Investigations of the molecular determinants of maize streak virus replication

by

Janet Anne Willment

Thesis Presented for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of Microbiology
UNIVERSITY OF CAPE TOWN

June, 1999

CERTIFICATE OF SUPERVISOR

In terms of paragraph GP8 of "General Rules for the Degree of Doctor of Philosophy (PhD)" I, supervisor of the candidate Janet A. Willment, certify that I approve of the incorporation in this thesis of material that has already been published or submitted for publication.

signature removed

Assoc. Prof. E. P. Rybicki
Department of Microbiology
University of Cape Town
South Africa

Abstract

Geminiviruses replicate via a rolling circle mechanism, which initiates at the origin of replication located within the long intergenic region (LIR). The viral replication associated-protein (Rep) in conjunction with the host's DNA replication machinery is responsible for the initiation and termination of the replication cycle from a stem-loop structure, located within the LIR and conserved throughout the three genera of *Geminiviridae*. The specific interaction between the Rep protein with sequences within the intergenic region has been well characterised for the begomoviruses and to some extent the curtoviruses; however, this interaction in the mastreviruses, and in particular maize streak virus (MSV), has yet to be fully explored. A theoretical model has been proposed based on sequence data and informed by the current understanding of replication specificity in begomoviruses. Due to the lack of conservation of the stem sequence of the stem-loop structure amongst mastreviruses, the model implicates a pair of nucleotide sequence repeats called iterons. These are located within the stem structure, and on the complementary sense side of the LIR. The former is the putative site of Rep interaction with the LIR. These iterons would therefore potentially act as the determinants of replication specificity amongst mastreviruses.

A sample population of African streak mastreviruses was assessed for the presence of conserved iterons. Forty-one isolates were analysed by degenerate primer PCR amplification and restriction fragment length polymorphism (RFLP) analysis of the entire RepA coding region, the LIR, and part of the movement protein gene. Although the majority of MSV-like isolates were found to have identical iteron sequences, two grass-infecting MSVs had unique iteron sequences: moreover within this MSV sample there was a considerable degree of sequence diversity as evidenced by phylogenetic analysis of RFLPs and nucleotide sequence data. The presence of distinct maize, wheat- and grass-subtypes of MSV was demonstrated, and amongst the samples both panicum streak (PanSV) and sugar cane streak viruses (SSV) were detected.

From the diversity study, two wheat-type MSVs – MSV-Tas and MSV-VW – were selected based on their overall sequence divergence from the previously sequenced maize-type MSVs. These viruses were completely sequenced and their host ranges were

determined by leafhopper transmissions. These were limited in maize, but broad amongst wheat and barley cultivars. The genomic sequences were compared with that of a typical maize-type MSV isolate (MSV-VM) that was fully sequenced during this study. While the sequence of MSV-VM was most similar to the previously sequenced South African isolates MSV-Komatipoort (Kom) and MSV-SA, MSV-Tas and MSV-VW were their own closest relatives. This is the first report of the full sequences of MSVs found infecting wheat.

Using a range of wild type (WT) viruses selected from the diversity study, the degree to which Rep proteins display specificity for their cognate LIRs was assessed by complementation of replication functions of three different MSV-Kom based Rep mutant constructs. T*rans*-replication of the mutant construct by WT MSV-Kom, MSV-Tas, MSV-VM, MSV-VW, MSV-Set (an MSV isolate from *Setaria*) and PanSV-Kar, was detected in a transient replication assay system. Because the stem sequences of MSV-Set and PanSV-Kar are not the same as those of MSV-Kom, these viruses should not have been able to complement Rep functions in the other viruses. Comparative analysis of their LIR sequences provided a possible explanation for MSV-Set's apparent inefficient and PanSV-Kar's relatively proficient complementation of replication functions.

To further characterise the nature of the replication specificity determinants of MSV-Kom and MSV-Set, regions within the LIR were exchanged between these viruses and replication was assessed by the emergence of symptoms on agroinfected sweetcorn. The results indicated that MSV does not require its cognate stem-loop sequence for replication and that the *cis*-elements responsible for replication specificity reside within the region containing the Rep proximal iteron. Transient *trans*-replication assays of a range of chimaeric constructs by WT viruses confirmed these results. Furthermore, the construction of LIR deletion mutants indicated that both the 5' side of the LIR, up to and including the Rep proximal iteron, and a region 3' of the stem-loop, were essential for replication. This is the first evidence that the MSV replication specificity determinants reside within the Rep proximal iteron region and are not defined by the stem-loop sequence as was previously expected. The MSV *cis*-elements responsible for controlling replication are thus more reminiscent of the begomo- and curtoviruses than was previously anticipated by the proposed model.

Contents

Abbreviations.....2

Chapter 1: Literature Review.....4

Chapter 2: Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data.....45

Chapter 3: Characterisation of three new MSVs.....73

Chapter 4: Investigation of the complementation of replication functions amongst mastreviruses.....118

Chapter 5: Identification of the MSV *cis*-elements involved in replication.....149

Conclusion.....180

Reference List.....183

Abbreviations

General Abbreviations

µg	microgram
µg/l	micrograms per litre
µl	microlitre
µM	micromolar
^{0}C	degrees Celcius
2,4-D	2,4-dichlorophenoxyacetic acid
aa	amino acid
ATP	adenosine triphosphate
bp	base pair
cm	centimeter
Dig	digoxygenin
DNA	deoxyribonucleic acid
dsDNA	double stranded DNA
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid
g	grams
in. Hg	inches mercury
kb	kilobase
l	liter
LB	Luria-Bertani
M	molar
mm	millimeter
mRNA	messenger RNA
N	nucleotide
ng	nanogram
nm	nanometer
nt	nucleotide
NTP	nucleotide triphosphate
OD	optical density
ORF	open reading frame
pg	picograms
psi	pascals per square inch
RLU	relative light units
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
20 x SSC	3 M NaCl, 300 mM Na-citrate solution
ssDNA	single stranded DNA
0.5 x TBE	0.045 M Tris-borate and 0.001 M EDTA buffer
Tris	tris(hydroxymethyl)aminoethane
v/v	volume per volume
w/v	weight per volume
WT	wild-type
X-Gluc	5-bromo-4-chloro-3indolyl glucuronide

Amino Acids

A	alanine
R	argenine
N	asparagine
D	aspartic acid
C	cysteine
Q	glutamine
E	glutamic acid
G	glycine
H	histidine
I	isoleucine
L	leucine
K	lysine
M	methionine
F	phenylalanine
P	proline
S	serine
T	threonine
W	tryptophan
T	tyrosine
V	valine
x	any residue
u	any hydrophobic residue

Nucleic Acids

A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil

Chapter 1

Literature Review

1.1: GENERAL INTRODUCTION **5**
1.2: GEMINIVIRUS OVERVIEW **5**
1.3: THE LIFE CYCLE OF GEMINIVIRUSES **7**
1.3.1: Plant - virus interactions **11**
1.3.2: Vector - virus interactions **13**
1.4: GEMINIVIRUS REPLICATION STRATEGIES **15**
1.4.1: The intergenic regions **16**
1.4.1.1: The virion-sense origin of replication ***19***
1.4.1.2: Replication specificity determinants ***26***
1.4.2: Complementary sense origin of replication **28**
1.4.3: Replication associated protein **29**
1.4.3.1: Post-transcriptional regulation of Rep ***34***
1.4.2.2: The role of Rep during rolling circle replication ***35***
1.4.3.3: Transcriptional control of the mastrevirus virion-sense genes ***36***
1.4.3.4: Regulation of complementary sense gene expression ***37***
1.4.3.5: Rep interactions with the host cell cycle regulators ***39***
1.4.4: Comparisons with other rolling circle replication systems **41**
1.5: PROJECT AIMS **43**

1.1: GENERAL INTRODUCTION

The replication strategy and genome structures of the *Geminiviridae* share features with both bacteriophages and ssDNA bacterial plasmids, yet they utilise eukaryotic systems to provide the necessary cellular components required for the virus lifecycle. This feature makes geminiviruses uniquely useful tools for the elucidation of molecular interactions within the eukaryotic cell, including DNA replication, gene expression and control of the cell cycle. Discovering how these circular ssDNA viruses replicate and interact with their respective hosts has formed the basis of many studies, which have in turn led to some successful strategies to control these plant pathogens. However, without fully understanding the intricacies of the virus lifecycle, it is unlikely that we will ever successfully overcome the devastating effect these viruses have on economically important crops world-wide. Geminiviruses are responsible for major crop losses of staple foods, such as cassava, potatoes and maize, as well as other cultured crops such as wheat, tomatoes, beans and cotton (Wyatt and Brown, 1996; Thottappilly *et al.*, 1993). How does a virus, which at a minimum is only 2.5 - 3kb in size, so efficiently regulate its replication, gene expression, movement throughout the host, and transmission by insect vectors, in order to cause disease so effectively? In answering this question for geminiviruses, many aspects of eukaryotic molecular biology will simultaneously be resolved.

1.2: GEMINIVIRUS OVERVIEW

There are three genera of Geminiviruses: these are the *Mastrevirus*, *Curtovirus* and *Begomovirus* genera, previously known as Subgroups I, II and III respectively (Mayo and Pringle, 1998; Briddon and Markham, 1995). All geminiviruses are morphologically similar with unique quasi-icosahedral virions, have a closed circular ssDNA genome, are obligately transmitted by insect vectors under natural conditions and rely on their host's replication machinery for replication. Although it is quite easy to assign newly-discovered viruses to a genus, the generic definitions have had to change over the years because as new viruses are identified, the parameters used to define the genera have needed to become broader. For instance, mastreviruses have a single component genome (see Fig. 1.1A) and are transmitted by various leafhopper species. They were thought to have narrow host ranges among monocotyledenous plants (monocots); however, two

mastreviruses, bean yellow dwarf virus (BeYDV) and tobacco yellow dwarf virus (TYDV), infect dicotyledenous plants (dicots) (see Table 1.1) (Liu, L. *et al.*, 1997; Morris *et al.*, 1992). Curtoviruses have a single genomic component (see Fig. 1.1B), infect a wide range of dicots, and were though to be only transmitted by leafhoppers - however tomato pseudo-curly top virus (TPCTV) was found to be transmitted by treehoppers (Briddon *et al.*, 1996a). The begomoviruses have narrow dicot host ranges, are all transmitted by the whitefly *Bemisia tabaci*, and were thought all to have two genomic components (named A and B); however it is now known that many begomoviruses have a single component genome (Briddon and Markham, 1995). Furthermore, a distinction between a group of "Old World" and "New World" begomoviruses has been made: all New World begomoviruses have two genomic components and lack an open reading frame (ORF), AV1, present in all the Old World geminiviruses (see Fig. 1.1C & D) (Palmer and Rybicki, 1998; Padidam *et al.*, 1996; Rybicki, 1994).

The genomic organisations of viruses in the three genera differ, but there are many conserved elements, such as the intergenic region (IR) from which bi-directional transcription of various genes occurs (see Fig. 1.1 for comparisons of genomic organisation). The only viral protein essential for replication is the replication associated protein (Rep) (Elmer *et al.*, 1988). Rep is presumed to be expressed off the spliced mRNA product of the complementary sense DNA (C1/C2) of the mastreviruses (Schalk *et al.*, 1989) or from a single transcript in the curto- (C1) (Frischmuth *et al.*, 1993) and begomoviruses (AC1) (Elmer *et al.*, 1988). All Rep-related transcripts are initiated on the left-hand side of the IR. For ease of reading, all functionally equivalent replication associated proteins regardless of the name of their ORF, will be designated Rep. The A or single component of begomoviruses contains all the necessary viral gene products for replication (the Rep protein) and for encapsidation (the coat protein, CP) (Elmer *et al.*, 1988; Rogers *et al.*, 1986). The CPs are transcribed from the right or virion sense side of the IR in all three genera (Frischmuth *et al.*, 1993; Kallender *et al.*, 1988; Morris-Krsinich *et al.*, 1985). Movement proteins (MP) are expressed off the virion sense DNA of mastre- and curtoviruses, while the two component begomoviruses have two unique movement proteins, BC1 and BV1, bi-directionally transcribed from the B component IR (see Sanderfoot and Lazarowitz, 1996, for a review).

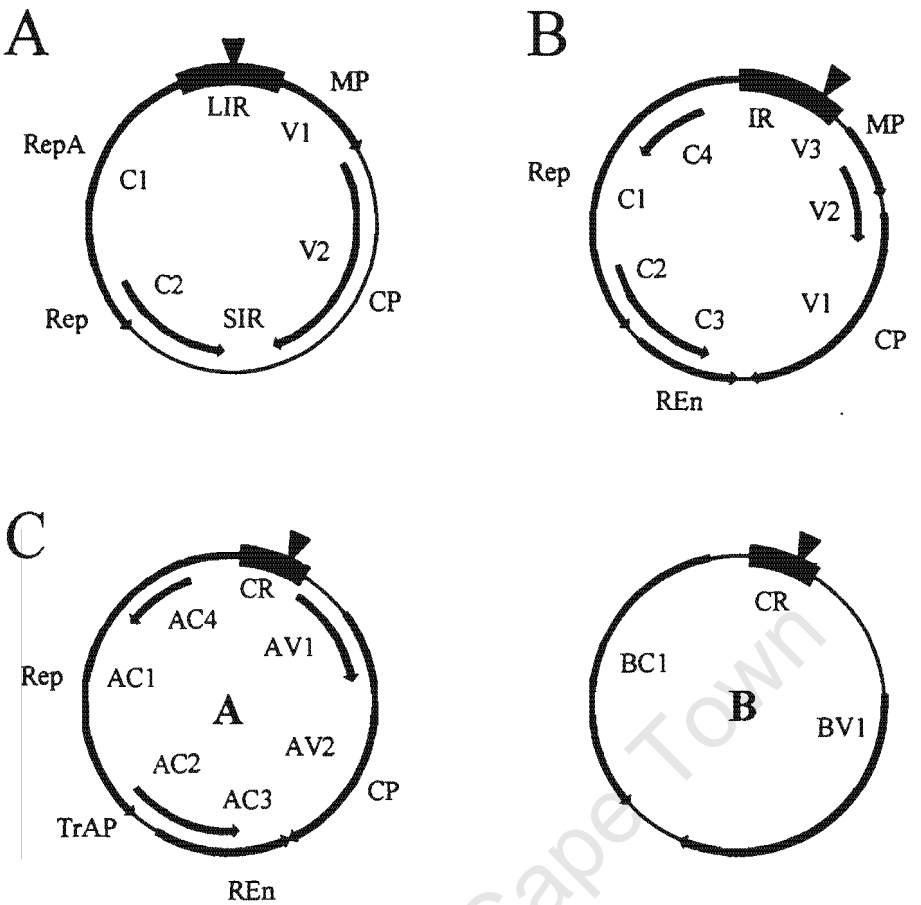


Figure 1.1: Genomic organisation of the three genera of geminiviruses A) *Mastrevirus*, B) *Curtovirus* and C) the A and B components of *Begomoviruses*, respectively. The various open reading frames (ORFs) of both the complementary (C) and virion (V) sense strands, with their respective protein products labelled outside the circle, are identified by the arrows. The intergenic region (IR), from which initiation of replication and the divergent promoters originate, is named the long intergenic region (LIR) in mastreviruses, the IR in curtoviruses, and in begomoviruses is called the common region (CR) due to it being conserved between both A and B components. It is highlighted by the large black arc. Complementary sense strand DNA synthesis initiates within the small intergenic region (SIR) in mastreviruses. Protein products with defined functions are named as follows: the MP, the CP, the viral virion-sense transcriptional activator (TrAP) (Sunter and Bisaro, 1997), the replication enhancer (REn) (Sunter *et al.*, 1990a) and Rep. The mastrevirus Rep is expressed from the spliced mRNA transcript of the C1 and C2 ORFs, while RepA is expressed from the C1 ORF. The B component of begomoviruses expresses two movement proteins, a cell-to-cell transporter, BC1, and a ssDNA binding nuclear transport protein, BV1 (Sanderfoot *et al.*, 1996; Sanderfoot and Lazarowitz, 1996; Noueiry *et al.*, 1994; Pascal *et al.*, 1994). The AC4 gene products of the two component begomoviruses seems to be redundant, while the single component viruses may require it for movement and symptom development within certain hosts (Latham *et al.*, 1997; Pooma and Petty, 1996; Rigden *et al.*, 1994; Jupin *et al.*, 1994). The equivalent curtovirus ORF, C4, has been implicated in inducing host cell division (Latham *et al.*, 1997). The curtovirus C2 ORF gene product does not have a defined function (Hormuzdi and Bisaro, 1995; Sunter *et al.*, 1994) and the V2 ORF affects the levels of ssDNA accumulation (Frischmuth *et al.*, 1993; Hormuzdi and Bisaro, 1993). The begomovirus AV1 ORF may act as an auxiliary movement protein (Padidam *et al.*, 1996).

Table 1.1: Representative species of the three genera of geminiviruses with their host and vector species[a]

Genus	Name[b]	Reference[c]	Host	Vector[d]
Begomovirus (Monopartite)				
	TYLCV-Is	Navot *et al.*, 1991	Solanaceae	*Bemisia tabaci* (W)
	TYLCV-Sar	Kheyr-Pour *et al.*, 1991	Solanaceae	*Bemisia tabaci* (W)
	TLCV-Au	Dry *et al.*, 1993	Solanaceae	*Bemisia tabaci* (W)
Begomovirus (Bipartite)				
	AbMV	Frischmuth *et al.*, 1990	Malvaceae	No vector
	ACMV	Stanley and Gay, 1983	Euphorbiaceae, Solanaceae	*Bemisia tabaci* (W)
	BDMV	Gilbertson *et al.*, 1991a	Leguminosae, Malvaceae	*Bemisia tabaci* (W)
	BGMV-DR	Faria *et al.*, 1994	Leguminosae	*Bemisia tabaci* (W)
	BGMV-GA	Faria *et al.*, 1994	Leguminosae	*Bemisia tabaci* (W)
	EACMV	Zhou *et al.*, 1997	Euphorbiaceae, Solanaceae	*Bemisia tabaci* (W)
	ICMV	Hong *et al.*, 1993	Euphorbiaceae	*Bemisia tabaci* (W)
	PHV	Torres-Pacheco *et al.*, 1993		*Bemisia tabaci* (W)
	SiGMV-Co	Hofer *et al.*, 1997b		*Bemisia tabaci* (W)
	SLCV	Lazarowitz, 1991	Cucurbitaceae	*Bemisia tabaci* (W)
	TGMV	Bisaro *et al.*, 1982; Hamilton *et al.*, 1982	Solanaceae	*Bemisia tabaci* (W)
	TLCV-In	Padidam *et al.*, 1995b	Solanaceae	*Bemisia tabaci* (W)
	ToMoV	Gilbertson *et al.*, 1993	Solanaceae	*Bemisia tabaci* (W)
Curtoviruses:				
	BCTV-Cal	Stanley *et al.*, 1986	44 dicot families	*Circulifer tenellus (L)*
	BCTV-Logan	Hormuzdi and Bisaro, 1993; Stenger *et al.*, 1990	44 dicot families	*Circulifer tenellus* (L)
	BCTV-Worland	Stenger and Ostow, 1996	44 dicot families	*Circulifer tenellus* (L)
	HrCTV	Klute *et al.*, 1996	narrow dicot range	*Circulifer tenellus (L)*
	TPCTV	Briddon *et al.*, 1996a		*Micrutalis malleifera* (T)
Mastreviruses:				
	SSV	Hughes *et al.*, 1993	Gramineae	*Cicadulina mbila (L)*
	MSV	Mullineaux *et al.*, 1984	Gramineae	*Cicadulina mbila* (L)
	PanSV	Briddon *et al.* 1992	Gramineae	*Cicadulina mbila (L)*
	WDV-S	MacDowell *et al.*, 1985	Gramineae	*Psammotettix alienus* (L)
	WDV-F	Bendahmane *et al.*, 1995	Gramineae	*Psammotettix alienus* (L)
	DSV	Donson *et al.*, 1987a	Gramineae	*Nesoclutha declivata* (L)
	TYDV	Morris *et al.*, 1992	Solanaceae, Legumineae	*Orosius argentatus* (L)
	CSMV	Andersen *et al.*, 1988	Gramineae	*Nesoclutha pallida* (L)
	MiSV	Ikegami *et al.*, 1989	Gramineae	Unknown
	BeYDV	Liu, L. *et al.*, 1997	Legumineae	Unknown

[a]: Table based upon Timmermans *et al.* (1994).
[b]: see Briddon and Markham (1995) for a comprehensive list of the members of the geminivirus family.
[c]: Reference given for viral genomic sequence.
[d]: Vector species are whiteflies (W), leafhoppers (L) and treehoppers (W).

The CP, expressed from the V2 or AV1 ORF (Kallender *et al.*, 1988; Morris-Krsinich *et al.*, 1985), is responsible for the characteristic geminate particles. The biological function of the CP, apart from its essential role in specific vector transmission of geminiviruses (Briddon *et al.*, 1990), is believed to be the control of the late stage of the virus lifecycle. The CP influences the amount of ssDNA production, either by sequestering the ssDNA into capsids and making it unavailable for conversion into ds replicative form (RF) DNA, or in a concentration dependent manner – causing a switch in the production of ds DNA to ssDNA (Qin *et al.*, 1998; Azzam *et al.*, 1994). The presence of CP, but not necessarily of capsids, is linked with the accumulation of ssDNA, as was evidenced by CP and TrAP mutants of tomato golden mosaic virus (TGMV), CP mutants of bean golden mosaic virus (BGMV), and the accumulation of squash leaf curl virus (SLCV) ssDNA in an environment not conducive for capsid formation (Qin *et al.*, 1998; Azzam *et al.*, 1994; Sunter *et al.*, 1990).

In the later stage of their life cycle, mastreviruses require the capsid protein for systemic spread; however, whether begomoviruses move as either unencapsidated ssDNA or dsDNA, or as a nucleoprotein complex, is unknown (Sudarshana *et al.*, 1998; Pascal *et al.*, 1994; Lazarowitz *et al.*, 1989; Gardiner *et al.*, 1988). Once encapsidated the virus is acquired by the insect vector, in which it is cycled through the digestive system to the salivary glands, and is finally introduced back into the plant host (Reynaud and Peterschmitt, 1992) (see Table 1.1 for a list of vector species). The precise nature of the interaction between virus and vector is detailed below.

Experimentally the necessity for an insect vector to transmit geminiviruses has been overcome by the discovery that *Agrobacterium tumefaciens* is able to transfer viral DNA into plant cells, allowing infection to occur (Grimsley *et al.*, 1987; Grimsley *et al.*, 1986). This has led to the practice of cloning dimers, complete or partial viral genomes, between the borders of the T-DNA, which is transferred as ssDNA into the recipient cell (Tinland *et al.* (1994) & see Tinland (1996) for a review). The use of *A. tumefaciens* to transmit plant viruses, a technique called agroinfection, has revolutionised the study of geminiviruses by allowing greater freedom of mutational analysis without the necessity of adhering to the strict requirements for vector transmission. Previously it had been noted that some geminiviruses, in particular the curto- and begomoviruses, could be

mechanically transmitted, albeit with some difficulty (Gilbertson *et al.*, 1993; Evans and Jeske, 1993b; Rogers *et al.*, 1986). Low frequency mastreviruses infection of maize seedlings has been achieved using vascular puncture (Louie, 1995). The use of a particle gun to accelerate inert particles coated with viral DNA causing viral infection, in a procedure loosely termed bombardment or biolistics, has been accomplished for numerous geminiviruses, although not very successfully for mastreviruses (Chen and Dale, 1992). This technique has the added advantage of simultaneously introducing the begomoviral A and B genomic components, or a mutant virus plus the DNA containing *trans*-acting gene sequences under investigation, into a single cell type (Jeffrey *et al.*, 1996).

1.3.1: Plant - virus interactions

Geminiviruses interact with the host cell on a number of levels, from requiring the hosts DNA replication machinery to replicate, to interacting with the cell plasmodesmata to allow cell-to-cell movement throughout the plant (Ward *et al.*, 1997; Noueiry *et al.*, 1994). The begomovirus B component genes, provide both a nuclear transport protein, encoded by BV1, and a cell to cell transporter, encoded by BC1 (see Fig. 1.1) (Sanderfoot *et al.*, 1996; Sanderfoot and Lazarowitz, 1996; Noueiry *et al.*, 1994; Pascal *et al.*, 1994). The mastrevirus MP is encoded on the virion sense DNA and is involved in cell-to-cell movement; more specifically, it is associated with the cell wall and plasmodesmata, and its expression correlates with the appearance of symptoms (Dickinson *et al.*, 1996; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989).

The maize streak virus (MSV) CP is responsible for systemic spread and the classic streak symptoms seen in infected plants furthermore, the virus lifecycle is disrupted by the absence of CP (Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989). Despite this, CP and/or MP gene replacements of MSV may still give rise to chlorotic streaks, implying that in the absence of encapsidation and movement, the replicating virus may retain a degree of cell and/or tissue specificity. When infection arises in the leaf primordia, as occurs during agroinfection, the symptoms seem to follow cell lineage, possibly due to the replicating virus damaging the emerging vascular tissue. However, unpublished results quoted by Lucy *et al.* (1996), indicated that there is no evidence of infection of cell lineages during a normal infection (Lazarowitz *et al.*, 1989).

The begomovirus CP is often not required for replication or movement, but it is necessary for systemic movement of some viruses in certain hosts - e.g. tomato yellow leaf curl virus (TYLCV) in *N. benthamiana* and tomato - and for the enhancement of disease progression (Liu, H. *et al.*, 1997; Wartig *et al.*, 1997; Michelson *et al.*, 1994; Rochester *et al.*, 1994; Boulton *et al.*, 1989). Complementation of function between the SLCV CP and its homologous BV1 movement protein indicates that in begomoviruses, the CP may have some role in movement, although it may be redundant in certain hosts (Ingham *et al.*, 1995; Rybicki, 1994). This is seen with TGMV, for which three distinct categories of movement have been observed: CP-independent local and cell-to-cell movement, CP-dependent systemic movement, occurring in a wide range of hosts, and CP-independent systemic movement found in host for which the particular virus is well adapted (Pooma *et al.*, 1996). The accumulation of ssDNA is also associated with the presence of CP, which *in vitro* is a sequence non-specific ss and dsDNA binding protein (Liu, H. *et al.*, 1997; Azzam *et al.*, 1994; Rigden *et al.*, 1993; Boulton *et al.*, 1989). Furthermore, the TYLCV CP binds co-operatively to ssDNA and may have an equivalent function to the BV1 protein due to the presence of a nuclear localisation signal (Kunik *et al.*, 1998; Palanichelvam *et al.*, 1998).

Symptom severity in mastreviruses is directly related to the ability of the virus to systemically infect its host, and is reflected in altered chlorophyll levels (Boulton *et al.*, 1991; McClean, 1947). The extent of chlorotic streaking is related to the stage of development at which plants become infected, and is correlated with the degree of stunting and yield loss experienced (Rose, 1978; Storey, 1928). With certain isolates of MSV the initial symptoms are severe, but later emerging leaves show milder symptoms. This is speculated to be due to the inability of the virus to pace its replication cycle with that of the plant cell division (Pinner *et al.*, 1988). MSV has been detected only in vascular tissue and not the apical meristem within the shoot apex (Lucy *et al.*, 1996). However, in mature leaves the virus is not restricted to the vasculature but found in the cells of leaves displaying the distinctive chlorotic symptoms. These cells include the mesophyll, vascular associated parenchyma and the bundle sheath cells (Lucy *et al.*, 1996). In *N. benthamiana* and sugarbeet infected by the curtovirus beet curly top virus (BCTV), viral DNA is predominantly located within the vascular tissue and infrequently in the adjacent cells. The virus has not been detected in the meristematic tissue or in the

vascular cambium of *N. benthamiana* thereby implying that actively dividing cells are not mandatory for infection (Latham *et al.*, 1997). In localisation studies of begomoviruses it has been shown that Abutilon mosaic virus (AbMV) is never present in any meristematic tissue, while African cassava mosaic virus (ACMV) is found in epidermal and cortical cells of *N. benthamiana* (Sequeira and Harrison, 1982). *In situ* hybridisation studies on AbMV infected leaf material demonstrated that only one phloem cell per 2000 leaf cells was infected (Horns and Jeske, 1991). This was unexpected considering the symptoms of large yellowing areas in mosaics. Horns and Jeske (1991) speculated that the colour pattern is caused by a degeneration of the plastid ultrastructure in the spongy and palisade parenchyma cells in the yellow areas. This was possibly due to a viral induced block in the carbohydrate translocation in the phloem leading to a down regulation of photosynthetically active cells (Horns and Jeske, 1991).

Specific interactions between geminiviruses and their hosts' cellular replication machinery are also required for effective infection and production of symptoms. Suboptimal replication in non-host plant species has been noted for a number of geminiviruses: for example, ACMV in maize protoplasts (Paszkowski *et al.*, 1993) and TYLCV in resistant tomato cultivars (Czosnek *et al.*, 1993). However, host range restrictions are not entirely due to suboptimal replication as was observed with agroinfectious TGMV, which although it does not produce systemic infections in certain tomato cultivars, sugar beet or *Arabidopsis thaliana*, it has been detected replicating in explant tissue (Stenger *et al.*, 1992).

1.3.2: Vector - virus interactions

The spread of geminiviruses is inextricably linked to the ability of their insect vectors to survive in the environment (see Table 1.1 for a complete list of vectors). The mastreviruses are transmitted by a variety of leafhopper vectors of the Homoptera family *Cicadellidae*. While six *Cicadulina* species have been implicated in transmitting MSV, *C. mbila* (Naudè) is the most efficient (van Rensburg and Giliomee, 1990; Rose, 1978). The geographical distribution of *Cicadulina* species covers the continent of Africa, parts of Asia, Australia, the Indian and Pacific Islands and northern parts of South America (Rose, 1978).

have confirmed the role of the capsid in transmission, while the protein sequence determines the vector specificity (Hofer *et al.*, 1997a; Hong and Stanley, 1995; Azzam *et*

al., 1994; Briddon *et al.*, 1990; Mullineaux *et al.*, 1984). (More specifically the TYLCV CP aa 129 – 134 are responsible for assembly of the virions into capsids (Noris *et al.*, 1998).) The high homology between the begomovirus CP sequences is reflected in their transmission by a single vector species, *Bemisia tabaci,* whereas the mastreviruses which have relatively lower degree of CP sequence homology, have a larger number of vector species (Briddon *et al.*, 1994; Briddon *et al.*, 1992; Boulton *et al.*, 1989). Understanding the transmission of two component begomoviruses was complicated by the recent finding that accumulation of ACMV particles within the vector was B component dependent, and that transmission was dependent on the presence of the A component over and above the requirement for the CP (Liu, S. *et al.*, 1997). Thus, for some geminiviruses there may be other virus-encoded factors which influence vector transmission other than the CP.

1.4: GEMINIVIRUS REPLICATION STRATEGIES

The virus life cycle is tightly regulated at the level of transcription, with both up and down regulation of the expression of proteins involved in either replicating or moving the virus. I will present detail of the regulation of the geminivirus life cycle, with particular emphasis on replication and the components of the genome involved in controlling it. The role of viral DNA *cis*-acting sequences essential for replication and control of the viral life cycle will be discussed with reference to *trans*-acting activity of the Rep protein performing these functions.

Geminivirus genomes replicate to high copy number in the nucleus of infected cells (Shen and Hohn, 1995; Timmermans *et al.*, 1992; Abouzid *et al.*, 1988; Dollet *et al.*, 1986), and also apparently in the plastids, at least in the case of AbMV (Frischmuth *et al.*, 1990). Viral genomes exist as minichromosomes (nucleoprotein complexes associated with histones) and are found in cells during the S-phase of cell cycle (Accotto *et al.*, 1993; Pilartz and Jeske, 1992; Abouzid *et al.*, 1988). Harrison *et al.* (1977) first obtained evidence that geminiviruses had single-stranded circular DNA genomes; it was later shown that the virus underwent a ds replicative intermediate during its lifecycle (Ikegami *et al.*, 1981). Replication via a tightly regulated rolling circle (RC) mechanism was subsequently confirmed (see Fig. 1.2) (Stenger *et al.*, 1991; Saunders *et al.*, 1991; Abouzid *et al.*, 1988). The region of the genome that is responsible for control of

replication, and ultimately the high copy number, is the origin of replication located within the viral intergenic region (IR).

Agroinfection or biolistics, generally using tandemly repeated viral sequences containing two IRs, is often performed to analyse the sequences within the IR required for replication (Stanley, 1995; Buragohain *et al.*, 1994; Heyraud *et al.*, 1993a; Heyraud *et al.*, 1993b; Lazarowitz *et al.*, 1992). Replicational release of monomeric circular ssDNA forms of the virus genome occurs by rolling circle replication (RCR) mechanism. Replication is initiated at the first IR by Rep together with the host replication machinery, and proceeds to the second IR where termination occurs, effectively resulting in one replication cycle (Stenger *et al.*, 1991) (see Fig. 1.2). The replicative release of monomeric circular forms occurs preferentially over homologous recombination, even in the presence of large duplicated regions of the viral genome (Heyraud *et al.*, 1993b; Stenger *et al.*, 1991).

1.4.1: The intergenic regions

There are two non-coding regions of the viral genome in mastreviruses, the larger of which is the long intergenic region (LIR), which is functionally equivalent to the begomovirus common region (CR). The smaller is known as the small intergenic region (SIR), and contains the transcription polyadenylation signals reside. The virion associated complementary sense oligonucleotide primer anneals to the virion-sense DNA within the SIR to initiate second strand DNA synthesis (see Fig 1.2) (Townsend *et al.*, 1985; Donson *et al.*, 1984). The LIR, SIR and the Rep protein are required for autonomous replication of mastreviruses (Kammann *et al.*, 1991; Lazarowitz *et al.*, 1989), whereas only the CR and Rep are essential for begomo- and curtovirus replication, which have no such primer binding region within a SIR-like region (Saunders *et al.*, 1992; Frischmuth and Stanley, 1992).

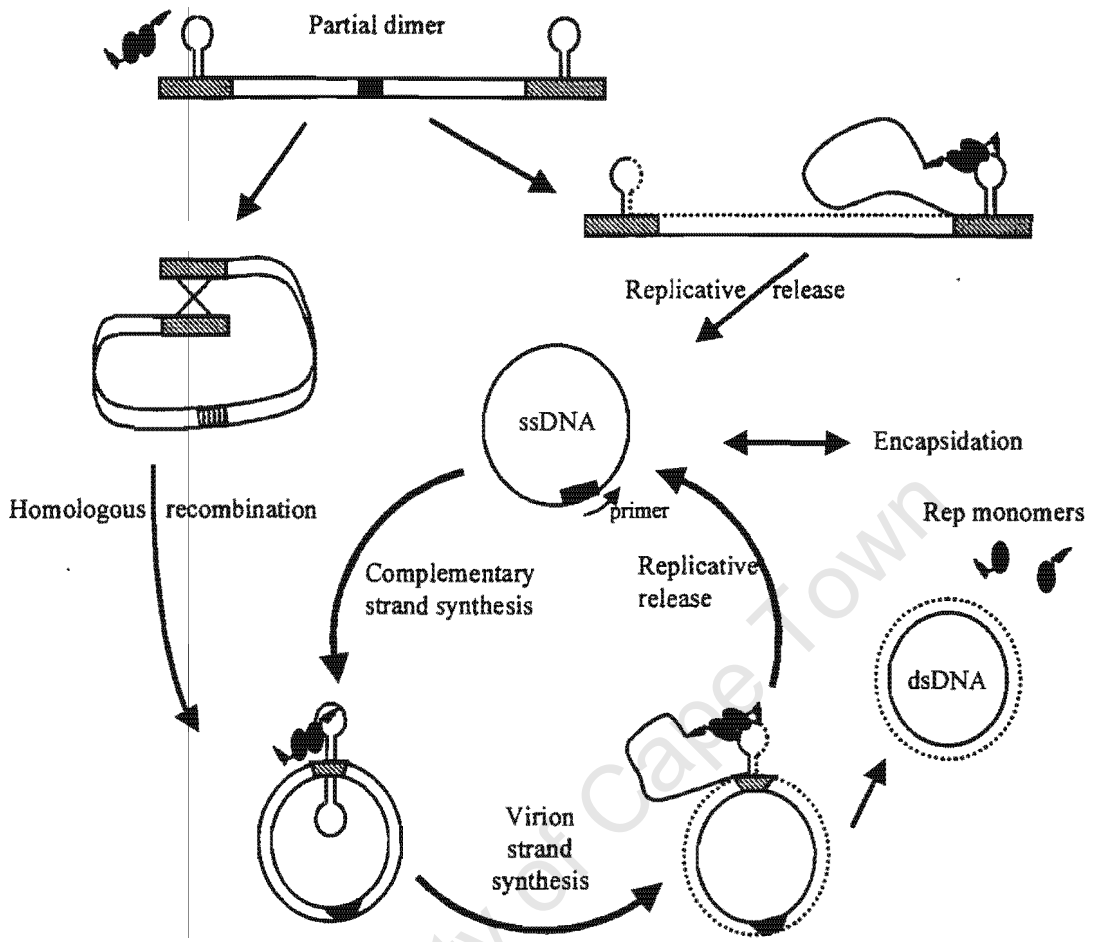


Figure 1.2: Model for the replication of geminiviruses via a rolling circle replication mechanism showing how the product of a replicative release or homologous recombination event arising from a partial dimer enters the replication cycle. The active site tyrosine of the Rep protein (grey oval) is shown interacting with the stem-loop structure within the LIR (obliquely hatched box) (Stanley, 1995; Laufs *et al.*, 1995b; Heyraud *et al.*, 1993b). In mastreviruses a small DNA primer is responsible for the initiation of complementary sense strand synthesis from the SIR (black box) (Hayes *et al.*, 1988; Donson *et al.*, 1984). Based on the diagrams of Stenger *et al.* (1991) and Khan (1997).

The sequence of the LIR is "hypervariable" between viruses when compared to the more conserved protein encoding sequences (Padidam *et al.*, 1995a & Rybicki, 1994). Conversely, the approximately 200 bp long CR of the A and B components of a single begomovirus isolate are highly conserved (Lazarowitz *et al.*, 1992; Hayes and Buck, 1989). There is a direct correlation between the phylogenetic relationships drawn up on the basis of begomovirus CR sequences and cognate Rep gene sequence, probably due to the specific nature of their interaction (Faria *et al.*, 1994; Rybicki, 1994). Sequence comparisons between the IRs even within the same genus are not good indicators of relatedness due to the high degree of sequence variability; however such comparisons may be used to differentiate between isolates / strains within a genus (Padidam *et al.*, 1995a; Rybicki, 1994).

Investigations of both the differences and similarities between the IR / CRs of various viruses have enabled the identification of the *cis*-elements involved in replication. Complementation of protein functions or substitutions of A and B components between different viruses, a process termed pseudorecombination (Stanley *et al.*, 1985), has been successfully used investigate the effect of mutations within the CR of begomoviruses. Pseudorecombination is possible due to the fact that the A component of begomoviruses can replicate autonomously (Rogers *et al.*, 1986). These types of exchanges have allowed investigations of the interaction of Rep proteins with their cognate and non-cognate IRs. Due to their single genomic components, investigations of both the curto- and mastrevirus IR requires alternative strategies to obtain the analogous data. These strategies are described in later chapters.

The LIR sequence contains the plus strand origin of replication (*ori*), and the divergent promoters for the virion and complementary sense genes (Stenger *et al.*, 1991; Dekker *et al.*, 1991; Sunter and Bisaro, 1989; Townsend *et al.*, 1985; Morris-Krsinich *et al.*, 1985). A distinction needs to be made between the *ori* and sequences within the IR responsible for replication specificity amongst geminiviruses. The *ori*, by definition, contains all the essential *cis*-acting features required for efficient RCR (see (Khan, 1997; Espinosa *et al.*, 1995), (del Solar *et al.*, 1993; Waters and Guiney, 1993; Koonin and Ilyina, 1993; Ilyina and Koonin, 1992) for reviews). Replication specificity determinants (RSDs), on the other hand, are elements identified within the LIR that allow interaction between the *cis*- and *trans*-acting components of homologous viruses and prevent heterologous

interactions. Some of these features are both RSDs and essential parts of the *ori* simultaneously. There are some significant variances in both the RSDs and the spatial arrangement of the *ori cis*-elements between the different genera of geminiviruses (see Fig. 1.3) (Argüello-Astorga *et al.*, 1994a; Argüello-Astorga *et al.*, 1994b). In an attempt to elucidate these features for MSV, the type species of the mastreviruses (Briddon and Markham, 1995), I will first present published data on the currently understood interactions in the IR of all three genera of geminiviruses and expand in later chapters of this work on MSV in particular.

1.4.1.1: The virion-sense origin of replication

Common to all geminivirus IRs is a structurally conserved region capable of forming a stem-loop (Sunter *et al.*, 1985): this sequence is necessary for the production of viral genome units (Kammann *et al.*, 1991). Mutants of the TGMV and BGMV IR confirmed that the stem-loop was necessary for DNA replication and was possibly part of the *ori* (Smith and Maxwell, 1994; Stenger *et al.*, 1991; Revington *et al.*, 1989). Lazarowitz *et al.* (1992) initially mapped the begomoviral *ori* to a region approximately 90 nt long, including the stem-loop region and an adjacent 60 nucleotides on the complementary sense side of the IR (see Fig. 1.3). Similar mapping of the mastrevirus *ori* localised it to a region between the Rep TATA box up to and including the stem-loop (Heyraud *et al.*, 1993b; Hofer *et al.*, 1992; Schneider *et al.*, 1992). The region immediately 3' of the stem-loop was not greatly affected by mutations (Schneider *et al.*, 1992). The begomovirus *ori* has been defined as the region starting from the stem-loop structure towards the Rep ORF, which includes direct repeats situated downstream of the Rep TATA box and a GC-box motif, the latter two acting as replication "efficiency elements" (Orozco *et al.*, 1998; Eagle and Hanley-Bowdoin, 1997; Revington *et al.*, 1989). Recent data obtained by Orozco *et al.* (1998) has further defined the TGMV minimal *ori* to an 89-bp region, which contains a complex set of *cis* elements, some of which are essential for replication while others serve only to enhance replication.

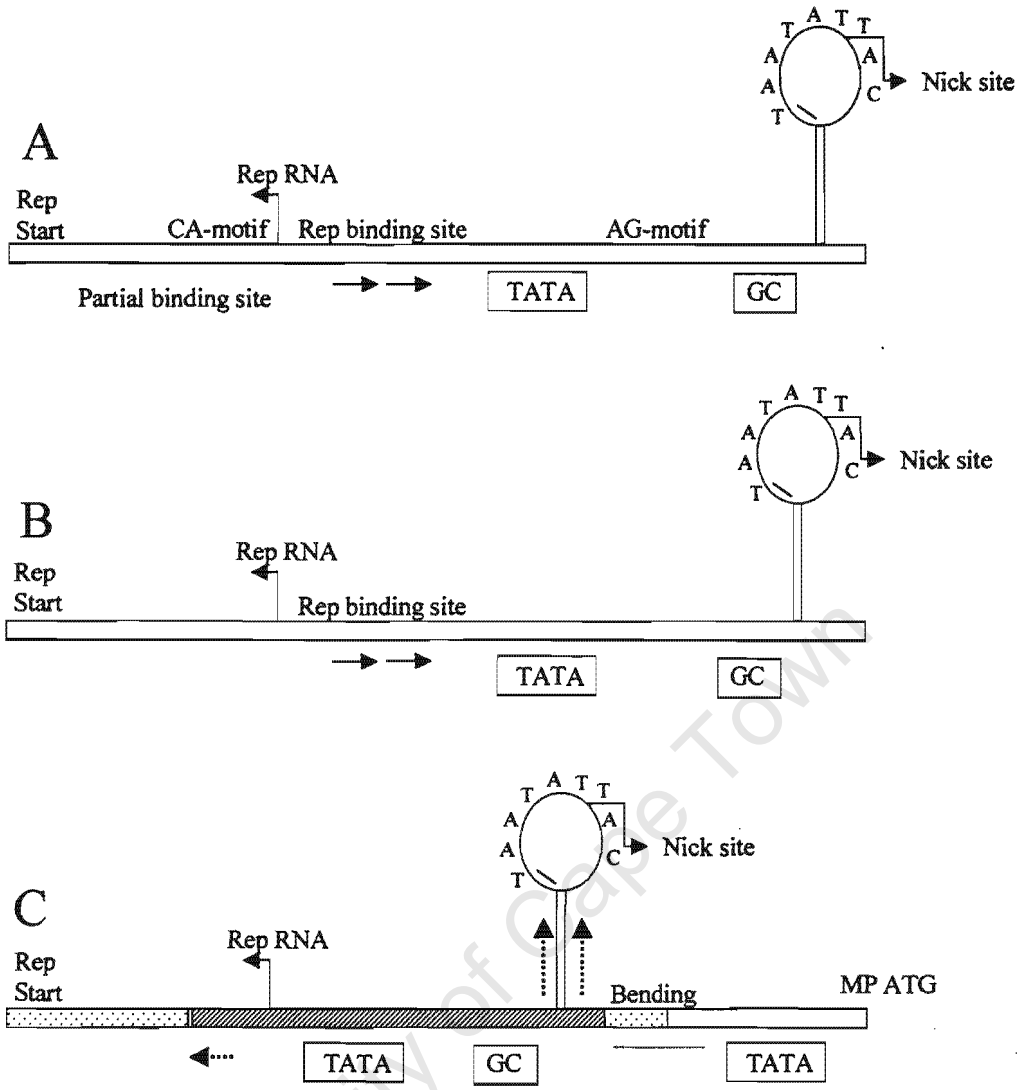


Figure 1.3: A schematic representation of the begomo- (A), curto- (B) and mastreviral intergenic regions (C). Elements essential for replication: the stem-loop structure with its invariant TAATATTAC loop sequence, and the Rep binding sites, shown as straight arrows. Elements involved in enhancing replication in begomoviruses are the host transcription factor binding areas, the TATA and GC boxes (boxed) and the CA-motif and AG-motif (Orozco *et al.*, 1998). Rolling circle replication is initiated by Rep, which nicks the loop sequence at the penultimate A nucleotide (Stanley, 1995; Laufs *et al.*, 1995c). The loop sequence also contains an AT-rich spacer region, shown by the line inside the loop, the sequence of which is not essential for replication (Fontes *et al.*, 1994b). The relative positions of the Rep RNA transcripts to the Rep complementary sense and the MP virion sense start codons (ATG) are indicated. The presence of direct repeats, or iterons, proven to be involved in specific Rep binding are shown in A and B (Choi and Stenger, 1996; Fontes *et al.*, 1994a), while the putative mastrevirus Rep binding sites are indicated in C (Argüello-Astorga *et al.*, 1994a), by the dashed arrows (not to scale). The minimal region required for replication of the mastrevirus wheat dwarf virus (WDV) (C), is shown by an obliquely hatched box, with the 5' and 3' auxiliary regions (dotted boxes) indicated (Sanz-Burgos and Gutiérrez, 1998). The presence of T- rich tracts 3' of the stem-loop in WDV, possibly causing DNA bending is underlined by a dashed line (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995).

The necessity for the begomovirus stem-loop stem structure for replication, rather than its sequence, was shown by mutational analysis (Orozco and Hanley-Bowdoin, 1996). The sequence and AT or GC content of the stem was found to be flexible, as long as a certain degree of base pairing occurred, as some mismatches are tolerated in ACMV (Roberts and Stanley, 1994). Mutations of the left-hand side of the stem were consistently more viable than those on the right. The stem of TGMV may be involved in sequence recognition by a viral protein, possibly REn (see below), and for this reason mutations in one side, the right side in particular, may be deleterious to replication (Orozco and Hanley-Bowdoin, 1996; Hanley-Bowdoin *et al.*, 1996). Alternatively a yet undefined further function of the stem may be required *in planta* (Orozco and Hanley-Bowdoin, 1996). *In vitro* cleavage of oligonucleotides by Rep requires neither the stem-loop structure nor the stem sequence, but the loop sequence is essential (Orozco and Hanley-Bowdoin, 1996).

The begomovirus stem-loop structure is 29- to 32-nucleotides long, with the consensus sequence GGCCA(T/A)CCGN(T/A)(A/T)TAATATTACCGG(A/T)TGGCC (Lazarowitz, 1987). Part of the loop sequence, underlined above, is absolutely conserved amongst all geminiviruses. The invariant nonanucleotide sequence also has high homology with the cleavage sites of ϕX174 (a ssDNA coliphage), pC194 (a bacterial plasmid), circoviruses and multi-component ssDNA nanoviruses of plants e.g. banana bunchy top virus (BBTV) (Burns *et al.*, 1995; Wu *et al.*, 1994). All these systems replicate via a RCR mechanism, which initiates at a stem-loop structure (see later for more detail) (Hafner *et al.*, 1997; Stenger *et al.*, 1991). Although the stem can tolerate some degree of mismatch, deletions or nucleotide substitutions within the loop sequence of ACMV are sometimes restored to wild type (WT) in the presence of a functional IR through homologous recombination with another A component (Roberts and Stanley, 1994). Mutations of the second adenosine to guanine in the invariant nonanucleotide sequence resulted in a non-infectious virus; however, the substitution of a cytosine for the third thymidine was tolerated (Roberts and Stanley, 1994; Schneider *et al.*, 1992).

The presence of a stem structure enhances the termination of the replication cycle. Termination *in planta* is the result of the two ends of the newly synthesised virion sense DNA being covalently joining to form a circular ssDNA genome (see Fig. 1.2). The joining of two oligonucleotides *in vitro*, was enhanced by pairing between the acceptor

strand (3' hydroxyl end of a cleavage site) and the donor strand (5' phosphoryl end bound to Rep) although it was not dependent on this happening (Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995c). The TGMV stem-loop structure and its cognate sequence are both necessary for efficient initiation and termination of replication; this is also presumably true for other begomoviruses (Orozco and Hanley-Bowdoin, 1996). A significant difference between mastre- and begomoviruses was confirmed by the *in vivo* experiments of Heyraud *et al.* (1993), where a second site of initiation of replication in the mastrevirus WDV upstream of the stem-loop in a region unable to form any secondary structural elements was identified. Although initiation of positive strand synthesis occurred from this initiation site, there was no termination of replication, resulting in accumulation of high molecular weight DNA (Heyraud *et al.*, 1993a; Matzeit *et al.*, 1991). Thus while, the requirement for a stem structure for initiation of replication (in this mastrevirus at least) was not absolute, it was essential for termination.

It has been confirmed, using oligonucleotides, that Rep's nicking and joining activity is localised at the penultimate nucleotide of the invariant nonanucleotide sequence (Stanley, 1995; Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995c; Heyraud *et al.*, 1993b). Exchanges between TGMV and BGMV stem-loops, which differ by two nucleotides in the loop sequence part of the AT-rich spacer region, did not affect replication, partly confirming that loop sequences outside the invariant sequence are not essential for replication (Fontes *et al.*, 1994a; Fontes *et al.*, 1994b). However, *in vitro* these sequences did affect efficiency of cleavage and although the presence of the ninth nucleotide in the nonanucleotide sequence is not essential the specific type of base does effect the efficiency of replication initiation (Laufs *et al.*, 1995c). Furthermore the complementary sequence of the invariant 5'-TAATATTAC-3' is 5-GTAATATTAN-3', therefore if the next nucleotide (N) in the sequence is C, this would not allow Rep to differentiate between the plus and minus DNA strand, which could also impact the efficiency of replication initiation (Laufs *et al.*, 1995c). The minimum sequence requirement for *in vitro* Rep cleavage of TGMV DNA is no more than 15 bp including the invariant sequence (Orozco and Hanley-Bowdoin, 1996). Hafner *et al.* (1997) supplied *in vitro* evidence that BBTV Rep was able nick and join oligomers containing only the nonanucleotide sequence with no functional stem. This indicates that neither the sequence nor the structure of the stem was essential for nicking to occur *in vitro*. It was hypothesised that cleavage of the viral *ori* by Rep occurs on ssDNA, based on the

similarities to the cruciform shapes formed due to the genomic and complementary stem-loop structures in bacterial plasmids or parvoviruses (see refs in Orozco and Hanley-Bowdoin, 1996). The inability of TYLCV Rep to cleave dsDNA oligomers *in vitro* is partial proof of this (Laufs *et al.*, 1995c).

The exact region where the TGMV Rep interacts with the IR during replication was first precisely defined by Fontes *et al.* (1994a) as a 5 bp direct repeat separated by a 3 bp central core (5'-GGTAGTAAGGTAG-3'). Furthermore, it was established that while a Rep interaction with the 3' repeat was essential for replication, the presence of a WT 5' repeat enhanced replication, and that the flanking sequences could also potentially influence the interaction (Fontes *et al.*, 1994a). Further sequence analysis identified similar direct repeats in the IRs of other begomoviruses, all of which contained a conserved GG dinucleotide sequence (Orozco *et al.*, 1998; Faria *et al.*, 1994; Fontes *et al.*, 1994a). The curtoviruses also have a direct repeat, proven to be necessary for replication, whose spatial arrangement is similar to those of the begomoviruses (Choi and Stenger, 1996) (see Fig. 1.3B).

Mastreviruses do not have the same types of repeats, and no direct evidence of any specific nucleotides sequences other than the stem-loop structure have been shown to be necessary for replication (Schneider *et al.*, 1992; Kammann *et al.*, 1991). Sequence analysis has, however, shown evidence of a conserved direct repeat in mastreviruses, with one repeat situated on the Rep side of the TATA box and the other forming part of the right-hand side of the stem structure (see Fig. 1.3C) (Argüello-Astorga *et al.*, 1994a; Argüello-Astorga *et al.*, 1994b). Based on analogy with the begomoviruses and other circumstantial evidence, the sequence between the TATA box and the Rep start site containing the Rep proximal half of the repeat could be the putative Rep binding region (Sanz-Burgos and Gutiérrez, 1998; Schneider *et al.*, 1992). The minimal origin of WDV replication was recently delineated to 170 nt upstream and 28 nt downstream of the stem-loop (see Fig. 1.3C). Furthermore, use of deletion mutants defined auxiliary regions on both the 3' and 5'-ends which enhanced replication, and could be involved in stabilisation of the replication complex (Sanz-Burgos and Gutiérrez, 1998). It was proposed that the mastrevirus Rep interacts directly with the stem-loop and that host factors binding at the transcriptional activator sequences also found within the IR, may be involved in maintaining or stabilising this interaction (Fenoll *et al.*, 1990; Fenoll *et al.*, 1988).

Host specific factors have been implicated in virus replication: for example, although the detection of replicating ACMV DNA in non-host cells (i.e. maize protoplasts) indicates that the *ori* is functional in a monocot as well as a dicot cell system; when compared to WDV replication in monocot protoplasts the ACMV replication was 80 – 90 % less efficient. This indicates that either specific host factors were required, or that Rep was not interacting optimally with the DNA replication machinery (Paszkowski *et al.*, 1993). Conversely, TGMV is able to replicate, albeit less efficiently, in cells of non-permissive hosts, with its systemic movement being hampered (Stenger *et al.*, 1992). Specific host nuclear factors are involved in binding to the LIR of mastreviruses within the rightward promoter element, *rpe1,* implying that host DNA binding factors affect replication (see later for more details) (Fenoll *et al.*, 1990; Fenoll *et al.*, 1988). The presence of the *rpe1* is conserved amongst the mastreviruses which have two imperfect GC rich direct repeats separated by a variable spacer region (see Fig 1.3C, where the *rpe1* is indicated as a GC-box). The distances between these *rpe1* elements and the stem-loop structure also differ amongst the various mastreviruses (Fenoll *et al.*, 1990; Lazarowitz *et al.*, 1989). The biological significance of these differences is uncertain; however, differences in this same region in begomoviruses correlate with different host ranges (Hughes *et al.*, 1993; Lazarowitz, 1987).

Spacing within the IR seems to be important, as deletion of certain regions has a negative impact on the interaction between *cis*-elements (Orozco *et al.*, 1998). Sequence alignments of the begomovirus CRs indicate that the distance between the Rep TATA box and start of the CR was conserved (about 87-nucleotides). The most variable region, in terms of size, falls between the TATA box and the conserved stem-loop (ranging from 23 to 82bp) (Fontes *et al.*, 1994a). Within this region there are three elements involved in enhancing replication, which are simultaneously host factor binding sites. The AT rich region around the TATA box was found to be necessary for TGMV origin function, possibly requiring the transcription machinery's role in DNA melting, also necessary for efficient replication (Orozco *et al.*, 1998). The second element is a GC-box region involved in the transcription of the complementary sense genes and contains spatial elements required for replication efficiency (Orozco *et al.*, 1998; Eagle and Hanley-Bowdoin, 1997). Between the GC-box and the TATA box, a third element (AG motif, 5'-GAAGTTC), was identified as essential for replication. Although its exact function is

unknown, it is tentatively presumed to interact with host factors, but it does not affect complementary sense transcription (Orozco *et al.*, 1998).

The potential role of the transcriptional regulators (the TATA- and GC-box motifs) in mastrevirus replication has not been experimentally confirmed (Sanz-Burgos and Gutiérrez, 1998; Fenoll *et al.*, 1990; Lazarowitz *et al.*, 1989). Furthermore, whether exact spatial arrangements are required or not, is also unknown. It has been demonstrated, however, that deletion mutations of the WDV LIR spanning the region between the TATA box of the Rep and the stem-loop abolished replication, implying that either sequences within this region or the distance between the motifs were important for replication (Hofer *et al.*, 1992). There are differences in the IR between the curtovirus BCTV strains. Some share the same direct repeats i.e. putative Rep binding sites, while others differ in the length and sequence of the region between the direct repeats and the stem-loop (Stenger, 1994). The region between the 3' end of the direct repeat and the stem-loop differ in sequence and length between BCTV-CFH and BCTV-Logan; however, evidence of secondary specificity determinants as are required for begomovirus replication was not discovered in BCTV (Choi and Stenger, 1996; Fontes *et al.*, 1994b).

Two further regions within the IR require the particular nucleotide sequences. In the begomovirus CR these are exact purine and pyrimidine sequences; in the mastrevirus LIR these are T tracts (Orozco *et al.*, 1998; Suárez-López *et al.*, 1995). The begomovirus TGMV requires both a 25-bp region on the Rep side of the TATA box, which includes the direct repeats with essential dual roles in replication and transcriptional regulation of the Rep, and a 5'-CCAAAA sequence element or CA motif found outside the minimal *ori,* which is required as an enhancer of replication (see Fig. 1.2A) (Orozco *et al.*, 1998; Eagle and Hanley-Bowdoin, 1997; Eagle *et al.*, 1994; Lazarowitz *et al.*, 1992). The sequence requirement in mastreviruses, and in particular WDV, is a series of T tracts shown to cause bending of the DNA, starting 14 bp 3' of the stem-loop with the centre of the bend 80 bp away from the loop (see Fig. 1.3C). Bending in DNA replication origins has been implicated in the activation of replication in ss plasmids, bacteriophages and yeast cells (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995). Bending within this region could potentially promote the formation of protein complexes necessary for the binding of *trans*-acting factors involved during replication or transcription (Suárez-López and Gutiérrez, 1997; Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995).

In summary, a large number of *cis*-elements are involved in the efficient replication of geminiviruses. The elements involved in begomovirus replication are: the stem-loop structure containing the cleavage site, the Rep binding domain (i.e. the direct repeats), an AT-rich sequence based at the TATA box, a CCAAAA sequence located outside the minimal *ori*, and a AG box and a GC-box 5' of the stem-loop (see Fig.1.3A). The arrangement of all of these elements is conserved amongst the begomoviruses; furthermore spacing between these elements seems to be critical, thereby implying an indirect interaction between two or more elements, probably through DNA binding proteins (Orozco *et al.*, 1998). The unique features of the mastrevirus LIR, compared to those of the begomovirus IR, are: the presence of virus specific stem-loop sequences, direct repeats positioned upstream of the Rep TATA box and in the left-hand side of the stem sequence, the necessity of a stem-loop structure only for termination of replication and not necessarily for initiation, and potentially a DNA bending region downstream of the stem-loop sequence. Not only are these *cis*-elements necessary for efficient replication, but some have a further role in virus-specific replication, a phenomenon shared by all three genera of geminiviruses.

1.4.1.2: Replication specificity determinants

A theoretical approach based upon sequence analysis for the elucidation of sequence specificity elements involved in replication defined the nucleotide sequences of repeated elements - called iterons - for the different genera of geminiviruses (Argüello-Astorga *et al.*, 1994a; Argüello-Astorga *et al.*, 1994b). The assumption was made that viruses with different iterons would not be able to *trans*-replicate each other, i.e. that the iterons acted as RSDs. These iterons differed in sequence, but were spatially conserved amongst the different lineages of geminiviruses.

Experimental evidence for the "iteron theory" was provided by numerous studies. Chimaeras of SLCV and TGMV, which have the same stem-loop but differ in IR sequence, were used to delineate the RSDs to the Rep proximal half of the IR (Lazarowitz *et al.*, 1992). This region constituted 57 bp in SLCV and 52 bp in TGMV 5' of the stem-loop. Although the sequence of the stem-loop was not essential, enhanced replication occurred with WT sequence (Orozco *et al.*, 1998; Orozco and Hanley-Bowdoin, 1996;

Fontes *et al.*, 1994b). More precise mutations of the direct repeats within the IR of a TGMV B component to match those of BGMV destroyed the ability of both the TGMV and BGMV Reps to replicate the B component. This inability of Rep to recognise and *trans*-replicate the mutant indicated that these repeats, as well as further undefined regions, were acting as secondary RSDs (Fontes *et al.*, 1994b). Further evidence regarding the presence of RSDs was supplied by the use of chimaeric Rep proteins, which simultaneously helped elucidate the functional Rep domains involved in replication (Gladfelter *et al.*, 1997). These discoveries are more fully explored in the section on the essential viral proteins involved in replication.

The RSDs of the curtoviruses were identified within a 97 bp region of the IR, which contained a direct repeat, which is believed to be the putative Rep binding site (Choi and Stenger, 1996; Choi and Stenger, 1995; Stenger, 1994). Putative mastrevirus RSDs were identified based on the iteron model due to the fact that these viruses have different stem sequences, and the relative position of the Rep proximal repeat is conserved, 25 – 35 nt downstream of the TATA box (Argüello-Astorga *et al.*, 1994a; Argüello-Astorga *et al.*, 1994b). The repeats (i.e. iterons) are GC rich and in all most all cases the right hand side iteron in the stem-loop is repeated once 5' of the Rep TATA box (see Fig. 1.3C).

The assumption that RSDs should contain sequence specific elements led Arguello-Astorga *et al.* (1994a & b) to identify the stem as being responsible for virus specificity in the replication of mastreviruses. Interestingly, the dicot-infecting mastreviruses BeYDV and TYDV have a similarly positioned direct repeat, a begomovirus-like feature, sharing 5 out of 8 nucleotides, (Liu, L. *et al.*, 1997). Exchanges between the C1/C2 or LIRs of MSV and BeYDV resulted in non-viable recombinants, indicating that the *cis*- and *trans*-acting elements necessary for complementation of replication functions were not conserved (Liu *et al.*, 1999). The only evidence of *trans*-replication between mastreviruses is the unpublished data of Heyraud-Nitschke and Gronenborn quoted in Bendahmane *et al.* (1995), which indicated that the Rep proteins of two WDV strains (WDV-S and WDV-ER) are able to complement replication functions. These two viruses have 16.4 % sequence divergence over the entire genome and have different stem-loop sequences, yet were able to recognise and replicate a heterologous stem-loop (Bendahmane *et al.*, 1995; Heyraud *et al.*, 1993b). Recent data from our lab indicates that MSV-Set, sharing only 78 % homology and having different iterons from another MSV

strain, MSV-Kom, was able to recognise and *trans*-replicate MSV-Kom, albeit inefficiently (Schnippenkoetter, 1998). The precise *cis*-elements, which would allow more efficient replication, were not identified.

1.4.2: Complementary sense origin of replication

Initiation of complementary strand DNA synthesis is a crucial early event in the virus lifecycle, without which the complementary sense genes cannot be expressed. The mastrevirus complementary strand origin is located in the smaller of the two intergenic regions, the SIR, where a DNA primer approximately 80bp long binds to its complementary sequence (positions 1123 – 1206 in MSV-N). This primer initiates DNA synthesis upon cell entry (Kammann *et al.*, 1991; Mullineaux *et al.*, 1984; Donson *et al.*, 1984). The region adjacent to the 3' end of the primer within the C2 ORF has the potential to form two stem-loops, which by analogy with other systems, could be involved in synthesis of ds replicative intermediates (see references within Donson *et al.*, 1984). The presence of a DNA primer has also been detected for WDV (Hayes *et al.*, 1988), digitaria streak virus (DSV) (Donson *et al.*, 1987b), chloris striate mosaic virus (CSMV) (Andersen *et al.*, 1988), sugarcane streak virus (SSV) (unpublished data cited in Hughes *et al.*, 1993) and TYDV (Morris *et al.*, 1992). The primer is encapsidated with the viral genome, potentially giving the virus an expedient way of initiating second strand synthesis and therefore replication (Hayes *et al.*, 1988; Donson *et al.*, 1984).

The distance between the LIR and SIR does not seem to be absolute: extending the distance considerably does not have a deleterious effect on replication; however, bringing the two intergenic regions closer seems to have adverse effects (Matzeit *et al.*, 1991; Lazarowitz *et al.*, 1989). Insertions of sequences of varying lengths (up to 32 bp) into certain sites within the SIR does not affect the ability of an agroinfected MSV mutant to become infectious; however the foreign DNA was often deleted, possibly by homologous recombination (Shen and Hohn, 1991).

The SIR also contains the polyadenylation sequences of the virion and complementary sense genes, similar to the homologous region in the curto- and begomoviruses (Dekker *et al.*, 1991). Unlike the mastreviruses, however, the initiation of complementary strand synthesis in begomoviruses occurs in the IR (Saunders *et al.*, 1992). The minimal origin

of replication within the IR of begomoviruses suggests that initiation of both virion and complementary strand DNA synthesis occurs from this region, while Saunders *et al.* (1992) proposed that an RNA primer initiates the synthesis of dsDNA intermediates within the IR (Lazarowitz *et al.*, 1992).

1.4.3: Replication associated protein

Numerous mutational studies have mapped the viral protein involved in replication of the begomoviruses to the AC1 ORF (Evans and Jeske, 1993a; Lazarowitz *et al.*, 1992; Hanley-Bowdoin *et al.*, 1990; Elmer *et al.*, 1988), of the mastreviruses to the C1/C2 ORF (Kammann *et al.*, 1991; Matzeit *et al.*, 1991) and of curtoviruses to the C1 ORF (Klute *et al.*, 1996; Stanley *et al.*, 1986). The AC1 gene product is equivalent to the Rep encoded from the spliced C1/C2 mRNA of the mastreviruses (Etessami *et al.*, 1991; Hanley-Bowdoin *et al.*, 1990; Schalk *et al.*, 1989) (see Fig. 1.4). The curtovirus Rep is functionally very similar to those of the begomoviruses as is reflected by their closer sequence homology (Rybicki, 1994). There seems to be functional conservation between the Reps of all three genera of geminiviruses, despite their apparent specificity for their cognate IRs (Gladfelter *et al.*, 1997; Jupin *et al.*, 1995; Choi and Stenger, 1995; Laufs *et al.*, 1995b; Fontes *et al.*, 1994b; Lazarowitz *et al.*, 1992). (See Laufs *et al.* (1995a) for a review on Rep protein functions and Hanley-Bowdoin *et al.* (1996 and 1999), and Bisaro (1996) for reviews on geminivirus replication in general.)

Rep proteins are multifunctional. They are responsible for the initiation of virion sense DNA synthesis (Elmer *et al.*, 1988), origin recognition (Fontes *et al.*, 1994b), inducing the accumulation of host proteins necessary for replication (Nagar *et al.*, 1995), interaction with cell cycle regulators (Horváth *et al.*, 1998; Ach *et al.*, 1997; Xie *et al.*, 1996; Xie *et al.*, 1995), and in begomoviruses at least, they negatively regulate their own transcription (Hong and Stanley, 1995; Sunter *et al.*, 1993; Haley *et al.*, 1992).

The Rep proteins share conserved motifs with the functionally similar proteins involved in RCR of bacterial plasmids. These motifs are: motif I, which has the aa sequence FLTY(P/S)xCx and has an unknown function; motif II, which contains two histidine residues positioned within a series of bulky hydrophobic residues, and is thought to bind to Mg^{2+} ions; motif III, which contains a conserved tyrosine residue involved in DNA

cleavage and linkage; and motif IV, which is a putative NTP binding site and has potential helicase activity (Koonin and Ilyina, 1993; Ilyina and Koonin, 1992; Koonin and Ilyina, 1992). Not all these identified motifs have been experimentally proven to play a role in Rep activity. Motif III, in particular the tyrosine residue, has however been proven to be responsible for nicking activity in TYLCV (Hoogstraten *et al.*, 1996; Laufs *et al.*, 1995b).

Rep must specifically interact with and bind to DNA in order for nicking to take place, and has been shown to bind both dsDNA and ssDNA, with higher affinity for the latter *in vitro* (Thommes *et al.*, 1993; Fontes *et al.*, 1992). The binding is sequence-specific as Rep preferentially binds ssDNA sequences from within the IR (Laufs *et al.*, 1995c). Furthermore, not only is the IR sequence specifically bound by Rep, but it has higher affinity for its cognate IR sequence (Jupin *et al.*, 1995; Fontes *et al.*, 1994a; Lazarowitz *et al.*, 1992), (Gladfelter *et al.*, 1997; Fontes *et al.*, 1994b).

The presence of the other RCR protein motifs are required for full Rep activity. The conserved motif IV, with the sequence $EGX_4GKTX_{32}DD$, is a known NTP binding motif, having the characteristic P-loops found in proteins with kinase and DNA helicase activity (Hanson *et al.*, 1995; Gorbalenya and Koonin, 1989). This motif is not required for nicking and joining of ssDNA nor DNA binding *in vitro*; however, it is essential for continuance of the replication cycle *in vivo* (Desbiez *et al.*, 1995; Hanson *et al.*, 1995; Heyraud-Nitschke *et al.*, 1995; Thommes *et al.*, 1993). Mutations within this motif of various begomovirus Reps have either abolished activity or detrimentally impaired its function (Desbiez *et al.*, 1995; Hanson *et al.*, 1995; Eagle *et al.*, 1994; Desbiez *et al.*, 1995). The putative helicase activity is yet to be confirmed using the native protein and not recombinant fusions; however circumstantial evidence, based on homology of the BGMV Rep's NTP binding domain with members of the helicase superfamily III, indicates a potential role of this motif in DNA replication activity (Desbiez *et al.*, 1995; Hanson *et al.*, 1995; Gorbalenya and Koonin, 1989).

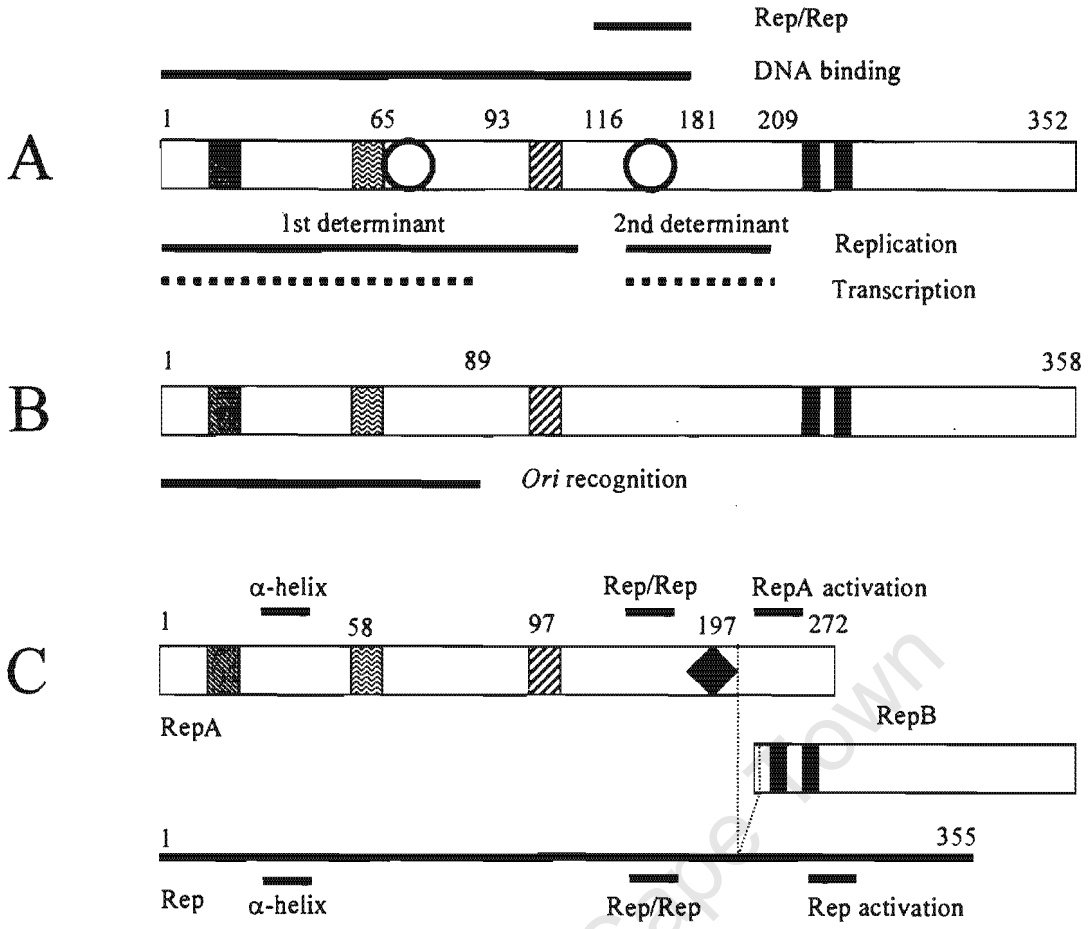


Figure 1.4: A schematic representation of the Rep protein of begomoviruses (A), curtoviruses (B) and mastreviruses (C). Conserved motifs shared with rolling circle replication proteins are indicated as follows, starting from the left: motif I, hatched box; motif II, wavy lined box; motif III, light hatched box; and motif IV, two narrow black boxes (see text for details) (Koonin and Ilyina, 1993; Ilyina and Koonin, 1992; Gorbalenya and Koonin, 1989). Domains found in *myb*-like gene products overlap with motif IV in the Rep protein of all three genera of geminiviruses (not shown) (Hofer *et al.*, 1992). The begomovirus Rep specificity determinants are indicated with respect to TGMV aa positions (numbers above the large boxes) as defined by Gladfelter *et al.* (1997). The domains and the two helix-loop-helix motifs implicated in Rep/Rep oligomerisation and DNA binding are indicated by the circles and solid lines, respectively (A) (Orozco *et al.*, 1997). The virus specific determinants of replication and transcriptional repression, indicated by solid lines under (A), are implicated in interacting with the direct repeats within the CR (see Fig. 1.3A). Secondary determinants necessary for efficient replication and transcriptional repression are underlined, a short solid line and dashed line, respectively (Gladfelter *et al.*, 1997). Amino acid residues 3 to 89 of BCTV (B) implicated in virus specific *ori* recognition and replication are indicated by the solid line (Choi and Stenger, 1995). Although the MSV Rep protein is represented (C), the cleavage and joining activity of RepA has only been confirmed for WDV (Heyraud-Nitschke *et al.*, 1995). A RepB protein product has never been detected *in vivo*, although the MSVs have a start codon for the C2 ORF (Lazarowitz, 1988). The presence of a conserved Rb motif, indicated by the black diamond, known to interact with viral cell cycle regulators, has been identified in most mastreviruses (Horváth *et al.*, 1998; Xie *et al.*, 1996; Collin *et al.*, 1996; Xie *et al.*, 1995). The protein product of the spliced mRNA of the mastrevirus C1/C2 ORF, Rep, is indicated by the long solid line under (C), with the parts contributed by RepA and RepB defined by the dotted line. The Rep and RepA domains identified for oligomerisation, putatively origin recognition (α-helix) and transcriptional activation are indicated by short solid lines (Horváth *et al.*, 1998).

The use of truncated proteins with N- and/or C-terminal deletions have identified the functional domains involved in the various Rep activities (see Fig. 1.4). The N-terminal domains of the begomo-and mastrevirus Rep are essential for cleavage and joining activity (Orozco *et al.*, 1997; Jupin *et al.*, 1995; Heyraud-Nitschke *et al.*, 1995). Cleavage activity of the TGMV Rep was delimited to the N-terminal first 120 residues, and similarly that of TYLCV to the N-terminal 116 residues (Orozco *et al.*, 1997; Jupin *et al.*, 1995).

Two specific domains of the begomovirus Rep protein have been identified as origin recognition determinants. The first domain providing origin specificity is between aa 1 and 116, which recognises the direct repeats within the origin-binding site. The second domain between aa 121 and 209 is not dependent on the first DNA binding domain. This second domain was not detected in BCTV or TYLCV, potentially due to the higher homology between the isolates used to make chimaeric proteins in those experiments (see Fig 1.4) (Gladfelter *et al.*, 1997; Jupin *et al.*, 1995; Choi and Stenger, 1995). The replication determinants overlap slightly with those responsible for the ability of Rep to repress complementary sense transcription, and vary slightly amongst the different begomoviruses (Eagle and Hanley-Bowdoin, 1997; Eagle *et al.*, 1994).

The N- terminal region of the ACMV Rep is involved in down regulation of both its own and of AC4 ORF expression. A larger region of the TGMV and BGMV Rep is required for replication than for specific transcriptional regulation, with aa residues 66 – 93 being required for efficient binding and repression and residues 121 – 209 also contributing to activity (see Fig. 1.4A) (Gladfelter *et al.*, 1997; Hong and Stanley, 1995; Fontes *et al.*, 1994a). The possible reason for this difference is that the initial Rep interaction with DNA is sufficient to repress transcription but more interactions are required for replication. These interactions could either be Rep-Rep or Rep-REn binding or even binding between Rep proteins and host factors (Gladfelter *et al.*, 1997). The Rep proteins from BGMV and TGMV are 82 % similar at the amino acid level, and have different DNA recognition and replication specificities (Fontes *et al.*, 1994b), yet they are able to form active heteroligomers, showing that the protein-protein binding domains are conserved between the two viruses (Settlage *et al.*, 1996). The protein-protein binding domains identified by Orozco *et al.* (1997), between aa 121 – 181, are also essential for DNA binding but not

cleavage activity, implying that Rep multimers are required for efficient replication. The α-helices, identified by computer analysis in TGMV between aa 25 – 52, and another within the oligomerisation domain, may have a role in the binding and cleavage activities (see Fig. 1.4A) (Orozco *et al.*, 1997). Within the MSV Rep are two predicted α-helices, one within the N-terminal domain and the other within the oligomerisation domain, with the former postulated to have a role in specific origin recognition (see Fig. 1.4C) (Horváth *et al.*, 1998).

Interactions between Rep and REn, the begomo- and curtovirus proteins necessary for enhancing replication, form hetero-oligomers which are thought to stabilise Rep's interaction with the IR (Settlage *et al.*, 1996; Hormuzdi and Bisaro, 1995; Sunter *et al.*, 1994; Fontes *et al.*, 1994a; Sunter *et al.*, 1990a). The Rep-REn interaction is not virus specific, although it preferentially occurs between proteins from the same virus, as it even extends to mixtures between curto- and begomoviruses (Gladfelter *et al.*, 1997; Sung and Coutts, 1995; Hormuzdi and Bisaro, 1995; Sunter *et al.*, 1994).

Although the N-terminal domains have been extensively mapped, no firm function other than the ATPase activity of motif IV has been assigned to the C-terminal region. Within this motif, however, are aa residues conserved with the N-terminal DNA binding domains of plant transcription factors. These transcription factors have homology with the avian myeloblastosis (*myb*) oncogene regulatory protein, which opens up further speculation on the numerous roles Rep has within the cell (Hofer *et al.*, 1992). The last 39 C-terminal aa of the TGMV Rep have been implicated as essential regions for replication and repression *in vivo*, as a larger deletion of 139 C-terminal aa enhanced DNA binding four fold *in vitro*. This provided evidence of a DNA binding negative effector within the 39 C-terminal aa (Gladfelter *et al.*, 1997). *In vivo* this part of the Rep may be involved in binding other factors, thereby cancelling the negative effect it may have on DNA binding (Orozco *et al.*, 1997). Similarly it has been hypothesised that the C-terminal region of the mastrevirus Rep may require interaction(s) with host factors to allow exposure of the activation domain (Horváth *et al.*, 1998).

In summary, the Rep protein is absolutely required for RCR of geminiviruses. There are conserved motifs present in the Reps of all three genera, of which only two have so far been confirmed as essential for activity: these are Motif III, containing the nicking tyrosine residue, and Motif IV, which is an NTP binding domain. The N-terminal region of Rep is responsible for DNA binding, protein–protein interactions, cleavage and joining activities. There are two regions within the begomovirus Rep N-terminal domain which contribute towards the specificity its activity in replication, and two regions involved in transcriptional repression. Rep is able to form oligomers, and to interact with REn and various other cell cycle regulators.

1.4.3.1: Post-transcriptional regulation of Rep

A feature of the mastreviruses not shared by the begomo- and curtoviruses is the use of introns for regulation of protein expression. The mastreviruses regulate their Rep gene expression, with Rep and RepA expressed from the same complementary-sense transcript (Wright *et al.*, 1997). The complementary sense ORF mRNAs are expressed at very low levels, and there is also proportionally more unspliced C1 mRNA than the spliced C1/C2 mRNA in MSV infected maize plants (Wright *et al.*, 1997; Mullineaux *et al.*, 1990; Accotto *et al.*, 1989; Morris-Krsinich *et al.*, 1985). The C1/C2 intron of DSV is not efficiently spliced in *Digitaria* and was not spliced at all in tobacco (Mullineaux *et al.*, 1990). The significance of the expression of more RepA relative to Rep is as yet unknown, but may it regulate the change from viral DNA replication to that of encapsidation.

The role of post-translational modifications, and in particular phosphorylation, on regulation of begomoviral Rep activity has been investigated (quoted as unpublished data by Laufs *et al.* 1995a). Regulation of protein activity by phosphorylation is well characterised in eukaryotic systems. A similar effect could regulate Rep activity, and there is evidence of Rep phosphorylation; however the specific domains involved have not yet been identified (Laufs *et al.*, 1995a). Unpublished data of Hartitz and Bisaro (cited in Bisaro 1996) also indicates that the TGMV and TYLCV Reps are phosphorylated, but whether or not this plays a role in Rep regulation is uncertain.

1.4.2.2: The role of Rep during rolling circle replication

Two models exist for Rep activity in the geminivirus IR: in the first Rep binds at the direct repeat, melts the DNA due to its putative helicase activity and then proceeds to nick at the stem-loop (Fontes *et al.*, 1994a). In the other, the combination of transcription factors binding at the GC-box and the TATA box, and the presence of REn binding to Rep and interacting both with Rep and the stem-loop, causes bending of the DNA to meet Rep bound at the direct repeat (Hanley-Bowdoin *et al.*, 1996; Argüello-Astorga *et al.*, 1994b). The resolution of the cycle of replication would require the presence of a second tyrosine residue, presumably from a Rep dimer, with the second Rep supplying the required residue. There are two models for the resolution of Rep's activity: in the first, Rep acts as a oligomer, with the first Rep becoming covalently attached to the 5' end of the nicked DNA by a phosphodiester bond, allowing the free 3'OH group to act as a primer for DNA synthesis (Settlage *et al.*, 1996; Laufs *et al.*, 1995c). Rep would remain attached until a full cycle of replication has occurred and the stem-loop is reached again at which point a second Rep nicks the newly formed structure (see Fig. 1.2). The new free 3'-OH acts as an acceptor of the ssDNA from the first Rep, while the second Rep becomes covalently attached to the new 5' end resulting in a continuous replication mechanism (Orozco and Hanley-Bowdoin, 1996; Laufs *et al.*, 1995b). The second model would require additional Rep tyrosine residues possibly at motif I, and the separation of the linked DNA strand from the Rep from one tyrosine residue to another. However, neither scheme has yet been proved (Ilyina and Koonin, 1992).

Evidence of begomo- and mastrevirus Rep proteins forming oligomers lends more weight to the first model (Sanz-Burgos and Gutiérrez, 1998; Horváth *et al.*, 1998; Settlage *et al.*, 1996). Dimerization of Rep would provide 2 active site tyrosines; furthermore, each of the subunits involved in the oligomer could be responsible for binding to one of the repeated motifs in the DNA binding site of the IR, binding at secondary sites identified in the IR by Arguello-Astorga *et al.* (1994 a & b), and/or the binding specificity regions identified by (Orozco *et al.*, 1998; Fontes *et al.*, 1994a; Fontes *et al.*, 1994b).

1.4.3.3: Transcriptional control of the mastrevirus virion-sense genes

The core MSV virion-sense promoter required for MSV virion sense transcription was delimited to the region spanning the start of the CP gene through to the 3' side of the stem-loop (see Fig. 1.3C). A further region within the LIR which includes the stem-loop and the *rpe1* element sequences, called the upstream activating sequence (UAS), has been shown to enhance expression from the CP promoter (Hofer *et al.*, 1992; Fenoll *et al.*, 1988). There is evidence from a transient expression study undertaken in suspension culture cell lines that the stem-loop structure itself is not required for virion sense transcription (Hofer *et al.*, 1992; Fenoll *et al.*, 1990). The region upstream of the virion sense TATA box has a number of CTTTT tracts and CCAAT box motifs, which have been implicated in causing DNA bending and are often characteristic of promoter regions (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995; Hofer *et al.*, 1992).

Viral sequence elements with homology to plant gene regulatory elements were also identified on the basis of sequence comparisons to known plant transcriptional *cis* regulatory elements (Argüello-Astorga *et al.*, 1994a). An invariant core sequence (CACGTC), which is known to be bound by eukaryotic transcription factors, was identified in the region between the TATA box and the stem-loop. The presence of this element indicates a role for host factors in the transcription of viral genes. Furthermore, levels of transcription from the CP promoter are dependent on the type of host cell (Matzeit *et al.*, 1991).

The evidence for mastreviral Rep proteins acting as transcriptional *trans*-activators was recently analysed by Palmer and Rybicki (1998). They discounted the observations made by Hofer *et al.* (1992) using WDV based constructs to prove Rep's role in *trans*-activation. This was due to the potential of these constructs to form high molecular weight RF DNA, meaning that expression from the virion sense promoter only due to *trans*-activation may have been miscalculated because of the additive effect of genome copy number on gene expression (Palmer and Rybicki, 1998). Other evidence for the *trans*-activation of the virion-sense promoter by Rep comes from a study involving WDV and the expression of the β-glucuronidase (GUS) gene as a V2 gene replacement (Collin *et al.*, 1996). GUS expression, in the context of replication proficient constructs, was far

higher when both Rep and RepA were present than when only Rep was present. *In vitro* studies have shown that RepA is able to nick and join DNA, both essential functions for replication of geminiviruses (Heyraud-Nitschke *et al.*, 1995); however, the presence of a RepA protein *in vivo* has yet to be confirmed. Unlike Rep, RepA lacks a nuclear targeting domain (Hofer *et al.*, 1992) therefore, in order to function in the nucleus RepA would need to interact with Rep. Horváth *et al.* (1998) provide evidence of weak interactions between hetero-oligomers of Rep and RepA, and stronger interactions as homo-oligomers.

It has been proposed that RepA could have numerous roles: these are the down regulation of replication, enhancement of virion-sense gene expression (*trans*-activation) and interactions with host factors (Palmer and Rybicki, 1998; Horváth *et al.*, 1998). An increase in viral replication was observed when Rep was expressed from an intronless C1/C2 ORF for both WDV and BeYDV, which seemed to indicate that the presence of RepA may have a negative effect on replication (Liu *et al.*, 1998; Collin *et al.*, 1996).

In begomoviruese, two different viral proteins are involved in replication and *trans*-activation: these are Rep and TrAP. While both the curto- and begomoviruses have a C2 ORF, only the latter (excepting TYLCV-Sar; Wartig *et al.*, 1997), express a transcriptional activator protein called TrAP. This has a confirmed role in *trans*-activation within specific host plants and tissue types (Sunter and Bisaro, 1997; Hormuzdi and Bisaro, 1995; Sunter *et al.*, 1994; Stanley *et al.*, 1992b; Etessami *et al.*, 1991). The precise mechanism of TrAP activity is unknown; however, *in vitro* it binds both ss and dsDNA non-specifically, and it is known to act at the level of transcription (Sunter and Bisaro, 1997; Noris *et al.*, 1996; Sung and Coutts, 1996; Sunter and Bisaro, 1992; Haley *et al.*, 1992; Sunter *et al.*, 1990). The minimal *cis*-element responsible for the *trans*-activation by TrAP of the pepper huasteco virus (PHV) CP promoter has been recently identified as a 16 bp sequence, which is able to confer responsiveness to heterologous promoters (Ruiz-Medrano *et al.*, 1999).

1.4.3.4: Regulation of complementary sense gene expression

Multifunctional proteins involved in both replication and transcription are not unique to geminiviruses: other viral proteins such as the papillomavirus E1 and E2 proteins and the large T antigen of SV40 also have dual roles (see refs. quoted in Eagle *et al.*, 1994). No

evidence exists for the regulation of the mastrevirus complementary-sense genes by Rep, however, a wealth of information exists on the ability of the begomovirus Rep to regulate its own expression, and also that of the other genes transcribed from the polycistronic mRNA (Mullineaux *et al.*, 1993). Expression of the complementary-sense begomovirus genes, apart from the essential Rep protein, ensures that the virus is able to infect the host systemically and accumulate sufficient levels of viral DNA (Etessami *et al.*, 1991). The necessity for both TrAP and REn expressed from the complementary sense DNA, has been well established (Sunter and Bisaro, 1991; Sunter *et al.*, 1990). Because the virion and complementary sense promoters overlap with the *ori,* a number of subtle switches must control replication and simultaneously co-ordinate expression of viral genes.

The full begomovirus complementary sense promoter, includes the intact IR plus some of the virion sense DNA (133 bp from the start (AC) (Zhan *et al.*, 1991). Regulation of the expression of complementary sense genes by the CP of ACMV was detected using an experimental system in which the CP was highly overexpressed. Other than this example, no other A component encoded protein besides Rep is involved in repression of the AC1 promoter (Hong and Stanley, 1995); furthermore, only the AC1 and not the BC1 promoter is regulated by Rep (Sunter *et al.*, 1993). Down regulation of the complementary-sense promoters of both ACMV and BCTV with the simultaneous repression of AC4 and C4, respectively, was shown to be due to Rep (Hong and Stanley, 1995; Mullineaux *et al.*, 1993). The promoter region involved in repression was mapped to within a 92 bp region immediately upstream of the Rep start codon including the TATA box and transcription start site (Hong and Stanley, 1995), and contains the 13 bp direct repeat essential Rep binding sequences for replication (Eagle *et al.*, 1994; Fontes *et al.*, 1994a). Although the sequences responsible for repression of transcription and replication are the same, these two functions are not linked, and repression was shown to be tissue specific (Eagle *et al.*, 1994).

Mapping the interaction between Rep and the direct repeats within the CR has delimited the regions of both the protein and the CR sequence involved in the repression event. Mutated Rep proteins missing the C-terminal portion of the protein (i.e. having a frameshift mutation from the putative helicase motif), are able to bind to the direct repeat and to repress expression but can not initiate replication. Heterologous Rep protein from BGMV is unable to repress the TGMV Rep promoter, indicating specific recognition of

the RSDs for transcriptional repression as for replication (Fontes *et al.*, 1994a). The exact nature of the Rep protein activity during repression is unknown, but it is hypothesised by Eagle *et al.* (1994) that it may be involved either in interacting with specific host transcription factors, or in competing for their binding sites.

The repression by Rep of its own promoter has been extensively studied using numerous begomoviruses as model systems, and given the presence of an analogous Rep proximal iteron in mastreviruses, a similar repression event may occur. However, this is unproven. Evidence of Rep oligomers binding to this region was recently obtained, where a WDV Rep complex was visualised binding between the TATA box and the start site of the C1 ORF (Sanz-Burgos and Gutiérrez, 1998).

1.4.3.5: Rep interactions with the host cell cycle regulators

Interactions between the mastrevirus Rep and plant cellular proteins, and particularly factors that are involved in regulating the cell cycle, have been recently examined (Horváth *et al.*, 1998; Xie *et al.*, 1996; Collin *et al.*, 1996; Xie *et al.*, 1995). Retinoblastoma proteins (Rb) are negative regulators of the cell cycle and control the mammalian cell cycle by sequestering transcription factors (TF) (e.g. E2F), which are essential for the cell to continue in development into S phase (Hirt, 1996). While hyposphosphorylated, the Rb protein arrests cells in G1 phase. Once either hyperphosphorylation or sequestered by interaction with an oncoviral protein, Rb is inactivated, which allows cellular replication to begin (Grafi *et al.*, 1996). Viral oncoproteins (e.g. SV40 T antigen; see Fanning and Knippers, 1992, for a review) are known to interact with Rb proteins and prevent their regulation of the transcription factors in mammalian systems (Xie *et al.*, 1995).

A motif with the sequence LXCXE, similar to those found in the expressed immediate early genes of oncoviruses, was identified in the C-terminal region of the spliced and unspliced products of the WDV C1/C2 mRNA, Rep and RepA (Xie *et al.*, 1995). It is present in all other sequenced mastreviruses sequenced to date, except SSV, MiSV and CSMV (Liu, L. *et al.*, 1997; Xie *et al.*, 1995). Xie *et al.* (1995) were able to show a direct interaction between the WDV RepA LXCXE motif and Rb (a human Rb homologue, p130) in yeast cells. The ability of a maize Rb homologue (ZmRb) to bind to

WDV RepA has also been demonstrated both *in vivo* and *in vitro,* and the presence of either ZmRb, or p130, in transfected cells causes a reduction of WDV replication (Xie *et al.*, 1996; Grafi *et al.*, 1996). In a yeast two-hybrid system the MSV RepA is able to interact with both the maize and human Rb. Rep is less proficient possibly indicating that it is conformationally changed due to its longer length, its interactions with other proteins or its phosphorylation state (Horváth *et al.*, 1998).

Recently two *Zea mays* genes, *RRB1* and *RRB2,* both encoding Rb-like proteins, were discovered. Although neither curto- nor begomovirus Rep proteins contain the LXCXE motif, TGMV Rep was shown to specifically complex with RRB1 (Ach *et al.*, 1997). This provides evidence that binding motifs other than LXCXE may be present on the geminivirus Rep proteins, allowing binding to cell cycle regulators. These motifs could involve the putative *myb* transcription factor-like binding domain in the C-terminal part of the WDV Rep (Hofer *et al.*, 1992).

Viral replication preferentially occurs in an environment conducive to actively replicating DNA, as evidenced by the replicative form of DSV viral DNA detected in S phase cells (Accotto *et al.*, 1993). Contrasting with these observations was the finding that neither MSV nor WDV replication was reliant on its host cell's DNA replication, but on cell division (Lucy *et al.*, 1996; Timmermans *et al.*, 1992). Replicating TGMV has also been detected in the nuclei of terminally differentiated cells. This was unexpected as geminiviruses, except for AbMV, were previously thought to replicate only in actively dividing plant cells (Nagar *et al.*, 1995; Horns and Jeske, 1991; Matzeit *et al.*, 1991; Boulton and Davies, 1990).

Further evidence that geminiviruses are capable of influencing the cell cycle was provided by the detection of PCNA protein in cells infected with TGMV (Nagar *et al.*, 1995). PCNA is involved in the replication machinery's DNA Polδ processivity (Loor *et al.*, 1997), and both the protein and mRNA had also only been previously detected in actively dividing cells (see refs. in Nagar *et al.*, 1995). PCNA was also detected in transgenic plants expressing the Rep protein, therefore specifically implicating Rep as the inducer of PCNA accumulation. There is evidence of other viral proteins influencing the cell cycle: when the C4 protein of BCTV is expressed in transgenic *N. benthamiana* the plants undergo abnormal development and produce tumorigenic growths (Latham *et al.*, 1997). WT BCTV and other monopartite begomoviruses cause vein swelling, which is seen as

WT BCTV and other monopartite begomoviruses cause vein swelling, which is seen as the familiar symptom of leaf curling due to the division and expansion of parenchyma cells. All of this evidence points towards geminiviruses influencing the host's cell cycle (Nagar *et al.*, 1995; Stanley *et al.*, 1992b).

Based on the evidence provided by both Nagar *et al.* (1995) and Latham *et al.* (1997) it can be inferred that geminiviruses are able to induce DNA replication in quiescent cells, which - due to the nature of plant cells - results in dedifferentiation. The ability to dedifferentiate cells is linked to the finding of WDV Rep interaction with Rb; however no evidence as yet has been found showing accumulation of host replication enzymes due to the presence of WDV Rep (Ach *et al.*, 1997).

1.4.4: Comparisons with other rolling circle replication systems

Although it had been known for some time that geminiviruses undergo rolling circle replication, similar to ssDNA plasmids, it had been assumed that the Rep proteins may not be able to interact with the prokaryotic DNA replication machinery. The discovery that a partially dimeric clone of the TLCV vial genome apparently replicates in *A. tumefaciens* C58 opened up speculation on the origin(s) of geminiviruses (Rigden *et al.*, 1996). This evidence led to speculation that geminiviruses may have evolved from a prokaryotic replicating episomal source, employing RCR as its mechanism for replication (Rigden *et al.*, 1996). The detection of AbMV in plastids, and its prokaryotic promoter-like elements, further supports this speculation (Frischmuth *et al.*, 1990).

The presence of conserved motifs within the Rep proteins of replicons as diverse as those from eukaryotic ssDNA viruses, ssDNA plasmids, phages and Ti plasmids, all of which replicate via a RCR mechanism, strongly implies that they have a common origin (Ilyina and Koonin, 1992; Koonin and Ilyina, 1992). They share similar arrangements of motifs, which are unlikely to evolve independently in several lineages (Koonin and Ilyina, 1993; Ilyina and Koonin, 1992). The simian virus 40 (SV40) T antigen shares several features with those of geminivirus Reps, such as regulation of the virus lifecycle, repression of its own transcription, initiation of viral replication and stimulation of the expression of viral capsid genes. Furthermore, it regulates the host cell cycle by binding to a number of cellular proteins, one of which is the Rb protein; has putative helicase and definite

ATPase activity, acts as a protein complex, and is regulated by phosphorylation (see Fanning & Knippers, 1992 for a review).

Similarities of geminiviral replication to the RCR strategies of bacterial plasmids are remarkable, both as regards the functions of Rep, and the mechanism of replication (see del Solar *et al.*, 1998; Khan, 1997 and Espinosa *et al.*, 1995 for recent reviews). The Rep protein of these plasmids introduces a nick at a specific site and termination involves a Rep-catalysed strand exchange to form a dsDNA and ssDNA product (Espinosa *et al.*, 1995). While the dependence on host factors for replication is absolute, the requirement for viral *cis*-elements both in the *ori* and in the complementary sense origin is more flexible (Espinosa *et al.*, 1995). Conserved features include the highly specific recognition of the *ori* by its cognate Rep and the presence of two regions within the *ori*, one bound by Rep and one nicked by the same protein. These regions are either adjacent to each other or separated by spacer regions between 13 – 91 nucleotides in length. Initiation of complementary sense DNA synthesis begins in a highly heterologous region, even amongst plasmids of the same family, a similarity shared by the SIR of mastreviruses (Schnippenkoetter, 1998; Khan, 1997).

In contrast to most plasmids undergoing rolling circle replication and the circo- and nonoviruses, the direction of geminivirus Rep transcription is in the opposite direction to that of replication. Moreover the *ori* is not encoded within the Rep ORF as is usually the case with plasmids (Khan, 1997). The binding regions are not conserved within plasmids of the same family (e.g. pT181 and pC194), but the nicking regions are, and they form part of a stem-loop structure. Nicking prior to an adenosine residue seems to be conserved amongst all the plasmid families, circo-, nanoviruses and phages, and geminiviruses (del Solar *et al.*, 1998). In contrast to geminiviruses, direct repeats or iterons are not found in the *ori* of most RCR plasmids (Khan, 1997). The sequence requirements for initiation of replication in plasmids are greater than for termination, with various nucleotides enhancing nicking and others involved in joining. Sequences in the left arm of the stem region are less critical for nicking of oligonucleotides, but are essential for joining (Khan, 1997). However, pT181 was shown to require specific sequences found in the stem even if mutations conserved base pairing (see refs. in (Orozco and Hanley-Bowdoin, 1996)). The secondary structures of the *ori* are more essential for efficient initiation than termination, indicating that the DNA structures

during these events may be different (Khan, 1997). A 100 nt DNA bending locus has been identified in pT181, which increases the interaction between the Rep protein and the *ori*. This is strongly reminiscent of the WDV bending locus (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995).

1.5: PROJECT AIMS

The overall objective of this project was the identification of the sequence determinants in the LIR which were involved in maize streak virus replication. As can be seen from the data presented above, extensive analysis of both the intra- and inter-generic interactions between viruses have been used to elucidate the specific nature of viral DNA-protein interactions. I have undertaken a similar approach for the identification of the mastrevirus RSDs. To this end, a diversity study of African streak mastreviruses was initiated, based on the amplification via polymerase chain reaction of most of the C1 ORF, the LIR and part of the V1 ORF. Comparative analysis of these viruses would allow the identification of sequence elements that were conserved, and which therefore could have a role in the non-specific interaction of Rep with the LIR, conversely non-conserved elements would potentially act as replication specificity element(s).

At the time this work was initiated, Argüello-Astorga *et al.* (1994a & b) proposed that "iterons", located within the mastrevirus LIR, were the theoretical specificity determinants of replication. The spatial arrangement of these iterons in mastreviruses is different to those of the begomo- and curtoviruses: specifically, the stem-loop structure contains two of the three iterons and may act as a RSD. I undertook to examine the sample of mastrevirus isolates available to determine the prevalence of viruses with similar iteron sequences, and to identify those with unique iterons. I subsequently expanded the study to give an overall indication of relatedness and diversity amongst the African streak mastreviruses.

All the maize isolates of MSV and a significant proportion of the wheat and grass isolates examined had identical iterons; therefore, two wheat-infecting MSVs and one maize-infecting MSV were selected on the basis of their overall genomic sequence uniqueness when compared to the larger group of ubiquitous maize-type MSVs. The host range of the two wheat MSV isolates was determined using leafhoppers and compared to those of a

maize-type MSV-Kom and a grass-type virus MSV-Set previously determined by Schnippenkoetter (1998). The three viruses were completely sequenced, and features shared with other mastreviruses particularly those involved in replication, were identified.

Theoretical speculations based on sequence data regarding the potential for heterologous LIR interactions has yet to be proven experimentally. To this end, a transient *trans*-replication study was undertaken. A range of WT viruses were assayed for their ability to *trans*-replicate a Rep-mutant virus expressing a marker gene. Once the limits of *trans*-replication had been tested, the regions of the LIR containing the *cis*-acting sequences involved in allowing or inhibiting replication required identification.

During the course of this study W. H. Schnippenkoetter obtained evidence of *trans*-replication occurring between two distinct MSVs (Schnippenkoetter, 1998). This provided an additional incentive for further investigation into the precise location of the *cis*-elements responsible for efficient replication. A two pronged approach was taken: first, LIR deletion mutants were constructed in order to define the minimal origin of replication, and second LIR exchanges were made between two MSV strains, MSV-Kom and MSV-Set, and replication of the chimaeric virus was assessed.

Chapter 2

Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data

2.1: SUMMARY **46**
2.2: INTRODUCTION **47**
2.3: METHODS AND MATERIALS **49**
2.3.1: DNA isolations **49**
2.3.2: PCR amplification and restriction endonuclease analysis **49**
2.3.3: Cloning and sequencing **52**
2.3.4: Phylogenetic analysis **53**
2.4: RESULTS **55**
2.4.1: PCR and RE analysis **55**
2.4.2: Identification of mixed infections **63**
2.4.3: Sequence and phylogenetic analysis **64**
2.5: DISCUSSION **69**

2.1: SUMMARY

DNA from geminiviruses belonging to the "African streak virus" cluster of the genus *Mastrevirus* was amplified by polymerase chain reaction using a single set of degenerate primers. The region amplified spans the C1 open reading frame (ORF), the long intergenic region (LIR) and part of the V1 ORF. The 41 samples analysed included a comprehensive range of maize streak viruses (MSVs), two *Panicum* streak viruses (PanSVs) and a sugarcane streak virus (SSV), all of which were collected over a period of 10 years from different geographical locations. The isolates could be grouped according restriction fragment length polymorphism (RFLP) analysis, which was sufficiently sensitive to reliably differentiate between closely related viruses. A distinction between predominantly maize infecting and grass/wheat infecting MSVs was also observed. Samples obtained from plants with mixed virus infections were analyzed and where possible the composition of the mixed population was determined. Based on published sequences and sequence data presented here, the observed restriction patterns were confirmed and these data was then used for establishing relatedness between streak viruses. A wide geographical distribution of the virus genotypes was noted, while a small group was only present in South Africa.

2.2: INTRODUCTION

MSV is an economically significant plant pathogen because of its detrimental effect on cereal crop yield. These crops include maize, wheat, oats, barley, rye, finger millet, pearl millet, sorghum and Napier fodder grass (Rose, 1978). The leafhopper *Cicadulina mbila* Naudé and its relatives have successfully spread MSV widely across Africa, with closely-related isolates having been identified from the north to the south ends of the continent, as well as from the neighbouring Indian ocean islands of Madagascar, Mauritius and Reunion (Thottappilly *et al.*, 1993; Peterschmitt *et al.*, 1991; Autrey and Ricaud, 1983).

There are a number of mastreviruses that constitute a cluster termed the "African streak viruses": these are the MSVs, *Panicum* streak viruses (PanSVs) and sugarcane streak viruses (SSVs), which differ by up to 40 % based on nucleotide sequence (Rybicki *et al.*, 1998; Hughes *et al.*, 1993; Briddon *et al.*, 1992; Pinner *et al.*, 1988). Originally these species were all assumed to be MSV strains and the differences in pathogenicity were postulated to be due to host-adaptation (Peterschmitt *et al.*, 1991; Pinner and Markham, 1990; Autrey and Ricaud, 1983; Rose, 1978; Bock *et al.*, 1974). Later, these isolates were recognised as distinct species, and it is these species that display the more extreme degrees of host adaptation, rather than strains of MSV (Mesfin *et al.*, 1992; Briddon *et al.*, 1992; Hughes *et al.*, 1991; Pinner *et al.*, 1988). These viruses were initially identified on the basis of their serological relationships, until recently, when analysis based on restriction fragment length polymorphisms (RFLPs), differential hybridisation and the use of PCR and sequence analysis became prevalent (Briddon *et al.*, 1994; Hughes *et al.*, 1992; Hughes *et al.*, 1991).

The C1/C2 ORF and the amino acid sequence, and the CP sequences of mastreviruses can be used to provide accurate estimations of the phylogenetic relatedness of virus isolates (Rybicki *et al.*, 1998; Padidam *et al.*, 1995; Briddon *et al.*, 1994; Rybicki, 1994; Rybicki and Hughes, 1990). A range of severe maize infecting MSVs, isolated from locations throughout Africa and the neighbouring islands were found to be only 2.0 % divergent at the CP amino acid level, most probably due to restraints placed on the virus by host and/or vector interactions with the CP (Briddon *et al.*, 1996a; Briddon *et al.*, 1994). The conserved nature of the CP makes it difficult to differentiate MSV isolates by serological methods (Hofer *et al.*, 1992; Clarke *et al.*, 1989). However, serology can be used to detect the different mastrevirus species, due to their greater CP sequence divergence (Pinner *et al.*, 1992; Hughes *et al.*,

1991). Another disadvantage of serologically based assays is that the detection of mixed populations of viruses will often be masked by the presence of a predominant species (Peterschmitt *et al.*, 1991). It would, therefore, be of considerable value to identify all the streak viruses in a single assay, which the use of degenerate PCR primers allows (Rojas *et al.*, 1993). Geminiviruses are ideally suited to PCR amplification as both their ssDNA and ds replicative forms (RF) can be used as template DNA.

A number of studies on the diversity of MSV and other geminiviruses using PCR have been undertaken (Rybicki *et al.*, 1998; Wyatt and Brown, 1996; Briddon *et al.*, 1994; Rojas *et al.*, 1993; Rybicki and Hughes, 1990). Asymmetric PCR and DNA sequencing of the "hypervariable" intergenic region of begomovirus BGMV isolates was successfully used in a study to determine begomovirus diversity within the Dominican Republic (Gilbertson *et al.*, 1991b). Assessing MSV diversity using PCR and subsequent sequencing has been achieved by amplifying various regions of the genome. The CP and SIR region of 12 MSV isolates were examined by Briddon *et al.* (1994) who identified two groups of viruses: a mainland African group and one from the Indian Ocean Islands; isolates from both groups were severe in maize. Sequence data obtained from a ± 250 bp PCR amplified region of the C2 ORF indicated that there are a number of closely related maize-type MSVs, and another distinct group of related viruses (4 isolates) from grasses or wheat (Rybicki *et al.*, 1998; Rybicki and Hughes, 1990).

Here I present the analysis of African streak viruses from South Africa, Kenya and Zimbabwe, on the basis of restriction endonuclease and partial sequence analysis of both a conserved and hypervariable region of the viral genome, the C1 ORF and LIR, respectively. The diversity of these African streak viruses is discussed with respect to geographical and evolutionary relationships.

2.3: METHODS AND MATERIALS

2.3.1: DNA isolations

Total nucleic acid extractions were performed on harvested leaf material using the methods described by Palmer *et al.* (1998). DNA concentrations were determined using a Beckman DU®-64 spectrophotometer and samples were stored in TE at -20^0C. All routine DNA manipulation techniques were performed as outlined in Sambrook *et al.* (1989). Briddon *et al.* (1994) kindly provided total DNA from a number of previously identified MSV isolates; all other isolates were provided by M.B. von Wechmar, except where indicated (Table 2.1). In some cases, sweetcorn cv. Jubilee was agroinfected with cloned isolates of MSV and compared symptomatically to the original isolate where possible (see Table 2.1) Viral DNA extracted from plant material agroinfected with MSV-MntKA, MSV-MatA and MSV-MatB was provided by D. P. Martin (D. P. Martin and E. P. Rybicki, unpublished data), and that from plants infected with MSV-Set and MSV-Kom by W. H. Schnippenkoetter (1998). The agroinfection and subsequent isolation of viral DNA from MSV-Tas, MSV-VW and MSV-VM infected plants is discussed in chapter 3.

2.3.2: PCR amplification and restriction endonuclease analysis

Degenerate primers were designed to direct the amplification of DNA fragments from all African streak viruses sequenced up until 1996 (E. P. Rybicki, unpublished data). These were derived from the sequences of the following viruses: MSV-SA (Y00514, Lazarowitz, 1988), MSV-Kom (AF003952, D. James, W.H. Schnippenkoetter, F. L. Hughes and E. P. Rybicki, unpublished data), MSV-Set (AF007881, Schnippenkoetter, 1998), PanSV-Ken (X60168, Briddon *et al.*,1994), PanSV-Kar (X60168, this study and W.H. Schnippenkoetter, M. Fyvie & E. P. Rybicki, unpublished data) and SSV-N (M82918, Hughes *et al.*, 1993). The sequence and relative positions of the primers (numbering starts from the penultimate nucleotide of the conserved TAATATTAC sequence of MSV-Kom) are as follows:

Primer 215-234	5'-CGG AGG AGC TGA (TAC)(AC)T TTG G-3'
Primer 1770-1792	5'-TTG G(CGA)C CG(AC) (ACG)GA TGT A(CG)AG-3'

Table 2.1: Origins of the African streak Mastreviruses

Origin	Name	Host	Symptoms	Year	Reference	Provided by
Kenya	MntKA (mix)	Maize*	Severe streak	1996	≈	E. P. Rybicki
Kenya	MSV-MntKA	Maize#	Severe streak	1998	≈	E. P. Rybicki
Kenya	MSV-MntKB	Maize*	Severe streak	1996	≈	E. P. Rybicki
South Africa – W. Cape	MSV-Raw	*Urochloa* sp.*	Mild streak	1998	≈	M.B. von Wechmar
South Africa – W. Cape	MSV-Lib	*Urochloa* sp.*	Mild streak	1998	≈	M.B. von Wechmar
South Africa – W. Cape	MSV-Stell	Maize cv. sweetcorn	Severe streak	1998	≈	M.B. von Wechmar
South Africa – N. KwaZulu-Natal	MSV-Mak	Maize*	Severe streak	1993	≈	M. Barrow
South Africa – N. KwaZulu-Natal	MSV-MakD	Maize*	Severe streak	1998	≈	M. Barrow
Zimbabwe – Matabeleland	MSV-MatA	Maize#	Severe streak	1994	≈	K. E. Palmer
Zimbabwe – Matabeleland	MSV-MatB	Maize#	Severe streak	1996	≈	K. E. Palmer
Zimbabwe – Matabeleland	MSV-MatC	Maize*	Severe streak	1998	≈	K. E. Palmer
South Africa – Transvaal (Tvl)	MSV-WES	Wheat	Severe dwarfing	1989	Hughes *et al.* (1992) & Rybicki *et al.* (1998)	M.B. von Wechmar
South Africa – Orange Free State (OFS)	MSV-VW	Wheat#	Severe streak	1989	Rybicki & Hughes (1990) & Rybicki *et al.* (1998)	M.B. von Wechmar
South Africa – N.Tvl	MSV-Koe	Maize*	Severe streak	1989	Hughes *et al.* (1992)	M.B. von Wechmar
South Africa – N.Tvl	MSV-Breedt	Maize*	$	1989	This work	M.B. von Wechmar
South Africa – N.Tvl	MSV-Thab	Maize*	Mild streak	1989	This work	M.B. von Wechmar
South Africa – E.Tvl	Kom (mix)	Maize*	Severe streak	1989	Schnippenkoetter (1998) & Hughes *et al.* (1992)	M.B. von Wechmar
South Africa – E.Tvl	MSV-Kom	Maize#	Moderate streak	1998	Schnippenkoetter (1998), Hughes *et al.* (1992) & Rybicki *et al.* (1998)	M.B. von Wechmar
South Africa – E. Cape	WW (mix)	Wheat*	Severe	1988	Rybicki & Hughes (1990)	M.B. von Wechmar
South Africa – W.Cape	MSV-Tas	Wheat#	Streak	1991	Rybicki *et al.* (1998)	M.B. von Wechmar
South Africa – OFS	MSV-VM	Maize#	Severe streak	1990	Rybicki & Hughes (1990) & Rybicki *et al.* (1998)	M.B. von Wechmar
South Africa – N. Kwazulu-Natal	Pas (mix)	*Paspalum* sp.*	Very mild	1993	This work	E. P. Rybicki

Table 2.1: Continued

South Africa	MSV-Tyg	Maize*	$	1988	This work	M. B. von Wechmar
South Africa	MSV-Osg	*Urochloa* sp.*	Mild streak	1989	This work	M. B. von Wechmar
South Africa –Kwazulu-Natal	MSV-WES(Elu)	*Eleusine* sp.*	Mild streak	1989	Hughes *et al.* (1992) & Rybicki *et al.* (1998)	M. B. von Wechmar
South Africa – W. Cape	MSV-Geo	Maize*	$	1988	Hughes *et al.* (1991)	M. B. von Wechmar
South Africa – W. Cape	MSV-RSE	Maize*	Severe streak	1988	Rybicki & Hughes (1990) & Hughes *et al.* (1992)	J. Laubscher
Namibia	MSV-SW	Maize*	Severe streak	1987	Clark *et al.* (1989), Rybicki & Hughes (1991), Hughes *et al.* (1992)	H. Hoffman
South Africa – W. Cape	MSV-Cit	Sweetcorn*	Severe	1991	This work	M.B. von Wechmar
Zimbabwe	MSV-Zi (B)	Millet	$	1991	Briddon *et al.* (1994)	
Ethiopia	MSV-E	Maize	Severe	1986	Briddon *et al.* (1994)	
Ghana	MSV-G (A)	Maize	Severe	1985	Briddon *et al.* (1994)	
Ghana	MSV-G (B)	Maize	Severe	1992	Briddon *et al.* (1994)	
Reunion	MSV-Reu	Maize	Severe	1987	Briddon *et al.* (1994)	
Nigeria	MSV-N (B)	Maize	Severe	1990	Briddon *et al.* (1994)	
Nigeria	MSV-N (C)	*Coix lacryma jobi* L.	$	1990	Briddon *et al.* (1994)	
Uganda	MSV-U	Maize	Severe	1991	Briddon *et al.* (1994)	
South Africa – Kwazulu-Natal	MSV-Set	*Setaria* sp.*[#]	Mild streak	1988	Hughes *et al.* (1992)	K. Harborne
South Africa - N. E. Tvl	PanSV-Kar	*Panicum* sp.*	Mild streak	1989	Hughes *et al.* (1992)	M.B. von Wechmar
South Africa – Tvl	PanSV-deK	*Panicum sp.**	Mild streak	1989	This work	M.B. von Wechmar
South Africa – Kwazulu-Natal	SSV-N	*Saccharum officinarum* cv. Uba*	Mild streak	1987	Hughes *et al.* (1992), Hughes *et al.* (1993) & Hughes *et al.* (1991)	K. Harborne

*: Original DNA isolation used for PCR

[#]: Nucleic acid obtained from agroinfected maize cv. Jubilee plants (D. P. Martin and E. P. Rybicki, unpublished results & Chapter 3)

[$]: No record of symptoms.

≈: J. A. Willment, D. P. Martin and E. P. Rybicki, unpublished data

Degeneracies are indicated by the bases in brackets. The PCR reaction contained the following components: 0.5 μM of each primer, 0.25 mM dNTPs, 1x PCR Buffer, which includes 1.5 mM $MgCl_2$, and 1U/50ul *Taq* polymerase (Boehringer Mannheim). Template DNA (10 ng) isolated from the original host was used at all times unless otherwise stated (see Table 2.1). The PCR amplification cycle allowed an initial denaturation of the template DNA for 60 seconds at 94^{0}C, and then the following cycle, repeated 30 times: 45 seconds at 93^{0}C, 30 seconds at 54^{0}C, 90 seconds at 72^{0}C and a final elongation time of 3 minutes at 72^{0}C (Perkin-Elmer GeneAmp® PCR System 9700). The amplification products were then subjected to 0.5 X TBE gel electrophoresis (0.8% agarose) (Sambrook *et al.*, 1989). The DNA was visualised by staining with ethidium bromide and compared with *Pst*I (Boehringer Mannheim) restriction endonuclease (RE) digested lambda phage DNA.

A 5 μl aliquot of the amplified DNA was subjected to RE analysis by incubation with one of the following enzymes: *Rsa*I, *Hpa*II, *Hae*III, *Sau*IIIa, *Bam*HI, *Cfo*I and *Hind*III (Boehringer Mannheim) in their respective buffers as recommended by the manufacturer. The DNA was analysed by gel electrophoresis (as above) except the agarose concentration was increased to 1.5 %. All the possible restriction patterns obtained for each RE were identified and named alphabetically. The different isolates were then grouped on the basis of the number of shared banding patterns obtained when comparing all seven RE patterns.

2.3.3: Cloning and sequencing

Subclones constructed in pBluescript (Stratagene) or pUC vectors, containing the regions covered by the PCR product, were made from original full length genomic clones of the following isolates: MSV-Tas, MSV-VM and MSV-VW (see Chapter 3; Rybicki and Hughes, 1990; F. L. Hughes and E. P. Rybicki, unpublished data). All standard DNA molecular techniques were performed as described by Sambrook *et al.* (1989) and according to the specific manufacturers instruction (Boehringer Mannheim, Promega and Amersham). Large scale DNA preparations were isolated from transformed *Escherichia coli* (JM105, JM109 or DH5α) using the Nucleobond™ Kit (Machery-Nagel, Germany). All sequence data, except those mentioned specifically, were obtained using universal

sequencing primers with the labeled reactions run on an ALF-Express automated sequencer (Pharmacia) (D. James, UCT Microbiology Dept Sequencing Service). A 352 bp *Pst*I *Sal*I subclone of PanSV-Kar (Hughes *et al.*, 1992) was constructed in pBluescript(KS). The clone was sequenced by the chain termination method of Sanger *et al.* (1977), using the Sequenase Version 2.0 (United States Biochemical) kit (as per the manufacturers recommendations) for the incorporation and detection of ^{35}S-dATP. Apart from the *Sal*I *Pst*I subclone, the remainder of the PanSV-Kar genome was sequenced by M. Fyvie & E. P. Rybicki (unpublished data). All sequence data was compiled using sequence from both strands of DNA, with a third independent read when necessary. Sequences were manipulated in either DNAMAN 4.0 (Lynnon Biosoft© 1994 – 1997) or the University of Wisconsin Genetics Computer Group (GCG) package (Devereux *et al.*, 1984).

2.3.4: Phylogenetic analysis

The use of restriction map data for the generation of relationship dendrograms is an established technique, having been successfully used to differentiate between MSV strains and other members of the African streak subgroup of mastreviruses (Hughes *et al.*, 1992; Hughes *et al.*, 1991; Clarke *et al.*, 1989). We have overcome the necessity to first clone the genome under investigation by using the +/- 1300 bp PCR product for RFLP analysis. The accuracy of this method was further increased by the use of sequence data, as the RE patterns observed experimentally and the positions of the restriction sites could be accurately mapped using DNAMAN (Version 4.0, Lynnon Biosoft©). The selection of restriction enzymes included those with 6 and 4 bp recognition sites to allow for, on the one hand, the wide diversity among African streak virus species, and on the other, the apparently limited diversity among the MSVs. The sequence amplified contained both a conserved sequence – the 5' end of the C1 ORF - as well as a highly variable region in the LIR and the 5' end of the V1 ORF (Padidam *et al.*, 1995; Rybicki, 1994; Hughes *et al.*, 1992; Dowling *et al.*, 1990).

Phylogenetic data drawn from the RE mapping results were obtained as follows: a consensus map was drawn for each RE based on the maps generated by DNAMAN. Where two or more RE sites were positioned closer than 20 bp apart for any individual virus, both sites were disregarded. However, if two mapped sites were within 10 bp of

each other when comparing two different sequence derived maps, the site was considered to be conserved between the two maps. This allowed the mapping of "significant" sites in a controlled manner, which is necessary due the asymmetry of losing sites versus gaining sites over time, and the lack of detection of closely positioned sites during gel electrophoresis (Swofford and Olsen, 1990). The number of significant sites per RE map was then compared to all the other possible mapped patterns. The transformation equations of Nei & Li (1979) for restriction data were used to estimate the proportion of ancestral sites conserved, S. Where the formula $S = 2n_{XY}/(n_X + n_Y)$ is based upon the number of identical sites (n) shared between two isolates (n_{XY}) and the number of sites in each isolate is n_X and n_Y, respectively. Where no sites were shared between two isolates, a value of 0.25 and 0.1667, for four and six base pair RE, respectively, was given for n. This is based on the assumption that at least one out of four nucleotides or one out of six nucleotides, for the four or six base pair RE, respectively, was shared between two isolates. The mean number of changes per nucleotide site (d) can then be calculated using the following formula: $d = -(3/2)\ln[(4S^{1/2r} - 1)/3]$, where r is the number of nucleotides in each restriction site. The average number of changes (F) ($F = [d_{RsaI} + d_{HpaII} + d_{HaeIII} + d_{CfoI} + d_{HindIII} + d_{BamHI} + d_{SauIIIa}]/7$) between two viruses was used to generate a distance matrix. The distance matrix was analysed using the neighbour-joining tree program, NJTREE version 2.0, which generates an output file for the construction of a phylogenetic tree using NJDRAW version 1.0 (Saitou and Nei, 1987).

Sequence data of the PCR-amplified fragments from MSV-MatA (AJ012633), MSV-MatB (AJ012634), MSV-MatC (AJ012635), MSV-MntKA (AJ012640), MSV-MakD (AJ012641), MSV-Raw (AJ012639) was obtained from D. P. Martin & E. P. Rybicki (unpublished data). Sequences of the following mastreviruses were also included in the phylogenetic analysis: MSV-SA (Y00514, Lazarowitz, 1988), MSV-Reu (X94330, Peterschmitt *et al.*, 1996), MSV-K (X01089, Howell, 1984), MSV-N (X01633, Mullineaux *et al.*, 1984), MSV-Set (AF007881, Schnippenkoetter, 1998), MSV-Kom (AF003952, Schnippenkoetter *et al.*, in preparation) and SSV-N (M82918, Hughes *et al.*, 1993). The sequences were aligned using the default parameters of the optimal alignment option in DNAMAN 4.0 (Lynnon Biosoft©). The programme utilizes the progressive multiple sequence alignment methods of Feng and Doolittle (1987) and Thompson *et al.* (1994). The output alignment file was used to construct a phylogenetic tree, using the Neighbour-Joining method of Saitou & Nei (1987), which was then subjected to a 1000

bootstrap trials. For presentation purposes the tree was manipulated in Treeview (Win 16©, Page, 1996). These results were compared to the similarity scores and phylogenetic analysis of the derived 261 aa sequence of the RepA protein from the C1 ORF. These data were obtained using DNAMAN 4.0, which utilises the BLOSUM protein weight matrix.

In order to verify the validity of the RFLP generated data for phylogenetic analysis a linear regression curve, comparing the distances generated by sequence data with those generated using restriction-site based analysis, was drawn using Sigma Plot™ (Jandel Scientific).

2.4: RESULTS

2.4.1: PCR and RE analysis

The PCR amplification products consistently gave a band of approximately 1300 bp, for all isolates tested, when electrophoresed on a 0.8% TBE gel (data not shown). The RE treated and agarose electrophoresed products gave the following banding patterns: full data are only shown for *Rsa*I restriction patterns, otherwise only the unique pattern obtained for the other RE are presented (See Table 2.2 for comparative analysis of the different patterns).

*Rsa*I restriction patterns

A large proportion of the MSVs, including some wheat, grass and maize infecting viruses, had the same RE pattern (A, see Fig. 2.1 & 2.2I). A second much smaller group contained seven viruses with an MSV-VM like pattern (Pattern B) with extra *Rsa*I sites, with only the one at position 647 being significant (see Fig. 2.4). MSV-Mak and MSV-MakD had a completely unique pattern (C) as did the Kenyan isolate, MSV-MntKA (H) compared to all the other MSVs. MSV-Set, MSV-Raw, PanSVs and SSV had unique and distinguishable patterns (D, G, F and E, respectively). DNA samples from five collections contained two superimposed banding patterns: WW(mix), Pas(mix), MSV-G(A), Kom(mix) and MntKA(mix) (see Fig. 2.1 lanes 12, 14, 17, 38 & 42).

Table 2.2: Composite restriction patterns for all isolates:

Isolate name[1]	No.	*Rsa*I	*Hpa*II	*Hae*III	*Cfo*I	*Hind*III	*Bam*HI	*Sau*IIIa
MSV-Stell	1	A	A	A	B	A	A	A
MSV-Thab	2	A	A	A	B	A	A	A
MSV-MatA	3	A	A	A	A	A	A	A
MSV-Zi (B)	4	A	A	D	B	G	A	A
MSV-Koe	5	B	A	A	B	A	A	A
MSV-N(C)	6	B	C	A	A	A	A	A
MSV-MatB	7	B	C	A	A	A	A	A
MSV-MakD	8	C	B	A	A	G	A	A
MSV-Mak	9	C	B	B	A	G	A	A
MSV- MntKA	10	H	A	A	A	A	A	A
MSV-Breedt	11	A	A	A	B	B	A	A
MSV-RSE	12	A	A	A	B	B	A	A
MSV-Tyg	13	A	A	A	B	B	A	A
MSV-SA[2]	14	A	A	A	B	B	A	A
MSV-E	15	A	C	A	A	B	A	A
MSV-N[2]	16	A	C	A	A	B	A	A
MSV-Kom	17	A	D	A	B	B	A	A
MSV-Geo	18	B	A	A	B	B	A	A
MSV-VM	19	B	A	A	B	B	A	A
MSV-SW	20	B	C	A	A	B	A	A
MSV-U	21	A	A	A	A	E	A	A
MSV-Cit	22	A	B	B	A	C	A	B
MSV-Lib	23	A	B	B	A	C	A	B
MSV-Tas	24	A	B	B	A	C	A	B
MSV-Osg	25	A	B	C	A	C	A	B
MSV-WES(Elu)	26	A	B	C	A	C	A	B
MSV-WES	27	A	B	C	A	C	A	B
MSV-VW	28	A	B	C	A	C	A	B
MSV-N(B)	29	A	A	A	A	E	A	C
MSV-G(B)	30	A	A	A	A	E	A	C
MSV-MntKB	31	B	A	A	A	A	A	A
MSV-Raw	32	G	I	H	F	D	A	H
MSV-MatC	33	A	B	B	A	A	A	D
MSV-Reu	34	A	E	D	A	D	A	D
MSV-Set	35	D	F	E	C	E	A	E
SSV-N	36	E	G	F	D	E	C	F
PanSV-Kar	37	F	H	G	E	E	B	G
PanSV-deK	38	F	H	G	E	E	B	G
PanSV-Ken[2]	39	I	H	I	E	F	C	I
MSV-G(A)	40	A & B	C	A	A	A	A	A
Mnt-Kenya '97	41	**A**[3] & H	A	A	A	A & B	A	A
Pas(mix)	42	A & C	B & E	**A** & B	A & B	A	A	A
Kom(mix)	43	**A** & C	D, B & A	A	**A** & B	B	A	A
WW (mix)	44	**A** & B	**A** & B	**A**, B &C	A & B	B & C	A	A & B

[1]: Isolates indicated in bold have been sequenced (this work and others, see Table 2.1)
[2]: RE patterns as determined from sequence data only
[3]: Letters in bold are the predominant pattern

Analysis of the composition of mixed samples is discussed more fully below. The distance matrix (d_{RsaI}) generated according to the transformation equations of Nei and Li (1979) based on the proportion of shared significant sites between *Rsa*I maps is presented in Table 2.3. Similar matrices were generated for each RE (data not shown).

*Hpa*II restriction patterns

Most of the maize infecting isolates could be differentiated from the isolates that had been found in wheat/grasses, except for the maize isolates MSV-Mak, MSV-MakD, MSV-MatC and MSV-Cit. A small group of isolates, none of which were found in Southern Africa, included viruses with RE pattern C (generated from sequence data), similar to MSV-N (Mullineaux *et al.*, 1984). MSV-Reu and MSV-Kom each had a unique pattern (E and D, respectively) along with that of MSV-Set, MSV-Raw, SSV and the PanSVs (see Fig. 2.2II).

*Hae*III restriction patterns

The isolates could be differentiated into a group of largely maize infecting MSVs (Pattern A), and a smaller group of wheat, grass and sugarcane infecting isolates (see Fig. 2.2 III). Isolates having pattern C were either grass or wheat isolates, while pattern B contained some *Zea mays* isolates (MSV-Mak, MSV-Cit and MSV-MatC), a grass isolate (MSV-Lib), and a wheat isolate (MSV-Tas). MSV-N(C) was the only grass-infecting isolate in the A group. As with *Hpa*II the PanSVs, SSV, MSV-Set and MSV-Raw were distinct from the other viruses (Patterns G, F, E and H, respectively). MSV-Reu and MSV-Zi(B) were both found to be missing one site compared to the maize infecting group, and thus had a slightly different banding pattern (Pattern D) (393 bp instead of 336 bp in pattern A), and were unique amongst all the isolates tested.

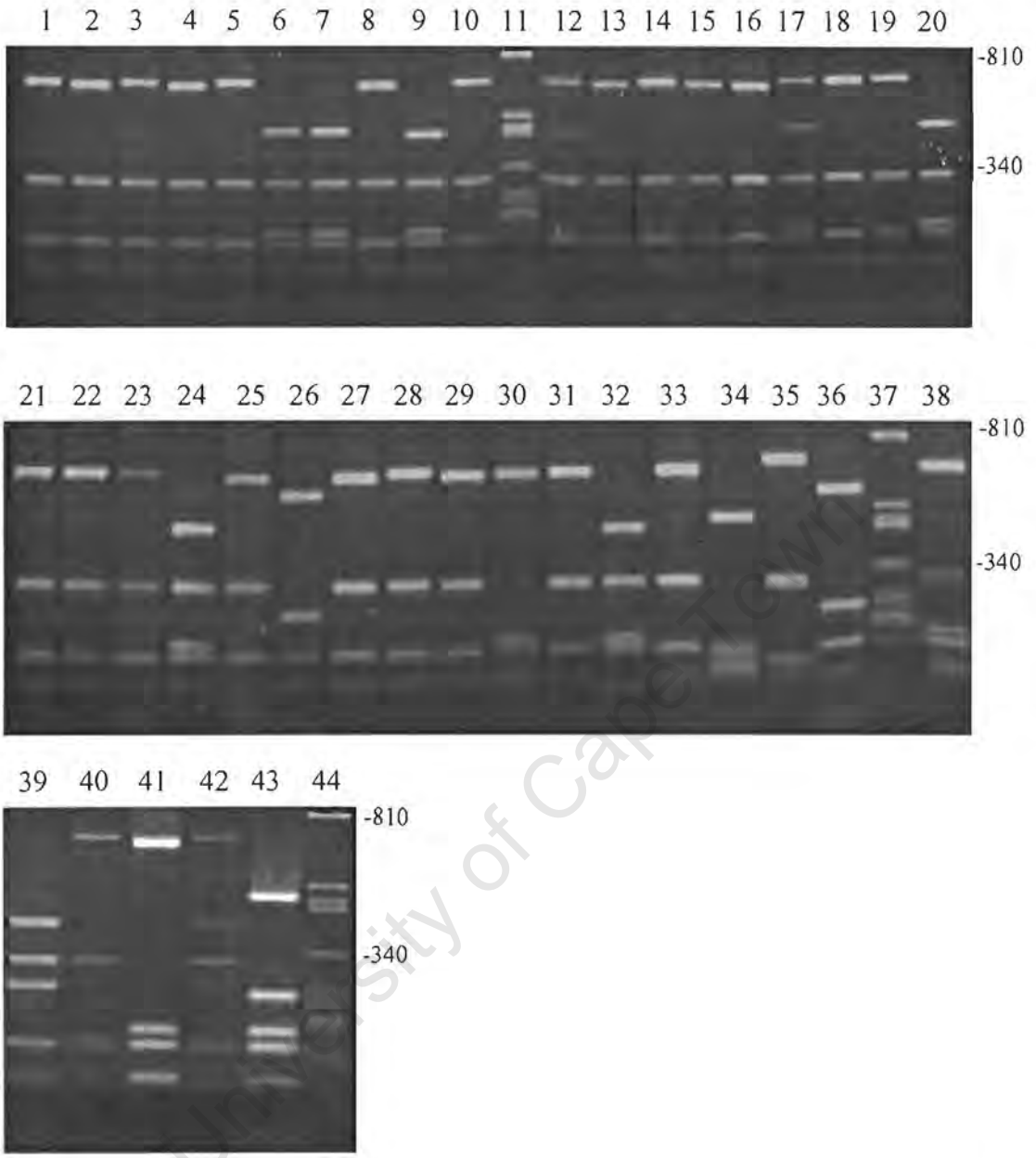


Figure 2.1: Restriction fragment length polymorphism typing of the samples amplified by PCR. All PCR products were restriction endonuclease treated with *Rsa*I. The samples were electrophoresed through a 1.5 % agarose gel in 0.5 x TBE buffer and visualized by ethidium bromide staining. Lane 1, MSV-Lib; lane 2, MSV-Stell; lane 3, MSV-Thab; lane 4, MSV-Tyg; lane 5, MSV-Breedt; lane 6, MSV-Koe, lane 7, MSV-SWA; lane 8, MSV-RSE; lane 9, MSV-Geo; lane 10, MSV-Cit; lane 13, MSV-Wes(Elu); lane 15, MSV-Osg; lane 16, MSV-WES; lane 18, MSV-G(A); lane 19, MSV-G(B); lane 20, MSV-N(C); lane 21, MSV-E; lane 22, MSV-U; lane 23, MSV-MatA; lane 24, MSV-MatB; lane 25, MSV-MatC; lane 26, PanSV-deK; lane 27, MSV-VW; lane 28, MSV-Zi(B); lane 29, MSV- Reu; lane 30, MSV-Mak; lane 31, MSV-Tas; lane 32, MSV-VM; lane 33, MSV-Kom; lane 34, MSV-Set; lane 35, SSV-N; lane 36, PanSV-Kar; lane 39, MSV-MntKA; lane 40, MSV-MntKB; lane 41, MSV-MakD; lane 43, MSV-Raw; lanes 12, 14, 38, and 42, WW(mix), Pas(mix), Kom(mix) and MntKA(mix), respectively, contained more than one banding pattern, and Lanes 11, 37 and 44 contain lambda *Pst*I digested DNA (λ).

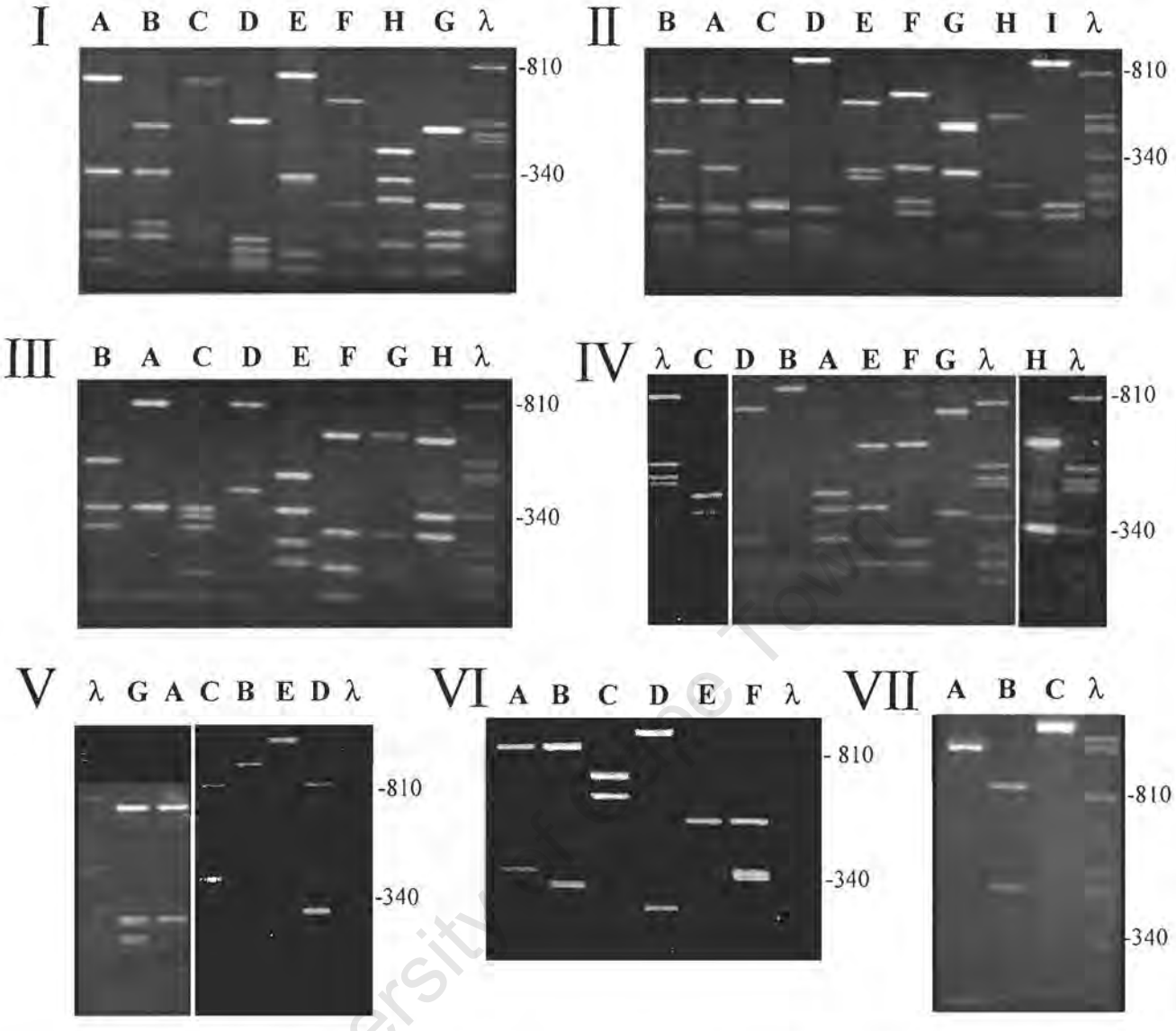


Figure 2.2: The characteristic banding pattern obtained after treatment of the PCR amplification products with the following restriction endonucleases **I)** *Rsa*I, **II)** *Hpa*II, **III)** *Hae*III, **IV)** *Sau*IIIa, **V)** *Hind*III, **VI)** *Cfo*I, and **VII)** *Bam*HI, is represented as follows: **I)** Pattern A, MSV-Kom; pattern B, MSV-VM; pattern C, MSV-MakD; pattern D, MSV-Set, pattern E, SSV-N; pattern F, PanSV-Kar; pattern H, MSV-MntKA and pattern G, MSV-Raw. **II)** Pattern B, MSV-Tas; pattern A, MSV-VM; pattern C, MSV-MatB; pattern D, MSV-Kom; pattern E, MSV-Reu; pattern F, MSV-Set; pattern G, SSV-N, pattern H, PanSV-Kar and pattern I, MSV-Raw. **III)** Pattern B, MSV-Tas; pattern A, MSV-Kom; pattern C, MSV-VW; pattern D, MSV-Reu; pattern E, MSV-Set; pattern F, SSV-N; pattern G, PanSV-Kar and pattern H, MSV-Raw. **IV)** Pattern C, MSV-N(C); pattern D, MSV-Reu; pattern B, MSV-Tas; pattern A, MSV-VM; pattern E, MSV-Set; pattern F, SSV-N; pattern G, PanSV-Kar and pattern H, MSV-Raw. **V)** Pattern A, MSV-MatB; pattern C, MSV-Tas; pattern B, MSV-Kom; pattern D, MSV-Reu, pattern E, MSV-Set and pattern F, MSV-Raw and pattern G, MSV-MakD. **VI)** Pattern A, MSV-Tas; pattern B, MSV-Kom; pattern C, MSV-Set; pattern D, SSV-N; pattern E, PanSV-Kar and pattern F, MSV-Raw. **VII)** Pattern A, MSV-Kom; pattern B, PanSV-Kar and pattern C, SSV-N. All DNA products were electrophoresed through an ethidium bromide stained 1.5 % agarose gel in 0.5 x TBE. λ *Pst*I digested DNA was used as a marker in all gels and sizes are indicated in base pairs.

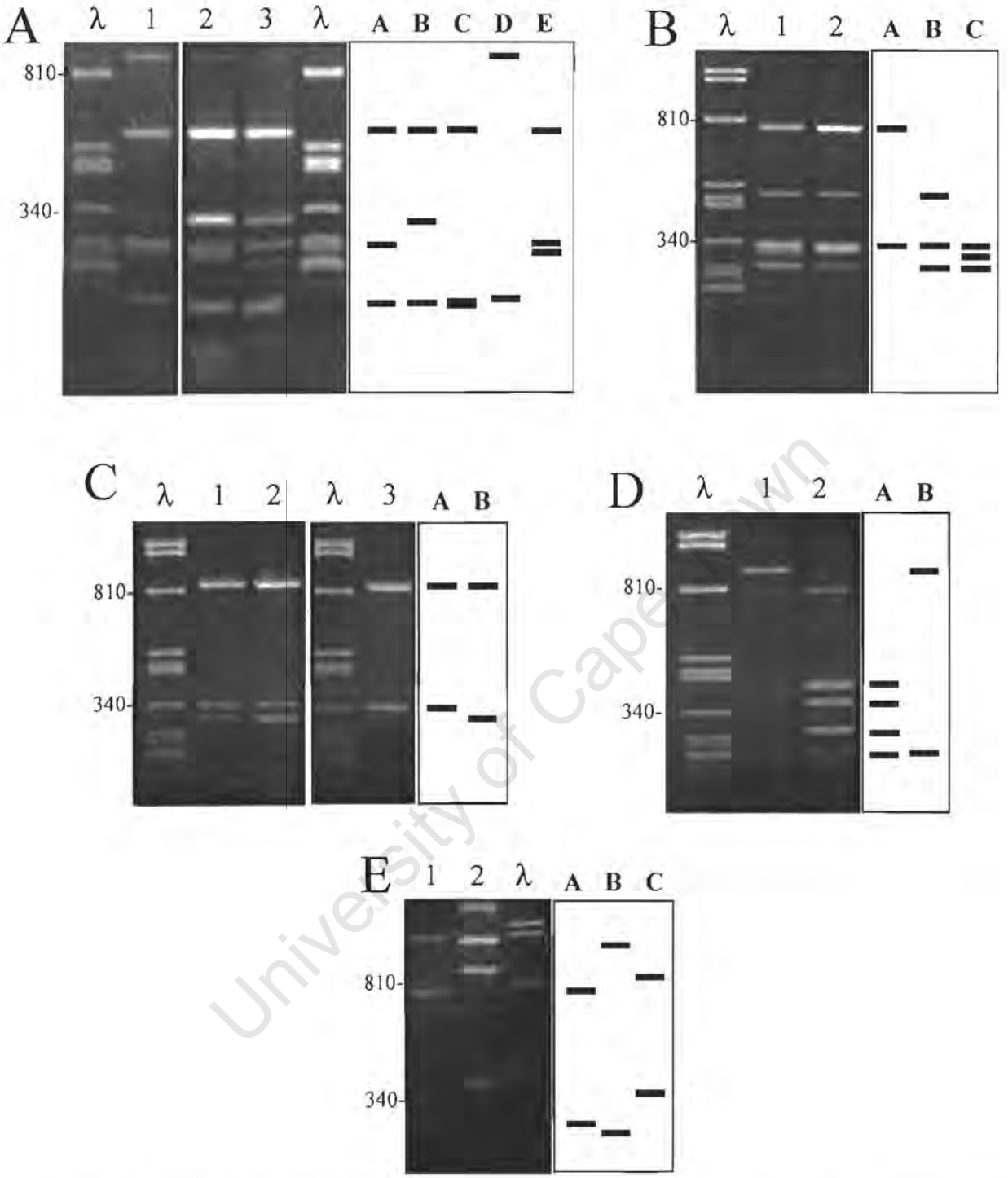


Figure 2.3: DNA isolations containing mixtures of viral DNA were detected after restriction endonuclease digestion of the PCR products with the following enzymes A) *Hpa*II, lane 1, Kom(mix); lane 2, Pas(mix), lane 3, WW(mix); B) *Hae*III; lane 1, WW(mix) and lane 2, Pas(mix); C) *Cfo*I; lane 1, WW(mix) lane 2, Pas(mix) and lane 3, Kom(mix); D) *Sau*IIIa, lane 1, WW(mix) and for reference lane 2, Pas(mix) with only one pattern; E) *Hind*III, lane 1, MntKA(mix) and lane 2, WW(mix). Partially digested bands were obtained for both the *Hind*III and *Sau*IIIa samples. All 1.5 % agarose gels were electrophoresed in 0.5x TBE and stained with ethidium bromide. λ *Pst*I DNA was used as a marker, with band sizes indicated as number of base pairs. A graphical representation of the individual banding patterns, which in certain combinations may contribute towards the complex patterns observed in each gel, is shown on the right (see Table 2.2).

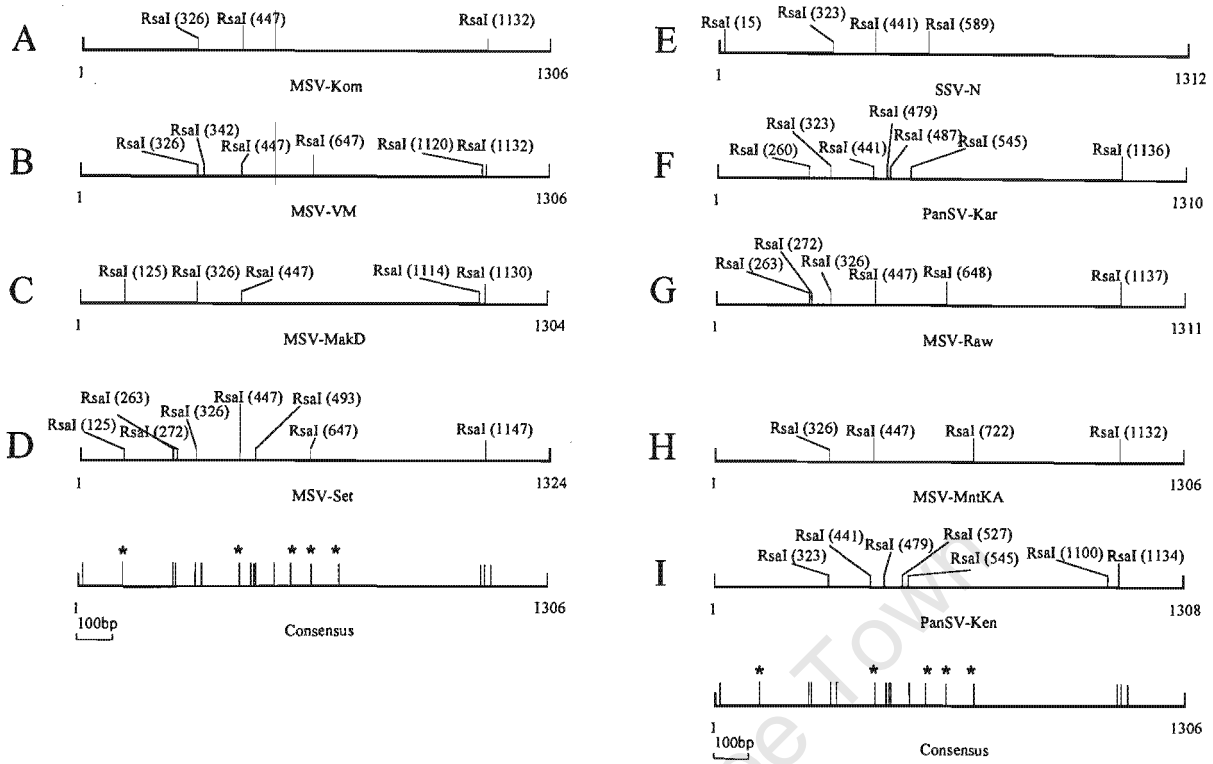


Figure 2.4: Isolates with unique *Rsa*I patterns were identified and consensus maps were generated (DNAMAN 4.0) to represent all possible restriction sites. Sites which were less than 20 bp apart were ignored for the purpose of generating the number of significant sites per individual isolate. The remaining significant sites, indicated by an asterisks in the consensus map, were used to calculate the proportion of sites shared between any two of the above patterns (A - I). The proportion of sites shared between two patterns was used to generate a *Rsa*I distance matrix (d_{RsaI}) (see Table 2.3) (Nei and Li, 1979). Similar maps and distance matrices were generated for each RE used (data not shown).

Table 2.3: Matrix indicating the proportion of sites shared and the mean number of substitutions per site between *Rsa*I patterns[1].

	A	B	C	D	E	F	G	H	I
A	-	0.667	0.667	0.500	0.667	1.000	0.667	0.667	1.000
B	0.102	-	0.500	0.800	0.500	0.667	1.000	0.500	0.667
C	0.102	0.176	-	0.800	0.500	0.667	0.500	0.500	0.667
D	0.176	0.056	0.056	-	0.400	0.500	0.800	0.400	0.500
E	0.102	0.176	0.176	0.234	-	0.667	0.500	0.500	0.667
F	0.000	0.102	0.102	0.176	0.102	-	0.667	0.667	1.000
G	0.102	0.000	0.176	0.056	0.176	0.102	-	0.500	0.667
H	0.102	0.176	0.176	0.234	0.176	0.102	0.176	-	0.667
I	0.000	0.102	0.102	0.176	0.102	0.000	0.102	0.102	-

[1]: The proportion of sites shared (S) between each of the patterns is indicated on the top right hand side of the table. The mean number of substitutions per nucleotide site (d_{RsaI}) is shown in the lower left-hand side. The S value for each pattern compared to itself is 1 and therefore, its d_{RsaI} value equals 0 as indicated by the dash (-).

*Sau*IIIa restriction patterns

The banding patterns observed after agarose gel electrophoresis could only be used to distinguish between the previously noted groups: maize (A), wheat or grass (B) infecting MSVs, and the more distantly related PanSVs, SSV, MSV-Set and MSV-Raw (G, F, E and H, respectively) (see Fig. 2.2 IV). No greater resolution of diversity between viruses belonging to the same groups could be seen, however, MSV-Reu and MSV-MatC shared a unique pattern (D), and MSV-G(B) and MSV-N(B) shared the other rare pattern (C). Apart from MSV-Zi(B), MSV-Cit and MSV-N(C) all MSV-isolates originating from *Zea mays* gave pattern A and all obtained from wheat or grass gave pattern B.

*Hind*III restriction patterns

Of the six different RE patterns, pattern C (see Fig. 2.2V), consisted almost exclusively of wheat or grass infecting isolates excepting MSV-Cit, the sweetcorn isolate. Most maize isolates gave patterns A and B, which differed by a *Hind*III site 3' of the stem-loop region in the former pattern. MSV-Reu and MSV-Raw had the same pattern D, while MSV-U, MSV-N(B), MSV-G(B), PanSV-Kar and SSV-N had no sites therefore were designated pattern E. The sequence data of PanSV-Ken indicates a site in the C1 ORF and has therefore been designated pattern F, unique amongst all isolates examined. Three isolates (MSV-MakD, MSV-Mak and MSV-Zi(B)) shared pattern G, which was only distinguishable from pattern A, by the presence of a 247 bp band due to the lack of a *Hind*III site within the C1 ORF.

*Cfo*I restriction patterns

The majority of the MSVs had one of two possible patterns. The largest group contained representatives of all the maize, wheat and some grass infecting viruses. The smaller group (10 isolates) consisted only of maize infecting MSVs, except MSV-Zi(B) which was isolated from millet (Briddon *et al.*, 1994). The two groups differed only slightly in that the 319 bp band was cleaved at a second *Cfo*I site found in the virion sense promoter region (Figure 2.2 V: Patterns A & B, respectively). The PanSVs, SSV-N, MSV-Set and MSV-Raw had unique banding patterns (patterns E, D, C and F, respectively), that were readily distinguishable from the other MSV patterns.

*Bam*HI restriction patterns

All MSVs, which contain a conserved *Bam*HI restriction site at the start of the MP ORF, have been assigned pattern A. The two PanSVs investigated have a unique site compared to the sequenced RE map generated for PanSV-Ken, designated patterns B and C, respectively. Pattern C is also shared by SSV-N (see Fig. 2.2 VII)

2.4.2: Identification of mixed infections

Identification of the isolates in the mixed infections was limited to those patterns already confirmed in this study. PCR samples containing two or more patterns after RE treatment are shown in Fig. 2.3 and listed in Table 2.2. The relative proportions of the various components of the mixed infections can be deduced from the intensity of the different banding patterns. This can be seen when identifying the components of the Kom(mix) sample. The PCR composite patterns inferred that there were three isolates present in this sample (three different patterns see for the *Hpa*II digests (see Fig. 2.3 A & Table 2.2). One isolate is the agroinfectious MSV-Kom strain (Schnippenkoetter *et al.* unpublished), which originated from this sample, the other was a putative MSV-SA-like isolate (AAABBAA) and the third may have one of a few possible combinations (CXAYBAA, where X represents A, B or D and Y represents A or B). The MSV-Kom type isolate was not the predominant strain as both the *Cfo*I B pattern and the *Hpa*II D pattern, both characteristic of MSV-Kom, are proportionally less evident than the other patterns (see Fig. 2.3 A & C). At least three isolates were identified in the WW (mix) PCR (see Fig. 2.3B) from the three different *Hae*III digests, with the predominant pattern being pattern A. The three viruses could be a typical MSV-VW type isolate (ABCACAB), a MSV-Tas type isolate (ABBACAB) and a MSV-VM type isolate (BAABBAA). The RE analysis of the MntKA(mix) PCR indicates that there are two viruses present, a MSV-MntKA composite pattern (HAAAAAA) and a putative MSV-MatA-type isolate (AAAAAAA). The agroinfectious MSV-MntKA isolate was originally cloned from the MntKA(mix) DNA extraction (D. P. Martin & E. P. Rybicki, unpublished data). Identifying the putative components of the Pas(mix) sample is somewhat complicated by the presence of a MSV-Reu like *Hpa*II pattern E, albeit at lower intensity than the A pattern (see Fig. 2.3 A), and the resultant *Rsa*I pattern could be a combination of A and C like patterns (see Fig. 2.1 lane 14). Pas(mix), MSV-Mak and MSV-MakD, were all obtained from the same region in South Africa (Kwazulu-Natal), and therefore the probability of the *Rsa*I C

pattern being present is strong. MSV-G(A) sample yielded one pattern for all RE, except *Rsa*I (see Fig. 2.1 lane 17). One of the viruses present within this sample could be a MSV-MatB like virus (BCAAAAA) and the other possible combination of (ACAAAAA) does not fit any composite pattern (see Table 2.2), and could represent an undiscovered virus type.

2.4.3: Sequence and phylogenetic analysis

Sequences were deposited in the Genbank under the following accession numbers: MSV-Tas (AJ012636), MSV-VW (AJ012638), MSV-VM (AJ012637) and PanSV-Kar (X60168). Sequence data obtained from the various viruses that reacted positively in the PCR and gave unique RE patterns, were aligned using DNAMAN 4.0 and an identity matrix was generated (Table 2.3). The following groupings of viruses can be determined from the matrix: the severe maize-infecting isolates all sharing between 95 – 100 % identity; two highly similar wheat-infecting isolates (MSV-Tas and MSV-VW), which are slightly more closely related to the maize types than they are to the sequenced grass isolates; MSV-Set and MSV-Raw; and finally the other members of the African streak virus group. The PanSVs and SSV-N are as divergent from each other as they are from the larger MSV group, consistent with previous observations (Schnippenkoetter, 1998; Rybicki *et al.*, 1998; Briddon *et al.*, 1992; Rybicki and Hughes, 1990). The groupings, based on the partial C1 ORF (RepA) amino acid percentage pairwise similarity matrix generated by DNAMAN 4.0, remain identical to those generated from the full DNA sequence data (Table 2.4).

The phylogenetic tree generated from sequence distance data indicates that the maize-type MSVs cluster together, and that there are two groups of viruses not found in maize, one of which infects grasses only and the other of which infects wheat and grasses (see Fig. 2.5). The last two groups are more distantly related to the maize type MSVs than is MSV-Reu, which until recently was the most distantly related MSV known (Peterschmitt *et al.*, 1996). Although the wheat and grass subgroups have been observed before, the previous evidence was based on a +/- 250 bp sequence from the C2 ORF, rather than on approximately half the genome as is presented here (Schnippenkoetter, 1998; Rybicki *et al.*, 1998; Hughes *et al.*, 1992). The PanSVs are more closely related to each other than to any of the other viruses, sharing 88.5 % nucleotide sequence identity over half the genome (see Table 2.4). Together with SSV-N, the PanSVs are basal to the MSV-related African streak viruses (Schnippenkoetter, 1998; Hughes *et al.*, 1992). The phylogenetic tree drawn up using the

partial C1 ORF amino acid sequence was virtually identical, excepting that lower bootstrap scores were achieved for branches within the large maize type MSV group, therefore less can be inferred about their relatedness (data not shown). The only significant grouping (bootstrap value of 801) was that of MSV-MntKA and MSV-K being strongly related. The same can be seen in the tree generated from the +/- 1300 bp sequence data (Fig. 2.5A).

Phylogenies based on RE mapping data are known to be less accurate than those based on sequence data, especially when more distantly related isolates are compared (Hughes *et al.*, 1992; Swofford and Olsen, 1990). We have overcome some of the experimental error using sequence data to firstly confirm the patterns evidenced by gel electrophoresis of the RE digests, and secondly, to construct the maps from which the distance matrix was generated. However the method is still generally less accurate especially when very few restriction sites are shared between the isolates under investigation (Hughes *et al.*, 1992; Nei and Li, 1979). Given these caveats, the tree generated from the distance matrix based on RE mapping data (Table 2.5) closely follows the overall topology of that of the sequence generated tree (see Fig. 2.5 A & B), with the following exceptions: the addition of another MSV subtype (MSV-U, MSV-G(B) & MSV-N(B)) and the separation of the two grass isolates, MSV-Raw and MSV-Set. The MSV-U subtype was previously noted by Briddon *et al.* (1994) based on the CP sequence data. The wheat/grass subtype of MSVs consists of seven isolates, with the exception of MSV-Cit, a sweetcorn isolate, being the only one found in maize.

Due to the low number of shared sites between MSV-Set and MSV-Raw they appear more distantly related when using RE mapping data for phylogenetic analysis (71 % versus 89 % obtained from sequence data). Conversely the PanSVs and SSV-N share greater identity with the MSV isolates when distance is measured using the RE map data than the sequence data (60 – 70 % versus 50 – 60 %) (see Tables 2.4 & 2.5). The linear regression analysis comparing the two sets of generated distance data indicates the close correlation between the highly related isolates with decreasing correlation at the higher divergence (see Fig. 2.6). The best fit to these points is a curve, becoming asymptotic at about 40 % sequence divergence: this is in accord with other work, which has shown that distances greater then 30 % are very inaccurately estimated by this technique (see Hughes *et al.*, 1992, and references therein).

Table 2.4: Pairwise percentage sequence identity between pairs of African streak mastreviruses[1]

Isolate name		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
SV-VM	1	100	98.9	98.9	99.2	98.5	98.5	99.2	98.1	96.2	94.6	97.7	86.8	85.7	77.4	75.9	56.3	57.8	60.6
MSV-SA	2	97.9	100	99.2	99.6	98.5	98.5	99.6	98.5	95.8	94.6	97.7	87.2	86.0	78.2	76.7	56.6	59.0	61.0
MSV-Kom	3	97.9	99.1	100	99.6	98.5	98.5	99.6	98.5	95.8	94.6	97.7	86.8	85.7	77.8	76.3	57.0	58.2	60.6
MSV-N	4	97.8	98.2	98.3	100	98.9	98.9	100	98.9	96.2	95.0	98.1	87.2	86.0	78.2	76.7	57.0	58.6	61.0
MSV-K	5	97.6	98.0	98.0	98.3	100	98.9	98.0	97.7	95.8	94.6	97.7	86.8	85.7	78.2	76.7	56.3	57.8	61.0
MSV-MntKA	6	97.4	97.9	97.8	97.9	98.8	100	98.9	97.7	96.2	95.0	97.7	86.8	85.7	78.2	76.7	56.6	58.2	61.4
MSV-MatA	7	97.1	97.5	97.6	98.2	98.1	97.8	100	98.9	96.2	95.0	98.1	87.2	86.0	78.2	76.7	57.0	58.6	61.0
MSV-MatB	8	96.8	97.0	97.0	97.8	97.6	97.2	98.5	100	95.8	96.2	98.5	86.8	86.0	78.6	77.0	57.4	59.0	61.8
MSV-MatC	9	95.2	95.3	95.2	95.6	96.1	95.9	95.9	96.0	100	95.0	96.2	87.2	86.0	79.0	76.3	57.0	58.2	61.8
MSV-MakD	10	95.0	95.3	95.2	95.7	96.1	95.8	96.5	97.2	97.7	100	96.2	86.0	85.3	76.7	74.3	54.7	56.3	61.0
MSV-Reu	11	95.5	95.3	95.2	95.2	95.4	95.5	95.2	94.9	94.7	94.9	100	88.0	87.2	78.6	76.3	56.3	57.8	61.8
MSV-Tas	12	87.5	87.2	87.0	87.0	87.2	87.3	87.3	86.7	87.2	87.1	87.9	100	96.2	81.2	78.5	54.7	57.4	61.1
MSV-VW	13	87.9	87.3	87.2	87.1	87.4	87.3	87.5	87.0	87.2	87.2	88.2	97.5	100	81.9	80.4	55.0	57.8	61.5
MSV-Raw	14	79.9	80.0	79.9	79.8	79.6	80.1	80.1	79.9	79.6	79.3	79.6	81.4	82.1	100	87.7	57.1	58.7	59.8
MSV-Set	15	76.4	76.4	76.5	76.4	76.3	76.8	77.1	76.6	75.9	75.9	76.8	78.1	79.2	89.3	100	58.3	59.1	62.1
Pansv-Ken	16	53.7	53.8	53.8	53.6	54.1	54.1	53.8	54.0	54.7	54.0	54.4	54.1	54.3	53.1	52.9	100	86.5	52.7
PanSV-Kar	17	54.1	54.6	54.2	54.4	54.1	54.1	53.8	54.0	54.0	53.4	54.4	53.2	53.3	53.7	52.1	88.5	100	55.4
SSV-N	18	57.5	57.0	57.0	57.1	57.2	57.2	57.5	57.9	57.3	57.7	58.8	58.4	58.9	57.5	57.2	58.6	59.4	100

[1]: Lower left-hand side of the table indicates the nucleotide sequence identity matrix based on the 1300 bp sequence amplified by PCR. Upper right hand side of the table indicates the percentage amino acid sequence identity of the RepA protein, as translated from the truncated C1 ORF generated during PCR. DNAMAN 4.0 was used to generate the matrix in both cases. The lightly shaded area demarcates the group of highly related maize-infecting MSV isolates and the darkly shaded area indicates the relationship the distinct strains of PanSV and SSV have with the other members of the African streak viruses.

Table 2.5: Distance matrix based on restriction map data[1]

		MSV-SA	MSV-N	MSV-Stell	MSV-Zi(B)	MSV-MatA	MSV-VM	MSV-MtKA	MSV-Kom	MSV-Koe	MSV-MatB	MSV-MtKB	MSV-SW	MSV-U	MSV-G(B)	MSV-MakD	MSV-Mak	MSV-MatC	MSV-Reu	MSV-Tas	MSV-VW	MSV-Raw	MSV-Set	SSV-N	PanSV-Kar	PanSV Ken
MSV-		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
SA	1		98.5	99.5	99.5	99.5	98.5	96.6	96.7	98.0	96.6	96.6	97.1	94.1	91.6	95.1	93.6	95.1	95.5	85.3	84.3	73.0	68.7	62.7	63.1	69.8
N	2	0.0146		98.0	98.0	99.5	96.6	98.0	95.2	96.6	98	98.0	98.5	95.6	93.1	97.2	95.1	96.6	96.9	86.8	85.7	72.4	70.2	68.4	61.1	70.0
Stell	3	0.0051	0.0197		99.5	98.5	98.0	97.1	96.1	98.5	97.1	97.1	96.6	92.5	90.0	97.1	94.2	95.6	94.6	84.6	83.6	77.6	67.1	61.1	61.5	69.8
ZiB	4	0.0051	0.0197	0.0051		98.5	98.0	97.1	96.1	98.5	97.1	97.1	96.6	94.0	90.0	95.6	94.2	95.6	96.6	84.6	83.6	77.6	79.7	61.1	61.5	62.7
MatA	5	0.0051	0.0051	0.0146	0.0146		96.6	98.5	94.7	97.1	97.1	98.5	98.0	94.0	91.5	97.1	94.2	97.1	98.0	86.1	85.0	73.5	68.6	62.7	62.7	70.9
VM	6	0.0146	0.0343	0.0197	0.0197	0.0343		95.5	95.2	99.5	98.0	98.0	98.5	92.7	90.2	94.1	92.6	93.6	94.0	83.9	82.8	78.0	68.8	60.1	67.2	69.0
MntKA	7	0.0343	0.0197	0.0292	0.0292	0.0146	0.0449		93.2	96.0	97.5	97.5	97.0	92.5	90.0	96.0	94.6	95.6	96.6	85.3	83.6	74.9	67.8	61.6	61.2	69.5
Kom	8	0.0334	0.048	0.0385	0.0385	0.0531	0.0480	0.0677		94.7	93.2	93.2	95.2	90.8	88.3	91.0	89.5	91.0	91.4	81.2	80.2	79.9	67.9	63.8	66.0	72.2
Koe	9	0.0197	0.0343	0.0146	0.0146	0.0292	0.0051	0.0397	0.0531		98.5	98.5	98.0	91.1	88.6	94.6	94.2	94.2	95.1	83.2	82.1	76.3	68.1	61.1	61.5	69.8
MatB	10	0.0343	0.0197	0.0292	0.0292	0.0292	0.0197	0.0251	0.0677	0.0146		100	99.5	92.5	90.0	96.0	94.6	95.6	96.6	84.6	83.6	73.4	70.3	62.7	59.6	67.9
MntKB	11	0.0343	0.0197	0.0292	0.0292	0.0146	0.0197	0.0251	0.0677	0.0146	0.0000		99.5	92.5	95.5	96.0	94.6	95.6	96.6	84.6	83.6	74.9	70.3	61.6	61.2	70.9
SW	12	0.0292	0.0146	0.0343	0.0343	0.0197	0.0146	0.0302	0.0480	0.0197	0.0051	0.0051		94.1	91.6	95.5	94.1	95.1	95.5	85.3	84.3	73.9	71.9	68.3	61.2	68.6
U	13	0.0586	0.0440	0.0746	0.0599	0.0599	0.0732	0.0746	0.0920	0.0892	0.0746	0.0746	0.0586		97.5	91.1	89.6	91.1	95.6	89.1	88.0	68.4	79.7	68.7	68.6	70.0
N(B)	14	0.0837	0.0691	0.0997	0.0997	0.0851	0.0983	0.0997	0.1171	0.1143	0.0997	0.0449	0.0837	0.0251		88.6	87.1	94.1	94.4	91.6	90.6	73.5	68.4	64.3	64.3	71.6
MakD	15	0.0489	0.0277	0.0292	0.0438	0.0292	0.0595	0.0397	0.0899	0.0543	0.0397	0.0397	0.0449	0.0892	0.1143		98.5	97.1	98.0	86.1	85.0	70.2	68.8	61.6	59.6	67.9
Mak	16	0.0635	0.0489	0.0584	0.0584	0.0584	0.0741	0.0543	0.1045	0.0584	0.0543	0.0543	0.0595	0.1038	0.1289	0.0146		98.5	96.6	87.5	86.7	73.9	67.7	65.3	65.3	67.2
MatC	17	0.0489	0.0343	0.0438	0.0438	0.0292	0.0635	0.0438	0.0899	0.0584	0.0438	0.0438	0.0489	0.0892	0.0595	0.0292	0.0146		98.0	86.3	85.5	80.3	64.1	66.3	65.3	67.2
Reu	18	0.0452	0.0306	0.0541	0.0343	0.0197	0.0598	0.0343	0.0862	0.0489	0.0343	0.0343	0.0452	0.0438	0.0557	0.0197	0.0343	0.0197		85.5	84.5	77.2	68.3	64.6	67.5	70.0
Tas	19	0.1468	0.1322	0.1539	0.1539	0.1392	0.1614	0.1468	0.1878	0.1685	0.1539	0.1539	0.1468	0.1095	0.0835	0.1392	0.1246	0.1370	0.1445		99.2	75.6	67.2	77.8	68.3	67.5
VW	20	0.1573	0.1427	0.1644	0.1644	0.1498	0.1719	0.1644	0.1983	0.179	0.1644	0.1644	0.1573	0.1200	0.0940	0.1498	0.1326	0.1450	0.1550	0.0080		70.4	71.6	73.0	67.5	63.1
Raw	21	0.2704	0.2761	0.2236	0.2236	0.2652	0.2199	0.2506	0.2011	0.2373	0.2662	0.2506	0.2615	0.3157	0.2652	0.298	0.2615	0.1969	0.2282	0.2436	0.2957		70.9	76.5	74.0	74.9
Set	22	0.3133	0.2977	0.3293	0.2026	0.3136	0.3121	0.3219	0.3209	0.3189	0.2965	0.2965	0.2805	0.2026	0.3160	0.3121	0.3234	0.3589	0.3169	0.3282	0.2835	0.2913		70.7	66.6	65.2
SSV-N	23	0.3729	0.3163	0.3888	0.3888	0.3732	0.3994	0.3837	0.3618	0.3888	0.3732	0.3837	0.3167	0.3133	0.3572	0.3837	0.3473	0.3367	0.3537	0.2222	0.2702	0.2353	0.2931		69.2	71.5
PanSV-Kar	24	0.3688	0.3891	0.3848	0.3848	0.3734	0.3281	0.3881	0.3398	0.3848	0.4037	0.3881	0.3877	0.3135	0.3575	0.4037	0.3475	0.3475	0.325	0.3167	0.325	0.2595	0.3343	0.3084		86.9
PanSV-Ken	25	0.3020	0.2996	0.3020	0.3734	0.2907	0.3099	0.3053	0.2776	0.302	0.3209	0.2907	0.3142	0.3002	0.284	0.3209	0.3285	0.3285	0.2996	0.3254	0.3688	0.2511	0.3475	0.2848	0.1305	

[1]: Lower left hand side of the table indicates the distance estimated between isolates, which is based on the proportion of shared restriction sites transformed into the average number of substitutions per nucleotide site (F) (see Methods for explanation). Upper right hand side of the table indicates the calculated percentage sequence identity between pairs of isolates i.e. (1300 - (F x 1300))/1300. Shading as for Table 2.4.

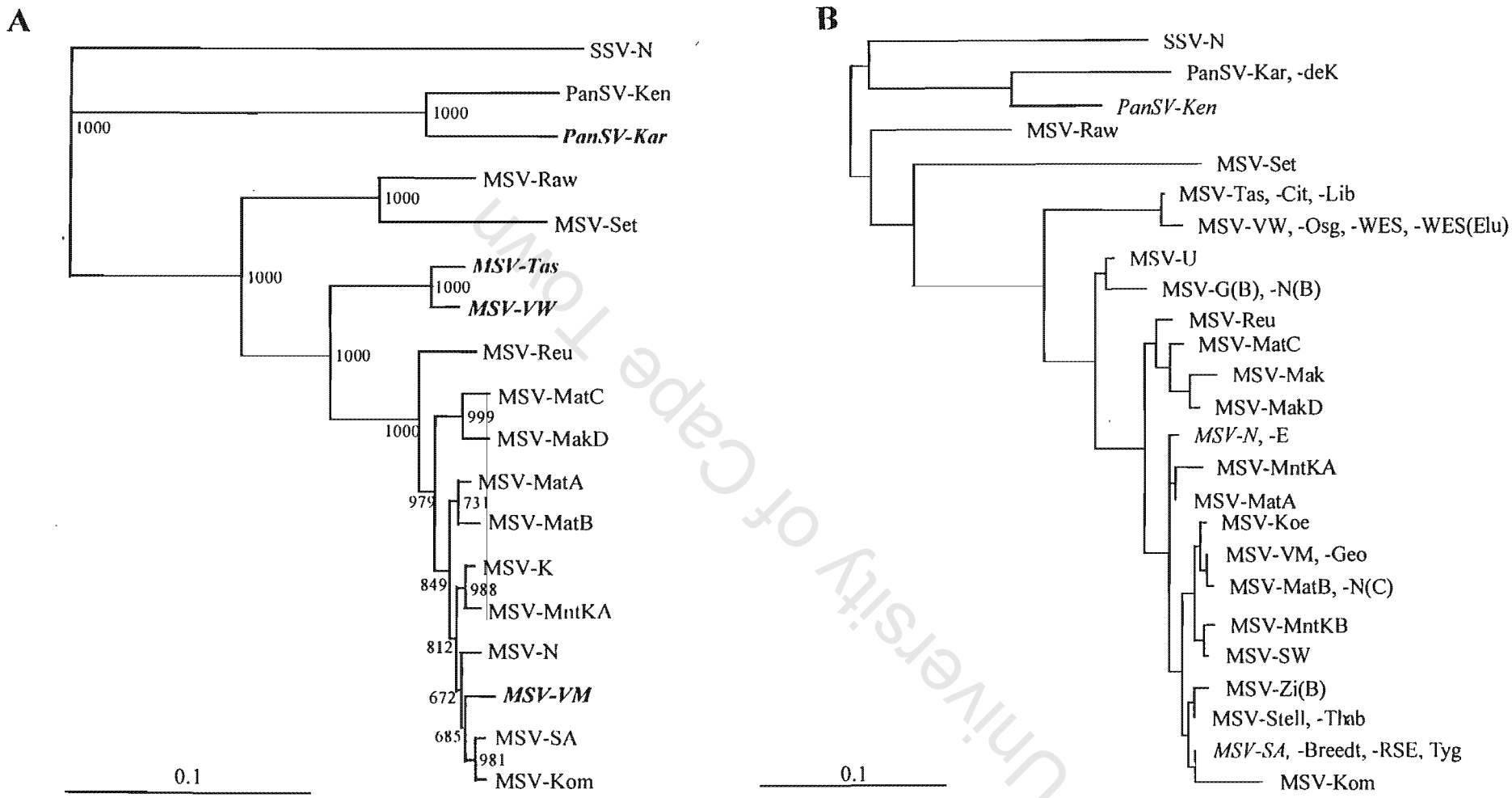


Figure 2.5: Relationship dendrograms generated from sequence data (A) and from the RE map distance matrix (B). A) DNA sequences were aligned and the output multiple sequence file was subjected to phylogenetic analysis with a 1000 bootstrap trials (DNAMAN 4.0), and rooted with SSV-N as an outgroup using TREEVIEW (Page, 1997). Names in bold italics indicate sequences to which I contributed data. B) The dendrogram was generated using NJTREE version 2 (Saitou & Nei, 1987), based on the distance matrix generated using the transformation equations of Nei and Li (1979). The tree was rooted with SSV-N as the outgroup using NJDRAW version 1. Where sequence data only was used to predict the RE maps, the virus names are indicated in italics. Bars indicate 10 % divergence. All vertical distances are arbitrary, while horizontal distances are proportional to actual or estimated sequence divergence. Bootstrap values are indicated in (A).

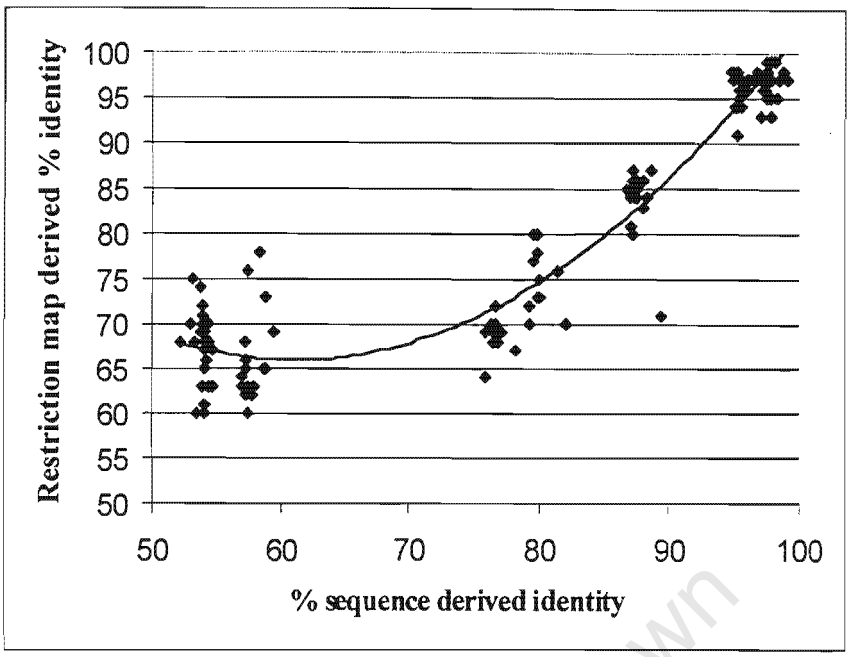


Figure 2.6: Regression analysis of the percentage identity between two isolates obtained from sequence data using DNAMAN 4.0, compared with distance data generated from restriction map analysis and transformation equations of Nei & Li (1979) (see Table 2.5). A value of $R^2 = 0.9245$ was achieved for the curve, whereas if drawn as a straight line $R^2 = 0.8124$ (data not shown) (Microsoft® Excel 97).

2.5: DISCUSSION

Mastreviruses, and in particular the MSVs, are known to cause significant damage to the yearly yield of cereal grains in Africa (Thottappilly *et al.*, 1993). Natural infections are often used as a screening method for the evaluation of resistant cultivars of maize (Pers. Comm. J. B. J. van Rensburg), therefore, it would be of considerable value to be able to distinguish in natural infections between viruses that are mild and those that are known to be more severe. In this chapter I have presented a comprehensive detection and identification method for the typing of African mastreviruses, and an effective means of differentiating between strains that have been known to infect wheat, maize, grasses and sugarcane.

Two major groupings of MSV isolates were identified based on the distance matrix generated by transformation of the RE mapping data. These two groups, the larger consisting mostly of maize infecting MSVs and the smaller of predominantly wheat and grass isolates, were verified based on sequence data of the same region (Tables 2.4 and

2.5 and Fig. 2.5A & B). These observations are similar to and expand on those made based on a 250 bp region of the C2 ORF (Rybicki *et al.*, 1998; Hughes *et al.*, 1992). Two subtypes of wheat/grass infecting isolates (see Table 2.2), a typical MSV-Tas type and an MSV-VW type, were distinguishable using the composite RE patterns, even though these isolates share 97.5 % sequence homology in the 1.3 kb fragment (see Fig. 2.2 III patterns B & C and Fig. 2.5). The majority of isolates analysed were maize infecting viruses, which could be differentiated effectively by RE analysis; even MSV-Kom and MSV-SA which share 99 % sequence homology in the PCR amplified region (Table 2.4) could be readily distinguished (*Hpa*II patterns D and A, respectively. See Fig. 2.2 II).

A common sequence was identified within the LIR of almost all MSVs: the *Cfo*I recognition site (GCGC) forms part of the putative replication specificity determinants of MSV replication, i.e. iterons (GCGCTTC), found in the left-hand side of the LIR and within the stem-loop (Argüello-Astorga *et al.*, 1994a; Argüello-Astorga *et al.*, 1994b). These iterons seems to be conserved amongst almost all the MSV isolates examined, which supports the hypothesis that replication functions may be shared amongst these isolates. Further work to test this hypothesis is presented in later chapters of this thesis. Interestingly MSV-Set, previously named SetSV (Hughes *et al.*, 1992), does not share the same iteron sequence (Schnippenkoetter, 1998), while MSV-Raw shares six out of eight nucleotides with the typical MSV iteron sequence despite being most closely related to MSV-Set (see Chapter 4 & 5). Further differences have been confirmed by whole genome sequence data and host range studies on MSV-Set (Schnippenkoetter, 1998) and the same will be undertaken for MSV-Raw (D. P. Martin, Pers. Comm).

Previously, based on restriction map analysis, MSV-Set's closest relative was MSV-WES, although it was only slightly more closely related than were the other MSVs (Hughes *et al.*, 1992). Our results consistently place them into separate subgroups, which is also confirmed by the recent data obtained by Rybicki *et al.* (1998) for a +/- 250 bp sequence from the C2 ORF, and those based on whole genomic sequence (Schnippenkoetter, 1998). The detection of a second grass isolate, MSV-Raw, whose closest relative is MSV-Set, indicates that there may be more of these isolates, but due to their mild symptoms in grasses may go undetected. These two viruses share 89.3 % homology (see Table 2.4), which is far greater than any previously noted between MSV-Set and other MSVs (+/- 78 % similarity with the MSVs) (Schnippenkoetter, 1998).

The isolates examined in this work were collected from across the African continent and included one isolate from Reunion. Interestingly, the isolates most similar to MSV-SA based on the composite restriction patterns, are common only in South Africa where they are fairly widely distributed. It must be noted that only a small number of the isolates analysed are northern African viruses; however, a recent study using these same degenerate primers performed on 18 MSV isolates from Kenya did not detect the presence of MSV-SA like viruses (R. Bilharz, B. Odhiambo, D. P. Martin and E. P. Rybicki, unpublished data) although the wheat-type MSVs were represented. It would be of significant epidemiological interest to monitor the natural spread, or lack thereof, of the MSV-SA like viruses on the continent of Africa over time. It also proved possible to monitor the changes a viral population undergoes in specific areas over time, albeit to a very limited extent. The first example of this was MSV-Mak and MSV-MakD, which were collected 6 years apart from the same region but are highly related (see Fig. 2.5 B & Table 2.5). Furthermore, the grass sample Pas(mix) also from the same region may have been a host or reservoir for a MSV-Mak–type virus at some time. Secondly, the same region in Matabeleland, Zimbabwe had three successions of resistance breaking strains of MSV over a similar time period (K.E. Palmer, Pers. Comm). The viruses identified as being responsible were three distinct isolates MSV-MatA, B and C, with the first two sharing greater homology (see Table 2.4).

The method presented here for the detection and typing of MSV isolates provides a means of identifying the more pathogenic isolates from wheat and grasses capable of infecting maize, especially if these hosts are acting as reservoirs (Pinner *et al.*, 1988; Rose, 1978). This method was also able to identify mixtures of isolates, as previously demonstrated by Rojas *et al.* (1993). The agroinfectious MSV-Kom previously cloned from a mixed population was readily identified in the Kom(mix) sample (F. L. Hughes, unpublished data; Schnippenkoetter, 1998). MSV-Kom was one of potentially three different viruses in a mixture shown by leafhopper transmissions to contain mild, moderate and severe isolates of MSV (M. B. von Wechmar, Pers. Comm.). Mixed viral populations were further noted in four other samples, WW(mix), Pas(mix), MntKA(mix) and MSV-G(A), with both maize and wheat types simultaneously present in WW(mix).

The identification of other members of the African streak virus group was possible using RFLP analysis of the same region amplified by the degenerate primers. The two South

African PanSVs differ from the sequenced PanSV-Ken with respect to 5 out of 7 restriction enzymes. Hughes *et al.* (1992) showed by restriction mapping of RF DNA that PanSVs are strains of the same virus but distinct from each other. Here we show two PanSVs from different regions of South Africa have identical restriction patterns. Although only one SSV isolate was included in this study the existence of others is known: SSV-M (Rybicki and Hughes, 1990), SSV-E (L. Bigarre, unpublished data. GenBank accession number AF037752) and SSV-Mill (Briddon *et al.*, 1996b). Based on the available sequence of SSV-E, the PCR primers required further degeneracies for amplification of SSV-E, one in primer 1770 – 1792 and three in primer 215 – 234 (data not shown).

I have presented here a method for the detection of diversity within a closely related population of viruses, which simultaneously includes those that are quite distinct (SSV and PanSV). Unlike the Briddon *et al.* (1994) study most of our PCR products were amplified from the DNA extracted from the original host plant(s) (see Table 2.1), meaning that no selection or "sieving" of isolates will have occurred. Virus mixtures present in plants could also be identified. These mixtures would not have been detected had the viruses, through a series of leafhopper transmissions, been maintained in non-permissive host plants. For example, had the WW(mix) viruses been propagated through maize cultivars the two putative wheat-type viruses (2.4.2), which are weakly pathogenic in maize (see Chapter 3), may have eventually been lost due to the greater probability of the more pathogenic maize-type virus being transmitted.

Although the use of this PCR product and subsequent restriction endonuclease analysis can be used to differentiate between virus isolates, we do not propose that these groupings are definitive, but only that they reflect patterns of relatedness. More conserved regions of the genome, i.e. the CP or C1/C2, may give a lower number of virus groups; however, the presence of the "hypervariable" LIR and V1 sequences in this study almost certainly allowed finer differentiation by RFLPs than would have been possible with the above-mentioned regions.

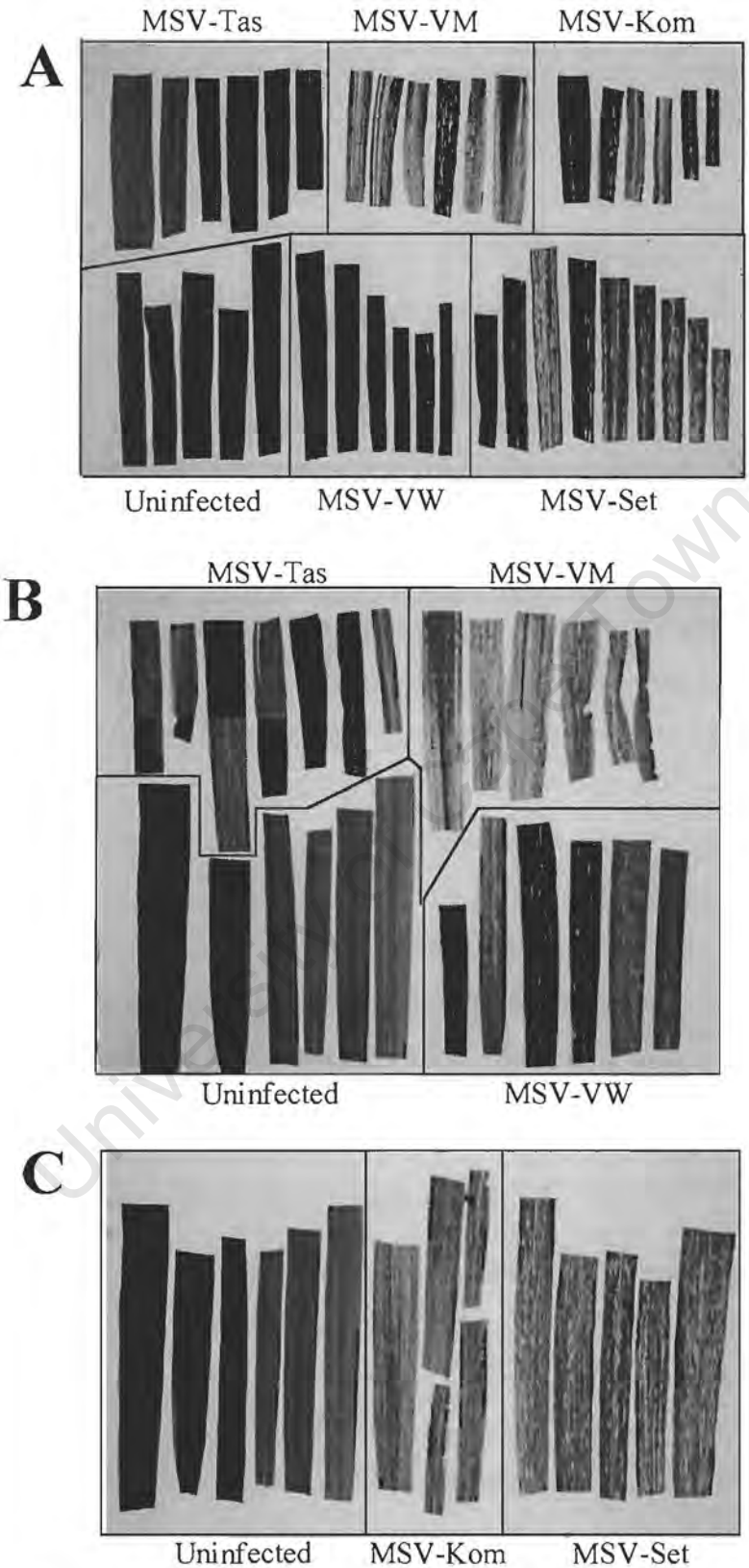


Figure 3.3: Streak symptoms on sweetcorn cv. Jubilee. Three day old seedlings were agroinfected with partial dimers of MSV-Tas, MSV-VM, MSV-VW, MSV-Kom and MSV-Set (pTas401, pVM601, pVW303, pKom603 and pSet107, respectively). Symptoms on the third leaf were assessed 15 days after agroinfection (A) and on the fourth leaf 22 days after agroinfection (B & C). The second quarter of each leaf is presented for comparative analysis of symptoms.

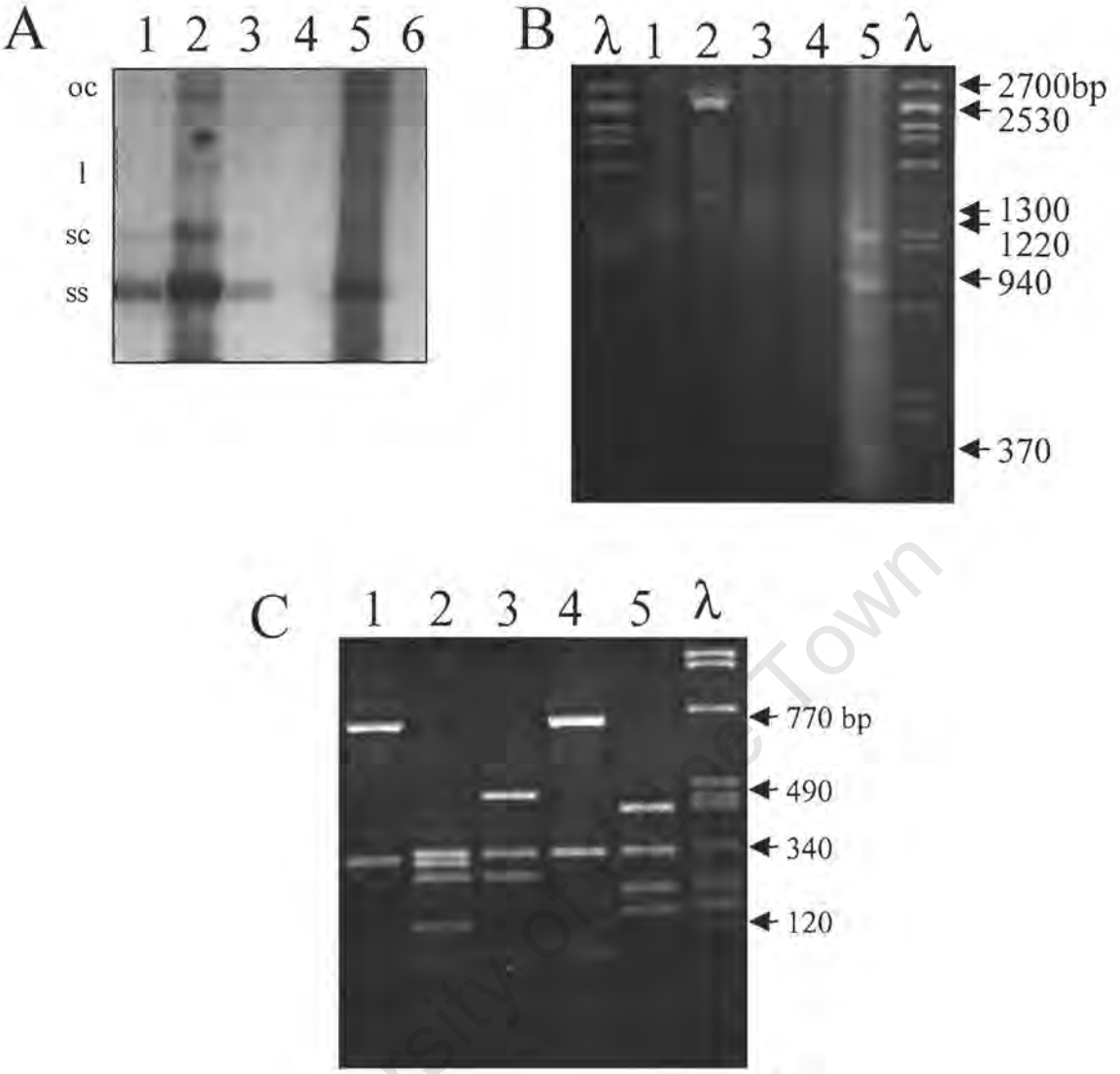


Figure 3.4: Analysis of viral DNA extracted from agroinfected plants by Southern blot using a MSV-Kom Dig-labeled probe (A), *Sac*I digestion of genomic DNA (B), and PCR and *Hae*III digestion (C). A & B) lanes 1, MSV-Tas; lane 2, MSV-VM; lane3, MSV-VW; lane 4, MSV-Set and lane 5, MSV-Kom. A) Lane 6 contained uninfected sweetcorn genomic DNA. C) Lane 1, MSV-VM; lane 2, MSV-VW; lane 3, MSV-Tas; lane 4, MSV-Kom; and lane 5, MSV-Set. DNA was electrophoresed through 1.5 % agarose gels, stained with ethidium bromide, in 0.5x TBE. Lambda *Pst*I digested DNA was used as a molecular weight marker (λ). The various replicative forms of viral genomic DNA are indicated as follows: open circular (oc), linear (l), supercoiled (sc) and single stranded (ss). Sizes of bands are indicated as number of base pairs (bp).

Chapter 3

Characterisation of three MSVs

3.1: SUMMARY 74
3.2: INTRODUCTION 75
3.3: METHODS AND MATERIALS 78
3.3.1: Virus origins 78
3.3.2: Agroinfectious constructs 79
3.3.3: Agroinfection of maize plants 80
3.3.4: Detection of MSV isolates in agroinfected plants 81
3.3.5: Leafhopper transmissions 82
3.3.6: Sequencing of MSV-VW, MSV-VM and MSV-Tas 83
3.4: RESULTS 84
3.4.1: Agroinfection of MSV isolates 84
3.4.2: Detection of viral DNA in agroinfected plants 88
3.4.3: Leafhopper transmissions 88
3.4.4: Sequence analysis 93
3.4.5: Phylogenetic analysis 109
3.5: DISCUSSION 115

3.1: SUMMARY

Full-length agroinfectious clones of two maize streak viruses (MSVs) originally isolated from wheat, MSV-Tas and MSV-VW, were characterised biologically by leafhopper transmissions to a range of host plants. The complete genome sequences of both viruses as well as that of a third maize-infecting isolate, MSV-VM, were obtained. Leafhopper transmissions of MSV-Tas and MSV-VW indicated differences in their capabilities to infect maize cultivars, while both infected a variety of wheat and barley cultivars. The characteristic mastrevirus genomic arrangement of open reading frames, with bi-directional transcription of the ORFs originating from the long intergenic region and terminating in the small intergenic region, was observed. Post-transcriptional modification signals for introns, in both the virion and complementary sense ORFs, were maintained. Phylogenetic analysis of the three viruses indicates that MSV-Tas and MSV-VW were more closely related to each other than to any of the other sequenced maize isolates, while MSV-VM was closely related to MSV-Kom, a typical maize isolate. These relationships were maintained when comparing individual predicted protein amino acid sequences, except for the movement protein where MSV-VM grouped with the two wheat isolates. Sequence analysis of the LIR identified the conserved MSV stem-like structure and invariant geminivirus loop sequence. The presence of the putative replication associated protein binding sites were conserved between all three isolates, suggesting similar replication strategies; however the Rep protein sequence of the two wheat isolates shares only 88 % identity with the maize isolate. Regardless of this, the Rep motifs required for rolling circle replication are conserved amongst all the sequenced MSVs. This is the first report of the characterisation of wheat infecting MSV isolates, which are distinct from the larger group of characterised maize infecting isolates.

3.2: INTRODUCTION

Maize streak virus (MSV), which is endemic on the continent of Africa, is a member of the "African streak virus" subgroup, of the genus *Mastrevirus* of the family *Geminiviridae* (Mayo and Pringle, 1998; Briddon and Markham, 1995; Hughes *et al.*, 1992; Soto *et al.*, 1982). Members of this group include the MSVs, sugarcane streak virus (SSV) and panicum streak virus (PanSV), all sharing their own genus of leafhopper vector, *Cicadulina mbila* (Briddon and Markham, 1995; Briddon *et al.*, 1994; Hughes *et al.*, 1993; Hughes *et al.*, 1992; Hughes *et al.*, 1991). Other mastreviruses infecting monocots are digitaria streak virus (DSV), wheat dwarf virus (WDV), chloris striate mosaic virus (CSMV) and miscanthus streak virus (MiSV), each having their own leafhopper genus vector. Two dicot infecting mastreviruses have recently been identified, tobacco yellow dwarf virus (TYDV) and bean yellow dwarf virus (BeYDV) (see Table 3.1). All mastreviruses are similar in terms of their genomic organisation and requirements for both their coat proteins (CP) and movement proteins (MP) for movement (Boulton *et al.*, 1993; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989; Woolston *et al.*, 1989). Furthermore, the absolute requirement of the viral replication associated protein (Rep), working in conjunction with the host's DNA replication machinery, is essential for replication of the geminivirus ssDNA genome via a rolling circle mechanism (Stenger *et al.*, 1991; Matzeit *et al.*, 1991; Schalk *et al.*, 1989; Elmer *et al.*, 1988).

Within the African streak virus group of isolates exists a large variety of strains of MSV (Storey, 1928). Historically, all streak viruses isolated from Gramineae which gave mild symptoms in maize were designated as "host-adapted" strains of MSV (Pinner and Markham, 1990; Bock and Bailey, 1989; Bock *et al.*, 1974; Storey and McClean, 1930). Differences in the pathogenicity of MSV isolates in maize was noted, with some being highly pathogenic, such as the MSV A form (an example of which is MSV-SA) and the milder MSV B form (Lazarowitz, 1988; McClean, 1947). The A and B forms could be distinguished based upon their different chlorotic streak patterns in susceptible hosts; however their host ranges were similar, complicating identification (McClean, 1947). Furthermore, streak isolates obtained from the wild grass species *Sporobolus*, *Eleusine*, *Paspalum*, and *Digitaria* had variable degrees of success at infecting maize cultivars (McClean, 1947; Storey and McClean, 1930). More recently serological assays and other diagnostic DNA molecular techniques identified these host-adapted isolates as distinct

mastreviruses, such as the PanSVs and SSVs (Hughes *et al.*, 1993; Hughes *et al.*, 1992; Mesfin *et al.*, 1992; Peterschmitt *et al.*, 1991).

Whether or not grasses could be acting as reservoirs for mastreviruses, where the viruses are mostly host-adapted and display mild disease, remains to be established definitively (Rybicki *et al.*, 1998; Mesfin *et al.*, 1992; Peterschmitt *et al.*, 1991; Rose, 1978). Although the existence of a distinct grass and wheat infecting subgroup of MSV isolates has been shown (Chapter 2; Rybicki *et al.*, 1998; Hughes *et al.*, 1992), no studies other than the early research by Storey & McClean (1930) and McClean (1947) has been undertaken to investigate the natural progression of these isolates in the field into different seasonal hosts. Most isolates that were obtained from non-maize hosts were closely related, more so than they are to the maize-infecting isolates (Chapter 2; Rybicki *et al.*, 1998).

A serological survey by Mesfin *et al.* (1992) of 24 virus isolates from maize and grasses in Nigeria found that 18 were severe MSV types, and the rest were PanSVs or sugarcane isolates, presumably SSV. Of the viruses tested, those which were isolated from *Axonopus compressus*, *Brachiaria lata*, and *Setaria barbata* were readily transmitted to maize: therefore, these grasses could act as reservoirs for maize-infecting MSV isolates (Mesfin *et al.*, 1992). Wild sorghum (*Sorghum verticilliflorum*), *Coix lachryma-jobi*, *Cenchrus echinatus*, and *B. reptens* are reported to be responsible for harbouring maize infecting MSV isolates in Mauritius. The latter two grasses are widely distributed annuals and may be the carry-over hosts between the perennial reservoirs (sorghum and *C. lachryma-jobi*) and maize (Autrey and Ricaud, 1983). Similarly, a serologically based study in Reunion identified a large collection of maize-infecting viruses constituting a serotype, originating from grasses and maize, and three serotypes from sugarcane (Peterschmitt *et al.*, 1991). The maize and *Brachiaria* sp. isolates were readily transmitted to maize, whereas other grass isolates were transmitted to maize at low frequency and caused milder symptoms. Although severe maize infecting isolates may use grasses as hosts, most of the isolates from these plants seem to be less severe in maize indicating that the grasses may be hosts of weakly pathogenic strains, very mild strains or host-adapted strains (Peterschmitt *et al.*, 1991; Autrey and Ricaud, 1983; Rose, 1978).

Classical biological assays such as host range and symptom severity are known to be less accurate at typing of MSV isolates than molecular techniques (Briddon *et al.*, 1994; Clarke *et al.*, 1989; Pinner *et al.*, 1988). To this end, numerous MSV isolates have now been sequenced: MSV-SA (Lazarowitz, 1988), MSV-K (Howell, 1984), MSV-Reu (Peterschmitt *et al.*, 1996), MSV-N (Mullineaux *et al.*, 1984), MSV-Kom (Schnippenkoetter *et al.*, unpublished) and MSV-Set (Schnippenkoetter, 1998). Of these, MSV-SA, MSV-K, MSV-N and MSV-Kom share greater than 98 % sequence identity (Schnippenkoetter, 1998), while MSV-Reu shares between 94.4 to 95.3 % sequence identity with this group (Peterschmitt *et al.*, 1996). MSV-Set is the most distantly related, sharing only 78 % sequence identity with MSV-Kom (Schnippenkoetter, 1998). Full genomic sequence provides the most conclusive evidence of virus relationships, with Padidam *et al.* (1995) proposing that viruses sharing 90 % or more sequence similarity should be regarded as strains of a species. Rybicki (1998) argues against this arbitrary cut off mark due to the differences in intra-generic sequence similarities, and proposes the ability to complement replication functions as a criteria for relatedness (van Regenmortel *et al.*, 1997). These criteria, along with full genome sequence data, were used when designating BGMV isolates, genus *Begomovirus*, into one of two subtypes (Faria *et al.*, 1994; Gilbertson *et al.*, 1991).

A large body of evidence now exists that indicates that maize isolates of MSV are more pathogenic in maize than the grass and indigenous crop infecting isolates, where their host-adaptation seems to limit pathogenicity in maize (Schnippenkoetter, 1998; Soto *et al.*, 1982). Maize, barley, wheat and oats are all crops introduced to Africa; therefore these hosts may have selected for the more pathogenic strains of MSV from the population of viruses present (Soto *et al.*, 1982). These pathogenic strains are maintained, via leafhopper transmission, from one maize generation to the next, whereas the milder less pathogenic strains are eventually lost (Mesfin *et al.*, 1992; Pinner *et al.*, 1988). The continued maintenance of these strains may in part be due to increased leaf chlorosis, which affects the normal vector feeding patterns on grasses by attracting them towards the infected maize (Page, 1997; Boulton *et al.*, 1991; Rose, 1978). It is uncertain whether or not these severe maize-infecting MSV isolates evolved to become severe over time: there is evidence of maize MSV isolates becoming slightly more severe over a period of four years; conversely, no changes were detected after 58 generations of the mild MSV B form in maize (Pinner *et al.*, 1988; McClean, 1947). The presence of wheat infecting MSVs is

well established; however, no extensive data exists on: the pathogenicity of these isolates in other Gramineae; characterisation of their genome; and their relationship to other mastreviruses (Rose, 1978). The purpose of this study was therefore to biologically characterise two new wheat isolates of MSV, MSV-Tas and MSV-VW, and compare their genomic sequences with other mastreviruses, including the sequence of a new maize-infecting isolate, MSV-VM.

3.3: METHODS AND MATERIALS

3.3.1: Virus origins

The two wheat isolates MSV-Tas and MSV-VW originated from the South-western Cape Province and the Vaalhaarts irrigation scheme in the Orange Free State in South Africa, respectively, while the maize isolate MSV-VM was collected from the same area as the latter a year later (see Table 2.1) (Rybicki *et al.*, 1998; Rybicki and Hughes, 1990). The following clones and preliminary restriction map data were provided by Dr F. L. Hughes (F. L. Hughes and E. P. Rybicki, unpublished data): pTas100, a monomer of MSV-Tas cloned into the *Eco*RI site of pBluescript(SK) (Stratagene); pVM100 and pVW100, monomers of MSV-VM and MSV-VW cloned in the *Bam*HI site of pBluescript(SK). The clones had previously been shown to be representative of the predominant genomic forms present in infected plants by means of preliminary RE mapping of viral RF DNA (F. L. Hughes and E. P. Rybicki, unpublished data). The agroinfectious constructs of MSV-Kom and MSV-Set (pKom603 and pSet107, respectively) are both 1.1mers containing two LIRs cloned into pBI121 (Clonetech, Palo Alto, CA), were provided by Schnippenkoetter (1998) (see Table 3.1).

3.3.2: Agroinfectious constructs

Monomeric clones of MSV are not infectious, so agroinfectious constructs require duplication of sequence, allowing escape from the binary vector either by recombination or replicative release. The latter is only possible, and occurs preferentially, when two LIRs are present (Heyraud *et al.*, 1993; Stenger *et al.*, 1991; Boulton *et al.*, 1989). The monomeric clones of MSV-Tas, MSV-VW and MSV-VM were partially dimerised as follows. Subclones containing the LIR were made for the three isolates (see Fig. 3.1 for restriction maps): pVM100 was digested with *Sac*I, pVW100 digested with *Hind*III and pTas100 with *Sal*I, as per manufacturer's instructions (Boehringer Mannheim).

Table 3.1: Sequenced Mastreviruses

Isolate Name	GenBank Accession No.	Reference	Isolate Name	GenBank Accession No.	Reference
MSV-SA	Y00514	Lazarowitz, 1988	DSV	M23022	Donson *et al.*, 1987
MSV-Reu	X94330	Peterschmitt *et al.*, 1996	PanSV-Kar	L39638	-[1]
MSV-K	X01089	Howell, 1984	PanSV-Ken	X60168	Briddon *et al.*, 1992
MSV-Set	AF007881	Schnippenkoetter, 1998	CSMV	M20021	Anderson *et al.*, 1988
MSV-N	X01633	Mullineaux *et al.*, 1984	BeYDV	Y11023	Liu, L. *et al.*, 1997
MSV-Kom	AF003952	-[1]	TYDV	M81103	Morris *et al.*, 1992
SSV-N	M82918	Hughes *et al.*, 1993	WDV-S	X02869	MacDowell *et al.*, 1985
SSV-E	AF037752	-[2]	MiSV	D01030	Chatani *et al.*, 1991
SSV-Mill[3]	X86705	Briddon *et al.*, 1996b			

[1]: Unpublished data: W. H. Schnippenkoetter F. L. Hughes, D. James, M. Fyvie, J. A. Willment & E. P. Rybicki. University of Cape Town, South Africa.
[2]: Unpublished data: L. Bigarre. CIRAD-CA, Phyma, BP 5035, Montpellier, 34080, France
[3]: Coat protein gene sequence only available.

All DNA manipulation techniques were performed as described by Sambrook *et al.* (1989), unless otherwise stated. Each sample was then electrophoresed through a 0.8 % agarose 0.5 x TBE gel. To recover the DNA fragment of interest, the gel was stained with methylene blue (Flores *et al.*, 1992), and the required band excised and separated from the agarose using the Geneclean® Kit (Bio101 Inc). The DNA fragment was then self ligated using T4 DNA ligase (Boehringer Mannheim) overnight and transformed into competent *E. coli* cells (Chung and Miller, 1988). Plasmid DNA was isolated from positive colonies and the resultant plasmids containing a 0.1mer, 0.25mer and 0.5mer of viral DNA were named pVM401, pVW103 and pTas201, respectively.

Plasmids pVM100 and pVW100 were digested with *Bam*HI and *Sca*I to release the full monomer, and separated from the vector by gel electrophoresis. Likewise, pTas100 was digested with *Eco*RI and *Sca*I and the viral genomic DNA was isolated as above. The constructs pVM401 and pVW103 were digested with *Bam*HI and pTas301 with *Eco*RI, all three were dephosphorylated using calf intestinal phosphatase (Boehringer Mannheim) and ligated to their respective full-length genomes (as above). All dimers were transformed into and maintained in *RecA*$^-$ strains of *E. coli* such as DH5α or JM109. These partial tandem dimers were then named pVM501, pVW203 and pTas301 (see Fig. 2.2). Large scale plasmid DNA preparations were obtained using Nucleobond AX100 columns (Machery-Nagel).

The *A. tumefaciens* binary vector pBin19 (Bevan, 1984), expressing the kanamycin resistance marker, was used as the cloning vector for the partial dimers. The partial dimer of MSV-VM was cloned into *Sac*I and *Eco*RI digested pBin19, by digesting pVM501 with *Eco*RI to completion and then partially digesting with *Sac*I. The required DNA fragment was isolated by gel electrophoresis and subsequent Geneclean® treatment (as above). The plasmid pVW203 was digested to completion with *Xba*I and partially with *Sal*I releasing the partial dimer which was then cloned into pBin19 previously prepared with the same restriction enzymes. Plasmid pTas301 was digested with *Xho*I and *Xba*I and cloned into pBin19 cut with *Sal*I and *Xba*I. The viral constructs in pBin19 (pVM601, pVW303 and pTas401) were then transformed using the freeze-thaw method (Holsters *et al.*, 1978) into the nopaline strain of *A. tumefaciens* C58C1 (pMP90) (Koncz and Schell, 1986). Transformants were selected for on LB agar plates containing kanamycin (100 µg/ml), gentamycin (40 µg/ml) and rifampacin (100 µg/ml) grown at 30^{0}C for three days (Sambrook *et al.*, 1989).

3.3.3: Agroinfection of maize plants

Cultures of *A. tumefaciens* were grown up in LB (Sambrook *et al.*, 1989) with gentamycin (40 µg/ml), rifampicin (100 µg/ml) and kanamycin (100 µg/ml) overnight at 30^{0}C and sub-inoculated into a 5 ml culture and grown until an 0.4 OD was reached as measured on a Beckman DU-64® spectrophotometer at 600 nm. A 1 ml aliquot was centrifuged at 5000 rpm in a benchtop microfuge (Eppendorf) for 2 minutes. The pelleted bacterial cells were resuspended in 100 µl sterile distilled water, centrifuged and the pellet resuspended in sterile distilled water. Of this suspension approximately 2 µl was injected directly into the coleoptilar node of three-day old sweetcorn cv. Jubilee seedlings using a 25 µl Hamilton syringe (Bonaduz, Switzerland) (Grimsley *et al.*, 1988; Grimsley *et al.*, 1987). The seedlings had been germinated in moist prebaked vermiculite and incubated at 30^{0}C for three days in a sealed bag prior to injection. The injected seedlings were planted directly into soil, four per pot, and kept in plant growth rooms with a 14hr on/10 hr off light cycle (VHO Growlux fluorescent light tubes), approximately 70 % humidity and temperatures between 18^{0}C and 24^{0}C. Symptom development was monitored every three days until 26 days after injection. The second quarter of the third and fourth leaves was photographed on day 15 and 22, respectively. For comparison with other agroinfectious MSVs, *A. tumefaciens* was transformed with 1.1mers of MSV-Kom and MSV-Set in

3.3.5: Leafhopper transmissions

Non-viruliferous leafhoppers (*C. mbila*) were maintained in gauze covered wooden cages as described by Clark *et al.* (1989) and Hughes (1991). The cages were placed in front of a window receiving late morning to early afternoon sun at ambient room temperatures (+/- 22^0C). Maize plants (*Zea mays* cv. Kalahari early pearl, commonly known as "Witplat"), were used for maintaining the colonies and were exchanged periodically for new plants. Leafhoppers were culled using the insecticide Kombat™ (Efecto). For the leafhopper transmission studies, agroinfected plant material (see above) was placed in a new cage and the leafhoppers were allowed to acquire the virus over a three-day period. A latency period of seven days was allowed after which the viruliferous leafhoppers were divided, approximately 20 leafhoppers per cage, into new cages containing the prospective host plant (7 – 10 days old) under investigation. The cages held eight different pots containing different varieties of two week old host plants, 4 – 5 maize plantlets per pot and 12 – 50 wheat or barley plants per pot. Every second day the leafhoppers were gently shaken off the plants and allowed to resettle ensuring that most plants would be fed upon. The leafhoppers were fed on the plants for a period of one to two weeks, after which they were culled (von Wechmar and Hughes, 1990). The plants were returned to the plant growth rooms and symptoms were assessed visually.

Symptoms were rated using the following scale: + = mild streak and recovery; ++ = moderate stippled streak to continuous streak (++[+]), +++ = severe continuous streak and stunting or distortion of growth and death (+++[+]) (Schnippenkoetter, 1998). Comparisons were made with the biologically characterised isolates MSV-Kom and MSV-Set, a moderately severe maize infecting and a mild *Setaria* infecting strain, respectively (Schnippenkoetter, 1998; Martin and Rybicki, 1998; Martin *et al.*, 1999).

All leafhopper transmissions were performed in isolation and only one MSV isolate was maintained in cages at any time to ensure no cross contamination of viruliferous leafhoppers. The acquisition and feeding times were longer than the necessary minimum periods as defined by Storey (1928) and Granados (1969): firstly, this allowed the establishment of a viruliferous leafhopper colony of all instars; secondly due to the high number of plants in some cages, it ensured adequate feeding on all plants. Each cage contained a susceptible host, sweetcorn cv. Jubilee, which was used as a positive indicator

of the successful transmission of virus (von Wechmar and Hughes, 1990). Where no symptoms were obtained on a cultivar the transmission study was repeated with only the cultivar under question present in the cage.

The choice of hosts for the transmission studies was limited to those used in previous studies (Schnippenkoetter, 1998; von Wechmar and Hughes, 1990). Furthermore, in order to make the results relevant and comparable to the transmission results obtained by these studies the same methodology and assessment criteria was followed. The cultivars used during the transmission studies were obtained as follows: the susceptible varieties of maize (von Wechmar and Hughes, 1990), including the commercial varieties of sweetcorn cv. Jubilee from Stark Ayres Nursery, South Africa, maize cv. Kalahari early pearl "Witplat" from Pioneer Seed Co. (Pty) Ltd. (now PANNAR Ltd.), and popcorn from the local supermarket. Tolerant or resistant varieties of maize: "Vaalhartz Geel" was obtained from the Grain Crops Institute (GCI), Potchefstroom, South Africa. Hybrid maize, PNR 6549 and PNR 6552, susceptible and tolerant varieties respectively, obtained from PANNAR Ltd. (Barrow, 1990; van Rensburg *et al.*, 1990), as were the lines Pan 6099, Pan 6363 and Pan 6364 (van Rensburg *et al.*, 1990). The Elsenberg Agricultural Development Institute, South Africa provided wheat cultivars SST44, SST66, Dias, Vloekstoot, Adam Tas, Chokka, Festiquay, Marquis, Agent, and barley cultivars Clipper, Schooner and Sterling. Sugarcane (*S. officinarum* cv. Uba and H44-3098) was provided by the South African Sugar Association Experimental Station, Mt. Edgecombe, Kwazulu-Natal.

3.3.6: Sequencing of MSV-VW, MSV-VM and MSV-Tas

Subclones of pVW100, pTas100 and pVM100 were made in pBluescript KS or SK, using conveniently positioned restriction endonuclease sites for the generation of subclones with overlapping ends (Fig. 3.1). Each subclone was sequenced in both directions using the universal sequencing primers:

Forward primer: 5'-CGCCAGGGTTTTCCCAGTCACGAC-3' and
Reverse primer: 5'-GAGCGGATAACAATTTCACACAGG-3'.

Mrs Di James of the Dept. of Microbiology Sequencing Service performed the sequencing reactions and provided me with the unedited sequence data generated on the ALF Express

automated sequencer (Pharmacia). DNA sequence manipulations were performed using the Genetics Computer Group (GCG) (Devereux *et al.*, 1984) and DNAMAN Version 3.0 (Lynnon BioSoft© 1994 –1997) software packages. Numbering of the genomic sequences is based on the recommendation of the 1st International Symposium on Geminiviruses, Almeria, Spain (1994) where the first base is the penultimate A of the geminivirus invariant TAATATTAC loop sequence.

Multiple sequence analysis was performed using PILEUP function, a progressive pairwise alignment program in the GCG package, with the subsequent manipulation of the alignments in CLUSTALW (Thompson *et al.*, 1994). The CLUSTALW program default parameters were used to generate a phylogenetic tree based on the Neighbour-Joining method of (Saitou and Nei, 1987). The phylogenetic tree was then subjected to 1000 bootstrap trials and the output file manipulated in TREEVIEW© 1.4 (Win32) (Page, 1996) for presentation purposes. Alternatively, the multiple sequence alignment and phylogenetic tree generating features of DNAMAN Version 4.0 (Lynnon BioSoft©, Quebec, Canada), also based on the Neighbour-Joining tree method (Saitou and Nei, 1987), were used. Reference to either method is given with the appropriate result. Comparisons of the putative protein products of various sequenced viruses was performed (see Table 3.1 for GenBank accession numbers) and the relevant features of both the DNA and protein are discussed.

3.4: RESULTS

3.4.1: Agroinfection of MSV isolates

Subclones of variable sizes containing the LIR were made from all three full-length genomic clones, using convenient restriction sites (see Fig. 3.1). Partial dimers were then constructed so that each contained two LIRs (pTas301, pVW301 and pVM501 (see Fig. 3.2)) which were then cloned into pBin19 (Bevan, 1984). These clones (pTas401, pVW302 and pVM601) were then transformed into *A. tumefaciens* and used for the agroinfection of sweetcorn cv. Jubilee seedlings. The three isolates, MSV-Tas, MSV-VW and MSV-VM all produced symptoms on the indicator host sweetcorn cv. Jubilee. The symptoms ranged from the relatively severe (MSV-VM), to moderate (MSV-Tas) and mild (MSV-VW) (see Fig. 3.3 A & B). Symptoms were the classical chlorotic stippled to continuous streak, with MSV-VM causing the most chlorosis and leaf stunting, while

MSV-Tas and MSV-VW produced finer streaks. These symptoms were reminiscent of the A and B forms of MSV, respectively (McClean, 1947). Comparisons were made with agroinfected MSV-Kom and MSV-Set, showing that in all three cases the symptoms differed visually, and that the infection kinetics varied slightly. The plants that became agroinfected were noticeably smaller than the negative control plants or those injected with transformed *A. tumefaciens,* but which remained uninfected (data not shown). Similar observations were made by Dickinson *et al.* (1996). The amount of stunting produced by the various isolates ranged from slight (MSV-Set and MSV-VW), to distinct (MSV-Tas) to dramatic (MSV-VM and MSV-Kom). Evidence of this can be seen in Fig. 3.3, where there is a strong inverse correlation between the width of the leaf segment and the severity of the isolate (D. P. Martin, Pers. Comm.). All five viruses infected 75 % or more (the average of two separate experiments) of surviving injected plants (see Table 3.2).

The percentage chlorosis on an infected leaf gives an indication of the severity of infection and a rough guide for the overall pathogenicity of the virus (Storey and McClean, 1930). The average percentage chlorosis of the second quarter of the fourth to sixth leaves of agroinfected plants (three separately performed experiments, with 14 plants per experiment) was quantified and performed by D. P. Martin using his Image09 computer based program (Martin and Rybicki, 1998). These data indicate that on sweetcorn cv. Jubilee MSV-Tas caused 31.1 % (+/- 3.0 %) chlorosis, MSV-VW caused 11.8 % (+/-1.7 %) chlorosis and MSV-VM caused 55.6 % (+/- 4.3 %) chlorosis, confirmation of the relative severity of these viruses (Martin *et al.*, 1999).

Table 3.2: The percentage (%) of agroinfected sweetcorn cv. Jubilee plants showing symptoms after 29 days.

	MSV-Tas		MSV-VW		MSV-VM		MSV-Kom		MSV-Set	
	#[1]	%	#	%	#	%	#	%	#	%
1	11/11	100	12/14	85.7	10/11	90.9	12/14	85.7	17/17	100
2	17/24	70.8	20/23	86.9	14/19	73.6	17/17	100	20/20	100

[1]: number of plants infected divided by the total number surviving the agroinfection procedure

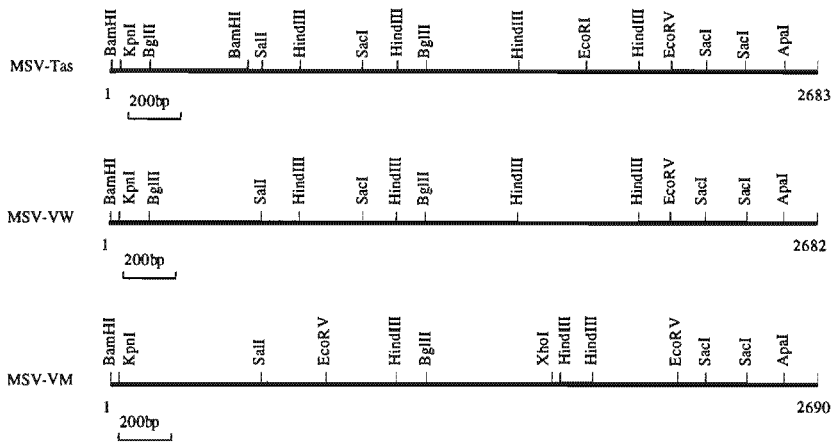


Figure 3.1: Restriction endonuclease maps of MSV-Tas, MSV-VW and MSV-VM linearised at the *Bam*HI site, two nucleotides after the start codon of the V1 ORF. The maps were generated in DNAMAN 4.0 using the sequence data obtained during this study.

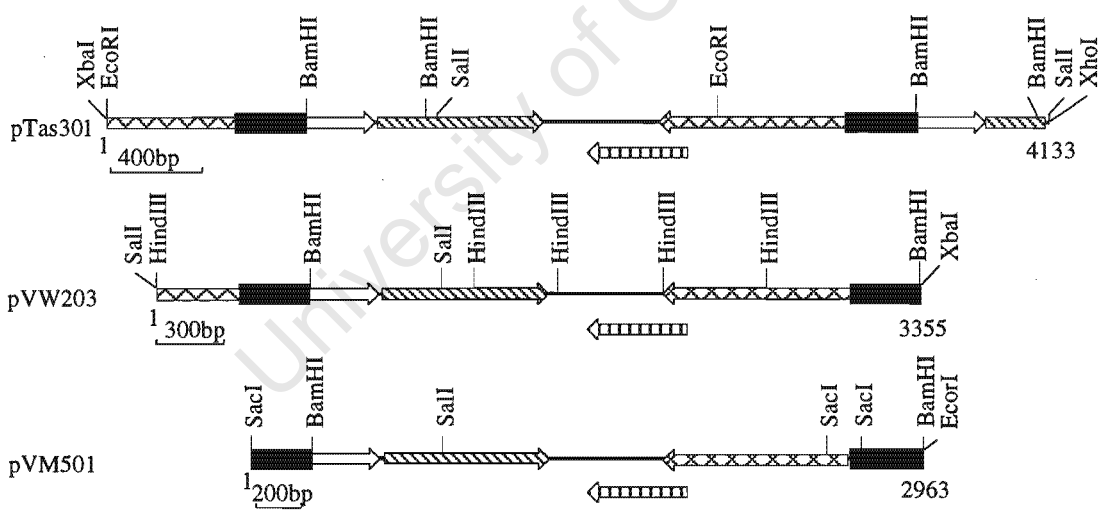


Figure 3.2: Partial tandem dimer constructs of MSV-Tas (top), MSV-VW (middle) and MSV-VM (bottom) in pBluescript(SK), which were used for cloning into the binary vector, pBin19 . The direction and relative size of the ORFs are shown as follows: the MP (ORF V1) is a white arrow, the CP (ORF V2) is an obliquely shaded arrow, the RepA (C1 ORF) is a obliquely hatched arrow, the C2 ORF is a vertically shaded arrow and the LIR is indicated by a black box. Restriction endonuclease sites used during cloning of the various constructs are indicated as well as the full size, in base pairs, of the partial dimer.

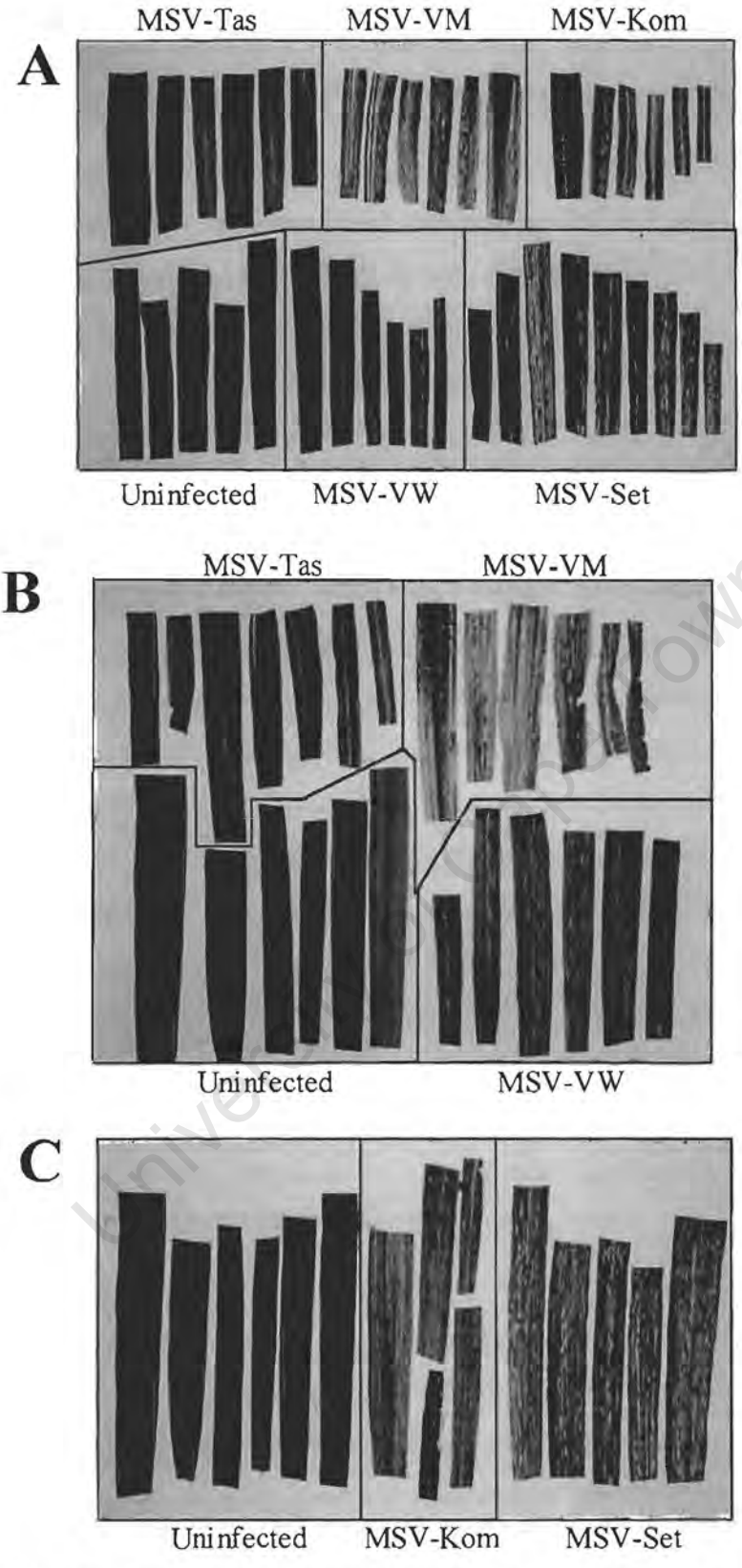


Figure 3.3: Streak symptoms on sweetcorn cv. Jubilee. Three day old seedlings were agroinfected with partial dimers of MSV-Tas, MSV-VM, MSV-VW, MSV-Kom and MSV-Set (pTas401, pVM601, pVW303, pKom603 and pSet107, respectively). Symptoms on the third leaf were assessed 15 days after agroinfection (A) and on the fourth leaf 22 days after agroinfection (B & C). The second quarter of each leaf is presented for comparative analysis of symptoms.

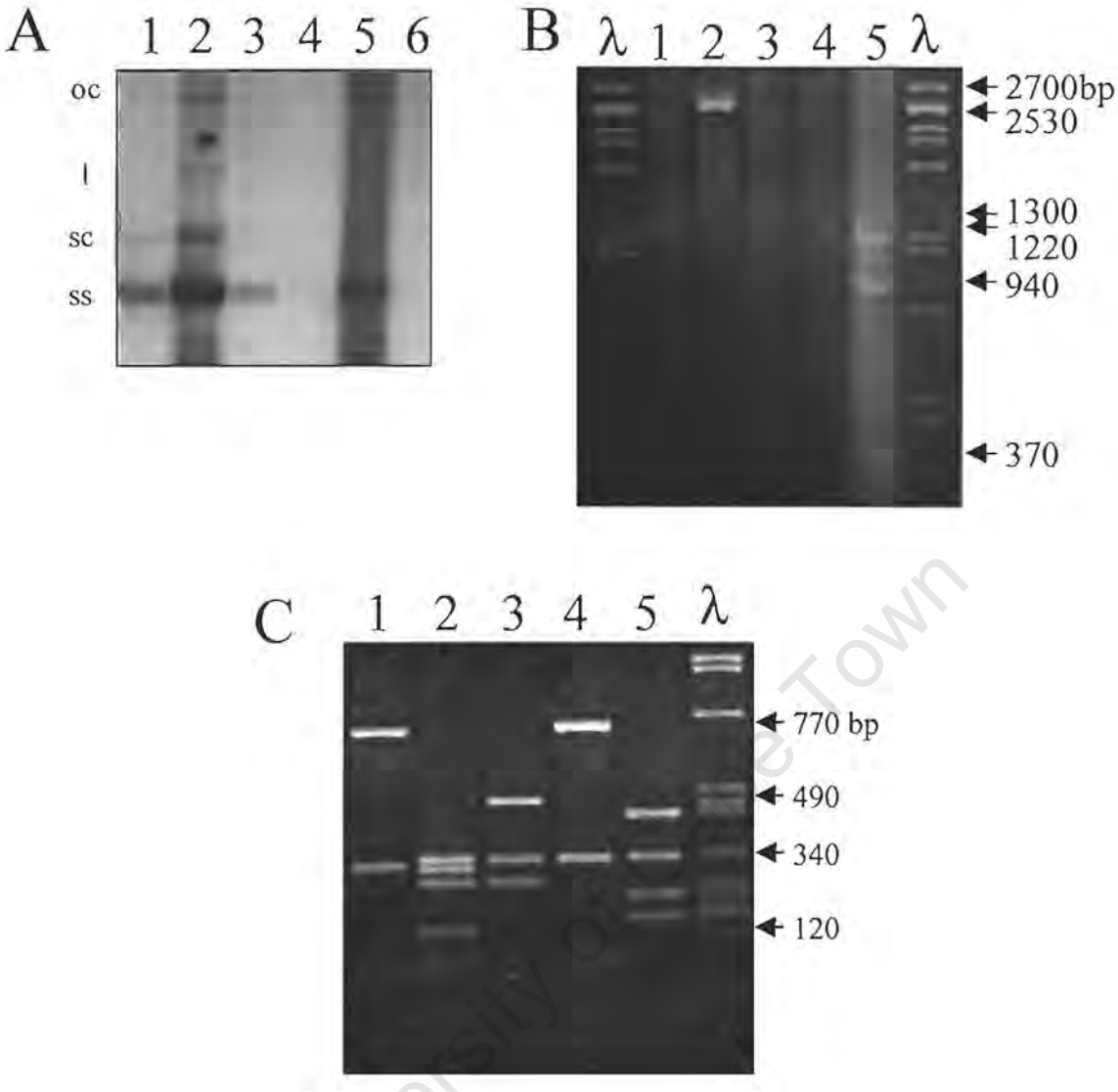


Figure 3.4: Analysis of viral DNA extracted from agroinfected plants by Southern blot using a MSV-Kom Dig-labeled probe (A), *Sac*I digestion of genomic DNA (B), and PCR and *Hae*III digestion (C). A & B) lanes 1, MSV-Tas; lane 2, MSV-VM; lane3, MSV-VW; lane 4, MSV-Set and lane 5, MSV-Kom. A) Lane 6 contained uninfected sweetcorn genomic DNA. C) Lane 1, MSV-VM; lane 2, MSV-VW; lane 3, MSV-Tas; lane 4, MSV-Kom; and lane 5, MSV-Set. DNA was electrophoresed through 1.5 % agarose gels, stained with ethidium bromide, in 0.5x TBE. Lambda *Pst*I digested DNA was used as a molecular weight marker (λ). The various replicative forms of viral genomic DNA are indicated as follows: open circular (oc), linear (l), supercoiled (sc) and single stranded (ss). Sizes of bands are indicated as number of base pairs (bp).

3.4.3.1: Host range studies

Viruliferous leafhoppers were able to transmit both MSV-Tas and MSV-VW to a variety of cultivars of wheat, barley, maize and a grass species (*Digitaria sanguinalis*) (see Table 3.3). Although the host range was limited in the number of cultivars of each type of host, comparative assessments with the well characterised isolate MSV-Kom and the grass type MSV-Set could be performed (Schnippenkoetter, 1998). These comparisons are necessary to fully appreciate the implications of the different host ranges: the fully characterised isolate MSV-Kom is known to be of moderate severity and MSV-Set is a mild virus in maize (Schnippenkoetter, 1998; Martin *et al.*, 1999). Due to the high sequence similarity of MSV-VM to the maize-type MSVs (see 3.4.4) and its similar severity of agroinfection with MSV-Kom (see Fig 3.3 and Martin *et al.* 1999) leafhopper transmission studies were not undertaken for this virus.

Both MSV-VW and MSV-Tas were successfully transmitted to the susceptible varieties of sweetcorn cv. Jubilee, popcorn and maize cv. Kalahari early pearl "Witplat", although in each case MSV-VW produced milder symptoms and affected fewer plants (see Table 3.3 & Fig. 3.5A). Neither isolate infected the resistant and tolerant varieties Pan 6099 or "Vaalhaarts Geel", respectively; however both maize varieties had previously been shown to be successfully infected by MSV-Kom (Schnippenkoetter, 1998). The susceptible hybrid PNR 6549 was more severely affected by MSV-Tas than MSV-VW (see Fig 3.5B), while the moderately resistant hybrid Pan 6364 was only moderately affected by MSV-Tas (Barrow, 1990; van Rensburg *et al.*, 1990). Both these hybrids were severely infected by MSV-Kom and not at all by MSV-Set (Schnippenkoetter, 1998). Relative to the ability of MSV-Kom to infect non-maize hosts, the increased host range of both MSV-Tas and MSV-VW into the wheat/barley cultivars, indicates some form of host-adaptation (see Fig.3.5C & D).

The wheat cultivars Festiquay, Marquis and Agent (all rust differentials) and SST44, an irrigation cultivar, obtained from Elsenberg Agricultural Development Institute, were the most resistant to infection by all four isolates, with the grass isolate MSV-Set being the most pathogenic (Schnippenkoetter, 1998). The grass species *Digitaria sanguinalis* was the most susceptible host with visible stunting and large chlorotic areas on all new leaves for both MSV isolates tested (data not shown).

Table 3.3: Comparative susceptibility of wheat, barley, maize and grass cultivars to leafhopper transmission of MSV.

	MSV-Tas		MSV-VW	
Cultivars	Symptoms[a]	% Infected[b]	Symptoms	% Infected
Triticum aestivum **cultivars**				
Adam Tas	+++	92.8 (52/56)	+++[+]	71.7 (33/46)
Agent	+	5.7 (3/52)	++	15.7 (6/38)
Dias	+++	54.1 (26/48)	+++[+]	64.4 (29/45)
Marquis	-	0 (0/50)	-	0 (0/34)
SS44	+	4 (3/75)	++	6.5 (6/91)
SS66	+++	91.6 (11/12)	+++	55.2 (21/38)
Festiquay	+[+]	8.3 (4/48)	++	23.0 (9/39)
Chokka	+++[+]	82.6 (86/104)	+++[+]	70.5 (24/34)
Vloekskoot	+++	33.3 (20/60)	++	11.1 (2/18)
Hordeum vulgare **cultivars**				
Clipper	+++[+]	100 (50/50)	+++[+]	87.8 (36/41)
Sterling	+++	ND[c]	+++[+]	91.6 (44/48)
Schooner	++	30.1 (19/63)	+	10.3 (6/58)
Zea mays **cultivars**				
sweetcorn "Jubilee"	+++	100 (20/20)	++[+]	92.3 (12/13)
sweetcorn "US More"	++	100 (13/13)	ND	ND
Popcorn	+++	92.5 (37/40)	++	87 (7/8)
Pan 6099	-	0 (0/43)	-	0 (0/15)
PNR 6549	+++	56.2 (18/32)	+	25 (2/8)
PNR 6552	-	0 (0/9)	ND	ND
Pan 6363	+[+]	25 (7/28)	+[+]	25 (2/8)
Pan 6364	+	25 (2/8)	-	0 (0/10)
"Vaalhaarts Geel"	-	0 (0/28)	-	0 (0/9)
Kalahari early pearl "Witplat"	++	84.3 (27/32)	+[+]	90 (6/10)
Digitaria sanguinalis	+++[+]	100 (10/10)	+++[+]	100 (23/23)
Saccharum officinarum **cultivars**				
H44-3098	-	0 (0/1)	ND	ND
Uba	++	100 (1/1)	ND	ND

[a]: The symptoms were rated as follows: - = no symptoms, + = mild streak and recovery; ++ = moderate stippled streak to continuous streak (++[+]), +++ = severe continuous streak and stunting or distortion of growth and death (+++[+]), +/- slight symptoms only seen on one or two plants. [b]: Combined results of two transmission experiments indicating total percentage plants infected. Numbers in brackets are the number of plants infected divided by the total number introduced to viruliferous leafhoppers. The sugarcane experiment was only performed once. [c]: ND = not determined

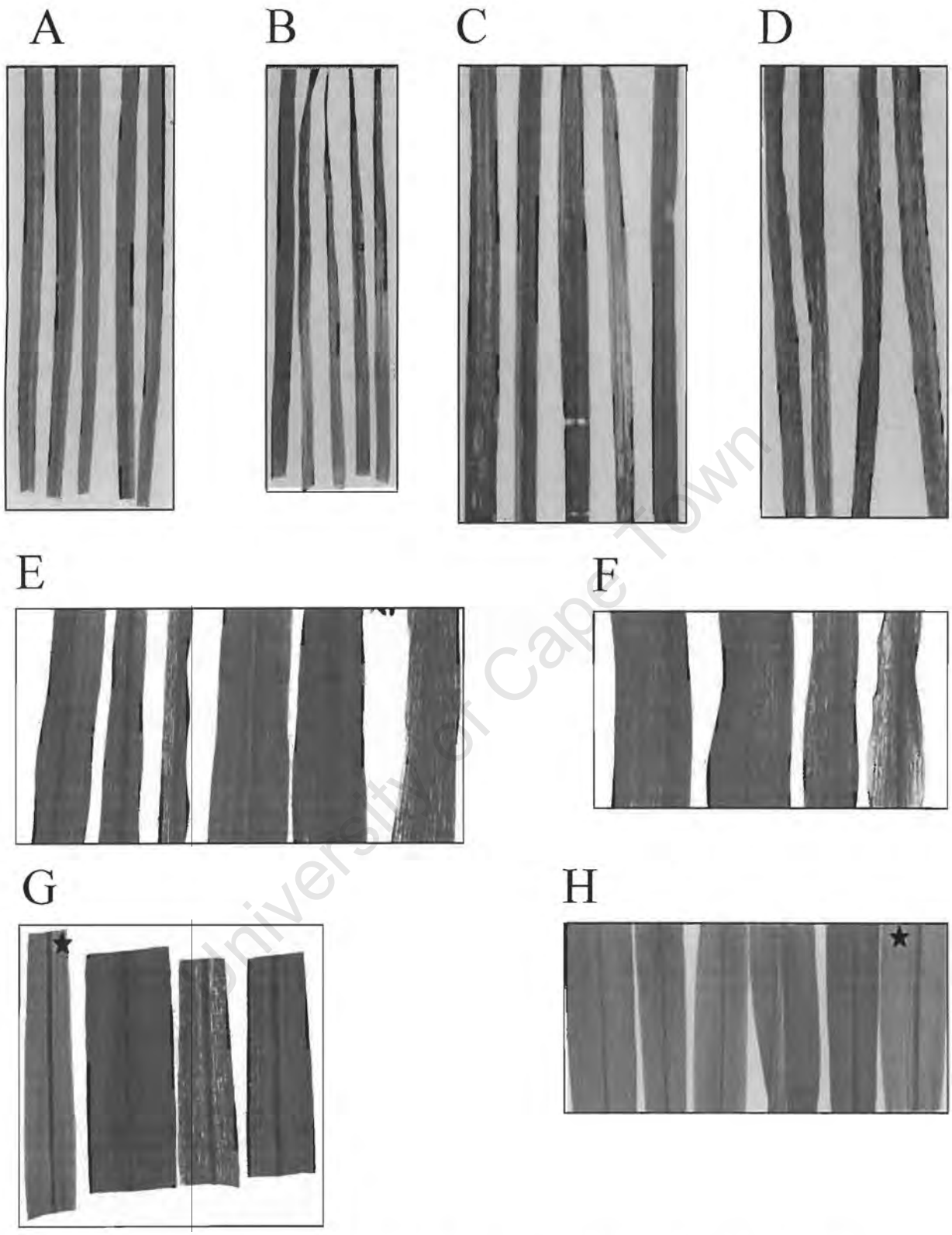


Figure 3.5: Leafhopper transmission studies of MSV-Tas and MSV-VW. Viruliferous leafhoppers were introduced to healthy two week old plants and the emergence of streak symptoms was recorded. MSV-VW (A) and MSV-Tas (B) symptoms on barley cv. Clipper. MSV-VW (C) and MSV-Tas (D) symptoms on wheat cv. SST66. MSV-Tas infected maize hybrid PNR6459 (E) and cv. Kalahari early pearl (F). MSV-VW infected maize hybrid PNR6459 (G) and cv. Kalahari early pearl (H). Uninfected leaves are indicated by a star (★).

MSV infection of sugarcane has been reported before (Thottappilly *et al.*, 1993; Storey and McClean, 1930); interestingly, in this study of the two sugarcane cultivars only *S. officinarum* cv. Uba became infected by MSV-Tas. All subsequent growth for well over a year showed symptoms, unlike the results obtained by Storey & McClean (1944), where MSV infections were transient.

3.4.4: Sequence analysis

The assembled sequence data for MSV-Tas, MSV-VW and MSV-VM indicated that the genome lengths of 2683, 2682, and 2690 bp, respectively were comparable to previously sequenced strains of MSV (see Fig. 3.6). Four ORFs was detected using the ORF predicting function of DNAMAN 4.0, which also estimates the size of the protein products (see Table 3.4). The pertinent features for RNA transcription and translation of the various protein products are indicated relative to their sequence positions starting from the penultimate nucleotide of the loop sequence (Table 3.4). The ORFs are identical in length, with the differences in genome sizes being due to insertions or deletions within the intergenic regions (see below).

Table 3.4: Predicted open reading frames for protein products with known function, based on sequence data of MSV-Tas, MSV-VW and MSV-VM[a].

Isolate	ORF	TATA[b]	Initiation codon[c]	Termination codon	Polyadenylation signal[f]	Protein (Daltons)
MSV-Tas	V1	120	150	453[d]	1203	10 876
MSV-VW	V1	120	150	453[d]	1203	10 946
MSV-VM	V1	121	150	453[d]	1203	10 821
MSV-Tas	V2	120	466	1198[d]	1203	26 936
MSV-VW	V2	120	466	1198[d]	1203	26 952
MSV-VM	V2	121	466	1198[d]	1203	26 973
MSV-Tas	C1	2622	2522	1706[e]	1222	31 404
MSV-VW	C1	2621	2522	1706[e]	1222	31 568
MSV-VM	C1	2629	2528	1712[e]	1223	31 340
MSV-Tas	C1/C2	2622	2522	1365[d]	1222	41 363
MSV-VW	C1/C2	2621	2522	1365[d]	1222	41 479
MSV-VM	C1/C2	2629	2528	1371[d]	1223	41 238

[a]: Table based on that of Lazarowitz (1988). [b]: Consensus sequence $T^{G}/_{C}TATA\ ^{T}/_{A}A_{1\text{-}3}$
[c]: AUG start codon. [d]: UAA termination codon. [e]: UAG termination codon. [f]: Probable polyadenylation signal $^{A}/_{G}ATAA$

```
MSV-Tas    ACCGCGCCTTCTTTTCCTGCGAGGGCCCG.TAGGGACCGAGCGATTTGATTTAAA    54
MSV-VW     ---------------------------.a----------------------    54
MSV-VM     ----------------------------g----------c----------    55

MSV-Tas    GTTCGGTTCTGCTTTGTCTGATTTATCTAAAGCAGCCCAATCTAAAGAAACCGGT   109
MSV-VW     ---------------------------------------------------   109
MSV-VM     ---t---c---------t-------------------------------------   110

MSV-Tas    CCCCTGCGACTATAAATTGTTTCACAAGTGCGATTCATTTATGGATCCACAGAAC   164
MSV-VW     -------------------c-----------------c---------------   164
MSV-VM     ---gg--.-----------cc-a---------------c---------------   164

MSV-Tas    TCCTTTCTTTTACAGCCGCGGGTACCCACAGCAGCTCCGACATCCGGAGGAGTGT   219
MSV-VW     -----ct--g---------------------------------------------   219
MSV-VM     -----ct--g---------------------------------------------   219

MSV-Tas    CGTGGAGTCGCGTCGGCGAGGTAGCTATTTTGAGCTTTGTTGGATTGATTTGCTT   274
MSV-VW     ------------a----------------------------ct------at---   274
MSV-VM     ------------a----------------------------c------------   274

MSV-Tas    TTACCTGCTTTACCTTTGGGTGCTGAGAGATCTTATCTTAGTTCTGAAGGCTCGA   329
MSV-VW     ---------------------------------------------------   329
MSV-VM     ----------------------------c----------------------   329

MSV-Tas    CAAGGCAGATCCACGGAGGAGCTGATATTTGGAATACAAGCTGTGGATAGGAGCA   384
MSV-VW     ---------------------------------------------------   384
MSV-VM     ---------------------------------gg-----g------------   384

MSV-Tas    ACCCTATCCCTAATACACAGGCACCACCAAGTCAGGGGAATCCCGGGCCGTTTGT   439
MSV-VW     ----------------------------t---------------------c--   439
MSV-VM     -------------ct--ca----------------c----------a-----   439

MSV-Tas    TCCAGGCACGGGATAAGCAATCAGCCATGTCCACGTCCAAGAGGAAGCGGGGAGA   494
MSV-VW     ----------------------------t--t----------a--------   494
MSV-VM     ---------------------------------------------------   494

MSV-Tas    TGATGCGAACTGGAATAAGCGGACGACTAAGAAGAAGCCATCTTCAGCTGGTCTG   549
MSV-VW     ----t-------------------------------t---------------   549
MSV-VM     ----t----t-----------gt-c------------------------------   549

MSV-Tas    AAGAAGGCTGGAAGCAAGGCCGAAAGGCCATCCCTTCAGATCCAGACACTCCAGC   604
MSV-VW     ---------------------------------------------------   604
MSV-VM     ----g-----c-----------t----------c--a----------------   604

MSV-Tas    ATGCTGGGTCCACCATGATAACAGTCCCCTCCGGAGGAGTATGTGACCTCATCAA   659
MSV-VW     ---------------------------------------------------   659
MSV-VM     --------a-------------g-----a-------------------------   659

MSV-Tas    CACCTATGCCCGAGGATCCGACGAGGGCAACCGCCACACCAGCGAGACTCTGACG   714
MSV-VW     ------------------a-----------------------------------   714
MSV-VM     ------------------t-----------------------------------   714

MSV-Tas    TACAAGATTGCCGTCGACTACCACTTCGTTGCAGACTCGCAAGCCTGCAAGTATT   769
MSV-VW     ---------------------------------------------------   769
MSV-VM     ----------------------------------t---g--gc-------cgc--c-   769

MSV-Tas    CTAACACCGGGACCGGTGTGATGTGGCTGGTGTACGACACCACTCCCGGCGGACA   824
MSV-VW     ---------------------------------------------------   824
MSV-VM     -c--------a--------a-------------t---------------------   824

MSV-Tas    AGCGCCGACCCCACAAACCATATTTGCATATCCGGACACGCTGAAAGCTTGGCCG   879
MSV-VW     ---------------------------------------------------   879
MSV-VM     ---t--------g-----t--------c--c--t--------a-----g------   879

MSV-Tas    GCAACATGGAAAGTGAGCCGGGAGCTGTGTCATCGCTTCGTGGTGAAACGGCGAT   934
MSV-VW     ---------------------------------------------------   934
MSV-VM     --c-----------------------------------------------------   934
```

```
MSV-Tas    GGTTGTTCAACATGGAGACCGACGGTCGAATTGGTTCGGACATCCCTCCTTCGAA    989
MSV-VW     ----------------------------------------------------------  989
MSV-VM     ----------------------------g-----------t--------------    989

MSV-Tas    TGCAAGTTGGAAGCCGTGCAAGCGCAACATCTACTTCCACAAGTTCACGAGTGGG   1044
MSV-VW     ----------------------------------------------------------  1044
MSV-VM     -a-------------t-----------------------------------------  1044

MSV-Tas    TTGGGAGTGAGAACGCAGTGGAAGAATGTAACGGACGGAGGAGTTGGTGCGATAC   1099
MSV-VW     ----------------------------------------------------------  1099
MSV-VM     ----------------------------------------------------c--c-  1099

MSV-Tas    AGAGAGGAGCTCTGTACATGGTCATTGCCCCTGGCAATGGCCTTACATTTACTGC   1154
MSV-VW     ----------------------------------------------------------  1154
MSV-VM     ----------g-----t-----t--------c-----------------------  1154

MSV-Tas    CCATGGGCAGACCCGTCTGTACTTTAAGAGTGTCGGCAACCAGTAATGAATAAAA   1209
MSV-VW     ----------------------------------------------------------  1209
MSV-VM     ---------------------------------t----------------------  1209

MSV-Tas    AC.TCCGTTTTATTATATCTGATGAATGCTCAAAGCTTACATTAATATGTCGTGC   1263
MSV-VW     --.------------------------------------------------------  1263
MSV-VM     --gc-------------------------g-------------------------  1264

MSV-Tas    GATGGCACGAAAAA.CACACTGGAACGCAGCCCCGGGCTAAAGACCGGGTCTCAA   1317
MSV-VW     --------------.--------------------------------------  1317
MSV-VM     --------------a-----gca-t-aata-agg---g--gtcgg----cgg-t-  1319

MSV-Tas    GAGACCCTGCG....ATACAAAACATCGAAAAATCAAGATCTATATGAATTACAC   1368
MSV-VW     -----------....------------------a---------------------  1368
MSV-VM     ag-gtgg---tcggcggg-----------------------------------------  1374

MSV-Tas    TTCCTCCGTAGGAGGAAGCACAGGGGGAGAATACCACTTCTCCCCCGGCGACATT   1423
MSV-VW     ----------------------------------------------t---------  1423
MSV-VM     ----------------------------------------------t--------a  1429

MSV-Tas    TTATAAATCATGCAGTTTGCCTCGAAATACTCCAGCTGCCCTGGAGTCATTTCAT   1478
MSV-VW     ---------------------------------t-----------------c-  1478
MSV-VM     a-g-----g-------------------------------------------c-  1484

MSV-Tas    TCATCCAGTCTTCATCCCAGTTGGCGAGGATGATTGTAGGCTTAGACTTCATCTG   1533
MSV-VW     ------------------------t-------------------------------  1533
MSV-VM     -------a---------g-------------t-----------------t----  1539

MSV-Tas    AACTTTCTTCTTCTTGCCATATTTTGGATTTACAACGAACTCTTTCTGACAGCCA   1588
MSV-VW     ----------------------------c---------------------------  1588
MSV-VM     c--c--t--------a-----c--g--g-------t---a--cc-----------  1594

MSV-Tas    ACTAACTGTTTCCAACAAGGACAGAATTTAAATGGAATATCATCTACGATGTTGT   1643
MSV-VW     ---------------------------------a---------------------  1643
MSV-VM     ---------------------------------c---------------------  1649

MSV-Tas    AGATTGCGTCTTCGTTGTAAGAAGACCAATCAACATTATTTTGCCAATAATTATG   1698
MSV-VW     -------------------c--------------------------------------  1698
MSV-VM     -------------------t--------------------------g--------  1704

MSV-Tas    AAGCCCTAAGCTTCTGGCCCAAGTAGATTTCCCAGTTCTTGTTGGACCGACGATG   1753
MSV-VW     ----------------------------t--t----g----------------  1753
MSV-VM     --c-----g--------------------t--g----------g---------  1759

MSV-Tas    TAGAGGCTCTGCTTTCTTGTTCCTTCATCTGTTGGCTGGATAATTTATCCATCCA   1808
MSV-VW     ------------------------------------------g---------  1808
MSV-VM     -------------------a--t--------a--a-------caga---------  1814

MSV-Tas    TTCAAGGTCAGAGATAGCATCTTCGAGGGTGTAACAGGTAGGTTGAAGAAGCATA   1863
MSV-VW     ------------------c-------------------t--------------g  1863
MSV-VM     --gg--------a--t-----c--------a-----------------g-----g  1869
```

```
MSV-Tas    TATGATTCGGGACTAACCTGAAAGATGTTAGGCTGGAGCCAATCTTTGATTGACT   1918
MSV-VW     --------------------g---------------------------------   1918
MSV-VM     --a-c-----c---------g----------------------a----------   1924

MSV-Tas    CATTACAAAGAAGATCAGGTGAAGAGGGTGGATGAGGACTGATGAATTCTTCCTG   1973
MSV-VW     ----------------------------------------------c--------   1973
MSV-VM     ----------t-a---------g---------------t--g----c--------   1979

MSV-Tas    AATCTCAGGGAACAATTTATTTGCAGAGTATTCAAAATACTGCAATTTTGTGGCC   2028
MSV-VW     ------------------------------------------------------   2028
MSV-VM     ---------a--a-gc-------------------------------------a-   2034

MSV-Tas    CAGTCATAGGGGAACTCTTTACGAATCATAGAGAGGTACTCTAGCTTTGAAGTGG   2083
MSV-VW     ------------------------c-----------------------------   2083
MSV-VM     --a---a----a-g------ct-g-----g------------tct--g-----ac   2089

MSV-Tas    AGTGTGAAATAATCTCTCTCATTATTTCATCTTTTGAAGGTTTCTTTTCAGAATT   2138
MSV-VW     ----------------------------------g-------------------   2138
MSV-VM     -------------g----g--------------g-----c--t-----ctttac   2144

MSV-Tas    TCCCTTGGAAGGATTTCCTTGGAAGCTTGACTTCCTAGGAATGAAAGTACCTCTC   2193
MSV-VW     -------------------a------------------------------------   2193
MSV-VM     ct-tgaatcg-at------a-----ggg---------------------------   2199

MSV-Tas    TCAAACACAGCCAGAGGTTCCTTGAGAATGTAATCCCTGACCTTGTTTACTGATT   2248
MSV-VW     -------ag------------------------------------------------   2248
MSV-VM     ------------------------------------c---c----a-----c-   2254

MSV-Tas    TTGCACTCTGAATGTTTGGGTGGAACCCTTCGATATCAAAGAACCTTGAGTCAGT   2303
MSV-VW     ----------------------------t-------------------------   2303
MSV-VM     -g----------a--------a-----a-tt----------------------a   2309

MSV-Tas    GATTCTTACCGGTTTCTCTGTCTGGAGTAATGCATGTAAATGCATATCTCCGTCT   2358
MSV-VW     -------------------------t-------------------------a---   2358
MSV-VM     t--c--------c----------a--c---------------a-ct---a---   2364

MSV-Tas    TTGTGTGCCTCTTGGGCACATATGATATACTTCGGGGTCCAACGACCAACGAGCT   2413
MSV-VW     c-a---------c------------------------------------------   2413
MSV-VM     --a---------c---------ga--g-----g--aa--------a-g-------   2419

MSV-Tas    CCCAAATCATCTGACAGACGATTTCTGGATTTTCTGGACAGTGTGGATATGTTAA   2468
MSV-VW     ----------------g------------------------------------   2468
MSV-VM     ----g----------g-------a--------------c-t------g----g   2474

MSV-Tas    GAACGTGTTAACGTTCCTGTGTGAGAATTGACGGTTGGATGAGGAGGAGGCCATA   2523
MSV-VW     -------------------------c--------------------------   2523
MSV-VM     ---------tg---------------c--------------------------   2529

MSV-Tas    TCGGACGACTCCGAGC.AGCTTGCGGGATGGCAGGATGGGAGCTCCAAACTCTAT   2577
MSV-VW     ---------------.------a----------------------------   2577
MSV-VM     g-c------gga-gttg--gc--a----------ac------------------   2584

MSV-Tas    ATCAACCGGTTTGCGCCTTCGAAATCCGCCGCTCCCCC.TTTTATAGTGGTTGTT   2631
MSV-VW     ----------.-------------------------.----------------   2630
MSV-VM     -gt-ta-ctg-.-----------------------ttg-c----------a   2638

MSV-Tas    TATGGGCCGGACCGGGCCGGCCCAGCAGGAAAAGAAGGCGCGCAATAATATT      2683
MSV-VW     ------------------------------------------------------   2682
MSV-VM     a-------------------------------------------c-------   2690
```


Figure 3.6: Multiple sequence alignment of the full genomes of MSV-Tas, MSV-VW and MSV-VM performed using DNAMAN 4.0. The full sequence of MSV-Tas is shown in uppercase. Where the sequences of MSV-VW, MSV-VM and MSV-Tas are identical the nucleotides are indicated by a dash (-), different bases are in lower case and for the sake of correct alignment dots (.) have been inserted in place of gaps. The genome sequence is linearised at the penultimate nucleotide of the invariant TAATTATAC loop sequence.

3.4.4.1: Control and expression of proteins

Although the direct role of the ORFs defined above in Table 3.4 have not been experimentally confirmed, sufficient sequence homology with other characterised MSVs exists, and therefore their functions can be presumed to be similar.

3.4.4.2: Virion sense genes

Movement protein:

The gene product of the V1 ORF, the MP, is a non-structural protein necessary for infectivity due to its role in virus movement (Boulton *et al.*, 1989; Mullineaux *et al.*, 1988; Morris-Krsinich *et al.*, 1985). The MP has a core region showing high homology with other Gramineae infecting mastreviruses; while the first 14 N-terminal aa do not seem to be essential for infectivity, the protein itself is essential for systemic spread and cell to cell movement (Boulton *et al.*, 1993; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989; Mullineaux *et al.*, 1988). The onset of viral lesions coincides with the expression of MP, while a substitution of C to T at nt 40 affected streak width and chlorophyll levels, even though this change does not affect the aa sequence (Dickinson *et al.*, 1996; Boulton *et al.*, 1991). A putative *trans*-membrane domain has been identified, spanning aa residues 33 – 57: this is predicted to be a hydrophobic β- sheet, which agrees with the role of the MP as a plasmodesmata associated protein (Dickinson *et al.*, 1996; Boulton *et al.*, 1993). A multiple sequence alignment of the amino acid sequences of the MPs of the three sequenced isolates compared to MSV-N confirms the presence of a variable region within the 14 aa N-terminal region (Peterschmitt *et al.*, 1996; Boulton *et al.*, 1989). The *trans*-membrane spanning region is absolutely conserved between MSV-VM and MSV-N, while MSV-Tas and MSV-VW both have one change relative to MSV-N, A to G in the former and C to I in the latter (see Fig. 3.7). Both these substitutions are non-polar conservative changes. The predicted sizes of the three MPs (see Table 3.4) are all approximately 10.8 kDa: experimental evidence obtained by Western blot analysis and immunoprecipitation of plant extracts detected maize-type MSV MP of 14 kDa and 11 kDa, respectively, indicating that post-translational modifications may occur (Dickinson *et al.*, 1996; Mullineaux *et al.*, 1988).

```
                1
                |
MSV-N           MDPQNALYYQPRVPTAAPTSGGVPWSRVGEVAILSFVALICFYLLYLWVL
MSV-Tas         -----sfll----------s----------g----------
MSV-VW          -----sffv----------s---------i---------
MSV-VM          -----sffv----------s-----------------------

                                                              101
                                                               |
MSV-N           RDLILVLKARQGRSTEELIFGGQAVDRSNPIPNIPAPPSQGNPGPFVPGTG
MSV-Tas         --------------------i----------tq---------------
MSV-VW          --------------------i----------tq---i-----------
MSV-VM          ----------------------g---------l----------------
```

Figure 3.7: Alignment of the MP amino acid sequence of MSV-Tas, MSV-VW and MSV-VM with respect to MSV-N (Mullineaux *et al.*, 1984). Conserved amino acids are indicated by a dash (-), while difference are indicated in lower case letters, with respect to the translated V1 ORF of MSV-N. The predicted *trans*-membrane domain identified in MSV-N (Boulton *et al.*, 1993), aa 32 - 57, is indicated in bold. Sequences were aligned using DNAMAN 4.0.

Coat protein

Mastreviruses require both the MP and CP in order to systemically infect their hosts, unlike most begomoviruses, which do not require a CP for systemic infection (Boulton *et al.*, 1993; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989; Gardiner *et al.*, 1988). This distinction between the two geminivirus genera is not due to the type of host infected, because BeYDV, a dicot-infecting mastrevirus, also requires its CP for systemic infection (Liu *et al.*, 1998). The CP genes of MSVs are highly conserved, presumably due to the multi-functionality of the CP in terms of vector specificity, movement within the plant and encapsidation (Briddon *et al.*, 1994; Briddon *et al.*, 1992; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989).

Sequence analysis of the begomovirus CP revealed that the N-terminal 60 – 70 aa were more variable than the rest of the sequence; however both the curto- and mastrevirus CPs are more uniformly variable throughout the whole aa sequence (Padidam *et al.*, 1995). There is a stretch of high homology between the 5 aa positioned around the N-terminal 110 region; however, the significance of this region is unknown (see Fig. 3.8) (Padidam *et al.*, 1995). Interestingly, and possibly due an ancient recombination event, the CP of BCTV is more similar to those of the mastreviruses; however the rest of the genome is more similar to those of the New World begomoviruses (Padidam *et al.*, 1995; Rybicki,

1994). Strong interrelationships exist between CPs and specific vectors as evidenced by BCTV and other leafhopper transmitted geminiviruses, with vector specificity being defined by the CP (Hofer *et al.*, 1997; Briddon *et al.*, 1990; Stanley *et al.*, 1986).

The DNA binding domain of the MSV-N CP has been localised to the N-terminal 104 aa, where it binds both ss and ds DNA non-specifically. However, these are *in vitro* results and may not necessarily reflect an *in vivo* function (Liu, H. *et al.*, 1997). A nuclear localising domain, between aa 5 – 22, could be responsible for the nuclear shuttling of viral DNA to the nucleus. The CP, together with the MP, is responsible for the movement of DNA both intra- and intercellularly, although the mechanism by which this occurs is uncertain (Liu, H. *et al.*, 1997). Alignments of the N-terminal 110 aa of the CP reveal that the most differences in sequence fall within this region, with most of the differences in the putative nuclear localisation signal. Other differences seem to be mostly conserved between the maize / wheat isolate groupings (see Fig. 3.8); however, without further investigation the implication of these conserved differences is unknown.

```
MSV-Tas    MSTSKRKRGDDANWNKRTTKKKPSSAGLKKAGSKAERPSLQIQTLQHAGSTMITV    55
MSV-VW     ----------s----------------------------------------------    55
MSV-VM     ----------s-----vp----------r-a---d-------------t-----    55
MSV-N      ----------s--s--v----------r-----d-------------t-----    55
MSV-SA     ----------s--vp----------r-----d-------------t-----    55
MSV-Reu    ----------s-----vs----------r-----d-------------t-----    55

MSV-Tas    PSGGVCDLINTYARGSDEGNRHTSETLTYKIAVDYHFVADSQACKYSNTGTGVMW   110
MSV-VW     ----------------------------------------------------   110
MSV-VM     ------------------------------------aa--r----------   110
MSV-N      ----------------------------i-------aa--r----------   110
MSV-SA     ------------------------------------aa--r----------   110
MSV-Reu    ----------------------------i-------aa--r----------   110
```

Figure 3.8: Multiple sequence alignment of the first 110 aa of the CP of various MSV isolates showing the putative DNA binding domain (in bold) and nuclear localisation signal (underlined) (Liu, H. *et al.*, 1997) (DNAMAN 4.0). Identical aa residues are indicated by a dash (-) and different aa by a lower case letter. The C-terminal half of the protein (data not shown) only differs in one position with MSV-VM and MSV-SA sharing an A to T change, with respect to MSV-Tas.

The presence of the CP has been linked to the accumulation of ssDNA (Azzam *et al.*, 1994), and therefore it must be expressed at the appropriate time in the virus lifecycle. Two possible mechanisms exist for the regulation of the expression of the CP: first is the *trans*-activation of the virion sense genes, more extensively studied in the begomoviruses;

and second is the use of introns (Groning *et al.*, 1994; Sunter and Bisaro, 1992; Hofer *et al.*, 1992; Luehrsen and Walbot, 1991; Sunter and Bisaro, 1991; Callis *et al.*, 1987). A typical plant intron, having a high U + A content relative to the adjacent exon sequences, has been identified in the V1 ORF for MSV-N severe isolate, DSV, PanSV-Ken, WDV, SSV and TYDV (Wright *et al.*, 1997). A series of T tracts have been identified in the DNA sequence of the 76 nt long intron in MSV-Tas, MSV-VW and MSV-VM. Typical plant intron features, such as the consensus donor and acceptor sites and the "lariat acceptor" sequence, the intron's A + T-rich sequence relative to the flanking exons and a series of T-rich nucleotides, have been identified (see Fig. 3.9 & 3.10 and Table 3.5) (Ko *et al.*, 1998; Wright *et al.*, 1997; Luehrsen and Walbot, 1994). These features correspond to those found for the MSV-N virion sense intron found within the V1 ORF, which has been shown experimentally to be inefficiently spliced (Wright *et al.*, 1997). The presence of this intron is therefore a possible mechanism for regulating the expression of the MP thoughout the late cycle of gene expression (Wright *et al.*, 1997). No spliced virion sense transcripts have been detected for WDV and TYDV, while the expression of the CP of WDV is potentially due to a ribosomal slippage mechanism. This was proposed due to the presence of rare codons for arginine and leucine 5' of the methionine initiation codon of the CP of WDV, and the lack of full consensus sequences for the intron splice junctions (Wright *et al.*, 1997; Dekker *et al.*, 1991).

Table 3.5: Nucleotide composition (DNA) of the 5' and 3' 40 nucleotides flanking the putative virion sense and complementary sense introns.

Isolate	Virion Sense[1] %T			Complementary Sense[2] %T		
	5' Exon	Intron	3' Exon	5' Exon	Intron	3' Exon
MSV-Tas	18	46	18	28	32	10
MSV-VW	18	47	18	25	28	10
MSV-VM	18	45	18	28	30	18
	Virion Sense[1] % A+T			Complementary Sense[2] % A+T		
MSV-Tas	32	62	46	58	58	53
MSV-VW	36	63	46	55	55	53
MSV-VM	36	61	46	56	53	58

[1]: See Figure 3.9 for the virion-sense nucleotide sequence
[2]: See Figure 3.10 for complementary-sense nucleotide sequences

```
          200                                          240
           |                                            |
MSV-Tas   TCCGACATCCGGAGGAGTGTCGTGGAGTCGCGTCGGCGAGgtagctattttgagcttt
MSV-VW    ------------------------------a-----------------------
MSV-VM    ------------------------------a-----------------------

                                                                 315
                                                  YTNAN           |
MSV-Tas   gttggattgatttgcttttacctgctttacctttgggtgctgagagatcttatcttag
MSV-VW    ----ct------at-------------------------------------------
MSV-VM    ----c----------------------------------------c----------

                                              355
                                               |
MSV-Tas   TTCTGAAGGCTCGACAAGGCAGATCCACGGAGGAGCTGAT
MSV-VW    -----------------------------------
MSV-VM    -----------------------------------
```

Figure 3.9: Identification of the DNA sequence of the probable virion sense introns of MSV-Tas, MSV-VW and MSV-VM. The intron (76 nt) sequence is represented in lower case letters with the 5' donor site and the 3' acceptor site underlined. The nucleotide sequence of the flanking 5' and 3' exon sequences are shown in capital letters. The T nucleotides of the intron are shown in bold. The relative position of the intron to the whole genome sequence is indicated above the sequence. Differences in sequence between the three isolates are shown with a lower case letter under the corresponding base in MSV-Tas, while conserved sequences are indicated by a dash (-). The consensus branch point sequence YTNAN is indicated above the sequence of MSV-Tas in italics (Wright *et al.*, 1997).

3.4.4.3: Complementary sense ORFs

The viral protein(s) essential for replication is the product of the complementary sense genes, in particular the C1 ORF (RepA gene) and the spliced product of the C1/C2 ORF or Rep gene (Schalk *et al.*, 1989). Not all mastreviruses have an ATG start codon for their C2 ORFs, and although WDV (start codon GUG, Lazarowitz, 1988), SSV, PanSV-Kar, and CSMV have a C2 start codon, they all have an in frame stop codon three codons later (Peterschmitt *et al.*, 1996). While both MSV-VW and MSV-VM have start codons in similar positions to the other sequenced MSVs, MSV-Tas is unique as it has a stop codon three codons later (see Fig. 3.10).

The necessity for the spliced and unspliced products of the complementary sense ORFs, RepA and Rep respectively, in the virulence of MSVs has been well proven (Wright *et al.*, 1997; Collin *et al.*, 1996). Both the DSV and WDV complementary sense introns have an A + U content of approximately 54 %, and both are inefficiently spliced, but splicing is essential for replication (Dekker *et al.*, 1991; Mullineaux *et al.*, 1990; Schalk *et al.*, 1989).

Further evidence of the inefficiency of splicing of the complementary sense RNA was obtained for MSV, where 80 % of the transcripts are unspliced and presumably produce RepA (Wright *et al.*, 1997). The MSV intron is 92 nt long, the length of which is conserved amongst the three isolates presented here (see Fig. 3.10) (Wright *et al.*, 1997; Schalk *et al.*, 1989).

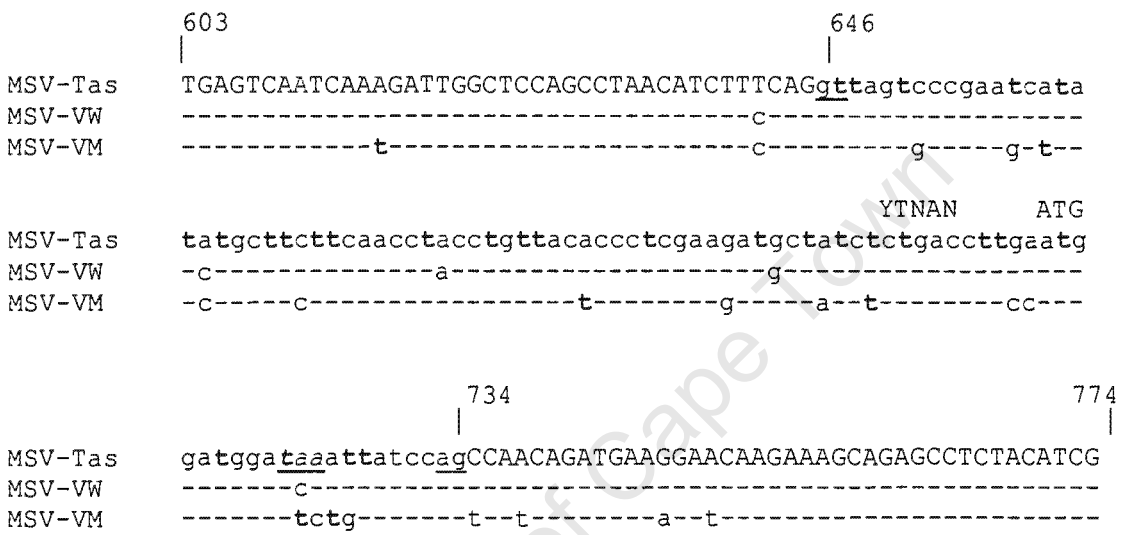


Figure 3.10: DNA sequence of the complementary sense introns and flanking exon sequences of MSV-Tas, MSV-VW and MSV-VM. The 40 nt flanking 5' and 3' sequences are shown in upper case, while the intron sequence (92 nt) is in lower. The 5' splice donor site and 3' splice acceptor sites are underlined and the branch point sequence (YTNAN) is indicated above the sequence of MSV-Tas in italics (Wright *et al.*, 1997; Schalk *et al.*, 1989). Sequences identical to MSV-Tas are indicated by a dash (-) and those that are different by lower case letters. The T nt of the intron are indicated in bold. Numbering above the sequence indicates the relative position of the intron with respect to the start codon of the C1 ORF, while the start codon (ATG) of the putative C2 gene product is indicated above the sequence. MSV-Tas has a stop codon (TAA) three after the C2 start codon (underlined italic).

The Rep protein:

The protein products of the complementary sense genes Rep and RepA are essential for infectivity, with the former being necessary for replication, while the latter possibly interacts with cell cycle regulators and may have a role in *trans*-activation of the virion sense gene expression (Horváth *et al.*, 1998; Xie *et al.*, 1996; Collin *et al.*, 1996; Xie *et al.*, 1995; Hofer *et al.*, 1992). Synonymous with its role in replication, the geminivirus Rep protein has a number of conserved motifs found in all rolling circle replication initiator proteins (see Fig. 1.4) (Koonin and Ilyina, 1993; Ilyina and Koonin, 1992; Koonin and Ilyina, 1992; Gorbalenya and Koonin, 1989). Although the function of Motif I is unknown, it is strongly conserved amongst all geminivirus (Ilyina and Koonin, 1992). Part of this motif's consensus sequence, FLTYP, is conserved amongst the MSVs, CSMV, TYDV, BeYDV, WDV and MiSV (see Table 3.6). Of the two classes of RCR initiator proteins, geminiviruses and other Rep proteins have a putative metal ion coordination motif II, containing two histidine residues surrounded by bulky hydrophobic residues, located on the N-terminus side of their nicking motif III. This situation is reversed in the second class of proteins, those that are involved in plasmid mobilisation (Koonin and Ilyina, 1993; Ilyina and Koonin, 1992).

The nicking motif III has a tyrosine residue to which the 5' end of the virion sense DNA becomes covalently attached during replication (Laufs *et al.*, 1995b; Laufs *et al.*, 1995c). Mutations to the tyrosine or lysine residue within the motif rendered the BGMV Rep incapable of initiating replication (Hoogstraten *et al.*, 1996). The nicking and joining functions can be supplied by the WDV RepA alone or the N-terminal region of the TYLCV Rep; however, the presence of motif IV, the NTP binding sequence, within the C-terminal half of Rep is essential for *in vivo* replication (Desbiez *et al.*, 1995; Hanson *et al.*, 1995; Heyraud-Nitschke *et al.*, 1995). Motif IV has two separate sites, A and B, which are both proceeded by a stretch of hydrophobic residues (see Table 3.6) (Gorbalenya and Koonin, 1989).

A further domain, with as yet an unconfirmed role in the virus lifecycle, is the DNA binding domain with some homology to the *myb*-related plant transcription factors. This domain overlaps with the NTP A and B sites (Hofer *et al.*, 1992). The identified MSV-N RepA transcriptional activation motif MLLQPTCYTL (aa 225 – 226) is conserved

amongst the three viruses sequenced here and MSV-N (data not shown; Horváth *et al.*, 1998). A similar motif in Rep (aa 252 – 270) has been identified - and is also conserved amongst these four viruses (data not shown) - it is thought to require a conformational change brought about by interactions with host proteins for exposure of the activation domain (Horváth *et al.*, 1998).

A motif present only in of most the mastrevirus Rep and RepA proteins is the putative Rb binding domain, with the sequence LXCXE (see Table 3.6) (Collin *et al.*, 1996; Xie *et al.*, 1995). It has been suggested that the interaction of RepA with Rb may regulate the cell cycle, while Rep may interact only under certain conformational conditions (Horváth *et al.*, 1998; Collin *et al.*, 1996). The LXCXE motif has not been found in curto- nor begomoviruses as well as some mastreviruses (SSV, CSMV and MiSV); however, an interaction of TGMV Rep with a maize type Rb (RRB1) was detected, suggesting that there are alternative Rb binding domains (Ach *et al.*, 1997; Xie *et al.*, 1995 & see Table 3.6). Neither of the two Rep sequences of SSV available (see Table 3.1 for GenBank accession numbers) display the classical LXCXE motif; however, the flanking amino acids around the identified motif for SSV-E closely resemble those of PanSV-Ken (Xie *et al.*, 1995; Briddon *et al.*, 1992) (see Table 3.6). Attempts at agroinfection of the cloned SSV-N isolate into maize were never successful, possibly due to the lack of Rb binding domain (E. P. Rybicki, pers. comm. & Hughes *et al.*, 1993).

The Rep proteins of MSV-Tas, MSV-VM and MSV-VW display all the consensus sequences essential for their activity as rolling circle replication proteins, sharing a high degree of identity to the other sequenced MSV isolates (see Table 3.6). The domains responsible for specific DNA binding identified in begomo- and curtoviruses are currently being defined in mastreviruses (Horváth *et al.*, 1998; Orozco *et al.*, 1997; Gladfelter *et al.*, 1997; Settlage *et al.*, 1996; Choi and Stenger, 1995). The MSV-N predicted protein-protein oligomerisation domains are present in both Rep and RepA (aa 175 – 187) and are conserved amongst the three viruses sequenced here and MSV-N (data not shown).

Table 3.6: Multiple sequence alignments of the Mastrevirus motifs I, II, III and Rb present in both RepA and Rep. The motif IV A and B sites are only present in Rep.

Viruses	Motif I	Motif II	Motif III	Rb Motif	Motif IV "A" Site	Motif IV "B" Site
MSV-VM	(18) FLTYPKCP	(58) SLHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESINDW	(224) SLYIVGPTRTGKSTWARSLGV	(262) IYNIVDDIPFKF
MSV-SA	(18) FLTYPKCP	(58) SLHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESINDW	(224) SLYIVGPTRTGKSTWARSLGV	(262) IYNIVDDIPFKF
MSV-Kom	(18) FLTYPKCP	(58) SLHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESINDW	(224) SLYIVGPTRTGKSTWARSLGV	(262) IYNIVDDIPFKF
MSV-N	(18) FLTYPKCP	(58) SLHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESINDW	(224) SLYIVGPTRTGKSTWARSLGV	(262) IYNIVDDIPFKF
MSV-K	(18) FLTYPKCP	(58) SLHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESINDW	(224) SLYIVGPTRTGKSTWARSLGV	(262) IYNIVDDIPFKF
MSV-Reu	(18) FLTYPKCP	(58) SLHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESINDW	(224) SLYIVGPTRTGKSTWARSLGV	(262) IYNIVDDIPFKF
MSV-VW	(18) FLTYPHCP	(58) DMHLHALIQ	(97) VRDYILKEPL	(193) PSSPDLLCNESIKDW	(224) SLYIVGPTRTGKSTWARSLGL	(262) IYNIVDDIPFKF
MSV-Tas	(18) FLTYPHCP	(58) DMHLHALLQ	(97) VRDYILKEPL	(193) PSSPDLLCNESIKDW	(224) SLYIVGPTRTGKSTWARSLGL	(262) IYNIVDDIPFKF
MSV-Set	(18) FLTYPQCP	(58) NMHLHALIQ	(97) VRDYITKEPL	(193) TSEPDLLCNESIKDW	(224) SLYIVGPTRTGKSTWARSLGR	(262) IYNIVDDIPFKY
SSV-N	(24) FLTYSRCP	(64) GYHIHVLAQ	(103) VRAYAMKNPV	(169) EWATKLQYFEYSANK	(228) SLYILGPTRTGKSSWARSLGR	(266) EYNIIDDIPFKY
SSV-E	(24) FLTYSKCH	(64) SWHIHALAQ	(103) VKEYVLKNPI	(198) PTEPDLVNFETIEDW	(228) SLYILGPTRTGKSTWARSLGR	(266) VMNIIDDIPFKF
PanSV-Ken[1]	(25) FLTYSKCP	(65) TWHCHALLQ	(104) VREYILKDPK	(199) TTEYDLHCNETIEDW	(229) SLYICGPTRTGKTSWARSLGR	(267) QYNVVDDIPFKF
PanSV-Kar[1]	(24) FLTYSKCP	(66) TWHCHALLQ	(105) VREYILKDPK	(200) ATDPDLLCNETLQDW	(230) SLYIVGPTRTGKTSWARSLGR	(268) AYNVVDDIPFKF
DSV	(15) FLTYSKCD	(55) SLHSHALVQ	(94) VRTYILKNPV	(188) PSEPDLLCQETITDW	(218) SLYILGPTRTGKSTWARGLGR	(256) QFNVIDDIPFKF
CSMV	(42) FLTYPRCP	(82) EPHLHAFVQ	(121) TLKYCMKHPE	-[3]	(235) SLYICGPTRTGKTSWARSLGT	(274) QFNIIDDIPFKF
TYDV	(19) FLTYPRCS	(58) TPHLHCLIQ	(98) VLEYISKDGN	(189) FTEEDLQCHEDLQLW	(221) SLYICGPSRTGKTTWARSLGR	(249) TYNVIDDIPFKF
BeYDV	(16) FLTYPKCS	(55) TTHYHALIQ	(95) VLDYISKDGD	(175) FTESDLRCHEDLHNW	(208) SLYICGPTRTGKTTWARSLGR	(246) TYNIIDDIPFKF
WDV	(17) FLTYPECT	(67) SPHLHVLVQ	(103) VRDYITKEVD	(189) FPTESLICHETIESW	(218) SIYICGPTRTGKTSWARSLGT	(256) KYNIIDDIPFKF
MiSV	(29) FLTYPHCN	(69) DPHLHVLIQ	(108) VFGYISKTNG	-[3]	(222) SIYICGPSRTGKTTWARNIGR	(260) IYNVIDDVPFKF
	**** *	* * *	* * *	* *	* ** * **** ***	* *** ***
Consensus[2]	Futltxxx uyp	xpHuHuuux u a	uxxYuxKxxx h	LxcxE	xuuuxgxxgxgKsxxxxxuxx a t	xxuuudexxxxx ed

Alignments were generated in CLUSTALW using the Rep protein sequence derived from the various sequenced viruses (see Table 3.1; isolates in bold were sequenced for this study). Numbers in brackets indicate the position of the first residue of the motif relative to the N-terminus of Rep. [1]: Neither SSV-N nor SSV-E have an exactly matching Rb motif sequence. An alternative site has been identified for SSV-N at starting at position 238 with the sequence ATSDLIWHETSRTL (Xie *et al.*, 1995). The aa best fitting with the Rb motif are indicated. [2]: Consensus sequences for Motifs I – III identified by Ilyina and Koonin (1992), Motif IV by Gorbalenya and Koonin (1989) and the Rb motif by Xie *et al.* (1995). The conserved residues are indicated in uppercase, while lower case are most likely present. u = are bulky hydrophobic residues (I, L, V, M, F, Y & W). x = no consensus. Residues under the consensus sequence are alternative identifying residues. * = conserved residues for all Mastreviruses. [3]: Neither CSMV nor MiSV have the matching aa consensus sequence (Xie *et al.*, 1995).

3.4.4.4: Intergenic regions

Long intergenic region

Alignments of the LIRs of the *Gramineae*-infecting mastreviruses have shown that the only conserved regions are the loop of the stem-loop structure and the TATA boxes; furthermore, phylogenetic analysis has shown that this region is hypervariable compared to the rest of the genome (Rybicki, 1994; Hughes *et al.*, 1992; Mullineaux *et al.*, 1990). The hypervariablity of the intergenic region extends to the begomoviruses (Faria *et al.*, 1994). The stem-loop sequences of the three viruses sequenced here, except for one nucleotide 5' of the invariant sequence, are absolutely conserved (see Fig. 3.11). The Rep proximal iteron sequence and its relative position with respect to the TATA boxes are also conserved. The presence of T tracts, possibly responsible for DNA bending within the LIR, has been identified, and these vary slightly between the wheat and maize viruses sequenced here (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995). More in depth analysis of the LIR sequences is discussed in Chapter 4 and 5 with respect to sequence requirements for replication and replication specificity amongst MSVs.

Short intergenic region

The mastrevirus complementary sense origin of replication is located in the SIR, where a variable sized DNA primer initiates second strand synthesis in MSV, DSV, TYDV, CSMV and WDV (Morris *et al.*, 1992; Hayes *et al.*, 1988; Andersen *et al.*, 1988; Donson *et al.*, 1987; Howell, 1984; Donson *et al.*, 1984). This region also contains the polyadenylation signals ($^{A}/_{G}$ATAA) (Messing *et al.*, 1983) for the virion and complementary sense genes and secondary structural elements of unknown function (Mullineaux *et al.*, 1984; Donson *et al.*, 1984). A comparison of the SIR and flanking C2 sequence of MSV-Tas, MSV-VW and MSV-VM with those of other sequenced MSV isolates is presented in Figure 3.12. The second strand primer binding site is highly conserved amongst the maize infecting MSVs, while the two wheat isolates and the grass isolate, MSV-Set show greater diversity. The C2 ORF secondary structural elements are highly conserved amongst all the isolates, differing only slightly in the stem length and loop sequence of the 3' structure. The function of these elements still remains to be elucidated, but considering the strong conservation amongst all the isolates, their role in the virus life-cycle may be significant.

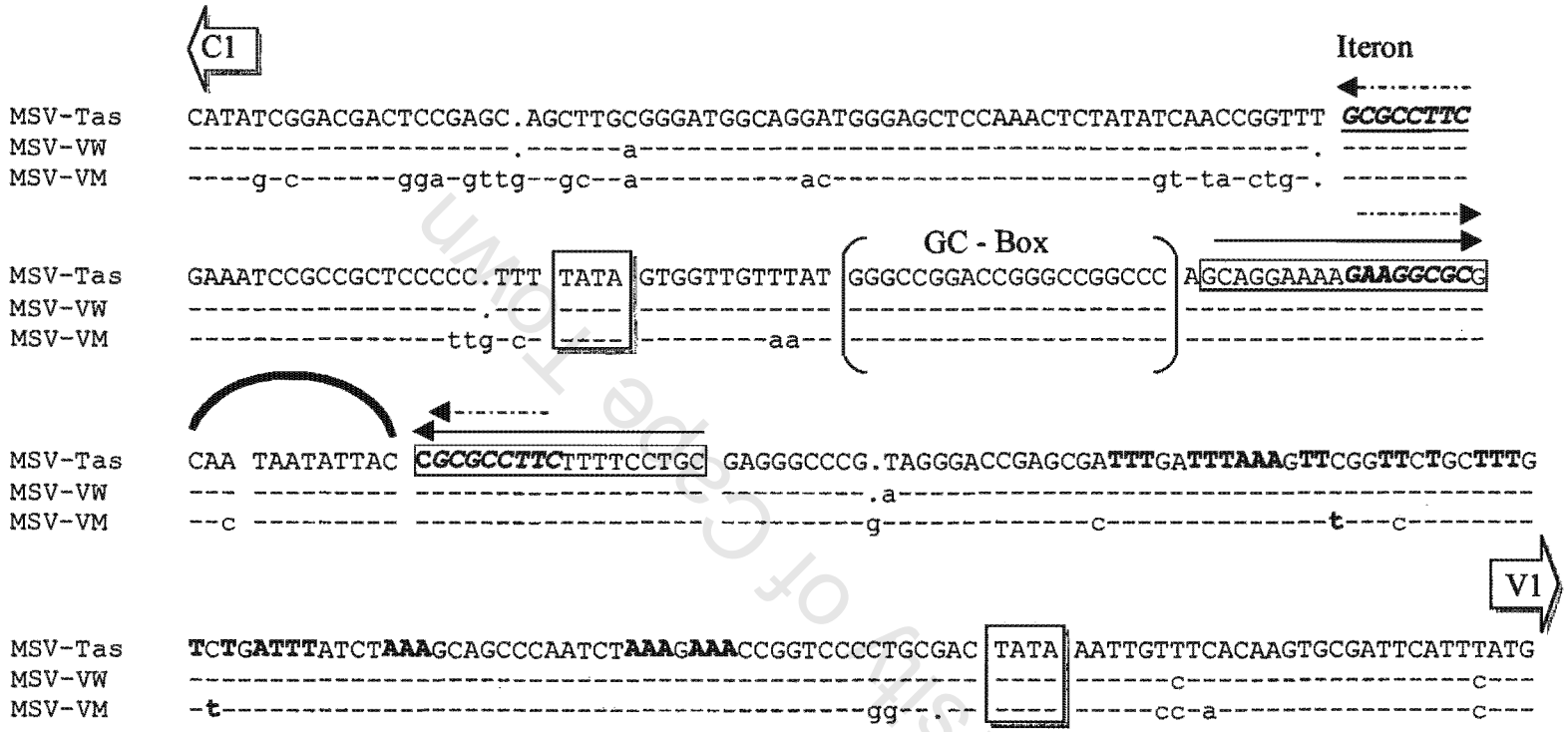


Figure 3.11: Multiple sequence alignment of the LIRs of MSV-Tas, MSV-VW and MSV-VM. The bi-directional transcription of the complementary and virion sense genes is indicated by the large arrows above the respective start codons (Morris-Krsinich *et al.*, 1985 & Dekker *et al.*, 1991). The transcription TATA boxes are highlighted by a box shape, while the presence of a GC-box element (also known as the *rpeI* element) essential for the virion sense promoter is bracketed (Fenoll *et al.*, 1988 & Fenoll *et al.*, 1990). The structurally conserved stem-loop is indicated by the smaller box and arrows around and above the stem sequence, while the loop sequence containing the invariant TAATATTAC sequence is indicated by an arc. The iterons, thought to be involved in replication specificity, are in bold italics, with the direction of the repeat highlighted by an arrow (Argüello-Astorga *et al.*, 1994 a & b). A series of AT tracts shown to be involved in DNA bending of WDV are presented in bold (Suárez-López *et al.*, 1995 & Gutiérrez *et al.*, 1995). All sequences identical to MSV-Tas are indicated by a dash (-) and those that are different are indicated by a lower case letter.

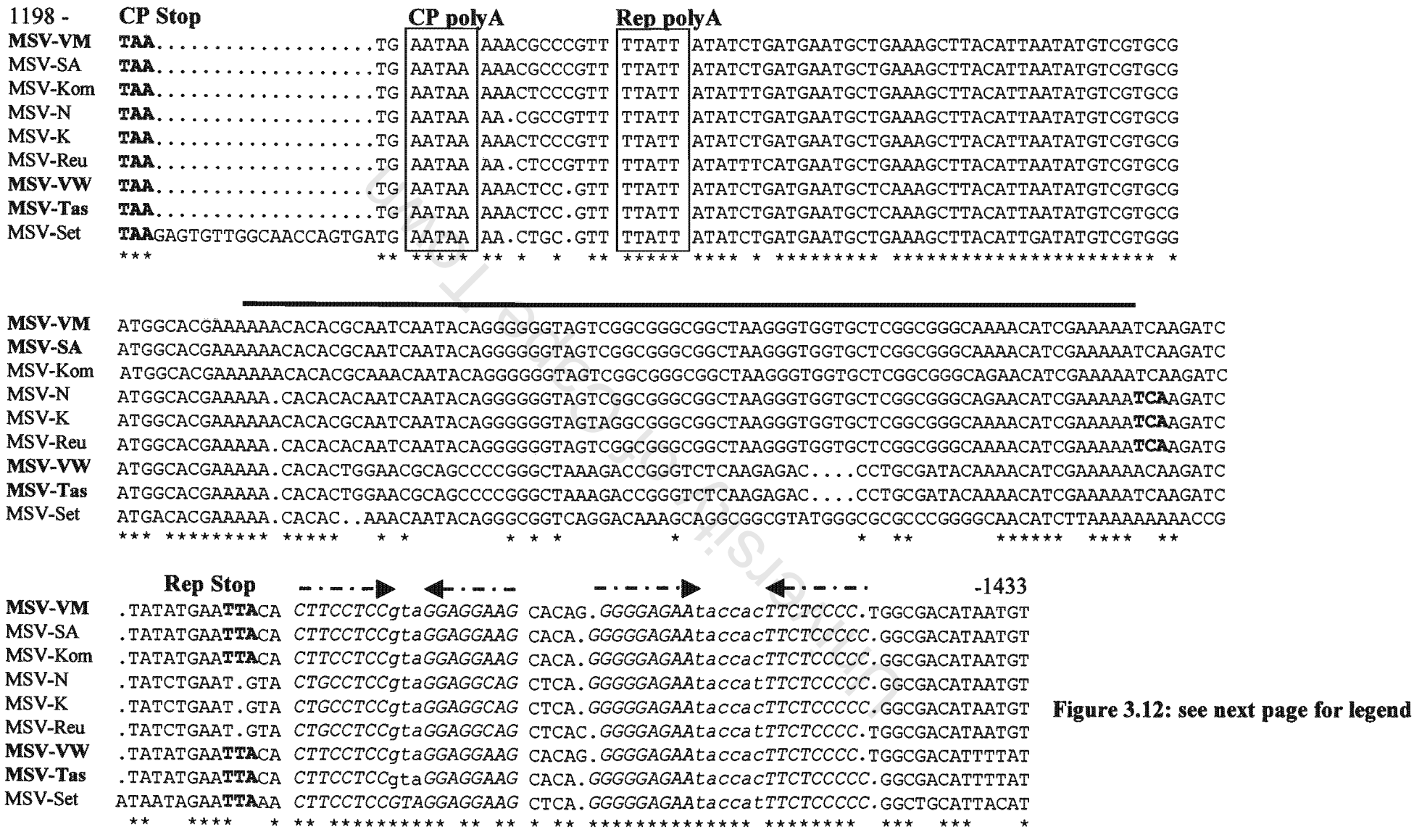


Figure 3.12: see next page for legend

Figure 3.12: Multiple sequence alignments of the virion sense sequence of the SIR and flanking CP and RepB gene sequences. The stop codons of the genes are shown in bold, while the polyadenylation signals are boxed. The MSV-Reu, MSV-N and MSV-K C2 stop codon is five codons after the other MSVs (Peterschmitt *et al.*, 1996). The putative primer binding site for dsDNA synthesis is indicated by a solid line above the sequence (Mullineaux *et al.*, 1984). Secondary structural elements, within the C2 coding region, are indicated in italics (Donson *et al.*, 1984). The stem sequences are shown by a broken arrow with the putative loop sequence in lower case. Conserved residues in the sequence alignment are indicated by an asterisk (*) and dots (.) are indicated where gaps were inserted for optimal alignment. (see Table 3.1 for the sequence accession numbers). The nucleotide sequence positions of the transcription products are shown in Table 3.4 and sequence positions are numbered with respect to MSV-VM.

3.4.5: Phylogenetic analysis

The overall sequence identity between MSV-VW and MSV-Tas is 98.1 %, while MSV-VM shares 89.9 % sequence identity with both MSV-VW and MSV-Tas. The homology between MSV-VW and MSV-Tas is similar to those shared between the maize isolates MSV-SA and MSV-N (97.8 %) (see Table 3.8). Prior to this study and the one performed by Rybicki *et al.* (1998), MSV-Reu was the most distinct of the characterised MSVs, and was more closely related to viruses from the neighbouring islands than it was to the African streak viruses (Peterschmitt *et al.*, 1996; Briddon *et al.*, 1994). The more recently sequenced MSV-Set is now the most divergent MSV, sharing +/- 84 % CP aa sequence identity with the larger group of maize infecting MSVs, and 64 % or less identity with the other African streak viruses (Schnippenkoetter, 1998; Table 3.7). The inclusion of MSV-Set (previously known as SetSV) into the MSV group of viruses has extended the range of diversity within this group considerably: the designation of the virus as an MSV was due to MSV-Kom being able to *trans*-replicate MSV-Set (Schnippenkoetter, 1998; Hughes *et al.*, 1992). While the grass and wheat subtypes of MSV have been observed previously, there was no CP sequence available to compare with previous studies (Rybicki *et al.*, 1998; Briddon *et al.*, 1994; Hughes *et al.*, 1992; Chapter 2). The CP sequence of MSV-Tas and MSV-VW indicates that they are intermediate in variation between MSV-Set and the maize infecting MSVs, with greater than the maximum 2 % CP amino acid divergence from the latter, previously observed by Briddon *et al.* (1994) for maize-type MSVs (see Fig. 3.13A and Table 3.7).

Phylogenetic analysis of the MP sequence for mastreviruses indicated that these group significantly when compared to the equivalent ORF of other geminivirus genera (Rybicki, 1994), even though the MP is the least conserved ORF within the *Mastrevirus* genus (Hughes *et al.*, 1992). Interestingly, the relationship dendrogram based on the MP amino acid sequence indicates that the MSV-VM MP is not part of the larger maize-type MSV subgroup (see Fig. 3.13B): this may be evidence of a recombination event. Analysis of the amino acid sequences indicated that differences amongst the three viruses occurred in similar regions, when compared to the N-terminal region of MSV-N (see Fig. 3.7).

Phylogenetic relationships based on the Rep sequence differentiate geminiviruses into two major groups: these are the mastreviruses, and a second group containing both curto- and begomoviruses (Rybicki, 1994). Within the mastrevirus group a distinction can be made between those viruses belonging to the African streak cluster and the other members of the genus (Rybicki, 1994). The recent discovery of BeYDV, a dicot-infecting virus means there is a dicot subgroup of TYDV and BeYDV as nearest neighbours within the mastrevirus genus. The two wheat infecting MSVs, MSV-Tas and MSV-VW, are each others closest relatives based on whole genomic nucleic acid sequence and Rep amino acid identity (see Fig. 3.13 A & B). MSV-VM consistently clusters with the maize-type MSVs, all sharing 96 % or greater Rep protein identity. The wheat-type MSVs MSV-Tas and MSV-VW are only slightly more similar to the maize-type MSVs (88 %) than they are to the sequenced grass isolate MSV-Set (86 %) (see Table 3.8).

The recent identification of the grass and wheat subtypes of MSVs, and the range of sequence data now available, indicates that a far greater diversity of MSVs exists than was originally considered to be the case (Rybicki *et al.*, 1998; Hughes *et al.*, 1992; Chapter 2).

Table 3.7: Pairwise identity matrix of the Mastrevirus movement and coat protein amino acid sequences[1]

Virus Name	#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MSV-VM	**1**	100	94	93	93	93	94	94	93	79	33	49	48	50	50	20	33	29	28	19	13
MSV-SA	**2**	98	100	99	99	99	96	90	91	81	34	50	48	51	50	19	33	29	28	20	13
MSV-Kom	**3**	99	99	100	98	98	95	89	90	81	34	49	47	50	49	18	33	29	28	20	13
MSV-N	**4**	98	98	98	100	100	97	90	91	80	34	50	48	51	50	19	33	29	28	20	13
MSV-K	**5**	97	98	98	98	100	97	90	91	80	34	50	48	51	50	19	33	29	28	20	13
MSV-Reu	**6**	98	98	98	99	98	100	91	91	78	34	50	48	50	49	21	33	29	28	20	13
MSV-VW	**7**	95	95	96	96	96	96	100	95	77	32	49	48	49	49	19	33	29	28	20	13
MSV-Tas	**8**	95	95	95	95	96	95	99	100	77	32	47	46	50	50	19	31	31	27	19	13
MSV-Set	**9**	83	84	84	83	84	83	84	85	100	34	47	45	48	47	13	31	28	28	20	10
DSV	**10**	64	65	64	66	66	65	64	64	62	100	37	37	38	40	15	22	20	25	15	9
SSV-N	**11**	69	69	69	69	69	69	68	68	66	71	100	54	56	56	16	30	25	27	18	13
SSV-E	**12**	69	70	69	70	70	69	69	69	69	69	78	100	62	57	18	29	25	23	17	9
PanSV-Ken	**13**	69	69	69	69	69	69	68	68	69	71	75	74	100	87	18	31	32	25	19	16
PanSV-Kar	**14**	70	71	70	70	70	70	69	69	69	72	77	78	89	100	18	31	32	25	17	15
CSMV	**15**	43	43	43	43	43	43	44	43	45	45	44	47	45	44	100	16	18	13	12	7
TYDV	**16**	35	35	35	35	35	35	35	35	35	37	37	37	34	36	37	100	62	47	19	11
BeYDV	**17**	36	36	36	36	36	36	36	36	37	38	40	39	36	37	37	57	100	46	19	7
WDV	**18**	32	32	32	33	33	32	31	31	34	35	33	34	33	34	31	32	34	100	15	8
MiSV	**19**	30	30	30	30	30	30	29	29	32	31	29	30	30	30	31	32	32	35	100	16
BCTV-Wor	**20**	24	24	24	24	24	24	24	24	24	26	22	23	22	22	23	21	24	24	22	100

1: Right-hand part of the table above the left-to-bottom-right diagonal indicates the percentage movement protein amino acid identity. The left-hand section below the diagonal indicates the percentage coat protein amino acid identity. Percentages were generated using DNAMAN 4.0 optimal sequence alignment parameters for the multiple sequence alignment option (see Table 3.1 for the Genbank accession numbers). Percentages have been rounded down to the nearest integer. The shaded columns and rows indicate data generated based on the sequences presented in this study. BCTV-Wor is included for comparison with the curtoviruses.

Table 3.8: Pairwise identity matrix of the whole genomic nucleic acid sequence and the Rep amino acid sequence of Mastreviruses[1]

Virus Name	#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
MSV-VM	1	100	99	98	99	98	97	88	88	82	67	64	67	63	65	45	47	48	44	45	30
MSV-SA	2	98	100	98	99	98	96	88	88	83	67	64	67	63	66	46	47	48	44	45	31
MSV-Kom	3	98	99	100	98	98	96	87	88	82	67	64	67	63	65	46	47	49	45	45	30
MSV-N	4	97	97	97	100	99	97	87	88	83	67	64	67	63	65	46	47	48	44	46	30
MSV-K	5	97	98	97	98	100	97	87	88	82	67	64	67	63	65	45	47	48	45	46	30
MSV-Reu	6	95	95	95	95	96	100	88	89	83	68	64	67	64	66	45	47	49	44	46	30
MSV-VW	7	89	89	89	88	89	89	100	96	86	67	64	67	62	65	46	47	48	43	47	29
MSV-Tas	8	90	89	89	89	89	89	98	100	85	67	64	67	61	65	45	47	48	42	46	30
MSV-Set	9	79	79	79	78	79	79	80	79	100	67	65	67	65	67	47	46	47	43	45	30
DSV	10	63	63	63	63	63	63	63	63	61	100	62	68	63	67	44	46	48	45	46	32
SSV-N	11	59	59	59	59	59	60	58	59	58	57	100	73	64	65	45	44	45	43	42	31
SSV-E	12	61	61	61	61	61	62	61	61	60	59	68	100	70	69	45	46	48	42	45	32
PanSV-Ken	13	58	59	59	59	59	59	58	58	59	58	62	65	100	86	46	45	45	43	45	31
PanSV-Kar	14	60	60	60	60	60	60	59	60	59	58	63	66	87	100	45	45	45	44	44	31
CSMV	15	48	48	48	48	48	48	47	47	47	46	47	48	47	46	100	44	44	45	44	29
TYDV	16	45	45	45	45	45	45	44	44	44	43	44	44	43	43	43	100	76	46	47	34
BeYDV	17	45	45	45	45	45	45	45	45	45	45	44	44	43	44	42	65	100	45	45	33
WDV	18	42	42	42	42	42	42	42	42	42	43	43	43	43	43	41	51	50	100	47	33
MiSV	19	43	43	43	43	43	43	43	43	43	43	43	45	44	44	42	41	41	40	100	36
BCTV-Wor	20	34	34	34	34	34	34	33	34	33	33	33	34	34	34	33	34	33	33	33	100

1: Right-hand part of the table above the left-to-bottom-right diagonal indicates the percentage Rep amino acid identity between each virus. Left-hand and below the diagonal indicates the percentage identity based on full genomic nucleic acid sequence. The matrix was generated using the DNAMAN 4.0 optimal sequence alignment default parameters for protein and nucleic acid multiple sequence alignments (see Table 3.1 for the Genbank Accession numbers). Percentages were rounded down to the nearest integer. The shaded columns and rows indicate data generated based on the sequences presented in this study. The curtovirus BCTV-Wor was included for comparisons sake.

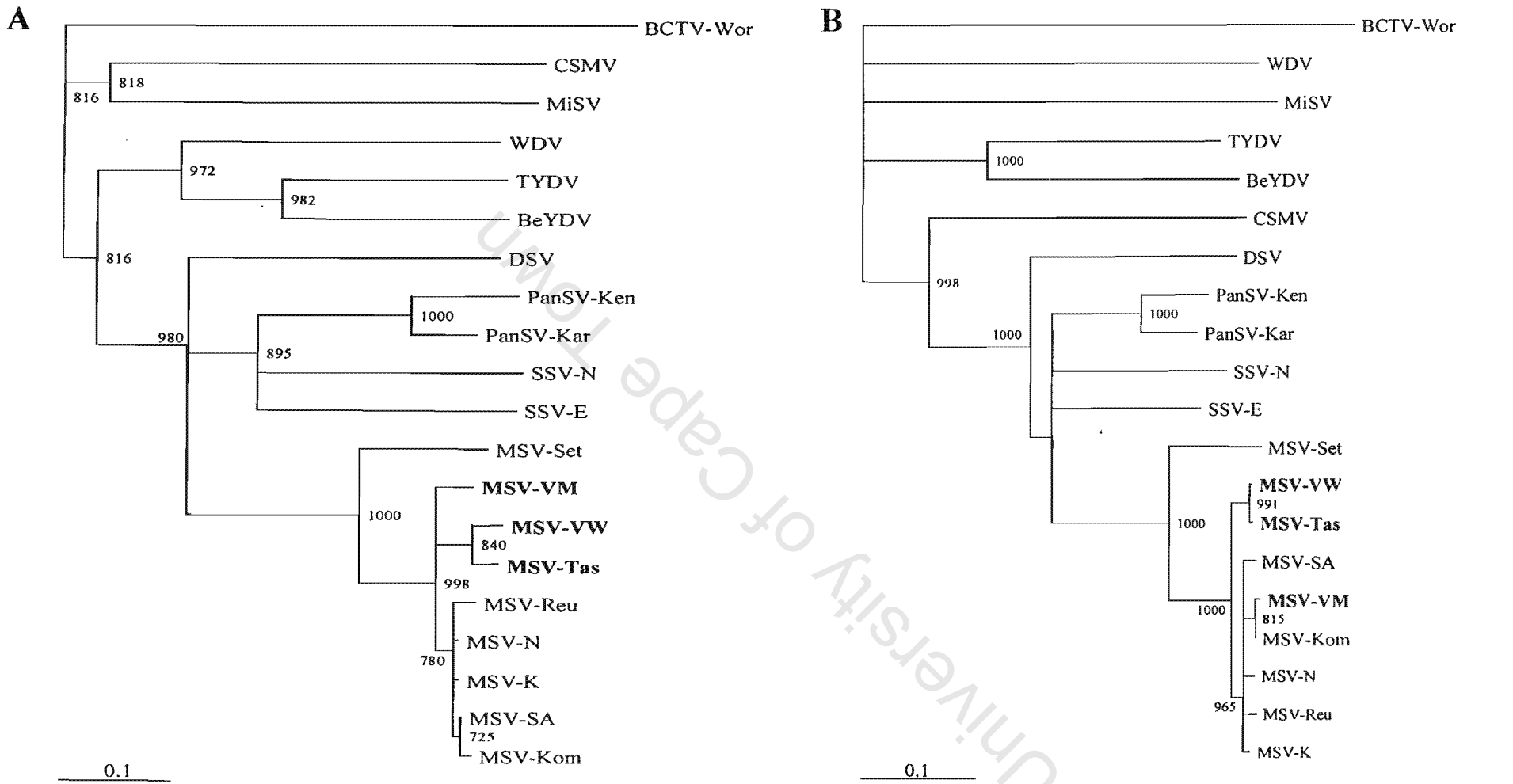


Figure 3.13: Relationship dendrograms generated from movement protein (A) and coat protein (B) amino acid sequence. Sequences were aligned using the GCG PILEUP program and the output file was used to create a phylogenetic tree, subjected to 1000 bootstrap trials, in CLUSTALW (Devereux *et al.*, 1984 & Thompson *et al.*, 1994). The tree was rooted with BCTV-Wor as the outgroup using TREEVIEW (Page, 1996) and all nodes with bootstrap values less than 700 were collapsed to form polytomies (E. P. Rybicki, Pers. Comm.). The bar indicates 10 % sequence divergence (i.e. a value of 0.1). All vertical distances are arbitrary, while horizontal distances are proportional to sequence divergence. Virus names in bold were sequenced during this study. See Table 3.1 for the GenBank accession numbers for all other viruses.

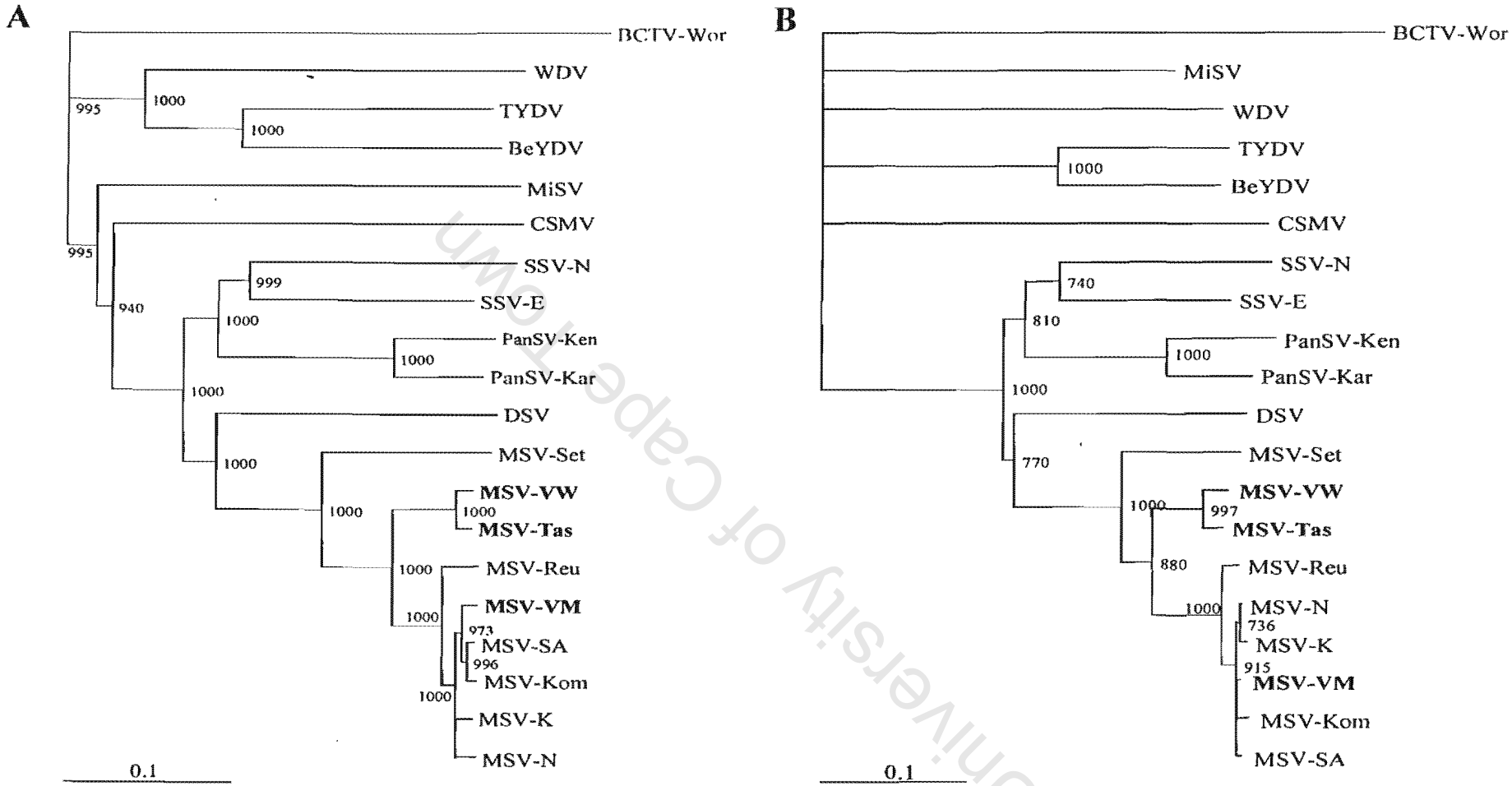


Figure 3.14: Relationship dendrograms generated from full genomic nucleic acid sequence (A) and Rep protein amino acid sequence (B). Sequences were aligned using the GCG PILEUP program and the phylogenetic tree, the product of a 1000 bootstrap trials, was generated using CLUSTALW (Devereux *et al.*, 1984 & Thompson *et al.*, 1994). The dendrogram was rooted using BCTV-Wor as an outgroup using TREEVIEW (Page, 1996). All nodes with bootstrap scores of 700 or less were collapsed to form polytomies (E. P. Rybicki, pers. comm.) The bar indicates 10 % sequence divergence (i.e. a value of 0.1). All vertical distances are arbitrary while horizontal distances are proportional to sequence divergence. Virus names indicated in bold were sequenced for the purpose of this study.

3.5: DISCUSSION

The full genomic sequences and partial host ranges have been obtained for two strains of MSV, MSV-VW and MSV-Tas. These were originally procured from infected wheat found in two different locations in South Africa, separated by over 1000 km. The genomic sequence of a third virus- MSV-VM, isolated from an infected maize plant from the same location as MSV-VW, was also obtained. Sequence data for all the viruses indicated the typical mastrevirus arrangement of ORFs, bi-directionally transcribed from the LIR and terminating in the SIR (Dekker *et al.*, 1991; Morris-Krsinich *et al.*, 1985). The putative translation products of the various ORFs were used in comparative analytical studies, the results of which indicated that MSV-VW and MSV-Tas form a distinct subgroup within the African streak virus group of mastreviruses (see Fig 3.7 & 3.8). Similar analyses of MSV-VM places this isolate firmly within the typical maize-infecting group, by all criteria except those based on MP sequence comparisons. It is possible that a recombination event occurred between a maize- and wheat-type MSV, arising in MSV-VM having a MP sharing more identity with the latter, but having a genome sharing greater identity with the former. There is circumstantial evidence that this event may have occurred: both MSV-VM and MSV-VW were collected from the same location in consecutive seasons (Rybicki *et al.*, 1998), and mixed populations of viruses are often present in maize, grasses and wheat (Chapter 2), providing a suitable environment for recombination. Viable intermolecular recombination between different species of geminiviruses, in particular, ToMoV and BDMV (Hou and Gilbertson, 1996) and East African cassava mosaic virus (EACMV) and ACMV (Zhou *et al.*, 1997; Hong *et al.*, 1993), is known to have resulted in an increase in pathogenicity of the recombinant virus. This has led to speculation that intermolecular recombination has a role in the evolution of geminiviruses (also see Bisaro (1994) for a review; Rybicki, 1994; Klute *et al.*, 1996; Briddon *et al.*, 1996a).

Functional protein domains and motifs identified in the various geminiviral proteins were conserved amongst all three sequenced viruses: all Rep proteins have the four motifs essential for their role in RCR and the Rb binding motif required for interaction with cell cycle regulators, and the MP hydrophobic *trans*-membrane spanning domain is present in all three proteins. Standard putative plant intron donor and acceptor splice sites were present in both the complementary and virion sense ORFs for all three viruses, indicating

shared post-transcriptional regulation of protein expression. Interestingly, the C2 ORF of MSV-Tas has a stop codon shortly after its putative start codon, similar to SSV, PanSV-Kar and CSMV (Peterschmitt *et al.*, 1996). Although the CPs are highly conserved amongst viruses sharing the same vector species, differences have been observed within in the putative DNA binding domain between the CP of the wheat isolates and the maize isolates (see Fig. 3.8) (Liu, H. *et al.*, 1997; Briddon *et al.*, 1994; Briddon *et al.*, 1990). This may reflect one of the CP functions that is virus specific, although *trans*-encapsidation and presumably also CP- DNA binding has been observed between distinct geminiviruses (Briddon *et al.*, 1990).

The putative determinants of replication specificity in mastreviruses - the iterons - are highly conserved between the three sequences presented here and those of all other MSVs, excepting MSV-Set and MSV-Raw (Schnippenkoetter, 1998; D. P. Martin, Pers. Comm & Chapter 2). Despite the conservation of the RSDs within the LIR, the Rep proteins of the two wheat viruses share only approximately 88 % identity with those of the maize viruses (see Table 3.8). Whether or not this is a reflection of host-adaptation in terms of Rep interaction with host factors, or that the amino acid sequence differences are within regions that are non-essential for function, remains unknown. Evidence exists of begomoviruses replicating but not producing systemic infections within certain hosts, implying that factors other than Rep-host interactions were limiting the host ranges (Hofer *et al.*, 1997; Buragohain *et al.*, 1994; Stenger *et al.*, 1992). Furthermore, analysis of Rep protein evolution determined that the outcome has probably not been directed by the host plant (Bradeen *et al.*, 1997).

Although a limited range of plant species were used in the leafhopper transmission studies, the pathogenicity of MSV-Tas and –VW in the wheat and barley cultivars was greater than the maize isolate, MSV-Kom (Schnippenkoetter, 1998). Conversely, they were significantly less pathogenic in maize than MSV-Kom, following the trend established by the grass isolate MSV-Set (Schnippenkoetter, 1998). I would like to propose that a uniform panel of plants be used in future for the proper biological characterisation of MSVs. This panel would include common grass varieties as well as cereal crops and must be of the same ecotype (Peterschmitt *et al.*, 1991). Unfortunately, due to the lack of grass isolates used in the leafhopper transmission studies, we are unable to establish whether or not these two wheat isolates are indeed similar to the grass isolates

in terms of host-adaptation to non-maize plants. All indications, in terms of the diversity study presented in Chapter 2, implied that the grass isolates may use wheat as an alternative host species. Phylogenetic analysis based on sequence data obtained from part of the C2 ORF also consistently groups the grass and wheat isolates together, apart from the larger group of characterised maize MSVs (Rybicki *et al.*, 1998; Hughes *et al.*, 1992): for this reason, and because of the clear biological differences between the wheat- and maize-infecting viruses, I have designated MSV-VM and –Tas as isolates of a distinct strain grouping (the wheat/grass strains) of MSV.

The presence of an intermediate group of viruses sharing only approximately 88 % sequence homology with the larger group of highly-related MSV isolates, allows for investigation using complementation of replication functions or the use of chimaeric proteins to establish the validity of using *trans*-replication as a criteria for relatedness (Rybicki, 1998; van Regenmortel *et al.*, 1997). These types of experiments will also allow elucidation of virus specific protein domains involved in genome replication, host-adaptation and pathogenicity, which have been lacking in MSV research to date.

Chapter 4

Investigation of the complementation of replication functions amongst Mastreviruses

4.1: SUMMARY **119**
4.2: INTRODUCTION **120**
4.3: METHODS AND MATERIALS **123**
4.3.1: Source of virus isolates and Rep proteins **123**
4.3.2: Construction of replication deficient MSV-Kom mutants **124**
4.3.3: Establishment of a transient replication assay system **127**
4.3.4: PCR confirmation of *trans*-replication **130**
4.3.5: Detection of protein expression **131**
4.3.6: Sequence analysis of RepA and the LIR **132**
4.4: RESULTS **133**
4.4.1: Construction of partial dimers of WT viruses **133**
4.4.2: Replication kinetics of WT viruses in BMS suspension culture cells **133**
4.4.3: *Trans*-replication of Rep mutant MSV-Kom based constructs **133**
4.4.4: Sequence analysis of the LIR and Rep proteins **141**
4.5: DISCUSSION **145**

4.1: SUMMARY

The extent to which mastreviruses display replication specificity is currently unknown; however, the ability to complement replication functions is thought to extend to those viruses sharing the same iterative elements within the long intergenic region (LIR). A transient replication assay system was established for the detection and quantification of mastrevirus *trans*-replication in a Black Mexican sweetcorn (BMS) cell line. The replication kinetics of MSV in BMS were assessed over a period of 2 to 7 days and found to plateau after 4 days. A range of maize streak viruses (MSV) and one panicum streak virus (PanSV-Kar) were chosen for *trans*-replication studies of the typical maize-type virus MSV-Kom: MSV-Kom, MSV-VW, MSV-Tas and MSV-VM share the same iterative elements, but their RepA proteins range in amino acid sequence identity with MSV-Kom from MSV-VM with 98 %, to MSV-VW with 87 %. MSV-Set and PanSV-Kar do not share the same LIR iterative elements with MSV-Kom, and their RepA proteins share 76 and 56 % sequence identity with that of MSV-Kom, respectively. These viruses were assessed for their ability to *trans*-replicate three different Rep defective MSV-Kom –derived constructs.

Co-bombardment of a CaMV 35S promoter and chloramphenicol acetyl tranferase (CAT) MSV-Kom C1/C2 gene replacement construct with the wild type viruses indicated that they were all able to complement replication of the MSV-Kom derived replicon, even though MSV-Set was very inefficient. *Trans*-replication studies of a β-glucuronidase (GUS) V2 replacement and C1 frameshift mutant of MSV-Kom, which had a luciferase expression cassette cloned in tandem, was done in an attempt to quantify changes in GUS expression due to an increase in genomic DNA resulting from complementation of replication functions. Results indicated that only the homologous Rep was able to significantly *trans*-replicate this construct above background levels. An unexpected result was observed: the two most heterologous viruses, MSV-Set and PanSV-Kar, inhibited the constitutive GUS expression from the virion sense promoter. The third Rep deficient construct co-bombarded with WT viruses into BMS cells was a CaMV 35S and GUS C1/C2 replacement construct. Assessment of the expression of GUS relative to that of the internal luciferase control indicated that all the MSVs *trans*-replicated MSV-Kom, although MSV-Set was slightly less proficient than the more homologous viruses; PanSV-Kar, the most heterologous virus, was unexpectedly also able to complement Rep

functions and *trans*-replicate MSV-Kom. Sequence alignments of the various LIRs revealed some conserved elements shared by all the viruses, which differ from the putative replication specificity elements previously identified.

4.2: INTRODUCTION

Geminivirus replication specificity determinants (RSDs) have been identified using begomovirus pseudorecombinants, chimaeric constructs and site directed mutants (Sanz-Burgos and Gutiérrez, 1998; Gladfelter *et al.*, 1997; Fontes *et al.*, 1994). The *cis*- and *trans*-acting elements essential for replication are in the process of being defined for begomoviruses, while similar research in curto- and mastreviruses has only recently started providing insight into these determinants, with considerable more data being available on the former (see Hanley-Bowdoin *et al.* (1997 and 1999) for reviews & Chapter 1).

The use of pseudorecombinants has been generally limited to isolates of a strain of a two component begomovirus, sharing common direct repeats called iterons within the left-hand side of the intergenic region (Faria *et al.*, 1994; Fontes *et al.*, 1994). Exceptions to this rule were the viable pseudorecombinants obtained from ToMoV and BDMV, which share most of the sequence requirements within the direct repeats (Gilbertson *et al.*, 1993). Furthermore, even if efficient *trans*-replication can occur using a certain A component and a heterologous B component, the converse is not necessarily true (Hofer *et al.*, 1997; Faria *et al.*, 1994). These direct repeats are only the primary determinants of begomovirus specific replication, as secondary determinants located between the repeats and the stem-loop are also required (Hou and Gilbertson, 1996; Fontes *et al.*, 1994). The correct spatial arrangement of these elements is also essential for efficient replication (Gladfelter *et al.*, 1997). The stem-loop sequence itself does not determine specificity, but the presence of a WT stem does enhance replication, possibly due to a REn specific interaction (Hanley-Bowdoin *et al.*, 1996; Fontes *et al.*, 1994).

Similar observations were made for the curtovirus BCTV, where the RSDs were initially mapped to within 97 bp of the intergenic region and later identified as a direct repeat, the sequence of which does tolerate small differences (Choi and Stenger, 1996; Choi and Stenger, 1995). Differences, both in sequence and length, between the direct repeat and

the stem-loop structures of BCTV isolates do not seem to act as secondary replication specificity elements, unlike the begomoviruses (Choi and Stenger, 1996; Fontes *et al.*, 1994).

A TLCV-associated satellite DNA, with the same Rep proximal iterons and the conserved geminivirus invariant loop sequence as TLCV, was *trans*-replicated with varying levels of success by TLCV, TYLCV, ACMV and BCTV. The putative Rep proximal iteron formed part of a small stem-loop structure approximately 280 bp upstream from the invariant TAATATAC loop sequence. This separation of *cis*-elements was found to be essential for *trans*-replication (Dry *et al.*, 1997). The *trans*-replication of the satellite virus is the first report of such a diverse range of viruses all being able to *trans*-replicate a heterologous origin, which does not share the same RSDs.

The theoretical identification of the RSDs for mastreviruses has yet to be conclusively challenged or confirmed (Argüello-Astorga *et al.*, 1994a & b). I have undertaken a study of the complementation of replication functions amongst mastreviruses sharing common RSDs, and have also included some which were more diverse. While this study was in progress, an agroinfectious construct of MSV-Set with a MSV-Kom LIR was made by Dr W. H. Schippenkoetter of this laboratory (Schnippenkoetter, 1998). This construct proved to be replication proficient, albeit highly attenuated, which indicates that these two viruses may not be sufficiently divergent for complete lack of recognition of each others RSDs. However, the putative iteron sequences are different between MSV-Kom and MSV-Set (Schnippenkoetter, 1998). Similarly, chimaeric constructs of WDV isolates with different iterons were replicatively released from hybrid stem-loop structures, and two WDV isolates with 16.4 % sequence diversity could complement replication functions (Heyraud *et al.*, 1993b & unpublished data cited in Bendahmane *et al.*, 1995). Chimaeric constructs of MSV and BeYDV (55 % sequence divergence, Table 3.8) with either C1/C2 or LIR exchanges were not replication proficient, suggesting that both *cis*- and *trans*-acting replication specificity elements differed considerably between these host adapted mastreviruses (Liu *et al.*, 1999). Studies using deletion mutants of the WDV LIR have shown that the minimal requirements for efficient replication are far greater than was previously expected, and were perhaps more similar to the other genera in the requirement for *cis*-elements (Sanz-Burgos and Gutiérrez, 1998). The same could perhaps be true of the *cis*-elements responsible for determining replication specificity.

Due to the necessity for all of the MSV ORFs to be expressed *in planta* for symptom development, I used microprojectile bombardment to simultaneously introduce a self-replication deficient construct and one providing Rep in *trans,* into Black Mexican sweetcorn (BMS) suspension culture cells, in order to assess complementation of replication functions. Similar approaches have used to study the expression of foreign genes as V2 replacements in WDV, and for SIR insertions in MSV (Shen and Hohn, 1994; Matzeit *et al.*, 1991). Transient expression assays have been used to investigate viral promoter activity in the presence of Rep or TrAP, where it was also observed that amplification of viral DNA also increased the transcription of viral mRNA (Brough *et al.*, 1992; Hanley-Bowdoin *et al.*, 1988). These studies used the expression of β-glucuronidase (GUS) linked to genome amplification as a means of quantifying viral replication. It was assumed that the increased levels of GUS were due to increased replication of the viral construct; i.e. that increased replication would result in the concomitant increase in transcription levels. Expression levels from replicating WDV V2 gene replacements constructs were enhanced 20 times compared to non-replicating constructs, indicating that perhaps mastreviruses could be used as gene vectors (Palmer and Rybicki, 1997; Timmermans *et al.*, 1994; Mullineaux *et al.*, 1992; Timmermans *et al.*, 1992; Hofer *et al.*, 1992; Matzeit *et al.*, 1991). To this end various dimers of MSV-Kom based constructs containing a MSV V2 translational fusion with GUS were made by K. E. Palmer and used to assay for the effect replication had on expression levels of the virion sense genes (Palmer, 1997). These experiments showed that increased GUS expression was linked to replication, similar to the findings of Suárez-López and Gutiérrez (1997), and that no *trans*-activation of the MSV-Kom virion sense genes occurred in BMS, a finding analogous to the lack of *trans*-activation activity of the begomovirus TrAP in callus tissue (Sunter and Bisaro, 1997; Palmer, 1997).

To date all *trans*-acting Rep replication specificity determinants have been identified within the N-terminal half of the protein, with the first 89 aa required for curtovirus specificity and the first 116 aa for the Rep protein of the begomovirus TYLCV (Jupin *et al.*, 1995; Choi and Stenger, 1995). More precisely, two specific domains of the begomovirus TGMV Rep have been identified as origin recognition determinants (Gladfelter *et al.*, 1997). No data are available on the Rep *trans*-acting determinants of mastrevirus replication. However, the WDV Rep protein is able to nick a TYLCV-like stem-loop structure *in vitro*, and *in vivo* can initiate replication from an upstream region

other than the stem-loop structure, indicating that these determinants may be more flexible than those in other geminivirus genera (Heyraud-Nitschke *et al.*, 1995; Heyraud *et al.*, 1993a).

4.3: METHODS AND MATERIALS

4.3.1: Source of virus isolates and Rep proteins

Our lab strain of MSV, MSV-Kom, was chosen for the *trans*-replication studies for two reasons: first, it had been completely sequenced, and second, its biological characterisation had nearly been completed when this study was initiated (Schnippenkoetter, 1998 & Schnippenkoetter *et al.*, unpublished data). A second isolate, MSV-Set, was in the process of being sequenced and characterised, which work has since been completed (Schnippenkoetter, 1998). MSV-Set was found to be a distinct strain of MSV, as had been previously predicted based on RFLP studies and part of the C2 sequence (Schnippenkoetter, 1998; Hughes *et al.*, 1992; Rybicki and Hughes, 1990 & Chapter 2). A range of viruses were selected for *trans*-replication studies. These shared the following percentage amino acid identity with the MSV-Kom Rep: MSV-VM 98 %; MSV-Tas 88 %; MSV-VW 87 %; MSV-Set 82 %, and PanSV-Kar 65 % (see Table 3.8). The Rep protein of the sequenced isolate of DSV Rep shares 67 % amino acid identity with that of MSV Kom (see Table 3.8) (Donson *et al.*, 1987), but our full agroinfectious dimeric clone of DSV, pDSV200 (Palmer, 1997), was separately isolated and cloned by Dr F. L. Hughes and is not sequenced (Hughes *et al.*, 1992; Rybicki and Hughes, 1990). A partial dimer of a chimaeric virus based on MSV-Set, which has its V1 and V2 ORF exchanged with those of MSV-Kom (pSV1V2Kδ, E. van der Walt, K. E. Palmer and E. P. Rybicki, unpublished data), was included in one of the *trans*-replication studies.

4.3.1.1: Construction of a WT PanSV-Kar partial dimer

The original agroinfectious clone of PanSV-Kar had been constructed as a partial dimer with only one LIR (Schnippenkoetter, 1998): therefore a 1.2 mer with two LIRs was reconstructed from the original monomeric clone pPS100, which is linearised at its unique *Bam*HI site within the C1 ORF (Hughes *et al.*, 1992). The plasmid pPS100 was digested

with *Xba*I before the larger of the two fragments, containing the 0.2 mer of PanSV-Kar and the pUC18 vector, was recovered from a 0.8 % TAE agarose gel using the Geneclean™ Kit (BIO101). The DNA was then self-ligated and the resultant clone named pPSV.1. The PanSV-Kar *Bam*HI fragment was recovered from the vector by digestion of pPS100 with *Bam*HI and *Pvu*I, allowing the larger vector band to be distinguished from the two smaller vector bands in an agarose gel. The *Bam*HI fragment was then ligated into a *Bam*HI and calf intestinal phosphatase treated pPSV.1 plasmid. The final plasmid containing the partial dimer was called pPSV1.1. All molecular manipulations of DNA were performed as recommended by Sambrook *et al.* (1989) and varied where necessary according to specific manufacturer's specifications.

4.3.2: Construction of replication deficient MSV-Kom mutants

4.3.2.1: A CAT marker gene replacement of the complementary sense genes

MSV-Kom-specific PCR primers were designed to amplify both virion sense genes as well as both the small and long intergenic region (designated as LVS):

LIR-Kom (5'-CTATCTAGACGACGACGGAGGTTGAG-3')

SIR-Kom (5'-CCGCTCGAGTAATTCATATAGATC-3') (Genosys, Cambridge, UK).

The primers contain the restriction sites *Xba*I and *Xho*I, respectively (underlined sequences) and were designed to amplify the region spanning from bp 2532, three nucleotides from the start codon of the C1 ATG (position 2528), to position 1372 (C2 stop codon at position 1198). This was achieved using DNA isolated from MSV-Kom agroinfected plants (see 3.3.4). A PCR master mix was made up as follows: 1.5 mM $MgCl_2$, 0.5 µM of each primer, 1 x PCR buffer, 0.5 units Taq (Applied Biotechnology) and 200 µM dNTPs (Boehringer Mannheim) per 100 µl reaction volume. The following cycling conditions were used on a waterbath-based JDI Model 8012 thermal cycler (JD Instruments, Cape Town): 94^0C for 60 seconds, followed by 30 cycles of 94^0C for 60 seconds, 60^0C for 60 seconds and 72^0C for 150 seconds, and finally 72^0C for 300 seconds. The PCR product was then ethanol precipitated, treated with Klenow polymerase (Boehringer Mannheim) and blunt cloned into the *Eco*RV site of the pBluescript(KS) multiple cloning site to produce the clone pKomLVS. This clone was sequenced in one

direction to ensure that no PCR generated errors had occurred (Mrs Di James, Department of Microbiology Sequencing Service, UCT).

A marker gene cassette (pSKCAT) was constructed in pBluescript(SK). The *Taq*I fragment containing the chloramphenicol acetyl transferase (CAT) gene, obtained from pBSG8-15 (gift to the department from Georges Rapoport, Institut Pasteur) was ligated into the *Cla*I restriction site of pBluescript(SK). The CaMV 35S promoter was obtained from pMF6 (Callis *et al.*, 1987) by digestion with *Pst*I and *Xba*I and ligation into the same restriction sites of pSKCAT to give pCamcat. A monomer backbone of the C1/C2 replacement cassette (pLVScc2) was then made by excising the LVS part of pKomLVS using *Xho*I and *Xba*I and then ligating the appropriate Geneclean™ treated fragment into a *Sal*I restriction endonuclease digested pSKCAT. After overnight incubation at 15^0C the remaining unligated ends were blunted using Klenow (Boehringer Mannheim), which allowed the ligation of the *Sal*I and *Xba*I ends. The polyadenylation signals for the CAT transcript are provided by the SIR (Dekker *et al.*, 1991).

A subclone containing the LIR was constructed by the digestion of pKomLVS with *Bam*HI, and the appropriate fragment was ligated into *Bam*HI and calf intestinal phosphatase treated pBluescript(KS). The resultant subclone, called pLIR1, contained flanking *Xba*I sites, one introduced by PCR primer Kom-LIR from clone pKomLVS, as well as the *Xba*I site from the vector. The monomeric clone, pLVScc2 was then partially dimerised by the insertion of the *Xba*I fragment from pLIR1 into the *Xba*I site flanking the CaMV 35S promoter. The resultant dimer was named pLVSccLIR.

4.3.2.2: Replication deficient GUS V2 replacement construct

Trans-replication of a Rep-mutant virus expressing β-glucuronidase (GUS) under the control of the virion sense promoter was used in an attempt to quantify the nature of the interaction between a heterologous Rep protein and origin of replication. To this end a transient replication assay comparing the expression of GUS relative to the expression of an internal control, the firefly luciferase gene (*luci*) (Ow *et al.*, 1986), was initiated. Linkage of these reporter genes onto the same plasmid would insure that experimental

variation would be kept to a minimum. All plasmids annotated with the letters KEP used in this study were constructed by Dr K. E. Palmer (Palmer, 1997).

A full tandem dimer of a β-glucuronidase (*gusA*) gene and *nos* 3' terminator sequence replacement of the V2 ORF of MSV-Kom, called pKEP171*gus*d, had been created by the replacement of the virion sense ORFs from pKEP171 with a V1, *gusA* and *nos* terminator cassette from plasmid pKEP176. The plasmid pKEP171 is a MSV-Kom based monomer with an insertion of a TG dinucleotide at position 2304, resulting in a frame shift mutation in the C1 ORF and created a *Pst*I site. Plasmid pKEP176 has a mutation at the V2 start codon, which was changed by site directed mutagenesis to a *Nco*I site, allowing the *gusA* and *nos* terminator from pGUSN358→S (Clonetech, Palo Alto, CA) to be inserted as a *NcoI* and *Eco*RI fragment and at the same time removing all other MSV sequences except the *Bam*HI to *Nco*I V1 gene containing fragment. I linked pKEP171*gus*d to a CaMV 35S promoter-driven luciferase expression cassette, pKEPluci, and named this plasmid p171luci. Construction of p171luci was achieved by the digestion of pKEPluci with *Bgl*II (which cuts at a unique site upstream of the CaMV35S promoter) which allowed the ligation of the MSV-Kom based dimer excised from the partially *BamH*I digested pKEP171*gus*d plasmid.

4.3.2.3: C1/C2 ORF replacement with a GUS expression cassette

The GUS gene was excised from pKEP176 by digestion with *Nco*I and *Eco*RI, treated with Klenow and ligated into an *Eco*RV digested pBluescript KS. The resultant plasmid with the start codon of *gusA* orientated on the *Sac*I side of the multiple cloning site, called pKSgus1, was digested with *Xho*I and *Pst*I to release the *gusA* and *nos* fragment allowing directional cloning into the *Xho*I and *Pst*I treated pLVSccLIR. This effectively replaces the 770 bp CAT gene with the approximately 2000 bp *gusA* and *nos* terminator, now under control of the CaMV 35S promoter. This partial dimer was named pLVScgLIR.

4.3.3: Establishment of a transient replication assay system

4.3.3.1: Maintenance of BMS suspension culture cells

The Black Mexican sweetcorn (BMS) suspension culture, provided by Pioneer Hi-Bred Inc. (Des Moines, Iowa), was maintained in 25 ml BMS medium in a 250 ml DeLong flask, and routinely sub-cultured once a week by a one-tenth dilution into fresh medium. BMS liquid media was made by the protocol of Murashige and Skoog (1962) using 4.324 g lyophilised MS salt powder (Highveld Biologicals, South Africa), 2 mg/ml 2,4-D (Sigma), 0.1 mg *myo*-inositol, 30 g sucrose and MS vitamins made up at pH 5.75 (Murashige and Shoog, 1962). For solid BMS media 8 g/l of agar (Sigma) was added to the above liquid media. The suspensions were maintained in the dark on a rotary shaker at 26^0C and 150 rpm, and three days prior to the introduction of DNA into the BMS cells by microprojectile bombardment, the suspension was diluted by one third.

4.3.3.2: Preparation of BMS cells for bombardment

Four hours prior to bombardment a 1 ml volume of the suspension was layered gently onto a Whatman #4 filter disk by vacuum filtration. The disks were then placed onto solid BMS media containing 0.2 M mannitol and 10 µg/ml $AgNO_3$ (Vain *et al.*, 1993; Armstrong and Songstad, 1993). The cells were maintained on this high osmoticum media for 24 hours after bombardment and were then returned to ordinary solid BMS media until harvested.

4.3.3.3: Particle bombardment of BMS suspension culture cells

A Du Pont PDS-1000/He particle gun was used to bombard each plate of prepared BMS tissue twice at 650 psi (all Bio-Rad components were used as per the manufacturer's instructions), using a 6 mm gap distance between rupture disk holder and macrocarrier, a 5 mm macrocarrier flight distance, and a particle flight distance of 6 cm. A vacuum of 27 in. Hg was maintained prior to firing the gun. The preparation of gold and precipitation of DNA onto the gold was performed according to the protocol of Dunder *et al.* (1995), similar to the procedure used by Xie *et al.* (1995) and Suárez-López & Gutiérrez (1997).

4.3.3.4: Precipitation of DNA

The replication kinetics of WT MSV-Kom and MSV-Set were assessed in BMS. Either pKom602 (1200 ng) or pSet105 (1200 ng) was added to a 3 mg gold particle suspension in 50 % (v/v) glycerol. While continuously vortexing the sample 50 µl of 2.5 M $CaCl_2$ and 20 µl 0.1 M spermidine (Sigma) were added. The gold preparation was then briefly centrifuged for three 4 seconds spins at 3000 rpm, washed and resuspended in 1 ml of absolute ethanol, centrifuged as before, and finally resuspended in 75 µl of ethanol (Dunder *et al.*, 1995). Each of six sterile macrocarriers was then coated with 12 µl of the gold suspension and allowed to air dry. Two precipitation reactions were performed for each plasmid, with six bombardments per precipitation achieved. Due to the high variability found between precipitation reactions, for this experiment 6 BMS plates were bombarded once with the first precipitation and a second set of 6 plates was bombarded with DNA from the second precipitation. One plate from each precipitation could then be compared on any individual day.

For experiments where protein was to be extracted from the bombarded BMS cells, either 6400 ng of p171luci or 1200 ng of pLVScgLIR and 1200 ng pKEPluci, and finally 1200 ng of the plasmid providing Rep in *trans,* was added to 3 mg of gold (see Table 4.1). As an alternative to the WT virus, 1200 ng of pKEP132, a MSV-Kom Rep gene cloned under the control of the rice *actin 1* promoter and *nos* terminator from pCOR113, was used (Palmer, 1997; McElroy *et al.*, 1991). For this experiment each plate was bombarded twice, resulting in three plates per precipitation (see Table 4.1).

The same protocol as above was followed for bombardment of samples from which DNA was to be extracted, except 800 ng of each plasmid was precipitated onto 2 mg of gold and a total of four macrocarriers were used per precipitation. Each plate of BMS was then bombarded twice, resulting in two plates each receiving a total of 200 ng /plasmid/bombardment. This experiment was repeated three times. Unmodified plasmid vector (pBluescript(SK) was co-bombarded with the Rep defective construct as a control for the expression of GUS in a non-replicating environment.

4.3.3.5: DNA extractions and analysis

The extraction protocol for the preferential isolation of low molecular weight DNA of Anat and Subramanian (1992) as quoted in Xie *et al.* (1995) was followed. The final preparation was treated with 10 µg/ml RNaseA and subjected to a final 3 M sodium acetate and ethanol precipitation step. The final DNA concentration was then determined by measuring the absorbance at 260 nm.

Table 4.1: Summary of the plasmids co-precipitated for microprojectile bombardment for the detection of *trans*-replication of MSV-Kom based Rep defective constructs.

Rep defective plasmid	Plasmid providing Rep in *trans*	WT Virus	Total number of precipitations
pLVSccLIR	none[#]	none	3
	pKom602	MSV-Kom	3
	pTas301	MSV-Tas	3
	pPSV1.1	PanSV-Kar	3
	pSet105	MSV-Set	3
	pVW203	MSV-VW	3
	pVM501	MSV-VM	3
p171luci	pBluescript(SK)[#]	none	3
	pKEP132[$]	MSV-Kom C1/C2	2
	pKom602	MSV-Kom	4
	pVW203	MSV-VW	4
	pVM501	MSV-VM	4
	pTas301	MSV-Tas	3
	pSet105	MSV-Set	4
	pPSV1.1	PanSV-Kar	3
	pSV1V2Kδ	MSV-Set/Kom recombinant	3
pLVScgLIR + pKEPluci	pBluescript(SK)[#]	none	3
	pKom602	MSV-Kom	4
	pVW203	MSV-VW	4
	pVM501	MSV-VM	4
	pSet105	MSV-Set	3
	pPSV1.1	PanSV-Kar	3
	pDSV200	DSV	3

[#]: No Rep was expressed from either construct present in the precipitation
[$]: Rep expression cassette

4.3.3.6: Southern blot analysis

A 7 µg aliquot of total DNA extracted as above was electrophoresed through a 0.8 % agarose gel in 0.5 x TBE at 30 volts overnight (Sambrook *et al.*, 1989). The DNA was capillary transferred in a solution of 20 x SSC (all solutions were made as specified in the

Boehringer Mannheim Guide to Dig Users) onto a nylon membrane (Hybond N+), after which the DNA was fixed onto the membrane by UV transillumination at 256 nm for 2 minutes. The membrane was pre-hybridised in Dig Easy Hyb buffer (Boehringer Mannheim) for 30 minutes prior to the addition of a heat denatured probe. After overnight hybridization at 42^0C, the membrane was washed stringently in a 0.1 x SSC solution and the presence of Dig detected using alkaline phosphatase conjugated anti-Dig Fab fragments (Boehringer Mannheim) at a 1 in 20 000 dilution. The presence of alkaline phosphatase was detected by the addition of the chemiluminescent substrate, CPD-Star (Boehringer Mannheim), and the exposure of membrane to X-ray film for the required amount of time.

For the detection of WT replicating DNA a MSV-Kom-specific Dig-labeled probe was made as before (3.3.4). To detect the *trans*-replicated construct pLVSccLIR, a Dig-labeled probe specific for CAT was made by the excision of the *cat* gene from pSKCAT and recovery of the DNA from an agarose gel slice (Geneclean™). The DNA was heat denatured and was labeled overnight at 37^0C by the addition of Klenow polymerase, random hexanucleotides, Digoxigenin-11-dUTP and dNTP mix (as per Boehringer Mannheim Dig DNA Labeling and Detection kit instructions). A working concentration of 1 µl/ml probe was added to 20 ml of Dig Easy Hyb (Boehringer Mannheim) for hybridisation at 42^0C.

4.3.4: PCR confirmation of *trans*-replication

The results obtained from the Southern blot of DNA isolated from pLVSccLIR bombarded cells were confirmed using a primer specific for the CaMV 35S promoter sequence, which could work in conjunction with either of the two degenerate primers 215 - 234 and 1770 - 1792 (Chapter 2). The CaMV 35S primer designed by Dr K. E. Palmer (unpublished data), with the sequence 5'-CAACCACGTCTTCAAAGC-3' (Genosys Biotechnologies), would only amplify a segment spanning the LIR if replication of the Rep mutant had resulted in a circularized DNA form. To each PCR reaction was added a 100 ng of isolated DNA and the same PCR cycle and components as in section 2.3.2 were used except the annealing temperature was decreased to 53^0C. The resultant PCR

amplification product was electrophoresed through a 0.8 % agarose gel and visualized by the addition of ethidium bromide.

Confirmation that the various plasmids carrying partial dimers of WT viruses were replicating, and therefore able to provide Rep in *trans,* was obtained by PCR amplification of 100 ng of DNA extracted from BMS cells (as above), using primers 215 – 234 and 1770 – 1792 (see 2.3.2).

4.3.5: Detection of protein expression

4.3.5.1: Histochemical assay

GUS activity in bombarded BMS cells was visualised by addition of the GUS substrate X-Gluc (Sigma) to cells harvested three days after bombardment (Jefferson *et al.*, 1987). The cells were incubated overnight at 37^0C.

4.3.5.2: Protein assays for the detection of GUS and luciferase activity

Three days after bombardment protein was extracted from the BMS cells in the following manner: the cells were collected into a 1.5 ml Eppendorf microcentrifuge tube and after the addition of diatomaceous earth (Celite), were ground using a microcentrifuge tube pestle. A 250 µl aliquot of luciferase lysis buffer (Luciferase Assay Kit, Tropix, Inc. USA) containing 1 mM DTT was added, the tubes were spun at 14 000 rpm at 4^0C for 2 minutes and the supernatant was kept for analysis as per manufacturers instructions. A 10 µl aliquot of the protein extraction was analysed for luciferase activity, using a Hewlett Packard Pico-lite luminometer, immediately after the proteins were extracted. Within 10 minutes of the addition of 100 µl of substrate A (Tropix) to the protein sample, 100 µl of substrate B (Tropix) was added and the luciferase activity was counted for 10 seconds after a 1 second delay. Activity was expressed as relative light units (RLU). The Tropix luciferase lysis buffer is compatible with the reagents of the GUS-Light™ (Tropix, Inc., USA) GUS activity detection kit and therefore both enzymes could be assayed from the same extraction. GUS activity was either assessed immediately after the luciferase assay

or the samples were frozen at -70^0C until required. A 10 µl protein aliquot was used for the GUS assay to which 90 µl of the reaction buffer (GUS-Light™, Tropix, Inc. USA) had been added. Exactly one-hour later 150 µl of the GUS enhancer (Tropix) was added and the luminometer readings counted for 5 seconds after a delay of 3 seconds. Each precipitation effectively gave three readings i.e. three plates per precipitation. Three DNA precipitations, except for one sample indicated in Table 4.1, were performed for each WT construct and the mean of 9 individual readings is given. The GUS and luciferase activity of the unbombarded BMS cells was analysed and if a precipitation gave values lower than these, all three readings were assumed to be anomalous and were disregarded. A serial dilution of a sample of both the highest luciferase and GUS expressing protein extractions were made, and the readings analysed and found to be within the linear part of the range of detection (data not shown).

4.3.6: Sequence analysis of RepA and the LIR

The LIR nucleic acid sequences of the following viruses; MSV-Tas, MSV-VW, MSV-VM, MSV-Kom, MSV-Set, MSV-Reu and PanSV-Kar were initially aligned using DNAMAN 4.0 (see Table 3.1 for GenBank accession numbers), and then manually realigned to highlight conserved stretches of sequence. The replication specificity domains of both the begomo- and curtovirus Rep protein have been localised to the N-terminus (Jupin *et al.*, 1995; Choi and Stenger, 1995). A multiple sequence alignment of the RepA portion of some of the various sequenced isolates that were analysed in Chapter 2 (see Table 2.4) was performed and the pairwise percentage identity was assessed using DNAMAN 4.0.

4.4: RESULTS

4.4.1: Construction of partial dimers of WT viruses

The relative efficiency of agroinfection of partially dimeric viral genome constucts with one LIR is markedly lower than those with two (D.P. Martin and E. P. Rybicki, unpublished data, and Schnippenkoetter, 1998). Both replicative release and homologous recombination occur when partial dimers with two LIRs are introduced into plant cells, whereas only the latter mechanism is possible with partial dimers with one LIR (Schnippenkoetter, 1998; Heyraud *et al.*, 1993b; Stenger *et al.*, 1991). A partial dimer of PanSV-Kar containing two LIRs was constructed and named pPSV1.1: this was done so that experiments in which PanSV was used to *trans*-replicate MSV-Kom would be directly comparable to experiments using MSV-derived infectious genomic constructs, all of which contained two LIRs.

4.4.2: Replication kinetics of WT viruses in BMS suspension culture cells

I separately bombarded the partial dimers of pKom602 and pSet105 into BMS, and harvested the DNA over a period of 2 to 7 days after bombardment. MSV-Kom dsDNA replicons were detected from day 2 to day 7 (see Fig 4.1), and MSV-Set from days 2 to 6 (data not shown). The level of replication reached a plateau at four days and maintained this level. The replicative forms were compared to MSV-Kom genomic DNA extracted from agroinfected plants (see 3.3.4). Genome copy number was not determined.

4.4.3: *Trans*-replication of Rep mutant MSV-Kom based constructs

4.4.3.1: Trans-*replication of a Rep mutant construct (pLVSccLIR)*

A partial MSV-Kom-derived dimer with a C1/C2 replacement with CaMV 35S-CAT was constructed (pLVSccLIR), which would release a genome of approximately 2.9 kb in the event of *trans*-replicative release or homologous recombination (see Fig. 4.2). No circular

DNA arising from homologous recombination could be detected after three days. Similar observations, using different WDV based constructs, have been made by Suárez-López and Gutiérrez (1997) (see Fig. 4.3). Co-bombardment of pLVSccLIR with partial dimers of wild type MSV isolates and PanSV-Kar indicated that complementation of replication functions could be provided by heterologous viruses with sequence divergence of greater than 20 %. However MSV-Set was repeatedly the least competent at *trans*-replication of the MSV-Kom based construct, as evidenced by both PCR and Southern blot analysis (see Fig. 4.3 & 4.4). PCR analysis using both primers 215 – 234 and 1770 - 1792 confirmed that the WT viruses were self-replicating and therefore expressing functional Rep proteins (see Fig. 4.3A). These results were confirmed in two further experiments.

4.4.3.2: Trans-*replication of a Rep⁻ and GUS V2 replacement MSV-Kom based construct (p171Luci)*

The effect of co-bombardment of WT partial dimers with p171luci on the levels of GUS expression was visualised by the addition of the histochemical substrate X-Gluc to cells harvested three days after bombardment. The virion sense promoter of pKEP171*gus*d is constitutively expressed in BMS (Palmer, 1997), which was evident in the BMS cells co-bombarded with p171luci and pBluescript(SK) (See Fig. 4.5A). Upon co-bombardment of p171luci with plasmids harbouring MSV-Kom, MSV-Tas, MSV-VW or MSV-VM partial dimers, an increase of GUS expression was observed, whereas the addition of PanSV-Kar or MSV-Set seemed to inhibit the expression of GUS to below that of constitutively expressed levels.

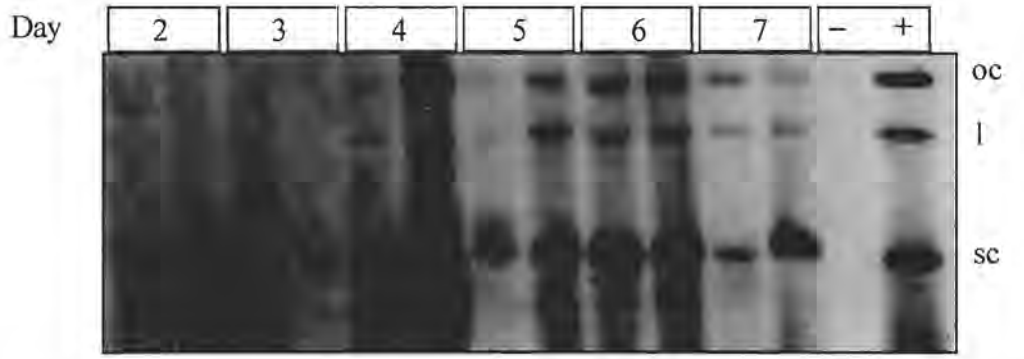


Figure 4.1: Detection of replication of MSV-Kom in BMS cells over a period of 2 to 7 days post bombardment, by Southern blot analysis using a MSV-Kom based Dig-labeled probe. The different dsDNA replicative forms are indicated; linear (l), supercoiled (sc) and open circular (oc). The experiment was performed in duplicate and compared to genomic DNA extracted from MSV-Kom agroinfected sweetcorn cv. Jubilee (+) and from unbombarded BMS cells (-). Two samples were extracted for each time point.

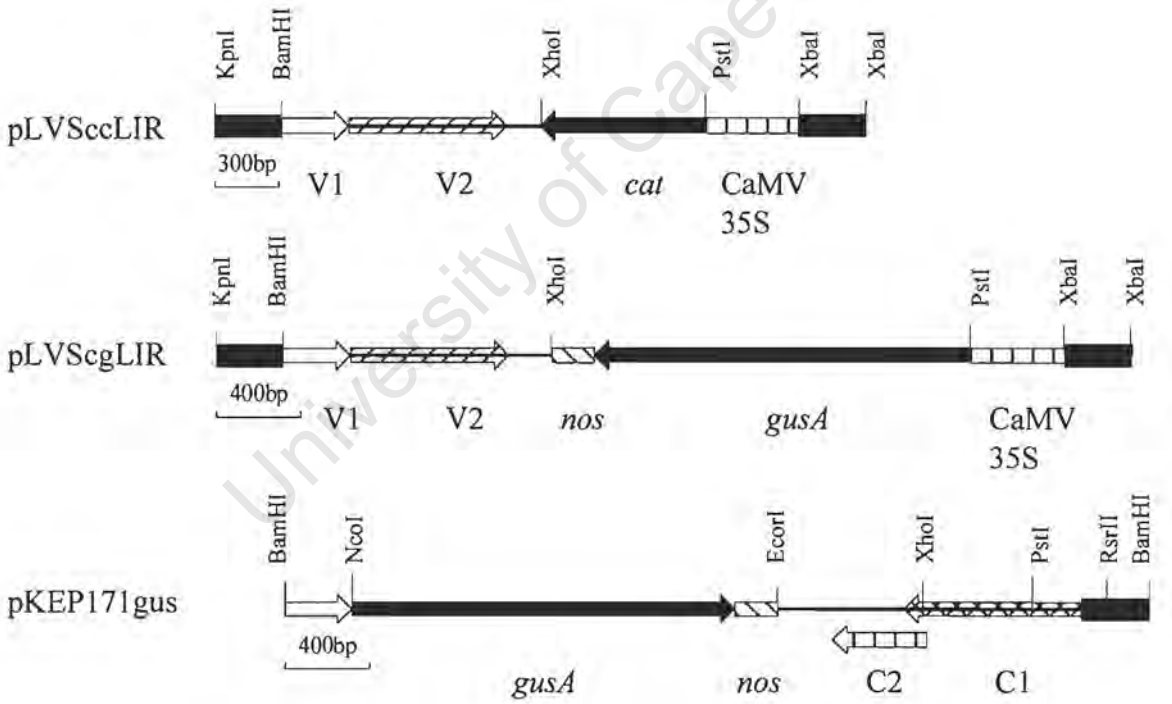


Figure 4.2: Replication deficient MSV-Kom-derived constructs used to assay for the complementation of replication functions by WT viruses. The 1.1 mers of MSV-Kom-based constructs with CaMV 35S and *cat* or *gusA* gene C1/C2 ORF replacements, pLVSccLIR and pLVScgLIR, respectively are represented. The monomer of the MSV-Kom *gusA* and *nos* 3' V2 ORF replacement, pKEP171*gus*, was completely dimerised to form pKEP171*gus*d (Palmer, 1997). All identical ORFs are shaded in the same manner and are labeled under the arrow indicating direction of transcription. The LIR is indicated by a black rectangle and the SIR by a black line, not overlapped by any arrows. The bars indicate the scale in base pairs (bp) of the clone, without vector DNA.

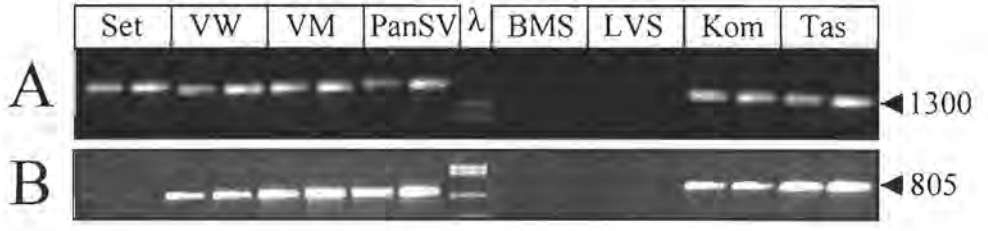


Figure 4.3: Detection by replicon-specific PCR amplification of the *trans*-replication by various WT viruses of a Rep-deficient MSV-Kom construct (pLVSccLIR). All fragments were electrophoresed through a ethidium bromide stained 0.8 % agarose gel. A) Detection of replicating WT viruses by PCR using the degenerate primers 215 - 234 and 1770 - 1792. B) Detection of pLVSccLIR replicative forms by PCR using primers 215 - 234 and CaMV 35S. DNA was isolated from 2 individually bombarded plates of BMS cells. The following plasmids were co-bombarded with pLVSccLIR: pSet105 (Set), pVW301 (VW), pVM501 (VM), pPSV1.1 (PanSV), pKom602 (Kom) and pTas301 (Tas). As controls, DNA was extracted from unbombarded cells (BMS) and cells bombarded with only pLVSccLIR (LVS). The approximate size (in base pairs) of the PCR products compared to the lambda *Pst*I digested DNA (λ), are indicated on the right.

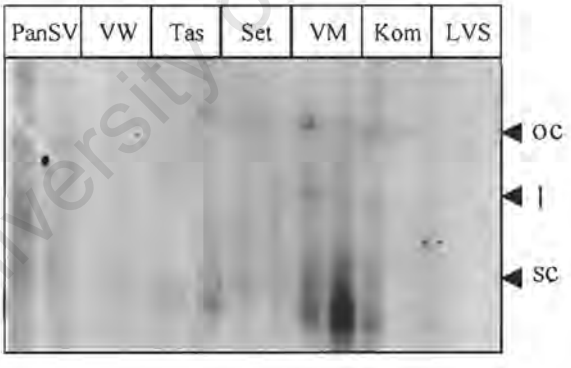


Figure 4.4: Detection of the replicative forms of the MSV-Kom based CaMV 35S and *cat* cassette C1/C2 ORF replacement (pLVSccLIR), by Southern blot analysis using a Dig-labeled CAT probe. The DNA extractions were electrophoresed through a 0.8 % agarose gel prior to capillary transfer to the nylon membrane. The samples are the same as as those subjected to PCR analysis (see Fig. 4.3), confirming the lower yield of replicating DNA with MSV-VW and MSV-Set co-bombardments with pLVSccLIR in this experiment. Due to high experimental variation even within the results obtained for one precipitation, the experiment was repeated three times. One set of precipitations are shown here.

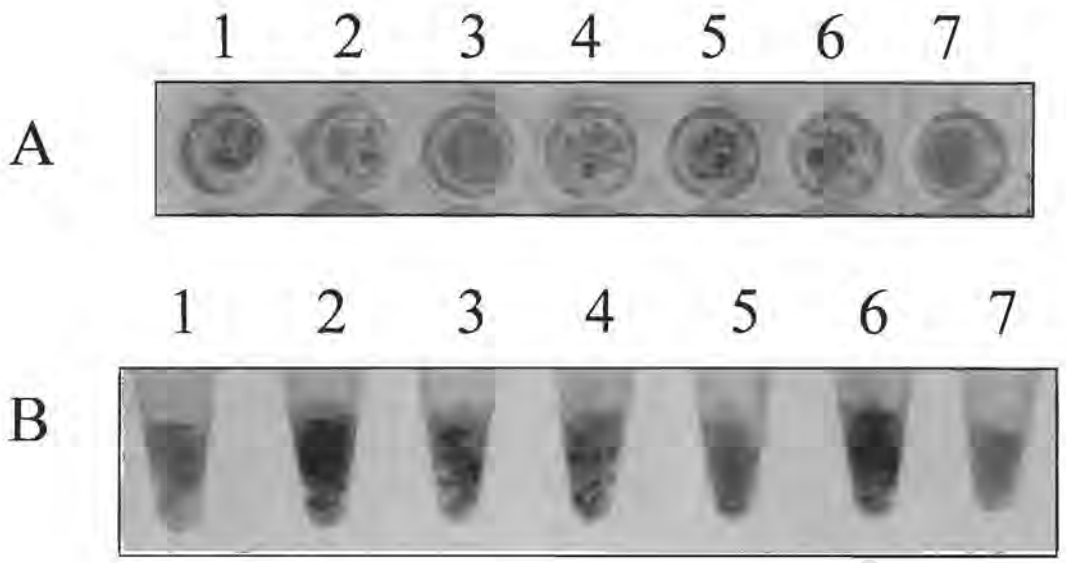


Figure 4.5: The effect of providing Rep in *trans* on the expression of GUS under the control of: A) the virion sense promoter in p171luci, and B) the CaMV 35 S promoter in pLVScgLIR. A) All wells contain BMS bombarded with p171luci and well 1, pBluescript(SK); well 2, pSet105; well 3, pPSV1.1; well 4, pTas301; well 5, pVW203; well 6, pVM501; and well 7, pKom602. B) All microcentrifuge tubes contain BMS bombarded with pLVScgLIR and 1, pBluescript(SK); 2, pKom602; 3, pVW302; 4, pVM501; 5, pSet105; 6, pPSV1.1 and 7, pDSV200.

The relative expression levels of GUS and luciferase in BMS cells co-bombarded with p171luci and various plasmids expressing Rep are tabulated below. These results indicated that only in the presence of high concentrations of Rep (pKEP132) was any large enhancement of GUS expression observed, this being a six fold increase over the non-replicating construct (see Table 4.2) (K. E. Palmer, 1997). Only Rep derived from the homologous virus (MSV-Kom) expressed as part of a replicating genome gave significant expression of GUS over the background levels, while significant inhibition of GUS expression, as was observed in the histochemical assay, was evident for the same two constructs (e.g. MSV-Set and PanSV-Kar). Co-bombardment of p171luci with MSV-Set which had both the virion sense ORFs replaced with the same MSV-Kom sequences, pSV1V2Kδ, increased the levels of GUS expressed back to the basal constitutively expressed levels compared to that of the WT MSV-Set construct (see Fig. 4.6).

Table 4.2: Mean GUS/luciferase ratio for co-bombardments of p171luci with various plasmids providing Rep in *trans*.

Plasmid	Mean GUS/Luc[1]	SD[2]	95 % CI[3]	No. of samples[4]
pKEP132[#]	18.35	6.62	5.29	6
pBluescript(SK)	2.89	0.63	0.41	9
pKom602	4.77	3.01	1.71	12
pVW203	2.95	0.89	0.50	12
pVM501	3.48	1.52	0.86	12
pTas301	2.21	0.95	0.62	9
pPSV1.1	0.76	0.25	0.17	9
pSet105	1.52	0.82	0.46	12
pSV1V2Kδ	2.48	0.83	0.54	9

[1]: Mean of the standardised GUS and luciferase expression ratios obtained for the number of samples indicated in column 4.
[2]: SD: Standard deviation
[3]: 95 % CI: 95 % confidence interval for the mean
[4]: Number of readings obtained for each *trans*-replication assay
[#]: Non-replicating construct

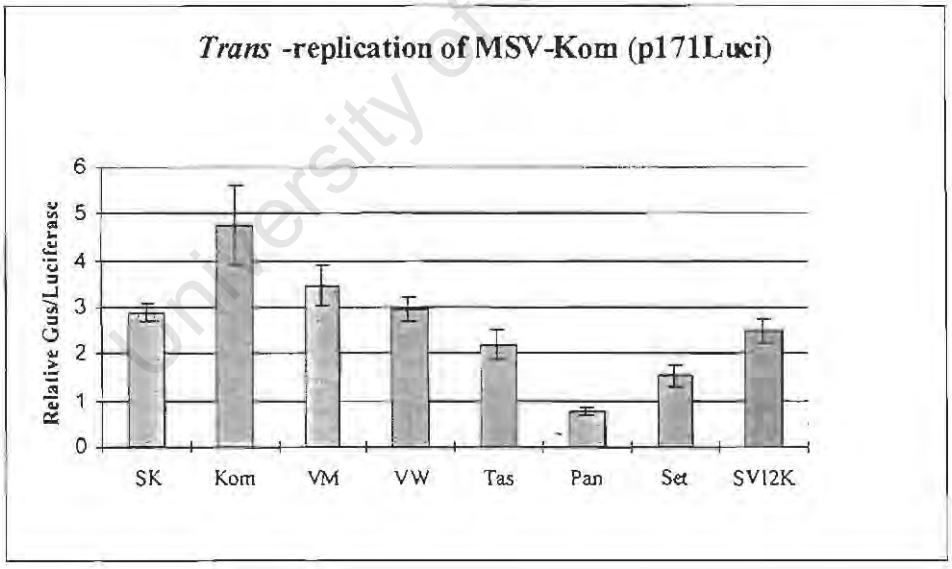


Figure 4.6: Expression of GUS in BMS cells after co-bombardment of p171luci with the various plasmids indicated. SK = pBluescript(SK); Kom = pKom602; VW = pVW203; VM = pVM501; Tas = pTas301, PanSV = pPSV1.1; Set = pSet105 and SV12K = pSV1V2Kδ. Error bars indicate the 95 % confidence intervals for the mean (see Table 4.2). Significant differences were determined using the Student's *t* test. MSV-Kom, MSV-Set and PanSV-Kar gave the only significantly different results ($p < 0.05$) compared to the non-replicating pSK samples.

4.4.3.3: Trans-*replication of MSV-Kom C1/C2 ORF replacement (pLVScgLIR)*

The plasmid pLVSccLIR, which contained all the necessary *cis*-elements required for replication, was modified to express the GUS gene under the control of the CaMV 35S promoter as a C1/C2 ORF replacement (see Fig 4.2). Any putative negative regulation of Rep transcription, analogous to that observed in the begomoviruses, therefore need not be taken into consideration (Eagle *et al.*, 1994; Sunter *et al.*, 1993). The plasmid pLVScgLIR would replicatively release a 4.1 kb replicon in the presence of a functional Rep supplied in *trans*. The relative ability of each of the various WT viruses to complement replication functions of either pLVScgLIR or pKEPluci would be directly correlated to the levels of *gusA* transcripts present and therefore the levels of GUS expression. The assessment of luciferase activity expressed from a separate plasmid, pKEPluci, was used as the internal experimental control. The histochemical assay for GUS expression correlates well with the ratios of GUS to luciferase activity measured using bioluminescent detection of the enzyme in the crude protein extracts (see Fig. 4.5B and 4.7). A threefold increase of GUS expression was achieved after co-bombardment of pLVScgLIR with the homologous virus, MSV-Kom (see Table 4.3). High standard deviation of the mean, possibly due to experimental variation was noticed when three plasmids were simultaneously introduced into the BMS. Nonetheless, significant *trans*-replication of the MSV-Kom Rep mutants by a range of WT viruses was achieved ($\rho < 0.05$) and the inclusion of DSV served as a non-replication complementing control.

Differences in the levels of GUS expression from the MSV-Kom virion sense promoter compared to expression from the CaMV 35S promoter were observed in the histochemical assays (compare Fig 4.5A1 with B1). No conclusions can be drawn from this because; although it appears that the virion sense promoter is expressing GUS at a higher level this could be due to the higher amount of input DNA used in the p171luci bombardments. Furthermore, p171luci contains a full dimer and therefore two copies of the virion sense promoter compared with the single CaMV 35S promoter present in the partial dimer of pLVScgLIR.

Table 4.3: Mean GUS/luciferase ratio for co-bombardments of pLVScgLIR with plasmids providing Rep in *trans*.

Plasmid	Mean GUS/Luc[1]	SD[2]	95 % CI[3]	No. of samples[4]
pBluescript(SK)[#]	4.56	2.82	1.84	9
pKom602	13.52	7.26	4.11	12
pVM501	13.63	13.44	7.60	12
pVW203	13.86	8.17	4.63	12
pSet105	10.93	3.24	2.12	9
pPSV1.1	12.96	4.96	3.24	9
pDSV200	4.41	2.85	1.86	9

[1]: Mean of the standardised GUS and luciferase expression ratios obtained for the number of samples indicated in column 4.
[2]: SD: Standard deviation
[3]: 95 % CI: 95 % confidence interval for the mean
[4]: Number of readings obtained for each *trans*-replication assay
[#]: Non- Rep expressing construct

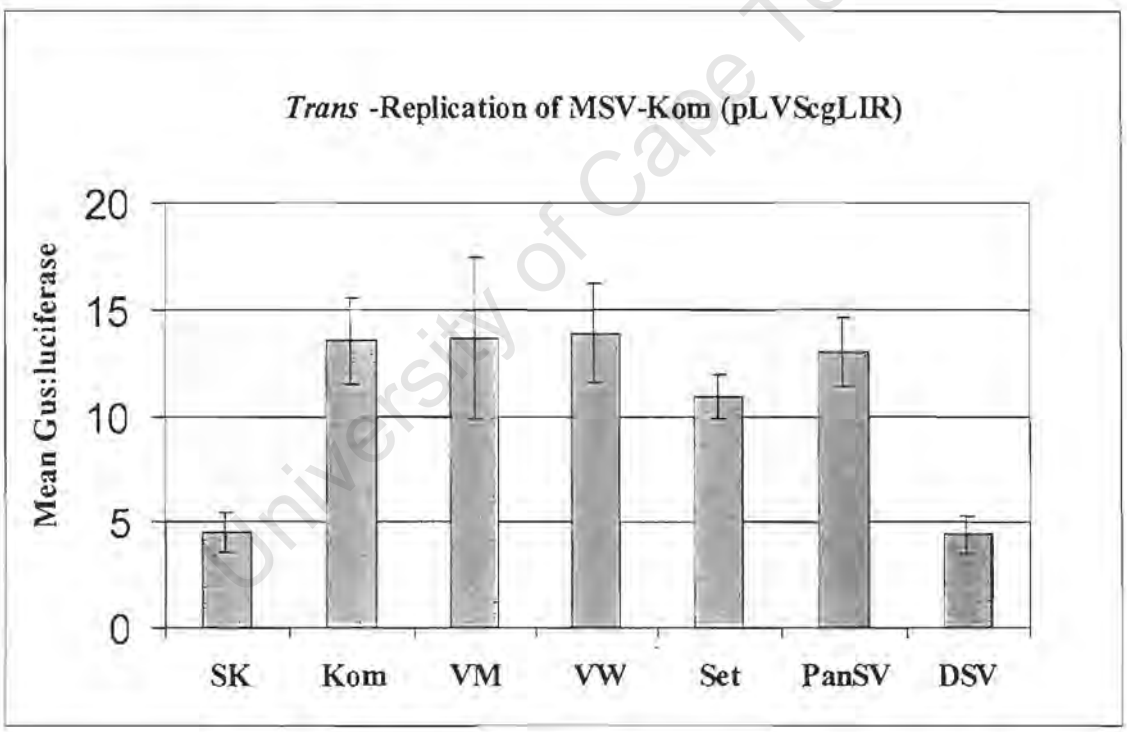
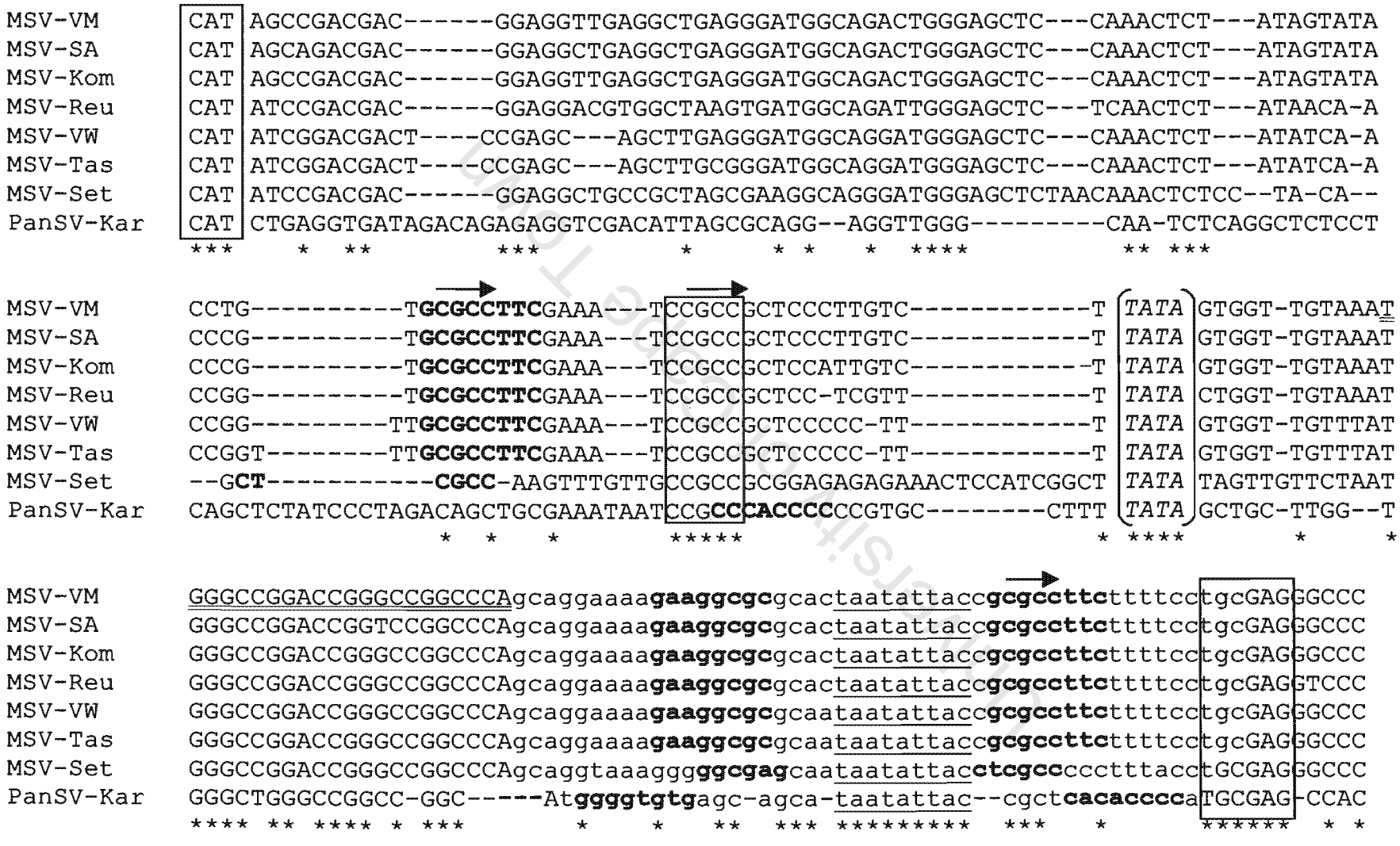


Figure 4.7: Expression of GUS in BMS cells after co-bombardment of pLVScgLIR and pKEPluci with the various plasmids indicated. SK = pBluescript(SK); Kom = pKom602; VW = pVW203; VM = pVM501; Set = pSet105, PanSV = pPSV1.1 and DSV = pDSV200. Error bars indicate the 95 % confidence interval for the mean (see Table 4.3). Significant differences in GUS expression due to *trans*-replication, determined using the Student's *t* test, were achieved for all WT viruses except DSV ($p < 0.05$). No significant differences were observed between the ability of the homologous virus, MSV-Kom, and those of the heterologous viruses, to *trans*-replicate pLVScgLIR ($p > 0.05$).

4.4.4: Sequence analysis of the LIR and Rep proteins

In an attempt to explain the complementation of *trans*-replication functions between virus as diverse as PanSV and MSV-Kom I undertook to identify common features within the origin of replication, other than the previously identified iterons, which may act alone or in conjunction as replication specificity determinants. The PanSV-Kar Rep is able to recognise the MSV-Kom origin of replication and *trans*-replicate MSV-Kom derived constructs to significantly detectable levels; therefore, according to current thinking the *cis*-element(s) that reside within the PanSV-Kar *ori* must be conserved to some extent within the MSV-Kom *ori*. The putative PanSV-Kar Rep proximal iteron sequence 5'-CCCACCCC-3' has the closest sequence identity to a region just downstream of the MSV-Kom putative iteron, where either of two overlapping positions share 4 out of 8 nucleotides (see Fig. 4.8). Conversely, the PanSV-Kar LIR contains the sequence 5'-CGTGCCTTT-3', just downstream of its own Rep proximal iteron (underlined nucleotides are part of the MSV-Kom Rep proximal iteron sequence). Analysis of the stem-loop sequences revealed that 3' of the TAATATTAC loop sequence was a highly conserved sequence C(**G**/T)C(**G**/T)C(**C**/A), with the majority of viruses sharing the nucleotides indicated in bold. The underlined nucleotides are also present in the Rep proximal iteron of all 16 sequenced MSV LIRs. The multiple sequence alignment used to generate the relationship dendogram (Fig. 2.5) in Chapter 2 confirms this (data not shown, J. A. Willment, D. P. Martin and E. P. Rybicki, unpublished data). This same CGCC sequence exists as a repeat (see Fig. 4.8), once in the MSV Rep proximal iteron, and once again just downstream, although PanSV-Kar only shares the latter overlapping with its own putative iteron sequence. Both the putative MSV-Set Rep proximal and stem-loop iteron sequence share 4 out of 8 nucleotides with the typical MSV iteron (Schnippenkoetter, 1998): these four nucleotides are the same CGCC sequences repeated downstream.

There are two further highly conserved groups of sequences on the virion sense side of the stem-loop, each six nucleotides long. The first is at the base of the MSV stem-loop structure and the second just prior to the putative DNA bending region (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995).



(This figure continues on the next page)

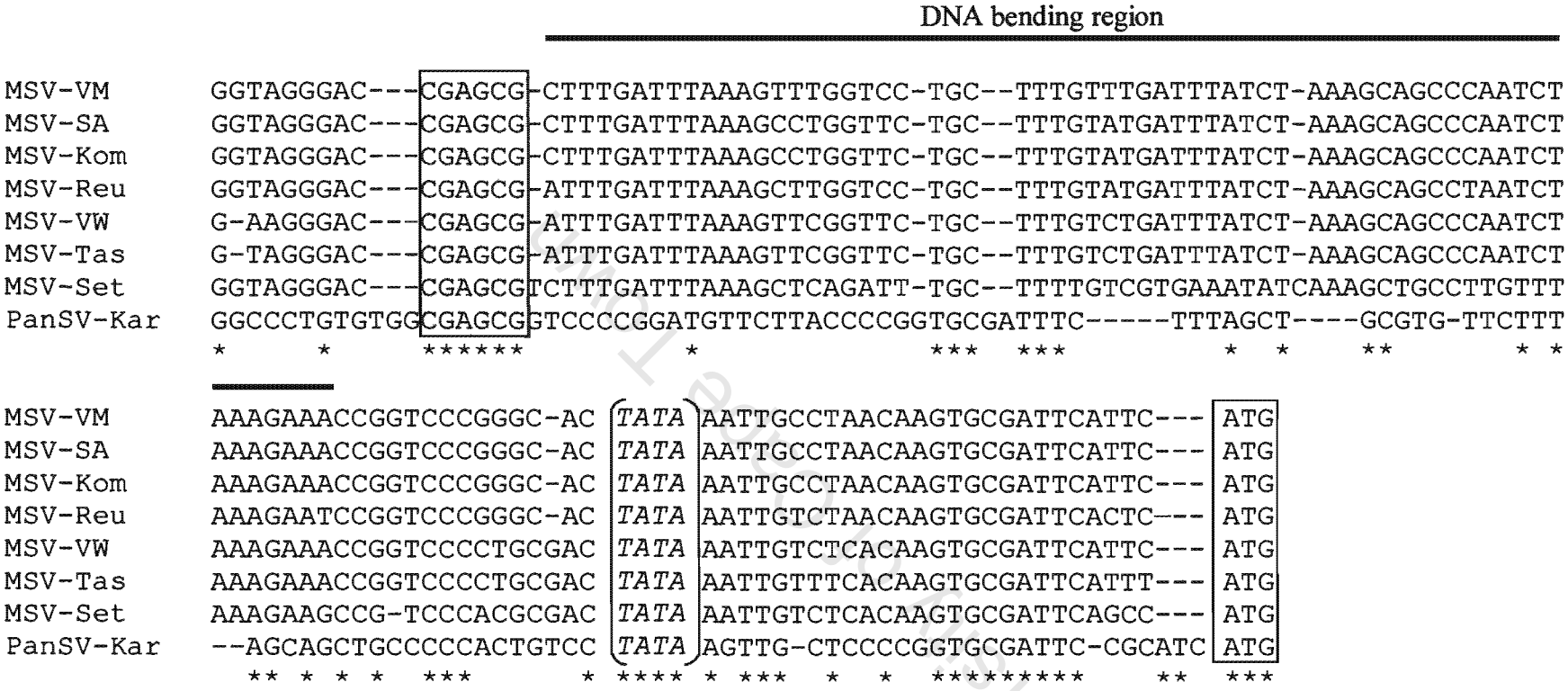


Figure 4.8: Multiple sequence alignment of the LIRs of selected mastreviruses. The sequences were initially aligned using DNAMAN 4.0 and then manually realigned to allow identification of stretches of homology (shaded boxes). Sequences conserved amongst all eight viruses are highlighted by an asterisks (*), while gaps within the alignment are indicated by a dash (-). The sequence is displayed from the start codon (boxed) of the C1 ORF on the left (position 2528 of MSV-VM) to the start codon of the V1 ORF (boxed) (position 150 of MSV-VM) (see Table 3.4). The stem-loop structure is indicated in lower case, with the invariant loop sequence underlined. The identified iterons are indicated in bold type (Schnippenkoetter, 1998; Argüello-Astorga *et al.*, 1994a & b). The TATA boxes are indicated by brackets and the rightward promoter element (*rpeI*) is double underlined on the MSV-VM sequence (Fenoll *et al.*, 1990). The presence of T tracts known to cause DNA bending are indicated by a bold line above the sequence (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995). The presence of a repeated sequence, CGCC, forming part of or overlapping with the identified iterons, is indicated by the arrows above the alignment. (For GenBank accession numbers see Table 3.1)

The significance of these conserved regions is unknown, but the WDV minimal *ori* does require about 28 nucleotides downstream of the stem-loop for replication to occur, and a further 22 nucleotides 3' of this enhances replication (Sanz-Burgos and Gutiérrez, 1998).This is in contrast to the begomoviruses, which do not require any sequences 3' of the stem-loop for replication (Lazarowitz *et al.*, 1992). A putative DNA bending region has been identified in both WDV and MSV, and occurs in all mastreviruses except MiSV and CSMV (as cited in (Suárez-López *et al.*, 1995; Gutiérrez *et al.*, 1995)). Although the MSVs are highly conserved in this region, this is not much sequence homology with PanSV-Kar (see Fig 4.8); thus, this region may not play a significant role in replication (Sanz-Burgos and Gutiérrez, 1998).

Table 4.4: Pairwise percentage amino acid identities of RepA proteins and nucleic acid identities of LIR sequences[1].

Isolate		1	2	3	4	5	6	7	8
MSV-VM	**1**	100	98.9	98.5	98.9	85.5	86.6	76.1	56.9
MSV-SA	**2**	98.3	100	98.9	97.1	85.9	87.0	76.9	58.1
MSV-Kom	**3**	98.1	99.0	100	96.7	85.1	86.2	76.1	57.3
MSV-Reu	**4**	95.6	95.6	95.4	100	87.0	87.4	76.5	57.3
MSV-VW	**5**	90.1	89.7	89.6	89.5	100	95.2	80.8	57.6
MSV-Tas	**6**	90.2	90.0	89.8	89.6	98.1	100	78.6	56.5
MSV-Set	**7**	79.1	78.9	78.9	78.9	80.3	79.9	100	59.3
PanSV-Kar	**8**	63.2	63.4	63.2	63.7	63.4	63.5	63.4	100

[1]: Upper right-hand side of the table above the diagonal indicates the RepA percentage amino acid sequence identity and the lower left-hand side indicates the LIR percentage nucleic acid sequence identity obtained using the default optimal alignment options in DNAMAN 4.0. The percentage identities MSV-Kom shares with the other viruses are highlighted.

The most variable part of the geminivirus genome is the large intergenic region, or common region, with the 5' end showing greater variability (Zhou *et al.*, 1998; Padidam *et al.*, 1995; Faria *et al.*, 1994; Rybicki, 1994). Phylogenetic relationships based on the intergenic region generally correspond well to those obtained from both the nucleic acid and amino acid sequences of the various expressed genes at least for more closely related viruses (Padidam *et al.*, 1995; Rybicki, 1994). Overall the Rep protein is less variable than the C1 (RepA) protein sequence (compare Table 3.8 with Table 4.4), which is expected considering that the *trans*-acting replication specificity domains within the geminivirus Rep protein reside in the N-terminus. The divergence observed between the MSV-Kom RepA sequence and those of the viruses which were able to complement

transcriptional repression, has been proven (Eagle and Hanley-Bowdoin, 1997; Eagle *et al.*, 1994; Sunter *et al.*, 1993).

A more representative assay of the ability of WT viruses to *trans*-replicate a Rep mutant MSV-Kom, although still not sufficiently sensitive enough for detection of subtle differences in the efficiency of *trans*-replication, was the use of the CaMV 35S promoter to express GUS as a C1/C2 ORFs replacement cassette. GUS expression normalized against the expression of the internal luciferase control gave a threefold increase in GUS expression due to homologous *trans*-replication. Results indicated that all WT viruses, with the exception of DSV, were able to *trans*-replicate MSV-Kom, duplicating the results obtained with the CAT construct. However, the high variablity between assays, even with the presence of an unlinked internal control, pKEPluci, did not allow any significant differences in replication efficiency to be detected.

Efficient *trans*-replication would presumably require the sharing of common RSDs within the LIR. The mastrevirus intergenic regions are very variable in terms of sequence diversity (Padidam *et al.*, 1995; Rybicki, 1994; Hughes *et al.*, 1992); however, the RSDs defined by Arguello-Astorga *et al.* (1994a & b) are conserved between all of these sequenced MSVs, excepting MSV-Set and now MSV-Raw (Schnippenkoetter, 1998 & D. P. Martin and E. P. Rybicki, unpublished data). The more distantly related PanSVs, SSV, DSV and WDVs share the overall spatial arrangement of RSDs with the MSVs, i.e. the iterons within the stem-loop and one C1 proximal iteron (Palmer and Rybicki, 1998). As can be observed from the multiple sequence alignment (see Fig. 4.8), there are considerable differences in the spacing of the TATA boxes and iterons, both Rep proximal and within the stem-loop, for PanSV-Kar and MSV-Set compared to the other viruses analysed. Only in the case of MSV-Set do these differences (potentially acting as secondary replication specificity elements) seem to impact its ability to recognise and *trans*-replicate MSV-Kom (Fontes *et al.*, 1994). Other than the BCTV direct repeats, no further secondary determinants of replication have been observed in curtoviruses, even between BCTV strains now defined as distinct viruses, differing in nucleotide sequence and spacing upstream of the stem-loop (Choi and Stenger, 1996). Conversely, the begomovirus TGMV requires up to six *cis*-elements for efficient replication, including elements with precise sequences, besides the direct repeats, and conserved spatial arrangements of these elements (Orozco *et al.*, 1998).

There was far less replication specificity displayed amongst the mastreviruses examined in this study than was shown for the curto- or begomoviruses analysed to date. Evidence that a chimaeric MSV-Set with MSV-Kom LIR sequences still produced symptoms in plants, although these were highly attenuated compared to the WT MSV-Set, adds further proof that the multifunctional mastrevirus Rep protein can engage all essential interactions with a highly heterologous LIR (Schnippenkoetter, 1998). The putative PanSV-Kar Rep proximal iteron region is very similar to the cognate MSV-Kom sequence (see Fig. 4.8), which may be the factor which allows such efficient Rep recognition of the MSV-Kom LIR compared to the case with MSV-Set, despite the fact that the PanSV genome shares 60 % overall sequence similarity with MSV-Kom, compared to the 79 % identity shared between MSV-Set and MSV-Kom (Table 3.8). The direct repeats responsible for curto- and begomovirus Rep recognition also display some flexibility in sequence, with conservation of the 3' repeat sequence being more important (Choi and Stenger, 1996; Fontes *et al.*, 1994).

The stem-loop structures of mastreviruses are presumed to be the site of Rep binding for its nicking and joining activity, and therefore act as the RSDs for these viruses (Laufs *et al.*, 1995c; Argüello-Astorga *et al.*, 1994a & b). I have detected *trans*-replication of viruses which have considerable differences within their stem-loop sequences; however there is a partly conserved CGCC sequence, which is repeated at least once near or at the Rep proximal iteron, in all the viruses investigated (see Fig. 4.8). The significance of this repeated sequence is presently undetermined. As yet no direct experimental evidence exists that the stem-loop sequence holds the RSDs, only that it is required for mastrevirus termination of replication and does tolerate some changes, for example the WDV hybrid stem-loop allowing replicative release (Heyraud *et al.*, 1993a & b).

The results presented here have led to the analysis of the regions within the LIR responsible for firstly, origin function, and secondly, for replication specificity, using MSV-Kom and MSV-Set based constructs. Although these viruses do complement replication functions, these are clearly sub-optimal and the identification of the *cis*-elements responsible for this impaired interaction will provide greater insight into the mechanisms involved in mastrevirus replication.

Chapter 5

Identification of the MSV *cis*-elements involved in replication

5.1: SUMMARY 150
5.2: INTRODUCTION 151
5.3: METHODS AND MATERIALS 153
5.3.1: Construction of partial dimers with LIR deletion mutants 154
5.3.2: Construction of agroinfectious clones with recombinant LIRs 159
5.3.3: Directed genomic exchanges between MSV-Kom and MSV-Set 160
5.4: RESULTS 166
5.4.1: Construction of the truncated LIR partial dimers 166
5.4.2: Transient replication results of LIR deletion mutants 166
5.4.3: Analysis of agroinfectious constructs with chimaeric LIRs 170
5.4.4: Transient *trans*-replication assays 175
5.5: DISCUSSION 175

5.1: SUMMARY

The main *cis*-acting control regions of maize streak virus replication reside within a approximately 350 bp region called the long intergenic region (LIR). The sequence determinants for initiation and termination of mastrevirus replication have yet to be fully elucidated, but the iterons - found on either side of the conserved stem-loop structure in the LIR and one upstream of the Rep TATA box - are believed to be involved. I have made series of deletion and directed mutants of the LIR in order to identify the minimal regions of the LIR necessary for replication, and those that are involved in replication specificity. The minimal LIR sequence required for replication included sequences 5' of the stem-loop up to the Rep proximal iteron and 25 bp 3' of the base of the stem structure. Specific LIR sequence exchanges between MSV-Kom and MSV-Set revealed that the stem-loop sequence is not involved in determining replication specificity, but the presence of a homologous wild type stem sequence enhances replication. A 80 bp region surrounding the theoretically defined Rep proximal iteron was essential for replication specificity. The presence of the same iteron sequence in both the stem-loop and Rep proximal region enhanced the pathogenicity of an agroinfectious MSV-Set mutant over and above levels achieved when only one or the other was present. Transient *trans*-replication assays indicated that a self-replication deficient mutant was preferentially replicated when its iteron region and the Rep protein provided in *trans* originated from the same virus. This interaction was confirmed by making reciprocal exchanges between the MSV-Kom and MSV-Set LIRs.

5.2: INTRODUCTION

The *cis*-elements essential for mastrevirus replication reside within the LIR and the SIR, which contain the respective origins of virion and complementary sense strand DNA synthesis (Kammann *et al.*, 1991; Hayes *et al.*, 1988; Donson *et al.*, 1984). The only viral protein necessary for mastrevirus replication is Rep, unlike the begomo- and curtoviruses, which have a further protein (REn) responsible for enhancing replication (Hormuzdi and Bisaro, 1995; Etessami *et al.*, 1991; Sunter *et al.*, 1990; Schalk *et al.*, 1989). The begomovirus Rep, the most extensively characterised Rep protein, is a sequence-specific DNA binding protein, which recognises and specifically binds to a direct repeat sequence within the *ori* (Behjatnia *et al.*, 1998; Fontes *et al.*, 1994a; Fontes *et al.*, 1992). Furthermore, these repeats, located between the Rep TATA box and translation start site in both the curto- and begomoviruses (see Fig. 1.3), are replication specificity determinants (RSDs), with viruses sharing the same direct repeats being able to *trans*-replicate each other (Choi and Stenger, 1996; Choi and Stenger, 1995; Faria *et al.*, 1994; Lazarowitz *et al.*, 1992). While further elements within the begomovirus CR, some of which are host factor binding sites, enhance or contribute towards efficient replication, their role in curto- or mastrevirus replication is not yet defined (Orozco *et al.*, 1998; Choi and Stenger, 1996) (also see Chapter 1). Apart from the research that culminated in the identification of the site of initiation of replication in WDV and BCTV, all then-current (1996) understanding of the *cis*-elements involved in geminivirus replication had been investigated using isolates of begomo- and to some extent curtoviruses (Bisaro, 1996; Hanley-Bowdoin *et al.*, 1996; Stanley, 1995; Heyraud *et al.*, 1993b; Kammann *et al.*, 1991; Stenger *et al.*, 1991).

Evidence to date implicates the region between the Rep TATA box and stem-loop of WDV as being necessary for replication (Schalk *et al.*, 1989). This region contains the binding site of host factors involved in rightward transcription, i.e. the *rpe*I element (Hofer *et al.*, 1992; Fenoll *et al.*, 1990; Fenoll *et al.*, 1988). During the course of this project, the WDV minimal *ori* was mapped to a sequence stretching from 170 bp upstream of the stem-loop to 28 bp downstream (Sanz-Burgos and Gutiérrez, 1998). The presence of auxiliary regions on either side of the minimal origin enhanced replication (Sanz-Burgos and Gutiérrez, 1998). K. E. Palmer from this laboratory found that a 86 bp

region downstream of the MSV stem-loop was required for replication (Palmer, 1997), which prompted my study of the MSV origin of replication.

The requirement for specific sequences within the stem of the conserved stem-loop structure is flexible in begomoviruses, although cognate stem sequences enhance replication possibly due to an interaction with REn and Rep-REn oligomers (Orozco and Hanley-Bowdoin, 1996; Hanley-Bowdoin *et al.*, 1996). The lack of a REn-like homolog, and the presence of the conserved iterons, has led to speculation that Rep may interact directly with these elements (see Chapter 1).

The various species of mastreviruses have unique stem sequences, unlike many begomoviruses, and it has been proposed that the stem sequence as well as the Rep proximal iteron constitute the RSDs (Argüello-Astorga *et al.*, 1994a & b). The necessity for these elements during replication was obtained by the mutational studies of Schneider *et al.* (1992). As well as providing evidence that the MSV stem-loop structure was essential for replication, the insertion of a linker into the *Asu*II site (which overlaps with the Rep proximal iteron sequence) abolished infectivity; however, this was thought to be due to interference with Rep transcription (Schneider *et al.*, 1992). Rep complexes have also been visualised binding to the WDV LIR in the same proximity as its upstream iteron, thereby further implicating this element's involvement in replication (Sanz-Burgos and Gutiérrez, 1998).

Termination but not initiation of replication requires the presence of a functional stem-loop structure in WDV (Heyraud-Nitschke *et al.*, 1995; Heyraud *et al.*, 1993a). However, two WDV isolates (sharing 75 % LIR sequence identity) used to examine the mechanism of replicative release during rolling circle replication have different stem sequences, yet wild type viruses were replicatively released from recombinants with hybrid stem structures (Heyraud *et al.*, 1993b). These results indicate that Rep recognised the 3' side of the stem-loop structure, and was able to initiate and terminate replication and from the invariant loop sequence, even though the stem sequence contained some non-complementary nucleotides (Heyraud *et al.*, 1993b).

While this study was in progress, an agroinfectious construct was obtained from a partial MSV-Set and a full MSV-Kom genome (sharing 78 % overall sequence identity) cloned in tandem, i.e. a partial heterodimer (Schnippenkoetter, 1998). The viral progeny was cloned back from the infected plants and partially sequenced by W. H. Schnippenkoetter (Schnippenkoetter, 1998). The predominant class of clones had hybrid LIRs arising from homologous recombination events between the two viral virion sense promoters (62 %), and a subset (11 %) were recombination events on the 5' side of the stem-loop. Full genomic heterodimers were also agroinfected and some of the viral progeny recovered had small insertions within the loop sequence, indicating that aberrant replicative release had probably occurred (Schnippenkoetter, 1998). These monomeric clones were made available to me for further investigations of the sequence requirements for maize streak virus replication. Evidence provided by the W. H. Schnippenkoetter study seemed to imply that the cognate stem-loop sequence, which differs between the two MSVs, was essential for efficient replicative release (Schnippenkoetter, 1998).

In summary, this work sought to identify the *cis*-elements responsible for controlling the replication cycle of MSV in particular, and perhaps of mastreviruses in general, and to define the elements that are involved in determining replication specificity between two MSV strains, MSV-Set and MSV-Kom.

5.3: METHODS AND MATERIALS

MSV-Kom and MSV-Set genomes were isolated from total genomic DNA from infected plants, and cloned into the *Bam*HI sites of pUC19 and pUC18 respectively (Schnippenkoetter, 1998; Hughes, 1991) (see Fig. 5.1). These two clones, pKom500 and pSet100, provided the foundation for construction of the chimaeric MSV-Kom and MSV-Set clones made during the course of this study. All monomers were cloned at the *Bam*HI site, the recognition pattern of which starts at the last nucleotide of the V1 start codon. The direction of translation of each ORF is indicated with respect to the LIR in Fig 5.1.

5.3.1: Construction of partial dimers with LIR deletion mutants

All deletions of the LIR were based upon pLIR1 (4.3.2.1), a full LIR clone from the C1 start codon to the *Bam*HI site at the start of the V1 ORF, 315 bp in length (see Fig. 5.2). Effects of these deletions on replication were investigated by cloning, as part of a 1.1 mer, the mutant LIRs either on the virion sense side or the complementary sense side of the WT MSV-Kom genome. Replication proficient constructs were then assayed for as described below.

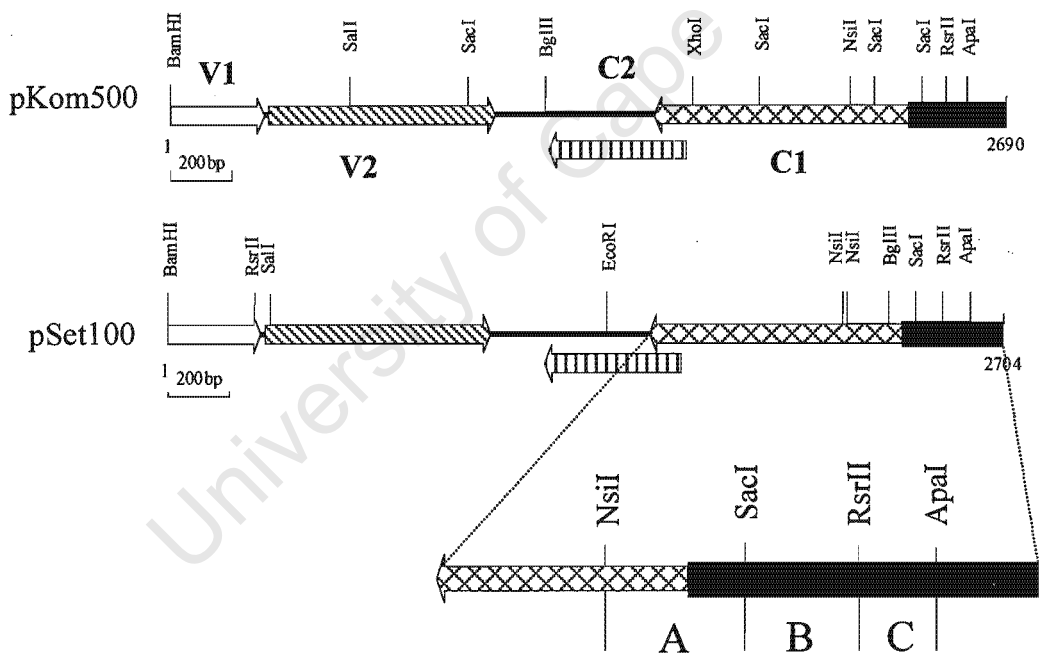


Figure 5.1: Restriction maps of the monomeric clones of MSV-Kom and MSV-Set, pKom500 and pSet100, respectively (Schnippenkoetter, 1998 & Hughes, 1991). Maps were generated from sequence data (see Table 3.1) using DNAMAN 4.0. The various ORFs, and the direction of transcription are indicated by the arrows and are named above or below, the MSV-Kom map. Regions (A - C) exchanged between the two viruses, are indicated with the flanking restriction sites.

A series of plasmids with deletions in either the virion or complementary sense side of the LIR were cloned into a virion sense gene replacement monomer, pKEP151, provided by Dr K. E. Palmer (Palmer, 1997). This resulted in a 1.1mer genome, with MSV ORFs flanked on the virion sense side (first LIR) by a mutant LIR and on the complementary sense side (second LIR) by a WT LIR (see Fig. 5.2). A genomic dimer based on a deletion mutant of pKEP151, called pKEP167, had previously been used to show that the 86 bp region between the *Apa*I and *Bam*HI site was essential for replication (Palmer, 1997). The plasmid pKEP151 was based on the monomeric clone of pKom500, which had the virion sense ORFs replaced by a *Xba*I - *Nco*I fragment containing the CaMV 35S and *bar* gene fragment from plasmid pJW7503. I had previously constructed pJW7503 from pPHP7503 (obtained from Dr Brad Roth, Pioneer Hi-Bred, International, Inc., Des Moines, IA) by deleting the TMV omega leader sequence and *AdhI* intron, which were flanked by *Pst*I and *Bam*HI. The sites were destroyed by Klenow treatment (Boehringer Mannheim) before self-ligation of the remaining parental DNA. The resultant plasmid pKEP151, which would produce a MSV-derived genome only 33 bp larger than that in pKom500, formed the monomer backbone from which deletions to the first LIR were investigated. The choice of the virion sense gene replacement cassette from pJW7503 was based on its appropriate size and not necessarily its translation product. The truncated LIR fragments were cloned into the unique *Xba*I site. All clones based on this V1 and V2 replacement construct were named pKorix2, where x corresponds to the name of the LIR deletion.

Partial dimers with a truncated second LIR were constructed using the complementary sense ORF replacement plasmid pLVScc2 (4.3.2.1), which has a single *Xba*I site for the insertion of the truncated LIRs (see Fig. 5.2). The plasmid was completely digested with *Xba*I and treated with calf intestinal phosphatase (Boehringer Mannheim) prior to ligation with the relevant LIR clone. A series of clones containing deletions to the second LIR were constructed and given the name pKorix3, where x corresponds to the name of the deletion series. All final 1.1mer pKori clones were sequenced in one direction across the mutated LIR (Di James, Dept of Microbiology Sequencing Service, University of Cape Town).

5.3.1.1: Construction of LIR deletions

A series of deletions to either the complementary or virion sense sides of the LIR was made by the digestion of pLIR with the appropriate enzymes as indicated in Figure 5.2. All fragments were initially recovered from pLIR1 digested first with *Pvu*II and the released fragment was separated by gel electrophoresis in a 0.8 % agarose gel. The DNA was recovered using the Geneclean™ kit and then redigested with the appropriate enzymes. The DNA was then either cloned into pBluescript(SK) digested with the same enzymes, or with enzymes allowing compatible ends to be ligated. Where necessary the LIR deletion clones were either re-orientated by digestion with *Kpn*I and *Sac*I and cloned into the same sites of pUC18, which resulted in flanking *Xba*I sites. Where both the original flanking *Xba*I sites were maintained the LIR mutant it was then cloned directly into the monomeric constructs pLVScc2 or pKEP151 (see Fig. 5.2).

Two constructs were made with partial deletions in both the first and second LIR. The LIR fragments from pKori21 and pKori1 (see Fig. 5.2) were excised by digestion with *Xho*I and *Bam*HI and ligated into the same sites of pKEP176 (see 4.3.2.2), which resulted in pKori24 and pKori14, respectively. These two clones have a partial LIR fragment upstream of the V1, GUS and *nos* 3' terminator. The *Sac*I fragment of pKori13 and pKori23 (see Fig. 5.2) both containing the SIR, *cat*, CaMV 35S promoter and mutated LIR was cloned into the *Sac*I site between the GUS and *nos* 3' terminator of pKori14 and pKori24, respectively. The resultant plasmids, with two mutated LIRs, were called pKori15 and pKori25, respectively.

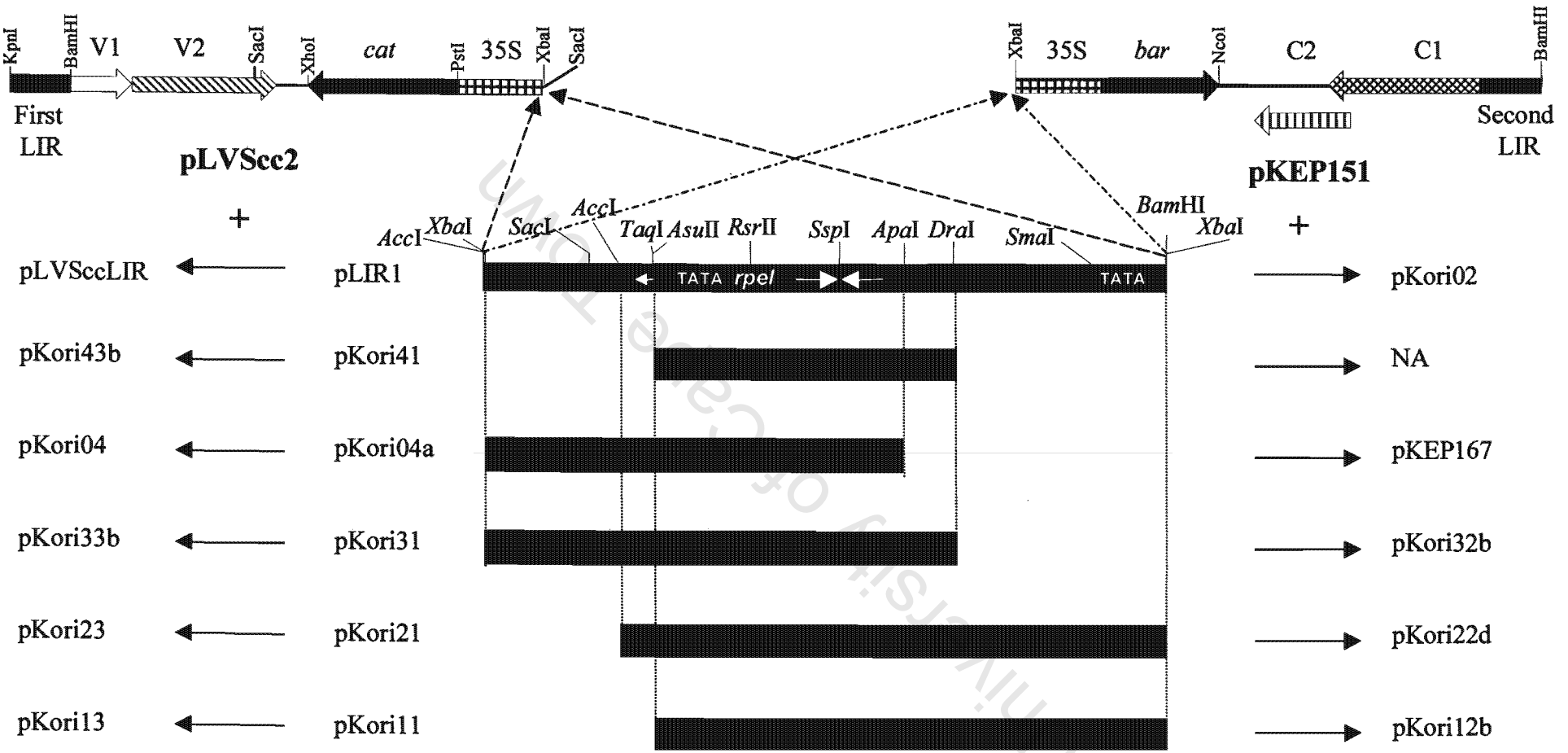


Figure 5.2: Construction of partial dimers with one deletion mutant LIR, either the second in pLVScc2 based constructs, or the first in constructs based on pKEP151. Plasmid pLIR1 was subjected to restriction digestion (enzymes used are highlighted by the dashed line) and the resultant subclones, after re-orientation in pUC18, were cloned into the *Xba*I site of the monomeric constructs pKEP151 and pLVScc2. The LIR is indicated by a black rectangle in all constructs, while the indirect repeats forming the stem-loop and the Rep proximal iteron are highlighted in white. The relative positions of the *rpeI* element and the TATA boxes are indicated in white in pLIR1. 35 S indicates the CaMV 35S promoter and NA is not applicable.

5.3.1.2: Transient replication assays

These were performed essentially as described in 4.3.3, using microprojectile bombardment of three-day old BMS suspension culture cells. Each bombardment carried 200ng of the pKori plasmid, and if required, 200 ng of the partial dimer pKom602 or the Rep expression cassette pKEP132 (see Table 5.1 for a full list). The bombardments were performed in duplicate and the experiment repeated at least twice. Three days after bombardment the low molecular weight DNA was harvested (4.3.3) and 15 µg of each sample was subjected to gel electrophoresis and Southern blot analysis using a Dig-labeled probe MSV-Kom (3.3.4) or CAT probe (4.3.3.6). Replication was also detected using the PCR primers 215 – 234 (2.3.2) and CaMV 35S primer (4.3.4) and the same amplification parameters as before.

Table 5.1: Plasmids bombarded into BMS for transient replication assays

Constructs	**First LIR**	**Second LIR**	**Co-bombarded plasmid**[1]	
pKori02	C1[2] – *Bam*HI	C1 – *Bam*HI	-	-
pKori22d	*Acc*I – *Bam*HI	C1 – *Bam*HI	-	-
pKori12b	*Taq*I – *Bam*HI	C1 – *Bam*HI	-	-
pKori32b	*Xba*I – *Dra*I	C1 – *Bam*HI	-	-
pKEP167[3]	C1 – *Apa*I	C1 – *Bam*HI	-	-
pKEP151	no LIR	C1 – *Bam*HI	-	-
pLVSccLIR	C1 – *Bam*HI	C1 – *Bam*HI	pKom602	pKEP132
pKori33b	C1 – *Bam*HI	*Xba*I – *Dra*I	pKom602	pKEP132
pKori04	C1 – *Bam*HI	*Xba*I – *Apa*I	pKom602	pKEP132
pKori23	C1 – *Bam*HI	*Acc*I – *Bam*HI	pKom602	pKEP132
pKori13	C1 – *Bam*HI	*Taq*I – *Bam*HI	pKom602	pKEP132
pKori43b	C1 – *Bam*HI	*Taq*I – *Dra*I	pKom602	pKEP132
pLVScc2	C1 – *Bam*HI	no LIR	pKom602	pKEP132
pKori15	*Taq*I – *Bam*HI	*Taq*I – *Bam*HI	pKom602	pKEP132
pKori25	*Acc*I – *Bam*HI	*Acc*I – *Bam*HI	pKom602	pKEP132

[1]: Each of the constructs listed in the first column was co-bombarded with pKom602 in one series of bombardments and with pKep132 in the next, unless this was not necessary (-) due to the presence of an intact C1/C2 ORF in the construct being investigated.
[2]: C1 indicates that the LIR sequence starts from the C1 ORF ATG
[3]: pKEP167 is a full monomer from pKEP151 cloned in tandem with the same monomer, which had been digested with *Apa*I , thereby removing most of the virion sense side of the LIR: this would result in a replication non-proficient viral clone (Palmer, 1997).

5.3.2: Construction of agroinfectious clones with recombinant LIRs

MSV-Kom based monomeric clones with varying amounts of MSV-Set LIR sequences were obtained from W. H. Schnippenkoetter (Schnippenkoetter, 1998). These clones were constructed using viral genomic DNA isolated from plants agroinfected with full tandem heterodimers in the case of pSek112, 113 and 118, or a 0.5mer of MSV-Set and 1 mer of MSV-Kom in the case of pSek009. These constructs were either the result of homologous recombination events between heterodimers of MSV-Kom and MSV-Set (pSek112 and pSek009) or presumptive aberrant replicative release events resulting in small insertions within the stem sequence (pSek315 and pSek318) (Schnippenkoetter, 1998). The sequences of the LIRs are presented in Fig. 5.3.

The monomeric *Bam*HI clones pSek112, pSek115, pSek118 and pSek009 were digested with *Sac*I and *Bam*HI to release the fragment containing the LIR. These fragments were then ligated into *Sac*I and *Bam*HI treated pUC18 to form the 0.1mers pSek212, pSek215, pSek219 and pSek009.1. The monomeric clones were treated with *Bam*HI and *Sca*I to release the monomer from the vector and to separate the similarly sized vector, which has a *Sca*I site, and viral DNA bands during electrophoresis. The viral DNA was purified from the gel using Geneclean™ and ligated to the appropriate *Bam*HI - and calf intestinal phosphatase - treated 0.1mer. The resultant clones were called pSek312, pSek315, pSek318 and pSek309. The binary vector pBI121 (Clonetech, Palo Alto, CA) was treated with *Eco*RI and *Xba*I, which released the GUS gene and *nos* terminator fragment, and ligated to the similarly treated and recovered pSek3 series of 1.1mer fragments. The resultant clones were named pSek412, pSek415, pSek418 and pSek409 (see Fig. 5.4).

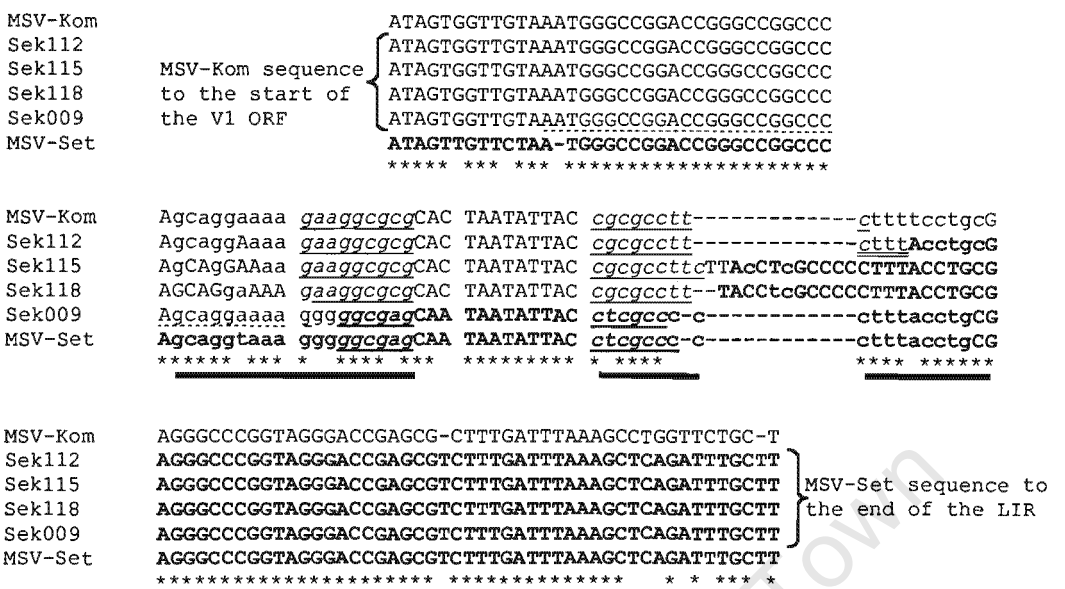


Figure 5.3: Partial LIR sequences of the wild type MSV-Kom and MSV-Set compared to those of the recombinant virus genomic clones Sek112, Sek115, Sek118 and Sek009. Sequences in bold indicate WT MSV-Set sequence. The putative iterons, as identified by W. H Schnippenkoetter (1998), are indicated in italics and are underlined. Nucleotides which may form part of the stem structure (bold line underneath the sequence) are in lower case, with non-complementary pairs in upper case. Asterisks indicate nucleotides conserved amongst all six sequences. The possible recombination site between MSV-Kom and MSV-Set resulting in Sek009, is indicated by a stippled line and that of Sek112 by a double line (Schnippenkoetter, 1998). Alignments were performed using DNAMAN 4.0.

5.3.3: Directed genomic exchanges between MSV-Kom and MSV-Set

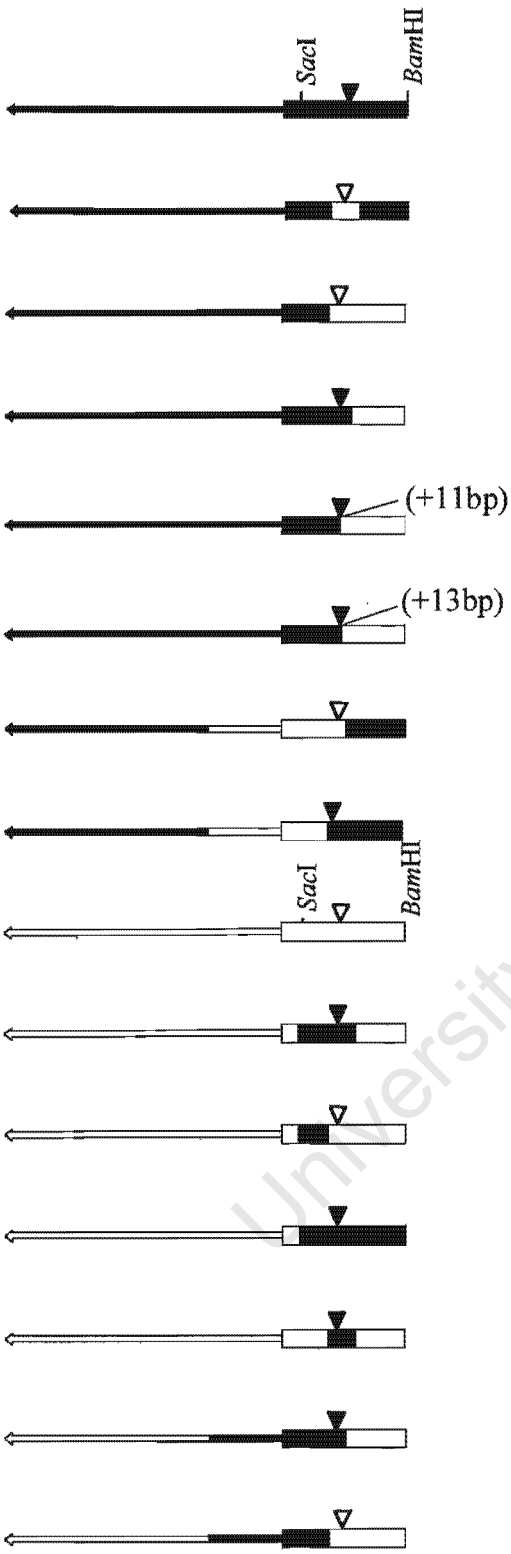
5.3.3.1: Complementary sense side of LIR and stem-loop exchanges

For the purposes of this study, regions flanked by convenient restriction enzymes, which were exchanged between the MSV-Kom and MSV-Set LIR, have been designated alphabetical letters for ease of reading (see Fig.5.1). The monomeric clones of MSV-Kom and MSV-Set (pKom500 and pSet100, respectively) were digested with *Nsi*I and *Apa*I, releasing a 377 bp band for MSV-Set and a 374 bp band for MSV-Kom from the monomer/vector fragment. The fragments were exchanged between the two resulting in MSV-Kom with a *Nsi*I to *Apa*I fragment of MSV-Set and *vice versa* (A – C exchanges). The chimaeric clones were named pKNAS and pSNAK, with the first letter indicating the

parental viral species present (i.e. MSV-**K**om or MSV-**S**et), the next two letters corresponding to the sites used for cloning and the last letter indicating the origin of the exchanged viral fragment (see Fig. 5.4 for a complete list of clones). The same convention for clone names was continued for all subsequent exchanges. The clone pKNAS is missing 10 bp at the reconstituted *Nsi*I site due to the presence of two closely positioned *Nsi*I sites in MSV-Set, and therefore has a frameshift mutation within the C1 ORF. The reciprocal exchange in pSNAK gives an in-frame C1 ORF.

The complementary sense strand side of the LIR up to the *rpeI* element (Fenoll *et al.*, 1990; Fenoll *et al.*, 1988), and the first 185 bp of the C1 ORF were exchanged between MSV-Set and MSV-Kom (i.e. regions A and B). The GC-rich *rpeI* element is specifically recognised by *Rsr*II in both MSV-Kom and MSV-Set; therefore, exchanging the 811 bp *Xho*I to *Rsr*II fragments between pKNAS and pKom500 resulted in the clones pKNRS and pKRAS, which have the sequences between the *Nsi*I and *Rsr*II site and the *Rsr*II to *Apa*I fragments exchanged from MSV-Kom to MSV-Set sequences, respectively. Plasmid pKNRS has the same *Nsi*I – *Nsi*I deletion as pKNAS. The same approach was used for pSNRK, except that MSV-Set has two *Rsr*II sites, therefore the *Rsr*II fragment 2227 bp fragment from pSet100 was exchanged for the same region in pSNAK. The resultant plasmids from the *Rsr*II exchanges were called pSNRK and pSRAK#21 (see Fig 5.4).

The MSV-Kom Rep gene 5' terminus of pSNRK and pSNAK based clones was restored to WT MSV-Set sequence by the exchange of the smaller fragment (A) from *Sac*I digested pSet100 into the corresponding sites of the chimaeric monomers. This was achieved by recovering the larger of the two bands, which includes the vector sequence, from the gel electrophoresed *Sac*I digestion products of pSNRK and pSNAK. The recovered DNA was then treated with calf intestinal phosphatase prior to ligation with the pSet100 *Sac*I smaller fragment. The final constructs were then named pSSRK and pSSAK. MSV-Kom has 4 *Sac*I sites (see Fig. 5.1) making similar exchanges between pKNRS and pKNAS with pKom500, using general cloning techniques, unfeasible.



Monomer	Partial LIR 0.1mer	Dimer 1.1 mer	Binary Vector
pKom500[1]	NA	pKom602[1]	pKom603[1]
pKRASb	pKRASb.1	pKRASb2.1	pKRASb3.1
pSek009[1]	pSek209	pSek309	pSek409
pSek112[1]	pSek212	pSek312	pSek412
pSek115[1]	pSek215	pSek315	pSek415
pSek118[1]	pSek218	pSek318	pSek418
pKNAS	pKNAS.1	pKNAS1.1	NA
pKNRS	pKNRS.1	pKNRS1.1	NA
pSet100[2]	NA	pSet105[1]	pSet107[1]
pSSAK	pSNAK.1	pSSAK2.1	pSSAK3.1
pSSRK	pSNRK.1	pSSRK2.1	pSSRK3.1
NA	NA	pKesek1[1]	pKesek2[1]
pSRAK7	pSRAK#21	pSRAK77	pSRAK773.1
pSNAK	pSNAK.1	pSNAK1.1	pSNAK2.1
pSNRK	pSNRK.1	pSNRK1.1	pSNRK2.1

Figure 5.4: The C1 ORF and LIR of MSV-Kom (black) and MSV-Set (white) chimaeric clones are represented on the left hand side with their respective names on the right. The partial LIR clones, the 1.1mers, and pBI121 clones are listed. The position of the *Sac*I site used for construction of all the partial LIR clones, relative to the conserved *Bam*HI site, is shown above pKom500 and pSet100. Where clones were either not made or not relevant to the study they are indicated by NA (not applicable). [1]: Clones provided by W. H. Schnippenkoetter (1998). [2]: Clone provided by F. L. Hughes (Hughes *et al.*, 1992). The stem-loop structure is represented by a triangle.

5.3.3.2: Construction of agroinfectious clones

The monomeric clones pSNAK, pSNRK, and pSRAK were digested with *Sac*I and the vector sequence with the remaining partial LIR (see Fig. 5.4) was self-ligated. The resultant 0.1mers were named pSNAK.1, pSNRK.1 and pSRAK#21. Partial dimers were then constructed using the unique *Bam*HI site in the 0.1mer as the site for the insertion of the monomeric viral sequence obtained after excision and purification from the *Bam*HI and *Pvu*I digested pSNAK, pSNRK and pSRAK7. No *Pvu*I sites are present in the MSV-Set sequence allowing the separation of the viral from vector sequence during gel electrophoresis. The clones pSNRK.1 and pSNAK.1 were used for the construction of partial dimers of pSSRK and pSSAK, respectively, as they share the same chimaeric LIR sequence. The resultant clones containing 1.1mers were named pSNAK1.1, pSNRK1.1, pSRAK77, pSSRK1.1 and pSSAK1.1. The 1.1mers were partially digested with *Eco*RI and completely with *Xba*I to release the viral DNA for ligation into the *Eco*RI, *Xba*I digested binary vector, pBI121 (Clonetech, Palo Alto, CA). The plasmids were then named pSNAK3.1, pSNRK3.1, pSRAK773.1, pSSRK3.1 and pSSAK3.1.

Due to the presence of the frameshift mutation within the C1 ORF of pKNRS and pKNAS only pKRASb was made agroinfectious. A 0.1mer clone was constructed by digestion of pKRAS with *Bam*HI, *Sac*I and *Eco*RV, which preferentially allows only the partial LIR sequence flanked by the *Bam*HI and *Sac*I sites to be religated back into the pUC vector. The resultant clone pKRASb.1 was then digested with *Bam*HI and treated with calf intestinal phosphatase. The *Bam*HI monomer fragment excised from the plasmid pKRASb (as above), was ligated to the *Bam*HI digested 0.1 mer pKRASb.1. The resultant plasmid was named pKRASb1.1. The partial dimer was then cloned into the *Xba*I and *Eco*RI sites of pBI121 resulting in plasmid pKRASb3.1 (see Fig. 5.4).

5.3.3.3: Transformation and agroinfection of A. tumefaciens

The following plasmids were transformed into *A. tumefaciens* C58C1 (pMP90) (Koncz and Schell, 1986) using the freeze-thaw method of Holsters *et al.* (1978) (3.3.2): pSek412, pSek415, pSek418, pSek409, pSNRK3.1, pSNAK3.1, pSRAK773.1 and pKRASb3.1. The same procedure for the preparation and injection of the transformed *A. tumefaciens*

into 3 day-old sweetcorn cv. Jubilee was used as described in Chapter 3 (3.3.3). The plants were monitored every three days and the symptoms recorded after 15 and 22 days. The percentage of infected plants after 29 days was recorded. For comparison the agroinfectious chimaeric 1.1 mer, pKesek2, a partial LIR replacement of the MSV-Kom *Sac*I to *Bam*HI fragment (regions B, C & D) into the MSV-Set backbone (Schnippenkoetter, 1998), was agroinfected under the same conditions as above. Each agroinfection experiment was repeated on three separate occasions.

5.3.3.4: Extraction and analysis of DNA from agroinfected plants

The viral genomic DNA was harvested from the agroinfected plants (3.3.4) and either subjected to Southern blot analysis or to degenerate primer PCR amplification and subsequent RFLP analysis using *Cfo*I as described previously (2.3.2). The expected banding patterns after restriction digest of the PCR product with *Cfo*I are shown in Table 5.2. These patterns were generated using DNAMAN 4.0 based on the GenBank sequences of MSV-Kom and MSV-Set (see Table 3.1).

Table 5.2: Predicted size fragments obtained after restriction digestion of the 1300 bp PCR amplification product with *Cfo*I

Plants agroinfected with:	Expected PCR RFLP band sizes:
pKom603	18, 37, 80, 319, 853 bp
pSek415	18, 80, 370, 856 bp
pSek418	18, 80, 368, 856 bp
pSek412	18, 80, 360, 856 bp
pSek409	455, 853 bp
pKRASb3.1	135, 319, 853 bp
pSet107	11, 25, 601, 684 bp
pSSAK3.1	11, 18, 80, 216, 357, 601 bp
pSSRK3.1	11, 216, 455, 601 bp
pSRAK773.1	11, 18, 25, 309, 358, 601 bp
pKesek2	11, 18, 25, 37, 80, 216, 319, 601 bp

5.3.3.5: Construction of Rep mutants for trans*-replication assays*

The MSV-Set based 1.1mers, pSRAK1.1, pSSRK1.1, pSSAK1.1, pKesek1 (the 1.1 mer from which the agroinfectious construct pKesek2 was made (Schnippenkoetter, 1998)) and pSet105 (Schnippenkoetter, 1998) were digested with *NsiI*, releasing the 10bp *Nsi*I - *Nsi*I fragment within the C1 ORF, which would cause the translation product to prematurely terminate after 60 amino acids. All resultant plasmids were screened for the deletion and positive clones given the same basic name as the parental clone with the suffix *Nsi*I e.g. pSRAK1.1 became pSRAK*Nsi*I.

The MSV-Kom based constructs pKNRS and pKNAS have the C1 *Nsi*I fragment missing, and are therefore self-replication incompetent. They were both dimerised using the following strategy: 0.1mer partial LIR *Sac*I to *Bam*HI subclones were first constructed in pUC19, resulting in plasmids pKNRS.1 and pKNAS.1, and then the full *Bam*HI – *Bam*HI genomic fragment from pKNRS and pKNAS was cloned into the *Bam*HI site of their respective partial LIR subclone. The final 1.1 mers are named pKNRS1.1 and pKNAS1.1. The MSV-Kom *Pst*I frameshift mutant pKEP177 (Palmer, 1997), a full dimer, was used as the WT LIR control. This plasmid, provided by K. E. Palmer, contains the same site directed mutation as pKEP171*gus*d (4.3.2.2) (Palmer, 1997). The same *Pst*I mutation was introduced into pKRASb2.1 by the replacement of its *Xho*I to *Rsr*II fragment with that of pKEP177. The mutation allows the translation of a truncated protein with 24 Rep amino acids followed by 10 missense amino acids before a premature termination codon is reached (Palmer, 1997). This Rep$^-$ plasmid was named pKRAS*Pst*I.

5.3.3.6: Transient trans*-replication assay in BMS cells*

The 1.1mers pKNRS1.1 and pKNAS1.1 were initially bombarded into BMS and the lack of replication was confirmed by both Southern blot analysis and PCR (data not shown). The insertion of the *Pst*I frame shift mutation had previously been confirmed to render pKEP171 and p171luci self-replication incompetent (Palmer, 1997 & Chapter 4).

The Rep$^-$ mutants were co-bombarded with either pKom602 or pSet107 into 3 day-old BMS suspension culture cells and the DNA was harvested as previously described, except

that for this experiment the cells were harvested after four days. The experiment was performed in duplicate using 200ng of each plasmid per bombardment. The extracted genomic DNA was subjected to degenerate primer PCR (see 2.3.2) and *Cfo*I digestion. A total of 5 µg of genomic DNA from the DNA from the MSV-Kom based bombardments was treated with *Dpn*I (Boehringer Mannheim) to remove any residual input methylated (i.e. bacterial derived) DNA. DNA samples obtained from the pKRAS*Pst*I and pKEP177 bombardments were also digested with *Sal*I and *Pst*I, while those obtained from the pKNRS1.1 and pKNAS1.1 bombardments were digested with *Bam*HI and *Bgl*II. The DNA was then electrophoresed in 0.5 x TBE through a 0.8 % agarose gel and subjected to Southern blot analysis using the Dig-labeled MSV-Kom PCR derived probe (3.3.4).

5.4: RESULTS

5.4.1: Construction of the truncated LIR partial dimers

The cloned monomer containing the CAT gene marker, pKomLVS, was just over 200bp larger in size than the monomer of MSV-Kom (pKom500), whereas the *bar* containing mutant, pKEP151, was only 33 bp larger (Palmer, 1997). The partial dimers, pKori02 – pKori43b, pKori15 and pKori25 were all sequenced through the mutated LIR, which was shown to be correct in sequence and orientation.

5.4.2: Transient replication results of LIR deletion mutants

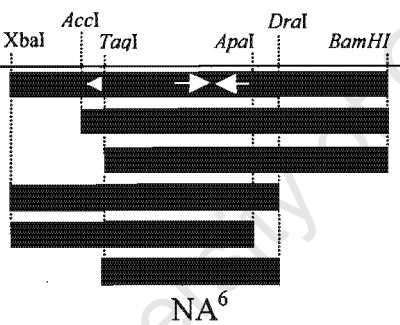
5.4.2.1: Replication of second LIR mutants

Three days after bombardment of the BMS cells the low molecular weight DNA was extracted and analysed for the presence of replicative form DNA. Replicative form DNA was detected by two different methods: firstly, low molecular weight DNA was detected by Southern blot analysis; and secondly, by PCR amplification using primers that would only amplify replicating DNA (see 4.3.4). The Dig-labeled CAT probe (4.3.4) detected RF DNA only for pKori04, pKori33b and the WT LIR construct pLVSccLIR, when co-bombarded with pKom602 (data not shown) (See Table 5.3 for full summary of results). Co-bombardment with pKEP132 gave the same positive results, except that faint bands

for pKori23, pKori13 and pKori43b were also obtained, indicating that in an excess of Rep these previously self-replication impaired constructs could weakly amplify (see Fig. 5.5A). This was also observed by PCR analysis (data not shown). Similar results were observed using WDV LIR mutations (Sanz-Burgos and Gutiérrez, 1998). No replicative form DNA was detected with the monomeric construct, pLVScc2, using either method. No high molecular weight DNA, indicating initiation and termination of replication at the unmutated LIR, was detected by Southern blot analysis, possibly due to the lower level of accumulation of these products and the relatively short time after bombardment that the DNA was harvested. Furthermore, the PCR assay would only detect replicating DNA that had reconstituted a continuous strand of DNA from the V1 ORF through the LIR.

Table 5.3: Summary of results obtained from analysis of the DNA harvested from the transient replication assay of the deletion LIR mutants.

Plasmid name	Second LIR deletion mutants[1]	Southern blot	PCR	Southern blot	PCR
		Co-bombarded with pKom602		Co-bombarded with pKEP132	
pLVSccLIR		+[2]	+	+	+
pKori13		-[3]	-	+/-[4]	(+)[5]
pKori23		-	-	+/-	(+)
pKori33b		+	+	+	+
pKori04		+	+	+	+
pKori43b		-	-	+/-	(+)
pLVScc2	NA[6]	-	-	-	-



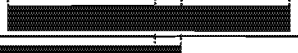
[1]: Restriction sites used to construct deletion mutants of pLIR1 (see Fig. 5.2); [2]: +, positive reaction; [3]: - negative reaction; [4]: +/-, limited detection; [5]: (+), weakly positive; and [6], NA, not applicable.

5.4.2.2: Replication of first LIR mutants

All of the 1.1 mers with deletions on the complementary sense strand side of the first LIR were replication proficient (see Table 5.4). Even if no replicative release was initiated from the first LIR, homologous recombination between the mutated LIR and the second LIR in each construct would reconstitute a WT LIR (See Table 5.5). This would therefore render the assay incapable of detecting the initial lack of replication due to inefficient initiation. The only construct with a 3' deletion which replicated as efficiently as pKori02, was pKori32b (see Fig. 5.5B). Two constructs did not replicate via either homologous recombination or replicative release events: these were pKEP151 and pKEP167. The latter had previously been shown to be replication incompetent (Palmer, 1997) and the

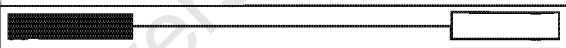
former has only one LIR. The 25 bp between the *Apa*I site at the base of the stem-loop and the *Dra*I site is therefore necessary for replication.

Table 5.4: Identification of replication proficient first LIR mutants from DNA harvested from the transient replication assays in BMS

Plasmid name	Deletion mutants of the first LIR[1]	Southern blot	PCR
	XbaI *Acc*I *Taq*I *Apa*I *Dra*I *Bam*HI		
pKori02		+[3]	+
pKori22d		+	+
pKori12b		+	+
pKori32b		+	+
pKEP167		-[4]	-
pKEP151	NA[2]	-	-

[1]: The restriction enzymes used to construct the various deletant mutants of pLIR1 are aligned above the schematic diagram of the LIR (see Fig. 5.2). [2]: NA, not applicable; [3]: (+) positive reaction; [4]: (-) negative reaction.

Table 5.5: Analysis of the type of LIR formed should either of the two mechanisms responsible for the release of a viral replicon occur.

Plasmid name	Relative positions of the WT (white) and mutant (black) LIRs	Homologous recombination	Replicative release
pKori02		WT	WT
pKori22d		WT	WT
pKori12b		WT	WT
pKori32b		mutant	mutant
pKEP167		mutant	mutant
pLVSccLIR		WT	WT
pKori13		mutant	mutant
pKori23		mutant	mutant
pKori33b		WT	WT
pKori04		WT	WT
pKori43b		mutant	mutant

5.4.2.3: Replication of double mutated LIRs

No replicative form DNA was detected after co-bombardment of pKori15 and pKori25 with either pKom602 or pKEP132. Replication was assayed for by both Southern blot analysis and PCR, using primer 215 – 234 and CaMV 35S primer (4.3.4), but in neither case was a positive reaction obtained (data not shown).

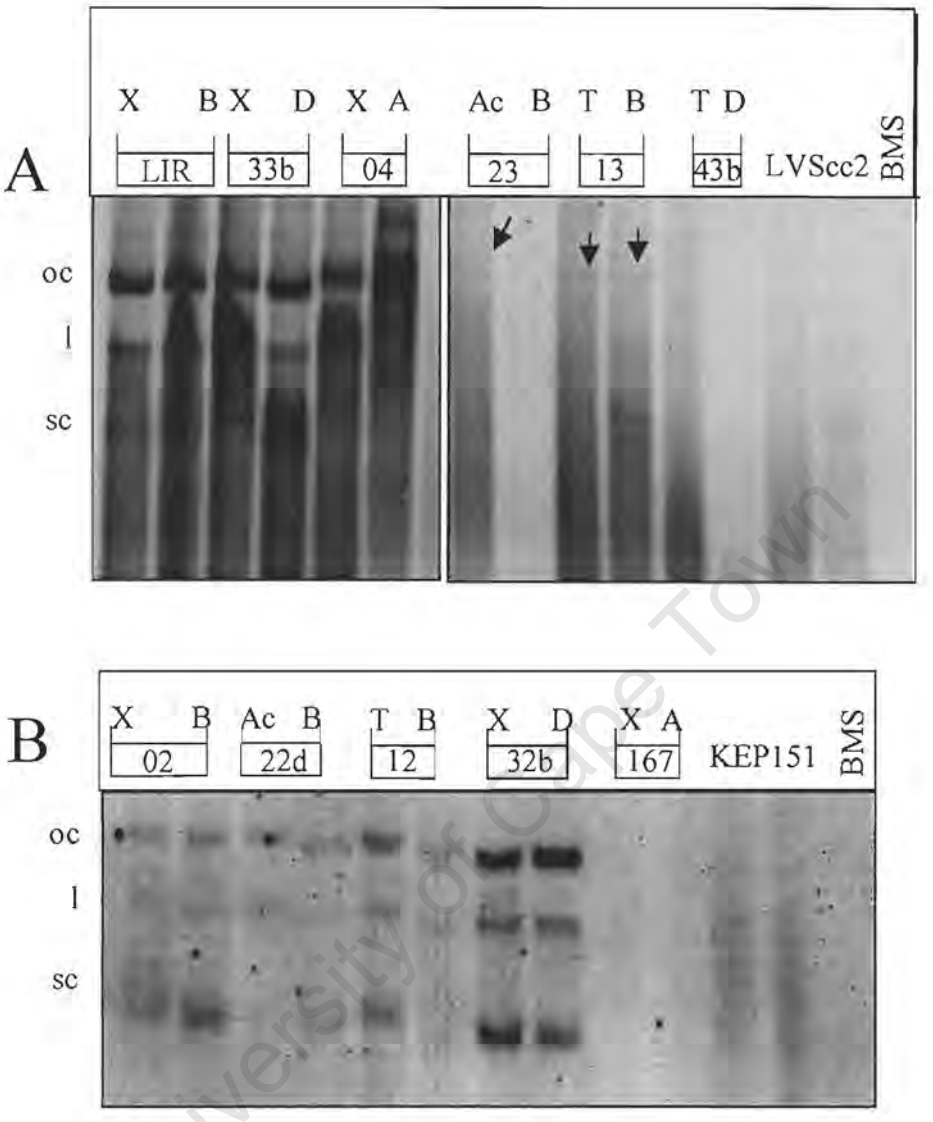


Figure 5.5: Southern blot of DNA extracted from BMS bombarded with second LIR mutant partial dimers and pKEP132 (A) and first LIR mutant partial dimers (B). Two BMS plates were bombarded for each precipitation and the extracted DNA electrophoresed through a 0.8 % agarose gel prior to Southern blot analysis. The various replicative forms associated with replicating DNA are indicated: oc, open circular; l, linear; and sc, supercoiled DNA. The Southern blot was probed with PCR generated Dig-labeled MSV-Kom (3.3.4). The arrows indicate the presence of faint replicative form DNA. The rectangles represent the proportional size of the mutated LIR, the number of the pKori clone or pKEP clone, and the restriction sites used to construct the partial LIRs are as follows: X, *Xba*I; B, *Bam*HI, D, *Dra*I; A, *Apa*I; Ac, *Acc*I; and T, *Taq*I. DNA was extracted from unbombarded BMS cells as a negative control.

5.4.3: Analysis of agroinfectious constructs with chimaeric LIRs

All plasmids constructed with LIR exchanges between MSV-Kom and MSV-Set were first cloned as 1.1 mers and then cloned into the binary vector, pBI121, and transformed in *A. tumefaciens* C58C1 (pMP90). The transformed *A. tumefaciens* was then injected into three day old sweetcorn cv. Jubilee seedlings. The number of surviving plants showing symptoms 29 days after agroinfection was counted in three separate experiments (see Table 5.6). The symptoms were compared after 15 days and 22 days (See Fig. 5.6 and data not shown). Plants injected with *A. tumefaciens* transformed with pSNRK3.1 and pSNAK3.1 remained symptomless and pSRAK773.1 gave extremely mild chlorotic symptoms down the mid-rib vein of the second leaf and occasionally on the third, in two separate experiments. No subsequent leaves showed any symptoms (see Fig. 5.6B). Infection rates were monitored every three days but proved to be unreliable measures of pathogenicity due to high standard deviation between experiments, especially with plants showing mild symptoms.

Table 5.6: The percentage (%) of sweetcorn cv. Jubilee plants showing symptoms after agroinfection with chimaeric MSV-Kom and MSV-Set constructs

	Experiment 1		Experiment 2		Experiment 3		
Name of dimer in pBI121	**#**	**%**	**#**	**%**	**#**	**%**	**Average %**
pSek412	4/8	50	7/12	58.3	11/13	84.6	64.3
pSek415	6/8	75	6/9	66.6	12/12	100	80.5
pSek418	10/12	83.3	8/11	72.7	11/12	91.3	80.4
pSek409	11/13	84.6	24/29	82.7	23/24	95.8	87.7
pKRASb3.1	9/10	90	24/26	92.3	22/23	95.6	92.6
pKesek2	6/6	100	4/6	66.6	12/12	100	88.8
pSSAK3.1	10/11	90.9	25/28	89.2	18/22	81.8	87.3
pSSRK3.1	5/6	83.3	16/26	61.5	9/18	80	74.9
pSNAK3.1	0/8	0	0/11	0	-	-	0
pSNRK3.1	0/10	0	0/9	0	-	-	0
pSRAK773.1	8/24	33.3	10/24	41.6	-	-	37.4
pBI121	0/12	0	0/12	0	0/12	0	0

#: Number of plants infected/number that survived agroinfection

Of all the plants infected with MSV-Kom based genomes, pSek412, pSek415 and pSek418 had the mildest symptoms with pSek412 being consistently the weakest. pSek415 agroinfected plants had milder symptoms than those agroinfected with pSek418, but symptoms became more pSek418 like in appearance in the older leaves (data not shown). The insertion of 11 to 13 bp within the stem-loop sequence potentially allows some base pairing to occur (see Fig. 5.3 & Schnippenkoetter, 1998), but these mutant viruses are vastly attenuated in symptoms compared to the wild-type MSV-Kom symptoms observed in Chapter 3.4.1 (see Fig. 5.6A). Approximately 800 bp of sequence from either side of the *Bam*HI site of the monomeric clones, pSek112, pSek115 and pSek118 was obtained, and except for the regions indicated in Fig. 5.3, all the clones have WT MSV-Kom sequence (Schnippenkoetter, 1998). It is not known whether the remainder of the genome has undergone any small rearrangements, which may further decrease pathogenicity (W. H. Schnippenkoetter, Pers. Comm). Similarly, with pSek009 only 270 bp from the *Bam*HI site of the LIR was sequenced and 832 bp of virion sense ORFs sequenced (Schnippenkoetter, 1998). The three mutant viral constructs pSek112, pSek115 and pSek118, differed further from pSek009, which had the entire stem-loop sequence of MSV-Set as well as the virion sense promoter. These results, together with the agroinfectious pKesek2 clone (Schnippenkoetter, 1998), are the first evidence that the MSV stem tolerates nucleotide changes and even potential changes in structure due to less base pairing occurring, similar to that observed for WDV (Heyraud *et al.*, 1993b).

The *Rsr*II to *Apa*I (region C) stem-loop exchange, from MSV-Set into MSV-Kom, a total of 68 nucleotides, has all sequence differences confined to the stem: in particular, pKRASb now has three out of eight nucleotides changed in its stem iteron sequence. The agroinfectious dimer, pKRASb3.1, infects an average of 92.6 % plants (see Table 5.6) and is only slightly milder in symptoms than the WT agroinfectious construct pKom603 (see Fig. 6A). This was confirmed by computer based analysis of the percentage chlorosis using the method of Martin and Rybicki (1998) (D. P. Martin Pers. Comm.). The reciprocal construct, pSRAK773.1, never developed full streak symptoms (see Fig. 5.6B) and only an average of 37 % of plants displayed the mild streak on the second or third leaf, and all plants recovered from the infection.

Exchanging MSV-Set sequences between the *Sac*I site and the *rpeI* element (region B) and the *Sac*I site and *Apa*I site (regions B and C) with those of the same regions in MSV-Kom, resulting in pSSRK3.1 and pSSAK3.1, rendered the symptoms observed in the mildly infectious MSV-Set even further attenuated (see Fig. 5.6B and compare Table 5.6 with Table 3.2). The stem-loop sequence of pSSAK3.1 is that of MSV-Kom, and the construct was able to replicate and cause symptoms. Exchanging the sequences between the *Sac*I site *Rsr*II sites alone (region B) decreased the mutant's pathogenicity compared to that of pSSAK3.1. Similarly, exchanging only the stem-loop sequence (region C) of a MSV-Set based construct also greatly affected its pathogenicity (see Fig. 5.6B and Table 5.6). The C1 *Nsi*I – *Apa*I (A, B & C) or *Nsi*I – *RsrII* (A & B) exchanges of MSV-Set with MSV-Kom sequences did not show symptoms.

5.4.3.1: Analysis of genomic DNA from agroinfected plants

Small-scale genomic DNA extractions by the method of Palmer *et al.* (1998) were performed on a number of leaf segments from plants showing symptoms. Southern blot analysis detected replicative form DNA in plant agroinfected with pSSAK3.1, pSSRK3.1, but not with pKesek2 or pSRAK773.1 (see Fig. 5.7). PCR amplification using the degenerate primers (2.3.2), which amplifies both MSV-Kom and MSV-Set, gave an approximately 1300 bp band for all infected plants except for those agroinfected with pSRAK773.1, pSNRK3.1 and pSNAK3.1 (data not shown). Four individual extractions from various leaves infected with pSRAK773.1, showing the mild streak, were also negative. The RFLP analysis of the PCR amplified DNA using *Cfo*I confirmed that the various exchanges between the LIR of MSV-Kom and MSV-Set were maintained during the agroinfection process (see Fig. 5.8).

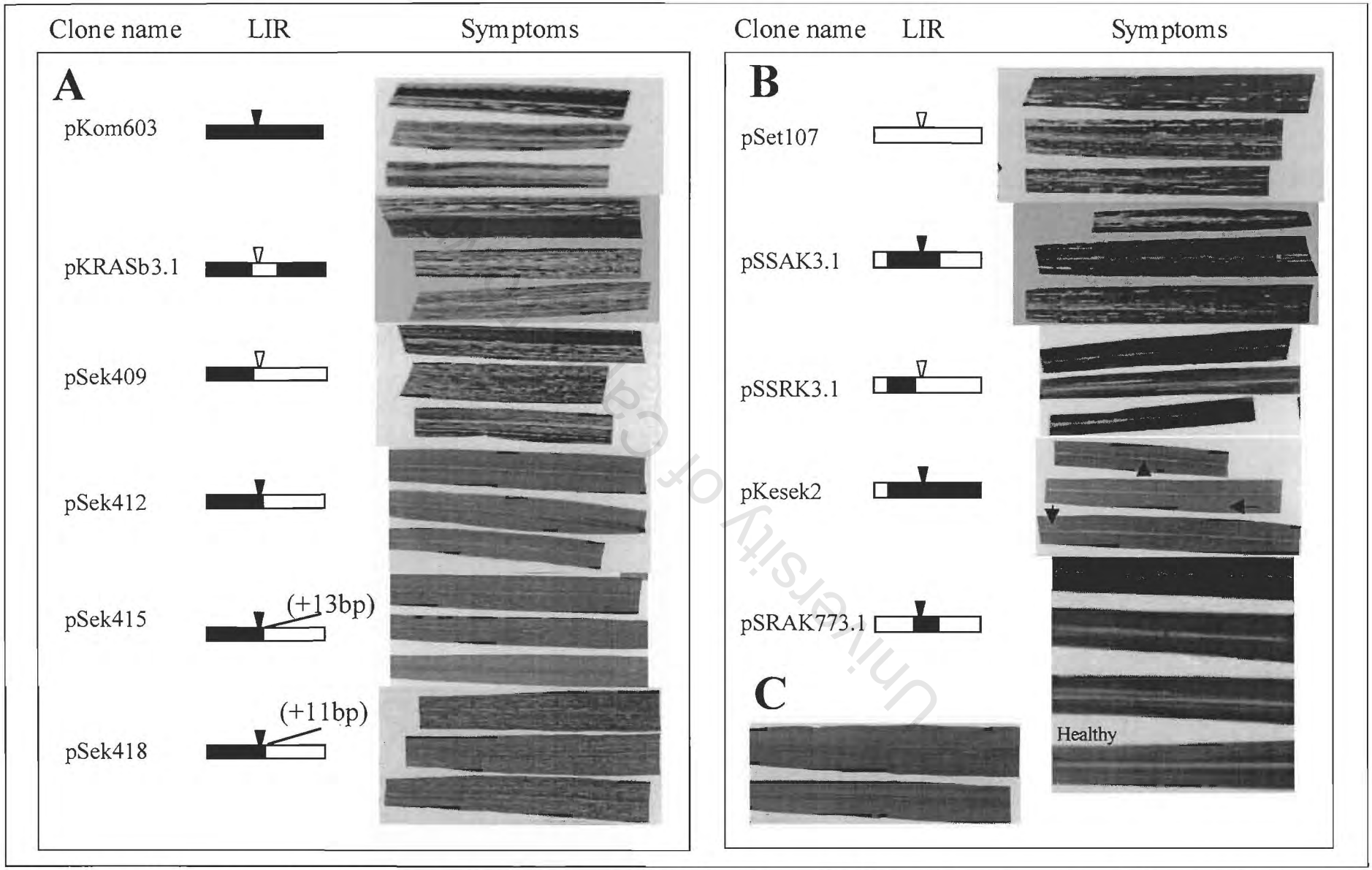


Figure 5.6: Sweetcorn cv. Jubilee agroinfected with chimaeric MSV-Kom (A) and MSV-Set (B) based constructs, indicated in black and white, respectively. Symptoms on the third leaf 15 days after agroinfection (See Table 5.6). C) Control plants agroinfected with pBI121. The triangle indicates the position of the stem-loop structure and the arrows highlight the faint symptoms.

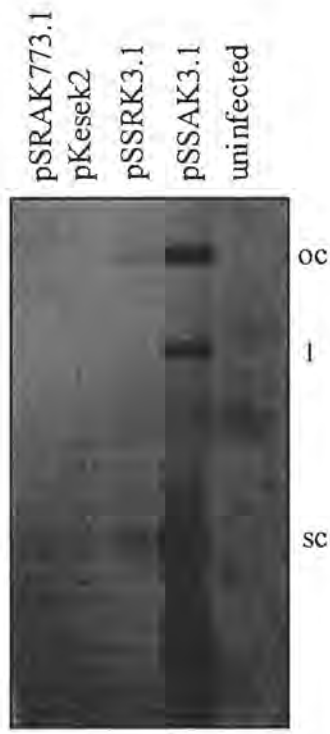


Figure 5.7: Southern blot of chimaeric MSV-Set based agroinfectious constructs. 1 µg of DNA, extracted from plants showing symptoms, was electrophoresed through a 0.8 % agarose gel. The DNA was transferred to a nylon membrane and hybridised with a MSV-Kom Dig-labeled probe, which was then washed under low stringency conditions (0.5 x SSC). The various DNA replicative forms are indicated as follows: oc, open circular; l, linear; and sc, supercoiled.

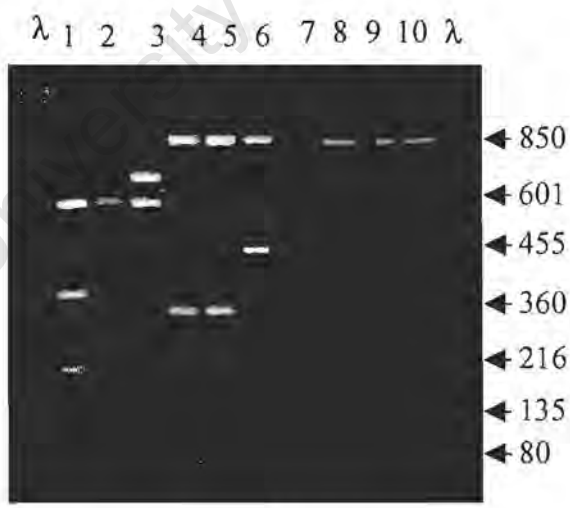


Figure 5.8: RFLP analysis of the various MSV-Kom and MSV-Set based chimaeric constructs. PCR amplification and restriction analysis of the amplification product with *Hha*I was performed on DNA was extracted from plants agroinfected with the following constructs: Lane 1, pSSAK3.1; lane 2, pSSRK3.1; lane 3, pSet107; lane 4, pKRASb3.1; lane 5, pKom603; lane 6, pSek409; lane 7, pKesek2; lane 8, pSek412; lane 9, pSek415; and lane 10, pSek418. The restriction digests were electrophoresed through a 1.5 % agarose gel stained with ethidium bromide. *Pst*I digested lambda DNA (λ) was included as a molecular weight marker and all band sizes are indicated as number of base pairs (See Table 5.2 for predicated restriction patterns).

5.4.4: Transient *trans*-replication assays

Both MSV-Set and MSV-Kom based 1.1mers were made replication incompetent; this was by the insertion of a *Pst*I site within the MSV-Kom C1 ORF, or by the deletion of the 10 bp *Nsi*I fragment from the C1 ORF of MSV-Set. Co-bombardment of each of these constructs with WT MSV-Kom or with WT MSV-Set allowed regions within the LIR containing *cis*-elements conferring replication specificity to be identified. The WT LIR in each co-bombardment experiment would presumably out-compete the chimaeric LIR for the Rep protein, unless the necessary *cis*-elements are present (Fontes *et al.*, 1994b). Neither the WT MSV-Kom LIR (pKEP177) nor the MSV-Kom stem-loop mutant (pKRASb3.1) successfully competed for the MSV-Set Rep, whereas both the Rep proximal iteron containing mutant (pKNRS1.1) and the dual iteron stem-loop mutant were recognised and replicated by MSV-Set (see Fig. 5.9A & B). The reciprocal experiments, based on MSV-Set WT LIR (pSet*Nsi*I) and the stem-loop mutant (pSRAK*Nsi*I), showed that while both were *trans*-replicated by MSV-Set Rep, neither was replicated by MSV-Kom to any detectable levels. Only once the B region, which contains the Rep proximal iteron, consisted of MSV-Kom sequence (pKesek*Nsi*I, pSSRK*Nsi*I, and pSSAK*Nsi*I) was MSV-Kom Rep able to *trans*-replicate the mutant virus (see Fig. 5.10).

5.5: DISCUSSION

Replication of geminiviruses requires the presence of specific *cis*-elements within the plus strand origin of replication found in the begomovirus common region or its mastrevirus equivalent, the LIR. The *cis*-elements required for mastrevirus replication, and in particular the replication specificity determinants, are the least well characterised of all the geminiviruses (Hanley-Bowdoin *et al.*, 1999). A minimal MSV LIR region similar to that found in WDV (Sanz-Burgos and Gutiérrez, 1998), which is essential for replication has now been delineated. This region includes a 25 bp region 3' of the stem-loop sequence, which is not involved in the DNA bending tracts identified by Gutiérrez *et al.* (1995), but is highly conserved amongst the various MSV isolates sequenced to date (see Fig. 4.8). The distance from the stem-loop of this 25 bp region could be extended by 11 - 13 bp, as evidenced with the viable agroinfectious constructs pSek418 and pSek415. This is further than the previously observed agroinfectious construct with a 4 bp linker inserted into the *Apa*I site (Schneider *et al.*, 1992).

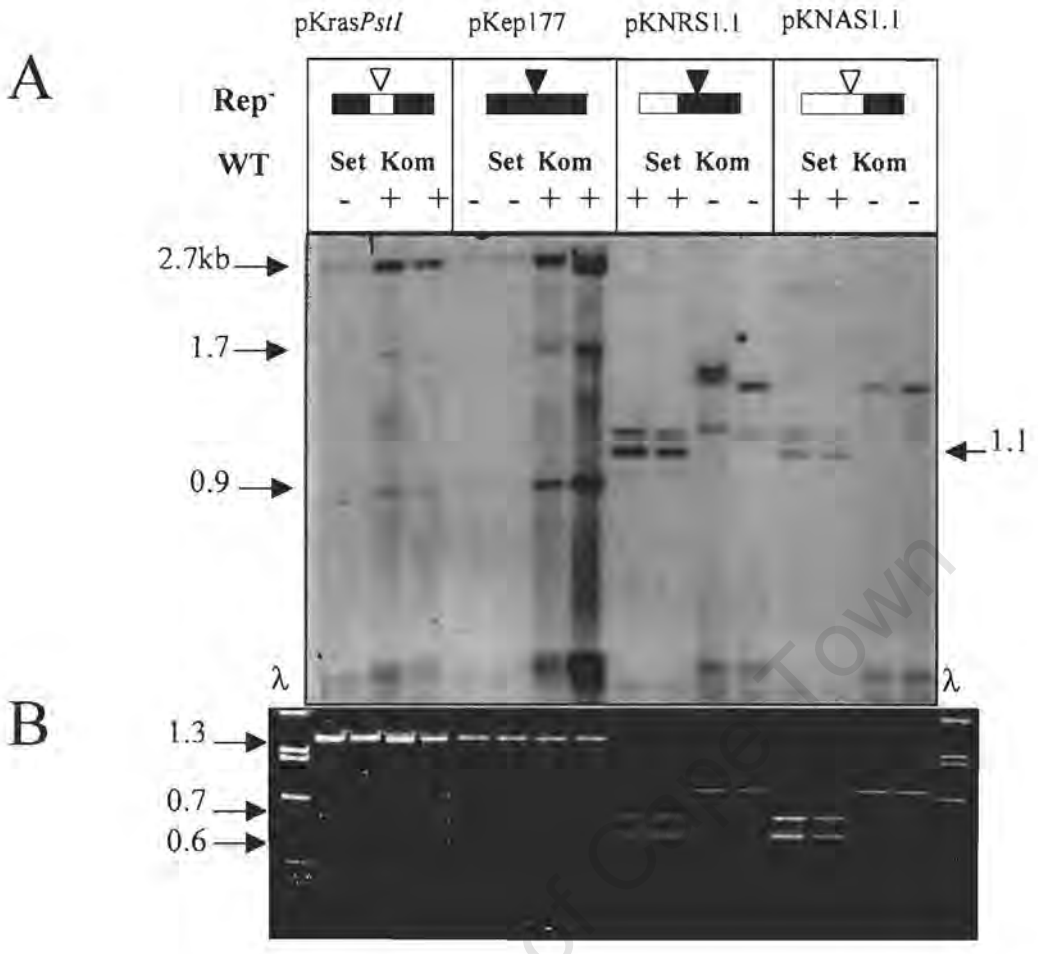


Figure 5.9: *Trans*-replication by MSV-Kom and MSV-Set of various Rep mutant MSV-Kom based chimaeric constructs. The chimaeric LIRs are represented by rectangles below its partial dimer name. Regions in black and white are MSV-Kom and MSV-Set derived sequences, respectively. The stem-loop structure is represented by a triangle. All Rep mutants were co-bombarded with either pKom602 (Kom) or pSet105 (Set) and the detection of complementation of replication functions (+), or lack thereof (-), is indicated above the respective lanes. Each bombardment was performed in duplicate. A) Southern blot, using a Dig-labeled MSV-Kom probe (3.3.4), of *Dpn*I treated DNA. The first seven lanes were digested with *Sal*I and *Pst*I. The presence of 971 and 1729 bp bands indicated that the *Pst*I Rep mutant was *trans*-replicated. The next eight DNA samples were digested with *Bam*HI and *Bgl*II, releasing a characteristic 1131 bp band from the *trans*-replicated chimaeric constructs (see Fig. 5.1). B) PCR amplification and RFLP analysis of the same DNA samples as in A, digested with *Pst*I (first eight lanes) or *Cfo*I (last eight lanes) and electrophoresed through a 1.5 % agarose gel stained with ethidium bromide. The various banding patterns either correspond to the characteristic WT virus patterns (uncut by *Pst*I, & see Table 5.2 for *Cfo*I banding patterns) or are a mixed population of WT and chimaeric virus (1304, 709 and 595 bp for *Pst*I Rep mutants and WT virus). Lambda *Pst*I treated DNA (λ) was used as a molecular weight marker.

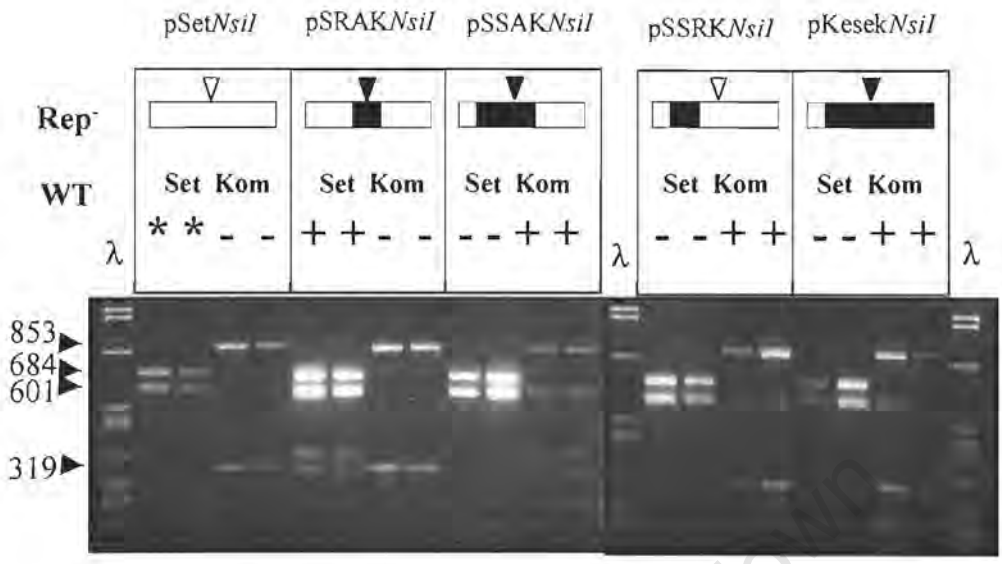


Figure 5.10: *Trans*-replication of various Rep mutant MSV-Set based chimaeric constructs by MSV-Kom and MSV-Set. The chimaeric LIRs are represented by rectangles below each partial dimer name. Regions in black and white are MSV-Kom and MSV-Set derived sequences, respectively. The stem-loop structure is represented by a triangle. All Rep mutants were co-bombarded with either pKom602 (Kom) or pSet105 (Set) and the detection of complementation of replication functions (+), or lack thereof (-), is indicated above the respective lanes. The mutant pSet*Nsi*I can not be distinguished from the wild type MSV-Set (*). Each bombardment was performed in duplicate. A 100 ng of the extracted DNA was subjected to PCR amplification, the product digested with *Cfo*I and electrophoresed through a 1.5 % agarose gel stained with ethidium bromide. The various banding patterns either correspond to the characteristic WT virus patterns (see Table 5.2 for *Cfo*I banding patterns) or are a mixed population of WT and chimaeric virus. Lambda *Pst*I treated DNA (λ) was used as a molecular weight marker.

Sequences up to and including the Rep proximal iteron are also required for replication: mutant LIRs truncated at the *Acc*I site with the full iteron sequence present, were weakly replication proficient, and then only in the presence of high levels of Rep. Auxiliary regions similar to those found in WDV may be required for efficient replication , which may be present between the *Acc*I site and the C1 start codon.

Evidence obtained by W. H. Schnippenkoetter using pKesek2 (Schnippenkoetter, 1998) and my results in Chapter 4, indicated that the MSV-Set Rep was able to recognise and *trans*-replicate a mutant virus with a MSV-Kom origin of replication. This mutant virus was highly attenuated, as can be seen in Fig. 5.6. The extent that the virion sense promoter, the stem-loop sequence or the Rep proximal iteron region contributed towards its impaired ability to replicate and infect were unknown (W. H. Schnippenkoetter, Pers. Comm.). I have provided a more precise analysis of the regions within the MSV LIR that contribute towards efficient replication. First, I have found that the 74 bp sequence from the *rpeI* element to the *Sac*I site (region B) determines replication specificity, with both pSSRK*Nsi*I and pKNRS1.1 successfully competing for the MSV-Kom and MSV-Set Rep, respectively (see Fig. 5.9 & 5.10). The mildly agroinfectious construct pSSRK3.1, which contains both regions B and C from MSV-Kom, compared with the essentially non-agroinfectious construct pSRAK773.1 (only region C), indicates a co-operative relationship between the stem-loop and the upstream region. This could be due to the distance between the stem-loop and the Rep proximal iteron of MSV-Kom, which is more ideally suited to the maize host, being required for an interaction with host factors (Orozco *et al.*, 1998). Both viruses have the same *rpeI* element (Schnippenkoetter, 1998) (see Fig. 4.8), but differ in the number of nucleotides between their stem-loop structures and the Rep proximal iteron elements. MSV-Set is a grass-type MSV and does not appear to be suited to infecting a variety of maize hybrids (Schnippenkoetter, 1998, & D. P. Martin and E. P. Rybicki, unpublished data). A reflection of this may be that specific host factors other than those that bind at the *rpeI* element are required for binding to elements within the LIR (Fenoll *et al.*, 1990). These elements may be similar to those described for begomoviruses, and may not be conserved in maize to stabilise a non-optimal Rep-LIR interaction (Orozco *et al.*, 1998).

The sequence of the MSV stem structure has been shown, by the use of a variety of mutant viruses (pSek409, pSRAK*Nsi*I, pSSAK3.1, and pKesek2), to be non-essential for replication. However, only a wild type stem allows optimal replication (pKRASb3.1 compared to pKom603). Furthermore, the stem-loop sequence is not, or at least not the main replication specificity determinant, similar to the situation with begomoviruses (Fontes *et al.*, 1994b). The size of the putative stem structure was decreased to 9 bp due to the insertion of non-complementary DNA, yet still is replication proficient and maintains the insertion during infection (e.g. pSek415 and pSek418) (see Fig. 5.3 and Fig. 5.8).

Exchanging the N-terminus (60 aa) region of MSV-Set's Rep with the corresponding region of MSV-Kom did not produce a viable infectious virus: this was also confirmed in a transient assay system (data not shown). Furthermore, even when both the LIR *cis*-elements and Rep N-terminal 60 aa from MSV-Kom were present in the chimaeric MSV-Set based construct (pSNAK3.1), this still resulted in a non-replicating virus (see Table 5.6). Therefore, it can be assumed that not all the *trans*-acting RSDs of the Mastrevirus Rep reside within these N-terminal 60 aa, perhaps similar to the begomoviruses where a secondary *trans*-acting replication specificity determinant is evident in AL1 (Gladfelter *et al.*, 1997). The N-terminal 116 aa of the TYLCV Rep and the first 89 aa of BCTV hold the *trans*-acting replication specificity determinants (Jupin *et al.*, 1995; Choi and Stenger, 1995), thus the MSV chimaeric Rep proteins may not have included all the necessary sequences, only having the first 60 aa. Alternatively, the chimaeric proteins could have been impaired in one of their other essential functions, such as transcriptional control of the virion sense genes (Collin *et al.*, 1996; Hofer *et al.*, 1992); however, because pSNAK1.1 could not replicate in a transient assay any effect due to lack of these functions should be minimal. All the Rep motifs essential for rolling circle replication are conserved between MSV-Kom and MSV-Set (see Table 3.6).

I have established that the *cis*-elements required for MSV replication show greater architectural homology to the other geminivirus genera than was previously expected (Argüello-Astorga *et al.*, 1994a & b). These similarities include the flexibility of the stem sequence and the requirement of elements on the complementary sense side, both for replication and replication specificity. Similar to WDV, but unlike the begomoviruses, sequences 3' of the stem-loop are required for replication (Orozco *et al.*, 1998). In contrast to the current hypothesis of mastrevirus RSDs (Argüello-Astorga *et al.*, 1994a & b), the stem-loop does not define replication specificity as was clearly observed in these studies. However, the Rep proximal iteron region (region B) does, as was observed in the transient replication assays. With the characterisation of MSV-Raw now underway (D. P. Martin, Pers. Comm) exchanges between this virus and MSV-Kom will further refine the nature of the replication specificity determinants.

Conclusion

The highly specific interaction of the begomo- and curtovirus Rep with its cognate origin of replication has allowed the elucidation of the precise nucleotide sequences as well as the protein domains responsible for this function (Orozco *et al.*, 1998; Gladfelter *et al.*, 1997; Settlage *et al.*, 1996; Choi and Stenger, 1996; Choi and Stenger, 1995; Fontes *et al.*, 1994; Fontes *et al.*, 1992). These studies were all primarily based on the use of different virus strains to determine which elements defined replication specificity. A similar approach was used to examine the MSV Rep protein's replication specificity. Firstly, the diversity of MSV isolates was assessed, using PCR, RFLP and sequence data. All but two MSVs (MSV-Set and MSV-Raw) shared the same putative replication specificity determinants as defined by Argüello-Astorga *et al.* (1994a & b): they all had identical stem sequences and Rep proximal iterons, however the wheat- and grass-infecting isolates shared far lower RepA aa homology. MSV-Tas and MSV-VW shared only approximately 87 % RepA sequence identity with the maize-type MSV (Table 3.7). Considering that the wheat-infecting virus Rep protein sequence is the most divergent of the three proteins analysed (Table 3.7 & 3.8), these differences may reflect specific interaction with host factors other than the plant retinoblastoma homologues, as the Rep:Rb binding motif is conserved (see Table 3.6) (Horváth *et al.*, 1998; Ach *et al.*, 1997; Xie *et al.*, 1996; Collin *et al.*, 1996; Nagar *et al.*, 1995; Hofer *et al.*, 1992).

A considerable diversity of isolates was detected within the MSV population, ranging from the highly homologous maize-infecting MSVs - all sharing approximately 96 % sequence identity - to the wheat-infecting MSVs - only sharing 89 % - and the grass infecting MSVs (78 %) (Table 2.4). From within this range of viruses, certain isolates were selected for *trans*-replication studies based on their relative Rep sequence diversity. Due to the low RepA identity of the wheat infecting isolates MSV-Tas and MSV-VW with the maize-types, full genome sequence was obtained prior to their inclusion in the replication study, and a study was made of their host ranges by leafhopper transmissions. Both the symptoms observed from agroinfections, and leafhopper transmission studies indicated that these two viruses were weak in maize hosts and infected a larger number of wheat cultivars than the previous characterised maize-type virus MSV-Kom (see Table 3.3) (Schnippenkoetter, 1998). Based on full sequence comparisons the maize-infecting

isolates were all highly related, sharing 96 % or greater whole genome sequence identity. MSV-Reu, the previously most distinct isolate, defined the outer limit of this subgroup of viruses (Peterschmitt *et al.*, 1996) (Table 3.8). The smaller subgroup, of wheat- and grass-infecting isolates, shared about 89 % sequence identity with the typical maize type, MSV-Kom, and about 98 % with each other. These diverse viruses - some sharing the same *cis*-elements yet putatively having different *trans*-acting elements, while others differed in both *cis*- and *trans*-acting elements - were used to assess the molecular determinants of replication specificity.

In light of the results presented here and those recently obtained by others (Sanz-Burgos and Gutiérrez, 1998; Schnippenkoetter, 1998), the replication mechanism of MSV is in some respects more similar to those of the other genera of geminiviruses than was previously anticipated. For instance: the *cis*-elements involved in replication specificity were found to reside within the complementary sense side of the LIR and not within the stem-loop (5.4.3 & 5.4.4), and although only a preliminary investigation was undertaken, evidence indicated that this same region as well as some sequence 3' of the stem-loop structure were essential for replication (5.4.2). In hindsight, an approach similar to that of Sanz-Burgos & Gutiérrez (1998) and Suárez-López & Gutiérrez (1997) using WDV replicons smaller than wild-type size may have given more conclusive results. However, my results support the former's conclusions on the *cis*-elements required for mastrevirus replication, and showed that viruses at the extreme of variation within a genus (50 % or greater nucleotide sequence difference), share the same basic replication mechanism.

The ability of a heterologous mastrevirus Rep to complement replication functions extends to more diverse set of viruses than was previously expected (4.4.3), even if these interactions were sub-optimal. However, the Rep protein has higher affinity for its cognate origin of replication (Fontes *et al.*, 1994). These interactions were greatly enhanced when sequences (74 bp) including and flanking the Rep proximal iteron region originated from the same virus as the Rep protein (5.4.4). However, as yet undefined elements conserved amongst these diverse viruses may allow *trans*-replication; for example, the identified sequence elements shared between MSV-Kom and PanSV-Kar (4.4.4). *Trans*-complementation of gene products has been proposed to be one of a few criteria for the designation of virus species along with > 90 % sequence identity, having different host ranges and vector species (Rybicki, 1998; van Regenmortel *et al.*, 1997). I

have shown that within the group of mastreviruses chosen for this study, the diversity of viruses able to *trans*-replicate extends from the MSVs to a PanSV (60 % total genome sequence identity (Table 3.8), but does not include DSV (63 %). Furthermore, it was recently shown that the dicot-infecting virus BeYDV (45 %) is unable to complement MSV's replication functions (Lui *et al.* 1999). While neither DSV nor BeYDV share the same vector species with MSV (see Table 1.1) and PanSV-Kar has a fairly unique host range (Schnippenkoetter, 1998), it seems that sharing of multiple criteria is the only reasonable way to designate isolates of mastreviruses to the same species, even if complementation of replication functions occurs.

The implications of a promiscuous replication strategy, considering the evidence of mixed populations of diverse viruses (2.4.2), could explain how mastreviruses maintain their diversity and improve host adaptation. For example there exists the possibility that MSV-VM arose from a recombination event between a wheat- and maize-type MSV: this would explain how, although the whole genome sequence identity is only 89 %, there is greater sequence identity within the normally even more variable MP protein, particularly in the N-terminus region (see Fig. 3.7 & 3.13). There is evidence of intermolecular recombination having occurred between geminiviruses during natural selection, as well as during experimental procedures. These types of recombination events may be the mechanism by which geminiviruses continue to evolve (Zhou *et al.*, 1998; Zhou *et al.*, 1997; Hou and Gilbertson, 1996; Klute *et al.*, 1996; Roberts and Stanley, 1994).

The use of mastrevirus isolates as described here has provided the tools for the elucidation of elements involved in replication, without the necessity for making forced mutations. The knowledge gained from this work will be of significant use to those undertaking transgenic resistance strategies based on the inhibition of mastrevirus replication, as well as to those using the more classical methods of breeding resistant maize. These methods require information regarding the specificity of the Rep protein for its cognate origin of replication and the diversity of MSV, respectively, as a means of overcoming this pathogen.

Reference List

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

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

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

Ach, R.A., T. Durfee, A.B. Miller, P. Taranto, L. Hanley-Bowdoin, P.C. Zambryski, and W. Gruissem. 1997. RRB1 and RRB2 encode maize retinoblastoma-related proteins that interact with a plant D-type cyclin and geminivirus replication protein Mol.Cell Biol. **17**:5077-5086.

Anant, S. and K.N. Subramanian. 1992. Isolation of low molecular weight DNA from bacteria and animal cells. Methods in Enzymology **216**:20-29.

Andersen, M.T., K.A. Richardson, S.-A. Harbison, B.A.M. Morris, S.A. Harbison, and B.A. Morris. 1988. Nucleotide sequence of the geminvirus chloris striate mosaic virus. Virology **164**:443-449.

Andersen, M.T., K.A. Richardson, S.A. Harbison, and B.A. Morris. 1988. Nucleotide sequence of the geminivirus chloris striate mosaic virus. Virology **164**:443-449.

Argüello-Astorga, G.R., R.G. Guevara-González, L.R. Herrera-Estrella, and R.F. Rivera-Bustamante. 1994. Geminivirus replication origins have a group-specific organization of iterative elements: a model for replication. Virology **203**:90-100.

Argüello-Astorga, G.R., L. Herrera-Estrella, and R. Rivera-Bustamante. 1994. Experimental and theoretical definition of geminivirus origin of replication. Plant Mol.Biol. **26**:553-556.

Armstrong, C. L. and Songstad, D. D. Method for transforming monocotyledenous plants. 93870173.7. 1993. European Patent.

Autrey, L.J.C. and C. Ricaud. 1983. The comparative epidemiology of two diseases of maize caused by leafhopper-borne viruses in Mauritius, p. 277-285. In R.T. Plumb and

J.M. Thresh (eds.), The spread and control of insect-borne viruses. Blackwell Scientific Publications, Oxford London Edinburgh Boston Melbourne.

Azzam, O., J. Frazer, D. de la Rosa, J.S. Beaver, P. Ahlquist, and D.P. Maxwell. 1994. Whitefly transmission and efficient ssDNA accumulation of bean golden mosaic geminivirus require functional coat protein. Virology **204**:289-296.

Barrow, M. R. Development of maize hybrids resistant to maize streak virus. Gevers, H. O. 232, 120-124. 1990. Pretoria, South Africa, Department of Africultural Development. Proceedings of the Ninth South African Maize Breeding Symposium.

Behjatnia, S.A.A., I.B. Dry, and M.A. Rezaian. 1998. Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl virus. Nucl.Acids Res. **26**:925-931.

Bendahmane, M., H.-J. Schalk, and B. Gronenborn. 1995. Identification and characterization of wheat dwarf virus from France using a rapid method for geminivirus DNA preparation. J.Phytopathology **85**:1449-1455.

Bevan, M. 1984. Binary *Agrobacterium* vectors for plant transformation. Nucl.Acids Res. **12**:8711-8721.

Bisaro, D.M. 1996. Geminivirus Replication, p. 833-854. In M.L. De Pamphilis (ed.), DNA replication in eukaryotic cells. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bisaro, D.M., W.D.O. Hamilton, R.H.A. Coutts, and K.W. Buck. 1982. Molecular cloning and characterisation of the two DNA components of tomato golden mosaic virus. Nucl.Acids Res. **10**:4913-4922.

Bock, K.R. and R.A. Bailey. 1989. Streak, p. 323-329. In C. Ricaud, B.T. Egan, A.G. Gillaspie Jr, and C.G. Hughes (eds.), Diseases of Sugarcane: Major Diseases. Elsevier, Amsterdam, Oxford, New York, Tokyo.

Bock, K.R., E.J. Guthrie, and R.D. Woods. 1974. Purification of maize streak virus and its relationship to virus associated with streak diseases of sugar cane and *Panicum maximum*. Annals Appl.Biol. 77:289-296.

Boulton, M.I., W.G. Buchholz, M.S. Marks, P.G. Markham, and J.W. Davies. 1989. Specificity of *Agrobacterium*-mediated delivery of maize streak virus DNA to members of the Gramineae. Plant Mol.Biol. **12**:31-40.

Boulton, M.I. and J.W. Davies. 1990. Monopartite geminiviruses: markers for gene transfer to cereals. Aspects of Appl.Biol. **24**:79-86.

Boulton, M.I., D.I. King, J. Donson, and J.W. Davies. 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., D.I. King, P.G. Markham, M.S. Pinner, and J.W. Davies. 1991. Host range and symptoms are determined by specific domains of the maize streak virus genome. Virology **181**:312-318.

Boulton, M.I., C.K. Pallaghy, M. Chatani, S. MacFarlane, and J.W. Davies. 1993. Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. Virology **192**:85-93.

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

Bradeen, J.M., M.C. Timmermans, and J. Messing. 1997. Dynamic genome organization and gene evolution by positive selection in geminivirus (Geminiviridae). Mol.Bio.Evo. **14**:1114-1124.

Briddon, R.W., I.D. Bedford, J.H. Tsai, and P.G. Markham. 1996a. Analysis of the nucleotide sequence of the treehopper-transmitted geminivirus, tomato pseudo-curly top virus, suggests a recombinant origin. Virology **219**:387-394.

Briddon, R.W., P. Lunness, I.D. Bedford, L.C.L. Chamberlin, T. Mesfin, and P.G. Markham. 1996b. A streak disease of pearl millet caused by a leafhopper-transmitted geminivirus. Eur.J.Plant Pathol. **102**:379-400.

Briddon, R.W., P. Lunness, L.C. Chamberlin, and P.G. Markham. 1994. Analysis of the genetic variability of maize streak virus. Virus Genes **9**:93-100.

Briddon, R.W., P. Lunness, L.C. Chamberlin, M.S. Pinner, H. Brundish, and P.G. Markham. 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 P.G. Markham. 1995. Family *Geminiviridae*, p. 158-165. In F.A. Murphy, C.M. Fauquet, D.H.L. Bishop, S.A. Ghabrial, A.W. Jarvis, G.P. Martelli, M.A. Mayo, and M.D. Summers (eds.), Virus Taxonomy: Classification and Nomenclature of Viruses. Springer-Verlag, Wein New York.

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

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

Buragohain, A.K., Y.K. Sung, R.S. Coffin, and R.H. Coutts. 1994. The infectivity of dimeric potato yellow mosaic geminivirus clones in different hosts. J.Gen.Virol. **75**:2857-2861.

Burns, T.M., R.M. Harding, and J.L. Dale. 1995. The genome organization of banana bunchy top virus: analysis of six ssDNA components. J.Gen.Virol. **76**:1471-1482.

Callis, J., M. Fromm, and V. Walbot. 1987. Introns increase gene expression in cultured maize cells. Genes Dev. **1** :1183-1200.

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

Chen, D.F. and P.J. Dale. 1992. A comparison of methods for delivering DNA to wheat: the application of wheat dwarf virus DNA to seeds with exposed apical meristems. Transgenic Res. **1**:93-100.

Choi, I.R. and D.C. Stenger. 1995. Strain-specific determinants of beet curly top geminivirus DNA replication. Virology **206**:904-912.

Choi, I.R. and D.C. Stenger. 1996. The strain-specific *cis*-acting element of beet curly top geminivirus DNA replication maps to the directly repeated motif of the *ori*. Virology **226**:122-126.

Chung, C.T. and R.H. Miller. 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucl.Acids Res. **16**:3580

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

Collin, S., M. Fernandez-Lobato, P.S. Gooding, P.M. Mullineaux, and C. Fenoll. 1996. The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma-binding proteins. Virology 324-329.

Czosnek, H., A. Kheyr-Pour, B. Gronenborn, E. Remetz, M. Zeidan, A. Altman, H.D. Rabinowitch, S. Vidavsky, N. Kedar, and Y. Gafni. 1993. Replication of tomato yellow leaf curl virus (TYLCV) DNA in agroinoculated leaf discs from selected tomato genotypes. Plant Mol.Biol. **22** :995-1005.

Dekker, E.L., C.J. Woolston, Y.B. Xue, B. Cox, and P.M. Mullineaux. 1991. Transcript mapping reveals different expression strategies for the bicistronic RNAs of the geminivirus wheat dwarf virus. Nucl.Acids Res. **19**:4075-4081.

del Solar, G., R. Giraldo, M.J. Ruiz-Echevarria, M. Espinosa, and R. Diaz-Orejas. 1998. Replication and control of circular bacterial plasmids. Microbiology and Molecular Biology Reviews **62**:434-464.

del Solar, G., M. Moscoso, and M. Espinosa. 1993. Rolling circle-replicating plasmids from Gram-positive and Gram-negative bacteria: a wall falls. Mol.Micro. **8**:789-796.

Desbiez, C., C. David, A. Mettouchi, J. Laufs, and B. Gronenborn. 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.

Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucl.Acids Res. **12**:387-395.

Dickinson, V.J., J. Halder, and C.J. Woolston. 1996. The product of maize streak virus ORF V1 is associated with secondary plasmodesmata and is first detected with the onset of viral lesions. Virology **220**:51-59.

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

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

Donson, J., B.A.M. Morris-Krsinich, P.M. Mullineaux, M.I. Boulton, and J.W. Davies. 1984. A putative primer for second-strand DNA synthesis of maize streak virus is virion-associated. EMBO J. **3** :3069-3073.

Dowling, T.E., C. Moritz, and J.D. Palmer. 1990. Nucleic Acids II: Restriction Site Analysis, p. 250-317. In D.M. Hillis and C. Moritz (eds.), Molecular Systematics. Sinauer Associates, Inc., Sunderland, Massachusetts, U.S.A.

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

Dry, I.B., J.E. Rigden, L.R. Krake, P.M. Mullineaux, and M.A. Rezaian. 1993. Nucleotide sequence and genome organization of tomato leaf curl geminivirus. J.Gen.Virol. **74**:147-151.

Dunder, E., J. Dawson, J. Suttie, and G. Pace. 1995. Maize transformation by microprojectile bombardment of immature embryos, p. 126-138. In I. Potrykus and G. Spangenberg (eds.), Gene Transfer to Plants. Springer-Verlag, Berlin and Heidelberg.

Eagle, P.A. and L. Hanley-Bowdoin. 1997. cis elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. J.Virol. **71**:6947-6955.

Eagle, P.A., B.M. Orozco, and L. Hanley-Bowdoin. 1994. A DNA sequence required for geminivirus replication also mediates transcriptional regulation. Plant Cell **6**:1157-1170.

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

Espinosa, M., G. del Solar, F. Rojo, and J.C. Alonso. 1995. Plasmid rolling circle replication and its control. FEMS Micro.Lett. **130**:111-120.

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

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

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

Evans, D. and H. Jeske. 1993b. DNA B facilitates, but is not essential for, the spread of abutilon mosaic virus in agroinoculated *Nicotiana benthamiana*. Virology **194**:752-757.

Fanning, E. and R. Knippers. 1992. Structure and function of simian virus 40 large T antigen. Annu.Rev.Biochem. **61**:55-85.

Faria, J.C., R.L. Gilbertson, S.F. Hanson, F.J. Morales, P. Ahlquist, A.O. Loniello, and D.P. Maxwell. 1994. Bean golden mosaic virus type II isolates from the Dominican Republic and Guatemala: Nucleotide sequences, infectious pseudorecombinants, and phylogenetic relationships. J.Phytopathology **84**:321-329.

Feng, D.F. and R.F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J.Mol.Evo. **25**:360

Fenoll, C., D.M. Black, and S.H. Howell. 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., J.J. Schwartz, D.M. Black, M. Schneider, and S.H. Howell. 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.

Flores, N., F. Valle, F. Bolivar, and E. Morino. 1992. Recovery of DNA from agarose gels stained with methylene blue. BioTechniques **13**:204-205.

Fontes, E.P., P.A. Eagle, P.S. Sipe, V.A. Luckow, and L. Hanley-Bowdoin. 1994a. Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. J.Biol.Chem. **269**:8459-8465.

Fontes, E.P.B., H.J. Gladfelter, R.L. Schaffer, I.T.D. Petty, and L. Hanley-Bowdoin. 1994b. Geminivirus replication origins have a modular organisation. Plant Cell **6**:405-416.

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

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

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

Frischmuth, T., G. Zimmat, and H. Jeske. 1990. The nucleotide sequence of abutilon mosaic virus reveals prokaryotic as well as eukaryotic features. Virology **178**:461-468.

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

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

Gilbertson, R.L., M.R. Rojas, D.R. Russell, and D.P. Maxwell. 1991b. Use of the asymmetric polymerase chain reaction and DNA sequencing to determine genetic variability of bean golden mosaic geminivirus in the Dominican Republic. J.Gen.Virol. **72**:2843-2848.

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

Gladfelter, H.J., P.A. Eagle, E.P.B. Fontes, L. Batts, L. Hanley-Bowdoin, and E.P. Fontes. 1997. Two domains of the AL1 protein mediate geminivirus origin recognition. Virology **239**:186-197.

Gorbalenya, A.E. and E.V. Koonin. 1989. Viral proteins containing the purine NTP-binding sequence pattern. Nucl.Acids Res. **17**:8413-8440.

Grafi, G., R.J. Burnett, T. Helentjaris, B.A. Larkin, J.A. DeCaprio, W.R. Sellers, and W.G.J. Kaelin. 1996. A maize cDNA encoding a member of the retinoblastoma protein family: involvement in endoreduplication. Proc.Natl.Acad.Sci.USA **93**:8962-8967.

Granados, R.R. 1969. Maize viruses and vectors, p. 327-359. In K. Maramorosch (ed.), Viruses, vectors, and vegetation. Interscience Publishers, A division of John Wiley & Sons, New York London Sydney Toronto.

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

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

Grimsley, N.H., C. Ramos, T. Hein, and B. Hohn. 1988. Meristematic tissues of maize plants are most susceptible to agroinfection with maize streak virus. Bio/Tech. **6**:185-188.

Groning, B.R., R.J. Hayes, and K.W. Buck. 1994. Simultaneous regulation of tomato golden mosaic virus coat protein and AL1 gene expression: expression of the AL4 gene may contribute to suppression of the AL1 gene. J.Gen.Virol. **75**:721-726.

Gutiérrez, C., P. Suárez-López, E. Ramirez-Parra, A.P. Sanz-Burgos, Ponninger.J, and Q. Xie. 1995. DNA bending as a potential regulatiory cis-acting element of the geminivirus intergenic region. agronomie **15**:415-420.

Hafner, G.J., M.R. Stafford, L.C. Wolter, R.M. Harding, and J.L. Dale. 1997. Nicking and joining activity of banana bunchy top virus replication protein *in vitro*. J.Gen.Virol. **78**:1795-1799.

Haley, A., X. Zhan, K. Richardson, K. Head, and B. Morris. 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.O., D.M. Bisaro, and K.W. Buck. 1982. Identification of novel DNA forms in tomato golden mosaic virus infected tissue. Evidence for a two component viral genome. Nucl.Acids Res. **10**:4901-4912.

Hanley-Bowdoin, L., P.A. Eagle, B.M. Orozco, D. Robertson, and S.B. Settlage. 1996. Geminivirus Replication, p. 278-292. In G. Stacey, B. Mullin, and P.M. Gresshoff (eds.), Biology of Plant-Microbe Interactions. ASPP, St. Paul, MN.

Hanley-Bowdoin, L., J.S. Elmer, and S.G. Rogers. 1988. Transient expression of heterologous RNAs using tomato golden mosaic virus. Nucl.Acids Res. **16**:10511-10528.

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

Hanley-Bowdoin, L., S.B. Settlage, B.M. Orozco, S. Nagar, and D. Robertson. 1999. Geminiviruses: models for plant DNA replication, transcription, and cell cycle regulation. CRC Crit.Rev.Plant Sci. **18**:71-106.

Hanson, S.F., R.A. Hoogstraten, P. Ahlquist, R.L. Gilbertson, D.R. Russell, and D.P. Maxwell. 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., H. Barker, K.R. Bock, E.J. Guthrie, G. Meredith, and M. Atkinson. 1977. Plant viruses with circular single-stranded DNA. Nature **270**:760-762.

Hayes, R.J. and K.W. Buck. 1989. 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. Nucl.Acids Res. **17**:10213-10222.

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

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

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

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

Hirt, H. 1996. In and out of the plant cell cycle. Plant Mol.Biol. **31**:459-464.

Hofer, J.M., E.L. Dekker, H.V. Reynolds, C.J. Woolston, B.S. Cox, and P.M. Mullineaux. 1992. Coordinate regulation of replication and virion sense gene expression in wheat dwarf virus. Plant Cell **4**:213-223.

Hofer, P., I.D. Bedford, P.G. Markham, H. Jeske, and T. Frischmuth. 1997. Coat protein gene replacement results in whitefly transmission of an insect nontransmissible geminivirus isolate. Virology **236**:288-295.

Hofer, P., M. Engel, H. Jeske, and T. Frischmuth. 1997a. Host range limitation of a pseudorecombinant virus produced by two distinct bipartite geminiviruses. Mol.Plant Microbe Interact. **10**:1019-1022.

Hofer, P., M. Engel, H. Jeske, and T. Frischmuth. 1997b. Nucleotide sequence of a new bipartite geminivirus isolated from the common weed *Sida rhombifolia* in Costa Rica. J.Gen.Virol. **78**:1785-1790.

Holsters, H., D. de Waele, A. Depicker, E. Messens, M. van Montague, and Schell. 1978. Transfection and transformation of *Agrobacterium tumefaciens*. Mol.Gen.Genet. **163**:181-187.

Hong, Y. and J. Stanley. 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.

Hong, Y.G., D.J. Robinson, and B.D. Harrison. 1993. Nucleotide sequence evidence for the occurrence of three distinct whitefly-transmitted geminiviruses in cassava. J.Gen.Virol. **74**:2437-2443.

Hoogstraten, R.A., S.F. Hanson, and D.P. Maxwell. 1996. Mutational analysis of the putative nicking motif in the replication- associated protein (AC1) of bean golden mosaic geminivirus. Mol.Plant Microbe Interact. **9**:594-599.

Hormuzdi, S.G. and D.M. Bisaro. 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.

Hormuzdi, S.G. and D.M. Bisaro. 1995. Genetic analysis of beet curly top virus: examination of the roles of L2 and L3 genes in viral pathogenesis. Virology **206**:1044-1054.

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

Horváth, G.V., A. Pettkó-Szandtner, K. Nikovics, M. Bilgin, M. Boulton, J.W. Davies, C. Gutiérrez, and D. Dudits. 1998. Prediction of functional regions of the maize streak virus replication- associated proteins by protein-protein interaction analysis. Plant Mol.Biol. **38**:699-712.

Hou, Y.M. and R.L. Gilbertson. 1996. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. J.Virol. **70**:5430-5436.

Howell, S.H. 1984. Physical structure and genetic organisation of maize streak virus (Kenyan isolate). Nucl.Acids Res. **12**:7359-7375.

Hughes, F.L. 1991. Molecular investigations of subgroup I geminiviruses. PhD Thesis. University of Cape Town, South Africa.

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

Hughes, F.L., E.P. Rybicki, and M.B. von Wechmar. 1992. Genome typing of southern African subgroup I geminiviruses. J.Gen.Virol. **73**:1031-1040.

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

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

Ikegami, M., K. Yazaki, M. Chatani, and H. Mizuta. 1989. Molecular cloning and physical mapping of the DNA of Miscanthus streak geminivirus. Intervirology **30**:340-345.

Ilyina, T.V. and E.V. Koonin. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucl.Acids Res. **20**:3279-3285.

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

Jefferson, R.A., T.A. Kavanagh, and M.W. Bevan. 1987. GUS fusions: ß-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. EMBO J. **6**:3901-3907.

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

Jupin, I., F. de Kouchkovsky, F. Jouanneau, and B. Gronenborn. 1994. Movement of tomato yellow leaf curl geminivirus (TYLCV): involvement of the protein encoded by ORF C4. Virology **204**:82-90.

Jupin, I., F. Hericourt, B. Benz, and B. Gronenborn. 1995. DNA replication specificity of TYLCV geminivirus is mediated by the amino-terminal 116 amino acids of the Rep protein. FEBS Lett. **362**:116-120.

Kallender, H., I.T.D. Petty, V.E. Stein, M. Panico, I.P. Blench, A.T. Etienne, H.R. Morris, R.H.A. Coutts, and K.W. Buck. 1988. Identification of the coat protein gene of tomato golden mosaic virus. J.Gen.Virol. **69**:1351-1357.

Kammann, M., H.J. Schalk, V. Matzeit, S. Schaefer, J. Schell, and B. Gronenborn. 1991. DNA replication of wheat dwarf virus, a geminivirus, requires two *cis*-acting signals. Virology **184**:786-790.

Khan, S.A. 1997. Rolling-circle replication of bacterial plasmids. Microbiology and Molecular Biology Reviews **61**:442-455.

Kheyr-Pour, A., M. Bendahmane, V. Matzeit, G.P. Accotto, S. Crespi, and B. Gronenborn. 1991. Tomato yellow leaf curl virus from Sardinia is a whitefly- transmitted monopartite geminivirus. Nucl.Acids Res. **19**:6763-6769.

Klute, K.A., S.A. Nadler, and D.C. Stenger. 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.

Ko, C.H., V. Brendel, R.D. Taylor, and V. Walbot. 1998. U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. Plant Mol.Biol. **36**:573-583.

Koncz, C. and J. Schell. 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.

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

Koonin, E.V. and T.V. Ilyina. 1993. Computer-assisted dissection of rolling circle DNA replication. BioSystems **30** :241-268.

Kunik, T., K. Palanichelvam, H. Czosnek, V. Citovsky, and Y. Gafni. 1998. Nuclear import of the capsid protein of tomato yellow leaf curl virus (TYLCV) in plant and insect cells. Plant J. **13**:393-399.

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

Laufs, J., I. Jupin, S. Schumacher, F. Heyraud-Nitschke, and B. Gronenborn. 1995a. Geminivirus replication: Genetic and biochemical characterization of Rep protein function, a review. Biochimie **77**:765-773.

Laufs, J., S. Schumacher, N. Geisler, I. Jupin, and B. Gronenborn. 1995b. Identification of the nicking tyrosine of geminivirus Rep protein. FEBS Lett. **377**:258-262.

Laufs, J., W. Traut, F. Heyraud, V. Matzeit, S.G. Rogers, J. Schell, and B. Gronenborn. 1995c. *In vitro* cleavage and joining 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 characterization of geminiviruses. Plant Mol.Biol.Reporter **4**:177-192.

Lazarowitz, S.G. 1988. Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus. Nucl.Acids Res. **16**:229-249.

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**:70-80.

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

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

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

Liu, H., M.I. Boulton, and J.W. Davies. 1997. Maize streak virus coat protein binds single- and double- stranded DNA in vitro. J.Gen.Virol. **78**:1265-1270.

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

Liu, L., J.W. Davies, and J. Stanley. 1998. Mutational analysis of bean yellow dwarf virus, a geminvirus of the genus *Mastrevirus* that is adapted to dicotyledonous plants. J.Gen.Virol. **79**:2265-2274.

Liu, L., M.S. Pinner, J.W. Davies, and J. Stanley. 1999. Adaptation of the geminivirus bean yellow dwarf virus to dicotyledenous hosts involves both virion-sense and complementary-sense genes. J.Gen.Virol. **80**:501-506.

Liu, S., I.D. Bedford, R.W. Briddon, and P.G. Markham. 1997. Efficient whitefly transmission of African cassava mosaic geminivirus requires sequences from both genomic components. J.Gen.Virol. **78**:1791-1794.

Loor, G., S.-J. Zhang, P. Zhang, N.L. Toomey, and M.Y.W.T. Lee. 1997. Identification of DNA replication and cell cycle proteins that interact with PCNA. Nucl.Acids Res. **25**:5041-5046.

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

Lucy, A.P., M.I. Boulton, J.W. Davies, and A.J. Maule. 1996. Tissue specificity of *Zea mays* infection by maize streak virus. Mol.Plant Microbe Interact. **9**:22-31.

Luehrsen, K.R. and V. Walbot. 1991. Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. Mol.Gen.Genet. **225**:81-93.

Luehrsen, K.R. and V. Walbot. 1994. Addition of A- and U-rich sequence increases the splicing efficiency of a deleted form of a maize intron. Plant Mol.Biol. **24**:449-463.

MacDowell, S.W., H. MacDonald, W.D.O. Hamilton, R.H.A. Coutts, and K.W. Buck. 1985. The nucleotide sequence of cloned wheat dwarf virus DNA. EMBO J. **4**:2173-2180.

Martin, D.P. and E.P. Rybicki. 1998. Microcomputer based quantification of maize streak virus symptoms in *Zea mays* . Phytopathology

Martin, D.P., J.A. Willment, and E.P. Rybicki. 1999. Quantification of maize streak virus pathogenicity in differentially resistant *Zea mays* genotypes. Phytopathology (In Press)

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

Mayo, M.A. and C.R. Pringle. 1998. Virus taxonomy - 1997. J.Gen.Virol. **79**:649-657.

McClean, A.P.D. 1947. Some forms of streak virus occuring in maize, sugar-cane and wild grasses. p.1-39. South Africa Government Printers, Pretoria.

McElroy, D., A.D. Blowers, B. Jenes, and R. Wu. 1991. Construction of expression vectors based on the rice actin 1 (*Act1*) 5' region for use in monocot transformation. Mol.Gen.Genet. **231**:150-160.

Mesfin, T., N.A. Bosque-Perez, I.W. Buddenhagen, G. Thottappilly, and S.O. Olojede. 1992. Studies of maize streak virus isolates from grass and cereal hosts in Nigeria. Plant Dis. **76**:789-795.

Messing, J., D. Geraghty, G. Heidecker, N.T. Hu, J.C. Kridl, and I. Rubenstein. 1983. p. 211-227. In T. Kasuge, C.P. Meredith, and A. Hollaender (eds.), Genetic Engineering of Plants. Plenum Press, New York.

Michelson, I., D. Zamir, and H. Czosnek. 1994. Accumulation and translocation of tomato yellow leaf curl virus (TYLCV) in a *Lycopersicon esculentum* breeding line containing the *L. chilense* TYLCV tolerance gene *Ty-1*. J.Phytopathology **84**:928-933.

Morris-Krsinich, B.A.M., P.M. Mullineaux, J. Donson, M.I. Boulton, P.G. Markham, M.N. Short, and J.W. Davies. 1985. Bidirectional transcription of maize streak virus DNA and identification of the coat protien gene. Nucl.Acids Res. **13**:7256

Morris, B.A.M., K.A. Richardson, A. Haley, X. Zhan, and J.E. Thomas. 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.

Mullineaux, P.M., M.I. Boulton, P. Bowyer, R. van der Vlugt, M. Marks, J. Donson, and J.W. Davies. 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., J.W. Davies, and C.J. Woolston. 1992. Geminiviruses as gene vectors, p. 187-215. In T.M.A. Wilson and J.W. Davies (eds.), Genetic Engineering with Plant Viruses. CRC Press, Boca Raton, FL.

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

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

Mullineaux, P.M., J.E. Rigden, I.B. Dry, L.R. Krake, and M.A. Rezaian. 1993. Mapping of the polycistronic mRNAs of tomato leaf curl geminivirus. Virology **193**:414-423.

Murashige, T. and F. Shoog. 1962. A revised medium for rapid growth and bio-assays with tobacco tissue cultures. Physiol.Plant **15**:473-497.

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

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

Nei, M. and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc.Natl.Acad.Sci.USA **76**:5269-5273.

Noris, E., I. Jupin, G.P. Accotto, and B. Gronenborn. 1996. DNA-binding activity of the C2 protein of tomato yellow leaf curl geminivirus. Virology **217**:607-612.

Noris, E., A.M. Vaira, P. Caciagli, V. Masenga, B. Gronenborn, and G.P. Accotto. 1998. Amino acids in the capsid protein of tomato yellow leaf curl virus that are crucial for systemic infection, particle formation, and insect transmission. J.Virol. **72**:10050-10057.

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

Orozco, B.M., H.J. Gladfelter, S.B. Settlage, P.A. Eagle, R.N. Gentry, and L. Hanley-Bowdoin. 1998. Multiple cis elements contribute to geminivirus origin function. Virology **242**:346-356.

Orozco, B.M. and L. Hanley-Bowdoin. 1996. A DNA structure is required for geminivirus replication origin function. J.Virol. **70**:148-158.

Orozco, B.M., A.B. Miller, S.B. Settlage, and L. Hanley-Bowdoin. 1997. Functional domains of a geminivirus replication protein. J.Biol.Chem. **272**:9840-9846.

Ow, D.W., K.V. Wood, M. De Luca, J.R. de Wet, D.R. Helinski, and S. Howell. 1986. Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. Science **234**:856-859.

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

Padidam, M., R.N. Beachy, and C.M. Fauquet. 1995b. Tomato leaf curl virus from India has a bipartite genome and coat protein is not essential for infectivity. J.Gen.Virol. **76**:25-35.

Padidam, M., R.N. Beachy, and C.M. Fauquet. 1996. The role of AV2 ("precoat") and coat protein in viral replication and movement in tomato leaf curl geminivirus. Virology **224**:390-404.

Page, R.D.M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences **12**:357-358.

Page, W.W. 1997. Behaviour in relation to maize streak virus epidemiology and possible antixonosis in maize. Proceedings of the 1st International Maize Streak Disease Symposium, Hazyview, Mpumalanga, South Africa.

Palanichelvam, K., T. Kunik, V. Citovsky, and Y. Gafni. 1998. The capsid protein of tomato yellow leaf curl virus binds cooperatively to single-stranded DNA. J.Gen.Virol. **79**:2829-2833.

Palmer, K. E.1997. Investigations into the use of maize streak virus as a gene vector. PhD Thesis, University of Cape Town, South Africa.

Palmer, K.E. and E.P. Rybicki. 1997. The use of geminiviruses in biotechnology and plant molecular biology, with particular focus on Mastreviruses. Plant Sci. **129**:115-130.

Palmer, K.E. and E.P. Rybicki. 1998. The molecular biology of mastreviruses. Adv.Virus Res. **50:183-234**:183-234.

Pascal, E., A.A. Sanderfoot, B.M. Ward, R. Medville, R. Turgeon, and S.G. Lazarowitz. 1994. The geminivirus BR1 movement protein binds single-stranded DNA and localizes to the cell nucleus. Plant Cell **6**:995-1006.

Paszkowski, U., S.B. Zhang, I. Potrykus, and J. Paszkowski . 1993. Replication of the DNA A component of African cassava mosaic virus in a heterologous system. J.Gen.Virol. **74**:2725-2729.

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

Peterschmitt, M., B. Reynaud, G. Sommermeyer, and P. Baudin. 1991. Characterization of maize streak virus isolates using monoclonal and polyclonal antibodies and by transmission to a few hosts. Plant Dis. **75**:27-32.

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

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

Pinner, M.S., P.G. Markham, R.H. Markham, and E.L. Dekker. 1988. Characterization of maize streak virus: description of strains; symptoms. Plant Path. **37**:87

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

Pooma, W., W.K. Gillette, J.L. Jeffrey, and I.T.D. Petty. 1996. Host and viral factors determine the dispensability of coat protein for bipartite geminivirus systemic movement. Virology **218**:264-268.

Pooma, W. and I.T. Petty. 1996. Tomato golden mosaic virus open reading frame AL4 is genetically distinct from its C4 analogue in monopartite geminiviruses. J.Gen.Virol. **77**:1947-1951.

Qin, S., B.M. Ward, and S.G. Lazarowitz. 1998. The bipartite geminivirus coat protein aids BR1 function in viral movement by affecting the accumulation of viral single-stranded DNA. J.Virol. **72** :9247-9256.

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

Reynaud, B. and M. Peterschmitt. 1992. A study of the mode of transmission of maize streak virus by *Cicadulina mbila* using an enzyme-linked immunosorbent assay. Annals Appl.Biol. **121**:85-94.

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

Rigden, J.E., I.B. Dry, P.M. Mullineaux, and M.A. Rezaian. 1993. Mutagenesis of the virion-sense open reading frames of tomato leaf curl geminivirus. Virology **193**:1001-1005.

Rigden, J.E., L.R. Krake, M.A. Rezaian, and I.B. Dry. 1994. ORF C4 of tomato leaf curl geminivirus is a determinant of symptom severity. Virology **204**:847-850.

Roberts, S. and J. Stanley. 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., J.J. DePaulo, C.M. Fauquet, and R.N. Beachy. 1994. Complete nucleotide sequence of the geminivirus tomato yellow leaf curl virus, Thailand isolate. J.Gen.Virol. **75**:477-485.

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

Rojas, M.R., R.L. Gilbertson, D.R. Russell, and D.P. Maxwell. 1993. Use of degenerate primers in the polymerase chain reaction to detect whitefly-transmitted geminiviruses. Plant Dis. **77**:340-347.

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

Ruiz-Medrano, R., R.G. Guevara-González, G.R. Argüello-Astorga, Z. Monsalve-Fonnegra, L.R. Herrera-Estrella, and R.F. Rivera-Bustamante. 1999. Identification of a sequence element involved in AC2-mediated transactivation of the pepper huasteco virus coat protein gene. Virology **253**:162-169.

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

Rybicki, E.P. 1998. A proposal for naming geminiviruses: a reply by the Geminiviridae Study Group Chair. Arch.Virol. **143**:421-424.

Rybicki, E. P., Dennis, S., Napier, G., and Hughes, F. L. 1998. Novel grass and wheat strains of maize streak virus detected by DNA amplification and sequencing. Molecular Plant Pathology Online.

Rybicki, E.P. and F.L. Hughes. 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.

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

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory Press, New York.

Sanderfoot, A.A., D.J. Ingham, and S.G. Lazarowitz. 1996. A viral movement protein as a nuclear shuttle. The geminivirus BR1 movement protein contains domains essential for interaction with BL1 and nuclear loacalization. Plant Physiol. **110**:23-33.

Sanderfoot, A.A. and S.G. Lazarowitz. 1996. Getting it together in plant virus movement: cooperative interactions between biparitite geminivirus movement proteins. Trends Cell Biol. **6**:353-358.

Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain terminating inhibitors. Proc.Natl.Acad.Sci.USA **74**:5467

Sanz-Burgos, A.P. and C. Gutiérrez. 1998. Organization of the cis-acting element required for wheat dwarf geminivirus DNA replication and visualization of a rep protein-DNA complex. Virology **243**:119-129.

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

Saunders, K., A. Lucy, and J. Stanley. 1992. RNA-primed complementary-sense DNA synthesis of the geminivirus African cassava mosaic virus. Nucl.Acids Res. **20**:6311-6315.

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

Schneider, M., E. Jarchow, and B. Hohn. 1992. Mutational analysis of the "conserved region" of maize streak virus suggests its involvement in replication. Plant Mol.Biol. **19**:601-610.

Schnippenkoetter, W. H. 1998. The use of agroinfectious clones to investigate recombination between distinct maize streak virus strains. PhD Thesis, University of Cape Town, South Africa

Sequeira, J.C. and B.D. Harrison. 1982. Serological studies on cassava latent virus. Annals Appl.Biol. **101**:33-42.

Settlage, S.B., A.B. Miller, and L. Hanley-Bowdoin. 1996. Interactions between geminivirus replication proteins. J.Virol. **70**:6790-6795.

Shen, W.-H. and B. Hohn. 1994. Amplification and expression of the β-glucuronidase gene in maize plants by vectors based on maize streak virus. Plant J. **5**:227-236.

Shen, W.-H. and B. Hohn. 1995. Vectors based on maize streak virus can replicate to high copy numbers in maize plants. J.Gen.Virol. **76**:965-969.

Shen, W.H. and B. Hohn. 1991. Mutational analysis of the small intergenic region of maize streak virus. Virology **183**:721-730.

Smith, D.R. and D.P. Maxwell. 1994. Requirement of the common region of DNA-B and the BL1 open reading frame of bean golden mosaic geminivirus for infection of *Phaseolus vulgaris*. J.Phytopathology **84**:133-138.

Soto, P.E., I.W. Buddenhagen, and V.L. Asnani. 1982. Development of streak virus-resistant maize populations through improved challenge and selection methods. Annals Appl.Biol. **100**:539-546.

Stanley, J. 1995. Analysis of African cassava mosaic virus recombinants suggests strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling circle DNA replication. Virology **206**:707-712.

Stanley, J. and M.R. Gay. 1983. Nucleotide sequence of cassava latent virus DNA. Nature **301**:260-262.

Stanley, J. and J.R. Latham. 1992a. A symptom variant of beet curly top geminivirus produced by mutation of open reading frame C4. Virology **190:**506-509

Stanley, J., J.R. Latham, M.S. Pinner, I. Bedford, and P.G. Markham. 1992b. Mutational analysis of the monopartite geminivirus beet curly top virus. Virology **191**:396-405.

Stanley, J., P.G. Markham, R.J. Callis, and M.S. Pinner. 1986. The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. EMBO J. **5**:1761-1767.

Stanley, J. and R. Townsend. 1985. Characterisation of DNA forms associated with cassava latent virus infection. Nucl.Acids Res. **13**:2189-2206.

Stanley, J., R. Townsend, and S.J. Curson. 1985. Pseudorecombinants between cloned DNAs of two isolates of cassava latent virus. J.Gen.Virol. **66**:1055-1061.

Stenger, D.C. 1994. Strain-specific mobilization and amplification of a transgenic defective-interfering DNA of the geminivirus beet curly top virus. Virology **203**:397-402.

Stenger, D.C., D. Carbonaro, and J.E. Duffus. 1990. Genomic characterization of phenotypic variants of beet curly top virus. J.Gen.Virol. 71:2211-2215

Stenger, D.C., K.R. Davis, and D.M. Bisaro. 1992. Limited replication of tomato golden mosaic virus DNA in explants of nonhost species. Mol.Plant Microbe Interact. **5**:525-527.

Stenger, D.C. and K.M. Ostow. 1996. Genetic complexity of a beet curly top virus population used to assess sugar beet cultivar response to infection. J.Phytopathology **86**:929-933.

Stenger, D.C., G.N. Revington, M.C. Stevenson, and D.M. Bisaro. 1991. Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling-circle replication of a plant viral DNA. Proc.Natl.Acad.Sci.USA **88**:8029-8033.

Storey, H.H. 1928. Transmission studies of maize streak virus. Annals Appl.Biol. **15**:1-25.

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 A.P.D. McClean. 1930. The transmission of streak disease between maize, sugarcane and wild grasses. Annals Appl.Biol. 17:691-719.

Suárez-López, P. and C. Gutiérrez. 1997. DNA replication of wheat dwarf virus geminivirus vectors: effects of origin structure. Virology **227**:389-399.

Suárez-López, P., E. Martinez-Salas, P. Hernandez, and C. Gutiérrez. 1995. Bent DNA in the large intergenic region of wheat dwarf geminivirus. Virology **208**:303-311.

Sudarshana, M.R., H.L. Wang, W.J. Lucas, and R.L. Gilbertson. 1998. Dynamics of bean dwarf mosaic geminivirus cell-to-cell and long-distance movement in Phaseolus vulgaris revealed, using ghe green fluorescent protein. Mol.Plant Microbe Interact. **11**:277-291.

Sung, Y.K. and R.H. Coutts. 1996. Potato yellow mosaic geminivirus AC2 protein is a sequence non-specific DNA binding protein. FEBS Lett. **383**:51-54.

Sung, Y.K. and R.H.A. Coutts. 1995. Pseudorecombination and complementation between potato yellow mosaic geminivirus and tomato golden mosaic geminivirus. J.Gen.Virol. **76**:2809-2815.

Sunter, G. and D.M. Bisaro. 1989. Transcription map of the B genome component of tomato golden mosaic virus and comparisons with the A component transcripts. Virology **137**:647-655.

Sunter, G. and D.M. Bisaro. 1991. Transactivation in a geminivirus: AL2 gene product is needed for coat protein expression. Virology **180**:416-419.

Sunter, G. and D.M. Bisaro. 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. and D.M. Bisaro. 1997. Regulation of a geminivirus coat protein promoter by AL2 protein (TrAP): evidence for activation and derepression mechanisms. Virology **232**:269-280.

Sunter, G., K.W. Buck, and R.H.A. Coutts. 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. Nucl.Acids Res. **13**:4645-4659.

Sunter, G., W.E. Gardiner, Rushing A.E, S.G. Rogers, and D.M. Bisaro. 1987. Independent encapsidation of tomato golden mosaic virus A component DNA in transgenic plants. Plant Mol.Biol. **8**:477-484.

Sunter, G., M.D. Hartitz, and D.M. Bisaro. 1993. Tomato golden mosaic virus leftward gene expression: autoregulation of geminivirus replication protein. Virology **195**:275-280.

Sunter, G., M.D. Hartitz, S.G. Hormuzdi, C.L. Brough, and D.M. Bisaro. 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.

Sunter, G., D.C. Stenger, and D.M. Bisaro. 1994. Heterologous complementation by geminivirus AL2 and AL3 genes. Virology **203**:203-210.

Swofford, D.L. and G.J. Olsen. 1990. Phylogeny Reconstruction, p. 411-501. In D.M. Hillis and C. Moritz (eds.), Molecular Systematics. Sinauer Associates, Inc., Sunderland, Massachusetts, U.S.A.

Thommes, P., T.A.M. Osman, R.J. Hayes, and K.W. Buck. 1993. TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region. FEBS Lett. **319**:95-99.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucl.Acids Res. **22**:4673-4680.

Thottappilly, G., N.A. Bosque-Perez, and H.W. Rossel. 1993. Viruses and virus diseases of maize in tropical Africa. Plant Path. **42**:494-509.

Timmermans, M.C., O.P. Das, and J. Messing. 1992. *Trans* replication and high copy numbers of wheat dwarf virus vectors in maize cells. Nucl.Acids Res. **20**:4047-4054.

Timmermans, M.C.P., O.P. Das, and J. Messing. 1994. Geminiviruses and their uses as extrachromosomal replicons. Annu.Rev.Plant Physiol. **45**:79-112.

Tinland, B. 1996. The integration of T-DNA into plant genomes. Trends in Plant Science **1**:178-184.

Tinland, B., B. Hohn, and H. Puchta. 1994. *Agrobacterium tumefaciens* transfers single-stranded DNA into the plant cell nucleus. Proc.Natl.Acad.Sci.USA **91**:8000-8004.

Torres-Pacheco, I., J.A. Garzon-Tiznado, L. Herrera-Estrella, and R.F. Rivera-Bustamante. 1993. Complete nucleotide sequence of pepper huasteco virus: analysis and comparison with bipartite geminiviruses. J.Gen.Virol. **74**:2225-2231.

Townsend, R., J. Stanley, S.J. Curson, and M.N. Short. 1985. Major polyadenylated transcripts of cassava latent virus and location of the gene encoding coat protein. EMBO J. **4**:33-37.

Vain, P., M.D. McMullen, and J.J. Finer. 1993. Osmotic treatment enhances particle bombardment-mediated transient and stable transformation of maize. Plant Cell Reports **12**:84-88.

van der Merwe, C. P. The maize jassid (*Balclutha mbila* Naude). Department of Agriculture. 12, 75-77. 1926. Pretoria, Union of South Africa Government Printer.

van Regenmortel, M.H.V., D.H.L. Bishop, C.M. Fauquet, M.A. Mayo, J. Maniloff, and C.H. Calisher. 1997. Guidelines to the demarcation of virus species. Arch.Virol. **142**:1505-1518.

van Rensburg, G.D.J. and J.H. Giliomee. 1990. A comparison of females and males of *Cicadulina anestae* and *C. mbila* (Homoptera:Cicadellidae) as vectors of maize streak virus. Phytophylactica **22**:241-243.

van Rensburg, G. D. J., Giliomee, J. H., and Pringle, K. L. 1990 Resistance of South African commercial maize hybrids to maize streak virus. Proceedings of the Ninth South African Maize Breeding Symposium. Ed. Gevers, H. O. Pretoria, South Africa, Department of Agricultural Development. 232, 124-129.

von Wechmar, M. B. and Hughes, F. L.1990. A rapid technique for typing streak virus isolates using a panel of differential hosts. Proceedings of the Ninth South African Maize Breeding Symposium. Ed. Gevers, H. O. Pretoria, South Africa, Department of Agricultural Development. 232, 115-119

Ward, B.M., R. Medville, S.G. Lazarowitz, and R. Turgeon. 1997. The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. J.Virol. **71**:3726-3733.

Wartig, L., A. Kheyr-Pour, E. Noris, F. de Kouchkovsky, F. Jouanneau, B. Gronenborn, and I. Jupin. 1997. Genetic analysis of the monopartite tomato yellow leaf curl geminivirus: Roles of V1, V2, and C2 ORFs in viral pathogenesis. Virology **228**:132-140.

Waters, V.L. and D.G. Guiney. 1993. Processes at the nick region link conjugation, T-DNA transfer and rolling circle replication. Mol.Micro. **9**:1123-1130.

Woolston, C.J., H.V. Reynolds, N.J. Stacey, and P.M. Mullineaux. 1989. Replication of wheat dwarf virus DNA in protoplasts and analysis of coat protein mutants in protoplasts and plants. Nucl.Acids Res. **17**:6029-6041.

Wright, E.A., T. Heckel, J. Groenendijk, J.W. Davies, and M.I. Boulton. 1997. Splicing features in maize streak virus virion- and complementary-sense gene expression. Plant J. **12**:1285-1297.

Wu, R.-Y., L.-R. You, and T.-S. Soong. 1994. Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus. J.Phytopathology **84**:951-957.

Wyatt, S.D. and J.K. Brown. 1996. Detection of subgroup III isolates in leaf extracts by degenerate primers and polymerase chain reaction. J.Phytopathology **86**:1288-1293.

Xie, Q., A.P. Sanz-Burgos, G.J. Hannon, and C. Gutiérrez. 1996. Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. EMBO J. **15**:4900-4908.

Xie, Q., P. Suárez-López, and C. Gutiérrez. 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.

Zhan, X., K.A. Richardson, A. Haley, and B.A. Morris. 1993. The activity of the coat protein promoter of chloris striate mosaic virus is enhanced by its own and C1-C2 gene products. Virology **193**:498-502.

Zhan, X.C., A. Haley, K. Richardson, and B. Morris. 1991. Analysis of the potential promoter sequences of African cassava mosaic virus by transient expression of the beta-glucuronidase gene. J.Gen.Virol. **72**:2849-2852.

Zhou, X., Y. Liu, L. Calvert, C. Munoz, G.W. Otim-Nape, D.J. Robinson, and B.D. Harrison. 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.

Zhou, X., Y. Liu, D.J. Robinson, and B.D. Harrison. 1998. Four DNA-A variants among Pakistani isolates of cotton leaf curl virus and their affinities to DNA-A of geminivirus isolates from okra. J.Gen.Virol. **79**:915-923.