN.

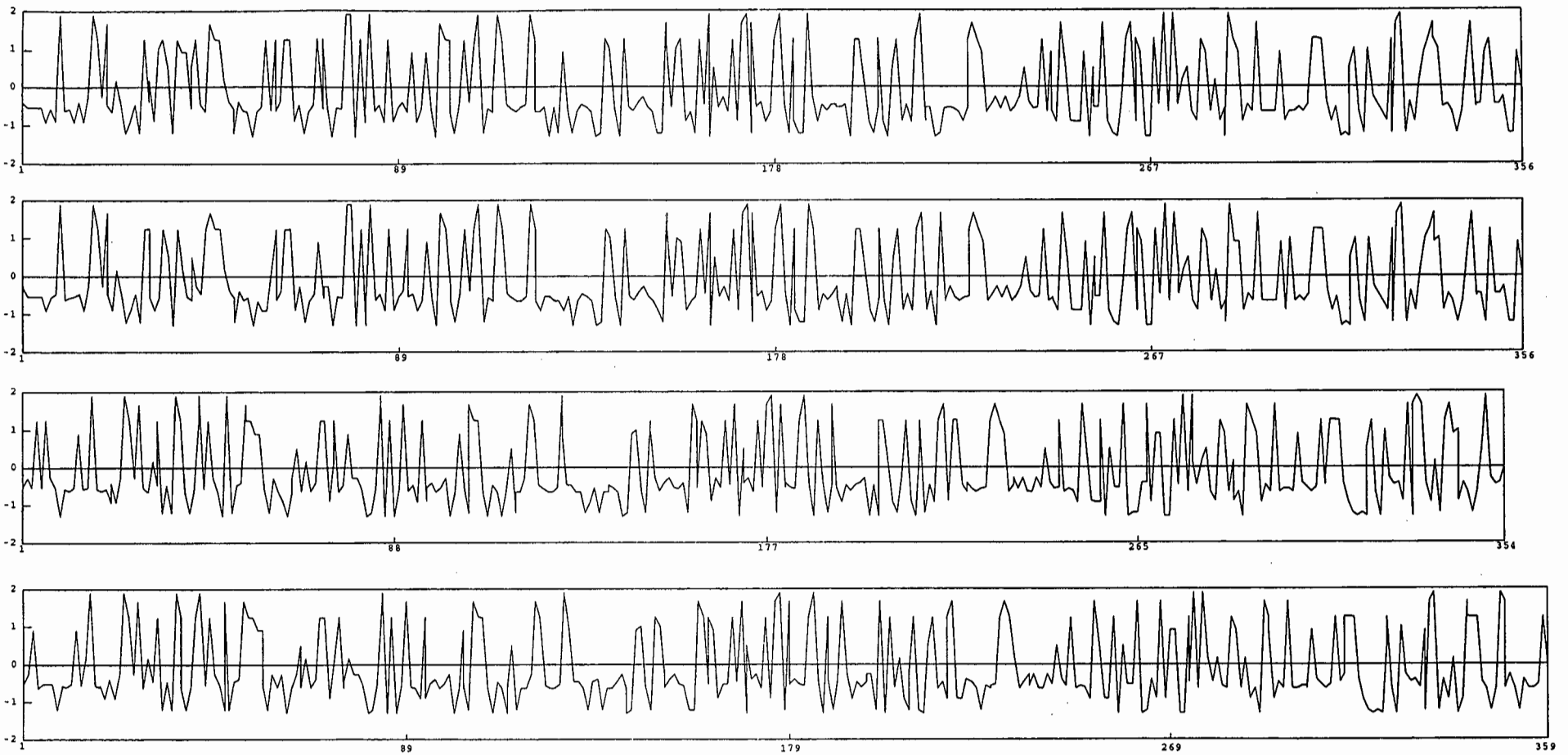

Figure 3.7 Hydrophobicity profiles for the Rep proteins

From the top; MSV-Kom (very similar to MSV-SA) , MSV-Set, PanSV-Kar and PanSV-Ken (Briddon *et al,* 1992). Hydrophobicity values are above zero and hydrophilicity values below zero. The X-axis indicates the amino acid count. Plots were generated using DNAMAN.

```
                1
MSV-SA      CCAG GT TAGTCCCGAAGCTTACATGCTCCTTCAACCTACCTGTTATACCCT
MSV-Kom     CCAG GT TAGTCCCGAAGCTTACATGCTCCTTCAACCTACCTGTTATACCCT
MSV-Set     CCAG GT CAGTCCACAAGCGTACCTGCTTCTTCAACCCAACTGTTATTCCAT
PanSV-Kar   CCAG GT CAGTCCTAAAGCATACATGCTCCTTGAACCTAGCTGCTTAAGCCT
PanSV-Ken   TCAG GT CAGTCCTCAAGCTTACAAACTCCTAGAACCTAGCTGCTTAAGCCT
SSV-N       CCAG GT TAGTCCTTCTGCATACATGCTAGCTAATCCCAGTTGCTTAACCCT
             *** **  *****    ** ***   **     * ** *  ** *    * *

                          62   66                          92
MSV-SA      CGAGGATGCAATCT CTGAC CTCCAATGGATGGATTCTGTATCC AG TCAT
MSV-Kom     CGAGGATGCAATTT CTGAC CTCCAATGGATGGATTCTGTATCC AG TCAT
MSV-Set     CGACGACGCTATTT CTGAT CTCGAATGGCTGGATGATTTAACC AG CAAG
PanSV-Kar   AGAGCAAGCCAAAG CTGAC TTAGAATGGTTGTCTGAAACAACC AG ATTA
PanSV-Ken   AGAACAGGCAATTG CTGAC TTAGAATGGTTGGATGATACAACC AG AATG
SSV-N       GGAAGAAGCAACTT CTGAT CTGATATGGATGCATGAAACATCC AG AACT
             **   * ** *    ****   *   **** **  *     * ** **
```

Figure 3.8 Aligned C1/C2 ORF intron sequences from selected Mastreviruses
Sequences are complementary to the virion sense sequence and in the 5'-3' direction. Position 1 of the intron is indicated at the GT splice donor site indicated in bold type. Position 62-66 indicates the "lariat acceptor" and position 92 of the intron is indicated at the AG splice acceptor site indicated in bold type (Schalk *et al*, 1989). Asterisks indicate conserved residues amongst. Their genomic sequence positions are as follows on the virion sense strand; MSV-SA and MSV-Kom 1795-1886; MSV-Set 1792-1884; PanSV-Ken 1812-1903; PanSV-Kar 1810-1901; SSV-N 1821-1912 (Hughes *et al.*, 1993).

Conserved sequence motifs in initiator proteins for rolling circle DNA replication were found in diverse replicons from eubacteria, eukaryotes and archaebacteria (Ilyana and Koonin, 1992). Computer analyses of the C1/Rep A ORF sequence of Mastreviruses revealed motifs resembling those of the rolling circle initiator proteins of eubacterial plasmids (Koonin and Ilyana 1992). The amino acid sequences of the putative Rep proteins of MSV-Kom, MSV-Set and PanSV-Kar were aligned using the DNAMAN computer package. The three conserved sequence motifs found in geminiviruses were identified in MSV-Kom, MSV-Set and PanSV-Kar and are presented in Figure 3.9. The C-terminal amino acid motif 3 (Figure 3.9) in MSV-Kom, MSV-Set and PanSV-Kar are all congruent with the conserved sequence Y(U)xK identified by Koonin and Ilyana (1992). Motif 3 is conserved in MSV-Kom and PanSV-Kar as Tyr-Ile-Leu-Lys in line with MSV-SA and PanSV-Ken and contains a Tyr(Y) residue thought to bind covalently to DNA. Motif 2 contains two invariant His(H) residues embedded in a run of bulky hydrophobic residues. The conserved H residues in motif 2 are speculated to function as ligands to Mg^{2+} or Mn^{2+} since this arrangement is reminiscent of the structure of metal binding centres in various metalloenzymes such as cytochrome *c*, oxidases and carbonic anhydrases (Koonin and Ilyana, 1992). The second RCR initiator motifs of MSV-Kom and PanSV-Kar are found to consist of the two H residues along with

other related viruses, however MSV-Set only has one H residue (Figure 3.9). MSV-Kom, MSV-Set and PanSV-Kar have the FLTYp motif 1 (Phe-Leu-Thr-Tyr-**Pro** with Pro not being absolutely conserved in geminiviruses) sequence in common with other geminiviruses, however in the PanSVs (-Kar and -Ken) were found to have Ser(S) in place of Pro (Figure 3.9).

Replication of the genomes of geminiviruses depends on the host replication proteins and on the viral Rep protein which is involved in initiating replication. Animal DNA tumour viruses such as SV40 are analogous to geminiviruses in this respect. The mechanism by which these viruses create an environment where viral DNA replication is permitted, involves the binding of the cellular retinoblastoma protein to a virally encoded oncoprotein through its LXCXE (**Leu**-X-**Cys**-X-**Glu**) binding motif. Xie *et al.*, (1995) identified a motif in the Rep protein of WDV similar to the retinoblastoma protein binding domain of animal DNA tumour viruses and showed, by making single amino acid changes in the LXCXE motif, that this motif is absolutely required for replication. This motif was found to be conserved in a variety of geminiviruses except for SSV-N (Xie *et al.*, 1995), and is conserved in the Rep protein sequences of MSV-Kom, MSV-Set and PanSV-Kar as **Leu**-Leu-**Cys**-Asn-**Glu** (LLCNE) (Figure 3.9).

MSV-Kom, MSV-Set and PanSV-Kar and other Mastrevirus Rep B ORF predicted amino acid sequence alignments (Figure 3.10) using the DNAMAN Version 2.2 computer package, revealed NTP-binding domains (Walker et al, 1982) resembling those found in kinases and DNA helicases (Gorbalenya *et al.*, 1990). NTP-binding motif, $EGX_4GKTX_{32}DD$ identified in the Rep protein of Begomoviruses (Gorbalenya and Koonin, 1989), have been shown to be required for geminiviral replication in $EGX_4GKTX_{32}DD$-mutant experiments of BGMV and TYLCV (Hanson *et al.*, 1995; Desbiez *et al.*, 1995). Based on this occurrence of the NTP binding motif, it was hypothesised that the geminivirus Rep protein may act as a helicase in addition to its RCR functions (Fontes *et al.*, 1992; Thömmes *et al.*, 1993; Gorbalenya, 1990; Koonin and Ilyina, 1992; Stanley, 1995).

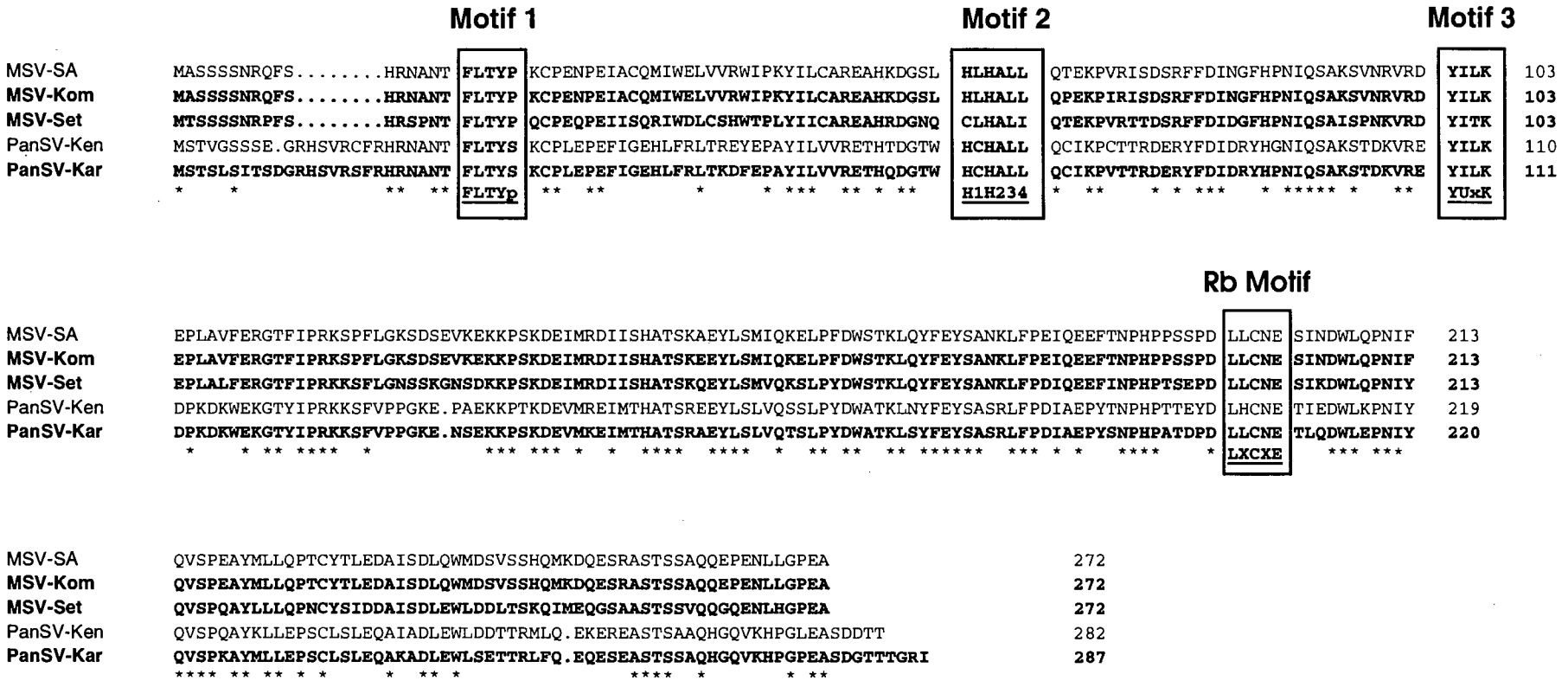

Figure 3.9 Amino acid sequence alignment of Rep A of selected mastreviruses

Amino acid sequences were obtained using the GCG and DNAMAN computer packages. Asterisks indicate amino acids that are conserved in the indicated viruses. The retinoblastoma protein binding motif and the RCR motifs 1, 2 and 3 are boxed. Sequence alignment signatures or conserved residues obtained for RCR motifs (Koonin and Ilyina, 1992) and the retinoblastoma protein binding site (Xie *et al.*, 1995) are underlined.

p indicates amino acid residue P which is not entirely conserved amongst geminiviruses; **x** indicates any amino acid residue; **U** indicates any of the bulky hydrophobic residues (I,L,V,M,F,Y or W); **1** = I,L,V,F,Y,W,S or T **2** = I,L,V,M,A or C; **3** = I,L,V,M,F,Y or W; **4** = I,L,V,M,F,Y,W,A or C.

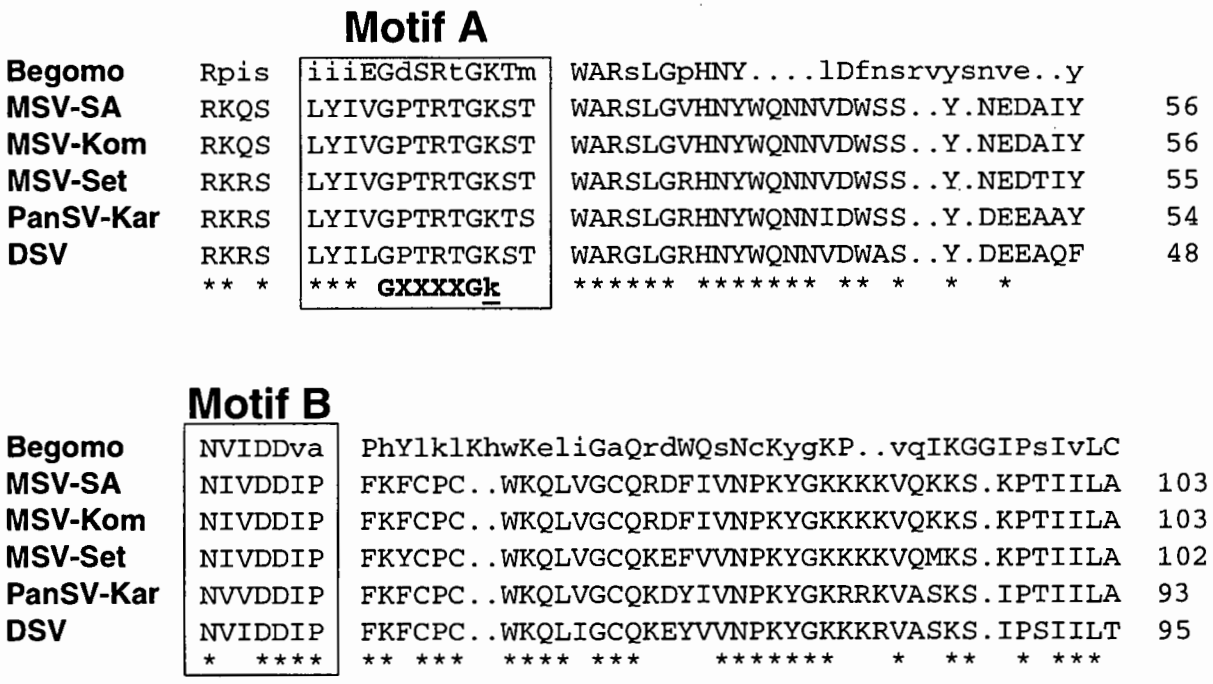

Figure 3.10: Rep B ORF sequence alignment of selected mastreviruses

Sequences were aligned using the DNAMAN Version 2.2 package. A portion of the Rep B sequences are shown including Motifs A and B putatively involved in helicase activity. Begomo = Begomovirus consensus sequence of AC1 protein product (Rep). Putative NTP-binding motif extends from Motif A to Motif B ($EGX_4GKTX_{32}DD$) and contains the conserved P-loop sequence **GXXXXGk** (Gly-Xaa_4-Gly-Lys/Ser).

3.4.2.3 *Non-coding and potential control sequences*

As is reported for other Mastreviral genomes, two apparent intergenic regions were found to be located on opposite sides of the genomes of MSV-Kom, MSV-Set and PanSV-Kar. The LIRs were found in nucleotide positions 2528-2690 and 1-150 for MSV-Kom, 2526-2701 and 1-152 for MSV-Set and 2564-2705 and 1-145 for PanSV-Kar. The LIRs for these viruses were aligned with those of MSV-SA and PanSV-Ken using DNAMAN (Figure 3.11). The three LIRs all feature the absolutely conserved nonanucleotide TAATATT↓AC. Flanking the nonanucleotide are the reverse complementary sequences that can potentially form a stem-loop structure or hairpin loop with the conserved nonanucleotide exposed as the loop (Figure 3.11). The free energies of formation (ΔG) for the three hairpin loop structures were obtained from the DNAMAN computer package; -28.70, -31.10 and -24.70 kcal/mol respectively for MSV-Kom, MSV-Set and PanSV-Kar. The stem sequences were found to be dissimilar in the three studied viruses and dissimilar to MSV-SA except for MSV-Kom.

Most likely promoter and polyadenylation sequences were identified and are summarised in Table 3.4. A GC-rich "upstream activating sequence" (UAS) in MSV-Kom immediately upstream from the stem-loop sequence was found to be essential for the transcription of V1

and V2 in MSV-Ken (Fenoll *et al.*, 1988, 1990). Potential UASs were identified in MSV-Kom, MSV-Set and PanSV-Kar (nucleotide positions 2641-2662, 2652-2673 and 2644-2670 respectively) which are in line with those of other MSV types (Fenoll *et al.*, 1990). It is interesting to note that the UASs found in MSV-Kom and MSV-Set are complete inverse repeats that can potentially form stem-loop structures (Figure 3.11). The GC rich upstream activating sequence of PanSV-Kar resembles that of MSV-Set (Figure 3.11). Maize nuclear factors were proposed to bind to G residues located on opposite faces, separated by one and a half turns of the DNA helix of double stranded RF-DNA in MSV-Ken (Fenoll *et al.*, 1990); however Hughes *et al.* (1993) showed for SSV-N that a similar mechanism would not work for this virus as the sequence between the C-rich regions was longer than in MSV. Rightward promoter element (*rpe*) equivalents were GGGCCGG double repeat sequences in MSV-Set and MSV-Kom - as is the case in MSV-SA and related MSV strains - and only a single GGGCCGG sequence for PanSV-Kar, similar to that of PanSV-Ken (Briddon *et al.*, 1992). It has been suggested that these UASs may be important determinants of host range in that they may bind host-specific factors (Hughes *et al.*, 1993). Variations in the GC-rich region of the UAS of the Begomoviruses (bipartite genome) geminivirus squash leaf curl virus, are associated with differences in host range (Lazarowitz, 1988).

Elements analogous to the proposed iterative sequence motifs (iterons; Argüello-Astorga *et al.*, 1994) found in geminivirus replication origins were identified in all three viruses (Figure 3.11). The eight base iterons for MSV-Kom (nucleotide positions 4-11, 2596-2603 and 2672-2679) were identical to those of MSV-SA which is in line with their sequence homology; however those of MSV-Set are two nucleotides smaller and found in very similar genomic regions (nucleotide positions 3-8, 2591-2596 and 2686-2691) to those of the above mentioned MSV strains. The iterons identified in PanSV-Kar are also eight bases in length and are found in nucleotide positions 7-14, 2632-2639 and 2684-2691. The potential iterons of PanSV-Ken are six or seven bases in length and are found in similar positions to those of PanSV-Kar; nucleotide positions 7-13, 2627-2634 and 2679-2685.

Two dimensional electrophoretic and electron microscope experiments have recently revealed that a defined 80-bp cluster of dT.dA tracts, located downstream from the putative stem-loop in the LIR of WDV confers DNA bending (Suàrez-Lòpez *et al.*, 1995). Stably bent DNA sequences arise when short dT.dA tracts occur periodically, a few bases apart, on a given length of DNA (Crothers *et al.*, 1990). dT.dA tracts have been shown to be associated with origins of DNA replication and are necessary for origin activity in a number of systems, e.g. bacteriophage λ, yeast and *Leishmania tarentolae* (Crothers *et al.*, 1990, Suàrez-Lòpez *et al.*, 1995). MSV-Kom, MSV-Set and PanSV-Kar in the locus depicted in Figure 3.11, have

dT.dA tracts of 2-4 bases in length. MSV-Kom was found to have eight dT or dA 3-bp tracts. MSV-Set was found to have eight 3-bp and one 4-bp tracts. In contrast PanSV-Kar along with PanSV-Ken only has three 3-bp tracts including TTCTT sequences found to detectably bend DNA (Crothers *et al.*, 1990).

3.4.2.4 *Potential DNA primer binding site*

The small intergenic regions (SIRs) of MSV-Kom, MSV-Set and PanSV-Kar are situated in nucleotide positions 1200-1371, 1183-1369 and 1225-1413. Alignments of the SIRs of selected mastreviruses and those of MSV-Kom, MSV-Set and PanSV-Kar are presented in Figure 3.12. Like MSV-Nig (Donson *et al.*, 1984) and MSV-SA, MSV-Kom has a potential primer binding site (PBS) of 84 bases whereas MSV-Set and PanSV-Kar have potential PBSs of 86 and 79 nt respectively. This compares with DSV (Donson *et al.*, 1987) and PanSV-Ken PBSs with sizes of 82 and 79 nt respectively. These sizes mentioned include the uracil ribonucleotide and ribonucleotides attached to the 3' end of the primer (Donson *et al.*, 1984, 1987).

3.4.3 Nucleotide and protein sequence comparisons and phylogenetic analysis

The amino acid sequences for the potential MP, CP and Rep proteins for MSV-Kom, MSV-Set and PanSV-Kar were deduced using the GCG (Genetics Computer Group) package and aligned with the sequences of other Subgroup I geminiviruses obtained from the GENBANK database. Alignments and phylogenetic analyses were done using the CLUSTALW computer package, Version 1.5 and trees were generated using the neighbour-joining option in the package. Distance matrices of amino acid sequences of the three proteins and the entire nucleotide sequences of all Mastreviruses were generated to establish the relatedness of MSV-Kom, MSV-Set and PanSV-Kar with the other already sequenced geminivirus genomes. Tables 3.2 and 3.3 present the distance matrices generated using the nucleotide sequences, MP, CP and Rep amino acid sequences, and Figures 3.13 and 3.14 present the phylogenies obtained. All the trees showed conclusively that MSV-Kom and MSV-Set cluster together with the other MSV strains that have been sequenced, but with MSV-Set unique amongst the other MSV strains. PanSV-Kar showed a relatively close homology to PanSV-Ken. This was shown to be the case in all four phylogenetic trees generated (Figures 3.13 and 3.14).

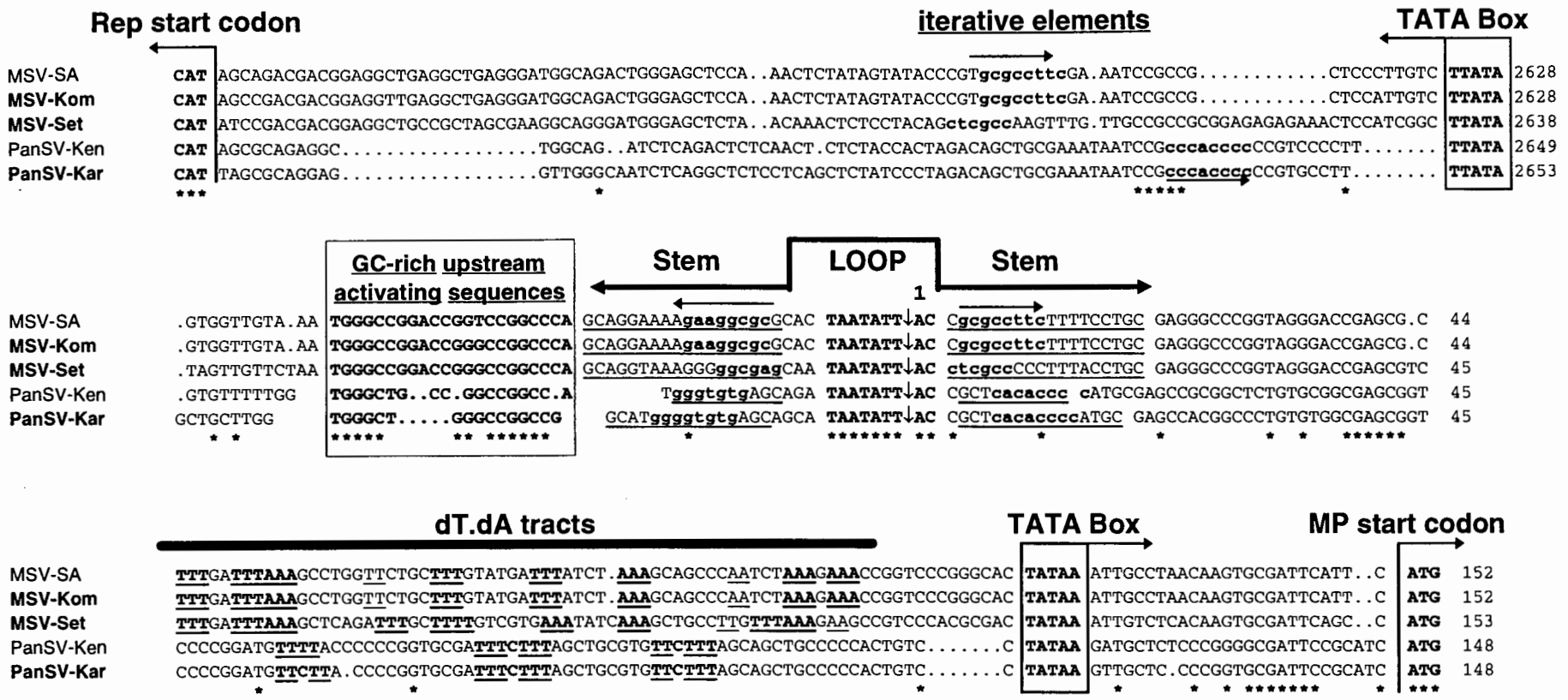

Figure 3.11 Large intergenic region alignments of selected related Mastreviruses

Virion sense LIR sequences aligned from the start codon (**CAT**) of the Rep protein through the non-coding region to the start codon (**ATG**) of the MP. TATA boxes are indicated in boxes with the directions (5'-3') indicated. Potential iterons (iterative elements) are indicated by bold lower case sequence and in the stem regions (arrows in opposite directions indicate reverse complementary iteron sequences). Stem of the stem-loop structure indicated as underlined sequences which are reverse complements. The loop structure has the absolutely conserved nonanucleotide motif, **TAATATT↓AC** indicating with an arrow (↓) the site where the Rep protein nicks the loop initiating replication of the viral genome. Asterisks indicate conserved nucleotides. The GC-rich upstream activating sequences (UAS) are indicated in bold uppercase sequence and resemble GC boxes, (Kadonaga *et al.*, 1986; Fenoll *et al.*, 1988, 1990). Putative dT.dA locus (in bold and underlined) comprising tracts normally associated with DNA bending in gene regulatory regions (Suàrez-Lòpez *et al.*, 1995).

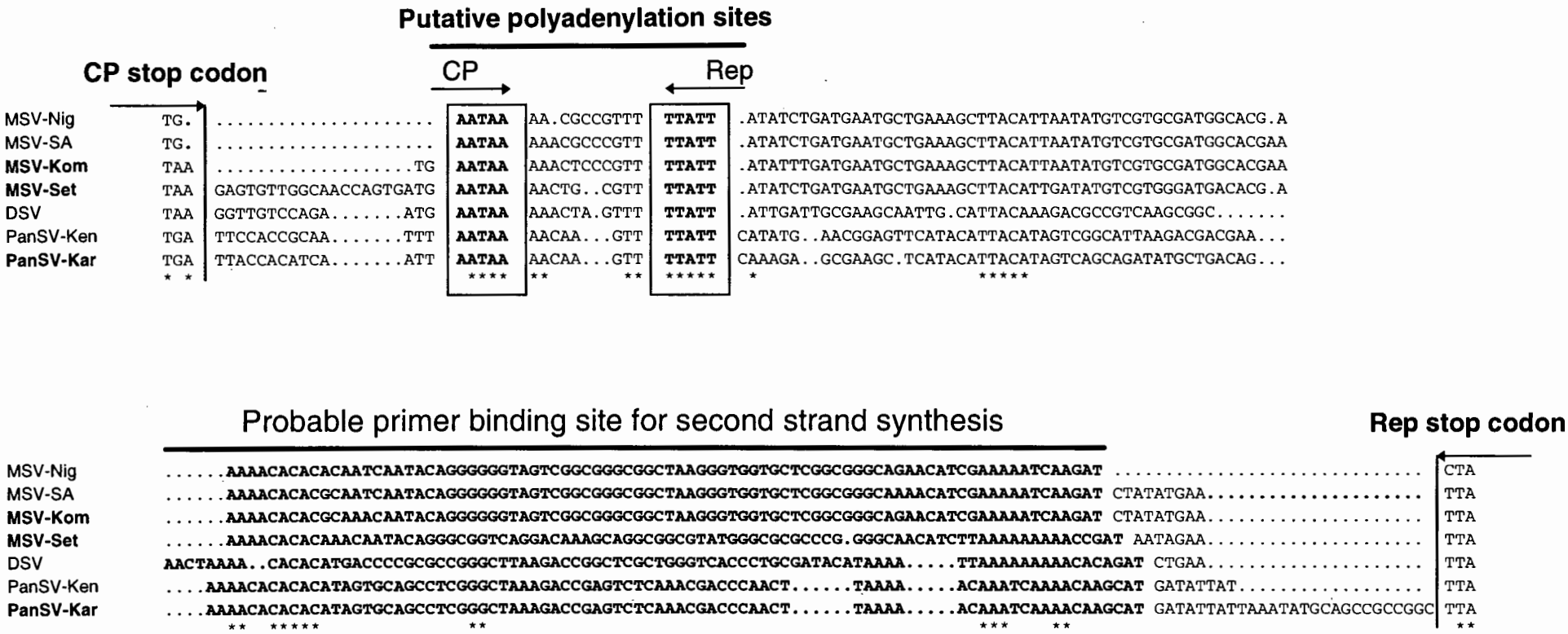

Figure 3.12 Small intergenic region alignments of selected mastreviruses and MSV-Kom, MSV-Set and PanSV-Kar

The virion sense(+) strand is shown, from the stop codons of the CP ORF to the stop codons of the Rep or C1/C2 ORF. Polyadenylations sites are indicated and the most probable primer binding sites for second strand synthesis are indicated and shown in bold type. See table 3.2 for nucleotide positions of stop codons and polyadenylation sites.

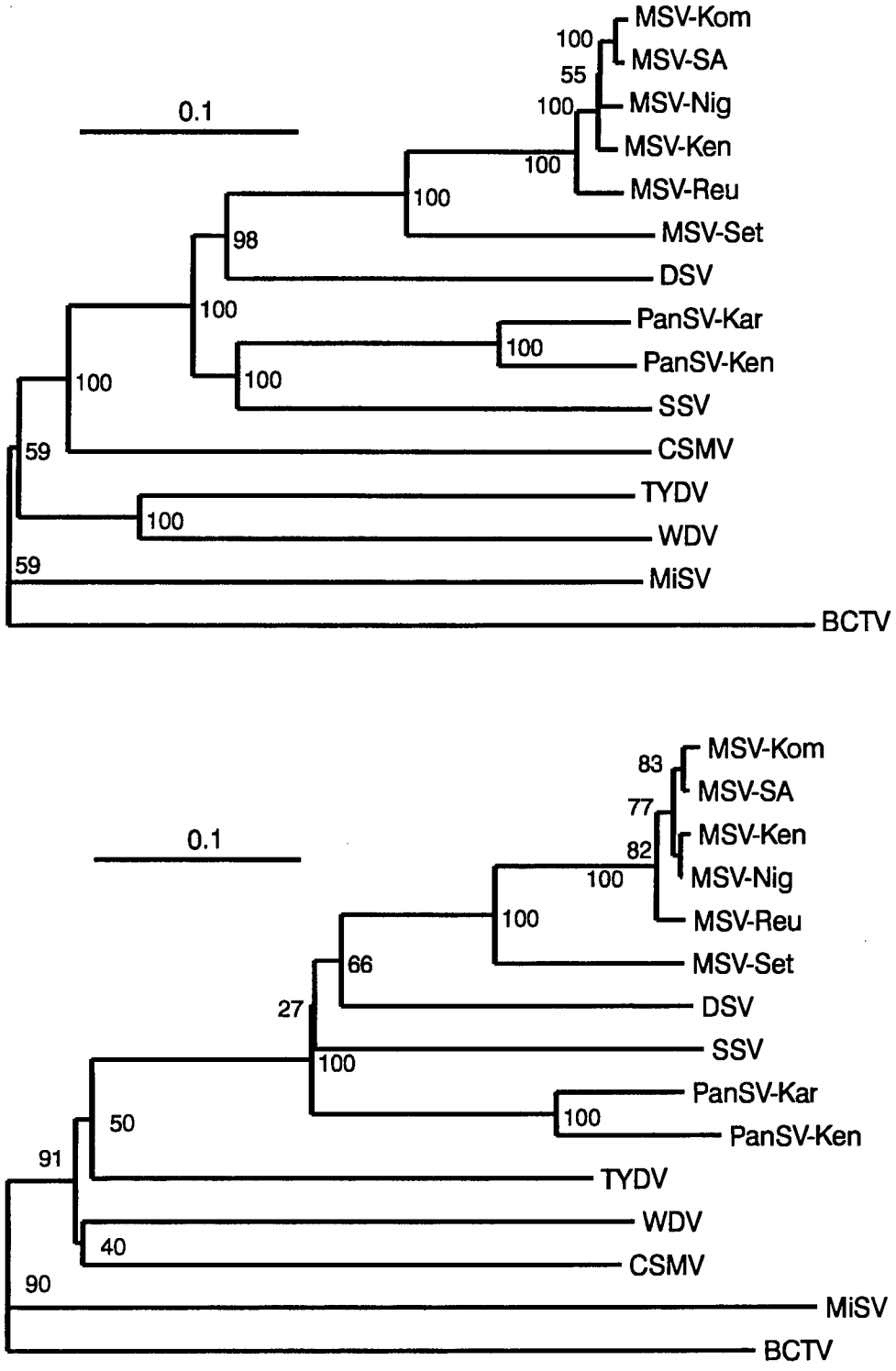

Figure 3.13 Phylogenies generated from entire nucleotide and Rep protein sequences

Calculations of pairwise percent sequence distance values from multiple sequence alignments were done using CLUSTALW for the Rep sequence (bottom) and DNAMAN for the nucleotide sequences (top). Neighbour-joining trees were generated using the CLUSTALW package. Top and bottom; nucleotide and Rep sequence phylogenies respectively. Horizontal lines indicate proportional genetic distance and vertical lines are arbitrary. Line bar indicates a 0.1 genetic distance value (i.e. 10% difference). Trees were rooted with BCTV sequence and bootstrap values are a percentage value of 100 bootstrap trials for the nucleotide sequence (above) and of 1000 bootstrap trials for the Rep protein sequence (below).

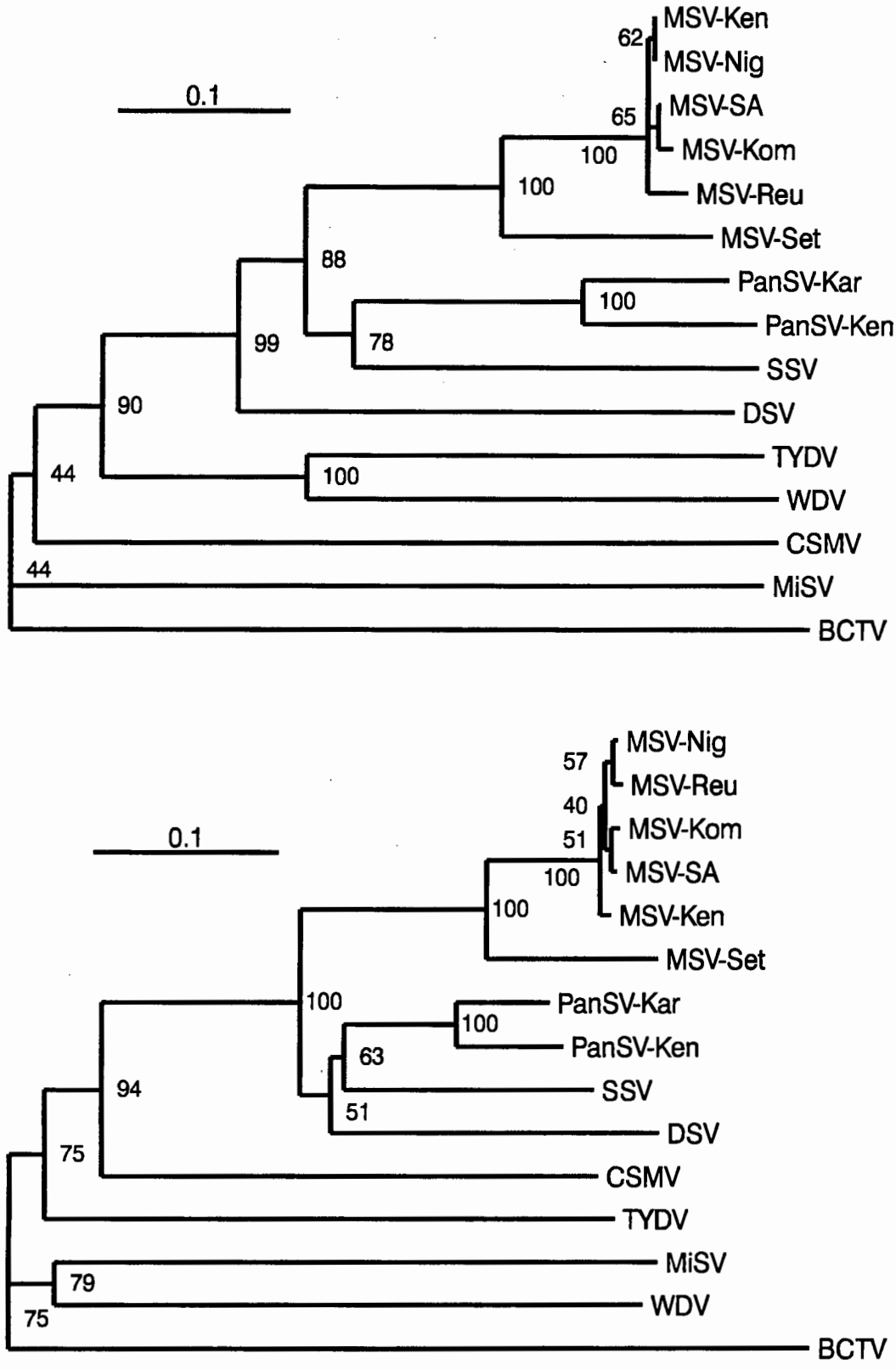

Figure 3.14 Phylogenies of MP and CP protein sequences

Calculation of pairwise percent sequence distance values from multiple sequence alignments was done using CLUSTALW. Neighbour-joining trees were generated using CLUSTALW. Top and bottom; MP and CP sequence phylogenies respectively. Horizontal lines indicate proportional genetic distance and vertical lines are arbitrary. Line bar indicates a 0.1 (10%) genetic distance value (i.e.. 10% difference) Trees were rooted using BCTV sequence and bootstrap values are a percentage value of 1000 bootstrap trials.

3.5 DISCUSSION

3.5.1 Sequence analyses

3.5.1.1 *Non-coding regions*

The sequences of MSV-Kom, MSV-Set and PanSV-Kar have revealed genomic characteristics typical of other Mastreviruses (Chapter 2). The absolutely conserved nonanucleotide TAATATTAC found in all geminiviruses is present in MSV-Kom, MSV-Set and PanSV-Kar. The conserved nonanucleotides are flanked by inverted repeats of 18 bases for MSV-Kom and MSV-Set (as for MSVs isolated from maize) and 15 bases for PanSV-Kar (as in PanSV-Ken) with free energy values (ΔG) of -28.70, -31.10 and -24.70 kcal/mol respectively. These inverted repeats are capable of forming stem-loop or hairpin structures and are probably the structures where replication is initiated by their respective Rep proteins. Dicot-infecting geminiviruses have highly conserved inverted repeat sequences forming the stem of the hairpin loops structures whereas those of the Mastreviruses are not conserved (Argüello-Astorga *et al.*, 1995). The differences of these inverted repeats between PanSV-Kar and MSV-Kom or MSV-Set are in line with the observation that they are generally not conserved in Mastreviruses suggesting that these sequences could include specific Rep binding sites and impose specificity on each virus (Argüello-Astorga *et al.*, 1995). This may be an indication that these differences help determine their different biological properties as determined in Chapter 2.

Other inverted repeats with ΔG values of below -10.00 in the nt sequences of MSV-Kom and MSV-Set are conserved. MSV-Kom has three such inverted repeats at nt positions 1397-1405/1412-1420, 2159-2169/2175-2188, and 2641-2648/2655-2662 (-13.00, -11.40 and -11.90 kcal/mol respectively) and those of the MSV-Set occur in nt positions 1395-1403/1411-1419, 2157-2166/2174-2183 and 2652-2659/2666-2673 (-13.00, -10.00 and -11.90 kcal/mol respectively). These inverted repeats seem to be conserved in other Mastreviruses isolated from maize such as MSV-SA, however they do not appear to be conserved in PanSV-Kar or PanSV-Ken (Briddon *et al.*, 1992). Only one of the three inverted repeats is situated in the non-coding region of the MSV-Kom and MSV-Set genomes which presumably is involved with regulation of transcription and/or replication of these viruses (Mullineaux *et al.*, 1984). These are the GC rich rightward promoter elements (*rpe*) in the UAS (Figure 3.11), presumably associated with virus-specific biological charasteristics such as the determinants of host range that bind host-specific host-factors (Hughes *et al.*, 1993; Lazarowitz, 1992). The UAS sequence of MSV-Set is identical to that of MSV-Kom and other Mastreviruses isolated from maize such as MSV-SA. The UAS

sequences may be an influencing factor of the Mastreviral host specificity as speculated by Lazarowitz, (1988) and Hughes *et al.*, (1993). If this is the case it is possible that the identical UASs of MSV-Kom and MSV-Set influence their relatively similar host ranges and moderate to severe streak symptoms. Further, the UAS sequence of PanSV-Kar resembles that of PanSV-Ken.

Iterative elements were identified in the LIRs of MSV-Kom, MSV-Set and PanSV-Kar and found to be dissimilar to each other. This is in line with the observations of Mastrevirus LIRs by Argüello-Astorga *et al.* (1995), where they identified different iterons in all the Mastrevirus (MSV, DSV, SSV, PanSV-Ken, WDV, TYDV, MiSV and CSMV) sequences they studied. The iteron sequences and positions are similar in dicot-infecting geminiviruses whereas in Mastreviruses they were found to be dissimilar in sequence, size and position. Mastrevirus iterons are 7-9 base GC-rich nt sequences situated as inverted repeats in the stem of the hairpin loop and another situated approximately 70 bases to the left of the conserved TAATATTAC nonanucleotide. MSV-Kom iterons were 8 bases, MSV-Set were 6 bases and for PanSV-Kar were 6 bases in length (Figure 3.11). Mastreviruses such as MSV-Nig and WDV, as well as MSV-Kom and MSV-Set in this study, have completely duplicated iterons whereas in most cases there is a single complete iteron in the left half of the stem and a partial duplication on the right replication (Argüello-Astorga *et al.*, 1995) as was the case for PanSV-Kar in this study (Figure 3.11). If the Rep protein can act as a repressor of its own synthesis as shown by Haley *et al.* (1992) and Sunter *et al.* (1993) in dicot-infecting geminiviruses, it may be that this function relies on binding elements located in different positions to those involved in replication (Argüello-Astorga *et al.*, 1995). This may be another virus specificity or host determining factor.

Other non-coding genome features in the genomes of MSV-Kom, MSV-Set and PanSV-Kar are the dT.dA tracts associated with origins of DNA replication (Suarez-Lopez *et al.*, 1995) and the potential DNA binding sites situated in the SIRs of Mastreviruses (Donson *et al.*, 1987). The arrangement and number of dT.dA tracts of PanSV-Kar resembled that of PanSV-Ken (Briddon *et al.*,1992) and the dT.dA tracts of MSV-Set resembled that of MSV-Kom along with other MSV strains such as MSV-SA (Lazarowitz., 1988). Primer binding sites are another distinguishing feature of Mastreviruses and the potential PBSs for MSV-Kom, MSV-Set and PanSV-Kar have been identified. The primers are evidently specific for these viruses.

3.5.1.2 *Coding regions*

The comparisons of the sequences of the MP, CP and Rep proteins of MSV-Kom, MSV-Set and PanSV-Kar were congruent with other Mastreviruses such as that of MSV-SA sequence determined by Lazarowitz (1988). Hydrophobicity profiles of the proteins of MSV-Kom, MSV-Set and PanSV-Kar were similar to those of MSV-SA (Figures 3.5, 3.6 and 3.7). The MP amino acid sequences for MSV-Kom, MSV-Set and PanSV-Kar share identities of 99, 75 and 46% respectively with MSV-SA. However their CPs share identities of 99, 84, amd 69% and their Rep proteins share higher identities of 99, 82 and 65 % respectively with MSV-SA. Thus the CP and Rep protein sequences are more conserved than the MPs in Mastreviruses with the Rep protein being the most conserved (Hughes *et al.*, 1993). The range of sequence identity for the CPs suggests a multifunctional role, although one might expect the African Mastreviruses to have higher levels of identity while they share the same vectors. For leafhopper-transmitted geminiviruses the CP has a number of proposed functions including the protection of the genome (possibly involving DNA binding), specificity for virus transmission (Briddon *et al.*, 1990) and movement within plants (Briddon *et al.*, 1989, 1990; Lazarowitz, 1989). The MP is the least conserved protein which interacts with host cell factors for cell-to-cell movement. This is an indication that the MP may be one of the more important determining factors influencing the host range of the Mastreviruses. The differences in the identities of the MPs of MSV-Kom, MSV-Set and PanSV-Kar as determined in this study is presumably a reflection of their different host ranges (Chapter 2). This does not rule out the possibility that the CP and Rep protein are also involved (possibly in concert) with the MP conferring specificity and/or the host ranges to these geminiviruses.

The MP and CP are considered late gene products required for systemic infection (Boulton *et al.*, 1989a; 1993; Chatani *et al.*, 1991; Lazarowitz, *et al.*, 1989; Mullineaux *et al.*, 1988; Woolston *et al.*, 1989). The hydrophobic region found in the sequences of MP have been identified in MSV-Kom, MSV-Set and PanSV-Kar (Figure 3.5). These hydrophobic regions presumably interact with host cell membranes which would be associated with virus cell-to-cell movement. The MP AT-rich intron (such as those identified by Wright *et al.*, [1997] in MSV-Nig and WDV) contains the hydrophobic region, thus complete MPs would not be translated from spliced transcripts (Wright, *et al.*, 1997). The MP intron have also been identified in MSV-Kom, MSV-Set and PanSV-Kar. It was suggested that this intron may be part of a gene regulatory mechanism to prevent expression of the MP throughout the late gene expression period (Wright *et al.*, 1997). It is possible that the differences of these introns (such as the different consensus YTNAN sequences) is a factor that may confer specific host ranges and/or symptom severity to each virus. If this is the case, this efficiency may dependent on the efficacy of intron splicing by the plant host.

Another Mastrevirus feature is the intron located in the C1/C2 ORFs which has been identified in MSV-Kom, MSV-Set and PanSV-Kar. The intron of MSV-Set has a lariat sequence unusual in Mastreviruses but which is, however, identical to that of SSV (Hughes *et al.*, 1993). How the infectivity of MSV-Set as determined in Chapter 2 is affected by this unusual lariat sequence CTGAT (as opposed to the usual CTGAC) is not evident and appears that this lariat sequence does not alter the function of the intron. The three RCR motifs, the Rb/LXCXE motif and the NTP-binding motifs on the Rep proteins of MSV-Kom, MSV-Set and PanSV-Kar have also been identified and found to have no striking differences to those identified in other Mastreviruses.

3.5.2 Phylogenetic analyses

The determination of the sequence of MSV-Set has elucidated the taxonomic ranking of this virus as a member of the "African streak group" of viruses. Alignments of the MP, CP and Rep amino acid sequences, as well as of the complete nucleotide sequences of MSV-Kom, MSV-Set and PanSV-Kar along with 11 other Mastrevirus sequences, grouped MSV-Set with the MSV group of geminiviruses isolated from maize, and PanSV-Kar grouped with PanSV-Ken isolated from *P. maximum* (Figures 3.13 and 14). Although MSV-Set was grouped with the MSVs, it is nevertheless distinct from the the MSVs isolated from maize. These results are in line with the nucleotide sequence identities of 78% between MSV-Set and MSV-Kom and 89% between PanSV-Kar and PanSV-Ken. These results are reflected in that the symptoms and host ranges of MSV-Set and MSV-Kom are more alike, and those of PanSV-Kar and PanSV-Ken (Briddon *et al.*, 1992) are similar (Chapter 2). MSV-Set is however the most distinct MSV strain yet sequenced, including partial sequences of a number of other wheat and grass strains (E.P. Rybicki, pers. comm.).

3.6 CONCLUSIONS

Since crops such as maize and sugarcane were introduced into Africa, and are affected by geminiviruses only in Africa, the progenitors of these viruses are likely to be viruses infecting African indigenous grasses, such as *P. maximum* which is widely distributed throughout Africa as natural vegetation and is also cultivated for grazing (Briddon *et al.*, 1992). If this is indeed the case PanSV-Kar would fall into this progenitor category. MSV-Set however will be a more likely progenitor or closer relative of the MSVs since it was isolated from an indigenous African *Setaria* species (Hughes *et al.*, 1992) and has more MSV-like genomic and biological properties (Chapters 2 and 3). Briddon *et al.* (1992) observed that the numerous chlorotic lesions of PanSV-Ken appeared to have no effect on these grasses.

Presumably, PanSV-Kar also appeared to have no adverse effects on its original *P. maximum* host since the symptoms appearing on the original host were described as "very mild streak" (Hughes, *et al.*, 1992). In the case of MSV-Set, the description of the streak appearing on the original *Setaria* host was also described as "very mild streak", thus it presumably also has no adverse effects on its natural African grass hosts. The attenuated symptoms caused by PanSV and MSV-Set on their original hosts can be explained by host/pathogen co-evolution.

The cloned genomes of MSV-Kom, MSV-Set and PanSV-Kar and their agroinfectious *Agrobacterium* clones are a useful tool in investigating the biological and genomic diversity of African Mastreviruses. The phylogenetic analyses of their sequences have clarified their taxonomic rankings with those of the other Mastreviruses. Previously called SetSV, MSV-Set partial sequence, hybridisation and mapping data, correlated best with MSV group data than with other Mastreviruses and thus the strain was considered "a grass strain" of MSV (Hughes, *et al.*, 1992).

The available sequences of three distinct Mastreviruses isolated from different hosts and from different geographical regions offers an attractive opportunity to study African Mastreviruses in gene replacement or recombinational experiments. This is especially the case between the MSV-Set and MSV-Kom viruses since they are distinct, although grouped in the same cluster. Whether they are sufficiently related to conduct recombinational or sequence replacement studies between them is investigated in Chapters 4 and 5.

CHAPTER 4

AGROINOCULATIONS OF HETERODIMERIC CONSTRUCTS OF MSV-Set AND MSV-Kom

---

4.1 SUMMARY

MSV-Kom isolated from maize in Komatipoort, Mpumalanga and MSV-Set isolated from a *Setaria* species in Mt. Edgecombe,Kwazulu/Natal share a nucleotide sequence identity of 78% and have different pathogenicities and overlapping, but non-identical host ranges (Chapters 2 and 3). Despite their genomic and biological differences, six of eight different partially heterodimeric constructs ranging from full-sized dimers to 1.1mers were shown to be agroinfectious. The infectious constructs contained at least two LIRs flanking a set of ORFs, however the non-infectious constructs contained only one LIR. All infectious heterodimer clones produced streak symptoms approximately 5 days later than the wild-type counterparts pKom603 and pSet107 (homologous 1.1mers containing two LIRs) suggesting that parental recombinants were released by random homologous recombination as opposed to replicational release. Partial sequence analyses on 43 resultant recombinants isolated from agroinoculated sweetcorn plants revealed 14 types of recombinants. 40 recombinant genomes consisted mainly of wild-type sequence but with sequence replacements of 2-176bp to the left of the *Bam* HI junction (in the LIR) between the two genomes of the construct, or 25bp or 176bp replacements to the right of that junction (in the V1 ORF). Although this method generated recombinants that were genomically wild-type-like, this method can be potentially used to generate viable Mastreviral recombinants of relatively distinct viruses. Such recombinants can be used in agroinoculation tests to determine the functions of focused or small regions of the genome that may be responsible for biological characters such as host range and/or pathogenicity. The implications of the sequence analyses are discussed.

4.2 INTRODUCTION AND RATIONALE

The introduction of head-to-tail dimers of geminivirus genomes into the host cell leads to the release of replicating unit-length virus molecules in infected nuclei. Cloned geminivirus DNAs are infectious as tandemly repeated copies within a plasmid DNA and may be delivered to plants by mechanical inoculation including biolistic particle delivery (Stenger *et al.*, 1990; Gilbertson *et al.*, 1991) or transfection of protoplasts (Matzeit *et al.*, 1991). Infection can also be achieved using the *Agrobacterium*-mediated methods; either by the construction of entire transgenic plants using *A.tumefaciens* pTi vectors in leaf disc transformations (Rogers *et al.*, 1986; Elmer *et al.*, 1988) or by agroinoculation of young

plantlets with similar *Agrobacterium* strains (Grimsley *et al.*, 1987). In all cases the release of unit-length geminivirus genomes from dimeric constructs leads to systemic infection.

Two models, which are not mutually exclusive, have been proposed for the release of unit-length geminivirus RF-DNAs from head-to-tail genome repeats (Figure 1.5; Elmer *et al.*, 1988): intramolecular homologous recombination resulting from a single crossover event at random locations within the tandem viral genome repeats; or replicational release by displacement DNA synthesis from one origin of replication to the next. Direct evidence for the former mechanism has been reported for MSV (Lazarowitz *et al.*, 1989). The latter process functionally resembles rolling circle replication (RCR) from native circular dsDNA form, where a linear ssDNA is generated that is subsequently cleaved from the replication intermediate and ligated to form unit-length circular ssDNA (Stenger *et al.*, 1991). The Rep proteins of TGMV and SLCV have been shown to interact specifically with their respective intergenic regions (Fontes *et al.*, 1992; Lazarowitz *et al.*, 1992). Rep initiates RCR by binding the viral strand origin of replication located in the LIRs or common regions, recognises the specific nonanucleotide TAATATT↓AC (situated within the hairpin loop), and introduces a nick (↓) thereby allowing DNA synthesis from the 3'-OH terminus (Heyraud-Nitschke *et al.*, 1995). Elongation proceeds while the parental viral-strand is displaced and termination occurs at the same site within the conserved nonanucleotide TAATATT↓AC.

It was not possible to determine whether the release of replicating virus molecules resulted from intramolecular homologous recombination or from replicational release when using homodimeric clones. However sequence analyses of progeny resulting from the agroinoculation or mechanical inoculation of tandem genome repeats derived from different strains of geminiviruses - if the strains were sufficiently closely related - provided evidence that the replicative release mechanism is favoured over homologous recombination (Stenger *et al.*, 1991; Heyraud *et al.*, 1993). Recombinant plasmids containing tandem copies of different CP replacement genomes of WDV were used to study the mechanism leading to the release of replicating unit-length molecules (Heyraud *et al.*, 1993). For plasmids containing two complete genomes, the viral unit bracketed by the two LIRs was preferentially released (Heyraud *et al.*, 1993). Sequence analyses of the intergenic regions of BCTV recombinants from the Worland and Logan strains allowed the mapping of the origin of replication to a 20 bp sequence within the hairpin situated in the intergenic region (Stenger *et al.*, 1991). Further, sequence analyses of unit-length progeny released from plasmids containing a single WDV genome flanked by two different LIRs - of the Swedish isolate and the barley-adapted strain - showed that the conserved nonanucleotide TAATATTAC is the region of the LIR within which the release of unit-length molecules occurs (Heyraud *et al.*,

1993). Despite the fact that the LIRs of the Swedish and barley-adapted WDV strains only share 75% sequence identity, partial heterodimeric 1.1mers containing two different LIRs flanking the rest of the genome were able to replicate in wheat protoplasts (Heyraud *et al.*, 1993).

Intermolecular recombination between extrachromosomal geminivirus DNAs has frequently been observed in complementation experiments with various combinations of mutants (Brough *et al.*, 1988; Etessami *et al.*, 1988; Lazarowitz *et al.*, 1989 Hormuzdi and Bisaro, 1993). These observations of intermolecular recombination occurred between DNAs which shared sequence identities of almost 100%. However sequence evidence strongly suggests that several geminiviruses have resulted from ancient or more recent recombinational events occurring between two or more different geminivirus progenitors with different sequences (Chapter 1). Geminiviruses can be seen as a collection of viruses with similar genome structure and replication mechanisms, thus it is not surprising that recombination is playing a role in mix of sequence conservation and divergence among the *Geminiviridae* (Bisaro, 1994).

MSV-Set and MSV-Kom are two biologically and genomically distinct Mastreviruses however they have been phylogenetically grouped with the MSV group of Mastreviruses isolated from maize (Chapters 2 and 3). They share nucleotide sequence identities of 78%, in line with their differences in biological features. It is possible that they are able to recombine to form new strains of viruses in mixed infections in nature. In this study however, "forced" recombination by agroinoculation of heterodimers, which would necessitate recombination between the different genomes for the release of infectious unit-length genomes was tested. Intramolecular homologous recombination resulting from a single crossover event could occur at random locations within the tandem viral genome repeats. Consequently, a number of recombinant genome progeny genotypes would be expected if the parental repeated genomes were derived from distinguishable strains such as MSV-Kom and MSV-Set.

The agroinoculation of heterodimeric constructs of the genomes of MSV-Set and MSV-Kom is reported in this Chapter. Sequence analyses of resultant MSV-Kom/MSV-Set recombinant viral genomes isolated from agroinoculated sweetcorn plants, and the determination of recombinational crossover sites, are reported. The rationale for this investigation was twofold: first; to shed some light on the mechanism of release of the genome of MSV during agroinfection, which would have implications in resolving the mechanism of MSV DNA replication in the host; and second, to see whether selected recombinant viruses could be used in leafhopper transmissions and agroinfectivity

tests to elucidate the genomic domains responsible for host range and symptom severity of MSV respectively. This was further investigated in Chapter 5.

4.3 MATERIALS AND METHODS

4.3.1 Construction of heterodimeric clones for agroinfection

Full and partial "heterodimeric" or chimaeric clones were constructed in a similar fashion to constructing homodimeric clones (2.2.2 and 2.2.3). Since MSV-Set and MSV-Kom RF-DNAs share the same unique *Bam* HI restriction site and were both cloned into the *Bam* HI site of the MCS of the pUC18 vector (Figure 2.1), dimers were constructed using the *Bam* HI site. Complete MSV-Set/MSV-Kom and MSV-Kom/MSV-Set dimers were constructed in pUC18 and designated pSeK100 and pKoS100 respectively. *Bam* HI-partially-digested pSet100 was ligated with *Bam* HI/*Eco* RI-double-digested pKom500 and *Bam* HI-partially-digested pKom500 was ligated with *Bam* HI-*Pst* I-double-digested pSet100 (Apendices A1-A4). *Eco* RI and *Pst* I sites are not present in the RF-DNA of MSV-Kom and MSV-Set respectively (Figure 2.1). *Eco* RI and *Pst* I sites, however, are present in the MCS of pUC18 and thus *Eco* RI- or *Pst* I-digested pKom500 or pSet100 partially digested with *Bam* HI will render linear pUC18 molecules less likely to interfere in these ligations. *E.coli* colonies were selected on the basis of their plasmid minipreparations yielding a 2.7kb fragment in *Apa* I digestions, since there are single *Apa* I sites in the RF-DNA of both MSV-Kom and MSV-Set (Appendices A1, A3, A5 and A6). The selected plasmids were further digested with *Eco* RI, *Xho* I and *Eco* RI/*Xho* I double digests to distinguish pSeK100 and pKoS100 plasmids since *Xho* I and *Eco* RI sites are not present in MSV-Set and MSV-Kom RF-DNA respectively. Maxipreparations of pSeK100 and pKoS100 were prepared as decribed in Appendix A11.

Xba I/*Pvu* I-digested pSeK100 and pKoS100 were partially digested with *Eco* RI and ligated with *Xba* I/*Eco* RI/*Sac* I-digested pBI121 to replace the GUS gene and NOS-terminator in pBI121 with the MSV-Set/MSV-Kom or MSV-Kom/MSV-Set RF-DNA dimers respectively (Appendices A1, A3, A5 and A6). Since there are two Pvu I sites in pUC18 and a *Sac* I site in the GUS/NOS terminator fragment, digesting pBI121 with *Sac* I in addition to *Xba* I and *Eco* RI, and digesting pSeK100 and pKoS100 with *Pvu* I in addition to *Xba* I and *Eco* RI (partially), renders the GUS/NOS terminator DNA fragment and pUC18 less likely to interfere in "shotgun" ligations. *E. coli* colonies harbouring pBI121 with inserted MSV-Set or MSV-Kom DNA were selected in colony blots and hybridisations using a DIG-labelled MSV-Set and/or MSV-Kom DNA probe (Appendices A7-A10). Plasmid minipreparations from selected colonies were subjected to *Apa* I, *Xba* I and *Eco* RI digestions including *Xba* I/*Eco* RI double-

digests to distinguish partial and complete heterodimers cloned into pBI121 (Appendices A1-A3). Maxipreparations were prepared (Appendix A11) and the partial and complete MSV-Set/MSV-Kom heterodimers in pBI121 were designated pSeK102 and pSeK101 respectively. An additional partial MSV-Kom/MSV-Set heterodimer was cloned into pBI121 by ligating *Bgl* II/*Pvu* I-digested pKoS100 with *Bam* HI-digested pBI121 (Appendices A1-A11). The resulting partial dimer was designated pKoS103. Heterodimeric constructs are depicted in Figure 4.2. Agroinoculations and subsequent leafhopper transmissions were performed as described in 2.2.4 and 2.2.5 and 2.2.6.

4.3.2 Construction of clones with two heterologous LIRs

Bam HI-digested pKom601 was ligated with *Bam* HI/*Pst* I-digested pSet100 and *Bam* HI/*Eco* RI-digested pKom500 was ligated with *Bam* HI-digested pSet100.10 to produce pKoS104 and pSeK 103; these were both hetero-1.1mers constructed in pUC18 (Appendices A1, A3-A6). The *Pst* I and *Eco* RI digestions reduced the probability of linear pUC18 interfering with these "shotgun" ligations. *E.coli* colonies on ampicillin/IPTG/X-gal plates were selected on the basis of *Apa* I-, *Sac* I- and *Bam* HI-plasmid minipreparation digestions (Appendices A1, A3 and A6). To confirm that the LIRs flanking the ORFs were from different viruses, both pKoS104 and pSeK103 were sequenced using the standard M13 reverse and forward primers since restriction endonuclease sites that would distinguish the desired clones were lacking in the LIRs (Appendix B). The hetero-1.1mers were cloned into pBI121 using the *Xba* I and *Eco* RI sites of *Xba* I/*Sac* I/*Eco* RI digested vector DNA in a similar fashion to that described for the construction of the homo-1.1mers in Chapter 2 (2.2.3; Appendices A1 and A3-A11). These clones were designated pKoS105 and pSeK104 respectively after selection from LA + kanamycin plates on the basis of colony blots using DIG-labelled MSV-Kom or MSV-Set DNA DNA in hybridisations and further *Sac* I, *Apa* I and *Bam* HI-plasmid minipreparation digestions as described in 4.2.1 above.

In addition to these hetero-1.1mers, pSet105 - a homo-1.1mer clone of MSV-Set in pUC18 (2.2.3) - was digested with *Sac* I (a unique restriction site in the MSV-Set RF-DNA) and *Pst* I (for lower interference probability) and ligated with with *Sac* I-digested pKom601 (2.2.3) to obtain pSeK106, an entire MSV-Set RF-DNA from the *Sac* I site attached to an additional MSV-Kom LIR attached to the *Sac* I site on the left of the MSV-Kom LIR up to the *Bam* HI site on the right of the stem-loop sequence (Appendices A1 and A3-A11; Figure 4.2). Hetero-1.1mers depicted in Figure 4.3 were subjected to agroinoculation and subsequent leafhopper transmissions as described in 2.2.4, 2.2.5 and 2.2.6.

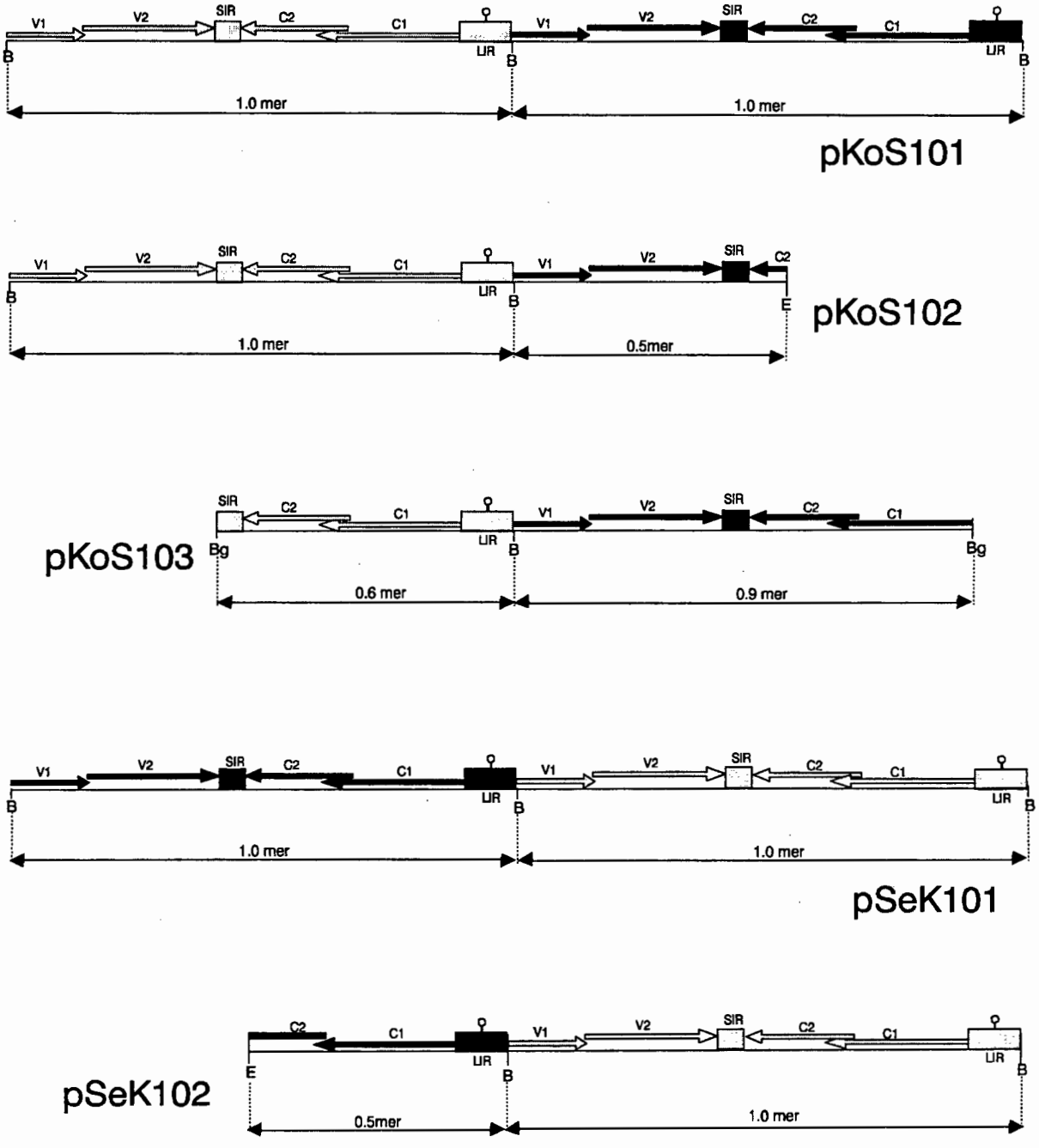

Figure 4.2 MSV-Kom/MSV-Set and MSV-Set/MSV-Kom partial and complete heterodimers

Heterodimers were constructed in pBI121. V1 = MP ORF; V2 = CP ORF; C1+C2 = Rep ORF; LIR = large intergenic region with hairpin loop; SIR = small intergenic region; B = *Bam* HI; E = *Eco* RI; Bg = *Bgl* II. Shaded and black parts represent MSV-Kom and MSV-Set sequences respectively.

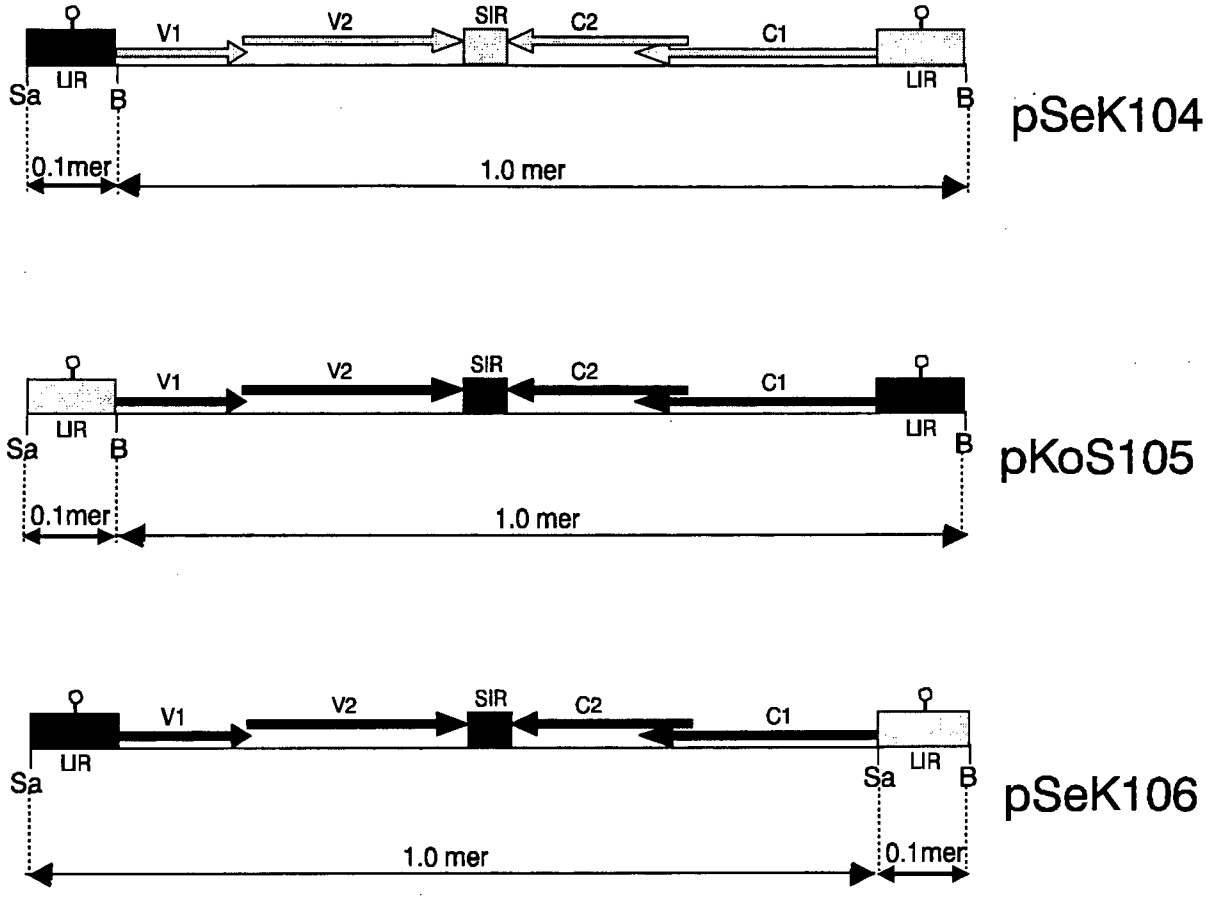

Figure 4.3 MSV-Kom/MSV-Set heterodimeric 1.1mers constructed in pBI121

V1 = MP ORF; V2 = CP ORF; C1+C2 = Rep ORF; LIR = large intergenic region with hairpin loop; SIR = small intergenic region; B = *Bam* HI; Sa = *Sac* I. Shaded and black parts represent MSV-Kom and MSV-Set sequences respectively.

4.3.3 Detection of the presence of recombinants in heterodimer-infected plants

Total DNA extractions and RF-DNA isolations, as described in Appendices A12 and A14, were performed on individual agroinoculated sweetcorn plants showing symptoms. RF-DNA was also subjected to a series of restriction endonuclease digestions and double digestions to ascertain whether recombinants were present, by analysing the maps of the RF-DNAs and comparing them to maps of MSV-Set and MSV-Kom as depicted in Figure 2.1. 1μg of each RF-DNA sample was digested with *Rsa* I, a four base cutter, or *Sac* I, a six base cutter, to identify MSV-Kom DNA from MSV-Set DNA on agarose gels and in Southern blots (Appendices A1, A3, A13). This was done to ascertain whether recombinant viral RF-DNAs were present in small numbers in heterodimer-agroinoculated plants, since Southern blots are more sensitive for detecting specific DNA sequences. DIG-labelled MSV-Kom and MSV-Set DNA were used in two separate hybridisations on the same blots, with these blots being stripped for each repeat hybridisation (Appendix A9).

4.3.4 Cloning and sequencing of recombinant RF-DNAs

RF-DNA DNA was isolated from infected plants as described above. In some instances RF-DNA was pooled, but only if the DNA came from plants that were agroinoculated with the same construct (Appendices A12 and A14). *Bam* HI-digested RF-DNA was ligated with *Bam* HI-digested pUC18 to obtain RF-DNAs cloned into the *Bam* HI site of the MCS of the vector (Appendices A1, A3-A6). White colonies from LA + ampicillin/X-gal/IPTG plates were selected for minipreparation of plasmid DNA on the basis of colony blots and hybridisations done using DIG-labelled MSV-Set or MSV-Kom DNA. Plasmid minipreparations were selected on the basis of fragmentation patterns obtained in *Bam* HI and *Sal* I, *Eco* RI, *Xho* I or *Apa* I digestions (Appendix A1). Maxipreparations of cloned recombinant RF-DNAs were made (Appendix A11) and subjected to reverse and/or forward sequencing reactions (Appendices B and C). Recombinant clones obtained from the agroinfection of pKoS101 were designated pKoS001, pKoS002, pKoS003...pKoS010; products of pSeK101 agroinoculations designated pSeK111, pSeK112, pSeK113...pSeK119, pSeK1110 and pseK1111; from pSeK102 agroinfections were designated pSeK001, pSeK002, pSeK003...pSeK011; products of pSeK104 agroinoculations were designated pSeK141, pSeK142...pSeK145; products of pSeK106 agroinoculations designated pSeK161, pSeK162 and pSeK163, and products of pKoS105 agroinfections were designated pKoS151, pKoS152 and pKoS153. Sequences were compared and aligned with the complete sequences of MSV-Kom and MSV-Set to identify sites of recombination between the two viral genomes, using GAP and FASTA options in GCG Version 7.1 on a VAX/VMS mainframe.

The primers presented in Table 4.1 were used in sequencing reactions on the recombinant clone pKoS005, which had a 180 base region of the V1 region switched, to ascertain whether this was the only recombinational event that occurred. These primers anneal to various regions of the MSV-Kom RF-DNA. These primers were use in sequencing reactions as described in Appendix B.

Table 4.1 Primers used to sequence a recombinant virus, pKoS005

Primer	Size (bases)	Sequence (5'-3')	Strand	Genome location*
C2KEP	25mer	ACGCAAACAATACAGGGGGGTAGTC	+	1283-1307
LIR-Kom	26mer	**ctatctaga**CGACGACGGAGGTTGAG	+	2532-2548
C1KEP	25mer	**ttagg**ATCCCTCAGCCTCAACCTCC	-	2538-2557
17KR	17mer	GGAAANNCTNCNTGGGC	-	1720-1736
SIR-Kom	24mer	**ccgctcga**gTAATTCATATAGATC	-	1356-1371

KEP obtained from Kenneth Palmer; Kom obtained from Janet Willment; 17KR is a degenerate primer obtained from Prof. E.P. Rybicki; Department of Microbiology, UCT. N = degenerate positions of the degenerate primer used. Lower case sequence in bold are sequences that are not homologous to MSV-Kom since these primers are overhanging primers. * = MSV-Kom genome location.

4.4 RESULTS

4.4.1 Agroinfectivities of heterodimeric constructs

The agroinfectivities of the partially homodimeric constructs with one LIR (pKom504 and pSet102) were totaled at 56% and 35% respectively in sweetcorn agroinoculations. The partial heterodimer pSeK102 and the complete heterodimers pSeK101 and pKoS101 had total agroinfectivity rates of 56%, 30% and 13% respectively. In individual popcorn and sweetcorn agroinoculations, however, pKoS101 agroinfectivity rates were as high as 20% and 31% respectively - thus the 13% agroinfection rate is presumably not a true reflection. Interestingly, the partial heterodimers pKoS102 and pKoS103 - which only have single LIRs - were not agroinfectious in popcorn or in two sweetcorn agroinoculation experiments. Agroinfectivity profiles of pSeK102, pSeK101, and pKoS101 are depicted in Figures 4.3 and Appendix D Figure D2. Streak symptoms appear 3-5 days later than those for the homodimeric (wild-type) clones.

Since homodimeric wild-type clones with two LIRs (pKom603 and pSet107) were significantly more agroinfectious than those with one LIR (Chapter 2), heterodimeric 1.1mer clones containing two different LIRs were tested in agroinoculations. The agroinfectivity rates of pSeK104, pSeK106 and and pKoS105 were determined and summarised in Table 4.2. In all heterodimeric agroinoculations including pSeK102 and pKoS101 (both with two LIRs), streak symptoms appeared 5-6 days later than those for the wild-type clones. In contrast to the wild-type agroinfectivity rates of 100% and 86% for pKom603 and pset107 respectively, those for pSeK104 and pKoS105 were 57% and 18%. Although streak symptoms in pSeK106 agroinoculated plants appeared 6 days later than the wild-type 1.1mer clones, the agroinfectivity rate was as high as 91%. In a repeat agroinoculation experiment the agroinfectivity rate for pSeK106 was 63% when the wild-type clones were 63% and 67% for pSet107 and pKom603 respectively. Thus pSeK106 has agroinfectivity rates similar to the wild-type counterparts although streak symptoms appeared 3-4 days later, along with the other partially heterodimeric clones. In addition the agroinfectivity rates of heterodimeric partial dimers with two LIRs (pSeK102 and pKoS101) are 49% and 9% were found to be significantly less than those of their smaller 1.1mer counterparts (pSeK104 and pKoS105 which were 57% and 18% respectively.

Table 4.1 Agroinfectivities of partial and complete heterodimeric constructs

Clone	Popcorn	Sweetcorn Experiment I	Sweetcorn Experiment II	Sweetcorn Total
pKom504	2/6 33%	8/11 73%	23/44 52%	31/55 56%
pSet102	4/10 40%	4/10 40%	15/45 33%	19/55 35%
pSeK101	3/15 20%	11/18 61%	7/43 16%	18/61 30%
pSeK102	0/9 0%	9/13 69%	26/44 59%	58/104 56%*
pKoS101	2/10 20%	4/13 31%	6/48 13%	14/107 13%*
pKoS102	0/13 0%	0/15 0%	0/50 0%	0/65 0%
pKoS103	0/17 0%	0/18 0%	0/50 0%	0/68 0%

Values indicate the number of seedlings showing streak symptoms by the number of seedlings injected which survived. Percentage values are also indicated. Experiments I and II are sweetcorn Jubilee seedling inoculations and the totals are indicated in the sweetcorn total column. Asterisks indicate the totals of three agroinfectivity tests which include the values of the sweetcorn agroinfectivity values in Table 4.2. Agroinfectivity profiles are depicted in Figures 4.3 and Appendix D, Figure D2 respectively.

Table 4.2 Agroinfectivities of heterodimeric 1.1mers with two LIRs

Clone	Experiment I	Experiment II
pKom603	19/19 100%	10/15 67%
pSet107	19/22 86%	12/19 63%
pSeK102*	23/47 49%	na
pSeK104	28/49 57%	na
pSeK106	20/22 91%	10/16 63%
pKoS101*	4/46 9%	na
pKoS105	9/49 18%	na

Values indicate the number of sweetcorn seedlings showing streak symptoms by the number of surviving seedlings. Percentage values are also indicated. The experiment on pSeK106 was repeated in a separate experiment along with the pKom603 and pSet107 agroinoculations as positive controls. Asterisks indicate that these values are pooled to obtain totals for three experiments in Table 4.1.

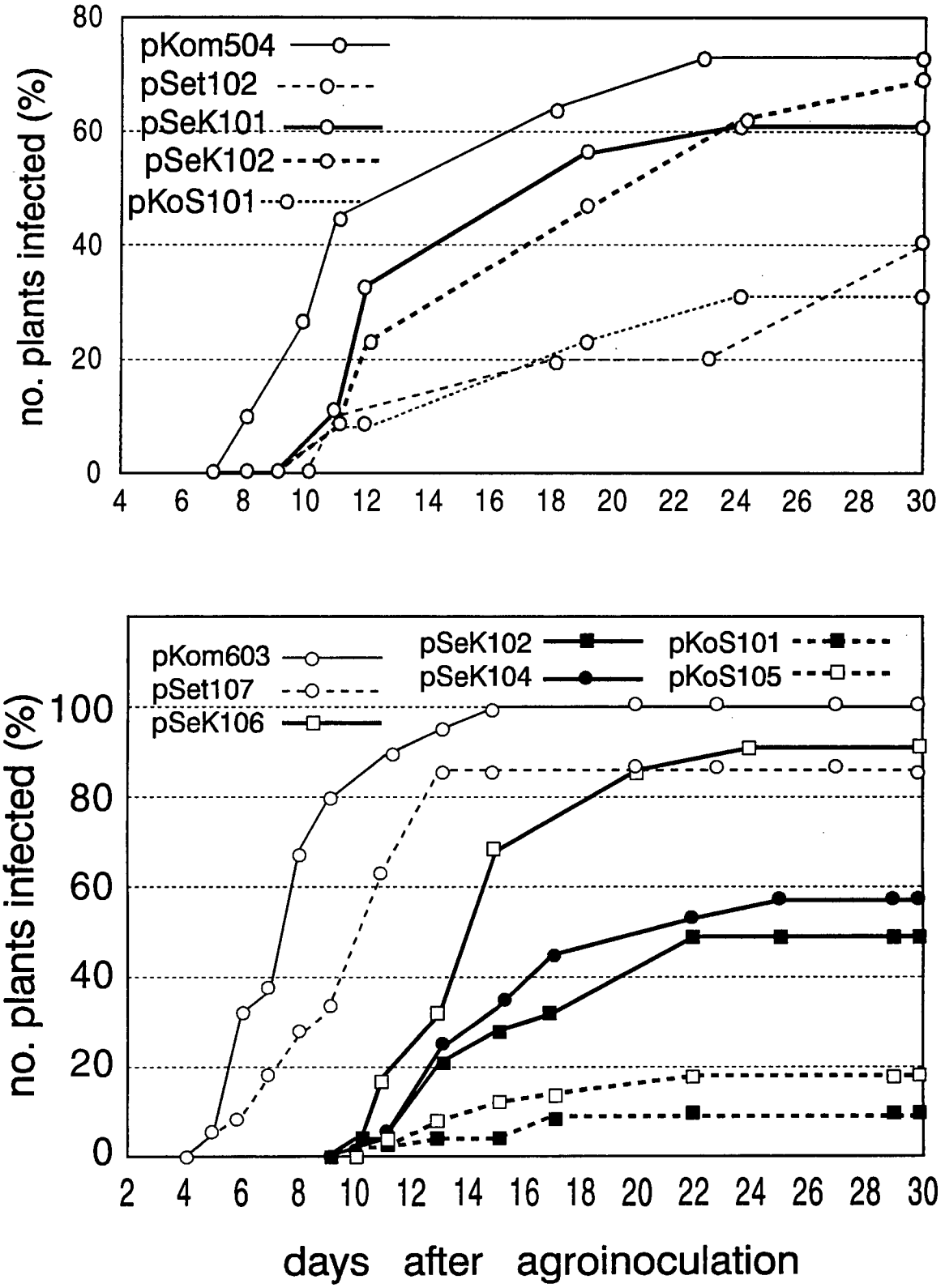

Figure 4.3 Agroinfectivities of heterodimeric constructs

Top diagram depicts the relative rates of agroinfection of the partial and complete heterodimeric constructs, pSeK101, pSeK102 and pKoS101; and those of the wild-types pSet102 and pKom504. Bottom diagram depicts the relative rates of agroinfection of the heterodimeric 1.1mers, pSeK104, pSeK106, pKoS105; and those of the wild-types, pSet107 and pKom603.

4.4.2 Symptoms and host ranges of heterodimeric constructs

pSeK102- and pKoS101-agroinoculated sweetcorn plants with streak symptoms were used in leafhopper transmission tests to determine the host ranges of the recombinant viruses in these plants. The results of these tests are summarised in Table 4.3. The symptoms of heterodimer-agroinoculated sweetcorn seedlings were initially milder than those of the wild-type-agroinoculated seedlings (Figure 4.4), presumably since the symptoms appeared 5 days later than those of the wild-types. The symptoms on agroinoculated plants later developed into MSV-Kom- and MSV-Set-like streak symptoms. The results for pKoS101 heterodimer leafhopper transmissions were similar to that of pSet102 (wild-type MSV-Set) and those for pSeK101 and pSeK102 were similar to that of pKom504 (MSV-Kom wild-type). For example this is well depicted in Figure 4.5 where the symptoms obtained from pSeK102-viruliferous leafhoppers were MSV-Kom-like with severe leaf curling and streak symptoms in Clipper barley and Wheat SST44 cultivars. Similarly, the symptoms obtained on Clipper barley from pKoS101-viruliferous leafhoppers resembled those of the wild-type MSV-Set (Figure 4.5). However severe symptoms were observed in maize cv.s such as Vaalharts wit, Vaalharts geel, PNR 6552 and PNR 7563 which are normally resistant to MSV-Set in one of the four cages with pKoS101 agroinoculated plants (leafhopper transmissions tests were quadrupled in four sets of cages; Table 4.3). This was regarded as an aberrant result, and possibly the result of a contamination: however, an alternative explanation is given later in the discussion (4.5.1). The symptom severities of the pSeK102 agroinoculations in wheat cv.s SST66 and Nantes were not as severe as in the MSV-Kom or MSV-Set wild-types.

4.4.3 RF-DNA restrictions and Southern blot analyses

4.4.3.1 *Mapping RF-DNA from pSeK102 agroinoculated plant DNA*

RF-DNA from pSeK102 agroinfected plants was subjected to a range of single and double restriction digests to establish if recombinants were present (Figure 4.6). Restriction digests indicated that the majority of the RF-DNA molecules were MSV-Kom-like rather than MSV-Set -like. It was shown that sufficient amounts of RF-DNA can be isolated to perform restriction digests for mapping for this recombinant virus or population of recombinants at least. Southern blots of digested DNA using restriction enzymes that would distinguish MSV-Kom from MSV-Set RF-DNA, were performed in an attempt to distinguish recombinants that may have been present at lower DNA levels.

Table 4.3 Host ranges of the heterodimeric constructs

Cultivar	pKom504	pSet102	pSeK102	pKoS101
maize				
Jubilee[a]	+++	++	+++	+++
US More[a]	+++	+++	+++	+++
Popcorn[b]	+++	++	+++	+++
Witplat	+++	++	+++	++
Vaalharts wit	+++	-	+++	$-^{(+++)}$
Vaalharts geel	+++	-	+++	$-^{(+++)}$
PNR 6549	+++	+[+]	+++	++
PNR 6552	+++	-	+++	$-^{(+++)}$
PNR 7541	+++	-	na	+++
PNR 7563	+++	-	+++	$-^{(+++)}$
Clipper[d]	+++[+]	+++	+++	++
wheat				
SST 16	+++	+++	na	+++
SST 33	++	++	na	++
SST 44	-	+	[+]	+
SST 55	++	+++	++	+++
SST 66	+++	+++	++	+++
Adam Tas	+++	+++	na	+++
Chokka	+	+++	+[+]	+++
Dias	+	+++	+	+++
Gamtoos	++	++	+	++
Nantes	++[+]	+++	++	+++
Schooner	++	++	++	+
Sterling	+++	++	++	-
Vloekskoot	-/+	++	-	++

Top, *Z.mays* cultivars and bottom, *T.aestivum* cultivars; [a] Sweetcorn variety; [b] purchased at local supermarket in Cape Town; [d] *H.vulgare* variety; PNR seed obtained from Pioneer Seed; na = leafhopper transmissions not performed. - = no streak symptoms; + = mild stippled streak and recovery(-/+); ++ = moderate stippled to continuous streak; +++ = severe continuous streak and stunting and/or distortion of growth and death(+++[+]). $-^{(+++)}$ = severe symptoms obtained in one of the four cages (experiments were quadrupled).

Figure 4.4: Streak symptoms on heterodimer-agroinoculated sweetcorn

3rd leaves are shown three weeks after agroinoculation.
From the top: a, heterodimers pKoS102 or pKoS103 agroinoculations produced no streak symptoms; b and c, wild-type homodimers pKom504 and pSet102 respectively; d, e and f; heterodimers pSeK101, pSeK102 and pKoS101 respectively.

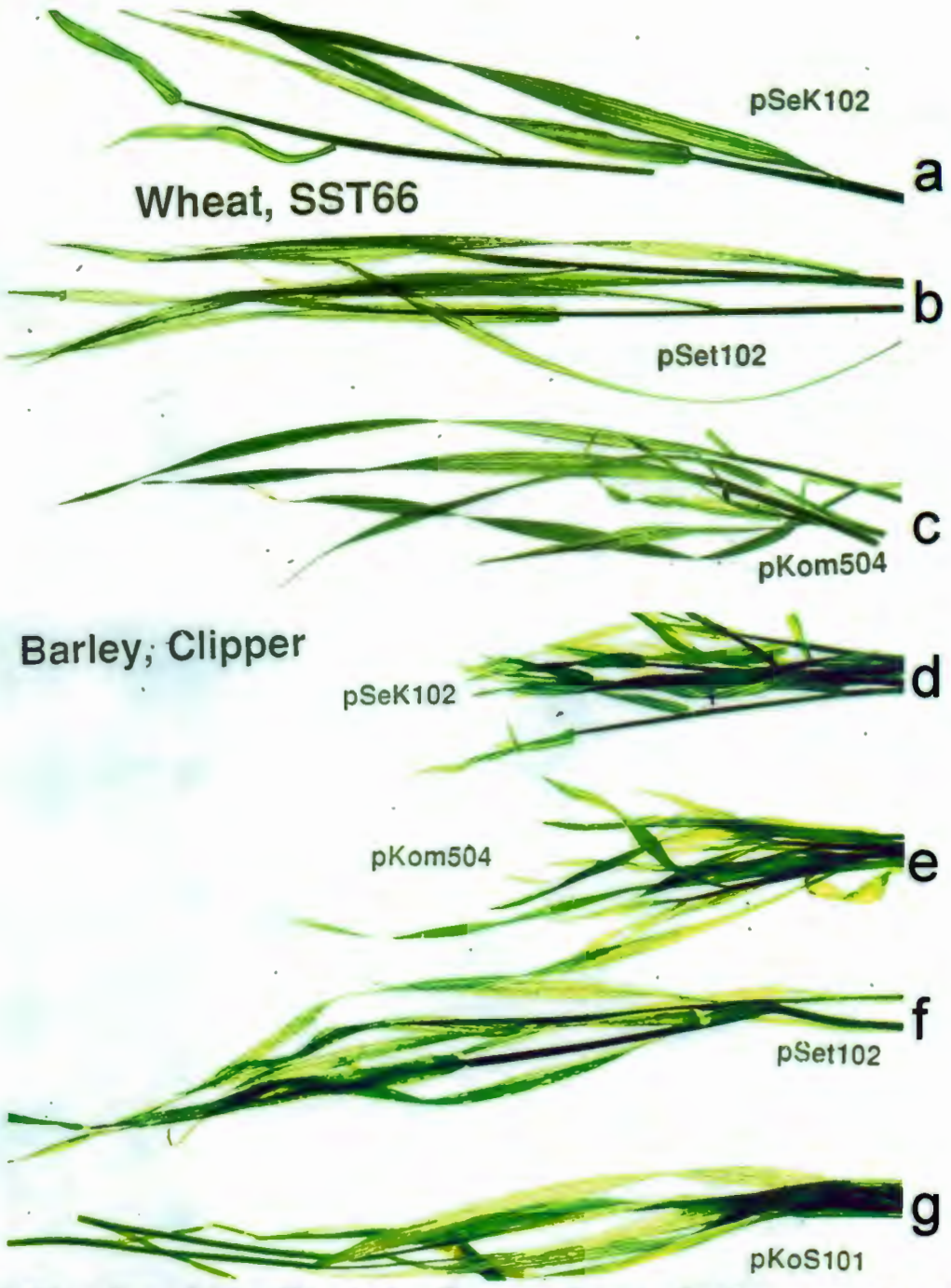

Figure 4.5: Streak and stunting in leafhopper transmissions on wheat and barley

Two-week-old plantlets were exposed to viruliferous leafhoppers that had fed off heterodimer-agroinoculated sweetcorn. Plants shown are 4 weeks old. a, b, and c are leafhopper transmissions on SST66 wheat; heterodimer pSeK102 agroinoculation, and wild-type homodimeric agroinoculations pSet102 and pKom504 respectively. d, e, f and g; leafhopper transmissions on Clipper barley; d and g from heterodimer agroinoculations pSeK102 and pKoS101 respectively; e and f, from wild-type homodimer agroinoculations pKom504 and pSet102 respectively.

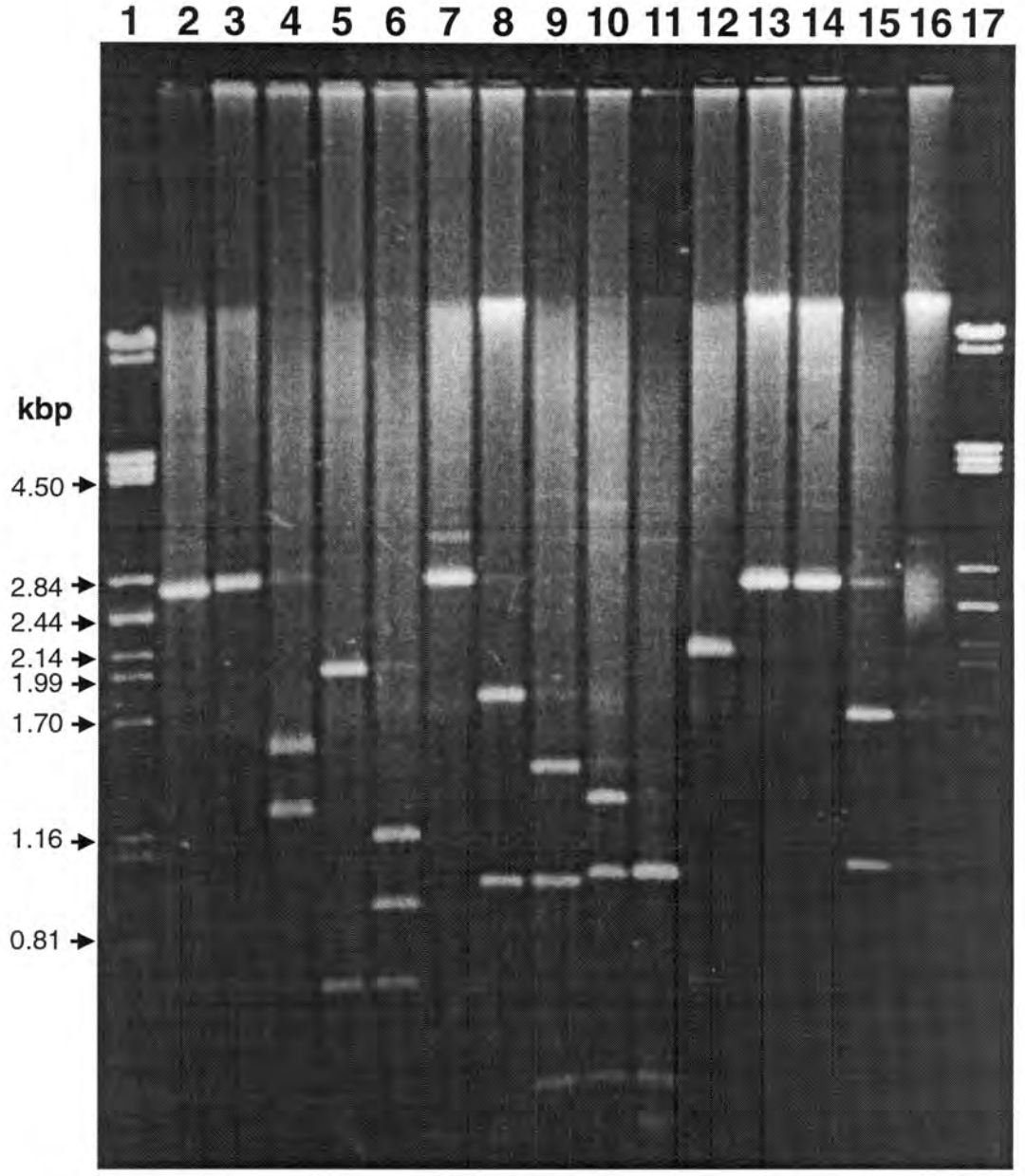

Figure 4.6 Restriction digests to determine recombinant population infecting a pSeK102-agroinoculated plant.

0.8% agarose gel run as described in appendix B3. RF-DNA isolated from total plant DNA extract as described in Appendices B12 and B14. Digestions were performed as described in Appendix B1. Lanes 1 and 17, λ DNA digested with Pst I (Molecular weight marker); **Digestions**: Lane 2, *Bam* HI = 2.7 kb; lane 3, *Bgl* II = 2.7 kb; lane 4, *Bam* HI and *Bgl* II = 1.5 kb + 1.2 kb; lane 5, *Hind* III = 2.0 kb + 0.7 kb; lane 6, *Hind* III and *Bam* HI = 1.12 kb + 0.88 kb + 0.7 kb; lane 7, *Kpn* I = 2.7 kb; lane 8, *Pvu* II = 1.82 kb + 0.92 kb; lane 9, *Pvu* II and *Bam* HI = 1.4 kb + 0.96 kb + 0.4 kb; lane 10, *Sac* I = 1.3 kb + 1.0 kb + 0.4 kb + 0.2 kb(not seen); lane 11, *Sac* I and *Bam* HI = 1.0kb + 1.0 kb + 0.35 kb + 0.40 kb; lane 12, *Sal* I = 2.7 kb; lane 13, *Xho* I = 2.7 kb; lane 14, *Xho* I and *Bam* HI = 1.65 kb + 1.05 kb; lane 15, *Bgl* I = 2.0 kb + 0.78 kb, lane 16, undigested.

Digestions mapped to the MSV-Kom map (See figure 2.1) thus recombinant population was MSV-Kom-like. A cloned pSeK recombinant cloned from this population was also found to be MSV-Kom-like (gel not shown). kbp = kilo base pairs.

4.4.3.2 *Southern blots to detect recombinant viruses*

DNA extracted from agroinoculated plants with and without typical Mastrevirus streak was subjected to *Rsa* I or *Sac* I digestion and Southern blotted to identify whether mixed infections of recombinants and/or wild-type viruses were replicating in the host plants. The Southern blots are depicted in Figure 4.7. "RF-DNA" (plasmid-like DNA isolated from total plant DNA extract using Qiagen columns) extracted from pBI121-agroinoculated (negative control), homodimer- or heterodimer-agroinoculated symptomless plants did not hybridise with either the MSV-Kom or MSV-Set DNA probes. This indicated (as in Chapter 2) that only plants with noticeable streak symptoms had Mastreviral RF-DNA at detectable levels in Southern blot tests (blots not shown). Along with the results obtained in Chapter 2, DNA isolated from plants with relatively severe streak displayed DNA banding patterns on autoradiographs of greater intensity than DNA from plants with milder streak symptoms. This is presumably due to the greater amount of RF-DNA available for hybridisation in plants with severe symptoms than DNA from plants with milder streak symptoms. For example, as indicated in Figure 4.7, RF-DNA banding patterns from plants with mild symptoms after agroinoculation with pSeK102, pSeK101 or pKoS101, were less intense than the banding patterns obtained from plants agroinoculated with the same clones, but with more severe streak symptoms.

The *Sac* I and *Rsa* I digestions of RF-DNA from pSeK101, pSeK102, and pKoS101 agroinoculated plants indicate that the recombinants infecting the plants were largely MSV-Kom-like or MSV-Set-like by Southern blot criteria (Figure 4.7). Banding patterns were found to generally agree with the MSV-Kom and MSV-Set wild-type patterns. However in the complete heterodimer (pSeK101 and pKoS101) agroinoculations, banding patterns of both wild-types were obtained which indicated mixed infections of MSV-Kom-like and MSV-Set-like recombinants rather than restriction patterns unique to the wild-type banding patterns.

On the blot with *Rsa* I-digested DNA, wild-type MSV-Set banding patterns were not obtained when probed with MSV-Kom DNA. The reason for this is that MSV-Kom tends to replicate at much higher levels in sweetcorn plants than MSV-Set and thus MSV-Set DNA is not as detectable in cross-hybridisations (Chapter 2). When probed with MSV-Set however, cross-hybridisation is detectable indicating that there are higher levels of more severe MSV-Kom-like recombinant DNA (that is only 78% homologous to MSV-Set; Chapter 3) than milder MSV-Set-like recombinant DNA in the same plant. Thus the blots probed with MSV-Kom DNA did not display any mixed infections of MSV-Set-like viruses.

In *Sac* I digestions however, MSV-Kom probes could be used to detect MSV-Set-like recombinants or wild-type MSV-Set RF-DNA by cross-hybridisation since the MSV-Set genome has only one *Sac* I whereas the MSV-Kom RF-DNA has four *Sac* I sites. This means that MSV-Set RF-DNA will migrate as a single 2.7 kb band on the gel whereas that of MSV-Kom with four *Sac* I sites would migrate as four separated fragments of smaller size - thus the MSV-Kom probe could detect MSV-Set RF-DNA by cross-hybridisation whereas MSV-Set probes could not detect the smaller *Sac* I fragments of MSV-Kom or MSV-Kom-like RF-DNA on the same blot. This phenomonon is depicted in lanes 5, 9 and 10 of the Southern blots with *Sac* I digestions where pSeK101- or pKoS101-agroinoculated plant DNA displayed both banding patterns (Figure 4.7). It is interesting to note that the streak symptoms were mild on these plants indicating the possibility that MSV-Set-like recombinants dominated in these populations.

Interestingly, DNA from plants agroinoculated with the complete heterodimers pSeK101 and pKoS101 displayed the banding patterns of both MSV-Kom and MSV-Set, indicating a mixed infection of MSV-Kom-like and MSV-Set-like recombinants. In contrast however, agroinoculations of pSeK102 only produced banding patterns that indicated MSV-Kom-like recombinants although a MSV-Set-like 0.50 kb fragment was observed along with the four MSV-Kom-like fragments (0.90, 0.69, 0.53, 0.40 kb; Figure 4.7). Bands not conforming to the wild-type banding patterns were the 0.3 kb fragments observed in DNA from pSeK101- and pKoS101- agroinoculated plant plants with severe symptoms (Lanes 2 and 7 in left blot, Figure 4.7). This indicated the possibility of the presence of recombinants that were not wild-type-like, however, at much lower levels than wild-type RF-DNA or wild-type-like RF-DNA.

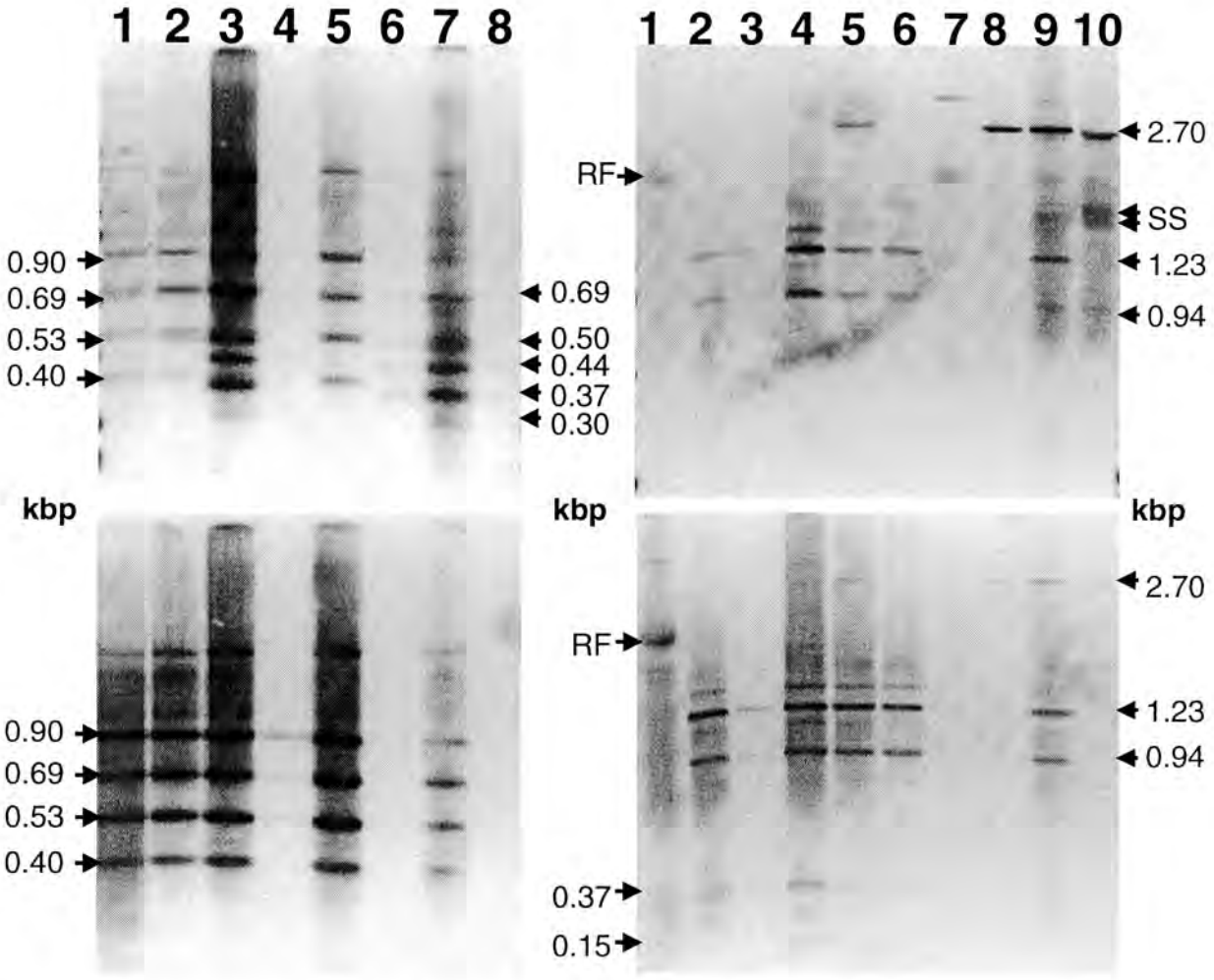

Figure 4.7 Southern blots for the detection of recombinant viruses

Southern blots with individual RF-DNA extracts from total plant DNA extractions from selected agroinoculated plants with streak symptoms.

Left: RF-DNA digested with *Rsa* I to determine the presence of recombinant viruses. Top left; DIG-labelled MSV-Set probed blot that was stripped for; bottom left, rehybridisation with MSV-Kom probe. Wild-type MSV-Kom will produce 60, 121, 398, 528, 685 and 897 bp fragments whereas MSV-Set wild-types will produce 9, 36, 46, 51, 54, 86, 121, 138, 154, 372, 442, 500 and 692 bp fragments. Only the fragments larger than 150 bp are detected on the blots. Lane 1; pKom504 agroinoculation with typical streak symptoms of MSV-Kom. Lane 2; pSeK101 agroinoculation with mild streak symptoms. Lane 3; pSeK102 agroinoculation with severe streak symptoms. Lane 4; pSeK102 agroinoculation with mild streak symptoms. Lane 5; pSeK102 agroinoculation with severe streak symptoms. Lane 6; pSet102 agroinoculation with typical moderate streak symptoms of MSV-Set. Lane 7; pKoS101 agroinoculation with severe streak symptoms. Lane 8; pKoS101 agroinoculation with mild streak symptoms.

Right: RF-DNA digested with *Sac* I. Top right; DIG-labelled MSV-Set probed blot that was stripped for, bottom right; rehybridisation with MSV-Kom probe. Wild-type MSV-Kom will produce 154, 369, 940 and 1227 bp fragments whereas MSV-Set containing only one *Sac* I site will produce a 2701 bp fragment. Lane 1 and 2 ; uncut and cut pKom504 agroinoculation with typical MSV-Kom streak symptoms respectively. Lanes 3 and 4; pSeK102 agroinoculations with mild or severe streak symptoms respectively. Lanes 5 and 6; pSeK101 agroinoculations with mild or severe streak symptoms respectively. Lanes 7 and 8; uncut and cut pSet102 agroinoculation with typical MSV-Set streak symptoms respectively. Lanes 9 and 10; pKoS101 agroinoculations with mild or severe streak symptoms respectively. Arrowed values indicate fragment sizes in kb units. SS = single stranded DNA. RF = RF-DNA. kbp = kilobase pairs

4.4.4 Sequence analyses of recombinant cloned RF-DNAs

Recombinant RF-DNAs were obtained from heterodimer agroinoculated plants with typical Mastreviral streak symptoms using Qiagen® columns. They were cloned into the *Bam* HI site of pUC18 cloning vectors, and sequenced into the complementary sense LIR or the virion sense V1 ORF. The results of this analysis are summarised in Tables 4.4, 4.5 and 4.6.

Eleven 2.7 kb recombinant RF-DNAs were cloned from pSeK102 agroinoculated plants with severe streak symptoms and partially sequenced to analyse and determine genomic recombinational sites between MSV-Kom and MSV-Set wild-types. The recombinants pSeK001-pSeK008 were found to consist of MSV-Kom wild-type sequence except for nucleotide positions (nt pos.) 147 and 148 (TT) which was apparently replaced by nt pos. 149 and 150 (GC) of MSV-Set (Figure 4.8). Similarly, three of the ten recombinants obtained from pKoS101-agroinoculated plants (pKoS001, pKoS002 and pKoS003) with milder symptoms had the same recombinational event occurring except the sequences were wild-type MSV-Set sequences with nt pos. 149 and 150 (GC) apparently replaced by the nt pos. 148 and 149 (TT) of MSV-Kom (Table 4.5; Figure 4.8). A recombinant cloned from a 1.1mer heterodimer pKoS105-agroinoculated plant, pKoS151 had the same sequence with this type of replacement. It appears that the recombinational crossover point resulting in this 2 bp replacement is situated within nt pos. 133-146 and 134-147 in the MSV-Kom and MSV-Set genomes respectively. This 14bp sequence is conserved in both viruses and a 7 b stretch (GCGATTC MSV-Kom nt pos. 139-145 or MSV-Set nt pos. 138-144) within this conserved region is conserved in most geminiviruses (K. Palmer pers. comm.).

In addition to the homologous site where crossover occurs it is of note that the ATG start codon/*Bam* HI site may be an aid in homologous recombination for the 2 bp replacement event. This is supported by the MSV-Set wild-type recombinants which were obtained from pKoS101 and pKoS105 agroinoculated plants: pKoS152 and pKoS153 obtained from pKoS105 agroinoculated plants and pKoS109 obtained from the pKoS101 agroinoculated plant were found to be wild-type MSV-Set sequences. For this recombination to occur the crossover site must have occurred at the ATG start codon/*Bam* HI homologous region which is 9 b in length resulting in no effective replacements between MSV-Kom and MSV-Set in the heterodimeric clones (Figure 4.8). This homologous region is situated in MSV-Kom nt pos. 149-157 or MSV-Set nt pos. 150-158. No wild-type MSV-Kom recombinants have thus far been found except for pSeK119 (Table 4.5).

Other instances of wild-type MSV-Set unit-length RF-DNAs cloned from heterodimer agroinoculated plants were pSeK161, pSeK162 and pSeK163 obtained from pSeK106-

agroinfected plants showing typically MSV-Set-like streak symptoms. In this case the homologous recombinational release of unit-length genomes appears to have occurred in the *Sac* I site used as a junction for recombining MSV-Set unit length RF-DNA with the MSV-Kom sequence stretching from the conserved *Sac* I site through the LIR to the conserved *Bam* HI site (Figure 4.3). The homologous region is 9 b in length (TGG**GAGCTC**) including the *Sac* I site (bold) and is situated in the C1 or Rep A ORF (MSV-Kom nt pos. 2566-2574 or MSV-Set nt pos. 2564-2572).

Most recombinants cloned from pSeK102 and pSeK101 agroinoculated plants contained MSV-Kom sequences with various LIR replacements of homologous MSV-Set sequences. These include pSeK009 and pSeK010 from pSeK102 agroininfected plants and pSeK111 from pSeK101 agroinoculated plants which had significant replacements of 174 bp from the UAS through the hairpin loop up to the *Bam* HI site replaced by MSV-Set homologous sequences (nt pos. 2666-2690 and 1-148 replaced by MSV-Set nt pos. 2678-2701 and 1-149). The identified homologous sequences where the crossover occurs is 29 bp in length and situated in MSV-Kom nt pos. 2635-2665 or MSV-Set nt pos. 2647-2677 which includes the sequences of the UASs and part of the invert repeat sequence of the hairpin loop (Figure 4.9).

The largest group of pSeK recombinants with LIR replacements were those with identical 105 bp LIR replacements; pSeK011 from pSeK102-agroinoculated plants, pSeK114 and pSeK1111 from pSeK101-agroinoculated plants and pSeK141, pSeK142, pSeK143, pSeK144 and pSeK145 from pSeK104-agroinoculated plants. They were all made up of MSV-Kom wild-type sequence with nt pos. 44-148 replaced by MSV-Set nt pos. 44-149. The site of recombination was identified as the 28 bp conserved region (nt pos. 16-43 of MSV-Set or MSV-Kom) containing the *Apa* I restriction site, and the replacement stretches to the *Bam* HI site (Figure 4.10).

pSeK112, the only recombinant from a pSeK101 agroinoculated plant with severe symptoms, consisted of wild-type MSV-Kom sequence with a 134 bp replacement stretching from the base of the hairpin loop sequence to the *Bam* HI site (MSV-Kom nt pos. 15-148 replaced by MSV-Set nt pos. 15-149). (Figure 4.11). The crossover site is identified as a conserved 4 bp CTTT region (MSV-Kom or MSV-Set nt pos. 11-14) in the hairpin loop to the right of the invariant nonanucleotide TAATATTAC sequence. This would presumably be able to disrupt the function of the normally 18 b inverted repeat that potentially forms a hairpin loop since it was effectively a 12 b inverted repeat rather than 18 b found in the wild-type MSV-Kom or MSV-Set (Figure 4.11).

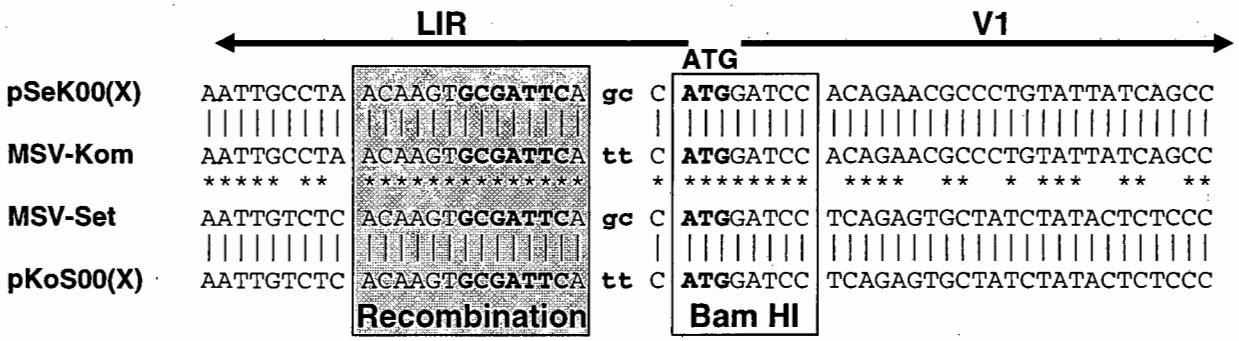

Figure 4.8 Homologous regions where crossover apparently occurs between MSV-Kom and MSV-Set resulting in wild-type-like or wild-type recombinants.

Shaded region indicates the region where the recombinational event occurred in pSeK101, pSeK102 and pKoS101 agroinoculations. Interestingly the 7 bp **GCGATTC** is conserved in most Mastreviruses. The start codon **ATG** is indicated and part of the GGATCC *Bam* HI site (site of MSV-Set/Kom junction in tandem repeats of agroinfectious heterodimers). Arrow right of the start codon indicates direction of the V1 ORF and direction of sequence obtained in the virion sense and left arrow indicates the LIR sequence obtained in the complementary sense. Asterisks indicate homology between wild-type MSV-Kom or MSV-Set and recombinants obtained from their respective agroinoculations homology are indicated by I. pSeK00(X) indicates recombinant clones SeK001-pSeK008 and pKoS00(X) indicates the recombinants pKoS001-pKoS003 and pKoS151. Sequence indicated in lower case indicates the only two nucleotides in recombinant genomes differing from the wild-types. Crossover sites for pKoS009 from a pKoS101-agroinoculated plant and pKoS152 and pKoS153 from pKoS105-agroinoculated plants occur at the white-boxed region containing the V1 ORF **ATG** start codon and the conserved Bam HI site resulting in 100% wild-type MSV-Set sequence.

pSeK115, pSeK1110 and pSeK118 were members of a curious group of recombinants from pSeK101 agroinoculated plants as they appear to include an extra 11bp or 13bp MSV-Set extra sequence in the unit-length recombinant genome. This seemed to be achieved by replacing MSV-Kom nt pos. 14-148 by MSV-Set nt pos. 1-149 in recombinants pSeK115 and pSeK1110; that is effectively 135 MSV-Kom bp replaced by 149 MSV-Set bp (Figure 4.12). pSeK118 has 11 bp extra, presumably by the same mechanism as that of pSeK115 and pSeK1110. In this case it appears that 137 MSV-Kom bp (nt pos. 11-148) is replaced by MSV-Set nt pos. 2701 + 1-10 (Figure 4.13).

Since the yield of RF-DNA from pKoS101-inoculated plants with mild streak symptoms was low, it complicated the cloning process of RF-DNA into a cloning vector for sequencing. However RF-DNA from pKoS101-agroinoculated plants with moderate symptoms were successfully cloned for sequencing and were designated pKoS001-pKos010. The results of the analysis of these sequences are tabulated in Table 4.6. As already mentioned,

pKoS001, pKoS002, pKoS003 and pKoS151 (from pKoS105-agroinoculated plant) were made up of MSV-Set wild-type sequence except for the nt pos. 149 and 150 (GC) which was replaced by MSV-Kom nt pos. 148 and 149 (TT) with the homologous crossover site of 14 bp (MSV-Kom nt pos. 133-146 or MSV-Set nt pos. 134-147) (Figure 4.5).

A group of recombinants were obtained that were MSV-Kom wild-type-like but had replacements in the V1 ORF which encodes the MP. These included pKoS005, pKoS006, pKoS007, pKoS008 and pKoS010 which were obtained from pKoS101-agroinoculated plants with relatively severe streak symptoms. pKoS005 and pKoS006 had more significant replacements of 176 bp stretching from the conserved *Bam* HI site into the V1 ORF (Figure 4.14). They consisted of MSV-Kom sequences with nt pos. 158-333 replaced with MSV-Set nt pos. 159-334. This replacement effectively replaces 62 N-terminus amino acids of a total 101 (60%) of the MSV-Kom MP sequence with that of MSV-Set (Figure 4.15). Less significant replacements of the MP ORF (V1) were found in pKoS007, pKoS008 and pKoS010 with only 25 bp in nt pos. 158-182 replaced by MSV-Set nt pos. 159-183 (Figure 4.15). Effectively only four amino acids are changed from the MP sequence of the MSV-Kom wild-type. pKoS005 was subjected to further sequencing using C1KEP, 17KR and SIR-Kom primers for virion-sense sequence (1032 b) and C2KEP and LIR-Kom primers for complementary-sense sequence (940 b). Sequences in different regions of the genome were all MSV-Kom wild-type.

Although pKoS004, pSeK113, pSeK116 and pSeK117 were selected for sequencing due to their sizes (2.7 kb) on agarose gels, the sequence data revealed non-MSV sequence. These were regarded as cloning artefacts and since the interest in this study was limited to unit-length genomes that were more likely to be infectious, no further investigations were done on these.

Table 4.4 Summary of recombinants found in pSeK-agroinoculated plants

Recombinant clone	obtained from plants agro-inoculated with	Complementary strand sequenced and read (LIR direction)	Virion sense strand sequenced and read (V1 direction)	Sequence replacements	Probable region of recombination
pSeK001		200 b	69 b		
pSeK002		370 b	487 b	MSV-Kom sequence with 2 bp TT (nt pos. 147 and 148) replaced with MSV-Set nt pos. 149 and 150 (GC).	14 bp in nt pos. 133-146 in MSV-Kom or 134-147 in MSV-Set. (Figure 4.8)
pSeK003	pSeK102	300 b	330 b		
pSeK004	(severe MSV-	320 b	298 b		
pSeK005	Kom-like	420 b	298 b		
pSeK006	symptoms)	320 b	370 b		
pSeK007		320 b	300 b		
pSeK008		300 b	410 b		
pSeK009	pSeK102 (severe symptoms)	270 b ALF	832 b ALF	MSV-Kom sequence with 174 bp nt pos. 1-148 and 2666-2690, including the hairpin loop, replaced with MSV-Set nt pos. 2678-2701 and 1-149	29 bp including the UAS; MSV-Kom nt pos. 2635-2665 or MSV-Set nt pos. 2647-2677. (Figure 4.9)
pSeK010		690 b ALF	786 b ALF		
pSeK111	pSeK101 (severe symptoms)	639 b ALF	905 b ALF		
pSeK011	pSeK102 (severe symptoms)	643 b ALF	687 b ALF	MSV-Kom sequence with 105 bp nt pos. 44-148 replaced with MSV-Set nt pos. 44-149	28 bp including the Apa I site; MSV-Kom or MSV-Set nt pos. 16-43. (Figure 4.10)
pSeK114	pSeK101 (severe symptoms)	825 b ALF	932 b ALF		
PSeK1111		446 b ALF	artefact seq.		
pSeK141	pSeK104 (severe symptoms)	245 b	na		
pSeK142		173 b	na		
pSeK143		290 b	na		
pSeK144		240 b	na		
pSeK145		293 b	na		

ALF = sequence obtained using the ALF express automated sequencer. na = sequencing not done.

Table 4.5 Summary of recombinants found in pSeK-agroinoculated plants continued.

Recomb-inant clone	obtained from plants agro-inoculated with	Complement-ary strand sequenced and read (LIR direction)	Virion sense strand sequenced and read (V1 direction)	Sequence replacements	Probable region of recombination
pSeK112	pSeK101 (severe symptoms)	901 b ALF	822 b ALF	MSV-Kom sequence with 134 bp nt pos. 15-148 replaced with MSV-Set nt pos. 15-149	4 bp MSV-Kom or MSV-Set nt pos. 11-14 in stem of hairpin loop. (Figures 4.10 and 4.11)
pSeK115 pSeK1110	pSeK101 (severe symptoms)	741 b ALF 807 b ALF	616 b ALF 1046 b ALF	MSV-Kom sequence with 135 bp nt pos. 14-148 replaced with MSV-Set nt pos. 1-149. (Figure 4.12)	Appears that the 13 bp inclusion occurs during replicative release when unit-length linear ssDNA ends are joined to form circular molecules.
pSeK118	pSeK101 (severe symptoms)	765 b ALF	1100 b ALF	MSV-Kom sequence with 137 bp nt pos. 11-148 replaced with MSV-Set nt pos. 2701 + 1-10. (Figure 4.13)	Appears that the 11 bp inclusion occurs during replicative release when unit-length linear ssDNA ends are joined to form circular molecules.
pSeK119	pSeK101 (severe symptoms)	353 b ALF	972 b ALF	MSV-Kom sequence with no apparent replacements in LIR.	Wild-type MSV-Kom
pSeK161 pSeK162 pSeK163	pSeK106 (moderate streak)	690 b ALF 680 b ALF 668 b ALF	na na na	wild-type MSV-Set sequence with no replacements	9 bp identity including Sac I site in MSV-Kom nt pos. 2566-2574 or MSV-Set nt pos. 2564-2572

ALF = sequence obtained using the ALF express automated sequencer. na = sequencing not done.

Table 4.6 Summary of recombinants obtained from pKoS-agroinoculationed plants

Recomb-inant clone	obtained from plants agro-inoculated with	Complement-ary strand sequenced and read (LIR direction)	Virion sense strand sequenced and read (V1 direction)	Sequence replacements	Probable region of recombination
pKoS001 pKoS002 pKoS003 pKoS151	pKoS101 (moderate streak) pKoS105 (moderate streak)	320 b 360 b 360 b 291 b	450 b 300 b 300 b na	MSV-Set sequence with 2 bp GC (nt pos. 149 and 150) replaced with MSV-Kom nt pos. 148 and 149 (GC).	14 bp in nt positions 133-146 (MSV-Kom) or 124-147 (MSV-Set). (Figure 4.8)
pKoS005* pKoS006	pKoS101 (moderate streak)	332 b, 348 b (C1KEP), 157 b (17KR), 195 b (SIR-Kom) 307 b	360 b, 280b (C2KEP), 300 b (LIR-Kom) 463 b	MSV-Kom sequence with 176 bp nt pos. 158-333 replaced with MSV-Set nt pos. 159-334 in V1 region	2bp identical in MSV-Kom nt pos. 334 and 335 or MSV-Set nt pos. 335 and 336. (Figure 4.14)
pKoS007 pKoS008 pKoS010	pKoS101 (moderate streak)	265 b 230 b 280 b	250 b 230 b 280 b	MSV-Kom sequence with 25bp nt pos. 158-182 replaced with MSV-Set nt pos. 159-183 in V1	23 bp identical in MSV-Kom nt pos. 183-205 or MSV-Set nt pos. 184-206. (Figure 4.15)
pKoS009 pKoS152 pKoS153	pKoS101 (moderate streak) pKoS105 (moderate streak)	290 b 662 b ALF 683 b ALF	270 b na na	wild-type MSV-Set sequence with no replacements	9 b including the V1 start codon and Bam HI site.. MSV-Kom nt pos. 149-157 or MSV-Set nt pos. 150-158. (Figure 4.8)

ALF = sequence obtained using the ALF express automated sequencer. na = sequencing not done

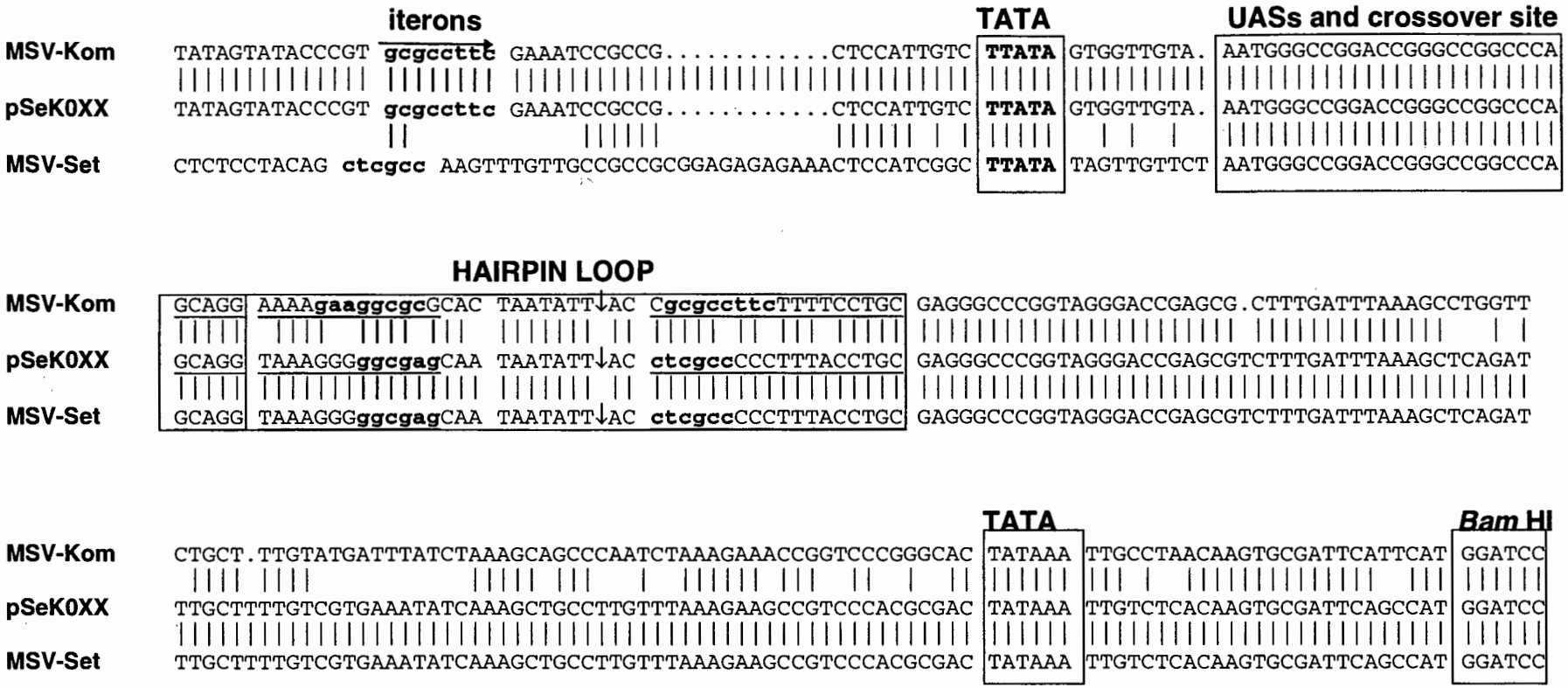

Figure 4.9 Sequence of pSeK009, pSeK010 and pSeK011 recombinant clones with 174 bp replacement in the LIR

The sequence of pSeK009, pSeK010 and pSeK 011 recombinants consists of MSV-Kom sequence with the sequences from the UAS to the *Bam* HI site replaced with the cognate sequences of MSV-Set. The recombinational site is shaded (UAS): this has 22 bp of complete homology between MSV-Kom and MSV-Set. The other crossover point is the *Bam* HI site which is the junction joining MSV-Set in tandem with MSV-Kom in the heterodimeric clone pSeK102 used in agroinoculations. TATA boxes, *Bam* HI site, hairpin loop structure with invariant nonanucleotide TAATATT↓AC, UAS (upstream activating sequences) and iterons (bold and lower case) are indicated. (see Figure 3.10). Homologous sequences between the recombinant and wild-types are indicated by lines (|) between the sequences.

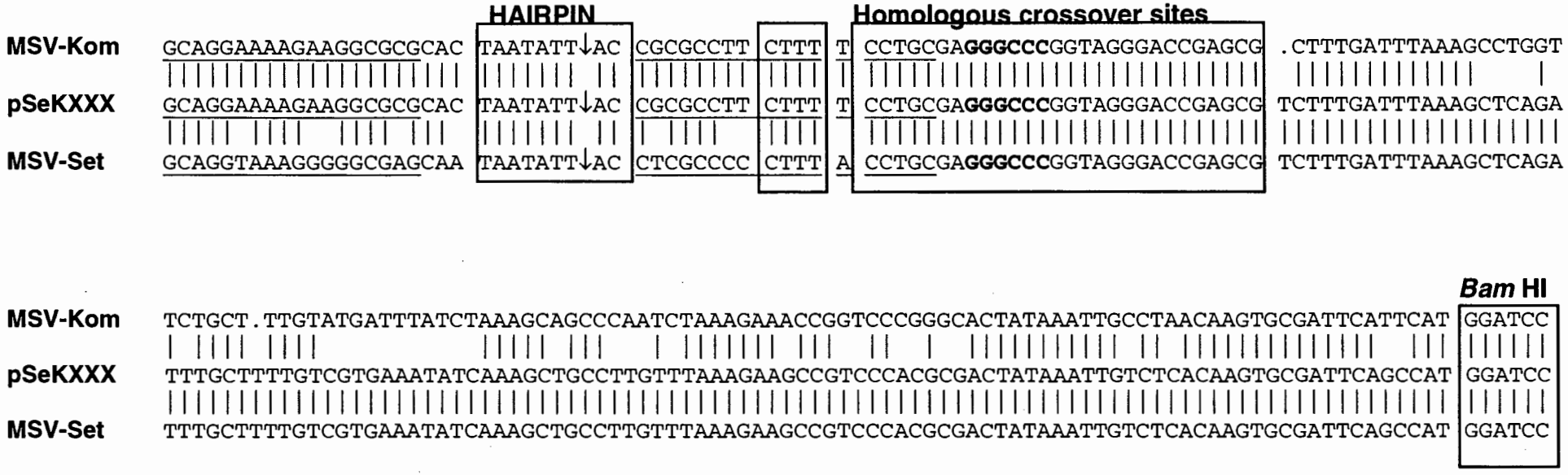

Figure 4.10 105 bp LIR sequence replacements of recombinants from pSeK102, pSeK104 and pSeK101 agroinoculated plants

Sequence from the homologous region(shaded) to the *Bam* HI site is replaced by the corresponding sequence from MSV-Set. Shaded region containing the GGGCCC (bold) *Apa* I site must be the site of recombination for recombinants pSeK141, pSeK142, pSeK143, pSeK144, pSeK145, pSeK011, pSeK114 and pSeK1111. The *Bam* HI site is the junction site of the MSV-Set/MSV-Kom tandem repeat cloned in pSeK102. The size of the homologous recombinational region is 28 bases. CTTT shaded region is the crossover site for the pSeK112 recombinant.

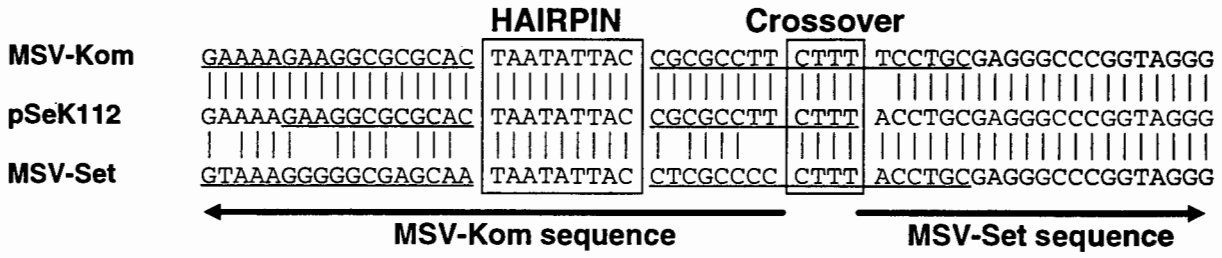

Figure 4.11 Replacement in pSeK112

pSeK112 recombinant with a 133bp replacement extending from the CTTT homologous crossover region (indicated with a shaded box) situated within the right stem sequence from nucleotide positions 15-148 up the the *Bam* HI site. Underlined sequence indicates inverted repeat in the stem.

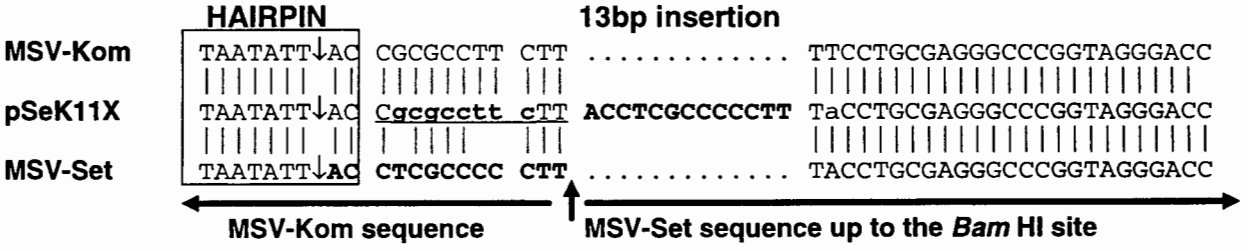

Figure 4.12 Sequence insertions in pSeK115 and pSeK1110

An inclusion of MSV-Set 13 bp (bold) is a repeat sequence of MSV-Set nt pos. 1-13. Underlined sequence indicates the inverted repeat on right side of TAATATTAC putatively forms an 11 bp stem of hairpin. Lower case bold type indicates a full MSV-Kom iteron.

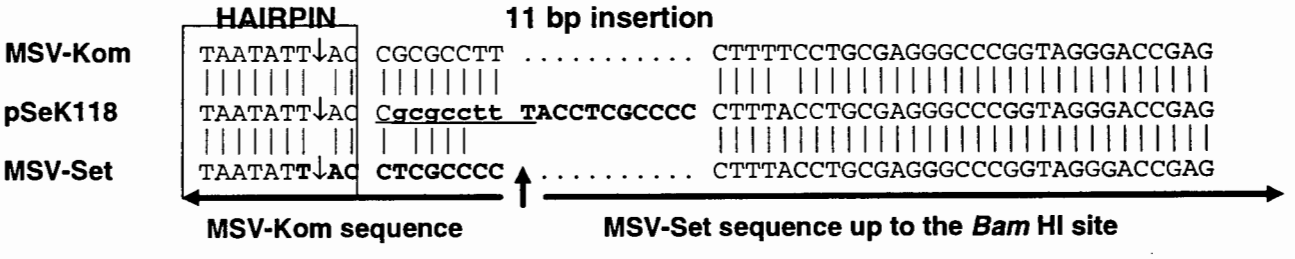

Figure 4.13 11bp insertion in pSeK118 recombinant

An inclusion of MSVSet 11bp (bold) is a repeat sequence MSV-Set nt pos. 2701 and 1-10. Underlined sequence indicates the right 9b inverted repeat flanking TAATATTAC forms a putative stem of hairpin loop. Lower case and bold type indicates a one base smaller MSV-Kom iteron in the stem.

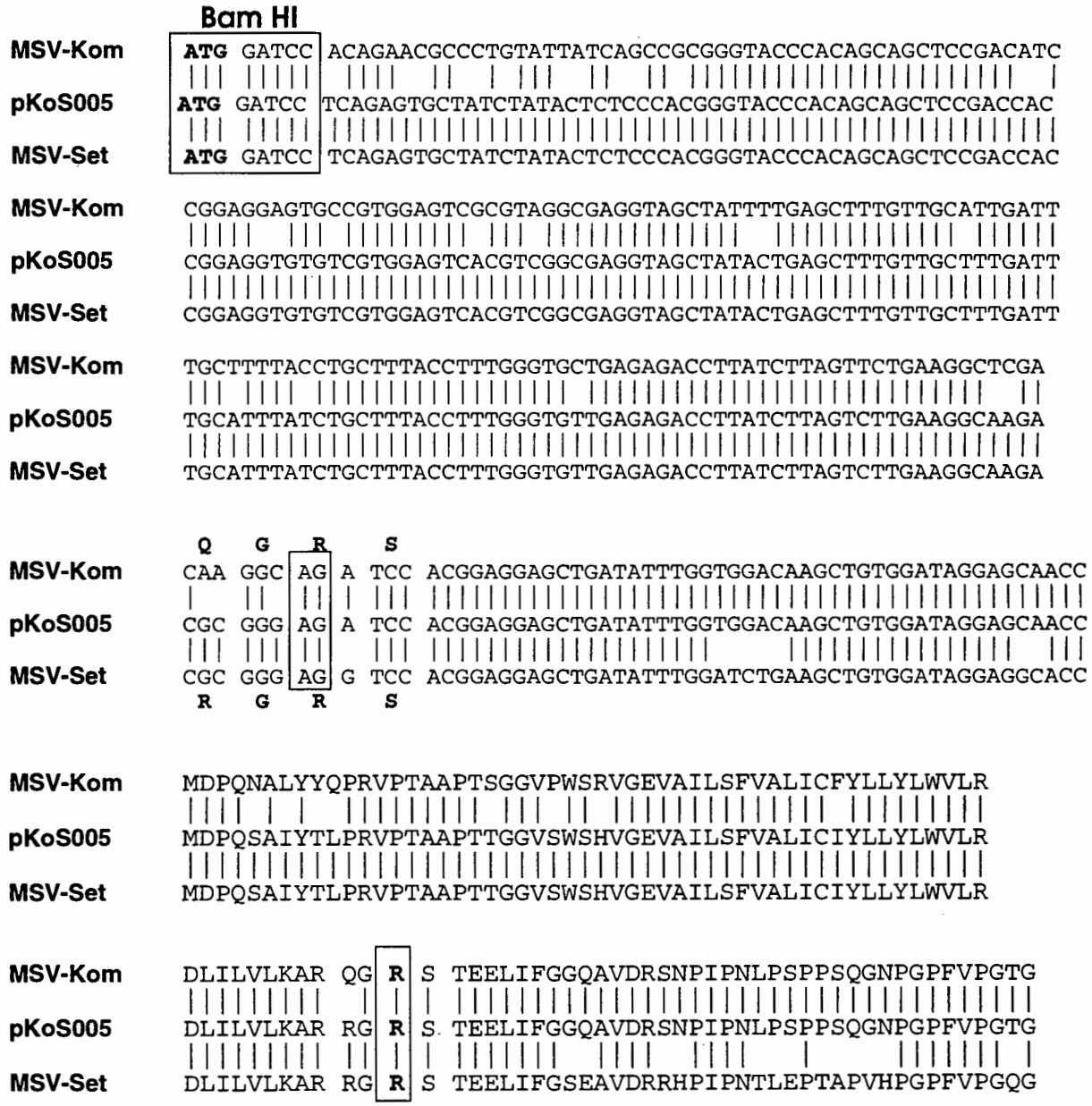

Figure 4.14 Replacement in the MP ORF of pKoS105 and pKoS006 recombinants.

The 176 bp replacement of pKoS005 and pKoS006 replaces the amino acid sequence of 62 residues of the MP of MSV-Kom with that of MSV-Set. The shaded AG region of the nt sequence is the presumptive site of recombination. Note that the sequences flanking the crossover point are conserved in MSV-Kom and MSV-Set. This recombination would not alter the hydrophobicity profiles significantly since they are similar for MSV-Kom and MSV-Set (Chapter 3). Above four rows are the nt sequences (incomplete) and the two rows below are the amino acid sequences (complete). Althought the codon is altered in the putative crossover site the resulting amino acid coded for is not altered. *Bam* HI site and start codon (**ATG**) is indicated in white box.

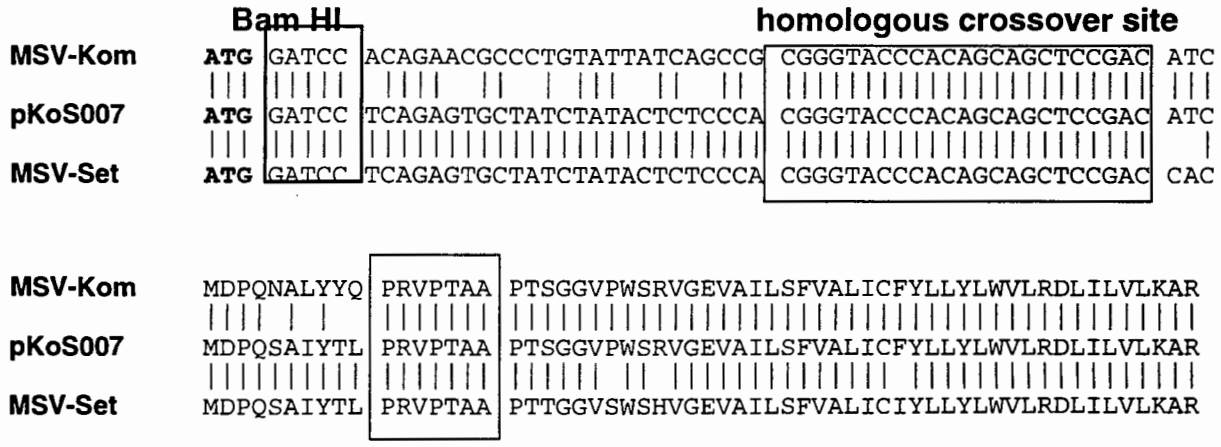

Figure 4.15 MP ORF replacement of pKoS007, pKoS008 and pKoS010 recombinants

The 25 bp replacement of pKoS007, pKoS008 and pKoS010 replaces six amino acid residues of the MP of MSV-Kom with those of MSV-Set. The shaded region of the nt sequence is the presumptive region where recombinational crossover occurs. Top are the nt sequences and below the amino acid sequences. Effectively only four amino acids differ in these MPs. *Bam* HI site and start codon (**ATG**) are indicated. the entire nucleotide sequences of the MP ORF and amino are not included but only the pertinent sequences.

4.5 DISCUSSION

All cloned MSV-Kom/MSV-Set heterodimers were agroinfectious, except for those *Agrobacterium* clones containing one LIR; that is pKoS102 and pKoS103. Despite the biological and genomic differences between MSV-Kom and MSV-Set (Chapter 2 and 3), these results indicate a degree of biological and genomic compatibility between the two viruses. Southern blot tests, leaf-dip preparations (Figure 4.16; M. Jaffer pers. comm.), and leafhopper transmission tests were the criteria used - as in Chapter 2 - to ascertain the infectivity and pathogenicity of the recombinants. Restriction digestions using restriction endonucleases that distinguish the MSV-Kom or MSV-Set genomes, and/or Southern blot tests indicated that resultant recombinant populations contained largely wild-type-like genomes (Figures 4.6 and 4.7). The cloning and sequencing of several recombinants indicated that they were indeed "wild-type-like" recombinants as the biological results suggested. The term "wild-type-like" in this section refers to unit-length genomes consisting mainly of wild-type MSV-Kom or MSV-Set sequence (> or = 93.5%) with minor MSV-Set or MSV-Kom sequence replacements (< or = 6.5%) respectively. Thus "MSV-Kom-like" and "MSV-Set-like" refers to genomes with more than 93.5% MSV-Kom or MSV-Set sequence respectively, contained in recombinant unit-length genomes or cloned RF-DNA.

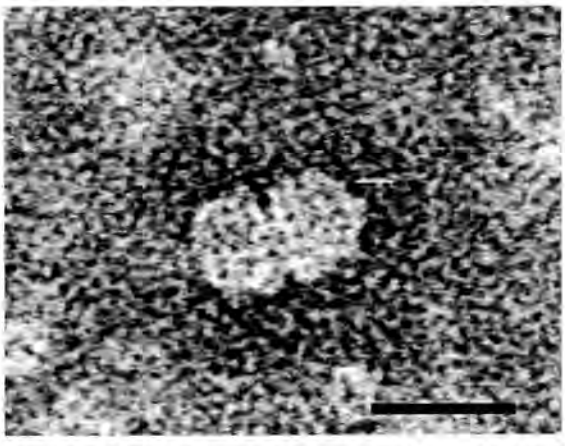

Figure 4.16 Geminate particle from leaf-dip preparation of pSeK102-agroinoculated plant (M.Jaffer) Stained with uranyl acetate and the barline indicates 30nm .

Previous investigations in complementation experiments with geminivirus mutants have shown that recombination occurs between mutants to produce wild-types which subsequently supplant the less viable mutants to dominate the population (TGMV, Brough *et al.*, 1988; ACMV, Etessami *et al.*, 1988; MSV, Lazarowitz *et al.*, 1989; BCTV, Hormuzdi and Bisaro, 1993). The present study specifically targeted more viable recombinants by allowing agroinoculated plants to grow up to 3-5 weeks before the RF-DNAs were isolated.

Presumably as the plant grows older, less viable recombinants would become dominated by more infectious viable recombinants in any infecting recombinant population. Fundamental differences between the complementation experiments and the present study are: i) that the complementation experiments demonstrated intermolecular as opposed to intramolecular recombination of geminiviruses; and, ii) recombination in the complementation experiments occurred between homologous mutant genomes as opposed to between genomes sharing only 78% identity as is the case in the present study. However intramolecular recombinational studies have been conducted in which heterodimers of different strains of geminiviruses were tested in either agroinoculations using the Worland and Logan strains of BCTV (Stenger *et al.*, 1991) or in wheat suspension cells using WDV-S and WDV-Er genomes (Heyraud, *et al.*, 1993). In these cases the nucleotide sequence identity between the BCTV strains is approximately 83% and between the WDV strains is 82.5%. In both cases the more efficient mechanism of release was determined to be the replicational release model which led to the mapping of the location of initiation and termination of unit-length genomes of geminiviruses by RCR (Chapter 1).

4.5.1 Biological features of partial or full heterodimers

It appears that the full heterodimer pKoS101 and the partial pSeK102 (partial heterodimer with two LIRs) behaved more like the wild-type homodimeric counterparts pKom504 and pSet102. Although their agroinfectivity results were erratic (Table 4.1; Figure 4.3 and Appendix D2), it appears that the severity of the symptoms and their host ranges (Table 4.3) are in line with the results of the one agroinoculation experiment (Experiment I, Table 4.1; Figure 4.3). That is, the host ranges and symptom severities obtained for pSeK102 and pKoS101 were comparable to those of wild-types pKom504 and pSet102 respectively (Table 4.3; Figure 4.5). Although pSeK101 *Agrobacterium* clones displayed erratic agroinfectivities (Table 4.1) they produced MSV-Kom-like streak symptoms on agroinoculated plants (Figure 4.4 d). These results suggest that the possible recombinant unit-length genomes released from pSeK102 are probably MSV-Kom-like and those from pKoS101 are probably MSV-Set-like.

Recombinant RF-DNA isolated from a pSeK102-agroinoculated plant and subjected to restriction digests (Figure 4.6) indicated that the recombinant population was MSV-Kom-like. Further a pSeK102 recombinant RF-DNA cloned and mapped also revealed that the clone was MSV-Kom-like. Since the MSV-Set genome contained in pKoS101 and that of MSV-Kom contained in pSeK102 are delimited by heterologous LIRs (Figure 4.2), one would expect MSV-Set-like and MSV-Kom-like recombinants released from pKoS101 and pSeK102 respectively (Heyraud *et al.*, 1993).

Restriction digests using enzymes that distinguish the MSV-Set from the MSV-Kom genomes on an agarose gel and subsequent Southern blot tests revealed that the majority of recombinants released from partial or full heterodimers were wild-type-like (Section 4.4.3.2 and Figure 4.7). However plants agroinoculated with the full heterodimers pKoS101 and pSeK101 were dominated by MSV-Set-like and MSV-Kom-like recombinants amongst detectably lower levels of MSV-Kom-like and MSV-Set-like recombinants respectively. This follows the strong suggestion that the more efficient mechanism of release from dimeric clones is that of replicational release from between two LIRs as opposed to random intramolecular recombination. Very low levels of recombinants with significantly large sequence replacements are apparent in Southern blot tests (Section 4.4.3.2 and Figure 4.7).

The relatively low amount of MSV-Kom-like recombinants in pKoS101-agroinoculated plants, as ascertained by restriction digestions and Southern blot tests on recombinant RF-DNA (Figure 4.7), was in line with subsequent leafhopper transmission tests. Streak disease was transmitted by pKoS101-viruliferous leafhoppers from pKoS101-agroinoculated plants with MSV-Set-like symptoms to the panel of hosts where one of the four cages (of quadrupled transmissions) developed unusually severe MSV-Kom-like symptoms on four different MSV-Set-resistant maize cultivars (Table 4.3). This indicated that leafhoppers were able to acquire and transmit MSV-Kom-like recombinants simultaneously with MSV-Set-like recombinants, despite the low levels of MSV-Kom-like recombinants in pKoS101-agroinoculated plants with distinctly MSV-Set-like streak symptoms. This is a reflection of the efficiency at which leafhoppers are able to transmit MSV, even from plants of a mixed infection where MSV-Kom is dominated by another population (in this case MSV-Set). These results are in line with the suggestion that replicational release from between two LIRs is more efficient than by random homologous intramolecular recombination and, as a consequence in this case, the latter type of recombinant is transmitted at lower levels by leafhoppers.

These results suggest a certain degree of compatibility between MSV-Kom and MSV-Set, however the partial heterodimers pKoS102 and pKoS103 - clones containing only one LIR each - were unable to produce Mastrevirus-like streak symptoms (Figure 4.4) presumably due to the lack of a second LIR. This result, along with the above, is an indication that the release of unit-length recombinant genomes from MSV-Kom/MSV-Set heterodimers is dependent on the more efficient replicational release mechanism as opposed to that of intramolecular homologous recombination. This does not rule out the possibility that unit-length recombinant genomes are released by random intramolecular homologous

recombination between the heterologous LIRs. It is also possible that sequence replacements in recombinant genomes can render themselves non-viable and/or replication deficient and thus are "sifted" out by the plant or supplanted by more viable wild-type-like recombinants with wild-type phenotypes during the growth of the host.

In light of the non-infectiousness of partial heterodimers pKoS102 and pKoS103, which only contain one LIR each, it was decided to clone heterodimeric 1.1mers such as pSeK104, pSeK106 and pKoS105; all containing two LIRs (Figure 4.3). The agroinfectivities of pSeK104 and pKoS105 were comparable to partial dimers containing two LIRs pSeK102 (1.5mer) and pKoS101 (2mer) whereas that of pSeK106 was comparable to the wild-type homodimeric 1.1mers pSet107 and pKom603. It is not understood why the agroinfectivity of the heterodimeric 1.1mer pSeK106 is as high as the wild-types. However all heterodimers produced symptoms on Jubilee sweetcorn plants six days later than those of the 1.1mer wild-types pSet107 and pKom603. This late infection may be a reflection of the duration for symptoms to appear (four days later than those of their 1.1mer counterparts with two LIRs, pSet107 and pKom603; Chapter 2) on pSet102- and pKom504-agroinoculated wild-types which only containing one LIR each. Unit-length genomes from homodimeric clones pSet102 and pKom504 can only be released by intramolecular homologous recombination since they only contain a single LIR in each (Chapter 2). Thus the six day delay for the symptoms to appear in the heterodimer 1.1mers (Figure 4.3) is presumably invoked by the less efficient intramolecular homologous recombination release mechanism. However, replicative release of unit-length recombinant genomes cannot be ruled out; and if it is occurring, the late appearance of these symptoms may be due to a degree of incompatibility between the MSV-Kom and MSV-Set genomes.

The streak disease symptoms obtained on pSeK106- and pKoS105-agroinoculated Jubilee sweetcorn seedlings were MSV-Set-like and those of pSeK104-agroinoculated seedlings were distinctly MSV-Kom-like. This is presumably due to pSeK106 and pKoS105 containing at least an entire MSV-Set genome with an additional MSV-Kom LIR in each; thus it is expected that resulting recombinants were MSV-Set-like. Similarly, it was expected that the symptoms obtained on pSeK104-agroinoculated seedlings would be MSV-Kom-like since pSeK104 contains at least an entire MSV-Kom genome with an additional MSV-Set LIR. Thus, in this case, MSV-Kom-like recombinants would be expected.

4.5.2 Sequence analyses of recombinants

Deductions from the biological results about what type of and how recombinants are released from heterodimers are largely speculative. It was thus decided to isolate, clone and

determine the partial sequences over the *Bam* HI junction site between the different genomes. The results of these are tabulated in Tables 4.4, 4.5 and 4.6. A summary of the types of recombinants obtained from these heterodimers are depicted in Figure 4.17.

The heterodimer agroinfection data indicate that some form of replicative release occurs, as only constructs with two LIRs were infectious. However, the data are also compatible with homologous recombination as a means of release, in that symptoms appeared later than with the wild-type 1.1mer clones pKom603 and pSet107. It is possible that both mechanisms contribute to the release of recombinant genomes: that is, first a partial release occurs by replication between two LIRs; and/or second, homologous recombination may occur to release a unit-length genome. This is further discussed below.

It appears that most recombinants are as a result of random homologous recombination events occurring between the two different genomes where the crossover sites have complete homology of 2bp in the V1 ORF to 29bp situated left of the hairpin loop in the LIR (Table 4.4, 4.5 and 4.6). Although a crossover site of only 2bp was identified, and can be argued to have occurred as a result of an illegitimate recombinational event between MSV-Kom and MSV-Set, the 78% homology between MSV-Kom and MSV-Set probably rules this event out as a likely explanation. The possible crossover sites are all localised in the vicinity of the *Bam* HI junction between the two genomes in the heterodimers, thus the recombinants released and isolated were all wild-type-like with sequence replacements of only 2-176bp (Figure 4.17 and Tables 4.4, 4.5 and 4.6). Since recombinants were isolated from plants as late as 3-5 weeks after agroinoculation it is likely that less-viable recombinants with supposedly larger sequence replacements may have been supplanted by the wild-type-like populations such as the recombinants obtained in this study.

A few MSV-Kom and MSV-Set wild-types were however identified (Tables 4.4, 4.5, and 4.6). For example, pSeK119 isolated from pSeK101(2mer)-agroinoculated plants, although only partially sequenced, appeared to be the progeny of an MSV-Kom wild-type genotype. Additionally, pKoS152, pKoS153 isolated from pKoS105(1.1mer)-agroinoculated plants and pSeK161, pSeK162 and pSeK163 from pSeK106-agroinoculated plants were all MSV-Set wild-types. In all these cases it appears that the crossover sites appeared to be the conserved *Sac* I or *Bam* HI sites as junctions between the two genomes in heterodimers which appear to be recombinational "hotspots" or preferred recombinational sites.

While 14% of the recombinants obtained were wild-types, <12% were found to have mixed sequence arrangements as a result of unknown recombinational events - presumably

illegitimate recombination. Although the genotypes of these supposed cloning artefacts were mostly that of MSV-Kom or MSV-Set sequences however unknown non-MSV sequences were also present such as of the clones pSeK1111, pSeK116 and pSeK117. These clones were approximately 2.7kb in size (size of Mastreviral genome), since cloned recombinants were selected on the basis of their molecular size on agarose gels and their homology to MSV-Kom and MSV-Set DNA probes in colony blot tests. Thus only 32 (i.e. 75% of the clones) were MSV-Kom/MSV-Set recombinant genotypes with MSV-Kom or MSV-Set sequence replacements. These were not further investigated since this study targeted viable whole genomes of recombinants.

It is assumed that recombinants in a population originate from unit-length genomes released from heterodimers by random homologous recombination or replicational release are subsequently replicated and infect the host. Of the 40 recombinants, including those consisting of wild-type sequences, 25 (or 62.5%) appear to be progeny originating from unit-length genomes released from the right of the hairpin loops or right of the conserved nonanucleotide in the LIRs from the heterodimers (Figures 4.8, 4.10, 4.11, and 4.17). However only 11 (or 27.5%) of the 40 appear to be recombinants originating from unit-length genomes released from the left of the hairpin loops and outside the region delimited by hairpin loops of the heterodimers (Figure 4.17). It is possible that the replicational mechanism of release - being the more efficient mechanism of release (Chapter 2) - may be playing a role leading to the recombinational events to the right of hairpin loop structures (Figure 4.17 and 4.18). However the possibility that replacements from random homologous recombinational events occurring to the left of the hairpin loops may render these genotypes less viable and thus susceptible to supplantation in the infection, cannot be ruled out. If this is indeed the case, such non-viable or less-viable or replication deficient recombinants would become dominated by more viable recombinants in a population (Bisaro, 1994).

Interestingly, three cloned recombinants had replacements from the TAATATTAC nonanucleotide sequence to the conserved *Bam* HI site which appeared to indicate that the parent unit-length genomes were probably released via replicational release. Additionally, pSeK115, pSeK1110 and pSeK118, all isolated from pSeK101(2mer)-agroinoculated plants, contained MSV-Kom sequence with the LIR from the TAATATT↓AC nonanucleotide which included additional 13 or 11 bp insertions which were homologous to the first 10 or 13 bases of the MSV-Kom sequence (Figures 4.12 and 4.13). It appears the parent unit-length genome was released via replicational release; however, due to a probable mis-recognition of the right stem-loop in the heterodimer - where termination supposedly occurs (Heyraud et

al., 1993) - the slight over-replication (11-13 bases) past the hairpin loop may have occurred (Figure 4.17 and 4.18).

Since the stems of the hairpin loops are different for MSV-Kom and MSV-Set (in this case 18 base inverted repeats; Chapter 3), perfect initiation at an hairpin and terminations at an heterologous hairpin during replicational release would result in unit-length genomes with non-complementary inverted repeats that are unable to form stems. It is for this reason that the insertions of recombinants pSeK115, pSeK1110 and pSeK118 would disrupt the 18 bp stem of the hairpin. Judging by the sequences of these insertions it appears that pSeK115 and pSeK1110 can only form a MSV-Kom 11 b inverted repeat (flanking the conserved nonanucleotide TAATATT↓AC) stem as opposed to the normal 18 bp inverted repeat of the wild-type MSV-Kom genotype (Figures 4.12). In addition, pSeK118 can only form a 9b inverted repeat (flanking the conserved nonanucleotide TAATATT↓AC) which would be half the size of that of the wild-type MSV-Kom genotype (Figure 4.13). It is likely that, due to mis-recognition of the right heterologous hairpin at termination for the release of unit-length recombinant genomes, a possible "over-replication" and subsequent production of overlapping complementary ssDNA during the resolution of circular single stranded DNA genomes, may result in the 13 or 11 bp insertions. Thus, as depicted in Figure 4.18, it is possible that the resolving events may occur beyond the termination site. Whether these are agroinfectious was not investigated here, however these are being made available for further investigation (J. Willment, pers. comm.), which potentially may reveal aspects of Rep/hairpin loop interactions. The nature of the "resolving event" is unclear: if the ssDNA is not cleaved at the second TAATATT↓AC, and the "stem" sequences are not compatible (i.e cannot base pair), then the second stem-loop may form - which will be part of the native LIR of the second of the "tandem" sequences. Since this will not allow circularisation of the ssDNA, this may result in the ssDNA being made double-stranded, after which event homologous recombination occurs in the "repeated sequence" to the right of the second stem-loop. However, this would not explain sequence repeats or insertions (11 or 13 b) such as those occurring in pSeK115, pSeK1110 and pSeK118 (Figures 4.11, 4.12 and 4.13).

The stems of the stem-loop structures - unlike the Begomo- and Curtoviruses- are not conserved in Mastreviruses, thus it is possible that the Rep protein does not readily recognise the hairpin loop of a different Mastrevirus. The possible involvement of Rep in the heterodimers would presumably be dependent on its specificity for the LIR. The full heterodimers pSeK101 and pKoS101 contain the entire C1/C2 ORFs in each of the two unit-length genomes within the clones. Presumably the Rep proteins would be specific to the LIRs from their own genomes. In pseudorecombination experiments Dry *et al.* (1997)

showed that TLCV sat-DNA can be supported by other taxonomically distinct geminiviruses such as TYLCV, ACMV and BCTV, although the Rep protein amino acid sequence identities of TYLCV, ACMV and BCTV with TLCV are only 79%, 78% and 61% respectively. The sequence identity of the MSV-Kom and MSV-Set Rep proteins is 82% (Chapter 3). Whether this is sufficient homology between the proteins or the LIRs for transreplication is not clear here, however this is investigated further in Chapter 5 where the pKeSeK agroinfectious clone is reported. Gilbertson *et al.* (1993) reported viable pseudorecombinants of ToMoV and BDMV, whose IRs shared 87% similarity, which allowed the trans-replication of the two viruses. Different IRs with <75% nucleotide identities are apparently not able to be recognised by a single Rep (Gilbertson *et al.*, 1991; Lazarowitz *et al.,* 1992; Stanley and Gay, 1983).

The possibility that different Reps were responsible for nicking their respectively homologous LIRs in full dimers or partial dimers with two complete Rep genes was eliminated by cloning heterodimeric 1.1mers pSeK104 and pKoS105. Of 8 recombinants generated from pSeK104 and pKoS105 agroinoculations examined, two (pKoS152 and pKoS153; Table 4.6) appeared to consist of wild-type MSV-Set genotypes, presumably due to the crossover events occurring at the conserved *Bam* HI restriction site , located to the right of the LIR (Figures 4.8). The recombinant pKoS151 generated from pKoS105 agroinoculations was found to have the two base pair replacement in an MSV-Set genome identical to those of pKoS001, pKoS002 and pKoS103 generated from pKoS101 agroinoculations (Table 4.6). All five progeny recombinants generated from pSeK104 agroinoculations (pSeK141-145; Table 4.4) contained 105 bp sequence replacements in the LIR at a common recombinational site right of the stem-loop region (Figure 4.10). From these results it appears that trans-replicational release does not occur readily during the agroinfectious process between MSV-Kom and MSV-Set. Thus it appears that in these cases, random intramolecular recombination may be the prevalent mechanism releasing unit-length recombinants. Other "hotspot" or preferred MSV-Kom or MSV-Set recombinational sites are identified aligned and depicted in Figures 4.9, 4.11, 4.13 and 4.14.

pSeK106, is another 1.1mer containing an entire MSV-Set with an MSV-Set LIR cloned to the right of the ORFs and an MSV-Kom LIR cloned to the right of the MSV-Set ORFs. Thus the MSV-Kom C1/C2 Rep promoter would drive the transcription of the MSV-Set C1/C2 Rep protein if replicational release occurred during agroinoculation of pSeK106. Since the progeny pSeK161, pSeK162 and pSeK163 contained wild-type sequence, it is concluded that the conserved 9 bp sequence containing the *Sac* I of MSV-Kom and MSV-Set is also a preferred crossover site (mentioned above).

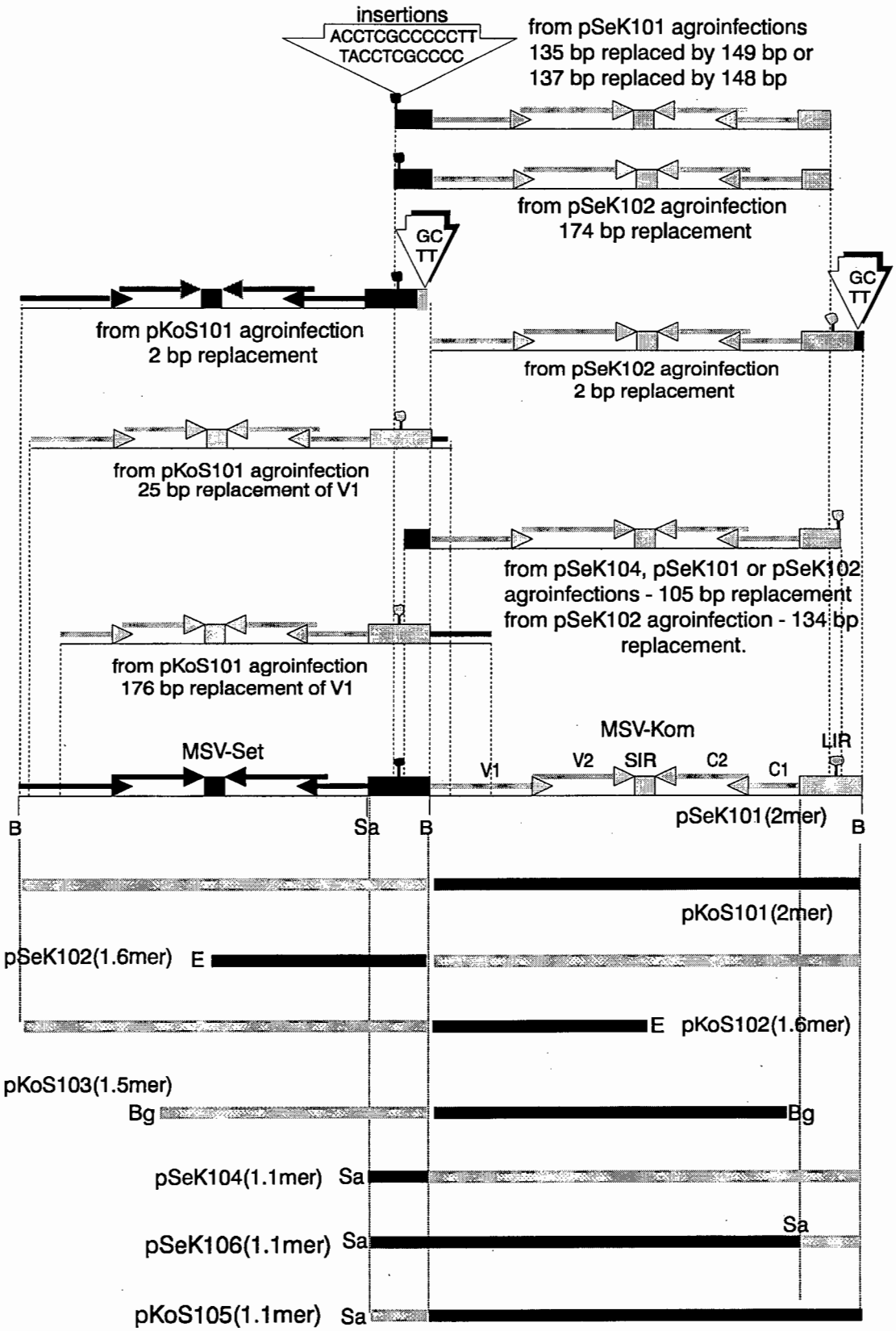

Figure 4.17 Types of recombinants from heterodimer-agroinoculated plants

Recombinants isolated from heterodimer-agroinoculated plants were partially sequenced and are depicted above. The heterodimers used are depicted below. Black regions indicate MSV-Set sequence and shaded regions indicate MSV-Kom sequence. Restriction endonuclease sites used to clone heterodimers, B = *Bam* HI; Bg = *Bgl* II and Sa = *Sac* I. (Tables 4.4, 4.5 and 4.6).

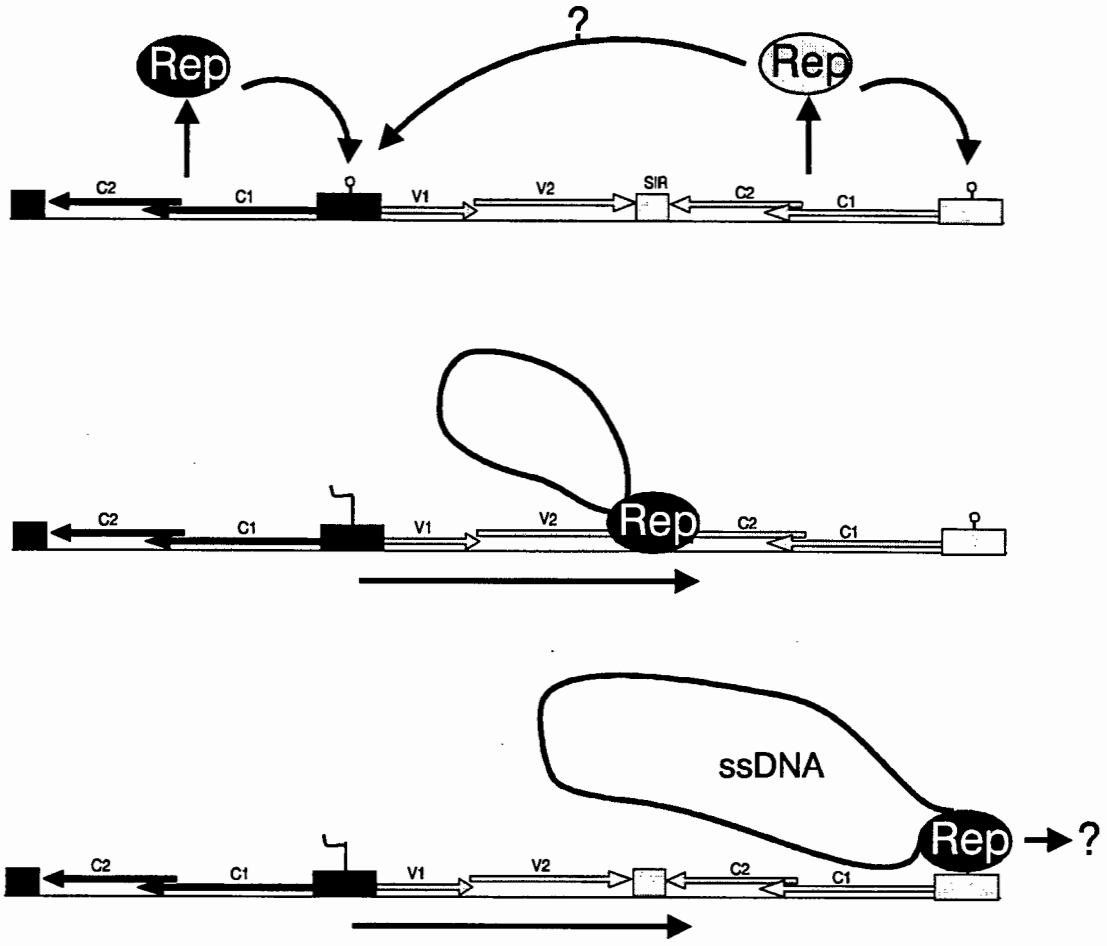

Figure 4.18 Model for the replicational release of recombinants from heterodimers

Top: Whether Rep from one genome is able to initiate replication in another heterologous LIR is investigated in Chapter 5. It is more likely that Rep of one genome will recognise its own LIR and initiate replication.

Middle: Rep covalently binds the 3'OH end of nicked strand while it interacts with the host polymerase and displaces ssDNA from RF-DNA during RCR. Host polymerase is reponsible for replication.

Bottom: Rep probably does not recognise the right hairpin loop thus replication may continue and circular ssDNA is extended beyond the right misrecognised hairpin loop resulting in sequences mapped to the right of the second hairpin loop appearing in recombinant progeny.

4.6 CONCLUSIONS

Since some recombinants such as the V1 ORF replacement appear to have only a 2bp homologous crossover site it is likely that several different genotypes may have been released from various regions of the heterodimers by random homologous recombinational events. It appears that large regions of complete homology are not required, however the

78% homology between MSV-Kom and MSV-Set may aid in the crossover events. Many recombinants released via this mechanism are presumably later supplanted by viable recombinants. However recombinants with sequence replacements of up to 176bp were generated and replicated in plant hosts. These can be used to determine how these replacements affect the biology of the virus in agroinoculation tests (investigated in Chapter 5). This is a potentially powerful tool to generate recombinants with replacements that are not possible to make or clone in the laboratory due to restriction recognition site limitations. Since recombinant progeny was isolated from 3-5 week old agroinoculated plants, only more viable infectious recombinants were targeted in these experiments. This limitation can presumably be overcome if recombinants were isolated from heterodimer-infected protoplasts which would generate initial and early recombinant unit-length genomes and thus a wide range of recombinants could potentially be obtained (Heyraud *et al.*, 1993).

It would be interesting to investigate the recombinants generated from heterodimers by using other restriction junctions such as the conserved *Sal* I or *Ssp* I sites in MSV-Kom and MSV-Set. Also investigations into intramolecular recombination between viruses with sequence identities lower than that of MSV-Kom and MSV-Set (for example MSV-Kom or MSV-Set and PanSV-Kar with sequence identity of only 60%) would presumably generate recombinants that may be useful in future studies in an effort to determine more closely what genomic regions are responsible for the pathogenicity of these Mastreviruses. The recombinants pKoS005 with a 176 bp replacement in the V1 ORF, pSeK142 with a 105 bp replacement in the LIR and the recombinants with only 2bp replacements are further investigated in Chapter 5.

CHAPTER 5

AGROINOCULATIONS OF MSV-Set/MSV-Kom RECOMBINANTS AND OF A CHIMAERIC CONSTRUCT WITH HETEROLOGOUS LIRS

5.1 SUMMARY

Cloned MSV-Kom/MSV-Set recombinant RF-DNAs pSeK142 and pKoS005, generated by pSeK104 and pKoS101 MSV-Kom/MSV-Set construct agroinoculations reported in Chapter 4, were selected for determination of their infectivity and biological characteristics in agroinoculation and leafhopper transmission tests. pSeK141 contained the MSV-Kom RF-DNA with 105 bp replaced by MSV-Set cognate sequence in the LIR from the right of the hairpin loop to the *Bam* HI site on the right side of the start V1 start codon, effectively replacing the DNA bending locus normally associated with origins of replication. pKoS005 contained the MSV-Kom sequence with 176 bp replaced by MSV-Set cognate sequence in the V1 ORF from the *Bam* HI site into the V1 ORF, effectively replacing 57% of the N-terminus of the MP. The agroinfectivities, symptomatology and host ranges were determined for these chimaeras in homodimeric agroinoculations and subsequent leafhopper transmission tests. Although the host ranges and leafhopper transmissibility were not altered with respect to MSV-Kom, the symptoms were attenuated and the agroinfectivities were significantly reduced. It appeared that these attenuations are attributed to the "fitness" of the chimaeric genomes containing sequences that are not specific to the rest of the genome in interactions and functions. The implications of these results are discussed.

pSeK001, a member of a large group of recombinants, analysis of which identified a preferred homologous recombinational crossover site left of the V1 start codon and containing only 2 bp replacements, was used in agroinoculations in an effort to determine how the preferred crossover site was influencing the release of unit-length genomes from MSV-Kom/MSV-Set heterodimers during agroinoculations. pGC/TT2 and pTT/GC2 are MSV-Kom 1.1mers containing two LIRs with a MSV-Set GC in the left LIR and and an MSV-Kom TT on the cognate site in the right LIR, and *vice versa* as indicated by the plasmid designations. The agroinfectivity and symptomatology were comparable to MSV-Kom wild-type. Isolated and partially sequenced progeny RF-DNAs indicated that one in ten releases appear to be mediated by homologous recombination at the 14bp preferred crossover sequence.

pKeSeK2, a heterologous 1.1mer containing two MSV-Kom LIRs flanking the four ORFs of MSV-Set, was found to be agroinfectious only in Jubilee plants growing under low light intensities. Symptom severity and agroinfection rates (10% as opposed to 79 and 78% for MSV-Kom and MSV-Set wild-types) and the duration of streak symptom development were greatly reduced. These results as well as PCR detections of replicating KeSeK chimaeras, indicated that the MSV-Set Rep protein is able to recognise and initiate *trans*-replication in

the MSV-Kom LIR sequence. The implications of this result with respect to possibilities for recombination between MSV strains are discussed.

5.2 INTRODUCTION AND RATIONALE

A variety of MSV-Kom/MSV-Set recombinants were generated by agroinoculations of partial and full heterodimers of MSV-Kom and MSV-Set (Chapter 4). Few cloned progeny RF-DNAs consisted of wild-type sequence with apparently no sequence replacements presumably released from conserved restriction sites (used for cloning heterodimers) during agroinoculations. Most recombinant progeny however, had sequence replacements ranging from 2 bases to 174 bases in the LIR and 25 or 176 bp replacements in the V1 ORF. Recombinants generated and replicating in heterodimer-agroinoculated plants were members of mixed populations of numerous different recombinant types which are probably in competition for replicational and systemic infection mechanism resources made available in the host (Bisaro, 1994; Chapter 1). Presumably unit-length genomes were randomly released from different regions of the heterodimers, however it appeared that most infectious recombinants were released from crossover regions located close to the junction of the two genomes in the heterodimer (Figure 4.17). Thus progeny recombinants had relatively small sequence replacements of up to 176 bp. It is possible that many of these recombinants may have been replication deficient, non-viable or non-infectious; probably trans-replicated by wild-type or more wild-type-like recombinant progeny.

Gene replacement studies is a potentially powerful method of determining the role of functions of ORFs, as in the elegant demonstration by Briddon *et al.* (1990), in which the CP ORF of ACMV (Begomovirus) was replaced by that of BCTV (Curtovirus). The chimaeric ACMV, although causing milder symptoms, was transmitted by the natural vector of BCTV. Such experiments cannot always be performed due to conserved restriction site limitations or sequence differences. Thus, as demonstrated in Chapter 4, distinct Mastreviruses with different sequences (78% identity) and biological characteristics (Chapters 3 and 2 respectively) can be used in agroinoculations of heterodimers to generate recombinants with sequence replacements that have retained their biological activities. How these replacements have affected the biological characters as opposed to the wild-types, or whether these were viable was not conclusively determined. Thus, cloned recombinants recovered from agroinfected plants were dimerised and investigated in this study to investigate their viability. This chapter reports on investigations of the biological activities of

selected recombinants pSeK142 and pKoS005 in agroinoculation tests (Tables 4.4 and 4.6 respectively).

Most plant virus MPs studied so far have been shown to be involved in mediating short-distance cell-to-cell spread and many have been shown to increase the size exclusion limit of plasmodesmata (Hull, 1989). Long distance movement however, occurs in the vasculature, seemingly - in most cases - through the phloem tissue. Since geminiviruses, unlike RNA plant viruses, replicate within the host nucleus it is expected that the mechanisms of geminiviral movement may differ from other plant viruses, due to the additional nuclear membrane barrier the geminivirus must get through. V1 mutant experiments confirmed that the MP was necessary for infection, but the mutants could be rescued by wild-types (Boulton, *et al.*, 1989), suggesting that the MP was not absolutely required for replication but was involved in an aspect of symptom development. Lazarowitz *et al.*, (1989) confirmed that the MP was essential for movement. Inoculation of protoplasts with MSV V1 gene mutants which are unable to produce systemic infections in plants resulted in DNA replication and encapsidation indistinguishable from that obtained with wild-types which implicated the MP in movement (Boulton *et al.*, 1993). However ssDNA was not detected following the inoculation of V2 CP mutants, thus it was suggested that the inability of these mutants to spread may be due to a lack of ssDNA, CP or a combination of the two. More recent unpublished results of MP replacement studies in MSV-Kom and MSV-Set conducted by E. van der Walt (unpublished) in this laboratory indicate that MPs interact with CPs in the infection process.

pKoS005, a recombinant generated from pKoS101 agroinoculations (Chapter 4), has a 176 bp V1 ORF region replaced by cognate MSV-Set sequence ranging from the *Bam* HI/ATG site into approximately 57% of the V1 ORF (Table 4.6). This 176 bp replacement effectively replaces 57 of the N-terminus amino acid residues for that of MSV-Set. The hydrophobicity profile of the putative chimaeric MP remains similar to that of the wild-type MPs, suggesting that the chimaeric MP may still be functional (Chapter 4). How this replacement biologically affects MSV-Kom or MSV-Set may reveal the function of the N-terminus of the MP. This is investigated in agroinoculation experiments here by comparing the biological characters of pKoS005 with those of the wild-type homodimeric *Agrobacterium* clones.

A putative DNA bending locus containing dT.dA tracts, first identified by Suàrez-Lòpez *et al.*, (1995) in WDV, was identified in MSV-Set and MSV-Kom in Chapter 3. This locus is situated in the LIR, right of the hairpin loop. These tracts have been shown to be associated with origins of DNA replication and are necessary for origin activity in a number of eukaryote

or phage systems (Crothers *et al.*, 1990). pSeK142 was generated in pSeK104-agroinoculated Jubilee sweetcorn plants and has 105bp of the LIR replaced by the cognate MSV-Set sequence: this stretches from the conserved *Apa* I site right of the stem-loop to the *Bam* HI/ATG upstream of the V1 ORF (Chapter 4). This replacement effectively replaces the MSV-Kom DNA bending region with that of MSV-Set, which has significantly different numbers of dT.dA tracts which are arranged differently (Figure 3.11). How this replacement affects the replication of MSV-Kom is also investigated in this Chapter - as for pSeK142 - by agroinoculation tests of the recombinant pKoS005 and comparing the biological characters with those of the wild-types.

Additionally, taking advantage of the conserved *Sac* I and *Bam* HI sites situated on either side of the LIR in MSV-Kom and MSV-Set, heterologous 1.1mers containing the four MSV-Set ORFs delimited by the LIRs of MSV-Kom was also investigated. The agroinoculation of this clone should effectively release chimaeric unit-length genomes - by random homologous recombination or replicational release between the MSV-Kom LIR copies - that consist of the four ORFs of MSV-Set with the LIR replaced with that of MSV-Kom. This non-coding region contains the origin of replication (with the hairpin loop structure containing the conserved nonanucleotide TAATATTAC) with Rep binding sites that are presumably specific for the Rep protein encoded by the same genome (Chapter 3). This region is also involved in the regulation of transcription of the virion and complementary sense ORFs (Chapters 1 and 3). How this LIR replacement affects the biological character of MSV-Set was investigated in agroinoculation tests. The implications of this result with respect to trans-replication and recombination between MSV-Kom and MSV-Set is discussed. In addition the results of this would indicate whether heterologous Rep proteins compete in the initiation of replicational release from MSV-Kom/MSV-Set heterodimers used in agroinoculation tests in Chapter 4 (Figure 4.18).

A surprisingly large proportion of the recombinants obtained from MSV-Kom/MSV-Set heterodimers constituted only a 2 bp replacement just left of the ATG/*Bam* HI site right of the LIR. The probable crossover site was apparently a 14 bp sequence left of the *Bam* HI site conserved in MSV-Kom and MSV-Set which includes a 7 bp stretch conserved in most geminiviruses (Chapter 4). The function of this conserved region is not known. Since eight of 10 recombinant progeny isolated from pSeK102-agroinoculated plants and three of 10 from pKoS101-agroinoculated plants, it is assumed that this region in both MSV-Kom and MSV-Set is a preferred recombinational site or "hotspot". Alternatively the two bp difference may not alter the pathogenicity of MSV-Kom or MSV-Set, and the genomes become dominant in recombinant populations (Bisaro, 1994; Chapter 1). The effects of these 2bp

replacements in agroinoculations and the biological effects thereof were investigated for this Chapter. The sequence data of progeny from these agroinoculations were analysed and the implications with respect to what the possible role of this preferred recombinational site may hold are discussed.

5.3 MATERIALS AND METHODS

5.3.1 Construction of agroinfectious MSV-Set/MSV-Kom recombinant clones

Homodimeric constructs of two recombinant clones - pKoS005 and pSeK141, that were isolated and cloned from symptomatic pKoS101- and pSeK104-agroinoculated sweetcorn plants respectively (Chapter 4) - were made in a similar fashion to the 1.1mer homo- and heterodimeric constructs as described in 4.3.2 and 2.3.3. pKoS005 has 176 bases of the V1 ORF region of MSV-Kom replaced by MSV-Set sequence starting from the ATG start codon at the *Bam* HI site and, pSeK141 has 105 bases of the MSV-Set LIR from the conserved region right of the hairpin loop structure at the *Apa* I site to the *Bam* HI site, at the ATG start codon of V1, replacing cognate MSV-Kom sequence. *Sac* I-digested pSeK141 was religated to excise the ORFs between the *Sac* I site in the MCS of pUC18 and the *Sac* I site on the left of the LIR of the pSeK141 to give pSeK421 (Appendices A1,A3-A6 and A11). *Bam* HI/*Pst* I double-digested pKoS005 and pSeK141 were "shotgun" ligated with *Bam* HI-digested pKom601 and pSeK421 respectively, to obtain homologous 1.1mers of the recombinants in pUC18 (Appendices A1, A3-A6 and A11). The extra *Pst* I digests of pKoS005 and pSeK141 rendered the pUC18 fragment less likely to interfere in the ligation. White *E.coli* colonies were selected on LA + ampicillin/X-gal/IPTG plates and plasmid minipreparations were selected on the basis of *Bam* HI, *Apa* I and *Xho* I digestions that distinguished the desired clones. Maxipreparations of the desired clones were made and designated pKoS025 and pSeK422 respectively. Xba I/Eco RI/Pst I-digested pKoS025 and pSeK422 were ligated with Xba I/Eco RI/Sac I-digested pBI121 respectively in "shotgun" ligations to replace the GUS/NOS-terminator with the 1.1mers (Appendices A1, A3-A11). The extra *Pst* I digestions of pKoS025 and pSeK422 rendered the pUC18 fragment less likely to interfere in their respective ligations and similarly, the extra *Sac* I digestion of pBI121 rendered the GUS/NOS terminator fragment less likely to interfere in the ligations. Plasmid minipreparations were prepared from *E. coli* colonies that were selected on the basis of colony blotting and hybridisations using DIG-labelled MSV-Kom DNA. Plasmids were selected on the basis of *Apa* I, *Bam* HI, *Xba* I, *Eco* RI and *Xba* I/*Eco* RI digestions. Maxipreparations were prepared and the 1.1mers constructed in pBI121 were designated

pKoS035 and pSeK423 respectively (Figure 5.1). Agroinfections and leafhopper transmissions were performed as described in 2.3.4, 2.3.5 and 2.3.6.

Two heterologous 1.1mers were constructed from the recombinants with only a GC or TT two base switch, (MSV-Set nt pos. 149 and 150 or MSV-Kom nt pos. 147 and 148 respectively), from the pKoS101, pSeK101 and pSeK102 agroinoculated plants (Tables 4.4, 4.5 and 4.6). Two configurations were constructed: one with the TT on the left LIR and GC on the right LIR and *vice versa,* as depicted in Figure 5.1. These constructs were again made similarly to the homologous and heterologous 1.1mers as described earlier. A pSeK001 recombinant with a GC in place of the MSV-Kom wild-type's TT (three bases left of the *Bam* HI site), was digested with *Sac* I to excise the ORFs between the *Sac* I sites between the MCS of pUC18 and just left of the LIR, and then religated to form pSeK/GC (Appendices A1, A3-A6 and A11). pKom601 with the wild-type TT (three bases left of the *Bam* HI site), was digested with *Bam* HI and ligated with *Bam* HI/*Pst* I-digested pSeK001 to form the 1.1mer, pTT/GC1. Similarly, *Bam* HI-digested pSeK/GC was ligated with *Bam* HI/*Pst* I-digested pKom500 (2.2.1) to form the 1.1mer construct, pGC/TT1 (Appendices A1, A3-A6 and A11). Plasmids were selected on the basis of digestions (Appendix A1 and A3) as described in 4.2.2 and 2.2.3. pTT/GC1 and pGC/TT1 were sequenced from both directions using standard reverse and forward M13 sequencing primers to confirm GC and TT were correctly placed (Appendix B). The recombinant hetero-1.1mers were cloned into pBI121 as decribed above (Appendices A1, A3-A11), and in 4.2.2 and the constructs were designated pGC/TT2 and pTT/GC2 respectively (Figure 5.1). Agroinfections were performed as described in 2.2.4 and 2.2.6 and agroinfectivity rates were compared with the wild-types. Recombinant genomes were isolated, cloned and sequenced as described in 4.3.3 and 4.3.4. Recombinant genomes were designated pTT/GC21, pTT/GC22...pTT/GC25 and pGC/TT21, pGC/TT22...pGC/TT25.

5.3.2 Construction of a chimaeric agroinfectious clone - pKeSeK2

In a similar fashion to that described in 5.3.1, 4.3.2 and 2.3.3, a chimaeric clone was constructed designated pKeSeK2 (Figure 5.1). The heterologous 1.1mer constructed in pUC18, pKoS104 (4.3.2), was double-digested with *Sac* I and *Pst* I and "shotgun" ligated with *Sac* I-digested pKom601, the construct with the MSV-Kom LIR, in a similar fashion to that described above and in 4.3.2 (Appendices A1, A3-A6 and A11) to form the desired clone, pKeSeK1. This chimaeric 1.1mer construct was constituted of two MSV-Kom LIRs flanking the ORFs of MSV-Set. This was then inserted into pBI121 in place of the GUS/NOS-terminator as described in 5.3.1 and 4.3.2. Agroinfections and leafhopper transmissions were performed as described in 2.3.4, 2.3.5 and 2.3.6.

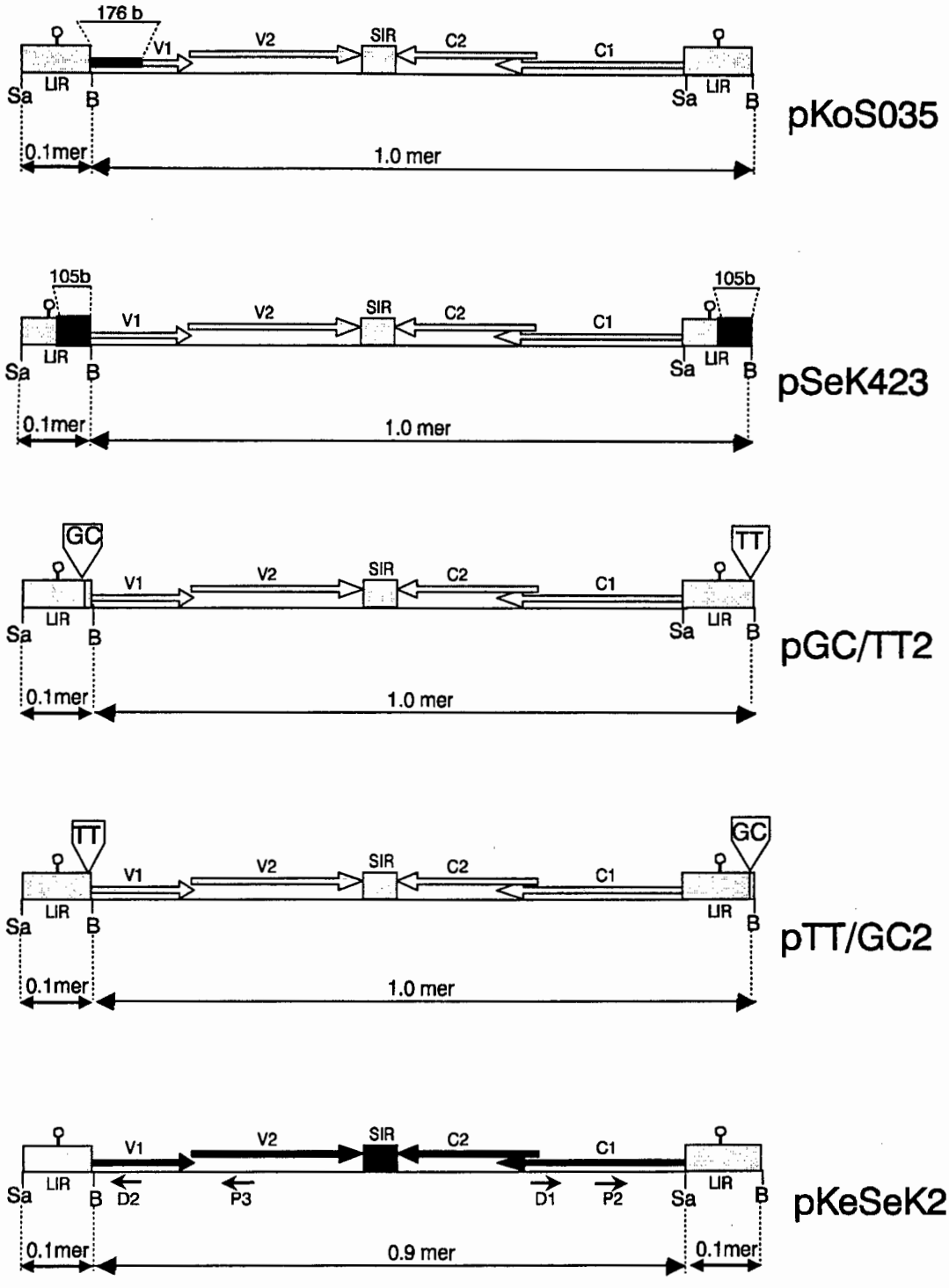

Figure 5.1 Chimaeric, homologous and heterologous 1.1mers constructed in pBI121

V1 = MP ORF; V2 = CP ORF; C1 + C2 = Rep; LIR = large intergenic region with stem-loop structure; SIR = small intergenic region; Sa = *Sac* I; B = *Bam* HI; GC and TT are the nucleotides three bases left of the Bam HI site which are specific to MSV-Set and MSV-Kom respectively; shaded regions are MSV-Kom sequences and black regions are MSV-Set sequences. P1 and P2 are MSV-Set-specific primers for the PCR amplification of circular RF-DNA of MSV-Set or the KeSeK chimaera (5.3.3; Table 5.1). D1 and D2 are degenerate primers used for the amplification of circular RF-DNA of African Mastreviruses (5.3.3; Table 5.1).

5.3.3 PCR detection of KeSeK chimaeric RF-DNA

Since the streak symptoms of the pKeSeK2 recombinant were very mild (Figure 5.5) it was thought that PCR detection of the chimaeric RF-DNA in the total DNA extraction of plant DNA would be more effective. Total plant DNA extraction was performed as decribed in Appendix A12. Thermostable *Taq* polymerase, nucleotides and reaction buffer were all obtained from Boehringer-Mannheim. A reaction master mix was made by adding the reagents into an eppendorf tube in the order: 730µl water, 85µl 10x reaction buffer with $MgCl_2$, 4.5µl of 100mM of each of two degenerate primers, D1 and D2 or P2 and P3 (Table 5.1); 8.5µl 25mM dNTPs and 0.2µl of 5U/µl Taq polymerase. Primers D1 (T_m = 50°C) and D2 (T_m = 52°C) amplify 1.3kb from all known African streak viruses, including MSV, PanSV and SSV, from the C2 ORF through the LIR to the V1 ORF (E.P. Rybicki, unpublished). MSV-Set sequencing primers P2 (T_m = 54°C) and P3 (T_m = 56°C) specifically amplify a 1.7kb fragment of MSV-Set, and were also used to amplify the KeSeK chimaeric RF-DNA. 49µl of the mix were aliquoted into thin walled PCR tubes that fit the Hybaid™ Omni E Thermocycler® with a hot top. 1µl of 50ng/µl, 5.0ng/µl or 0.5ng/µl total plant DNA extracts were added to the mixes to obtain a final volume of 50µl. Thermal cycling started with 60 seconds at 94°C for initial denaturation, followed by 30 denaturing/annealing/extension cycles of 45 seconds at 93°C, 30 seconds at 54°C and 90 seconds at 72°C respectively. Final extension was set at 72°C for 200 seconds. 10µl of amplified DNA with loading buffer were run on 0.8% TBE agarose gels in TBE containing 20ng/µl ethidium bromide (Sambrook, *et al.* 1989). Pictures of gels on a UV transilluminator were recorded using the UVP™ GDS image capture apparatus and computer programme. KeSeK PCR products were subjected to *Bgl* II and/or *Nco* I restriction digests to detect whether recombinant KeSeK viral genomes were amplified rather than the wild-type viruses. PCR products were digested with *Bgl* II and *Kpn* I (common sites flanking the LIR) and cloned into the *Bam* HI and *Kpn* I MCS restriction sites of pUC18 (Appendices A1, A3-A6 and A11). Three cloned PCR products were designated pKeSeK21, pKeSeK22 and pKeSeK23 and sequenced on the ALFexpress automated sequencer (Appendix B) to confirm the successful agroinfection of pKeSeK2.

Table 5.1: Primers used for detecting the KeSeK chimaeric RF-DNA

Primer name	Size (bases)	Sequence (5'-3')	Genome location	for viruses	PCR products
P2	20mer	CTCGAAGTACTGAAGTTTAG	2008-2027 2006-2025	MSV-Set KeSeK	1729 bases (MSV-Set)
P3	20mer	ACTTGTGGAAGTCAACGTTC	1016-1035 1014-1033	MSV-Set KeSeK	1714 bases (KeSeK)
D1	19mer	TTGG**V**CCG**MV**GATGTA**S**AG	1746-1764 1744-1762 1742-1760	MSV-Kom MSV-Set KeSeK	1306 bases (MSV-Kom)
D2	18mer	CGGAGGAGCTGA**HM**TTGG	343-360 342-359 343-361	MSV-Kom MSV-Set KeSeK	1321 bases (MSV-Set) 1306 bases (KeSeK)

The sequences of the degenerate primers, D1 and D2 follow the IUPAC degenerate codes as described by the DNAMAN computer package: **V** = A, C or G; **M** = A or C; **S** = C or G; **H** = A, C or T. D1 and D2 were obtained from Prof. E. Rybicki for detecting African streak viruses including MSV, PanSV and SSV (unpublished) . P2 and P3 were designed to complete the MSV-Set sequence (see 3.3.1).

5.4 RESULTS

5.4.1 Agroinfectivities of recombinant and chimaeric clones

All recombinants that were made were agroinfectious, but some with reduced agroinfection rates compared to wild-type pKom603 and pSet107 *Agrobacterium* clones. Agroinfectivities and agroinfectivity profiles are summarised in Table 5.2 and Figure 5.2 respectively. The agroinfection rates of pTT/GC2 and pGC/TT2 were not significantly reduced at 92% and 88%, compared to the wild-types pKom603 and pSet107, which were 100% and 86% respectively. In contrast however, the agroinfectious recombinant 1.1mer pKoS035, which has a 176 b V1 ORF replacement was found to have a significantly lower agroinfectivity than the wild-type counterparts, at 79% as opposed to 100% and 86% for pKom603 and pSet107 respectively (Figure 5.2).

In a separate agroinfection experiments pSeK423 with a 105 bp replacement of the LIR had agroinfection rates of 68%, 29% and 59% while wild-type pKom603 had agroinfection rates of 67%, 79% and 92% respectively (Table 5.1). pSeK423 therefore seemed to have a reduced agroinfection rate (Figure 5.2).

The pKeSeK2 chimaeric construct with the MSV-Set ORFs flanked by MSV-Kom LIRs (Figure 5.1) was only weakly agroinfectious and produced very mild streak symptoms more than three weeks after agroinoculation. The agroinfectivity rate of pKeSeK2 was initially 0% in a preliminary agroinoculation test, but when agroinoculations were repeated in an experiment where the plants were fortuitously placed in lower light intensities due to a lack of growth room space, a 10% agroinfection rate was obtained. Lower light intensities appeared to stress the sweetcorn plants, and reduced their growth rate. The stressed plants apparently allowed the severely attenuated KeSeK to replicate and display very mild symptoms. Symptoms were not detectable on the younger leaves (i.e. fourth leaf onwards, Figure 5.5).

5.4.2 Symptoms and host ranges of recombinants and pKeSeK2

Extensive host range tests were performed using pKom603-, pSet107-, pKoS035- and pSeK423-agroinoculated plants and virus-free leafhoppers to ascertain whether the host ranges and symptoms differed between the recombinant viruses and the wild-types. pKoS035 is a homo-1.1mer constructed from the pKoS005 unit-length genome of 176 b MSV-Kom sequence replaced by MSV-Set sequence in the V1 or MP ORF. pSeK423 is a homo-1.1mer constructed from the pSeK142 unit-length genome with MSV-Kom sequence and 105 b in the LIR replaced by homologous MSV-Set sequence. The host ranges

remained wild-type-like (i.e. MSV-Kom-like) except the symptoms were significantly attenuated in both Jubilee sweetcorn and popcorn (Figure 5.2 and Figure 5.3A). The host range results of these leafhopper transmission tests are summarised in Tables 5.3 and 5.4. More dramatic attenuations were observed in resistant or tolerant maize cv.s such as PAN 6141, PAN 6099, PAN 6191 and PAN 6552 where the symptoms were greatly attenuated or even absent on some tolerant maize lines compared to the MSV-Kom wild-type infected counterparts (Figure 5.4)

Table 5.2 Agroinfectivity rates of recombinants and the pKeSeK2 clone

Clone	Experiment I	Experiment II	Experiment III	Experiment IV
pKom603	19/19 100%	10/15 67%	19/24 79%	23/25 92%
pSet107	19/22 86%	12/19 63%	14/18 78%	20/25 80%
pGC/TT2	38/43 88%	na	na	na
pTT/GC2	35/38 92%	na	na	na
pKoS035	30/38 79%	na	na	na
pSeK423	na	21/31 68%	12/42 29%	19/32 59%
pKeSeK2	na	0/13 0%	5/50 10%	na

Values indicate the number of Jubilee sweetcorn plants infected out of the number of seedlings that survived the injection. 25 seedlings were agroinoculated for pKom603 and pSet107 in all experiments and for pKeSeK2 agroinoculation in experiment II. 50 seedlings were agroinoculated in all other instances. na = not applicable.

pKeSeK2 agroinoculated sweetcorn plants with streak symptoms were used in carefully controlled leafhopper transmission tests. Plants with pKeSeK2 streak symptoms were the only plants initially placed in the cages so that the leafhoppers had no choice but to feed on the infected plants in an attempt to mediate acquisition and obtain viruliferous leafhoppers. It was noted that the leafhoppers fed from regions of the leaves which had streak symptoms. The plants tested for susceptibility included Jubilee sweetcorn, popcorn, PAN 6552, Clipper barley, SST44 wheat, STT66 wheat, and Festiquay wheat. The leafhoppers were allowed to feed in the cages for up to three weeks. No symptoms were observed in any of the plants exposed to "viruliferous" leafhoppers.

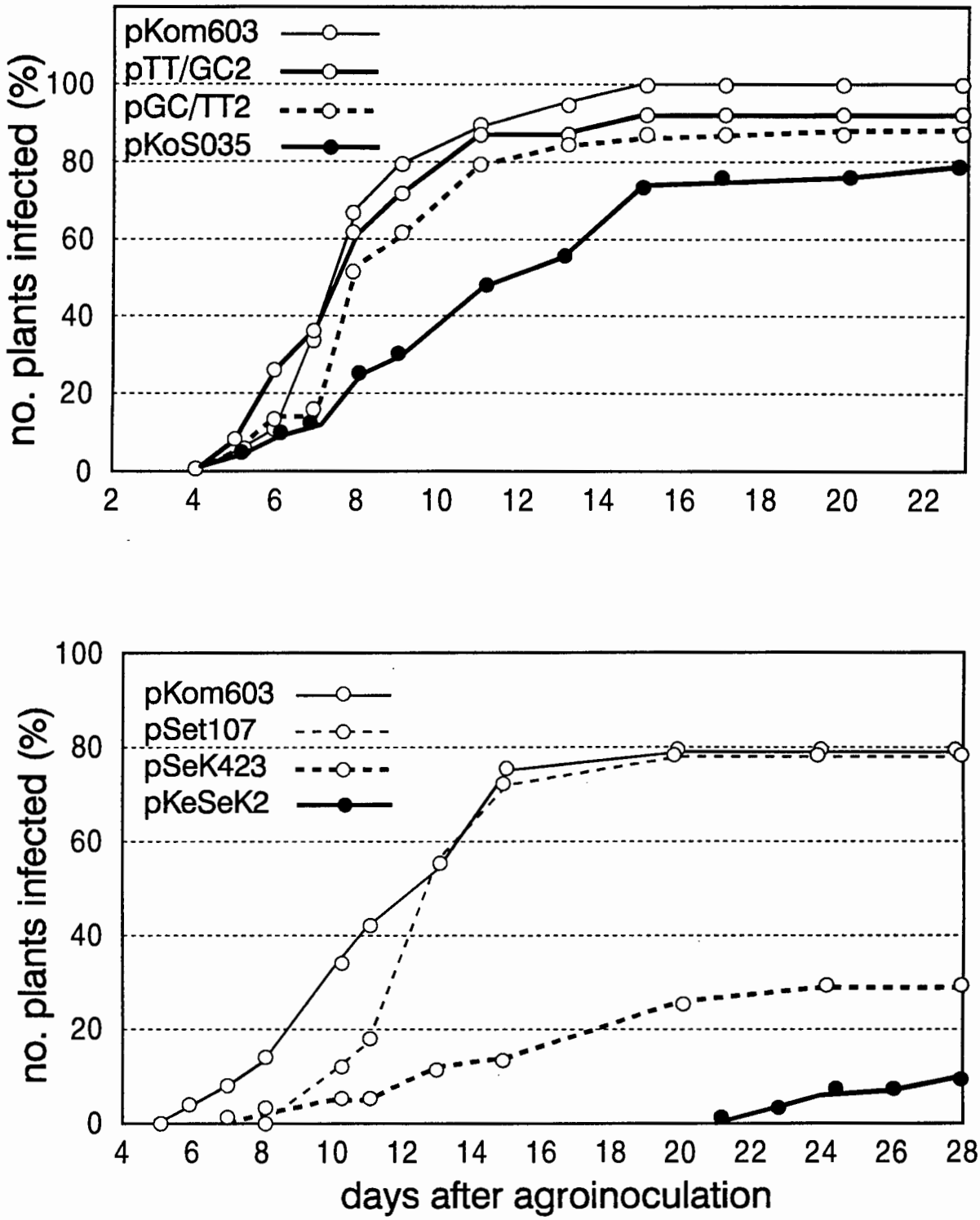

Figure 5.2 Agroinfectivities of homodimeric recombinant clones and pKeSeK2

Top depicts the relative agroinfectivity rates of pTT/GC2, pGC/TT2 and the recombinant clone, pKoS035. Bottom depicts the relative rates of agroinfectivity of pSeK423 and the chimaera, pKeSeK2 in a separate experiment. These agroinoculations were not repeated; however, the wild-type pKom603 and pSet107 agroinfectivity profiles remained similar to those in previous experiments.

Table 5.3 Maize host ranges and symptoms of recombinants

Cultivar	pSet107	pKom603	pSeK423	pKoS035
Jubilee[a]	++[+]	+++[+]	++[+]	+++
Popcorn[b]	++	+++	++	+++
Witplat	++	+++	++	+++
Vaalharts wit	-	+++	[+]	+
Vaalharts geel	-	+++	+	[+]
PNR 6552	-	+++	+	++
PNR 7563	-	+++	+	++
PAN 6043	-	+++	++	++[+]
PAN 6099	-	+++	+	-
PAN 6133	-	+++	++	+++
PAN 6141	-	+++	++	+
PAN 6166	-	+++	+	+++
PAN 6191	-	+++	-/+	-/+
PAN 6195	-	+++	-	+
PAN 6363	-	+++	+	++
PAN 6364	-	+++	+	+
PAN 6462	-	+++	++	na
PAN 6473	-	+++	++	+++
PAN 6479	-	+++	+	+
PAN 6480	-	+++	++	+++
PAN 6481	-	+++	++	++
PAN 6496	-	+++	+	+
PAN 6528	-	+++	++	++
PAN 6549	-	+++	++	++
PAN 6552	-	+++	+	-/+
PAN 6564	-	+++	++	++
PAN 6578	-	+++	++	+++

[a] Sweetcorn variety; [b] purchased at local supermarket in Cape Town; PAN seed obtained from Pannar Seed. PNR seed obtained earlier from Pannar Seed; na = leafhopper transmissions not performed. - = no streak symptoms; + = mild stippled streak and recovery(-/+); ++ = moderate stippled to continuous streak; +++ = severe continuous streak and stunting and/or distortion of growth and death(+++[+]).

Table 5.4 Wheat host ranges of recombinants

Cultivar	pSet107	pKom603	pSeK423	pKoS035
Clipper[d]	+++[+]	+++	+++	++
SST 16	+++	+++	++	na
SST 33	++	++	+	na
SST 44	+	-	-	-
SST 66	+++	+++	++	++
Agent	[++]	-	-	-
Adam Tas	+++	+++	++	++
Chokka	+++	+	++	++
Dias	+++	+	++	++
Festiquay	+++	-	-	-
Gamtoos	+++	+++	++	na
Marquis	-[+++]	-	-	na
Nantes	+++[+]	+++	++	na
Schooner	++	++	+	na
Sterling	++	++	++	na
Vloekskoot	++	+	+	na
D.singuinalis	+++[+]	+++	++[+]	na
P.maximum	-	-	-	-

[d] *H.vulgare* variety; na = leafhopper transmissions not performed. - = no streak symptoms; + = mild stippled streak and recovery (-/+); ++ = moderate stippled to continuous streak; +++ = severe continuous streak and stunting and/or distortion of growth and death(+++[+]). -[+++] = symptoms appeared in single single plantlets.

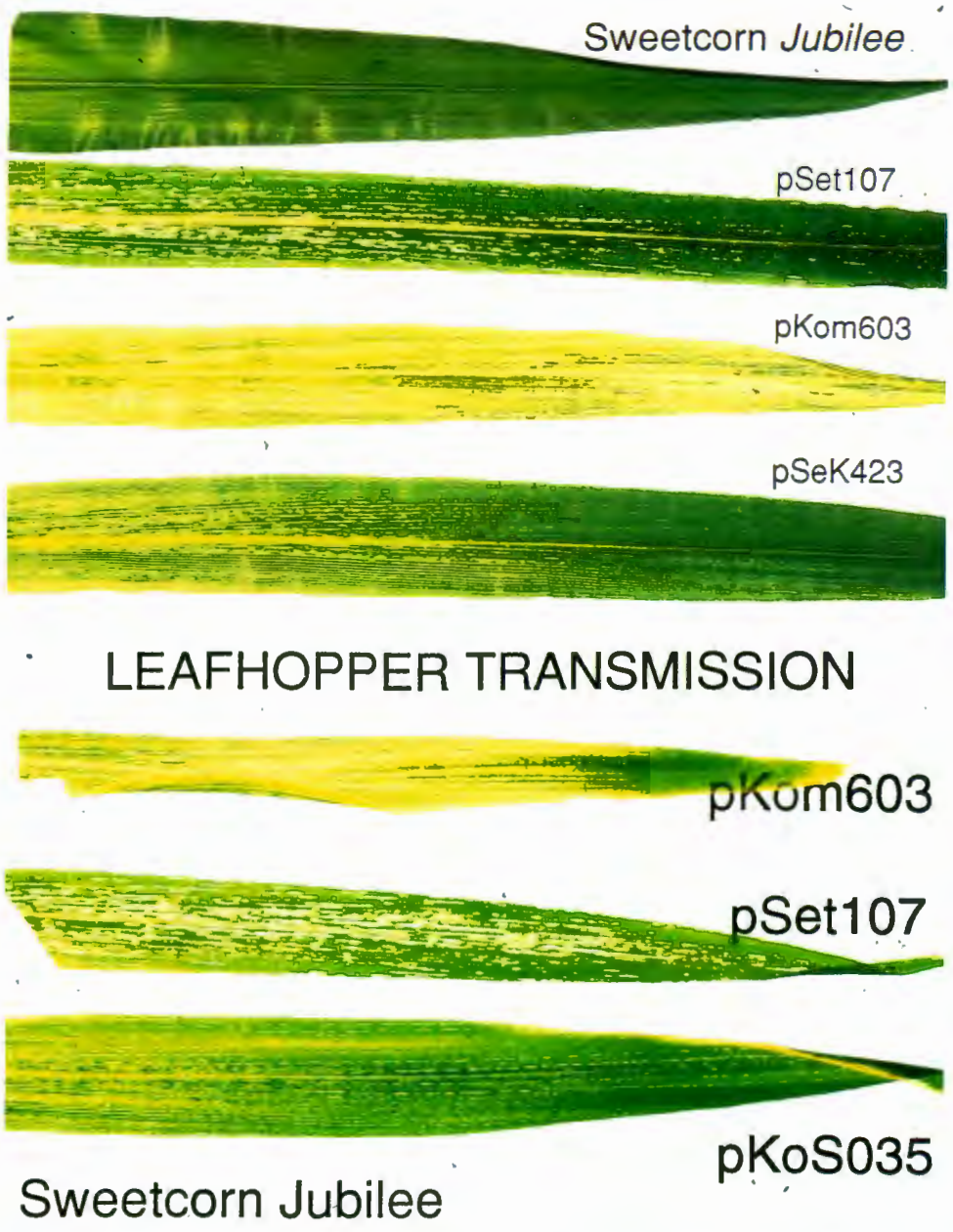

Figure 5.2: Streak symptoms of selected recombinants on sweetcorn

Top; 3rd leaves of agroinoculated sweetcorn. From the top; healthy plant wild-type infections of pSet107 and pKom603 respectively and agroinoculated recombinant pSeK423. Below are 3rd leaves from sweetcorn plants exposed to viruliferous leafhoppers that fed from wild-type pKom603- and pSet107-agroinoculated plants, and recombinant pKoS035-agronoculated plants respectively.

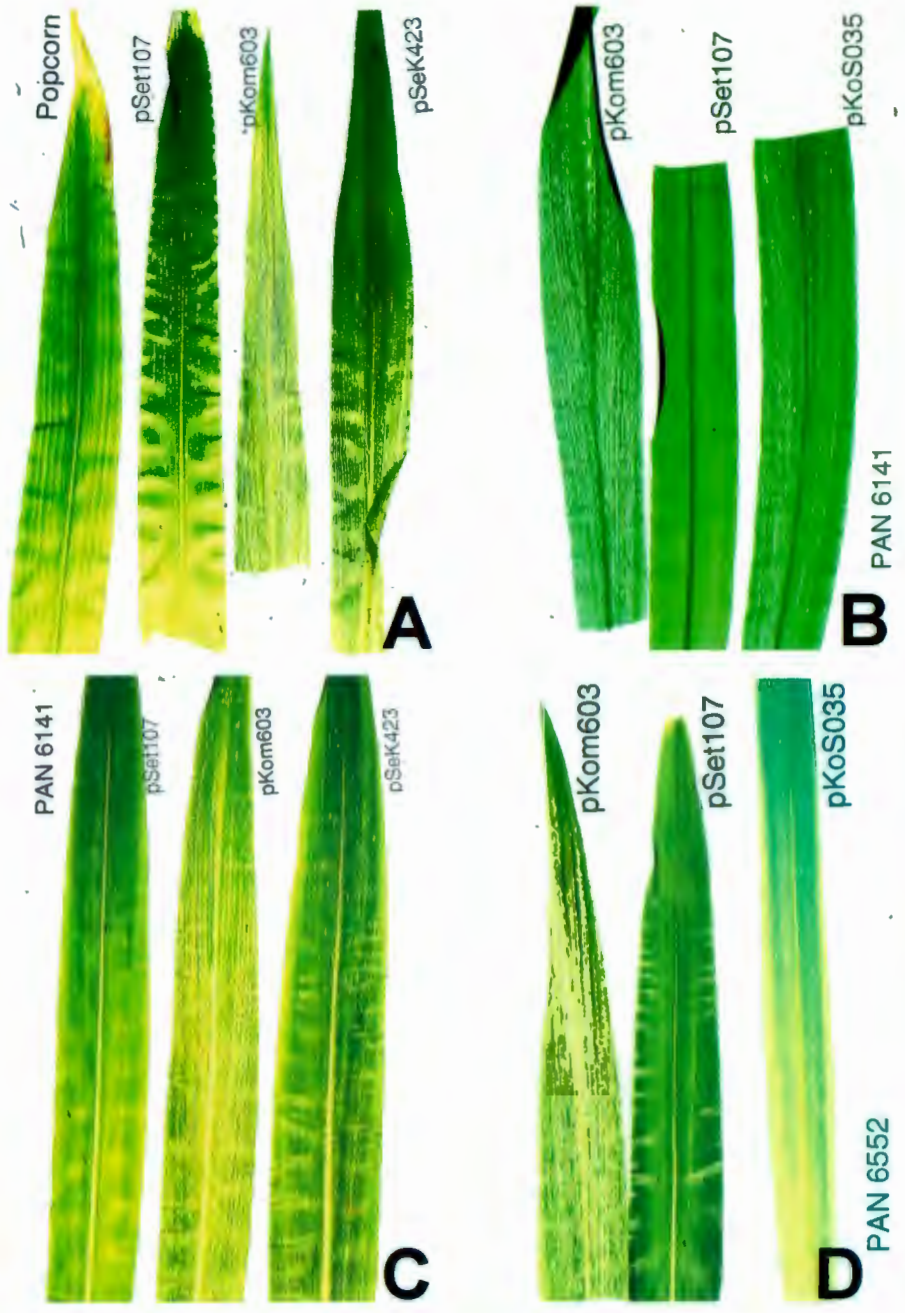

Figure 5.3: Streak symptoms of recombinants on maize cultivers

Viruliferous leafhoppers were allowed to feed on two-week old maize plantlets. All leaves are the 4th leaves from 5-week-old plants except for leaf in D on the left (2nd leaf). A: from the left, popcorn infected with wild-types pSet107 and pKom603, and LIR recombinant pSeK423 respectively. B and C: "resistant" maize cv.s. B: infected with wild-types pKom603 and pSet107 (no streak), and MP recombinant pKoS035 respectively. C: infected with wild-types pSet107 (no streak) and pKom603, and pSeK423 respectively. D: "resistant" PAN6552 infected with pKom603 and pSet107 (no streak) wild-types respectively; and 2nd leaf with MP recombinant pKoS035.

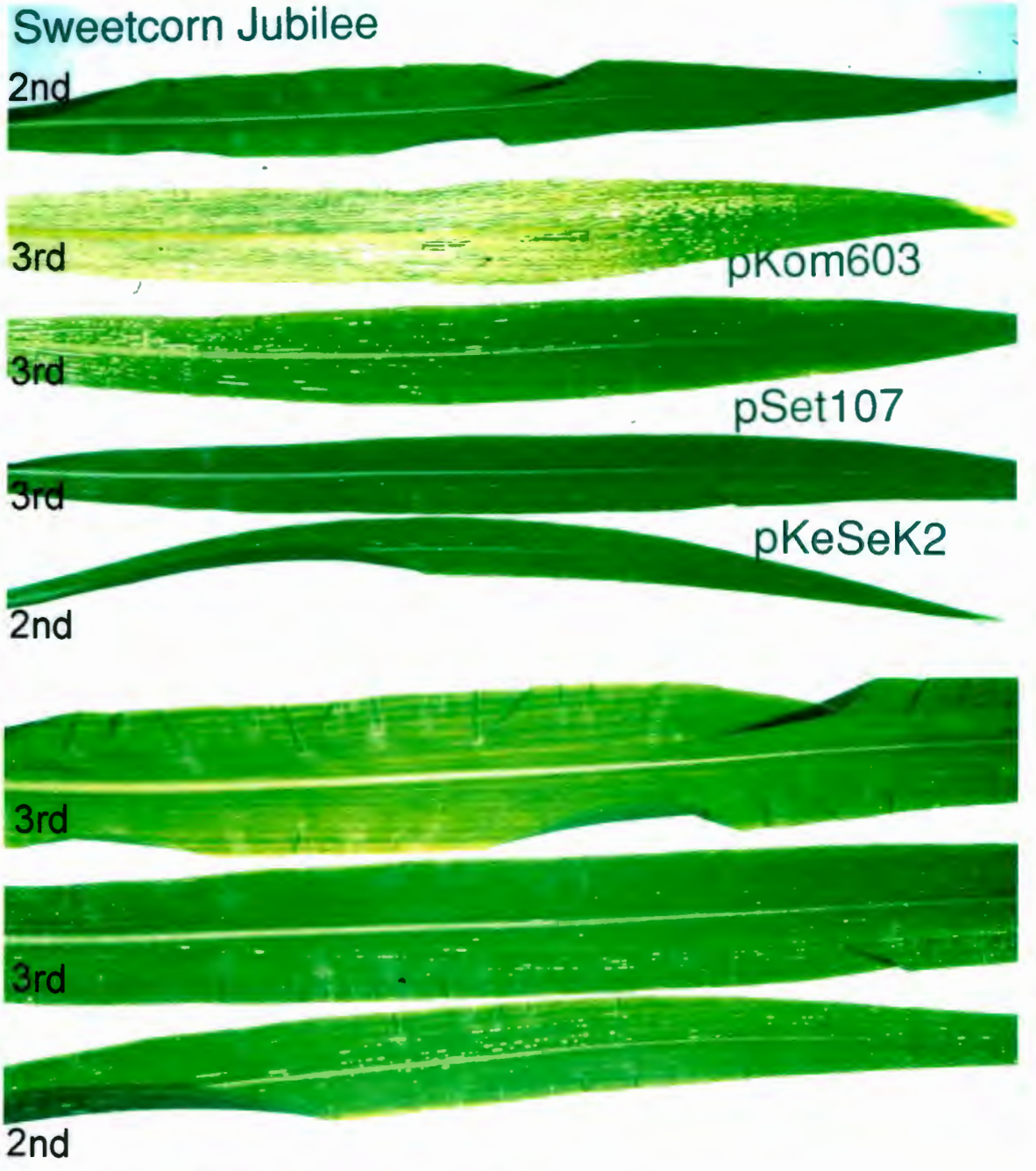

Figure 5.4: Mild symptoms of pKeSeK2 on sweetcorn

From the top: healthy sweetcorn pBI-agroinoculated, wild-type pKom603-agroinoculated, wild-type pSet107-agroinoculated, cloned LIR recombinant pKeSeK2-agroinoculated. Bottom three are healthy leaf and pKeSeK2 mild streak. Leaf numbers are indicated.

5.3.3 PCR detection of infectious pKeSeK

Since the pKeSeK2-induced symptoms were extremely mild it was thought that PCR detection would be a better alternative to Southern blotting for detection of replicating virus. MSV-Set-specific primers designed for the extension of MSV-Set sequencing (Chapter 3) and the African Mastrevirus-specific primers could be used to ascertain whether MSV-KeSeK was infectious or at least replicating at low levels in sweetcorn plants. Both sets of primers amplify across the LIR of RF-DNA: this means that no amplification could occur without replicational release of circular unit-length genomes since the LIR/V1 (*Bam* HI site) and the C1/LIR (*Sac* I site) junctions are interrupted in the 1.1mer clones used for agroinfection. The MSV-Set specific primers P2 amd P3 tests produced 1.7 kb PCR DNA products in MSV-Set wild-type infected plant DNA, and in pKeSeK2 agroinoculated plants whether they had streak symptoms or not (Table 5.1, Figure 5.5). Plants not showing streak symptoms showed very low levels of PCR product, indicating that the KeSeK chimaera was replicating at very low levels. The PCR test was repeated using the degenerate primers D1 and D2, which produced 1.3 kb PCR products in pKeSeK2-agroinoculated plants and both the wild-type- (pKom603 and pSet107)-agroinoculated plants. No DNA was amplified from any of the negative controls, such as DNA from pBI121-agroinoculated plants and the water controls. The P2 and P3 primers were specific to both the MSV-Set wild-type and pKeSeK recombinant, indicating that the pKeSeK chimaera was replicating in symptomatic sweetcorn host plants, and at very low levels in agroinoculated symptomless plants. Digestions using *Nco* I and *Bgl* II indicated that the PCR products obtained from pKeSeK2-agroinoculated plants were KeSeK chimaeras (See Figure 5.6 and Table 5.5). pKeSeK retained the *Bgl* II site from MSV-Set and lost the *Nco* I site in the LIR of MSV-Set, indicating that it was the chimaera that was amplified. Wild-type MSV-Kom RF-DNA has neither of the restriction sites within the PCR product obtained using D1 and D2 primers.

5.4.4 Sequencing analyses of RF-DNA from agroinoculated plants

RF-DNA from pGC/TT2- and pTT/GC2-agroinoculated plants with streak symptoms were cloned as sequenced as described in Chapter 4 (4.3.4). This was done to ascertain whether the GC or TT (MSV-Set nt pos.149 and 150 or MSV-Kom nt pos. 148 and 149 respectively) nucleotides are normally switched as a mechanism of release of unit-length MSV genomes during agroinoculation. Of the ten progeny genome clones that were partially sequenced (pTT/GC21 - pTT/GC25 and pGC/TT21 - pGC/TT25), only pGC/TT25 showed that GC was indeed switched by TT, presumably by the same mechanism in which 12 of the 43 recombinants sequenced (Chapter 4) were released in heterodimer agroinoculations.

Two RF-DNAs that were cloned from pKoS035-agroinoculated plants and designated pKoS351 and KoS352, were both sequenced into the V1 ORF and found to have retained the MSV-Set 176b replacement.

Table 5.5 Digestions of the PCR products to determine whether KeSeK was infectious

PCR product	PCR product nt pos.	*Nco* I nt pos. in LIR	*Bgl* II nt pos. in C1 ORF	*Nco* I product sizes	*Bgl* II product sizes
MSV-Kom D1+D2 1306 bp	1746-2690 & 1-360	none	none	1306bp (Lanes 6 & 7)	1306bp (Lanes 6 & 8)
MSV-Set D1 + D2 1321 bp	1762-2701 & 1-359	150	2485	1321bp	723bp & 598bp (Lanes 12&13)
MSV-Set P2 + P3 1729 bp	2008-2701 & 1-1035	150	2485	930bp & 799bp (Lanes 10&11)	1729bp
MSV-KeSeK D1 + D2 1306 bp	1742-2686 & 1-361	none, since contains MSV-Kom LIR	2483, since contains MSV-Set C1 ORF	1306bp	741bp & 564bp (Lanes 3 & 4)
MSV-KeSeK P2 + P3 1714 bp	2025-2686 & 1-1033	none, since contains MSV-Kom LIR	2483, since contains MSV-Set C1 ORF	1714bp (Lanes 1 & 2)	473 bp &1241 bp

D1 & D2 = African Mastrevirus-specific degenerate primers used to amplify MSV-Kom, -Set or -KeSeK; P2 & P3 = MSV-Set-specific primers used to amplify MSV-Set or the KeSeK chimaera. Digestions that were performed are indicated by lanes mentioned and are depicted in Figure 5.6.

Bgl II appears in the C1 ORF of MSV-Set genome but not in the C1 ORF of MSV-Kom.

Nco I appears in the LIR of MSV-Set but not in the LIR of MSV-Kom.

Thus KeSeK chimaera would have the Bgl II but not the *Nco* I site.

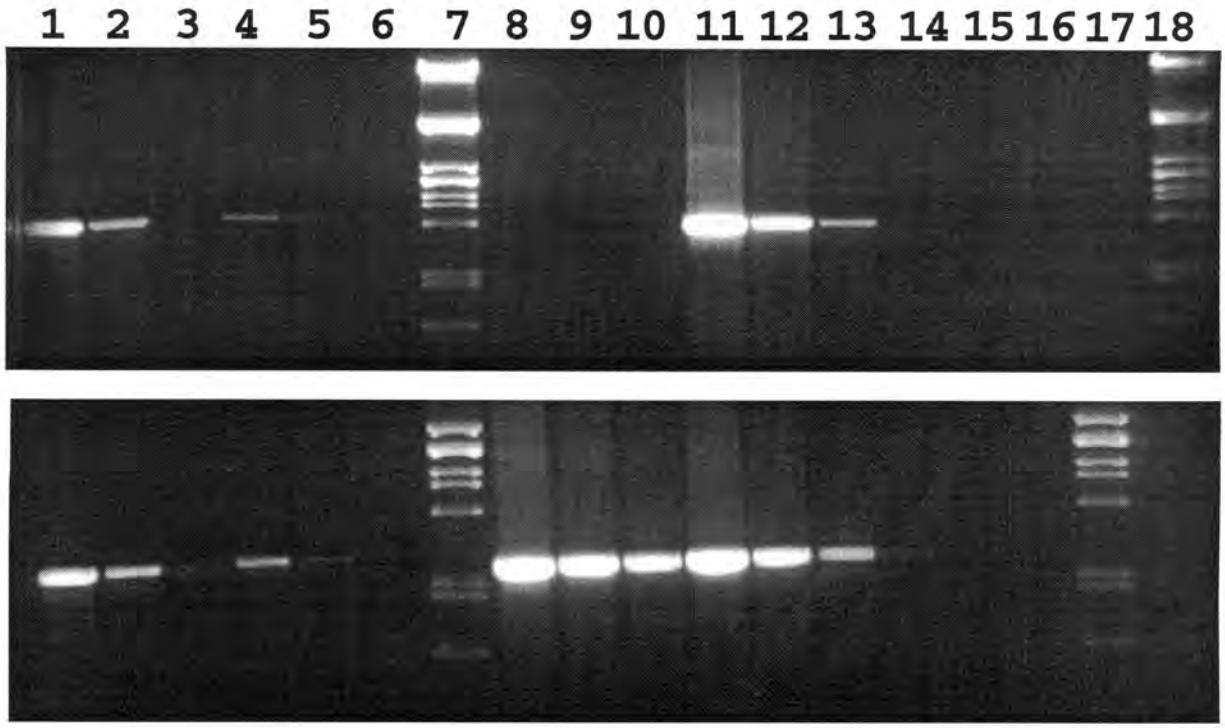

Figure 5.5 PCR detection of chimaeric KeSeK in agroinoculated plant DNA extract

Top: MSV-Set specific primers used in PCR reactions generate 1731 bp product.

Bottom: Degenerate primers used in PCR reactions generate 1321 bp product from MSV-Set DNA and 1306 bp products from MSV-Kom and chimaeric KeSeK DNAs.

10 µl PCR products were run on a TBE 0.8% agarose gel with 20 ng/ml ethidium bromide stain. Images were obtained using the UVP image processor. Lanes 1, 2 and 3 contain 50ng, 5ng and 0.5ng total plant DNA respectively isolated from a pKeSeK2-agroinoculated sweetcorn plant with streak symptoms: Lanes 4, 5 and 6; 50ng, 5ng and 0.5ng total plant DNA respectively isolated from a pKeSeK2-agroinoculated plant with no symptoms. Lanes 8, 9 and 10; 50ng, 5ng and 0.5ng respectively isolated from a pKom603-agroinoculated plant with typical MSV-Kom streak symptoms. Lanes 11, 12 and 13; 50ng, 5ng and 0.5ng total plant DNA respectively isolated from a pSet107-agroinoculated sweetcorn plant with typical MSV-Set streak symptoms.. Lanes 14, 15 and 16; 50ng, 5ng and 0.5ng respectively of total plant DNA extract isolated from pBI121-agroinoculated plants. Lanes 17 (top) and 18 (bottom); negative water control. Lanes 17 (bottom) and 18 (top); λ DNA digested with *Pst* I (molecular weight marker).

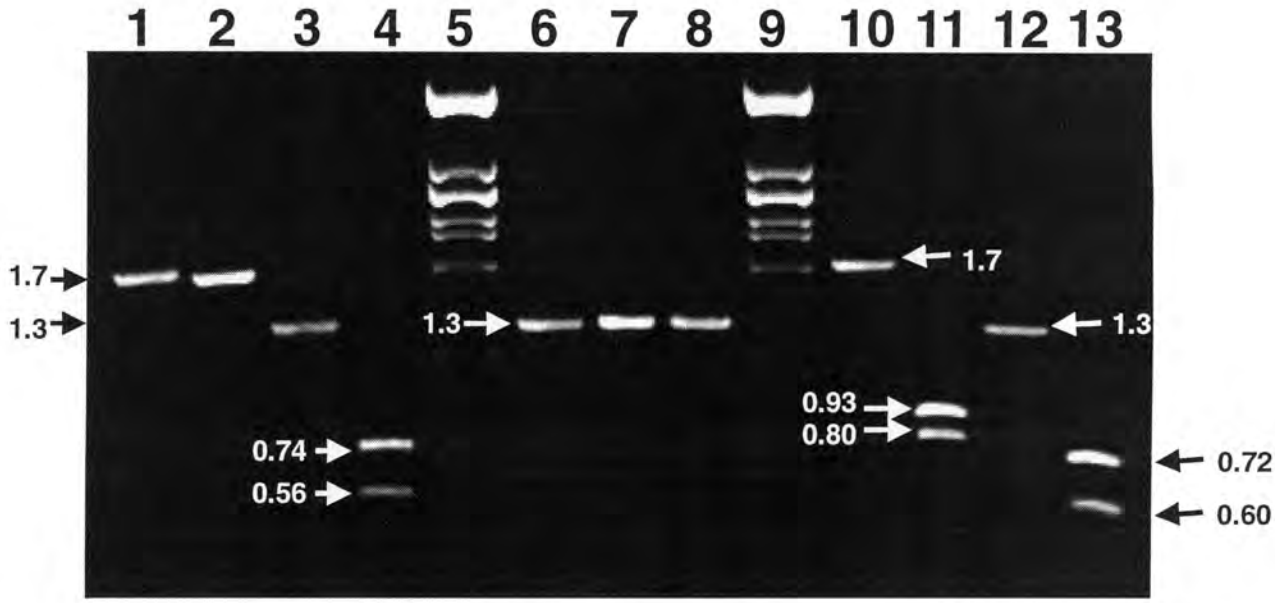

Figure 5.6 Digestions of PCR products to distinguish chimaeric KeSeK

Lanes 1 and 2; 5µl PCR product amplified using P2 and P3 on total plant DNA extracted from pKeSeK2-agroinoculated sweetcorn, undigested and *Nco* I-digested respectively. Lanes 3 and 4; 5µl undigested and 10µl *Bgl* II-digested PCR product amplified using D1 and D2 on total plant DNA extracted from pKeSeK2-agroinoculated sweetcorn. Lanes 5 and 9; *Pst* I-digested λ DNA molecular weight marker. Lanes 6, 7 and 8; 5µl undigested, *Nco* I-digested and *Bgl* II-digested PCR products amplified using D1 and D2 on total plant DNA extracted from pKom603-(wild-type-)agroinoculated sweetcorn. Lanes 10 and 11; 5µl undigested and 10µl *Nco* I-digested PCR product amplified using P2 and P3 on total plant DNA extracted from pSet107-(wild-type-)agroinfected sweetcorn. Lanes 12 and 13; 5µl undigested and 10µl *Bgl* II-digested PCR product amplified using D1 and D2 on total plant DNA extracted from pSet107 agroinfected sweetcorn. (see Table 5.5)

5.5 DISCUSSION

The successful agroinfections of recombinants generated from the heterodimer-agroinoculations reported in Chapter 4 indicated that the recombinant genomes contained in pSeK142, pKoS005 and pSeK001 were indeed infectious despite the sequence replacements of 105, 176 and 2 bp respectively. The agroinfectivities and the transmissibility of streak disease by leafhoppers (fulfilling Koch's postulates) allowed the determination of their host ranges and symptomatology, which were subsequently compared to those of the wild-types to ascertain the biological effects of the replacements. Thus the DNAs cloned in pSeK142, pKoS005 and pSeK001 are sufficient for infection and appear to contain whole Mastrevirus genomes. The chimaera pKeSeK, although produced very mild streak in stressed (lower photosynthetic rate due to low daylight intensity) slow-growing Jubilee sweetcorn plants, was not shown to be leafhopper-transmissable in carefully conducted transmission tests; thus not fulfilling Koch's postulates. However PCR detections revealed that pKeSeK was replicating at low levels. The implication of this result is discussed.

5.5.1 pGC/TT2 and pTT/GC2 agroinoculations

The agroinfectivies of pGC/TT2 and pTT/GC2 were comparable to those of the MSV-Kom wild-type pKom603. Symptoms and agroinfection rates appeared to be unaltered in agroinoculation tests (Table 5.2 and Figure 5.2). These results indicate that the 2bp replacement in nt pos. 147 and 148 in MSV-Kom does not affect the virus at the levels of agroinfectivity and symptomatology on Jubilee sweetcorn. Thus it appears that there is no important function for these nucleotides in MSV-Kom, and presumably in MSV-Set as well.

Ten progeny RF-DNAs from pGC/TT2 and pTT/GC2 (five from each) were isolated and sequenced in an effort to determine the possible function of, or whether the region immediately left of the *Bam* HI site with the conserved 14 bp recombinational "hotspot" (Chapter 4), was involved in the replicational release of unit-length genomes from tandem dimers during agroinoculations. The fact that only one progeny RF-DNA revealed that the two bases (nt pos. 147 and 148 for MSV-Kom) was switched is further indication that this particular site is a preferred recombinational site rather than that it has a function in the release of these unit-genomes during agroinoculation or Mastreviral RCR. These results also suggest that replicational release - as the results suggest in Chapter 2 and 4 - is nine times more effective than the homologous recombination method. Since there are other preferred homologous sites at which crossovers occur in agroinoculations (such as the UAS and the 28bp conserved site containing the conserved *Apa* I site right of the hairpin loop;

Figure 4.10), it is likely that homologous recombinational releases occur more often than the suggested 10% from these results, since replicational release would not be distinguishable from recombinational release in progeny RF-DNAs from pGC/TT2 and pTT/GC2 agroinoculations.

5.5.2 pSeK423 and pKoS053 agroinoculations

The reduced agroinfectivities (Table 5.2; Figure 5.2) and symptom severities (Figures 5.2 and 5.3) of pSeK142 and pKoS005 is an indication that although there is a degree of incompatibility between MSV-Kom and MSV-Set (Chapter 3), the recombinants still retain their Mastreviral properties such as leafhopper transmissibility. Observations of the streak symptoms on agroinoculated plants are reflected by host/virus interactions and is not likely to be entirely dependent on viral factors only. This is reflected by the fact that MSV-Kom and MSV-Set have distinct symptoms from each other on the same plant cultivar but their symptoms vary on different cultivars (Chapter 2). In addition they have varying host ranges. Although the host ranges of pSeK142 and pKoS005, as reflected by the agroinoculations of the homodimeric 1.1mers pSeK423 and pKoS035, were largely unchanged, their streak symptom severities were generally attenuated. Some cultivars developed dramatically decreased streak severity or no streak at all after being exposed to viruliferous leafhoppers (Tables 5.3 and 5.4; Figures 5.2 and 5.3). For example, tolerant lines PAN 6191 and PAN 6195 developed no streak symptoms although severe streak was obtained from wild-type MSV-Kom-viruliferous leafhoppers. It is unlikely that the N-terminus of the MP per se is directly involved in host range determination, since the symptom severity of these recombinants are reduced in all cultivars. However, inherently tolerant lines could well be resistant to these recombinants as is often seen in this laboratory in comparisons of maize and grass-isolated strains of MSV (E.P. Rybicki pers. comm.; unpublished)

pSeK423 and pKoS035 produced continuous streaks on leaves of agroinoculated sweetcorn plants, however the streaks are much narrower than those of the MSV-Kom and MSV-Set wild-types. It appears that while the long-distance movement along the vasculature of the leaves is comparable with the wild-types the lateral or cell-to-cell movement appears to be hindered (Figure 5.2). A mutation on nucleotide 42 of the V1 ORF of MSV, observed by Boulton *et al* (1991), caused the reduction of the width of chlorotic streaks in infected maize plants. This result along with those of pSeK423 and pKoS035 in this study seems to confirm the importance of the V1 gene or MP in the width of the streak symptoms or cell-to-cell movement. Long-distance movement is associated with the longitudinal movement of the virus in phloem tissue from cell to cell through sieve plates as opposed to through plasmodesmata in lateral movement. Whether the N-terminus - more specifically - of the MP

is involved in the cell-to-cell movement is not clear here. However the 176 bp replacement effectively replaces 57% of the N-terminus of the MP which includes a large portion of the hydrophobic region (Boulton *et al.*, 1989), of MSV-Kom with that of MSV-Set. Since the MP was observed to be the least conserved Mastreviral protein, and it interacts with host cell factors for cell-to-cell movement, it could be an important factor in determining host range (Chapter 3). Whether this host range determining factor includes the hydrophobic region of the MP (which has been suggested to have affinities with host cell membranes and thereby assisting cell-to-cell movement or lateral movement through plasmodesmata) is also not clear. The hydrophobicity profile of the putative chimaeric protein closely resembles those of the wild-types (which also have similar hydrophobicity profiles; Chapter 3), thus it is not surprising that this 176 bp V1 ORF sequence replacement in pSeK142 does not "cripple" or render the recombinant non-viable. However this reduction in symptom severity and agroinfectivity is more likely to be due to a less fit putative chimaeric MP with the N-terminus interactions with the host or the MSV-Kom machinery being less specific than with its wild-type counterpart. This is evident since MP and CP replacement experiments performed by Eric van der Walt (Honours thesis) in this laboratory indicated that the exchange of V1 (MP) or V2 (CP) ORFs consistently resulted in lower virulences than those with complete V1/V2 replacements, indicating specific interactions between these proteins. Thus it is possible that the replacement of the N-terminus of the MP of MSV-Kom with that of MSV-Set renders the chimaeric MP less able to interact with MSV-Kom at the DNA and protein levels thereby causing the attenuated streak disease and agroinfectivities. Whatever the case, the 176 bp V1 ORF replacement, effectively replacing 57% N-terminal amino acid sequence of the MSV-Kom MP with cognate MSV-Set sequence, does not result in the MSV-Set host range being conferred on MSV-Kom.

The LIR 105 bp replacement upstream of the V1 ORF in pKoS105 appears to have a similar effect than that of the V1 ORF replacement above. Similarly disrupted specificities in possible interactions with the bending locus containing dT.dA tracts in the LIR right of the hairpin loop replaced in pKoS035 may also be disrupted. Presumably as a result of these replacements and reduced virulences, the symptoms appear approximately three days later than those of the wild-types (Figure 5.2). This requires further investigation and indeed the MSV-Kom and MSV-Set CP and MP replacement experiments and their reciprocals will possibly reveal more details of these interactions (E. van der Walt, pers. comm.).

Although the sequence replacements in pSeK142 and pKoS005 recombinants did not conclusively reveal possible functions or what biological characters they may confer, the fact that recombinants are visibly infectious at the level of streak disease development and thus

easily detectable, is further indication that the heterodimer agroinfection method, as described in Chapter 4, is potentially convenient in generating viable recombinants with significant sequence replacements. These results strongly suggest that random intermolecular "homologous" recombination events that may occur between wild-type MSV strains in mixed infections in nature can indeed generate new natural viable strains. This is further supported by sequence data that indicate that certain geminiviruses, such as the HrCTV genome which has part sequence affinities with SLCV and part sequence affinities with BCTV (Klute *et al.*, 1996), are as a result of an ancient recombinational event. Numerous examples of recombination between different geminiviruses are cited in Chapter 1.

5.5.3 pKeSeK agroinoculations and detection

PCR detections of replicating RF-DNAs in total DNA extracts of pKeSeK2-agroinoculated Jubilee sweetcorn and subsequent restriction digests of the products indicated that the MSV-Set chimaera with its entire LIR region replaced with that of MSV-Kom was indeed replicating at low levels. This was substantiated by the development of extremely mild streak disease symptoms on Jubilee sweetcorn plants (Figure 5.4). This was, however, only on plants growing at lower light intensities. As explained in Chapter 2, the efficacy of acquisition of the leafhopper depends on the extent of the streak caused, since virus distribution correlates with symptom expression in the host (Briddon *et al.*, 1992). Thus it is possible that the inability of leafhoppers to acquire and transmit these "chimaeric viruses" may be due to the inadequate levels of viruses available from very mild streak (von Wechmar and Hughes, 1992). Determining whether this chimaera is in fact leafhopper transmissible will require more work; however it appears to have retained its ability to replicate, though at low levels, and indeed its ability to produce streak disease in stressed slow growing Jubilee sweetcorn plants.

PCR products were cloned and sequenced to determine whether the monomeric viral sequences indeed had the LIR of MSV-Kom replacing that of MSV-Set between the *Sac* I and *Bam* HI sites of the MSV-Set genome. Repeated PCR amplifications and the sequences of three PCR clones cleared the doubt that chimaeric KeSeK was present in plant DNA extracts of pKeSeK2-agroinoculated Jubilee sweetcorn plants replicating at low levels.

This result is interesting since it suggests that the MSV-Set Rep protein (with an amino acid sequence identity with that of MSV-Kom of 82%) would be able to initiate replication of MSV-Kom wild-types. Thus in nature it is likely that in mixed infections, MSV-Kom and MSV-Set

are able to *trans*-replicate each other. This result was further confirmed in *trans*-replication experiments conducted by Janet Willment (pers. comm.) in co-bombardments of Black Mexican Sweet cultured callus where MSV-Set Rep mutants were *trans*-replicated by MSV-Kom replicons. Bidirectional transcription of the virion- and complementary-sense ORFs is likely to be attenuated due to a degree of loss of specificity of the MSV-Set proteins in interactions with the MSV-Kom LIR, which is presumably reflected by the three-week duration of the development of the streak symptoms on leaves of agroinoculated stressed slow growing Jubilee sweetcorn. This may also be a factor in the pKoS035-agroinoculations (where there is a 105 bp replacement of the right part of the LIR), in which attenuated symptoms and lower agroinfectivity rates were observed.

Since the UASs of MSV-Kom and MSV-Set are identical (Chapter 4), it appears unlikely that the V1/V2 transcriptional factors will be a factor in reducing the virulence of the KeSeK chimaera. However the iterative elements situated in the LIR left of the hairpin loop and within the stem of the hairpin have disimilar sequences, are different in size, and are not equidistant in MSV-Kom and MSV-Set LIRs (Chapter 4). Iterative elements have been proposed to specifically bind Rep proteins as a prerequisite for the initiation of replication of geminiviruses (Argüello-Astorga *et al.*, 1994). The reduced agroinfectivities, lengthened time for symptoms to appear on stressed Jubilee sweetcorn and attenuated streak symptoms thus may be as a result of the inability or reduced ability of MSV-Set Rep to bind to the iterons of MSV-Kom and initiate replication of the KeSeK chimaera. However since replication is evidently occurring for KeSeK, the iteron model for Mastreviruses proposed by Argüello-Astorga (1994) is in question and is under further investigation in this laboratory (Janet Willment, pers. comm.). More recombinants (pSeK009, pSeK010 and pSeK111; Table 4.4) with the entire hairpin loops along with the right side of the hairpin of the LIR up to the conserved *Bam* HI site replaced were found, and although not shown to be infectious in this study, are being further investigated in this laboratory. pSeK115, pSeK1110 and pSeK118 generated from heterodimer-agroinoculations with the LIR right of the stem-loop replaced including insertions disrupting the formation of the hairpin loop (Chapter 4), are also being further investigated in this laboratory (J. Willment, pers. comm.).

Pseudorecombination experiments have been done where combinations of closely related Begomovirus strains isolated from the same plant species were able to *trans*-replicate each other as long as their intergenic regions shared identities of equal to or greater than 87% and were not infectious when the intergenic regions shared identities of <75% (Stanley *et al.*, 1985; Morris *et al.*, 1990; Faria *et al.*, 1994; Lazarowitz, 1991 and von Arnim and Stanley, 1992b; Gilbertson *et al.*, 1993; Gilbertson *et al.*, 1991; Lazarowitz *et al.*, 1992; Stanley and

Gay, 1983). It is thus surprising that pKeSeK2 was found to be infectious since while the complete sequence identities of MSV-Kom and MSV-Set is 78% (Chapter 3), alignments of their LIRs only share an identity of 71% along with the iteron differences discussed above. This is reflected in the extremely attenuated symptoms in plants growing in low light intensities (Figure 5.4). Although the identity of the LIRs of MSV-Kom and MSV-Set are relatively low, this is in line with the ability of heterodimeric constructs of WDVs with complete sequence identities of 82.5% but with LIR identities of 75%, to release unit-length genomes in transfected wheat protoplasts (Heyraud *et al.*, 1993). Whether the progeny of these WDV heterodimers were isolated and tested in agroinoculations to determine their biological effects and on whether in fact they were infectious, has not been reported.

According to the guidelines to the demarcation of virus species of virus species by the ICTV (van Regenmortel *et al.*, 1997), genome features and sequence are two of six criteria used in demarcating virus isolates as distinct species in the *Geminiviridae* family. More specifically, to demarcate viruses into distinct species there should be no transcomplementation of gene products and no pseudorecombination between components and in addition, the genome sequences should not have more than 90% identity. In the case of MSV-Kom there is no doubting that it is a strain of MSV since it has been isolated from maize, and has sequence identities with the MSV group of viruses isolated from maize of greater than 98%. However, although the results detailed above indicate strongly that there is transcomplementation between the MPs, Rep proteins and LIRs of MSV-Kom and MSV-Set, their sequence identity is only 78%. According to the sequence criteria MSV-Set can probably be considered a distinct species rather than a strain of MSV, however the transcomplementation criterium would class MSV-Set as an MSV strain. This is further discussed in Chapter 6.

5.6 CONCLUSIONS

The recombinants obtained from heterodimer-agroinoculations reported in Chapter 4 were found to be infectious, however with reduced virulences. Thus the use of heterodimer-agroinoculated plants to "sift-out" crippled non-useful recombinants for further research is indeed a potentially effective method in obtaining viable recombinants for further investigation. This work has demonstrated that viable recombinants can be "naturally" selected, presumably on the basis of their fitness in complementary interactions between different parts of the heterologous genomes and/or host/viral interactions. Since recombinants such as pKoS005 and pSeK423 which have been shown to maintain their

Mastreviral characters, they can be useful in determining the biological functions and/or genomic interactions of those genomic domains. Additionally, recombinants such as these can be difficult to construct in the laboratory due to sequence or conserved restriction site limitations thus it is possible to generate them using the heterodimer agroinoculation method.

Although the sequence replacements of 105 bp in the LIR and 176 bp in the V1 ORF of the chimaeric RF-DNAs contained in pSeK141 and pKoS005 respectively are reduced, they have indeed maintained their Mastrevirus biological characters such as the leafhopper transmissibility and the ability to develop streak symptoms on host plants. This complementation of heterologous sequences is reflected in the fact that MSV-Set, although isolated from a grass species, is grouped with the MSVs isolated from maize in phylogenetic studies (Chapter 3). The reduced virulences are presumably due to the level of unfitness of the replacements in heterologous genomes where complementary specific interactions are disrupted to a certain degree.

The infectivity of the chimaera pKeSeK, in which the the LIR of MSV-Set is replaced by that of MSV-Kom, is an indication that the MSV-Set Rep is able to *trans*-replicate the genome of MSV-Kom. This is evident in co-inoculation experiments in which MSV-Set Rep mutants and MSV-Kom mutants were bombarded into Black Mexican Sweet callus (K. Palmer and J. Willment, pers. comm.). This is the first case among geminiviruses in which *trans*-replication of heterologous intergenic regions with sequence identities of less than 75% has been reported. Although a greater than 90% genome sequence identity is one of the guidelines for demarcating geminiviruses (van Regenmortel *et al.*, 1997), MSV-Set with only a 78% identity with that of MSV-Kom evidently is able to *trans*-replicate the genome of MSV-Kom. However transcomplementation of genes is another criterion for demarcating these viruses, therefore the demarcation of MSV-Set as an MSV strain is questionable and is further discussed in Chaper 6. Since the iterons of MSV-Kom and MSV-Set are different, this has implications for the model proposed by Argüello-Astorga *et al.*, (1994).

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

6.1 GENERAL DISCUSSION AND CONCLUSIONS

This thesis presents the biological and genomic characterisation of three related Mastreviruses that were previously isolated in different geographical regions in South Africa. The biological characterisation of MSV-Kom, MSV-Set and PanSV-kar indicated that they were distinct Mastreviruses. The genome sequence of MSV-Set was determined in entirety and comparatively analysed along with those of MSV-Kom and PanSV-Kar; all genomes feature typical Mastreviral genomic characteristics. The three Mastreviruses were taxonomically ranked by phylogenetic analyses of their genome and putative protein sequences; the results reflect their biological differences.

In addition, the possibility that two distinct viruses can recombine to form viable chimaeras is investigated in this project. This was achieved by constructing hetero-tandem genomes in *Agrobacterium* plasmids and injecting these into very young susceptible maize seedlings, and then allowing agroinoculated plants to grow and "sift" non-viable recombinants out. Whether recombinant progeny were infectious and subsequently acquired different biological features to those of the wild-types, was also investigated in this study. Three recombinant progeny as well as a constructed chimaera were shown to be infectious and their biological characteristics were determined in this study.

6.1.1 The classification of MSV-Set

Criteria that demarcate geminiviruses into distinct species include differences in genome or protein features and sequences, biological features such as transmission, the effects in infected tissue and host range (van Regenmortel *et al.*, 1997). MSV-Kom, MSV-Set and PanSV-Kar share properties such as leafhopper transmissibility, and streak symptomatology on a number of cereals and grasses, and feature ssDNA and RF-DNA in infected plants, as well as sharing identical genome organisations. These factors group them unequivocally as Mastreviruses. This was also supported by the fact that geminate particles 18x30nm in size were observed in leaf-dip preparations.

The genome sequences of MSV-Kom and PanSV-Kar are shown to have sequence identities of 98% and 89% with MSV-SA (Lazarowitz, 1988) and PanSV-Ken (Briddon et al., 1992) respectively. Although a genome sequence criterion for placing geminiviruses as distinct species is a less than 90% sequence similarity (van Regenmortel *et al.*, 1997), the fact that PanSV-Kar was isolated from *P. maximum* and causes very mild non-persistent streak in Jubilee sweetcorn comparable to that of PanSV-Ken (Briddon *et al.*, 1992), means that PanSV-Kar is almost undoubtely a strain of *Panicum* streak virus. The 11% genomic

sequence difference is thus a reflection of their geographical origins; PanSV-Kar being isolated in South Africa and PanSV-Ken, isolated in Kenya.

An additional criterion for demarcation is the differences in CP amino acid sequences, which to a large extent are related to how virions react differently with key antibodies (van Regenmortel *et al.*, 1997). The CP sequence identities shared between MSV-Kom, MSV-Set or PanSV-Kar and MSV-SA are 99%, 84% and 69% respectively. While MSV-Kom falls in the MSV group by this criterion, PanSV-Kar however shares an 89% CP sequence identity with that of PanSV-Ken (Briddon *et al.*, 1992). This CP difference is similar to that of their genomic sequence differences; thus, for the same reasons mentioned above PanSV-Kar falls into the *Panicum* streak virus group.

Criteria that may class MSV-Set as a different species to the MSV group are the relatively low genome sequence and CP sequence identities of 78% and 84% between MSV-Kom and MSV-Set, as well as their host range and streak symptom differences. Also MSV-Set was isolated originally from a *Setaria* grass species (Hughes *et al.*, 1992) and, although able to cause moderately severe streak in susceptible lines of maize such as Jubilee sweetcorn, does not infect most (even moderately tolerant) maize varieties (Chapter 2).

On the other hand there are criteria that may class MSV-Set as a strain in the same species as the MSVs. These are the transcomplementation of gene products and pseudorecombination criteria (van Regenmortel *et al.*, 1997). Pseudorecombination cannot be a criterion for the demarcation of Mastreviruses since they only have monopartite genomes. However, an experiment was performed and presented in Chapter 5 where the pKeSeK2 construct containing the four MSV-Set ORFs flanked by two MSV-Kom LIRs effectively mimics Begomovirus pseudorecombination experiments. Pseudorecombinants are made by exchanging cloned Begomovirus A or B components with those of a related virus (Bisaro, 1994; Stanley, 1985; Gilbertson *et al.*, 1993) where the Rep of the A component would transreplicate the B component of a closely related Begomovirus. In the pKeSeK2 experiment, the MSV-Set Rep (which is transcribed from a MSV-LIR with identical UASs to that of MSV-Set) is able to initiate replication despite the fact that the iterons (which are though to bind Rep as a prerequisite for replication) are different in MSV-Set and MSV-Kom. Although mixtures of the components of geminiviruses with different common IRs of <75% nucleotide identity are apparently not infectious (Gilbertson *et al.*, 1991; Lazarowitz *et al.*, 1992; Stanley and Gay, 1983), the LIR sequence identity between MSV-Kom and MSV-Set is only 71%. This is the lowest IR identity between geminiviruses where transreplication has been reported.

In addition, this thesis reports the infectivity of recombinant progeny, pSeK423 and pKoS005, both comprising MSV-Kom wild-type sequences but with MSV-Set sequence replacements in the LIR of 105 bp and in the V1 ORF of 176 bp (effectively replacing 57% of the N-terminus of the MP). Although their infectivities and symptom severity have been reduced, it appears that transcomplementation - a criterion for demarcating geminiviruses - does occur between genes or regions of MSV-Kom and MSV-Set. Further, MSV-Kom and MSV-Set CP and MP gene replacement experiments were shown to be infectious despite their genome sequence and biological experiments (Eric van der Walt, Honours thesis 1996, and pers. comm.).

These transcomplementation data indicate that MSV-Set is an MSV strain rather than than a new *Setaria* grass virus species, according to the criteria for demarcating geminiviruses. This is in line with the overlapping host ranges enjoyed by MSV-Kom and MSV-Set, but however, conflicts with the sequence criteria as presently constituted (e.g. Padidam *et al.*, 1995). It can be argued, according to the sequence criteria, that since the sequences have identities significantly below the suggested 90% criterion, this, along with their biological differences, indicates that the virus should be classed as a new species. On the other hand, the corollary of this would be that the sequence criteria may need to be reconsidered for Mastreviruses, as perhaps these can be transreplicated by a wider range of sequence variants than can Begomovirus genomes. This is presently under investigation by Janet Willment in this laboratory.

6.2 The role of the two LIRs in agroinfection

pKom504 and pSet102 contain homodimers of 1.4 and 1.8 unit-length genomes of MSV-Kom and MSV-Set respectively; however they have only one LIR each. The agroinfection rates of the 1.1mer counterparts with redundancies of a LIR each (pKom603 and pSet107), were dramatically increased from 38% and 28% for the 1.4- and 1.8mers, to 100% and 90% respectively for the 1.1mers. Since the common differences in these clones were that the more agroinfectious 1.1mers contained 2 LIRs as opposed to one, one could assume that the presence of two LIRs separated by one genome length is an important factor in these agroinoculation tests. In addition only those MSV-Kom/MSV-Set heterodimers containing at least a set of four ORFs delimited by two LIRs were infectious. These results are in line with heterodimer inoculations of the Worland and Logan strains of BCTV performed by Stenger *et al.*, 1991) and of the WDV-S and WDV-Er strains performed by Heyraud, *et al.*, (1993).

This phenomenon can be explained by the model proposed by Elmer *et al.*, (1988) and Stenger *et al.*, (1991) in which the mechanism of release of unit-length genomes is more

efficient than the random homologous recombination model. This was evident since the recombination sites of most recombinant progeny that were mapped and/or partially sequenced were mapped to the region delimited by hairpin loops of the different genomes. In fact, Stenger *et al.* (1991) were able to map initiation and termination of geminiviruses within 20 bp in the stem of the hairpin. Further, Heyraud *et al.*, (1993) mapped the initiation nick site within the conserved nonanucleotide TAATATTAC, which had implications of determining the current model for RCR in geminiviruses. In this study, analyses of partial sequences of recombinant progeny from heterodimer agroinoculated plants revealed that the majority of them appeared to originate from the right of or from within the stem of the hairpin and slightly outside the region delimited by the hairpin loops, suggesting an important role for this region in the replicational release mechanism. However the possibility that release was occurring as a result of "hotspot" random homologous recombination within this region could not be ruled out.

Additional agroinoculations (Chapter 5) of 1.1mer clones containing MSV-Kom wild-type sequence, but with only two MSV-Set bases replaced just left of the identified ATG/*Bam* HI preferred recombinational site, either in the left LIR (pGC/TT2; Figure 5.1) or in the right LIR (pTT/GC2; Figure 5.1), indicated that replicative release in MSV appeared to occur nine times more frequently than homologous recombinational release from the region delimited by this preferred crossover site. This was evident from the partial sequences of 10 progeny genomes from pGC/TT2 or pTT/GC2 where GC or TT was switched with TT or GC respectively. This 9/10 replicational release frequency is probably an overestimate, since this experiment did not consider other preferred crossover sites in the LIR, however is a good indication that replicational release from dimers occurs more efficiently. The precise mechanism of release of the original recombinants was not established (see Chapter 3); however, it may very well have involved a mixture of replicational release, followed by over-elongation (i.e. a lack of second LIR nicking) of ssDNA, followed by a "resolving event" which may have involved homologous recombination (see 4.5.1).

With the establishment of the role of the LIRs in agroinoculation tests, heterodimer and chimaeric 1.1mers were subsequently cloned in this thesis in further investigations. For example, all the agroinfectious clones reported in Chapter 5 are 1.1mers containing a set of four ORFs delimited by two LIRs. In addition, in all bombardment inoculations and agroinoculations of wild-type, gene replacement or mutant infection studies in this laboratory, clones are now constructed containing at least two LIRs delimited by the ORFs (K. Palmer, J. Willment, D. Martin and E. van der Walt, pers. comm.).

On a more practical side, the wild-type pKom603 and pSet107 1.1mer clones are frequently used in this laboratory as positive controls or used in gene replacement experiments as comparative controls (K.palmer, J. Willment, D Martin and E van der Walt). In addition these wild-type agroinfectious constructs (pKom603, pSet107 and pPS103) can be potentially used in susceptibility or tolerance tests on new cv.s or prospective transgenic plants.

6.3 Generating viable recombinant progeny from heterodimers

MSV-Set is the most distinct MSV strain reported so far and the only MSV strain to be isolated from a grass species. It was thus decided to investigate recombination and their effects between these viruses in forced recombination tests using the heterodimer agroinoculation technique. The recombinant progeny was isolated from agroinfected plants 3-5 weeks after agroinoculation to specifically target viable recombinants as opposed to less viable or crippled recombinants by "natural selection". It is evident that less viable genomes are supplanted by more viable genomes in mixed populations within a plant (Bisaro, 1994). As a consequence of this selection, recombinants with sequence replacements of up to 176bp were identified.

Agroinoculation tests of these chimaeras with either a 105bp replacement in the LIR or a 176bp replacement in the V1 ORF confirmed that recombinants were indeed infectious and thus can be used in determining the function of cognate sequences of two different viruses from a variety of recombinants. This work has demonstrated that it may be possible to pinpoint very focussed regions of the Mastreviral genomes responsible for some of their biological activities and properties by using recombinants in agroinoculation tests as described in this thesis. For example, these studies showed that it is possible to replace half the MP gene of MSV-Kom without affecting host range; however, the symptoms and agroinfectivities were affected (i.e. reduced, Chapter 5).

Previous heterodimer studies have generated recombinant progeny which, however, have nothing reported on whether they were infectious, and if so, how these sequence replacements affected the biology of these viruses as compared to the wild-types. The heterodimer studies of Stenger *et al.* (1991) and Heyraud *et al.* (1993) were perfomed on the Worland and Logan strains of BCTV which shared a genome sequenced identity of 80% or WDV-S and WDV-Er which shared a genome sequence identity of 82.5% respectively. This study however investigates the "forced" recombination between MSV-Set and MSV-Kom, which share a genomic sequence identity of only 78%, via heterodimer agroinoculation. From the results obtained in these experiments including the determination of the infectiousness of recombinant progeny, it seems likely that recombination can occur between

MSV-Kom and MSV-Set in mixed infections in nature, and likely between any other strains of the virus.

6.4 Future investigations

The biological and genomic characterisation of MSV-Kom, MSV-Set and PanSV-Kar has laid down a basis for further investigations into the Mastreviruses in this laboratory. Apart from providing wild-type agroinfectious clones for comparative and/or positive controls in agroinoculations of gene replacement studies and/or transcomplementation studies, the heterodimer agroinoculation technique provides a potentially useful method for generating sequence replacement progeny that may be difficult to clone due to sequence and/or restriction endonuclease site limitations. It also establishes a method to select viable as opposed to less viable recombinants that may not be useful since sequence replacements could render them crippled. It would however be interesting to study initial unit-length releases that would probably have more substantial replacements as a result of random homologous recombination. The isolation and cloning of such recombinants could probably be achieved in protoplasts (Heyraud *et al.*, 1993): one could probably generate useful initial recombinants with more substantial replacements (that are more likely to be less viable), in callus bombardment, or protoplast transfection experiments, with progeny genomes selected at an early stage of infection (K. Palmer, J. Willment, pers. comm.; Heyraud *et al.*, 1993).

Agroinfectious clones of chimaeras of MSV-Kom and MSV-Set such as pKeSeK2, pSeK423, pGC/TT2, pTT/GC2 and pKoS035 have been provided for further investigations into different aspects of these viruses (J. Willment, D. Martin and E. van der Walt, pers. comm.). Interesting isolated and cloned recombinant progeny generated in heterodimer agroinoculations, which have not been made agroinfectious as yet, have also been provided for further investigation. These include pSeK009, which has a 174 bp replacement in the LIR including the hairpin loop, and effectively has MSV-Set iterons in the stem of the MSV-Set hairpin, and an iteron located to the left of the hairpin from MSV-Kom; thus pSeK009 and pKeSeK2 are clones that can be used to further investigate the proposal that the iterons are specific sites for Rep binding in initiating replication of Mastreviruses (Argüello-Astorga *et al.*, 1994). pSeK011, which has a 105 bp replacement right of the hairpin loop in the LIR effectively replacing the DNA bending locus normally associated with origins of replication (Suàrez-Löpez *et al.*, 1995), is a clone that could be used to investigate the role of this bending site in Mastreviral LIRs. pKoS005 and pKoS007, which have sequence replacements of 176 bp and 25 bp in the V1 ORF effectively replacing 57% and 8% amino acid sequences of the N-terminus of the MP with cognate sequence respectively, could be

used in partial gene replacement studies to identify the functions of the N-terminus of the MP.

MSV-Kom and PanSV-Kar heterodimers, pKoP1 and pPaK1 constructs have already been constructed, and as yet have not been used in agroinoculation tests. It is hoped to continue recombinational studies between these viruses, which only share genomic sequence identity of 60%.

APPENDICES

APPENDIX A: DNA TECHNIQUES

A1. restriction endonuclease digestions

For single digests the following constituents were placed into a 1.5 ml eppendorf-type vial in sequence:

1μg plasmid DNA

2μl appropriate restriction buffer [10x](supplied with restriction endonuclease)

MilliQ®-filtered distilled water to 19.5μl

0.5μl restriction endonuclease [2-5U] (Boehringer-Mannheim®)

The eppendorf vial was incubated at the restriction endonuclease's optimal temperature in a waterbath or incubator for 1-2 hours. Most enzymes function optimally at 37°C. The procedures used were essentially those described by Sambrook *et al.* (1989). Boehringer-Mannheim restriction endonucleases are usually supplied in a concentration of 10U/μl. Double and triple digests were carried out simultaneously if the salt and temperature requirements of the enzymes were compatible. If not, sequential digests were done using the enzyme with the lowest salt optimum followed by adjusting the salt concentration and adding the second enzyme. Salt concentrations were adjusted by adding 2μl of the optimal restriction buffer for the second enzyme and so on. If the optimal temperatures for the enzymes differed, the mixture was incubated at appropriate temperatures corresponding to that current enzyme's optimals. If the digested DNA products were required for further manipulation such as ligations, they were subjected to ammonium acetate / isopropanol / ethanol precipitation as described below. If the digested DNA, such as total plant DNA extract was required for Southern blots (Appendix A13) the digestions would be loaded directly into agarose gels (Appendix A3)

Ammonium acetate / isopropanol / ethanol precipitation of DNA

One fifth the volume of the digestion of 5M ammonium acetate was added to the digestion and mixed. Twice the total digestion/ammonium acetate volume of isopropanol was added to the mixture, which was then left for five minutes at room temperature. The DNA was centrifuged down for 15 minutes in a desktop microfuge at (14 000rpm). The supernatant was discarded and 100 - 500 μl 70% ethanol was added and the DNA pelleted for 5 minutes in a centrifuge. The DNA pellet was allowed to air dry. The DNA was resuspended in water, or TE buffer (10mM Tris-HCl pH 7 or 8; 1mM EDTA) was added to obtain a DNA concentration of 100ng/μl for further manipulation or cloning.

A2. Partial DNA restriction digests

Partial digests were performed when a single cleavage was required on a plasmid construct containing more than one restriction site for that particular enzyme. Partial digestions would ensure that a sufficient proportion of the plasmids would be cut at the required site/s and the remaining proportion/s would be uncut or cut at the site/s that should be intact. The method followed is described in Sambrook *et al.* (1989). A "master mix" was made up in the following sequence in an eppendorf tube:

1100ng DNA

11µl appropriate restriction buffer

MilliQ®-filtered distilled water up to 110µl

19µl master mix was aliquoted into an eppendorf tube labelled 1 and 10µl into eppendorf vials labelled 2 to 10. Tubes 1 to 10 were placed on ice to ensure that the restriction enzyme was inactive. 10U or 1µl restriction enzyme [10U/µl] was added to tube 1 and mixed. 10µl of the the contents of tube 1 were transferred to tube 2 and the contents of tube 2 were mixed. 10µl of the contents of tube 2 were transferred to tube 3 and the contents of tube 3 were mixed. This procedure was followed up to tube 10 and the remaining 10µl was discarded. The tubes 1 to 10 were a 1:2 serial dilution of the restriction enzyme digesting 100ng of DNA. The tubes were incubated at 37°C for 1 hour. 2µl loading buffer (see Appendix A4) was added to each of the tubes 1 to 10 and the contents of the tubes were loaded into ten wells of a TBE or TAE agarose gel with ethidium bromide, and electrophoresed. See Appendix A3 for agarose gel electrophoresis (AGE). DNA bands were visualised on a 305nm wavelength UV transilluminator and photographed using a Polaroid 667 or 665 film or printed on a UVP™ GDS image analyser and printed. The lane with the most intense DNA band of the required size was chosen and amount of the enzyme required to partially restrict DNA accordingly was calculated as follows:

$$1/20 \times [1/2]^{(\text{no. tube} - 1)} = \text{dilution factor for enzyme required for optimum partial digestion}$$

1/20 is the dilution of enzyme in tube 1 and 1/2 to the power of the number of tube minus 1 is the dilution factor of the enzyme in the 1:2 dilution series. The digest was then scaled up by 10 to 50 to partially digest 1 to 5µg DNA in the appropriate dilution of enzyme that would digest partially in one hour. The digest was then ammonium acetate / isopropanol precipitated, as described in Appendix A1, and used for further construction of clones or manipulation.

A3. Agarose gel electrophoresis (AGE) of DNA

Electrophoresis of DNA was performed using horizontal slab-gel systems, as described in Sambrook *et al,* 1989. 1xTris-borate-EDTA (TBE) or 1x Tris-Acetate-EDTA (TAE) running buffers was used as the gel and tank buffer. 10 x TBE consisted of 108g Tris-HCl, 55g boric acid and 20ml 0.5M EDTA made up to 1ℓ. 50 x TAE consisted of 242g Tris-HCl, 57ml glacial acetic acid and 100ml 0.5M EDTA in 1ℓ. Molten agarose was prepared using a microwave oven. Agarose concentration varied between 0.6 and 1.2% depending on the sizes of DNA fragments being analysed. Ethidium bromide (1mg/ml aqueous stock solution) was added to the gel mix before pouring to a concentration of 25ng/ml. DNA samples were mixed with 1/5 volume of loading buffer [0.25% (w/v) bromophenol blue, 40% (w/v) sucrose in 1 x TBE] and electrophoresed at 4-8V/cm for 1-5 hours depending on the size of the gel apparatus used. DNA bands were visualised using a 305nm or 295nm wavelength UV transilluminator and photographed with Polaroid 667, 665 film or the UVP™ GDS image analyser which was able to print and store images as bit-map files on computer "stiffy disks". DNA fragments were sized by comparing the distances the bands migrated to the bands of known size of Pst I-digested λ DNA which was used as the molecular weight marker.

A4. DNA ligations

1/5th or 1/10 the amount of appropriately digested vector (pUC18 or pBI121 respectively) was added to an amount of appropriately digested or partially digested insert to make a final amount of 200-250ng of DNA in a ligation experiment. The ligation mixes were made up as follows:

20-50ng digested vector DNA
180-200ng digested insert DNA
2µl 10x ligation buffer (Boehringer-Mannheim)
MilliQ®-filtered water to19.5µl
0.5µl T_4 ligase [10U/µl] (Boehringer-Mannheim)

The mixture was incubated at 15°C overnight and 3-5µl of the mixture was used in *E.coli* transformations immediately or stored at -20°C and used to transform *E.coli* at a later date. 10x ligation buffer consisted of 0.5M tris-HCl pH 7.6, 0.1M $MgCl_2$, 10mM ATP and 10mM dithiothreitol.

For religations of digested clones the ligation mix was prepared as above except that 50-100ng appropriately digested plasmid DNA was added to the above mix. Religation circularises digested plasmid DNA in the event that the restriction sites are complementary.

Religations using single-restricted (e.g. *Bam* HI) pUC18 were put up in a similar fashion and subsequently used in transformations (Appendix A5) as a positive ligation control. Negative ligation controls were performed by transforming *E.coli* cells using unligated single-restricted pUC18.

A5. Introduction of recombinant plasmids into *E. coli* by transformation

E.coli strain LK112 was used for constructing agroinfectious clones and strains JM105, JM109 or DH5α were used for clones that were to be sequenced. Two methods were used to make competent *E.coli* cells and tranform them; the method described by Chung and Miller (1988) or the method described by Armitage *et al*, (1988).

The Chung and Miller method

A 1:100 dilution of an overnight culture of *E.coli* in Luria-Bertani (LB) broth was made in 25ml LB. The culture was incubated at 37°C with constant agitation until an OD_{600} of 0.3-0.5 was obtained. The culture flask was rapidly chilled in ice-water and the culture pelleted by centrifugation at 5000rpm at 4°C for 10 minutes in a precooled Sorvall SS34 rotor. The bacterial pellet was resuspended in 2.5ml volume cold transformation storage buffer (TSB) (LB pH6.1, PEG-4000, 10mM $MgCl_2$, 10mM $MgSO_4$, 5% dimethylsulphoxide [DMSO]) and the suspension kept on ice for 10 minutes to allow cells to develop competence. Competent cells could be stored in TSB at -70°C, if required, with approximately 10-fold decrease in competence.

3-5µl ligation mix (Appendix A4) was dipensed to a sterile eppendorf tube and prechilled on ice. 100µl of the competent cell suspension was added to the 3-5µl DNA ligation in the eppendorf tube. Tubes were kept on ice for 20-30 minutes. The 0.9ml TSB containing 20mM glucose was added and the transformation mix was incubated at 37°C for one hour to allow transformed cells to express the pUC18-encoded ampicillin-resistance gene, β-lactamase. When pUC18 plasmid was used, 5µl of 200mg/ml IPTG and 50µl X-gal were spread onto Luria Agar (LA) plates containing 100µg/ml ampicillin. Ampicillin, IPTG and X-gal are required for screening for acquired ampicillin resistance and blue/white *E.coli* phenotypes that have acquired the ligated plasmids. Colonies containing the desired recombinant plasmid grow and appear white on LA + Ampicillin + X-gal + IPTG plates when pUC18 vector is used (Appendix A1). This screening procedure is based on the insertional inactivation of the α-fragment of the *lacZ* gene which is in pUC19 and pUC18 (Yanisch-Perron *et al,* 1985). These vectors are transformed into E.coli strains that contain a second defective *lacZ* gene. The *lacZ* genes complement each other in transformants and isopropylthio-β-D-galactoside (IPTG) induces the genes to produce an active β-galactisidase

that metabolises the chromogenic X-Gal, producing blue colonies. When working with pBI121, cultures were plated onto LA + kanamycin (50μg/ml) since this plasmid has a kanamycin resistance gene and does not have the blue/white selection feature (Appendix A2). 50-100μl of the transformed cultures were plated onto Luria Agar (LA) ampicillin, IPTG and X-gal. The agar plates were incubated at 37°C overnight. Positive transformation controls were prepared by transforming 10ng intact covalently closed circular pUC18 vector as described above and spreading onto a separate LA plate. A negative transformation control was prepared by spreading an equal amount of competent cells not treated with DNA on a separate LA ampicillin plate. Required clones were found by colony blotting (Appendix A7) and/or restriction digestions of plasmid minipreparations (Appendix A6) of white colonies using restriction endonucleases that would distinguish the desired clones.

Armitage, Walden and Draper method

5ml sterile broth was inoculated with antibiotic sensitive colony and incubated and aerated at 37°C overnight. The starter culture was inoculated into 100ml prewarmed sterile LB in a conical flask. The cells were aerated at 37°C until the OD_{600} reading reached 0.3-0.5 when the culture was in its logarithmic stage. The cultures were transferred to centrifuge tubes and cells were spun down at 5000rpm at 4°C for 5 minutes in a SS34 Sorvall rotor. The supernatant was discarded and the cells were resuspended in 10mls ice-cold TFB1 (100mM RbCl, 50mM $MnCl_2$, 30mM KOAc, 10mM $CaCl_2$, 15% glycerol), pooled and put on ice for 90 minutes. The cells were spun down as decribed above. Cells were resuspended in 3ml ice-cold TFB2 (10mM MOPS pH 7, 10mM $RbCl_2$, 75mM $CaCl_2$, 15% glycerol). 50-100μl were aliquoted into 1.5ml eppendorf tubes. These cells were either used immediately in transformations or flash-frozen in liquid nitrogen and stored at -70°C for later use. Competent cells were thawed at room temperature until the suspension was liquid and incubated on ice for 10mins. Up to 10ng DNA was added to the cells (including DNA required for appropriate positive and negative ligation and transformation controls), mixed and incubated on ice for 20 minutes. Cells were heat-shocked in a 40°C waterbath for 2 minutes and put on ice for 2 minutes. 900μl sterile LB was added to each transformation and aerated by shaking for an hour. 50-100μl of the transformed cultures were spread onto LA + Ampicillin + X-gal + IPTG plates for pUC18 vector plasmids and LA + kanamycin for pBI121 vector plasmids. Desired colonies were found by colony blotting and/or restricting plasmid minipreparations with restriction endonucleases that would distinguish the desired clones (Appendices A7 and A6 respectively).

A6. DNA plasmid minipreparation

Single colonies harbouring putative recombinant plasmids were picked and inoculated into 5ml LB containing selective antibiotics (i.e. 100μg/ml ampicillin for pUC18 recombinants and 50μg/ml for kanamycin recombinants) and aerated at 37°C overnight. 1.5ml of culture were transferred into eppendorf tubes and microfuged at 12000-14000 rpm for 2 minutes. Supernatant was discarded and cells were resuspended in 100μl lysis solution (20mM Tris-HCl pH 8, 10mM EDTA, 0.5M glucose). 200μl alkaline SDS solution (0.2M NaOH, 1% SDS) was added to the suspensions and mixed until clear. 150μl 3M NaOAc (pH 4.8) was added and mixed and the white flocculant was spun down at 12000-14000rpm for 10 minutes. The supernatant was then transferred to a clean eppendorf tube containing 0.5ml isopropanol. After mixing well, the plasmid DNA was pelleted in a microfuge at 12000-14000rpm for 15 minutes. The supernatant was discarded and 0.2ml 70% ethanol was used to clean the DNA pellet and microfuged for 5 minutes. The pellet was airdried and resuspended in 50μl TE buffer.

Plasmid DNA minipreparations were selected on the basis of restriction digests which were performed as follows. The following constituents were dispensed into an eppendorf tube;

1μl minipreparation DNA
1μl appropriate restriction buffer (see B1)
0.5μl appropriate restriction endonuclease
7.5μl distilled MilliQ water

The mixture was incubated for 30-60 minutes and 2μl loading buffer (see B3) was added to the mix. This mix was then run on AGE (see B3).

A7. Preparation of probe by Digoxigenin-labeling of DNA

DNA was labelled using the Boehringer-Mannheim Random Primed DIG-labeling and Detection Kit. The desired DNA (i.e. MSV-Set, MSV-Kom or PanSV-Kar genomic DNA) was *Bam* HI-restricted and isolated from the pUC18 vector by TAE or TBE AGE (see Appendix A3) and subsequent elution of DNA using the Biolab101 Geneclean® kit.

Elution of DNA from agarose

The desired DNA fragment visualised on a 305nm UV transilluminator was excised from the gel, using a clean scalpel blade. The agarose block was placed in a 1.5 or 2.2ml eppendorf tube. The weight of the agarose was determined by the difference between the weight of an empty eppendorf tube with the one that has the agarose excised with desired DNA fragment. Three volumes NaI solution (90.8g NaI and 1.5g Na_2SO_3 in a final 100ml volume) provided in the kit, was added to the agarose if TAE was used in AGE or four volumes NaI solution if

TBE was used. Agarose was then incubated in a 55°C waterbath until the agarose had dissolved completely. If TBE AGE had been used TBE modifier solution was added. 5µl glass milk suspension (nitric acid treated silica powder, 325 mesh) was mixed in the eppendorf tubes and incubated at room temperature for 10 minutes and 5 minutes on ice with occasional mixing. The beads were microfuged down for 20 seconds and the supernatant discarded. The glass beads were resuspended in 500µl ethanol wash (50% ethanol, 100mM NaCl, 10mM Tris-HCl, pH7.5, 1mM EDTA) was two or three times. The beads were then resuspended in an appropriate volume MilliQ® water or TE buffer and incubated at 55°C for 2 minutes. The beads were spun down in a microfuge for 20 seconds and the supernatant containing desired DNA was transferred into a new tube.

DIG-labelling of DNA

DNA to be used as probes Southern or colony blots were labelled using the random primed DIG DNA labeling and detection kit from Boehringer-Mannheim. The following reagents were added to a sterile eppendorf tube;

10ng-1µg DNA template

MilliQ®-filtered distilled water to 15µl

DNA was denatured by placing the eppendorf tube in 95°C waterbath for 10 minutes then the following was added after the eppendorf tube was put on ice;

2µl 10x random hexanucleotide mixture

2µl 10x dNTP labelling mixture with DIG-11-dUTP

1µl Klenow enzyme (1U/µl)

The eppendorf tube was incubated at 37°C overnight to 20 hours. Labelled DNA was stored at -20°C. Labeled DNA was used in hybridisation tests described in Appendix B9.

A8. Colony blotting

Colonies were picked using sterile toothpicks and spotted, equally spaced, onto LA plates with appropriate antibiotics (i.e. 100µg/ml ampicillin if pUC18 plasmid vector was used and 50µg/ml kanamycin if pBI121 plasmid vector was used) in duplicate. One of the replica plates were subjected to colony blotting screening for desired colonies. The method for colony blotting described in "Protocols for nucleic acid blotting and hybridisation" for Hybond-N+® from Amersham was employed. Hybond-N+ nylon membrane was cut to fit LA plates. Membrane was carefully placed onto the colony/agar surface and marked to ensure correct orientation of colony side. After a minute the membrane was removed and placed, colony side up, on a pad of absorbant filter paper soaked in denaturing solution (1.5M NaCl, 0.5M NaOH) for 7 min. The membrane was placed, colony side up, on a pad of absorbant paper

soaked in neutralising solution (1.5M NaCl, 0.5M Tris-HCl pH 7.2, 1mM EDTA) for 3 min. and this step was repeated on a fresh pad soaked in nuetralising solution. Membrane membrane was briefly washed in 2x SSC (20x SSC; 3M NaCl, 0.3M Na_3 citrate) and placed, colony side up, on a pad of absorbant filter paper soaked in 0.4M NaOH for 20-30 min. to fix the DNA on the membrane. The membrane was washed in 5x SSC and airdried and stored in Gladwrap® at 4°C or immediately used in hybridisation tests (Appendix B9).

A9. Hybridisation tests

The hybridisation method described in the DIG-labelling and detection kit manual from Boehringer-Mannheim was followed. The filter treated as described in Appendix A8 or A14 colony blots or capillary blots respectively), was sealed in an hybridisation bag or box containing approximately 20ml prehybridisation solution (5x SSC, 1% [w/v] Boehringer-Mannheim blocking reagent, 0.1% N-lauroylsarcosine, 0.02% SDS) for each 10cm^2 nylon membrane. The blots were prehybridised with gentle shaking for 3-6 hours at hybridisation temperature of 68-70°C. 3-5µl of DIG-labelled DNA probe, prepared in Appendix A7, was denatured at 95°C for 5 minutes, and added to hybridisation solution (same constituents as prehybridisation solution described above). Prehybridisation solution was replaced by hybridisation solution containing 5-25ng/ml single-stranded DIG-labeled DNA probe, and hybridisation was allowed for 16 hours in the same conditions as described for the prehybridisation. Hybridisation solution was stored at -20°C and reused, after heating at 95°C for 10-20 minutes, in other experiments. Membranes were washed twice in 2x wash solution (2x SSC, 0.1% SDS) for 5 minutes with gentle agitation at room temperature and twice in 0.1x wash solution (0.1x SSC, 0.1% SDS) with gentle agitation for 15 minutes at the hybridisation temperature. Hybridisation signals were detected immediately as described in Appendix A10.

A10. Chemiluminescent and NBT/X-phosphate colourimetric detection

After hybridisation and post hybridisation washes the nylon membrane was equilibrated in buffer 1 (100mM Tris-HCl, 150mM NaCl, pH 7.5) for 1-2 minutes at room temperature. The membrane was blocked by gentle agitation in buffer 2 (1% [w/v] blocking reagent from Boehringer-Mannheim DIG-labelling and detection kit, dissolved in buffer 1) for 30 minutes at room temperature. The membrane was subjected to gentle agitation in buffer 2 containing goat anti-DIG-alkaline phosphatase conjugate (150U/µl) diluted 1:10 000 for 30 minutes. The membrane was then washed for 15 minutes in buffer 1 containing 0.3% Tween® 20 twice. The membrane was then equilibrated in buffer 3 (100mM Tris-HCl pH9.5, 100mM NaCl, 50mM $MgCl_2$) for 1-2 minutes.

For chemiluminescent detection Lumigen™PPD from Boehringer-Mannheim was diluted 1:100 in buffer 3 and sealed in a plastic bag with the nylon membrane and exposed to Agfa Curix P1 X-ray film in an X-ray cassette for 5-60 minutes. The X-ray film was then developed in an OKAMOTO™ Medical X-ray film automatic processor X2.

For NBT/X-phosphate colourimetric detection, 45µl NBT(75mg/ml nitroblue tetrazolium salt in 70% [v/v] dimethylformamide) and 35µl X-phosphate (50mg/ml 5-bromo-4-chloro-3-indolyl phosphate toluidinium salt in 100% dimethylformamide) was added to 10ml buffer 3. This freshly prepared colour substrate was sealed in a plastic bag with the hybridised nylon membrane. Since the reagents for the colour substrate are light sensitive the colour reaction was left in a dark cupboard or drawer. The colour precipitate forms from 5 minutes to overnight directly on the nylon membrane. Once the desired spots or bands were detected the membrane was washed in water or TE buffer to stop the colour reaction.

In cases where the Southern blot was reprobed using a different DNA probe the blot was not allowed to dry. The nylon membrane was stripped by incubating twice for 20 minutes in a 0.2M NaOH, 0.1% [w/v] SDS solution at 37°C. The membrane was then washed in 2x SSC. The membrane was then dried or immediately used in a hybridisation test (Appendix A9) and subsequent detection as described above.

A11. Maxipreparations of plasmid DNA

Deletion plasmid templates for sequencing, subclones and plasmid constructs were prepared using silica-based anion-exchanger columns, a commercially available plasmid isolation kit (Nucleobond® AX100 columns manufactured by Macherey-Nagel), according to the manufacturer's instructions. 100-150ml LB containing the appropriate selective antibiotic in a conical flask was inoculated with the desired *E. coli* clone and aerated at 37°C overnight. Cells were centrifuged at 5000rpm at 4°C for 5 minutes using a Sorvall SS34 rotor. Cells were resuspended in 4ml resuspension buffer, S1 (50mM Tris-HCl, 10mM EDTA, RNase A 400µg/ml, pH 8). 4ml of lysis buffer, S2 (200mM NaOH, 1% SDS), was added and mixed. 4ml neutralising buffer, S3 (2.55M potassium acetate pH 4.8), was added and mixed. The resulting white flocculant was spun down at 15 000rpm at 4°C for 40 minutes. 2ml equilibration buffer, N2 (100mM Tris/H_3PO_4, 15% ethanol, 900mM KCl, pH 6.3) was used to equilibrate a silica based anion exchange AX100 column. The supernatant was loaded into the column and allowed to pass through. 4ml wash buffer, N3 (100mM Tris/H_3PO_4, 15% ethanol, 1300mM KCl, pH 6.3), was passed through the column and this wash was repeated. Plasmid DNA was eluted by passing 2ml elution buffer, N5 (100mM

Tris/H_3PO_4, 15% ethanol, 1000mM KCl, pH 8.5), through the column. 1ml aliquotes of the elute was collected in 2.2ml microfuge tubes containing 1ml isopropanol and the tubes mixed. The plasmid DNA was microfuged for 15 min. The supernatant was discarded and the DNA pellet was washed by adding 0.5ml 70% ethanol and microfuging down for 5-10 minutes. The DNA was airdried in the microfuge tubes or 20-30 minutes and resuspended in 50-100µl TE buffer. DNA was stored at 4°C.

Nucleobonded DNA was quantified by diluting 1µl DNA in 99µl TE buffer in a 100µl quartz cuvette and scanning absorbance reading in the UV wavelength range from 210-310nm. The reading at 260nm was used in the following formula (Sambrook *et al*, 1989); $1Abs_{260}$ = 50µg/ml. (i.e the absorbance reading at 260nm was multiplied by 5000 for the concentration in ng/µl). 100ng DNA was digested with 3 or 4 selected restriction endonucleases to distinguish and confirm the desired clone or construct on an agarose gel (Appendix A6 and A3).

A12. Plant DNA extraction

Total DNA was isolated from healthy and infected plants by a modification of the method of Ikegami *et al.* (1981). 5-10g of plant material was frozen in liquid nitrogen and ground to a fine powder in a mortar and pestle or in a Bosch® kitchen coffee grinder. An equal volume of extraction buffer (0.1M Tris-HCl pH 7.0, 0.1M NaCl, 0.1M EDTA, 1% [w/v] SDS) was added to the fine powder in a beaker and stirred until it reached room temperature. The homogenate was centrifuged in 30ml polypropylene tubes at 12 000 x g (10 000 rpm in a Sorvall SS34 fixed angle rotor) for 10 minutes to pellet undissolved plant debris. The resulting supernatant was transferred into polypropylene tubes containing 10-15ml tris-buffered phenol, mixed and centrifuged for another 10 minutes to separate the aqueous and organic phases. The top aqueous phase was transferred into a polypropylene tube containing 10-15ml chloroform, mixed and centrifuged as above. The aqueous phase was transferred to another polypropylene tube containing a 10-15mls isopropanol and mixed gently. The DNA was spun down at 17 000 x g (12 000 rpm) for 20 minutes, the supernatant was discarded and the pellet was allowed to air dry for 10 minutes. The plant DNA extract was then resuspended in TE buffer. DNA was quantified as described in Appendix A11.

A13. Capillary transfer of dna from agarose gels to nylon membranes

Agarose gels were submerged in denaturing solution (0.4M NaOH) for 30 minutes with gentle agitation and at room temperature. The gel was placed, face-down, on the platform or bridge of a capillary transfer system filled with denaturing solution. The system was made up of a platform placed in a reservoir containing denaturing solution. A wick made of 3MM

paper was placed on the platform soaked in denaturing solution with the ends hanging submerged in the denaturing solution of the reservoir. The width and length of the platform usually corresponded to the size of the gel and the wick was cut to the same width. The platform and reservoir were made to the same height. Positively charged Hybond-N+® nylon membrane obtained from Amersham was cut to the same size of the agarose gel. This was placed directly on the face-down gel on the platform. The membrane was wetted with denaturing solution and all air bubbles removed between the 3MM absorbent layers, the gel and especially between the gel and membrane to ensure complete transfer of entire gel. The membrane was marked to ensure which was the DNA side. 2-3 pieces of 3MM paper cut to the size of the gel were placed onto the membrane. Open parts of the resovoir surrounding the platform and exposed parts of the platform or wick were sealed using Gladwrap®. A stack of absorbent paper towels was placed on top of the 3MM paper and a 1-2kg weight was placed on top of the stack. Capillary transfer of DNA was allowed overnight. The system was disassembled and the nylon membrane washed in 2x SSC. The blot was then used in hybridisation tests or airdried and stored in Gladwrap® at 4°C for later hybridisation tests s described in appendix A9.

A14. Isolation of RF-DNA from total plant DNA extractions

RF-DNA was isolated from the total plant DNA extract by alkaline denaturation and anion-exchange chromatography using the commercially available Qiagen® plasmid isolation kit (Qiagen GmbH and Qiagen Inc.) with Qiagen tip-20 columns. The total plant DNA pellet was resuspended in 0.3ml resuspension buffer, P1 (100µg/ml Rnase A, Tris-HCl, 10mM EDTA, pH 8.0, stored at 4°C). 0.3ml of lysis buffer, P2 (200mM NaOH, 1% SDS), was added to the DNA suspension, mixed and incubated for 5 min. 0.3 ml neutralising buffer, P3 (3.0M Kac, pH5.5, stored at 4°C), was added to the mixture and placed onto ice for 10min after mixing. The resulting white flocculant was centrifuged down for 15 min at 12 000rpm in a microfuge. A Qiagen-tip 20 was equilibrated by passing 1ml equilibration buffer, QBT (750mM NaCl, 50mM MOPS, 15% ethanol, pH 7.0, 0.15% Triton X-100, stored at room temperature) through it. The supernatant was applied to the equilibrated column and allowed to pass through the resin by gravity flow. The column was washed four times with 1ml of wash buffer, QC (1.0M NaCl, 50mM MOPS, 15% ethanol, pH 7.0, stored at room temperature). RF-DNA was eluted by adding 0.8ml elution buffer, QF (1.25M NaCl, 50mM Tris-HCl, 15% ethanol, pH 8.5, stored at room temperature), and the eluted DNA was collected in a microfuge tube. 0.7ml isopropanol was added and mixed, and the DNA was spun down at 12 000rpm for 30min in a microfuge. The supernatant was removed and the DNA pellet was washed in 70% ethanol, spun down for 5min in a microfuge and airdried. DNA was

resuspended in MilliQ®-filtered water or TE buffer. DNA was quantified using a spectrophotometer as described in Appendix A11.

A15. Direct *Agrobacterium* transformation: freeze-thaw method

Method used was from An *et al.*, (1988). Once a desired clone is constructed in *E. coli*, the dimer cloned in pBI121 was transferred directly into *Agrobacterium* by the freeze-thaw method. Agrobacterium was grown overnight in 5 ml LB at 28°C. 2 ml of this was overnight culture was added to 50 ml of LB in a 200 ml flask and aerated at 28°C until the culture grew to an $O.D_{600}$ of 0.5 to 1.0. The culture was chilled on ice and centrifuged for 5 min at 5000rpm and 4°C. The supernatant was discarded and the cells were resuspended in 1 ml of ice-cold 20 mM $CaCl_2$ solution. 100 µl aliquotes were dipensed into eppendorf vials containing the 1 µg of desired plasmid. The cells were quick frozen in liquid nitrogen and then thawed at 37°C for 5 min. 1 ml LB was added and incubated for 2 to 4 h at 28°C to allow bacteria to express Km resistance. Cells were centrifuged and for a minute, supernatant was discarded and cells were resuspended in 50 µl LB. The cells were spread on Km LA plates containing 50 µg/ml Km. Plates were incubated at 28°C for two to three days for transformed Km resistant *Agrobacterium* colonies to grow.

APPENDIX B: RADIOACTIVE DNA SEQUENCING

Most templates were sequenced by the Sanger dideoxy termination method using the Sequenase® Version 2.0 DNA Sequencing Kit and protocols (Amersham LIFE SCIENCE) supplied by United States Biochemical Corporation (USB™). This method of sequencing uses the T7 polymerase labeling reactions. Alternatively, the TaqTrack® kit supplied by Promega™ was used where DNA polymerase isolated from *Thermus aquaticus* (*Taq* polymerase) was used in labelling reactions.

Some of the sequencing in this study was done using the ALF Express DNA Sequencer (Pharmacia) for automated detection of fluorescently-labelled DNA molecules separated by electrophoresis. Fluorescent-labelled primers were used in sequencing reactions instead radioactively labelled dNTPs (Appendix C).

B1. Alkaline denaturation of dsDNA templates

5-10μg DNA template purified on Nucleobond columns (Appendix A11) was diluted to 18μl using MilliQ® deionised distilled water in an eppendorf tube. 2μl 2M NaOH was added to the microfuge and mixture was incubated at 37°C for 30 minutes. The eppendorf was placed on ice and 3μl 3M sodium acetate and 100μl 96% ethanol was added. The contents of the eppendorf tube was vortexed and the eppendorf tube was placed at -70°C for 10 minutes. The denatured DNA was microfuged for 10 minutes. The supernatant was discarded and 500μl 70% ethanol was added and the pellet was microfuged for 5 minutes to wash the denatured DNA pellet. The pellet was subsequently airdried.

B2. Priming reaction

The following was added to the denatured pellet in the eppendorf tube prepared in B1, and the mix subsequently incubated in the eppendorf 37°C for 10-30 minutes, and mixed occasionally to ensure that the denatured DNA is resuspended:

2μl primer [20ng/μl or 2pmol] (1212F or 1233R)

2μl 5x Sequenase buffer (supplied by kit)

6μl MilliQ® deionised distilled water (supplied by kit)

Sequenase buffer (5x concentrate); 200mM Tris-HCl pH 7.5, 100mM $MgCl_2$, 250mM NaCl.

When TaqTrack® was used the following was added to the denatured DNA prepared in C1:

2μl primer [20ng/μl or 2pmol] (1212F or 1233R)

2μl *Taq* DNA polymerase 5X buffer (supplied by TaqTrack® sequencing kit)

made up to a volume of 20µl with MilliQ®-filtered sterile distilled water.

Taq DNA polymerase 5X buffer; 250mM Tris-HCl pH 9.0, 50mM $MgCl_2$.

B3. Extension/Labeling reactions

Before starting the labelling reactions, the termination tubes were prepared as follows: 2.5µl (for the Sequenase method) and 1µl (for the TaqTrack kit) ddNTP/dNTP (appropriately proportioned) termination/extension mix obtained from the kit was added to their respective labelled eppendorf tubes: tubes labelled A, C, G and T will contain ddATP, ddCTP, ddGTP and ddTTP termination mixes respectively. These four tubes per template were put on ice for the termination reactions described in B4.

Sequenase kit termination mixes:
ddA termination mix:
80µM each of dATP, dCTP, dGTP and dTTP, 8µM ddATP, 50mM NaCl
ddC termination mix:
80µM each of dATP, dCTP, dGTP and dTTP, 8µM ddCTP, 50mM NaCl
ddG termination mix:
80µM each of dATP, dCTP, dGTP and dTTP, 8µM ddGTP, 50mM NaCl
ddT termination mix:
80µM each of dATP, dCTP, dGTP and dTTP, 8µM ddTTP, 50mM NaCl

TaqTrack kit termination mixes:
ddA termination mix: 25µM dATP, 250µM each of dCTP, dGTP and dTTP, 350µM ddATP
ddC termination mix: 25µM dCTP, 250µM each of dATP, dGTP and dTTP, 160µM ddCTP
ddG termination mix: 25µM dGTP, 250µM each of dATP, dCTP and dTTP, 25µM ddGTP
ddT termination mix: 25µM dTTP, 250µM each of dATP, dCTP and dGTP, 300µM ddTTP

As suggested by the Sequenase kit protocol, the supplied 25µl inorganic pyrophosphatase (5U/ml in 10mM Tris-HCl pH 7.5, 0.1mM EDTA, 50% glycerol) was added with the 25µl modified T7polymerase (13U/µl in 20mM KPO_4 pH7.4, 1mM DTT, 0.1mM EDTA, 50% glycerol) along with 150µl of supplied glycerol enzyme dilution buffer (20mM Tris-HCl pH 7.5, 2mM DTT, 0.1mM EDTA, 50% glycerol). Band-weakening caused by slow, sequence-dependent reversal of the DNA polymerase reaction (pyrophosphorolysis) was eliminated by the use of pyrophosphotase. This mix and the kit were stored at -20°C. To each denatured DNA template/primer complex (from Appendix B2) the following ingredients supplied by the kit were added:

1µl 0.1M dithiothreitol (DTT)

2.8µl 1X labelling mix (5x labelling mix diluted 1 in 5)

1.7µl DMSO

2µl T7 polymerase/pyrophosphatase mix

0.5-1µl [α-^{35}S]dATP (Redivue® or Isoblue® 1Ci/mmol from Amersham LIFE SCIENCE™; not supplied with the kit) at 10µCi/µl.

Labelling mix (5X concentrate): 7.5µM each of dGTP, dCTP and dTTP

When the TaqTrack kit was used the following was added to the primed reaction tube prepared in B2:

2.5µl DMSO

2µl 1X extension/labelling mix (from TaqTrack kit)

0.5µl Taq polymerase

0.5-1µl [α-^{35}S]dATP (Redivue® or Isoblue® 1Ci/mmol from Amersham LIFE SCIENCE™; not supplied with the kits) at 10µCi/µl.

Extension/labelling mix: 7.5µM each of dGTP, dCTP and dTTP

This mixes were scaled up proportionally depending on how many templates were being sequenced, in which case 8µl of the Sequenase labelling reactions or 6µl of the TaqTrack labelling mixes were added to each template. The labelling reactions were incubated at room temperature for Sequenase and 37°C for TaqTrack for 10-20 minutes.

B4. Termination reactions

The four A-, C-, G-, and T-labelled termination tubes for each template was prewarmed at 40-48°C for Sequenase reactions and 70°C for TaqTrack reactions in a dry heating block. 3.6µl Sequenase labelling reaction (from B3) aliquotes was added to each of the four prewarmed labelled termination tubes and mixed for each template. In the case of TaqTrack, 6µl of the labelling reactions were added to the termination tubes. These tubes were incubated at the respective 40-48°C for Sequenase and 70°C for TaqTrack for 10-15 minutes. 5µl stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05 xylene cyanol FF) supplied with the kit was added to each termination tube. The reaction mixes were stored at -20°C or denatured at 95°C in a waterbath before loading sequencing acrylamide gels.

B5. Gel preparation and pouring

The Omeg Scientific™ (Cape Town) or the Hoefer™ Pokerface® sequencing apparati were used for the electrophoresis of the sequencing reactions (Appenidix C4). A pair of glass

plates were washed in warm water using a wash sponge and all-purpose non-scratch detergent, Chemico Liguicleen®. One glass plate (surface that will make contact with the gel) was made hydrophobic using SeeThru® (car windscreen water repellant). The inside surface of each plate was wiped with ethanol to get rid of any dust particles. A spacer (0.5mm thickness) was placed along each long edge and on the bottom edge of the unnotched plate. The notched plate was placed onto the notched plate, clean sides facing each other kept apart by the spacers. The plates were clamped using foldback paper clips or "bulldog clips". 40% acrylamide was made up by dissolving 38g acrylamide (molecular grade or Electran-BDH) and 2g N,N'-methylene bis-acrylamide (Electran-BDH) in 100ml deionised distilled water. Alternatively, Acrylogel® (Electran-BDH) was used. One of two electrophoretic buffers was used in preparing the gel mix:

a) 5x TBE 6% gel mix was prepared as follows;

15ml 40% acrylamide or Acrylogel®

50mls 10X TBE (Appendix A3)

46g molecular grade urea

made up to 100mls with deionised distilled water.

b) 5x Taurine 6% gel mix was prepared as in (a) above except 50mls 10X TBE was replaced by the same volume of 10X taurine buffer (108g Tris-HCl, 36g taurine and 4g EDTA made up to 1ℓ with deionised distilled water)

50µl freshly prepared 50% ammonium persulphate and 50µl N,N,N',N'-tetramethyl-1,2-diaminoethane (TEMED) were added to freshly prepared 5x TBE or taurine 6% gel mix. The mix was filtered through a 0.45µm Millipore® filter and sucked up into a 50ml needle-less syringe. The gel mix was injected between the cleaned glass plates, maintaining a steady flow. Once the gel mould was filled, the comb was inserted, closing the open end. The comb was also clamped using paper foldback or "bulldog clips". The gel was left to polymerise for an hour, or when left overnight, the gel was sealed with Gladwrap® to avoid drying.

B6. Electrophoresis of sequencing reactions

Gladwrap®, paper foldback clips, the spacers at the top and bottom ends of the gel were all removed. The plates containing polymerised gel was assembled in the electrophoresis apparatus with the well recess of the glass plate facing the upper buffer chamber. The upper and lower buffer resovoirs were filled with 1x TBE or 1X taurine buffer depending on whether the gel mix was made with taurine buffer or TBE. The "shark's tooth" comb was carefully inserted between the top of the plates 0.5-1mm into the gel surface and evenly across the gel surface. Sequencing reactions were denatured at 95°C for 5 minutes and placed on ice. The wells were flushed using a 20ml syringe with a needle to remove the

diffused urea. Using a P20 Gilson® micropipettor, 3µl of the denatured sequencing reactions (A, C, G, and T; 4 per template) into respective wells. When the samples were loaded a connected high voltage power supply was switched on to 30V/cm. For ten templates the power supply was set at 2000-2300V, 30-35mA and 80-85W for 1.5-2 hours to read 100-200bases, 3.5-4 hours to read 150-300 bases and on a metre-long Omeg® sequencing apparatus up to 450 bases. The metre-long apparatus was usually run for 20 hours.

B7. Gel drying and autoradiography of DNA sequencing gels

When the runs were complete the gel apparati were dismantled, with radioactive electrophoretic buffer disposed of as recommended by the laboratory manual. With the notched plate uppermost the plates were pried apart using a thin wedge or spatula. The gel remained on the unnotched glass plate. A 3MM absorbant paper sheet cut to the size of the gel was placed onto the 0.5mm thick acrylamide gel and gently pressed over the entire surface ensuring complete contact. The gel was lifted from the unnotched glass by carefully removing the paper. The gel was covered with Gladwrap® and place on a 70°C slab gel drier attached to a vacuum pump for 1 hour. The Gladwrap® was removed from the dried gel and placed in a light-tight X-ray film cassette in direct contact with a sheet of Agfa Curix P1 X-ray film (Agfa-Gevaert™) in a darkroom. Exposure was allowed overnight to 7 days depending on the strength of the ^{35}S label radioactive signal. The X-ray film was developed in an OKAMOTO Medical X-ray film automatic processor X2. The A, C, G and T lanes of each template were read and stored and analysed in the GCG Version 7.1 computer package on a VMS mainframe.

APPENDIX C: AUTOMATED DNA SEQUENCING

The Pharmacia™ ALFexpress® sequencer was designed for automated detection of fluorescently labelled dideoxy chain-termination fragments by electrophoresis. The non-radioactive approach in this study was based on the detection of Cy5-labelled primers used in Thermo Sequenase® cycle sequencing reactions (see the Pharmacia ALF Express manual and the Thermo Sequenase fluorescent -labelled primer cycle sequencing kit). Primers were obtained that were labelled with Cy5 amidite on oligonucleotide synthesisers. Thermo Sequenase is a exonuclease-free thermostable DNA polymerase specifically engineered for cycle sequencing.

C1. Thermo Sequenase cycle sequencing with fluorescent dye primer labeling

5µg of template DNA was put in a volume of 20µl MilliQ-filtered water in an eppendorf tube. 4 lidded thermocycler tubes were labelled A, C, G and T for each template. Reactions were prepared as follows:

Methyl violet-labelled primer (1-2 pmol/µl)	1µl
[Standard M13 reverse and forward primers were used]	
Template DNA	5µl
A, C, G or T reagent	2µl

A, C, G and T reagents have the following components in addition to their respective ddATP, ddCTP, ddGTP or ddTTP: Tris-HCl (pH 9.5), $MgCl_2$, Tween™ 20, Nonidet™ P-40, 2-mercaptoethanol, dATP, dCTP, dGTP, dTTP and ddATP, thermostable pyrophosphatase and Thermo Sequenase DNA polymerase. (concentrations not provided in the kit manual)

The contents were mixed by pipetting up and down twice. The following thermocycler conditions were used for the cycle reactions: pre-denaturation at 98°C for 5 min; denaturation, annealing and extension/terminations at 98°C for 30sec, 60°C for 30sec and 72°C for 40 sec respectively for 25-30 cycles. The reactions were ethanol precipitated as follows:

1/10th volume 3M NaoAc, pH 5.2 (1µl)

2.5 volumes of ethanol and mix (23µl)

The mixture were spun down at 12 000rpm for 10min at room temperature. 70% ethanol was added to the tubes and the DNA was spun down as described previously. The DNA pellets were airdried and 5µl formamide loading dye (components: formamide, EDTA and methyl violet, concentrations are not given in the kit) was added to each tube. DNA was denatured at 90°C for two minutes prior to gel electrophoresis.

C2. Electrophoresis and detection system

Electrophoresis was carried out in a vertical gel cassette. Temperature was controlled by circulation of water through the integrated thermoplate of the ALFexpress automated sequencer. Samples were loaded into the wells at the top of the gel. A fixed laser beam passes through the glass spacer located between the glass and the thermoplate of the gel cassette. The beam is directed into the gel perpendicular to the direction of band migration. The laser beam excites the fluorescently-labelled DNA sequence bands and the light emmitted is detected by photodetectors located behind the gel. The signals were collected, digitised and sent to a computer for storage and processing. During electrophoresis, raw data was displayed as a chromatogram on a computer monitor.

C3. Computer Software and analyses

The AM V3.0 software, the ALF Express during electrophoresis, stores the data and performs the post run evaluation of the sequencing data. Raw signals from runs were processed and were analysed using the GCG Version 7.0 computer package loaded on a VMS mainframe computer.

APPENDIX D: REPEAT AGROINFECTIVITY PROFILES

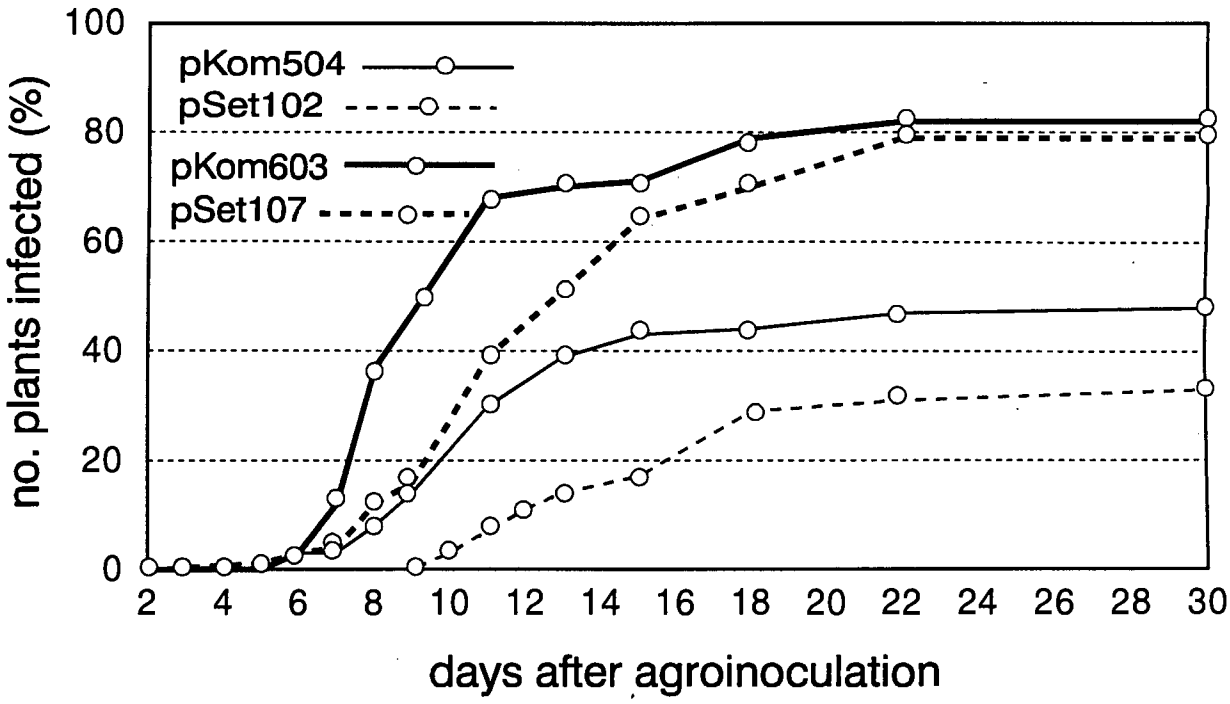

Figure D1. Agroinfectivity rates of homodimeric dimers (repeated)

Diagram depicts the relative rates of agroinfection of the partial dimers (pKom504 and pSet102) and those of the 1.1mers constituting two LIRs (pKom603 and pSet107). Number of sweetcorn Jubilee 3-4 day old seedlings agroinoculated with the following clones pKom504, pSet102, pKom603, pSet107 and pBI121 (negative control) were 83, 79, 85, 75 and 23 respectively. Plants agroinoculated with pBI121 were symptomless. (See 2.3.1)

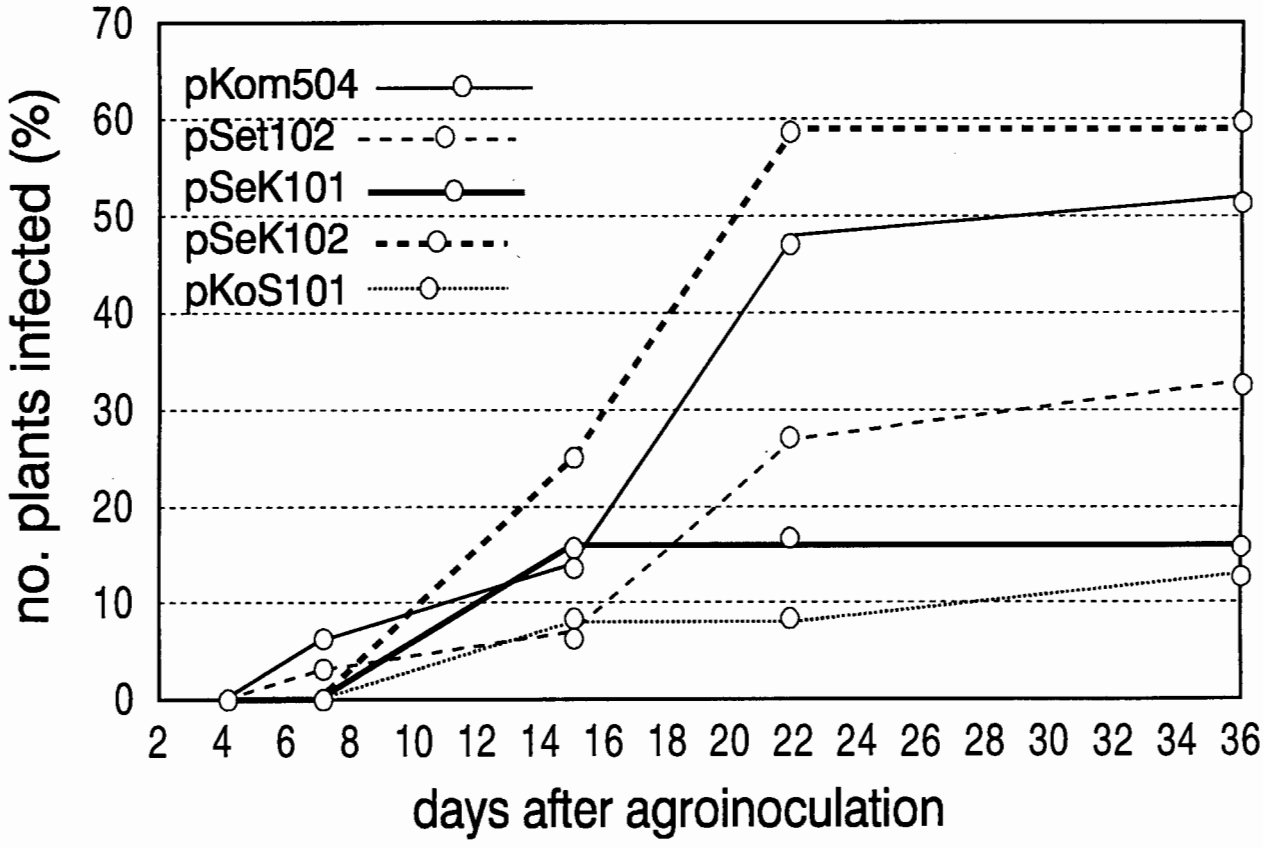

Figure D2 Agroinfectivity rates of heterodimeric clones of MSV-Kom and MSV-Set

Diagram depicts the relative agroinfectivity rates of the heterodimeric clones pSeK101, pSeK102, pKoS101 realtive to those of the wild-type homodimeric clones, pKom504 and pSet102. The number of plants agroinculated for pKom504, pSet102, pSeK101, pSeK102 and pKoS101 were 12, 12, 24, 24 and 24 respectively. pKoS102 and pKos103 were not agroinfectious in this repeat experiment.

REFERENCES

Abouzid, A.M., Frischmuth, T., and Jeske, H. (1988). A putative replicative-form of the Abutilon mosaic virus (gemini group) in a chromatin-like structure. *Mol. Gen. Genet.* 212: 252-258.

Accotto, G-P., Mullineaux, P.M., Brown, S.C. and Marie, D. (1993). Digitaria streak geminivirus replicative forms are abundant in S-phase Nuclei of infected calls. *Virology.* 195, 257-259.

Accotto, G-P., Donson, J., and Mullineaux, P.M. (1989). Mapping of *Digitaria* streak virus transcripts reveals different RNA speciesfrom the same transcription unit. *EMBO J.* 8, 1033-1039.

Ammar, E.D. (1994). Cytopathology and ultrastructure of some African isolates of maize streak and sugarcane streak viruses. *J. Phytopathology.* 141: 153-158.

An, G., Ebert, P.R., Mitra, A. and Ha, S.B. (1988). Binary vectors. In *Plant Molecular Biology Manual* Eds. S.B. Gelvin and R.A. Schilperoort. Kluwer Publishers, Dordrecht, The Netherlands.

Anderson, M.T., Richardson, K. A., Harrison, S.A., and Morris, B. A. M. (1988). Nucleotide sequence of the geminivirus chloris striate mosaic virus. *Virology* 164, 443-449.

Argüello-Astorga, G.R., Guevara-Gonsàlez, R.G., Herrerra-Estrella, L.R. and Rivera-Bustamante, R.F. (1994). Geminivirus replication origins have a group specific organisation of iterative elements: A model for replication. *Virology.* 203, 90-100.

Argüello-Astorga, G., Herrerra-Estrella, L. and Rivera-Bustamante, R. (1994). Experimental and theoretical definition of geminivirus origin of replication. *Plant Molecular Biology.* 26: 553-556.

Armitage, P., Walden, R. and Draper, J. (1988). Vectors for the transformation of plant cells using *Agrobacterium* (Chapter 1). In *Plant Genetic Transformation and Gene Expression: A Laboratory Manual.* Draper, J., Scott, R. and Armitage, P. (Eds). Blackwell Scientific Publications, Oxford London Edinburgh Boston Palo Alto Melbourne.

Bisaro, D.M. (1994). Recombination in the geminiviruses: mechanisms for maintaining genome size and generating genomic diversity. In: *Homologous recombination in plants*, edited by Paszkowski, J.Kluwer Academic Publishers. pp. 39-60.

Bisaro, D.M. (1996). Geminivirus replication. In: *DNA replication in eukaryotic cells*, edited by Paszkowski, J. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Bock, K.R., Guthrie, E.J. and Woods, R.D. (1974). Purification of maize streak virus and its relationship to viruses associated streak symptoms of sugarcane and *Panicum maximum. Annals of Applied Biology.* 77: 289-296.

Bock, K.R. (1982). Geminivirus diseases in tropical crops. *Plant disease.* 66: 266-270.

Bolton, G.W., Nester, E.W. and Gordon, M.P. (1986). Plant phenolic compounds induce expression of *Agrobacterium tumefaciens* loci neede for virulence. *Science.* 232: 983-985.

Boulton, M. I., Steinkellner, H., Donson, J., Markham, P. G., King, I. D., and Davies, J. W. (1989a). Mutational analysis of the virion-sense genes of the maize streak virus genome. *J. Gen. Virol.* 70, 2309-2323.

Boulton, M.I., Buchholz, W.G., Marks, M.S., Markham, P.G. and Jeffries, W.D. (1989b). Specificity of *Agrobacterium*-mediated delivery of maize streak virus DNA to members of the Graminae. *Plant Mol. Biol.* 12: 31-40.

Boulton, M.I., King, D.I., Donson, J., and Davies, J.W. (1991). Point substitutions in a promoter-like region and the V1 gene affect the host range and symptoms of maize streak virus. *Virology* 183:114-121.

Boulton, M.I., Pallaghy, C.K., Chatani, M., MacFarlane, S. and Davies, J.W. (1993). Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. *Virology.* 19: 85-93.

Briddon, R.W., Watts, J., Markham, P.. and Stanley, J. (1989). The coat protein of beet curly top virus is essential for infectivity. *Virology.* 172: 628-633.

Briddon, R.W., Pinner, M.S., Stanley, J. and Markham, P.G. (1990). Geminivirus coat protein replacement alters insect specificity. *Virology.* 177: 85-94.

Briddon, R.W., Lunness,P., Chamberlin,L.C., Pinner,M.S., Brundish,H. and Markham,P.G., (1992). The nucleotide sequence of an infectious insect-transmissible clone of the geminivirus *Panicum* streak virus. *J. Gen. Virol.* 73, 1041-1047.

Briddon, R.W. and Markham, P. (1995). Family *Geminiviridae.* In Classification and nomenclature of viruses. *Sixth Report of the International Committee on Taxonomy of Viruses.* Edited by F.A. Murphy, C.M. Fauquet, D.H.L. Bishop, S.A. Ghabrial, A.W. Jarvis, G.P. Martelli, M.A. Mayo and M.D. Summers. Vienna and New York: Springer-Verlag. pp. 158-165.

Brough, C.L., Hayes, R.J., Morgan, A.J., Coutts, R.H.A. and Buck, K.W. (1988). Effects of mutagenesis *in vitro* on the ability of cloned tomato golden mosaic virus DNA to infect *Nicotiana benthamiana* plants. *J. Gen Virol.* 69: 503-514.

Brough, C.L., Sunter, G., Gardiner, W. and Bisaro, D. (1992). Kinetics of tomato golden mosaic virus replication and coat protein promoter activity in *Nicotiana tabacum* protoplasts. *Virology.* 187: 1-9.

Chatani, M., Matsumoto, Y.,Mizuta, H., Ikegami, M., Boulton, M.I. and Davies J.W., (1991). The nucleotide sequence and genome structure of the geminivirus miscanthus streak virus. *J. Gen. Virol.* 72: 2325-2331.

Chung, C.T. and Millar, R.H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nuc. Acids Res.* 16, 3580.

Citovsky, V. and Zambryski, P. (1993). Transport of nucleic acids through small holes. *Annu. Rev. Microbiol.* 47: 167-197.

Clarke, B.A., Rybicki, E.P., Hughes, F.L., Kirby, R., and von Wechmar, B., (1989). Characterisation of Southern African isolates of maize streak virus: Typing of three isolates by restriction mapping. *Intervirology.* 30, 86-95.

Coutts, R.H.A., Buck, K.W., and Hayes, R.J. (1990). Development of geminivirus-based gene vectors for dicotyledonous plants. *Austr.J.Plant Physiol.* 17:365-375.

Crothers, D.M., Haran, T.E. and Nadeau, J.G. (1990). Intrinsically bent DNA. *J. Biol. Chem.* 265 (13): 7093-7096.

Damsteegt, V. D. (1983). Maize streak virus. 1. Host range and vulnerability of germ plasm. *Plant Dis.* 67: 734-737.

Davies, J.W. and Stanley, J. (1989). Geminivirus genes and vectors. *Trends Gen.* 5(3):77-81.

Dekker, E.L., Pinner, M.S., Markham, P. G., and van Regenmortal, M.H.V., (1988). Characterisation of maize streak virus isolates from different plant species by polyclonal and monoclonal antibodies. *J. Gen. Virol.* 69, 983-990.

Desbiez, C, David, C., Mettouchi, A., Laufs, J. and Gronenborn, B. (1995). Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proc. Natl. Acad. Sci. USA.* 92: 5640-5644.

Dollet, M., Accotto, G. P., Lisa, V., Menissier, J. and Boccardo, G. (1986). A geminivirus serologically related to maize streak virus, from *Digitaria sanguinalis* from Vanuatu. *J. Gen. Virol.* 67: 933-937.

Donson, J., Accotto, G-P, Boulton, M.I., Mullineaux, P.M. and Davies, J.W. (1987). The nucleotide sequence of a geminivirus from *Digitaria sanguialis. Virology.* 161: 160-169.

Donson, J., Morris-Krsinich, B.A.M., Mullineaux, P.M., Boulton, M. I., and Davies, J. W. (1984). A putative primer for second-strand synthesis of maize streak virus is virion -associated. *EMBO J.* 7:1589-1596.

Dry, I.B., Krake, L.R., Rigden, J.E. and Rezaian, M.A. (1997). A novel subviral agent associated with a geminivirus: The first report of a DNA satellite. *Proc. Natl. Acad. Sci. USA.* 94: 7088-7093.

Elmer, S. and Rogers, S.G. (1990). Selection for wild type size derivatives of tomato golden mosaic virus during systemic infection. *Nucl.Acids Res.* 18:2001-2006.

Elmer, J.S., Brand, L., Sunter, G., Gardiner, W.E., Bisaro, D.M., and Rogers, S.G. (1988a). Genetic analysis of the tomato golden mosaic virus. II. The product of the AL1 coding sequence is required for replication. *Nucleic. Acids.Res.* 16(14B):7043-7060.

Elmer, J.S., Sunter, G., Gardner, W.E., Brand, L., Browning C.K., Bisaro, D.M., and Rogers, S.G., (1988b). *Agrobacterium*-mediated inoculation of plants with tomato golden mosaic virus DNAs. *Plant. Mol. Biol.* 10: 225-234.

Escudero, J., Neuhaus, G., Schläppi, M. and Hohn, B. (1996). T-DNA transfer in meristematic cells of maize provided with intracellular *Agrobacterium. Plant Journal.* 10: 355-360.

Etessami, P., Callis, R., Ellwood, Sand Stanley, J. (1988). Delimitation of essential genes of cassava latent virus DNA 2. *Nuc. Acids Res.* 16: 4811-4829.

Etessami, P., Watts, J., and Stanley, J. (1989). Size reversion of African cassava mosaic virus coat protein gene deletion mutants during infection of Nicotiana benthamiana. *J.Gen.Virol.* 70 (Pt 2): 277-289.

Etessami, P., Saunders, K., Watts, J. and Stanley, J. (1991). Mutational analysis of complementary-sense genes of African cassavamosaic virus DNA A. *J. Gen. Virol.* 72: 1005-1012.

Evans, D. and Jeske, H. (1993). Complementation and recombination between mutants of complementary sense genes of DNA A of abutilon mosaic virus. *Virology.* 197: 492-496.

Faria, J.C., Gilbertson, R.L., Morales, F.J., Russel, D.R., Ahlquist, P.G., Hanson, S.F. and Maxwell, D.P. (1990). Sequence of infectious clones of two mechanically transmissable isolates of Bean Golden mosaic geminivirus. *Phytopathology.* 80: 983.

Faria, J.C., Gilbertson, R.L., Hanson, S.F., Morales, F.J., Ahlquist, P., Loniello, A.O., and Maxwell, D.P. Bean (1990). Golden mosaic virus type II isolates from the Dominican Republic and Guatemala: nucleotide sequences, infectious pseudorecombinants, and phylogenetic relationships. *Phytopathology.* 84:321-329.

Fauquet, C., Dejardin, J. and Thouvenel, J.-C. (1986). Evidence that the amino acid composition of the particle proteins of plant viruses is characteristic of the virus group. *Intervirology.* 25: 1-13.

Feng and Doolittle, (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* 25:351-360.

Fenoll, C., Black, D.M. and Howell, S.H. (1988). The intergenic region of maize streak virus contains promoter elements involved in rightward transcription of the viral genome. *EMBO J.* 7, 1589-1596.

Fenoll, C., Schwarz, J.J., Black, D.M., Schneider, M. and Howell, S.H. (1990). The intergenic region of maize streak virus contains a GC-rich element that activates rightward transcription and binds maize nuclear factors. *Plant Mol. Biol.* 15, 865-877.

Fontes, E.P.B., Luckow, V.A. and Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell.* 4: 597-608.

Fontes, E.P.B., Eagle, P.A., Sipe, P.S., Luckow, V.A. and Hanley-Bowdoin, L. (1994a). Interaction between a Geminivirus Replication Protein and Origin DNA Is Essential for Viral Replication. *J. Biol. Chem.* 269 (11): 8459-8465.

Fontes, E.P.B., Gladfelter, H.J., Schaffer, R.L., Petty, I.T.D. and Hanley-Bowdoin, L. (1994b). *Plant Cell.* 6: 495-416.

Francki, R.I.B., Hatta, T., Boccardo, G. and Davies, J.W. (1979). The composition of chloris striate mosaic virus, a geminivirus. *Virology.* 101: 233-241.

Frischmuth, S, Frischmuth, T., Latham, J.R. and Stanley, J. (1993). Transcriptional analysis of the virion-sense genes of the geminivirus beet curly top virus. *Virology.* 197: 312-319.

Frischmuth, T. and Stanley, J. (1991). Atrican cassava mosaic virus DI DNA interferes with the replication of both genomic components. *Virology.* 183: 539-544.

Frischmuth, T. and Stanley, J. (1992). Characterization of beet curly top virus subgenomic DNA localizes sequences required for replication. *Virology* 189(2): 808-811.

Frischmuth, T. and Stanley, J. (1993). Strategies of the control of geminivirus diseases. *Seminars in Virology.* 4: 329-337.

Frischmuth, T. and Stanley, J. (1994). Beet curly top virus symptom amelioration in *Nicotiana benthamiana* transformed with a naturally occurring viral subgenomic DNA. *Virology.* 200: 826-830.

Fuller, C. (1901). Mielie variegation. In *1st Report of the Government Entomologist, Natal, 1899-1900.* pp. 17-19.

Gardiner, W.E., Sunter, G., Brand, L., Elmer, J.S., Rogers, S.G. and Bisaro, D.M. (1988). Genetic analysis of tomato golden mosaic virus: the coat protein is not required for systemic spread or symptom development. *EMBO J.* 7: 899-904.

Gardner, R. and Knauf, V., (1986). Transfer of *Agrobacterium* DNA to plants requires a T-DNA border but not the *vir*E locus. *Science* 231: 725-727.

Ghai, J., and Das, A. (1989). The virD operon of *Agrobacterium tumefaciens* Ti plasmid encodes a DNA-relaxing enzyme. *Proc. Natl. Acad. Sci. USA.* 86: 3109-3113.

Gilbertson, R.L., Faria, J.C., Hanson, S.F., Morales, F.J., Ahlquist, P., Maxwell, D.P., and Russell, D.R. (1991). Cloning of the complete DNA genomes of four bean-infecting geminiviruses and determining their infectivity by electric particle discharge acceleration. *Phytopathology* 81: 980-985.

Gilbertson, R.L., Hidayat, S.H., Paplomatas, E.J., Rojas, M.R., Hou, Y.M., and Maxwell, D.P. (1993). Pseudorecombination between infectious cloned DNA components of tomato mottle and bean dwarf mosaic geminiviruses. *J.Gen.Virol.* 74(Pt 1):23-31.

Goodman, R.M., (1981a). Geminiviruses. *J. Gen. Virol.* 54, 9-21.

Goodman, R.M., (1981b). "Handbook of plant virus infection and comparative diagnosis" (E.Kurstak, ed.), 879-910.

Gorbalenya, A.E. and Koonin, E.V. (1989). Viral proteins containing the purine NTP-binding sequence pattern. *Nuc. Acids Res.* 17: 23-31.

Gorbalenya, A.E., Koonin, E.V. and Wolf, Y.I. (1990). A new superfamily of putative NTP-binding domains encoded by genomes os small DNA and RNA viruses. *FEBS Lett.* 262: 145-148.

Greber, R.S. (1989). Biological characteristics of grass geminiviruses from eastern Australia. *Ann. Appl. Biol.* 114: 471-480.

Grimsley, N., and Bisaro, D., (1987). Agroinfection. In: Hohn, T and Schell, J. (eds) Plant DNA infectious agents. Springer-Verlag, Vienna. pp. 88-107.

Grimsley, N., Hohn, B., Hohn, T., and Walden, R., (1986). "Agroinfection", and altenative route for viral infection of plants using the Ti plasmid. *Proc. Natl. Acad. Sci. USA.* 83: 3282-3286.

Grimsley, N., Hohn, T., Davies, J. W., and Hohn, B. (1987). *Agrobacterium*-mediated delivery of infectious maize streak virus into maize plants. *Nature (London)* 325, 177-179. Gronenborn and Messing (1978).

Groning, R.R., Frischmuth, T. and Jeske, H. (1990). Replicative form DNA of Abutilon mosaic virus is present in plastids. *Mol. Gen. Genet.* 220: 485-488.

Grylls, N.E. (1963). A striate mosaic virus disease of grasses and cereals in Australia, transmitted by the cicadellid Nesoclutha obscura. *Aust. J. Agr. Res.* 14: 143-153.

Haley, A., Richardson, K., Zhan, X.C. and Morris, B. (1995). Mutagenesis of the BC1 and BV1 genes of African cassava mosaic virus identifies conserved amino acids that are essential for spread. *J. Gen. Virol.* 76: 1291-1298.

Haley, A., Zhan, X., Richardson, K., Head, K., and Morris, B. (1992). Regulation of the activities of African cassava mosaic virus promoters by the AC1, AC2, and AC3 gene products. *Virology.* 188: 905-909.

Hamilton, W.D., Bisaro, D.M., and Buck, K.W. (1982). Identification of novel DNA forms in tomato golden mosaic virus infected tissue. Evidence for a two component viral genome. *Nucleic.Acids.Res.* 10(16): 4901-4912.

Hamilton, W.D.O., Bisaro, D.M., Coutts, R.H.A. and Buck, K.W. (1988). Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. *Nuc. Acids Res.* 11: 7387-7396.

Hanley-Bowdoin, L., Elmer, J.S. and Rogers S.G. (1990). Expression of functional replication protein from tomato golden mosaic virus in transgenic tobacco. *Proc. Natl. Acad. Sci. USA.* 87: 1446-1450.

Hanley-Bowdoin, L., Eagle, P.A., Orozco, B.M., Robertson, D., and Settlage, S.B. (1996). Geminivirus replication. In: *Biology of Plant-Microbe Interactions*, edited by Stacey, G., Mullin, B., and Gresshoff, P.M.St. Paul, MN: Int Soc Mol Plant-Microbe Interactions. pp. 287-292.

Hanson, S.F., Hoogstraten, R.A., Ahlquist, P., Gilbertson, R.L., Russell, D.R. and Maxwell, D.P. (1995). Mutational analysis of a putative NTP-binding domain in the replication-associated protein (AC1) of bean golden mosaic geminivirus. *Virology.* 211: 1-9.

Harrison, B.D., Barker, H., Bock, K.R., Guthrie, meredith, G. and Atkinson, M. (1977). Plant viruses with circular single-stranded DNA. *Nature*, 270: 760-762.

Harrison, B. (1985). Advances in geminivirus research. *Annu. Rev. Phytopathol.* 23: 55-82.

Hatta, T. and Francki, R.B.I. (1979). The fine structure of chloris striate mosaic virus. *Virology.* 92: 428-435.

Hayes, R.J. and Buck, K.W. (1989a). Replication of tomato golden mosaic virus DNA B in transgenic plants expressing open reading frames (ORFs) of DNA A: requirement of ORF AL2 for production of single-stranded DNA. *Nuc. Acids Res.* 17(24): 10213-10222.

Hayes, R.J., Coutts, R.H., and Buck, K.W. (1989b) Stability and expression of bacterial genes in replicating geminivirus vectors in plants. *Nucleic.Acids.Res.* 17(7):2391-2403.

Hayes, R.J., MacDonald, H., Coutts, R. H. A., and Buck, K. W. (1988a). Priming of complementary DNA sequences *in vitro* by small DNA molecules tightly bound to virion DNA of wheat dwarf virus. *J. Gen. Virol.* 69: 1345-1350.

Hayes, R.J., Petty, I.T., Coutts, R.H., and Buck, K.W. (1988b). Gene amplification and expression in plants by a replicating geminivirus vector. *Nature* 334:179-182.

Hayes, R.J., Brough, C.L., Prince, V.E., Coutts, R.H.A. and Buck, K.W. (1988c). Infection of *Nicotiana benthamiama* with uncut cloned tandem dimers of tomato golden mosaic virus DNA. *J. Gen. Virol.* 69: 209-218.

Hayes, R.J., Coutts, R.H.A. and Buck, K.W. (1988d). Agroinfection of *Nicotiana* spp. with cloned DNA. *J. Gen. Virol.* 69: 1487-1496.

Heyraud, F., Matzeit, V., Kamman, M., Schaefer, S., Schell, J., and Gronenborn, B., (1993a). Identification of the initiation sequence for viral-strand DNA synthesis of wheat dwarf virus. *EMBO J.* Vol. 12(11): 4445-4452.

Heyraud, F., Matzeit, V., Schaefer, S., Schell, J., and Gronenborn, B., (1993b). The conserved nonanucleotide motif of the geminivirus stem-loop sequence promotes replicational release of virus molecules from redundant copies. *Biochemie.* 75: 605-615.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., and Gronenborn, B., (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nuc. Acid. Res.* Vol. 23. (6): 910-916.

Hofer, J.M.I., Dekker, E.L., Reynolds, H.V., Woolston, C.J., Cox, B.S. and Mullineaux, P.M. (1992). Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. *Plant Cell.* 4: 213-23.

Hohn, B., Hohn, T., Boulton, M.I., Davies, J.W., and Grimsley, N. (1987). Agroinfection of *Zea mays* with maize streak virus DNA. *In* "Plant Molecular Biology" (D. von Wettstein and N-H. Chua, Eds.), pp. 459-468. Plenum, New York/London.

Hormuzdi, S.G. and Bisaro, D.M. (1993). Genetic analysis of beet curly top virus: evidence for three virion sense genes involved in movement and regulation of single- and double-stranded DNA levels. *Virology.* 193: 900-909.

Hong, Y.G. and Stanley, J. (1995). Regulation of African cassava mosaic virus complementary-sense gene expression by N-terminal sequences of the replication-associated protein AC1. *J. Gen Virol.* 76: 2415-2422.

Horns, T. and Jeske, H. (1991). Localisation of Abutilon mosaic virus (AbMV) DNA within leaf tissue by *in situ* hybridisation. *Virology.* 181: 580-588.

Howard, E.A., Winsor, B., De Vos, G., Zambryski, P. (1989). Activation of the T-DNA transfer process in *Agrobacterium* results in the generation of a T-strand-protein complex: tight association of VirD2 with the 5' ends of T-strands. *Proc. Natl. Acad. Sci. USA.* 86: 4917-4021.

Howarth, A.J.and Goodman, R.M. (1982). Plant viruses with genomes of single-stranded DNA. *Trends Biochem. Sci.* 7, 180-182.

Howarth, A.J. and Goodman, R.M. (1986). Divergence and Evolution of Geminivirus genomes. *J. Mol. Evol.* 23: 313-319.

Howarth, A.J., and Vandemark, G.J., (1989). Phylogeny of Geminiviruses. *J. Gen. Virol.* 70: 2717-2727.

Howell, S.H. (1985). Physical structure and genetic organisation of the genome of maize streak virus (Kenyan isoalate). *Nucleic Acids Res.* 13: 3018-3019.

Hughes, F.L. (1991). Molecular investigations of subgroup I geminiviruses. PhD thesis.

Hughes, F.L., Rybicki, E.P., Kirby, R. and von Wechmar, M.B. (1991). Characterisation of the sugarcane streak agent as a distinct geminivirus. *Intervirolgy.* 32: 19-27.

Hughes, F.L. (1992). Genome typing of Southern African subgroup I geminiviruses. *J. Gen. Virol.* 73: 1031-1040.

Hughes, F.L., Rybicki, E.P., and Kirby, R., (1993). Complete nucleotide sequence of sugarcane streak monogeminivirus. *Arch. Virol.* 132, 171-182.

Hull, R. (1989). The movement of viruses in plants. *Annu. Rev. Phytopathol.* 27: 213-240.

Ikegami, M., Haber, S. and Goodman, R.M. (1981). Isolation and characterisation of a virus-specific double stranded DNA from tisues infected by bean golden mosaic virus. *Proc. Natl. Acad. Sci. USA.* 78(7): 4102-4106.

Ikegami, M., Yazaki, K., Chatani, M. and Mizuta, H. (1990). Molecular cloning and physical mapping of the DNA of miscanthus streak geminivirus. *Intervirology.* 30: 340-345.

Ilyina, T.V. and Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA repliaction encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nuc. Acid. Res.* 20, 3279-3285.

Ingham, D.J., Pascal, E. and Lazarowitz, S.G. (1995). Both bipartite geminivirus movement proteins define viral host range, but only BL1 determines viral pathogenicity. *Virology.* 207: 191-204.

Jeffrey, J.L., Pooma, W. and Petty, I.T.D. (1996). Genetic requirements for local and systemic movement of tomato golden mosaic virus in infected plants. *Virology.* 223: 208-218.

Julia, J.F. and Dollet, M. (1989). *Nesoclutha-declivata* Homoptera Cicadellidae vector of digitaria streak virus (geminivirus) Vanuata. *J. Phytopathology.* 127: 42-48.

Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Sp 1. *Trends Biochem. Sci.* 11, 20-23.

Kikuno, R., Toh, H., Hayashida, H., and Miyata, T. (1984). Sequence similarity between putative gene products of geminiviral DNAs. *Nature.* 308(5959):562.

Klinkenberg, F.A., Ellwood, S., and Stanley, J. (1989). Fate of African cassava mosaic virus coat protein deletion mutants after agroinoculation. *J.Gen.Virol.* 70:1837-1844.

Klinkenberg, F.A. and Stanley, J. (1990). Encapsidation and spread of African cassava mosaic virus DNA A in the absence of DNA B when agroinoculated to *Nicotiana benthamiama. J. Gen. Virol.* 71: 1409-1417.

Klute, K.A., Nadler, S.A. and Stenger, D.C. (1996). "Horseradish curly top virus is a distinct subgroup II geminivirus species with *Rep* and C4 genes derived from a subgroup III ancestor" *J. Gen. Virol.* 77, 1369-1378.

Koonin, E.V., and Ilyina, T.V., (1992). Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. *J. Gen. Virol.* 73, 2763-2766.

Koncz, C. and Schell, J., (1986). The promoter of T_L-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of *Agrobacterium* binary vector. *Mol. Gen. Genet.* 204: 383-396.

Latham, J.R., Saunders, K., Pinner, M.S. and Stanley, J. (1997). Induction of plant cell division by beet curly top virus gene C4. *The Plant Journal.* 11(6): 1273-1283.

Laufs, J., Jupin, I., David, C., Schumacher, S., Heyraud-Nitschke, F. and Gronenborn, B. (1995). Geminivirus replication: Genetic and biochemical characterisation of Rep protein function, a review. *Biochemie.* 77: 765-773.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J., and Gronenborn, B., (1995). *In vitro* cleavage and cloning at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc. Natl. Acad. Sci. USA.* 92, 3879-3883.

Lazarowitz, S.G., (1987). The molecular characterisation of geminiviruses. *Plant Mol. Biol. Reptr.* 4, 177-192.

Lazarowitz, S. G., (1988). The complete nucleotide sequence of the South African strain of maize streak virus. *Nucleic Acids Res.* 16, 229-249.

Lazarowitz, S. G., Pinder, A. J., Damsteegt, V. D., and Rogers S. G. (1989). Maize streak virus genes essential for systemic spread and symptom develoment. *EMBO J.* 8, 1023-1032.

Lazarowitz, S.G. (1991). Molecular characterization of two bipartite geminiviruses causing squash leaf curl disease: role of viral replication and movement functions in determining host range. *Virology* 180(1):70-80.

Lazarowitz S.G., (1992). Geminiviruses: Genome structure and gene function. *CRC Crit. Rev. Plant Sci.* 11(4): 327-349.

Lazarowitz, S.G., Wu, L.C., Rogers, S.G. and Elmer, J.S. (1992). Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell.* 4(7):799-809.

Leroux, B., Yanofsky, M.F., Winans, S.C., Ward, J.E., Ziegler, S.F. and Nester, E.W. (1987). Characterisation of the *vir*A locus of *Agrobacterium tumefaciens*: a transcriptional regulator and host range determinant. *EMBO. J.* 6: 849-856.

Lindsten, K., Lindsten, B., Abdelmoeti, M. and Juntti, N. (1980). Purification and some properties of wheat dwarf virus. *Proc. 3rd Conf. on Virus Diseases of Graminae in Europe, Rothamsted Exptl Stn, Harpenden, Herts., U.K.* 27-31.

Liu, L., van Tonder, T., Pietersen, G., Davies, J.W., and Stanley, J. (1997). Molecular characterisation of a subgroup I geminivirus from a legume in South Africa. *J. Gen. Virol.* 78: 2113-2117.

Louie, R. (1995). Vascular puncture of maize kernels for the mechanical transmission of the maize white line mosaic virus and other viruses of maize. *Phytopathology.* 85(2): 139-143.

Lucy, A.P., Boulton, M.I., Davies, J.W. and Maule, J. (1996). Tissue specificity of *Zea mays* infection by Maize Streak Virus. *Mol. Plant-Microbe Int.* 9(1): 22-31.

MacDonald, H., Coutts, R.H.A., and Buck, K.W. (1988). Characterization of a subgenomic DNA isolated from *Triticum aestivum* plants infected with wheat dwarf virus. *J.Gen.Virol.* 69:1339-1334.

MacDowell, S.W., MacDonald, H., Hamilton, W.D.O., Coutts, R.H.A.and Buck, K.W. (1985). The nucleotide sequence of cloned wheat dwarf virus DNA. *EMBO J.* 8: 1023-1032.

Matthews, R.E.F. (1979). Classification and nomenclature of viruses. *Intervirology* 12, 129-296.

Matzeit, V., Schaefer, S., Kamman, M., Schalk, H-J., Schell, J. and Gronenborn, B. (1991). Wheat dwarf virus vectors replicate and express foreign genes in cells of monocotyledonous plants. *Plant Cell.* 3: 247-258.

McClean, A.P.D. (1947). Some forms of streak virus occurring in maize, sugarcane and wild grasses. *South African Department of Agriculture Science Bulletin.* 265: 1-33.

Meyer, P., Heidmann, I. and Niedenhof, I. (1992). The use of African cassava mosaic virus as a vector system for plants. *Gene.* 110: 213-217.

Morinaga, T., Ikegami, M., Arai, T., Yazaki, K., and Miura, K. (1988). Infectivity of cloned tandem dimer DNAs of bean golden mosaic virus. *J.Gen.Virol.* 69:897-902.

Morris, B.A.M., Richardson, K.A., Haley, A., Zhan, X.C. and Thomas J.E., (1992). The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants. *Virology.* 187: 633-642.

Morris, B., Coates, L., Lowe, S., Richardson, K. and Eddy, P. (1990). Nucleotide sequence of the infectious cloned DNA components of African cassava mosaic virus (Nigerian strain). *Nuc. Acids. Res.* 18: 197-198.

Morris, B.A.M., Richardson, K.A., Anderson, M.T. and Gardner, R.C. (1988). Cassava latent virus infections mediated by the Ti-plasmid of Agrobacterium tumefaciens containing either monomeric or dimeric viral DNA. *Plant Mol. Biol.* 11: 795-803.

Morris-Krsinich, B.A.M., Mullineaux, P.M., Donson, J., Boulton, M. I., Markham, P. G., Short, M. N., and Davies, J. W. (1985). Bidirectional transcription of maize streak virus DNA and identification of the coat protein gene. *Nucleic Acid Res.* 13, 7237-7256.

Mullineaux, P. M., Donson, J., Morris-Krsinich, B.A.M., Boulton, M. I., and Davies, J. W. (1984). The nucleotide sequence of maize streak virus DNA. *EMBO J.* 3, 3063-3068.

Mullineaux, P. M., Boulton, M. I., Bowyer, P., van der Vlugt, R., Marks, M., Donson, J., and Davies, J. W. (1988). Detection of a non-structural protein of M_r 11 000 encoded by the virion DNA of maize streak virus. *Plant Mol. Biol.* 11, 57-66.

Mullineaux, P.M., Guerineau, F., and Accotto, G.P. (1990). Processing of complementary sense RNAs of Digitaria streak virus in its host and in transgenic tobacco. *Nucleic.Acids.Res.* 18(24):7259-7265.

Nagar, S., Pedersen, T.J., Carrick, K.M., Hanley-Bowdoin, L., and Robertson, D., (1995). A geminivirus induces expression of a host DNA synthesis protein in terminally differentiated plant cells. *The Plant Cell.* Vol 7. 705-719.

Navot, N., Pichersky, E., Zeidan, M., Zamir, D. and Czosnek, H. (1991). Tomato yellow leaf curl virus: a whitefly-transmitted geminivirus with a single genomic component. *Virology.* 185, 151-161.

Noueiry, A.O., Lucas, W.J. and Gilbertson, R.L. (1994). Two proteins of a plant DNA virus coordinate nuclear and plasmadesmal transport. *Cell.* 76: 925-932.

Okoth, T., and Inouye, T. (1978). Resemblance in morphology and intranuclear appearance of viruses isolated from yellow dwarf diseased tomato and leaf curl diseased tobacco. *Ann. Phytopathol. Soc. Jpn.* 44: 167-178.

Padidam, M., Beachy, R.N., and Fauquet, C.M., (1995). Classification and identification of geminiviruses using sequence comparisons. *J. Gen. Virol.* 76, 249-263.

Palmer, K., (1997). Investigation into the use of MSV as a gene vector in maize. PhD thesis.

Pascal, E., Goodlove, P.E., Wu, L.C. and Lasarowitz, S.G. (1993). Transgenic tobacco plants expressing the geminivirus BL1 protein exhibit symptoms of viral disease. *Plant Cell.* 5: 795-807.

Peterschmitt, M. (1992). Quiet, J.B., reynaud, B. and Baudin, P. (1992). Detection of maize streak virus antigens over time in different parts of maize plants of a sensitive and a so-called tolerant cultivar by ELISA. *Ann. Appl. Biol.* 121: 641-653.

Peterschmitt, M., Granier, M., Frutos, R. and Reynaud, B. (1996). Infectivity and complete nucleotide sequence of the genome of a genetically distinct strain of maize streak virus from Reunion Island. *Arch. Virol.* 141 (9), 1637-1650.

Pilartz, M. and Jeske, H (1992). Abutilon mosaic geminivirus double-stranded DNA is packed into minichromosomes. *Virology.* 189: 800-802.

Pinner, M.S., Markham, P.G., Markham, R. H. and Dekker, E.L. (1988). Characterisation of maize streak virus: *Plant Pathol.* 37: 74-87.

Pinner, M.S. and Markham, P.G. (1990). Serotyping and strain identification of maize streak virus isolates. *J. Gen. Virol.* 71: 1635-1640.

Pinner, M.S., Rybicki, E.P., Greber, R.S., and Markham, P.G. (1992). Serological relationships of geminivirus isolates from Gramineae in Australia. *Plant Pathol.* 41:618-625.

Pinner, M.S., Medina, V., Plaskitt, K.A. and Markham, P.G. (1993). Viral inclusions in monocotyledons infected by maize streak and related geminiviruses. *Plant Pathol.* 42: 75-87.

Pinner, M.S., Markham, P.G., Markham, R.H., and Dekker, E.L., (1988). Characterisation of maize streak virus: Description of strains, symptoms. *Plant Pathol.* 37, 74-87.

Pooma, W., illette, W.K., Jeffrey, J.L. and Petty, I.T.D. (1996). Host and viral factors determine the dispensability of the coat protein for bipartite geminivirus systemic movement. *Virology.* 218: 26-268.

Rao, J.K. (1985). Sequence homology among the coat proteins of gemini viruses. *Biochem. Biophys. Res. Commun.* 130(2):892-896.

Revington, G.N., Sunter, G. and Bisaro, M. (1989). DNA sequences essential for replication of the B genome component of tomato golden mosaic virus. *Plant Cell.* 1: 985-992.

Rigden, J.E., Dry, I.B., Krake, L.R. and Rezaian, M.A. (1996). Plant virus DNA replication processes in *Agrobacterium*: Insight into the origins of geminiviruses? *Proc. Natl. Acad. Sci. USA.* 93: 10280-10284.

Roberts, I.M., Robinson, D.J. and Harrison, B.D. (1984). Serological relationships and genome homologies among geminiviruses. *J. Gen. Virol.* 65: 1723-1730.

Roberts, I.M. (1989). Indian Cassava mosaic virus: ultrastructure of infected cells. *J. Gen. Virol.* 70: 2729-2739.

Roberts, S. and Stanley, J. (1994). Lethal mutations within the conserved stem-loop of African cassava mosaic virus DNA are rapidly corrected by genomic recombination. *J. Gen. Virol.* 75: 3203-3209.

Rochester, D.E., Kositratana, W. and Beachy, R.N. (1990). Systemic movement and symptom production following agroinoculation with a single DNA of tomato yellow leaf curl geminivirus (Thailand). *Virology.* 178: 520-526.

Rogers, S.G., Bisaro, D.M., Horsch, R.B., Fraley, R.T., Hoffman, N.L., Brand, L., Elmer, J.S. and Lloyd, A.M. (1986). Tomato golden mosaic virus A component DNA replicates autonomously in transgenic plants. *Cell.* 45: 593-600.

Rose, D.J.W. (1978). Epidemiology of maize streak disease *Ann. Rev. Ent.* 23: 259-282.

Rossi, L., Hohn, B. and Tinland, B. (1996). Integration of complete T-DNA units is determined on the activity of VirE2 protein of *Agrobacterium tumefaciens. Proc. Natl. Acad. Sci. USA.* 93: 126-130.

Rushing, A.E., Sunter, G., Gardiner, W.E., Dute, R.R. and Bisaro, D.M. (1987). Ultrastructural aspects of tomato golden mosaic virus infection in tobacco. *Phytopathology.* 77: 1231-1236.

Rybicki, E.P., (1994). A phylogenetic and evolutionary justification for three genera of Geminiviridae. *Arch. Virol.* 139: 49-77.

Rybicki, E.P., and Hughes, F.L., (1990). Detection and typing of maize streak virus and other distantly related geminiviruses of grasses by polymerase chain reaction amplification of a conserved viral sequence. *J. Gen. Virol.* 71: 2519-2526.

Rybicki, E.P. (1988). Maize streak virus: an African pathogen come home? *South African Journal of Science.* 84: 30-32.

Rybicki, E.P., Hughes, F.L., Clarke, B.A., and von Wechmar, M.B. (1989). Detection and differentiation of isolates of maize streak virus by use of cloned DNA hybridisation probes. *In Proceedings of the 8th Maize Breeding symposium, 1988, Technical communications* No 222, pp. 27-32. Edited by J.G. du Plessis. Pretoria: The Government Printer.

Rybicki, E.P. and von Wechmar, B. (1985). Serology and immunochemistry. *In* The Plant Viruses Vol.1: Polyhedral virions with tripartite genomes. Ed., Francki, R.I.B., Plenum Press, New York, London. pp.207-244.

Saitou, N. and Nei, M. (1987). The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406-425.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular cloning, A laboratory manual. Second edition. *Cold Spring Harbor Laboratory Press, New York, USA.*

Saunders, K., Lucy, A., and Stanley, J., (1991). DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nuc. Acids Res.* 19, 2325-2330.

Schalk, H. J., Matzeit, V., Schiller, B., Schell, J., and Gronenborn, B. (1989). Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication. *EMBO J.* 8, 359-364.

Schläppi, M. and Hohn, B. (1992). Competence of immature maize embryos for Agrobacterium-mediated gene transfer. *The Plant Cell.* 4: 7-16.

Settlage, S.B., Miller, A.B. and Hanley-Bowoin, L. (1996). Interactions between geminivirus replication proteins. *J. Virol.* 70 (10): 6790-6795.

Shukla, D.D. and Ward, C.W. (1988). Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *J. Gen. Virol.* 69: 270-2710.

Shen, W-H., Das, S. and Hohn, B. (1992). Mechanism of Ds1 excision from the genome of maize streak virus. *Mol. Gen. Genet.* 233: 388-394.

Shen, W-H. and Hohn, B. (1995). Vectors baed on maize streak virus can replicate to high copy numbers in maize plants. *J. Gen. Virol.* 76: 965-969.

Shen, W-H. and Hohn, B. (1994). Amplification and expression of the β-glucoronidase gene in maize plants by vectors based on maize streak virus. *The Plant Journal.* 5(2): 227-236.

Shen, W-H. and Hohn, B. (1992). Excision of a transposable element from a viral vector introduced into maize plants by agroinfection. *The Plant Journal.* 2(1): 35-42.

Shen, W-H. and Hohn, B. (1991). Mutational analysis of the small intergenic region of maize streak virus. *Virology.* 183: 721-730.

Smith, D.R., and Maxwell, D.P. (1994). Requirement of the common region of DNA-B and the BL1 open reading frame of bean golden mosaic geminivirus for the infection of *Phaseolus vulgaris*. *Phytopathology.* 84: 133-138.

Song, J-Y. and Jeske, H. (1994). The kavel of Abutilon mosaic geminivirus in leaf discs and wound callus. *J. Phytopathology.* 141: 45-52.

Stachel, S.E., Messens, E., van Montagu, M. and Schell, J. (1985). Identifiation of the signal moleculaes produced by wounded cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. *Nature*, 318: 624-629.

Stanley, J., (1995). Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs within the conserved nanonucleotide motif during the initiation of rolling circle DNA replication. *Virology.* 206, 707-712.

Stanley, J. (1993). Geminiviruses: plant viral vectors. *Current opinion in genetics and development*, 3: 91-96.

Stanley, J., (1985). The molecular biology of geminiviruses. *Ad. Virus Res.* 30, 139-177.

Stanley, J. (1983). Infectivity of cloned geminivirus genome requires sequences from both DNAs. *Nature, London* 305: 643-645.

Stanley, J. and Gay, M.R. (1983). Nucleotide sequence of cassava latent virus DNA. *Nature, London.* 301, 260-262.

Stanley. J. and Latham, J.R. (1992). A symptom variant of beet curly top geminivirus producd by mutation of open reading frame C4. *Virology.* 190: 506-509.

Stanley, J., Latham, J., Pinner, M.S., Bedford, I. and Markham, P.G. (1992). Mutational analysis of the monopartite geminivirus beet curly top virus. *Virology.*191: 396-405.

Stanley, J., Markham, P.G., Callus, R.J. and Pinner, M.S. (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. *EMBO J.* 5(8): 1761-1767.

Stanley, J. and Townsend, R. (1985). Characterisation of DNA forms associated with cassava latent virus infection. *Nucleic.Acids.Res.* 13(7):2189-2206.

Stanley, J. and Townsend, R. (1986). Infectious mutants of cassava latent virus generated *in vivo* from intact recombinant clones containing single copies of the genome. *Nuc. Acids Res.* 14: 5981-5998.

Stenger, D.C., Carbonaro, D., and Duffus, J.E. (1990). Genomic Characterization of Phenotypic Variants of Beet Curly Top Virus. *J.Gen.Virol.* 71:2211-2215.

Stenger, D.C., Davis, K.R. and Bisaro, D.M. (1994). Recombinant beet curly top virus genomes exhibit both parental and novel pathogenic phenotypes. *Virology.* 200: 677-685.

Stenger, D.C. and McMahon, C.L. (1997). Genotypic diversity of beet curly top virus populations in the Western United States. *Phytopathology.* 87(7): 737-744.

Stenger, D.C., Revington, G.N., Stevenson, M.C., and Bisaro, D.M., (1991). Replicational release of geminivirus genomes from tandemly repeated copies: Evidence for rolling-circle replication of plant viral DNA. *Proc. Natl. Acad. Sci., USA.* 88, 8029-8033.

Stenger, D.C., Stevenson, M.C., Hormuzdi, S.G., and Bisaro, D.M. (1992). A number of subgenomic DNAs are produced following agroinoculation of plants with beet curly top virus. *J.Gen.Virol.* 73(Pt 2):237-242, 1992.

Storey, H.H., (1928). Transmission studies of maize streak disease. *Ann. Appl. Biol.* 12, 442-439.

Storey, H.H. (1938). Investigations of the mechanism of the transmission of plant viruses by insect vectors. II. The part played by puncture in transmission. *Proc. Roy. Soc. London Ser. B.* 125: 455-477.

Storey, H.H and Howland, A.K. (1967). Inheritance of resistance in maize to the virus of streak disease in East Africa. *Ann. App. Biol.* 59: 429-436.

Storey, H.H. and MacLean, A.P.D. (1930). The transmission of streak disease between maize, sugarcane and wild grasses. *Ann. Appl. Biol.* 17: 691-719.

Suàrez-Lòpez, P., Marìnez-Salas, E., Hernàndez, P. and Guièrrez. (1995). Bent DNA in the large intergenic region of wheat dwarf geminivirus. *Virology.* 203: 303-311.

Studier, J.A., and Keppler, K.L. (1988). A note on the neighbour-joining algorithm of Saitou and Nei. *Mol. Biol. Evol.* 5, 729-731.

Sunter, G. and Bisaro, D.M. (1991). Transactivation in a geminivirus: AL2 gene product is needed for coat protein expression. *Virology.* 180: 416-419.

Sunter, G. and Bisaro, D.M. (1992). Transactivation of geminivirus AR1 and BR1 gene expression by the viral AL2 gene product occurs at the level of transcription. *Plant Cell.* 4: 1321-1331.

Sunter, G., Buck, K.W. and Coutts, R.H.A. (1985). S1-sensitive sites in the supercoiled double-stranded form of tomato golden mosaic virus DNA component B: identification of regions of potential alternative secondary structure and regulatory function. *Nuc. Acids Res.* 13(13): 4645-4659.

Sunter, G., Gardner, W.E., Rushing, A.E., Rogers, S.G. and Bisaro, D.M. (1987). Independent encapsidation of tomato golden mosaic virus A component DNA in transgenic plants. *Plant mol. Biol.* 8: 477-484.

Sunter, G., Hartitz, M.D. and Bisaro, D.M. (1993). Tomato golden mosaic virus leftward gene expression: autoregulation of geminivirus replication protein. *Virology* 195: 275-280.

Sunter, G., Hartitz, M.D., Hormuzdi, S.G., Brough, C.L. and Bisaro, D.M. (1990). Genetic analysis of tomato golden mosaic virus. ORF AL2 is required for coat protein accumulation while ORF AL3 is necessary for efficient DNA replication. *Virology.* 179: 69-77.

Thomas, J.E. and Bowyer, J.W. (1980). Properties of tobacco yellow dwarf and bean summer death viruses. *Phytopathology.* 70: 214-217.

Thomas, J.E., Massalski, P.R. and Harrison, B.D. (1986). Production of monoclonal antibodies to African Cassavamosaic virus and differences in their reactivities with other white-fly-transmitted geminiviruses. *J. Gen. Virol.* 67: 2739-2748.

Thömmes P., Osman, T.A.M., Hayes, R.J. and Buck, K.W. (1993a). Tomato golde mosaic virus leftward gene expression: autoregulation of geminivirus replication protein. *Virology* 195: 275-280.

Thömmes P., Osman, T.A.M., Hayes, R.J. and Buck, K.W. (1993b). TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region. *FEBS Lett.* 319: 95-99.

Thompson, J.D., Higgins, D.G. and Gibson, T.J.(1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix. *Nucleic Acids Res.* 22: 4673-4680.

Timmermans, M.C.P., Das, O.P., and Messing, J. (1992). *Trans* replication and high copy numbers of wheat dwarf virus vectors in maize cells. *Nucleic Acids Research.* 20, 4047-4054.

Tinland, B. (1996). The integration of T-DNA into plant genomes. *Trends in Plant Sci.*1(6):178-184.

Toro, N., Datta, A., Carmi, O.A., Young, C., Prusti, R.K., Nester, E.W. (1989). The *Agrobacterium tumefaciens vir*C1 gene product binds to overdrive, a T-DNA transfer enhancer. *J. Bacteriol.* 171: 6845-6849.

Torres-Pacheco, I., Garzon-Tiznado, J.A., Herrera-Estrella, L., and Rivera-Bustamante, R.F. (1993). Complete nucleotide sequence of pepper huasteco virus: analysis and comparison with bipartite geminiviruses. *J.Gen.Virol.* 74 (Pt 10):2225-2231.

Townsend, R., Watts, J. and Stanley, J. (1986). Synthesis of a viral DNA forms in Nicotiana plumbaginifolia protoplasts inoculated with cassava latent virus (CLV); evidence for the independent replication of one component of the CLV genome. *Nuc. Acids Res.* 14: 1253-1265.

van Regenmortel, M.H.V., Bishop, D.H.L., Fauquet, C.M., Mayo, M.A., Maniloff, J. and Calisher, C.H. (1997). Guidelines to the demarcation of virus species. *Arch. Virol.* 142(7): 1505-1518.

von Arnim, A. and Stanley, J. (1992a). Inhibition of African cassava mosaic virus systemic infection by a movement protein from the related geminivirus tomato golden mosaic virus. *Virology* 187(2):555-564.

von Arnim, A. and Stanley, J. (1992b). Determinants of tomato golden mosaic virus symptom development located on DNA B. *Virology* 186(1):286-293, 1992.

von Wechmar, B. and Hughes F. L., (1992) A rapid technique for typing streak virus isolates using a panel of differential hosts. *Proceedings of the Ninth South African Maize Breeding Symposium 1990.* Gevers, O, H. (ed). pp.115-119.

Ward, A., Etessami, P., and Stanley, J. (1988). Expression of a bacterial gene in plants mediated by infectious geminivirus DNA. *EMBO J.* 7:1583-1587.

Winans, S.C., Alenza, P., Stachel, S.E., Mc Bride, K.E. and Nester, E.W. (1987). Characterisation of the virE operon of the *Agrobacterium* Ti plasmids pTiA6. *Nucleic Acids Res.* 15: 825-837.

Woolston, C.J., Reynolds, H.V., Stacey, N.J. and Mullineaux, P.M. (1989). Replication of wheat dwarf virus DNA in protoplasts and analysis of coat protein mutants in protoplasts and plants. *Nucleic Acids Res.* 17: 6029-6041.

Wright, E.A., Heckel, T., Groenendijk, J., Davies, J.W. and Boulton, M.I. (1997). Splicing features in maize streak vrus virion- and complementary-sense gene expression. *The Plant Journal.* 12: 1285-1297.

Xie, Q., Suàrez-Lòpez, P. and Gutiérrez, C. (1995). Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO J.* 14, 4073-4082.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1885). Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors. *Gene* 33:103-119.

Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G.W., Robinson, D.J. and Harrison, B.D. (1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J. Gen. Virol.* 78: 2101-2111.