12

Maize Streak Virus: Diversity and Virulence

Darren Patrick Martin

Thesis presented for the degree of Doctor of Philosophy in the Department of Microbiology,
University of Cape Town
December, 2000

Certification by Supervisor

In terms of paragraph GP9 of the regulations for the degree of Doctor of Philosophy at the University of Cape Town, I certify that I approve of the inclusion in this thesis of material already published, or submitted for publication by candidate Darren Patrick Martin

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

December, 2000

Contents

Acknowledgements

Ed Rybicki. For being the best possible PhD supervisor I could have hoped for. After providing me with a project he allowed me to do whatever I wanted with it. He was always excited about everything I did and always open to persuasion whenever he had reservations about the directions I wanted to take my work. Without his day to day encouragement and enthusiasm most of this thesis would not exist.

Jo Ann Passmore. For her constant encouragement, indulgence, financial support, and love. Her consistently constructive criticism of every part of this thesis has substantially improved its quality.

My mother and father, **Patricia** and **Tony Martin**. For being absolutely perfect parents.

Barbara Von Wechmar. For providing most of the MSV samples analysed during the coarse of my project.

Donald Solomons and Melvin Keating. For invaluable help in growing the many thousands of maize seedlings involved in this study.

Di James. For her tireless help sequencing the many full length MSV genomes presented in this thesis

Justin Keyser and Trevor Sewell. For providing the encouragement, information and means required for the development all software presented in this thesis.

Eric van der Walt, Janet Willment, Kenneth Palmer, Dionne Miles, Wendelin Schnippenkoetter and **Fiona Hughes** for the many enthusiastic and often extremely productive talks about various aspects of MSV biology. In particular, Eric, Janet and Kenneth are directly responsible for the ideas behind much of the work in this thesis. I am also extremely grateful to Eric, Janet, Fiona, and Dionne for their enormously constructive criticism of various parts of this thesis.

Patrick, Mike, Grant, Dan, Andrew, Kevin, Callem, Justin and **Glen**. For having absolutely no real interest in MSV but still spending the time to both grasp what it is I spent my days doing, and being impressed by it.

Benjamin Odhiambo and **Jackson Njaguna**. For my two trips to Kenya, collecting virtually every Kenyan MSV sample analysed in our labs, and doing much of the DNA isolation and initial PCR-RFLP typing of most of the Kenyan samples that are included in this thesis.

Rosie Billharz and **Renate Velters**. For their assistance in the PCR-RFLP typing of MSV samples.

J. B. J. Van Rensberg, Kevin Pixly and **Dave Nowell**. For generously providing me with seed from their most MSV resistant maize genotypes.

The **FRD**. For providing both the funding for this work and my subsistence money during my PhD.

Chapter 1

Literature Review

Contents

1.1 INTRODUCTION

Zea mays was first introduced to Africa in Ghana by Portuguese traders in the 16th century (Gorter 1953, Fajemisin and Shoyinka, 1976). The steady spread of maize cultivation since then has made it the most important cereal crop in Africa today. Whereas improved maize genotypes and agricultural techniques enable yearly yields above 10 tons hectare^{-1} in the developed world, yearly yields across Africa have remained low at ~1 ton hectare^{-1} in most countries (Efron *et al.*, 1989). Although outmoded agricultural practices are the main reason for poor yields, maize pathogens inflict substantial additional losses (Efron *et al.*, 1992). Of the many pathogens currently confronting maize farmers in Africa, *Maize streak virus* (MSV) is the most significant (Bosque-Perez and Allem, 1992).

It is probably safe to assume that indigenous grass-infecting viruses resembling modern MSV began infecting maize plants from the moment they were first planted in sub-Saharan Africa. Despite this, the first definitive reports of maize streak disease (MSD) were only documented 350 years later in Natal, South Africa (Fuller, 1901). Fuller (1901) mistakenly identified MSD as a soil nutrient deficiency and it was not until 1924 that its causal agent was first determined to be a virus, named MSV, that was obligately transmitted by leafhopper species in the genus *Cicadulina* (Storey, 1924).

MSV particles were first visualised in 1974 and were found to be 18 X 30 nm with an unusual twinned quasi-isomeric or "geminate" morphology (Bock *et al.*, 1974). The particles are presumed to consist of 2 incomplete icosohedra with T=1 surface lattice, comprising 22 pentameric capsomers each containing 5 coat protein molecules (Hatta and Francki, 1979). Another surprising discovery was that the nucleic acid within these unusual virus particles was circular single stranded DNA (ssDNA) – this was the first time that this had ever been observed for a plant virus (Harrison *et al.*, 1977). Based on these observations, MSV and other viruses sharing a geminate particle morphology and ssDNA genomes were assigned to a new taxonomic group called the geminiviruses (Mathews, 1979). This taxon has subsequently been upgraded to family status: The family *Geminiviridae* was established in 1995 (Briddon and Markham, 1995).

As with most other virus groups, the age of geminivirus molecular biology dawned in the 1980s with the cloning and full sequence analysis of the first viral genomes (Howell, 1984; Mullineaux

et al., 1984; Stanley and Gay, 1983;). In the 16 years since the first MSV genome was sequenced, substantial advances have been made in our understanding of MSV's population diversity, life cycle, evolution, and pathology. There are, however, still major gaps in our knowledge of this virus. Although the primary focus of this review is MSV, it will also draw heavily on information gathered on other geminivirus species to bridge some of these gaps.

1.2 GEMINIVIRUS TAXONOMY AND EVOLUTION

There are currently 4 recognised genera within the *Geminiviridae* (Table 1.1). Besides substantial phylogenetic support for the existence of these genera (Fig 1.1) they may also be differentiated from one another based on genome organisation (Fig 1.2), vector specificity and host range (Fig 1.1). While viruses in the genera *Mastrevirus* (type member: *Maize streak virus*), *Curtovirus* (type member: *Beet curly top virus*), and *Topocuvirus* (type member: *Tomato pseudo curly top virus*) all have monopartite (one component) genomes, those in the genus *Begomovirus* (type member: *Bean golden mosaic virus*) have either monopartite or bipartite (two component) genomes (Pringle, 1999; Rybicki, 2000).

In addition to the two begomovirus genomic components that are currently recognised, it has recently emerged that there are an array of additional DNAs associated with certain monopartite begomovirus infections (Fig 1.2; Dry *et al.*, 1997; Mansoor *et al.*, 1999; Saunders and Stanley, 1999; Saunders *et al.*, 2000). While most of these are apparently satellite DNAs (DNA-1 and DNA-sat; Fig 1.2) that do not appreciably contribute to the survival of their associated begomovirus (Dry *et al.*, 1997, Mansoor *et al.*, 1999; Saunders and Stanley, 1999), an instance has been described where an otherwise weakly virulent begomovirus isolate interacts with one of these DNAs (DNA-β; Fig 1.2) to form an extremely severe disease complex (Saunders *et al.*, 2000).

A general rule among the geminiviruses is that whereas genes encoded on the complementary virus strand ("C" genes in Fig 1.2) are involved in replication and transcriptional regulation, those encoded on the virion strand ("V" genes in Fig 1.2) are involved in movement and encapsidation. While there are differences between the geminivirus genera in the number and arrangement of the

TABLE 1.1 A selection of geminivirus species for which either full genomic sequences (for monopartite viruses) or full component A sequences (for bipartite viruses) have been determined

Genus	Name	Species (Major Strains/Variants)	Genome Size
Mastrevirus	*Maize streak virus*	MSV (Ns, Set, Tas)	2684 – 2701
	Sugarcane streak virus	SSV-Asw	2706
		SSV-N	2758
	Digitaria streak virus	DSV	2701
	Panicum streak virus	PanSV (Ken, Kar)	2700 – 2705
	Wheat dwarf virus	WDV	2749
	Chloris striate mosaic virus	CSMV	2750
	Miscanthus streak virus	MiSV	2672
	Bean yellow dwarf virus	BeYDV	2561
	Tobacco yellow dwarf virus	TYDV	2580
Curtovirus	*Horseradish curly top virus*	HrCTV	3080
	Beet curly top virus	BCTV (Cfh, Wor, Cal)	2927 – 2993
Topocuvirus	*Tomato pseudo-curly top virus*	TPCTV	2861
Begomovirus	*Cotton leaf curl virus*	CLCuV (802a, Okr, PK1)	2725 – 2747 2750
		CLCuV-PK2	
	Tomato leaf curl virus	ToLCV-Ban	2744 – 2749
		ToLCV-I	2739
		ToLCV-AU	2766
		ToLCV-TW	2739
		ToLCV-PA	2584
	Papaya leaf curl virus	PaLCV	2746
	Okra yellow vein mosaic virus	OYVMV (201, 301)	2739 – 2741
		ITmLCV	
	Tobacco leaf curl virus	TbLCV	2734
	Tomato yellow leaf curl virus	TYLCV-TH	2743
		TYLCV-Sar (Sic, Es)	2770 – 2777
		TYLCV-IL (mld)	2787 – 2790
	Ageratum yellow vein virus	AYVV	2741
	Indian cassava mosaic virus	ICMV	2815
		ChaMV	2787
	Althea rosea enation virus	AREV	2755
	East African cassava mosaic virus	EACMV (TZ, UG, CM)	2799 – 2808
	South African cassava mosaic virus	SACMV	
	African cassava mosaic virus	ACMV	2779
	Cowpea golden mosaic virus	CPGMV	2728
	Abutilon mosaic virus	AbMV	2629
	Sida golden mosaic virus	SiGMV-HO (YV)	2603 – 2612
		SiGMV-CR	2605
		SiGMV-Flo	2642
	Tomato mottle virus	ToMoV (Tai)	2597 – 2601
	Bean dwarf mosaic virus	BDMV	2615
	Pepper yellow mosaic virus	PYMV (VE, TT)	2582 – 2593
	Bean golden mosaic virus	BGMV-BR	2617
		BGMV-PR	2647
	Tomato golden mosaic virus	TGMV	2588
	Cabbage leaf curl virus	CaLCuV	2583
	Pepper golden mosaic virus	PepGMV	2613
	Squash leaf curl virus	SqLCV	2634
	Cotton leaf crumple virus	CLCrV	2630
	Pepper hausteco virus	PHV	2631

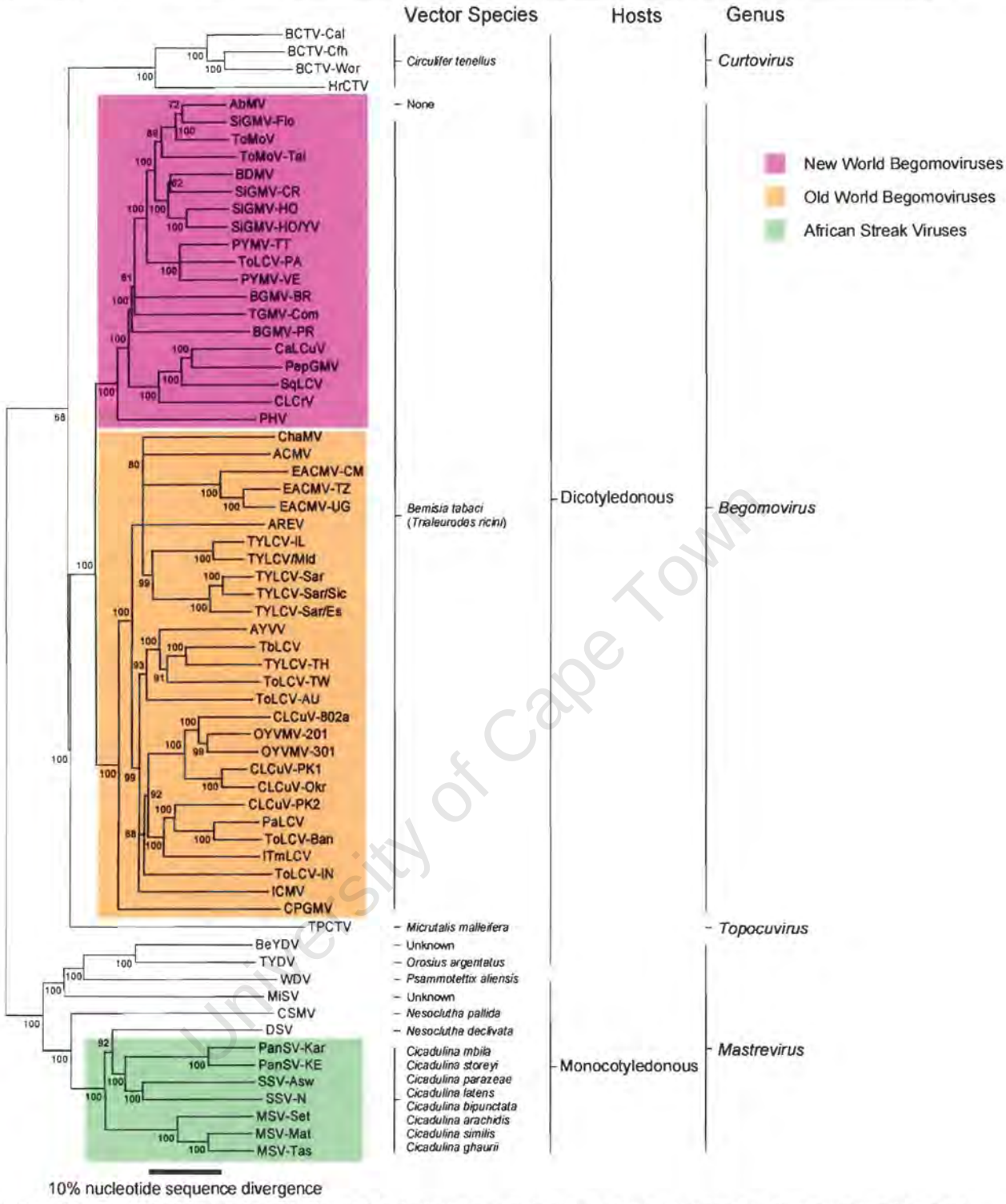


FIGURE 1.1 The phylogenetic relationships between a selection of begomovirus (A components only), mastrevirus, curtovirus and topocuvirus genome sequences. Mastreviruses and curtoviruses have leafhopper vectors, begomoviruses have whitefly vectors and topocuviruses have treehopper vectors. While the begomovirus AbMV is not vector transmissible, the mastreviruses BeYDV and MiSV have no known vector. Whereas all begomoviruses, curtoviruses and topocuviruses have dicot hosts, mastreviruses have either monocot or dicot hosts. Old and new world begomoviruses (in pink and orange respectively) form distinct species groupings. The depth of diversity amongst the African streak virus group (in green) is greater than that found amongst either the old or new world begomoviruses. Sequences were aligned in Clustal V (Higgins *et al.*, 1992) and used to construct a rooted neighbour joining tree (Saitou and Nei, 1987) with 1000 bootsrap iterations. Numbers associated with nodes represent percentages of bootstrap support for the nodes. Nodes with less than 60% bootstrap support were collapsed. Whereas vertical distances are arbitrary, horizontal distances reflect relationships between the sequences. GenBank accession numbers for the sequences and full virus names can be found in Table A.1 (see Appendix A).

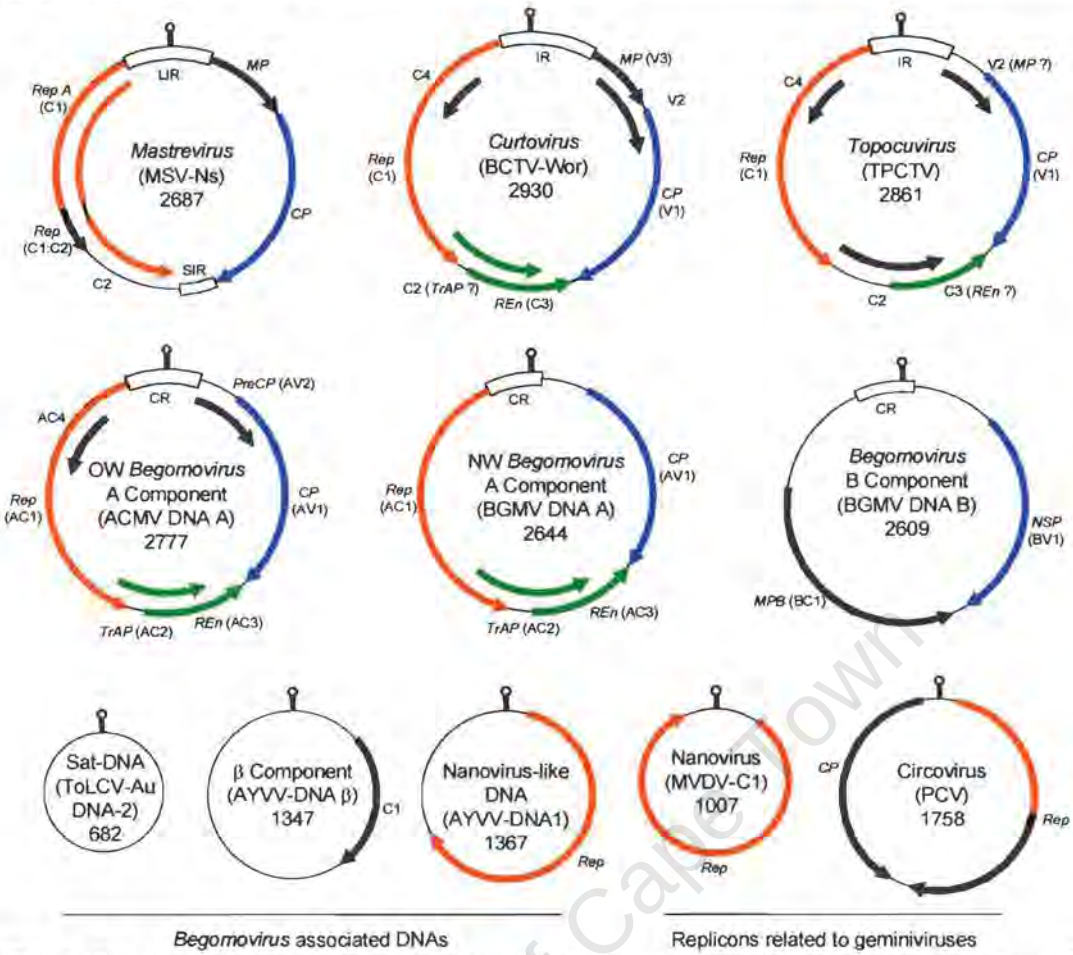


FIGURE 1.2 The genomic arrangements of geminiviruses and some other related replicons. Origins of (+) strand synthesis are indicated by the stem loop symbol at 12 o'clock. OW = Old World. NW = New World. IR = intergenic region. LIR = long/large intergenic region. SIR = short/small intergenic region. CR = common region that is nearly identical in the A and B components of bipartite begomoviruses. Genes and parts of genes in the same colour either express or potentially express proteins with detectable sequence homology. *Rep* = replication associated/initiator protein gene found in all geminiviruses, circoviruses and nanoviruses. In mastreviruses, it is expressed from the post-translationally spliced transcript of ORFs C1 and C2; *RepA* = variant of *Rep* potentially expressed from an unspliced complimentary sense transcript and encoding a multifunctional regulatory protein that is unique to the mastreviruses. *MP* = movement protein gene of mastreviruses and curtoviruses. The V2 ORF of topocuviruses may also encode a movement protein but this remains to be proven. *CP* = coat protein gene. *REn* = replication enhancer gene found in begomoviruses and curtoviruses (Bisaro, 1996). While an ORF encoding a protein with substantial sequence homology with the begomovirus and curtovirus REn exists in topocuviruses, its function in this genus remains to be proven. *TrAP* = transcription activator protein gene found in begomoviruses. While a homologue exists in curtoviruses, its function in this genus remains to be determined (Hormuzdi and Bisaro, 1995). While the C4 ORF of curtoviruses is apparently involved in tumour induction (Latham *et al.*, 1997), in the monopartite begomovirus, TYLCV, it is apparently involved in movement (Jupin *et al.*, 1994). While one study on the C4 ORF of the NW begomovirus, TGMV, indicated it is potentially involved in transcriptional regulation of *Rep* (Groning *et al.*, 1994), another study has found that the TGMV C4 is non functional and is probably not expressed (Pooma and Petty, 1996). While a C4 ORF exists in topocuviruses and certain mastreviruses, its function in both these virus groups is unknown (Palmer *et al.*, 1998; Pooma and Petty, 1996). *PreCP* = pre coat protein gene that in OW begomoviruses is apparently involved in ssDNA accumulation (Wartig *et al.*, 1997) and/or movement (Padidam *et al.*, 1996). In curtoviruses the V2 gene is apparently involved in regulation of ssDNA accumulation (Stanley *et al.*, 1992; Hormuzdi and Bisaro, 1993). *NSP* = nuclear shuttle protein gene is unique to bipartite begomoviruses but expresses a protein with some sequence homology to geminivirus CPs (Rybicki, 1994). *MPB* = movement protein gene found on the B component of bipartite begomoviruses. DNAs associated with certain monopartite begomovirus infections are mostly satellites (Dry *et al.*, 1997; Mansoor *et al.*1999; Saunders and Stanley, 1999; Saunders *et al.*, 2000), but the β component DNA detected by Saunders *et al.* (2000) massively enhances the virulence of its associated begomovirus.

genes that they possess, all encode a replication initiator/associated protein (Rep) and a coat protein (CP). Besides Rep and CP, mastreviruses express a movement protein (MP), and may also express a variant of Rep, called RepA, that is unique to this genus. Whereas the curtoviruses also possess a MP gene in a similar position to that found in mastreviruses, the MP gene of bipartite begomoviruses, called MPB, is expressed on the complementary strand of their B component. Other geminivirus encoded proteins with established functions include a transcription activator protein (TrAP) found in begomoviruses, a nuclear shuttle protein (NSP) found in bipartite begomoviruses, and a replication enhancer protein (REn) found in both begomoviruses and curtoviruses. In addition to these genes, there are a number of other curtovirus, topocuviruses, and begomovirus open reading frames (ORFs) with either unknown or only suspected functions (Fig 1.2).

While there is easily detectable homology between the Rep proteins of geminiviruses, nanoviruses and circoviruses (Fig 1.2 ; Boevink *et al.*, 1995; Meehan *et al.*, 1997), amino acid sequence motifs within these Rep proteins share highly conserved spatial and sequence relationships with motifs found in replication associated proteins from a number of ssDNA prokaryotic replicons (Koonin and Illyina, 1993). It is therefore believed that the portions of geminivirus Rep genes encoding these motifs probably have a prokaryotic origin (Fig 1.3; Ilyina and Koonin, 1992; Koonin and Ilyina, 1992, 1993). In support of this view, certain geminivirus Rep proteins appear to have retained their capacity to facilitate replication in prokaryotic hosts (Rigden *et al.*, 1996) and prokaryote-like plastid environments (Groning *et al.*, 1987, 1990). Also, the discovery of geminivirus Rep-related sequences within an extra-chromosomal DNA of a phytoplasma may indicate that occasional geminiviral infections of prokaryotes still occur (Rekab *et al.*, 1999).

The genus *Mastrevirus* contains more diversity than all the other geminivirus genera combined, with the depth of diversity amongst the African streak virus group alone being greater than that found amongst either the New or Old world begomoviruses (Fig 1.1). This has prompted speculation that the mastreviruses are the oldest geminivirus group (Rybicki, 1994). It is unusual therefore that despite the occurrence of all other geminivirus genera in the Americas, no "New World" mastreviruses have yet been discovered there. The degree of sequence diversity amongst the New World geminiviruses is substantially lower than that found in the Old World and it is feasible that the New World viruses were more recently introduced there from Asia. If

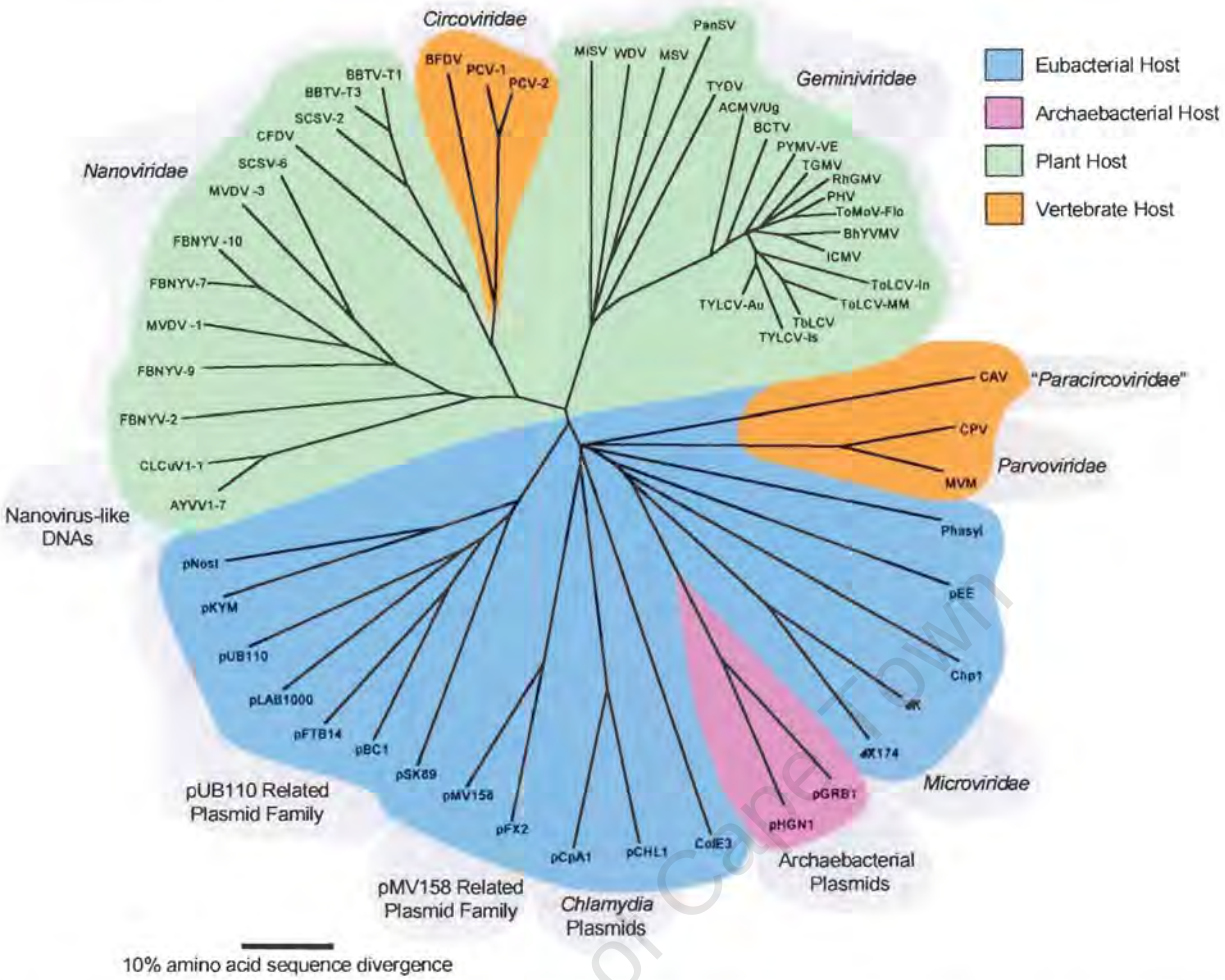


FIGURE 1.3 The possible evolutionary relationships of rolling circle replication (RCR) domains within the replication-associated proteins encoded by a number of ssDNA replicons. Based on a published alignment of the ~80 amino acids spanning the three RCR motifs from a diverse group of ssDNA replicons (Koonin and Ilyina, 1993), this unrooted dendrogram was constructed by the neighbour joining method (Saito and Nei, 1987) using 1000 bootstrap iterations. While bootstrap support for nodes near the centre of the dendrogram is relatively poor (all nodes with less than 30% support have been collapsed), support for all the presented nanovirus, circovirus and geminivirus groupings was in excess of 50%. While the nanoviruses, circoviruses and geminiviruses probably share a common plant virus origin, the circoviruses now have vertebrate hosts (Gibbs and Weiller, 1999). The circovirus, CAV (*Chicken anaemia virus*), probably has an independent prokaryotic origin and is classified here in the "*Paracircoviridae*" according to the suggestion of Takahashi *et al.* (2000). There was 48% bootstrap support for the pUB110 plasmid family being more closely related to the geminivirus/circovirus/nanovirus grouping than to other prokaryotic replicons. The unabbreviated versions of virus names and GenBank accession numbers for the sequences used to construct this dendrogram can be found in Table A.2 (see Appendix A).

mastreviruses are the oldest geminivirus group, then the earliest geminiviruses must have evolved following the separation of Africa and South America 130 million years ago (Rybicki, 1994).

As the number of completely sequenced geminivirus genomes has increased in recent years, so too has the realisation that recombination has been a major contributor to sequence diversity within this family (Padidam *et al.*, 1999a; Rybicki, 1994; Torres-Pacheco *et al.*, 1993; Zhou *et al.*, 1997). It has even been speculated that the *Curtovirus* and *Topocuvirus* genera may have

arisen through ancient recombination events between mastreviruses and begomoviruses (Briddon *et al.*, 1996a; Klute *et al.*, 1996). This speculation was based on the observation that whereas the Rep genes of curtoviruses and topocuviruses more closely resemble those of begomoviruses, the curtovirus and topocuvirus CP genes more closely resemble the CP genes of mastreviruses (Klute *et al.*, 1996; Briddon *et al.*, 1996a). However, more sophisticated recombination analysis of geminivirus genomes has revealed that, while recombination is almost certainly the cause of curtovirus and topocuvirus Rep proteins appearing begomovirus-like, the recombination events involved were most likely far more recent than the divergence of these genera from the begomoviruses and involved only the 5' portion of *Rep* (Fig 1.4; Padidam *et. al.*, 1999a; personal observation). Apart from its involvement in inter-genus recombination, the 5' portion of *Rep* is also the only coding region that has been transferred between both Old and New world begomoviruses, and Asian and African begomoviruses (Padidam *et. al.*, 1999a). Because the ~400 5' nucleotides of so many currently sequenced begomovirus, curtovirus and topocuvirus Rep genes have a recombinant origin (Padidam *et al.*, 1999a), phylogenetic analysis using only this region of these viruses' genomes yields virtually no information on the relationships of the rest of their genomes to one another (Fig 1.4).

1.3 THE EPIDEMIOLOGY OF MAIZE STREAK DISEASE

As is the case with most virus diseases, MSD is erratic: It can be devastating in some years and insignificant in others (Efron *et al.*, 1989). It occurs throughout Africa from northern Egypt (Ammar *et al.*, 1987) to the southernmost parts of South Africa (Willment, 1999), on the Indian Ocean islands of La Réunion and Mauritius (Briddon *et al.*, 1994), the Atlantic Ocean island of Cape Verde (Buchen-Osmond, 1998b), the Middle East (Walkey, *et al.*, 1990) and in India (Seth *et al.*, 1972).

Despite its name, MSD can not be fully understood in terms of the isolated relationship between maize plants, a single maize-adapted MSV strain, and an insect species that transmits viruses between maize plants. It is instead the product of extremely complex interactions between at least 8 species of cicadellid leafhoppers in the Genus *Cicadulina* (Fig 1.1; Dabrowski, 1987; Fennah, 1959; Nielson, 1986; Ruppel, 1965; Soto, 1978), over 80 grass species (Buchen-Osmond, 1998a; Damsteegt, 1983; Konate and Traore, 1992), and an unknown number of virus strains (Konate

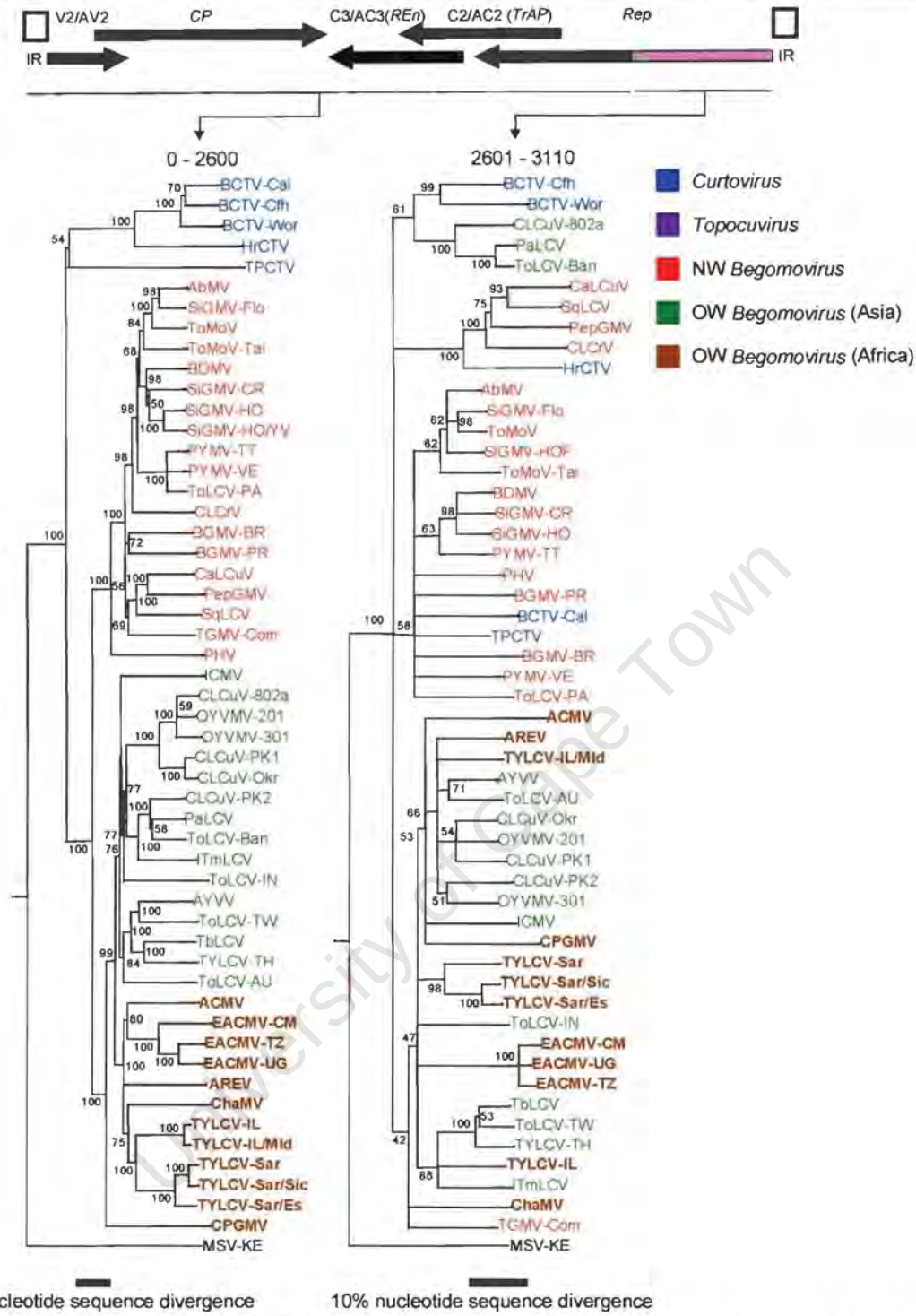


FIGURE 1.4 Recombination within the Rep genes of begomoviruses, topocuviruses and curtoviruses. 52 genomic sequences (using only DNA A components of begomoviruses) were aligned in Clustal W (Higgins *et al.*, 1992). The first nucleotides of every sequence in the alignment were the conserved geminivirus (+) origin sequence, TAATATTAC. The total length of the alignment with insertions was 3304 nucleotides. The alignment was split into three pieces at positions 2600 and 3110. The piece corresponding with nucleotides 3111-3304 in the origional alignment was discarded because of extreme sequence variablity in this region. The pieces of the alignment corresponding to nucleotides 0-2600 and 2601-3110 were each used to construct a rooted neighbour joining tree (Saito and Nei, 1987) with 1000 bootstrap iterations. The phylogenetic relationships of the 510 5' nucleotides of begomovirus, curtovirus and topocuvirus Rep genes (the 2601-3110 tree) differ remarkably from the relationships between the rest of these viruses' genomes (the 0-2600 tree). Numbers associated with tree nodes represent percentages of bootstrap support for the nodes. Nodes with less than 40% bootstrap support were collapsed. Whereas vertical distances are arbitrary, horizontal distances reflect relationships between the sequences. Genbank accession numbers and full virus names are provided in Table A.1 (see Appendix A).

and Traore, 1994; Mesfin *et al.*, 1992; Schnippenkoetter, 1998; Willment, 1999). Compounding this complexity are environmental factors that influence the species composition of grass and leafhopper populations (Dabrowski *et al.*, 1987).

It is not surprising therefore that accurate predictive modelling of MSD epidemics has not yet been achieved. However, loose correlations between disease incidence and both environmental conditions and agricultural practices have been noted. Because there is a strong correlation between rainfall and leafhopper population densities (Asanzi *et al.*, 1994; Okoth and Dabrowski, 1987a; Rose, 1972), drought conditions followed by irregular rains at the beginning of growing seasons tend to be associated with severe MSD outbreaks (Efron *et al.*, 1989; Rossel and Thottappilly, 1985). It has also been noted that in regions with two maize growing seasons a year, the incidence of MSD is higher during the second season (Dabrowski *et al.*, 1991; Efron *et al.*, 1989; Fajemisin *et al.*, 1976).

The primary focus of future epidemiological modelling of MSD will probably be the dynamics of MSV transmission by leafhopper populations. Like most other infectious plant diseases with an insect vector, serious MSD outbreaks are absolutely governed by leafhopper acquisition and movement of severe MSV isolates from infected plants (wild grasses or crop plants) to sensitive, uninfected crop plants. Of the 8 *Cicadulina* species that are capable of transmitting MSV (Fig 1.1), *Cicadulina mbila* is considered to be the most important. It is the most widely distributed species and a larger proportion of *C. mbila* populations have the ability to transmit MSV, compared with other *Cicadulina* species (Story, 1928, 1933; Markham *et al.*, 1984). The relative abundance of the different *Cicadulina* species in different parts of Africa is influenced by altitude, temperature and rainfall (Dabrowski *et al.*, 1987; Okoth and Dabrowski, 1987a)

The virus is acquired by the leafhoppers during the course of their feeding on infected host plants. *C mbila* can feed on more than 138 grass species, ~70% of which are also known to be potential MSV hosts (Buchen-Osmond, 1998a; Damsteegt, 1983; Konate and Traore, 1992). When feeding on an infected plant, a leafhopper can only acquire the virus by feeding directly on chlorotic lesions (Peterschmitt *et al.*, 1992; Storey, 1928). It is believed that MSV infected plants displaying substantial chlorosis may actually attract leafhoppers because of their yellow colour and that leafhoppers may preferentially feed on infected cells within chlorotic lesions because they are a richer supply of nutrients (P. Markham, personal communication).

Uptake of virus particles by a leafhopper during feeding on a MSV infected cell does not automatically enable transmission of MSV by the insect. The ability of leafhoppers to transmit MSV is an inherited, dominant sex-linked characteristic (Story, 1932) and the different *Cicadulina* species vary enormously in both the proportion of individuals capable of transmitting the virus and in the efficiency with which transmission occurs (Asanzi *et al.*, 1995; Dabrowski, 1987; Markham *et al.*, 1984; Van Rensburg and Giliomee, 1990). Once ingested by a leafhopper, the mechanisms by which MSV particles are first transported into its haemocoel, and then into its salivary glands, are unknown.

Following acquisition of the virus by a leafhopper, it becomes viruliferous within 30 hours (Okoth *et al.*, 1987b) and is able to transmit for the rest of its life (Reynaud and Peterschmitt, 1992). Whereas there is some evidence that the begomovirus, *Tomato yellow leaf curl virus* (TYLCV), may be able to replicate within its whitefly vector (Rubinstein and Czosnek, 1997), it is unlikely that MSV replication occurs within leafhoppers since virus titres within an insect decrease over the lifetime of the insect (Reynaud and Peterschmitt, 1992). Another difference between TYLCV in whiteflies and MSV in leafhoppers is that MSV is not transmitted transovarially (Storey; 1928; Van Rensburg and Giliomee, 1990) whereas TYLCV is (Ghanim *et al.*, 1998).

The distance that MSV spreads from a source of innoculum is determined by the movement behaviour of leafhoppers. Distinct long and short distance flight morphs have been detected amongst *Cicadulina* populations in Zimbabwe (Rose, 1972). It is believed that the long flight morphs are a migratory form (Rose, 1972) and as such they may play an important part in the rapid long distance spread of virulent MSV variants (Rose, 1978; Dabrowski *et al.*, 1987). Migratory movement is more common in certain *Cicadulina* species than in others (Dabrowski *et al.*, 1991) and it is probably influenced by environmental conditions (Downham and Cooter, 1998; Rose, 1972).

The dynamics of primary infection following leafhopper invasion of a susceptible maize crop is influenced by leafhopper population densities, the proportion of viruliferous individuals in populations, and the virus titre within these individuals (Rose, 1978; Van Rensburg and Giliomee, 1990). Disease spread is apparently linear when only a few viruliferous leafhoppers are involved in transmission, but becomes exponential once the number of insects exceeds one individual per three plants (Rose, 1978).

Although attempts to understand the dynamics of MSD epidemics have focused primarily on vector population dynamics and behaviour, an important component of any realistic MSD epidemiological model will be the influence of the grass species that are both MSV and vector hosts. Because *Cicadulina* species have favoured hosts for mating and oviposition, the species composition of grass populations in any particular area will directly influence leafhopper population densities in that area (Odhiambo, 1995; Van Rensburg, 1982). Also, many preferred leafhopper hosts are annual grasses, and seasonal alterations in the species composition of grass populations strongly affect leafhopper population densities and feeding behaviours (Okath and Dabrowski, 1987a; Van Rensburg, 1979a).

The species composition and age distribution of grasses (including cultivated crops) in an area may also strongly influence the amount of MSV inoculum available for transfer in that area (Konate and Traore, 1992; Rodier *et al.*, 1995). While MSV infects at least 80 of the 138 grass species that leafhoppers feed on (Buchen-Osmond, 1998a; Damsteegt, 1983; Konate and Traore, 1992), both the susceptibility of these grasses to MSV infection and the severity of symptoms that occur following their infection, may be strongly influenced by a number of factors. While sensitivity to infection can vary substantially from species to species (Damsteegt, 1983), it can also vary within a species with genotype and plant age at the time of inoculation: For example, plants from many species, including maize, generally become more resistant to MSV infection with age (Damsteegt, 1983; Peterschmitt *et al.*, 1992).

Once a viruliferous leafhopper has successfully transmitted virus to a susceptible plant at a suitable growth stage, the time taken for symptoms to develop is correlated with the virus strain transmitted, the virus titre within the insect (Rose, 1978) and the sex of the insect (females are apparently better transmitters than males; Van Rensburg and Giliomee, 1990). Between 3 and 7 days after leafhopper inoculation of plants with MSV, roughly circular spots (0.5 - 2mm in diameter), become visible on the lowest exposed portions of the youngest leaves (Bock *et al.*, 1974). Later symptoms range from narrow veinal streaks a fraction of a millimeter in width to complete leaf chlorosis, stunting and sterility, depending on the MSV strain involved, the host plant species and genotype, and the age of the plant at the time of infection.

Whereas effective control of MSD in cultivated crops is possible with the use of carbamate insecticides (Drinkwater *et al.*, 1979; Van Rensburg, 1988; Van Rensburg *et al.*, 1978; Van Rensburg and Giliomee, 1991), it is widely believed that the development and use of MSV

resistant crop genotypes is the best way to minimise the impact of MSD on African agriculture (Efron *et al.*, 1989; Rodier *et al.*, 1995; Tang and Bjarnason, 1993). Great successes have been achieved in the development of MSV resistant maize genotypes that tolerate infection without significant yield loss (Barrow, 1992; Efron *et al.*, 1989; Rodier, 1995; Van Rensburg *et al.*, 1991). Besides providing relatively inexpensive security against crop losses, MSV infected tolerant maize genotypes contain 10 to 90 times fewer virus particles than infected sensitive genotypes (Peterschmitt *et al.*, 1992) and are therefore much poorer sources of innoculum during secondary disease spread (Rodier *et al* 1995).

While efforts are underway to promote the widespread cultivation of MSV resistant maize in Africa, surprisingly little is known about the MSV populations that will confront these new genotypes. Attempts to determine the depth of MSV's genetic diversity have just begun and to date only three major MSV strain groupings have been discovered (Fig 1.1; Rybicki *et al.*, 1998; Schnippenkoetter, 1998; Willment, 1999). The other African streak virus species, of which only PanSV and a few SSV species have yet been described (Bigarre *et al.*, 1999; Briddon *et al.*, 1992; Briddon, 1996b; Hughes *et al.*, 1992), have received even less attention than MSV.

Although MSV is by far the most studied African streak virus species, virtually all research efforts have been focused on the characterisation of only one MSV strain – MSV-A (Fig 1.5). The nucleotide sequence divergence of MSV-A isolates obtained from opposite ends of the continent is very low (< 3% divergence; Briddon *et al.*, 1994) which may indicate either that the rate at which MSV-A is evolving is fairly slow, or that continent wide spread and dominance of new MSV genotypes with enhanced fitness is very rapid. Experimental approximations of the rates at which a MSV-A isolate evolves when maintained in a susceptible maize genotype, a resistant maize genotype and a non-maize host (*Coix lacryma-jobi*) are respectively, 9.5×10^{-5}, 17.3×10^{-5}, and 26.5×10^{-5} substitutions per site per year (Isnard *et al.*, 1998). These rates of evolution are relatively low and imply that only one nucleotide per MSV genome becomes fixed every ~1.5 to ~4 years (depending on the selection pressures exerted by host species).

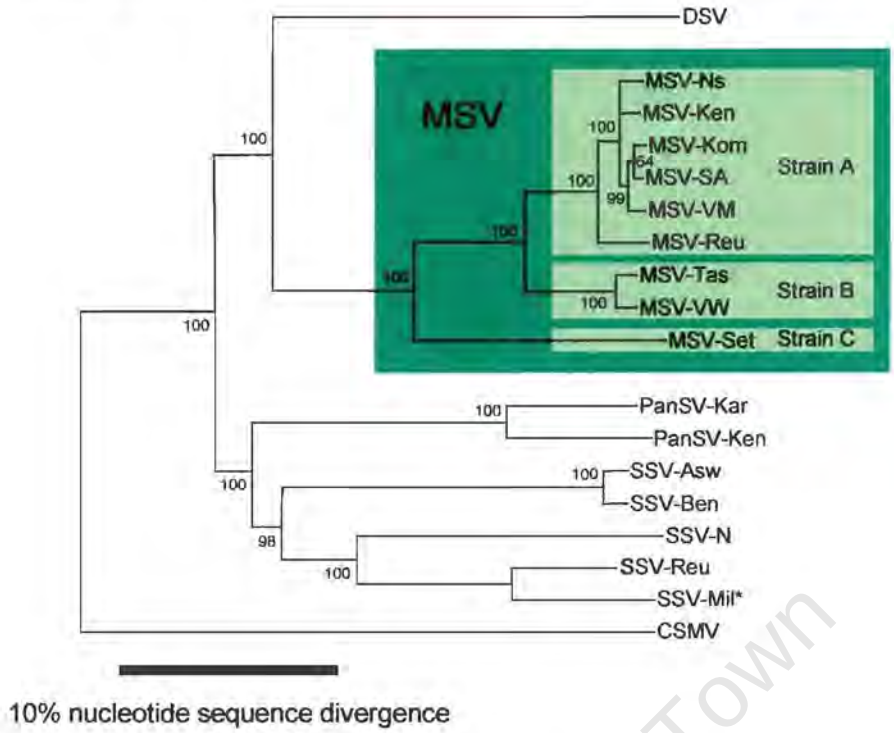


FIGURE 1.5 The African streak viruses. Three MSV strains are currently known to exist. Strain A contains isolates that are known to produce severe streak symptoms in maize. Whereas strain B viruses have been isolated from wheat and wild grasses the only discovered strain C virus, MSV-Set, was isolated from a *Setaria* sp. The other African streak viruses are PanSV and at least 3 distinct SSV species. DSV and CSMV are included here for reference purposes. While DSV shares remarkable similarities with the African streak viruses, the isolate presented here was obtained from the island of Vanuatu and is therefor technically not an African streak virus. Full genome sequences of all viruses (except SSV-Mil for which only a *CP* and SIR sequence is available) were aligned in DNAMAN (Lynnon BioSoft, Quebec) and used to construct a rooted neighbour joining tree (Saito and Nei, 1987) with 1000 bootsrap iterations. Numbers associated with tree nodes represent percentages of bootstrap support for the nodes. Nodes with less than 60% bootstrap support were collapsed. Whereas vertical distances are arbitrary, horizontal distances reflect relationships between the sequences. GenBank accession numbers for these sequences and full virus names are listed in Table A.1 (see Appendix A).

Besides MSV-A, isolates of two other MSV strains, MSV-B and -C, have been cloned and fully sequenced (Fig 1.5; Schnippenkoetter, 1998; Willment, 1999). While MSV-B isolates are common in streak-infected grasses and wheat (Rybicki *et al.*, 1998; Schnippenkoetter, 1998; Willment, 1999), they have only rarely been detected in maize (Willment, 1999). Only one MSV-C isolate (from a *Setaria* sp.) has ever been detected. Because MSV-A isolates are so infrequently found in MSV infected non-maize hosts (Willment, 1999), it is possible that they may have been as rare as MSV-C isolates before the arrival of maize in Africa.

It has been demonstrated that cloned isolates of all 3 MSV strains are capable of establishing infections in a range of host species and genotypes (Schnippenkoetter, 1998; Willment, 1999). However, the severity of symptoms resulting from these infections differ according to strain, with different strains producing their severest symptoms in different hosts (Schnippenkoetter, 1998; Willment, 1999). It would therefore appear that while evolutionary selection has yielded MSV

strains adapted to survival in certain species, it has not abolished the capacity of these strains to infect a variety of sub-optimal hosts. The reason that MSV strains have not become very narrowly host adapted is probably related to the feeding behaviours of *Cicadulina* species. Because these insects feed on an enormous range of grass species, there would be only a small probability that an extremely host-adapted virus once acquired by a leafhopper would be transmitted into a host that it could infect.

It is unknown whether any MSV strains other than MSV-A play an important role in the epidemiology of MSD. Both MSV-B and -C isolates have been found to produce mild symptoms only in maize genotypes that are very MSV susceptible and are therefore unlikely to pose any significant threat to maize production (Schnippenkoetter, 1998; Willment, 1999). It has been directly demonstrated that mixed MSV-A and -B infections occur in nature and there is also strong evidence of recombination between these strains (Willment, 1999). It is therefore possible that MSV-B, -C and other as yet undiscovered MSV strains may indirectly influence MSD epidemiology through recombination with MSV-A type viruses. Recombination has been linked with the emergence of a number of geminivirus diseases (Umaharan *et al.*, 1998; Zhou *et al.*, 1997; Zhou *et al.*, 1998) and it is quite conceivable that it may also eventually contribute to the evolution of MSV genotypes with elevated virulence.

1.4 THE MOLECULAR BIOLOGY OF MAIZE STREAK VIRUS

There has been a substantial drive in recent years to determine how geminiviruses replicate and move their genomes from cell to cell during infections. While the primary motivation behind these studies has been to shed some light on the still obscure molecular details of plant DNA replication and intercellular trafficking, they have also provided fascinating insights into the intricate survival mechanisms that these viruses have evolved.

1.4.1 The Genome Organisation of Maize Streak Virus

In stark contrast to the complexity of MSD, the genome of its causal agent is very simple. It is composed of only 4 genes, and two intergenic regions that contain sequence elements required for transcription and replication (Fig 1.2). Despite this, however, MSV is still capable of extremely elaborate behaviour within the cells of its hosts. Before an attempt is made to describe

the molecular mechanisms that operate to produce an infection, the small collection of genomic features comprising a MSV genome will be examined in detail.

1.4.1.1 The long intergenic region

All geminivirus genomes contain at least one intergenic region (IR) from which virus genes diverge in both the complementary and virion senses. Mastreviruses have two intergenic regions; a short or small one (SIR) which is unique to this genus and a long or large one (LIR) which is analagous to the single IR of the curtoviruses, topocuviruses and the A component of begomoviruses (Fig 1.2). Besides containing divergent RNA polymerase II-type promoters and other transcriptional regulatory features that are necessary for the expression of the complementary and virion sense genes, the LIRs of mastreviruses and the IRs of other geminiviruses also contain sequence elements that are essential for the replication of viral DNA (Gutierrez, 1999; Palmer and Rybicki, 1998; Fig 1.3).

The most striking feature within the LIR of the mastreviruses and IRs of other geminiviruses is an inverted repeat sequence that is capable of forming a stable hairpin loop structure in duplex DNA (Fig 1.6; Sunter *et al.*, 1985). All geminiviruses sequenced to date have the highly conserved nonanucleotide sequence, TAATATT↓AC, within the loop: This contains the (+) strand origin of replication (↓; Heyraud *et al.*, 1993a; Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995a; Stanley, 1995; Stenger *et al.*, 1991). Related nonanucleotide sequences have also been found in similar hairpin loop structures identified in nanoviruses, circoviruses, and a range of geminivirus associated DNAs, (Fig 1.2; Dry *et al.*, 1997; Mansoor *et al.*, 1999; Saunders *et al.*, 2000; Saunders and Stanley, 1999). The exact sequence of the nonanucleotide is not absolutely required in mastreviruses and it has been demonstrated that a MSV mutant possessing the sequence, TAATACTAC, is only slightly less fit than a wild-type virus (Schneider *et al.*, 1992).

Whereas the sequence of the stem is highly conserved amongst the begomoviruses, topocuviruses, and curtoviruses, substantial differences exist in the stem sequences found in the mastreviruses. While the stem loop structure is absolutely required for geminivirus replication (Laufs *et al.*, 1995a; Orozco and Hanley-Bowdoin, 1996), the sequence of the stem does not appear to be as crucial and only contributes to the efficiency of replication (Orozco and Hanley-Bowdoin, 1996; Willment, 1999). There is, however, some confusion as to whether the stem-loop structure is

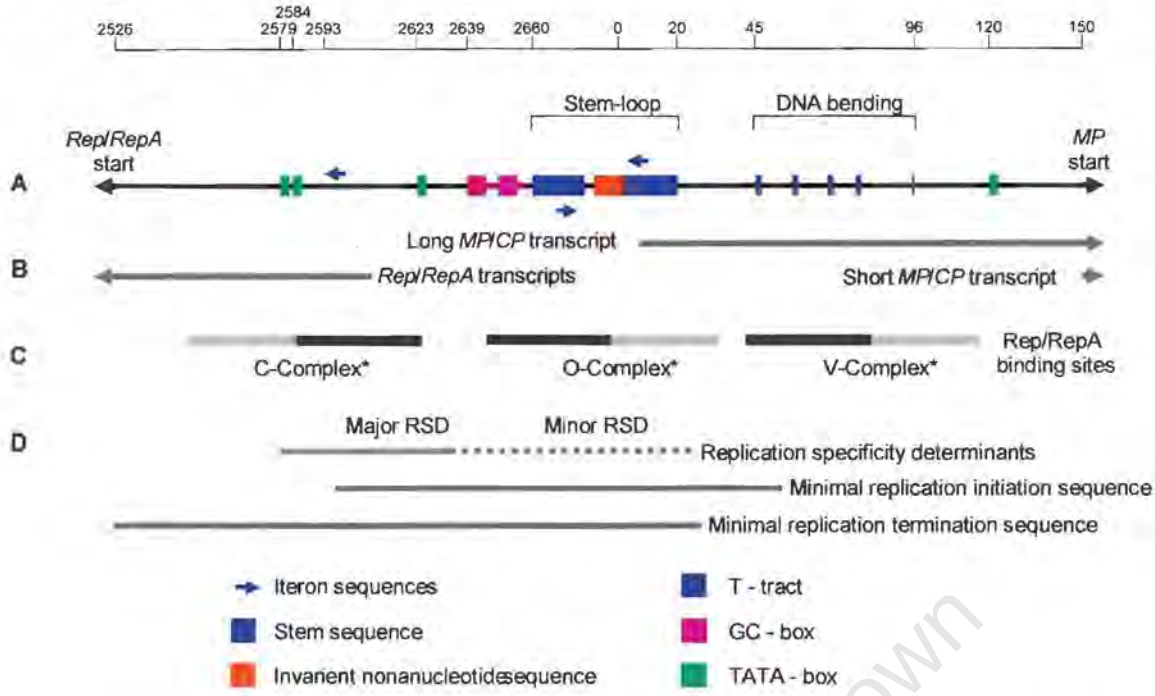


FIGURE 1.6 A schematic representation of the MSV-Ns long intergenic region (LIR). (**A**) The main landmarks within the LIR. These include a stem-loop structure and highly conserved nonanucleotide at the origin of (+) strand replication, iterated sequences (iterons) potentially involved with Rep recognition or binding to the LIR, TATA and GC-boxes boxes involved in gene transcription, and a region of static DNA curvature. (**B**) Initiation points of complementary and virion sense transcripts. (**C**) Possible positions of Rep binding within the LIR inferred by analogy to those detected in *Wheat dwarf virus* (WDV). (**D**) Minimal sequence requirements for MSV replication initiation and termination. The regions that contain replication specificity determinants (RSDs) most likely contain Rep binding sites analagous to those detected in WDV. Numbering is relative to the (+) strand replication initiation site. References are within the text.

required for both initiation and termination of virion strand replication or whether it is only required for termination (Gutierrez, 1999, Heyraud *et al.*, 1993b, Kammann *et al.*, 1991, Palmer and Rybicki, 1998). This confusion stems from the demonstration with wheat dwarf virus (WDV) that while iterated sequences found in the stem and a sequence resembling the conserved nonanucleotide are essential for initiating replication, the stem-loop structure is not explicitly required (Kammann *et al.*, 1991). There is, however, a rather fortuitous and apparently unique arrangement of nucleotides within the WDV LIR that enables replication initiation from a second site (Heyraud *et al.*, 1993b). It is therefore likely that WDV is an exceptional case and that the stem-loop is generally required for both initiation and termination of geminiviral replication.

A sequence occurring between the TATA box believed to direct *Rep/RepA* transcription (Boulton *et al.*, 1991a) and the *RepA* initiation codon is directly repeated in the stem of the stem-loop (Fig 1.6). This repeated sequence, called an iteron, occurs in all mastreviruses sequenced to date and ranges in size from 6 to 12 nucleotides for different viruses. By analogy with similar iterated sequences identified as Rep binding sites in begomoviruses and curtoviruses, these sequences are

also believed to be involved in Rep binding (Arguello-Astorga *et al.*, 1994a; Choi and Stenger, 1996; Fontes *et al.*, 1994a, b). Electron microscopic visualisation of WDV Rep and RepA binding complexes and DNAse I footprinting studies have indicated that spherical Rep and Rep A oligomers comprising 8 molecules form ~144 nucleotides upstream from the replication initiation site in the region of the 5' iteron (Castellano *et al.*, 1999; Missich *et al.*, 2000; Sanz-Burgos and Gutierrez, 1998). These "C-complexes" are approximately the same size for both Rep and RepA and span ~75 nucleotides (Missich *et al.*, 2000). Unlike begomovirus Rep proteins (Fontes *et al.*, 1992), WDV Rep also binds at both the stem loop to form an "O-complex" and ~90 nucleotides downstream from the replication initiation site to form a "V-complex" (Castellano *et al.*, 1999). While RepA also forms a V-complex its footprint is 10 nucleotides smaller than that of the Rep V-complex. Whereas the WDV iteron upstream of the *Rep* initiation codon was encompassed by the DNAse I footprint of the Rep and RepA C-complexes, the footprints of both their C and V-complexes encompassed the consensus sequence, GTGTGaN$_{22-23}$GTG(G)TC (on the (+) sense strand at the site of the C-complex and on the (-) strand at the site of the V-complex), which Castellano *et al.* (1999) propose is the actual Rep recognition sequence. Only part of the consensus is present in the WDV stem sequence encompassed by the O-complex and it is believed that this may explain the low affinity with which this complex forms relative to the other two (Castellano *et al.*, 1999).

It is unknown whether similar Rep complexes form within the LIRs of MSV and other mastreviruses. Nucleotide sequence analysis of these other LIRs has not revealed any arrangements of directly and inversely repeated sequences analogous to the Rep recognition sequences found in the WDV LIR. It is worthwhile noting that there are quite substantial differences between WDV and MSV LIRs. Firstly, the LIR between the replication initiation site and *RepA* start codon in WDV is substantially larger (239 nucleotides) than the corresponding region in MSV (162 nucleotides in MSV-Ns). The distance between the *RepA* TATA box and the *RepA* proximal iteron is 43 nucleotides in WDV and only 22 in MSV. Secondly, relative to the *RepA* start codon, complimentary sense transcription in MSV can be initiated from three TATA boxes located at nucleotide positions -101, -62 and -57, whereas in WDV it can only be initiated from 2 overlapping TATA boxes at positions -131 and -129. Lastly, the distance between the WDV *RepA* proximal iteron and the -129 TATA box and replication initiation site is 43 and 158 nucleotides respectively, whereas the distance between the iteron and -101 TATA box and replication initiation site in MSV-Ns is only 22 and 87 nucleotides respectively.

Assuming that the Rep oligomer complexes forming on the LIR of MSV will be approximately the same size as those in WDV, a far greater proportion of the MSV LIR would be contained within these complexes than has been noted for WDV (Fig 1.6). A MSV C-Complex could conceivably cover all three of the *RepA* proximal TATA boxes. While it is not possible to predict the exact position within the MSV LIR where Rep oligomer complexes might form, indirect evidence for the existence of C and O complexes has been obtained. Major and minor replication specificity determinants (RSDs) in MSV have been mapped to regions in the MSV LIR corresponding to the positions of the WDV C- and O- complex sites, respectively (Fig 1.6; Willment, 1999). It is probable that these RSDs span sites of Rep binding in the vicinity of the MSV RepA proximal iteron and stem loop (Willment, 1999).

An approximately 80bp region of static DNA curvature between the stem-loop and the *MP* start codon has been identified in WDV (Suárez-López *et al.*, 1995), MSV and other mastreviruses (Gutierrez *et al.*, 1995). The curvature is produced by a series of "A-T tracts" (Fig 1.4) which, on the (+) strand of WDV, are comprised of groups of three to four repeated T residues separated from one another by between 7 and 17 nucleotides (Suárez-López *et al.*, 1995). By analogy with demonstrations in other experimental settings (reviewed by Perez-Martin and de Lorenzo, 1997), it is believed that bending of this region may play a role in transcriptional regulation of mastrevirus virion sense genes (Castellano *et al.*, 1999). Besides the detection of Rep and/or RepA enhancement of virion sense gene expression (Collin *et al.*, 1996; Hofer *et al.*, 1992), it has been noted that the Rep V-complex spans 4 out of 6 of the A-T tracts in WDV (Castellano *et al.*, 1999). It is not inconceivable therefore that V-complex formation may influence virion sense gene transcription through alteration of DNA curvature within the V-sense promoter (Castellano *et al.*, 1999).

In the absence of both Rep and RepA expression, the MSV virion sense gene promoter is only about 20% as strong in maize protoplasts as the CaMV35s promoter (Fenoll *et al.*, 1988). The promoter required for optimal CP expression has been localised to a region 530 nucleotides upstream of the *CP* start codon (Fenoll *et al.*, 1988, 1990). This region, called the upstream activator sequence (UAS), includes the stem loop and two directly repeated GC boxes at the base of the stem proximal to the *RepA* start codon. One or two GC boxes at this position is a conserved feature of all the geminivirus genomes yet sequenced. In MSV, the region containing the GC boxes binds maize nuclear factors to the UAS and has been named the rightward promoter element (*rpe1*; Fig 1.4; Fenoll *et al.*, 1990). Besides functioning in transcription of

virion sense genes, it is also possible that *rpe1* may recruit nuclear factors to this region for purposes of replication (Arguello-Astorga *et al.*, 1994b). Nuclear factors bind non-cooperatively on opposite faces of the DNA helix to both of the MSV GC boxes (Fenoll *et al.*, 1990). The importance of binding on opposite faces is probably not significant as there is only one box in PanSV whereas distances between the repeated boxes in SSV, CSMV and WDV differ and so therefore will the relative position of factor binding (Hughes *et al.*, 1993; Palmer and Rybicki, 1998).

1.4.1.2 The short intergenic region

A short intergenic region (SIR) exists between the termination codons of the CP and Rep genes of all mastreviruses. Despite containing both the origin of (-) strand synthesis and the polyadenylation and termination signals of the virion and complementary sense transcripts (Fig 1.7), the SIR has not been very thoroughly investigated.

A small complementary sense primer-like molecule ~80 nucleotides in size and of unknown origin has been found bound within the SIR of enapsidated MSV, WDV, *Digitaria streak virus* (DSV), *Chloris striate mosaic virus* (CSMV), and *Tobacco yellow dwarf virus* (TYDV) DNA (Donson *et al.*, 1987; Donson *et al.*, 1984; Andersen *et al.*, 1988; Hayes *et al.*, 1988; Morris *et al.*, 1992). The 52 3' nucleotides of the primer binding site and 8 additional nucleotides past its 3' end are, together with most of the LIR, the only sequences always present in naturally occurring MSV subgenomic DNAs (Fig 1.7; Billharz, Martin, and Rybicki unpublished). At the onset of an infection, the primer molecule probably enables synthesis of a dsDNA replicative forms from newly uncoated (+) strand DNA. While it is possible that a similar primer is produced during active geminivirus replication (Gutierrez, 1999), its occurrence in virions may simply be a consequence of encapsidation prematurely arresting (-) strand synthesis (Donson *et al.*, 1984).

It appears as though altering the size of the SIR has an impact on the fitness of MSV. While addition or removal of small oligonucleotides within the SIR have been found to result in slight reductions of fitness correlated with the size of insertions and deletions, there is a strong selection

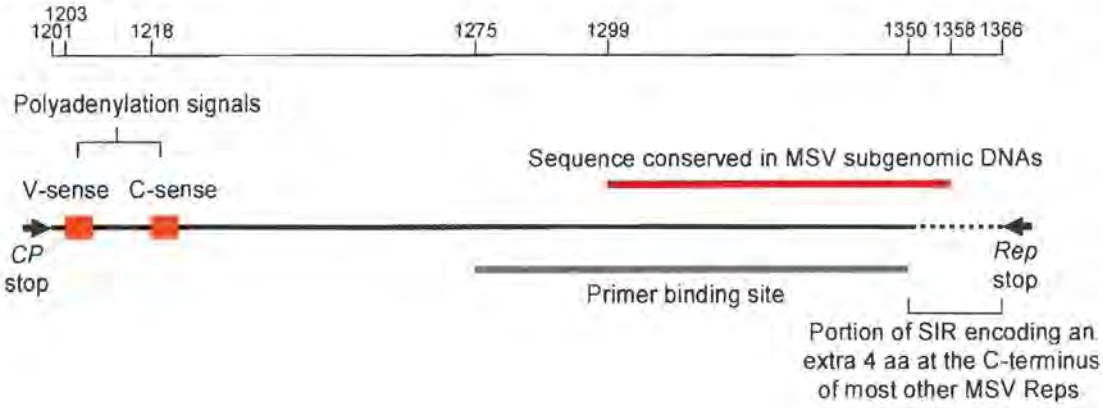


FIGURE 1.7 A schematic representation of the short intergenic region (SIR) of MSV-Ns. Orange boxes indicate the positions of polyadenylation signals for both virion and complementary sense transcripts. The primer binding site is the region of the SIR that binds a small complementary DNA molecule encapsidated within virions. The region conserved in MSV subgenomic DNAs is likely to contain the minimal origin of (-) strand replication. *Rep* terminates at slightly different positions in different MSV isolates. Numbering is relative to the (+) strand replication initiation site. References are within the text.

pressure for reversion to near wild-type SIR size (Shen and Hohn, 1991). It is possible that subtle constraints on SIR size in autonomously replicating viruses are related to the mechanics of replication. Within a group of 28 naturally occurring MSV subgenomic DNAs it was noted that a 60 nucleotide portion of the SIR (52 of the nucleotides overlapping the 3'portion of the primer binding site) and a 239 nucleotide portion of the LIR (from – 164 to 75 relative to the (+) strand origin of replication) were consistently present (Billharz, Martin, and Rybicki unpublished). However, the fewest number of nucleotides separating these two "core regions" was 112 nucleotides 5' and 141 nucleotides 3' of the SIR core region, implying that a degree of spatial separation between the minimal LIR and SIR sequences may be required for efficient replication (Billharz, Martin, and Rybicki unpublished).

Despite the overwhelming retention of an SIR core region amongst MSV subgenomic DNAs, the exact sequence requirements for the initiation of (-) strand synthesis are almost certainly very lax. It has been demonstrated that replacing the SIR of the dicot infecting mastrevirus, BeYDV, with that of MSV has only a slight effect on the ability of the chimeric BeYDV to replicate in tobacco protoplasts (Liu *et al.*, 1999b). While the sequences of the MSV and BeYDV SIRs share 55% sequence identity, they only share 48% sequence identity in the core SIR region conserved in all the MSV subgenomic DNAs.

1.4.1.3 The complementary sense genes (*Rep and RepA*)

All geminiviruses produce a replication initiator/associated protein (Rep) that is the only viral gene product that is absolutely required for replication (Elmer *et al.*, 1988; Hanley-Bowdoin *et al.*, 1990; Hayes *et al.*, 1989; Schalk *et al.*, 1989). In mastreviruses, *RepA* (the C1 ORF) and the C2 ORF (also called *RepB*), respectively encode the amino- and carboxy-terminal parts of Rep (Fig 1.2 and Fig 1.8; Mullineaux *et al.*, 1990; Schalk *et al.*, 1989). Because the C2 ORF lacks both a detectable transcript and a methionine start codon in all mastrevirus species except MSV, it is unlikely that it encodes any autonomous protein (Palmer and Rybicki, 1998).

In MSV, complementary sense (C-sense) transcription can potentially be initiated from any of three TATA boxes located 57, 62 and 101 nucleotides upstream from the *RepA* start codon (Fig 1.6). Beginning at the same transcription initiation site, two C-sense transcripts (1.5Kb and 1.2Kb; Fig 1.8B) are produced during MSV infections (Dekker *et al.*, 1991; Mullineaux *et al.*, 1990). Splicing of the larger transcript removes an intron which permits expression of full length Rep from the two ORFs (Wright *et al.*, 1997). Introns have also been directly demonstrated in the Rep genes of DSV (Accotto *et al.*, 1989), WDV (Schalk *et al.*, 1989), and TYDV (Morris *et al.*, 1992). It is highly likely that analogous introns also occur in the Rep genes of all other mastreviruses so far sequenced.

It is very probable, although as yet unproven *in vivo*, that RepA, a variant of Rep that is unique to the mastreviruses, is translated from both the unspliced 1.5Kb and the 1.2Kb C-sense transcripts (Dekker *et al.*, 1991; Mullineaux *et al.*, 1990; Gutierrez, 1999). If expressed in infected cells, MSV RepA would have the same N-terminal 214 amino acid sequence as Rep, but would have a different C-terminus (Fig 1.8C). Efficient replication of WDV in *Triticum monococcum* protoplasts with its native Rep replaced with an "intronless" cDNA of the spliced transcript has demonstrated that expression of Rep alone is sufficient for replication (Schalk *et al.*, 1989). While not absolutely required for replication, there is substantial evidence that RepA is a multifunctional protein with unique features (Table 1.2) that are required at different stages during the replicative cycles of mastreviruses (Gutierrez, 1999).

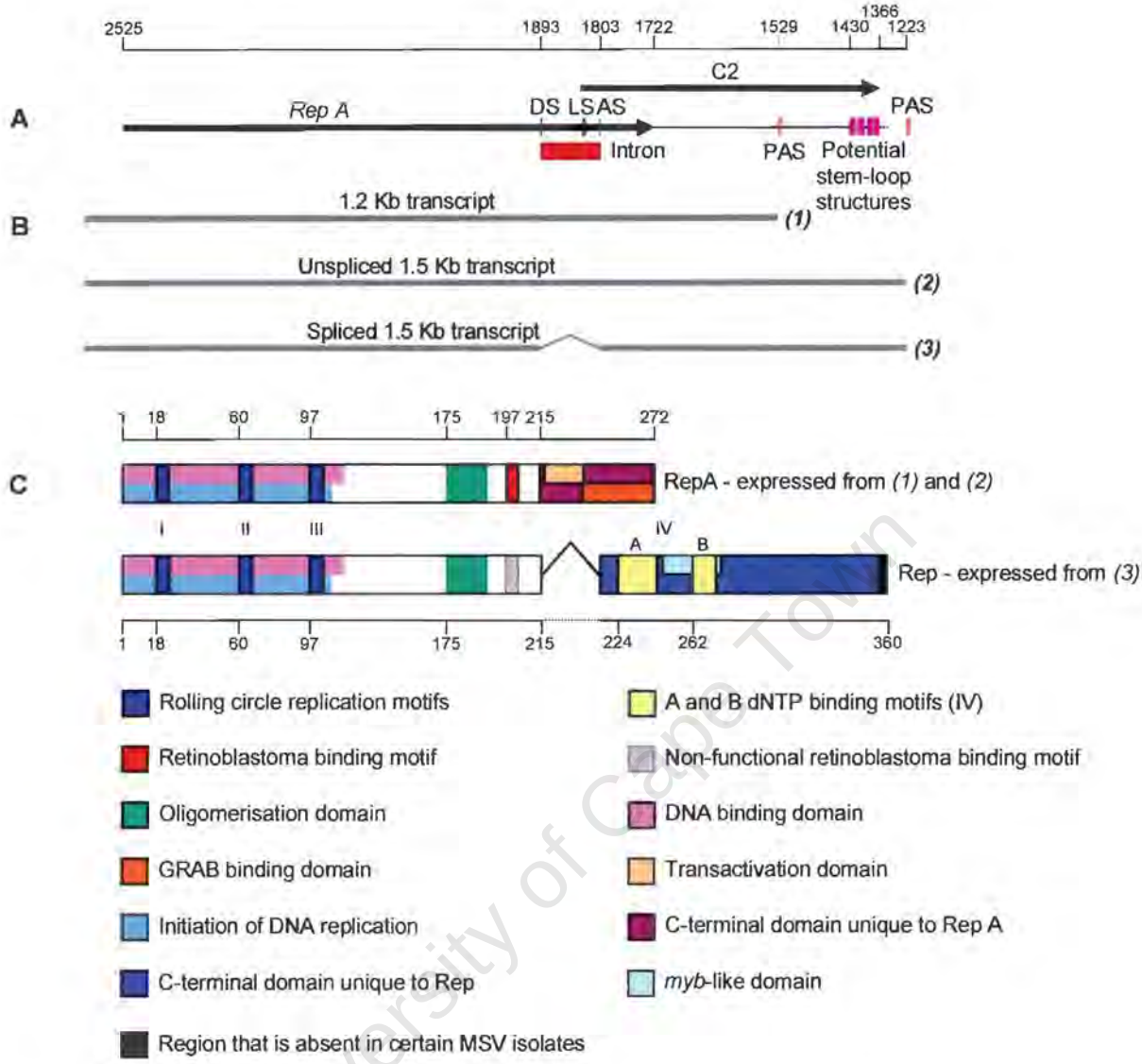


FIGURE 1.8 The MSV-Ns *RepA*(C1) and C2 ORFs, their expression and the functional domains of their products. (**A**) The C-sense intron (in red) is bounded by donor and acceptor sites (DS and AS respectively) and contains a consensus branch point or lariat sequence (LS). Orange boxes represent potential C-sense transcript polyadenylation signals (PAS). At the 3' end of the C2 ORF are a series of sequences that are conserved between MSV-isolates and are potentially capable of forming stable stem-loop structures (Donson *et al.*, 1984). (**B**) Small (1) and large (2) C-sense transcripts are produced with 20% of the large transcripts being spliced to produce the mRNAs from which Rep is expressed. RepA is potentially expressed from either long or short unspliced transcripts. (**C**) The sequence motifs and functional domains that have been detected in mastrevirus Reps (After Gutierrez, 1999). Note that no mastrevirus DNA binding domain has yet been mapped and the region indicated here is inferred by analogy to those mapped for begomovirus and curtovirus Reps (Choi and Stenger, 1995, 1996; Jupin *et al.*, 1991). Also, there is currently no evidence that the GRAB binding activity of WDV's RepA is shared by MSV's RepA. Numbering of DNA is relative to the (+) strand replication initiation site. References are within the text.

TABLE 1.2 The charateristics of Rep and RepA

Activity	Rep	RepA	Reference
Rb binding	-	+	Horvath *et al.*, 1998
GRAB binding	-	+	Xie *et al.*, 1999
Downregulation of replication	-	+	Collin *et al.*, 1996
Required for replication	+	-	Schalk *et al.*, 1989
Sequence specific DNA binding	+	+	Castellano *et al.*, 1998; Missich *et al.*, 2000
Oligomerisation	+	+	Horvath *et al.*, 1998; Sanz-Burgos and Gutierrez, 1998;
ATPase activity	+[a]	-	Hanson *et al.*, 1995; Desbiez *et al.*, 1995
Transactivation activity in yeast	+[b]	+	Horvath *et al.*, 1998
Transactivation of V-sense genes	+	+	Collin *et al.*, 1996; Hofer *et al.*, 1992; Zhan *et al.*, 1993)

[a]Inferred by analogy with the begomoviruses, BGMV and TYLCV
[b]Transactivation activity required truncation of the protein

The N-terminal portions of Rep and RepA that are common to both proteins contain a number of conserved amino acid sequence motifs with either known, unknown or suspected activities (Fig 1.8C). Three of these motifs, named I, II and III, were identified through their apparent conservation with those found in replication-associated proteins from ssDNA phages and plasmids that replicate via a rolling circle mechanism (Fig 1.3; Koonin and Ilyina, 1993). Whereas motif I has an unknown function, motif II is possibly involved in binding Mn^{2+} or Mg^{2+} (Koonin and Ilyana, 1992; Laufs *et al.*, 1995a) and motif III participates in DNA cleavage and linkage at the virion sense origin of replication (Laufs *et al.*, 1995b).

A motif similar to the retinoblastoma protein (Rb) binding motif found in a range of viral oncoproteins (Ludlow, 1993; Moran, 1993; Voesden, 1993) is found in the Rep and RepA proteins of most mastreviruses (Xie *et al.*, 1995). Whereas WDV and MSV RepA interact with human Rb proteins and a homologous maize Rb related protein (RbR) via this motif (Grafi *et al.*, 1996; Horvath *et al.*, 1998; Xie *et al.*, 1996), it appears as though MSV and BeYDV Rep proteins do not interact with RbR (Horvath *et al.*, 1998; Liu *et al.*, 1999c). It is possible that the failure of these and possibly other mastreviral Rep proteins to interact with RbR, despite the presence of a binding motif, may be a consequence of the motif being obscured by protein folding (Horvath *et al.*, 1998; Gutierrez, 1999).

Besides the amino acid sequence motifs common to Rep and RepA, a number of domains potentially participating in the various activities of these proteins have been identified. It is highly likely that mastrevirus Rep and RepA proteins perform certain of their functions in an oligomerised form (Sanz-Burgos and Gutierrez, 1998; Missich *et al.*, 2000). A 13 amino acid region between the Rb binding domain and motif III has been implicated in Rep-Rep, RepA-RepA and RepA-Rep interactions (Fig 1.8C; Horvath *et al.*, 1998).

While DNA recognition and binding domains have been mapped to the N-terminal 116 and 89 amino acids of begomovirus (Jupin *et al.*, 1995) and curtovirus (Choi and Stenger, 1996) Rep proteins, respectively, the observed sequence-specific DNA binding activities of mastrevirus Rep and RepA (Castellano *et al.*, 1999; Sanz-Burgos and Gutierrez, 1998; Missich *et al.*, 2000) are also likely to reside in this region (Fig 1.8C; Gutierrez, 1999).

Besides sharing common amino acid sequence motifs and functional domains, Rep and RepA both have unique C-termini that potentially enable these proteins to perform very different roles in the life cycles of mastreviruses. Conserved amongst geminivirus Rep proteins (but not found in RepA) is a dNTP binding domain (also referred to as motif IV) with A and B motifs similar to those found in proteins with kinase and helicase activities (Fig 1.8C; Gorbalenya and Koonin, 1989). It has been demonstrated that while in begomoviruses the motif has an associated ATPase activity that is apparently required for efficient replication (Desbiez *et al.*, 1995; Hanson *et al.*, 1995), the motif is not required for the Rep-mediated nicking reaction at the (+) strand origin during replication (Heyraud-Nitschke *et al.*, 1995). Although the exact purpose of the ATPase activity is uncertain, it is possible that the motif may have an associated helicase activity that is operational during replication (Breret *et al.*, 1999; Desbiez *et al.*, 1995)

A region with similarity to the DNA binding domains of the *myb*-related class of plant transcription factors is also conserved amongst all geminivirus Rep proteins and overlapping the dNTP binding motif (Hofer *et al.*, 1992). For MSV, the transactivation activity of this region has been demonstrated in yeast (Horvath *et al.*, 1998) and it is conceivable that it may enable Rep to induce virion sense gene transcription (Hofer *et al.*, 1992; Zhan *et al.*, 1993) or else to directly transactivate expression of host genes during an infection (Palmer and Rybicki, 1998). Interestingly, full length Rep is unable to activate transcription in yeast reporter gene constructs, implying that the transactivation domain may be masked during protein folding. Such a masking effect has been detected for the ATF-2 transcription factor (Li and Green, 1996) and may regulate the transactivation activity of Rep by ensuring that the domain is only exposed following binding of Rep to appropriate host proteins (Horvath *et al*, 1998).

Associated with the unique C-terminus of MSV RepA is another transactivation domain with demonstrable activity in yeast (Fig 1.7C; Horvath *et al.*, 1998). In WDV there is some evidence to suggest that RepA but not Rep may be responsible for transactivation of V-sense gene

transcription (Collin *et al.*, 1996). It is feasible that a domain in the C-terminal portion of RepA may be involved in this activity.

Also unique to the C-terminus of mastrevirus RepA proteins is a region that, in WDV at least, has been found to interact with a group of host GRAB (geminivirus RepA binding) proteins (Xie *et al.*, 1999). These GRAB proteins resemble members of a family of plant proteins containing a NAC (non-apical-meristem, ATAF and CUC2 genes) domain. It appears that NAC proteins are involved in diverse plant developmental pathways and it is possible that RepA-GRAB binding may influence these pathways for the benefit of viral processes (Gutierrez, 2000).

1.4.1.4 The virion sense genes (*MP* and *CP*)

All mastrevirus genomes contain movement protein (MP) and coat protein (CP) genes that are absolutely required for systemic infection of host plants (Boulton *et al.*, 1989a; Lazarowitz *et al.*, 1989; Woolston *et al.*, 1989). Whereas CP genes are similarly positioned in every geminivirus genome yet sequenced, the genomic location of MP genes is more varied. Curtovirus and monopartite begomovirus genes with movement functions are found in the same position as that of mastreviruses but bipartite begomoviruses have two genes encoding MPs (MPB and NSP) on their DNA B components (Fig 1.2). Unlike the substantial amino acid sequence identities shared by all geminivirus CPs, the MPs of viruses in the different genera have no detectable sequence homology and it is a distinct possibility that they all have different evolutionary origins (Rybicki, 1994).

Transcription of the MSV virion sense genes is probably directed by two TATA boxes 26 and 214 nucleotides upstream from the *MP* start codon (Wright *et al.*, 1997). Two co-terminal bicistronic transcripts are produced – a less abundant transcript initiated ~145 upstream from the *MP* start within the stem sequence of the stem loop and a more abundant transcript initiated only 1 nucleotide upstream from the *MP* start codon (Fig 1.9; Wright *et al.*, 1997). While it is probable that MP and CP can be expressed from both virion sense (V-sense) transcripts (Wright *et al.*, 1997), CP is generally present in infected cells at much higher concentrations than MP (Mullineaux *et al.*, 1988). In MSV and certain other mastreviruses, splicing of an intron within the *MP* portion of V-sense transcripts appears to be an important determinant of relative MP and CP expression levels (Fig 1.9B; Wright *et al.*, 1997). *In vitro* expression analysis has revealed

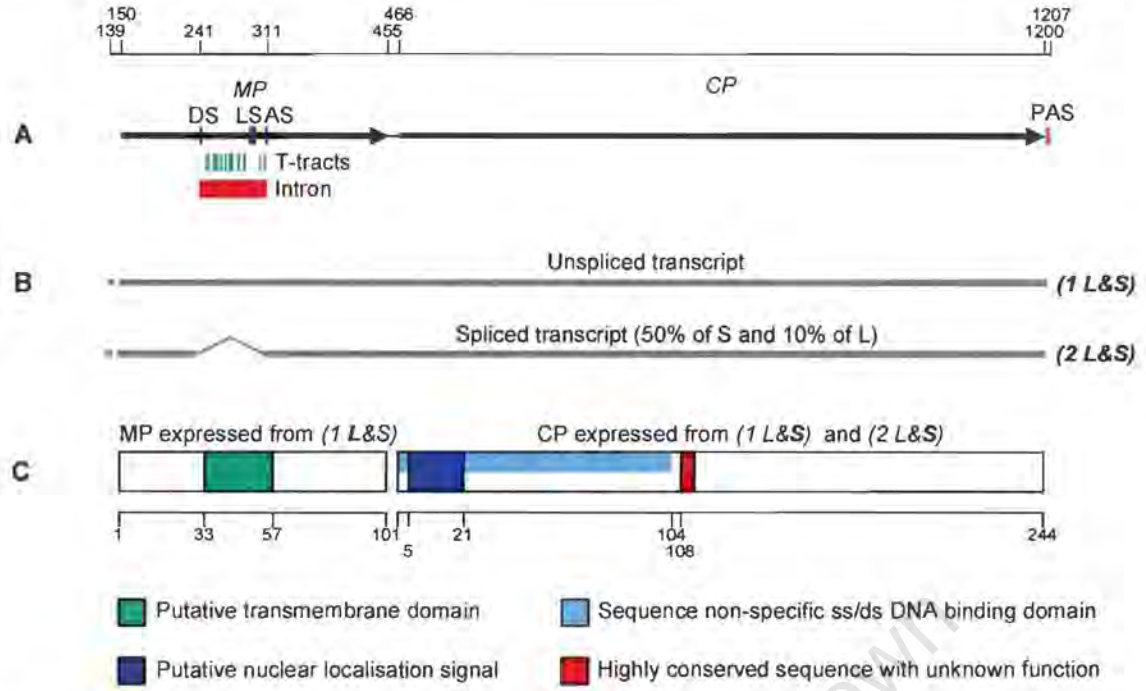


FIGURE 1.9 The MSV-Ns movement protein (MP) and coat protein (CP) genes, their expression and the functional domains of their products. (**A**) The *MP* intron is bounded by donor and acceptor sites (DS and AS respectively), and contains both a series of T-tracts and a consensus branch point or lariat sequence (LS). Shown in orange is the V-sense transcript polyadenylation signal (PAS). (**B**) Large (L) and small (S) V-sense transcript are produced with the small transcripts being more common. 10% of large and 50% of small transcripts are spliced. CP can be expressed from all transcripts but is most efficiently expressed from the spliced and unspliced small transcripts. MP can be expressed from large or small unspliced transcripts. (**C**) The known or suspected functional domains of MP and CP. The location of the *MP* intron corresponds directly with the suspected MP transmembrane domain. Numbering of DNA is relative to the (+) strand replication initiation site. References are within the text.

that CP could be expressed from all spliced and unspliced V-sense transcripts, whereas MP could only be expressed from the unspliced transcripts. While it is unclear how the presence of the intron might impact CP expression, it has been suggested that it may act as an enhancer element analogous to introns used in cereal transgene expression cassettes (Wright *et al.*, 1997; Palmer and Rybicki, 1998).

It is likely that post translational modification of the MSV MP occurs since its apparent size of 14kDa is larger than that predicted from its sequence (10.9kDa; Dickinson *et al.*, 1996). Mastrevirus MPs have a hydrophobic domain (Fig 1.8C) that may either facilitate their interaction with host cell membranes or be involved in homo- or hetero-oligomerisation with the CP (Boulton *et al.*, 1993; Kotlizky *et al.*, 2000). MPB, encoded by the begomovirus BC1 gene (Fig 1.2), appears to be functionally similar to the mastrevirus MP in that both localise to the cell wall in infected plant cells, both probably interact with plasmodesmata to facilitate virus cell to cell movement and both have been shown to move from cell to cell (Dickinson *et al.*, 1996; Kotlizky, *et al.*, 2000; Noueiry *et al.*, 1994; Pascal *et al.*, 1993). The MPB of the begomovirus,

Squash leaf curl virus (SqLCV), is apparently involved in the formation of endoplasmic reticulum-derived cytoplasmic tubules in infected cells (Ward *et al.*, 1997). However, no analogous structures have yet been observed in mastrevirus infected cells.

The CP obviously oligomerises during capsid formation but the CP-CP interaction domains enabling oligomerisation have yet to be determined for mastreviruses. The MSV CP binds both ds and ssDNA in a sequence non-specific manner (Liu *et al.*, 1997). Whereas the binding domain lies within the first 104 CP amino acids, the binding site probably falls between amino acids 5 – 22 (Fig 1.9C; Liu *et al.*, 1997). In accordance with the identification of potential nuclear localisation signals within the N-terminal regions of all mastrevirus CPs (Fig 1.9C; Liu *et al.*, 1997), it has been demonstrated that MSV CP expressed in insect cells, tobacco protoplasts and maize cells is imported into the nuclei of these cells (Kotlizky *et al.*, 2000; Liu *et al.*, 1999a). Nuclear targeting of a transiently expressed CP has also been demonstrated in maize cells (Kolizky *et al.*, 2000). The DNA binding activity and nuclear localisation of the MSV CP may imply that it is functionally equivalent to the bipartite begomovirus nuclear shuttle protein (NSP encoded by BV1). Begomovirus NSPs are believed to shuttle viral DNA across nuclear membranes (Lazarowitz and Beachy, 1999). Interestingly, there is phylogenetic evidence that the begomovirus NSP gene (BV1) may have originated from an ancient geminivirus CP gene (Fig 1.2; Rybicki, 1994).

1.4.2 Maize Streak Virus Infection of Maize

1.4.2.1 Initiation of infection following leafhopper delivery of virions

Very little information is available on the stages of MSV infection between leafhopper delivery of virus particles into permissive host plants and the onset of rolling circle replication in the nuclei of permissive cell types. Leafhoppers feed from both the mesophyll and phloem of healthy host plants (Mesfin *et al.*, 1995). Virus particles delivered into the phloem of leaves are apparently transported within sieve tubes to regions removed from the leafhopper feeding site (Peterschmitt *et al.*, 1992; Storey, 1938). The ultimate destination of these particles remains to be determined but a sufficient number move out of an inoculated leaf within a few hours to initiate a productive infection (Story, 1938).

A MSV infection is presumably initiated when at least one virus particle enters a nucleated plant cell such as a phloem companion or mesophyl cell. It is unknown whether a productive infection

can be initiated by entry of a particle into any nucleated cell or whether it requires that a particle enters a cell undergoing active division (Hanley-Bowdoin *et al.*, 1999). It is also unknown whether infection of these first cells is by individual particles (completely monoclonal) or by groups of particles (potentially polyclonal).

Once an individual virus particle is within a permissive cell it presumably uncoats either in the nucleus or in the cytoplasm. In the latter case, capsomeres or CP monomers derived from virion disassembly would probably bind to and transport the released DNA into the nucleus (Liu *et al.*, 1999a). DNA released from virions is single stranded except for an approximately 80 nucleotide region in the SIR which is believed to be vital for priming the initiation of (-) strand DNA synthesis (Donson *et al.*, 1984; Hayes *et al.*, 1988). The short complementary nucleic acid molecule bound to the otherwise ssDNA is primer-like in that it is predominantly DNA but also contains a few ribonucleotides on its 5' end (Donson *et al.*, 1984). While similar primer molecules have also been found in a number of other mastrevirus species (Andersen *et al.*, 1988; Donson *et al.*, 1987; Hayes *et al.*, 1988; Morris *et al.*, 1992), the mechanism of their synthesis remains to be determined (Gutierrez, 1999).

Synthesis of (-) strand to produce a dsDNA replicative form (RF) is believed to be initiated from the primer (Donson *et al.*, 1984; Hayes *et al.*, 1988). While this process is almost certainly exclusively host mediated, the enzymes and cofactors involved are unknown. It is assumed that no modification of host cell gene expression is required to convert the newly uncoated DNA into the dsDNA RF. It is therefore likely that in cells that are not actively dividing, constitutively expressed nick repair enzymes and cofactors are responsible for synthesising the (-) strand (Palmer and Rybicki, 1998). As mentioned previously, however, it is currently unknown whether an infection can be initiated by invasion of a cell that is not actively dividing (Hanley-Bowdoin *et al.*, 1999).

The newly synthesised relaxed covalently closed dsDNA, called RFII, is most likely converted into a supercoiled form containing approximately 16 superhelical turns through interaction with histones from the host cell (Gutierrez, 1999; Hamilton *et al.*, 1982). Histone associated geminivirus mini-chromosomes have been detected for the begomovirus, AbMV (Abouzid *et al.*, 1988; Pilartz and Jeske, 1992) and it is probable that they also occur in all other geminivirus infections (Gutierrez, 1999). The supercoiled histone associated form is called RFI and is both

the template from which the virus' genes are expressed and the starting point from which rolling circle replication can begin. (Gutierrez, 1999) .

1.4.2.2 Preparation of the cellular environment for replication

Besides an absolute dependence on Rep expression, geminivirus replication also relies on expression of the host cell's DNA replication machinery. The sequence of events within the nuclei of infected cells necessary for rolling circle replication (RCR) to occur are largely unknown, but Rep and probably RepA almost certainly play a central role in both the mechanics of the replication process and the preparation of the nuclear environment for replication to occur.

Under normal healthy conditions, the DNA replication machinery of a plant cell is only expressed immediately before mitosis during the S-phase of the cell cycle. Potential requirements for cell cycle specific host factors during replication has been noted for both MSV and DSV. Replication levels of both these viruses correlate with the division state of their host cells (Accotto *et al.*, 1993; Lucy *et al.*, 1996). However, MSV is not restricted to replicating in actively dividing cells (Lucy *et al.*, 1996) and a WDV based gene vector has been found to initiate replication independently of cell division (Timmermans *et al.*, 1992).

In newly infected non-dividing cells, Rep and/or RepA may induce the expression of host enzymes and cofactors required for rolling circle replication (RCR). Rep from the begomovirus, TGMV, is able to induce expression of proliferating cell nuclear antigen (PCNA) in differentiated cells (Nagar *et al.*, 1995). PCNA is the processivity factor of DNA polymerase δ and its expression does not normally occur in cells that are not actively dividing. While this indicates that a geminivirus Rep is capable of inducing at least part of a differentiated host cell's replication machinery, the exact mechanism by which this induction occurs is unclear.

Rep and/or RepA may either directly induce the promoters of certain host genes required for virus replication or they may interfere with cell cycle regulatory systems to indirectly induce the host genes required for virus replication (Gutierrez, 2000, Hanley-Bodoin *et al.*, 1999, Palmer and Rybicki, 1998). The C-terminal portion of all geminivirus Rep proteins share a substantial degree of sequence homology with the DNA binding domains of the *myb*-related class of plant transcription factors (Fig 1.8, Hofer *et al.*, 1992). Both the portion of Rep spanning this domain and an additional domain in the unique C terminus portion of RepA have transactivation activity in yeast (Horvath *et al.*, 1998). While it is conceivable that geminivirus Rep and mastrevirus

RepA proteins may function as transcription factors that directly modify the expression of host genes, there is currently no evidence that they have this capacity

There is, however, substantial evidence that geminivirus Rep and RepA proteins may interact with key components of host cell regulatory pathways to induce expression of host factors required for their replication. mastrevirus RepA proteins and begomovirus Rep proteins bind both human retinoblastoma (Rb) protein (Xie *et al.*, 1995; Collin *et al.*, 1996) and a maize retinoblastoma related (RbR) protein (Grafi *et al.*, 1996; Horvath *et al.*, 1998; Xie *et al.*, 1996). In mammals, the Rb family of proteins control cell cycling by sequestering transcription factors necessary for entry of cells into the S-phase (Hamel *et al.*, 1992; Lam and LaThangue, 1994). Mammalian viruses such as simian virus 40 (Ludlow, 1993), adenoviruses (Moran, 1993) and human papillomavirus (Voesden, 1993) encode gene products that bind Rb proteins via an Rb-binding motif (Nevins, 1992). This Rb-binding results in the release of transcription factors that are necessary for expression of the DNA replication machinery of host cells (Hamel *et al.*, 1992; Lam and La Thangue, 1994). The Rb-binding domains of these mammalian viruses closely resemble Rb-binding motifs found in most mastrevirus Rep and RepA proteins (Horvath *et al.*, 1998; Liu *et al.*, 1999c; Xie *et al.*, 1995).

Besides interacting with RbR proteins, WDV RepA also interacts via its ~170 C-terminus amino acids with a group of GRAB (geminivirus RepA binding) proteins (Xie *et al.*, 1999). While these GRAB proteins have unknown functions, they share substantial sequence identity to a family of plant proteins containing the so-called NAC (non-apical-meristem, ATAF and CUC2 genes) domain (Xie *et al.*, 1999). NAC domain proteins are involved in a variety of plant developmental pathways and it is conceivable that RepA interaction with GRAB proteins may be part of a second mechanism used by mastreviruses to modify cellular physiology to favour virus replication (Gutierrez, 2000).

As yet there is no direct *in vivo* experimental evidence that mastrevirus Rep proteins are capable of altering host regulatory pathways to specifically favour viral replication. No mastrevirus induced host cell proliferation has yet been observed so it is unlikely that these viruses manipulate cell cycle controls to achieve total progression from G1-phase through S-phase to mitosis. Accordingly, S-phase associated factors such as H2b are not always detectable in cells infected by MSV (Lucy *et al.*, 1996). It is possible that geminiviruses only induce partial

progression of the cell cycle into an S-phase like state in which only the DNA replication factors that these viruses require are expressed (Gutierrez, 2000).

1.4.2.3 Replication

Once RFI has been produced, sufficient quantities of Rep and RepA have been expressed, and the necessary host enzymes and cofactors are available, RCR of virus genomes can begin (Fig 1.10). During the coarse of replication in tomato golden mosaic virus (TGMV) infected *Nicotiana benthamiana* cells, it has been demonstrated that substantial alteration of nuclear architecture occurs. Whereas virus DNA accumulates within the centre of nuclei as variably sized inclusions, host chromosomal DNA is often condensed and excluded to the periphery of nuclei. Under these conditions, however, host genes can still be transcribed and there is also evidence that host DNA is replicated (Bass *et al.*, 2000; Hanley-Bowdoin *et al.*, 1999).

During replication Rep functions as a site and strand specific endonuclease (Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995a). Within the minimal origin of WDV replication (Sanz-Burgos and Gutierrez, 1998), Rep oligomers comprising 6 to 8 molecules (Sanz-Burgos and Gutierrez, 1998) bind with low affinity at the stem of the stem-loop sequence (O-complex) and with high affinity ~140 nucleotides upstream from the (+) strand replication initiation site (C-complex; Castellano *et al.*, 1999). It is possible that Rep binding at the stem sequences in RF DNA is required for extrusion of the stem-loop in a manner similar to that which occurs during DNA binding and replication initiation in dsDNA plasmids such as pT811 (Noirot *et al.*, 1990). While it is feasible that different Rep molecules within a single Rep oligomer bind the C and O sites to produce a higher order O:C complex at the origin (Gutierrez, 1999), an O-complex alone is capable of nicking the (+) strand between the last T and A residues of the invariant nonanucleotide (+) strand origin sequence (TAATATT↓AC; Castellano *et al.*, 1999; Laufs *et al.*, 1995a; Stanley, 1995). The initiation reaction involves nucleophilic attack on the phosphodiester bond between the T and A nucleotides by the OH-group of a conserved tyrosine residue within the RCR III motif (Fig 1.8) of a Rep molecule within the O/O:C complex (Laufs *et al.*, 1995b). The Rep molecule conducting the nicking reaction remains attached at the 5' end of the nicked strand, while the 3' OH-terminus is used as a primer for synthesis of a new plus strand (Laufs *et al.*, 1995b). The new strand is synthesised by host replication proteins and displaces the parental (+) strand from the intact (-) strand template. Begomovirus Rep proteins have a detectable ATPase

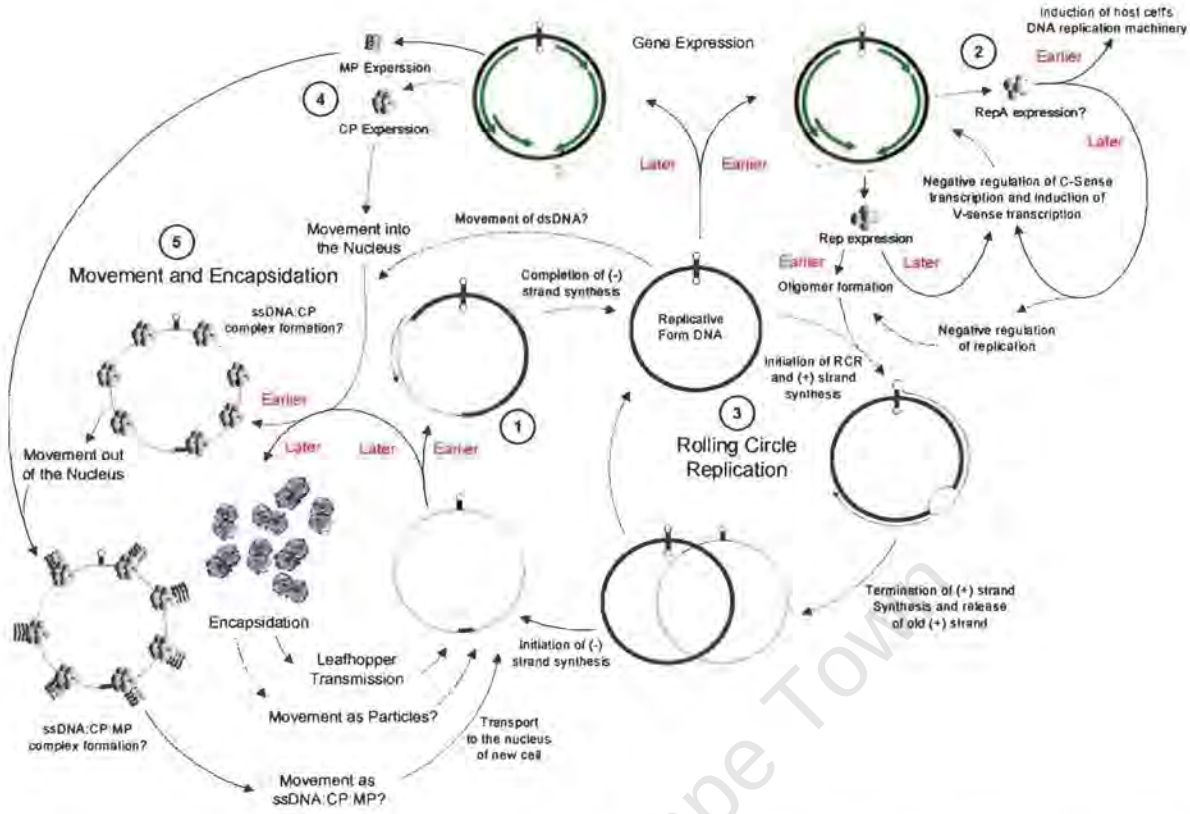


FIGURE 1.10 Summary of the MSV infection process. Early during an infection following the synthesis of a dsDNA replicative form (RF; 1) RepA is most likely expressed and induces a cellular state in which viral DNA replication can occur (2). Rep is also expressed early and rolling circle replication begins (3). At a later point in the infection process, following genome amplification and possibly Rep and/or RepA induction of the V-sense promoter, MP and CP are expressed (4) and movement and encapsidation occur (5). Represented here is movement of unencapsidated ssDNA but it should be noted that it is possible that dsDNA and/or encapsidated ssDNA may also be moved either cell to cell or systemically within the phloem of plants. Whereas the involvement of MSV CP and MP in movement has been demonstrated (Kotlizky *et al.*, 2000; Liu *et al.*, 1997, 1999a), the mechanics of the process are obscure and certain details have been borrowed from a cell to cell movement model proposed for the begomovirus, SqLCV (Lazarowitz and Beachy, 1999; Qin *et al.*, 1998). While the probable timing of events is indicated, it is unlikely, for example, that absolutely no MP and CP expression occurs during the earlier stages of the infection process. ssDNA is represented by blue lines, dsDNA by bold black lines and RNA by orange lines. References are within the text.

activity, indicating that these and other geminivirus Rep proteins may also have a helicase activity that is operational during this phase of (+) strand replication (Breret *et al.*, 1999; Desbiez *et al.*, 1995).

By analogy with other rolling circle replicons, it is believed that release of the old (+) strand probably occurs by one of two mechanisms (Palmer and Rybicki, 1998). The first potential mechanism is continuous. Following completion of the new (+) strand the newly synthesised nonanucleotide at the origin of replication is nicked by a second Rep molecule - possibly another Rep molecule within the same bound oligomer as the Rep molecule attached to the 5' end of the

old strand. This Rep molecule acts as a terminase by becoming covalently linked with the 5' AC end of the newly nicked nonanucleotide and the phosphoryl group of the displaced strand is transferred from the initial Rep's active site tyrosine to the newly generated 3'OH to simultaneously recircularise and release the old (+) strand. The second potential mechanism is discontinuous. Closing and release of the old (+) strand is mediated by a non-tyrosine residue in the active site of the same Rep protein and re-initiation of the next cycle of replication is mediated by a different molecule (Palmer and Rybicki 1998).

Once released, it is assumed that a site within the SIR of the old (+) strand must become the substrate of a priming event which will enable synthesis of a new (-) strand. As mentioned previously, the origin of the primer molecule associated with encapsidated mastrevirus DNA is unknown but the presence of a few ribonucleotide monophosphates on its 5' end indicates that its synthesis is probably initiated within the SIR by a host DNA primase (Gutierrez, 1999). Its length of ~80 nucleotides is a mystery and is possibly due to initial extension of the RNA primer by a DNA polymerase complex with limited processivity (Gutierrez, 1999). During active replication, it is also feasible that there is no break in primer DNA extension around nucleotide 80 and that the small RNA/DNA primer observed in encapsidated DNA is the result of primer extension being arrested by encapsidation (Donson *et al.*, 1984). No primer molecule associated with (-) strand synthesis has yet been detected in any of the other geminivirus genera and it would follow that for viruses in these genera to successfully initiate an infection they must be delivered into cell types actively expressing a DNA primase. It is possible that the presence of an encapsidated primer molecule in mastreviruses is an evolutionary adaptation enabling these viruses to initiate new infections in terminally differentiated cell types that do not express DNA primases.

During RCR, a newly released (+) strand has three possible fates which are largely determined by the stage during the replication process when it is released. Early on during replication it is highly likely that it will be converted into another RF DNA in order to maintain an exponential increase in viral DNA accumulation. Later on, however, an increasing number of released (+) strands will be removed from the replication pool either by encapsidation or movement to surrounding cells.

1.4.2.4 Regulation of the infection process

It is unknown whether discreet switches between replication, movement, and encapsidation occur within a cell at certain stages during the infection process. To operate successfully, such switches require mechanisms whereby the virus is able to "perceive" the appropriate stages during an infection when different activities should occur. Conceivably, these switches could operate at the levels of viral gene transcription, post-transcriptional mRNA processing, and post-translational modification of gene products. While advances have been made in identifying the geminiviral gene products and genes that may interact to control virus infections within host cells, very little is known about how these interactions converge to control the infection process.

The multiple activities that have been determined for geminivirus Rep proteins and mastrevirus RepA proteins (Table 1.2) indicate that they are the viral gene products that completely control and direct viral replication. Tight self regulation of Rep and/or RepA expression could be the central mechanism by which geminiviruses control their life cycles within host cells. Through its involvement in replication, Rep is directly responsible for multiplying the copy number of its own gene. Once replication commences the number of Rep genes available for expression will, for a time at least, increase exponentially. Despite gene expansion, *Rep* transcripts are relatively rare in infected cells (Wright *et al.*, 1997), indicating that Rep expression is probably tightly controlled. It has been demonstrated that mastrevirus, begomovirus, and curtoviruses Rep proteins (Akbar Behjatnia *et al.*, 1998; Fontes *et al.*, 1992, 1994b; Sanz-Burgos and Gutierrez, 1998) and mastrevirus RepA proteins (Missich *et al.*, 2000) bind near TATA boxes that direct *Rep* transcription. Because it has been demonstrated for begomoviruses that Rep is capable of repressing its own promoter (Eagle *et al.*, 1994; Sunter *et al.* 1993), it has been proposed that Rep binding may interfere with the initiation of *Rep* transcription (Arguello-Astorga *et al.*, 1994a, b). While this would provide an obvious negative feedback mechanism whereby constant Rep expression levels could be retained in the face of massive Rep gene expansion, in mastreviruses there is no evidence that Rep is able to auto-regulate its own expression. Although there is evidence that RepA expression down-regulates replication (Collin *et al.*, 1996), the mechanism by which this is achieved remains unknown.

Because Rep and RepA have different activities in mastreviruses it has been suggested that mechanisms altering the relative proportions in which they are expressed may control the progression of the infection cycle from replication initiation through to movement and encapsidation (Wright *et al.*, 1997; Palmer and Rybicki, 1998). In mastreviruses two complementary sense transcripts are produced. One transcript terminates within the C2 ORF and

the other terminates within the SIR (Fig 1.8B; Dekker *et al.*, 1991; Mullineaux *et al.*, 1990). Whereas only RepA could be expressed from the smaller transcript, both Rep and RepA could be expressed from the larger one. Because the larger transcript must first be spliced to enable Rep expression, one obvious way of modulating the relative amounts of Rep and RepA that are produced may be through control of intron splicing (Wright *et al.*, 1997). There is, however, anecdotal evidence that the mechanism controlling the relative amounts of Rep and RepA expressed may operate at the level of transcription. Mutational analysis of the three TATA boxes 101, 62, and 57 nucleotides upstream from the *RepA* start codon in MSV-Ns indicated that transcripts initiated from the two TATA boxes closest to the start codon were predominantly of the shorter type (personal communication with M. I. Boulton, 1997, cited by Palmer and Rybicki, 1998). In MSV, the exact positions of Rep and RepA oligomer binding within the LIR have yet to be determined but it is possible that Rep and or RepA binding in the proximity of the -101 TATA box may force the initiation of C-sense transcription from -57 or -62 TATA boxes. In so doing RepA and/or Rep C-complex formation may enhance the expression of RepA while at the same time inhibiting the expression of Rep.

One might expect that during replication, MP and CP expression levels will also increase as a result of gene amplification. While this probably does occur (Palmer and Rybicki, 1998), there is evidence that the expression of MP and CP may also be enhanced at an appropriate point during the infection process by the interaction of Rep and/or RepA with the virion sense (V-sense) promoter. In the mastreviruses, WDV and CSMV, it has been observed that Rep may mediate transactivation of the V-sense promoter (Hofer *et al.*, 1992; Zhan *et al.*, 1993). It has also been shown in WDV constructs unable to express RepA, that the activity of the V-sense promoter is far lower than in constructs retaining the potential to express RepA (Collin *et al.*, 1996). Both Rep and RepA oligomer complexes of WDV have been found to assemble immediately upstream of the TATA box that is believed to initiate *MP* and *CP* transcription (Fig 1.6; Castellano *et al.*, 1999; Missich *et al.*, 2000). It is possible that transactivation domains in Rep and/or RepA may play a role in enhancing V-sense transcription following complex formation (Horvath *et al.*, 1998; Castellano *et al*, 1999; Missich *et al.*, 2000). It is also possible that the mechanism by which the Rep and/or RepA binding complexes may regulate V-sense gene transcription is related to their covering 4 of 6 A-T tracts upstream of the *MP* start (Fig 1.6). The tracts, believed to confer static curvature of the DNA in this region, are possibly involved in transcriptional regulation. By covering most of the tracts and influencing the curvature of this region it is

feasible that Rep and/or RepA V-complexes may enhance expression of the V-sense genes (Castellano *et al.*, 1999).

The relative proportions of MP and CP expressed may also be vital in determining the specific times at which movement and encapsidation occur during the infection process. Whereas CP can be expressed from both spliced and unspliced versions of a long and short V-sense transcript, MP can only be expressed from the two unspliced transcripts (Fig 1.9; Wright *et al.*, 1997). If Rep and RepA V-complexes form on the LIR at similar positions to those that occur in WDV (Castellano *et al.*, 1999; Missich *et al.*, 2000), they may potentially alter the ratio of long and short V-sense transcripts produced by preventing formation of the larger transcript. *In vitro* expression assays have indicated that whereas MP is apparently expressed with similar efficiencies from both long and short unspliced transcripts, CP is expressed at higher levels from the spliced and unspliced short transcripts (Wright *et al.*, 1997). Additional regulation of relative MP-CP expression levels might be achieved through differential transcript splicing (Wright *et al.*, 1997). Apart from decreasing the amount of transcript available for MP expression (50% of the smaller transcript and 10% of the larger transcript is spliced), it is also possible that splicing of the *MP* intron may greatly enhance CP expression *in vivo* (Palmer and Rybicki, 1998; Wright *et al.*, 1997).

Future models describing the dynamics of mastrevirus replication from the initiation of infection through to late stages of the infection process will need to reconcile all of the regulatory activities of RepA and Rep. Of particular importance is information regarding the expression of RepA at different stages during infection. While RepA is possibly involved in preparing the cellular environment for replication early during the infection process, it apparently also directly influences viral replication levels (Collin *et al.*, 1996) and is potentially active in the transactivation of CP and MP expression during the later stages of infection (Collin *et al.*, 1996; Horvath *et al.*, 1998; Zhan *et al.*, 1993). If all these RepA activities are confirmed *in vivo*, it will imply that the mechanism that regulates RepA expression (transcriptional, post-transcriptional intron splicing) or activity (post-tratanslational modification) is likely to be the "master control" of the entire infection process.

1.4.2.5 Encapsidation and movement from cell to cell

Apart from being the primary location of replication, the nucleus is also the site of virus particle assembly. It is probable that later in the replication process, when virion sense gene expression is

underway, increasing nuclear concentrations of CP result in the accumulation of ssDNA (Boulton *et al.*, 1989a; Lazarowitz *et al.*, 1989). CP molecules in the nucleus possibly bind (+) strands released during RCR and arrest the synthesis of new RF DNAs (Donson *et al.*, 1984). There is apparently no encapsidation signal within geminivirus genomes and it is possible that any circular ssDNA of an appropriate size will be encapsidated (Mansoor *et al.*, 1999; Saunders and Stanley, 1999). MSV (+) strands, each with an attached ~80 nucleotide primer molecule (Donson *et al.*, 1984), are packaged into particles that congregate to form large paracrystalline nuclear inclusions (Pinner *et al.*, 1993). Crystalline arrays of MSV particles have also been detected outside nuclei within physiologically active phloem companion cells, and inside the vacuoles of dead and dying cells within chlorotic lesions (Pinner *et al.*, 1993).

At some point in the infection process after the onset of replication and probably before the completion of encapsidation, viral DNA must be moved into neighbouring cells. The mechanistic details of this process are still obscure but in MSV it appears to involve both the CP and MP (Boulton *et al.*, 1989a; Kotlizky, *et al.*, 2000; Lazarowitz *et al.*, 1989).

The movement functions of the CP may involve both its stimulation of ssDNA accumulation and its activity as a nuclear localised ss and dsDNA binding protein. While it has been observed that MSV *CP* mutants fail to accumulate ssDNA (Boulton *et al.*, 1989a; Lazarowitz *et al.*, 1989), it is unknown whether this is due to lack of ssDNA sequestration within capsids or whether the mutants lack a specific genetic switch that shifts the infection process from replication to ssDNA accumulation (Palmer and Rybicki, 1998). Lack of encapsidation would almost certainly account for a decrease in the amount of ssDNA in infected nuclei. However, curtovirus and begomovirus encoded proteins other than the CP are also involved in ssDNA accumulation (Jeffrey *et al.*, 1996; Padidam *et al.*, 1996; Rigden *et al.*, 1993) which suggests that, outside the context of encapsidation, the accumulation of ssDNA may be an important aspect of the geminiviral infection process (Ingham *et al.*, 1995; Jeffery *et al.*, 1996; Padidam *et al.*, 1999b). Although cell to cell movement of geminivirus genomes as both ssDNA and dsDNA apparently occurs in certain species (Noueiry *et al.*, 1994; Rojas *et al.*, 1998), most of the evidence currently available supports the view that ssDNA is the most common form in which geminiviruses move their genomes (Fig 1.10; Padidam *et al.*, 1999b; Pascal *et al.*, 1994; Rojas *et al.*, 1998; Sanderfoot *et al.*, 1996). While it is possible that the MSV CP initiates the onset of cell to cell movement by making ssDNA available for transport, a causal link between CP mediated ssDNA accumulation

and virus movement has only been established for the begomoviruses (Padidam *et al.*, 1999b; Qin *et al.*, 1998).

The second proposed movement function of the MSV CP is that it is directly involved in the transport of MSV DNA between cells. Based on observations that the CP binds both ds and ssDNA (Liu *et al.*, 1997) and is capable of transporting this DNA into nuclei (Liu *et al.*, 1999a), it has been suggested that it may be the functional equivalent of the bipartite begomovirus nuclear shuttle protein (NSP; encoded by BV1; Fig 1.2). NSP, which is also nuclear localised and binds both ss and dsDNA, is believed to mediating the movement of ssDNA into and out of nuclei (Rojas *et al.*, 1998; Sanderfoot and Lazarowitz, 1995; Sanderfoot *et al.*, 1996; Ward and Lazarowitz, 1999).

Similar parallels have been drawn between the movement functions of the mastrevirus MP and the movement functions of the bipartite begomovirus MPB (encoded by BC1; Fig 1.2). Both types of MPs localise to the cell wall in infected cells, interact with plasmodesmata, and have the ability to move independently from cell to cell (Dickinson *et al.*, 1996; Kotlizky *et al.*, 2000; Noueiry *et al.*, 1994; Pascal *et al.*, 1993). Interaction between MP and CP of mastreviruses (Kotlizky, *et al.*, 2000) is also paralleled by interactions between the MPB and NSP of biparpartite begomoviruses (Schaffer *et al.*, 1995; Sanderfoot and Lazarowitz, 1995; Sanderfoot *et al.*, 1996). While the MPB of SqLCV has been shown to bind NSP and move it across cell walls (Sanderfoot and Lazarowitz, 1995; Sanderfoot *et al.*, 1996), MSV's MP is apparently also able to interact with and redirect CP from the nucleus to the cell periphery (Kotlizky, *et al.*, 2000).

There are currently two proposed models of bipartite begomovirus movement. The first model is based on studies of the movement of the phloem limited virus, SqLCV. During the early stages of replication, CP in the nucleus binds (+) strand DNA and removes it from the replication pool. CP binding at this stage does not result in encapsidation because CP concentrations early during replication are too low for capsid formation (Qin *et al.*, 1998). Viral NSP moving between the cytoplasm and nucleus binds accumulated viral ssDNA in the nucleus and displaces the CP molecules (Qin *et al.*, 1998). NSP molecules complexed with viral ssDNA (ssDNA:NSP) move between the nucleus and the cytoplasm (Sanderfoot *et al.*, 1996; Ward and Lazarowitz, 1999) and, while in the cytoplasm, are bound by MPB. MBP transports the ssDNA:NSP complex through the cytoplasm, across the cell wall and releases it into a neighbouring cell (Noueiry *et al.*,

1994; Sanderfoot and Lazarowitz, 1995; Lazarowitz and Beachy, 1999). As infection in the first cell progresses, however, the concentration of CP in the nucleus eventually becomes sufficient for encapsidation to occur. Once encapsidation takes place ssDNA is no longer available for NSP binding and movement to surrounding cells ceases. By the time that this happens, however, transmission of virus genomes into every adjacent cell has already occurred (Qin *et al.*, 1998).

A second, less developed model has been proposed to describe the movement of BDMV – a begomovirus which is not phloem limited. The major differences between this and the first model is that it describes the movement of both ss and dsDNA during which MPB interacts directly with the vial DNA. In this model, NSP recognises and binds both ss and open circular dsDNA in a size and form dependent manner (Rojas *et al.*, 1998). Because mutants displaying dramatic decreases in ssDNA accumulation suffer little or no fitness loss (Azzam *et al.*, 1994; Ingham *et al.*, 1995), it is proposed that, in BDMV at least, dsDNA is the predominant form in which genomes are moved (Noueiry *et al.*, 1994; Rojas *et al.*, 1998). Following movement into the cytoplasm, the viral ds/ssDNA is transferred from NSP to MPB and the DNA:MPB complex moves through the cytoplasm and across the cell wall into a neighbouring cell (Rojas *et al.*, 1998).

While it is feasible that similar mechanisms of cell to cell movement may occur in mastreviruses, important differences between the movement of these viruses and begomoviruses are likely to exist. MSV, like BDMV, is not phloem limited and the ds and ssDNA binding activity of the MSV CP superficially resembles that of BDMV's NSP. It is, however, very unlikely that the BDMV model of movement will be applicable to MSV movement because, unlike the BDMV MPB, the MSV MP does not bind DNA (Liu *et al.*, 1997). Another important difference between mastreviral and begomoviral movement, is that the mastrevirus CP will need to simultaneously perform the movement functions of the begomovirus NSP and CP. While the MSV CP binds MSV DNA (Liu *et al.*, 1997) and is able to mediate its transport into the nucleus (Liu *et al.*, 1999a), its ability to act as a nuclear shuttle protein has not yet been proven.

Besides requiring the co-ordinated interactions of viral gene products and DNA, the successful movement of MSV genomes from infected to uninfected cells is strongly dependent on the extent of plasmodesmatal connections between neighbouring cells (Lucy *et al.*, 1996). Also, successful delivery of viral DNA into uninfected cells does not guarantee infection will occur. In maize, it appears as though certain cell types are more sensitive to MSV infection than others (Lucy *et al.*,

1996; Pinner *et al.*, 1993). For example, in maize leaves the virus infects all photosynthetic cell types (e.g. mesophyll and bundle sheath cells) but despite abundant plasmodesmatal connections between photosynthetic, epidermal and parenchyma cells, MSV is only rarely detectable in the latter two cell types (Lucy *et al.*, 1996).

It is unknown whether systemic movement of geminiviruses within plants simply relies on normal cell to cell movement to deliver genomic DNA into the phloem, or whether viral DNA is specifically packaged for long distance transport. It is possible that cell to cell movement might involve unencapsidated ss or dsDNA but that long distance movement in the phloem might require encapsidation (Palmer and Rybicki, 1998). However, leafhopper feeding experiments have determined that no virus acquisition occurs unless leafhoppers feed directly on symptomatic tissues (Peterschmitt *et al.*, 1992; Storey, 1928). This is interesting because MSV CP is detectable at low levels in the asymptomatic leaf tissues of infected plants (Peterschmitt *et al.*, 1992) indicating that movement of MSV throughout the plant is possibly in a form that is not transmissible. While it may be that leafhopper transmission requires that MSV particles form crystalline complexes to survive the ingestion process, it is also possible that the CP detected in asymptotic tissues is in the form of unencapsidated CP:DNA complexes in the process of movement.

In MSV-infected maize, the correlation between patterns of virus accumulation and translocation pathways between tissues is strong evidence that long distance movement occurs via phloem elements (Lucy *et al.*, 1996). It is believed that MSV is incapable of invading the root apical, shoot apical and reproductive meristems due to the absence of developed vasculature in these tissues (Lucy *et al.*, 1996). It is also noteworthy that in all maize organs except the leaves, MSV infection is almost completely restricted to the vasculature (Lucy *et al.*, 1996; Pinner *et al.*, 1993). There is a degree of correlation between MSV replication and the expression of the mitosis-specific histone gene, H2b, in the developing vasculature and it is therefore possible that infection of host cells in all organs except the leaves requires active cell division (Lucy *et al.*, 1996).

Within the shoot apex, MSV first enters developing leaves at approximately plastochron five. While the virus is restricted to the developing leaf vasculature before plastochron 12, it is likely that the development of metaphloem elements at approximately plastochron 12 provides an opportunity for MSV to escape the vasculature into the photosynthetic cells of the leaf. Metaphloem develops with the abundant plasmodesmatal connections required for efficient

loading of photoassimilates once the leaf emerges from the whorl. Before emergence, however, the developing photosynthetic tissues are still net importers of photoassimilates and the virus most likely moves into these cells through their plasmodesmatal connections with the metaphloem (Lucy *et al.*, 1996).

While MSV replication apparently ceases within mature leaf sheaths, it continues indefinitely within the infected cells of leaf laminae (Lucy *et al.*, 1996). Because there is obviously an upper limit to the amount of encapsidated genomes that a single nucleus can contain, indefinite replication of MSV in maize leaves implies that genomes must either be constantly exported or degraded within these cells. Given that MSV in the leaf preferentially infects photosynthetic cells, it is very interesting that replication in these cells neither ceases nor correlates with H2b expression (Lucy *et al.*, 1996). The features of photosynthetic leaf cells that make them more sensitive to MSV infection than surrounding epidermal and parenchyma cells, and enable stable long term MSV replication in the absence of host cell division certainly warrant further study.

On the leaves, the pattern of chlorotic streak-like lesions that characterises MSV infections is directly correlated with the pattern of virus accumulation within the leaves (Lucy *et al.*, 1996) and the virus can only be acquired by leafhoppers from these lesions (Peterschmitt *et al.*, 1992; Storey, 1928). The degree of chlorosis that occurs within lesions is MSV strain-dependent and is related to different degrees of infection-induced chloroplast malformation (Pinner *et al.*, 1993). It is interesting that while leafhoppers generally feed from the phloem sieve tubes of hosts such as *Digitaria sanguinalis*, . they mainly feed on the mesophyll cells of maize plants (Mesfin *et al.*, 1995). In an infected maize leaf, virus particles are undetectable in phloem sieve tubes whereas they occur at high concentrations within mesophyll cells (Lucy *et al.*, 1996) and it is possible that colonisation by MSV of photosynthetic cells in host species such as maize is an evolutionary adaptation to leafhopper feeding behaviour.

1.5 PROJECT AIMS

If the work that is about to be described in this thesis could be said to have had a single major objective it was the attempted marriage of traditional plant pathological and molecular biological approaches to improve our understanding of why MSV, a remarkably simple organism, is the most significant pathogen of Africa's most important food crop. Because very little is currently known about the distribution and geographical localisation of economically significant MSV

genotypes, an obvious starting point for this project was an investigation of the molecular diversity and pathogenicity in maize of MSV isolates from different parts of Africa. At the beginning of the study, no fully quantitative methods for rapidly assessing the pathogenicity of MSV isolates were available and the first goal of the project was therefore to develop a rapid, accurate and precise means of quantifying MSV symptoms. Towards this end an image analysis system making use of specifically written analysis software was devised to enable the automated and objective quantification of chlorotic areas on the leaves of MSV infected maize. The system was then rigorously tested for its ability to accurately and precisely assess symptoms in the hands of different users.

Having developed an accurate and precise means of determining chlorotic areas on the leaves of MSV infected maize plants, the next goal of the project was to devise a means of evaluating the virulence of cloned MSV isolates. Apart from being useful for the assessment of MSV virulence, it was realised that such a technique would also be of great potential value to maize breeders involved in producing MSV resistant genotypes. During this part of the study it was discovered that techniques used to construct agroinfectious MSV clones could influence both the infectivity of clones and the initial severity of the infections they produced. For testing the virulence of cloned MSV isolates it was therefore necessary to identify measures of symptom severity that were not influenced by the manner in which agroinfectious clones were constructed.

The MSV diversity aspect of this project was carried out as part of a long-term ongoing study investigating the diversity of MSV isolates infecting maize and other grass species from different parts of Africa. For descriptive purposes, a classification system for MSV isolates was devised and the prevalence of isolate groupings in maize and non-maize hosts in different regions of Africa was used as a basis for the selection of 12 MSV isolates for cloning and full genomic sequencing. Together with 10 previously described African streak virus isolates, the relative virulence of 22 virus isolates in maize was determined using the MSV symptom evaluation technique developed earlier in the study. While the main aim of this part of the project was identification of groups of MSV isolates that are most virulent in maize, it also culminated in the discovery of new MSV strains and yielded evidence of extensive recombination both amongst MSV isolates, and between MSV isolates and other African streak virus species. Because existing recombination analysis software was very poorly suited to the detection of both intra-strain and inter species recombination, custom software was written to achieve this goal.

Continuing with the recombination and MSV pathogenicity themes, the final aim of this project was to identify the genomic features of extremely virulent MSV isolates that were responsible for their elevated fitness in maize. To achieve this, a series of chimaeric viruses were constructed that contained genomic components of three less pathogenic virus isolates mixed with those of a highly virulent isolate. Apart from identifying pathogenicity determinants by analysing the virulence of the chimeras, a secondary aim of this part of the project was to determine the degree to which the different components of a MSV genome rely on co-evolved sequences in other genomic components for optimal functionality.

Chapter 2

Microcomputer Based Quantification of *Maize Streak Virus* Symptoms in *Zea mays*

Contents

ABSTRACT

The use of computer assisted image analysis techniques for the objective quantification of maize streak virus (MSV) symptoms in *Zea mays* was investigated. Comparisons were made between independent duplicate evaluations of chlorotic lesion areas occurring on MSV infected leaves using visual assessment, a commercial image analysis system, and a custom image analysis system employing software developed specifically for this study. Relative to visual assessments of disease severity, computer assisted image analysis employing both the commercial and custom systems provided significant enhancements in the accuracy and precision of chlorotic area estimations. The commercial image analysis system afforded no significant improvement in precision or accuracy over the custom system. An important advantage of examining images using the custom-written software was that the software permitted a high degree of analysis automation. Digitised images of maize leaves could be automatically analysed by the custom software five times faster than and with the same precision and accuracy as when the same images were analysed with the commercial software. Because of the flexibility of the image analysis techniques described they should be applicable to the measurement of symptom severity in other plant host-pathogen combinations.

2.1 INTRODUCTION

Quantification of disease severity is central to the evaluation of gross host-pathogen interactions in plant pathology and breeding. Methods used in the quantification of disease severity should ideally be simple, cheap, fast, flexible, precise, and accurate. Currently employed methods of symptom assessment for many plant diseases are either qualitative or only semi-quantitative and often rely heavily on visual assessment by a single scorer. While many of these techniques have been successfully employed in the assessment of disease control practices and the identification of disease resistant plant varieties, various studies have demonstrated the unreliability of the human eye for the objective determination of symptoms (Horsfall and Barratt, 1945; Lindow and Webb, 1983; Parker *et al.*, 1995; Sherwood *et al.*, 1983). These studies have found that data based on visual assessment of disease symptoms are inaccurate when percentages of leaf injury are close to 50% (Horsefall and Barratt, 1945) and contain biases correlated with the numbers (Sherwood et al., 1983), areas (Parker *et al.*, 1995, Sherwood *et al.*, 1983), and shapes of disease lesions (Lindow and Webb, 1983; Webb and Lindow, 1981).

Almost since their inception, personal computers have been employed in the analysis of plant disease severity (Blanchette, 1982; Lindow and Webb, 1983). The response of a video camera to incident light is unaffected by amounts of disease injury, and objective computational quantification of diseased areas is not prone to biases related to lesion number and shape (Lindow and Webb, 1983). Many commercial image analysis and processing systems, often backed by powerful software, are currently available and are being extensively used in plant disease assessment (Kokko *et al.*, 1993; Newton, 1989; Price *et al.*, 1993; Shaw and Royle, 1989). An important factor in the use of these dedicated systems is, however, their high cost. Depending on the complexity of the disease assessment task at hand, cheap alternative protocols may be devised using the personal computers and digital image capture equipment commonly found in many laboratories and homes (Kampmann and Hansen, 1993; Lindow and Webb, 1983; Sah and Fehrmann, 1992). Useful in this regard is the availability, via the Internet, of a wealth of relatively sophisticated image processing and analysis software that is in most cases inexpensive and often free.

For purposes of assessing the MSV resistance of individual maize plants, breeders have developed and currently use a semi-quantitative five- to twelve-point symptom severity rating systems (Barrow, 1992; Efron *et al.*, 1989; Govinden and Rummun, 1996; Rodier *et al.*, 1995; Van Rensburg *et al.*, 1991). However, because the MSV pathology studies that were to be carried out during this project would require more precise evaluation of symptoms in differentially resistant maize genotypes infected with differentially virulent MSV strains, it was necessary to develop a more accurate quantitative technique of symptom determination. It was realised that maize streak disease (MSD) symptoms are particularly suited to quantification by microcomputer assisted image analysis because MSV produces well defined chlorotic lesions on the leaves of afflicted host plants that directly correspond with the pattern of virus accumulation within the leaves (Efron *et al.*, 1989; Lucy *et al.*, 1996).

In this chapter the suitability of computer based image analysis techniques for the objective quantification of maize streak symptoms in maize was investigated. Approximation of percentage chlorotic areas occurring on MSV-infected maize leaves were made both visually and with image analysis techniques. The latter techniques involved the use of both a commercial image analysis system, and a custom developed system using software written specifically for this study. The

symptom assessment techniques investigated were compared for accuracy, precision and requisite analysis time.

2.2 MATERIALS AND METHODS

2.2.1 Virus Strains

Agroinfectious constructs of differentially virulent MSV strains were either produced during this study (MSV-MatA – see Chapter 4), or obtained from Dr W. H. Schnippenkoetter (University of Cape Town, Cape Town, South Africa; MSV-Kom and MSV-Set), Dr J. A. Willment (University of Cape Town, Cape Town, South Africa; MSV-Tas, MSV-VW, and MSV-VM) and Dr M. Boulton (John Innes Centre, Norwich, U.K. ; MSV-N).

2.2.2 Specimen Preparation

Inoculation of 3 d old seedlings of the MSV-sensitive *Zea mays* cultivar, Jubilee, was carried out by agroinfection (See Chapter 3 for details). Fifteen days after inoculation, blades of the third leaves of six uninfected control plants and 110 plants displaying symptoms ranging from almost complete chlorosis to small isolated chlorotic leaf spots, were cut transversely into four segments of equal length. Colour photographs were taken of the second segment from the base of each leaf blade. Once photographed, leaf segments were wrapped in damp paper towels and stored on ice until use.

2.2.3 Visual Analysis of Symptoms

Visual assessment of symptoms was carried out on all the leaf segments involved in the study. Percentage chlorotic area estimates were made from projections of the colour slide photographs taken during preparation of leaf segments. Four scorers, three of whom had extensive experience in visually assessing maize streak symptoms, independently estimated percentage chlorotic areas of the leaf segments under identical conditions. Each scorer repeated their visual assessment of chlorotic areas twice on consecutive weeks.

2.2.4 Image Capture Systems

Image analysis in this study utilised both a Joyce Loebl IV120 image-analysis system (JL Automation Ltd., Sunderland, U.K.) and GDS 5000 image capture equipment (Ultra-Violet Products, Upland, California). The commercial Joyce Loebl system comprised a Hitachi KP-140 monochrome video camera (Hitachi Denshi LTD, Tokyo, Japan) connected to a Joyce Loebl Control Processor. Near uniform top lighting was supplied by two fluorescent lamps. The GDS 5000 image capture equipment used in the custom image analysis system comprised an Ultra-Violet Products monochrome video camera connected to an Image Store 5000 processor box. Two opaqued 100W light bulbs arranged to provide near uniform top lighting were used for specimen illumination.

2.2.5 Image Analysis Protocols

Between 15 and 18 leaf segments, including the six symptomless control leaves, were grouped in the same orientation without touching one another on black felt within a 15 X 25 cm area. Images of leaf segments containing 256 grey shades with dimensions of 512 X 512 pixels and 756 X 512 pixels were captured with the Joyce Loebl and GDS 5000 systems respectively. Also included in every image was an internal colour control comprising a contact print of a photographic graded grey scale or "step-wedge" (Fig. 2.1) (Stouffer Graphic Arts, Inc., South Bend, Indiana). Optimum video camera aperture settings permitting the widest possible spread of pixel shades in captured images were established based on the examination of pixel shade frequency histograms from images containing a series of step-wedges and a test group of leaf segments (with MSV symptoms ranging from mild to severe). These images were captured using a variety of aperture settings with constant top lighting. Establishment of an optimum lighting intensity standard exploited the phenomenon of image flooding - which results in uniform whitening of bright areas of an image that are excessively illuminated - of a selected step-wedge that was then employed as a standard lighting intensity indicator in all future analyses with both of the image analysis systems used.

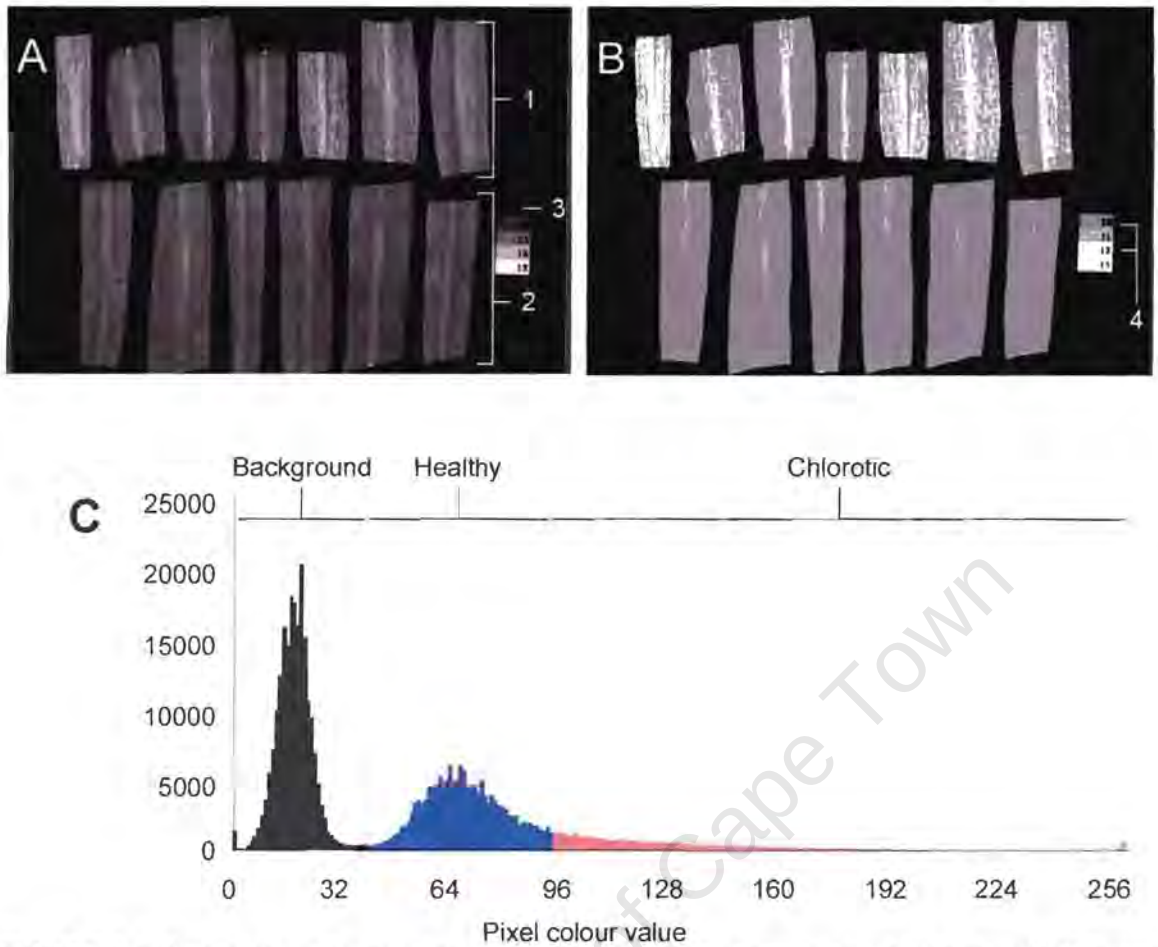


FIGURE 2.1 Computer assisted quantification of MSV symptoms on maize leaves. (A) Captured image containing: (1) symptomatic leaf segments, (2) asymptomatic control leaf sections, and (3) a step-wedge internal colour control. (B) Thresholding of image pixels into background (black), healthy leaf area (grey) and chlorotic leaf area (white) categories based on the gray shades of (4) specific step-wedge segments. (C) Histogram describing the distribution of pixel colour frequencies in image (A). While a pixel colour value of 1 represents black, a pixel colour value of 256 represents white. Differently coloured sections of the histogram indicates pixels grouped into the different threshold categories in image (B). Black represents background, blue represents healthy leaf area and pink represents chlorotic leaf area.

A pixel thresholding strategy, based on that described by Lindow and Webb (1983), was used for analysis of images captured with both the Joyce Loebl and GDS 5000 systems (Fig 2.1). Images obtained with the Joyce Loebl system were analysed using JLGenias (version 3.01), a commercial software package designed specifically for use with this image analysis system. Analysis was partially automated with the use of a custom "macro" based analysis sequence. With the custom system, captured images were analysed using Image09, an image analysis program written specifically to automatically set pixel threshold values based on the shade of designated step-wedge segments (see Appendix B for a description of the software; Set-up files for the program are available on the CD ROM packaged with this thesis in the directory "D:Image09/setup.exe"). Image capture and analysis of leaf segments using the commercial system was carried out by a single scorer with extensive experience in the use and development of image analysis techniques.

Image capture and analysis of leaf segments using the custom system was carried out independently by the same four scorers that visually assessed chlorotic areas. Prior to image capture and analysis scorers received approximately five minutes of instruction on setting-up the image capture equipment and using the Image09 software. For both the custom and commercial systems scorers repeated the entire image capture and analysis sequence twice.

The degree of analysis automation achievable with Image09 was investigated. For images examined using the custom image analysis system, the pixel shade values corresponding to both the background/healthy leaf area and the healthy/chlorotic leaf area threshold cut-off values determined automatically by Image09 were compared with corresponding values manually selected by the four scorers. To assess the feasibility of using Image09 to determine thresholding cut-off values without any human input, the software was used to analyse all of the images captured with the Joyce Loebl system without any scorer input other than image selection. Automatically determined threshold cut-off values, percentage chlorotic leaf areas and analysis times were recorded.

The mean chlorotic area values of the asymptomatic control leaf segments included in each analysed image were subtracted from the chlorotic area estimate for each symptomatic leaf segment analysed in that image. This yielded chlorotic area estimates adjusted to account for pixels incorrectly grouped in the chlorotic leaf area category.

2.2.6 Analysis of Symptoms by Leaf Tracing

A hybrid of analogue and digital techniques based on the acetate image analysis procedure described by Nutter *et al.* (1993) was used for quantification of the actual chlorotic areas of a random subset of the leaf segments involved in this study. For each selected leaf segment five- to ten-fold magnified tracings from projections of colour slides were made. Analysis of a single leaf segment involved making tracings of both total leaf area and total area of achlorotic leaf tissue. A digital image containing these was captured using the GDS 5000 system. Each image was processed to obtain a 2-color (black and white) image using Paint Shop Pro version 4.12 (JASC Inc., Eden Prairie, Minnesota). Calculation of percentage chlorotic leaf area from a processed image was carried out using IMPROCES version 4.2 (John Wagner, San Diego, California) to separately count the number of black pixels comprising the total and achlorotic leaf area tracings.

For each selected leaf segment, tracings of total and achlorotic areas and assessment of percentage chlorotic areas were repeated twice on consecutive weeks.

2.2.7 Assessment of Symptom Quantification Techniques

Linear regression analysis as described by Nutter *et al.* (1993) was used to estimate the relative precision of all the analysis techniques described. Reproducibility of chlorotic area estimations using visual symptom assessment and the image analysis techniques was determined for each scorer by regressing their first and second chlorotic area ratings of leaf segments against one another. To determine inter-scorer precision of symptom assessments made visually and with the custom image analysis system, means of each scorer's duplicated chlorotic area estimates were linearly regressed against those of all the other scorers and analysed accordingly. The accuracy of individual scorers employing the image analysis and visual assessment techniques were determined by linear regression analysis of the means of duplicated chlorotic area estimates against the means of duplicate percentage chlorotic area measurements made using the leaf tracing technique. To determine the practical implications of employing the different symptom quantification approaches, we analysed visual and image analysis derived chlorotic area estimates in relation to the seven MSV isolates used. The mean percentage chlorotic areas of leaves infected with the different MSV strains as determined by the four scorers visually and with the custom image analysis system, were compared. For all the analysis techniques total times taken to obtain chlorotic area estimates were recorded and averaged to obtain mean analysis times per leaf segment.

2.3 RESULTS

2.3.1 Assessment of Precision

In repeated chlorotic area assessments by single scorers (Table 2.1) leaf tracing had the highest reproducibility of the techniques evaluated. While repeated visual assessment of chlorotic areas by individual scorers yielded the least reproducible results, the repeatability of a single scorer's chlorotic area estimates using either the commercial or custom image analysis systems were similar. Linear regression analysis of repeated symptom assessments by all scorers using either image analysis system yielded regression lines with y-intercepts and slopes which did not differ

TABLE 2.1 Determination of intra-scorer precision by linear regression analysis of repeated chlorotic area estimates made using the various symptom assessment techniques

Analysis Technique	Scorer	Regression Parameters		
		R^2	Slope	Y-int[a]
Leaf Tracing	Manual	0.990	0.976	1.140
Commercial Image Analysis System	Manual	0.967	1.000	1.376
	Automatic	0.972	0.990	1.360
Custom Image Analysis System	Scorer 1	0.966	1.019	-1.270
	Scorer 2	0.962	0.965	1.031
	Scorer 3	0.972	0.999	0.033
	Scorer 4	0.969	0.979	0.539
Visual Assessment	Scorer 1	0.946	0.955	5.705*[b]
	Scorer 2	0.965	0.995	-3.023*
	Scorer 3	0.945	0.983	2.819*
	Scorer 4	0.896	1.087*[c]	0.999

[a] Y-intercept
[b] Y-intercept values which are significantly different from zero ($P < 0.01$) are followed by *.
[c] Slope values which are significantly different from one ($P < 0.01$) are followed by *.

significantly from zero and one respectively. Although linear regressions of repeated visual symptom estimations yielded regression lines with slopes which did not differ significantly from one, for three scorers, y-intercepts of the regression lines differed significantly from zero.

The inter-scorer precisions of chlorotic area estimates were generally higher using the custom image analysis system than when estimates were made visually (Table 2.2). In all cases, when estimates were made using the image analysis technique, y-intercepts and slopes of regressions did not differ significantly from zero and one respectively. Despite generally high coefficients of determination when regressing visually determined estimates of different scorers against one another, all of the y-intercepts and four of the six slopes derived from these regressions were significantly different from zero and one respectively.

2.3.2 Assessment of Accuracy

Regression analysis of estimated against actual chlorotic areas indicated that the commercial and custom image analysis systems were comparable in accuracy (Table 2.3). In all cases the slopes and y-intercepts from these regressions did not differ significantly from one and zero respectively (Fig. 2.2). Conversely, regressions of visual estimates against actual chlorotic areas yielded both y-intercepts which were significantly greater than zero for scorers one, three, and four and slopes which were significantly greater than one for scorers one, two and three (Table 2.3).

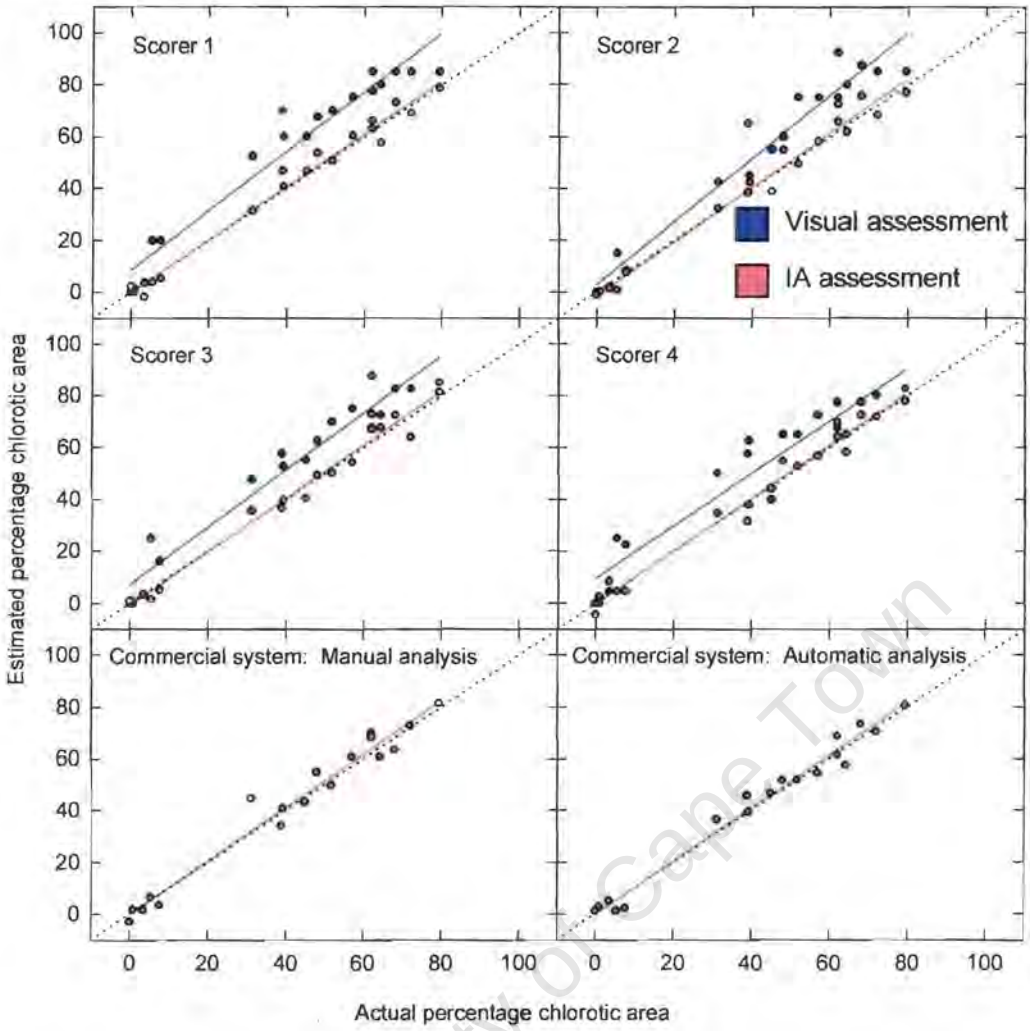


FIGURE 2.2 Regression of image analysis (in pink) and visually derived (in blue) percentage chlorotic area estimates against actual percentage chlorotic areas of 20 randomly selected leaf sections. Image analysis carried out by scorers one, two, three, and four employed a custom system. Images captured using a commercial image analysis system were analysed both manually using the commercial software (JLGenias version 3.01) and automatically using custom-written software (Image09). The dotted lines have slopes = 1 and y-intercepts = 0. The coefficients of determination, y-intercepts and slopes of the regressions presented here are given in Table 2.3.

TABLE 2.2 Determination of inter-scorer precision by linear regression analysis of chlorotic area estimates made by different scorers both visually and using the improvised image analysis system

Regression	Visual Assessment			Custom Image Analysis System		
Comparison	R^2	Slope	Y-int	R^2	Slope	Y-int[a]
Scorer 1 vs. Scorer 2	0.932	1.098[b]	-9.237[c]	0.948	0.983	0.948
Scorer 1 vs. Scorer 3	0.929	0.994	-3.028*	0.960	1.030	-0.841
Scorer 1 vs. Scorer 4	0.914	0.876*	3.467*	0.953	1.015	-0.742
Scorer 2 vs. Scorer 3	0.983	0.900*	5.621*	0.953	0.953	0.034
Scorer 2 vs. Scorer 4	0.937	0.780*	11.656*	0.952	0.996	0.727
Scorer 3 vs. Scorer 4	0.972	0.876*	6.371*	0.962	0.962	1.319

[a] Y-intercept

[b] Slope values which are significantly different from one ($P \# 0.01$) are followed by *.

[c] Y-intercept values which are significantly different from zero ($P \# 0.01$) are followed by *.

TABLE 2.3 Determination of disease assessment technique accuracy based on linear regression analysis of chlorotic area estimates obtained using the various analysis techniques plotted against actual chlorotic areas

Analysis Technique	Scorer	Regression Parameters		
		R^2	Slope	Y-int[a]
Commercial Image Analysis System	Manual	0.973	1.040	-0.877
	Automatic	0.976	1.045	0.346
Custom Image Analysis System	Scorer 1	0.985	1.013	0.275
	Scorer 2	0.971	1.032	-0.782
	Scorer 3	0.980	1.020	-0.190
	Scorer 4	0.979	0.999	0.610
Visual Assessment	Scorer 1	0.913	1.066*[b]	12.958*[c]
	Scorer 2	0.960	1.245*	1.064
	Scorer 3	0.957	1.135*	5.201*
	Scorer 4	0.927	1.029	8.385*

[a] Y-intercept
[b] Slope values which are significantly different from one ($P < 0.01$) are followed by *.
[c] Y-intercept values which are significantly different from zero ($P < 0.01$) are followed by *.

When comparing the mean percentage chlorotic areas of leaves infected with the different MSV strains as determined by the four scorers visually and with the custom image analysis system, it was noted that for all of the MSV strains except MSV-VM, visually derived estimates were higher than estimates obtained by image analysis (Fig. 2.3). In all cases except for MSV-VW, there was less variation between the mean chlorotic area estimates of the scorers when image analysis was used. Results obtained by image analysis indicated that, in terms of average chlorotic areas resulting from infection, MSV-Kom and MSV-VW did not differ significantly from one another in severity but both were significantly less severe than MSV-N. Alternatively, results obtained by visual assessment indicated that the severities of MSV-N and MSV-Kom were not significantly different from one another but that both were significantly more severe than MSV-VW.

2.3.3 Assessment of Automated Symptom Quantification

Of the 112 images captured using the custom image analysis system and analysed using Image09, the step-wedge was correctly identified in 111 (99.1%), the segment corresponding to healthy leaf area was correctly identified in 110 (98.2%) and the segment corresponding to chlorotic leaf area was correctly identified in 108 (96.4%). For the background/healthy and healthy/chlorotic leaf area pixel thresholding cut-off values, the mean absolute difference between the manually and automatically determined values for all scorers were 2.3 (± 2.3 SD) and 4.4 (± 4.0 SD)

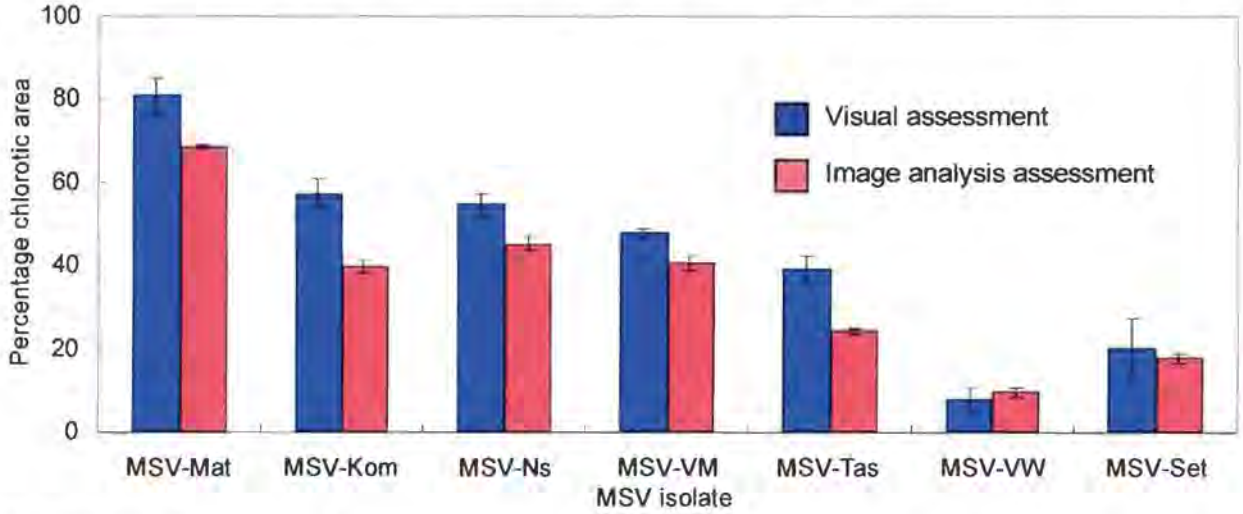


FIGURE 2.3 Comparison of mean chlorotic leaf areas resulting from infection of maize with the various MSV isolates. The bars represent means of the four scorers' individual estimations of average chlorotic areas resulting from infection with the different isolates, made both visually (blue bars) and with the use of the custom image analysis system (pink bars). The error bars represent standard deviations of the mean chlorotic area estimates made by the scorers.

respectively. Of the 28 images captured using the commercial system and examined with Image09, the step-wedge was successfully detected and used to set thresholding cut-off values in 27 (96.4%). Linear regression analysis of chlorotic area estimates from the first and second groups of images captured with this system indicated that automatic thresholding and symptom calculation was at least as precise as that carried out manually on the same images using JLGenias, the commercial image analysis software (Table 2.1). The means of duplicated chlorotic area estimates obtained manually using the commercial software and automatically using Image09 were highly correlated and, when regressed against one another, yielded a regression line with a y-intercept of -1.453 ($P > 0.99$ that this value is not significantly different from zero), a slope of 1.025 ($P > 0.99$ that this value is not significantly different from one) and a coefficient of determination of 0.990. The accuracy of chlorotic area estimates determined automatically with Image09 compared very favourably with those determined manually using the commercial software (Fig. 2.2).

2.3.4 Duration of Analyses

Of the techniques assessed, visual quantification of symptoms was fastest, requiring, on average, 7.6 (± 2.6 SD) seconds per leaf segment. The leaf tracing technique, which required 35.7 (± 15.4 SD) minutes per leaf segment, was the most time consuming. Mean analysis times required to manually determine the percentage chlorotic areas on individual leaf segments using the custom and commercial image analysis systems were 36.2 (± 2.6 SD) and 52.9 (± 6.1 SD) seconds

respectively. Automated analysis of images captured with the commercial system using Image09 required only 10.1 (± 0.3 SD) seconds per leaf segment.

2.4 DISCUSSION

In this chapter the potential use of commonly available resources for the development of a fast, precise, accurate, and objective symptom quantification technique has been demonstrated. The pixel thresholding approach used for the analysis and quantification of percentage chlorotic leaf areas has been the basis of most published protocols for symptom quantification by image analysis (Blanchette, 1982; Kampmann and Hansen, 1993; Lindow and Webb, 1983; Newton, 1989; Price *et al.*, 1993) . Because of the simplicity of pixel thresholding it was possible to devise a variety of symptom quantification techniques utilising software either obtained at trivial cost via the Internet or written specifically to suit the purposes of this study. An example of one of these techniques is presented here together with an analysis of its precision and accuracy relative to both commercial image analysis and visual symptom quantification techniques.

Two components of precision have been identified as significant when attempting to establish the overall precision of disease symptom quantification techniques. These are (i) intra-scorer precision, which refers to the reproducibility of disease severity estimates when made by a single scorer on different occasions; and (ii) inter-scorer precision, which refers to the reproducibility of symptom severity estimates when made independently by two different scorers (Nutter *et al.*, 1993).

Using linear regression analysis, we have demonstrated that the custom image analysis technique developed during this study yielded percentage chlorotic area estimates which were both comparable for intra-scorer precision with estimates obtained using a commercial image analysis system, and significantly better for intra- and inter-scorer precision than visual estimates. Regression analysis indicated that while relative differences between the estimated chlorotic leaf areas remained consistent within the two groups of visual chlorotic area estimates made by a single scorer, there were generally consistent differences between the estimates made by three of the four scorers.

The accuracy of the custom image analysis system was similar to that of the commercial system. Regression analysis revealed that there were neither constant nor systematic errors in estimation related to actual chlorotic areas. All scorers estimated chlorotic areas significantly more accurately with the custom system than when estimating chlorotic areas visually. While the overall accuracy of the visual estimates was high relative to previous reports for other diseases (Nutter *et al.*, 1993; Parker *et al.*, 1995; Price *et al.*, 1993; Sherwood *et al.*, 1983), regression analysis showed that visual estimations of chlorosis significantly exceeded actual areas across the entire range of symptoms for three out of four scorers, a situation commonly reported (Parker *et al.*, 1995; Price *et al.*, 1993; Sherwood *et al.*, 1983). In addition, and in accordance with previous reports, the degree of overestimation by three out of four scorers increased significantly as actual chlorotic areas increased (Price *et al.*, 1993; Sherwood *et al.*, 1983). Despite these relatively consistent errors, however, examination of residuals (differences between estimated and actual chlorotic areas) indicated that no mathematical transformation of the visual estimates to achieve greater linearity in the regressions was warranted. Also, the slopes of the regression lines did not differ from one in a consistent manner indicating that no single mathematical transformation to linearity could be uniformly applied to the chlorotic area estimates made by all scorers.

It was found that differentiating between strains based on of the degree of chlorosis they induced in infected plants, was influenced by whether the symptoms had been assessed visually or with image analysis. This indicates the need for accurate symptom quantification in studies dealing with the differentiation among MSV isolates that have subtly different pathologies.

An important feature of the image analysis protocols employed in this study was the inclusion of healthy leaf segments and a graded grey-scale step-wedge as internal controls in every digital image examined. The inclusion of at least six asymptomatic control leaves in all the captured images permitted estimation of the number of pixels grouped in the chlorotic category that should have been grouped in the healthy category: these pixels generally represented the lightly coloured leaf midrib. The background level of incorrectly identified chlorosis on healthy leaves was used to correct chlorotic area estimates of symptomatic leaves to discount leaf midribs included in the "diseased" pixel group.

The step-wedge played a dual role in the image analysis procedures, functioning both as a lighting standard and as an internal colour control. Its use obviated the need to specify camera aperture

settings, and location of lighting source(s) relative to specimens, since these would not be relevant to any system other than the one used in this study. With both the custom and commercial systems, the potential usefulness of an incident lighting intensity standard for recreating consistent lighting under different conditions was demonstrated. This represents a viable solution to the problem of standardising image analysis protocols in different locations using varied equipment and software.

Inclusion of the step-wedge as an internal colour control in images also provided a disease symptom independent guide for both the manual and automatic setting of pixel threshold cut-off values. The potential for negating the input of a human scorer at the analysis stage of symptom assessment was demonstrated with the use of a step-wedge as an internal colour control, coupled with the inclusion of step-wedge identification and examination algorithms in the Image09 software. Pixel threshold cut-off values automatically determined by the custom-written software compared well with those specified manually by scorers. Similarly, images captured with the commercial system and analysed with Image09 without any scorer input other than image selection, yielded chlorotic area estimates which were as precise and accurate as chlorotic area estimates obtained manually using JLGenias, the analysis software developed specifically for the commercial system. This result is significant in that it demonstrates that human judgement, which is an important subjective element in any analysis, need not be relied upon to determine which pixels should be grouped in the background, healthy and diseased categories.

Image09 has been developed with an in-built calibration capacity enabling the software to automatically determine pixel cut-off values from the graded grey scale segments of any step-wedge. Therefore, use of this software in conjunction with an appropriate step-wedge will enable increased standardisation and automation of analysis in image analysis protocols for quantifying the symptoms of other plant pathogens. The direct customisation of the software has also been facilitated by making the source code, written in Visual Basic and Visual C++ (Microsoft Corporation, Redmond, Washington), freely available (the source code can be found on the CDROM packaged with this thesis in the directories "D:Image09-SourceCode/VB" and "D:Image09-SourceCode/C++"). It should be possible for anybody reasonably competent at programming in BASIC or C to customise the program to accommodate any foliar disease symptom assessment task. These customisations could include: (i) adaptation of the software to analyse symptoms in direct conjunction with image capture equipment; (ii) adding symptom

identification algorithms that are more complex than the pixel thresholding ones which are currently in wide use; and (iii) modification of the pixel examination and manipulation algorithms to permit the analysis of colour images.

Because the image analysis protocols described in this chapter were found to be suitable for the accurate and precise estimation of chlorotic areas in MSV-infected maize leaves, they were subsequently used to both analyse the virulence of MSV isolates and rapidly screen maize genotypes for MSV resistance (see Chapters 3, 4, and 5) . With regard to the analysis of MSV symptoms in pathology studies, the image analysis techniques described here could provide either a viable alternative to or a valuable refinement of the semi-quantitative MSV symptom rating scales currently employed for breeding purposes (Barrow, 1992; Efron *et al.*, 1989; Govinden and Rummun, 1996; Rodier *et al.*, 1995; Van Rensburg *et al.*, 1991).

Chapter 3

Evaluation of *Maize Streak Virus* Pathogenicity in Differentially Resistant *Zea mays* Genotypes

Contents

ABSTRACT

A rapid technique was devised for the objective and precise assessment of *Maize streak virus* (MSV) resistance amongst maize genotypes, and of the pathogenicity of cloned MSV isolates. The technique involves the use of agroinoculation to infect maize seedlings and objective symptom evaluation by quantification of infection rates, stunting, and chlorotic leaf areas. In assessing the MSV resistance of 19 maize genotypes, a description is given of how differentially virulent virus isolates can be used to quantify MSV resistance phenotypes ranging from extremely susceptible to completely immune. It is demonstrated how quantification of chlorotic leaf areas by image analysis permits differentiation between degrees of MSV resistance that are indistinguishable from one another using currently employed symptom assessment approaches. The viability of using the technique to compare the pathogenicity of cloned MSV isolates was also investigated. It was found that the orientation of cloned viral genomes relative to the CaMV35S promoter of the binary cloning vector pBI121 can significantly affect the infectivity of agroinfectious constructs, the amount of stunting induced in infected plants and the degree of leaf chlorosis on the earliest symptomatic leaves of infected plants. The influence of cloning vector sequences on the efficiency of the agroinfection process is discussed and measures of MSV pathogenicity that are not influenced by these sequences are identified.

3.1 INTRODUCTION

The belief that the best solution to the African MSV problem is the widespread use of high yielding MSV resistant varieties has led to the initiation of maize breeding programmes aimed at developing such varieties throughout Africa (Barrow, 1992; Efron *et al.*, 1989; Rodier, 1995; Van Rensburg *et al.*, 1991). An important part of these programmes and related MSV pathology studies has been the development of a number of techniques for the quantification of MSV symptoms. The techniques currently employed by breeders are semi-quantitative and most commonly involve the use of five to twelve point rating systems based on the intensity of streaking (Barrow, 1992; Kim *et al.*, 1989; Rodier, 1995; Van Rensburg *et al.*, 1991). While these systems have the advantage of being rapid enough to permit the evaluation of large numbers of infected plants, their reliance on visual estimation of chlorotic areas makes them fundamentally subjective. Besides visual estimation of chlorotic areas, MSV pathology studies carried out in conjunction with breeding programs have employed measurement of infection rates (Boulton *et al.*, 1991b; Isnard *et al.*, 1997), chlorophyll concentrations, viral DNA concentrations (Boulton *et*

al., 1991b), chlorotic leaf areas (Rodier, 1995) and stunting (Isnard *et al.*, 1997) to objectively quantify the virulence of MSV isolates.

While natural transmission of MSV is mediated exclusively by a number of leafhopper species in the genus *Cicadulina* (Storey, 1928), Grimsley *et al.* (1987) have developed an alternative means of transmitting MSV using *Agrobacterium tumefaciens* - mediated transfer of tandemly cloned complete or partial dimers of MSV replicative forms in binary cloning vectors (called agroinfectious constructs). This technique, called agroinfection or agroinoculation, enables the highly efficient and reproducible infection of maize plants with precise concentrations of single MSV genotypes.

It has been demonstrated that infection rates achieved by injecting plants with agroinfectious MSV constructs are affected both by the *Agrobacterium* species and strain used to deliver agroinfectious constructs, and by the number of long intergenic regions (LIRs) within the constructs (Schnippenkoetter, 1998). While agroinfection is potentially a powerful tool for comparing the virulence of cloned viruses, it is essential that differences in symptoms can be wholly attributed to differences between the virus genomes being compared.

In this chapter a technique for accurately and objectively evaluating the MSV resistance of maize genotypes is described and modifications are investigated that enable the technique to be used for comparing the relative virulence of cloned MSV isolates. Also described are the potential effects that cloning vector sequences can have on the apparent virulence of agroinfectious MSV clones. The technique uses agroinoculation to transmit specific virus isolates and measurements of chlorotic leaf areas, stunting and infection rates, to allow rapid differentiation between the relative MSV resistance of maize genotypes with a degree of resolution and objectivity not attainable with currently employed resistance evaluation techniques.

3.2 METHODS AND MATERIALS

3.2.1 Maize Genotypes and Virus Strains

Seed of the inbred maize lines, Z391, Z446, Z459, Z470, Z471, and Z491, was supplied by Dr K. Pixley (CIMMYT, Harare, Zimbabwe). Seed of the inbred lines P608, P612, CML204, Mo17,

and B73 was provided by J. B. J. van Rensburg [Summer Grains Centre (SGC), Potchefstroom, South Africa]. Seed of the hybrids PAN6099, PAN6191, PAN6193, PAN6195, PAN6363, PAN6364, and PAN6549 was obtained from D. Nowell (PANNAR Ltd., Greytown, South Africa). Sweetcorn (cv. Jubilee) seed was obtained from Starke Ayres Nursery (Cape Town, South Africa). While cloning of the full length infectious genome of MSV-MatA is described in Chapter 4, construction of the chimaeric MSV genomes, MSV-MatMPKom, MSV-MatCPKom, and MSV-MatMPCPKom is described in Chapter 5. During preliminary studies based on visual assessment of symptoms in sweetcorn (cv. Jubilee), these viruses were identified as being very severe (MSV-MatA, MSV-MatCPKom), moderately severe (MSV-MatMPCPKom) and moderate (MSV-MatMPKom) in their virulence in maize. *Agrobacterium tumefaciens* C58C1 [pMP90] (Koncz and Schell, 1986) was used to deliver constructs during agroinoculation.

3.2.2 Production of Agroinfectious Constructs

Agroinfectious constructs of the infectious MSV isolates MSV-Kom and MSV-Set (pKom and pSet respectively) were obtained from Dr W. Schnippenkoetter (University of Cape Town, Cape Town, South Africa) and a non-infectious construct of the isolate MSV-WW (pWW) was obtained form Dr J. A. Willment (University of Cape Town, Cape Town, South Africa) .

Construction of agroinfectious clones was carried out according to Grimsley *et al.* (1987). To permit efficient replicative release upon agroinfection, all constructs contained one full MSV genome bounded by two full long intergenic regions (Schnippenkoetter, 1998). Entire genomes of the four MSV isolates MSV-MatA, MSV-MatMPKom, MSV-MatCPKom, and MSV-MatMPCPKom were all cloned in both orientations into the *BamHI* site of pUC19 using standard cloning techniques (Sambrook *et al.*, 1989). Each of the 8 clones was dimerised according to Palmer (1997), and cloned into the *EcoRI* and *XbaI* sites of pBI121 (CLONTECH, CA) to obtain MSV-AI, MSV-AII (MSV-MatA in both orientations in pBI121), MSV-BI, MSV-BII (MSV-MatMPKom in both orientations in pBI121), MSV-CI, MSV-CII (MSV-MatCPKom in both orientations in pBI121), MSV-DI, and MSV-DII (MSV-MatMPCPKom in both orientations in pBI121, Fig. 3.1). Dimerised MSV-MatA genomes were additionally cloned in both orientations into the *EcoRI* and *XbaI* sites of pBIN19 (CLONTECH, CA) to obtain MSV-AIII and MSV-AIV (Fig. 3.1). Constructs were transformed into *Agrobacterium tumefaciens* C58C1 (pMP90) (Koncz and Schell, 1986) by the method of An *et al.* (1988).

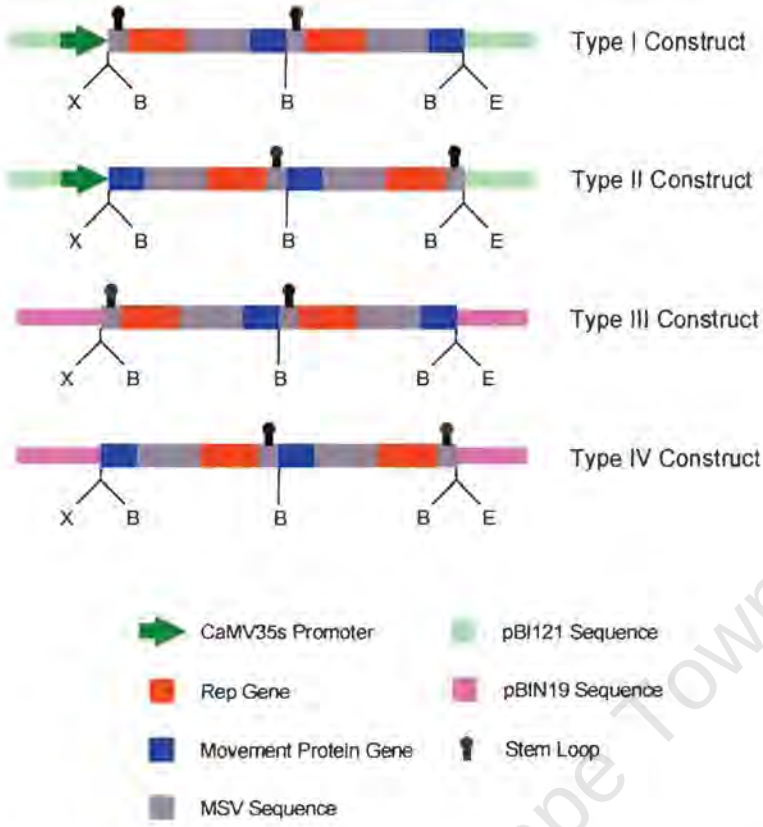


FIGURE 3.1 Types of agroinfectious MSV constructs evaluated in this study. Type I (MSV-AI, MSV-BI, MSV-CI and MSV-DI) and type II (MSV-AII, MSV-BII, MSV-CII and MSV-DII) constructs are tandemly dimerised full length MSV genomes cloned in opposite orientations into the XbaI and EcoRI sites of pBI121. Type III (MSV-AIII) and type IV (MSV-AIV) constructs are tandemly dimerised full length MSV genomes cloned in opposite orientations into the XbaI and EcoRI sites of pBIN19. B, E, and X represent BamHI, EcoRI, and XbaI restriction sites, respectively.

3.2.3 Agroinoculation

Agroinoculation was performed according to Grimsley *et al.* (1987) with the following modifications: (i) standardised inoculum was obtained from actively growing *A. tumefaciens* cultures at OD_{600}=0.4 which were concentrated 10-fold after two washes with sterile distilled water and stored on ice for no more than 30 minutes before injection; (ii) 3d old seedlings were injected; (iii) injection of 2:1 of inoculum 1mm beneath the coleoptilar node of seedlings was achieved using a 25:1 Hamilton syringe (Bonaduz, Switzerland), with its bevelled tip inserted to a standardised depth of 1.5 mm (aided by a specifically designed collar) into the coleoptile; (iv) only seedlings with a coleoptile between 15 and 30 mm in length were selected for injection. Agroinfection experiments involving specific constructs were carried out on groups of 14 plants. For each agroinfection experiment, a group of 14 seedlings injected with sterile distilled water served as uninfected controls. Seedlings were grown under near uniform conditions in a plant

growth room maintained at between 21 and 22°C, at 80% relative humidity, with 16 hours of light per day.

3.2.4 Analysis of Symptoms

Disease severity was measured in terms of (i) percentages of agroinoculated plants that became infected, (ii) percentages of leaf areas covered by chlorotic lesions in infected plants, and (iii) heights of infected plants relative to uninfected control plants. The proportion of plants showing symptoms was determined at three-day intervals between the fifth and fourteenth days after injection. For each MSV isolate-plant genotype combination these measurements were integrated into an infection rate (IR) value calculated as the mean percentage of infected plants observed at the four assessment times. Percentage leaf areas covered by chlorotic lesions in symptomatic plants were estimated for leaves two through six using the microcomputer-based image analysis technique described in Chapter 2. The image analysis protocol required the use of a digital image capture device and analysis software developed in our laboratory (set up files for the softwre, Image09 can be found on the CD-ROM supplied with this thesis in the directory D:Image09/setup.exe). The percentage chlorotic areas of leaves two and three were assessed 15 days after agroinoculation; the percentage chlorotic areas of leaves four, five and six were assessed 22, 29 and 35 days after agroinoculation respectively. For each virus isolate - plant genotype combination the mean percentage chlorotic area of the 2nd to 6th leaf (designated $C_{2\text{-}6}$), was used as a representative measure of chlorosis. Heights of symptomatic plants and uninfected control plants were measured 15 days after agroinoculation as the distance from their coleoptilar nodes to the tip of their fourth leaves. For specific MSV isolate - maize genotype combinations a value designated S (representing stunting occurring as a result of infection) was calculated as the mean height of symptomatic plants expressed as a proportion of the mean height of uninfected control plants. The value of 1-S was used as a description of stunting. Resistance was rated on a commonly used 6-point scale (0 = immune to 5 = highly susceptible) for each of the maize genotypes by the breeders that provided them for use in this study.

3.2.5 Evaluation of MSV Isolates and Maize Genotypes

The MSV-A (I-V), MSV-B (I and II), MSV-C (I and II), and MSV-D (I and II) agroinfectious constructs were used to infect sweetcorn (moderately susceptible) and PAN6099 (moderately

resistant). Agroinfection experiments involving each maize genotype – agroinfectious construct combination were repeated at least three times.

The MSV-AI, pSet and pKom constructs (representing MSV-MatA, MSV-Kom and MSV-Set, respectively), were used in the evaluation of MSV resistance in 19 maize lines. Each maize line was challenged with these three MSV isolates in at least three agroinfection experiments. For each agroinfection experiment involving a specific maize genotype the values, *IR*, *1-S*, and $C_{2\text{-}6}$, respectively representing the mean IR, 1-S and $C_{2\text{-}6}$ values determined for the three test MSV isolates, were used as generalised measures of resistance.

3.3 RESULTS

Chlorotic lesions on the second or third leaves of agroinfected plants first appeared between four and eleven days after injection. With very few exceptions, all plants that became infected showed symptoms by the 11th day after agroinoculation. Neither chlorotic streaking nor stunting were observed in plants injected with either distilled water, *A. tumefaciens* C58C1 [pMP90] containing pWW (a non-agroinfectious construct of the MSV isolate, MSV-WW), or *A. tumefaciens* C58C1 [pMP90] containing the binary cloning vector pBI121 without an MSV genome.

3.3.1 The Effect of Vector Sequences on Agroinfectious MSV Clones

An investigation was carried out to determine the effect of MSV cloning orientation in a series of agroinfectious constructs on the symptoms resulting when these constructs were used in agroinfection of highly MSV sensitive (Jubilee) and moderately MSV resistant (PAN6099) maize genotypes. The type I agroinfectious constructs (Fig. 3.1) containing the four MSV-isolates were, based on infection rate (IR) values, all more infectious than the type II constructs (Fig. 3.2). Both the degree of stunting 15 days post inoculation and the chlorotic areas on leaves two and three were greater for the type I than for the type II constructs (Fig.s 3.2 and 3.3). This was probably due to the difference in the frequency with which plants became symptomatic when injected with the two types of construct.

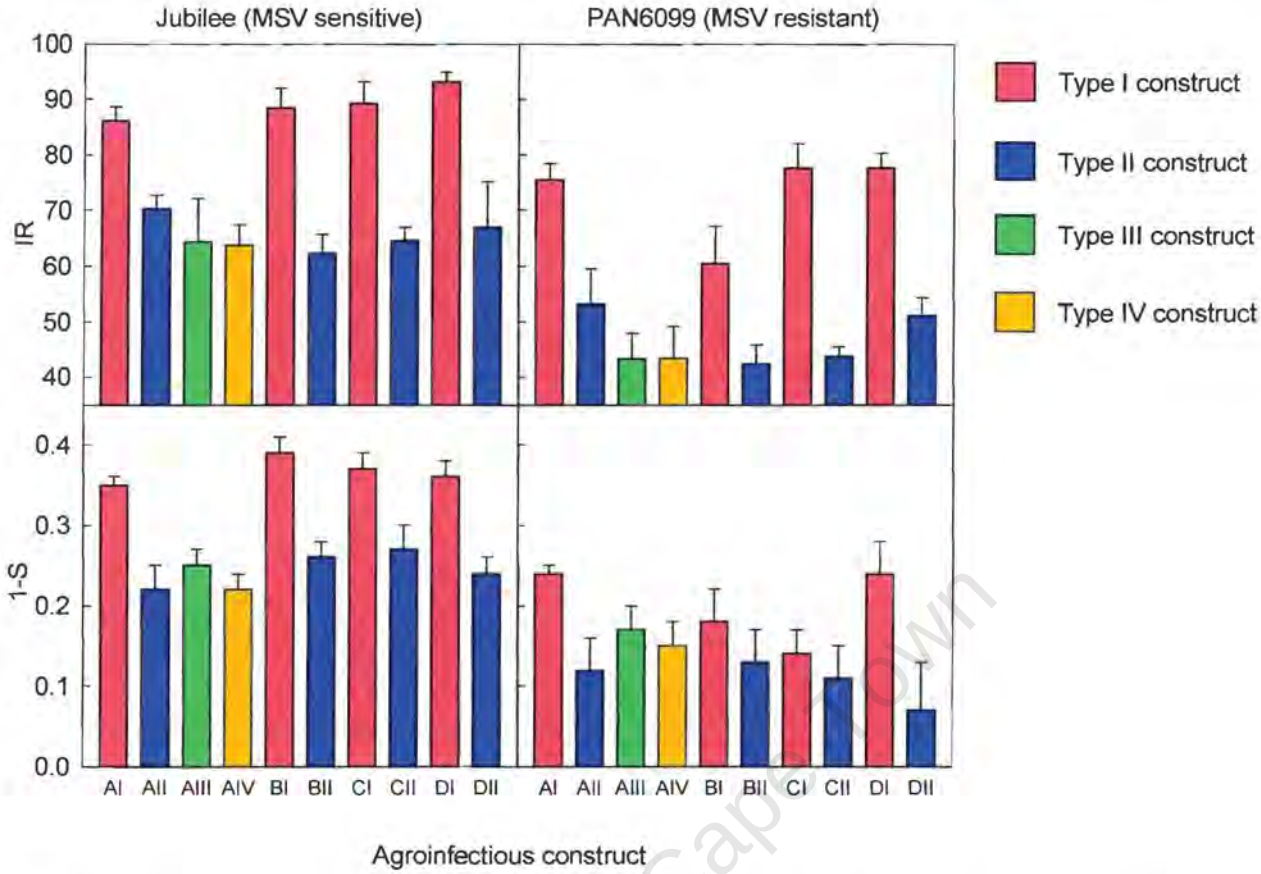


FIGURE 3.2 Relative rates at which plants became symptomatic (IR) and the amount of stunting symptomatic plants exhibited (1-S) when infected with different agroinfectious constructs. Type I constructs (AI = MSV-AI, BI = MSV-BI, CI = MSV-CI and DI = MSV-DI) are represented by pink bars, type II constructs (AII = MSV-AII, BII = MSV-BII, CII = MSV-CII and DII = MSV-DII) by blue bars, the type III construct (AIII = MSV-AIII) by green bars and the type IV construct (AIV = MSV-AIV) by orange bars. Error bars represent 95% confidence intervals of the mean.

It was hypothesised that the different rates at which plants infected with type I and type II constructs became symptomatic was due to either the sense of the single-stranded DNA (ssDNA) transferred during the agroinoculation process or the orientation in the constructs of viral genomes relative to the CaMV35S promoter. Because geminiviruses are ssDNA viruses it is feasible that the sense of the ssDNA delivered into maize cells by *A. tumefaciens* during agroinfection may have influenced the infectivity of agroinfectious constructs.

Rep is the only viral gene product required for replication (Lazarowitz *et al.*, 1989) and at the onset of an infection it can be assumed that the first viral gene product expressed is Rep (Palmer and Rybicki, 1998). There are believed to be two mechanisms by which infectious unit length genomes are derived from cloned geminivirus DNA following agroinfection. These are known as

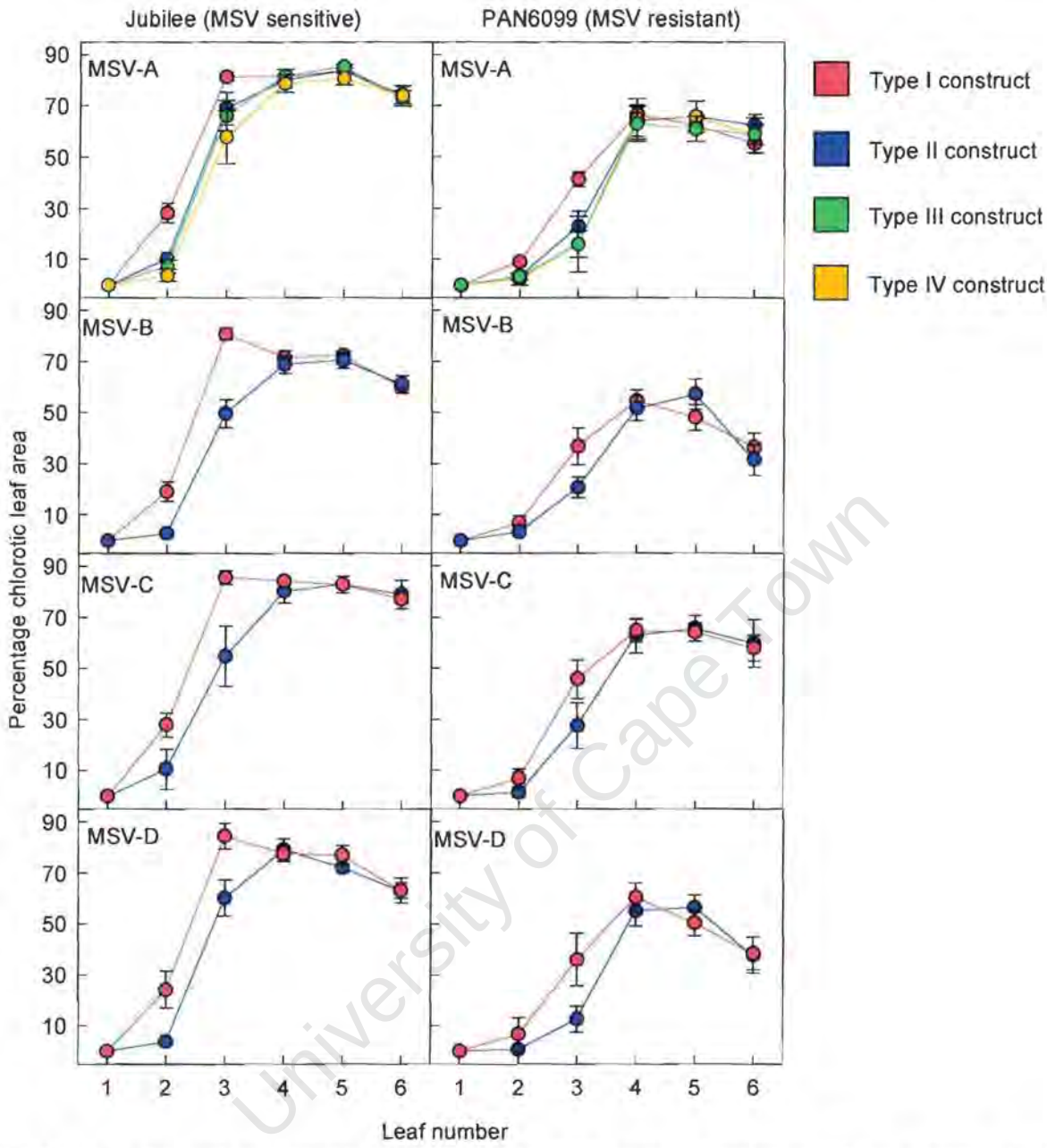


FIGURE 3.3 Percentage chlorotic areas recorded on leaves 2 through 6 of two maize genotypes (Jubilee and PAN6099) agroinoculated with the type I (pink symbols and lines) and type II (blue symbols and lines) agroinfectious constructs. The type III and type IV constructs, MSV-AIII (green symbols and lines) and MSVA-IV (orange symbols and lines), respectively, are represented in the MSV-A panel using broken lines. Error bars on either side of symbols represent 95% confidence intervals of the mean.

homologous recombination and replicational release (Stenger *et al.*, 1991), and, as is the case with natural MSV transmission, both require expression of Rep for the initiation of an infection. It is possible that the CaMV35S promoter upstream from the viral Rep gene in type I constructs (Fig. 3.1) may contribute to expression of Rep. In so doing, the promoter may enable earlier and/or more rapid infection of the maize cells into which *A. tumefaciens* delivers MSV DNA.

TABLE 3.1 Averaged MSV symptoms in maize genotypes infected with the test isolates MSV-MatA, MSV-Kom and MSV-Set

Maize Genotype	*IR* [a](CI [b])	*1-S*[c](CI)	$C_{2\text{-}6}$ [d](CI)	Rating
B73	82.2 (2.4)	0.351 (0.038)	50.4 (0.7)	5
Sweetcorn	81.6 (4.8)	0.262 (0.017)	45.7 (1.5)	5
M017	41.0 (11.7)	0.289 (0.087)	35.4 (3.2)	5
PAN6363	51.1 (5.0)	0.162 (0.017)	31.7 (1.5)	5
PAN6364	46.6 (3.2)	0.138 (0.009)	30.0 (1.4)	5
PAN6549	49.9 (4.6)	0.110 (0.068)	30.6 (0.7)	5
PAN6193	54.5 (4.5)	0.177 (0.006)	22.7 (2.8)	2
PAN6191	45.8 (3.8)	0.163 (0.021)	22.7 (2.0)	2
PAN6195	45.1 (1.6)	0.127 (0.011)	25.3 (1.6)	2
PAN6099	41.9 (2.6)	0.121 (0.011)	23.3 (1.3)	1
Z459	59.4 (3.6)	0.116 (0.056)	26.0 (0.9)	0
P612	38.7 (10.4)	0.086 (0.051)	28.2 (0.3)	0
Z391	48.1 (1.9)	0.110 (0.056)	15.1 (1.3)	0
CML204	28.0 (6.7)	0.140 (0.052)	15.8 (3.6)	0
P608	25.4 (10.9)	0.067 (0.012)	19.2 (1.4)	0
Z446	50.9 (1.8)	0.049 (0.007)	12.7 (2.0)	0
Z471	21.6 (5.0)	0.022 (0.023)	8.1 (3.8)	0
Z470	4.7 (5.1)	0.049 (0.047)	1.3 (0.8)	0
Z491	7.6 (6.8)	0.010 (0.035)	2.1 (1.6)	0

[a]Average infection rate.
[b] 95 percent confidence interval.
[c]Average degree of stunting.
[d]Average chlorotic area occuring on leaves two through six.

To investigate whether it was the position of the CaMV35S promoter relative to the viral Rep gene or the sense of the ssDNA transferred during agroinfection that was responsible for the differences in type I and type II construct infectivity, MSV-A was cloned in both orientations into the second binary cloning vector pBIN19. Except for the absence of the CaMV35S promoter sequence, the resulting type III (MSV-AIII) and type IV (MSV-AIV) constructs resembled MSV-AI and MSV-AII, respectively (Fig. 3.1). In both maize genotypes, MSV-AIII and MSV-AIV were not significantly different from one another in infectivity, the degree of stunting they induced (Fig. 3.2), or the chlorotic areas they produced on leaves 2 through 6 (Fig. 3.3). There were also no significant differences in infectivity or virulence between these constructs, and the MSV-AII construct. It is therefore unlikely that the sense of the transferred virus ssDNA is responsible for the different type I and type II construct infectivities. Rather, it appears to be the presence of the CaMV35S promoter upstream from the Rep gene in the Type I constructs that is responsible for their greater infectivity.

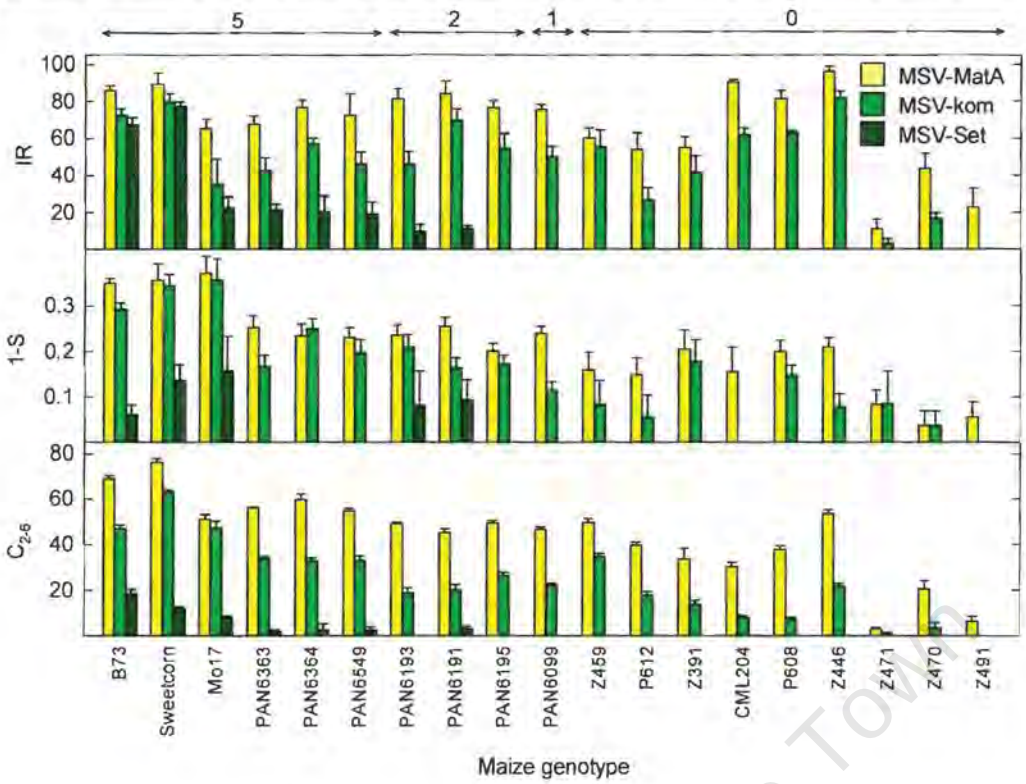


FIGURE 3.4 Relative MSV resistance of maize genotypes with regard to the rates at which plants became infected (IR), the amount of stunting symptomatic plants underwent (1-S), and the chlorotic areas which occurred on leaves two through six (C_{2-6}). Yellow, light green and dark green bars indicate results from agroinfections involving MSV-MatA, MSV-Kom, and MSV-Set respectively. Error bars represent 95% confidence intervals. Numbers and arrowed lines above the graph indicate, on a the commonly used 6 point rating scale (Rodier *et al.*, 1995), the ratings given to these lines by breeders at CYMMIT (Z491, Z471, Z470, Z459, Z446 and Z391), PANNAR (PAN6549, PAN6364, PAN6363 PAN6195, PAN6193, PAN6191, and PAN6099), and SGC (B73, Mo17, Sweetcorn, P612, P608 and CML204).

3.3.2 Assessing the Susceptibility of Maize Genotypes to MSV

Based on measurements of mean infection rates (*IR*), stunting (*1-S*), and chlorotic areas (C_{2-6}), the three most resistant maize lines examined were Z491, Z470, and Z471 (Table 3.1). With the exception of Z459 and P612, ranking of the lines in order of MSV resistance using the resistance screening technique developed in this study, corresponded well with rankings based on MSV resistance ratings made by the breeders who produced the lines (Table 3.1, Fig. 3.4). While P612 and Z459 were rated as entirely immune by the breeders at SGC and CIMMYT respectively, the C_{2-6} values of these lines when infected with MSV-MatA and MSV-Kom were similar to those of the susceptible PANNAR hybrids (PAN6364, PAN6363 and PAN6549; Fig. 3.4). It was noted, however, that P612 and Z459 had definite resistance characteristics. They were both immune to MSV-Set and, particularly in the case of P612, were significantly less prone to MSV-induced stunting than the susceptible PANNAR hybrids. Additionally, mean chlorotic areas on individual

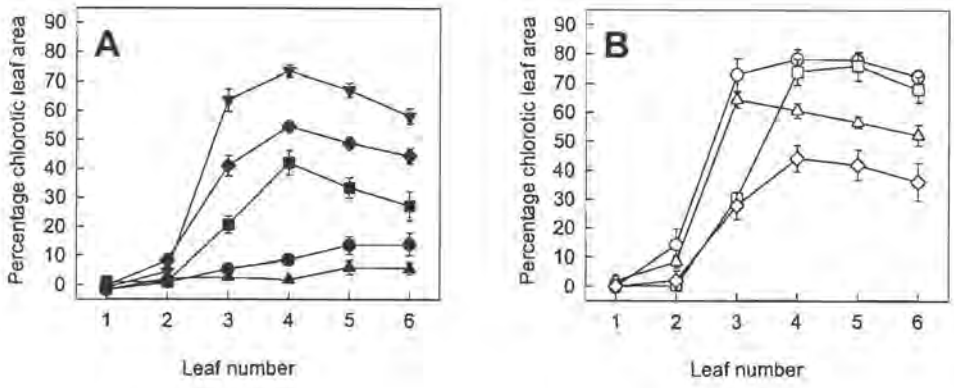


FIGURE 3.5 Use of percentage chlorotic area measurements on the first six leaves of agroinfected plants to determine the relative resistance of maize genotypes with degrees of MSV resistance which are not readily distinguishable using currently employed resistance screening approaches. A: Lines rated as immune to MSV infection by breeders at CIMMYT evaluated using MSV-MatA, a severe isolate; Z491(▲), Z471 (■), Z470 (●), Z459 (▼), Z391 (◆). B: Genotypes rated as having no resistance to MSV by breeders at SGC [B73 (○), Mo17 (□), and sweetcorn (△] and PANNAR [PAN6364 (◇)] tested with MSV-Kom, a moderately severe isolate. Error bars represent 95% confidence intervals.

leaves of MSV-MatA and MSV-Kom infected P612 indicated that while percentage chlorotic areas were greatest on leaf five (78.5 % ± 6.1 CI_{95} for MSV-MatA and 67.5 % ± 4.3 CI_{95} for MSV-Kom), they had significantly declined by leaf six (58.4 % ± 3.4 CI_{95} for MSV-MatA and 24.8 % ± 4.3 CI_{95} for MSV-Kom).

The analysis of symptoms carried out in this study revealed significant differences in MSV resistance within the groups of lines rated by breeders as either completely immune (a score of 0) or completely susceptible (a score of 5). Lines rated as immune using resistance screening methods at CIMMYT (Z391, Z446, Z459, Z470, Z471, and Z491) and SGC (CML204, P612 and P608) were all susceptible to agroinfection with both MSV-MatA and, with the exception of Z491, MSV-Kom. Differentiation between disease symptoms that occurred in all these "immune" lines was possible using the resistance evaluation technique developed in this study (Table 3.1). The most striking differences between these lines were observed by analysis of chlorotic areas following infection with MSV-MatA (Fig. 3.5). Similarly, chlorotic area measurements in MSV-Kom infected lines which had been rated as uniformly susceptible using current assessment approaches, varied significantly in their relative susceptibilities to this isolate (Fig. 3.5).

3.4 DISCUSSION

Precise information on the symptomatic responses of plants to their pathogens is extremely useful both for studying the virulence of different pathogen strains, and for breeding plants with elevated pathogen resistance. Although breeding for MSV resistant maize has yielded germplasm with significantly enhanced resistance (Barrow, 1992; Rodier *et al.*, 1995; Van Rensburg *et al.*, 1991), current procedures for testing the resistance of maize genotypes do not have sufficient resolution to differentiate between genotypes with similar resistance and are not exacting enough to properly examine highly resistant genotypes. Additionally, these resistance screening procedures yield information only on the resistance of maize lines to specific MSV isolates - information which is not necessarily accurate or relevant when the lines are challenged with different isolates (Rodier *et al.*, 1995). A technique is described here that both addresses these problems and may also be used with slight modification for analysing the virulence of MSV isolates.

It was demonstrated that, for the technique to be used to analyse the relative virulence of different agroinfectious MSV clones, it is necessary that the potential effects of sequences within cloning vectors be taken into consideration. While type II, III and IV agroinfectious constructs (Fig. 3.1) containing identical MSV genomes were equally infectious and induced indistinguishable stunting and chlorotic symptoms on all leaves tested, a type I construct containing the same genome was consistently more infectious, induced a greater degree of stunting and produced more extensive chlorotic areas on leaves two and three. However, it was also found that in plants infected with identical MSV genomes in type I and II constructs, indistinguishable chlorotic areas consistently occurred on leaves four, five and six. This indicated that regardless of cloning orientation in pBI121 or pBIN19, the pahogenicity of cloned MSV isolates agroinoculated into three day old maize seedlings can be directly compared using chlorotic area measurements on leaves four through six.

Inheritance of MSV resistance in maize is complex, involving at least three major genes and an undetermined number of minor genes (Kim *et al.*, 1989; Rodier *et al.*, 1995; Van Rensburg *et al.*, 1991; Welz *et al.*, 1998). Because of its sensitivity, the resistance evaluation technique developed in this study is ideally suited to analysing the continuum of resistance phenotypes that one would expect with quantitatively inherited resistance. Using the technique it was possible to resolving differences between the resistance of lines rated uniformly immune by breeders, indicating that

the lines tested here possibly contained a highly heterogeneous mix of MSV resistance genes. Using the technique it was also possible to detect differences in the MSV susceptibilities of lines rated as completely susceptible by breeders, indicating that a number of potentially useful minor resistance genes may exist undetected in lines previously considered to have no MSV resistance.

With two exceptions the resistance assessments made in this study were essentially in accordance with those of the breeders that had produced the maize genotypes used. While the lines P612 and Z459 were rated entirely immune by breeders at SGC and CIMMYT respectively, they displayed streak symptoms characteristic of susceptible lines when agroinoculated with MSV-MatA and MSV-Kom. These lines were possibly (i) resistant to infection at a later developmental stage, (ii) resistant to different MSV strains to those used in this study (iii) resistant to MSV infection via the leafhopper transmission route, or (iv) not resistant to MSV but either completely resistant to leafhopper feeding or less palatable to leafhoppers than other lines being tested at the same time at CIMMYT or SGC. This latter point is highly relevant since at SGC and CIMMYT different lines are tested beside one another in rows and in neither case are viruliferous leafhoppers restricted to feeding on specific plants. While SGC relies on natural infection (Van Rensburg *et al.*, 1991), CIMMYT relies on the use of anaesthetised leafhoppers which feed on plants onto which they are placed (Bosque-Perez and Allem, 1992; Welz *et al.*, 1998). PANNAR overcome problems with leafhopper feeding preferences by placing viruliferous leafhoppers in small cages attached to the second leaf of plants being tested for resistance (Barrow, 1992). Regardless of the reasons P612 and Z459 appeared immune to MSV during testing at SGC and CIMMYT respectively, the lines do appear to have a measure of MSV resistance in that they are immune to MSV-Set and experience less severe stunting than the susceptible lines. Differences in the resistance traits of P612 and Z459 from the other resistant lines examined in this study highlight the fact that various mechanisms for MSV resistance probably exist.

In this study mild, moderate, and severe agroinfectious MSV isolates were used for screening maize lines in order to demonstrate the utility of the technique described here for analysing the broadest possible range of MSV infection severities. For other applications, selection of the strains used would depend on the task at hand. Thus, while the analysis of MSV infection responses in highly resistant maize lines would warrant the use of a panel of genetically diverse but highly virulent MSV isolates, analysis of marginally resistant maize lines for low levels of polygenically inherited resistance would be best achieved with the use of mild isolates.

Using agroinoculation as a means of transmitting viruses has several advantages over leafhopper transmission. (i) Not only does agroinfection of three day old seedlings allow virus transmission rates as high as leafhopper transmission of severe MSV isolates, but it has also been found to yields higher transmission rates for milder isolates such as MSV-Set (Schnippenkoetter, 1998). (ii) Agroinoculation allows infection of plants at very specific time-points with a precise amount of inoculum and avoids the issue of differential feeding preferences of leafhoppers. (iii) Agroinoculation avoids logistical problems associated with mass rearing of viruliferous leafhoppers for simultaneous controlled transmission of different virus isolates. The main limitation of agroinoculation, however, is that *A. tumefaciens* transformed with an agroinfectious MSV construct constitutes a genetically modified organism that cannot be freely released into the environment. This inconvenience is largely offset by the fact that in an adequate containment facility, there are no constraints on the severity of MSV isolates that can be used for testing maize lines. Highly virulent isolates could be obtained either from naturally infected maize in regions know to have severe MSV strains, or by purposefully evolving resistance breaking isolates through controlled laboratory selection (Isnard *et al.*, 1998).

Because of the objectivity afforded by the symptom evaluation technique developed in this study, it could be put to use in standardised schemes for the quantification of both the MSV resistance of maize lines and, with minor modification, the virulence of MSV isolates. While it would not be convenient to use the technique for assessing the large numbers of plants commonly examined in resistance screening programs, we propose its use in a standard supplementary screening protocol for the analysis of promising maize lines. Besides a standard means of assessing resistance to MSV-isolates, the virulence of MSV isolates determined using a standard assessment protocol would also be of use to maize breeders who would have at their disposal virulence profiles of MSV genotypes found in particular locations. With this information, panels of MSV isolates specifically suited to particular resistance screening projects could be conveniently chosen. Plant pathologists studying the epidemiology of MSV would also find data of this nature useful. If resistance and virulence estimates were expressed relative to one or a few standard maize lines or virus isolates respectively, it should be possible to compare directly the results obtained by different research groups that use the technique described here.

Relative to techniques in current use, the approach described in this study provides greatly enhanced resolution and objectivity in the comparative assessment of MSV resistance amongst

maize genotypes. In this regard the technique could prove invaluable for the analysis of both marginal and extreme degrees of MSV resistance in interesting lines obtained using conventional screening approaches. While in Chapter 4 a modified version of the technique is used to assess the virulence of an extensive range of African streak virus isolates, in Chapter 5 it is used to analyse MSV pathogenicity determinants in a group of chimeric MSV constructs.

Chapter 4

Sequence Diversity and Virulence in *Zea mays* of *Maize Streak Virus* Isolates

Contents

ABSTRACT

Full genomic sequences were determined for 12 *Maize streak virus* (MSV) isolates obtained from *Zea mays* and wild grass species. These and 10 other publicly available full-length sequences were used to classify a total of 66 additional MSV isolates that had been characterised by PCR-restriction fragment length polymorphism and/or partial nucleotide sequence analysis. A description is given of the host and geographical distribution of the MSV strain and subtype groupings identified. Agroinoculation using agroinfectious clones of 21 fully sequenced MSV, *Panicum streak virus*, *Sugarcane streak virus* and *Digitaria streak virus* isolates was used to examine the relationship between the genotypes of virus isolates and their virulence in differentially MSV resistant *Zea mays* genotypes. While all isolates producing severe symptoms in maize shared greater than 95% nucleotide sequence identity, isolates obtained from grasses and wheat produced less severe symptoms in maize and shared less than 90% sequence identity with the maize isolates. Evidence that certain of the isolates investigated may be the products of either intra- and inter specific recombination is examined. The sequences determined in this study provide some insights into the recent evolution of MSV and a speculative account is given of the nucleotide substitution patterns that have occurred since the divergence of five contemporary MSV strains from their common MSV ancestors.

4.1 INTRODUCTION

Since 1984 the full genomic sequences of nine MSV isolates have been published (see Chapter 1, Fig 1.5). Until recently all of the sequenced isolates had been obtained from severely symptomatic *Zea mays* and shared greater than 95% sequence identity. The most divergent of these severe isolates, MSV-Reu, was obtained from the Indian Ocean island of La Réunion and its uniqueness is believed to reflect its evolution in isolation from the mainland African MSVs (Peterschmitt *et al.*, 1996). Full genomic sequences have recently been determined for a number of MSV isolates obtained from wild annual grass species and wheat (Schnippenkoetter, 1998; Willment, 1999). These sequences share less than 90% identity with those of the isolates obtained from maize and shed some light on the possible evolutionary origins of economically significant maize-adapted MSV genotypes.

A number of studies have attempted to determine both the extent of MSV diversity and the distribution of MSV genotypes in Africa, using partial nucleotide sequencing of MSV genomes and/or restriction fragment length polymorphisms (RFLPs) of MSV-derived PCR products (Briddon *et al.*, 1994; Rybicki *et al.*, 1998; Willment, 1999). Collectively these studies have involved the analysis of only 28 isolates obtained from maize throughout Africa and 11 isolates from annual grasses and wheat, predominantly from South Africa. These investigations have confirmed that while virtually all isolates from maize were of the MSV-A type (also known as the "maize type") and shared greater than 95% nucleotide sequence identity with one another (Briddon *et al.*, 1994, Rybicki *et al.*, 1998; Willment *et al.*, 1999), isolates from annual grasses and wheat were generally of the MSV-B and -C types (also collectively known as the "grass/wheat type") which respectively shared ~89% and ~78% nucleotide sequence identity with MSV-A isolates (Rybicki *et al.*, 1998; Willment, 1999). All of these studies, however, concentrated on inferring phylogenies of isolates from nucleotide sequence data obtained from less than half of the isolates' genomes and none fully explored the relationship between the genotypes of isolates and their ability to cause serious disease in maize.

Described in this chapter is the PCR-RFLP typing of 49 MSV isolates and complete sequencing and analysis of 12 new MSV genomes, selected from amoung the 49 as being representative of major virus groupings. These and 10 other publicly available full-length MSV sequences were used to classify a total of 88 MSV isolates (typed in this and other published MSV diversity studies) into strain and subtype groupings for which geographical and host distributions were determined. Agroinoculation of differentially resistant maize genotypes with 18 agroinfectious MSV isolates and symptom quantification by image analysis were used to identify specific groups of MSV genotypes that are likely to pose serious threats to maize production. Recombination between streak virus isolates was assessed using custom written analysis software and the potential contribution of less virulent virus strains and species to the evolution of economically significant MSV genotypes was investigated. The nucleotide sequences determined in this study were also used, together with already published streak virus sequences, to reconstruct the recent evolutionary history of the different MSV strain groupings.

4.2 MATERIALS AND METHODS

4.2.1 Maize Genotypes, Bacterial Strains, Plasmids and Virus Isolates

Seed of the inbred maize line B73 was provided by J. B. J. van Rensburg (Summer Grains Centre, Potchefstroom, South Africa). Seed of the hybrid PAN6099 was obtained from D. Nowell (PANNAR Ltd., Greytown, South Africa). Sweetcorn (cv. Jubilee) seed was obtained from Starke Ayres nursery (Cape Town, South Africa). pBluescriptSK+ (Stratagene, La Jolla, CA) was used as a vector in routine cloning. The binary cloning vectors, pBI121 (CLONTECH, CA) and pBIN19 (CLONTECH, CA) were used to construct agroinfectious clones. The RecA⁻ *Escherichia coli* strains JM109 and DH5α were used for routine cloning. *Agrobacterium tumefaciens* C58C1 [pMP90] (Koncz and Schell, 1986) was used to deliver agroinfectious clones during agroinoculation. The MSV isolates involved in this study and their origins are presented in Table 4.1. The origin of cloned full-length virus genomes that were completely sequenced and used to produce agroinfectious constructs are presented in Table 4.2. Unless otherwise stated all enzymes were obtained from Boehringer Mannheim (Mannheim, Germany)

4.2.2 Detection of MSV Diversity

DNA from the leaves of maize, *Urochloa* sp., *Eleusine* sp., and *Digitaria* sp. infected with 49 of the MSV isolates presented in Table 4.1 (listed as "this study" in the references column) was isolated by the method of Palmer *et al.* (1997). The PCR and restriction fragment length polymorphism (RFLP) MSV typing technique of Willment (1999) was used to identify virus genotypes that (1) occurred at high frequency throughout Africa and were likely to pose a significant threat to maize production on the continent, and (2) were likely to represent new MSV strains (Table 4.1). Briefly, for each of the 49 MSV isolates examined, this typing technique involved: (1) the PCR amplification (using the degenerate primer set: 5'-CAAAKDTCAGCTCCTCCG-3', 5'-TTGGVCCGMVGATGTASAG-3') of a ~1300 nucleotide sequence spanning *RepA*, the LIR and the 5' half of *MP* (co-ordinates 1897 - 361 for MSV-Ns when the A nucleotide at the origin of (+) strand replication is taken as position 1), (2) digestion of the PCR product with the restriction enzymes *RsaI*, *HpaII*, *HaeIII*, *SauIIIa*, *BamHI*, *CfoI*, and *HindIII*, and (3) matching restriction fragment lengths observed following electrophoretic separation to those determined for a reference set of fully and partially sequenced isolates with known phylogenetic relationships to one another (Willment, 1999).

TABLE 4.1 MSV isolates involved in this study, the hosts they were isolated from and their classification

Isolate Name	Isolated from	Region	Typed by[a]	$Strain_{subtype}$	Reference
		Southern Africa			
MSV-Moz	*Zea mays*	Mozambique	R/PS	A_1	Briddon *et al.*, 1994
MSV-SW	*Zea mays*	Namibia	R	A_1	Willment, 1999
MSV-1968	*Zea mays*	South Africa	R	A_4	This Study
MSV-1977	*Zea mays*	South Africa	R	A_4	This Study
MSV-Bre	*Zea mays*	South Africa	R	A_4	Willment, 1999
MSV-Cit	*Zea mays*	South Africa	R	B	Willment, 1999
MSV-CT	*Zea mays*	South Africa	R/PS	A_4	Rybicki *et al.*, 1998
MSV-Geo	*Zea mays*	South Africa	R	A_4	Willment, 1999
MSV-Koe	*Zea mays*	South Africa	R	A_4	Willment, 1999
MSV-Kom	*Zea mays*	South Africa	S	A_4	Schnippenkoetter, 1998
MSV-Mak	*Zea mays*	South Africa	R	A_5	Willment, 1999
MSV-MakC	*Zea mays*	South Africa	R	A_1	This Study
MSV-MakD	*Zea mays*	South Africa	S	A_5	This Study
MSV-MakE	*Zea mays*	South Africa	R	A_1	This Study
MSV-MakF	*Zea mays*	South Africa	R	A_1	This Study
MSV-RSE	*Zea mays*	South Africa	R/PS	A_4	Rybicki *et al.*, 1998
MSV-SA	*Zea mays*	South Africa	S	A_4	Lazarowitz, 1988
MSV-StelA	*Zea mays*	South Africa	R	A_4	This Study
MSV-Thab	*Zea mays*	South Africa	R	A_4	Willment, 1999
MSV-Tyg	*Zea mays*	South Africa	R	A_4	Willment, 1999
MSV-VM	*Zea mays*	South Africa	S	A_4	Willment, 1999
MSV-Wid	*Zea mays*	South Africa	R	A_4	This Study
MSV-BakA	*Urochloa* sp.	South Africa	R	B	This Study
MSV-DaiC	*Digitaria* sp	South Africa	R	A_1	This Study
MSV-Dig	*Digitaria* sp.	South Africa	PS	B	Rybicki *et al.*, 1998
MSV-Lib	*Urochloa* sp.	South Africa	R	B	This Study
MSV-Ond	*Setaria* sp.	South Africa	R	B	This Study
MSV-Osg	*Urochloa* sp.	South Africa	R	B	Willment, 1999
MSV-Ouw	*Setaria* sp.	South Africa	R	B	This Study
MSV-Pas	*Paspalum* sp.	South Africa	R	B	Willment, 1999
MSV-Pat	*Digitaria* sp.	South Africa	S	E	This Study
MSV-Raw	*Urochloa* sp.	South Africa	S	D	This Study
MSV-Set	*Setaria* sp.	South Africa	S	C	Schnippenkoetter, 1998
MSV-Tas	*Triticum aestivum*	South Africa	S	B	Willment, 1999
MSV-VW	*Triticum aestivum*	South Africa	S	B	Willment, 1999
MSV-WES	*Triticum aestivum*	South Africa	R	B	Willment, 1999
MSV-WESE	*Eleusine* sp.	South Africa	R/PS	B	Willment, 1999
MSV-Wil	*Secale cereale*	South Africa	R	B	This Study
MSV-WW	*Triticum aestivum*	South Africa	R/PS	B	Willment, 1999
MSV-MatA	*Zea mays*	Zimbabwe	S	A_1	This Study
MSV-MatB	*Zea mays*	Zimbabwe	S	A_1	This Study
MSV-MatC	*Zea mays*	Zimbabwe	S	A_5	This Study
MSV-ZimA	*Zea mays*	Zimbabwe	R/PS	A_1	Briddon *et al.*, 1994
MSV-ZimB	*Pennisetum sp.*	Zimbabwe	PS	A_4	Briddon *et al.*, 1994
		East Africa			
MSV-E	*Zea mays*	Ethiopia	R/PS	A_1	Briddon *et al.*, 1994
MSV-Ama	*Zea mays*	Kenya	S	A_1	This Study
MSV-Bah	*Zea mays*	Kenya	R	A_1	This Study
MSV-Emb	*Zea mays*	Kenya	R	A_1	This Study
MSV-Gat	*Zea mays*	Kenya	S	A_1	This Study
MSV-JamD	*Zea mays*	Kenya	R	A_1	This Study
MSV-Kag	*Zea mays*	Kenya	R	A_1	This Study
MSV-Kang	*Zea mays*	Kenya	R	A_1	This Study
MSV-Kany	*Zea mays*	Kenya	R	A_1	This Study

TABLE 4.1 Cont.

Isolate Name	Isolated from	Region	Typed by[a]	$Strain_{subtype}$	Reference
		East Africa			
MSV-Kar	*Zea mays*	Kenya	R	A_1	This Study
MSV-Ken	*Zea mays*	Kenya	S	A_3	Howell, 1984
MSV-Ker	*Zea mays*	Kenya	R	A_1	This Study
MSV-Kia	*Zea mays*	Kenya	R	A_1	This Study
MSV-Kit	*Zea mays*	Kenya	R	A_5	This Study
MSV-Man	*Zea mays*	Kenya	R	A_1	This Study
MSV-Mit	*Zea mays*	Kenya	R	A_5	This Study
MSV-Mom	*Zea mays*	Kenya	S	B	This Study
MSV-MtkA	*Zea mays*	Kenya	S	A_3	This Study
MSV-MtKB	*Zea mays*	Kenya	R	A_1	Willment, 1999
MSV-Mug	*Zea mays*	Kenya	R	A_1	This Study
MSV-Nju	*Zea mays*	Kenya	R	A_1	This Study
MSV-Oyu	*Zea mays*	Kenya	R	A_1	This Study
MSV-Sag	*Zea mays*	Kenya	S	A_1	This Study
MSV-Jam	*Digitaria* sp.	Kenya	S	B	This Study
MSV-JamB	*Digitaria* sp.	Kenya	R	B	This Study
MSV-JamC	*Digitaria* sp.	Kenya	R	B	This Study
MSV-JamE	*Digitaria* sp.	Kenya	R	B	This Study
MSV-MulA	*Zea mays*	Malawi	R	A_1	This Study
MSV-MulB	*Zea mays*	Malawi	R	A_1	This Study
MSV-MulC	*Zea mays*	Malawi	R	A_1	This Study
MSV-NkhA	*Zea mays*	Malawi	R	A_1	This Study
MSV-NkhB	*Zea mays*	Malawi	R	A_1	This Study
MSV-U	*Zea mays*	Uganda	R/PS	A_1	Briddon *et al.*, 1994
MSV-Zai	*Zea mays*	Zaire	PS	A_1	Briddon *et al.*, 1994
		West Africa			
MSV-Cam	*Zea mays*	Cameroon	R	A_2	This Study
MSV-GA	*Zea mays*	Ghana	R/PS	A_2	Briddon *et al.*, 1994
MSV-GB	*Zea mays*	Ghana	R/PS	A_1	Briddon *et al.*, 1994
MSV-Ns	*Zea mays*	Nigeria	S	A_2	Mullineaux *et al.*,1984
MSV-NB	*Zea mays*	Nigeria	R/PS	A_1	Briddon *et al.*, 1994
MSV-NC	*Coix lacryma*	Nigeria	R/PS	A_2	Briddon *et al.*, 1994
		Indian Ocean Islands			
MSV-Reu	*Zea mays*	La Réunion	S	A_6	Peterschmitt *et al.*, 1996
MSV-R2	*Zea mays*	La Réunion	S	A_6	Isnard *et al.*, 1998
MSV-RB	*Zea mays*	La Réunion	R/PS	A_6	Briddon *et al.*, 1994
MSV-Mau	*Zea mays*	Mauritius	R/PS	A_6	Briddon *et al.*, 1994

[a]Typing by RFLP analysis (R), partial sequencing (PS), or full genome sequencing (S)

TABLE 4.2 Agroinfectious *Mastrevirus* isolates involved in this study

Virus isolate	Source of Isolate	Year of Isolation	Region and Country of Origin	Reference
MSV-Ama	*Zea mays*	1998	Amagoro, Kenya	This Study
MSV-Gat	*Zea mays*	1998	Gathuke-ini, Kenya	This Study
MSV-Kom	*Zea mays*	1989	Komatipoort, South Africa	Schnippenkoetter, 1998
MSV-MakD	*Zea mays*	1998	Makatini, South Africa	This Study
MSV-MatA	*Zea mays*	1994	Matabeleland, Zimbabwe	This Study
MSV-MatB	*Zea mays*	1996	Matabeleland, Zimbabwe	This Study
MSV-MatC	*Zea mays*	1998	Matabeleland, Zimbabwe	This Study
MSV-Mom	*Zea mays*	1998	Mombasa, Kenya	This Study
MSV-MtKA	*Zea mays*	1997	Mt. Kenya, Kenya	This Study
MSV-Ns	*Zea mays*	1984	Nigeria	Mullineaux *et al.*,1984
MSV-R2X2	*Zea mays*	1997	La Réunion	Isnard *et al.*, 1998
MSV-Sag	*Zea mays*	1998	Sagana, Kenya	This Study
MSV-VM	*Zea mays*	1993	Vaalharts, South Africa	Willment, 1999
MSV-Jam	*Digitaria* sp	1999	Jamaica, Kenya	This Study
MSV-Pat	*Digitaria* sp.	1999	Durban, South Africa	This Study
MSV-Raw	*Urochloa* sp.	1998	Rawsonville, South Africa	This Study
MSV-Set	*Setaria* sp.	1988	Durban, South Africa	Schnippenkoetter, 1998
MSV-Tas	*Triticum aestivum*	1991	Elseberg, South Africa	Willment, 1999
MSV-VW	*Triticum aestivum*	1993	Vaalharts, South Africa	Willment, 1999
DSV	*Digitaria sanguinalis*	1987	Vanuatu	Donson *et al.*, 1987
PanSV-Kar	*Panicum maximum*	1989	Karina, South Africa	Schnippenkoetter, 1998
SSV-Mil	*Pennisetum* sp.	1994	Kenya	Briddon *et al.*, 1996b
SSV-Reu	*Saccharum Officinarum*	1998	La Réunion	Bigarre *et al.*, 1999

4.2.3 Cloning of Full-length MSV Genomes

Replicative form (RF) DNAs of the isolates MSV-MatA, MSV-MatB, MSV-MatC, MSV-MtKA, MSV-Sag, MSV-Ama, MSV-Gat, MSV-MakD, MSV-Jam, MSV-Mom, MSV-Raw and MSV-Pat were isolated from dried leaf tissue as described by Palmer *et al.* (1997). RF DNAs were linearised using either *BamHI* (MSV-MatA, MSV-MatB, MSV-MatC, MSV-MtKA, MSV-Sag, MSV-Ama, MSV-Gat, MSV-MakD, MSV-Jam, and MSV-Pat) or *SalI* (MSV-Mom, MSV-Raw) and cloned, using standard techniques (Sambrook *et al.*, 1989), into similarly linearised pBluescriptSK+.

4.2.4 Sequencing of Full-length MSV Genomes

Sequencing of the cloned MSV isolates (MSV-MatA, MSV-MatB, MSV-MatC, MSV-MtK, MSV-Sag, MSV-Ama, MSV-Gat, MSV-Mom, MSV-MakD MSV-Pat, MSV-Jam and MSV-Raw) was carried out using an ALF Express automated sequencer (Pharmacia Corporation, Peapack, NJ). The full nucleotide sequences of all these isolates were determined in both orientations. Sequence assembly, alignment and phylogenetic analyses were carried out using DNAMAN

(version 4.0; Lynnon BioSoft, Quebec) and PHYLIP (version 3.57c; Felsenstein, Washington). Open reading frames (ORFs) and Intergenic Regions (IRs) within the sequenced genomes were inferred by analogy with those determined for MSV-Ns (see Chapter 1, Fig 1.2; Howell, 1984; Lazarowitz, 1988; Mullineaux *et al.*, 1984). Nucleotide sequence elements and amino acid sequence motifs and functional domains with either known or suspected functions within predicted gene products were also inferred by analogy with MSV-Ns and other geminiviruses (see Chapter 1.4.1; Arguello-Astorga *et al.*, 1994a; Boulton *et al.*, 1989a, 1993; Dekker *et al.*, 1991; Donson *et al.*, 1984; Fennol *et al.*, 1990; Gorbalenya and Koonin, 1989; Hofer *et al.*, 1992; Horvath *et al.*, 1998; Koonin and Ilyina; Liu *et al.*, 1997; Mullineaux *et al.*, 1990; Saurez-Lopez *et al.*, 1995; Wright *et al.*, 1997; Xie *et al.*, 1995, 1999).

4.2.5 Generation of Agroinfectious Constructs

Agroinfectious clones were constructed as described by Grimsley *et al.* (1987) according to the method of Palmer (1998). *XbaI – EcoRI* fragments containing full tandem repeats of MSV-MatA, MSV-MatB, MSV-MatC, MSV-MtK, MSV-Sag, MSV-Ama, MSV-Gat, and MSV-MakD genomes were cloned from pBluescriptSK into the *XbaI* and *EcoRI* sites of pBI121 to obtain pMatA, pMatB, pMatC, pMtK, pSag, pAma, pGat, and pMakD, respectively. *XbaI – XhoI* fragments containing full tandem repeats of MSV-Pat, MSV-Mom and MSV-Raw genomes were cloned from pBluescriptSK into the *XbaI – SalI* sites of pBIN19 to obtain pPat, pMom and pRaw, respectively. Agroinfectious clones were transformed into *A. tumefaciens* C58C1 using the method of An *et al.* (1988).

Besides the agroinfectious MSV clones constructed during this study, agroinfectious virus constructs were obtained from M. I. Boulton (MSV-Ns, DSV; John Innes Centre, Norwich, U.K.), M. Peterschmitt (MSV-R2X2; CIRAD, Montpellier, France), R. Briddon (SSV-Mil; John Innes Centre, Norwich, U.K.), W. Schnippenkoetter (MSV-Kom, MSV-Set, PanSV-Kar; University of Cape Town, Cape Town, South Africa) and J. Willment (MSV-VM, MSV-VW, MSV-Tas; University of Cape Town; Table 4.2).

4.2.6 Agroinoculation and Evaluation of Isolate Pathogenicity

Agroinoculation was performed according to Grimsley *et al.* (1987) with modifications described previously in Chapter 2. Agroinfectious constructs of all the MSV, DSV, SSV, and PanSV isolates involved in this study were used to infect three maize genotypes - B73 (highly susceptible), sweetcorn (moderately susceptible) and PAN6099 (moderately resistant). Agroinfection experiments involving each maize genotype-MSV isolate combination were repeated at least three times. Agroinfection experiments involving specific constructs were carried out on groups of 14 plants. For each agroinfection experiment, a group of 14 seedlings injected with sterile distilled water served as uninfected controls. Seedlings were grown under near uniform conditions in a plant growth room maintained at 22°C, at 80% relative humidity, with 16 hours of light per day.

Percentage leaf areas covered by chlorotic lesions in symptomatic plants were estimated for leaves four through six using a microcomputer-based image analysis technique described previously in Chapters 2 and 3. Percentage chlorotic areas of leaves four, five and six were assessed 22, 29 and 35 days after agroinoculation, respectively. For each virus isolate-plant genotype combination, the mean percentage chlorotic area of the 4th to 6th leaves, (designated C_{4-6}), were used as representative measures of chlorosis. For each MSV isolate, a single value designated C_{4-6}, representing the mean of the C_{4-6} values calculated for B73, sweetcorn and PAN6099, was used as a generalised measure of pathogenicity.

4.2.7 Detection of Recombinant Genomes

28 full-length African streak virus genomes were aligned in DNAMAN (optimal alignment with a gap open penalty of 5 and a gap extension penalty of 2). Detection of potentially recombinant sequences, identification of likely parent sequences and localisation of possible recombination breakpoints was carried out using computer software, called RDP (Recombination Detection Program), that was developed during this study (Appendix C; Installation files for the program are available on the CD ROM packaged with this thesis in the directory "D:/RDP/setup.exe").

RDP utilises a simple pairwise scanning approach for the detection of recombination. Beginning with a multiple sequence alignment in either DNAMAN, CLUSTAL (Higgins *et al.*, 1992),

PHYLIP, GCG (Genetics Computer Group, Wisconsin), or FASTA (Pearson and Lipman, 1988) formats, the software examines every possible combination of three sequences for evidence of recombination in a three-step procedure (Fig 4.1). Once the analysis is complete, an interactive graphical interface (see Appendix C, Figs C.1 to C.3) for examination of the analysis results enables the user to access stored information on all the evidence for recombination detected.

The RDP settings used were: multiple comparison correction off, internal reference selection and a window size of 10. The 28 full-length African streak virus sequences were analysed for evidence of inter-subtype, inter-strain and inter-species recombination in two sets, each using a different highest acceptable *P*-value setting. The first set contained 15 MSV-A sequences (MSV-Ns, MSV-MatA, MSV-MatB, MSV-MatC, MSV-MakD, MSV-Ama, MSV-Sag, MSV-Gat, MSV-MtKA, MSV-Ken, MSV-Kom, MSV-SA, MSV-VM, MSV-Reu, and MSV-R2) and was analysed with a highest acceptable *P*-value of 1×10^{-3}. The second contained 22 MSV sequences (the 15 MSV-A sequences, MSV-Tas, MSV-Mom, MSV-VW, MSV-Jam, MSV-Pat, MSV-Set, MSV-Raw, PanSV-Ken, PanSV-Kar, SSV-Asw, SSV-Ben, SSV-Reu, SSV-N; GenBank accession numbers for all nucleotide sequences not determined in this study are given in Table A.1 in Appendix A) and was analysed with a highest acceptable *P*-value of 1×10^{-4} (Fig 4.1).

Neighbour joining trees (Saitou and Nei, 1987) were constructed using regions within the aligned genomes that RDP identified as having a potentially recombinant origin and in all cases bootstrap analysis was used to confirm the program's findings. To facilitate this analysis, RDP allows the portion of sequences within the input alignment that are encompassed by potential breakpoints to be saved to disk. Sequences within all regions of the alignment that corresponded with sets of potential recombination breakpoints were realigned in DNAMAN (to ensure that the detected recombinant regions were not the result of an alignment error) and were used to construct a series of neighbour joining trees, each with 1000 bootstrap iterations. Acceptable confirmation that a sequence was of recombinant origin was "movement" with over 70% bootstrap support of the sequence from its position in a tree constructed with full-length sequences to a divergent branch of a tree constructed using only the potentially recombinant region within the alignment.

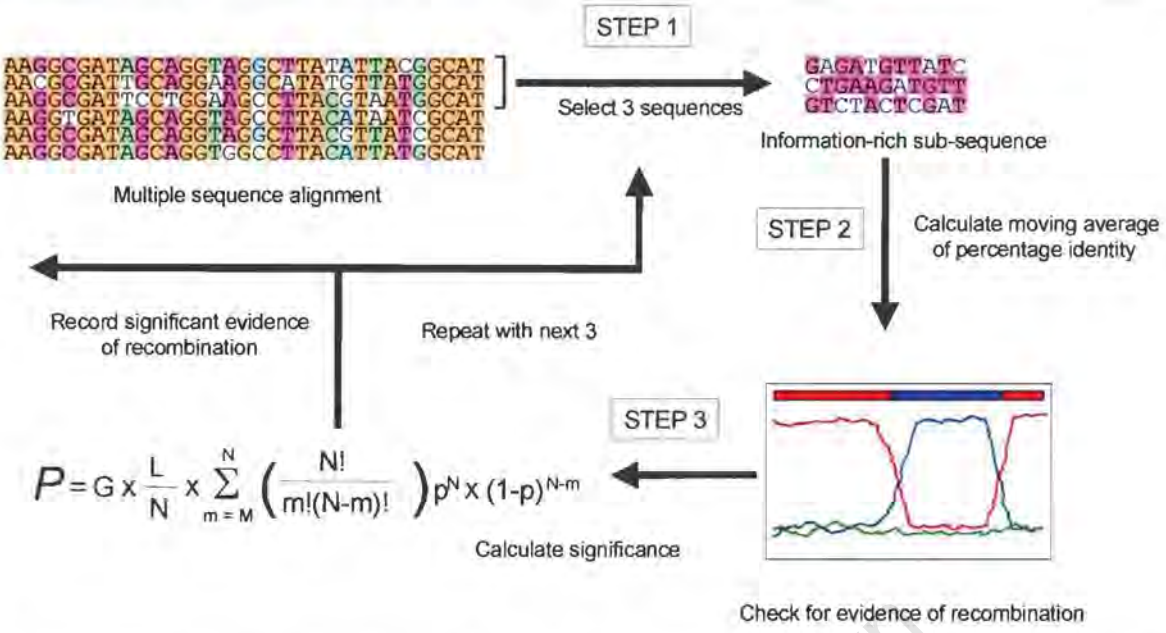


FIGURE 4.1. The three step analysis procedure performed by RDP (Recombination Detection Program). In the first step all phylogenetically non-informative sites are discarded from the group of three sequences to obtain three information-rich sub-sequences. In every group of three sequences there are two sequences, A and B, that are more closely related to one another than to a third sequence, C. Non-informative sites are (1) identical in all three sequences, (2) different in all three sequences, and (3) unique to A or B and are not present in any member of a group of reference sequences. The method of reference sequence selection is user-defined and is based on the relative positions of the three selected sequences in an UPGMA dendrogram. In the second analysis step, a window of user-defined width is moved along the aligned sub-sequences one nucleotide at a time and an average percentage identity for each of the three possible sequence pairs is calculated at each position. Sequences of possibly recombinant origin are identified as regions where the percentage identities of sequences A and C or B and C are higher than for sequences A and B. In the third analysis step, the probability that the nucleotide arrangement in the identified region that results in sequences A or B appearing more closely related to sequence C, may have occurred by chance, is approximated using the binomial distribution adapted from Rice (1995). G is the number of possible combinations of three sequences, L is the length of the information rich subsequences, N is the length of the putatively recombinant region, M is the number of nucleotides in common between either A or B and C in the putatively recombinant region, and p is the proportion of nucleotides in common between either A or B and C in the entire subsequence. If the value of *P* is lower than a user-definable cut-off figure, information on the potential recombination is stored for later access before the next combination of three sequences is selected and analysed.

4.2.8 Recent MSV Evolution Through Nucleotide Sequence Substitution

29 full-length African streak virus genome sequences (including the 12 sequenced during this study), DSV and CSMV sequences were used to infer the nucleotide sequences of ancestral African streak and maize streak viruses. The 31 full-length sequences were aligned in DNAMAN and were used to construct phylogenetic trees by the neighbour joining (using DNAMAN) and maximum parsimony (using PHYLIP) methods. A most parsimonious tree topology corresponding to that obtained using the neighbour joining method was used during the evolutionary reconstruction. The sequences of all nodes in the phylogenetic tree were inferred using DNAPARS (a component of PHYLIP). Where necessary, inferred sequences were edited

by replacing sites with completely unknown identities (ie those sites that could potentially contain a space or a nucleotide) with a nucleotide (N) to force genes into frame.

Five genomic regions corresponding to the MP gene, CP gene, RepA and C2 ORFs, LIR and SIR of MSV-Ns were examined separately. The changes that had occurred in each of these regions in MSV-MatA, -Kom, -MatC, -Reu, -Tas, -VW, -Jam, -Set, -Raw, and -Pat between each of the nodes separating them from their most recent common MSV progenitors were calculated. This was achieved by aligning actual sequences with inferred nodal progenitors and counting the number of nucleotide substitutions that occurred between them. During all pairwise comparisons, insertions and deletions (regardless of size), were considered a single nucleotide substitution and ambiguous nucleotides were ignored. Portions of sequence that had previously been found to have a recombinant origin were also ignored. For direct comparisons of evolution rates within the five selected genomic regions, the number of nucleotide substitutions occurring within a region was expressed as a percentage of the total number of nucleotides of non-recombinant origin within that region.

Because in certain cases (particularly in the depths of the phylogenetic tree) one or both of the branches diverging from a node contain more than one sequence, it was necessary to calculate an average measure of divergence for the virus genomes on each of the branches. To avoid biasing the average due to an over-representation of closely related sequences (such as MSV- MatA, -Kom -MatC, and -Reu), the average divergence of those branches containing more than two sequences was calculated as follows: (1) the percentage of nucleotide substitutions occurring between a nodal sequence and all its descendants (not including inferred sequences) was determined as described above; (2) the average percentage was calculated for all most closely related sequence pairs on a branch; (3) these pairs were then replaced by a symbolic sequence with the calculated average divergence of the pair representing its divergence from the ancestral sequence; (4) the process was then repeated with the next most closely related sequence pairs until finally only a single symbolic sequence and average percentage nucleotide substitution was obtained. An example of this scheme in practice is the calculation of the average nucleotide sequence divergence of the MSV-A and -B isolates occurring on the branch that diverges from the common ancestral MSV sequence. The average percentage nucleotide substitution occurring between this ancestral sequence and the sequence pairs MSV-MatA/Kom (A1) and MSV-Tas/VW (B1) was calculated as a1 and b1, respectively. The averages of the next two pairs,

A1/MSV-MatC (A2) and B1/MSV-Jam (B2) were calculated as a2 and b2, respectively. The average percentage of substitutions for A2/MSV-Reu (A3) was calculated as a3. The average percentage nucleotide substitution occurring on the entire branch was the average a3 and b2.

4.3 RESULTS

4.3.1 Full Genome Sequencing and Classification of MSV Isolates

To enable the classification of all the MSV isolates characterised in this and other studies a group of isolates that represented all of the observable MSV diversity was selected for full genome sequencing. The 49 MSV isolates typed using the PCR-RFLP technique developed by Willment (1999) could be divided into 4 major groups. While the largest group consisting of 37 isolates was obtained almost exclusively from maize (one sample, MSV-DaiC, was isolated from a *Digitaria* sp.), another large group consisting of 10 isolates was obtained from a range of wild grass species and rye. Two isolates, MSV-Raw and MSV-Pat, had unique RFLPs for six of the seven enzymes used and were each assigned to their own group. To determine the actual nucleotide sequence relationships between the viruses in the different groups, eight isolates from the large "maize group" (MSV-MatA, MSV-MatB, MSV-MatC, MSV-Sag, MSV-Ama, MSV-Gat, MSV-MtKA, and MSV-MakD), two isolates from the large "grass group" (MSV-Mom and MSV-Jam), MSV-Raw, and MSV-Pat were cloned and fully sequenced. The isolates selected for cloning and sequencing each represented a group of isolates with either identical or nearly identical RFLPs (nearly identical isolates differed by only one RFLP). Because the PCR-RFLP typing technique is able to differentiate between MSV isolates that share up to 99% nucleotide sequence identity (Willment, 1999), the isolates selected within the maize group were in most cases expected to share between 95 and 99% sequence identity but be representative of distinct, never-before sequenced variants within the group. The Zimbabwean isolate MSV-MatB had RFLPs that were identical to those detected for the Kenyan isolate MSV-Sag, and both were selected for sequencing to confirm their relatedness.

A pairwise nucleotide sequence identity matrix was constructed using the 12 sequences determined in this study and 10 other full-length MSV sequences currently available in GenBank (Fig 4.2). To classify the sequenced isolates, a strain was arbitrarily defined as a group of isolates sharing greater than 94% nucleotide sequence identity. Based on this definition, the 22 sequences

assessed contain representatives of five MSV strains (Fig 4.2). Three of these have already been identified as MSV-A (the maize-type MSV strain), MSV-B (the grass/wheat type MSV strain) and MSV-C (the strain represented by MSV-Set; Willment, 1999). In keeping with this MSV naming convention, the two new strains, represented by MSV-Raw and MSV-Pat, were named MSV-D and MSV-E respectively.

To further classify the large group of MSV-A isolates, a subtype was arbitrarily defined as a group of isolates sharing greater than 98% sequence identity. By this definition the MSV-A sequences currently available could be divided into six subtypes (Fig 4.2) named MSV-A_1 (containing MSV-MatA, MSV-Ama, MSV-MatB, MSV-Sag, and MSV-Gat), MSV-A_2 (containing MSV-Ns), MSV-A_3 (containing MSV-Ken and MSV-MtKA), MSV-A_4 (containing MSV-SA, MSV-Kom, and MSV-VM), MSV-A_5 (containing MSV-MatC and MSV-MakD) and MSV-A_6 (containing MSV-Reu and MSV-R2). These subtype groupings were supported by phylogenetic analysis of the sequences (Fig 4.3). It is notable that MSV-MatC and MSV-MakD were classified as being of the same subtype despite their sharing only 97.9% sequence identity. The reason for this is that MSV-MakD is probably a recombinant (see below for details) with ~75% of its genome sharing greater than 98% sequence identity with MSV-MatC.

PCR-RFLP typed isolates from both this study and that carried out by Willment (1999) were assigned to different strains and subtypes by comparing their RFLPs to those of the sequenced isolates belonging to the different groupings. Similarly, isolates previously typed by partial nucleotide sequence analysis (Briddon *et al.*, 1994; Rybicki *et al.*, 1998) were also classified based on their relatedness to the fully sequenced genomes. Collectively, 88 MSV isolates were typed and used to assess both the prevalence of different MSV strains in non-maize hosts and the distribution of different MSV-A subtypes in maize throughout Africa (Fig 4.4; Table 4.1).

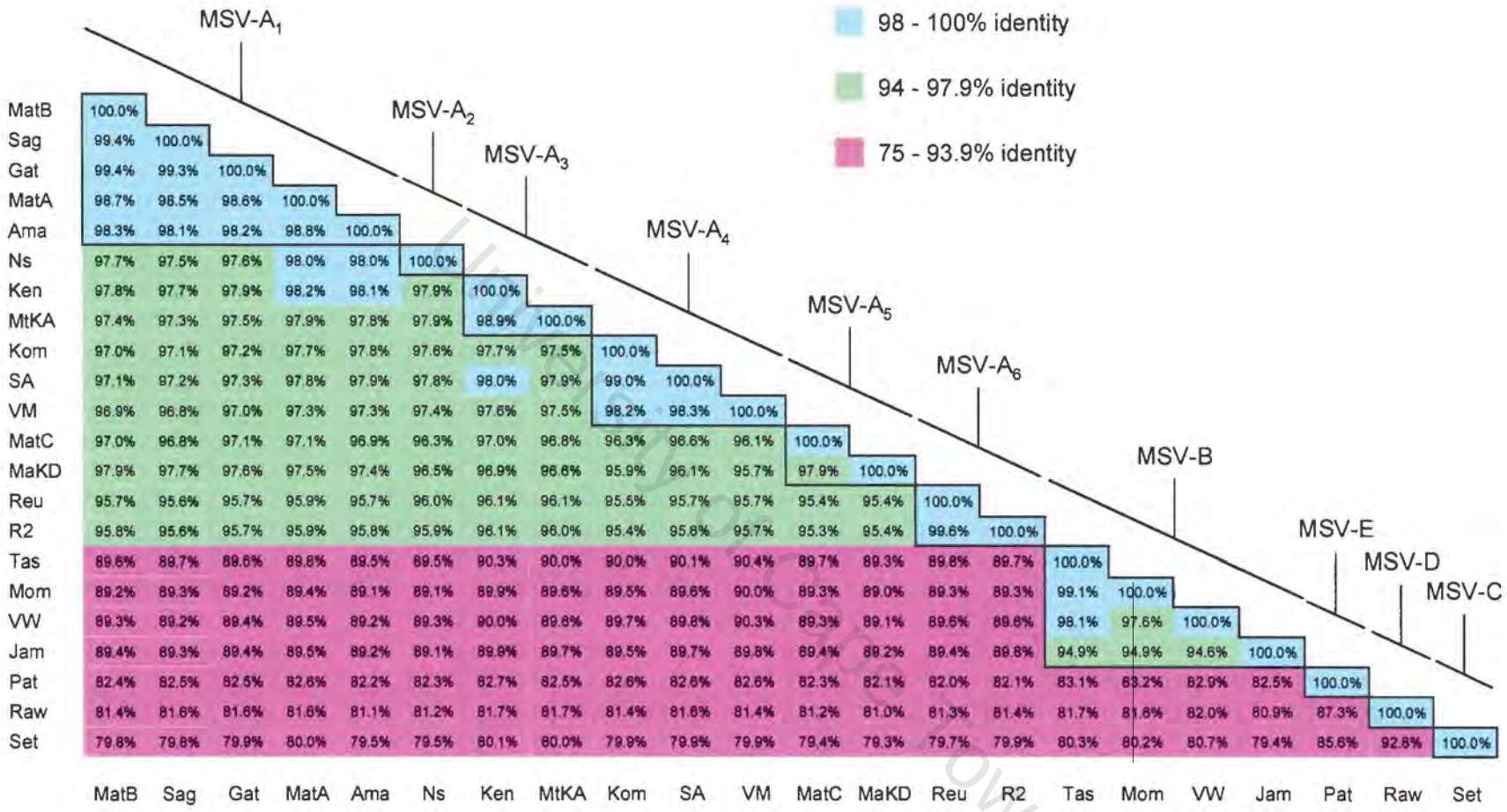


FIGURE 4.2 Percentage identity matrix of 22 currently available full length MSV genomic sequence. Sequences were classified into 5 MSV strains (a strain was arbitrarily defined as a group of sequences sharing greater than 94% sequence identity) and the isolates in strain A were grouped into 6 subtypes (a subtype was arbitrarily defined as a group of isolates sharing greater than 98% sequence identity). MatC and MakD were grouped into subtype A_6 for reasons described in the text. The Matrix was derived from an optimal DNAMAN alignment of the 22 sequences (Gap open penalty of 5 and gap extension penalty of 2).

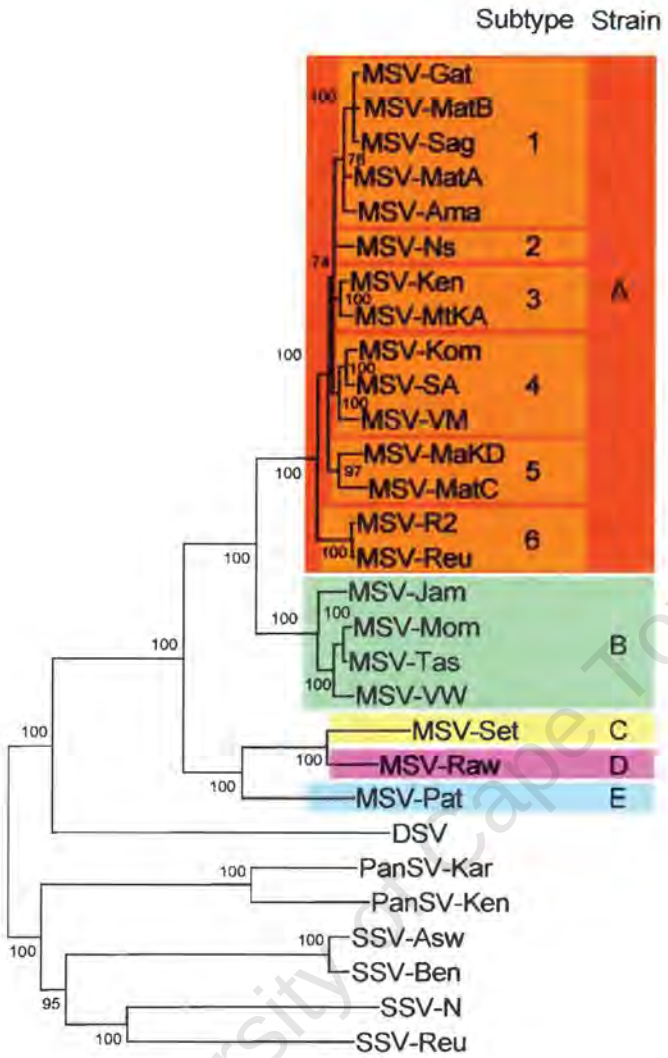


FIGURE 4.3 A rooted neighbour joining tree (Saito and Nei, 1987) indicating the phylogenetic relationships between the MSV genomes sequenced in this study and all other publicly accessible full-length African streak virus genomic sequences. GenBank accession numbers for the sequences not determined in this study can be found in Appendix A. The MSV sequences can be divided into five major strain groupings (designated A-E). Whereas all the MSV-A isolates (in orange) were obtained from MSV-infected maize, the MSV-B isolates (in green) were predominantly obtained from non-maize hosts (Wheat and *Digitaria* sp.) with only one isolate from maize (MSV-Mom). The MSV-C,-D, and -E isolates (in yellow, pink, and blue, respectively) were all obtained from wild grass species. The MSV-A isolates could be subdivided into six subtypes, each with different degrees of geographical localisation. The tree was constructed from an optimal multiple alignment of the sequences in DNAMAN (gap open and extension penalties of 5 and 2, respectively) and rooted on the sequence of the mastrevirus, *Chloris striate mosaic virus* (not shown). Numbers associated with nodes represent the percentage of 1000 bootstrap iterations supporting the nodes. Nodes with less than 60% bootstrap support were collapsed. Whereas vertical distances are arbitrary, horizontal distances reflect relationships between the sequences.

While it is possible that certain isolates typed by the PCR-RFLP technique may have been misclassified due to chance correspondence of their RFLPs with those of a fully sequenced member of a MSV-A subtype, only a small proportion of these fully or partially sequenced MSV-A isolates examined using the technique would have either been assigned to an incorrect subtype (1/26) or been unclassifiable (2/26) had a portion of their sequence not also been known. Based on the RFLPs, MSV-U would have been assigned to MSV-A_2 whereas its CP and SIR sequence indicates it is actually an A_1 isolate. Because MSV-GB and MSV-NB differed by at least two RFLPs from those of the representative MSV-A subtype sequences, they would have been unclassifiable had their CP and SIR sequences not been known.

4.3.2 Geographical and Host Distribution of MSV Strains and Subtypes

MSV-B was both the only detectable strain found in cultivated species such as wheat and rye, and the most commonly detected strain in wild grasses (70%; Fig . 4.4). Although strain B viruses were also detected in maize, they represented only 3% of all isolates detected in this host. Whereas MSV-A subtypes accounted for 16% of the isolates detected amongst wild grasses (not one was from wheat or rye), strain C, D, and E viruses are apparently quite rare and collectively comprised only 12% of isolates obtained from non-maize hosts.

The overwhelming majority of MSVs obtained from maize were strain A isolates (Fig 4.4). There appear to be substantial differences in the subtype composition of the MSV-A populations infecting maize in different parts of Africa (Fig 4.4). While subtype A_1 isolates occur throughout Africa from Nigeria to South Africa, subtype A_2, A_3, and A_4 MSVs are apparently restricted to Western, Eastern, and Southern Africa, respectively (Fig 4.4). Only subtype A_6 isolates have been detected in MSV- infected maize on the Indian Ocean islands of La Réunion and Mauritius (Briddon *et al.*, 1994) and no MSV-A_6 isolates have yet been detected on the African continent.

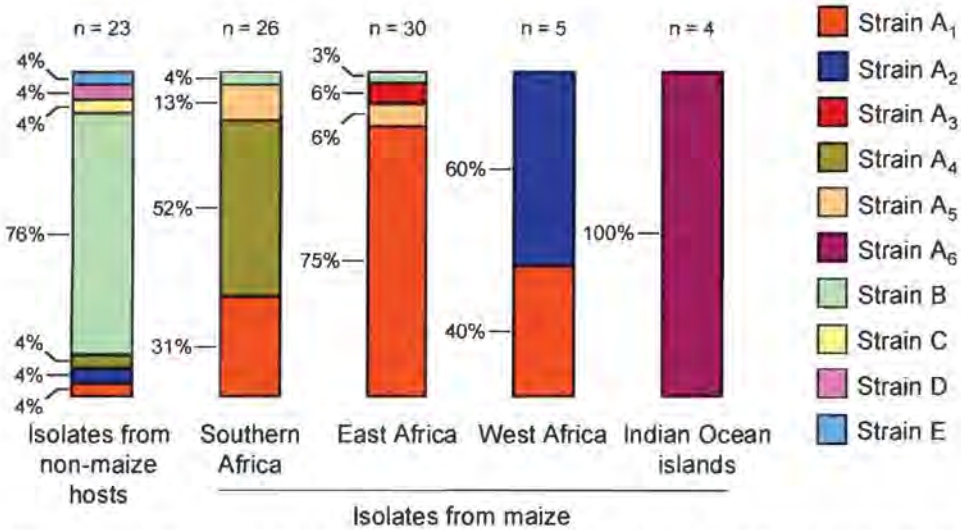


FIGURE 4.4 The geographical and host distribution of 88 MSV isolates that have currently been typed in this and other MSV diversity studies (Briddon *et al.*, 1994; Rybicki *et al.*, 1998; Willment, 1999). While 56% of the isolates used to construct this figure were typed exclusively by PCR-RFLP analysis, the rest were typed by partial and full-length genomic sequencing (see Table 4.1). Numbers above bars (n) indicate the number of isolates used to construct the bars. The vast majority of isolates detected in streak-infected non-maize hosts belong to the B strain. While a few MSV-A isolates have also been detected in non-maize hosts, only single isolates each of MSVC, -D, and -E have been isolated from wild grasses. While MSV-A_1 is the most prevalent MSV-A subtype found in maize throughout the African continent, it is only the predominant MSV-A subtype in East Africa (Kenya, Malawi, Ethopia, Uganda and Zaire) . The predominant subtypes in Southern (South Africa, Namibia, Zimbabwe, and Mozambique) and West Africa (Nigeria, Ghana, and Cameroon) are MSV-A_4 and MSV-A_2, respectively. While MSV-A_5 and MSV-A_3 are relatively minor subtypes, MSV-A_5 isolates have a relatively wide geographical distribution and have been found in both Southern and East Africa. The MSV-A_6 subtype has only been found on the Indian Ocean islands of La Reunion and Mauritius. MSV-B isolates are the only non MSV-A isolates that have been found in maize.

4.3.3 Analysis of the MSV Sequences

Examination of the 12 full MSV genomic sequences determined in this study indicated that they have features in common with all other previously published MSV sequences (Howell, 1984; Lazarowitz, 1988; Mullineaux *et al.*, 1984). Intergenic regions, ORFs, introns, primer binding sites, regulatory elements and potential secondary structures were identified in all 12 sequences by analogy with those determined for MSV-Ns (Arguello-Astorga *et al.*, 1994a; Boulton *et al.*, 1989a; Dekker *et al.*, 1991; Donson *et al.*, 1984; Fennol *et al.*, 1990; Mullineaux *et al.*, 1990; Saurez-Lopez *et al.*, 1995; Wright *et al.*, 1997) and MSV-SA (Lazarowitz, 1988) and are presented in Fig D.1 (in Appendix D). The predicted MP, CP, Rep and RepA amino acid sequences encoded by these 12 genomes closely resemble those determined for other MSV isolates (Figs D.2, D.3, D.4 and D.5 respectively in Appendix D). Also indicated in these figures are all known or suspected functional domains and sequence motifs inferred by analogy with those identified in MSV-Ns and other mastreviruses (Boulton *et al.*, 1993; Gorbalenya and Koonin,

1989; Hofer *et al.*, 1992; Horvath *et al.*, 1998; Koonin and Ilyina; Liu *et al.*, 1997; Xie *et al.*, 1995, 1999).

There were a few minor differences in potentially important sequence elements between certain of the sequences determined in this study and those published elsewhere. Most of these differences occur within sequences of the LIR that are involved in (+) strand replication and transcription of virion and complementary sense (V- and C-sense, respectively) genes.

The nucleotide sequence of the stem portion of the stem-loop structure at the (+) strand origin of replication varies slightly between the different MSV strains. The most notable variation is that the strain B isolates (MSV-Mom and MSV-Jam) contain a single nucleotide mismatch that may slightly destabilise the stem-loop structure. The mismatch has apparently resulted from a C to T substitution 5 nucleotides from the base of the stem downstream from the origin of replication (CGCGCC**TTCTTTT**CCTGC for MSV-A isolates to CGCGCC**TTCTTTT**TCTGC for MSV-B isolates; underlined sequence; Fig D.1). The most variable region of the stem is the A/T rich sequence between 6 and 12 nucleotides from its base (bold sequence above). In both MSV-Raw and MSV-Pat (Strain D and E isolates, respectively), single nucleotides have been deleted (at different positions in each of the viruses) within these sequences in both strands of the stem. Two paired nucleotide substitutions in MSV-Set (a strain C isolate) and one paired substitution in MSV-Raw have also occurred within this A/T rich sequence on both sides of the stem. The stem sequence of MSV-Raw also contains one unpaired nucleotide immediately downstream from the AC residues at the origin of (+) strand replication.

While the iterated stem sequence (called an iteron) that has been found to occur between the stem loop and RepA start codon in all mastreviruses (Arguello Astorga *et al.*, 1994a) is eight nucleotides long in the MSV-A,-B and E isolates, it is only six nucleotides long in MSV-Set (an MSV-C isolate; Schnippenkoetter, 1998) and MSV-Raw (an MSV-D isolate). Because of the nucleotide mismatch at the top of the stem sequence in MSV-Raw, the direct repeat of the iteron is not completed in the stem downstream from the (+) strand origin whereas the inverted repeat extends one nucleotide into the loop.

It has been noted that MSV-Ns, MSV-SA and MSV-Ken (MSV-A_2 -A_3 and A_4 isolates, respectively) all possess 3 TATA boxes 57, 62 and 101 nucleotides upstream from the RepA start codon (Boulton *et al.*, 1991a; Palmer and Rybicki, 1998). Of the three TATA boxes, the one at position –101 is the only one that is conserved in all the MSV strains. The TATA box at position –57 occurs in all currently sequenced MSV isolates except MSV-Set and -Pat (strain C and E isolates respectively), while the TATA box at position –62 only occurs in MSV-A_1, -A_2 -A_3, and -A_4 isolates. C-sense transcription in MSV-Ns is predominantly directed by the TATA box at position –101 but can also be directed from one or both of the other TATA boxes (Boulton *et al.*, 1991a; M. I. Boulton, 1998, unpublished results cited by Palmer and Rybicki, 1998). The TATA box at position –101 may also be involved in V-sense transcription in MSV-Ns (Wright *et al.*, 1997).

A GC-box sequence has been identified as an important element in the V-sense promoter of MSV (see Chapter 1, Fig 1.6; Fenoll *et al*, 1990) and analogous sequences were found to occur in the same position at the base of the stem loop in all of the MSV sequences determined in this and other studies (Schnippenkoetter, 1998; Willment, 1999). In all MSV isolates except MSV-Pat and MSV-MatB (strain E and A_1 isolates, respectively), the GC box contains a direct repeat of the sequence GGGCCGG. While the second repeat of this sequence is interrupted in MSV-MatB by substitution of the first C with a T, it is interrupted after the second C by a nine-nucleotide GC rich insertion in MSV-Pat.

Apart from variations in sequence elements within the LIR, another potentially significant difference between the MSV strains is the size of the Rep proteins they express. All MSV-A subtypes (except subtype A_4) potentially express Rep proteins that have 5 more amino acids at their C-termini than the Rep proteins expressed by all of the other MSV strains (Fig D.4 in Appendix D). It appears as though a nucleotide deletion within the Rep termination codon of a MSV-A progenitor has resulted in the Rep gene of most MSV-A isolates terminating 15 nucleotides downstream from the point where termination occurs in other MSV strains (see Appendix D, Fig D.1).

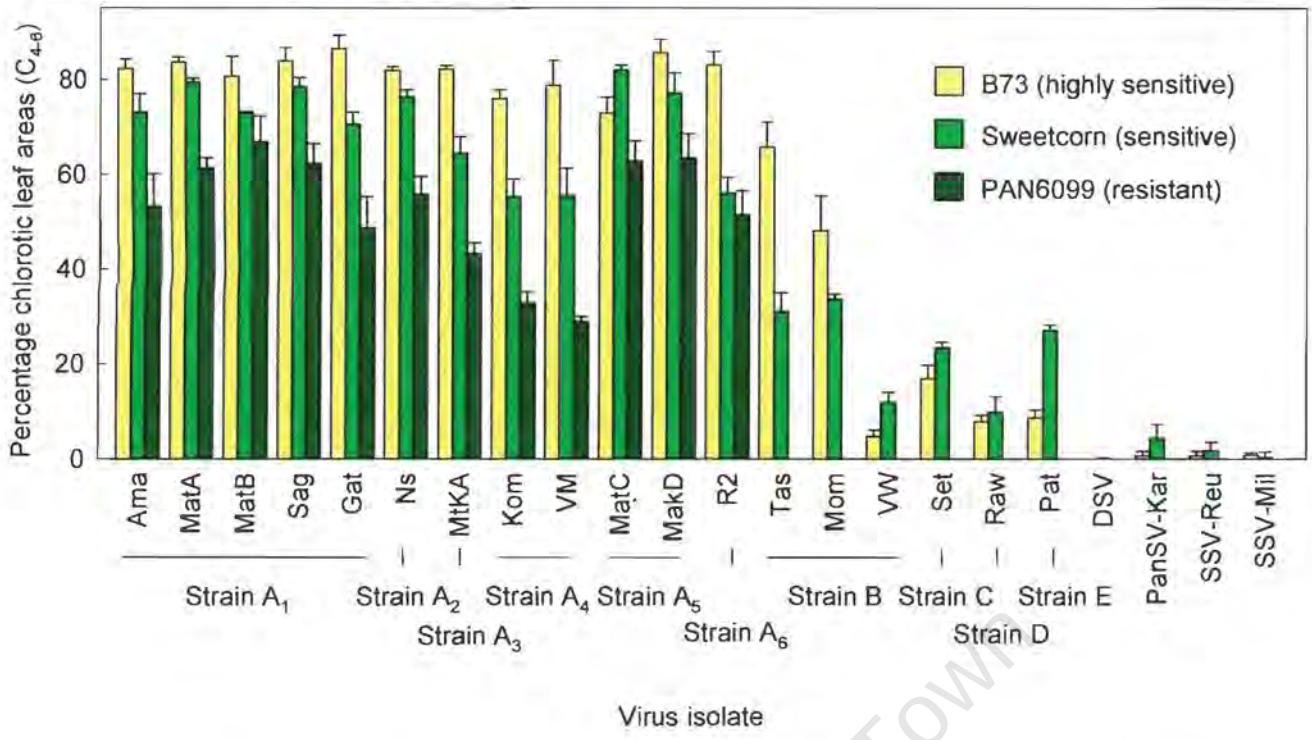


FIGURE 4.5 Average percentage chlorotic leaf areas occurring between leaves 4 and 6 (C_{4-6}) of different maize genotypes agroinfected with a range of MSV, PanSV, SSV and DSV isolates. The MSV isolates have been labeled according to their strain and subtype groupings. Maize genotypes were either highly sensitive (B73, yellow bars), moderately sensitive (sweetcorn, light green bars) or moderately resistant (PAN6099, dark green bars) to MSV infection. Error bars represent 95% confidence intervals.

4.3.4 Pathogenicity of Virus Isolates In *Zea mays*

In order to determine the relationship between the genotypes of streak virus isolates and their pathogenicity in maize, differentially MSV-resistant maize genotypes were agroinoculated with 21 fully sequenced MSV, PanSV, DSV and SSV isolates. MSV Strain A isolates produced the most severe symptoms in all the maize genotypes tested (Fig 4.5). While all these isolates share greater than 95% nucleotide sequence identity, there are significant differences in their pathogenicity in maize. Whereas subtype A_1 (MSV-Ama, -Gat, -MatA, -MatB, and -Sag), A_2 (MSV-Ns) and A_5 (MSV-MakD and -MatC) isolates produced the severest symptoms in all the maize genotypes (C_{4-6} of 68.6 - 75.5), subtype A_4 isolates (MSV-Kom and -VM) produced the least severe symptoms (C_{4-6} of 54.0 - 54.4). Although subtypes A_3 (MSV-MtKA) and A_6 (MSV-R2) isolates generally produced intermediate symptoms in all the maize genotypes (C_{4-6} of 62.8 - 63.2), the subtype A_6 isolate produced symptoms as severe as the subtype A_1, A_2 and A_5 isolates in the MSV-resistant maize genotype, PAN6099 (Fig 4.5).

The MSV-B, -C, -D and -E isolates, obtained predominantly from wild grasses and wheat, were all substantially less severe (C_{4-6} of 32.3 – 5.5) than the MSV-A isolates in all the maize genotypes. The single maize-derived strain B isolate, MSV-Mom, was significantly less severe in maize than the maize-derived strain A isolates and produced symptoms very similar to the wheat-derived strain B isolate, MSV-Tas (Fig 4.5). While no strain B, C, D, or E isolates produced any symptoms in the moderately resistant maize genotype, PAN6099, all except the strain B isolates MSV-Tas and -Mom produced significantly more severe symptoms in the moderately sensitive maize genotype, sweetcorn, than in the very sensitive genotype, B73.

The PanSV, DSV and SSV isolates produced the mildest symptoms in maize (C_{4-6} of 0.06 to 1.69). As with all non-strain A MSV isolates, none of these viruses produced any symptoms in PAN6099. PanSV, DSV and SSV isolates also produced slightly more severe symptoms in sweetcorn than in B73. Whereas the PanSV isolate was slightly more severe than the rest of these viruses (particularly in sweetcorn), DSV was by far the mildest and was unable to produce any symptoms on B73.

4.3.5 Detection of Recombination Between Virus Isolates

Because recombination has potentially resulted in the recent emergence of a number of serious begomovirus diseases (Umaharan *et al.*, 1998; Zhou *et al.*, 1997; Zhou *et al.*, 1998), the prevalence of recombination amongst MSVs was investigated. Using a custom-written recombination detection program (called RDP), substantial evidence was found for recombination both between different MSV isolates and between MSV isolates and the other African streak viruses (Fig 4.6). Of 16 potentially unique recombination events detected amongst the 28 African streak virus sequences analysed by RDP, nine were confirmed by bootstrap analysis. The confirmed recombinant regions ranged in size from 23 to 501 nucleotides with inter-species recombination apparently involving smaller regions (23 - 55 nucleotides) than intra- species recombination (56 - 501 nucleotides).

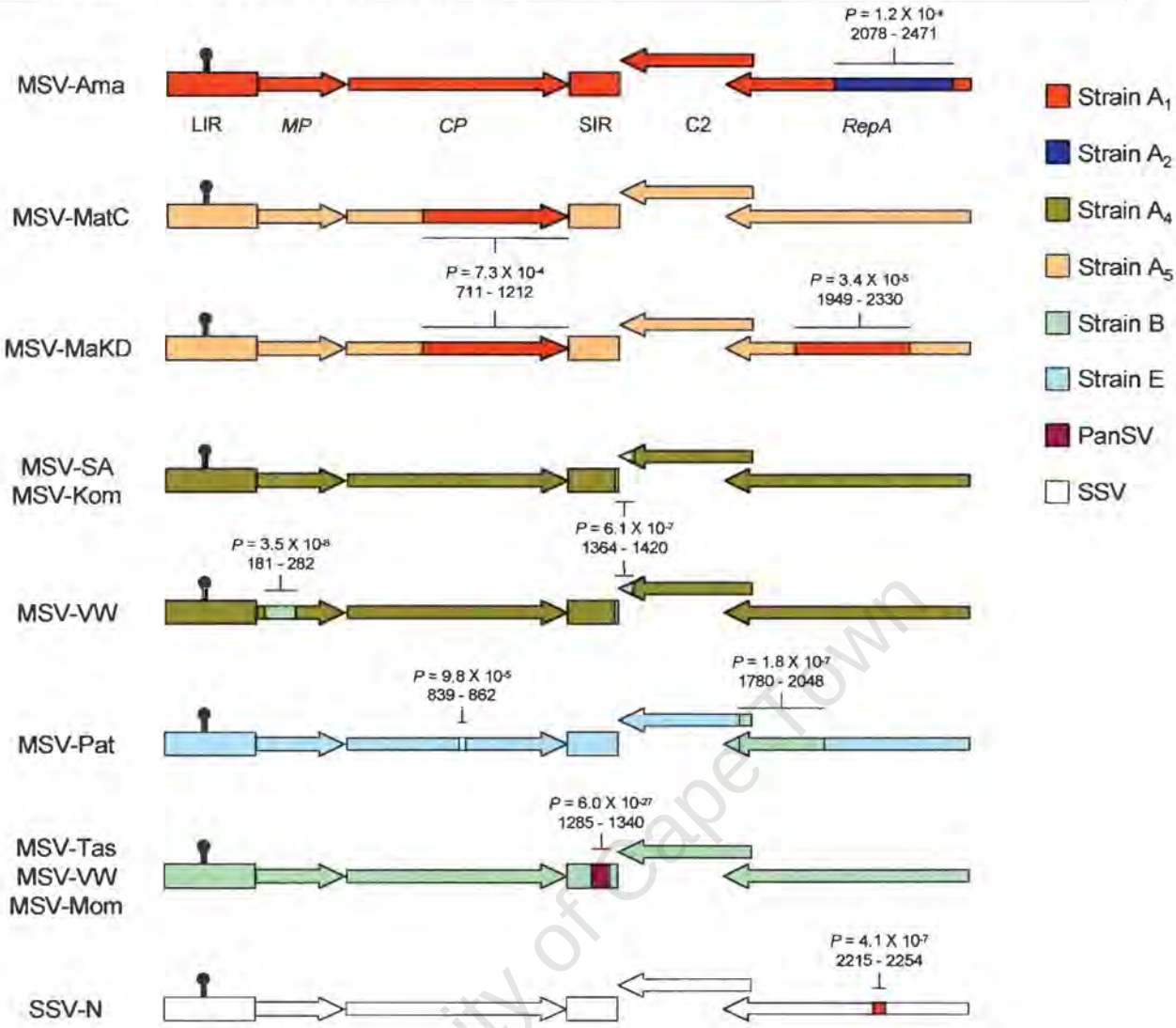


FIGURE 4.6 A schematic representation of the recombinant regions detected in the African streak virus genomes analysed during this study. Indicated are the major genomic features of the virus isolates (labeled for MSV-Ama). The origin of different genomic regions is indicated by their colour. Note that while the recombinant region detected within the *RepA* of SSV-N is presented here as being MSV-A_1 - like, it could be more accurately described as being MSV-like. Potential recombination breakpoints and the probability (*P* ;using the binomial distribution) that the indicated regions do not have a recombinant origin are presented.

Of the nine unique recombination events detected, six were between MSV isolates. Four of these were within *RepA* and/or the C2 ORF. Recombination amongst the MSV-A isolates occurred in *RepA* between subtype A_1 and A_2 viruses and between subtype A_1 and A_5 viruses (Fig 4.6). A portion of the *RepA* sequence of the MSV-A_1 isolate, MSV-Ama, resembles that found in the subtype A_2 isolate, MSV-Ns. The potentially recombinant region spans sequences encoding the three rolling circle replication (RCR) motifs and putative DNA recognition and binding domains of Rep and RepA (see Chapter 1, Fig 1.8). The portion of *RepA* in the MSV-A_5 isolate, MSV-MakD, that is potentially of MSV-A_1 origin includes sequences encoding the RCR motif III and the oligomerisation domain of Rep and RepA (see Chapter 1, Fig 1.8; Horvath *et al.*, 1998;

Koonin and Ilyana, 1993). Approximately 70% of the 3' end of the CP genes of both the MSV-A_5 isolates that were examined (MSV-MakD and -MatC) have apparently also been derived through recombination with a subtype A_1 virus. The potentially recombinant region contains sequences encoding the entire C terminal portion of the CP and includes part of the ss/ds DNA binding domain (see Chapter 1, Fig 1.9; Liu *et al.*, 1997).

Recombination between strain A and strain B isolates was only detected amongst subtype A_4 isolates (MSV-Kom, -SA, and -VM). One of these MSV- A_4 isolates (MSV-VM) has two genomic regions that are probably of MSV-B origin. All the subtype A_4 sequences examined contained strain B-like sequences at the 3' termini of their C2 ORFs (Fig 4.6). Interestingly, recombination in this region may have been responsible for shifting the C2 termination codon slightly upstream of its position in other MSV-A isolates. Subtype A_4 viruses would therefore be predicted to express Rep proteins with five fewer amino acids at their C-termini than viruses belonging to the other MSV-A subtypes (see Appendix D, Fig D.4).

Besides the recombinant region detected at the 3' end of the C2 ORF, the 5' portion of the MP gene of the subtype A_4 isolate, MSV-VM, also has a probable MSV-B origin (Fig 4.6). The potential recombinant region includes part of the *MP* intron and encodes a portion of the putative MP trans-membrane domain (see Chapter 1, Fig 1.9; Boulton *et al.*, 1993).

The only other evidence of inter-strain recombination detected was between the strain E isolate, MSV-Pat, and a MSV-B isolate. A 268 nucleotide region at the 3' end of *RepA* and spanning the start of the C2 ORF of MSV-Pat has a striking resemblance to the corresponding sequences found in MSV-B isolates (Fig 4.6). The potentially recombinant region contains the *Rep* intron and sequences encoding the retinoblastoma (Rb) binding motifs and oligomerisation domains of Rep and RepA, and the suspected transactivation domain of RepA (see Chapter 1, Fig 1.8; Horvath *et al.*, 1998; Xie *et al.*, 1995).

Strong evidence of three recombination events between different African streak virus species was also detected (Fig 4.6). While a 23 nucleotide portion of the MSV-Pat CP gene with no identified functionality appears to have an SSV origin (it is most closely related to the sequence found in SSV-Ben), a 55 nucleotide region of the SIR within the putative primer binding site (see Chapter 1, Fig 1.7, Donson *et al.*, 1984) of MSV-VW, -Tas, and -Mom has an apparent PanSV origin.

The finding that SSV-N contains 39 nucleotides of MSV-like sequence within its RepA gene provides evidence that MSV sequences have also been transferred to other African streak virus species. It is noteworthy that this very small recombinant region completely encompasses the sequence encoding the Rep motif involved in the nicking and joining activities during RCR (see Chapter 1, RCR motif III in Fig 1.8; Koonin and Ilyina, 1993; Laufs *et al.*, 1995b).

4.3.6 Recent MSV Evolution Through Nucleotide Substitution

The many new full-length MSV sequences determined in this study presented an opportunity to investigate recent events in the evolution of this species. DNAPARS (version 3.5c; a component of PHYLIP) was used to infer ancestral MSV nucleotide sequences from an alignment of all currently available African streak virus sequences, DSV and CSMV. A single most parsimonious tree topology which corresponded to that obtained using the neighbour joining method (Fig 4.3; Saito and Nei, 1987) was used to infer the changes in nucleotide sequence that occurred between the nodes of the tree. Sequences thus inferred were checked for their capacity to encode appropriate gene products. The quality of the inferred sequences occurring at each of the nodes in the dendrogram presented in Fig 4.7 is described in Table 4.3. The evolution of *MP, CP, Rep* LIR and SIR sequences were individually examined.

The relative rates at which these genomic features have evolved have apparently varied quite substantially in different lineages. From the common MSV ancestor (MSV-1 in Fig 4.7), it appears that the LIR has evolved at a greater rate than any of the other genomic components (compare the L columns to the others in the graph at the MSV-1 node). Whereas the average percentage nucleotide substitutions that have occurred in the LIRs of all currently sequenced MSV isolates since their evolution from MSV1 has been 11.4%, in these same viruses an average of only 8.2%, 7.2%, 8.7%, and 6.9% of nucleotides have been substituted in MP genes, CP genes, SIRs and Rep genes, respectively. The highest rate of LIR evolution has occurred in the lineage diverging from MSV-1 that has given rise to MSV-Set, -Raw and -Pat (the strain C, D an E isolates, respectively). The rate of LIR evolution relative to that occurring in the other genomic features has apparently been particularly high in MSV-Set and MSV-Raw following their divergence from the MSV-CD inferred ancestral sequence (compare the L columns to the others in the graph at nodes MSV-CD in Fig 4.7). The LIR of MSV-Set has probably undergone

quite substantial nucleotide sequence rearrangement since its divergence from MSV-Raw, and the 5' portion between its stem loop at the origin of (+) strand replication and the *RepA* start codon has increased in size by 14% relative to that of the inferred MSV-CD sequence.

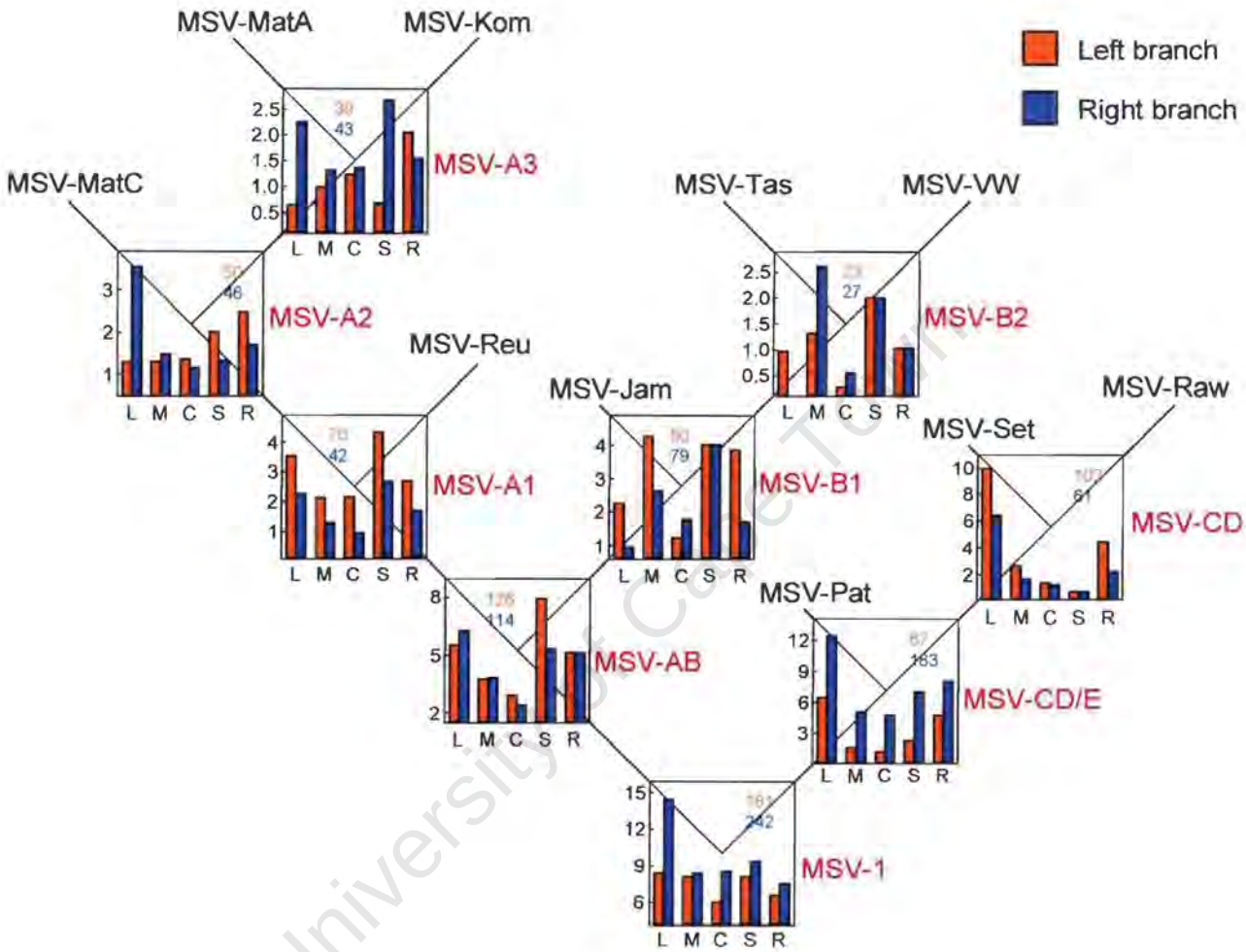


FIGURE 4.7 The reconstructed evolution of 10 currently sequenced MSV Isolates. Ancestral sequences were predicted using DNAPARS (version 3.5c; Felsenstein, Washington) and the average percentage nucleotide divergence was calculated for 5 different genomic regions of all the viruses evolving from each branch point. The bar graphs indicate the average percentage nucleotide sequence divergence (on Y-axis) of the genomic regions (on the X-axis; L = LIR; M = MP gene; C = CP gene; S = SIR; R = Rep gene) that has occurred during their evolution into the isolates (named in black) represented on the left (orange bars) and right (blue bars) hand sides of branch points. The names of inferred sequences at the nodes are given to the right of the graphs in red. The orange and blue numbers associated with bar graphs at nodes are the average number of unambiguous inferred nucleotide substitutions per genome that was used to construct the blue and orange bars, respectively. The topology of the tree represents that obtained with both the neighbour joining and maximum parsimony tree drawing methods but the distances between the nodes are arbitrary.

TABLE 4.3 The quality of ancestral MSV sequences inferred during this study.

	Number of ambiguous nucleotides (percentage)					
Sequence	LIR	*MP*	*CP*	SIR	*Rep*[a]	Total
MSV-1	20 (6.4)	18 (5.9)	60 (8.1)	13 (8.7)	77 (6.7)	188 (7.1)
MSV-CD/E	19 (6.1)	9 (2.9)	21 (2.9)	4 (2.7)	61 (5.3)	114 (4.3)
MSV-CD	15 (4.7)	5 (1.6)	3 (0.4)	1 (0.7)	29 (2.5)	53 (2.0)
MSV-AB	10 (3.2)	6 (2.0)	21 (2.9)	15 (10.1)	44 (3.8)	96 (3.6)
MSV-B1	3 (1.0)	2 (0.7)	5 (0.7)	11 (7.4)	17 (1.5)	38 (1.4)
MSV-B2	2 (0.6)	2 (0.7)	0(0)	0 (0)	6 (0.5)	10 (0.4)
MSV-A1	7 (2.3)	4 (1.3)	11 (1.5)	5 (3.3)	13 (1.1)	40 (1.5)
MSV-A2	4 (1.3)	2 (0.7)	8 (1.1)	4 (2.6)	10 (0.9)	28 (1.1)
MSV-A3	2 (0.6)	2 (0.7)	6 (0.8)	2 (1.3)	4 (0.4)	16 (0.6)

[a]LIR = Long/large intergenic region
[a]*MP* = Movement protein gene
[a]*CP* = Coat protein gene
[a]SIR = Short/small intergenic region
[a]*Rep* = Replication associated/initiation protein gene (from the initiation codon of *RepA* to the termination codon of the C2 ORF)

Interestingly, the rate at which the Rep genes of MSV-Pat, -Set, and -Raw have evolved has also been consistently higher than the rates at which the MP genes, CP genes and SIRs of these viruses have evolved (compare the R columns with the M, C and S columns in graphs at MSV-CD/E and MSV-CD in Fig 4.7). The elevated rate of nucleotide substitution in the Rep genes of the MSV-Pat and MSV-Set/Raw lineages has apparently taken place simultaneously and independently as it is undetectable at the MSV-1 node (compare the blue columns in the graph at the MSV-1 node in Fig 4.7). Most of the changes in *Rep* following the divergence of lineages from the MSV-CD/E node have been concentrated in the 5' portion of the gene. Whereas 67% of nucleotide substitutions have occurred in the 5' 450 nucleotides of MSV-Pat's *RepA* since its divergence from the MSV-CD/E node, following the joint divergence of MSV-Set and -Raw from this node 63% of the nucleotide substitutions that have occurred in their Rep genes have been within the 5' 450 nucleotides of *RepA*.

In contrast with the MSV-1 lineage that has produced strain C, D and E isolates, evolution of the LIR, SIR, and MP gene in the lineage giving rise to strain A and B isolates has, on average, proceeded at a slightly faster rate than evolution of the CP and Rep genes (compare the orange bars in the graph at the MSV-1 node in Fig 4.7). While evolution of the B-strain isolates is apparently characterised by a greater proportion of SIR and *MP* nucleotide replacements (graphs at nodes MSV-B1 and MSV-B2 in Fig 4.7), the relative rate of *MP* nucleotide replacements during evolution of the MSV-A isolates has remained approximately equal to that of the *CP* with

the SIR and LIR accumulating a proportionally greater number of nucleotide substitutions (compare C, M, L and S columns of graphs at nodes MSV-A1,-A2-and -A3 in Fig 4.7).

Remarkable genome-wide differences in evolution rates were noted in the viruses diverging from certain nodes (compare the orange and blue bars in the graphs at nodes MSV-A1, -CD/E, and -CD in Fig 4.7). Particularly interesting is that the Indian Ocean island isolate, MSV-Reu, has apparently diverged less from MSV-A1 (the inferred common MSV-A ancestor) than have the MSV isolates obtained from Africa (MSV-MakD, -MatA, and -Kom; compare orange and blue bars in the graph at node MSV-A1 in Fig 4.7). The greatest difference in genome-wide evolution rates of viruses diverging from a single node is between the MSV-Pat and MSV-Set/Raw lineages (compare orange and blue bars in graph at the MSV-CD/E node in Fig. 4.7). Following their divergence from MSV-CD/E (the inferred common ancestor of all strain C, D and E isolates), MSV-Raw and -Set have apparently evolved at an average rate 2.7 times higher than that of MSV-Pat.

4.4 DISCUSSION

Whereas a number of studies have begun to reveal the extent of MSV diversity (Briddon *et al.*, 1994, Hughes *et al.*, 1992; Rybicki *et al.*, 1998; Schnippenkoetter, 1998; Willment, 1999), virtually nothing is known of the geographical and host distributions of the virus genotypes that have been detected. Throughout Africa less than 3% nucleotide sequence divergence has been detected amongst MSV isolates obtained from severely deseased maize (Briddon *et al.*, 1994; Lazarowitz, 1988; Willment, 1999) and no attempt has been made to determine whether the small degree of sequence diversity that exists is responsible for significant differences in the virulence of these isolates. The aim of the work presented in this chapter was to provide an expanded view of MSV diversity, relating it to both the host and geographical distribution of MSV isolates and the relative virulence of these isolates in maize.

For descriptive purposes, 22 unique full-length MSV genome sequences were classified into strain and subtype groupings. While this classification was not intended to be definitive, the subtype and strain groupings were supported directly by phylogenetic analysis of the sequences and indirectly by both their pathogenicity in differentially MSV resistant maize genotypes and their geographical and host distributions. There is currently no established measure of what constitutes

a geminivirus species. It is potentially controversial that the isolates classified as belonging to strains MSV-C, -D, and -E were grouped together in the same species as MSV-A and -B isolates since they share less nucleotide sequence identity (between 78 and 83%) than many distinct *Begomovirus* species share with one another. It has been proposed that a species be defined as any group of isolates capable of *trans*-replicating one another (Rybicki *et al.*, 2000). *Trans*-replication of MSV-A and -C isolates has been demonstrated (Schnippenkoetter, 1998; Willment, 1999). The fact that these strains are more dissimilar than are MSV-A and either of MSV-D or -E implies that MSV-A isolates may also be able to *trans*-replicate and be *trans*-replicated by both MSV-D and -E isolates.

MSV is known to infect over 80 wild and cultivated grass species (Buchen-Osmond, 1998a; Damsteegt, 1983; Konate and Traore, 1992). This and other studies have indicated that severe streak disease in one of these species, *Zea mays*, is predominantly caused by MSV-A isolates (Briddon *et al.*, 1994; Willment, 1999). Collectively, all major MSV-diversity studies that have typed MSV isolates on the basis of their nucleotide sequences (by either partial sequencing or RFLP analysis) have examined only 23 isolates from eight non-maize host species. Given such a small sample size, it is interesting that four unique MSV strains have been detected. It is not inconceivable that the overwhelming predominance of MSV-A isolates in maize might reflect what is a common MSV host adaptation strategy and that there may be nearly as many host-adapted MSV strains as there are host species. While small diverse samples and host adaptation may explain why only single MSV-C, -D and -E isolates have been detected, the relatively frequent occurrence of MSV-A and -B isolates amongst a wide range of grass species demands an alternative explanation. Whereas MSV-A isolates have been detected in three non-maize grass species (*Coix lacryma jobi*, a *Pennisetum* sp. and a *Digitaria* sp.), MSV-B isolates have been detected in at least seven species (maize, wheat, rye, *Digitaria* sp., a *Paspalum* sp., *Eleusine* sp., a *Setaria* sp. and *Urochloa* sp.). MSV-A and -B are also the only MSV strains that have currently been detected in exotic cultivated grass species such as maize and wheat, which may indicate that they have a fundamentally different survival strategy from that of the other MSV strains. The ability of MSV-A and MSV-B isolates to effectively use a wide range of grass species as hosts was possibly an important factor enabling them to successfully infect and produce severe symptoms in different cultivated species that have been introduced into Africa from other parts of the world.

While too few MSV-B, -C, -D and -E isolates were analysed to determine the geographical distribution of subtypes within these strains, enough MSV-A samples were characterised to obtain a picture of subtype distributions in different parts of Africa. The six MSV-A subtypes each displayed varying degrees of geographical localisation. Two of the subtypes, MSV-A_1 and MSV-A_5 have a wide distribution with both being found in Eastern and Southern Africa and MSV-A_1 also being found in West Africa. Currently MSV-A_2, MSV-A_3, and MSV-A_4 isolates have only been detected in West Africa, East Africa, and Southern Africa, respectively. MSV-A_6 isolates have only been found on the Indian Ocean islands of Mauritius and La Réunion.

There were potentially important differences between the virulence of different MSV-A subtypes in maize (Fig 4.8). The degree of symptom severity produced by MSV-A_1, -A_2, and -A_5 isolates was significantly greater than that produced by the MSV-A_3, -A_4, and -A_6 isolates. The wider distribution of MSV-A_1 and MSV-A_5 relative to that of other MSV-A subtypes is possibly related to their greater virulence in maize. MSV-A_1 and MSV-A_5 collectively comprise 44% and 81% of the MSV isolates detected in streak-infected maize in Southern and East Africa, respectively. Despite only 5 samples from West Africa having been analysed, MSV-A_1 has also been detected in this region and is, together with MSV-A_2, likely to be one of the dominant subtypes found there. Because the virulence of the only MSV-A_2 isolate analysed during this study is equivalent to that of the MSV-A_1 isolates, it might be interesting to determine the prevalence of these isolates in West Africa over time using substantially larger sample sizes. The relative virulence of other MSV-A_2 isolates should, however, first be determined to ensure that MSV-Ns is not merely an exceptionally virulent MSV-A_2 variant. Also, if greater virulence is determined to be a prerequisite of wider distribution, it may be important to determine why MSV-A_2 isolates have only been identified in West Africa.

The virulence of all non MSV-A isolates tested appears to be correlated with the degree of nucleotide sequence identity they share with MSV-A isolates. Therefore while MSV-B isolates which share ~ 89% nucleotide sequence identity with MSV-A isolates are on average more virulent in maize than MSV-C, -D, and -E isolates that share only ~80% nucleotide sequence identity with MSV-A, all MSV isolates are more virulent than the other African streak virus isolates that share ~65% nucleotide sequence identity with MSV-A (Fig 4.8). Whereas virtually every MSV isolate detected to date has been obtained from either cultivated or wild annual grass species, with the exception of SSV-Mil (Briddon, 1996b), every non-MSV African streak virus

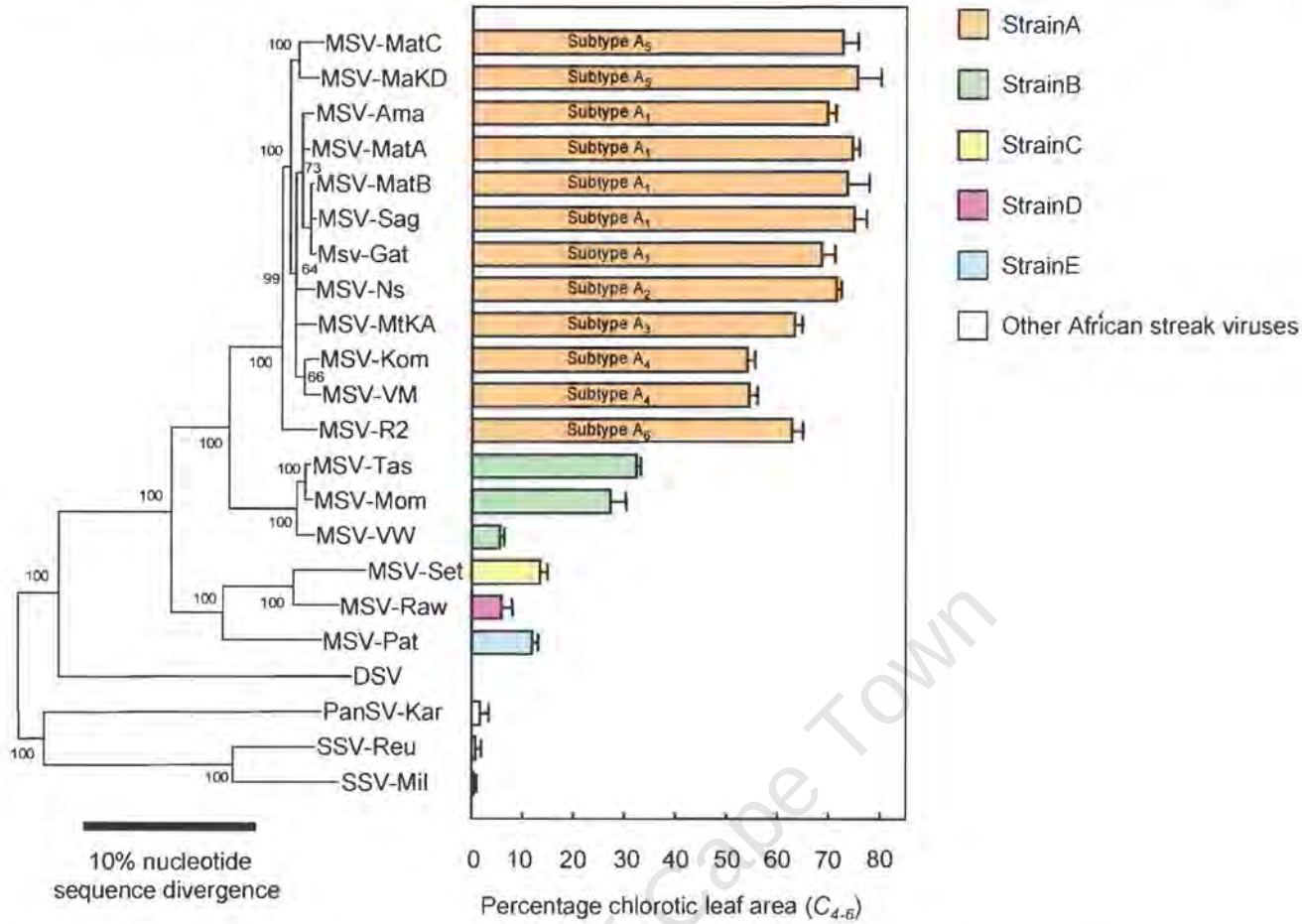


FIGURE 4.8 The relationship between virus genotype and pathogenicity in maize. The neighbour-joining tree presented here was constructed from aligned full genomic sequences of all viruses except SSV-Mil (for which only a partial nucleotide sequence is available). The numbers associated with the nodes of the tree represent the percentage bootstrap support for the nodes using 1000 bootstrap iterations. All nodes with less than 60% bootstrap support were collapsed. While horizontal distances reflect the relationships between the sequences, vertical distances are arbitrary. The bar graph indicates the average chlorotic areas produced on leaves 4-6 (C_{4-6}) of all the maize genotypes agroinfected by the virus isolates. Whereas bars are coloured according to strain, subtype classifications have been indicated within the bars. Error bars in the graph represent 95% confidence intervals of the mean.

currently characterised has been obtained from perennial grass species. This may indicate that adaptation of MSV isolates to infecting annual grass species – which perhaps requires the virus to produce more striking symptoms more quickly, to guarantee leafhopper transmission – has predisposed them to be more virulent in maize (which is also an annual species) than the other African streak virus species.

Although one of the MSV-B isolates examined in this study (MSV-Mom) was obtained from a mild streak infection of maize, it is unlikely from the results presented here that any non MSV-A isolate would pose any direct threat to maize production. While none of the MSV-B, -C, -D and -E isolates or any of the DSV, PanSV and SSV isolates produced any symptoms in PAN6099, every MSV-A isolate tested produced severe to moderate symptoms in this MSV-resistant maize

genotype. Of the MSV-A subtypes, A_1 and A_5 possibly pose the most significant threat to maize production because of their wide distribution throughout Africa and their extreme virulence.

Because widespread recombination has been detected amongst geminiviruses in the genera *Curtovirus*, *Topocuvirus* and *Begomovirus* (Padidam *et al.*, 1999a) and has potentially been responsible for the emergence of highly pathogenic virus genotypes (Umaharan *et al.*, 1998; Zhou *et al.*, 1997, 1998), an investigation of recombination amongst all available streak virus sequences was carried out. Towards this end custom recombination detection software, called RDP, was written. Although a number of recombination detection methods had been published at the time RDP was being written (Grasslly and Holmes, 1997; Hein, 1990; Maynard Smith and Smith, 1998; McGuire *et al.*, 1997; Salminen *et al.*, 1995; Sawyer, 1989; Siepel *et al.*, 1995; Weiller, 1998), none is suitable for simultaneously analysing reasonably large data sets, detecting recombination that has occurred between both very closely and very distantly related genomes, indicating likely parental sequences and estimating recombination breakpoints. Also, without prior knowledge of which genomes are likely to be non-recombinant, many of these methods are totally unusable. RDP was written to address all these problems and, unlike any other recombination analysis program that is currently available, it also runs under Windows 95/98/NT, is accessible to casual users and couples a high degree of analysis automation with an interactive and detailed graphical user interface (see Appendix C for a detailed description of the software and its use).

A recent study has indicated that recombination amongst mastreviruses occurs at a lower frequency than that which occurs amongst the begomoviruses, curtoviruses and topocuviruses (Padidam *et al.*, 1999a). The results presented here indicate that this observation may not be accurate. It is probable that the reason Padidam *et al.* (1999a) discovered so little evidence for recombination amongst mastreviruses compared to other geminivirus genera, was because the sequences they examined were both too few and too diverse. There are far fewer mastrevirus sequences deposited in GenBank than begomovirus sequences. The mastrevirus sequences that are available are also considerably more diverse than those available for the begomoviruses and curtoviruses. The detection of a recombinant sequence generally relies on the identification of at least one sequence resembling one of its parents (it is very unlikely that for natural recombinants the actual parental genomes will have been isolated and sequenced), and the probability of

detecting recombination in any particular group of sequences is therefore directly related to both the diversity and number of sequences analysed. Although there were only six sequences in the African streak virus group included in the analysis carried out by Padidam *et al.* (1999a), the depth of diversity in this group is comparable to that found in the entire *Begomovirus* genus. It is therefore hardly surprising that their analysis yielded so little evidence of recombination within the African streak virus group. In the present study enough MSV sequences were examined to detect a substantial degree of recombination both within this species and with other African streak viruses. There are, however, presently still too few non MSV African streak virus sequences available to properly detect recombination between them and any species other than MSV.

For the recombinant sequences detected in this study, there appears to be a direct correlation between the size of the recombinant regions present and the relatedness of their parental sequences. While recombination detected between MSV-A isolates involved exchanges of DNA fragments that averaged 425 nucleotides, recombination detected between different MSV strains and between African streak virus species involved an average of 142 and 46 nucleotides, respectively. The reason smaller recombinant regions were not detected amongst intra-subtype recombinants is probably because the high levels of sequence identity between parental genomes precluded detection of recombinant regions smaller than ~200 nucleotides. On the other hand, the absence of larger recombinant regions among inter-species recombinants probably indicates that there are fitness constraints associated with the relatedness of two parental viruses which limit the size of DNA fragments that can be transferred between them. In support of this, previous reports have shown that laboratory-constructed geminivirus recombinants are invariably less fit than either one or both of the parental viruses used to construct them (Liu *et al.*, 1999b; Schaffer *et al.*, 1995; Willment, 1999; see Chapter 5). Mastreviruses have extremely compact genomes and all of the proteins they encode are known to specifically interact with either viral DNA (Rep and RepA; Castellano *et al.*, 1999; Missich *et al.*, 2000) or other viral proteins (MP and CP; Kotlizky, *et al.*, 2000). Because of this, there is probably a substantial degree of co-evolution between different regions of mastrevirus genomes. When a recombination event occurs between two divergent mastrevirus genomes, the likelihood of an important interaction between viral components being disturbed will be directly proportional to both the degree of nucleotide sequence divergence between the genomes and the size of the sequences transferred.

It is unknown whether any of the recombinant viruses detected have modified host ranges or enhanced virulence under certain conditions, but none of them was detectably more pathogenic in maize than closely related but non-recombinant viruses. In three cases, nearly identical recombinant regions were noted in more than one isolate. MSV-A_5 isolates (MSV-MatC and MSV-MakD) contain a large proportion of coat protein sequence that resembles that found in MSV-A_1 isolates. MSV-A_4 isolates (MSV-SA, -Kom, and –VM) contain what appears to be MSV-B derived sequences spanning the 3' termini of their C2 ORFs. The MSV-B isolates, MSV-Tas, -Mom, and -VW all contain PanSV sequences within the predicted primer binding site at the origin of (-) strand replication. Because these recombinants are likely to represent major circulating MSV forms, one might surmise that they are significantly more fit than both of the parental viruses from which they arose. However, the possibility that these recombinants are no more or less fit than their parental viruses and that their prevalence today is merely the fortuitous result of random genetic drift can not be discounted. Alternatively, recombination may have only provided these viruses with a selective advantage under certain circumstances. It is noteworthy that while the recombinant MSV-A_4 isolates, MSV-VM and MSV-Kom, both produce significantly milder symptoms in maize than all other MSV-A isolates (Fig 4.8), MSV-A_4 is currently the dominant MSV-A subtype in Southern Africa (Fig 4.3). Both MSV-A_1 and MSV-A_5 are distributed throughout Africa and together comprise 44% of the MSV isolates detected in Southern Africa. It is surprising therefore that despite MSV-A_1 and MSV-A_5 isolates being more virulent in maize than MSV-A_4 isolates, neither subtype has displaced MSV-A_4 in Southern Africa. While recombination has restored the C2 termination codon of MSV-A_4 isolates to the same position as in all non MSV-A strains, it has shifted it 15 nucleotides upstream of where the C2 termination codon is found in all other MSV-A subtypes. Whether this has provided MSV-A_4 isolates with a subtle selective advantage over their non-recombinant highly pathogenic counterparts is unknown and may be difficult to test experimentally if, for example, the advantage is only apparent in certain host species or cell types.

The reconstructed evolution of the MSV strain groupings indicates that remarkably different patterns of nucleotide substitution have occurred in the genomes of viruses in the two lineages diverging from the common MSV ancestral sequence. Most notable is the very high rate at which sequences have evolved within the LIRs of MSV-C, -D and -E. Coupled with this high rate of LIR evolution has been a rate of *Rep* evolution that exceeds that of the *MP*, *CP* and SIR of these viruses. The disproportionate concentration of most nucleotide substitutions within the 5' 450

nucleotides of *RepA* possibly indicates that elevated rates of *Rep* evolution have been a response to the high rates at which the LIRs of these viruses have evolved. This portion of *RepA* encodes the 150 N-terminal amino acids of Rep which corresponds with DNA binding domains that have been determined for other geminivirus Rep proteins (Choi and Stenger, 1995, 1996; Jupin *et al.*, 1991). It is therefore possible that as changes in the sequence and spacing of Rep binding domains have occurred in the LIRs of MSV-Set, -Raw and -Pat, compensatory changes have needed to occur in the DNA binding and recognition domains of these viruses' Reps.

Another potentially important finding from the reconstructed evolution was evidence of substantial genome-wide differences in evolution rates of lineages diverging from certain nodes. Varying rates of nucleotide fixation in MSV genomes related to the degree to which the genomes are adapted to infection of different host genotypes have been observed experimentally (Isnard *et al.*, 1996). Perpetual infection of a suboptimal host apparently leads to the fixation of a greater number of mutations than occurs during perpetual infection of an optimal host (Isnard *et al.*, 1996). It is therefore possible that the higher evolution rates noted for MSV-Raw, and -Set relative to MSV-Pat may reflect the recent adaptation of MSV-Pat and MSV-Set to the infection of new grass species. If this were the case it is perhaps surprising that there is no evidence of increased evolution rates in the MSV-A lineage. Ancestral MSV-A isolates would have begun the process of becoming maize adapted after the introduction of maize into Africa a little over 400 years ago and one might have expected to observe a detectable difference in the rate at which this lineage had evolved relative to MSV-B.

It is interesting that the Indian Ocean island MSV-A_6 isolate, MSV-Reu, has apparently diverged substantially less from the inferred ancestral MSV-A sequence, MSV-A1, than any of the African MSV-A isolates currently sequenced. MSV-A_6 isolates have only ever been detected in samples collected from the Indian Ocean islands of Mauritius and La Réunion (Briddon *et al.*, 1994; Peterschmitt *et al.*, 1996). Whereas there is ~ 4.3% nucleotide sequence divergence between MSV-Reu and the African MSV-A isolates, the African MSV-A isolates generally share greater than ~ 97 % nucleotide sequence identity. While it is possible that the same distinct MSV isolate was introduced simultaneously to Mauritius and La Réunion in the recent past (the islands are geographically closer to one another than either are to the mainland), the evolutionary reconstruction indicates that MSV-A_6 isolates may be close relatives of a relatively uniform population of ancestral MSV-A genotypes that was once present on both these islands and in

Africa – possibly at the time maize was first introduced in Africa. While it is unlikely that individual MSV genomes are evolving at faster rates on the African continent than on the Indian Ocean islands, the apparent differences in evolution rates could reflect differences in African and island MSV population sizes. The number of individual MSV genomes on the African continent must be vastly greater than that found on any of the different Indian Ocean islands and it would therefore follow that the evolution and widespread dissemination of MSV genotypes with increased fitness might occur more frequently on the mainland. The absence of MSV sequences from Africa that resemble the ancestral MSV-A sequence might indicate that the spread of new, fitter genotypes involves displacement of pre-existing genotypes to the extent that they become virtually undetectable. Complete rapid displacement of one geminivirus strain by another has been observed for *Tomato yellow leaf curl virus* (TYLCV) in Spain (Sánchez-Campos *et al.*, 1999) and it is conceivable that similar displacements could occur during the epidemic spread of new MSV variants. While purely speculative, this feature of the MSV-A evolutionary reconstruction can be tested experimentally. Streak viruses also occur on the Atlantic Ocean islands of Sao Tomé and Cape Verde (Buchen-Osmond, 1998b). If it were discovered that MSV-A isolates from these islands more closely resemble MSV-A_6 isolates than any of the African MSV-A isolates, strong independent confirmation will have been provided that MSV-A_6 isolates are close extant relatives to the last common ancestral MSV-A sequence.

Through an analysis of MSV genotypic diversity, distribution and virulence, MSV-A subtypes have been identified which are likely to pose the greatest threat to maize production in Africa. While it is hoped that this preliminary description of MSV genotype distributions will aid future epidemiological investigations of changes in MSV population compositions over time, many of the agroinfectious MSV constructs produced in this study are already being used by Kenyan breeders to evaluate the MSV resistance of promising new maize lines. Whereas it was determined that non strain A isolates are unlikely to have any direct impact on maize production, the identification of inter-MSV strain recombination indicates that they may have an indirect impact on the epidemiology of maize streak disease. It has been widely speculated that recombination may be responsible for the emergence of certain begomoviral diseases (Padidam *et al.*, 1999; Umaharan *et al.*, 1998; Zhou *et al.*, 1997, 1998) and it is quite conceivable that either inter-strain or inter-species recombination might yield MSV-A variants with enhanced pathogenic properties.

ABSTRACT

A series of agroinfectious chimaeric *Maize streak virus* (MSV) genomes were constructed from four differentially virulent isolates in order to identify genomic regions that are important in determining the pathogenicity of MSV in maize. This involved reciprocally exchanging combinations of genes and intergenic regions between a highly virulent MSV isolate (MSV-MatA) and three less virulent MSV isolates (MSV-Kom, MSV-R2 and MSV-VW). The pathogenicity of chimaeras was analysed in differentially MSV-resistant maize genotypes. In total, 54 chimaeric genomes were constructed and analysed. Relative to the parental isolates used in their construction, chimaeras had either increased, unchanged or reduced virulence. Whereas none of the chimaeras was more pathogenic than MSV-MatA, many were more pathogenic than the less virulent isolates from which most of their sequences had been derived. Chimaeras with elevated virulence contained the MSV-MatA movement protein gene (*MP*) alone, the MSV-MatA *MP* and coat protein gene (*CP*) together, the MSV-MatA long intergenic region (LIR) alone, or the MSV-MatA *MP*,*CP*, and LIR together. The degree of nucleotide sequence identity between MSV-MatA and the less virulent isolates appeared to be a major factor determining the pathgenicity of the chimaeras. Of the less virulent isolates included in this study, MSV-VW shares the least nucleotide sequence identity with MSV-MatA and every genomic region exchanged between these isolates had a significant impact on the pathogenicity of the resulting chimaeras. Evidence was found of extensive interactions between both coding and intergenic regions. These interactions may aid in explaining the pattern of recombination observed in natural MSV recombinants.

5.1 INTRODUCTION

When investigating the relationships between the genotypes of pathogens and the severity of the disease phenotypes they induce in suitable hosts, attempts are often made to identify virulence determinants which are present in highly pathogenic isolates but absent from attenuated or milder forms. For cellular pathogens and viruses with large, complex genomes these pathogenicity determinants are often entire genes which can be deactivated or removed without having any effect on their viability other than to reduce their capacity to cause disease. For a simple virus such as MSV, however, removal or deactivation of any of its genes completely prevents it from producing systemic infections in host plants (Boulton *et al.*, 1989a; Lazarowitz *et al.*, 1989).

It has been noted in MSV that small alterations in the nucleotide sequences of its genes and intergenic regions can have major impacts on the disease symptoms produced by mutant viruses (Boulton *et al.*, 1989a, 1991a, b; Schneider *et al.*, 1992; Shen and Hohn, 1991, Willmet, 1999). These investigations have revealed that modified virulence phenotypes can often be related to the alteration of either amino acid sequences within gene products (Boulton *et al.*, 1989a) or highly conserved nucleotide sequences involved in replication or transcription (Boulton *et al.*, 1991a; Schneider *et al.*, 1992; Willment, 1999). Mutation studies have, however, also revealed that apparently minor nucleotide sequence substitutions that do not obviously alter gene products or regulatory sequences can also significantly effect the virulence of mutant genomes (Boulton *et al.*, 1991a, Shen and Hohn, 1991). An excellent example of this can be found in studies of the MSV movement protein (MP) gene. While completely altering the sequence of the MP N-terminal 14 amino acids has nearly no influence on the virulence of MSV (Boulton *et al.*, 1989a), a single conservative mutation within the nucleotide sequence encoding the 14th MP amino acid has a major impact on the symptoms produced by the virus (Boulton *et al.*, 1991a). This "conservative" mutation has since been found to effect pathogenicity by significantly altering the efficiency with which the MP gene intron is spliced (M. I. Boulton, personal communication).

Although mutation studies have provided interesting insights into some of the molecular mechanisms used by MSV during an infection, they have yielded no information on how the small amounts of nucleotide sequence variation between different naturally occurring MSV isolates, contributes to the differential virulence of these isolates. In the previous chapter 18 MSV isolates that had been obtained from naturally streak infected grasses and maize were evaluated for their ability to produce chlorotic streak symptoms in differentially MSV-resistant maize genotypes. MSV-MatA was found to be amongst the most pathogenic of these naturally occurring isolates, and was identified as belonging to what is potentially the most widespread and economically significant group of MSV genotypes – the MSV-A subtype, A_1. Despite being relatively closely related to MSV-MatA, MSV-Kom (an MSV-A_4 isolate sharing 97.7% nucleotide sequence identity with MSV-MatA) and MSV-R2 (an MSV-A_6 isolate sharing 95.9% nucleotide sequence identity with MSV-MatA) both produced significantly less severe symptoms in differentially MSV-resistant maize genotypes than MSV-MatA. The MSV-B isolate MSV-VW (sharing 89.5% nucleotide sequence identity with MSV-MatA) was the least pathogenic MSV isolate tested in maize and was unable to produce a symptomatic infection in the moderately MSV-resistant maize genotype, PAN6099.

To identify the nucleotide sequence determinants responsible for the pathogenicity differences between MSV-MatA, -Kom, -R2 and -VW, a series of chimaeric genomes was constructed which contained different combinations of genes (movement, coat and replication associated protein genes or *MP*, *CP* and *Rep*, respectively) and intergenic regions (long and short intergenic regions or LIR and SIR, respectively) from these MSV isolates. The virulence of 54 resulting chimaeras was examined in differentially MSV-resistant maize genotypes. These data were used to determine the individual contributions of the various genomic components to the differential pathogenicities of their cognate isolates were determined.

5.2 METHODS AND MATERIALS

5.2.1 Virus Isolates, Maize Genotypes, Plasmids and Enzymes

Chimaeras were constructed from the full-length infectious clones of MSV-MatA (Chapter 4), MSV-Kom, MSV-VW (obtained from F. Hughes, University of Cape Town, Cape Town, South Africa) and MSV-R2A (obtained from M. Peterschmitt, CIRAD, Montpellier, France). Seed of the hybrid maize genotype, PAN6099, was obtained from D. Nowell (PANNAR Ltd., Greytown, South Africa) and sweetcorn (cv. Jubilee) seed was obtained from Starke Ayres nursery (Cape Town, South Africa). The *RecA*$^-$ *Escherichia coli* strains JM109 and DH5α and the plasmid cloning vectors pBluescriptSK+ (pSK+; Stratagene, La Jolla, CA) and pUC19 were used during all routine cloning. *Agrobacterium tumefaciens* C58C1 [pMP90] (Koncz and Schell, 1986) was used to deliver agroinfectious clones during agroinoculation. The binary cloning vectors, pBI121 (CLONTECH, CA) and pBIN19 (CLONTECH, CA) were used to construct agroinfectious clones. Unless otherwise stated all enzymes were obtained from Boehringer Mannheim (Mannheim, Germany)

5.2.2 Construction of Chimaeric MSV Genomes

Three groups of chimaeric genomes were constructed using standard techniques (Sambrook *et al.*, 1989). These groups were MSV-MatA/ MSV-Kom, MSV-MatA/ MSV-R2, and MSV-MatA/ MSV-VW chimaeras. Each group contained a set of 18 chimaeric genomes. Each set of 18 genomes consisted of nine reciprocal gene and/or intergenic region swaps between MSV-MatA and either MSV-Kom, MSV-R2 or MSV-VW and are described in Figs 5.1, 5.2 and 5.3, respectively. The nine pairs of recombinant genomes produced in each group were (where X refers to either VW, R2 or Kom): (1) a *MP* recombinant (MatMPX/ XMPMat), (2) a *CP*

recombinant (MatCPX/ XCPMat), (3) a *MP* and *CP* recombinant (MatMPCPX/ XMPCPMat), (4) a LIR recombinant (MatLIRX/ XLIRMat), (5) a LIR, *MP*, and *CP* recombinant (MatLIRMPCPX/ XLIRMPCPMat), (6) a SIR recombinant (MatSIRX/XSIRMat), (7) a SIR, *MP*, and *CP* recombinant (MatMPCPSIRX/ XMPCPSIRMat), (8) a LIR and SIR recombinant (MatLIRSIRX/ XLIRSIRMat), and (9) a *Rep* recombinant (MatRepX/ XRepMat) (Figs 5.1 – 5.3). In almost all cases the naming of the chimaeras reflected the nucleotide sequence relationship between the chimaeras and their parental genomes. Therefore MatMPVW refers to a chimaera with greatest nucleotide sequence identity to MSV-MatA, containing the LIR, *CP*, SIR and *Rep* of MSV-MatA and the *MP* of MSV-VW. However the *Rep* recombinants containing, for example, the *Rep* of MSV-VW and the LIR, *MP*, *CP* and SIR from MSV-MatA were named MatRepVW for the sake of brevity despite sharing greater nucleotide sequence identity with MSV-VW.

MSV-MatA/Kom and MSV-MatA/R2 *MP* and *CP* chimaeras were constructed from full-length infectious *BamHI* clones of these viruses in pUC19, making use of naturally occurring *AsuII* and *NcoI* sites at the 3' termini of their MP and CP genes, respectively, and an *XbaI* site within the vector (Fig. 5.4). MSV-MatA/VW *MP* and *CP* chimaeras were constructed from full-length infectious *BamHI* clones of MSV-MatA and –VW in pSK+ using *XmaI* and *NcoI* sites at the 3' termini of their MP and CP genes, respectively, and an *XmaI* site within the vector.

While MSV-MatA/Kom and MSV-MatA/R2 LIR chimaeras were constructed from full-length *BamHI* clones using an *XbaI* site in the vector and an *AsuII* site 72 nucleotides upstream from the *RepA* initiation codon, the MatLIRMPCPX and XLIRMPCPMat chimaeras were constructed using the same sites in the cloned MPCP chimaeras (MatMPCPX and XMPCPMat) of these viruses. Because MSV-VW contains an extra *AsuII* site within its CP gene (Fig 5.4), MatLIRVW and VWLIRMPCPMat were constructed using an *EcoRI* site within the vector and the *AsuII* site within the LIR of the full-length MSV-MatA *BamHI* clone in pSK+ and the VWMPCPMat chimaera, respectively. MatLIRMPCPVW was constructed by cloning an *XbaI-NcoI* fragment of MatLIRVW (spanning the LIR, *RepA*, C2 and SIR; Fig 5.4) between the corresponding sites of the full-length MSV-VW clone. VWLIRMat was constructed by cloning an *EcoRI-NcoI* fragment from the full-length VW clone between the corresponding sites of VWLIRMPCPMat.

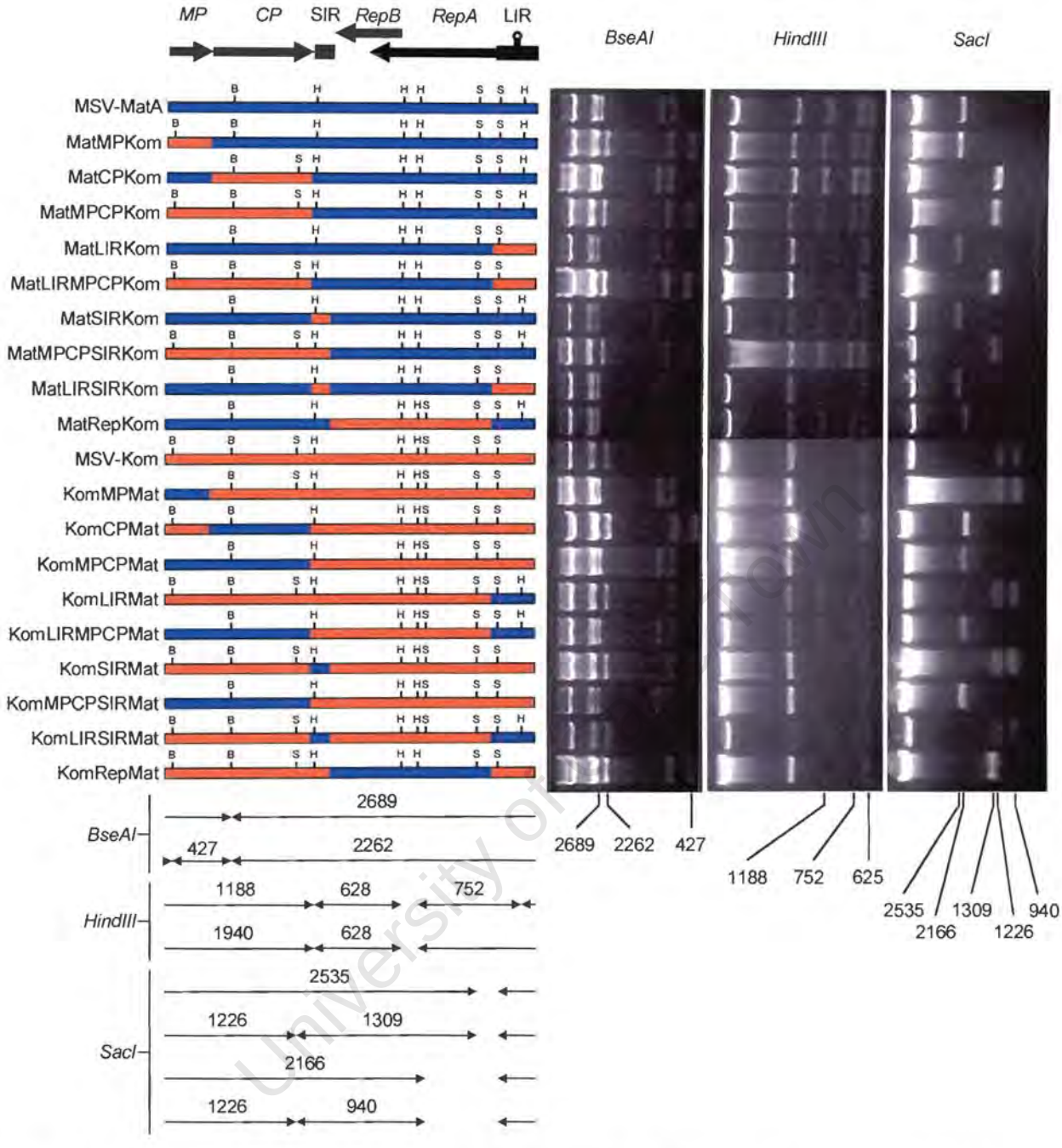


FIGURE 5.1 The MSV-MatA/Kom chimeras produced during this study. The identity of dimerised chimeric genomes cloned in pBI121 was confirmed by restriction endonuclease digestion with *BseAI* (B), *HindIII* (H), and *SacI* (S).

In order to enable construction of the MSV-MatA/R2 SIR chimaeras, it was necessary to introduce a *BglII* restriction site at the 3' terminus of the MSV-R2 C2 ORF at a position directly corresponding to that which occurs in MSV-MatA, -Kom and -VW (Fig 5.4). This was accomplished using the QuickChange Site Directed Mutagenesis Kit (Stratagene) using the following primers (modified nucleotide in bold and BglII site underlined): 5'-CAGATA**G**ATCTTGATTTTTCG-3'; 5'-CAAGAT**C**TATCTGTACTGCC-3'. Two *HindIII*

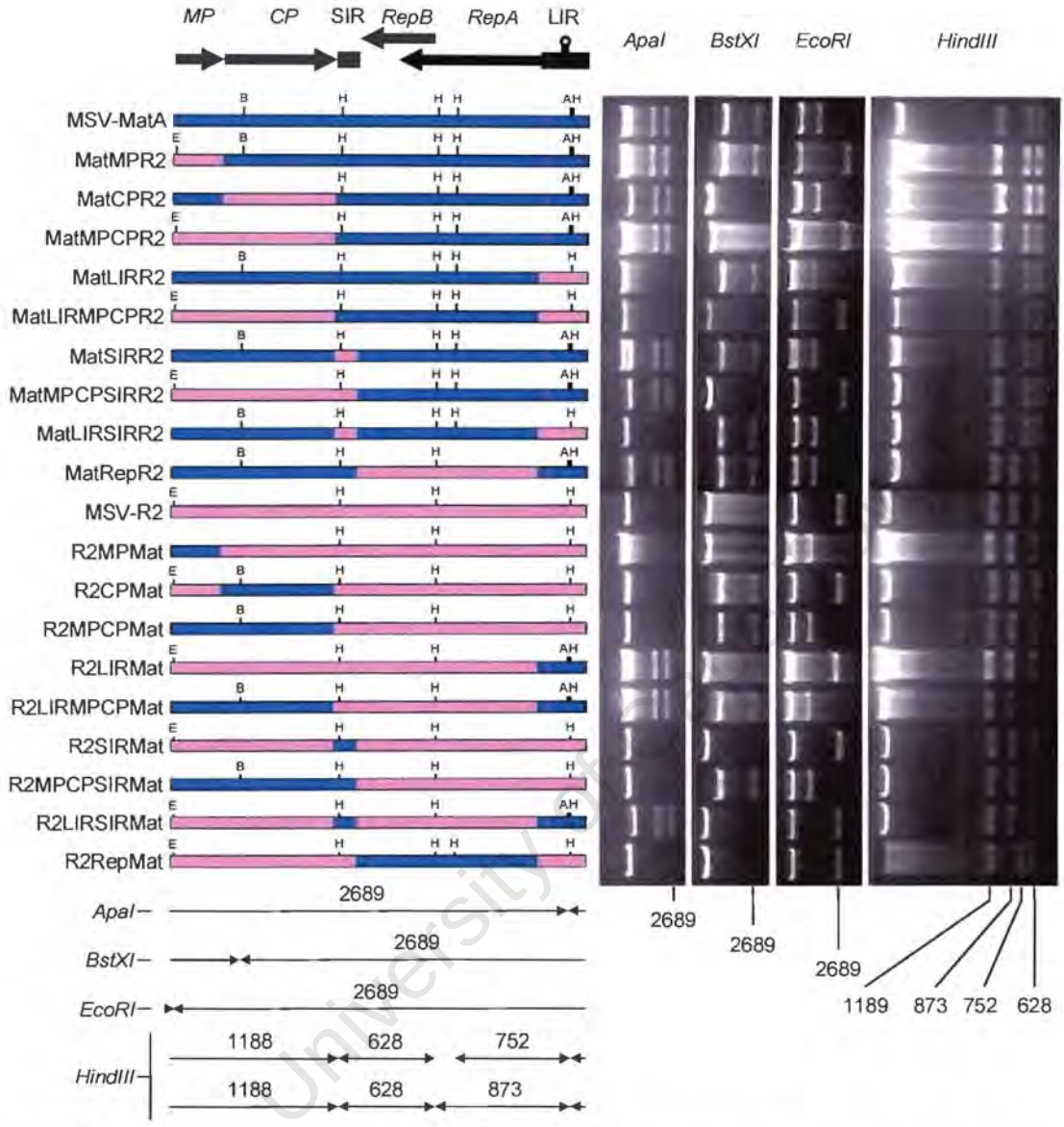


FIGURE 5.2 The MSV-MatA/R2 chimeras produced during this study. The identity of dimerised chimeric genomes cloned in pBIN19 was confirmed by restriction endonuclease digestion with *ApaI* (A), *BseAI* (B), *EcoRI* (E) and *HindIII* (H).

subclones were constructed from the clone obtained following PCR mutagenesis (pR2BglII^{+}) and sequenced to confirm that the mutated sequence was correct and that no other unwanted PCR-induced mutations had occurred in the SIR and the 3' portion of the MSV-R2 C2 ORF. An *NsiI-NcoI* fragment of pR2BglII^{+} (Fig 5.4) that had been fully sequenced and was known to be free of unwanted mutations was then cloned into the corresponding sites of both the full-length MSV-R2 clone and R2MPCPMat to obtain MSV-R2B^{+} and R2MPCPMatB$^{\pm}$, respectively.

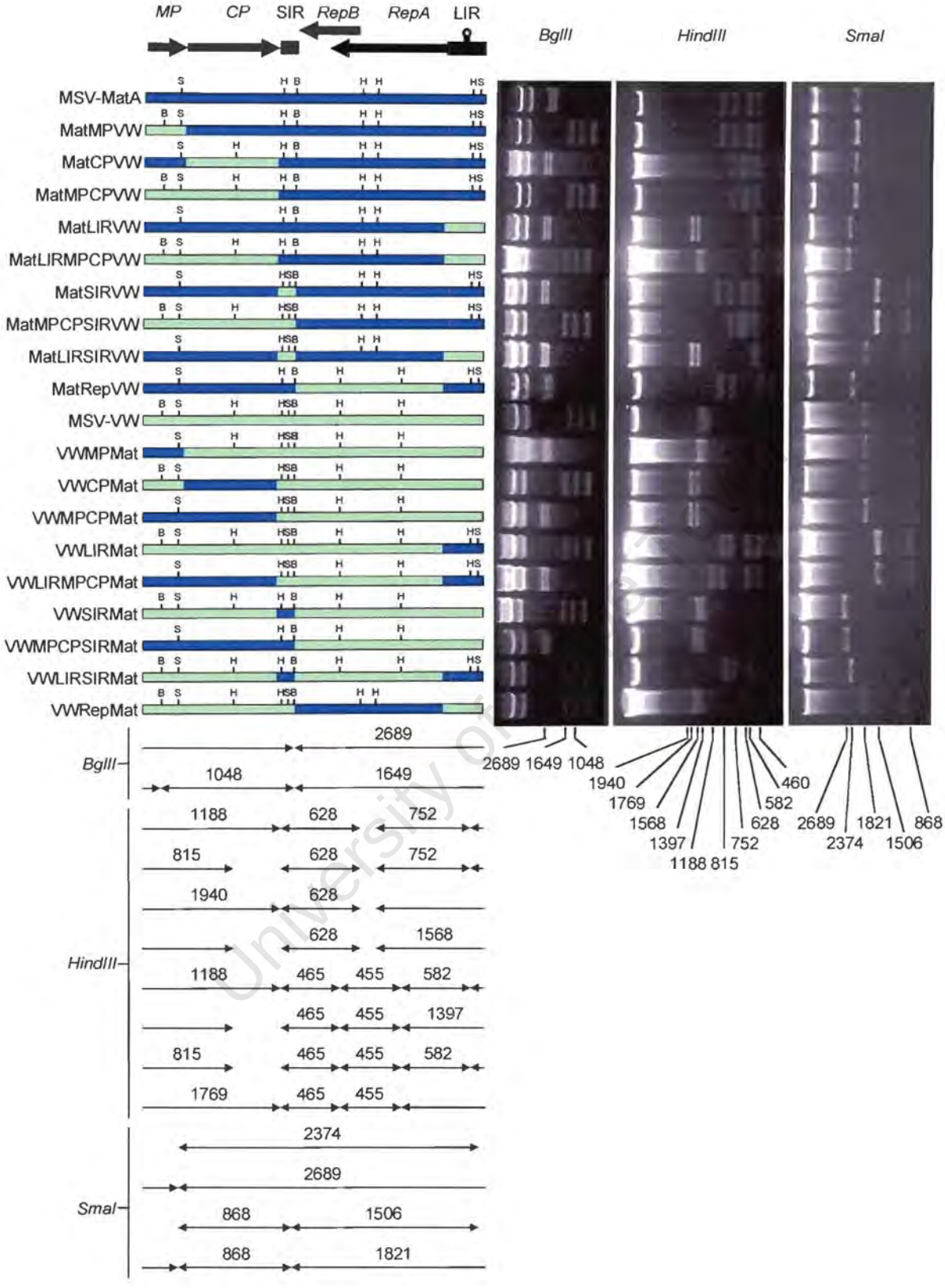


FIGURE 5.3 The MSV-MatA/VW chimeras produced during this study. The identity of dimerised chimeric genomes cloned in pBI121 was confirmed by restriction endonuclease digestion with *BglII* (B), *HindIII* (H), and *SmaI* (S).

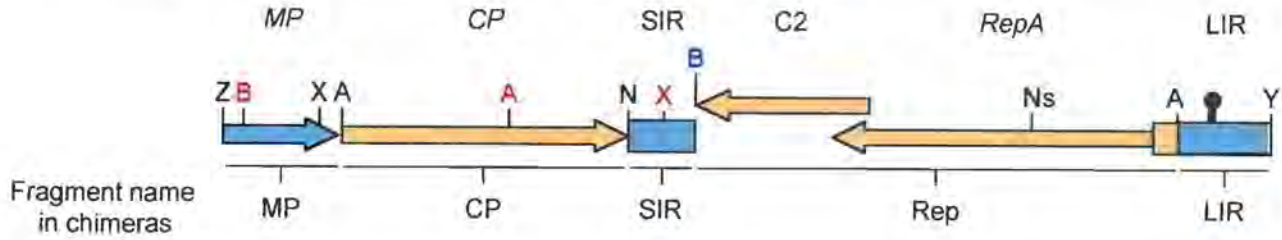


FIGURE 5.4 Restriction map of MSV-MatA, -Kom, -R2, and -VW. A = *AsuII*, B = *BglII*, N = *NcoI*, Ns = *NsiI*; X = *XmaI*. When cloning the MSV-MatA/Kom and MSV-MatA/R2 chimeras, site Y = *EcoRI* and Z = *XbaI*. When cloning the MSV-MatA/Kom and MSV-MatA/R2 chimeras, site Y = *XbaI* and Z = *EcoRI*, *SalI* or *XmaI*. These fortuitously arranged sites enabled transfer of intact genomic components between MSV-MatA and MSV-Kom, -R2 and -VW. Sites in red are unique to MSV-VW and the site in blue is absent in wt MSV-R2 and had to be introduced by site directed mutagenesis. *MP* = movement protein gene. *CP* = coat protein gene. SIR = short/small intergenic region. C2 = the 3' portion of the replication associated protein gene. *RepA* = the 5' portion of the replication associated protein gene. LIR = long/large intergenic region. Whereas the *AsuII* site between the *MP* and *CP* were used to produce *MP* and *CP* MSV-MatA/Kom and MSV-MatA/R2 chimeras, the *XmaI* site within the *MP* was used to produce the MSV-MatA/VW *MP* and *CP* chimeras. The labels beneath the restriction map indicate the names of the DNA fragments used during chimera construction. Indicated within the LIR is the position of the stem loop structure at the origin of (+) strand replication.

MatSIRX and XSIRMat chimaeras were constructed by replacing the *XbaI-BglII* fragments of the full-length clones (MSV-R2B$^+$ for the R2 chimaeras) with the *XbaI-BglII* fragment of MatMPCPX and XMPCPMat (R2MPCPMatB$^+$ for the R2 chimaeras), respectively (Fig 5.4). MatMPCPSIRR2 and R2MPCPSIRMat clones were constructed by replacing the *XbaI-BglII* fragments of MSV-R2B$^+$ and MSV-MatA with the *XbaI-BglII* fragments of MatMPCPR2 and R2MPCPMatB$^+$, respectively. Similarly, MatMPCPSIRKom and KomMPCPSIRMat clones were constructed by replacing the *XbaI-BglII* fragments of MSV-Kom and MSV-MatA with the *XbaI-BglII* fragments of MatMPCPKom and KomMPCPMat, respectively. MatSIRVW and VWMPCPSIRMat were constructed by cloning the *EcoRI-BglII* fragment of VWMPCPMat into the *EcoRI-BglII* sites of the full-length MSV-MatA and MSV-VW *BamHI* clones, respectively. VWSIRMat and MatMPCPSIRVW were constructed by cloning the *EcoRI-BglII* fragment of MatMPCPVW into the *EcoRI-BglII* sites of the full-length MSV-VW and MSV-MatA *BamHI* clones, respectively (Fig 5.4).

MatLIRSIRX and XLIRSIRMat clones were constructed by replacing the *BglII-EcoRI* (for MSV-MatA/Kom chimaeras), *BglII-SalI* (for MSVMatA/R2 chimaeras) or *BglII-XbaI* (for MSV-MatA/VW chimaeras) fragments of MatLIRX and XLIRMat into the corresponding sites of MatSIRX and XSIRMat, respectively. MatRepX and XRepMat clones were constructed by replacing the *BglII-EcoRI* (for MSV-MatA/Kom chimaeras), *BglII-SalI* (for MSV-MatA/R2 chimaeras) or *BglII-XbaI* (for Msv-MatA/VW chimaeras) fragments of XLIRMat and MatLIRX into the corresponding sites of XMPCPSIRMat and MatMPCPSIRX, respectively.

5.2.3 Generation of Agroinfectious Constructs

Agroinfectious clones were constructed essentially as described by Grimsley *et al.* (1987) according to the method of Palmer (1997). In total, 61 genomes were dimerised and cloned into either pBI121 (MSV-MatA/Kom and MSV-MatA/VW chimaeras) or pBIN19 (MSV-MatA/R2 chimaeras) . Besides the 54 chimaeras, these genomes included the wild-type (wt) MSV-R2, MSV-Kom and MSV-VW genomes, the mutant MSV-R2B$^+$ and wt MSV-MatA cloned into both pBIN19 (for comparison with the MSV-MatA/R2 chimaeras) and pBI121 (in two different orientations for comparison with the MSV-MatA/VW and MSV-MatA/Kom chimaeras). *XbaI – EcoRI* fragments containing full tandem repeats of MSV-MatA/Kom and MSV-MatA/VW chimaeras were respectively cloned from pUC19 and pSK+ into the *XbaI* and *EcoRI* sites of pBI121. *XbaI–XhoI* fragments containing full tandem repeats of MSV-MatA/R2 chimaeras were cloned from pBluescriptSK into the *XbaI – SalI* sites of pBIN19. Agroinfectious clones were transformed into *A. tumefaciens* C58C1 using the method of An *et al.* (1988).

5.2.4 Agroinoculation

Agroinoculation was performed according to Grimsley *et al.* (1987) with modifications described previously in Chapter 2. Each agroinfectious construct was used to infect both sweetcorn (a moderately MSV-susceptible maize genotype) and PAN6099 (a moderately MSV-resistant maize genotype). Agroinfection experiments involving each maize genotype-agroinfectious construct combination were repeated at least three times. Agroinfection experiments involving specific constructs were carried out on groups of 14 plants. For each agroinfection experiment, a group of 14 seedlings injected with sterile distilled water served as uninfected controls. Seedlings were grown under near-uniform conditions in a plant growth room maintained at 22^{o}C, at 80% relative humidity, with 16 hours of light per day.

5.2.5 Analysis of Symptoms

The proportion of plants showing symptoms was determined at three-day intervals between the fifth and fourteenth days post inoculation (dpi). For each MSV isolate-plant genotype combination, these measurements were integrated into an infection rate (IR) value calculated as the mean percentage of symptomatic plants observed at the four assessment times.

Heights of symptomatic plants and uninfected control plants were measured 15 dpi as the distance from their coleoptilar nodes to the tip of their fourth leaves. For specific agroinfectious construct - maize genotype combinations, a value designated S (representing stunting occurring as a result of infection) was calculated as the mean height of symptomatic plants expressed as a proportion of the mean height of uninfected control plants. The value of 1-S was used as a description of stunting.

Percentage leaf areas covered by chlorotic lesions in symptomatic plants were estimated for leaves two through six using the microcomputer-based image analysis technique described previously in Chapters 2 and 3. Whereas percentage chlorotic areas on leaves two and three were determined 15 dpi, those on leaves four, five and six were determined 22, 29 and 35 dpi, respectively. For each agroinfectious construct-plant genotype combination, the mean percentage chlorotic area of the fourth to sixth leaves, (designated $C_{4\text{-}6}$), was used as a representative measure of chlorosis that could be used to directly compare the virulence of MSV-MatA/Kom, MatA/R2 and MatA/VW chimaeras (see Chapter 3).

An interaction index (I_n) was devised to describe the degree to which a genomic component could be transferred into a divergent genomic background and still retain its ability to function as well as it did within the context of its native genomic background. This index was calculated for pairs of reciprocal chimaeras as follows:

$$I_n = \frac{(MXC_{4\text{-}6} + XMC_{4\text{-}6})}{(MC_{4\text{-}6} + XC_{4\text{-}6})}$$

where $MXC_{4\text{-}6}$ is the $C_{4\text{-}6}$ value of the predominantly MSV-MatA chimaera, $XMC_{4\text{-}6}$ is the $C_{4\text{-}6}$ value of the reciprocal, predominantly MSV-Kom, -R2 or -VW chimaera, $MC_{4\text{-}6}$ is the $C_{4\text{-}6}$ value of MSV-MatA, $XC_{4\text{-}6}$ is the $C_{4\text{-}6}$ value of MSV-Kom, -R2 or -VW. The I_n of a pair of reciprocal chimaeras therefore illustrates the degree to which the average virulence of the chimaeras has been altered relative to the average virulence of the two wild-type genomes used in their construction. This measure assumes that if the genomic component reciprocally exchanged between MSV-MatA and either MSV-Kom, -R2 or -VW influences the virulence of chimaeras due solely to the genomic component's functionality, then the value of I_n should = 1. "Functionality" of a genomic component is defined here as the average direct contribution that two different versions of the component make to the virulence of the chimaeric viruses in which they are present. If $I_n > 1$ then it is assumed that the functionality of the components has increased

in the reciprocal chimaeras relative to their functionality in their original genomic backgrounds. If $I_n < 1$ then it is assumed that the functionality of the two genomic components has decreased in the reciprocal chimaeras relative to their functionality in their origional genomic backgrounds. An increase in I_n above 1 would imply that a synergistic effect had occurred that enabled both versions of the genomic component being examined to, on average, function more efficiently in their non-native genomic backgrounds. Conversely, a decrease in I_n below 1 would imply that an antagonistic effect had occurred in the chimaeras that, on average, resulted in both versions of the genomic component being examined functioning less efficiently than they had in their native genomic backgrounds. The value of I_n should relate directly to the magnitude of these synergistic and antogonistic effects. It is also important to note that I_n is calculated using $C_{4\text{-}6}$ values which enables the direct comparison of I_n of analogous MSV-MatA/Kom, MSV-MatA/R2 and MSV-MatA/VW chimaeras (see Chapter 3).

5.3 RESULTS

5.3.1 Viability of Chimaeric and Mutant Genomes

Prior to construction of the MSV-MatA/R2 chimaeras, it was necessary to introduce a restriction site between the C2 ORF and SIR of MSV-R2 to enable the cloning of SIR and Rep gene replacements (Fig 5.4). Agroinfectious clones of MSV-R2 and the mutant genome, MSV-R2B$^+$, had identical infection rates and induced indistinguishable degrees of stunting and chlorotic areas on leaves two through six of agroinoculated sweetcorn and PAN6099 seedlings. This implied that the introduced *BglII* site at the 3' terminus of the C2 ORF of MSV-R2B$^+$ had no detectable effect on this virus' fitness in these maize genotypes.

All of the chimaeric genomes that were constructed could infect and produce symptoms in sweetcorn. Examples of streak symptoms occurring on the fifth leaves of sweetcorn and PAN6099 plants infected with the chimaeras are presented in Appendix E (Figs E.1 through E.6) Relative to the virulence of wt viruses, the virulence of chimaeras (as measured by infection rates, stunting, and chlorotic areas in the differentially MSV-resistant maize genotypes) could be divided into three categories: (1) no change in virulence relative to the most closely related wt virus, (2) increased virulence relative to the most closely related wt virus, and (3) reduced virulence relative to the most closely related wt virus. Alterations in the functional interactions of different genomic components could be detected when either one or both chimaeras in a reciprocal chimaera pair were less virulent than their most closely related wt viruses.

5.3.2 The Virion Sense Genes (*MP* and *CP*)

The virion sense genes of MSV-MatA apparently contain the major sequence elements responsible for this isolate being more pathogenic in maize than MSV-Kom, -R2, and -VW. MSV-MatA/Kom and MSV-MatA/R2 chimaeras containing either the MSV-MatA *MP* alone (KomMPMat and R2MPMat) or the MSV-MatA *MP* and *CP* together (R2MPCPMat) produced greater chlorotic areas and degrees of stunting in symptomatic sweetcorn plants than wt MSV-Kom and MSV-R2, respectively (Fig 5.5A, panels 2:1, 2:2, and 4:2, respectively; Fig 5.5 B). The chimaera, R2MPCPMat, was almost as virulent as wt MSV-MatA (Fig 5.5A panel 4:2). Whereas KomMPCPMat had a lower infection rate, induced significantly less stunting and produced smaller chlorotic areas on the first symptomatic leaves than wt MSV-Kom in sweetcorn, by leaf six it was producing greater chlorotic areas than wt MSV-Kom (Fig 5.5A panel 4:1). Of these MP and MPCP chimaeras, only the KomMPMat chimaera retained its ability to produce greater chlorotic areas than wt MSV-Kom in PAN6099 (Fig 5.5A panel 2:4). While R2MPCPMat and MSV-MatA had indistinguishable infection rates, induced similar degrees of stunting and produced nearly identical chlorotic areas on leaves three and four in infected PAN6099, chlorotic areas produced by R2MPCPMat decreased sharply after leaf four so that they were smaller than those produced by MSV-R2 on leaf six (Fig 5.5A panel 4:5, Fig 5.5B). The chimaeras, MatMPKom, MatMPR2 and MatMPCPR2 produced substantially smaller chlorotic areas in both sweetcorn and PAN6099 than did wt MSV-MatA (Fig 5.5A panels 2:1, 2:2 and 4:1, respectively). However, only in the case of MatMPR2 and MatMPCPR2 were these reductions in chlorotic areas also coupled with a significantly reduced infection rate and degree of stunting in symptomatic plants relative to MSV-MatA (Fig 5.5B). Corresponding with only a marginal increase in virulence of KomMPCPMat relative to MSV-Kom in sweetcorn, the reciprocal chimera, MatMPCPKom, produced only slightly reduced chlorotic areas than did MSV-MatA (Fig 5.5A panel 4:1). In PAN6099, however, MatMPCPKom produced substantially smaller chlorotic areas (particularly by leaf six) than did MSV-MatA (Fig 5.5A panel 4:4)

While the MSV-MatA/VW chimaera containing only the MSV-MatA *MP* (VWMPMat) did not produce significantly greater chlorotic areas on the leaves of symptomatic plants than those produced by wt MSV-VW (Fig 5.5A, panel 2:3), it had a higher infection rate and induced significantly more stunting than MSV-VW in sweetcorn (Fig 5.5B). VWCPMat in sweetcorn produced both significantly more stunting and slightly greater chlorotic areas than did MSV-VW

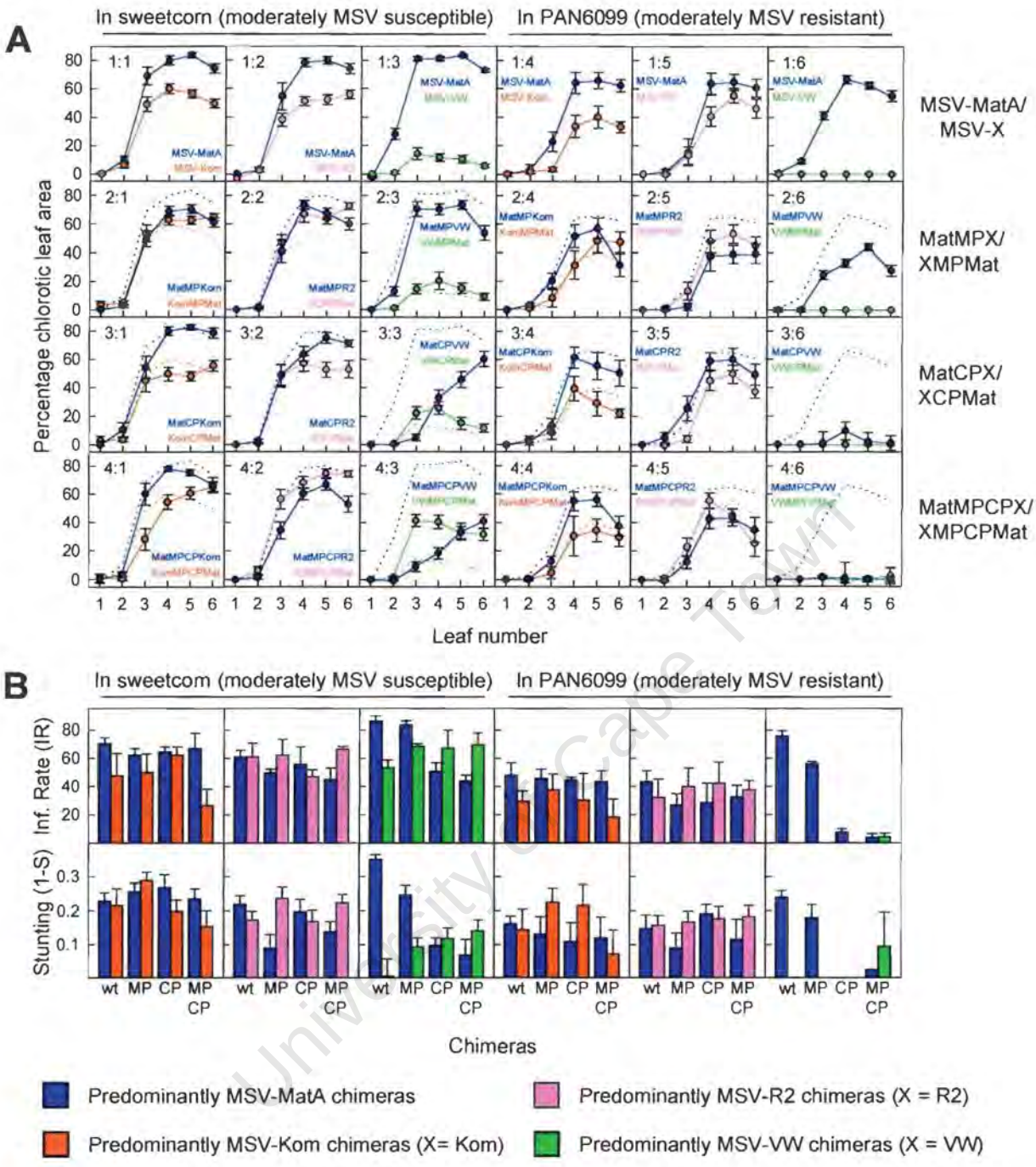


FIGURE 5.5 The virulence of movement protein gene (*MP*), coat protein gene (*CP*) and *MP* + *CP* chimeras. (A) The chlorotic areas produced in sweetcorn (a moderately MSV-susceptible maize genotype) and PAN6099 (a moderately MSV-resistant maize genotype) agroinoculated with either chimeras or wt genomes. MSV-MatA and chimeras containing predominantly MSV-MatA sequences are in blue, MSV-Kom and chimeras containing predominantly MSV-Kom sequences are in orange, MSV-R2 and chimeras containing predominantly MSV-R2 sequences are in pink, and MSV-VW and chimeras containing predominantly MSV-VW sequences are in green. Broken lines representing the chlorotic areas produced by wt viruses are presented for comparative purposes. (B) Infection rates and stunting induced by wt and chimeric viruses. All error bars represent 95% confidence intervals (Student's t-test).

(Fig 5.5A, panel 3:3; Fig 5.5B). However, when both the MSV-MatA *MP* and *CP* were present together in VWMPCPMat, the chimaera was substantially more infectious and produced far greater chlorotic areas and degrees of stunting than produced by MSV-VW, VWMPMat, and VWCPMat (Fig 5.5A, compare green plot in panel 4:3 with green plots in panels 2:3 and 3:3). Infection rates, chlorotic symptoms and stunting produced in sweetcorn by all other MSV-MatA/VW chimaeras containing the MSV-VW *Rep* and MSV-MatA *MP* and *CP* (VWLIRMPCPMat, VWMPCPSIRMat, and MatRepVW) differed only slighly from those produced by VWMPCPMat (compare green plot in Fig 5.5A panel 4:3 with green plots in Fig 5.6A, panel 2:3 and Fig 5.7A panel 2:3, and blue plot in Fig 5.7A panel 4:3). The converse of what was observed with VWMPMat, VWCPMat and VWMPCPMat was observed with MatMPVW, MatCPVW and MatMPCPVW, respectively. Relative to MSV-MatA in sweetcorn, MatMPVW produced slightly smaller chlorotic areas and less stunting, while MatCPVW and MatMPCPVW were remarkably less infectious and produced much reduced stunting and chlorotic areas (Fig 5.5A panels 2:3, 3:3 and 4:3, respectively; Fig 5.5B). The delayed development of symptoms was particularly interesting in plants infected with MatCPVW and MatMPCPVW (Fig 5.5A blue plots in panels 3:3 and 4:3; Fig 5.5B). On the earliest symptomatic leaves in plants infected with these chimaeras, chlorotic areas were approximately equivalent to those produced by wt MSV-VW, but by leaf six chlorotic areas had increased to become more extensive than those recorded for VWMPCPMat. The chlorotic areas produced by MatCPVW on leaf six were also greater than those produced by MatMPVW.

5.3.3 The LIR

The LIR of MSV-MatA apparently also contains sequence elements that are partially responsible for this isolate being more pathogenic than MSV-Kom. However, while the MSV-MatA/Kom chimaera, KomLIRMat, produced significantly greater chlorotic areas in sweetcorn than MSV-Kom (Fig 5.6A panel 1:1), the analogous MSV-MatA/R2 chimaera, R2LIRMat, was significantly less infectious and produced smaller chlorotic areas than MSV-R2 in both sweetcorn and PAN6099 (Fig 5.6A panels 1:2 and 1:5, respectively; Fig 5.5B). It therefore appears as though the MSV-MatA LIR interacts poorly with some other component of the MSV-R2 genome. To identify the nature of this interaction, it was necessary to investigate the pathogenicity of chimaeras containing paired genome replacements. Both KomLIRMPCPMat and R2LIRMPCPMat produced stunting and chlorotic symptoms that were almost indistinguishable from those of MSV-MatA in sweetcorn (Fig 5.6A panels 2:1 and 2:2; Fig 5.6B) indicating that the

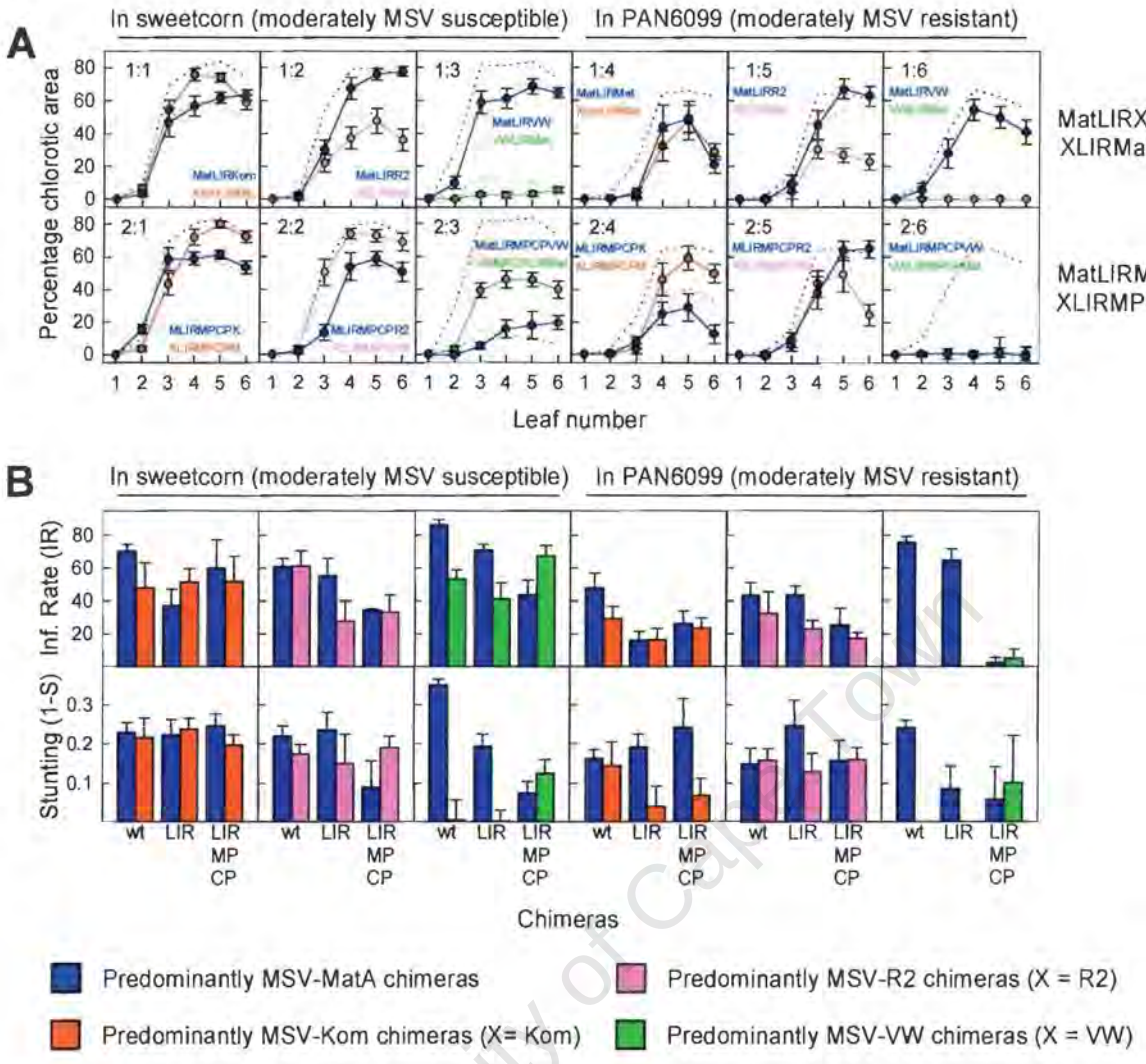


FIGURE 5.6 The virulence of long intergenic region (LIR) and LIR + *MP* + *CP* chimeras. (A) The chlorotic areas produced in sweetcorn (a moderately MSV-susceptible maize genotype) and PAN6099 (a moderately MSV-resistant maize genotype) agroinoculated with the chimeras. Chimeras containing predominantly MSV-MatA sequences are in blue, chimeras containing predominantly MSV-Kom sequences are in orange, chimeras containing predominantly MSV-R2 sequences are in pink, and chimeras containing predominantly MSV-VW sequences are in green. Broken lines representing the chlorotic areas produced by wt viruses are presented for comparative purposes. In certain cases the names of chimeras presented within each of the panels have been abbreviated due to space constraints. For example MatLIRMPCPKom has been abbreviated to MLIRMPCPK in panel 2:4. (B) Infection rates and stunting induced by wt and chimeric viruses. All error bars represent 95% confidence intervals (Student's t-test).

virulence properties of the LIR are probably influenced by sequences within the MP and/or CP genes. It appears, however, that host factors may also be involved in this interaction because R2LIRMPCPMat was substantially less virulent than MSV-MatA in PAN6099 and even produced smaller chlorotic areas than MSV-R2 in this host (Fig 5.6A panel 2:5). Relative to MSV-MatA in PAN6099, KomLIRMPCPMat induced as much stunting but had a slightly lower infection rate and produce slightly smaller chlorotic areas (Fig 5.6A panel 2:4; Fig 5.6B).

Examining the reciprocal chimaeras containing predominantly MSV-MatA sequences, it is apparent that the MSV-MatA and -R2 LIRs function better in the maize genotypes tested than does the MSV-Kom LIR. Whereas MatLIRKom had a much lower infection rate and produced substantially smaller chlorotic areas than wt MSV-MatA in both sweetcorn and PAN6099 (Fig 5.6A panels 1:1 and 1:4; Fig 5.6B), MatLIRR2 induced approximately the same degree of stunting as MSV-MatA and produced chlorotic areas on leaves four through six that were indistinguishable from those produced by MSV-MatA in both sweetcorn and PAN6099 (Fig5.6A panels 1:2 and 1:5; Fig 5.6B). In sweetcorn, the chimaeras MatLIRMPCPKom and MatLIRMPCPR2 produced chlorotic areas similar to those produced by MSV-Kom and MSV-R2, respectively (Fig 5.6A panels 2:1 and 2:2). In PAN6099, however, MatLIRMPCPKom produced significantly smaller chlorotic areas than those produced by MSV-Kom, while MatLIRMPCPR2 produced chlorotic areas on leaves five and six that were similar to those produced by MSV-MatA in this host (Fig 5.6A, 2:4 and 2:5 respectively). The differential virulence of MatLIRMPCPR2 relative to MSV-R2 in sweetcorn and PAN6099 again implies the existence of a host specific modification of the interaction between sequences in the MSV-R2 LIR and MP and/or CP genes.

The MSV-MatA/VW chimaera, VWLIRMat, produced significantly smaller chlorotic areas than did MSV-VW in sweetcorn and, like MSV-VW, was not infectious in PAN6099 (Fig 5.6A panels 1:3 and 1:6, respectively; Fig 5.6B). VWLIRMPCPMat, however, produced slightly more severe symptoms than did VWMPCPMat in sweetcorn and, like VWMPCPMat, produced mild symptoms in PAN6099 (Fig 5.5A panels 4:3 and 4:6; Fig 5.6A panels 2:3 and 2:6, respectively; Fig 5.6B). Considering the apparent interactions between the LIRs and MP and/or CP genes of MSV-MatA and -R2, it is perhaps surprising that MatLIRVW and MatLIRMPCPVW (the predominantly MSV-MatA chimaeras containing the MSV-VW LIR) were not substantially less fit than they were determined to be. While both MatLIRVW and MatLIRMPCPVW were respectively less pathogenic than MSV-MatA and MatMPCPVW in both sweetcorn and PAN6099 (Fig 5.6A panel 1:3 and compare panel 2:3 with Fig 5.5A panel 4:3), the differences in chlorotic areas, stunting and infection rates were only slight. The chlorotic areas produced by MatLIRVW were similar to those produced by MatLIRKom in sweetcorn, but were significantly greater in PAN6099 (Fig 5.6A, compare blue plots panels 1:1 and 1:4 with green plots in panels 1:3 and 1:6, respectively).

5.3.4 The SIR and Complementary Sense ORFs (*RepA* and C2)

The genomic region that apparently had the least influence on MSV pathogenicity in maize was the SIR. MSV-MatA/Kom and MatA/R2 chimaeras differing in only their short intergenic regions (MatSIRX and MatA, XSIRMat and X, MatLIRSIRX and MatLIRX, XLIRSIRMat and XLIRMat, MatRepX and XLIRMPCPMat, and XRepMat and MatLIRMPCPX where X is either Kom or R2) were equally infectious and produced indistinguishable chlorotic area and stunting symptoms in both sweetcorn and PAN6099 (Fig 5.7A panels 1:1, 1:2). While these MSV-MatA/Kom chimaeras also produced indistinguishable chlorotic areas in PAN6099, the MSV-Mat/R2 chimaeras, MatSIRR2 and R2SIRMat consistently produced slightly smaller chlorotic areas on leaves three through six than those produced by wt MSV-MatA and R2, respectively (Fig 5.7A panel 1:5).

For the MatA/VW chimaeras differing in only their SIRs, slight but significant reductions in chlorotic areas were produced in sweetcorn by MatSIRVW relative to MSV-MatA (Fig 5.7A panel 1:3), MatLIRSIRVW relative to MatLIRVW (compare blue plots in Fig 5.7A panel 3:3 and Fig 5.6A 1:3), MatRepVW relative to VWLIRMPCPMat (compare blue plot in panel Fig 5.7A panel 4:3 and green plot in Fig 5.6A panel 2:3), MatMPCPSIRVW relative to MatMPCPVW (compare blue plots in Fig 5.7A panel 2:1 and Fig 5.5A panel 4:3), and VWRepMat relative to MatLIRMPCPVW (compare green plot panel Fig 5.7A panel 4:3 and blue plot in Fig 5.6A panel 2:3). While VWSIRMat only produced smaller chlorotic areas than MSV-VW on leaf 3 (Fig 5.7A panel 1:3), chlorotic areas produced by VWLIRSIRMat and VWMPCPSIRMat were indistinguishable from those produced by VWLIRMat and VWMPCPMat, respectively (compare green plots in Fig5.7A panels 3:3 and 2:3 with those in Fig 5.6A panel 1:3 and Fig 5.5A panel 4:3, respectively). The infection rates achieved in sweetcorn by the chimaeras VWSIRMat, VWLIRSIRMat, VWMPCPSIRMat, MatRepVW (all containing a MSV-MatA SIR and a MSV-VW *Rep*; Fig 5.7B) did not differ significantly from those achieved by MSV-VW, VWMPCPMat (Fig 5.5B), VWLIRMat, and VWLIRMPCPMat (Fig 5.6B; all containing the MSV-VW SIR and *Rep*), respectively. However, the infection rates in sweetcorn of the chimaeras MatSIRVW, MatMPCPSIRVW, MatLIRSIRVW, VWRepMat (all containing the MSV-VW SIR and MSV-MatA *Rep*; Fig 5.7B) were significantly reduced relative to those of MSV-MatA, MatMPCPVW (Fig 5.5B), MatLIRVW, and MatLIRMPCPVW (Fig 5.6B; all containing the MSV-MatA SIR and *Rep*), respectively.

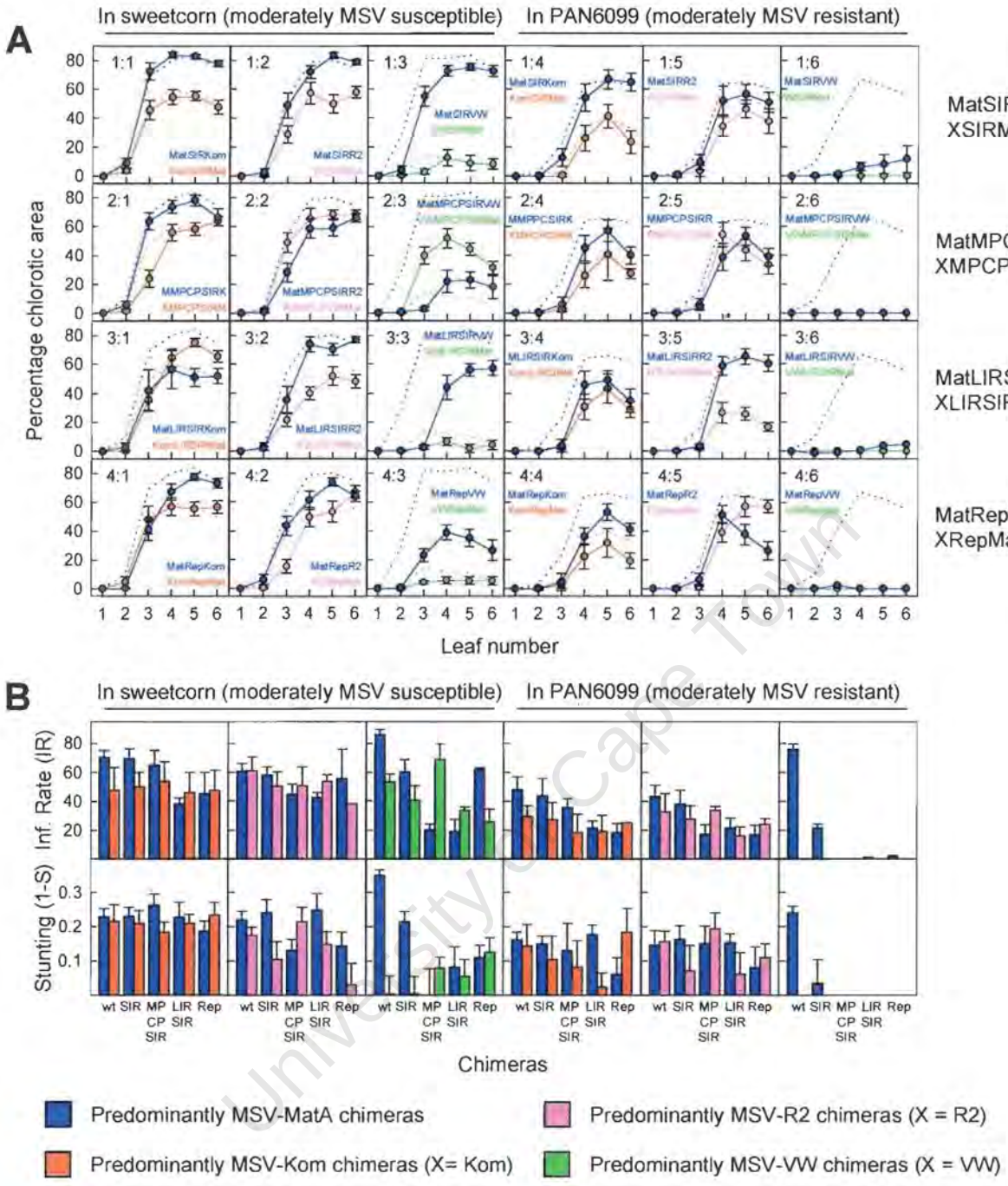


FIGURE 5.7 The virulence of short intergenic region (SIR), SIR + *MP* + *CP*, LIR + SIR, and Rep chimeras. (A) The chlorotic areas produced in sweetcorn (a moderately MSV-susceptible maize genotype) and PAN6099 (a moderately MSV-resistant maize genotype) agroinoculated with the chimeras. Chimeras containing predominantly MSV-MatA sequences are in blue, chimeras containing predominantly MSV-Kom sequences are in orange, chimeras containing predominantly MSV-R2 sequences are in pink, and chimeras containing predominantly MSV-VW sequences are in green. Broken lines representing the chlorotic areas produced by wt viruses are presented for comparative purposes. In certain cases the names of chimeras presented within each of the panels have been abbreviated due to space constraints. For example MatMPCPSIRKom has been abbreviated to MMPCPSIRK in panel 2:4. (B) Infection rates and stunting induced by wt and chimeric viruses. All error bars represent 95% confidence intervals (Student's t-test).

In PAN6099, the chimaeras MatSIRVW (Fig 5.7A panel 1:3) and MatLIRSIRVW (Fig 5.7A panel 3:3) produced substantially less streaking than did MSV-MatA and MatLIRVW (Fig 5.6A, panel 1:3), respectively. Whereas MatMPCPVW, VWMPCPMat and VWLIRMPCPMat, were able to produce symptoms in PAN6099, no symptoms were produced in this maize genotype by MatMPCPSIRVW, VWMPCPSIRMat, and VWRepMat.

5.3.5 Factors Determining the Host Range Limitation of MSV-VW

MSV-VW, like all non MSV-A isolates investigated in Chapter 4, is unable to infect the moderately MSV-resistant maize genotype, PAN6099. MSV-MatA/VW chimaeras were tested for their ability to produce symptoms in PAN6099 to determine which genomic components were primarily responsible for the inability of MSV-VW to infect this maize genotype. All MSV-MatA/VW chimaeras that contained one MSV-VW and four MSV-MatA genomic regions (MatLIRVW, MatMPVW, MatCPVW, MatSIRVW, and MatRepVW) were able to produce symptoms in PAN6099 indicating that no single MSV-VW component is solely responsible for MSV-VW being unable to produce symptoms in this host. Chimaeras containing more than one MSV-VW genomic region which produced very mild symptoms in PAN6099 were MatMPCPVW, VWMPCPMat (Fig 5.5B), MatLIRMPCPVW, VWLIRMPCPMat (Fig 5.6B) and MatLIRSIRVW (Fig 5.7B). The latter two genomes were the only ones which contained predominantly MSV-VW sequences. With the exception of MatLIRVW, MatMPVW, and MatSIRVW, the symptoms produced by all MSV-MatA/VW chimaeras in PAN6099 were extremely mild and chlorotic areas on individual leaves never exceeded 5%.

5.3.6 The Modularity of Genomic Components

In an attempt to understand the pattern of recombination observed amongst naturally occurring MSV chimaeras identified in Chapter 4, the interaction index (I_n) values of analogous MSV-MatA/Kom, MSV-MatA/R2 and MSV-MatA/VW reciprocal LIR, *MP*, *CP*, SIR and *Rep* chimaeras were compared. These comparisons were made in both sweetcorn and PAN6099. In almost all cases, the I_n values of analogous reciprocal chimaera pairs decreased with increasing nucleotide sequence divergence of the wt parental MSV isolates used in their construction (Fig 5.8). The rate at which I_n decreased with increasing divergence of parental genotypes differed for the different genomic components both within and between the different maize genotypes. The

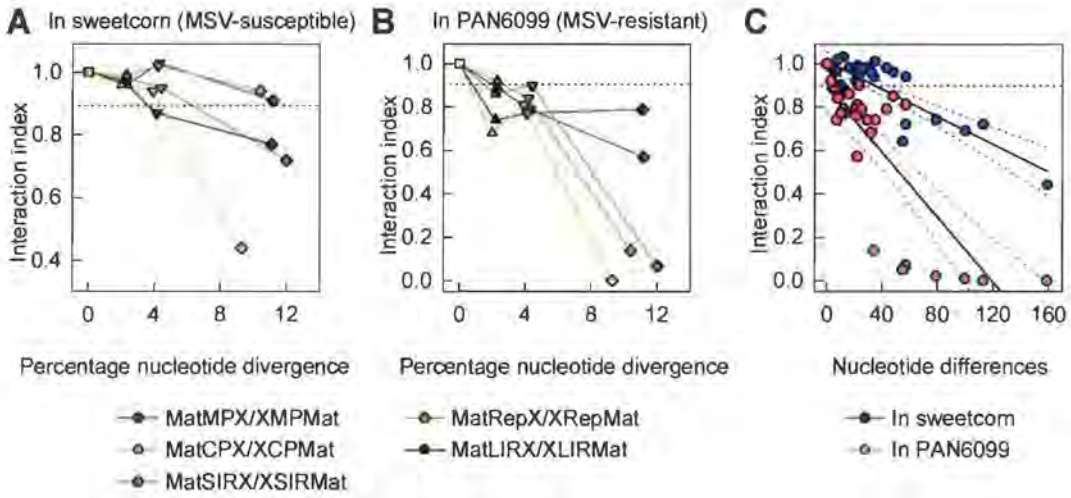


FIGURE 5.8 The modularity of MSV genomic components. Reciprocal chimera pairs containing individually transferred genomic components (either *MP*, *CP*, LIR, SIR or *Rep* sequences) differ in the rate at which their interaction indices (I_n - calculated as described in the text) decrease with increasing divergence of parental genomes. The percentage nucleotide divergence values used in plots in graphs A and B were calculated from pair-wise alignments of parental genomes with the sequence fragment transferred between the genomes during chimera construction removed. The value is therefore a measure, from the perspective of transferred sequence fragments, of the differences between the "background" genomic sequences of chimera and parental genomes. (A) Different symbols represent different chimera pairs: □ = wt genomes (by definition every pair has a I_n = 1); △ = MSV-MatA/Kom reciprocal chimera pairs; ▽ = MSV-MatA/R2 reciprocal chimera pairs; ◇ = MSV-MatA/VW reciprocal chimera pairs. In sweetcorn the *MP* (blue plot) and SIR (green plot) are the most modular MSV genomic components, i.e. the average virulence of reciprocal chimera pairs containing these components from divergent parents is similar to the average virulence of the parental genomes even when the parental genomes are fairly distantly related. Whereas the LIR (brown plot) and *CP* (pink plot) have an intermediate modularity in sweetcorn, *Rep* (yellow plot) has the lowest modularity in this host, i.e. the average virulence of reciprocal chimera pairs containing Rep genes from divergent parental genomes is much lower than the average virulence of the parental genomes. (B) Coloured plots represent the same genome components as in graph A. Symbols represent the same chimera pairs as in graph A (C) Linear regression of I_n values for all chimera pairs in sweetcorn (blue points with blue broken lines indicating the 95% confidence intervals) and PAN6099 (pink points with pink broken lines indicating the 95% confidence intervals) against the number of nucleotide differences between chimera and wt genomes. All genomic components are substantially less "modular" (their I_n values decrease more rapidly with increasing divergence of parental genomes) in PAN6099 than they are in sweetcorn. The broken black line at I_n = 0.9 in graphs A, B and C represents a hypothetical threshold below which naturally occurring chimeras might be too defective to compete effectively with parental genomes. The threshold value of 0.9 is arbitrary and is included here for purely descriptive purposes.

rates at which I_n decreased in these sets of LIR, *MP*, *CP*, SIR and *Rep* chimaera pairs were used as indicators of the relative modularity of the different MSV genomic components. The modularity of a genomic component as it is defined here, describes the functionality of an MSV gene or intergenic region when it is transferred into genomic backgrounds that differ by varying degrees from the genomic background in which the gene or intergenic region evolved. In sweetcorn, the rate of I_n decrease with increasing divergence of genomic backgrounds is lowest for the *MP* and SIR chimaeras (Fig 5.8A) indicating that, in this host at least, the MSV *MP* and SIR have a greater degree of modularity than other genomic components. Similarly, it is apparent that whereas the MSV *CP* and LIR have intermediate degrees of modularity in sweetcorn, *Rep* is the least modular MSV genomic component in this maize genotype (Fig 5.8A). In PAN6099, however, the LIR is most modular followed by the *MP* and the SIR, *CP* and *Rep* have the lowest

degrees of modularity (Fig 5.8B). Also apparent in Fig 5.8 is that the average rates of I_n decrease with increasing parental genotype divergence is far greater in PAN6099 than it is in sweetcorn (Fig 5.8C).

5.4 DISCUSSION

MSV-A_1 is the most widely distributed MSV-A subtype and contains isolates that are among the most virulent maize infecting MSV isolates yet discovered (see Chapter 4). Because of this it is likely that the MSV-A_1 isolate, MSV-MatA, has a genome that contains what are among the most highly maize-adapted MSV components that have yet evolved. The primary aim of this study was to determine which features of the MSV-MatA genome are responsible for its extreme virulence in maize. Towards this end, MSV-MatA genomic components were exchanged with those of three other MSV isolates that have been determined to be less virulent than MSV-MatA in maize. While two of these isolates, MSV-Kom and MSV-R2 (also MSV-A isolates in subtypes MSV-A_4 and -A_6, respectively), are moderately virulent in maize, the third, MSV-VW (a MSV-B isolate), only produces very mild symptoms in MSV-susceptible maize genotypes (see Chapter 4). As with all other non MSV-A isolates so far examined, MSV-VW does not produce any symptoms in the moderately MSV-resistant maize genotype, PAN6099 (see Chapter 4). A secondary aim of this study was therefore to determine which features of the MSV-VW genome prevent it from infecting PAN6099. During these investigations substantial genetic evidence of extensive functional interactions between MSV genomic components was accumulated (Fig 5.9).

The minimum portion of MSV-MatA's genome required to increase the virulence of MSV-MatA/Kom and MSV-MatA/R2 chimaeras to the same level of virulence as MSV-MatA differed between both the two groups of chimaeras and the two maize genotypes in which they were tested. Replacing the MSV-R2 *MP* and *CP* with those of MSV-MatA was sufficient to produce a chimaera (R2MPCPMat) that was as virulent as MSV-MatA in sweetcorn (Fig 5.5). To achieve a similar result in sweetcorn with a MSV-MatA/Kom chimaera, it was necessary to replace the LIR, MP and CP of MSV-Kom with those from MSV-MatA (to produce KomLIRMPCPMat; Fig 5.5). In PAN6099, however, no predominantly MSV-R2 or MSV-Kom derived chimaera produced symptoms approaching those of MSV-MatA. Whereas KomLIRMPCPMat was still significantly more virulent than wt MSV-Kom in PAN6099, R2MPCPMat in this host only produced symptoms as severe as MSV-MatA during the early stages of an infection. While these data suggest that the greater virulence of MSV-MatA relative to MSV-R2 and MSV-Kom in sweetcorn is almost entirely due to nucleotide variations in MSV-MatA's LIR, *MP* and *CP*, either host

factors alone or an interaction between host factors and unique sequences in MSV-MatA's *Rep* and/or SIR, are responsible for this isolate's greater virulence in PAN6099.

The individual contributions of MSV-MatA's LIR, *MP* and *CP* to its virulence is somewhat difficult to determine. Because the MSV-MatA *CP* has no detectable effect on the virulence of the chimaera R2CPMat relative to wt MSV-R2, a synergistic interaction between the MSV-MatA *MP* and *CP* in sweetcorn could explain why R2MPCPMat is more virulent than R2MPMat. Evidence of a synergistic interaction between MSV-MatA's *MP* and *CP* was also detected amongst the MSV-MatA/VW chimaeras. While VWMPMat and VWCPMat are both only slightly more virulent than MSV-VW, in sweetcorn VWMPCPMat produces substantially more severe symptoms than either VWMPMat or VWCPMat (Fig 5.9).

Whereas direct binding of MSV MP to CP has not been clearly demonstrated, MSV MP transiently expressed together with a CP-green fluorescence protein (GFP) fusion in maize redirects a small proportion of the CP:GFP fusion from the nucleus to the cell periphery (Kotlizky *et al.*, 2000). It is therefore likely that the apparent synergistic interaction detected between the MSV-MatA *MP* and *CP* in R2MPCPMat and VWMPCPMat reflects a direct protein-protein interaction between MSV CP and MP. It can not be discounted, however, that specific nucleotide sequences within the MSV-MatA *MP* that are absent in the MSV-R2 or MSV-VW *MP*s may be required for optimal MSV-MatA CP expression. The *CP* promoter contains the entire MP gene and has been mapped to a region 530 nucleotides upstream of the *CP* start codon (Fenoll *et al.*, 1988,1990).

The results presented here provide evidence that in MSV the nucleotide or amino acid sequence requirements for optimal *MP*/MP interaction with *CP*/CP are relatively strict. The MSV-MatA CP amino acid sequence differs from those of MSV-R2 and -VW by only 2 and 10 amino acids, respectively, with every difference occurring in the N-terminal 100 amino acids of these proteins. MP sequences are slightly more variable and, relative to the MSV-MatA MP, those of MSV-R2 and -VW differ by 5 and 12 amino acids, respectively. All except two of these differences in the MSV-VW MP occur at the N-terminus between amino acids 6 and 23, and at the C-terminus between amino acids 82 and 97. If the *MP-CP* interactions detected in this study were the result of amino acid sequence specific MP:CP binding, then it follows that at least one of the following regions is involved in this interaction: the N-terminal 23 amino acids of the MP; the C-terminal 17 amino acids of the MP; the N-terminal 100 amino acids of the CP. It has been demonstrated that altering the 14 N-terminal MP amino acids of MSV-Ns has almost no impact on the virulence of

this MSV isolate (Boulton *et al.*, 1989) and it is therefore likely that any amino acid sequence specific interaction between the MSV MP and CP will involve the MP C-terminal and/or the CP N-terminal regions.

There is also evidence of interaction between the MSV-MatA LIR and *MP* in the MSV-MatA/Kom, MSV-MatA/R2 and MSV-MatA/VW chimaeras. In sweetcorn, the MSV-MatA LIR in the chimaera KomLIRMat and the MSV-MatA MP in the chimaera KomMPMat can apparently increase the virulence of chimaeras relative to wt MSV-Kom in the absence of any other MSV-MatA sequences. In PAN6099, however, the individual contributions of these genomic regions in the absence of other MSV-MatA sequences are either slight (KomMPMat) or undetectable (KomLIRMat). Whereas the MSV-MatA-like virulence of KomLIRMPCPMat in sweetcorn could simply be the additive result of this chimaera having individually improved LIR and *MP* sequences, its significantly increased virulence relative to MSV-Kom in PAN6099 (where these sequences are not individually better than MSV-Kom sequences) is possibly the result of a synergistic interaction between the MSV-MatA sequences within KomLIRMPCPMat.

Interaction between the MSV LIR and *MP* and/or *CP* sequences was also detected amongst the MSV-MatA/R2 and MSV-MatA/VW chimaeras. Unlike the enhancement of virulence observed with KomLIRMat relative to wt MSV-Kom, R2LIRMat and VWLIRMat were both significantly less virulent than wt MSV-R2 and MSV-VW, respectively. However, in sweetcorn R2LIRMPCPMat was as virulent as R2MPCPMat and VWLIRMPCPMat was slightly more virulent than VWMPCPMat, indicating that virulence can be recovered and even enhanced when the MSV-MatA *MP* and *CP* are present together with the MSV-MatA LIR .

While these results indicate that sequences within the MSV LIR interact either positively or negatively with sequences in the MSV virion sense genes, they do not indicate whether it is *MP* or *CP* alone, or *MP* and *CP* sequences together that are involved. If the MSV-MatA LIR was interacting poorly with MSV-R2 and MSV-VW *MP* and/or *CP*, one might expect to find evidence of this by examining the virulence of the chimaeras MatMPR2, MatCPR2, MatMPVW and MatCPVW relative to MSV-MatA. However, only the absence of a negative interaction would be demonstrable because any decrease in the virulence of these chimaeras relative to MSV-MatA could simply be due to the MSV-R2 and MSV-VW genes functioning less efficiently than those of MSV-MatA. Because the chimaera MatCPR2 was determined to be as virulent as MSV-MatA, it is unlikely that the MSV-MatA LIR interacts negatively with the MSV-R2 *CP*. It follows therefore that it is probably a negative interaction between the MSV-MatA LIR and MSV-R2 *MP*

that is responsible for the reduced virulence of R2LIRMat relative to wt MSV-R2. This could also explain why in PAN6099 the chimaera MatMPR2 is significantly less virulent than wt MSV-R2.

There is, however, one slight complication to this scheme. The difference in virulence between MSV-R2 and any of the chimeras containing the MSV-MatA LIR together with the MSV-R2 *Rep* (MatRepR2, R2LIRMat, R2LIRMPCPMat, and R2LIRSIRMat) is much greater in PAN6099 that it is in sweetcorn. However, the difference in virulence between MSV-R2 and any of the chimeras containing the MSV-R2 LIR together with the MSV-MatA *Rep* (MatLIRR2, MatLIRMPCPR2, MatLIRSIRR2, and R2RepMat) is either the same in both sweetcorn and PAN6099 (MatLIRR2, MatLIRSIRR2, and R2RepMat) or is greater in sweetcorn than it is in PAN6099 (MatLIRMPCPR2). This indicates that, in PAN6099 there is possibly a negative interaction between the MSV-MatA LIR and the MSV-R2 *Rep* that either does not occur or is not detectable in sweetcorn.

It is possible that negative interactions detected between the MSV-MatA LIR and the MSV-R2 Rep and *MP* may be due to an interference in either viral replication, or in the expression of *MP*, *Rep* and *RepA*. In the mastrevirus *Wheat dwarf virus* (WDV), Rep and RepA oligomers both bind within the LIR ~144 nucleotides upstream and ~90 nucleotides downstream of the (+) strand origin of replication to form C- and V-complexes, respectively (Castellano *et al.*, 1999; Missich *et al.*, 2000). Rep also forms an O-complex at the origin of (+) strand replication (Castellano *et al.*, 1999). Whereas it is assumed that O-complex formation is involved in replication, it is presently unknown what roles C- and V-complex formation play in the WDV life cycle. It has been speculated that C-complex formation may be part of a mechanism whereby Rep and/or RepA autoregulate their expression and that V-complex formation is involved in the induction of virion sense gene expression late during an infection cycle (Castellano *et al.*, 1999).

While it is still unknown whether MSV Rep binds at similar positions within the LIR to those noted for WDV, the portion of the MSV-MatA LIR that was transferred during chimaera construction contains sequences corresponding to half of a potential C-complex binding region and the entire O- and V-complex binding regions. Negative interaction between the MSV-MatA LIR and the MSV-R2 *Rep* and *MP* could, therefore, be the result of an interference in MSV-R2 Rep/RepA complex formation on the MSV-MatA LIR which subsequently affects virion and/or complementary sense gene transcription. There are 6 nucleotide differences between the MSV-MatA and MSV-R2 LIRs in the region where MSV V-complex formation might occur – between

the TATA box believed to direct most virion sense gene transcription (Wright *et al.*, 1997) and the base of the stem loop. Downstream of the TATA box believed to direct complementary sense gene transcription (Boulton *et al.*, 1991a) there are three nucleotide differences between MSV-MatA and MSV-R2 (including one additional nucleotide in MSV-MatA six nucleotides downstream from the TATA box) in the region where MSV Rep C-complex formation might occur. Potential Rep recognition sequences have been identified in WDV (Castellano *et al.*, 1999), and any of the nucleotide differences between the MSV-MatA and MSV-R2 LIRs might be in analogous MSV Rep recognition sequences.

The reason that the MSV-MatA LIR and the MSV-R2 *Rep* interact negatively only in PAN6099 may be related to the activity of the MSV resistance genes that this maize genotype possesses. It is currently not known how any of the resistance genes that have been pooled by breeders into maize genotypes such as PAN6099 functions. It is conceivable that one of these genes in PAN6099 operates by interfering with some aspect of Rep and/or RepA interaction with the LIR, and that this interference exaggerates the effects of an already slightly suboptimal MSV-MatA LIR – MSV-R2 *Rep* interaction.

The only detectable effect of different SIR sequences on the virulence of chimaeras was observed with the MSV-MatA/VW chimaeras in both sweetcorn and PAN6099. In these chimaeras there is apparently a negative interaction between the MSV-MatA *Rep* and MSV-VW SIR. With all chimaeras containing the MSV-MatA *Rep* and the MSV-VW SIR (MatSIRVW, MatLIRSIRVW), MatCPMPSIRVW, and VWRepMat), there was both a substantial reduction in infection rates and a slight decrease in the chlorotic areas produced in symptomatic plants relative to related chimaeras containing MSV-MatA SIR sequences. However, in VW chimaeras containing the MSV-VW *Rep* and MSV-MatA SIR there was either an undetectable (VWMLIRSIRMat, VWMPCPSIRMat, MatRepVW) or only minor (VWSIRMat) decrease in virulence relative to related chimaeras containing the MSV-VW SIR. It was demonstrated in Chapter 4 that, along with certain other MSV-B isolates, MSV-VW contains a 55 nucleotide sequence within its SIR that is apparently of PanSV origin (see Chapter 4, Fig 4.6). Because of this recombinant region, the SIRs of MSV-MatA and -VW differ at 32 nucleotide positions, 29 of which occur within the putative recombinant region. Interestingly, this region falls within the ~80 nucleotide portion of the SIR to which a small complementary primer-like DNA has been found bound in encapsidated MSV-Ns virions (Donson *et al.*, 1984). Because this primer-binding region is believed to be the origin of (-) strand synthesis it is possible that the reduced fitness of chimaeras containing the

MSV-MatA *Rep* and the MSV-VW SIR is due to reduced efficiency in the initiation or termination of (-) strand synthesis during viral DNA replication in maize.

Very little is known about the origin of (-) strand replication. It has been demonstrated that a *Bean yellow dwarf virus* (BeYDV; a dicot-infecting mastrevirus) chimaera containing a MSV SIR is able to replicate in tobacco protoplasts (Liu *et al.*, 1999b). Because the MSV and BeYDV SIRs share only 49% nucleotide sequence identity in the region believed to be involved in primer binding, it is highly unlikely that there are very strict virus isolate-specific nucleotide sequence requirements within the SIR for the initiation of (-) strand synthesis. It is possible, however, that (-) strand synthesis initiated from the MSV-VW SIR may occur slightly less efficiently than it would from the MSV-MatA SIR. While there is no evidence for the direct involvement of any geminivirus Rep in (-) strand synthesis (Gutierrez, 1999), less efficient (-) strand replication in maize from the MSV-VW SIR could explain both its apparent interaction with the MSV-MatA *Rep* and the lack of a detectable interaction between the MSV-VW *Rep* and MSV-MatA SIR. In chimaeras containing the MSV-MatA Rep and the MSV-VW SIR the rate-limiting step during rolling circle replication (RCR) is possibly the suboptimal rate of (-) strand synthesis. However, in chimaeras containing the MSV-VW Rep and either the MSV-MatA or MSV-VW SIR, the inefficient mediation of (+) strand replication by MSV-VW Rep in maize is possibly rate-limiting regardless of whether (-) strand replication occurs at an optimal rate from the MSV-MatA SIR or a suboptimal rate from the MSV-VW SIR. The virulence difference between MSV-MatA and MatSIRVW is much larger in PAN6099 than it is in sweetcorn. If (-) strand replication in maize is less efficient from the MSV-VW SIR, than it is from the MSV-MatA SIR, then the much lower infectivity of MatSIRVW relative to MSV-MatA in PAN6099 may indicate another possible point of interference in the MSV life-cycle by the MSV resistance genes of PAN6099.

The inability of MSV-VW to infect PAN6099 is probably due to a combination of sequence determinants primarily situated in its *CP*, SIR and *Rep*. Replacing any one of these regions with the corresponding genomic region from MSV-MatA (VWCPMat, VWSIRMat and VWRepMat) was, however, not sufficient to enable infection of PAN6099. Conversely, predominantly MSV-MatA chimaeras containing any one of these components from MSV-VW (MatCPVW, MatSIRVW and MatRepVW respectively) all produced mild symptoms in PAN6099, meaning that none of the components absolutely prevented infection of this maize genotype. MatCPVW, MatSIRVW and MatRepVW were, however, all severely defective in PAN6099 relative to MatMPVW and MatLIRVW (predominantly MSV-MatA chimaeras containing the MSV-VW *MP* and LIR, respectively). While the MSV-VW *CP* and *Rep* had the greatest individual effects on

the virulence of chimaeras in sweetcorn, part of the *CP*'s impact may have been due to the disturbance of a MSV-VW *CP* - *MP* interaction analogous to that detected between the MSV-MatA *MP* and *CP*. However, no increase in virulence was detected in MatMPCPVW relative to MatCPVW in either sweetcorn or PAN6099, and it is possible that sub-optimal functioning of both the MSV-VW MP and CP in maize are additively responsible for the decreased virulence of MatMPCPVW relative to MatMPVW and MatCPVW.

Both VWMPCPMat and VWLIRMPCPMat contain the MSV-VW *Rep* and SIR and both are able to infect PAN6099. This implies that the MSV-VW *CP* probably contains the most significant sequence determinants preventing this isolate from infecting PAN6099. Every one of the 10 amino acids which differ between the MSV-MatA and -VW CP sequences occurs between amino acids 14 and 99 in the region believed to be involved in sequence non-specific ss and ds DNA binding (Liu *et al.*, 1997). Also, three of these 10 differeing amino acids are within the predicted MSV CP nuclear localisation signal (NLS; Liu *et al.*, 1997). Because the DNA binding domain and NLS of the MSV CP are probably involved in the nuclear import of viral DNA during movement (Kotlizky *et al.*, 2000; Liu *et al.*, 1999a), it is possible that the host range limitation imposed by the MSV-VW *CP* is due to inefficient CP interaction with PAN6099 cellular components involved in the transport of viral DNA between the cytoplasm and nucleus.

It has been demonstrated here that different components of MSV genomes have differing degrees of modularity and that the relative modularity of components can vary with host genotype. This observation may provide some insight into the pattern of recombination noted in natural MSV recombinants. Recombinant genomes that arise in nature probably contain recombinant fragments of random length (but less than half the genome length) in random positions within their genomes. One might assume that following its generation, the long-term survival of a recombinant genome would depend on it being at least as virulent as the most virulent of its parental genomes. A selection process would almost certainly result in only a minute portion of all the possible recombinants that get produced existing for long enough to be detected. The results presented here indicate that the position and size of a recombinant region, the relatedness of a recombinant's parents and the host genotype in which a recombinant is generated, are all likely to be significant factors determining the survival of a recombinant genome.

Differences in the interaction indices (I_n) of the various genomic components examined imply that recombination events occurring at certain positions within an MSV genome are more likely to yield defective chimaeras than those which occur at other positions. Sequences within genomic

regions with low I_n values apparently require co-evolved sequences in other parts of their native genomes for optimal activity. An example of this is the suboptimal activity of MSV-MatA CP gene in the absence of the MSV-MatA MP gene. While larger recombinant regions might encompass a set of interacting sequence determinants with low individual I_n values but which have a high joint I_n (an example of this is the MSV-MatA *MP* and *CP* in chimaeras such as R2MPCPMat and VWMPCPMat), the chance of a recombinant region spanning a sequence with a low I_n will probably increase with the size of the region. The decrease in I_n values observed with increasing divergence of parental genomes implies that as parental genomes become more divergent, fewer and smaller genomic regions retain I_n values high enough to ensure their survival within any recombinants that are produced (see hypothetical I_n thresholds in Fig 5.8).

The finding that the decrease of I_n values with increasing divergence of parental genomes occurs at a more rapid rate in PAN6099 than it occurs in sweetcorn may indicate that the survival prospects of recombinant genomes are far better in certain host species or genotypes than in others. It is possible that part of the reason reciprocal chimaeras containing MSV-MatA components had higher I_n values in sweetcorn than in PAN6099 is that MSV-MatA may be "over-virulent" in this host, i.e. MSV-MatA has possibly reached and surpassed the sequence requirements for the production of maximal symptoms in sweetcorn. If this were the case then predominantly MSV-MatA chimaeras with substantially reduced fitness relative to wt MSV-MatA could appear as virulent as wt MSV-MatA in sweetcorn due to their still being virulent enough to produce maximal symptoms in this host. Because MSV-MatA does not produce maximal symptoms in PAN6099 the fitness difference between MSV-MatA and these chimaeras would be instantly apparent in this host. The net result of this would be the significant reduction that was observed in the I_n values of reciprocal chimaera pairs between sweetcorn and PAN6099. The possibility that over-virulence of MSV isolates in MSV-susceptible maize genotypes facilitates the generation of recombinant genomes may have serious evolutionary and epidemiological implications and certainly warrants further investigation.

While none of the chimaeras produced in this study was significantly more virulent than both of its parental genomes, this does not imply that other MSV-MatA/Kom, -MatA/R2 or -MatA/VW chimaeras that are more virulent than MSV-MatA could not exist. Only a very small fraction of the possible chimaeric genomes that could have been constructed were produced in this study. The naturally occurring recombinant genomes described in Chapter 4 are possibly descendents of rare chimaeras that were fitter (at least under certain circumstances) than both of their parental

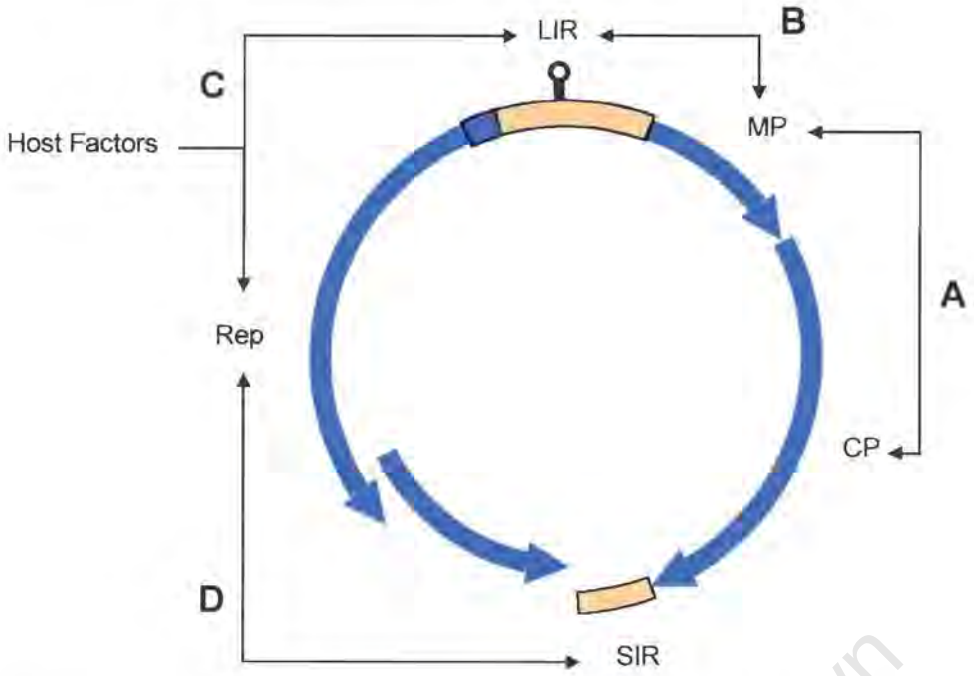


FIGURE 5.9 A schematic representation of the nucleotide sequence dependent genomic component interactions that have been detected in this study. LIR = genomic component containing most of the long intergenic region (the position of the (+) strand origin of replication is indicated), SIR = genomic component containing the short intergenic region, MP = genomic component containing the movement protein gene, CP = genomic component containing the coat protein gene, and Rep = genomic component containing the RepA and C2 ORFs and a 72 nucleotide fragment of the LIR. **A** MSV-MatA/R2 and MSV-MatA/VW chimaeras containing both the MSV-MatA MP and CP genes are substantially more virulent than chimeras containing either gene alone. **B** MSV-MatA/Kom, chimaeras containing the MSV-MatA LIR, *MP*, and *CP* are more virulent than chimeras containing either the LIR or the MP and CP genes alone. The impaired virulence of MSV-MatA/R2 and MSV-MatA/VW chimeras containing the MSV-MatA LIR alone was recovered in chimaeras containing the MSV-MatA LIR, *MP* and *CP* together. **C** Relative to wt MSV-MatA and MSV-R2, MSV-MatA/R2 chimaeras containing the MSV-MatA LIR and MSV-R2 *Rep* were considerably less virulent in a MSV-resistant maize genotype (PAN6099) than in a MSV-susceptible genotype (sweetcorn). **D** MSV-MatA/VW chimeras containing both the MSV-MatA *Rep* and SIR had significantly greater infection rates than related chimeras containing either of these components alone. The interaction between *Rep* and the SIR is almost certainly not direct and probably operates through their respective involvement in (+) and (-) strand synthesis during viral rolling circle replication.

genomes. Nevertheless, there is a definite correlation between the virulence of chimaeras generated here and the pattern of natural recombination that has been observed. The inter-subtype recombination events noted amongst MSV-A isolates in Chapter 4 all involved the exchange of either *CP* (MSV-MatC and MSV-MakD) or *RepA* (MSV-Ama, MSV-MakD) sequences. It is demonstrated here that individually exchanging *CP* and *Rep* sequences between MSV-A isolates (MSV-MatA, MSV-Kom and MSV-R2) has very little impact on the virulence of chimaeras in sweetcorn and yields reciprocal chimaera pairs with I_n values greater than 0.94 in this host (Fig 5.8A). Inter-strain recombinants that were detected in Chapter 4 included MSV-A isolates with MSV-B sequences at either the 3' end of their C2 ORFs (MSV-Kom, MSV-SA and MSV-VW) or within their MP gene (MSV-VW). Replacement of MSV-MatA's *Rep* with the "naturally" recombinant *Rep* of MSV-Kom in the chimaera MatRepKom had only a slight effect on the

virulence of this chimaera in sweetcorn relative to MSV-MatA. This implies that the original recombination event that produced the chimaeric MSV-Kom ancestor probably had little or no effect on the virulence of the ancestor in MSV-susceptible maize genotypes. The reciprocal chimaera pair MatMPVW and VWMPMat has an I_n value of 0.91 in sweetcorn (Fig 5.8A) and accordingly only a slight reduction in virulence relative to MSV-MatA was noted in this host for MatMPVW – a chimaera which resembles the naturally occurring MSV-A/MSV-B *MP* recombinant, MSV-VM (see Chapter 4, Fig 4.6). It was demonstrated in Chapter 4 that MSV-VW is an inter-species recombinant containing a piece of PanSV-like sequence within the putative primer binding region of its SIR. Despite containing this already recombinant SIR, the chimaera MatSIRVW was only slightly less virulent than MSV-MatA in sweetcorn. An I_n of 0.94 in sweetcorn (Fig 5.8A) for the chimaera pair MatSIRVW and VWSIRMat indicates that the original MSV/PanSV SIR recombinant which has evolved into MSV-VW was probably as virulent in maize (and possibly other grass species) as its non-recombinant parental MSV genome.

The pathogenic properties of MSV-MatA that are responsible for it being amongst the most virulent MSV isolates ever examined, have been mostly localised to a portion of its genome containing the LIR, *MP* and *CP*. While increasing the virulence of the predominantly MSV-R2, -Kom and -VW chimaeras, these MSV-MatA genome components also enabled predominantly MSV-VW chimaeras to produce a symptomatic infection of PAN6099. Genetic evidence was obtained for functional interactions between sequence determinants within the *MP* and *CP*; the LIR and *MP*; the LIR, host factors and *Rep;* and the SIR and *Rep* (Fig 5.9). While the exact nature of these interactions remains to be determined, these results are reminiscent of those which first demonstrated intergenic region - Rep interactions in curtoviruses (Stenger *et al.*, 1994) and movement protein - nuclear shuttle protein interactions in begomoviruses (Schaffer, *et al.*, 1995). Because the genomic interactions detected here are apparently most productive when genomic components have co-evolved in the same isolate, it is proposed that maintenance of optimal genomic component interactions is a major contributing factor to the pattern of recombination observed amongst natural MSV recombinants.

Chapter 6

Conclusion

As the causal agent of maize streak disease (MSD), MSV is currently the most significant pathogen of maize in Africa. It is believed that the generation and widespread use of high yielding MSV-resistant maize genotypes is a viable solution to the African MSD problem. Maize breeders throughout Africa have risen to this challenge and have very successfully produced maize genotypes that are either immune to, or highly tolerant of, MSV infections. The most pressing current problem experienced by MSV resistance breeders is probably best described by J. B. J. Van Rensburg who has commented that: "The problem with producing high yielding MSV resistant maize genotypes is not the generation of maize lines that are immune to MSV infection. That is easy. The problem is that genes for resistance to MSV are tightly linked with genes for resistance to yield" (Van Rensberg, 1997). Because of the erratic nature of MSV epidemics, few farmers are willing to accept the yield losses that are coupled with the security of having a resistant crop in the relatively unlikely event of an epidemic.

As with most diseases vectored by an insect, the epidemiology of MSD is complex and has currently defied successful modelling. While the broad host range of MSV amongst African grass species almost certainly compounds this complexity, the work presented in Chapter 4 of this thesis indicates that the genotypic and phenotypic diversity of MSV populations may also substantially increase the difficulty of predictively modelling MSD epidemics. While five major MSV strain groupings have been discovered (two of which are described for the first time in this thesis), it has been apparent for quite some time that only one of these strains, MSV-A, is responsible for serious disease in maize. It has, however, been demonstrated in Chapter 4 that not only is there a measurable degree of genotypic diversity amongst MSV-A isolates, but that this diversity is reflected in differences in the severity of disease symptoms produced by these isolates. A preliminary investigation into the distribution of MSV-A subtypes throughout Africa also revealed that these subtypes have a degree of geographical localisation. Together, these findings have potential significance to both MSV epidemiologists and maize breeders. If the relative virulences of dominant MSV-A subtypes differ between regions, it may be necessary to substantially alter epidemiological models according to the regions where they are applied. For breeders these findings may imply that the resistance of maize genotypes produced against

certain MSV-A subtypes endemic to the regions where maize genotypes are screened for resistance, may be broken when these maize genotypes are challenged with different MSV-A subtypes. While these results might be bad news for breeders attempting to produce maize varieties that are generally suitable for growth throughout Africa, they may encourage breeders who are attempting to specifically tailor the MSV resistance of maize genotypes to suit particular regions. Knowing, for example, that the predominant MSV-A subtype in Southern Africa, MSV-A_4, is only moderately virulent may inspire the growth in certain parts of South Africa of maize varieties which are only moderately MSV resistant but which have significantly better yield potential than varieties that are completely immune.

The rates at which MSV genotypes are able to move from one region to another are currently unknown. The demonstration in Chapter 4 that certain MSV-A subtypes are found throughout Africa while others are apparently restricted to certain regions indicates that while movement of MSV genotypes must occur, it is possible that factors related to the genotype of viruses may influence their rate of spread across the continent. While it was found that MSV-A_1 and -A_5 isolates were amongst the most virulent tested in maize, these subtypes also have the widest distribution in Africa. Whereas the data presented in Chapter 4 is too preliminary to determine whether there is any real correlation between increased virulence and increased rate of spread, it should provide a valuable starting point for future studies tracking the growth, decline and spread of African MSV subtypes over time.

If the rates at which MSV-A subtypes move are high and the subtype composition of MSV populations in particular regions varies substantially from year to year, then breeders would need to abandon efforts to produce maize genotypes with region-specific resistance in favor of genotypes with multiple MSV-A subtype resistance. The task of screening maize genotypes against a range of differentially virulent MSV-A subtypes will be greatly simplified by employing a resistance evaluation technique such as the one described in Chapters 2 and 3 of this thesis. This technique provides a substantially expanded view of MSV resistance phenotypes than are detectable with currently used resistance screening approaches as it allows differentiation between both extreme resistance phenotypes conventionally considered uniformly immune by breeders, and marginal resistance phenotypes collectively identified by breeders as being extremely MSV sensitive. Use of such techniques could aid the future development of extremely durable MSV resistance in maize.

The battle between maize breeders and MSV will, no doubt, be fought many years into the future, and while it will involve the development and implementation of new technologies on the side of the breeders, it will also almost certainly involve the relentless evolution of ever increasing virulence on the side of the virus. It was demonstrated in Chapter 5 that the extreme virulence of the MSV-A_1 isolate, MSV-MatA, relative to MSV-A_4, MSV-A_6 and MSV-B isolates is primarily the result of minor differences in their LIR, *MP* and *CP* genes. There is no reason to believe that any of the MSV genotypes that are currently in existence are approaching a fitness peak in maize and it would be reasonable to expect that degrees of virulence greater than that of isolates such as MSV-MatA are possible. As has been demonstrated in Chapter 3, an MSV isolate such as MSV-MatA has already evolved sufficient virulence to produce symptoms in what are possibly the most MSV resistant maize genotypes ever produced. With the selection pressures that widespread use of resistant maize varieties will provide, the evolution of increased MSV virulence may be considerably hastened.

The process whereby increased virulence evolves will probably also involve the continual transfer of genetic material between MSV strains and subtypes through recombination. In Chapter 4, it was found that recombination between MSV genomes does occur. Although the occurrence of recombination does not necessarily equate with the evolution of increased virulence and the emergence of devastating new viral genotypes, from an evolutionary and epidemiological perspective it increases the significance of MSV strains and other African streak viruses that are not themselves very pathogenic in maize. Results presented in Chapter 5 indicate that variables such as the relatedness of parental viruses, the genomic regions they exchange and the host species or genotype within which they recombine will potentially influence the survival and spread of recombinant genomes that are generated. While there is absolutely no way to predict if or when inter-strain or inter-species recombinant MSV genomes with substantially increased virulence in maize might emerge, the data presented in Chapters 4 and 5 does suggest how these genomes might arise and what they could look like. They will probably be generated within maize varieties or grass species that are extremely sensitive to infection with MSV-A and will probably consist almost entirely of MSV-A sequences with a small fragment of *MP*, SIR or LIR sequence from another MSV-strain or African streak virus species.

Appendix A

Sequence Accession Numbers

Contents

TABLE A.1 GenBank accession numbers of sequences used in the construction of Figs 1.1 and 1.5 (see Chapter 1).

Genus	Name	Species-Major Strains/Variants	Accession Number
Mastrevirus	Maize streak virus	MSV -Ns,	KO2026
		MSV-Set	AF007881
		MSV- Tas	AF239962
		MSV-Ken	X01089
		MSV-Kom	AF003952
		MSV-SA	Y00514
		MSV-VM	AF239961
		MSV-Reu	X94330
		MSV-VW	AF239960
	Sugarcane streak virus	SSV-Asw	AF039528
		SSV-Ben	AF039529
		SSV-N	M82918
		SSV-Reu	AF072672
	Digitaria streak virus	DSV	M23022
	Panicum streak virus	PanSV-Ken	X60168
		PanSV-Kar	L39638
	Wheat dwarf virus	WDV	X02869
	Chloris striate mosaic virus	CSMV	M20021
	Miscanthus streak virus	MiSV	D00800
	Bean yellow dwarf virus	BeYDV	Y11023
	Tobacco yellow dwarf virus	TYDV	M81103
Curtovirus	Horseradish curly top virus	HrCTV	U49907
	Beet curly top virus	BCTV-Cfh	U02311
		BCTV-Wor	U56975
		BCTV-Cal	X04144
Topocuvirus	Tomato pseudo-curly top virus	TPCTV	X84735
Begomovirus	Cotton leaf curl virus	CLCuV-802a	AJ002452
		CLCuV-Okr	AJ002459
		CLCiV-PK1	AJ002448
		CLCuV-PK2	AJ002458
	Tomato leaf curl virus	ToLCV-Ban	Z48182
		ToLCV-I	U15016
		ToLCV-AU	S53251
		ToLCV-TW	U88692
		ToLCV-PA	Y15034
	Papaya leaf curl virus	PaLCV	AF134484
	Okra yellow vein mosaic virus	OYVMV-201	AJ002451
		OYVMV-301	AJ002453
		ITmLCV	Z48182
	Tobacco leaf curl virus	TbLCV	D88773
	Tomato yellow leaf curl virus	TYLCV-TH	X63015
		TYLCV-Sar	X61153
		TYLCV-Sar-Sic	Z28390
		TYLCV-Sar-Es	Z25751
		TYLCV-IL	X15656
		TYLCV-IL-mld	X76319
	Ageratum yellow vein virus	AYVV	X74516
	Indian cassava mosaic virus	ICMV	Z24759
	Chayote mosaic virus	ChaMV	AJ223191
	Althea rosea enation virus	AREV	AF014881
	East African cassava mosaic virus	EACMV –TZ	Z83256
		EACMV-UG	Z83257
		EACMV-CM	AF112354
	South African cassava mosaic virus	SACMV	AF155806
	African cassava mosaic virus	ACMV	X17095

TABLE A.1 Cont.

Genus	Name	Species-Major Strains/Variants	Accession Number
	Cowpea golden mosaic virus	CPGMV	AF029217
	Abutilon mosaic virus	AbMV	X15983
	Sida golden mosaic virus	SiGMV-HO	Y11097
		SiGMV-HO-YV	Y11100
		SiGMV-CR	U77964
		SiGMV-Flo	AF049336
	Tomato mottle virus	ToMoV	L14460
		ToMoV-Tai	AF012300
	Bean dwarf mosaic virus	BDMV	M88179
	Potato yellow mosaic virus	PYMV-VE	D00940
		PYMV-TT	AF039031
	Bean golden mosaic virus	BGMV-BR	M88686
		BGMV-PR	D00200
	Tomato golden mosaic virus	TGMV	K02029
	Cabbage leaf curl virus	CaLCuV	U65529
	Pepper golden mosaic virus	PepGMV	U57457
	Squash leaf curl virus	SqLCV	M38183
	Cotton leaf crumple virus	CLCrV	AF076851
	Pepper hausteco virus	PHV	X70418

TABLE A.2 GenBank accession numbers of sequences used in the construction of Fig 1.3 (see Chapter 1).

ssDNA replicons	Abbreviation	Virus/Plasmid Name (variant)	Accession Number
Geminiviridae			
	ACMV-UG	*African cassava mosaic virus* (Ug)	AAF42737
	BCTV	*Beet curly top virus* (Cal)	S28360
	BhYVMV	*Bhendi yellow vein mosaic virus*	AF241479
	ICMV	*Indian cassava mosaic virus*	AF241479
	MiSV	*Miscanthus streak virus*	JQ1358
	MSV	*Maize streak virus* (Ns)	A05158
	PanSV	*Panicum streak virus* (Kar)	JQ1552
	PHV	*Pepper hausteco virus*	JQ2300
	PYMV-VE	*Pepper yellow mosaic virus* (VE)	AF155806
	RhGMV	*Rhynchosia Golden Mosaic Virus*	AAF44669
	TbLCV	*Tobacco leaf curl virus*	BAA34016
	TGMV	*Tomato golden mosaic virus*	P03567
	ToLCV-In	*Tomato leaf curl virus* (In)	CAA88229
	ToLCV-MM	*Tomato leaf curl virus* (MM)	AF206674
	ToMoV-Flo	*Tomato mottle virus*	AF241479
	TYDV	*Tobacco yellow dwarf virus*	D42452
	TYLCV-Au	*Tomato yellow leaf curl virus* (Au)	JQ1887
	TYLCV-Is	*Tomato yellow leaf curl virus* (Is)	AF239671
	WDV	*Wheat dwarf virus*	B24356
Paracircoviridae			
	CAV	*Chicken anaemia virus*	M81223
Parvoviridae			
	CPV	*Canine parvovirus*	A29962
	MVM	*Minute virus of mice*	A29510
Nanoviridae			
	BBTV-T1	*Banana bunchy top virus* (T1)	BAA33981
	BBTV-T3	*Banana bunchy top virus* (T3)	CAA06791
	CFDV	*Coconut foliar decay virus*	A46353
	FBNYV-10	*Faba bean necrotic yellows virus* (10)	BAA34048
	FBNYV-2	*Faba bean necrotic yellows virus* (2)	CAB44020
	FBNYV-7	*Faba bean necrotic yellows virus* (7)	CAB44025
	FBNYV-9	*Faba bean necrotic yellows virus* (9)	CAB44027
	MVDV-1	*Milk vetch dwarf virus* (1)	BAA33980
	MVDV-3	*Milk vetch dwarf virus* (3)	BAA33982
	SCSCV-2	*Subterranean clover stunt virus* (2)	AAA68018
	SCSV-6	*Subterranean clover stunt virus* (6)	AAA68022
Circoviridae			
	BFDV	*Beak and feather disease virus*	AAC69861
	PCV-1	*Porcine circovirus* (1)	AF166528
	PCV-2	*Porcine circovirus* (2)	AAC98885
Microviridae			
	ϕK	-	P25244
	ϕ174	-	A04239
pUB110 Related Plasmid Family			
	pBC1	-	M64604
	pFTB14	-	S01098
	pKYM	-	M38574
	pLAB1000	-	B35390
	pNost	-	M81381
	pSK89	-	M37889
	pUB110	-	M19465

TABLE A.2 Cont.

ssDNA replicons	Abbreviation	Virus/Plasmid Name (variant)	Accession Number
pMV158 Related Plasmid Family			
	pMV158	-	S05981
	pFX2	-	X54310
Chlamydia Plasmids			
	pCcpA1	-	X62475
	pCHL1	-	S02220
Archaeal Plasmids			
	pHGN1	-	S06780
	pGRB1	-	S10152
Nanovirus-like DNAs			
	CLCuV1-1	*Cotton leaf curl virus* DNA1	AJ132344
	AYVV1-7	*Ageratum yellow vein virus* DNA1	AJ238493
Unclassified Replicons			
ColE3		-	S04456
Chp1		-	JU0348
Phasyl		-	S03290
pEE		-	M81382

Appendix B

Image Version 0.9

Contents

B.1 INTRODUCTION

Image09 is a simple Windows 95/98/NT image analysis program for quantifying chlorotic disease/nutrient deficiency symptoms on plant leaves. In 1997 when this program was written I was surprised to discover that there was no freely available software on the internet for quantifying discolored areas on digitised leaf images. While relatively powerful image processing and analysis programs (such as Paint Shop Pro and IMPROC) were available online, it was very difficult to use them for this purpose - the time taken to analyse even a modest number of leaves using them was completely prohibitive (no more than 15-20 leaves could be examined per hour). I therefore decided to produce a simple, stripped down version of these programs that at the push of a single button would automatically detect the parts of images that represented leaves, the parts of leaves that represented lesions, and calculate the percentage chlorotic areas on the leaves. I realised that a high degree of analysis automation with little or no requirement for human input would not only make the software ideal for research purposes (where human subjectivity can be a major problem) but would also make the software an attractive alternative to commercial general purpose image analysis programs.

B.2 OPENING IMAGE FILES

The only image file format currently supported by Image09 are grayscale windows ".bmp" (bitmap) files. To open a file click on the "Open file" button in the control button panel (Fig. B.1 – A) and select the file to be opened. Once loaded the program will automatically divide the individual pixel elements of the image into three categories: (1) background pixels, (2) healthy leaf area pixels and (3) chlorotic leaf area pixels. This process is called "pixel thresholding". A

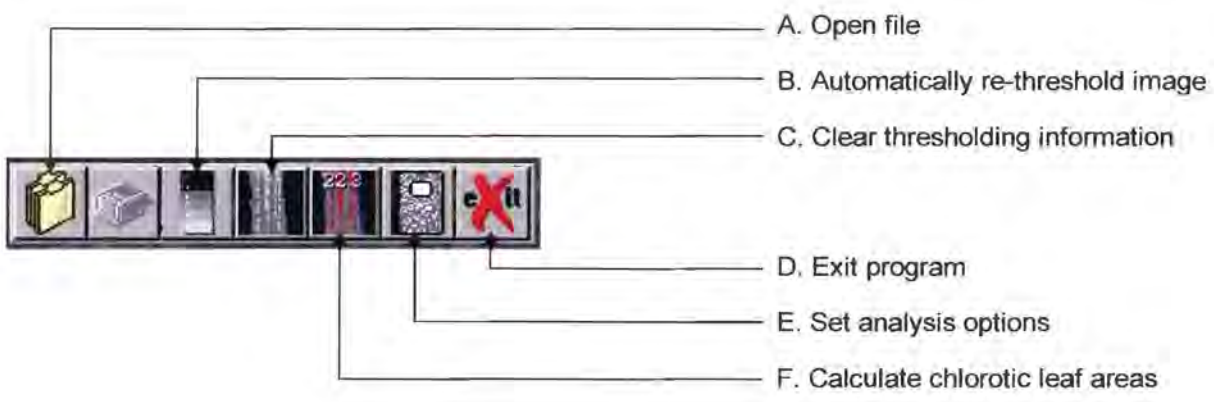


FIGURE B.1 The Image09 control button panel. Different buttons become active as the analysis proceeds.

grayscale image consists of rows and columns of pixels, each with a gray value between 0 and 255. Pixels with a value of 0 are black and pixels with a value of 255 are white. The image analysis technique used by Image09 is called pixel thresholding in that all pixels that have less than a certain gray value (say 50) may be considered black enough to be part of the background whereas pixels that have greater than a certain gray value (say 100) may be considered part of chlorotic lesions. Pixels in-between these threshold values will then be considered part of healthy leaf areas.

B.3 SETTING ANALYSIS OPTIONS

There are two possible ways in which the program can automatically set the two threshold values required to partition an image into background, healthy leaf areas, and chlorotic leaf areas. The program can also be forced to use threshold values specified by a user. The thresholding method used by the program can be set by pressing the "Analysis Options" button in the command button panel (Fig B.1 - E).

1. Manual thresholding: When an image is loaded the program always uses the pixel threshold values speciefied by the user. You can change the set threshold values that the program will automatically use once you select this option.
2. Enable automatic thresholding without a step-wedge: When an image is loaded the program evaluates the pixel gray level frequencies within the image and attempts to identify a large dark group of what it assumes to be background pixels, a smaller, lighter group of what it supposes to be healthy leaf area pixels and a much smaller group of lightly coloured pixels which it identifies as chlorotic leaf area pixels.
3. Enable automatic thresholding with a step-wedge: This is the program's default setting. When an image is loaded the program scans the image for the presence of a rectangular collection of pixels with a colour value of 255. These pixels are taken to represent the "pure white" segment of a graded grayscale internal shading control or "step-wedge" (Fig B.2 - B). The program then attempts to calculate the average grayscale values in the individual steps of the step wedge and, based on user defined settings, equates the gray value of particular steps with background/healthy leaf area (Fig B.2 - C) and healthy leaf area/diseased leaf area (Fig B.2 - D) threshold values. To determine which steps of the step wedge should be considered representative of healthy leaf areas and which should be representative of chlorotic leaf areas you will need to run some trials where you: (1) manually set the threshold values and get good separation of healthy/chlorotic leaf areas and background/healthy leaf areas and (2)

correlate these settings with the steps on the step wedge you are using that fall into these background, healthy and chlorotic categories.

B.4 ADJUSTING THE PROGRAM'S THRESHOLD SETTINGS

It will often be necessary to slightly adjust the program's automatic threshold settings to properly select the healthy/background cutoff and occasionally the healthy/diseased cutoff – this will be common when you are attempting to "calibrate" the program's automated analysis settings. The threshold adjust buttons (Fig B.2 – E and F) can be used to achieve this. Pressing the "up" arrows will increase the threshold settings and pushing the down arrows will decrease them. The lowest healthy/background setting is 0 and the lowest healthy/chlorotic setting is whatever the current healthy/background setting is. The highest healthy/chlorotic setting is 255. When you press these buttons different things will happen on the screen. When you adjust the healthy/background setting the blue borders surrounding objects in the image will begin to move about the screen. By decreasing this threshold value you will be including progressively darker pixels within the blue borders. By increasing this threshold value you will be excluding darker pixels from within the blue borders. The blue borders surround parts of the image that the

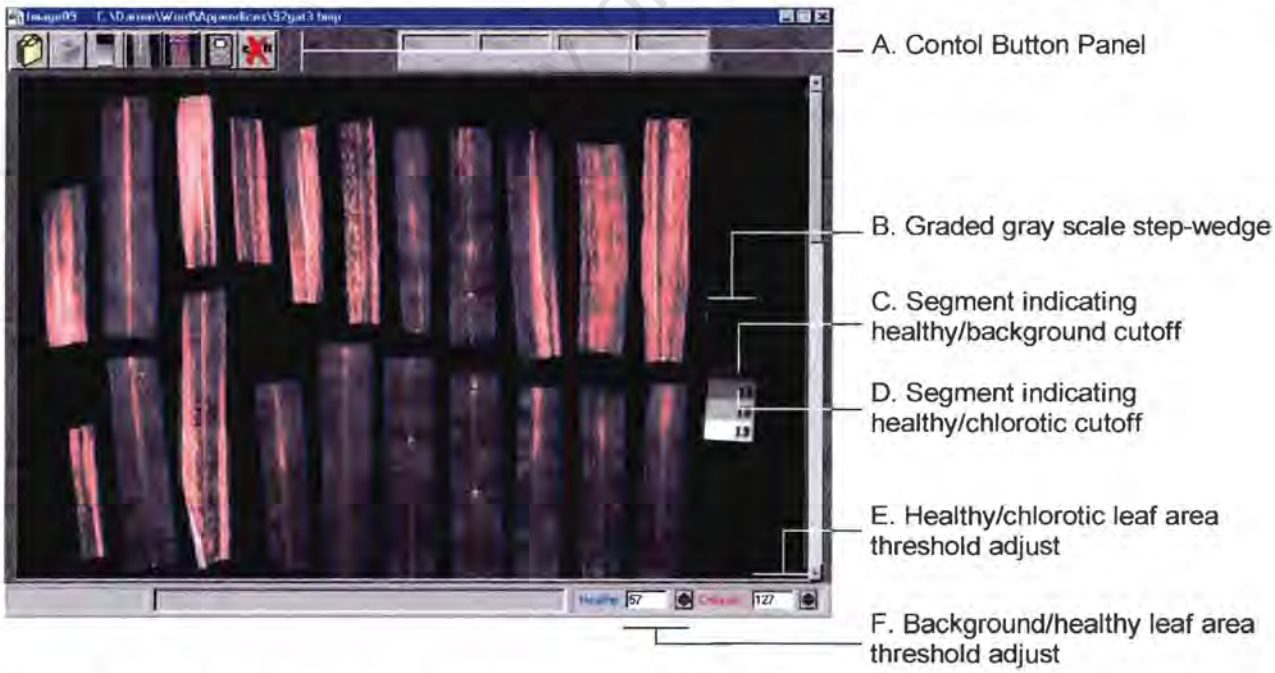


FIGURE B.2 The user interface once an image containing diseased leaves has been loaded. Options that are available to the user at this point are (1) to calculate chlorotic leaf areas (2) change analysis options, or (3) re-threshold the image (using the options selected in the analysis options). Leaf segments are surrounded by a blue border and all pixels with a gray value higher than 57 are considered to be within the leaf segments. All pixels with a gray value greater than 127 (indicated in pink) are considered to be within chlorotic lesions.

program will assume are leaves. The purpose of adjusting this value is to tell the program exactly where the leaves are in the image. Occasionally parts of the background will be included within these borders. This is not serious as the program will not interpret bright background spots as being leaves. It will however interpret the background between leaves as being part of a leaf area if, for example, the borders around two adjacent leaves merge. The pink parts of the image represent leaf areas that the program will interpret as being chlorotic lesions. If you alter the settings to an extent where you have lost confidence in the objectivity of the settings you have made, the "re-threshold" button (Fig B.1 – B) can be pressed and the program will reexamine the image and reset the threshold values.

B.5 QUANTIFYING CHLOROTIC LEAF AREAS

Press the "calculate chlorotic areas" button (Fig B.1 - F) to obtain information on the chlorotic areas of leaves within the image. Numbers representing percentage chlorotic leaf areas are placed on each leaf in the image. If during the previous analysis step the healthy/background setting had not been set properly, two adjacent leaves will share a single percentage chlorotic leaf area value. It is important to understand that this value is not simply the average value of these two leaves but is instead the average chlorotic leaf areas of these two leaves AND the background separating them in the image. To obtain a proper estimate of chlorotic areas on these leaves it will be necessary to reset the background/healthy threshold value so that the leaves are properly separated and then re-press the "calculate chlorotic leaf areas" button.

B.6 POSSIBLE PROBLEMS

B.6.1 Poor contrast between background, healthy and chlorotic leaf areas

The background used to capture images must appear much darker in the grayscale image than the health areas of the leaf. Similarly the healthy areas of leaves must be much darker in the grayscale image that the chlorotic lesions being examined. Unfortunately many diseases produce a range of symptom colourations such as chlorosis in certain parts of leaves to red/brown discolouration of veins. Image09 is not suitable for the analysis of these symptoms unless the different symptoms are examined separately. A common problem that is experienced by users of the program is that the symptoms they are examining appear to disappear when a grayscale image is captured. The reason for this is simple and so is the solution. In a grayscale there is no

differentiation between colours – there is only differentiation between shades – an identical shade of red and green will appear as the same shade of gray in an image. To analyse symptoms that disappear when an image is captured it will be necessary to use a colour filter on the lens of the image capture device. Match the filter you use to the colour of the disease symptom you are analysing – ie a red filter for reddish-brown symptoms, a yellow filter for chlorotic symptoms etc.

B.6.2 It is impossible to get single optimum threshold values for an entire image

If different parts of an image seem to require different optimum threshold values you may have to examine how you are lighting your specimens during image capture. All pixel thresholding image analysis techniques are extremely sensitive to errors caused uneven specimen lighting. Top lighting with a fluorescent ring lamp is probably the best way to illuminate specimens. Parallel fluorescent tubes at equal distances and angles on either side of the specimens is also a good option. Another cause of this problem could be the digital camera that is being used to capture images. Each individual pixel colour value captured in any particular part of an image should independently represent a corresponding fragment of a scene being captured. Unfortunately the colour values of each pixel are never completely independent of the colour values of the pixels that surround them. Depending on the camera used to capture an image, any individual pixel's colour will be influenced by that of pixels up to 50 or more pixel lengths away. The distance of this influence is directly related to the quality of the camera used to capture images. Better cameras minimise this influence. When analysing chlorotic symptoms on leaves this effect manifests itself when the background surrounding leaves with bright chlorotic symptoms is brighter than the background surrounding leaves with no symptoms. The result is that leaves with bright symptoms appear to require a lower healthy/background threshold values than leaves that are symptomless.

B.6.3 Software crashes/file incompatibility

While Image09 in my hands is relatively stable (I've corrected all the bugs that I've encountered during its use), I can not vouch for its stability in the hands of others. Also, while I am able to load all 8-bit ".bmp" files described in the Microsoft ".bmp" file format descriptions, I can not be certain that the formatting of files produced by other programs (or even versions of the software

that I use but am unfamiliar with) will work with Image09. Should you encounter any technical problems with the software I would really appreciate you telling me at darren@molbiol.uct.ac.za. I can only fix the problems that I am aware of and I promise to sort them out as quickly as my programming skills allow.

Appendix C

RDP Version 1.07

Contents

C.1 INTRODUCTION

RDP is a Windows 95/98/NT program for aiding in the detecting of recombination amongst a set of aligned DNA sequences. While a number of other programs have been written to carry out the same task, my motivation for writing RDP has been to produce a comparable analysis tool that is both accessible to users who are uncomfortable with the use of UNIX/DOS command lines and permits a more interactive role in the analysis of recombination.

C.2 OPENING ALIGNMENT FILES

A number of different alignment file formats are recognized by RDP including GDE, FASTA, CLUSTAL, GCG, NEXUS and DNAMAN. To open a file click on the "Open" button (Fig. C.1 - A) in the top panel and select the file to be opened. The directory from which the file is selected is "remembered" by the program when it is shut down. Once loaded the aligned sequences, together with their names are displayed in the "sequence display panel" (Fig. C.1 - C). Also displayed is the percentage identity of different regions of the aligned sequences in a "percentage identity display panel" (Fig. C.1 - B).

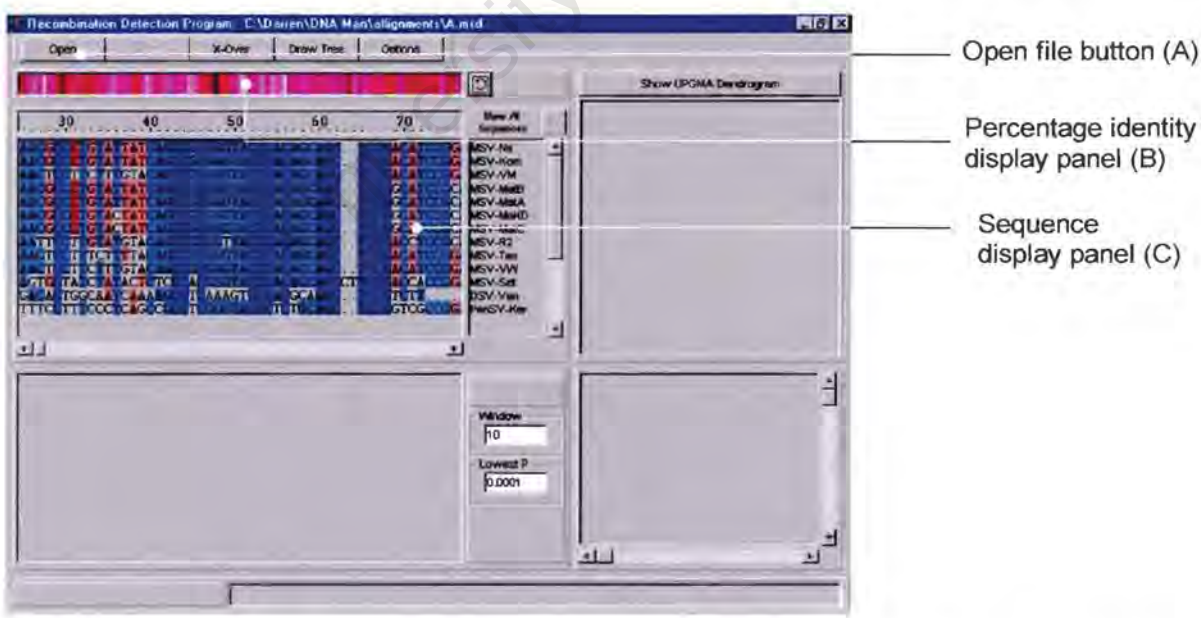


FIGURE C.1 The user interface once a group of aligned sequences has been loaded. Options that are available to the user at this point are to begin the analysis (pressing the "X-Over" button, draw an UPGMA dendrogram for the sequences (pressing the "Draw Tree" button), or setting analysis options (pressing the "Options" button).

Besides alignment files RDP project files (with a ".rdp" extension) may also be loaded. Apart from aligned sequences these files contain information on putative recombinations detected in former analysis sessions.

C.3 SETTING ANALYSIS OPTIONS

A number of analysis options may be set by pressing the "Options" button in the top panel.

1. For the "percentage identity display panel" or "homology box" the size of the window used for calculating mean percentage identities for every nucleotide site in the sequences may be changed. The display of percentage identity may be in the form of colored shading or a histogram. The calculation of percentage identities can be set to weight sequences according to their prevalence in the alignment (antibiasing).
2. The software can perceive sequences as being either linear or
3. The window size (the number of nucleotides looked at in a sliding window moved along the length of sequences comprised exclusively of phylogenetically significant sites) and maximum acceptable probability during detection of recombinant regions may be set. While Larger window sizes will lower signal:noise ratios but decrease the sensitivity of the analysis, smaller window sizes will increase the sensitivity but also increase the likelihood of false positives. The maximum acceptable probability is the highest probability that sequences could share high identities in potentially recombinant regions due to chance alone. The recommended setting for this parameter is a high figure (0.05 – 0.001) for large window sizes (50 –100 nucleotides) and a lower figure (0.001-0.000001) for small window sizes (5-20 nucleotides). Since version 1.03 of RDP there is an option for multiple comparison correction of P-values to be used. The default is that multiple comparison corrections are used as this makes the calculated P-values "experiment-wide" rather than "currently selected sequence triplicate wide" estimates of probability. A fourth option that can be set is to display/not display identity plots during analysis. It is recommended that for large numbers of long sequences this be set to the "don't show plots" setting as the speed of the analysis can be severely effected if the identity plots are displayed under such circumstances. The "spacer" or "reference" sequences used for determining phylogenetic significance of sites during analysis can be set to one of five different settings. While it is recommended that analyses be performed using all 5 different settings, we have found that the "using internal

spacers only" setting provides the best unambiguous estimates of recombination break-points.

C.4 BEGINNING AN ANALYSIS

When large numbers of sequences are to be analysed certain sequences in the data set can be either "masked" or completely removed from the analysis by clicking (with the left mouse button) on the name of the sequence in the sequence display panel (Fig C.1 - C). Masking does not stop the sequence being used in either tree construction or as a reference sequence in determining informative sites. Disabling a sequence is similar to masking it except that the sequence is no longer used by RDP in determining phlogenetically informative sites. Once the appropriate settings have been selected, clicking on the "X-Over" button in the top command button panel starts the analysis. A progress bar and either the number of detected potential recombination events or percentage identity plots indicate the progress of the analysis.

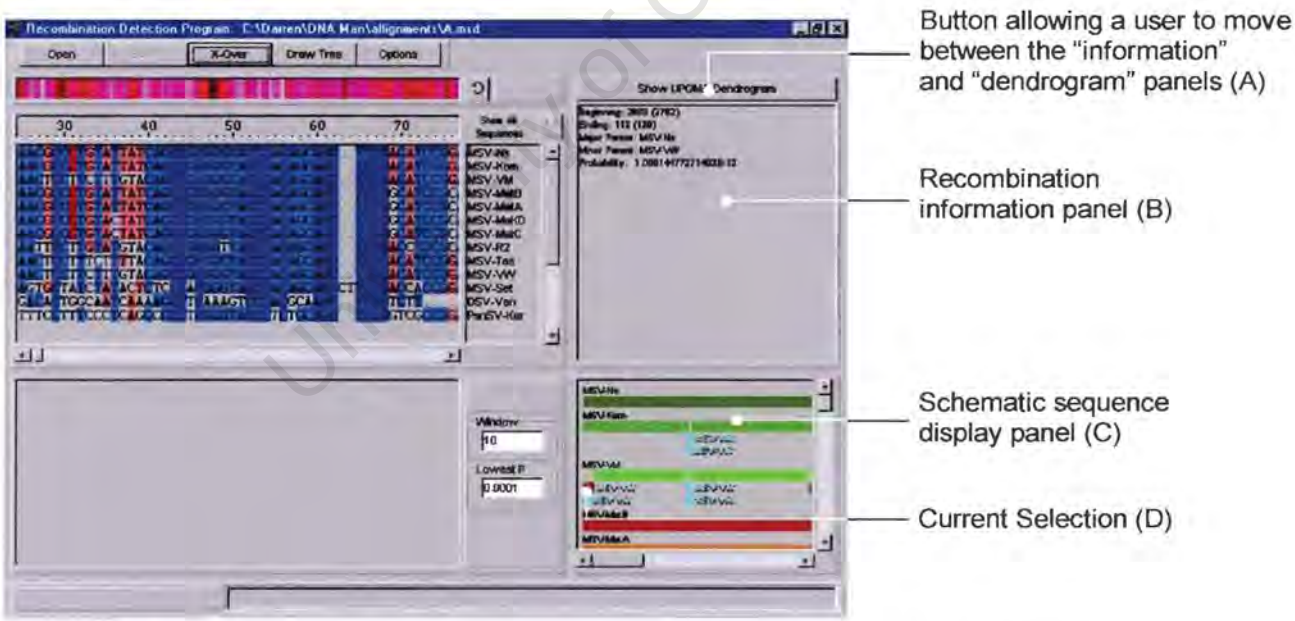


FIGURE C.2 Following analysis of the aligned sequences by RDP a schematic representation of the sequences is displayed in a "schematic sequence display panel"(A). Included is a diagrammatic representation of the sequences and the regions that have a potentially recombinant origin. Information regarding the currently selected recombinant region (D; which was selected by moving the mouse pointer over it) is displayed in the "recombination information display panel" (B).

C.5 EXAMINING RESULTS

Once the analysis has concluded, a schematic representation of the aligned sequences showing potentially recombinant regions is displayed in the "schematic sequence display panel" (Fig C.2 - C). The regions that are potentially recombinant are sensitive to the mouse pointer and when it is moved over these regions information regarding the region, is displayed in the "Recombination information panel" (Fig C.2 - B). The information displayed includes the possible breakpoints (in the sequence and in the alignment), identities of sequences that are closely related to the parental sequences and the probability that the recombinant sequence could have been more closely related to the "minor parent" than the "major parent" in the specified region by chance alone. Also displayed in versions after 1.07 are warnings on whether there is either only a single suitable parent-like sequence in the set of aligned sequences and/or a fair likelihood that the software could have misidentified the daughter sequences. The latter warning is only meant as a prompt for the user to properly examine the data presented before making a judgement on whether the program's decision regarding the daughter sequence is correct. Unfortunately the program is still capable of misidentifying daughter sequences without giving a warning when a substantial proportion of the reference sequences used are recombinants. It is therefor advisable to carefully examine all putative recombinants by both using the information

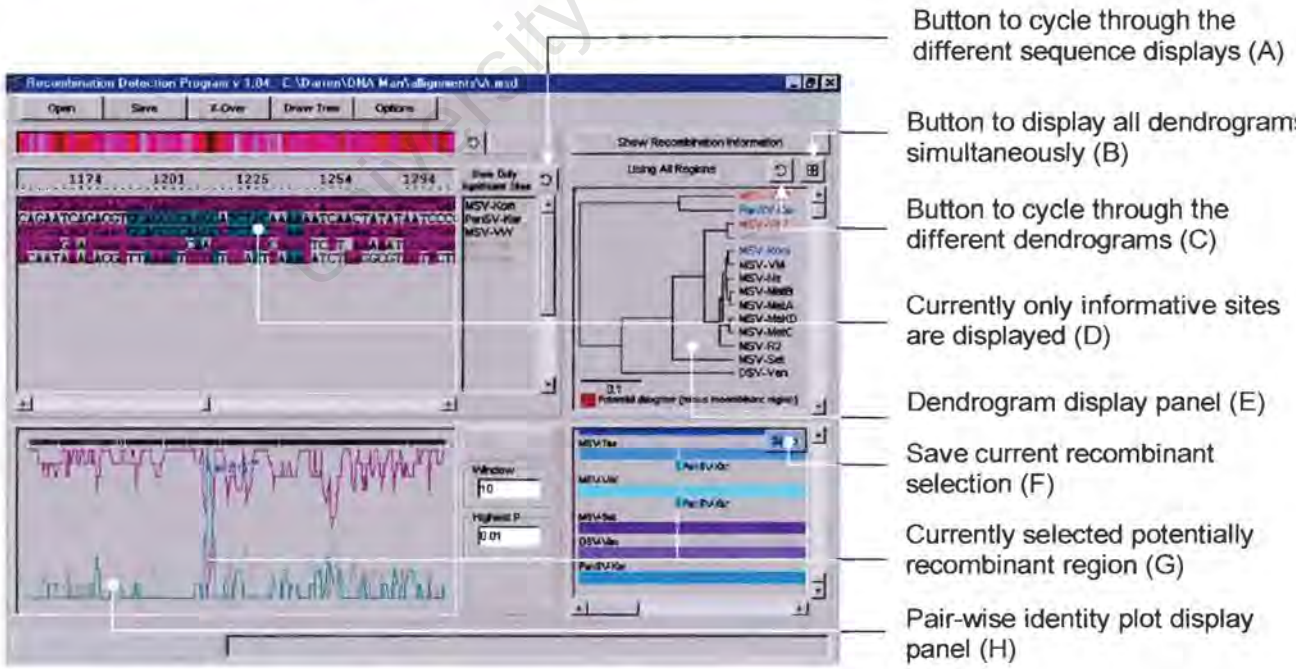


FIGURE C.3 Once a selected region is "clicked on" (G), information specific to the selection is displayed in the form of dendrograms (E), pair-wise sequence plots (H) and alignments of phylogenetically informative sites (D).

displayed by the program and by performing supplementary bootstrap analyses using the recombinant regions.

Clicking on the sensitive region with either the left or right mouse button displays the pair wise identity plots used for identifying this region as recombinant (Fig C.3 - G and H). The plot is sensitive to the mouse pointer and double clicking anywhere in this panel will take you to the corresponding region in the sequence display panel. At the top of the "pair-wise identity plot display" (Fig C.3 - H) is a graphical representation of the distribution of phylogenetically informative sites in the sequences involved in producing the plots. The sequence display window can be cycled (Fig C.3 - A) to show only the sequences involved in identifying the recombinant region either with or without all of their uninformative sites removed (Fig C.3 - D). By clicking on the "Show UPGMA Dendrogram" button (Fig C.2 - A) a number of different trees expressing the relationships of the different regions of the identified recombinant to the other sequences in the alignment are displayed in the dendrogram display panel. These trees can be cycled by pressing the cycle button (Fig C.3 - C). The trees can either be drawn by using all regions of recombinant sequences examined separately, only the identified recombinant region (the region related to the "minor" parent in the selected sequence), only the identified "non-recombinant" region (the region related to the "major" parent in the selected sequence) or all regions ignoring recombination. To display all dendrograms side-by side in a new window the button to display all dendrograms simultaneously can be pressed (Fig C.3 - B). From version 1.07 onwards when all dendrograms are displayed simultaneously it is possible to mark sequences in one dendrogram and have the corresponding sequences in all the other panels marked at the same time. I have found this feature very useful while tracking the movement of putatively recombinant sequences to different positions in the dendrograms constructed using putatively recombinant and non-recombinant regions of the sequences. The colour used to mark the sequences can be changed by right clicking in the dengrogram display panels and the sequences can be marked/unmarked by left clicking on them in the display panels.

Depending on which mouse button was pressed in the schematic sequence display panel, the program will in some cases use slightly different regions of the multiple alignment to draw trees. If the left mouse button is used the recombinant region is interpreted to be all overlapping or directly adjacent recombinant regions that have been detected together. If the right mouse button was used to select the recombinant region only the sequences within the selected recombinant region is interpreted to be recombinant.

Once an individual recombinant region has been selected there is an option to save all the sequences in the alignment within this region for supplementary analysis (Fig C.3 - F). Sequences are saved both individually and in DNAMAN.msd format. All sequences are saved in the same directory as the RDP.exe file.

C.6 SAVING RESULTS

Analysis results can be saved in one of two different formats. Saving results to an RDP project file (a file with a ".rdp" extension) enables an analysis to be reloaded at a later date for further scrutiny. Saving results to a text file allows raw data to be scrutinised using a spreadsheet program. The data written to the text file is unfiltered and will contain some redundancy, as all significant evidence of recombination (even that obtained using sub-optimal parents) will be saved. For the different fields of the text file to be read correctly by a spreadsheet program (such as Excel) you will need to specify when loading the file that columns are delimited by TABS. (note that for versions before version 1.07 the columns were delimited by spaces).

C.7 SUPPLEMENTARY ANALYSIS

It is recommended that once a recombinant region has been identified and appears to represent evidence of a genuine recombination event, this region in all the sequences in the alignment be saved to disk for further phylogenetic analysis by pressing the "Save" button in the schematic sequence display panel (Fig C.3 - F). The sequences are saved to a sub-directory within the RDP start-up directory that has the name of the selected recombinant and the nt positions of the potential breakpoints. It is recommended that these sequences be realigned and then subjected to phylogenetic analysis using the phylogenetic analysis software of your choice. The definitive evidence for recombination will be movement of the recombinant sequence from its usual branch in a tree to another branch with significant bootstrap support.

C.8 POSSIBLE PROBLEMS

C.8.1 Poor Alignments

Badly aligned sequences (ie a group of 10 sequences with more than 1000bp's that has been aligned in under 10 minutes on a P166) will probably result in incorrect identification of recombinants. Carrying out the recommended supplementary analysis should identify such errors.

C.8.2 Recombinant Spacer/Reference Sequences

If sequences that are used as references for phylogenetic significance during the analysis were themselves recombinant, the software may incorrectly identify parental and daughter sequences. The software will, however, still identify the correct region in which recombination has occurred. This error can be detected if the supposedly recombinant sequence is in the same tree position regardless of which part of the sequence has been used to draw the tree. Looking for changes in the tree position of one or a group of reference sequences will identify the recombinant reference sequence. In certain instances this "indirect" evidence for recombination in the reference sequence may be the only evidence the software is able to find that the reference sequence is recombinant (ie it will not be able to give any probability measures, descriptions of parents and precise break-points). Carrying out the recommended supplementary analysis will be the only means of certifying whether sequences identified in this way are recombinant. Since version 1.07 I have included a checking routine to detect incorrectly identified daughter sequences. The software now gives a warning if there is a fair likelihood that the daughter indicated is not the correct daughter. The routine is, however, not infallible and incorrect identification of daughter sequences is still possible. If you notice that results obtained with the same analysis setting using versions of the program before 1.07 differ slightly from those obtained with versions after 1.07, it is highly likely that the software has now correctly identified daughter sequences that it had formerly misidentified as one of the parental sequences.

C.8.3 Adjacent Recombinant Regions of Different Origin

When drawing trees and saving recombinant regions, the software now has the capacity to separately deal with directly adjacent regions of sequence that have originated from different

"minor" parents. The software determines which pieces of sequence to use when drawing trees depending on the mouse button pressed when selecting recombinant regions. If the left mouse button was used the software examines the part of the sequence in the specified region that did not originate in the "major" parent as though it has a single phylogenetic origin. In practice this is not always the case in that the recombinant section may have itself originated either from a recombinant sequence with a breakpoint in that region or when two overlapping recombination events occurred in the same region. If the right mouse button is pressed only the sequence in the specified recombinant region are identified as recombinant.

C.8.4 Degeneracies

Degeneracies that are expressed as letters of the alphabet will not be identified as degeneracies by the software. Use of "N" in sequences to describe sites that are known to exist but for which sequence has not yet been determined can be a potential problem. If a string of N's exist in corresponding regions of divergent sequences, the corresponding region will be identified by the software as having a recombinant origin. It is recommended that "N's" either be deleted before aligning the sequences or replaced in already aligned sequences (using a text editor) with "*" symbols.

C.8.5 Software Crashes/File Incompatibility

While RDP in my hands is relatively stable (I've corrected all the bugs that I've encountered during its use), I can not vouch for its stability in the hands of others. Also, while I am able to load files in all the alignment formats that I frequently use, I can not be certain that the formatting of files produced by other programs (or even versions of the software that I use but am unfamiliar with) will work with RDP. Should you encounter any technical problems with the software I would really appreciate you telling me at darren@molbiol.uct.ac.za. I can only fix the problems that I am aware of and I promise to sort them out as quickly as my programming skills allow.

APPENDIX D

Nucleotide and Amino Acid Sequence Alignments

Contents

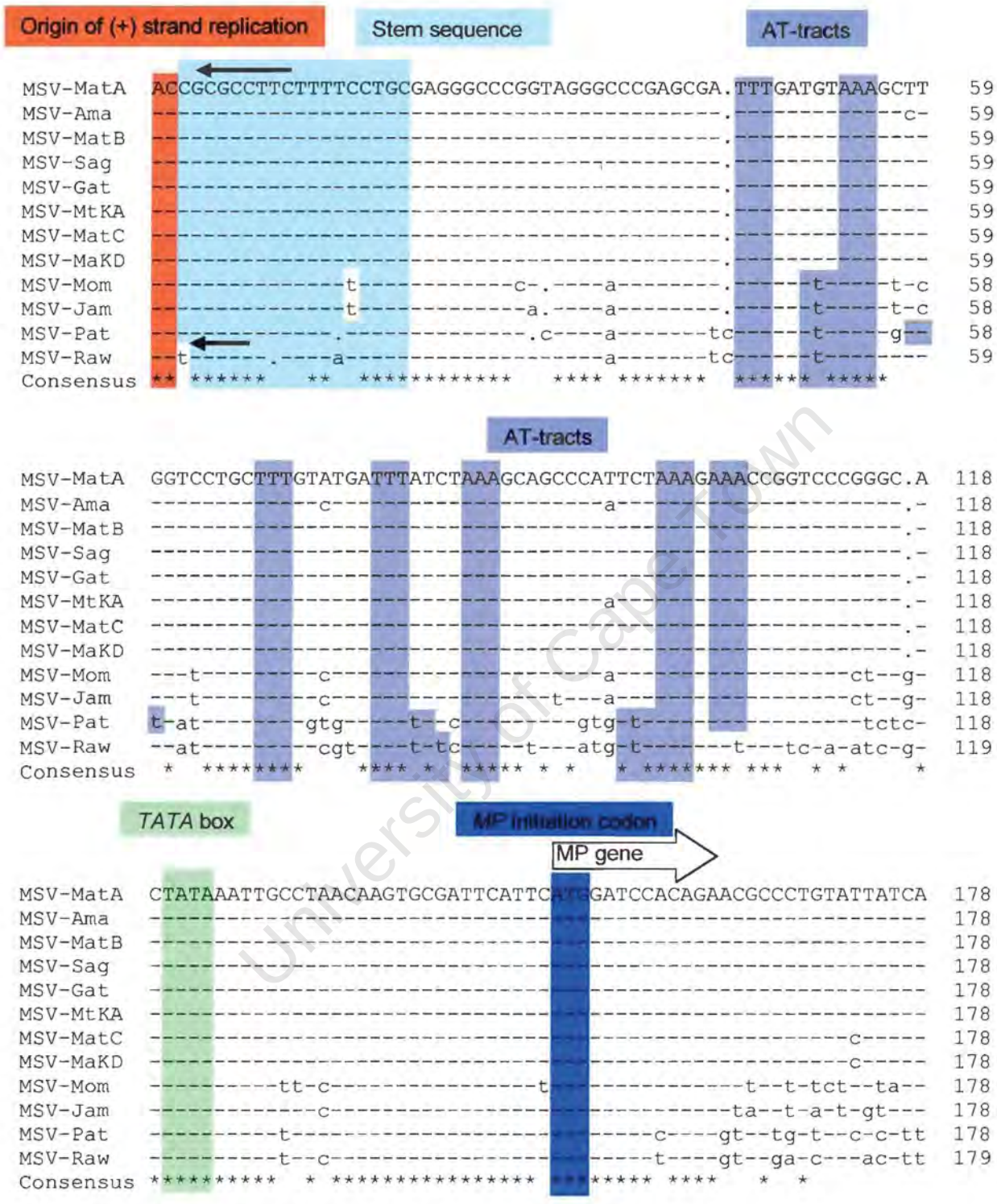


FIGURE D.1 Alignment of full length genomic sequences (Cont)

Cryptic donor site

```
MSV-MatA   GCCGCGGGTACCCACAGCAGCTCCGGCATCCGCAGGAGTGCCGTGGAGTCGCG.TAGGCG  237
MSV-Ama    ----------------------------------------------------.------  237
MSV-MatB   ----------------------------------------------------.------  237
MSV-Sag    ----------------------------------------------------.------  237
MSV-Gat    ----------------------------------------------------.------  237
MSV-MtKA   ------------------------a------g--------------------.------  237
MSV-MatC   ----------------------------------------------------.------  237
MSV-MaKD   ----------------------------------------------------.------  237
MSV-Mom    ------------------------a------g-------t-----------.-c----  237
MSV-Jam    ------------------------ag-c---g---------------.---c-c----  237
MSV-Pat    ------------------------c-caa-------------------a--.-c----  237
MSV-Raw    ---t--------------------a-ca---g---t---t---------a--.-c----  238
Consensus  *** ********************     ** *** *** ********* ** * ****
```

Donor site | T-tracts | Intron | Branch point

```
MSV-MatA   AGGTAGCTATTTTGAGCTTTGTTGCACTGATTTGCTTTTACCTGCTTTACCTTTGGGTGC  297
MSV-Ama    ---------------------------------------------------------  297
MSV-MatB   ---------------------------------------------------------  297
MSV-Sag    ---------------------------------------------------------  297
MSV-Gat    ---------------------------------------------------------  297
MSV-MtKA   ----------------------------t----------------------------  297
MSV-MatC   ----------------------------t----------t----------------  297
MSV-MaKD   ---------------------------------------------------------  297
MSV-Mom    ---------------------------g-t---------------------------  297
MSV-Jam    --------a------------------g-t---------------------------  297
MSV-Pat    ---------g-----------------tt---------a-----t-------------  297
MSV-Raw    ----------ac---------------tt---------a----t-------------t  298
Consensus  *******  *  ************   ******** ****  ****************
```

Branch point | Intron | Acceptor site | Cryptic acceptor site

```
MSV-MatA   TGAGAGACCTTATCTTAGTTCTGAAGGCTCGACAAGGCAGATCCACGGAGGAGCTGATAT  357
MSV-Ama    ---------------------------------------------------------  357
MSV-MatB   ---------------------------------------------------------  357
MSV-Sag    ---------------------------------------------------------  357
MSV-Gat    ---------------------------------------------------------  357
MSV-MtKA   ---------------------------------------------------------  357
MSV-MatC   ---------------------------------------------------------  357
MSV-MaKD   ---------------------------------------------------------  357
MSV-Mom    --------t------------------------------------------------  357
MSV-Jam    --------t----------t----------------------------------------  357
MSV-Pat    --------t----------ct-------aa---gc--g--g-------------------  357
MSV-Raw    --------t----------------------aa---gc--g--g-------------------  358
Consensus  ******* ***********  *******  ***  ** ** ******************
```


FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   TTGGAGGACAAGCTGTGGATAGGAGCAACCCTATCCCTAATCTACCAGCACCACCAAGTC  417
MSV-Ama    ------------------------------------------------------------  417
MSV-MatB   ------------------------------------------------------------  417
MSV-Sag    ------------------------------------------------------------  417
MSV-Gat    ------------------------------------------------------------  417
MSV-MtKA   ---------------------------------------a--------------------  417
MSV-MatC   ------------------------------------------------------------  417
MSV-MaKD   ------------------------------------------------------------  417
MSV-Mom    -----at-------------------------------------ac--ag------------  417
MSV-Jam    -----at-----------------t--t------------ac--a---------------  417
MSV-Pat    -----tctg---------------g-gt------------actttg-a----a--gc--  417
MSV-Raw    -----tctg---------------gc-t------------actttg-t----a--gc--  418
Consensus  *****    ****************    ************      * **** **  **
```



```
MSV-MatA   AGGGCAATCCCGGGCCATTTGTTCCAGGCACGGGATAAGCAATCAGCCATTTCCACGTC  476
MSV-Ama    ------------------------------------t------------------------  476
MSV-MatB   ------------------------------------t------------------------  476
MSV-Sag    ------------------------------------t------------------------  476
MSV-Gat    t-----------------------------------t------------------------  476
MSV-MtKA   ------------------------------------------------------------  476
MSV-MatC   ------------------------------------------------------------  476
MSV-MaKD   ------------------------------------t------------------------  476
MSV-Mom    ----g----------g-----------------------------------------------  476
MSV-Jam    ----g--------a--g-----a--c--g-------------------------t---g-  476
MSV-Pat    ca-tgc-------a----------t--tt-------------g-g-------------g-----  476
MSV-Raw    c--tgc-------a-----c-----t--tca-----------g-a-------------g-----  477
Consensus    *   ******* ** ** ** ** **   ********** * ********* *** *
```

```
MSV-MatA   CAAGAGGAAGCGGGGAGATGATTCGAATTGGAGTAAGCGGGTGCCTAAGAAGAAGCCCTC  536
MSV-Ama    ---------------------------------------a----------------------  536
MSV-MatB   -------------------------g---------------ta-------------------  536
MSV-Sag    ----------------------------------------ta-------------------  536
MSV-Gat    ----------------------------------------ta-------------------  536
MSV-MtKA   ----------------------g--------------------t-------------t--  536
MSV-MatC   ------------t---------g----c---------------a-------------t--  536
MSV-MaKD   ---------------g-----cg--------------------a-------------t--  536
MSV-Mom    ----------------------g----c----a-------ac-a-------------a--  536
MSV-Jam    ------------a---------g----c----a-------tc-a-------------tg-  536
MSV-Pat    ------------t-cc--c--ggt-c-g----a----a--tcca-c--------ag-t--  536
MSV-Raw    ------------c-cc-----ag--c-a----a-------tcca-c--------aggt--  537
Consensus  ************ *  ** **   *   **** **** **    * ********     *
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   TTCAGCTGGGCTGAAGAGGGCTGGAAGCAAGGCCGATAGGCCATCCCTCCAAATCCAGAC  596
MSV-Ama    g-------------------------------------------at----------------  596
MSV-MatB   ---------t--------------c-----------------------------------  596
MSV-Sag    ---------t--------------c-----------------------------------  596
MSV-Gat    ---------t--------------c-----------------------------------  596
MSV-MtKA   ---------------------------------------------------------  596
MSV-MatC   ---------------------------------------------------------  596
MSV-MaKD   ------------------------------------a----------------------  596
MSV-Mom    ---------t-------a------------------a-----------t--g--------  596
MSV-Jam    ---------t-------a-------gcg-----------------------g--------  596
MSV-Pat    -g--c-gcc-g------a-a----gg-t-----t--c-----t--------g--a-----  596
MSV-Raw    -g-gc-gca-gc-----aac----gg-t----gt--g-----t--------g--a-----  597
Consensus   * *      *****   ****    ****  ** ****    **** ** ** *****

MSV-MatA   ACTCCAGCATGCTGGGACCACCATGATAACTGTCCCATCCGGAGGAGTATGTGACCTC.A  655
MSV-Ama    ---------------------------------------------------------.-  655
MSV-MatB   ------------------------------g--------------------------.-  655
MSV-Sag    ------------------------------g---------t-----------------.-  655
MSV-Gat    ------------------------------g--------------------------.-  655
MSV-MtKA   ------------------------------g-----c--------------------.-  655
MSV-MatC   ------------------------------g--------------------------.-  655
MSV-MaKD   ------------------------------g-----c--t-----g-----------.-  655
MSV-Mom    ----------------t-------------a-----c--------------------.-  655
MSV-Jam    ----------------t-------------a-----c--------------------.-  655
MSV-Pat    t--a-tc--ct-g---ga-----------c--a--g--a-.t--g-gc---------c-  655
MSV-Raw    tt-a-tc--ct-a--tga---g-----c--c--a--g--a--t--g--c---------.-  656
Consensus    * *  **  * **   *** ***** ** ** ** ** *  ** *  ********* *

MSV-MatA   TCAACACCTATGCCCGAGGATCTGACGAGGGCAACCGCCACACCAGCGAGACTCTCACGT  715
MSV-Ama    ---------------------------------------------------------  715
MSV-MatB   ---------------------------------------------------------  715
MSV-Sag    ---------------------------------------------------------  715
MSV-Gat    ---------------------------------------------------------  715
MSV-MtKA   ---------------------------------------------------g----  715
MSV-MatC   ---------------------------------t-----------------------  715
MSV-MaKD   ----------c-----------g---------------------------------  715
MSV-Mom    ----------------------c-----------------------------g----  715
MSV-Jam    ----------------------c-----------------------------g----  715
MSV-Pat    ---gt--a-------------c--t----------t----------a--c--a----  715
MSV-Raw    ----t--a-----------g-----t----------g----------a--c--a----  716
Consensus  ***  ** ** ******** ** ** ******** ** ********** ** ** ****
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   ACAAGATCGCCATCGACTACCACTTCGTTGCCGACGCGGCAGCCTGCCGCTACTCCAACA  775
MSV-Ama    ------------------------------------------------------------  775
MSV-MatB   ------------------------------------------------------------  775
MSV-Sag    ----------g-------------------------------------------------  775
MSV-Gat    ----------g-------------------------------------------------  775
MSV-MtKA   -------t---g-------------------------------------------------  775
MSV-MatC   ------------------------------------------------------------  775
MSV-MaKD   ------------------------------------------------------------  775
MSV-Mom    -------t---g-------------------a---t--ca-------aag--t--t----  775
MSV-Jam    -------t--gg----------------c--a--gcg-ca-------aag----------  775
MSV-Pat    ----ag-t-ggg----------------c--------a-g-------a-a---------c  775
MSV-Raw    ----ag-t-ggg----------------c--t-----c---t-----aag--------tc  776
Consensus  ****  * *   **************** ** **      * *****   ** ** **

MSV-MatA   CCGGAACCGGTGTAATGTGGCTGGTGTATGACACCACTCCCGGCGGACAAGCTCCGACCC  835
MSV-Ama    ---------tg------------------------------------------------  835
MSV-MatB   ------------------------------------------------------------  835
MSV-Sag    ------------------------------------------------------------  835
MSV-Gat    ------------------------------------------------------------  835
MSV-MtKA   ------------------------------------------------------------  835
MSV-MatC   ------------------------------------------------------------  835
MSV-MaKD   ------------------------------------------------------------  835
MSV-Mom    ----g--------g-------------c----------------------g-------  835
MSV-Jam    ------------g-------------------------------------g-------  835
MSV-Pat    g------a-----g-------------c-----t--c---------a-ct-a---t--a  835
MSV-Raw    g------a-----g-------------c-----t------------a-c--a--c---a  836
Consensus   *** ** *  ** ************** ***** ** ********* *  * **  **

MSV-MatA   CGCAAACTATATTTGCCTACCCTGACACGCTAAAAGCGTGGCCGGCCACATGGAAGGTGA  895
MSV-Ama    ------------------------------------------------------------  895
MSV-MatB   ------------------------------------------------------------  895
MSV-Sag    ------------------------------------------------------------  895
MSV-Gat    ------------------------------------------------------------  895
MSV-MtKA   ------------------------------------------------------a-----  895
MSV-MatC   ------------------------------------------------------------  895
MSV-MaKD   ------------------------------------------------------------  895
MSV-Mom    -a-----g--------a--t--g--------g-----t--------a--------a----  895
MSV-Jam    -------c--------------g--------g-----t--------a--------a----  895
MSV-Pat    --a-ggac--c--c-----t--g--tg-a---gt---t------a-g--t-----a--t-  895
MSV-Raw    -c--gga---t--c--a-----ctc-g-a-----g--c------a-t--t-----a--t-  896
Consensus  *  *    ** ** ** ** **       **    ** ****** * ** ***** ** *
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   GCCGGGAGCTGTGTCATCGCTTCGTGGTGAAACGGCGATGGTTGTTCAACATGGAGACCG  955
MSV-Ama    ------------------------------------------------------------  955
MSV-MatB   ------------------------------------------------------------  955
MSV-Sag    ------------------------------------------------------------  955
MSV-Gat    ------------------------------------------------------------  955
MSV-MtKA   ------------------------------------------------------------  955
MSV-MatC   ------------------------------------------------------------  955
MSV-MaKD   ------------------------------------------------------------  955
MSV-Mom    ------------------------------------------------------------  955
MSV-Jam    ------------------------------------------------------------  955
MSV-Pat    -t------t-------c-------------g---------c-c----ca----------  955
MSV-Raw    -t------t----------------------g---------c-c----ca----------  956
Consensus  * ****** ******* ************** ********* * ****  **********

MSV-MatA   ACGGGCGGATTGGTTCGGACATTCCTCCCTCGAATGCAAGTTGGAAGCCTTGCAAGCGCA 1015
MSV-Ama    ------------------------------------------------------------ 1015
MSV-MatB   ------------------------------------------------------------ 1015
MSV-Sag    --------------------t----------------------------------------- 1015
MSV-Gat    --------------------t----------------------------------------- 1015
MSV-MtKA   ----t--------------t--------a------------------------------- 1015
MSV-MatC   ----t--------------t--c-----------------------c---------- 1015
MSV-MaKD   ----------------------------t------------------------------- 1015
MSV-Mom    ----t--a--------------c------t--------------------g---------- 1015
MSV-Jam    ----t--a--------------c------t--------------------g---------- 1015
MSV-Pat    ----t--a-----------t--c-----ga-c---cag------cc---a--t-----g- 1015
MSV-Raw    ----t-----a--c-----t-ca--a--gagc--ccag------cc---a--t-----g- 1016
Consensus  **** ** ** ** ***** *  ** **    **    ******  *** ** ***** *

MSV-MatA   ACATCTACTTCCACAAGTTCACGAGCGGGTTGGGAGTGAGAACGCAGTGGAAGAATGTAA 1075
MSV-Ama    ------------------------------------------------------------ 1075
MSV-MatB   ------------------------------------------------------------ 1075
MSV-Sag    ------------------------------------------------------------ 1075
MSV-Gat    ------------------------------------------------------------ 1075
MSV-MtKA   ------------------------------------------------------------ 1075
MSV-MatC   ------------------------------------------------------------ 1075
MSV-MaKD   ------------------------------------------------------------ 1075
MSV-Mom    -------------------------t---------------------------------- 1075
MSV-Jam    -------------------------t---------------------------------- 1075
MSV-Pat    ----ag----------------t----------------g------------------- 1075
MSV-Raw    --g-tg----------------t----------------g-------------c--c- 1076
Consensus  ** *  *************** ** ************* ************** ** *
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   CGGACGGAGGAGTTGGTGCCATCCAGAGAGGAGCGTTGTACATGGTCATTGCCCCCGGCA 1135
MSV-Ama    ------------------------------------------------------------ 1135
MSV-MatB   ------------------------------------------------------------ 1135
MSV-Sag    ----------------------------------t------------------------- 1135
MSV-Gat    ----------------------------------t------------------------- 1135
MSV-MtKA   ---------------------------------cc------------------------- 1135
MSV-MatC   ----------------------------------t-------------------a----- 1135
MSV-MaKD   ------------------------------------------------------------ 1135
MSV-Mom    --------------------g--a----------tc-------------------t---- 1135
MSV-Jam    --------------------g--a----------c--------------------t---- 1135
MSV-Pat    -a-----t--g--c-----g--a--------g--tc----t--------------t---- 1135
MSV-Raw    -a-----t--------a--g--a-----g--t--tc----tc-a---------------- 1136
Consensus  * ***** ** ** ** ** ** ***** ** **  ****  * ********** **** 
```



```
MSV-MatA   ATGGTCTTACATTTACTGCCCATGGGCAGACCCGTCTGTACTTTAAGAGTGTTGGCAACC 1195
MSV-Ama    ------------------------------------------------------------ 1195
MSV-MatB   ------------------------------------------------------------ 1195
MSV-Sag    ------------------------------------------------------------ 1195
MSV-Gat    ------------------------------------------------------------ 1195
MSV-MtKA   ----------t------------------------------------------------- 1195
MSV-MatC   ----c------------------------------------------------------- 1195
MSV-MaKD   ------------------------------------------------------------ 1195
MSV-Mom    ----c-------------------------------------------c------------ 1194
MSV-Jam    ----c------------------------------------------------------- 1196
MSV-Pat    -----g-a--g-------------------a----------c--------g------- 1195
MSV-Raw    -----a----g-------------------a------------------------------ 1196
Consensus  ****  * ** ******************** ********** ******** ******* 
```

V-sense transcript polyadenylation signal

CP termination codon

C-sense transcript polyadenylation signal

```
MSV-MatA   AGTAATGAATAAAAACTCCCGTTTTATTATATCTGATGAATGCTGAAAGCTTACATTAAT 1255
MSV-Ama    ---------------------------------------------------------- 1255
MSV-MatB   ---------------------------------------------------------- 1255
MSV-Sag    ---------------------------------------------------------- 1255
MSV-Gat    ---------------------------------------------------------- 1255
MSV-MtKA   ----------.-g--gt----------g------------------------------ 1254
MSV-MatC   ---------------------------t-c------------------------- 1255
MSV-MaKD   ------------------------ctgc------------------------- 1255
MSV-Mom    ----------.---------------------------------------------- 1254
MSV-Jam    ----------.-------------------c---------------- 1254
MSV-Pat    ----------.-c--.----------c-----------c---------------- 1253
MSV-Raw    ----------.--g-.-------------------------------g-- 1254
Consensus  ********** * *  *********    ********  ********** ** 
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

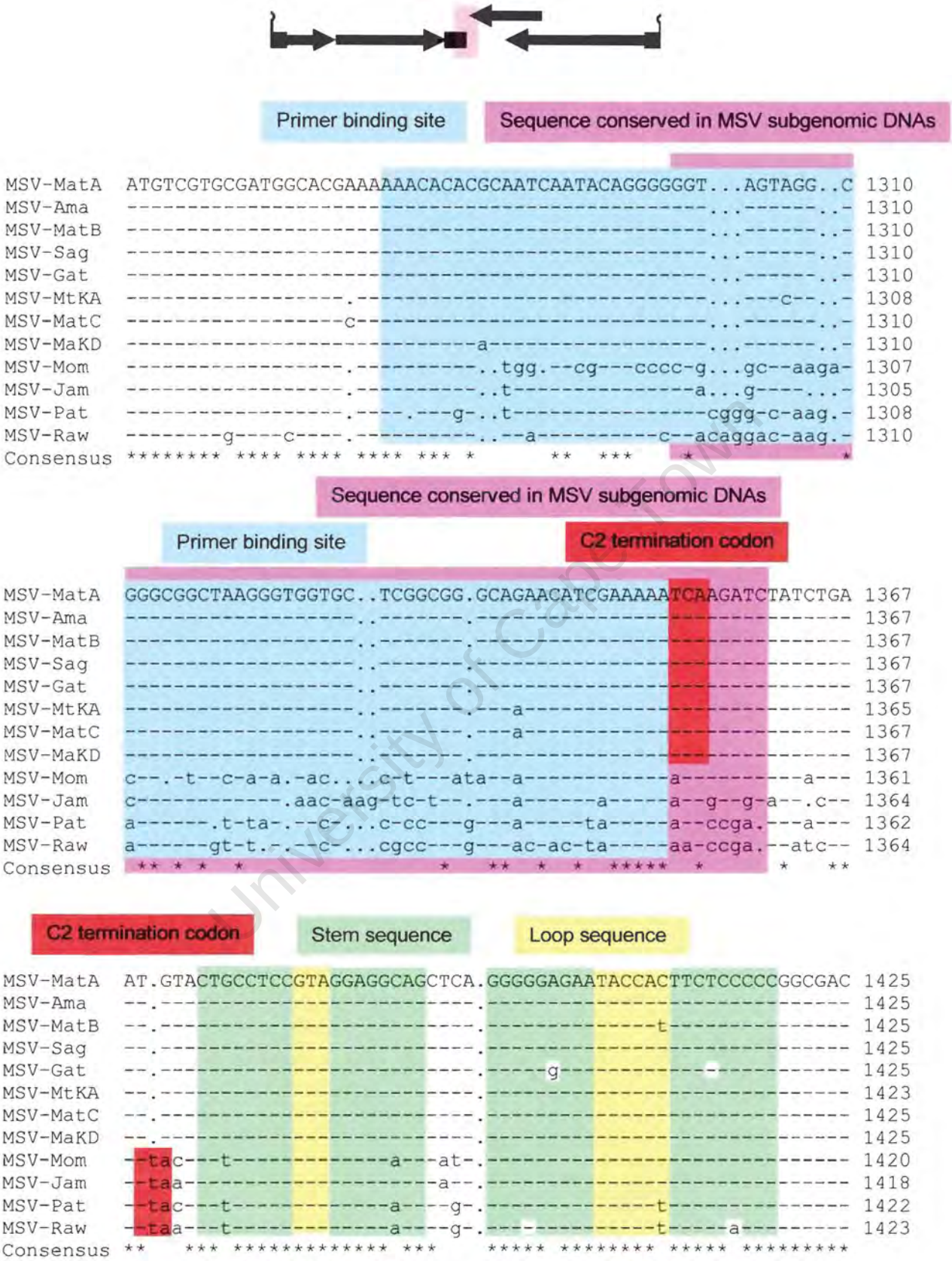


FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   ATAATGTAAATGACGCAGTTTGCCTCGAAATACTCCAGCTGCCCTGGAGTCATTTCCTTC 1485
MSV-Ama    -----------------a---------------------------------------- 1485
MSV-MatB   ------------------------------------------------------------ 1485
MSV-Sag    ------------------------------------------------------------ 1485
MSV-Gat    ------------------------------------------------------------ 1485
MSV-MtKA   ------------t-----------------a--------------t------------- 1483
MSV-MatC   ------------gt---a---------------------------------------- 1485
MSV-MaKD   -----cc----------a---------------------------------------- 1485
MSV-Mom    --tt-a-----c-t------g-----------------------------------a--- 1480
MSV-Jam    -----a-----c-t---------t------------------------------------ 1478
MSV-Pat    ----ca-----c-t---------t--------------t--------------------- 1482
MSV-Raw    -----------c-t------------------------------------------------ 1483
Consensus  **     ****   *** ** ** ******** ***** ********* ******* ***

MSV-MatA   ATCCAATCTTCATCCGAGTTGGCGAGGATTATTGTAGGCTTAGACTTCTTCTGCACCTTC 1545
MSV-Ama    --------------------------t--------------------------------- 1545
MSV-MatB   ------------------------------------------------------------ 1545
MSV-Sag    ------------------------------------------------------------ 1545
MSV-Gat    ------------------------------------------------------------ 1545
MSV-MtKA   ------------------------------------------------------------ 1543
MSV-MatC   ------------------------------------------------------------ 1545
MSV-MaKD   ------------------------------------------------------------ 1545
MSV-Mom    -----g---------c-------------g-----------g------a----a--t--- 1540
MSV-Jam    -----g-----------------------a-----------------a-------t--- 1538
MSV-Pat    -----g-----------------------------t---------a----a------ 1542
MSV-Raw    --------c-------------------g--a--t--t---------a----a------ 1543
Consensus  ***** ** ****** ********** ** ** ** ** ** ****** **** ** ***

MSV-MatA   TTCTTCTTACCATACTTGGGGTTCACAATGAAATCCCTCTGACAGCCAACTAACTGTTTC 1605
MSV-Ama    ----------------------t---------------------------------- 1605
MSV-MatB   ----------------------t---------------------------------- 1605
MSV-Sag    ----------------------t-----------------g------------------ 1605
MSV-Gat    ----------------------t---------------------------------- 1605
MSV-MtKA   -------------t--------t---------------------------------- 1603
MSV-MatC   ----------------------t---------------------------------- 1605
MSV-MaKD   ----------------------t---------------------------------- 1605
MSV-Mom    --t-----g-----t--t--a--t----c---c--tt---------------------- 1600
MSV-Jam    --------t-----t--t--a--t----c---c--tt---------------------- 1598
MSV-Pat    --------g-----t--a--a--t----c---c---t-t--g------------------ 1602
MSV-Raw    --t-----c--g--t--t--a--t----c---t--tt-t----------c--------- 1603
Consensus  ** ***** ** ** ** ** ** **** * * **  * ** ******** *********
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   CAACAAGGACAAAATTTAAACGGAATATCATCTACGATGTTGTAGATTGCGTCTTCGTTG 1665
MSV-Ama    ----------g---------------------------------------------- 1665
MSV-MatB   ----------g----------g------------------------------------ 1665
MSV-Sag    ----------g----------g------------------------------------ 1665
MSV-Gat    ----------g----------g------------------------------------ 1665
MSV-MtKA   ----------g---------------------------------------------- 1663
MSV-MatC   ----------g----------g----------a------------------------- 1665
MSV-MaKD   ----------g--------a--g-------------------------------a--- 1665
MSV-Mom    ----------gt-------a---------------------------------------- 1660
MSV-Jam    ----------gt-------t---------------------------------------- 1658
MSV-Pat    ----------t-------a--g-----------------a------------------- 1662
MSV-Raw    --g---------t-------a--g---------------------------------- 1663
Consensus  ** ***** **  ******* ** ********** ***** ************** ***
```

RepA termination codon

Potential polyadenylation signal for small C-sense transcript

```
MSV-MatA   TATGAAGACCAATCAACATTATTTTGCCAGTAATTATGAACCCCTAGGCTTCTGGCCCAA 1725
MSV-Ama    ------------------------------------------------------------ 1725
MSV-MatB   ------------------------------------------------------------ 1725
MSV-Sag    ------------------------------------------------------------ 1725
MSV-Gat    ----------------------------------------t------------------- 1725
MSV-MtKA   ------------------------------------------------------------ 1723
MSV-MatC   ----------------------------------------------c------------- 1725
MSV-MaKD   ------------------------------------------------------------ 1725
MSV-Mom    --a-----------------------a----------g-----a--------------- 1720
MSV-Jam    --a-------------------------------------------t-------------- 1718
MSV-Pat    --g-----------------------------------g-----aa-------------- 1722
MSV-Raw    ------------------------------------cgt-----a--------------- 1723
Consensus  ** ************************** *********    ***  * * ***** **
```

```
MSV-MatA   GTAGATTTTCCGGTTCTTGTTGGGCCGACGATGTAGAGGCTCTGCTTTCTTGATCTTTCA 1785
MSV-Ama    ---------------------c------------------------------------- 1785
MSV-MatB   ----------t------------------------------------------------- 1785
MSV-Sag    ----------t------------------------------------------------- 1785
MSV-Gat    ----------t------------------------------------------------- 1785
MSV-MtKA   ------------------------------------------------------------ 1783
MSV-MatC   ------------------------------------------------------------ 1785
MSV-MaKD   ----------t------------------------------------------------- 1785
MSV-Mom    --------c--a----------a-------------------------------c----- 1780
MSV-Jam    ----------a----------a-----------------------------t--t----- 1778
MSV-Pat    --------c--t--c--------a-----------------------------c----- 1782
MSV-Raw    ----------t--c--------a-----------------------------ca-t- 1783
Consensus  * *  *** *  *  ******** ** ** ** ** **** **** ****** **  * *
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

Acceptor site Intron Branch point

```
MSV-MatA  TCTGATGACTGGATACAGAATCCATCCATTGGAGGTCAGAAATTGCATCCTCGAGGGTAT 1845
MSV-Ama   --------------------------------------------------------- 1845
MSV-MatB  ---------------tt---------------------------------------- 1845
MSV-Sag   ---------------tt---------------------------------------- 1845
MSV-Gat   --g-------------tt---------------------------------------- 1845
MSV-MtKA  ------------------------------------g-------------t------ 1843
MSV-MatC  ------------------------------------------------t--------- 1845
MSV-MaKD  --------------------------------------------------------- 1845
MSV-Mom   ----t--g--------attt-----------ca--------g--a-----t--------g- 1840
MSV-Jam   ----t--g--------attt-----------ca--------g--a-----t--------- 1838
MSV-Pat   ----c--g--------agtt----------gtt--a--t--------t--a----c----- 1842
MSV-Raw   ----c-tg----t--a-tc-----g------a--a---------a--g--g----c----- 1843
Consensus  * * *  **** **    ***** ****   ** ** ** ** ** ** ****  ** *
```

Intron Donor site

```
MSV-MatA  AACAGGTAGGTTGAAGGAGCATGTAAGCTTCGGGACTAACCTGGAAGATGTTA.GGCTGG 1904
MSV-Ama   ------------------------t-------------------------.------ 1904
MSV-MatB  ---------------------------------------------------.------ 1904
MSV-Sag   ---------------------------------------------------.------ 1904
MSV-Gat   ---------------------------------------------------.------ 1904
MSV-MtKA  -----t--------------------------------a---------.------ 1902
MSV-MatC  -----t-----------------------g------g----------a---.------ 1904
MSV-MaKD  ------------------------c-------------------------.------ 1904
MSV-Mom   ----------------a-----a--t-a----------------a---------.------ 1899
MSV-Jam   -------t--------a-----a--t-a---t--a---------at-------t.------ 1897
MSV-Pat   ----------------a--------t-----------g-----at--------.------ 1901
MSV-Raw   -----t-t--------a-------------g------g-----t-t------.--t--c 1902
Consensus ***** * ******** *** * *  * ** *** ** *****  * ** **  ** **
```

```
MSV-MatA  AGCCAATCATTGATTGACTCATTACAAAGTAAATCAGGTGAGGAGGGTGGATGAGGATTG 1964
MSV-Ama   --------------------------------------------------------- 1964
MSV-MatB  ------------------------------------------tt-c------------ 1964
MSV-Sag   ------------------------------------------tt-c------------ 1964
MSV-Gat   --------------------------------------------------------- 1964
MSV-MtKA  -------------------------g--c---------tg------------------ 1962
MSV-MatC  ----------------t---------------------------------------- 1964
MSV-MaKD  --------------------c------------------t---tt-c---------- 1964
MSV-Mom   --------t--------------------a-g---------a---------------c-- 1959
MSV-Jam   --------t--------------------a-g---------a--t------------c-- 1957
MSV-Pat   --------t--------------------c-g---------a---------------c-- 1961
MSV-Raw   -a------t--------t--g--g--t--c--g--t---tc-----ta--g--------- 1962
Consensus * ****** *******   *  * ** ** *  ** ***   **    ** ****** **
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   GTGAACTCTTCCTGAATCTCAGGAAAAAGCTTATTTGCAGAGTATTCAAAATACTGCAAT 2024
MSV-Ama    ------------------------------------------------------------ 2024
MSV-MatB   ------------------------------------------------------------ 2024
MSV-Sag    -----t------------------------------------------------------ 2024
MSV-Gat    ------------------------------------------------------------ 2024
MSV-MtKA   ------------------------------------------------------------ 2022
MSV-MatC   ------------------------------------------------------------ 2024
MSV-MaKD   ------------------------------------------------------------ 2024
MSV-Mom    a----t----------------g--c-at------------------------------- 2019
MSV-Jam    a------------------------c-at------------------------------- 2016
MSV-Pat    a----------------g--------g--t--g-------------g------------- 2021
MSV-Raw    a----------------a--------c--t--g-----t-----c--g--g-----g-g- 2022
Consensus   ****  **********  **** ** *  *  ***** ***** ** ** ***** * *

MSV-MatA   TTTGTGGACCAATCAAAGGGAAGCTCTTTCTGGATCATGGAGAGGTACTCTTCCTTGGAA 2084
MSV-Ama    ------------------------------------------------t------- 2084
MSV-MatB   ---------------t-------------------------------------------- 2084
MSV-Sag    ---------------t-----------------------------------a-------- 2084
MSV-Gat    ---------------t-------------------------------------------- 2084
MSV-MtKA   ------------------------------------------------------------ 2082
MSV-MatC   -------c--------t------a-----------a-----------c---------- 2084
MSV-MaKD   ------ac--------t------a-------------------------------------- 2084
MSV-Mom    -------c---g---t----g-a-----ac-a-----a------------ag---t--- 2079
MSV-Jam    -------c---g---t----g-at-----ac-a-c--a------------aag---t--- 2076
MSV-Pat    --g----c---g--gt-a--g-at--c---c-a-c--a-c------------g------- 2081
MSV-Raw    --a----c--------t-a--g-atga---tc---c----------------agg---t--- 2082
Consensus ** ***  *** **  * ** *     **  * * **   **********   * * ***

MSV-MatA   GTAGCGTGTGAAATAATGTCTCGCATTATTTCATCTTTGGAAGG.TTTTTTTTCCT..TA 2141
MSV-Ama    ------------------------------------a-----.c----------..-t 2141
MSV-MatB   ----a-------------------------------------.c----------..-- 2141
MSV-Sag    ------------------------------------t-----.c----------..-- 2141
MSV-Gat    ----a-------------------------------------.c----------..-- 2141
MSV-MtKA   ----a-------------------------------------.c----------..-t 2139
MSV-MatC   ---ca----------------t--------------------.c----------..-t 2141
MSV-MaKD   ----a-------------------------------------.c----------..-- 2141
MSV-Mom    --g-a-----------c----t--------------------.---c-----agaa-t 2138
MSV-Jam    --g-a-----------c----t---------------t-----.---c-----agaa-t 2135
MSV-Pat    --g-----a----------g-gc----------------ttt--g---------gggaag 2141
MSV-Raw    -----a--a-----------t----------------a--t--.---------tgaa-t 2141
Consensus **    ** ******** * *  ***************    **  ** *****
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   ACTTCTGAATCAGATTTTCCTAGGAAGGGGGACTTCCTAGGAATGAAAGTACCTCTCTCA 2201
MSV-Ama    --c---------------------------------------------------------- 2201
MSV-MatB   ---------------------------------------------------------- 2201
MSV-Sag    ---------------------------------------------------------- 2201
MSV-Gat    ---------------------------------------------------------- 2201
MSV-MtKA   ---------------------------------------------------------- 2199
MSV-MatC   --c------------------------------------------------------- 2201
MSV-MaKD   ---------------------------------------------------------- 2201
MSV-Mom    t-cct--g-..--ga------t-----ctt------------------------------ 2196
MSV-Jam    c-cct--g-..--ga----------ctt----c---g-------------------- 2193
MSV-Pat    g......gggg--ta------t-----cat-------g---------------------- 2195
MSV-Raw    t-cct-tg-gg--..----------t-tttt----------c---------------g 2199
Consensus            **  ****** ****     ** **  **** ***************

MSV-MatA   AAAACAGCCAGAGGTTCCTTGAGAATGTAATCCCTCACCCTGTTTACTGATTTGGCACTC 2261
MSV-Ama    --c-------------------------------------------c--------- 2261
MSV-MatB   --c-t----------------g-------------------------c--------- 2261
MSV-Sag    --c-t----------------g-------------------------c--------- 2261
MSV-Gat    --c-t--------------t--g----------t-----------c--------- 2261
MSV-MtKA   --c-------------------------------------a-----c--------- 2259
MSV-MatC   --c-----------------------------------a--g--c--------- 2261
MSV-MaKD   --c-t---------------------------------------c--------- 2261
MSV-Mom    --c--------g-------------------g---t-------------t------ 2256
MSV-Jam    --c-a----------------------g--tt-------a-----t------ 2253
MSV-Pat    --t--t-----t-------------g--g-gt---t-------g--c--a------ 2255
MSV-Raw    --c-a---------------g--------t--t---t----agggctca----g--- 2259
Consensus  ** *  ***** ******** ** ***** ** *  **  ****        * ** ***

MSV-MatA   TGAATGTTTGGGTGAAACCCATTAATATCAAAGAACCTTGAGTCGGATATCCTTACCGGC 2321
MSV-Ama    -----a----------------t-------------------a--------------- 2321
MSV-MatB   ---------------------------------------c-----t--------- 2321
MSV-Sag    ---------------------------------------c-----t--------- 2321
MSV-Gat    ---------------------------------------c-----t--------- 2321
MSV-MtKA   -----a----------------t-------------------a--------------- 2319
MSV-MatC   ----------------------t-------------------c-----t--------- 2321
MSV-MaKD   ---------------------------------------c-----t--------- 2321
MSV-Mom    --------------g-----t-cg-------------------a-tg--t--------t 2316
MSV-Jam    --------------g-----g--g-------------------a-tg--t--------- 2313
MSV-Pat    --------------g--t--t-ct--g--g--a-t---g-ca----tcg----c--t--t 2315
MSV-Raw    --------------g--a----ct--g--g--------g-------tgg-t--c--tt-t 2319
Consensus  ***** ******** ** ** *  ** ** ** * *** *  ** *   * ** **  *
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

```
MSV-MatA   TTCTCTGTCTGAAGTAATGCATGTAAATGCAAACTTCCATCTTTATGTGCCTCTCGGGCA 2381
MSV-Ama    ----------c---------------------------------------------- 2381
MSV-MatB   --------t-----c---------------------------------------------- 2381
MSV-Sag    --------t-----c----------------------g--c------------------ 2381
MSV-Gat    --------t-----c---------------------------------------------- 2381
MSV-MtKA   ----------c---------------------------------------------- 2379
MSV-MatC   --------t-----c-----------------------------g--------------- 2381
MSV-MaKD   --------t-----c-----------------------------g--------------- 2381
MSV-Mom    ----------g-------------------t-tc---g---c----------t------ 2376
MSV-Jam    --a--------g-c-----------------t-tc---------g---------------- 2373
MSV-Pat    --a--c-----t--------------g----t-t-------acg----------------- 2375
MSV-Raw    --a--c-----t-t-----------------t-t-------c-c----------t------ 2379
Consensus  ** ** ** ** *  ********** **** *  *** **    ********* *****

MSV-MatA   CATAGAATGTATTTGGGAATCCAACGAACAACGAGCTCCCAGATCATCTGACAGGCGATT 2441
MSV-Ama    --------a-------------------------------------g---------- 2441
MSV-MatB   ----------c---------------------------------------------- 2441
MSV-Sag    ----------c---------------------------------------------- 2441
MSV-Gat    ----------c---------------------------------------------- 2441
MSV-MtKA   --------a---------------------------------------------- 2439
MSV-MatC   --g-c---a--c--c--t-------------------------t------------------ 2441
MSV-MaKD   --g-c---a--c-----t---------------------------------------- 2441
MSV-Mom    ----tg--a--c--c--gg--------c-------------a------------a----- 2436
MSV-Jam    ----t---a--c--a--gg--------c-g-----------a------------a----- 2433
MSV-Pat    ----cg-----caga---t---------g-g-----a-----t-gtt------t-at---- 2435
MSV-Raw    ----tg-----caga---t---------c-------a---------c------t-at---- 2439
Consensus  ** *  ** **    **  ******** * ***** ***** *   **** * *  ****

MSV-MatA   TCAGGATTTTCTGGACACTTTGGATAGGTGAGGAACGTGTTAGCGTTCCTGTGTGAGAAC 2501
MSV-Ama    ----------------------------t------------------------------ 2501
MSV-MatB   ---------------------------------------------g---------- 2501
MSV-Sag    ---------------------------------------------g---------- 2501
MSV-Gat    ---------------------------------------------g---------- 2501
MSV-MtKA   ----------------------g--c--t------------------------------ 2499
MSV-MatC   ----------------------g-----t-a---------ca----------------- 2501
MSV-MaKD   ----------------------------t-a---------ca----------------- 2501
MSV-Mom    --t--------------g-g------t--t-a----------a----------------- 2496
MSV-Jam    --t--------------g-g------t--t-a----------a----------------- 2493
MSV-Pat    --t--------------g-g------t--------t-----t--a---------a----- 2495
MSV-Raw    -------g-----------g------a--------t-----t--ag----a---ag---- 2499
Consensus  ** **** ********* * *** ** ** * *** *****  *  ***  ***  ****
```

FIGURE D.1 Alignment of full length genomic sequences (Cont)

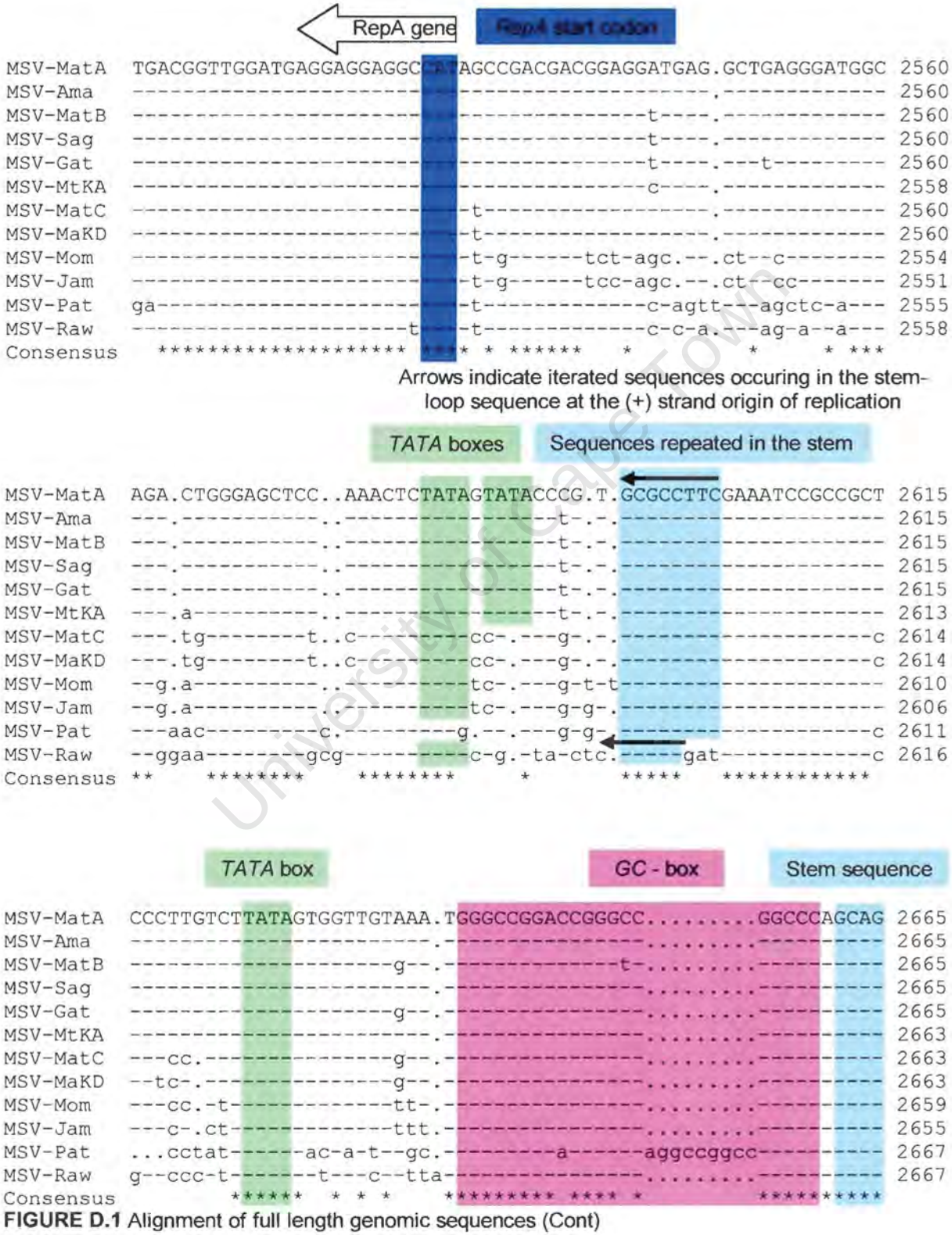


FIGURE D.1 Alignment of full length genomic sequences (Cont)

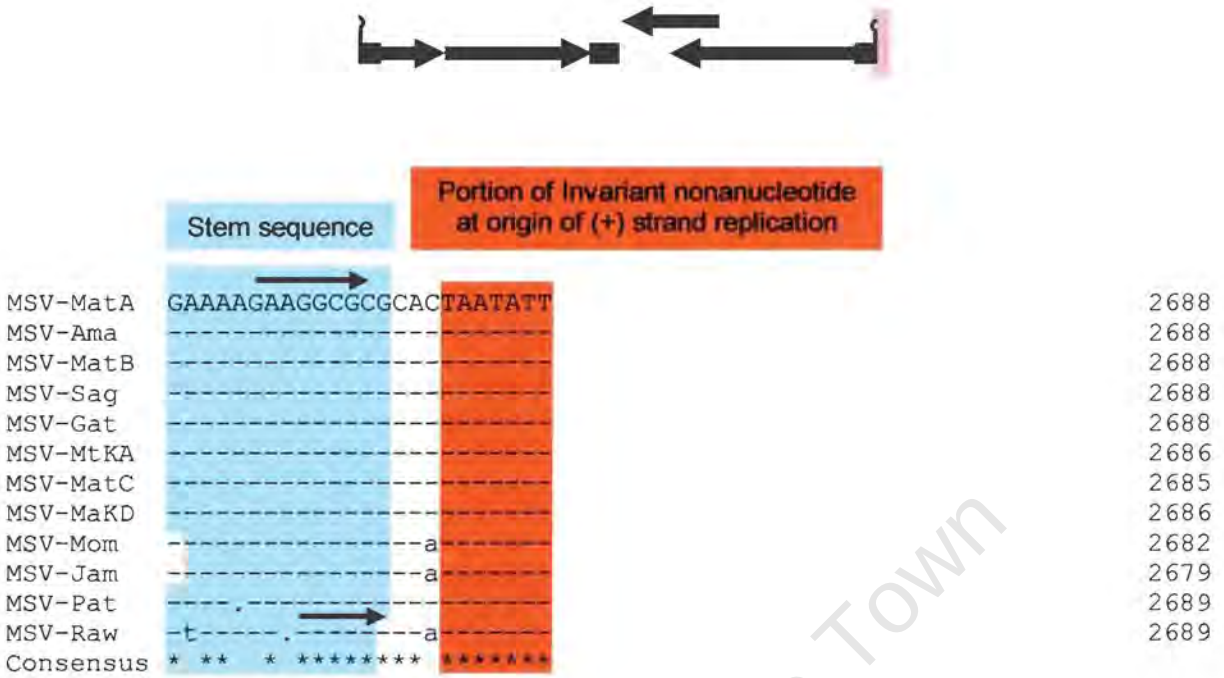


FIGURE D.1 Multiple alignment of the full genome sequences of 12 MSV isolates sequenced in this study. At positions where sequences are identical to MSV-MatA nucleotides are represented by a dash (-). Dots (.) represent spaces inserted into the alignment to compensate for insertions and deletions in certain of the sequences. Numbering begins at the origin of (+) strand replication. A schematic genome diagram at the top of each page of the figure indicates the portion of the genome represneted on that page. Highlighted are sequence elements with either known or suspected functionality (see Chapter 1.4.1 for details).

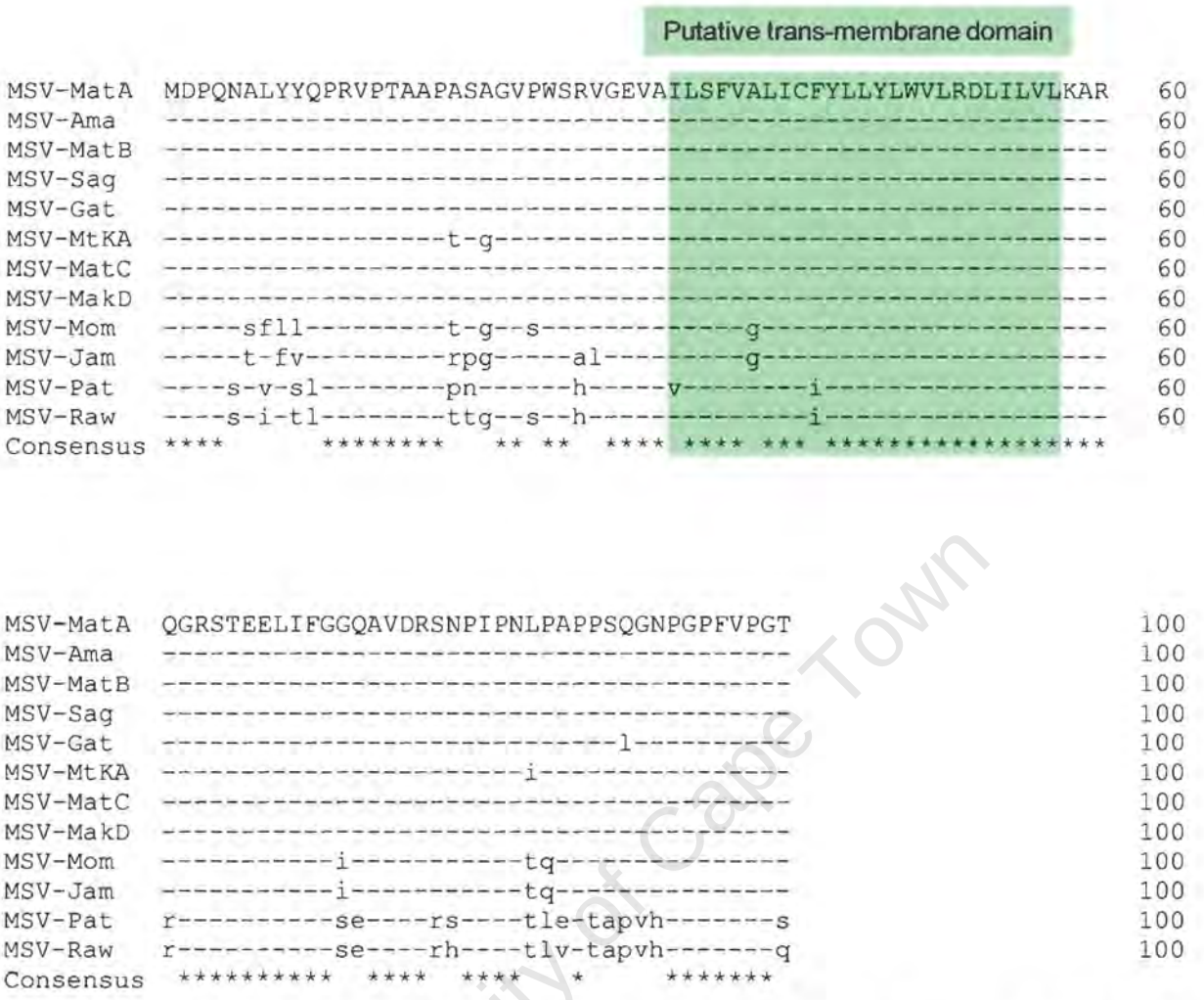


FIGURE D.2 Multiple alignment of movement protein amino acid sequences predicted from the MSV genomic sequence determined in this study. Highlighted is a portion of the sequence that is predicted to be within a trans-membrane domain (see Chapter 1.4.1.4 for details).

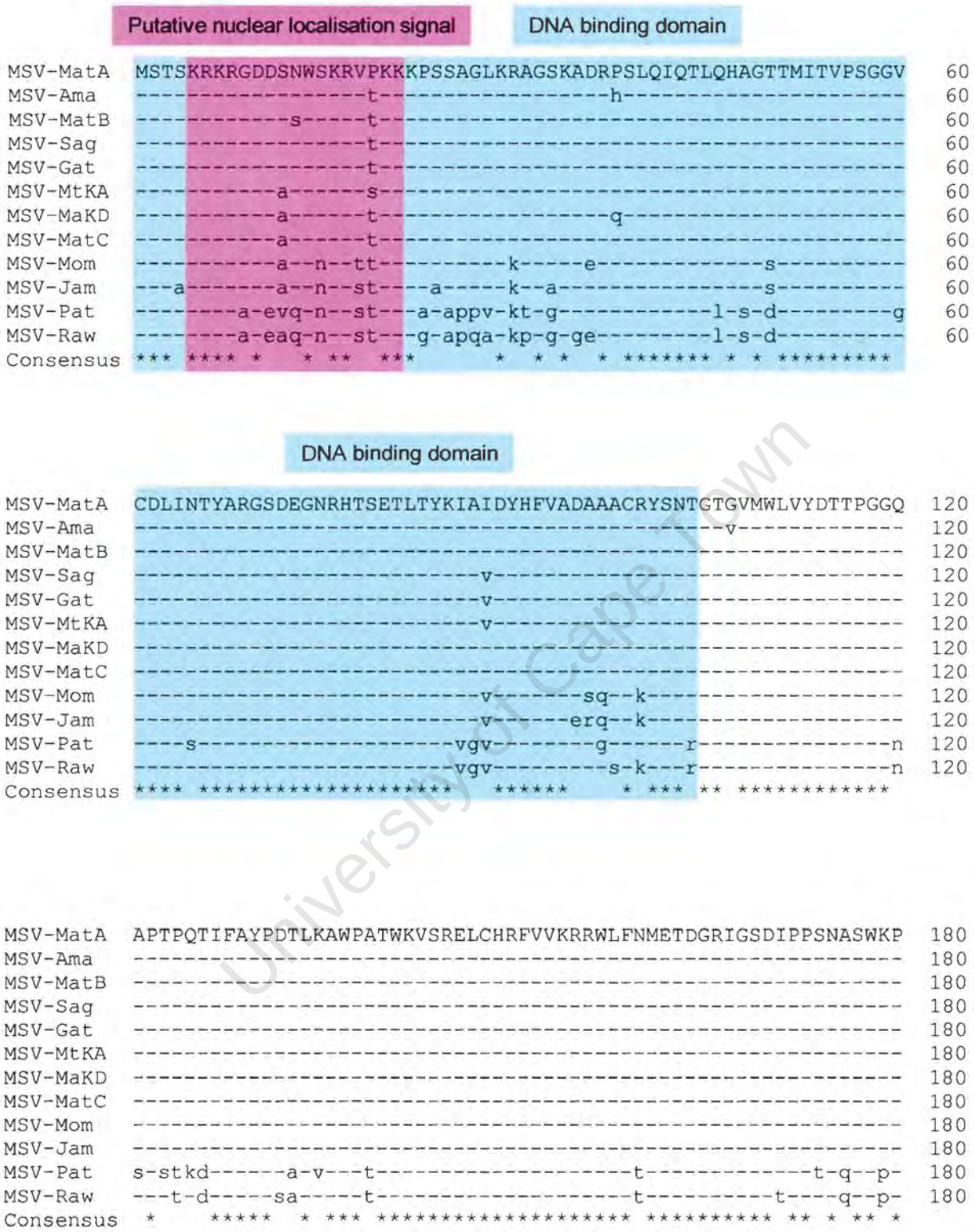


FIGURE D.3 Alignment of predicted coat protein sequences (Cont)

```
MSV-MatA   CKRNIYFHKFTSGLGVRTQWKNVTDGGVGAIQRGALYMVIAPGNGLTFTAHGQTRLYFKS  240
MSV-Ama    ------------------------------------------------------------  240
MSV-MatB   ------------------------------------------------------------  240
MSV-Sag    ------------------------------------------------------------  240
MSV-Gat    ------------------------------------------------------------  240
MSV-MtKA   ------------------------------------------------------------  240
MSV-MaKD   ------------------------------------------------------------  240
MSV-MatC   ------------------------------------------------------------  240
MSV-Mom    ------------------------------------------------------------  240
MSV-Jam    ------------------------------------------------------------  240
MSV-Pat    -----d-----------------------------------------v-------------  240
MSV-Raw    ----vd-------------------------------l-------i-------------  240
Consensus  ****  ******************************* ******* **************

MSV-MatA   VGN                                                           243
MSV-Ama    ---                                                           243
MSV-MatB   ---                                                           243
MSV-Sag    ---                                                           243
MSV-Gat    ---                                                           243
MSV-MtKA   ---                                                           243
MSV-MaKD   ---                                                           243
MSV-MatC   ---                                                           243
MSV-Mom    ---                                                           243
MSV-Jam    ---                                                           243
MSV-Pat    ---                                                           243
MSV-Raw    ---                                                           243
Consensus  ***
```

FIGURE D.3 Multiple alignment of coat protein amino acid sequences predicted from the MSV genomic sequence determined in this study. Highlighted is a portions of the sequence with either known or suspected functionality (see Chapter 1.4.1.4 for details).

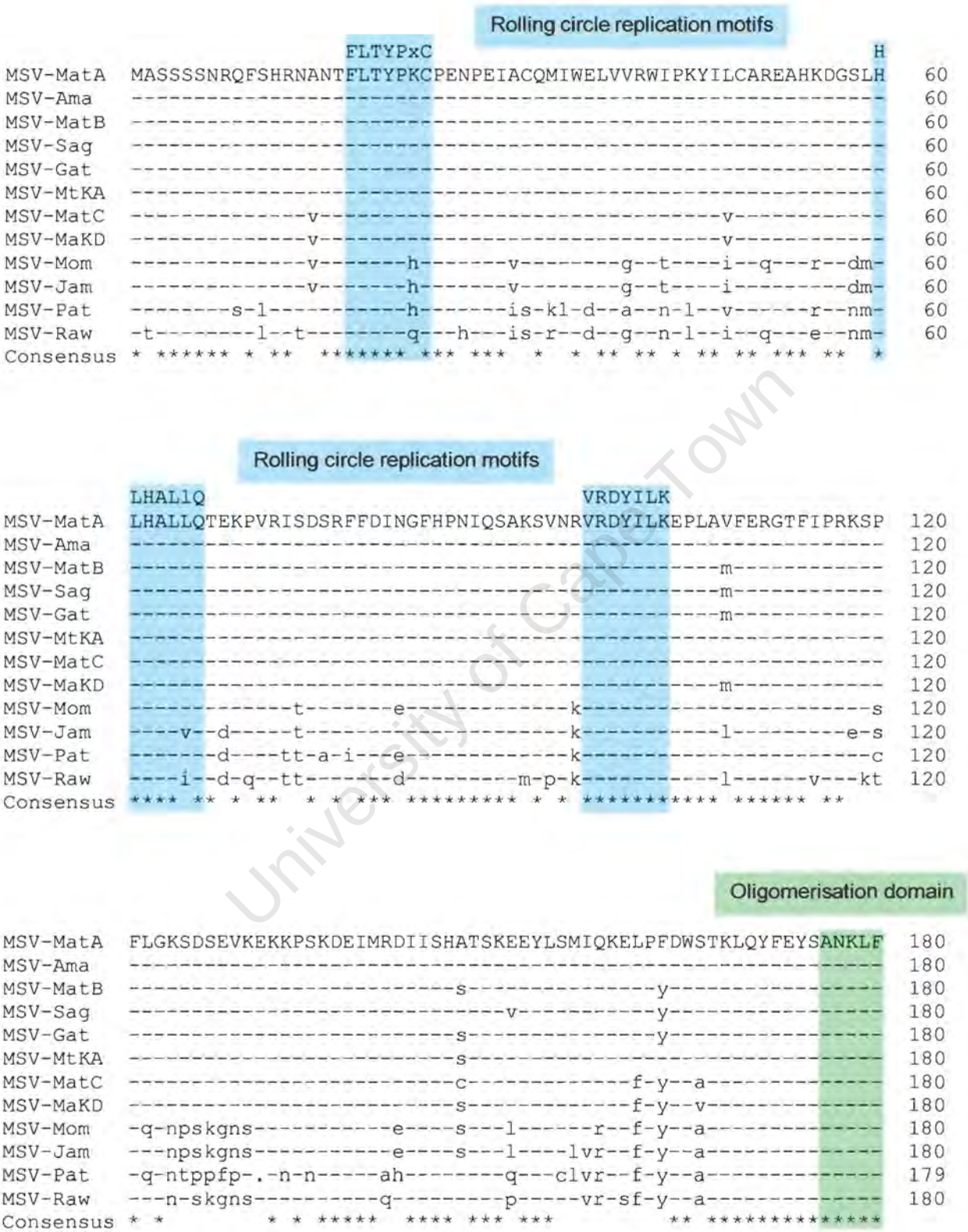


FIGURE D.4 Alignment of predicted relication associated protein sequences (Cont)

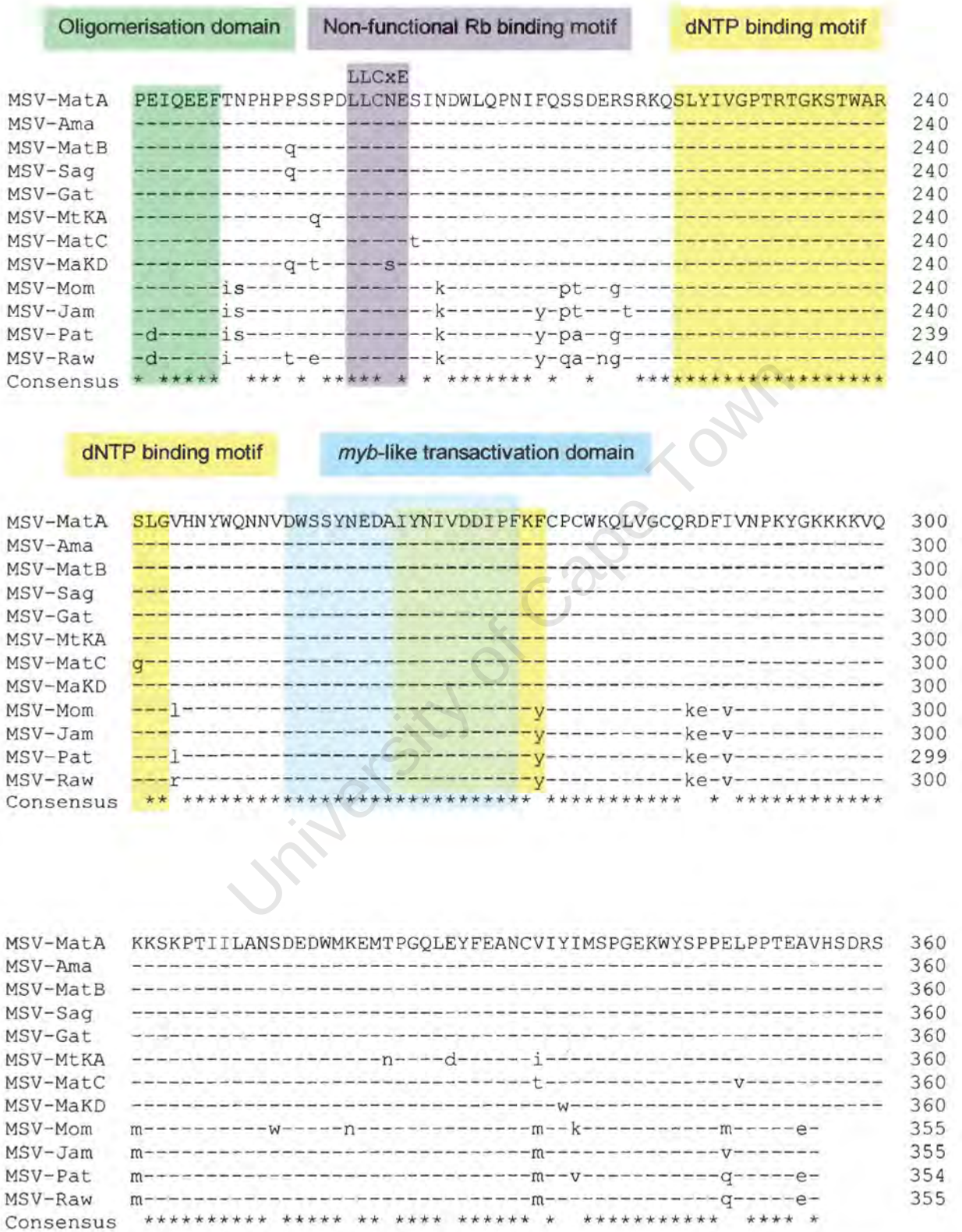


FIGURE D.4 Multiple alignment of replication associated protein amino acid sequences predicted from the MSV genomic sequence determined in this study. Highlighted is a portions of the sequence with either known or suspected functionality (see Chapter 1.4.1.3 for details).

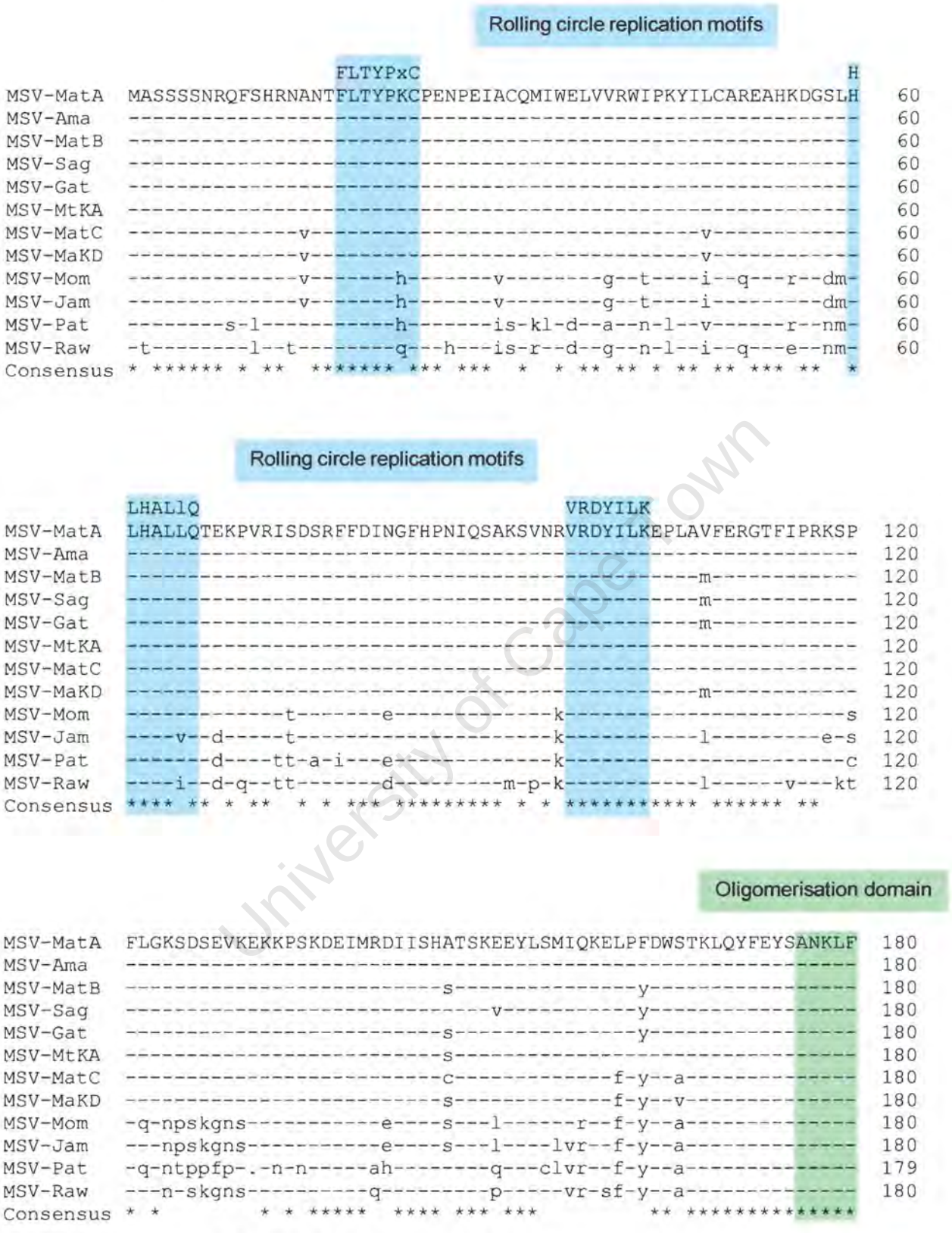


FIGURE D.5 Alignment of predicted RepA sequences (Cont)

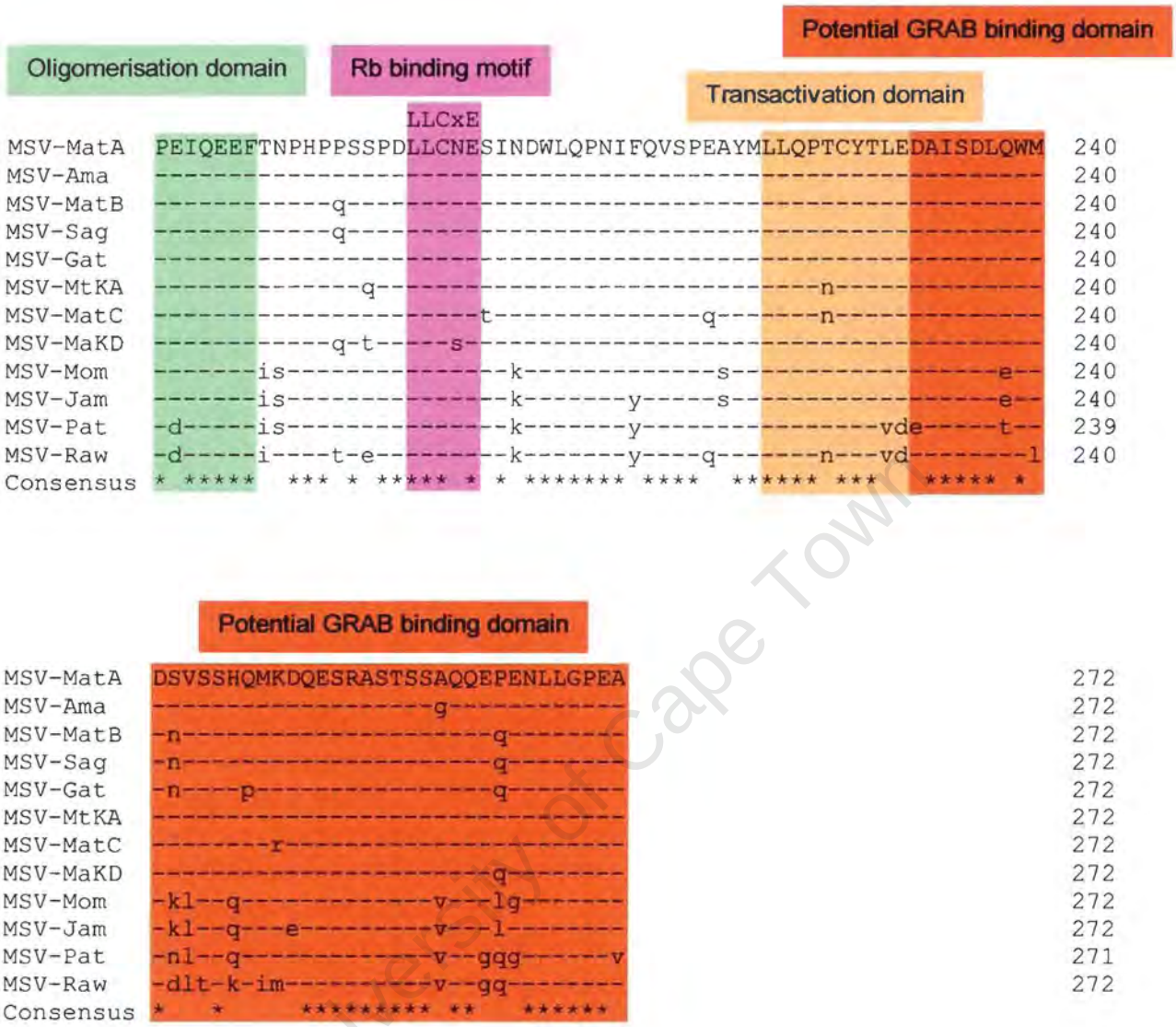


FIGURE D.5 Multiple alignment of RepA amino acid sequences predicted from the MSV genomic sequence determined in this study. Highlighted is a portions of the sequence with either known or suspected functionality (see Chapter 1.4.1.3 for details).

APPENDIX E

Symptoms Produced by Chimeric MSV-MatA, -Kom, -R2, and -VW genomes in sweetcorn and PAN6099

Contents

FIGURE E.1 Symptoms occurring on the 5th leaf of sweetcorn (a MSV-sensitive maize genotype) agroinoculated with MSV-MatA/Kom chimeras.

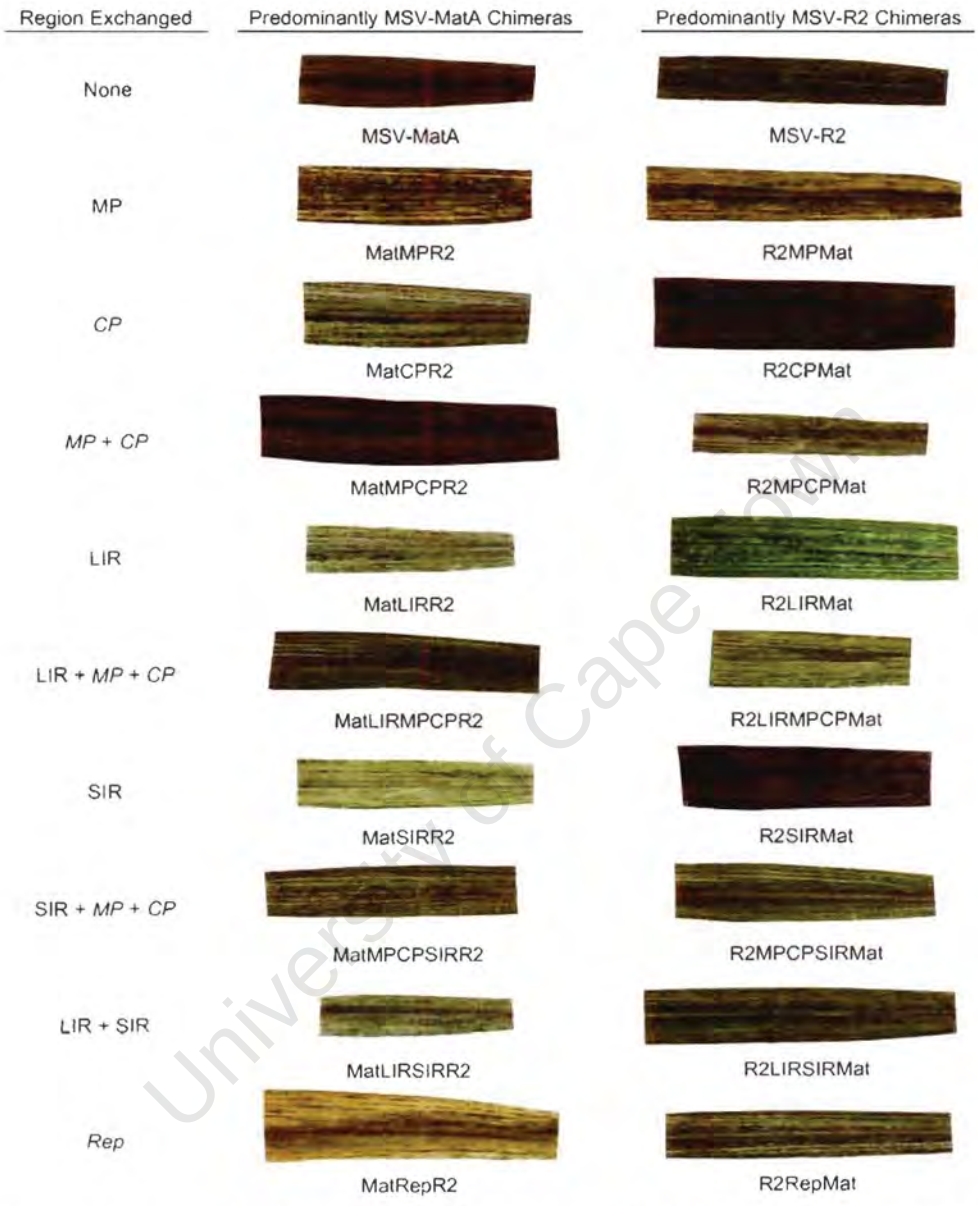


FIGURE E.2 Symptoms occurring on the 5th leaf of sweetcorn (a MSV-sensitive maize genotype) agroinoculated with MSV-MatA/R2 chimeras.

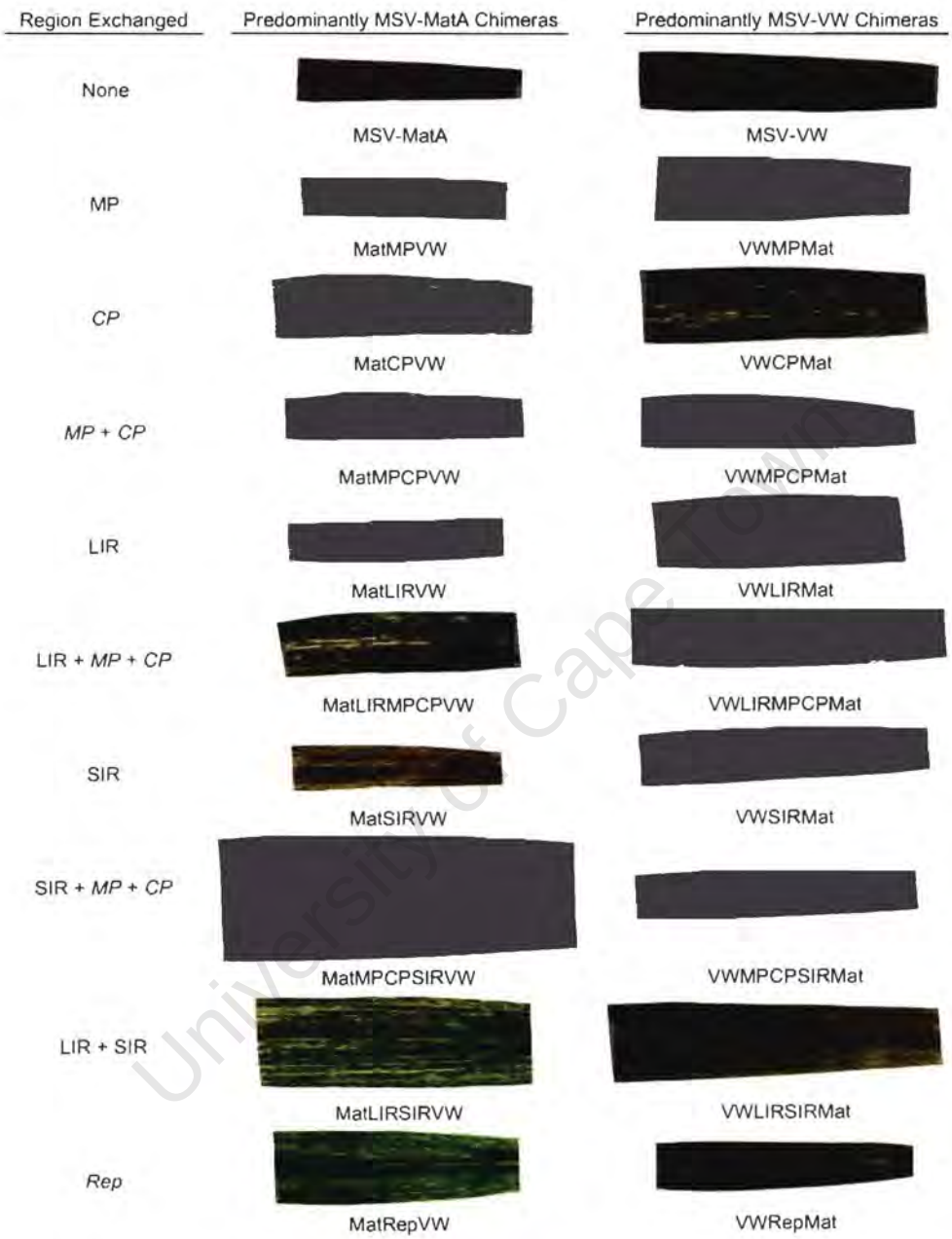


FIGURE E.3 Symptoms occurring on the 5th leaf of sweetcorn (a MSV-sensitive maize genotype) agroinoculated with MSV-MatA/VW chimeras.

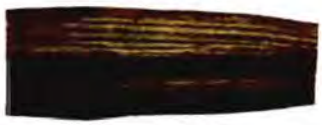
Region Exchanged	Predominantly MSV-MatA Chimeras	Predominantly MSV-Kom Chimeras
None	 MSV-MatA	 MSV-Kom
MP	 MatMPKom	 KomMPMat
CP	 MatCPKom	 KomCPMat
MP + CP	 MatMPCPKom	 KomMPCPMat
LIR	 MatLIRKom	 KomLIRMat
LIR + *MP* + *CP*	 MatLIRMPCPKom	 KomLIRMPCPMat
SIR	 MatSIRKom	 KomSIRMat

Fig E.4 Cont.

FIGURE E.4 Symptoms occurring on the 5th leaf of PAN6099 (a moderately MSV-resistant maize genotype) agroinoculated with MSV-MatA/Kom chimeras.

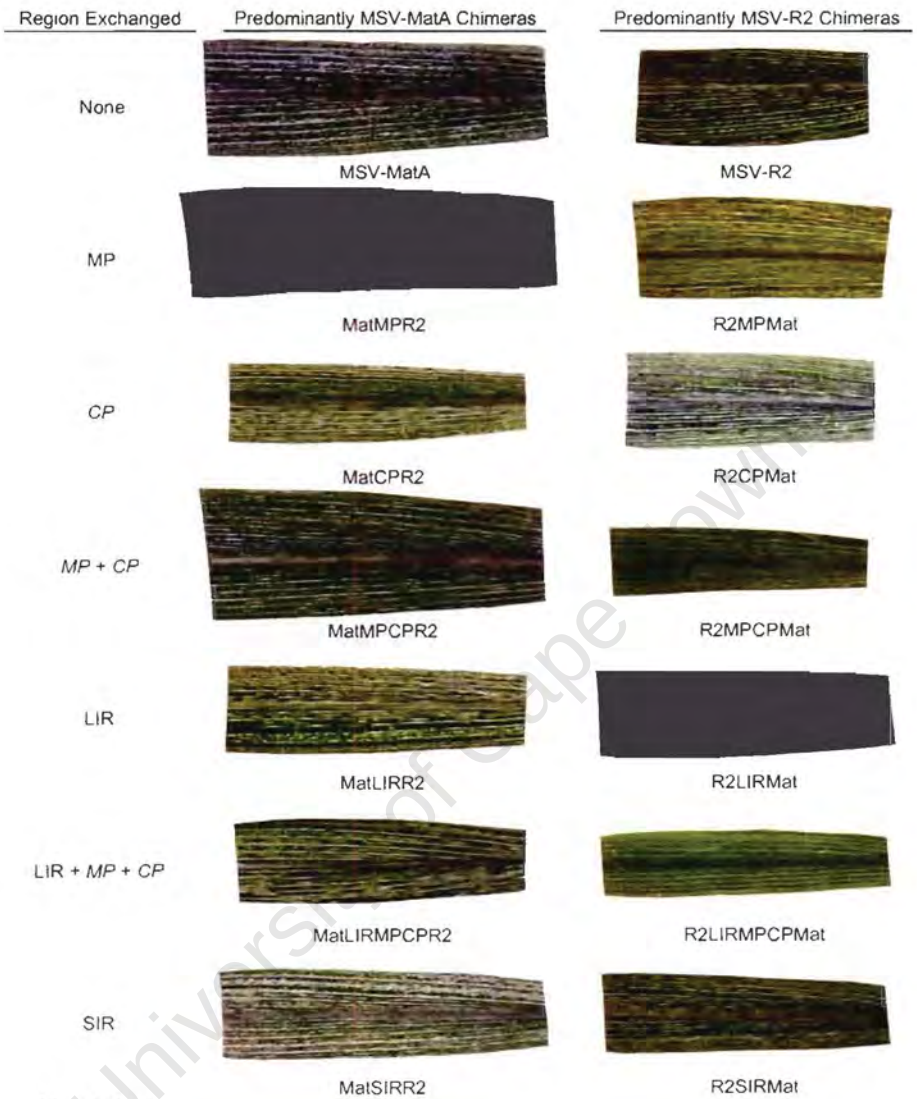


Fig E.5 Cont.

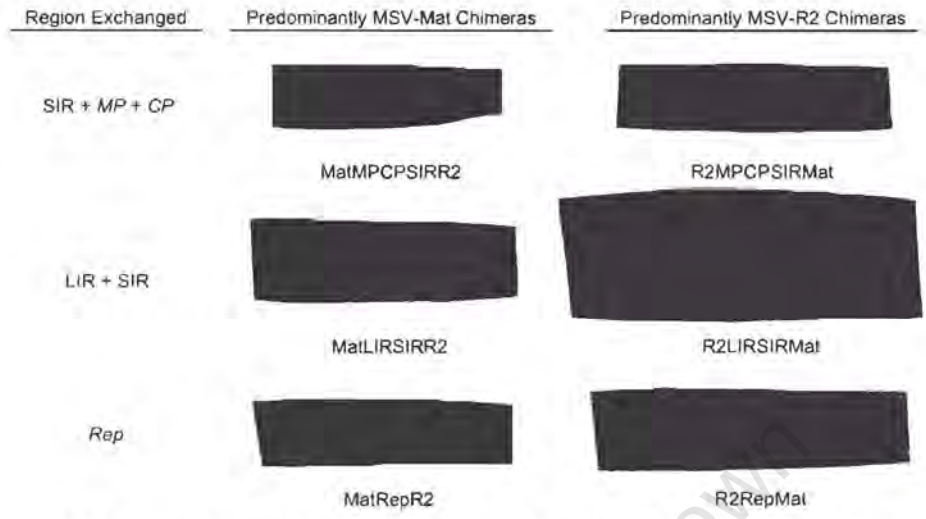


FIGURE E.5 Symptoms occurring on the 5th leaf of PAN6099 (a moderately MSV-resistant maize genotype) agroinoculated with MSV-MatA/R2 chimeras.

Region Exchanged	Predominantly MSV-MatA Chimeras	Predominantly MSV-VW Chimeras
None	 MSV-MatA	**NS** MSV-VW
MP	 MatMPVW	**NS** VWMPMat
CP	 MatCPVW	**NS** VWCPMat
MP + CP	 MatMPCPVW	**NP** VWMPCPMat
LIR	 MatLIRVW	**NS** VWLIRMat
LIR + *MP* + *CP*	**NP** MatLIRMPCPVW	**NP** VWLIRMPCPMat
SIR	 MatSIRVW	**NS** VWSIRMat

Fig E.6 Cont.

Region Exchanged	Predominantly MSV-Mat Chimeras	Predominantly MSV-VW Chimeras
SIR + *MP* + *CP*	**NS**	**NS**
	MatMPCPSIRVW	VWMPCPSIRMat
LIR + SIR		**NS**
	MatLIRSIRVW	VWLIRSIRMat
Rep	**NP**	**NS**
	MatRepVW	VWRepMat

FIGURE E.6 Symptoms occurring on the 5th leaf of PAN6099 (a moderately MSV-resistant maize genotype) agroinoculated with MSV-MatA/VW chimeras. NS = No symptoms ever observed. NP = symptoms rarely occur but have not been photographed.

References

Abouzid, A.M., Frischmuth, T., and Jeske, H.(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., Donson, J., and Mullineaux, P.M.(1989). Mapping of *Digitaria* streak virus transcripts reveals different RNA species from the same transcription unit. *EMBO J.* **8**, 1033-1039.

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

Akbar, B.S., Dry, I.B., and Ali, R.M.(1998). Identification of the replication-associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. *Nucleic.Acids.Res.* **26**, 925-931.

Ammar, E.D., Abul, A.E., El-Sheikh, M.A, and Sewify, G.H.(1987). Incidence of virus and viruslike disease syndromes on maize and sugarcane in Middle and Lower Egypt. *Egypt. J. Phytopath.* **19**, 97-107.

An, G, Ebert, P.R., Mitra, A., Ha, S.B. (1988) Binary vectors. Pages 87-88. *In*: Plant Molecular Biology Manual (Gelvin S.B. and Schilperoort R.A., eds). Kluwer Dordrecht, The Netherlands.

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

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

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

Asanzi, C.M., Bosque-Perez, N.A., Buddenhagen, I.W., Gordon, D.T., and Nault, L.R. (1994). Interactions among maize streak virus disease, leafhopper vector populations and maize cultivars in forest and savanna zones of Nigeria. *Plant-Path.* **43**, 145-157.

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

Barrow, M.R.(1992). Development of maize hybrids resistant to maize streak virus. *Crop Protect.* **11**, 267-271.

Bigarré, L., Salah, M., Granier, M., Frutos, R., Thouvenel, J.-C., and Peterschmitt, M.(1999). Nucleotide sequence evidence for three distinct sugarcane streak mastreviruses. *Arch Virol.* **144**, 2331-44.

Bisaro, D.M.(1996). *DNA replication in eukaryotic cells* (De Pamphilis, M.L., Ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 854-883.

Blanchette R.A.(1982). New technique to measure tree defect using an image analyser. *Plant Dis.* **66**, 394-397.

Bock, K.R., Guthrie, E.J., and Woods, R.D.(1974). Purification of maize streak virus and its relationship to viruses associated with streak diseases of sugarcane and *Panicum maximum*. *Ann.Appl.Biol.* **77**, 289-296.

Boevink, P., Chu, P.W., and Keese, P.(1995). Sequence of subterranean clover stunt virus DNA: affinities with the geminiviruses. *Virology* **207**, 354-361.

Bosque-Perez, N.A., and Allem, M. S.(1992). Mass rearing of *Cicadulina* leafhoppers to screen for maize streak virus resistance. A manual. International Institute of Tropical Agriculture, Ibadan, Nigeria.

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

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

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

Boulton, M.I., Pallaghy, C.K., Chatani, M., Macfarlane, S., and Davies, J.W.(1993). Replication of Maize Streak Virus Mutants in Maize Protoplasts - Evidence for a Movement Protein. *Virology* **192**, 85-93.

Breret, J., Clerot, D., and Gronenborn, B.(1999). ATP regulates the quarternery structure of geminivirus Rep proteins. XI International Congress of Virology. Sydney, Australia.

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

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

Briddon, R.W. and Markham, P.G.(1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses.* (Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A., and Summers, M.D., Eds.) Springer-Verlag, Wien. 158-165.

Briddon, R.W., Bedford, I.D., Tsai, J.H., and Markham, P.G.(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., Lunness, P., Bedford, I.D., Chamberlin, L.C., Mesfin, T., and Markham, P.G.(1996b). A streak disease of pearl millet caused by a leafhopper-transmitted geminivirus. *Eur.J.Plant Pathol.* **102**, 397-400.

Buchen-Osmond, C.(1998a). ICTVdB. http://www.ncbi.nlm.gov/ICTVdB/ICTV/dB/29010006.htm, *Maize streak virus*

Buchen-Osmond, C.(1998b). ICTVdB. http://www.ncbi.nlm.gov/ICTVdB/ICTV/dB/29010010.htm, *Sugarcane streak virus*

Castellano, M.M., Sanz-Burgos, A.P., and Gutierrez, C.(1999). Initiation of DNA replication of a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin. *Journal of Molecular Biology* **290**, 639-52.

Choi, I.R. and Stenger, D.C.(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.

Collin, S., Fernandez-Lobato, M., Gooding, P.S., Mullineaux, P.M., and Fenoll, C.(1996). The two nonstructural proteins from wheat dwarf virus involved in viral gene expression and replication are retinoblastoma- binding proteins. *Virology* **219**, 324-329.

Dabrowski, Z.T.(1987). *Cicadulina ghaurii* (Hem., Euscelidae): distribution, biology and maize streak virus (MSV) transmission. *J. Appl. Entomol.* **103**, 489-496.

Dabrowski, Z.T., Wilson, M.R.(ed.), and Nault, L.R.(1987) Comparative studies of *Cicadulina* leafhoppers in West Africa. *Proceedings of 2nd International Workshop on Leafhoppers and Planthoppers of Economic Importance*. p35-39. Brigham Young University, Provo, Utah, USA, 28th July-1st August 1986.

Dabrowski, Z.T., Nwilene F., and Kumar, R.(1991). First regular observations on leafhoppers, Cicadulina spp., vectors of maize streak virus (MSV) in southeastern Nigeria. *Insect Science and its Application* **12**, 249-261.

Damsteegt, V.D.(1983). Maize Streak Virus: I. Host range and vulnerability of maize germ plasm. *Plant Disease* **67**, 734-737.

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

Desbiez, C., David, C., and Mettouchi, A.(1995). Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. *Proc.Nat. Acad. Sci U.S.A* **92**, 5640-5644.

Dickinson, V.J., Halder, J., and Woolston, C.J.(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.

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

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

Downham, M.C.A., and Cooter R.J.(1998). Tethered flight and morphometric studies with *Cicadulina storeyi* and *C. mbila* leafhoppers (*Hemiptera: Cicadellidae*) vectors of maize streak virus in Uganda. *Bull. Entomol.Res.* **88**, 117-125.

Drinkwater, T.W., Walters, M.C., and Van Rensburg, J.B.J.(1979). The application of systemic insecticides to the soil for the control of the maize stalk borer, *Busseola fusca* (Fuller) (*Lep.: Noctuidae*), and of *Cicadulina mbila* (Naude) (*Hem.: Cicadellidae*), the vector of maize streak virus. *Phytophylactica* **11**, 5-11.

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

Efron, Y., Kim, S.K., Fajemisin, J.M., Mareck, J.H., Tang, C.Y., Dabrowski, Z.T., Rossel, H.W., Thottappilly, G., and Buddenhagen, I.W.(1989). Breeding for resistance to maize streak virus: A multidisciplinary team approach. *Plant Breed.* **103**, 1-36.

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

Fajemisin, J.M., and Shoyinka, S.A.(1976). Maize streak and other virus disease in West Africa. Proc. International Maize Virus Disease Colloquium and Workshop, Wooster Ohio, USA p52-61.

Fennah, R.G.(1959). A new species of *Cicadulina* (*Homoptera: Cicadelidae*) from East Africa. *Ann. Mag. Nat. Hist.* **2**, 757-758.

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

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

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

Fontes, E.P., Gladfelter, H.J., Schaffer, R.L., Petty, I.T., and Hanley-Bowdoin, L.(1994a). Geminivirus replication origins have a modular organization. *Plant Cell* **6**, 405-416.

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

Fuller, C.(1901). *1st Report of the Government Entomologist, Natal, 1899-1900* (Anonymous

Ghanim, M., Morin, S., Zeidan, M., and Czosnek, H.(1998). Evidence for transovarial transmission of tomato yellow leaf curl virus by its vector, the whitefly Bemisia tabaci. *Virology* **240**, 295-303.

Gibbs M.J., and Weiller, G.F. (1999) Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. *Proc. Natl. Acad. Sci. USA*, **96**, 8022-8027.

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

Gorter, G.J.M.A.(1953). Studies on the spread and control of streak disease of maize. *Sci. Bull. Dept. Agric. & For. Union of S. Afr.* **134**: 20.

Govinden, N., and Rummun, K.(1996). Registration of MSIRI 3B streak-resistant maize germplasm. *Crop Sci.* **36**, 824.

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

Grassly, N.C., and Holmes E.C.(1997). A likelihood method for the detection of selection and recombination using nucleotide sequences. *Mol. Biol. Evol.* **14**, 239-247.

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

Groning, B.R., Abouzid, A., and Jeske, H.(1987). Single-stranded DNA from abutilon mosaic virus is present in the plastids of infected Abutilon sellovianum. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8996-9000.

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

Groning, B.R., Hayes, R.J., and Buck, K.W.(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.

Gutierrez, C., Suarez-Lopez, P., Ramirez-Parra, E., Sanz-Burgos, A.P., Ponninger, J., and Xie, Q.(1995). DNA bending as a potential regulatory cis-acting element of the geminivirus intergenic region. *Agronomie* **15**, 415-420.

Gutierrez, C.(1999). Geminivirus DNA replication. *Cell. mol. Life Sci.* **56**, 131-329.

Gutierrez, C.(2000). DNA replication and cell cycle in plants: learning from geminiviruses. *EMBO J* **19**, 792-799.

Hamel, P.A., Gallie, B.I., and Phillips, R.A.(1992). The retinoblastoma protein and cell cycle regulation. *Trends Genet.* **8**, 180-185.

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

Hanley-Bowdoin, L., Elmer, J.S., and Rogers, S.G.(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., Settlage, S.B., Orozco, B.M., Nagar, S., and Robertson, B.M.(1999). Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Plant Sciences* **18**, 71-106.

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

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

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

Hayes, R.J., MacDonald, H., Coutts, R.H., and Buck, K.W.(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.

Hayes, R.J. and Buck, K.W.(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. *Nucleic.Acids.Res.* **17**, 10213-10222.

Hein, J.(1990). Reconstructing evolution of sequences subject to recombination using parsimony. *Math. Biosci.* **98**, 185-200.

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

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

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

Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) CLUSTAL V:improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, **8**, 189-191.

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

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

Horsfall, J.G., and Barratt, R.W.(1945). An improved grading system for measuring plant diseases. (Abstr.) *Phytopathology* **35**, 655.

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

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

Hughes, F.L., Rybicki, E.P., and von Wechmar, M.B.(1992). Genome Typing of Southern African Subgroup-1 Geminiviruses. *J.Gen.Virol.* **73**, 1031-1040.

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

Ilyina, T.V. and Koonin, E.V.(1992). Conserved Sequence Motifs in the Initiator Proteins for Rolling Circle DNA Replication Encoded by Diverse Replicons from Eubacteria, Eucaryotes and Archaebacteria. *Nucleic.Acids.Res.* **20**, 3279-3285.

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

Isnard, M., Granter, M., Frutos, R., Reynaud, B., and Peterschmitt, M.(1997). The quasi-species nature of the maize streak virus DNA. Maize Streak Disease Symposium 9 – 11 September. South African Agricultural Research Council, South Africa.

Isnard, M., Granter, M., Frutos, R., Reynaud, B., and Peterschmitt, M.(1998). Quasi-species nature of the maize streak virus isolates obtained from a population used to assess maize cultivar response to infection. J. Gen. Virol. 79:3091-3099.

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

Jupin, I., Hericourt, F., Benz, B., and Gronenborn, B.(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.

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

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

Kampmann, H.H., and Hansen, O.B.(1993). Using colour image analysis for quantitative assessment of powdery mildew on cucumber. *Euphytica* **79**, 19-27.

Kim, S-K., Efron, Y., Fajemisin, J. M., and Buddenhagen, I.W.(1989). Mode of gene action for resistance in maize to maize streak virus. *Crop Sci.* **29**, 890-894.

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

Kokko, E.G., Conner, R.L., Kozub, G.C., and Lee, B.(1993). Quantification by image analysis of subcrown internode discoloration in wheat caused by common root rot. *Phytopath.* **83**:976-981.

Konate, G., and Traore, O.(1992). Reservoir hosts of maize streak virus (MSV) in the Sudan-Sahel zone: identification and spatio-temporal distribution. *Phytoprotection* **73**, 111-117. 26 ref.

Konate, G., and Traore, O.(1994). Variability of the maize streak virus (MSV) in the Sudano-Sahelian region. *Phytoprotection* **75**, 91-99.

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

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

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

Kotlizky, G., Boulton, M.I., Pitaksutheepong, C., Davies, J.W., and Epel, B.L.(2000). Intracellular and intercellular movement of maize streak geminivirus V1 and V2 proteins transiently expressed as green fluorescent protein fusions. *Virology* **274**, 32-38.

Lam, E.W.F., and La Thangue, N.B.(1994). DP and E2F proteins: coordinating transcription with cell cycle progression. Curr. Opin. Cell. Biol. **6**, 859-866.

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

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S.G., Schell, J., and Gronenborn, B.(1995a). *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 U.S.A.* **92**, 3879-3883.

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

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

Lazarowitz, S.G. and Pinder, A.J.(1989). Molecular genetics of maize streak virus. *Molecular Biology of Plant-Pathogen Interactions* 167-183.

Lazarowitz, S.G., and Beachy, R.N.(1999). Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *Plant Cell.* **11**, 535-548.

Li, X., and Green, M.R.(1996). Intramolecular inhibition of activating transcription factor-2 function by its DNA-binding domain. *Genes Devel.* **10**, 517-527.

Lindow, S.E., and Webb, R.R.(1983). Measurement of foliar plant disease using microcomputer controlled digital video image analysis. *Phytopath.* **73**, 520-524.

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

Liu, H., Boulton, M.I., Thomas, C.L., Prior, D.A.M., Oparka, K.J., and Davies, J.W.(1999a). Maize streak virus coat protein is karyophyllic and facilitates nuclear transport of viral DNA. *Mol. Plant. Microb. Interact.* **12**, 894-900.

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

Liu, L., Saunders, K., Thomas, C.L., Davies, J.W., and Stanley, J.(1999c). Bean yellow dwarf virus RepA, but not Rep, binds to maize retinoblastoma protein, and the virus tolerates mutations in the consensus binding motif. *Virology* **256**, 270-279.

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

Ludlow, J.W.(1993). Interactions between SV40large-tumor antigen and the growth suppressor proteins pRB and p53. *FASB J.* **7**, 866-871.

Mansoor, S., Khan, S.H., Bashir, A., Saeed, M., Zafar, Y., Malik, K.A., Briddon, R., Stanley, J., and Markham, P.G.(1999). Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* **259**, 190-199.

Markham,P.G., Pinner, M.S., and Boulton, M.I.(1984). *Dalbulus maidis* and *Cicadulina* species as vectors of diseases in maize. Maize Virus Diseases Newsletter *1*, 33-34.

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

Maynard Smith, J., and Smith, N.H.(1998). Detecting recombination from gene trees. *Mol. Biol. Evol,* **15**, 590-599.

McGuire, G., Wright, F., and Prentice, M.J.(1997). A graphical method for detecting recombination in phylogenetic data sets. *Mol. Biol. Evol.* **14**, 1125-1131.

Meehan, B.M., Creelan, J.L., McNulty, M.S., and Todd, D.(1997). Sequence of porcine circovirus DNA: affinities with plant circoviruses. *J.Gen.Virol.* **78**, 221-227.

Mesfin, T., Bosque-Perez, N.A., and Buddenhagen, I.W.(1992). Studies of maize streak virus isolated from grass and cereal hosts in Nigeria. *Plant Disease* **76**, 789-795.

Mesfin, T., Den Hollander, J., and Markham, P.G.(1995). Feeding activities of *Cicadulina mbila* (*Hemiptera: Cicadellidae*) on different host-plants. *Bull. Entomol. Res.* **85**, 387-396.
1995

Missich, R., Ramirez-Parra, E., and Gutierrez, C.(2000). Relationship of oligomerisation to DNA binding of *Wheat dwarf virus* RepA and Rep proteins. *Virology* **273**, 178-188.

Moran, E.(1993). Interactions of adenoviral proteins with pRB and p53. FASB **7**, 880-885.

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

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

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

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

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

Nevins, J.R.(1992). E2F-a link between the Rb tumor suppressor protein and viral oncoproteins. *Science* **258**, 424-429.

Newton, A.C.(1989). Measuring the sterol content of barley leaves infected with powdery mildew as a means of assessing partial resistance to *Erysiphe graminis* f.sp. *hordei*. *Plant Path.* **38**, 534-540.

Nielson, M.W.(1986). The leafhopper vectors of phytopathogenic viruses. Taxonomy, biology and virus transmission. *U.S. Dept. Agric. Tech. Bull.* **1382**.

Noirot, P., Bagonetti, J., and Novick, R.(1990). Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc.Natl.Acad.Sci USA* **67**, 8560-8564.

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

Nutter, F.W., Gleason, M.L., Jenco, J.H., and Christians, N.C.(1993). Assessing the accuracy, intra-scorer repeatability, and inter-scorer reliability of disease assessment systems. *Phytopath* **83**, 806-812.

Odhiambo, O.B.(1995). Occurrence of wild grass geminiviruses in Kenya and their importance as source of inoculum for maize. *PhD Thesis* 151pp.

Okoth, V.A.O, and Dabrowski, Z.T.(1987a). Population density, species composition and infectivity with maize streak virus (MSV) of Cicadulina spp. leafhoppers in some ecological zones in Nigeria. *Acta Oecologica Oecologia Applicata* **8**, 191-200.

Okoth, V.A.O., Dabrowski Z.T., Thottappilly, G.,Van Emden, H.F. (1987b). Comparative analysis of some parameters affecting maize streak virus (MSV) transmission of various Cicadulina spp. populations. *Insect Science and its Application* **8**, 295-300.

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

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

Padidam, M., Sawyer, S., and Fauquet, C.M.(1999a). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225.

Padidam, M., Beachy, R.N., and Fauquet, C.M.(1999b). A phage single-stranded DNA (ssDNA) binding protein complements ssDNA accumulation of a geminivirus and interferes with viral movement. *J Virol* **73**, 1609-1616.

Palmer, K.E. (1997). Investigations into the use of *Maize streak virus* as a gene vector. *PhD Thesis*, University of Cape Town 215 pp.

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

Palmer, K.E., Schnippenkoetter, W.H., and Rybicki, E.P.(1998). Geminivirus isolation and DNA extraction. *In*: Foster G, Taylor S (eds) Plant virology protocols: From virus isolation to transgenic resistance. Humana Press, Totawa

Parker, S.R., Shaw ,M.W., and Royle, D.J.(1995). The reliability of visual estimates of disease severity on cereal leaves. *Plant Path.* **44**, 856-864.

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

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

Perez-Martin, J., and deLorenzo, V.(1997). Clues and consequences of DNA bending in transcription. Ann. Rev. Microbiol. **51**, 593-628.

Peterschmitt, M., Quiot, J.B., and Reynaud, B.(1992). Detection of maize streak virus antigens over time in different parts of maize plants of a sensitive and a so-called tolerant cultivar by ELISA. *Annals of Applied Biology* **121**, 641-653.

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

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

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

Pooma, W. and Petty, I.T.(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.

Price, T.V., Gross, R., Ho Wey, J., and Osborne, C.F.(1993). A comparison of visual and digital image-processing methods in quantifying the severity of coffee leaf rust (*Hemileia vastatrix*). *Australian J. Exp. Agric.* **33**, 97-101.

Pringle, C.R.(1999). Virus Taxonomy - 1999. The Universal System of Virus Taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. *Arch Virol* **144**, 421-429.

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

Rekab, D., Carraro, L., and Schneider, B.(1999). Geminivirus-related extrachromosomal DNAs of the X-clade phytoplasmas share high sequence similarity. *Microbiology* **145**, 453-459.

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

Rice, J.A. (1995) Mathematical statistics and data analysis. Duxbury Press, Belmont, pp. 36-38.

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

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

Robertson, D.L., Hahn, B.H. and Sharp, P.H. (1995) Recombination in AIDS Viruses. *J. Mol. Evol.*, **40**, 249-259.

Rodier, A.(1995). Déterminisme génétique de la résistance du maïs (*Zea mays* L.) au maize streak virus (MSV). *Thèse de Doctorat*, de l'Ecole Nationale Supérieure Agronomique de Rennes, 172pp.

Rodier, A., Assi0, J., Marchand, J. -L., and Herv0, Y.(1995). Breeding maize lines for complete and partial resistance to maize streak virus (MSV). *Euphytica* **81**, 57-70.

Rojas, M.R., Noueiry, A.O., Lucas, W.J., and Gilbertson, R.L.(1998). Bean Dwarf mosaic geminivirus movement proteins recognize DNA in a form- and size-specific manner. *Cell* **95**, 105-113.

Rose, D.J.W.(1972). Times and sizes of dispersal flights by *Cicadulina* species (*Homoptera*: *Cicadellidae*), vectors of maize streak disease. *J. Anim. Ecol.* **41**, 495-506.

Rose, D.J.W.(1978). Epidemiology of maize streak disease. *Ann. Rev. Ent.* **23**, 250-282.

Rossel, H.W., and Thottappilly, G.(1985). Virus disease of important food crops in tropical Africa. IITA Publication Series, 61pp.

Rubinstein, G. and Czosnek, H.(1997). Long-term association of tomato yellow leaf curl virus with its whitefly vector *Bemisia tabaci*: effect on the insect transmission capacity, longevity and fecundity. *J.Gen.Virol.* **78**, 2683-2689.

Ruppel, R.F.(1965) A review of the genus *Cicadulina* (*Homoptera, Cicadellidae*). *Michigan St. Univ. Museum Biol.* Ser. **2**, 385-428.

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

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 On-Line* http://194.247.68.33/mppol/1998/0929rybicki

Rybicki, E.P., Briddon, R.W., Brown, J.E., Fauquet, C.M., Maxwell, D.P., Harrison, B.D., Markham, P.G., and Stanley, J.(2000). *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses.* (Fauquet, C.M., Bishop, D.H.L., Carsten, E.B., Estes, M.K., Lemon, S.M., Mayo, M.A., McGeoch, D.J., Pringle, C.R., and Wickner, R.B., Eds.) Academic Press, San Diego. 285-297.

Sah, D. N., and Fehrmann, H.(1992). Virulence patterns of geographically differing isolates of *Pyrenophora tritici-repentis* and sources of resistance in wheat. *Plant Dis.* **76**, 712-716.

Salminen, M., Carr, J.K., Burke, D.S. and McCutchan, F.E. (1995) Identification of breakpointsin intergenotypic recombinants of HIV type 1 by bootscanning. *AIDS Res. Hum. Retrovirus.* **11**, 1423-1425.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning, a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Sánchez-Campos, S., Navas-Castillo, J., Camero, R., Soria, C., Díaz, J.A., and Moriones, E.(1999). Displacement of Tomato yellow leaf curl virus (TYLCV)-Sr by TYLCV-Is in Tomato Epidemics in Spain. *Phytopathology* **89**, 1038-1043.

Displacement of Tomato Yellow Leaf Curl Virus (TYLCV)-Sr by TYLCV-Is in Tomato Epidemics in Spain. S.
Sánchez-Campos, J. Navas-Castillo, R. Camero, C. Soria, J. A. Díaz, and E. Moriones. Estación Experimental "La Mayora,"

Consejo Superior de Investigaciones Científicas, 29750 Algarrobo-Costa, Málaga, Spain. Phytopathology 89:1038-1043.
Accepted for publication 23 July 1999. Copyright 1999 The American Phytopathological Society.

Sanderfoot, A.A. and Lazarowitz, S.G.(1995). Cooperation in viral movement: the geminivirus BL1 movement protein interacts with BR1 and redirects it from the nucleus to the cell periphery. *Plant Cell* **7**, 1185-1194.

Sanderfoot, A.A. and Lazarowitz, S.G.(1996). Getting it together in plant virus movement: Cooperative interactions between bipartite geminivirus movement proteins. *Tr.Cell Biol* **6**, 353-358.

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

Sanz-Burgos, A.P. and Gutierrez, C.(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. and Stanley, J.(1999). A nanovirus-like DNA component associated with yellow vein disease of *Ageratum conyzoides*: evidence for interfamilial recombination between plant DNA viruses. *Virology* **264**, 142-152.

Saunders, K., Bedford, I.D., Briddon, R.W., Markham, P.G., Wong, S.M., and Stanley, J.(2000). A unique virus complex causes Ageratum yellow vein disease. *Proc Natl Acad Sci U S A* **97**, 6890-6895.

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

Sawyer, S. (1989) Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**, 526-538.

Schaffer, R.L., Miller, C.G., and Petty, I.T.(1995). Virus and host-specific adaptations in the BL1 and BR1 genes of bipartite geminiviruses. *Virology* **214**, 330-338.

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

Schneider, M., Jarchow, E., and Hohn, B.(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 213 pp.

Seth, M.L., Raychaudhuri, S.P., and Singh, D.V.(1972)..Bajra (pearl millet) streak: a leafhopper-borne cereal virus in India. *Plant Dis. Rep.* **56**, 424-428.

Shaw, M.W., and Royle, D.J.(1989). Estimation and validation of a function describing the rate at which *Mycoshaerella graminicola* causes yield loss in winter wheat. *Ann. Appl. Biol.* **115**, 425-442.

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

Sherwood, R.T., Berg, C.C., Hoover, M.R., and Zeiders, K.E.(1983). Illusions in visual assessment of Stegonospora leaf spot of orchardgrass. *Phytopath.* **73**, 173-177.

Siepel, A.C., Halpern, A.L., Macken, C. and Korber, B.T.M. (1995) A computer program designed to screen rapidly for HIV type 1 intersubtype recombinant sequences. *AIDS Res. Hum. Retrovirus*. **11**, 1413-1416.

Soto, P.E.A.(1978). A new vector of maize streak virus. *E. African Agric. Forest.* J. **44**, 70-71.

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

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

Stanley, J.(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.

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

Stenger, D.C., Davis, K.R., and Bisaro, D.M.(1994). Recombinant beet curly top virus genomes exhibit both parental and novel pathogenic phenotypes. *Virology* **200**, 677-685.

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

Storey, H.H.(1924). The transmission of a new plant virus disease by insects. *Nature* **144**, 245-245.

Storey, H.H.(1928). Transmission of maize streak disease. *Ann. Appl. Biol.* **15**, 1-25.

Storey, H.H.(1938). Investigation of the mechanism of the transmission of plant viruses by insect vectors. II. The part played by puncture in transmission . *Proc. Royal Soc.* **125**, 455-477.

Suarez-Lopez, P., Martinnez-Salas, E., Hernandez, P., and Gutierrez, C.(1995). Bent DNA in the large intergenic region of wheat dwarf geminivirus. *Virology* **208**, 303-311.

Sunter, G., Buck, K.W., and Coutts, R.H.(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. *Nucleic.Acids.Res.* **13**, 4645-4659.

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

Takahashi, K., Iwasa, Y., Hijikata, M., and Mishiro, S.(2000). Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. *Arch Virol.* **145**, 979-993.

Tang, C.Y., and Bjarnason, M.S.(1993). Two approaches for the development of maize germplasm resistant to maize streak virus. *Maydica.* **38**, 301-307.

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

Torres-Pacheco, I., Garzon-Tiznado, J.A., Herrera-Estrella, L., and Rivera-Bustamante, R.F.(1993). Complete nucleotide sequence of pepper huasteco virus: analysis and comparison with bipartite geminiviruses. *J.Gen.Virol.* **74**, 2225-2231.

Umaharan, P., Padidam, M., and Phelps, R.H.(1998). Distribution and diversity of geminiviruses in Trinidad and Tobago. *Phytopath.* **88**, 1262-1268.

Van Rensburg, G.D.J.(1982). The ecology and culture of leaf hoppers. *Proceedings of the Fifth South African Maize Breeding Symposium* (Editor J. G. Du Plessis), 100-103.

Van Rensburg, G.D.J., and Giliomee, J.H.(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., and Giliomee, J.H.(1991). Chemical prevention of maize streak virus transmission by the maize leafhopper, *Cicadulina mbila* (*Homoptera: Cicadellidae*).

Phytophylactica **23**, 141-144.

Van Rensburg, G.D.J., Giliomee, J.H., and Pringle K.L.(1991). Resistance of South African maize hybrids to maize streak virus. *S. Afr. J. Plant Soil* **8**, 38-42.

Van Rensburg, J.B.J., and Walters, M.C.(1978). The efficacy of systemic insecticides applied to the soil for the control of *Cicadulina mbila* (Naude) (*Hem: Cicadellidae*), the vector of maize streak disease, and the maize stalk borer *Busseola fusca* (Fuller) (*Lep: Noctuidae*). *Phytophylactica* **10**, 49-52.

Van Rensburg, J.B.J.(1988). Efficacy of cloethocarb for control of *Busseola fusca*, *Cicadulina spp.* and nematodes in maize. *App. Plant Sci.* **2**, 63-67.

Van Rensburg, J.B.J.(1997). Maize streak virus resistance breeding. Maize Streak Disease Symposium 9 – 11 September. South African Agricultural Research Council, South Africa.

Voesden, K.(1993). Interactions of human papillomavirus transforming proteins with the products of tumour suppressor genes. *FASB J.* **7**, 872-879.

Walkey, D.G.A., Alhubaishi, A.A., and Webb, M.J.W.(1990). Plant virus diseases in the Yemen Arab Republic. *Trop. Pest Manage.* **36**, 195-206.

Ward, B.M., Medville, R., Lazarowitz, S.G., and Turgeon, R.(1997). The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. *J.Virol.* **71**, 3726-3733.

Ward, B.M. and Lazarowitz, S.G.(1999). Nuclear export in plants. Use of geminivirus movement proteins for a cell-based export assay. *Plant Cell* **11**, 1267-1276.

Webb, R.R., and Lindow, S.E.(1981). Evaluation of *Asochyta pteridium* as a potential biological control agent of bracken fern. (Abstr.) *Phytopath.* **71**:911.

Weiller, G.F. (1998) Phylogenetic profiles: A graphical method for detecting recombinations in homologous sequences. *Mol. Evol. System.*, **15**, 326-335.

Welz, H.G., Schechert, A., Pernet, A., Pixley, K.V., and Geiger, H.H.(1998). A gene for resistance to the maize streak virus in the African CIMMYT maize inbred line CML202. *Mol. Breed.* **4**, 147-154.

Willment, J. A. (1999) Investigation of the molecular determinants of *Maize streak virus* replication. *PhD Thesis*, University of Cape Town 207 pp.

Woolston, C.J., Reynolds, H.V., Stacey, N.J., and Mullineaux, P.M.(1989). Replication of wheat dwarf virus DNA in protoplasts and analysis of coat protein mutants in protoplasts and plants. *Nucleic.Acids.Res.* **17**, 6029-6041.

Wright, E.A., Heckel, T., Groenendijk, J., Davies, J.W., and Boulton, M.I.(1997). Splicing features in maize streak virus virion- and complementary-sense gene expression. *Plant J.* **12**, 1285-1297.

Xie, Q., Suarez-Lopez, P., and Gutierrez, C.(1995). Identification and analysis of a retinoblastoma binding motif in the replication protein of a plant DNA virus: requirement for efficient viral DNA replication. *EMBO J.* **14**, 4073-4082.

Xie, Q., Sanz-Burgos, A.P., Hannon, G.J., and Gutierrez, C.(1996). Plant cells contain a novel member of the retinoblastoma family of growth regulatory proteins. *EMBO J.* **15**, 4900-4908.

Xie, Q., Sanz-Burgos, A.P., Guo, H., Garcia, J.A., and Gutierrez, C.(1999). GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein [In Process Citation]. *Plant Mol Biol* **39**, 647-656.

Zhan, X., Richardson, K.A., Haley, A., and Morris, B.A.(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.

Zhou, X., Liu, Y., Calvert, L., Munoz, C., Otim-Nape, G.W., Robinson, D.J., and Harrison, B.D.(1997). Evidence that DNA-A of a geminivirus associated with severe cassava mosaic disease in Uganda has arisen by interspecific recombination. *J.Gen.Virol.* **78**, 2101-2111.

Zhou, X., Liu, Y., Robinson, D.J., and Harrison, B.D.(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.