EXPERIMENTAL INVESTIGATIONS OF MASTREVIRUS MOLECULAR BIOLOGY AND EVOLUTION

Eric van der Walt

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Molecular and Cell Biology.

University of Cape Town

2008

CERTIFICATION BY SUPERVISORS

In terms of paragraph GP9 of the regulations for the degree of Doctor of Philosophy at the University of Cape Town, we certify that we approve of the inclusion in this thesis of material already published, or submitted for publication by candidate Eric van der Walt.

Edward P. Rybicki, PhD
Associate Professor in Virology
Department of Molecular and Cell Biology
University of Cape Town
February 2008

Darren P. Martin, PhD
Institute of Infectious Diseases and Molecular Medicine
University of Cape Town
February 2008

UNIVERSITY OF CAPE TOWN

ABSTRACT

EXPERIMENTAL INVESTIGATIONS OF MASTREVIRUS MOLECULAR BIOLOGY AND EVOLUTION

Eric van der Walt

Supervisors: Prof. E.P. Rybicki and Dr. D.P. Martin

Maize streak virus (MSV) is a small, single stranded (ss) DNA geminivirus which causes significant economic losses through the destruction of maize crops throughout sub-Saharan Africa. This dissertation describes three major sets of experiments, all of which involved the construction and use of various reciprocal chimaeric MSV constructs. First, chimaeric viruses were used in genetic complementation-type experiments to investigate the biological significance of interactions between the two virion-sense open reading frames (ORFs) of MSV, their products, and the rest of the genome. Six chimaeric MSV constructs were made by reciprocally exchanging the ORFs encoding movement protein (MP) and coat protein (CP) individually, and in pairs, between MSV-Kom and MSV-Set, which share just 78% overall nucleotide identity. Analysis of symptomatology and infection efficiency of chimaeras and wild-type parental viruses revealed evidence of functionally relevant specific interactions between MSV MP and CP.

In the second set of experiments the chimaeric viruses described above were ~~passaged~~ maintained through twelve ~~generations~~ passages in~~of~~ sweetcorn over a period of one year, with the aim of investigating the evolution of the viruses during that time. In addition, an MSV-infected sugarcane plant was maintained for four years before cloning and sequencing full length virus progeny. Consistent with an hypothesis that polymerase fidelity influences evolution rate, RNA viruses have

generally been shown to evolve at rates of ~10^{-4} substitutions per site per year (subs/site/year), while dsDNA viruses seem to display substitution rates of ~10^{-8} subs/site/year. In contrast, substitution rates closer to those of RNA viruses have been reported for various small ssDNA viruses. While no signs of strong positive selection were observed, high mutation rates in the order of 10^{-4} subs/site/year and biased mutation frequencies were recorded. These biased mutation frequencies indicate that damage to viral genomic ssDNA may contribute significantly to high mutation rates in MSV.

The third set of experiments described here was aimed at investigating the mechanisms and adaptive power of recombination in MSV evolution. Once again, reciprocal MSV chimaeras comprising precise exchanges of ORFs encoding MP and CP, and displaying reduced fitness, were employed in these experiments. MSV-resistant maize was co-infected with both chimaeras and natural selection was used to preclude parental chimaeric viruses and to enrich fitter recombinant progeny viruses arising during co-infection. Full genome sequences of twelve recombinant viruses were obtained, and the fitness of selected virus clones was assessed and compared with that of both wild-type parents and original chimaeras. These experiments demonstrated that recombination is a powerful mechanism for rapidly reverting deleterious mutations and/or combining advantageous alleles during geminivirus infection. In addition to full genome sequences, a number of subgenomic clones were sequenced. Apart from deletions, many of these subgenomic clones displayed both direct and inverse repeats, while some contained host DNA sequences. Together, the full- and subgenome-length clones revealed recombination hot-spots that suggest particular mechanisms that may be responsible for driving recombination in MSV, and these are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to humbly and most sincerely thank my extraordinarily long-suffering supervisors, colleagues, family, friends, and lovers for their undeserved generosity, forbearance and kindness throughout the painfully slow process of completing the work presented here. Polite words cannot describe the magnitude of the personal tragedy represented by this dissertation.

It is customary to furnish an exhaustive list of creditors here, but if I begin to name people then I will inevitably offend someone by omitting them or by misrepresenting their contribution. Those who deserve credit will know, and it must therefore suffice for me to say:

My wonderful parents and my inspirational brother have each inspired and supported me in their unique ways throughout my life. I remain heart-broken at the thought that my magnificent father did not survive to see me lay this demon to rest.

I owe more than I could ever repay to my true friends, of all shapes, sizes, colours and forms, who never admitted to losing faith in me, nor openly despised me for dragging this thing out.

Finally, I wish to thank my two supervisors and many colleagues for their invaluable contributions to both theoretical and practical aspects of the work presented here.

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Viruses of some form or another have almost certainly been ubiquitous since very early in the evolution of life, and they are an abundant and highly diverse collection of organisms (reviewed by Koonin *et al.*, 2006). Viruses are the only organisms known to have anything other than double stranded DNA as genetic material, and they seem to represent every conceivable alternative. This diversity is similarly reflected in the variety and number of organisms that viruses infect: neither plants, animals, fungi, nor bacteria are spared. However, viruses by definition all share one trait with respect to their status in ecology – they are obligate parasites.

Although many viruses do not have noticeably destructive effects on their hosts (there is generally an obvious selective advantage in allowing the host to remain healthy), viruses are commonly notorious for their pathogenicity. Not only do they cause disease in man, but also in his livestock and crops. The extent of the impacts of virus diseases on crop yields have yet to gain general public recognition, but some might argue that the starvation and economic losses caused by plant viruses places them on par with any other threat to human contentment in the developing world (Varma & Malathi, 2003; Christou & Twyman, 2004; Bokanga *et al.*, 2006). Currently, plant viruses are controlled almost exclusively by poisoning their vectors through the application of insecticides (mostly organophosphates), which has negative effects on the environment and on human health. Plant viruses are socio-economically destructive, and as such deserve attention from the global scientific and technological research community, especially in the current era of spectacular advances in molecular biology.

Apart from helping to understand virus diseases and thereby suggesting ways of reducing their destructive effects, there are other perspectives from which virology yields useful information. Because of their relative simplicity, the ease with which

they are manipulated, their brief life-cycles, and their high degree of dependence on host cellular processes, viruses are exceptionally useful tools for studying a wide range of phenomena of general biological interest. Over the course of several decades viruses have been indispensable for investigating fundamental biochemical processes such as gene expression, assembly of macromolecular structures, replication, and evolution. In this regard geminiviruses are perhaps one of the most appropriate of the plant-infecting viruses: their genomes are encoded in DNA, and they are amongst the smallest known viruses.

This dissertation describes experimental investigations into the molecular biology and evolution of one particular geminivirus: maize streak virus (MSV).

The *Geminiviridae*

Although some are mechanically transmissible in laboratories, geminiviruses are naturally transmitted by homopteran insect vectors and infect both dicotyledonous and monocotyledonous plants (Stanley *et al.*, 2005). The family name *Geminiviridae* originates from the distinctive, paired-icosahedral capsid morphology – "gemini" means "the twins" in Latin (Figure 1-1). A single geminate virion (ca. 18x30nm; S_{20w} = 70) consists of two incomplete, joined T=1 icosahedra formed by a total of 22 pentameric coat protein (CP) capsomers (Zhang *et al.*, 2001). No other protein is known to form part of the virus particle, and virions apparently contain neither carbohydrates nor lipids.

Each geminate virion is thought to contain a single, covalently closed, circular, single-stranded DNA (ccc ssDNA) molecule (Francki *et al.*, 1980), which may be between 2.5kb and 3.0kb, depending on the strain of virus. Some geminivirus genomes comprise two separate ccc ssDNAs, which are both required for the infection of a plant (Stanley, 1983), and are individually referred to as DNA A and DNA B. In the case of such bipartite geminiviruses, the DNA A component encodes replication-

associated functions and the CP, while DNA B encodes movement functions including nuclear localisation. Although DNA A and DNA B encode different functions, both DNAs replicate via double-stranded (ds) DNA intermediates by the same mechanisms and are structured in analogous ways.

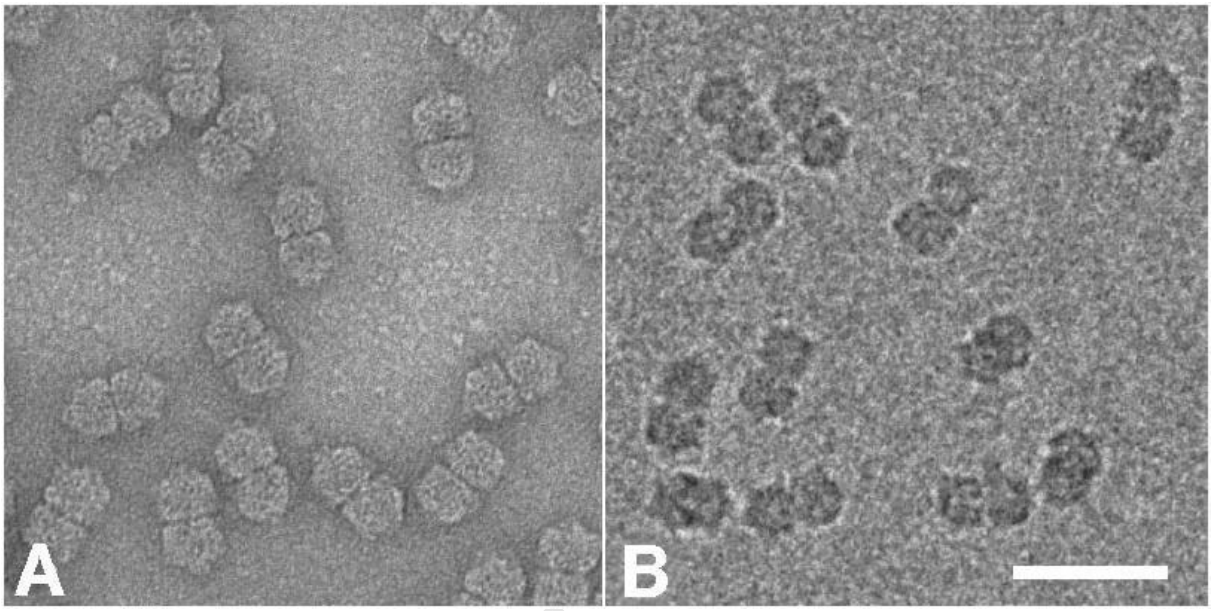


Figure 1-1. Transmission electron micrographs of MSV particles (taken from Zhang *et al.*, 2001).

A Negative staining with 1% uranyl acetate; **B** vitrified in a thin layer of 0.1 M sodium acetate buffer, pH 4.8, recorded on a $1K^2$ slow-scan, CCD camera under low-dose conditions (~20 e^- /$Å^2$) and 0° tilt. Scale bar = 500 Å.

In monopartite and bipartite geminivirus replicative ds DNA molecules, genes are encoded in both strands of DNA, and diverge from an intergenic region which contains the virion-sense origin of replication (Figure 1-2). Transcription is thus bidirectional, with transcripts originating in the intergenic region and converging on the diametrically opposite side of the circular genome (Morris-Krsinich *et al.*, 1985). Conventionally, geminivirus genomes are represented with the virion sense genes expressed in a clockwise direction – or to the right of the intergenic region – and *vice-versa* for the complementary sense genes (Figure 1-2). As a result, open reading

frames (ORFs) have been named according to two systems (Davies & Stanley, 1989), leading to some confusion in the literature.

One system distinguishes between genes firstly as being encoded either in the virion or complementary sense (they are assigned either a V or C, respectively), while the other scheme denotes genes as R (rightward, or virion sense) or L (leftward, or complementary sense). For bipartite genomes, both systems name genes according to whether they are on the DNA A or DNA B component. Both systems designate individual ORFs by number, although no strict convention for the order in which genes are numbered has been adhered to in either system. For the sake of clarity, I will use the system which distinguishes between ORFs as being encoded either in the virion or complementary sense (designated V or C, respectively), and I will use functional names for gene products where appropriate (Stanley *et al.*, 2005).

Geminivirus replication

All geminiviruses apparently share a common replication strategy (reviewed by Gutierrez, 1999), and certain features of the DNA molecule that are involved in replication are conserved accordingly. Geminiviruses are known to employ rolling circle replication (RCR; Saunders *et al.*, 1991; Saunders *et al.*, 1992; Heyraud *et al.*, 1993a; Heyraud *et al.*, 1993b; Laufs *et al.*, 1995b), and the majority of research into geminivirus replication has focussed on RCR. However, there is evidence that "recombination-dependent replication" (RDR) mechanisms play an important role in geminivirus replication (Saunders *et al.*, 1991; Jeske *et al.*, 2001; Preiss & Jeske, 2003). Geminiviruses are also known to interfere with the regulation of host cell cycles in order to induce expression of host replication machinery in quiescent cells (Gutierrez *et al.*, 2004; reviewed by Hanley-Bowdoin *et al.*, 2004).

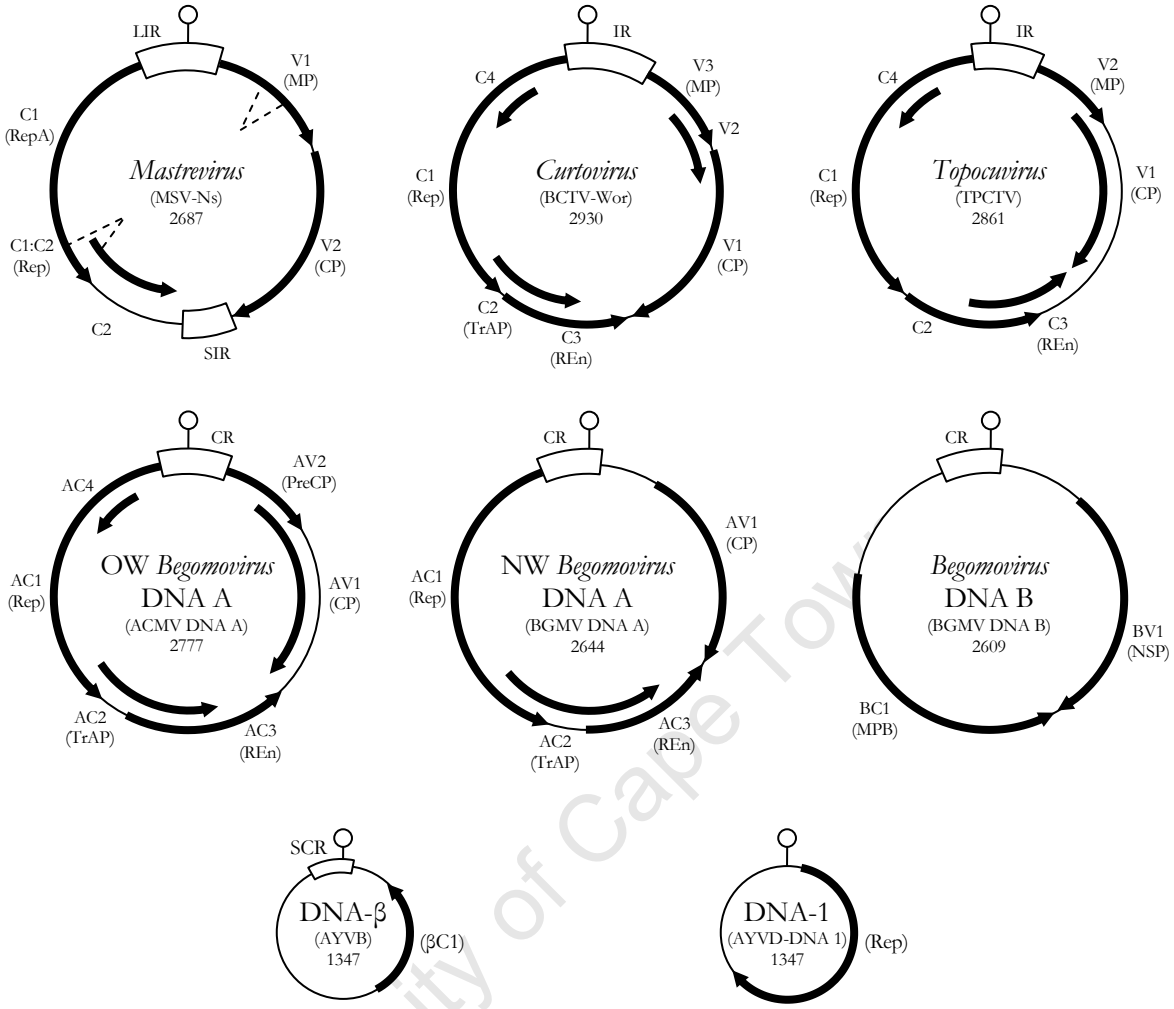


Figure 1-2. Genomes of selected representative geminiviruses and associated satellite viruses.

Genomes of representatives of each geminivirus genus are shown, along with the relatively recently described satellite viruses found in association with many begomoviruses. Genus names are italicised (OW = old world; NW = new world), with virus names in parenthesis and genome sizes provided below. The virion-sense origins of replication are represented with a stem-loop symbol. ORFs are depicted by arrows in the direction of transcription, with dashed lines indicating introns. Non-coding conserved genomic regions are shown with open boxes. ORFs are named according to the system described in the text, while functional names of proteins (in parenthesis below the ORF name) and names of non-coding conserved genomic regions are abbreviated as follows: MP, movement protein; CP, coat protein; SIR, short intergenic region; Rep, replication associated protein; LIR, long intergenic region; IR, intergenic region; REn, replication enhancer protein; TrAP, transcription activator protein; CR, common region; PreCP, pre-coat protein; NSP, nuclear shuttle protein; MPB, movement protein B; SCR, satellite common region.

According to the canonical geminivirus RCR model, virus replication proceeds through three phases, all taking place within the plant cell nucleus. First, the ccc ssDNA genomic form is made double-stranded by host enzymes, presumably acting in a capacity analogous to the "DNA repair" functions described in a wide variety of prokaryotes and eukaryotes. The host enzymes produce a ccc dsDNA by RNA-primed synthesis of the complementary strand. More specifically, Donson *et al.* (1984) and Howell (1984) provided evidence that a short DNA primer having several ribonucleotides at its 5'-end is associated with the ssDNA in the virion in mastreviruses. This putative DNA synthesis primer is about 80 bases long, and is bound to a region in the mastrevirus "short intergenic region" (SIR). Interestingly, equivalent primers have not been found to be encapsidated in the other geminivirus genera (Stanley & Townsend, 1985), and it is likely that second-strand synthesis is initiated within the LIR in begomoviruses (Saunders *et al.*, 1992), and probably also in curtoviruses. The ccc dsDNA is known as the "replicative form" (RF), which is thought to take the form of "minichromosomes" (i.e. is assembled into nucleosomes; Abouzid *et al.*, 1988; Pilartz & Jeske, 1992; Pilartz & Jeske, 2003) and acts as a template for viral transcription and further replication.

The second stage of replication results in an increase in the number of RF DNAs, and is dependent on the viral Rep protein. Rep recognises sequences in the viral dsDNA, and nicks the virion sense strand at a specific position in the LIR, in a sequence (TAATATTA↓AC; ↓ indicates the nick site) that is almost invariantly conserved in all geminivirus genomes sequenced to date (Stenger *et al.*, 1991; Heyraud *et al.*, 1993b; Heyraud-Nitschke *et al.*, 1995; Stanley, 1995; Laufs *et al.*, 1995b). Rep then binds covalently to the newly formed 5'-end, via a phosphotyrosine linkage (Laufs *et al.*, 1995a). This allows the synthesis of the virion sense strand by host replication proteins, with the 3'-end acting as a DNA primer. As the nascent virion sense strand is extended, the parental strand is displaced from the intact complementary strand template. When synthesis has progressed

completely around the circular complementary strand, the conserved origin of replication is regenerated, and Rep can once again bind the dsDNA, nick the virion sense strand, and attach to the nascent 5'-end. In this way Rep presumably acts as a terminase by releasing the newly synthesised virion sense strand, while it simultaneously ligates the ends of the new virion-sense ssDNA to produce a circular form by its closing activity (Heyraud-Nitschke *et al.*, 1995; Laufs *et al.*, 1995b).

This process represents a type of rolling-circle replication, and may ultimately yield one of two products. When viral DNA is at relatively low concentrations, the ccc ssDNA is generally used as a template for the synthesis of complementary strand DNA, yielding additional ccc dsDNA RFs. Alternatively, later in the viral replication cycle the ssDNA may accumulate, becoming encapsidated in virions. The production of virion sense ccc ssDNA from the RF represents the third stage in geminivirus replication.

The family Geminiviridae *comprises four genera*

The family *Geminiviridae* is divided into four genera according to host range, type of insect vector, genome organisation, and phylogenetic analyses. Each genus is briefly described below.

The Mastrevirus *genus*

MSV is the type species of the genus *Mastrevirus*, which includes at least eleven other characterised species (Stanley *et al.*, 2005). Most characterised mastreviruses have host ranges that are restricted to members of the family *Poaceae* (grasses), which are monocotyledonous, and the genus has therefore traditionally been considered to comprise mostly monocot-infecting viruses. However, two dicot-infecting mastrevirus species were identified some time ago – *Tobacco yellow dwarf virus* (Morris *et al.*, 1992), and *Bean yellow dwarf virus* (Liu *et al.*, 1997b; Halley-Stott *et al.*, 2007). More recently, we have described *Chickpea chlorotic dwarf virus* (CpCDV) and *Chickpea*

chlorotic dwarf Pakistan virus (Nahid *et al.*, 2008), thus bringing the total number of dicot-infecting mastrevirus species to four, which represents a third of the total species assigned to the genus.

Most mastreviruses are naturally transmitted by single species of leafhoppers (Homoptera: Cicadellidae) in a persistent (circulative, non-propagative) manner. Mechanical inoculation of plants has been described (Redinbaugh, 2003) but mastreviruses are more commonly transmitted in the laboratory by *Agrobacterium*-mediated transfer (hereafter referred to as "agroinoculation"; Grimsley *et al.*, 1987) using recombinant DNA methods. Although more labour intensive, the agroinoculation technique is considerably more efficient than mechanical inoculation, and was used extensively in the experiments presented in this dissertation.

Mastreviruses possess genomes consisting of a single component between 2.6 kb and 2.8 kb in size. The presence of a small (~80 nt.) complementary-sense, primer-like molecule bound in the small intergenic region has been detected in MSV, and in at least four other species in this genus (Gutierrez, 1999). This molecule is thought to prime complementary-sense synthesis, and has been notably absent from viruses that have been investigated in the other geminivirus genera. Mastrevirus genomes typically comprise four functional ORFs; two are encoded on the virion sense strand (named V1 and V2; Figure 1-2), and two on the complementary strand (C1 and C2). The C2 ORF lacks a methionine start codon in most species (Palmer & Rybicki, 1998), and in several of these viruses it has been shown that translation of the C2 ORF is dependent on splicing of the C1 and C2 transcripts (Schalk *et al.*, 1989; Mullineaux *et al.*, 1990) .

The Curtovirus *genus*

The type species for this genus is *Beet curly top virus* (BCTV; Stanley *et al.*, 2005), which was the sole member of the genus until relatively recently, and there are currently

only four species assigned to the genus (ICTVdB Management, 2006a). Curtoviruses seem to have broader host ranges than other geminiviruses with, for example, BCTV capable of infecting over 80 species in 21 plant families (ICTVdB Management, 2006b). These host plants are exclusively dicotyledonous. Like mastreviruses, curtoviruses are persistently (circulative, non-propagative) transmitted by leafhoppers. It is possible to transmit BCTV by mechanical inoculation with very low efficiency (Stanley *et al.*, 1986; Stenger *et al.*, 1990) and some curtoviruses have been transmitted by agroinoculation (Stanley *et al.*, 1992).

The genomes of these viruses are monopartite and 2.7 - 3.0 kb in size. BCTV has seven genes, four encoded in the complementary sense, and three in the virion sense (Figure 1-2). While the curtovirus complementary-sense ORFs are organised in a similar way to those of monopartite begomoviruses, the arrangement of the virion-sense half of the genome is more reminiscent of that of mastreviruses. In addition to the ORFs encoding MP (V3) and CP (V1), curtoviruses have an additional ORF (V2), which seems to be involved in regulating the relative levels of ssDNA vs dsDNA during curtovirus infection (Stanley *et al.*, 1992; Hormuzdi & Bisaro, 1993). In the complementary-sense portion of the genome, ORF C1 encodes Rep; C2 encodes a protein that is a homologue of the begomovirus transcription activator protein (TrAP) and enhances pathogenicity in some hosts (Hormuzdi & Bisaro, 1995); C3 encodes a homologue of the begomovirus replication enhancer protein (REn); and C4 is a symptom determinant that may be involved in cell-cycle control (Stanley & Latham, 1992).

The Begomovirus *genus*

The *Begomovirus* type species is *Bean golden yellow mosaic virus* (BGMV; Stanley *et al.*, 2005). There are currently 117 additional species within the genus (ICTVdB Management, 2006c). The host ranges of individual begomovirus species are generally narrower than those of curtoviruses, and are restricted to dicotyledonous

plants. *Bemisia tabaci* (Genn.)(Homoptera: Aleyrodidae), a single species of whitefly, is responsible for the natural transmission of all known begomoviruses (Fauquet *et al.*, 2003). It has been shown that at least some begomoviruses are capable of both transovarial (Ghanim *et al.*, 1998) as well as sexual transmission between whiteflies (Ghanim & Czosnek, 2000). In the laboratory, some begomoviruses have been successfully transmitted to plants by mechanical inoculation (Elmer *et al.*, 1988; Etessami *et al.*, 1988; Gilbertson *et al.*, 1993; Ascencio-Ibanez & Settlage, 2007) and a number of species have been transmitted using recombinant DNA techniques via agroinoculation (Ferreira *et al.*, 2008) or microprojectile bombardment (Lapidot *et al.*, 2007).

Many of the described begomoviruses have bipartite genomes like that of BGMV, with each component being 2.5 - 2.8 kb (Figure 1-2). However, at present most are more similar to tomato yellow leaf curl virus (TYLCV; Navot *et al.*, 1991) in having monopartite genomes of approximately 2.8 kb which encode two genes in the virion sense and four genes in the complementary sense (Stanley *et al.*, 2005). The components of bipartite genomes are designated DNA A, and DNA B. Generally, DNA A contains four ORFs – one in the virion sense and three in the complementary sense – which encode all the functions required for replication, as well as the coat protein. DNA B generally encodes two genes – one in each strand – which are required for cell to cell and systemic movement of the virus (Etessami *et al.*, 1988; Brough *et al.*, 1988; Noueiry *et al.*, 1994). Thus both genome components of bipartite begomoviruses are usually required for infectivity, although exceptions have been reported in which DNA-As are independently capable of establishing systemic infections (Saunders *et al.*, 2002; Galvao *et al.*, 2003). Within each species, the A and B components share an almost identical common region of 200-250 bp, which encompasses the intergenic region including the conserved TAATATTAC sequence. This common region is distinct between different viruses.

It has, relatively recently, been discovered that many monopartite begomoviruses require ssDNA satellite viruses – called DNA-β – for full expression of disease symptoms (Saunders *et al.*, 2000; Briddon *et al.*, 2001; Zhou *et al.*, 2003; Saunders *et al.*, 2003; Jose & Usha, 2003). A second type of ssDNA satellite virus called DNA-1 (Mansoor *et al.*, 1999; Saunders & Stanley, 1999) has also been shown to be associated with most or all begomovirus—DNA-β disease complexes (Briddon *et al.*, 2004), although DNA-1 satellites do not appear to affect virulence or pathogenicity. Both of these satellite viruses (reviewed by Stanley, 2004; Briddon & Stanley, 2006) are approximately half the size of a typical full-length begomovirus genome component (~1350 nt.; see Figure 1-2) and are thought to be encapsidated in half-geminate (single-icosahedron) forms of the helper begomovirus virions and to be transmitted via the relevant whitefly vectors (Briddon *et al.*, 2004).

DNA-1 satellite viruses encode nanovirus-like Reps (Figure 1-2), contain a conserved nanovirus-like replication origin sequence (TAGTATTAC), and are capable of autonomous replication in plant host cells, but require helper virus functions for movement and transmission (Stanley, 2004). Although DNA-1 satellite sequences are most closely related to nanoviruses, they are much larger than a typical ~1100 nt nanovirus genome component, and are thought to have incorporated additional "stuffer" sequences to enable efficient encapsidation by begomovirus CP (Stanley, 2004). DNA-1 satellite viruses are therefore thought to have originated through capture of an autonomously replicating nanovirus genome component by a begomovirus during a mixed infection.

In contrast to DNA-1, DNA-β satellites contain a conserved geminivirus replication origin sequence (TAATATTAC), do not encode a Rep, and are trans-replicated by the helper begomovirus Rep. Apart from the conserved stem-loop sequence at the origin of replication, DNA-β satellites do not share significant sequence homology with their helper viruses. While DNA-β genomes typically contain multiple small

ORFs in both directions, only one – encoding the ~13.8 kDa βC1 protein – has been shown to be expressed. Consistent with the important role of DNA-β satellites in the aetiology of many begomovirus diseases (Saunders *et al.*, 2000; Briddon *et al.*, 2001), βC1 has been shown to be a pathogenicity determinant (Saeed *et al.*, 2005). The protein is thought to facilitate virus movement (Saeed *et al.*, 2007) and the accumulation of viral DNA (Saunders *et al.*, 2000), is capable of binding DNA, and is a suppressor of post-translational gene silencing (Cui *et al.*, 2005). Over 260 full-length DNA-β sequences are now available in online databases, necessitating a recent effort to develop a standardised system for nomenclature and classification of these satellite viruses (Briddon *et al.*, 2008).

The Topocuvirus *genus*

Tomato pseudo-curly top virus (TPCTV) is the type species – and currently the only member – of the *Topocuvirus* genus (ICTVdB Management, 2006d). Unlike any of the viruses in the other geminivirus genera, TPCTV is transmitted by a species of treehopper (Homoptera: Membracidae), *Micrutalis malleifera* (Fowler) (Simons & Coe, 1958). The TPCTV genome (sequenced by Briddon *et al.*, 1996) is arranged similarly to those of the curtoviruses, monopartite begomoviruses and the A-components of bipartite begomoviruses (Figure 1-2).

Maize streak virus

The most economically significant disease caused by MSV is known as maize streak disease (MSD), and was first recorded in South Africa in 1901, by Claude Fuller (1901), the Government Entomologist. However, Fuller mistakenly ascribed MSD to a soil deficiency, and it was more than two decades later that Story (1924) identified MSV as the causal agent – and leafhoppers in the genus *Cicadulina* – as the obligate vectors – of MSD.

The disease occurs across sub-Saharan Africa and on the nearby islands in the Indian Ocean. At an early stage, it was known that a variety of distinct serotypes of "African streak viruses" exist that infect wild grasses, sorghums, millets, barley, wheat, oats, maize, and sugarcane (Bock *et al.*, 1974). Of these, MSV has the greatest economic importance, placing a major constraint on the production of maize and other cereal crops across Africa (Bosque-Perez, 2000; Varma & Malathi, 2003; Bokanga *et al.*, 2006). Although the host range of MSV is confined to the *Poaceae*, the virus is capable of infecting over 100 species within this family, including many cultivated cereals (Damsteegt, 1983). Among these species, individual strains of MSV may have distinct host ranges, which can include differences in the maize cultivars they are capable of infecting (Martin *et al.*, 2001).

Similarly, within a given host plant species, different strains of MSV may produce symptoms that vary over a wide range of severity, from barely visible to severe chlorosis and stunting. Typically, between 3 and 7 days after infection with a virulent strain of MSV, chlorosis manifests as almost circular, pale spots 0.5-2mm in diameter in the lowest exposed portions of the youngest leaves (Bock, 1974). When the disease is fully developed, the leaves display narrow veinal streaks ranging from a few millimetres to several centimetres in length, and varying in width from 0.5 to 1 mm. Depending on the strain of MSV, and the stage of the infection, diseased plants may be sterile and/or severely stunted, resulting in reduced or absent cobs.

The mechanisms by which geminiviruses replicate have been studied and reviewed in some depth (Gutierrez, 1999; Gutierrez *et al.*, 2004; Hanley-Bowdoin *et al.*, 2004) and the general principles of geminivirus replication (described briefly above) would seem to apply to MSV. However, more recently there has been renewed interest in the potential role of recombination (Saunders *et al.*, 1991; Jeske *et al.*, 2001; Preiss & Jeske, 2003; Jovel *et al.*, 2007) in geminivirus replication which – together with numerous reports of various forms of subgenome-length geminivirus DNAs

(reviewed by Patil & Dasgupta, 2006) – suggests that geminivirus replication is somewhat less tidy than a simplistic rolling-circle model might suggest.

Comparatively less progress has been made in understanding the determinants of vector specificity, host range, and symptom severity, especially with respect to mastreviruses. The current status of mastrevirus molecular biology has been most recently comprehensively and authoritatively reviewed by Boulton (2002), Gutierrez (2002), and others but some of the relevant literature is reviewed and summarised below for the sake of completeness.

Genes and gene products of the MSV virion sense strand

The virion sense of the MSV genome gives rise to two overlapping, co-terminal transcripts: a major 0.9 kb transcript and a less abundant (~10% of total V-sense transcripts) 1.05 kb mRNA (Morris-Krsinich *et al.*, 1985; Wright *et al.*, 1997). Both transcripts comprise two ORFs – V1 (306 nt.) and V2 (732 nt.) – and both are apparently polyadenylated. However, the 0.9 kb transcript starts just 1 nt. before the V1 start codon, making it unlikely to serve as a template for translation of the V1 ORF (Wright *et al.*, 1997).

Whereas analysis of the first complete MSV DNA sequence showed that ORF V2 encodes the 27 kDa viral capsid or coat protein (CP; Mullineaux *et al.*, 1984), the V1 ORF encodes a 10.9 kDa protein which has been more difficult to characterise functionally. This 10.9 kDa protein was produced as a product of *in vitro* translation by Mullineaux *et al.* (1988), and the addition of an in-frame translation stop codon in the V1 ORF prevented infection of maize plants via agroinoculation. This suggested that the V1 ORF produces a functional protein product that is essential for MSV infection. The V1 ORF and its protein product have since been studied in more detail, and its product has been identified as a type of plant virus movement protein (MP; reviewed by Boulton, 2002, and below), although MSV MP seems to facilitate

nuclear shuttling in addition to movement across cell walls. Similarly, although the most obvious function of the CP remains encapsidation of viral DNA, CP has been shown to be multifunctional: CP is a determinant of vector specificity, is required for the accumulation of ss virus DNA, and participates in both nuclear shuttling and systemic virus movement (reviewed by Boulton, 2002, and below).

Wright *et al.* (1997) showed that the larger virion sense transcript is occasionally spliced, with the removal of a 76-nucleotide intron (from nt.81 to nt.158). Since this splicing event disrupts the V1 reading frame and removes a large part of the coding sequence, it almost certainly blocks the production of MP. Interestingly, an upstream C→T transition at position 40 (i.e. in the exon) dramatically increased splicing of the intron, with the result that infection symptoms were severely attenuated. Wright *et al.* (1997) identified putative analogous intron sequences in five additional mastrevirus genomes (*Panicum* streak virus, PanSV; *Digitaria* streak virus, DSV; sugarcane streak virus, SSV; wheat dwarf virus, WDV; and tobacco yellow dwarf virus, TYDV), but experimentally verified the predicted intron in only one of them (DSV). Although the role that transcript splicing plays in MSV biology has not been demonstrated experimentally, it is likely to contribute to regulating (quantitatively, spatially, and/or temporally) the differential expression of the transcriptionally linked V1 and V2 ORFs.

V1 - movement protein

The V1 ORF in MSV (Figure 1-2) includes bases 2689 (AUG start codon) to 302 (UAA termination codon), with a likely TATA promoter box at position 2659 (Lazarowitz, 1987). A putative polyadenylation signal (consensus with G/AATAA) was identified at position 300.

Boulton *et al.* (1989) confirmed that the V1 product is necessary for symptomatic infection of maize by deletion and insertion (frame-shift) mutagenesis of the V1 gene

at nucleotide 34. However, the deletion mutant was capable of some symptomless infectivity, and *in planta* replication of deletion and insertion mutants was rescued by wild-type virus. This suggested that V1 and its product were not absolutely required for replication, but were involved in some other facet of symptom development. The authors speculated that V1 might be required for the spread of the virus through the plant. This investigation did not succeed in defining a more precise function for the V1 ORF.

Using deletion, gene replacement, and point mutants, Lazarowitz *et al.* (1989) demonstrated that V1 is not required for replication, and is essential for viral movement. Boulton *et al.* (1993) showed that DNA replication and encapsidation of V1 gene mutants, that were unable to produce systemic infections in plants, occurred in maize protoplasts with the same efficiency as for wild-type virus. This strongly implied a movement function for the V1 gene product, which is now generally referred to as a movement protein (MP). In addition, the authors made a prediction of secondary structure and hydrophobicity for the MPs of MSV and TYDV according to the parameters of Chou and Fasman (1974). These analyses predicted a large (hydrophobic) transmembrane region in the MPs of both viruses with a high degree of confidence, and this feature is conserved among a variety of mastreviruses (personal observation, and Boulton, 2002).

It was noted that the dicotyledonous host range of TYDV implies that any virus-host interactions in this region are probably not specific to cereals. A hydrophobic region in these proteins is not surprising, since many movement proteins already characterised in a variety of plant viruses (e.g. tobacco mosaic virus, TMV) are associated with membrane structures. Indeed, Dickinson *et al.* (1996) showed by immuno-electron microscopy that the MSV MP was localised to the cell walls, and was in close association with secondary plasmodesmata. Kotlizky *et al.* (2000) provided strong direct evidence that MP can affect plasmodesmal gating – they

bombarded constructs directing the expression of either green fluorescent protein (GFP) or GFP-MP fusion proteins into maize leaves, and showed that GFP-MP fusion proteins spread more efficiently from bombarded cells into their neighbours. By analogy with MPs of unrelated viruses, Boulton (2002) has suggested that the helical hydrophobic domain may result in oligomerisation of MP, although no functional requirement for MP oligomers has been demonstrated.

It has also been shown *in vitro* that MP binds specifically to CP (Liu *et al.*, 2001a) and that CP binds nonspecifically to ss and ds DNA (Liu *et al.*, 1997a), and this is the suggested mechanism whereby MP facilitates movement of an undefined CP-DNA nucleoprotein virus complex out of the nucleus and to adjacent cells (Liu *et al.*, 2001a; Boulton, 2002). Further evidence of a nuclear export function for MSV MP was provided by Kotlizky *et al.* (2000), who used microprojectile bombardment of maize leaves with expression constructs to show that MP prevented the accumulation of a CP-GFP fusion protein in the nuclei of bombarded cells.

V2 - coat protein

After the initial identification of the CP coding region following the determination of the first complete MSV genomic sequence (Mullineaux *et al.*, 1984), the ORF encoding the MSV CP was mapped on the genome by Morris-Krsinich *et al.* (1985) using a hybrid-arrested translation strategy. The CP produced by this *in vitro* translation experiment was specifically immunoprecipitated by antiserum to whole virus, and had an electrophoretic mobility indicating a molecular weight of 28 kDa; the predicted (according to DNA sequence) molecular weight is 26 969 Da. In the same experiments which determined that MP was dispensable for MSV replication *in planta*, it was shown that replication was independent of CP (Boulton *et al.*, 1989; Boulton *et al.*, 1993).

Using cryo-electron microscopy and three-dimensional image reconstruction, Zhang *et al* (2001) determined the structure of the MSV geminate particle to a resolution of 25 Å (Figure 1-3). This study revealed that the MSV geminate particle is 220 x 380 Å in size, and consists of two incomplete T=1 icosahedra, comprising a total of 22 capsomers each made up of five CP molecules. The CP model structure used in this study (Figure 1-3) was based on that of satellite tobacco necrosis virus (STNV), and consisted of an eight-stranded, anti-parallel β-barrel motif with an N-terminal α-helix, which the authors noted as being common to all known ssDNA viruses.

In addition to its obvious role in the encapsidation of the viral ssDNA, mastrevirus CP performs a multitude of other functions: the CP of at least some leafhopper-transmitted geminiviruses determines vector specificity, implying specific interactions with unknown vector factors (Boulton *et al.*, 1989; Briddon *et al.*, 1990); MSV CP is capable of binding non-specifically to both ss and ds DNA (Liu *et al.*, 1997a), and thereby facilitates the nuclear import of virus DNA (Liu *et al.*, 1999a); MSV CP is required for cell-to-cell and systemic spread of the virus in plants (Boulton *et al.*, 1989; Lazarowitz *et al.*, 1989; Boulton *et al.*, 1993; Liu *et al.*, 2001b); and MSV CP interacts specifically with MP to shuttle virus DNA out of the nucleus (Kotlizky *et al.*, 2000; Liu *et al.*, 2001a). It is probably at least partly due to its multifunctional nature that the CP amino acid sequence is highly conserved among a number of geographically distinct strains of MSV (Briddon *et al.*, 1994; Martin *et al.*, 2001; Willment *et al.*, 2001). The amino acid sequence of the CP diverged by no more than 2% (five substitutions in 224 residues) across the whole geographic range of MSV isolates tested by Briddon *et al.* (1994), including isolates from the African mainland and two Indian Ocean islands.

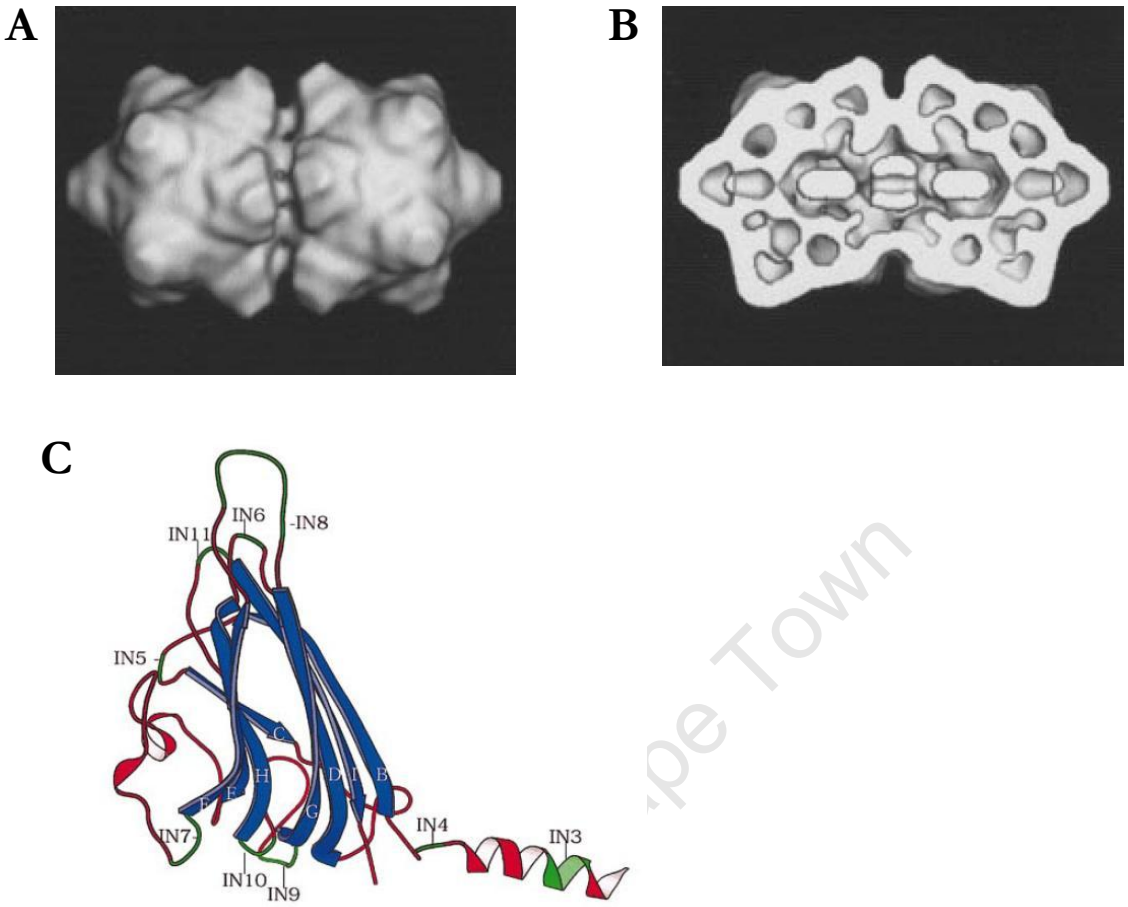


Figure 1-3. MSV particle reconstruction and CP model structure (taken from Zhang *et al.*, 2001).

A The model presented was computed with 52-dihedral-point-group symmetry, using cryo-EM of MSV particles and an MSV CP model structure derived from the known structure of STNV CP. **B** The model shown in A, with the nearest half cut away to reveal the interior of the virion. **C** The MSV CP model structure, based on 3-dimensional alignment with the known structure of STNV CP.

Genes and gene products of the MSV complementary sense strand

There are three TATA boxes located 57, 62 and 101 nt. upstream of the C1 ORF start codon, and any of these may potentially facilitate transcription of the complementary-sense ORFs. Two co-originating complementary-sense transcripts of 1.5 kb and 1.2 kb have been identified in mastrevirus-infected plants (Mullineaux *et al.*, 1990; Dekker *et al.*, 1991). While the shorter transcript is thought to direct expression of the C1 ORF to produce a ~31 kDa protein unique to mastreviruses termed RepA (Mullineaux *et al.*, 1990; Dekker *et al.*, 1991; Gutierrez, 1999), the

longer transcript may be spliced to remove an intron and enable translation of the full-length Rep protein (~40 kDa) comprising N- and C-terminal portions encoded by the C1 and C2 ORFs, respectively (Wright *et al.*, 1997). The splicing of equivalent C-sense introns has been demonstrated for WDV (Schalk *et al.*, 1989), DSV (Accotto *et al.*, 1989) and TYDV (Morris *et al.*, 1992). The C2 ORF is not thought to be translated independently of C1, since no C2-specific transcript has been detected and all known mastrevirus spp. except MSV lack a C2 start codon (Palmer & Rybicki, 1998). While Rep alone appears to be both necessary and sufficient for replication of mastreviruses in appropriate host cells (Schalk *et al.*, 1989; Liu *et al.*, 1998a), RepA is thought to perform a variety of important additional functions during the mastrevirus life cycle (Gutierrez, 1999; Boulton, 2002).

The RepA and Rep proteins of MSV share 214 amino acids at their N-termini and many functionally relevant sequence motifs are thought to reside in this region. Three motifs were identified as a result of the observation that they are conserved in RCR-associated proteins encoded by a variety of ssDNA phages and plasmids (Koonin & Ilyina, 1992). Arguello-Astorga and Ruiz-Medrano (2001) showed that Motif I is probably a conserved element of a DNA-binding domain (also called the "iteron recognition domain"; IRD) responsible for recognising the so-called "iteron" DNA sequences in the LIR. RCR Motif II is thought to be involved in binding Mn^{2+} or Mg^{2+} (Koonin & Ilyina, 1992; Laufs *et al.*, 1995b). By analogy with TYLCV, the mastrevirus RCR Motif III is assumed to facilitate DNA cleavage and re-ligation at the virion sense origin of replication (Laufs *et al.*, 1995b). Furthermore, the shared N-terminal portions of Rep and RepA contain a stretch of 13 amino acids that was shown by yeast two-hybrid assays to enable the formation of both homo- and/or heter-oligomers of Rep and RepA (Horvath *et al.*, 1998), and the oligomerisation of Rep and/or RepA has been linked to DNA sequence specific binding functions via electron microscopy studies (Sanz-Burgos & Gutierrez, 1998; Castellano *et al.*, 1999; Missich *et al.*, 2000). Finally, mastrevirus Rep and RepA both contain a motif that

bears resemblance to the retinoblastoma protein binding motif common to a variety of viral oncoproteins (Xie *et al.*, 1995), and WDV Rep and RepA (Collin *et al.*, 1996) as well as MSV RepA interact with maize retinoblastoma-related protein (RbR; Grafi *et al.*, 1996; Xie *et al.*, 1996; Horvath *et al.*, 1998). Interestingly, Rep proteins of both MSV and BeYDV do not bind to RbR (Horvath *et al.*, 1998; Liu *et al.*, 1999c), possibly due to protein folding differences between MSV Rep and RepA (Horvath *et al.*, 1998; Gutierrez, 1999).

Collin *et al.* (1996) showed that RepA, but not Rep, is capable of transactivating virion-sense gene expression in WDV, and the C-terminal, unique portion of MSV RepA comprises a region associated with transactivation activity in yeast cells (Horvath *et al.*, 1998). RepA also uniquely contains a region that facilitates interactions with a group of host factors termed "geminivirus RepA binding" (GRAB) proteins (Xie *et al.*, 1999), and is thereby thought to influence host cell cycle regulation, and possibly other developmental pathways (Gutierrez, 2002).

The geminivirus Rep C-terminal region contains a bipartite NTP binding motif (Gorbalenya & Koonin, 1989; Gorbalenya *et al.*, 1990) that has been associated with an ATPase activity required for efficient replication (Desbiez *et al.*, 1995; Hanson *et al.*, 1995). While no specific roles for NTP-binding and/or ATPase activity in Rep function have been demonstrated, Gorbalenya *et al.* (1990) noted that some other proteins containing these conserved motifs function as helicases. This portion of Rep also comprises a *myb*-like domain that has been implicated in transactivation of the virion-sense genes of WDV (Hofer *et al.*, 1992) and *Chloris* striate mosaic virus (CSMV; Zhan *et al.*, 1993), and Horvath *et al.* (1998) demonstrated that Rep is capable of transactivating heterologous promoters in yeast.

The aims and rationale of this work

As with most research efforts spanning multiple years of experimentation, the objectives of the work presented here shifted and evolved with each attempted experiment, and in response to work published by others; to pretend that a coherent plan of work existed at the outset, and that such a plan was followed to its conclusion, would be farcical. Nevertheless, if the investigations described in this dissertation followed a central theme, then that theme could be said to be MSV evolution.

At the time that the experiments of Chapter 2 were undertaken, very little was known about the biological implications of MSV diversity, the determinants of MSV host specificity, or the biochemical details of the functions of the MSV virion-sense genes. Liu *et al.* (1997a; 1999a; 2001a) were yet to publish their work describing the interactions among the MSV CP, MP, and DNA which are now thought to facilitate intra- and intercellular movement of MSV DNA. Moreover, from the naïve perspective of an undergraduate education that focussed on bacterial and phage genetics, it seemed striking to me that such an apparently simple virus – comprising just three genes expressing four proteins – was the pathogen responsible for the severe, wide-spread and complex disease that is MSD. My instinct was that the complexity of MSD might represent a typical example of "emergence" (Corning, 2002), i.e. that subtle and manifold interactions within and among the genomes and proteins of the virus, insect vector and plant hosts were likely to explain the complexity and diversity of the disease.

The laboratory in which I enrolled for my doctoral studies possessed the most diverse collection of MSV isolates in the world, and the bulk of the resident technical expertise at the time was centred on recombinant DNA techniques and phylogenetic analyses of DNA sequences. Thus it was an obvious, albeit speculative, progression to conduct genetic exchange and complementation studies using diverse MSV

isolates in order to probe both intra-virus as well as virus-host genetic interactions. The second chapter of this dissertation therefore describes the biological effects of genetic exchanges between two diverse MSV isolates, and thereby confirms the biological importance of specific interactions between the MSV MP and CP genes that were subsequently more directly demonstrated by Liu *et al.* (1997a; 1999a; 2001a).

Following the work presented in Chapter 2, two distinct paths forward seemed obvious. The most obvious option was to characterise the implied interactions between the MP, CP, and the rest of the MSV genome in more detail, with the aim of elucidating the relevant biochemical mechanisms. Various experiments in this vein were attempted but technical obstacles prevented me from achieving anything of significant value, and it eventually became apparent that teams in other laboratories were in a better position to conduct these investigations. With the subjects of geminivirus phylogenetics and evolution already a major focus of my supervisors and colleagues, it was apparent that the infectious chimaeric viruses constructed in Chapter 2 represented a unique model system for studying MSV evolution in a controlled manner in the laboratory. This realisation laid the foundation for the work presented in Chapters 3 and 4.

At a conceptual level, evolution of species occurs via two processes: diversification, and selection. Diversification, in turn, is driven by two fundamentally distinct mechanisms: mutation, and recombination. Since we had demonstrated that chimaeric MSV genomes could be infectious, but suffered significant reductions in fitness relative to their wild-type parental viruses, it was of interest to me to study the frequency, distribution, and types of mutations which might occur in these viruses over an extended period of selection *in planta*, and this is the work presented in Chapter 3.

Recombination, the other major origin of genetic diversity in the course of evolution, enjoyed increased interest amongst geminivirologists when Jeske *et al.* (2001) drew attention to the likely mechanistic role of recombination during geminivirus replication. Recombination also became increasingly relevant following the mounting number of reports of sub-genome-length geminivirus DNAs (Mansoor *et al.*, 1999; Saunders & Stanley, 1999; Saunders *et al.*, 2000; Briddon *et al.*, 2001), and with the publication of evidence of the significant contribution that recombination has made to geminivirus evolution (Padidam *et al.*, 1999). In Chapter 4 I describe the use of complementary (reciprocal) chimaeric MSV genomes in experiments to investigate the adaptive power and mechanistic aspects of recombination during MSV infection of maize plants.

CHAPTER 2: VIABLE CHIMAERIC VIRUSES CONFIRM THE BIOLOGICAL IMPORTANCE OF MOVEMENT PROTEIN–COAT PROTEIN INTERACTIONS

ABSTRACT

Protein-protein and protein-DNA interactions have been shown to mediate the movement of begomoviruses within their plant hosts, and evidence for specific binding between MSV movement and capsid proteins has been demonstrated *in vitro*. The primary aim of this study, therefore, was to investigate the biological significance of interactions between the two virion-sense open reading frames (ORFs) of maize streak virus (MSV), their products, and the rest of the genome. Six partially dimerised (agroinfectious), chimaeric MSV constructs were made by reciprocally exchanging ORFs V1 and V2 individually, and in pairs, between MSV-Kom and MSV-Set. These two very distinct isolates share just 78% nucleotide identity, with MSV-Kom representing the well-characterised MSV-A strain responsible for the vast majority of infections in maize crops, while MSV-Set is the sole example of the MSV-C strain and induces relatively mild symptoms in maize. All chimaeras were infectious in *Zea mays* c.v. Jubilee and were characterised in terms of symptomatology and infection efficiency. Compared with their parental viruses, all chimaeric viruses were attenuated in symptom severity, infection efficiency, and the rate at which symptoms appeared. The exchange of individual V1 and V2 ORFs resulted in lower infection efficiency and reduced symptom severity in comparison with exchanges of matched V1-V2 pairs, suggesting that specific interactions between these ORFs and/or their gene products are biologically significant. The potential utility of infectious chimaeric viruses in the investigation of *in planta* viral mutation and/or recombination is discussed.

Introduction

Looking back on the history of genetics, it is clear that the very earliest experiments aimed at investigating genetic heredity made use of recombination. More than 150 years ago, in the famous experiments that ultimately gave birth to the concept of the gene, Mendel unknowingly utilised the recombination that occurs during meiosis to exchange alleles between several varieties of pea. A century later, Edward B. Lewis helped to lay one of the cornerstones of modern experimental molecular genetics when he devised the genetic complementation study. In these tests, a wild-type allele may be shown to restore a wild-type phenotype when it is crossed into a mutant organism, thus confirming the function of the gene in question. Such obvious and intuitive experimental approaches that involve mutating, exchanging, and/or transplanting genetic elements between related organisms have spread throughout modern molecular biology.

Mutation studies are often employed in attempts to identify the genetic basis of important aspects of a pathogen's phenotype. For example, in order to understand the genomic determinants of pathogenicity, genetic elements may be altered in, deleted from, or exchanged between virulent and benign pathogen isolates. During the last two decades, geminivirus molecular biologists have made extensive use of genetic exchange in a wide variety of experiments. Briddon *et al.* (1990) replaced the coat protein gene of the whitefly-transmitted (*Bemisia tabaci*) African cassava mosaic virus (ACMV) with that of beet curly top virus (BCTV) and successfully transmitted the recombinant ACMV via the BCTV-specific leafhopper vector *Circulifer renellus* (Baker), thereby demonstrating that insect vector specificity is determined by the coat protein. Similarly, Liu *et al.* (1999b) constructed chimaeras of the dicot-infecting bean yellow dwarf virus (BeYDV) and MSV with the aim of identifying host specificity determinants. Although none of these chimaeras were able to systemically infect host plants of either parental virus, the study demonstrated the importance of intragenomic interactions in mastreviruses, and exposed the consequent limitations

of genetic swaps between such diverse members of the genus. Among the bipartite begomoviruses, pseudorecombination of A and B components has formed the basis of many useful studies illuminating various *trans*-acting functions important in replication (Lazarowitz *et al.*, 1992; Frischmuth *et al.*, 1993), symptom development (von Arnim & Stanley, 1992a), and *in planta* virus movement (von Arnim & Stanley, 1992b).

Findings from a large number of studies have led to a fairly detailed model of bipartite begomovirus movement (Hehnle *et al.*, 2004) involving interactions between viral DNA and the nuclear shuttle protein (NSP; encoded by ORF BV1), and between a viral DNA-NSP complex and the movement protein (MP; encoded by ORF BC1). There is good evidence of analogous interactions involving the MP (encoded by ORF V1) and coat protein (CP; encoded by ORF V2) of mastreviruses (Boulton *et al.*, 1993; Dickinson *et al.*, 1996; Liu *et al.*, 1997a; Liu *et al.*, 1999a; Liu *et al.*, 2001a) but the biological importance of these interactions has not yet been demonstrated in a natural host plant. We felt that it may be possible to employ genetic complementation to illustrate the functional relevance of specific interactions between the mastrevirus MP and CP.

In the small and presumably informationally compact MSV genome, deletion or inactivation of any genes results in asymptomatic infections or loss of infectivity (Boulton *et al.*, 1989; Lazarowitz *et al.*, 1989) and in some cases even small alterations in coding or intergenic regions have resulted in dramatic attenuation of virulence (Boulton *et al.*, 1989; Shen & Hohn, 1991; Boulton *et al.*, 1991a; Boulton *et al.*, 1991b; Schneider *et al.*, 1992; Willment, 1999; Shepherd *et al.*, 2005). While some of these mutations obviously altered amino acid sequences (Boulton *et al.*, 1989; Shepherd *et al.*, 2005) or disrupted conserved DNA sequences required for replication or transcription (Schneider *et al.*, 1992; Willment, 1999) the deleterious effects of other mutations has been more difficult to explain (Boulton *et al.*, 1991a). In one instance,

11 of the 14 N-terminal amino acids of the MSV MP were altered without causing a noticeable loss of virulence (Boulton *et al.*, 1989) but this is an exceptional case in the literature. Notwithstanding the apparent fragility of mastreviruses in the face of mutation, we reasoned that relatively substantial genetic changes might be tolerated if effected via the exchange of homologous genomic modules, rather than through the introduction of isolated point mutations or deletions. We took an ambitious stance and set out to exchange the virion-sense ORFs between the two most divergent of known MSV strains – MSV-Kom and MSV-Set.

MSV-Kom was originally isolated from maize grown in Komatipoort, South Africa, while MSV-Set was obtained from a *Setaria* sp. grass sample collected from Mt. Edgecombe (Schnippenkoetter *et al.*, 2001a), also in South Africa. Both strains have been characterised in terms of host range, transmission, and symptomatology. The viruses share 78% nucleotide sequence identity overall, with the V1 and V2 ORFs being 80% and 79% identical, respectively. MSV-Kom is an isolate of the MSV-A strain, which is the predominant MSV strain infecting maize in Africa (Martin *et al.*, 2001). MSV-Set, on the other hand, is the sole characterised representative of the MSV-C strain, and produces considerably milder symptoms in maize than does MSV-Kom (Schnippenkoetter *et al.*, 2001a). This chapter describes the construction of a series of six infectious MSV-Kom/MSV-Set chimaeric genomes comprising all the possible combinations of parental virus, V1, and V2 regions. Both parental viruses and all six recombinant viruses were assessed in terms of infectivity and symptomatology, and evidence of biologically important specific interactions between the MSV CP and MP is presented.

Methods and materials

Virus isolates, plasmids, bacterial strains, enzymes, and maize genotypes

The plasmid pKEP170 consists of a site-directed MSV-Kom mutant (T→G at nt. 468 relative to the virion strand *ori*, to introduce an *Nco*I restriction site at the 5' end

of the V2 ORF) cloned into the *Bam*HI site of pUC19, and was obtained from K. Palmer (University of Cape Town, South Africa). The plasmids pSET100, pSET107 and pKOM603 were obtained from Darren Martin (University of Cape Town) and are described in Schnippenkoetter *et al.* (2001a); pSET100 consists of a single MSV-Set genome cloned into the *Bam*HI site of pUC18; pSET107 comprises a partial dimer of the MSV-Set genome cloned between the right and left T-DNA borders of pBI121; pKOM603 consists of a partial dimer of MSV-Kom cloned between the right and left borders of pBI121.

The cloning vectors pBluescriptSK+ (pSK+; Stratagene, La Jolla, CA) and pUC19 (Stratagene), and the RecA- *Escherichia coli* strains DH5α and JM109 were used in all standard cloning procedures. The *E. coli*/*A. tumefaciens* binary vector pBI121 (Clontech, CA, U.S.A.) was used to produce agroinfectious DNA constructs, and the *Agrobacterium tumefaciens* strain C58C1[pMP90] (Koncz & Schell, 1986) was used for all agroinoculations. Restriction enzymes and DNA ligase were obtained from a variety of commercial suppliers and were used according to the manufacturers' instructions.

Sweetcorn maize cv. Jubilee seeds were purchased from Starke Ayres nursery (Rosebank, Cape Town, South Africa).

Construction of chimaeric viral genomes

Eight chimaeric viral genomes were constructed by reciprocally exchanging ORFs V1 and V2 either singly, or in pairs, between MSV-Kom and MSV-Set. To facilitate these ORF exchanges, it was necessary to first introduce an *Nco*I restriction site at nt. position 471 near the V2 start codon in the MSV-Set genome. The pUC19 sequencing primer 1212 (5' – CGCCAGGGTTTTCCCAGTCACGAC – 3') and primer SetNcoRev (5' – **CCATGG**CTTACTGCTTATCCC – 3') were used to amplify a 0.36 Kbp fragment of MSV-Set from pSET100 using Taq DNA

polymerase in a standard PCR (94 °C for 180s; 30 cycles of 30s at 94 °C, 30s at 52.8 °C, 60s at 72 °C; 72 °C for 300s). The primer SetNcoRev introduced the required *Nco*I recognition sequence (bold type) via a T→G transversion (underlined). Similarly, the remaining 2.4 Kbp portion of the MSV-Set genome was amplified using the pUC19 sequencing primer 1212 (5' – CGCCAGGGTTTTCCCAGTCACGAC – 3') and primer SetNcoFwd (5' – **CCATGG**CGACGTCCAAGAGG – 3'; the introduced *Nco*I site is in bold type and the T→G transversion underlined). Thermal cycling parameters were as follows: 94 °C for 30s; 30 cycles of 94 °C for 30s, 56.5 °C for 30s, 72 °C for 180s; 72 °C for 300s.

Both amplified DNA products were purified by agarose gel electrophoresis (GENECLEAN® III kit, Qbiogene, CA, U.S.A). The Klenow fragment of DNA polymerase I was used to fill in the ends of both PCR products as described by Sambrook *et al.* (1989). The blunt-ended PCR products were then purified by agarose gel electrophoresis.

The purified, blunt-ended PCR products were digested with *Bam*HI and cloned into the *Bam*HI and *Hind*II sites of pSK+ to produce pERC1 and pERC2 containing the 2.4 Kbp and 0.36 Kbp fragments, respectively. Restriction enzymes *Kpn*I and *Xba*I were used to transfer the insert from pERC1 into pUC18 to yield pERC3.

Using a ratio of 0.25U *Nco*I per microgram of DNA for one hour at 37 °C with the recommended buffer, *Bam*HI-linearised pERC3 plasmid was partially digested and the 2.4kbp fragment purified by agarose gel electrophoresis. This insert was ligated simultaneously into the *Bam*HI site of pUC18 along with the 3.2kbp *Bam*HI/*Nco*I fragment of pERC2, to yield pSNco – a full-genome clone of MSV-Set containing an *Nco*I site at the start of the V2 ORF.

Using vector sequencing primers, the inserts of both pKEP170 and pSNco were sequenced unidirectionally through the V1 ORFs into the V2 ORFs by D. James at the Sequencing Service, Dept. Molecular & Cell Biology, University of Cape Town. The sequences obtained for pKEP170 and pSNco were aligned with those of MSV-Kom and MSV-Set, respectively, using DNAMan™ Version 2.2 (default settings).

*Pst*I was used to digest pKEP170 and pSNco to completion and the linearised plasmids were purified by agarose gel electrophoresis. The purified, linearised plasmids were partially digested with *Nco*I using a ratio of 0.25 U *Nco*I : 1 μg DNA for one hour at 37 °C with the recommended buffer. Fragments of both plasmids of sizes 0.33kbp, 0.69kbp, and 4.4kbp (respectively referred to as K1, K2, and K3 for pKEP107 and S1, S2, and S3 for pSNco) were purified by agarose gel electrophoresis. Ligations were performed using equimolar quantities of the purified fragments of pKEP107 and pSNco, using the following combinations of fragments: 1) S1, K2, K3; 2) K1, S2, K3; 3) S1, S2, K3; 4) K1, S2, S3; 5) S1, K2, S3; 6) K1, K2, S3.

*Bgl*I and *Sma*I were used separately to screen transformants for recombinant plasmids comprising the appropriate chimaeric MSV-Kom/MSV-Set genomes cloned into the *Bam*HI site of pUC18. These plasmids were named as follows, corresponding to the numbers of the ligation reactions described above: 1) pKV1S; 2) pKV2S; 3) pKV1V2S; 4) pSV1K; 5) pSV2K; 6) pSV1V2K.

Construction of agroinfectious plasmids

A double restriction digest was used to delete *Bam*HI/*Bgl*II fragments of 1.25kbp and 2.34kbp from pKEP107 and pSNco respectively. The resulting plasmid comprising the MSV-Kom LIR was called pKBBg, and the plasmid containing the MSV-Set LIR was called pSBBg. The full-length viral genomes from pKEP107, pKV1S, pKV2S, and pKV1V2S were cloned into the *Bam*HI site of pKBBg to yield

the partially dimeric constructs pKNcoδ, KV1Sδ, pKV2Sδ, and pKV1V2Sδ in pUC18 (Figure 2-1~~Figure 2-1~~). Similarly, inserts from pSNco, pSV1K, pSV2K, and pSV1V2K were used with pSBBg to generate the partially dimeric constructs pSNcoδ, pSV1Kδ, pSV2Kδ, and pSV1V2Kδ.

To construct agroinfectious plasmids, the partially dimerised viral genomes described above were transferred to the *E. coli*/*A. tumefaciens* binary vector pBI121 using *Eco*RI and *Xba*I. In the case of the MSV-Kom-based constructs (pKNcoδ, KV1Sδ, pKV2Sδ, and pKV1V2Sδ) both restriction sites are unique. However, an *Eco*RI site in the MSV-Set genome made it necessary to perform partial *Eco*RI digests of the MSV-Set-based plasmids (pSNcoδ, pSV1Kδ, pSV2Kδ, and pSV1V2Kδ) in order to obtain the 5.04kbp fragments for ligation into pBI121; a ratio of 0.4U *Eco*RI:µg DNA was used for one hour at 37 °C with the recommended buffer. Plasmids consisting of partially dimerised virus genomes in the binary vector pBI121 were named as follows: pKNcoB, KV1SB, pKV2SB, pKV1V2SB, pSNcoB, pSV1KB, pSV2KB, and pSV1V2KB (Figure 2-1~~Figure 2-1~~).

Agroinoculation

The modified freeze-thaw method of Höfgen and Willmitzer (1988) was used to make transformation-competent *A. tumefaciens*. Agroinoculation was performed essentially as described by Grimsley *et al.* (1987): *A. tumefaciens* cells were grown overnight in 5ml liquid cultures (Luria broth with kanamycin and rifampicin at 10µg/ml; gentamycin at 40µg/ml) at 30 °C, with vigorous shaking. Liquid subcultures (5ml) were inoculated with 1ml from the overnight culture, and grown at 30 °C with vigorous shaking until an OD_{600} between 0.4 and 0.8 was reached. The OD_{600} of each culture was adjusted to 0.4 by diluting with culture medium. Cells were washed once in ddH_2O and resuspended in ddH_2O.

Maize seeds were germinated in wet vermiculite in sealed containers in the dark at 30 °C. Individual 3-day old seedlings were injected with 2µl *A. tumefaciens* suspension immediately below the colleoptillar node, using a Hamilton Microliter™ syringe. Inoculated seedlings were planted in damp, sterilized soil immediately following agroinoculation, and grown indoors at 25 °C under Grolux™ fluorescent lights with 16 hours of light per day. The syringe was rinsed thoroughly with 70% (v/v) ethanol followed by ddH_2O between different inocula. At least three separate agroinoculation experiments were conducted for each virus assessed. In each experiment at least 14 plants – and up to 76 plants in the case of less infectious chimaeric viruses – were inoculated with the relevant *A. tumefaciens* culture, and then 14 plants were each injected with 2µl ddH_2O to serve as negative controls. Plants were also infected with wild-type MSV-Kom and MSV-Set by using *A. tumefaciens* harbouring pKOM603 and pSET100, respectively.

Analysis of symptoms

Each plant was inspected for symptoms of virus infection regularly until twenty days post inoculation (dpi; day 0 = day of inoculation) and thereafter every week until forty-five dpi. Symptoms on the first emergent leaf were disregarded to avoid confusion with physical damage inflicted during injection. Plants that did not survive agroinoculation and subsequent planting were disregarded for all subsequent analyses. The percentage of symptomatic plants was used as a measure of infection efficiency and disease progression. Calculations of area under the disease progress curve (AUDPC) were performed using the simple trapezoidal rule for calculating areas. The relative symptom severity for each virus was ranked subjectively on a scale of 1 to 10, with 1 being the mildest and 10 the most severe, based on the extent of chlorotic area, stunting, and leaf curling and malformation.

Leaf numbers 3 and 4 were photographed 24 days after agroinoculation and leaves 5 and 6 were photographed 31 days post infection. Each leaf to be photographed was

removed from the plant and the portion of the leaf corresponding to the middle half was excised and photographed. Leaves were placed under a non-reflective glass sheet on a fluorescent-tube light-box in a darkened room and photographed from a distance of 27cm using a Pentax K1000 camera with a Pentax 50mm lens. Colour slide film (Fujichrome™ 100ASA) was used throughout, with exposures of 1/60s and the aperture set to 2.8.

Results

Viability of chimaeric genomes

To facilitate the exchange of the V1 and V2 regions, PCR-mediated mutagenesis was used to create *Nco*I restriction sites at the start codons of the V2 ORFs in the MSV-Kom and -Set genomes (KNco and SNco respectively; Figure 2-1~~Figure 2-1~~ and Figure 2-3~~Figure 2-3~~); the cloned PCR products were sequenced to ensure that no unintentional mutations had been introduced (data not shown). The corresponding T→G mutations resulted in the substitution of alanine for serine at the second positions of the CP amino acid sequences (Figure 2-3~~Figure 2-3~~). While these substitutions were expected to be conservative, mutant and wild-type viruses were compared by infecting *Z. mays* c.v. Jubilee (a sweetcorn) to confirm that the *Nco*I mutation does not affect either infectivity or symptomatology in this host. Both *Nco*I mutants were indistinguishable from their wild-type counterparts in terms of their infectivity and the symptoms they produced (data not shown). The eight virus constructs used in this study are shown in Figure 2-1~~Figure 2-1~~. In addition to KNco and SNco, all six chimaeric viruses produced symptoms in sweetcorn plants following agroinoculation.

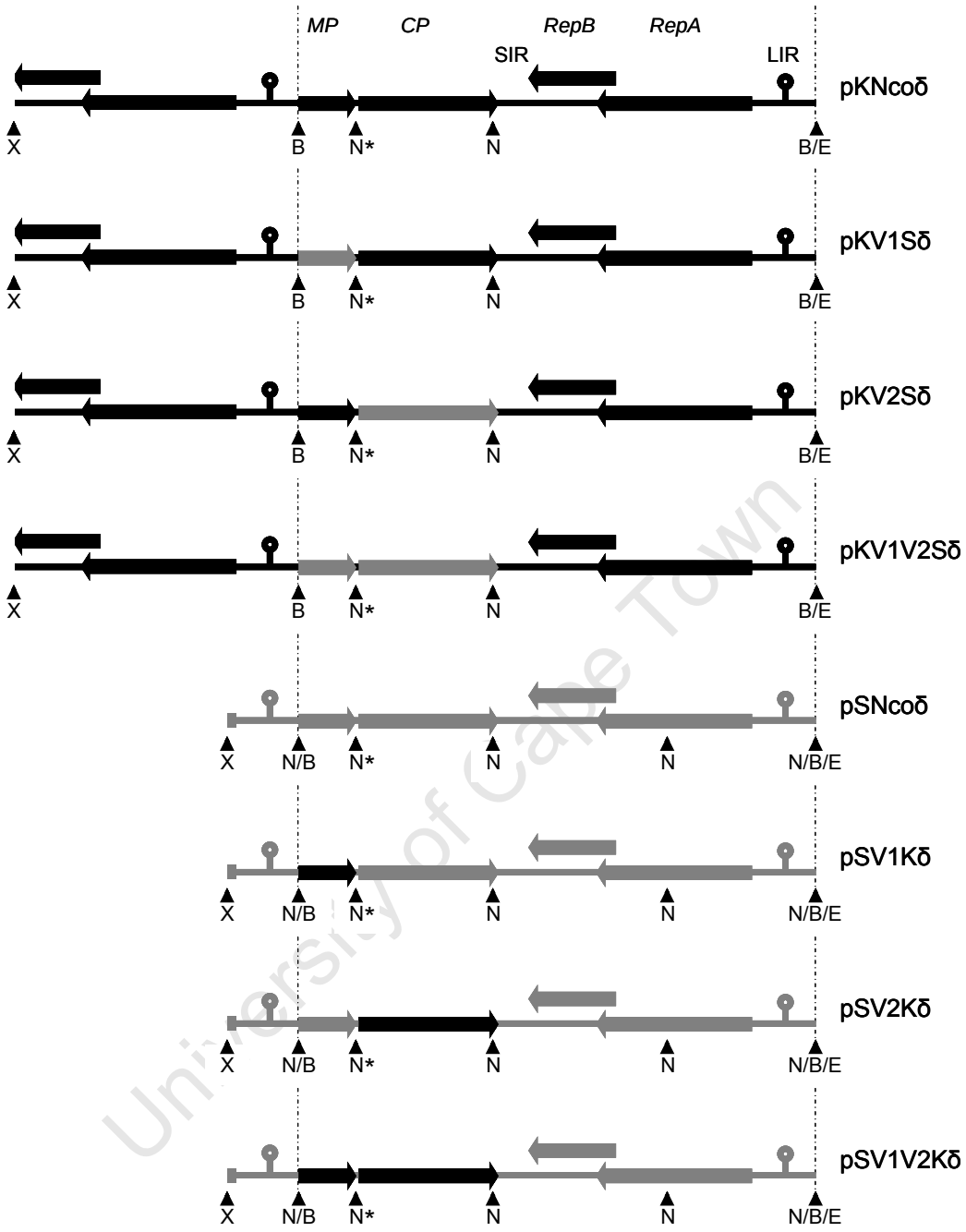


Figure 2-1. MSV-Kom/MSV-Set chimaeric infectious plasmid constructs.

Vector sequences are not shown. Arrows indicate ORFs in the direction of transcription; MSV-Kom sequences are shown in black and MSV-Set sequences in grey. Complete genomes are bounded by vertical dashed lines. The repetition of the stem-loop structure in the LIR allows replicational release of the genomes upon agroinoculation as described by Stenger *et al.* (1991). Restriction sites are indicated by▲; B = *Bam*HI, E = *Eco*RI, N = *Nco*I, X = *Xba*I. ***** The *Nco*I sites between the V1 and V2 ORFs were introduced via PCR-mediated mutagenesis.

Figure 2-2~~Figure 2-2~~ and Figure 2-3~~Figure 2-3~~ show sequence alignments of the MSV-Kom and -Set V1 and V2 ORFs, respectively, and of the corresponding MPs and CPs. Altogether, exchanging V1 ORFs produced changes at 62 out of 320 nucleotide positions, resulting in the alteration of 22 out of 101 MP amino acid residues; switching the 697 bp V2 regions led to 153 nucleotide changes affecting 39 out of 232 possible CP residues. Using a PAM250 substitution matrix (Dayhoff *et al.*, 1978) score of less than one as a guide, ten of the MP differences could be considered to be non-conservative, of which eight appear within the C-terminal quarter of the sequence. Using the same criterion, seventeen of the thirty-nine differences in the CP sequences represent non-conservative substitutions, none of which occurs among the fifty-eight C-terminal residues. Of the known splicing features in the V1 ORF (Wright *et al.*, 1997), only the putative branch point sequence is different between MSV-Kom and –Set, but both variants comply with the requisite intron branch point consensus sequence (YUNAN; Simpson *et al.*, 1996).

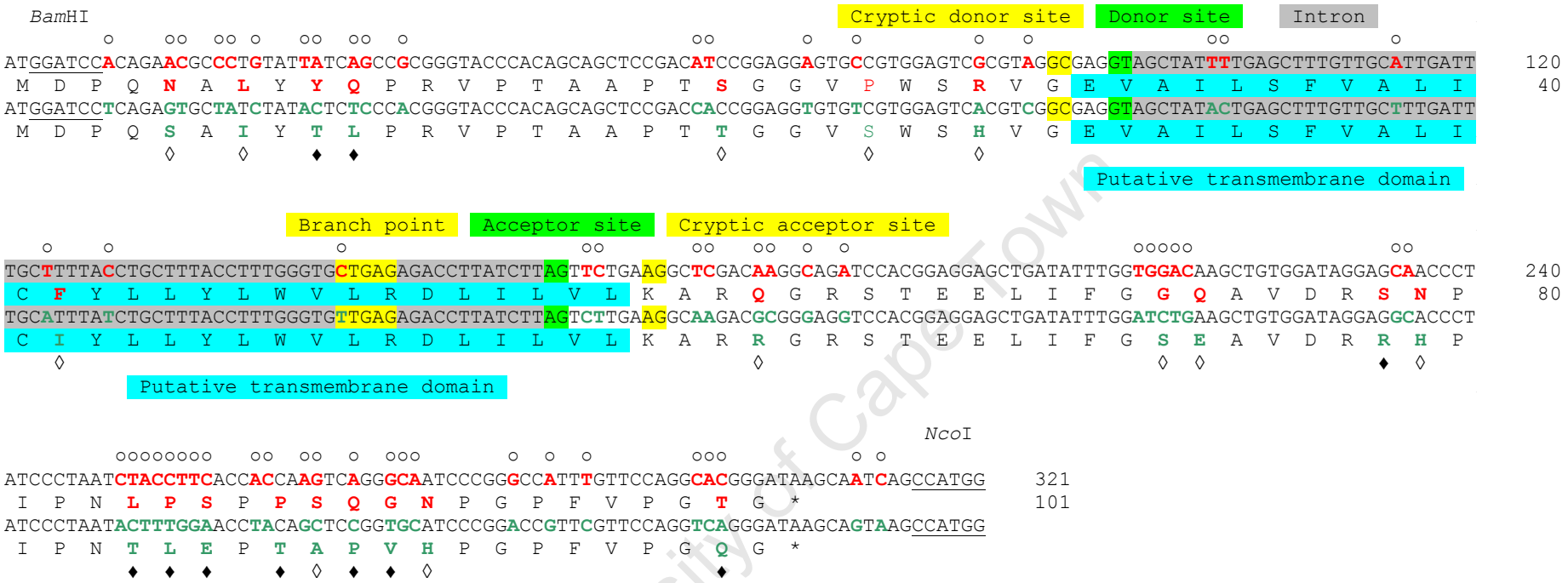
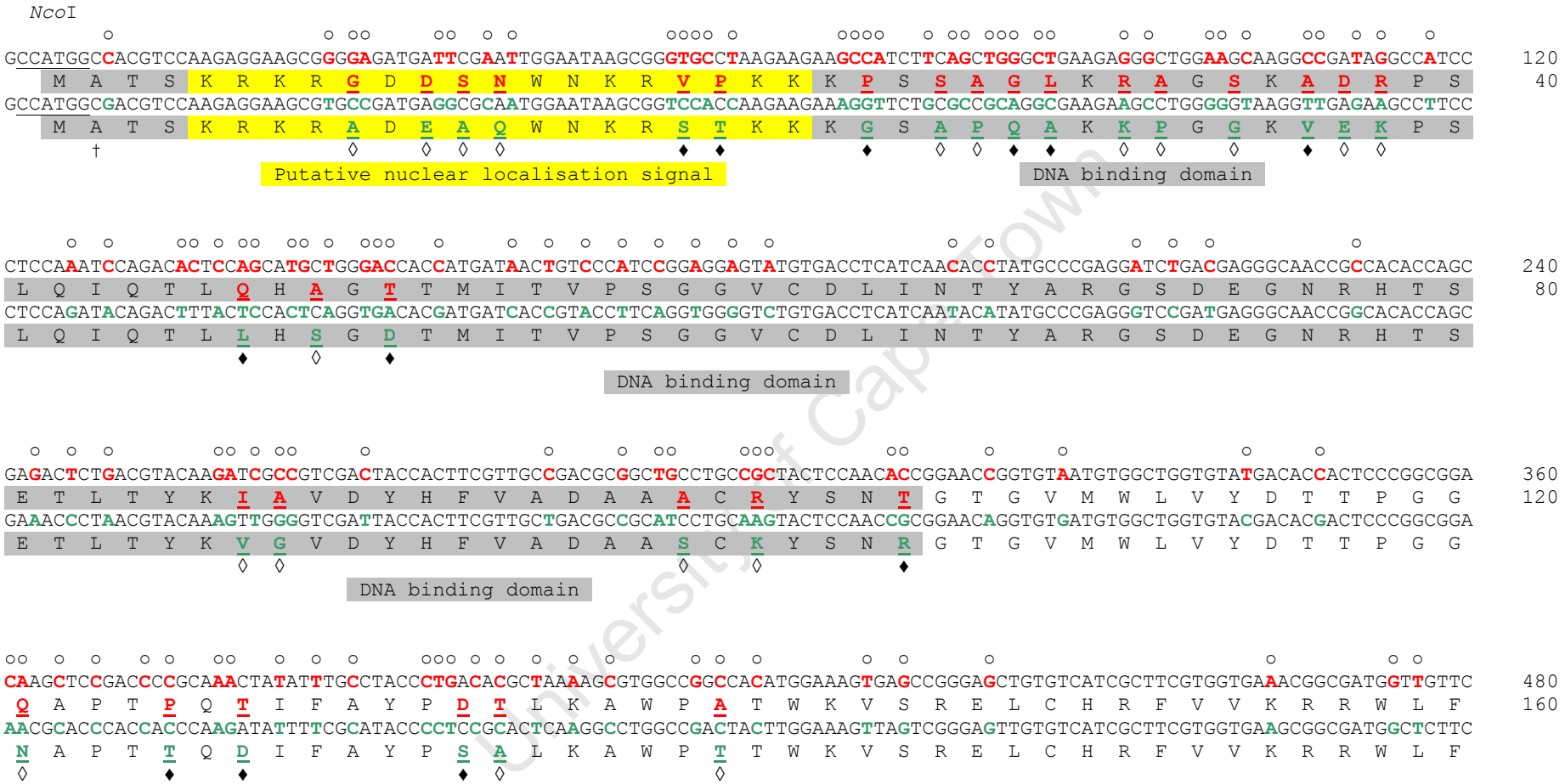


Figure 2-2. Nucleotide and amino acid changes resulting from V1 region exchanges.

Upper line: MSV-Kom V1 region, nucleotide sequence with corresponding MP amino acid sequence below (unique residues in bold, red typeface). Lower line: MSV-Set V1 region, nucleotide sequence with corresponding MP amino acid sequence below (unique residues are in bold, green typeface). Nucleotide differences are indicated with ○ and amino acid differences with ◊ or ♦; amino acid differences with scores < 1 in the PAM250 substitution matrix are marked with ♦; * indicates a stop codon. The predicted transmembrane domain (Boulton *et al.*, 1993; Cserzo *et al.*, 1997) and splicing features (Wright *et al.*, 1997) are highlighted and labelled in the diagram. Restriction sites used for exchanging sequences are underlined. Total nucleotide changes in exchanged region: 62/320 positions (19.4%). Total nucleotide changes in ORF: 60/306 positions (19.6%). Total amino acid changes: 22/101 positions (21.8%).



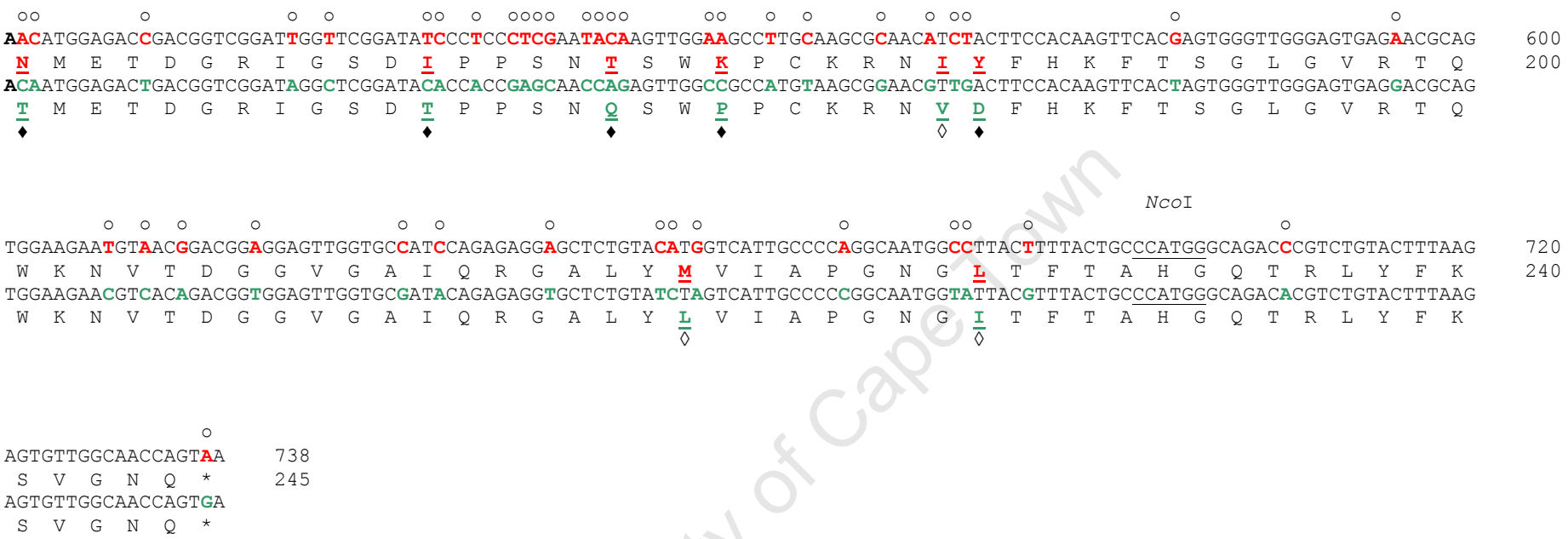


Figure 2-3. Nucleotide and amino acid changes resulting from V2 region exchanges.

Upper line: MSV-Kom V2 region, nucleotide sequence with corresponding CP amino acid sequence below (unique residues in bold, red typeface). Lower line: MSV-Set V2 region, nucleotide sequence with corresponding CP amino acid sequence below (unique residues are in bold, green typeface). Nucleotide differences are indicated with ○ and amino acid differences with ◊ or ♦; amino acid differences with scores < 1 in the PAM250 substitution matrix are marked with ♦; * indicates a stop codon. The predicted nuclear localisation signal (Liu *et al.*, 1999a) and DNA binding domain (Liu *et al.*, 1997a) are highlighted and labelled in the diagram. Restriction sites used for exchanging sequences are underlined. The S→A mutation resulting from the introduction of the *Nco*I site is shown with †. Total nucleotide changes in exchanged region: 153/697 positions (22.0%). Total amino acid changes in the exchanged region: 39/232 positions (16.8%).

Symptom severity and streak morphology

Both symptom severity and streak character varied markedly among the viruses tested. In sweetcorn, KNco typically and consistently produced extensive, yellowish chlorotic streaks which, in extreme cases, were almost as wide as the leaf, and usually extended unbroken for several centimetres (Figure 2-4~~Figure 2-4~~). In severe KNco infections, plants and leaves were noticeably stunted and in some cases leaves were malformed and curled. In contrast, the symptoms of SNco infection were milder in most respects (Figure 2-5~~Figure 2-5~~): the total chlorotic area per leaf was smaller and more variable among SNco infected plants than among plants infected with KNco; SNco did not cause severe stunting, curling or malformation of infected leaves; SNco streaks tended to be shorter and narrower than those of KNco, resulting in a more stippled appearance. However, in one respect SNco appeared to be more pathogenic than KNco in that SNco caused more acute chlorosis, giving rise to whiter streaks. In severe instances the chlorotic tissue eventually disintegrated, leading to fine perforations in the leaves of some SNco-infected plants.

All the chimaeric daughter viruses displayed less virulence than either KNco or SNco (Table 2-1~~Table 2-1~~). With the notable exception of the relatively severe symptoms of KV1S (Figure 2-4~~Figure 2-4~~), chimaeras containing unmatched V1-V2 pairs (KV2S, Figure 2-4~~Figure 2-4~~; SV1K and SV2K, Figure 2-5~~Figure 2-5~~) produced the mildest symptoms: chlorotic lesions were confined to short, narrow streaks that were sparsely distributed across the leaf. In contrast, the two reciprocal chimaeras containing matched V1-V2 pairs (KV1V2S and SV1V2K) were significantly more pathogenic, with SV1V2K showing particularly severe streak symptoms.

As with the parental KNco and SNco viruses, the lesions produced by the chimaeras differed in their degree of chlorosis, and could be classified as either yellow (MSV-Kom-like) or white (MSV-Set-like; Table 2-1~~Table 2-1~~). Chimaeric viruses carrying the MSV-Set V1 ORF – KV2S, SV1K, and SV1V2K – produced yellowish streaks, while those of KV1S, KV1V2S, and SV2K appeared more severely chlorotic and were correspondingly distinctly white.

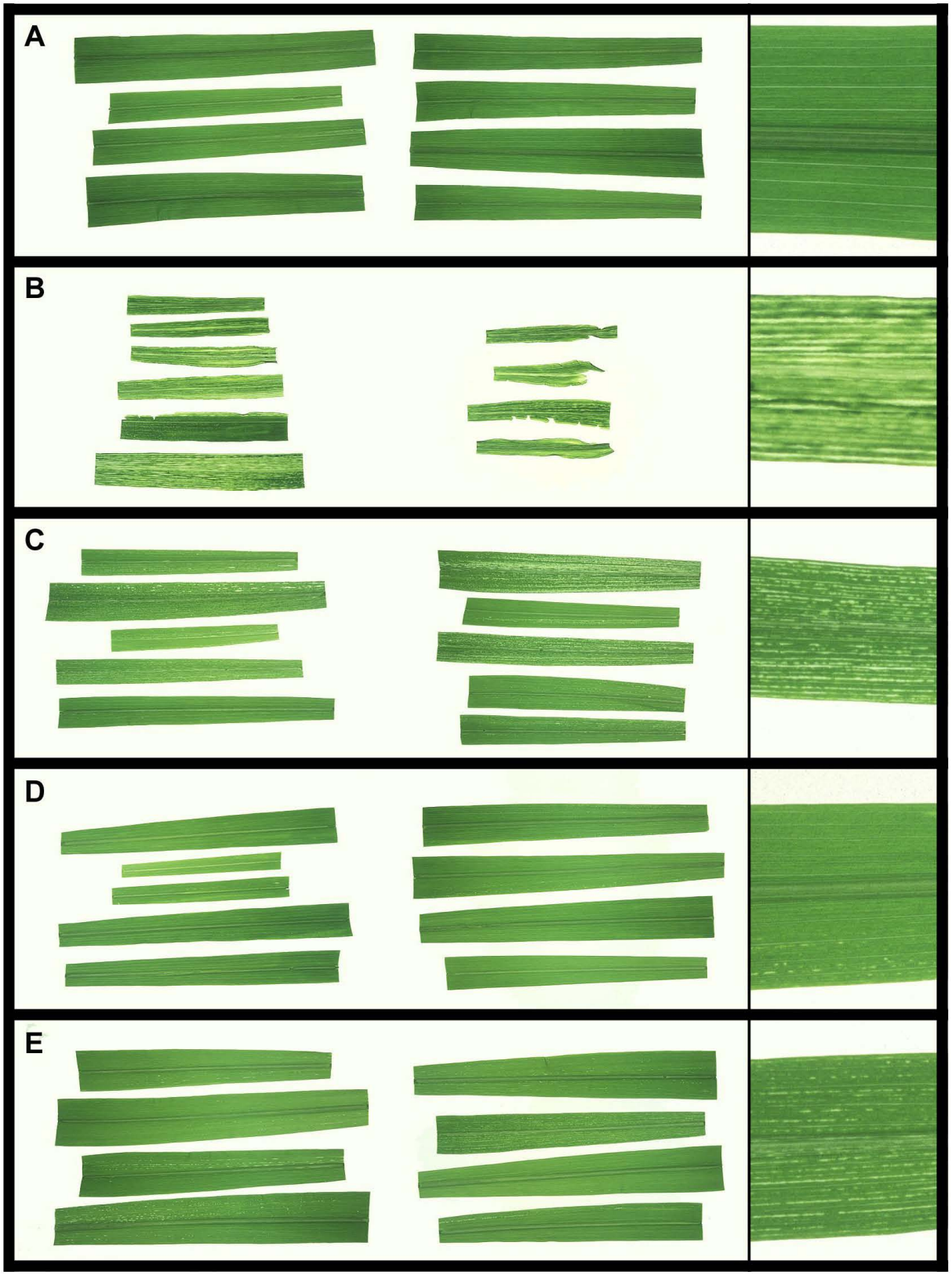


Figure 2-4. Streak symptoms produced by MSV-Kom-based constructs.

Equivalent portions (third quarter from leaf tip) of fourth leaves of infected plants are shown. **A** Mock-infected controls; **B** KNco; **C** KV1S; **D** KV2S; **E** KV1V2S. Insets show magnified views of streak symptoms. All photographs are scaled proportionately.

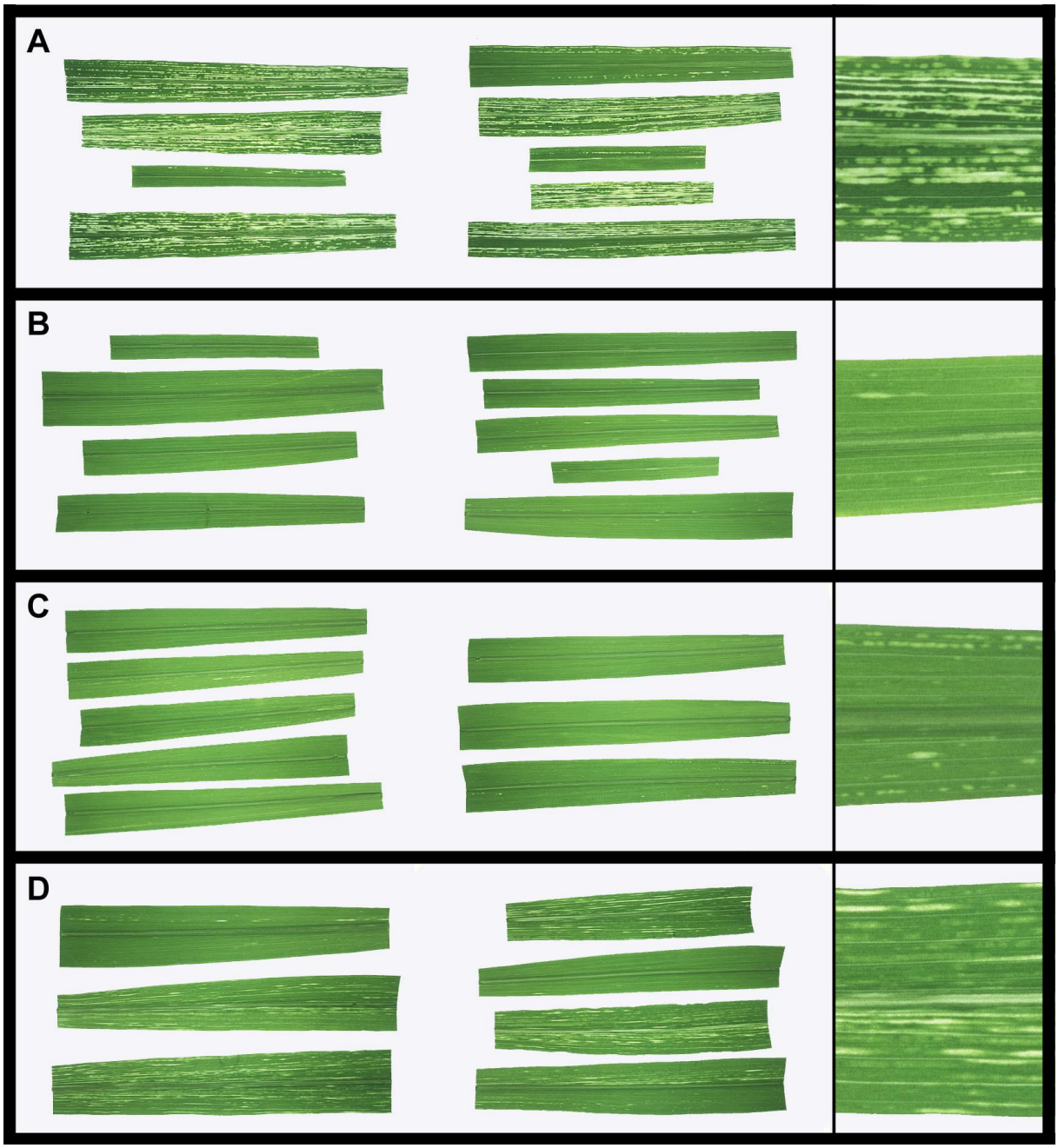


Figure 2-5. Streak symptoms produced by MSV-Set-based constructs.

Equivalent portions (third quarter from leaf tip) of fourth leaves of infected plants are shown. **A** SNco; **B** SV1K; **C** SV2K; **D** SV1V2K. Mock-infected controls are shown in Figure 2-4~~Figure 2-4~~. Insets show magnified views of streak symptoms. All photographs are scaled proportionately.

Table 2-1. Symptomatology of MSV-Kom and -Set chimaeras.

Virus	Origin of ORF: (MSV-**K**om / MSV-**S**et)		Streak colour: (**Y**ellow / **W**hite)	Symptom severity (1 = mild;10 = severe)
	V1	V2		
KNco	K	K	Y	10
KV1S	S	K	W	7
KV2S	K	S	Y	1
KV1V2S	S	S	W	5
SNco	S	S	W	9
SV1K	K	S	Y	2
SV2K	S	K	W	3
SV1V2K	K	K	Y	8

Infection efficiencies and rates of symptom development

To provide additional indications of viral fitness, infectivity and the rate of symptom appearance were determined for each virus by inoculating plants with agroinfectious constructs and then monitoring them for symptom development. Figure 2-6~~Figure 2-6~~ shows the rate at which symptoms appeared in plants following agroinoculation with each virus. As expected, plants inoculated with SNco developed symptoms slightly later than did plants inoculated with KNco, but both viruses showed very few new infections after fifteen days post inoculation (dpi). Over the course of twenty-five days, SNco infected a somewhat smaller percentage of plants than did KNco (SNco, 81% ± 7%; KNco, 87% ± 7%).

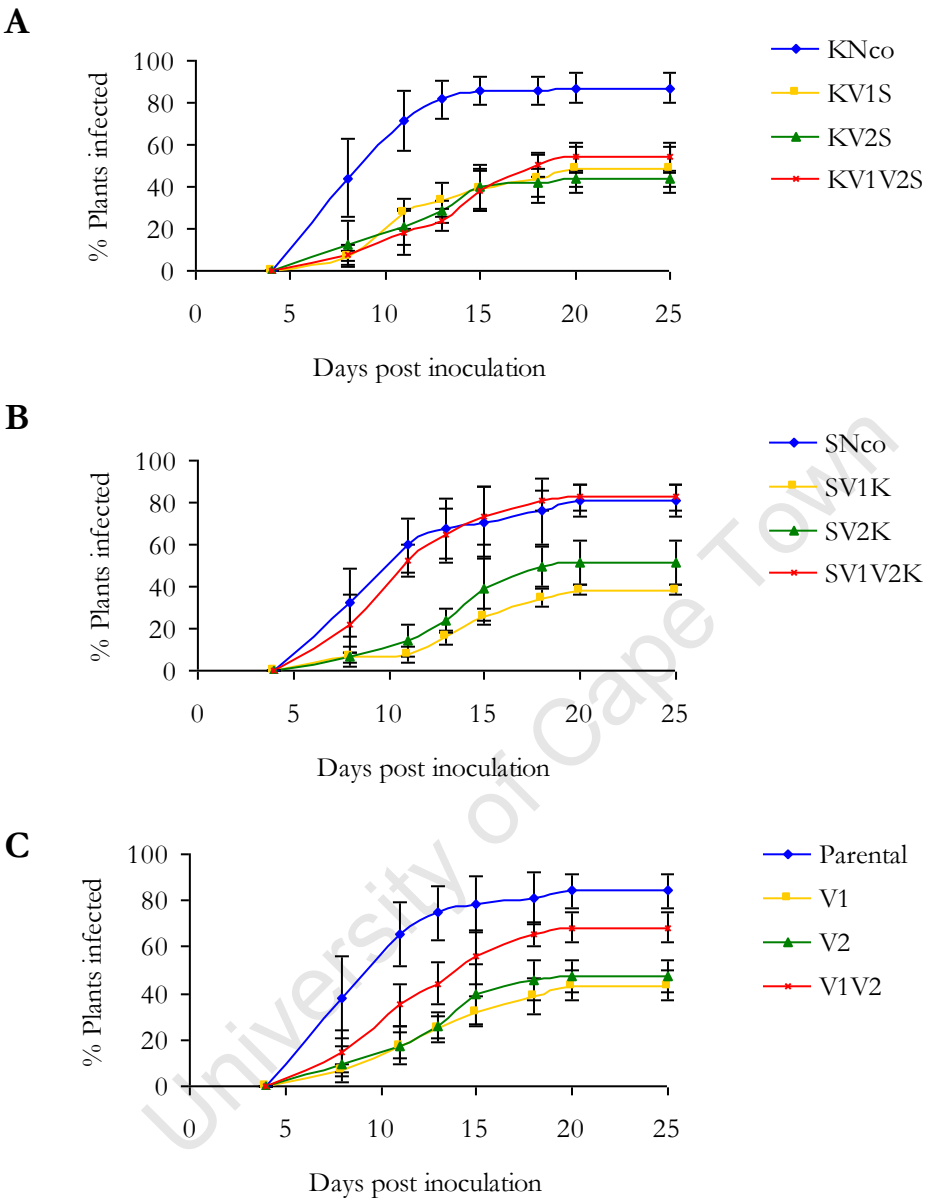


Figure 2-6. Average infection rates of chimaeras compared with parental viruses.

A – Chimaeras based on MSV-Kom; B – chimaeras based on MSV-Set; C – averaged, combined data for MSV-Kom and -Set based chimaeras. Error bars represent standard deviations (n ≥ 3).

Despite considerable differences in symptom severity, inoculation with all of the KNco-based chimaeras gave rise to symptomatic plants at similar rates, which were much slower than that of either parental virus. Accordingly, these chimaeras

infected a significantly smaller percentage of plants over a twenty day period than did either parent: KV1S, 48% ± 11%; KV2S, 44% ± 2%; and KV1V2S, 54% ± 7%. No new infections were observed later than 25 dpi (data not shown).

Compared with the KNco-based chimaeras, infection rates among SNco-based viruses were more distinct: while SV1K and SV2K both showed reductions in infectivity similar to those seen in KNco-based chimaeras, plants inoculated with SV1V2K became symptomatic at a similar rate to those inoculated with SNco. Of all the viruses in this study, SV1K infected the lowest percentage of plants (38% ± 2%), while SV2K appeared to be slightly more infectious (51% ± 12% of plants infected).

The fitness of each of the chimaeras and their parental viruses is summarised in Figure 2-7~~Figure 2-7~~, which shows the average area under the disease progress curve (AUDPC) and symptom severity for each agroinfectious construct. The chimaera comprising both MSV-Kom V1 and V2 ORFs exchanged into the MSV-Set genome showed the highest AUDPC of all the chimaeric constructs, achieving 80% and 85% of the AUDPC figures of KNco and SNco respectively. The remaining recombinant viruses all showed large reductions in infectivity, resulting in AUDPC figures less than half that of either parent. Symptom severity followed a similar pattern, except that KV1S appeared to be relatively more virulent and KV2S relatively less virulent than the infectivity data would suggest.

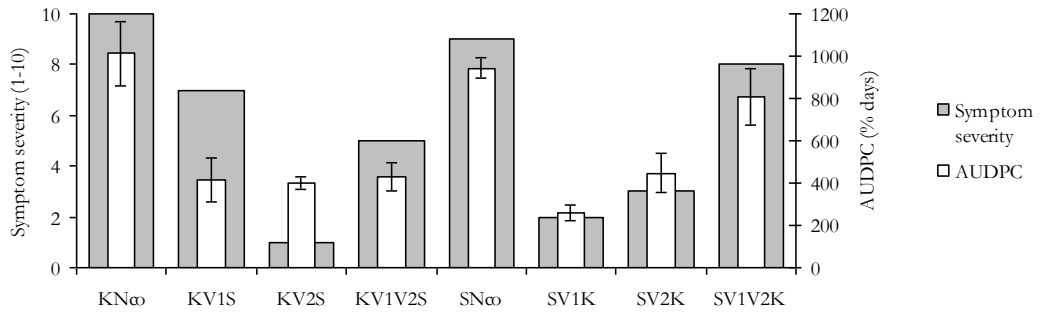


Figure 2-7. Average area under the disease progress curve and symptom severity for each chimaera, compared with the parental viruses KNco and SNco.

Error bars represent standard deviations (n ≥ 3). Symptom severity and AUDPC are positively correlated (R^2=0.75; P=0.005).

Discussion

MSV-Kom/MSV-Set chimaeric viruses are infectious

Few directed mutagenesis studies of mastreviruses have been reported, and of those most have focussed on knocking out entire genes with the aim of establishing their functions (Boulton *et al.*, 1989; Lazarowitz *et al.*, 1989; Liu *et al.*, 1998a). Where genetic variants of MSV have been compared, relatively small differences – such as single nucleotide substitutions – have often been found to be responsible for rather large phenotypic disparities (Boulton *et al.*, 1991a; Boulton *et al.*, 1991b; Martin *et al.*, 2001; Schnippenkoetter *et al.*, 2001b; Liu *et al.*, 2001b; Shepherd *et al.*, 2005). Considering the apparent sensitivity of MSV to mutation and the inability of similar BeYDV/MSV chimaeras to produce systemic infections (Liu *et al.*, 1999b) it may seem surprising that all the MSV-Kom/MSV-Set chimaeras described here are infectious. However, it should be borne in mind that directed mutagenesis studies are usually aimed at interrogating sequences suspected or known to be functionally critical, and neutral mutations are unlikely to be specifically reported because they are not generally considered interesting. Moreover, while numerous, the point mutations made in this study are of a special type – they comprise a set of mutations

known to function well together within the context of the original parental viruses. That is to say, because entire homologous ORFs were exchanged, no intra-ORF or intra-protein interactions were disrupted in the chimaeras.

Determinants of chlorotic severity

In view of the number of known – as well as the many likely but as yet unknown – *trans*-acting mechanisms engaged in functions such as gene regulation, virus replication, virus movement, virus-host interactions, *et cetera*, it is unsurprising that simple correlations between genotype and phenotype were not observed among the chimaeras described here: neither V1 nor V2 ORF, nor both together, could be considered wholly responsible for an MSV-Kom- or Set-like phenotype. However, the data do suggest that the V1 ORF was the determinant of the severity of chlorosis, with the MSV-Set V1 region inducing whiter chlorotic streaks than that of MSV-Kom (Figure 2-4~~Figure 2-4~~, Figure 2-5~~Figure 2-5~~, and Table 2-1~~Table 2-1~~). The V1 ORF not only encodes the MP, but also comprises an intron (Figure 2-2~~Figure 2-2~~) which is thought to affect V2 expression levels (Wright *et al.*, 1997) so it is possible that variations in chlorosis were mediated via differences in V2 expression.

Since the underlying causes of the chlorosis seen in MSV-infected tissue are not known, the significance of the varying degrees of chlorosis noted here is not obvious. It has been shown that chlorosis occurs only in infected cells (Lucy *et al.*, 1996) so it seems evident that the causal link between virus and chlorosis is fairly direct. One possibility is that chlorosis arises from the simple toxicity of one or more viral gene products. It is known that the MSV MP seems sufficiently toxic in *E. coli* to require strict control of expression for the generation of stable MP-expressing recombinants (personal observation, and personal communication from M.I. Boulton). Expression of geminivirus MP in transgenic plants can negatively affect plant development, necessitating the use of defective *MP* transgenes to

regenerate healthy plants (Hou *et al.*, 2000; Covey & Al-Kaff, 2000). In contrast, geminivirus CPs have readily been over-expressed both in transgenic plants (TYLCV, Kunik *et al.*, 1994) and in *E. coli* (MSV, Liu *et al.*, 1997a) with no apparent adverse effects and is therefore unlikely to be inherently toxic. Thus, one hypothesis is that MP causes chlorosis as a result of its toxicity, and that the MSV-Set MP is inherently more toxic than that of MSV-Kom in sweetcorn. Alternatively, it is possible that the MSV-Set and -Kom MPs are similarly toxic, but that the MSV-Set V1 ORF facilitates higher concentrations of MP.

A second hypothesis is that the MSV MP and/or CP modulate a hypersensitive response (van Wezel *et al.*, 2002; Hong *et al.*, 2003; HR reviewed by Kiraly *et al.*, 2007) or other innate defence pathway in infected cells, resulting in chlorosis. Mastrevirus CPs have detectable homology to begomovirus nuclear shuttle proteins (NSPs; Rybicki, 1994), and NSPs have been shown to elicit the HR (Garrido-Ramirez *et al.*, 2000) and to interact specifically with membrane-localised receptor-like kinases that are likely to play a role in defence responses (Fontes *et al.*, 2004). It would then follow that the MSV-Set MP is a more potent elicitor – or attains higher concentrations – than that of MSV-Kom; or the MSV-Kom MP is a more effective inhibitor of the defence response; or the V1 ORF influences CP levels, which in turn modulates the hypothetical defence response.

V1 and V2 interact specifically to facilitate infection and symptom development

Symptom severity and infection efficiency were roughly correlated (R^2=0.75, P=0.005; Figure 2-7~~Figure 2-7~~) although KV1S displayed relatively severe symptoms in relation to its infection efficiency, and KV2S displayed comparatively mild symptoms. In both parental backgrounds, the exchange of paired V1 and V2 ORFs rescued much of the loss of fitness seen in single-ORF exchanges: KV1V2S was considerably more infectious and pathogenic than KV2S; similarly, SV1V2K was

almost as infectious and virulent as SNco, whereas both SV1K and SV2K were drastically compromised in both respects.

These observations provide strong evidence in support of the importance of specific V1-V2 interactions in natural infections of maize. Specific binding of MP and CP has been demonstrated *in vitro* and *in vivo* (Liu *et al.*, 2001a) and some progress has been made in drawing parallels between the mechanisms underlying MSV cell-to-cell movement and the rather more developed models of movement in begomoviruses: MSV CP is localised to the nucleus and facilitates nuclear transport of viral DNA (Liu *et al.*, 1999a), which may be analogous to the nuclear localisation and/or shuttling functions performed by begomovirus CPs and/or NSPs (Lazarowitz & Beachy, 1999; Gafni & Epel, 2002); and MSV MP is localised at the cell periphery and binds to CP (Kotlizky *et al.*, 2000), which is reminiscent of at least some begomovirus MPs that have been shown to associate with plasma membranes and cell walls (Pascal *et al.*, 1993; von Arnim *et al.*, 1993) and to co-operate with NSP in moving viral DNA out of the nucleus to adjacent cells (Sanderfoot & Lazarowitz, 1995; Sanderfoot & Lazarowitz, 1996; Ward *et al.*, 1997). As others have noted, the roles that CP, NSP, and MP play in intra- and intercellular movement seem to differ somewhat among various geminiviruses (Gafni & Epel, 2002; Hehnle *et al.*, 2004) and a detailed model of mastrevirus movement has yet to be elucidated.

The limitations of mutation studies

Deleting or substituting entire genetic elements provides an intuitive and often technically simple starting point for investigating the relationship between genotype and phenotype. However, the results seldom shed much light on the detailed molecular basis of the phenotype(s) in question. First, it is often difficult or impossible to define precisely the bounds of the genetic elements that should be exchanged – amino acids distant from one another in primary sequence may lie adjacent in the folded protein, and regulatory elements are frequently far removed

from the ORFs whose expression they govern. Second, many biological molecules are involved in multiple interactions with a number of partners, and these interactions may be wholly or partly unknown, which complicates the interpretation of results considerably. Thus, studies like this one often yield more questions than answers, and complementary experiments are required to address specific biochemical questions in order to provide a full understanding of the expression of the phenotype under consideration. Moreover, care should be exercised in interpreting the results of such mutational investigations, because any evidence thus derived is inherently indirect, implied, and circumstantial, rather than explicit. In this study reciprocal chimaeras were used in an attempt to address, at least to some extent, these uncertainties.

Modularity of genetic elements

Although the idea of the modularity of genetic elements is inherent in the traditional concept of the gene, this notion of neatly delineated, modular genes has become increasingly blurred by the discovery of the myriad complex interactions governing gene expression and protein function. Thus it has become clear that relatively few genes act independently, while some phenotypic characters arise from intricate webs of highly specific interactions between numerous distinct genetic sequence elements. One might imagine that the structures of these genetic interaction networks define the boundaries of functional genetic modules, which may range in size from a few nucleotides in the case of some regulatory sequences to many megabases in the case of an entire genome.

This study provides some interesting perspectives on the varying degrees of modularity among the genetic regions studied here – V1, V2, V1-V2, and the remainder of the MSV genome. The results imply that V1 is modular with respect to the degree of chlorosis elicited in infected tissues, but not with respect to infection efficiency or chlorotic area. Similarly, exchanging V2 alone was insufficient to

maintain high infection rates or extensive virus movement. In contrast, the V1-V2 cassette behaved in a far more modular fashion, in that exchanging this region disrupted virus infectivity and *in planta* movement to a relatively small degree; it follows that the remainder of the genome reflected the same degree of modularity.

In demonstrating the viability of MSV chimaeras comprising the two most distinct strains known to date, the work presented here pointed the way towards the construction of additional MSV recombinant genomes. Martin *et al.* (2002) subsequently conducted an exhaustive study of pathogenicity determinants using MSV chimaeras, and together we later used both sets of virus constructs in an analysis of MSV genetic modularity and its potential role in natural recombination and evolution (Martin *et al.*, 2005a).

Using chimaeras to investigate virus evolution

The chimaeric constructs presented here offered obvious and intriguing opportunities for experimental investigations of mastrevirus evolution. First, a set of six viable, mutant viruses had been created and characterised – each virus comprised a relatively large number of known mutations in comparison with its wild-type parental viruses, and the fitness of each virus was clearly compromised to a different extent. If passaged *in planta* for sufficient periods of time, would these viruses evolve and regain a portion of their wild-type parental virulence? This question is addressed in the following chapter.

Second, reciprocal chimaeric genomes present the unique possibility of generating pairs of mirror-image mutants which, between them, comprise all the genetic information of both wild-type parental viruses. It has long been established that geminivirus genomes readily recombine during co-infections (Boulton *et al.*, 1989) so it seemed likely that co-infections of pairs of reciprocal chimaeras that were significantly functionally compromised would yield recombinants in which a degree

of wild-type fitness was regained. This reasoning formed the basis for the investigation presented in Chapter 4.

CHAPTER 3: HIGH MUTATION RATES AND BIASED MUTATION FREQUENCIES DURING THE DIVERSIFICATION OF CLONAL MSV POPULATIONS

ABSTRACT

While the majority of arising human infectious diseases are caused by RNA viruses, the extremely diverse ssDNA geminiviruses represent the greatest viral threat to edible crops. Nevertheless, most research on plant virus evolution has focussed on RNA viruses, which are thought to have high mutation rates due to their dependence on error-prone RNA polymerases. Animal and plant RNA viruses have been shown to evolve at rates of ~10^{-4} substitutions per site per year (subs/site/year). Consistent with an hypothesis that polymerase fidelity influences evolution rate, dsDNA viruses seem to display much lower substitution rates of ~10^{-8} subs/site/year. In contradiction to this hypothesis, substitution rates closer to those of RNA viruses have been reported for the small ssDNA circoviruses, parvoviruses, and geminiviruses. To investigate substitution rates and patterns in the geminivirus maize streak virus (MSV), we conducted evolution experiments using virus populations initiated from well-characterised clones and maintained *in planta* for between one and four years. In one experiment, we passaged chimaeric viruses representing a range of compromised fitness, and their constituent wild-type parental viruses, serially through 12 generations of sweetcorn host plants during a period of one year before cloning and sequencing full-length progeny virus genomes. We hypothesised that selection for restored epistatic interactions in artificial chimaeric viruses might give rise to obvious compensatory or back-mutations. In a second experiment, full-length MSV genomes were isolated from a single sugarcane plant and sequenced, four years after it was infected using a cloned virus inoculum. While no signs of strong positive selection were observed, we report high mutation rates – 7.4 x 10^{-4} subs/site/year for wild-type and chimaeric viruses in sweetcorn, and 1.6 x 10^{-3} subs/site/year for MSV in sugarcane – and biased mutation frequencies in line with those reported elsewhere for geminiviruses.

Introduction

Although there are many examples of extant host-pathogen interactions with ancient histories, the *status quo* in the relationship between a pathogen and its host represents a mere snapshot in the course of a dynamic process of co-adaptation. Thus, a more complete understanding of the mechanisms of adaptation and the rates at which they act is generally considered indispensable to a full appreciation of phenomena such as disease epidemics, the emergence of new diseases, and the appearance of new pathogen strains capable of overcoming control measures such as immunisation, drugs or resistant host genotypes. The numerous factors thought to influence the evolution of plant viruses – from the mutation, recombination, and reassortment rates that drive their diversification to selection processes such as migration and cyclical transmission- and climate-related population bottlenecks – have been extensively studied, and are reviewed elsewhere (Roossinck, 1997; Gibbs *et al.*, 1999; Garcia-Arenal *et al.*, 2001; Garcia-Arenal *et al.*, 2003). The primary conclusion of this work is that, as with all other life on earth, the evolution of plant viruses is profoundly dependent on the relentless generation and maintenance of sufficient genetic diversity to feed the ongoing processes of natural selection.

Given a reasonable number of genetic variants upon which to act, homologous and/or non-homologous mechanisms of recombination may produce an astounding variety of novel sequences, but point mutation is usually considered to be the ultimate fountainhead of genetic diversity. Mutation is essentially a serial one-step-at-a-time process and, given the unimaginably large unexplored potential sequence spaces represented even by small genomes, evolution via mutation alone can be a very slow process. However, probably because viruses have relatively small genomes, short generation times, enormous population sizes and – at least for RNA and some ssDNA viruses – high mutation rates, they can evolve relatively rapidly (Roossinck, 1997; Domingo & Holland, 1997; Garcia-Arenal *et al.*, 2001; Elena & Sanjuan, 2007).

Most research on plant virus evolution has focussed on RNA viruses (Roossinck, 1997; Garcia-Arenal *et al.*, 2001) which are generally considered to be subject to relatively high rates of mutation due to their dependence on error-prone RNA polymerases. Mammalian RNA viruses have been shown to evolve at rates of 10^{-3} to 10^{-5} substitutions per site per year (subs/site/year; Jenkins *et al.*, 2002) and plant RNA viruses exhibit similar evolution rates (Schneider & Roossinck, 2000; Schneider & Roossinck, 2001). In contrast – and consistent with the hypothesis that polymerase fidelity influences mutation rate – dsDNA papillomaviruses and polyomaviruses seem to display much lower substitution rates of ~10^{-8} to ~10^{-9} subs/site/year (Bernard, 1994; McGeoch & Gatherer, 2005). Intriguingly, and possibly in contradiction to the premise that polymerase fidelity is the major universal determinant of mutation rates, figures closer to those of RNA viruses have been reported for the small ssDNA circoviruses (Gallian *et al.*, 2002; Biagini, 2004) and parvoviruses (Shackelton *et al.*, 2005; Lopez-Bueno *et al.*, 2006; Shackelton & Holmes, 2006). Indeed, Drake *et al.* (1998; 1999) have suggested that the spontaneous mutation rate per genome per replication cycle (μ_g) is approximately constant within major groups of organisms: RNA viruses, $\mu_g \approx 1$; retroelements, $\mu_g \approx 0.2$; DNA viruses and cellular microbes, $\mu_g \approx 0.0034$; higher eukaryotes, $\mu_g \approx 0.01$.

While the ssDNA geminiviruses are generally acknowledged to represent extremely important threats to commercial agriculture and/or basic subsistence, and the threat of emergent geminivirus diseases has been widely noted (Anderson & Morales, 1994; Moriones & Navas-Castillo, 2000; Morales & Anderson, 2001; Mansoor *et al.*, 2003b; Rojas *et al.*, 2005) our understanding of geminivirus evolution remains fairly limited. Extensive regional diversity has been reported for both begomoviruses (Stenger & McMahon, 1997; Sanz *et al.*, 1999; Ndunguru *et al.*, 2005; Ge *et al.*, 2007; Lefeuvre *et al.*, 2007) and mastreviruses (Martin *et al.*, 2001; Varsani *et al.*, 2008) but the timescales over which this diversification occurred are uncertain.

Computational analyses of genome sequences have led a number of people to note the important role that recombination has played in the evolution of geminiviruses, but until very recently insufficient information was available to allow useful estimations of substitution rates (Duffy & Holmes, 2008). Using Bayesian coalescent methods and a dataset of tomato yellow leaf curl virus (TYLCV) sequences from viruses isolated between 1988 and 2006, Duffy and Holmes (2008) calculated an average evolution rate of 2.88×10^{-4} subs/site/year. Only two publications describing experimental investigations of geminivirus evolution rates appear in the literature. The first attempted to limit the diversity of the starting population by initiating infection via a single leafhopper and restricted acquisition and feeding times, and furnished mutation frequencies of the order of 10^{-3} to 10^{-4} subs/site for MSV that were generated over periods of up to 4 years (Isnard *et al.*, 1998). The second is a recent study of the begomovirus tomato yellow leaf curl China virus (TYLCCNV) in which mutation frequencies around 10^{-4} subs/site were observed within sixty days of agro-inoculation with infectious cloned DNA (Ge *et al.*, 2007). Two additional reports of high-frequency reversion of non-lethal deleterious mutations in the Rep proteins of a mastrevirus (Shepherd *et al.*, 2005) and begomoviruses (Arguello-Astorga *et al.*, 2007) highlight the ability of geminiviruses to adapt rapidly in at least some circumstances.

Mutation frequencies and/or rates obtained from any population are necessarily the product of both mutation as well as selection; strictly lethal mutations are most unlikely to be sampled, while very beneficial nucleotide substitutions would be expected to become well represented in a population given appropriate time, population size, and the absence of severe bottleneck events (Roossinck, 1997). Experimental investigations of both viral and cellular microbial (reviewed in Elena & Lenski, 2003; and in Elena & Sanjuan, 2007) and enzyme (reviewed in Poelwijk *et al.*, 2007) evolution provide opportunities to control and dissect at least some of the numerous factors at play, and have provided useful empirical perspectives on the

popular theoretical concept of fitness landscapes (Wright, 1932; Ao, 2005; Poelwijk *et al.*, 2007; Elena & Sanjuan, 2007). For example, a number of experiments have revealed a rapid initial rate of adaptation of organisms to new environmental conditions, followed by a gradual deceleration, consistent with the idea of scaling an adaptive peak or plateau (Lenski *et al.*, 1991; Lenski & Travisano, 1994; Novella *et al.*, 1995; Bull *et al.*, 1997; Elena *et al.*, 1998; Cooper & Lenski, 2000; de Visser & Lenski, 2002).

Upon consideration of the results presented in the preceding chapter, it was clear that we were in possession of a series of infectious mutant viruses exhibiting broad ranges of reduced fitness levels and sequence similarity, relative to their wild-type parental viruses. In addition, we had a population of MSV-Tas that had been initiated from a clone and then maintained in a single sugarcane plant for four years. MSV-Tas was originally isolated from wheat (Willment *et al.*, 2002) so we assumed that the initial virus was maladapted to sugarcane. Together, these materials constituted an opportunity to conduct controlled MSV evolution experiments: we wanted to test whether viruses displaced to varying extents – and in different ways – from the assumed fitness peaks occupied by their wild-type parents would evolve differently due to different selection pressures. We also hypothesised that selection for re-optimised epistatic interactions in chimaeric viruses might give rise to obvious compensatory or back-mutations. Similarly, full-length clones of MSV-Tas were re-isolated from the sugarcane plant and sequenced. While no signs of strong positive selection were observed, we report high mutation rates as well as biased mutation frequencies, in line with those reported elsewhere for geminiviruses.

Methods and materials

Virus isolates, plasmids, bacterial strains, maize genotypes, and leafhoppers

The agroinfectious plasmids pKNcoB, KV1SB, pKV2SB, pKV1V2SB, pSNcoB, pSV1KB, pSV2KB, and pSV1V2KB are described in Chapter 2 and the viruses they

give rise to are referred to as KNco, KV1S, KV2S, KV1V2S, SNco, SV1K, SV2K, and SV1V2K, respectively. The plasmids pKEP170 and pSNco are described in Chapter 2. *A. tumefaciens* C58C1 [pMP90] was used to deliver viral DNA to *Zea mays* cv. Jubilee (sweetcorn) seedlings by agroinoculation as described in Chapter 2. The *Cicadulina mbila* leafhoppers used to transmit viruses were obtained from a colony originally established by M.B. von Wechmar in 1978 and maintained in the Microbiology/Molecular and Cell Biology Department, University of Cape Town, South Africa.

The sugarcane plant infected with MSV-Tas (Willment *et al.*, 2002) was a kind donation from J. Willment (then Microbiology Department, UCT). At the time of sampling, the plant had been maintained (25 °C under Grolux™ fluorescent lights with 16 hours of light per day) in an infected state by D. Solomons in an enclosed plant room within a building for a period of four years.

Leafhopper cages and virus transmission by leafhopper

Leafhoppers and infected plants were kept at approximately 21 °C and were isolated in purpose-built cages (410mm x 410mm x 710mm, w x d x h) constructed from 6mm welded stainless steel rod clad in fine mesh curtain netting. The cages were maintained next to windows in the laboratory; thus lighting consisted of indirect natural light augmented by Grolux™ fluorescent tubes for 12 hours per day. Each cage contained plants infected with a single virus. Initially three 25-day-old plants infected by agroinoculation with each of the eight viruses KNco, KV1S, KV2S, KV1V2S, SNco, SV1K, SV2K, and SV1V2K were placed in separate isolation cages with c.a. 100 adult leafhoppers and three uninfected 8-day-old maize seedlings per cage. Thereafter the leafhopper population in each cage was monitored and culled as necessary using sticky traps, to prevent excessive feeding damage to plants. When symptoms became visible on new plants the older plants were removed from the cage and replaced with seedlings; this cycle was repeated approximately monthly.

The entire experiment lasted a period of 12 months, during which the viruses were passaged through 12 generations of maize plants.

Extraction, purification and cloning of viral DNA

Replicative form, double-stranded virus DNA was extracted from plants as described by Palmer *et al.* (1998). The isolated DNA was cloned into the *Bam*HI site of pUC18 using standard techniques (Sambrook *et al.*, 1989).

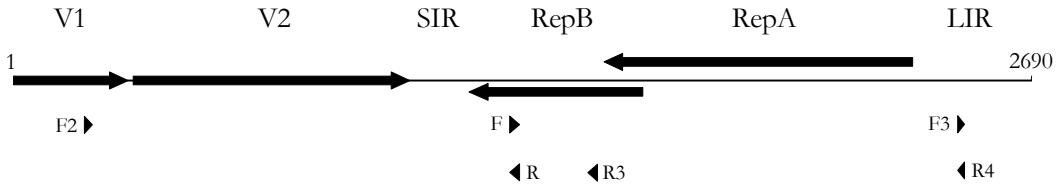
Using universal M13 sequencing primers (M13-Fwd, 5' – CCCAGTCACGACG TTGTAAAACG – 3'; M13-Rev, 5' – AGCGGATAACAATTTCACACAGG – 3') plasmid inserts were amplified by PCR directly from colonies (Gussow & Clackson, 1989), and agarose gel electrophoresis was used to determine the sizes of the amplified products. Products of 2.7 kbp were subsequently digested with a variety of restriction enzymes to confirm that the clones contained full-length mastrevirus genomes.

DNA sequencing and sequence analysis

DNA for sequencing was purified from *E. coli* cultures using the QIAprep Spin Miniprep Kit (Qiagen). Sequencing was performed by the DNA Sequencing Service (Molecular and Cell Biology Department, UCT). In addition to the universal M13 vector sequencing primers described above, six mastrevirus-specific primers were designed to allow sequencing of full-length clones without the need for subcloning (Table 3-1~~Table 3-1~~). These primers were designed to anneal to sequences that are widely conserved in members of the *Mastrevirus* genus, and consequently the same set of primers was used for sequencing MSV-Kom, -Set, and -Tas in this chapter and MSV-Mat and -VW in Chapter 4, as well as all chimaeric or recombinant genomes comprising elements of these viruses. Sequences were assembled and aligned using DNAMan5.2.9 and trace files were viewed using BioEdit6.0.7 (Isis Pharmaceuticals, Inc.). All mutations were verified by at least two sequencing runs and by

examination of the sequencing chromatograph files. All parental virus clones were re-sequenced in both directions.

Table 3-1. "Universal" MSV sequencing primers used in this study.



Primer	Position †	Primer sequence (5' – 3')
MSV-F	1337	CTGCCCTGGA GTCATTTCCT TCATCC
MSV-F2	209	TCCACGGAGG AGCTGATATT TGG
MSV-F3	2514	GACCGGGCCG GCCCAGCAGG
MSV-R	1312	GGATGAAGGA AATGACTCCA GGGCAG
MSV-R3	1517	TGGCAAAATA ATGTTGATTG GTCTTC
MSV-R4	2495	CCTGCTGGGC CGGCCCGGTC

† Position of the 3' end of the primer on MSV-Mat sequence, relative to the conserved, unique *Bam*HI site at the V1 start codon.

Point mutation diagrams were constructed using the column chart drawing function of Microsoft Excel. Briefly, macros were written for importing multiple sequence alignments into Excel, with sequences in columns and aligned residues in adjacent cells. In the case of KNco/SNco chimaeras, logical functions were used to generate arbitrary integer values according to whether the isolate sequence at a particular position corresponded with SNco (value = -1); KNco and SNco (value = 0); KNco (value = 1); or, neither KNco nor SNco (i.e. novel mutation; value = 2). Similarly, the sequences of MSV-Tas isolates were compared to that of the original parental clone such that positions containing mutations were assigned a value of 2. Column charts were then drawn using each nucleotide position in turn as a separate column with a column length equal to: 0 (non-informative site); 1 or -1 (informative site and

identity with relevant parental sequence); or -2 (novel mutation not seen in parental sequences).

Additional sequence analyses

The expected frequency for a given substitution of nt. X for nt. Y ($f^{E}_{X\rightarrow Y}$) was calculated assuming all substitution types were equally likely, as follows:

$$f^{E}_{X\rightarrow Y} = (P_X \times M)/3$$

where

P_X is the fractional proportion of nt. X in the parental sequence, and

M is the total number of observed mutations.

Expected transition and transversion frequencies were calculated by summing the expected frequencies of the relevant substitutions.

The total numbers of synonymous sites ($\sum$[S sites]) and non-synonymous sites ($\sum$[N sites]) in all the coding regions of each parental virus (KNco, SNco, and MSV-Tas) were obtained by SLAC analysis (Kosakovsky Pond & Frost, 2005b) carried out on the Datamonkey web-server (http://www.datamonkey.org/; Kosakovsky Pond & Frost, 2005a). The number of non-synonymous mutations per non-synonymous site (dN) was calculated as follows:

$$dN = N/\sum[\text{N sites}]$$

where

N is the observed number of non-synonymous mutations, and

$\sum$[N sites] is the total number of non-synonymous sites.

Similarly, the number of synonymous mutations per synonymous site is given by:

$$dS = S/\sum[S\ sites]$$

where

S is the number of synonymous mutations, and

$\sum[S\ sites]$ is the number of synonymous sites.

Expected frequencies of synonymous (S^E) and non-synonymous (N^E) mutations were calculated by multiplying the total number of substitutions by the ratio of synonymous sites to total sites, or non-synonymous sites to total sites, respectively:

$$S^E = (N+S) \times \sum[S\ sites] \ / \ (\sum[S\ sites] + \sum[N\ sites])$$

and

$$N^E = (N+S) \times (1 - (\sum[S\ sites] \ / \ (\sum[S\ sites] + \sum[N\ sites])))$$

where

N is the observed number of non-synonymous substitutions,

S is the observed number of synonymous substitutions,

$\sum[S\ sites]$ is the number of synonymous sites, and

$\sum[N\ sites]$ is the number of non-synonymous sites.

Results

MSV-Kom / MSV-Set parental and chimaeric viruses

With the intention of studying mutation rates and patterns in as many of the KNco/SNco chimaeras as possible, sweetcorn plants were initially inoculated with all eight virus constructs described in Chapter 2 – KNco, KV1S, KV2S, KV1V2S, SNco, SV1K, SV2K, and SV1V2K – and in each case they exhibited the expected, characteristic symptoms of infection. Three plants infected with each virus were used to initiate subsequent serial transmissions via leafhopper, with each

transmission lasting several days and involving tens of leafhoppers, with the intention of avoiding severe bottlenecks. All viruses were successfully transmitted through multiple passages via leafhopper, but KV2S and SV1K apparently failed to maintain infections somewhere between the fifth and eighth passages (i.e. between five and eight months into the experiment; data not shown). However, following twelve passages over a one-year period, no obvious changes in symptomatology were observed for any of the remaining passaged viruses (data not shown). At the end of the one-year period, viral genomes were successfully cloned from three of the remaining four chimaeric viruses, in addition to the two parental viruses, KNco and SNco: KV1S, KV1V2S, and SV2K. Full-length genomic sequences were obtained for two individual genomes from each of these viruses, except for KV1S, for which just one genome was sequenced.

Figure 3-1~~Figure 3-1~~ shows the positions of all of the mutations identified in the KNco/SNco viruses, while Table 3-2~~Table 3-2~~ details the nucleotide and protein sequence context and the specific sequence changes in each case. All of the genomes sequenced contained at least one mutation with respect to the original parental viruses; the most mutations in any single genome was four (E1-01, KNco; E2-01, KV1S) and fifteen unique mutations were identified in total. Both SNco genomes (E5-01 & E5-02) were identical, as were the two SV2K genomes (E7-01 & E7-02), resulting in seven unique full-length genome sequences. The two KNco clones had no mutations in common, whereas the KV1V2S clones shared one mutation, and had one unique mutation each. Of the fifteen polymorphisms, thirteen were unique with respect to both parents. Only two mutations (4 & 13 in Figure 3-1~~Figure 3-1~~) out of eight in the chimaeric viruses represented changes from one wild-type parental sequence to the other. The distribution of the mutations among the various genomic elements was as follows: V1, 2; V2, 9; SIR, 0; C2, 1; C1/C2, 1; LIR, 2. The average mutation rate for the nine MSV-Kom / MSV-Set genomes analysed here is 7.4×10^{-4} ($\pm 4.5 \times 10^{-4}$) subs/site over the one year period of the experiment.

Only one, single-base deletion – resulting in the premature termination of the C2 ORF in clone E1-02 by 284 nt. – was observed, and there were no insertions (Table 3-2~~Table 3-2~~). Where mutations occurred in sequence motifs or regions known or suspected to be associated with specific functions, these have been indicated, although it should be noted that relatively little is known about the sequence – structure – function relationships of the mastrevirus virion-sense genes and gene products. Most substitutions did not occur within any known functional sub-elements of the ORFs or intergenic regions, and most amino acid changes appeared to be relatively conservative. Notable exceptions are the replacement of the small, neutral proline with the relatively larger, polar glutamine in the putative GRAB-interaction domain of the RepA protein (by analogy with WDV; Xie *et al.*, 1999); and the disruption of the most distal of three potential C-sense TATA boxes.

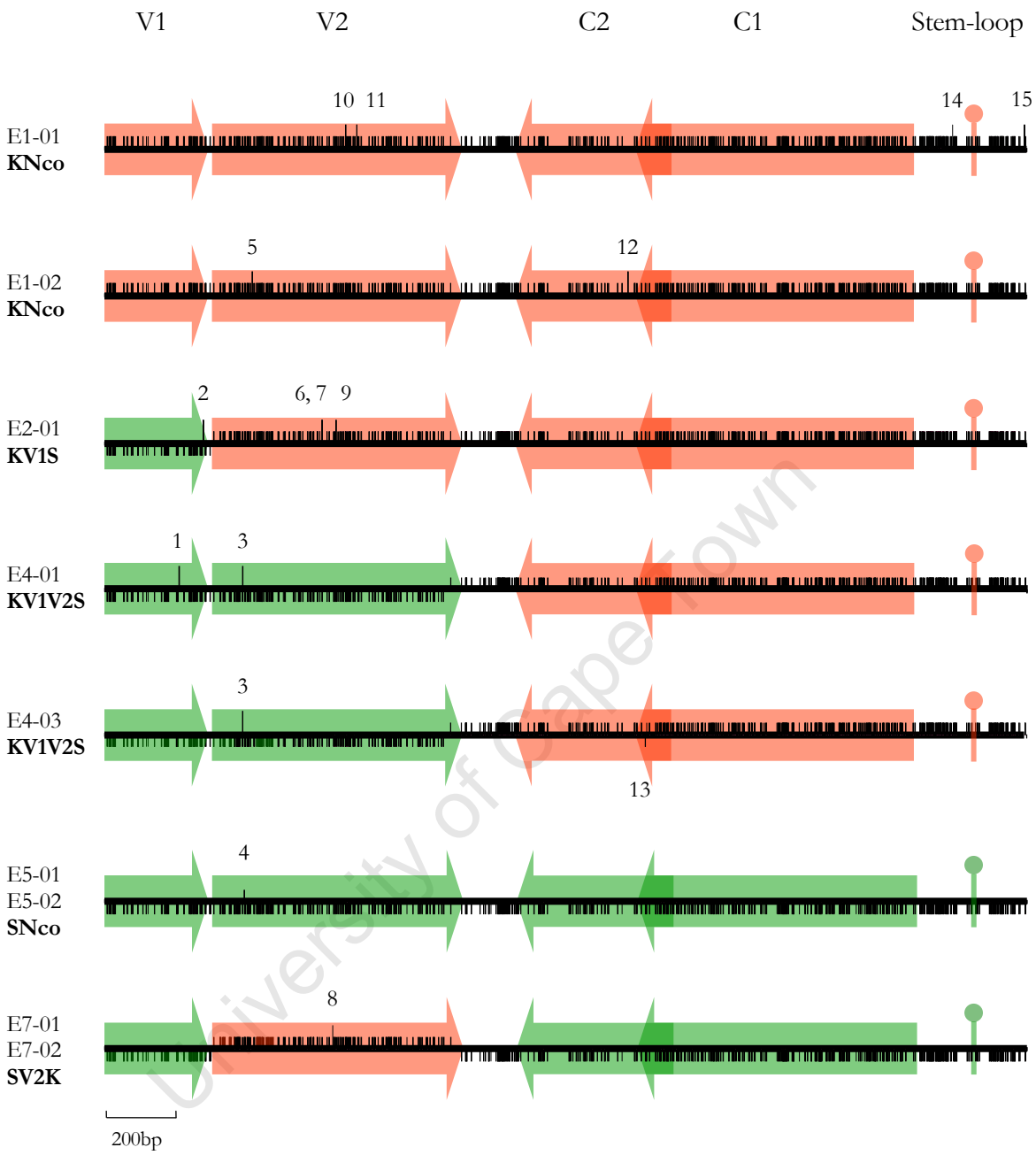


Figure 3-1. Mutations in MSV-Kom/-Set parental and chimaeric viruses.

Short vertical lines above or below the centre line indicate homology at informative sites to either KNco or SNco, respectively. Long vertical lines above the centre line represent positions not homologous to either KNco or SNco sequence. Mutations are numbered, and refer to those listed in Table 3-2~~Table 3-2~~. The positions of ORFs and the virion-strand replication origin stem-loop sequence are indicated in shaded red (KNco) or green (SNco). The diagrams are to scale.

Table 3-2. Mutations in MSV-Kom/MSV-Set chimaeras passaged in maize.

[a]	Nt [b]	**Parent** Clone(s)	Genomic Region (Protein)	Sequence changes	PAM250 [c]
1	221	**KV1V2S** E4-01	V1 (MP)	F G S E **A** V D R R TTTGGATCTGAAG**C**TGTGGATAGGAGG ↓ TTTGGATCTGAAG**T**TGTGGATAGGAGG F G S E **A** V D R R	2
2	292	**KV1S** E2-01	V1 (MP)	G P F V **P** G Q G * GGACCGTTCGTT**C**CAGGTCAGGGATAA ↓ GGACCGTTCGTT**G**CAGGTCAGGGATAA G P F V **A** G Q G *	1
3	407	**KV1V2S** E4-01 E4-07	V2 (CP) (DNA-binding region)	Q A K K **P** G G K V CAGGCGAAGAAGC**C**TGGGGGTAAGGTT ↓ CAGGCGAAGAAGC**A**TGGGGGTAAGGTT Q A K K **P** G G K V	6
4	412	**SNco** E5-01 E5-02	V2 (CP) (DNA-binding region)	A K K P **G** G K V E GCGAAGAAGCCTGG**G**[A1]GGTAAGGTTGAG ↓ GCGAAGAAGCCTGG**A**GGTAAGGTTGAG A K K P **G** G K V E	5
5	436	**KNco** E1-02	V2 (CP) (DNA-binding region)	D R P S **L** Q I Q T GATAGGCCATCCCT**A**CAAATCCAGACA ↓ GATAGGCCATCCCT**C**CAAATCCAGACA D R P S **L** Q I Q T	6
6, 7	640, 641	**KV1S** E2-01	V2 (CP) (DNA-binding region)	T G T G **V** **M** W L V Y ACCGGAACCGGTGT**AA**TGTGGCTGGTGTAT ↓↓ ACCGGAACCGGTGT**TG**TGTGGCTGGTGTAT T G T G **V** **V** W L V Y	4, 2
8	672	**SV2K** E7-01 E7-02	V2 (CP)	T T P G **G** Q A P T ACCACTCCCGGCG**G**ACAAGCTCCGACC ↓ ACCACTCCCGGCG**A**ACAAGCTCCGACC T T P G **G** Q A P T	5
9	682	**KV1S** E2-01	V2 (CP)	G G Q A **P** T P Q T GGCGGACAAGCTCC**G**ACCCCGCAAACT ↓ GGCGGACAAGCTCC**T**ACCCCGCAAACT G G Q A **P** T P Q T	6
10	710	**KNco** E1-01	V2 (CP)	F A Y P **D** T L K A TTTGCCTACCCT**G**ACACGCTAAAAGCG ↓ TTTGCCTACCCT**C**ACACGCTAAAAGCG F A Y P **H** T L K A	1

[a]	Nt [b]	**Parent** Clone(s)	Genomic Region (Protein)	Sequence changes	PAM250 [c]
11	743	**KNco** E1-01	V2 (CP)	A T W K **V** S R E L GCCACATGGAAA**G**TGAGCCGGGAGCTG ↓ GCCACATGGAAA**T**TGAGCCGGGAGCTG A T W K **L** S R E L	2
12	1540	**KNco** E1-02	C2 (Rep) Putative C1 polyadenylation signal (*myb*-like transactivation domain)	N Y W Q **N N V D W - - - *** AATTACTGGCAAA**A**TAATGTTGATTGG-N^{320}-TAA ↓ AATTACTGGCAAA**.**TAATGTTGATTGG-N^{36} -TAA N Y W Q **I M L I G - - ***	-8
13	1590	**KV1V2S** E4-03	C1 (RepA) (Putative GRAB binding domain)	RepA, 26 bp upstream of C1 STOP: A Q Q E **P** E N L L GGCCCAACAAGAAC**C**[A2]GGAAAATCTACT ↓ GGCCCAACAAGAAC**A**GGAAAATCTACT A Q Q E **Q**[A3] E N L L	0
			C1-C2 (Rep) (dNTP binding motif)	Rep, 57 bp downstream of splice junction: G P T R **T** G K S T GGCCCAACAAGAAC**C**GGAAAATCTACT ↓ GGCCCAACAAGAAC**A**GGAAAATCTACT G P T R **T** G K S T	3
14	2492	**KNco** E1-01	LIR C-sense TATA box	CTCCATTGTCTT**A**TAGTGGTTGTA ↓ CTCCATTGTCTT**G**TAGTGGTTGTA	
15	2704	**KNco** E1-01	LIR V1 START	M D P CAAGTGCGATTC**A**TTCATGGATCC ↓ CAAGTGCGATTC**G**TTCATGGATCC	

[a] Numbers correspond to those in Figure 3-1~~Figure 3-1~~; [b] nucleotide positions are given relative to the conserved *Bam*HI restriction site at the V1 start codon; [c] PAM250 scores ≥ 1 may indicate conservative amino acid changes. For each mutation, the following information is given on successive lines, from top to bottom: parental amino acid sequence (N→C); parental nucleotide sequence (5'→3', coding strand, or virion strand if not coding); arrow(s) indicating mutation(s); mutant nucleotide sequence; mutant amino acid sequence. Residues in red indicate changes from parental sequences; affected codons are underlined. Where functional motifs are known or suspected to form a part of the region shown, they are indicated by highlighting. Stop codons are indicated by *.

MSV-Tas maintained in sugarcane for four years

A single sugarcane plant was originally infected with MSV-Tas via agroinoculation with a sequenced clone, and had been maintained in a plant growth room for a period of four years before being generously donated to us by J. Willment. MSV-Tas was originally isolated from wheat, so it was expected that the virus might be sub-optimally adapted to sugarcane as a host plant. With this in mind, we decided to sequence the progeny of that initial clonal infection with the aim of investigating the extent and character of any evolution that may have taken place.

Two full-length genomes were cloned from this plant, and then sequenced (Figure 3-2~~Figure 3-2~~). It should be noted, however, that one genome (SC-E-02) was discovered to have suffered a significant deletion of sixteen nucleotides from the C1 ORF (event 19, Table 3-3~~Table 3-3~~), resulting in the loss of the intron acceptor site, and premature termination of C1 some thirty codons before the normal stop site. This was the only deletion in either clone, and no insertions were present. In all of the subsequent calculations, this deletion of sixteen nucleotides was counted as a single mutation event. A total of twenty-seven unique mutations were identified in both clones, with SC-E-02 having fifteen and SC-F-01 displaying twenty-one; the two clones shared nine mutations in common – five of them in the 3'-half of the V2 ORF (Figure 3-2~~Figure 3-2~~). The mutations were distributed among the major genomic regions as follows: V1, 2; V2, 9; SIR, 1; C2, 6; C1, 5; LIR, 4. Clone SC-F-01 displayed a mutation rate of 5.2×10^{-3} subs/site over the four year period, while SC-E-02 had accumulated 7.5×10^{-3} subs/site. This yields an average mutation rate of 6.3×10^{-3} subs/site over four years, or 1.6×10^{-3} subs/site/year.

Of the twenty-seven mutations, five were in the non-coding SIR and LIR regions, and four of these were common to both genomes. Using a PAM250 (Dayhoff *et al.*, 1978) substitution score of one or more as a guide, fifteen of the remaining mutations in coding regions may be considered conservative. Genome SC-F-01

contained eight conservative coding substitutions, while the SC-E-02 sequence revealed twelve; the two genomes shared five of these in common. In contrast, none of the relatively radical coding-region mutations were shared between the two genomes sequenced here. Clone SC-F-01 had three non-conservative coding mutations: S48I in the MP, a change from a very small polar residue to a large, aliphatic, hydrophobic residue; P322L in Rep, constituting a change from a small amino acid to a large, aliphatic, hydrophobic amino acid; and K292T in Rep, which represents a change from a large, positively charged residue to a small, polar amino acid. Clone SC-E-02 had four non-conservative coding mutations, one of which was the 16 nt deletion described above. A second mutation would have resulted in a premature stop codon at codon position 257 in *Rep,* but it should be noted that for this particular genome it is unlikely that the full-length *Rep* transcript would be produced, considering the deletion of the intron splice acceptor site described above. The two remaining non-conservative mutations in clone SC-E-02 resulted in G9V in the CP – a larger, aliphatic, hydrophobic, residue replacing a very small, hydrophobic residue – and H12Y in the Rep and RepA proteins, which constitutes the substitution of one hydrophobic, aromatic residue for another.

Where mutations occurred in known or suspected regions of functional significance, these have been indicated in Table 3-3~~Table 3-3~~. Apart from the 16 nt. deletion (event 19 in Table 3-3~~Table 3-3~~) and the downstream nonsense mutation (event 17) in the Rep sequences of clone SCE-02 (described above), they were as follows: a T→A transversion in a T-tract within the SC-E-02 V1 intron (event 1); G9V within the eleven-amino acid spacer in the SC-E-02 CP bipartite nuclear localisation signal (event 3); D253H at the start of the myb-like transactivation domain of the SC-E-02 Rep protein; and a C→T transition within the right-hand side of the stem-sequence that forms the stem-loop structure at the virion-sense origin of replication. It is noteworthy that this latter mutation was not accompanied by a complementary mutation in the other half of the stem sequence (not shown).

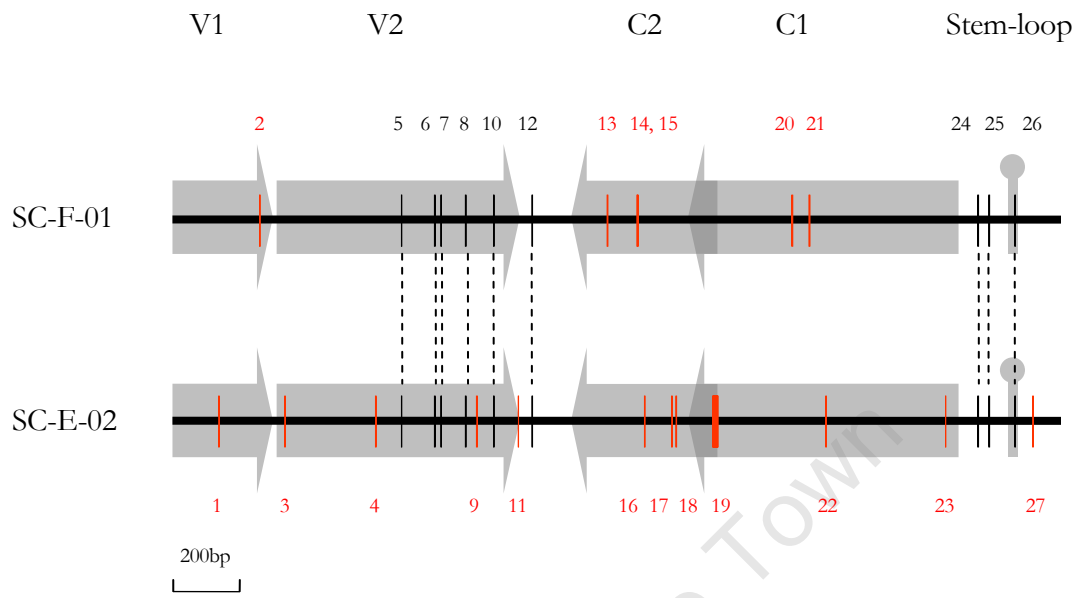


Figure 3-2. Mutations in MSV-Tas passaged in sugarcane.

The positions of ORFs and the virion-strand replication origin stem-loop sequence are indicated in shaded grey. Vertical lines represent nucleotide positions that differ from the parental MSV-Tas DNA sequence. Black lines (connected by broken lines) and numbers indicate mutations shared by both clones. Red lines and numbers distinguish mutations unique to either clone. Mutations are numbered according to their positions in Table 3-3~~Table 3-3~~. The diagrams are to scale.

Table 3-3 Mutations in MSV-Tas passaged in sugarcane

[a]	Pos[b]	Clone	Genomic Region	Sequence changes	PAM250[c]
1	139 (48)	SC-E-02	V1 (MP) V1 intron T-tracts Branch point (Putative transmembrane domain)	L L Y L **W** V L R D CTGCTTTACCTTTGGGTGCTGAGAGAT ↓ CTGCTTTACCTTAGGGTGCTGAGAGAT L L Y L **R** V L R D	2
2	263 (89)	SC-F-01	V1 (MP)	Q A P P **S** Q G N P CAGGCACCACCAAGTCAGGGGAATCCC ↓ CAGGCACCACCAATTCAGGGGAATCCC Q A P P **I** Q G N P	-1
3	339 (9)	SC-E-02	V2 (CP) (Bipartite nuclear localisation signal) (DNA-binding region)	K R K R **G** D D A N AAGAGGAAGCGGGGAGATGATGCGAAC ↓ AAGAGGAAGCGGGTAGATGATGCGAAC K R K R **V** D D A N	-1
4	613 (100)	SC-E-02	V2 (CP) (DNA-binding region)	S Q A C **K** Y S N T TCGCAAGCCTGCAAGTATTCTAACACC ↓ TCGCAAGCCTGCAATTATTCTAACACC S Q A C **N** Y S N T	1
5	691 (126)	SC-E-02 SC-F-01	V2 (CP)	P T P Q **T** I F A Y CCGACCCCACAAACCATATTTGCATAT ↓ CCGACCCCACAAACGATATTTGCATAT P T P Q **T** I F A Y	3
6	792 (159)	SC-E-02 SC-F-01	V2 (CP)	R W L F **N** M E T D CGATGGTTGTTCAACATGGAGACCGAC ↓ CGATGGTTGTTCAGCATGGAGACCGAC R W L F **S** M E T D	1
7	811 (165)	SC-E-02 SC-F-01	V2 (CP)	E T D G **R** I G S D GAGACCGACGGTCGAATTGGTTCGGAC ↓ GAGACCGACGGTCGGATTGGTTCGGAC E T D G **R** I G S D	6
8	886 (191)	SC-E-02 SC-F-01	V2 (CP)	F H K F **T** S G L G TTCCACAAGTTCACGAGTGGGTTGGGA ↓ TTCCACAAGTTCACCAGTGGGTTGGGA F H K F **T** S G L G	3
9	919 (202)	SC-E-02	V2 (CP)	T Q W K **N** V T D G ACGCAGTGGAAGAATGTAACGGACGGA ↓ ACGCAGTGGAAGAACGTAACGGACGGA T Q W K **N** V T D G	2

[a]	Pos[b]	Clone	Genomic Region	Sequence changes	PAM250[c]
10	969 (219)	SC-E-02 SC-F-01	V2 (CP)	A L Y M **V** I A P G GCTCTGTACATGGTCATTGCCCCTGGC ↓ GCTCTGTACATGGTTATTGCCCCTGGC A L Y M **V** I A P G	4
11	1045 (244)	SC-E-02	V2 (CP) V2 STOP V-transcript polyadenylation signal	S V G N **Q** * AGTGTCGGCAACCAGTAATGAATAAAA ↓ AGTGTCGGCAACCATTAATGAATAAAA S V G N **H** *	3
12	1087	SC-E-02 SC-F-01	SIR	TCTGATGAATGCTCAAAGCTTACATTA ↓ TCTGATGAATGCTGAAAGCTTACATTA	
13	1314 (322)	SC-F-01	C2 (Rep)	N E M T **P** G Q L E AATGAAATGACTCCAGGGCAGCTGGAGT ↓ AATGAAATGACTCTAGGGCAGCTGGAGT N E M T **L** G Q L E	-3
14, 15	1404, 1406 (291), (292)	SC-F-01	C2 (Rep)	F V V N **P** **K** Y G K K TTCGTTGTAAATCCAAAATATGGCAAGAAG ↓ ↓ TTCGTTGTAAATCCTACATATGGCAAGAAG F V V N **P** **T** Y G K K	6, 0
16	1427 (284)	SC-E-02	C2 (Rep)	L V G C **Q** K E F V TTAGTTGGCTGTCAGAAAGAGTTCGTT ↓ TTAGTTGGCTGTCACAAAGAGTTCGTT L V G C **H** K E F V	3
17	1508 (257)	SC-E-02	C2 (Rep) (*myb*-like transactivation domain)	D W S S **Y** **N** **E** **D** **A** GATTGGTCTTCTTACAACGAAGACGCA ↓ ↓ CATTGGTCTTCTTAAAACGAAGACGCA H W S S *	-8
18	1522 (253)	SC-E-02	C2 (Rep) Putative C1 polyadenylation signal (*myb*-like transactivation domain)	Q N N V **D** W S S Y CAAAATAATGTTGATTGGTCTTCTTAC ↓ ↓ CAAAATAATGTTCATTGGTCTTCTTAA Q N N V **H** W S S *	1
19	1633-1648 (242)-(246)	SC-E-02	C1 (RepA) Rep intron Acceptor site (Putative GRAB binding domain)	E W M **D** K L S S Q **Q** **M** **K** **-** **-** * GAATGGATGGATAAATTATCCAGCCAACAGATGAAG-N[69]-TAA ↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓↓ GAATGGATGGA.....CAGATGAAG E W M **D** **R** *	-8

[a]	Pos[b]	Clone	Genomic Region	Sequence changes	PAM250[c]
20	1873 (166)	SC-F-01	C1 (RepA) C1-C2 (Rep)	P Y D W **A** T K L Q CCCTATGACTGGGC**C**ACAAAATTGCAG ↓ CCCTATGACTGGGC**T**ACAAAATTGCAG P Y D W **A** T K L Q	2
21	1924 (149)	SC-F-01	C1 (RepA) C1-C2 (Rep)	S H S T **S** K L E Y TCACACTCCACTTC**A**AAGCTAGAGTAC ↓ TCACACTCCACTTC**T**AAGCTAGAGTAC S H S T **S** K L E Y	2
22	1975 (132)	SC-E-02	C1 (RepA) C1-C2 (Rep)	G N S E **K** K P S K GGAAATTCTGAAAA**G**AAACCTTCAAAA ↓ GGAAATTCTGAAAA**T**AAACCTTCAAAA G N S E **N** K P S K	1
23	2337 (12)	SC-E-02	C1 (RepA) C1-C2 (Rep)	R Q F S **H** R N V N CGTCAATTCTCA**C**ACAGGAACGTTAAC ↓ CGTCAATTCTCA**T**ACAGGAACGTTAAC R Q F S **Y** R N V N	0
24	2435	SC-E-02 SC-F-01	LIR TATA box Inverse repeat of partial ori stem sequence	CTATATCAACCGG**T**TTGCGCCTTCGAA ↓ CTATATCAACCGG**C**TTGCGCCTTCGAA	
25	2468	SC-E-02 SC-F-01	LIR	CGCTCCCCCTTTT**T**AGTGGTTGTTTA ↓ CGCTCCCCCTTTT**A**TAGTGGTTGTTTA	
26	2547	SC-E-02 SC-F-01	LIR Right-hand side of stem sequence for ori stem-loop structure	CGCGCCTTCTTTTT**C**CTGCGAGGGCCCG ↓ CGCGCCTTCTTTTT**T**CTGCGAGGGCCCG	
27	2601	SC-E-02	LIR AT-tracts	TCGGTTCTGCTTT**G**TCTGATTTATCTA ↓ TCGGTTCTGCTTT**T**TCTGATTTATCTA	

[a] Numbers correspond to those in Figure 3-2~~Figure 3-2~~; [b] nucleotide positions are given relative to the conserved *Bam*HI restriction site at the V1 start codon, while amino acid positions are shown in parentheses; [c] PAM250 scores ≥ 1 may indicate conservative amino acid changes. For each mutation, the following information is given on successive lines, from top to bottom: parental amino acid sequence (N→C); parental nucleotide sequence (5'→3', coding strand, or virion strand if not coding); arrow(s) indicating mutation(s); mutant nucleotide sequence; mutant amino acid sequence. Residues in red indicate changes from parental sequences; affected codons are underlined. Where functional motifs are known or suspected to form a part of the region shown, they are indicated by highlighting. Stop codons are indicated by *.

Mutation frequency bias

Despite the small size of the data-set, we attempted to determine whether there were any substitution patterns that were significantly over-represented. Table 3-4~~Table 3-4~~ lists the number of observed mutations of each type, as well as the expected frequencies taking initial nt. frequency into account. We found that G→T transversions were over-represented in both experiments. Using $p < 0.05$ as the criterion for significance, this higher-than-expected frequency was not statistically significant in the MSV-Kom / MSV-Set dataset (X^2 p-value = 0.09), but was significant in the MSV-Tas ($p = 0.006$) data, and in the combined dataset for both experiments ($p = 0.0002$). Though not statistically significant in our relatively small dataset, C→A and T→G changes appeared to be consistently under-represented. Similarly, there seems to be some evidence of a bias towards a shift from GC to AT basepairs (Table 3-4~~Table 3-4~~), especially in the MSV-Tas dataset ($p = 0.06$), although this is also not strictly supported by $p < 0.05$. There was no evidence of a bias in the ratio of transitions to transversions.

dN/dS

Mutations were distributed among coding and non-coding sites more or less as expected, given their relative numbers (Table 3-4~~Table 3-4~~). The ratio of non-synonymous to synonymous substitutions (dN/dS) was similar in both experiments, and was considerably less than one (dN/dS = 0.36 for the combined dataset; p=0.002).

Table 3-4. Mutation frequencies for chimaeric MSV-Kom/Set and MSV-Tas clones.

Subs.	MSV-Kom/Set			MSV-Tas			All mutants analysed		
(V-sense)	Obs.	Exp.	X^2 p-value	Obs.	Exp.	X^2 p-value	Obs.	Exp.	X^2 p-value
				Transversions					
A → C	2	1.18	0.43	0	2.26	0.12	2	3.45	0.38
A → T	1	1.18	0.86	1	2.26	0.38	2	3.45	0.38
C → A	0	1.08	0.28	1	2.01	0.46	1	3.09	0.20
C → G	1	1.08	0.94	4	2.01	0.14	5	3.09	0.27
G → C	1	1.22	0.84	1	2.13	0.42	2	3.35	0.41
G → T	3	1.22	0.09	6	2.13	0.006	9	3.35	0.0002
T → A	0	1.19	0.25	3	2.26	0.61	3	3.45	0.73
T → G	0	1.19	0.25	1	2.26	0.38	1	3.45	0.15
				Transitions					
A → G	3	1.18	0.08	2	2.26	0.86	5	3.45	0.40
C → T	1	1.08	0.94	2	2.01	0.99	3	3.09	0.85
G → A	2	1.22	0.46	3	2.13	0.53	5	3.35	0.36
T → C	0	1.19	0.25	2	2.26	0.86	2	3.45	0.38
Ins	0			0			0		
Del	1			1			2		
Total	15			27			42		
A/T→G/C	5	4.7	} 0.50	5	9.0	} 0.06	10	13.8	} 0.08
G/C→A/T	6	4.6		12	8.3		18	12.9	
Tv	8	9.3	} 0.45	17	17.3	} 0.89	25	26.7	} 0.58
Ts	6	4.7		9	8.7		15	13.3	
Ts/Tv	0.75	0.5		0.53	0.5		0.60	0.5	
Coding	13	12.4	0.67	21	22.3	0.52	34	34.7	0.79
N	6	10.1	} 0.006	13	16.3	} 0.079	19	26.5	} 0.002
S	7	2.9		8	4.7		15	7.5	
dN/dS	0.24	1.0		0.46	1.0		0.36	1.0	

Distribution of mutations by genomic region

Because epistatic interactions within the V1-V2 cassette appeared to account for a large proportion of the reduced fitness of the chimaeric MSV-Kom/MSV-Set genomes (see Chapter 2), we compared the distribution of mutations among the virion- and complementary-sense coding regions in the MSV-Kom/-Set chimaeras with the distributions seen in their KNco and SNco parents (Table 3-5). In the parental viruses, neither the virion-sense nor the complementary-sense ORFs contained a statistically significant disproportionate density of mutations in comparison to the rest of the genome. However, the chimaeric genomes contained eleven times more substitutions per site in the V1-V2 cassette than in the rest of the genome (Yates' X^2 p-value = 0.014), while the density of mutations in the C-sense genes was similar to that seen in the parental viruses, and not statistically different from that seen in the rest of the genome.

Similarly, we investigated the distribution of mutations across the MSV-Tas genome, and found no significant differences in the density of mutations in either the V- or the C-sense genes in comparison to the rest of the genome (Table 3-5). However, since a subset of the mutations were shared between the two genomes that we sequenced, with the remainder of the mutations being unique to either one or the other, we further tested whether there were significant differences in the clustering of shared versus unique mutations. Indeed, none of the shared substitutions occurred in the ~1.2 kbp C1-C2 region versus nine in the remaining ~1.5 kbp of the genome (Yates' X^2 p-value = 0.014). In contrast, no significantly skewed distribution was detectable for the sub-set of mutations unique to either genome.

Table 3-5. Distribution of mutations by genomic region.

Genomic region	Mutations in region (region size)	Mutations in rest of genome (region size)	Fold difference in sub/site	X^2 p-value
	MSV-Kom/Set – parental viruses			
V1 + V2	4 (1039)	3 (1658)	2.1	0.534 †
C1 + C2	1 (1159)	6 (1538)	0.2	0.250 †
	MSV-Kom/Set – chimaeras			
V1 + V2	7 (1039)	1 (1658)	11.2	0.014 † *
C1 + C2	1 (1159)	7 (1538)	0.2	0.164 †
	MSV-Tas – all mutations			
V1 + V2	11 (1039)	16 (1644)	1.1	0.830
C1 + C2	11 (1159)	16 (1524)	0.9	0.797
	MSV-Tas – shared mutations			
V1 + V2	5 (1039)	4 (1644)	2.0	0.489 †
C1 + C2	0 (1159)	9 (1524)	0.0	0.023 † *
	MSV-Tas – unshared mutations			
V1 + V2	6 (1039)	12 (1644)	0.8	0.640
C1 + C2	11 (1159)	7 (1524)	2.1	0.126

† Yates' chi-squared p-values; * p-value < 0.05.

Discussion

With the aim of investigating mutation frequencies and patterns in MSV under differing conditions of natural selection, mutations arising in various MSV populations initiated from clones and passaged *in planta* were determined by re-isolating and sequencing whole virus genomes. Two hypothetical selection pressures were applied: in the first series of experiments, various chimaeric viruses that were less fit than their constituent parents were passaged through twelve generations of sweetcorn plants over a one-year period; in the second experiment, a clone of the wheat-adapted MSV-Tas strain was used to infect a sugarcane plant, and the infected plant was maintained four years before analysing the progeny virus. Here we present experimental evidence confirming that our MSV variants display mutation rates at least as high as previously reported for the Reunion MSV variant (Isnard *et al.*, 1998) and some begomoviruses (Ge *et al.*, 2007; Duffy & Holmes, 2008). We also show

that MSV experiences some of the biased mutation frequencies recently observed in some begomoviruses (Ge *et al.*, 2007; Duffy & Holmes, 2008).

In addition to the wild-type parental viruses, we managed to obtain genomic clones from three of the original six chimaeric virus constructs with which we started the experiment. Although we attempted to avoid severe bottleneck events during leafhopper transmission by allowing long transmission periods with multiple leafhoppers, the apparent extinction of two of the chimaeric virus populations during the experiment can most likely be ascribed to their low initial fitness as well as the possibility of the negative effects of genetic drift arising from relatively small population sizes, high mutation rates (see below) and repeated bottle-neck events.

MSV exhibits high mutation rates

People have noted high population diversity among geminiviruses – and the attendant implication of relatively high mutation rates – for some time (Gilbertson *et al.*, 1991; Stenger & McMahon, 1997; Sanz *et al.*, 1999) and one early experimental study relying on RFLP and partial sequence analysis supported this conclusion for MSV (Isnard *et al.*, 1998). Although those infections were not initiated from cloned virus constructs, an attempt was made to limit the diversity of the starting population by using a single leafhopper vector and a short acquisition time, and the resulting estimated mutation frequencies were ~10^{-5} subs/site within single plants after infection periods of up to four years, equivalent to a rate of 2.6 x 10^{-4} subs/site/year. Recently, Ge *et al.* (2007) sequenced PCR-amplified partial TYLCCNV genomes from virus populations in individual plants at two and four months following inoculation with a single virus clone, and reported similar mutation frequencies of ~4 x 10^{-4} subs/site, or around 2 x 10^{-3} subs/site/year. However, their control experiment indicated that up to 20% of the mutations they observed were PCR-derived (taken into account in their calculations), despite their use of a type-B DNA polymerase, high template concentrations, and only twenty PCR cycles. We

attempted to overcome some of these confounding experimental conditions by initiating infections using cloned, characterised virus constructs and by avoiding PCR amplification. We also sequenced full-length progeny genomes, with the hope of obtaining a more complete view of mutation frequencies and potential mutation hotspots across the entire genome. Finally, our wild-type and chimaeric virus constructs were well characterised in terms of both genomic sequence and virulence, so we had a clear idea of a subset of possible mutations that might restore fitness, as well as the potential extent to which these viruses could regain virulence (represented by the wild-type MSV-Kom) via evolution.

Among the nine MSV-Kom/MSV-Set genomes we sequenced, the average mutation rate was ~7.4 x 10^{-4} subs/site/year, which is in line with other published figures for geminiviruses (Isnard *et al.*, 1998; Ge *et al.*, 2007; Duffy & Holmes, 2008). In contrast, the mutation rates we derived from the two MSV-Tas clones isolated after four years of infection in sugarcane were about 2-fold higher (average 1.6 x 10^{-3} subs/site/year) than the MSV-Kom/MSV-Set rates in maize, and about six-fold higher than the mutation rates Isnard *et al.* (1998) observed after four years of MSV infection of a single perennial host plant. Of course, we analysed just two of these genomes, so the uncertainty in the mutation rate we calculated is relatively large, and difficult to estimate. Moreover, one of the two clones contained a significant deletion at the *Rep* splice site, implying that it is unlikely to be able to independently maintain an infection; if this genome was already dependent on factors supplied by the co-infecting virus population then purifying selection against some of the mutations it had acquired may have been somewhat relaxed, allowing the accumulation of more substitutions. Indeed, Duffy and Holmes (Duffy & Holmes, 2008) calculated a substitution rate of 1.7 x 10^{-3} subs/site/year for the noncoding regions of TYLCV, so the high rates we report here are not without precedent.

Based on the few, recently published findings discussed above and the rates we report here, it is becoming increasingly apparent that geminiviruses display mutation rates at least as high as some plant RNA viruses (Kearney *et al.*, 1999; Schneider & Roossinck, 2000; Schneider & Roossinck, 2001; Stenger *et al.*, 2002) and orders of magnitude greater than rates determined for dsDNA viruses (Bernard, 1994; McGeoch & Gatherer, 2005). This represents a significant departure from the natural assumption that the synthesis of geminivirus genomes by their hosts' machinery (Palmer & Rybicki, 1998; Hanley-Bowdoin *et al.*, 1999) implies relatively error-free virus replication, in line with that seen in plants (Hanley-Bowdoin *et al.*, 1996; Roossinck, 1997). At least two other diverse ssDNA viruses have been shown to have nucleotide substitution rates in the range of 10^{-4} subs/site/year – canine parvovirus (Shackelton *et al.*, 2005), and B19 erythrovirus (Shackelton & Holmes, 2006) – which implies that high mutation rates are likely to be a common, if not universal, feature among ssDNA viruses.

Because of our relatively scant understanding of plant DNA replication in general, and more specifically the host factors involved in geminivirus replication (Hanley-Bowdoin *et al.*, 1999; Gutierrez *et al.*, 2004), the mechanisms underlying the surprisingly high mutation rates seen in geminiviruses remain a topic of speculation. There are, however, some clues about where to start looking. As early as 1997 Roossinck noted that since replicating geminivirus DNA is apparently not methylated (Brough *et al.*, 1992), it is possible that normal host mechanisms for mismatch repair may not operate on replicating geminiviruses (Inamdar *et al.*, 1992), and both Ge *et al.* (2007) and Duffy and Holmes (2008) make the same proposal. Duffy and Holmes (2008) suggested two additional possibilities: i) because geminivirus DNA is only transiently double-stranded during rolling-circle replication, it may not be suitable for base-excision repair; ii) the biased mutation patterns (Ge *et al.*, 2007; Duffy & Holmes, 2008) may be explained either by spontaneous deamination – potentially more likely to occur in ssDNA (Frederico *et al.*, 1990;

Caulfield *et al.*, 1998; Xia & Yuen, 2005) – or by the action of deaminating host enzymes (Stasolla *et al.*, 2003). While we were unable to assess relatively small differences in mutation frequencies, the elevated frequency of G→T transversions that we observed was also reported by Ge *et al.* (2007) and Duffy and Holmes (2008) for TYLCCNV and TYLCV, respectively.

There is one additional clue regarding the mechanistic origins of the high mutation rates seen in geminiviruses: we observed very high mutation rates – 60 subs/genome, or 2 x 10^{-2} subs/site – associated with evidence of extremely frequent recombination cross-over events in the same genome (Chapter 4). Viewed together with existing evidence of frequent illegitimate recombination † (and see Chapter 4) and the important role played by recombination-dependent replication in geminivirus biology (Jeske *et al.*, 2001; Preiss & Jeske, 2003), this constitutes evidence that mutations may arise with high frequency in the course of either homologous or non-homologous recombination, or both, which occur(s) constantly during geminivirus infection.

High mutation rates do not necessarily lead to rapid adaptation

In the context of reports that some MSV mutants either revert or achieve compensatory mutations at high rates to restore fitness (Shepherd *et al.*, 2005; Arguello-Astorga *et al.*, 2007) and that MSV can adapt to overcome host resistance within a period of about a year (Isnard *et al.*, 1998), we were surprised to see no changes in the symptomatology of any of our chimaeric viruses, even after one year of constant infection and twelve cycles of transmission. Indeed, sequencing and analysis of mutations in the progeny populations revealed low global dN/dS ratios indicating that little or no positive selection had occurred, although it should be noted that our data did not permit a site-by-site analysis. More striking was the fact that only two out of eight unique mutations among the chimaeric viruses represented

† (MacDowell *et al.*, 1986; MacDonald *et al.*, 1988; Stanley *et al.*, 1990; Stenger *et al.*, 1992; Frischmuth & Stanley, 1994; Kenton *et al.*, 1995; Bejarano *et al.*, 1996; Stanley *et al.*, 1997; Paximadis & Rey, 2001; Bull *et al.*, 2004; Stanley, 2004; Casado *et al.*, 2004; Patil *et al.*, 2007)

changes from one wild-type parental sequence to the other, which is close to the expected number (8/3 = 2.7) if one assumes random mutation. The only indication of positive selection was the higher rate of substitutions in the V1-V2 region of the chimaeras compared to the rest of the genome (Table 3-5). However, considering that only two of these substitutions resulted in (relatively conservative) non-synonymous changes (mutations 2 and 7, Table 3-2~~Table 3-2~~) any positive selection that may have occurred was likely to have been acting on noncoding functional aspects of the DNA sequences.

Although we did not monitor the symptomatology of the MSV-Tas infection in the sugarcane plant, the sequence data also revealed a low dN/dS ratio, and thus little or no positive selection acting on protein functions. However, the two clones that we sequenced shared, on average, about half of their substitutions in common, which implies that the consensus sequence of the population had diverged from that of the clone that was used to initiate the infection four years earlier – probably the result of drift more than of selection. Isnard *et al.* (1998) also observed changes in consensus sequences following passage of MSV for four years in a single perennial host plant, and following passage in partially MSV-resistant maize for an undisclosed period of a year or more. This contrasts with the conclusions of Ge *et al.* (2007), who reported stable populations of the begomovirus TYLCCNV with unchanging consensus sequences during their experiments, which lasted up to four months.

In hindsight, the lack of obvious evolutionary adaptation in our chimaeric viruses may be ascribed to a number of factors, the most obvious having to do with the relatively extreme evolutionary distances that they would have had to traverse in order to converge appreciably on either of the original wild-type parental sequences. If specific pairs or sets of mutations were required to rescue fitness in the chimaeras, then the probabilities of those mutations arising within the one-year duration and the relatively small population sizes of our experiment were almost certainly miniscule,

even in the face of high mutation rates. Moreover, it seems reasonable to imagine that selection of protein function is relatively weak in an organism that is essentially massively polyploid: in any one infected cell there are likely to be hundreds or thousands of geminivirus genomes capable of simultaneously directing protein expression. This would imply that defective mutants might be maintained via complementation by their fit siblings, and indeed the prevalence of "satellite" viruses, "subgenomic" genomes and genomes carrying apparently lethal mutations (e.g. SC-E-02 in this study, and see Chapter 4) bears this out. Similarly, one would expect the fitness benefits of an advantageous coding mutation in a single genome to be diluted by the presence of unimproved proteins expressed from the overwhelming majority of its more mediocre peers, and by the fact that any individual, improved, proteins would most likely benefit all of the genomes in a single cell more or less equally. Selection of mutations affecting the entire virus population in a plant – for example by modulating systemic host defence responses – would be even further diluted by this effect, since selection would effectively only occur upon transmission to new host plants. Conversely, mutations affecting DNA structures required for efficient replication would be differentiated at each round of replication.

This "stratification" of selection pressure may at least partially explain some paradoxical reports regarding the reversion of geminivirus mutants: Boulton *et al.* (1991a; 1991b) described two polymorphisms in MSV-Nm that were solely responsible for negatively affecting both symptom severity and host range when compared with MSV-Ns. Moreover, they demonstrated that reversion of each individual mutation by *in vitro* site-directed mutagenesis rescued a significant portion of the lost fitness. Since their host-range determinations entailed leafhopper transmission of the viruses from maize between eighteen and twenty-five days after agroinoculation, it seems likely that MSV-Nm had not reverted within that time-period. In contrast, there are two reports of high-frequency reversion of mutations within time-frames as short as ten days, and no longer than twenty-four days

(Shepherd *et al.*, 2005; Arguello-Astorga *et al.*, 2007). The relative stability of MSV-Ns compared with the extremely efficient reversion of other mutants may be partially due to differences in the mechanisms of selection that were at work, as described above.

Geminiviruses as a model system for evolution experiments

RNA viruses are often considered to be especially useful model systems for studying evolution, because of their high mutation rates and small genomes (Holmes, 2004) but it is becoming apparent that small ssDNA viruses like MSV may exhibit similarly high rates of mutation. There are at least four more good reasons why ssDNA viruses like MSV appear to be perfectly suited to serve as model systems for studying evolution at the molecular level: i) in contrast to RNA viruses, they have the advantage of not requiring error-prone, expensive, and time-consuming reverse-transcription before cloning and sequencing; ii) their circular genomes have proven particularly amenable to amplification using commercial kits based on phi29 DNA polymerase and rolling-circle amplification (Inoue-Nagata *et al.*, 2004; Haible *et al.*, 2006; Owor *et al.*, 2007b), which don't seem to suffer from the same fidelity or *in vitro* recombination issues documented for PCR; iii) it is relatively easy to make infectious DNA, and there is no need for expensive transcription steps as is the case with many RNA viruses; and iv) while geminiviruses appear to match the mutation rates of RNA viruses, they also seem to recombine prodigiously (see Chapter 4), so they offer opportunities to study both of the fundamental driving forces behind diversification in evolution in a single system.

CHAPTER 4: IN PLANTA RECOMBINATION BETWEEN CO-INFECTING MSV GENOMES OCCURS FREQUENTLY AND IS A POWERFUL MECHANISM FOR REVERTING DELETERIOUS MUTATIONS

ABSTRACT

Genetic recombination is ubiquitous among all major taxonomic groups, and forms the basis of important processes like DNA repair and sexual reproduction. By unlinking alleles, recombination allows at least partial escape from Muller's ratchet, and unlocks the potential to combine beneficial alleles that are spread throughout a population within a single individual. Among viruses, recombination plays an important role in the emergence of new diseases, and viruses can be useful experimental tools for studying the mechanisms and effects of recombination. In addition to having a relatively well-documented impact on geminivirus evolution, recombination is thought to play an important mechanistic role in geminivirus replication. However, relatively little effort has been made to establish experimental systems for studying recombination in geminiviruses. We have established such an experimental model system based on infectious, reduced-fitness maize streak virus (MSV) reciprocal chimaeras. These reciprocal chimaeras comprised precise exchanges of the entire virion-sense coding regions between MSV-Mat and MSV-VW, which are isolates of distinct MSV strains that share 89% nucleotide sequence identity. MSV-Mat efficiently infects the moderately MSV-resistant maize cultivar PAN6099, but MSV-VW cannot, and both reciprocal chimaeras that we used here are barely infectious in this host. We co-infected PAN6099 plants with both chimaeras and utilised natural selection to preclude parental chimaeric viruses and to enrich recombinant progeny viruses arising during co-infection. Here I describe full genome sequences of twelve recombinant viruses, and show that recombination is a powerful mechanism for rapidly reverting deleterious mutations and/or combining advantageous alleles during geminivirus infection. I further provide evidence of illegitimate recombination, as well as the resultant incorporation of host genetic sequences into MSV DNA.

Introduction

Both the phenomenon of genetic recombination and the mechanisms by which it occurs are common to all major phylogenetic groups, including viruses (Cromie *et al.*, 2001), and in cellular organisms recombination forms the basis of important processes such as DNA repair and sexual reproduction. For these and other reasons, it has been convincingly argued that genetic recombination is at least as ancient as cellular life, and probably arose very early in the evolution of replicating biopolymers (Lehman, 2003). Because of this pervasiveness, many people have expended a great deal of effort in attempting to explain the implicit underlying selective advantage(s) that must be offered by recombination, and consequently a number of hypotheses have been put forward and tested (reviewed in Hey, 1998; Barton & Charlesworth, 1998; de Visser & Elena, 2007). At the most fundamental level, most of these explanations boil down to a single, obvious feature of recombination: it enables the combinatorial re-assortment of otherwise perpetually linked alleles.

It has long been recognised that recombination is important for its role in unlinking deleterious mutations from those that may be neutral or beneficial, thus allowing small populations at least partial escape from the negative fitness effects of high mutation rates and the resulting phenomenon of "Muller's ratchet" (Muller, 1964; Chao, 1990; Duarte *et al.*, 1992). Similarly, recombination allows otherwise asexual populations to avoid the evolution-retarding effects of "clonal interference", which results from competition among distinct beneficial alleles that reside concurrently in multiple clones – eventually one dominates the population at the expense of all the rest (Gerrish & Lenski, 1998; Elena & Sanjuan, 2007; de Visser & Elena, 2007). In other words, recombination can bring advantageous alleles residing in separate, competing organisms together in a single progeny genome which may be fitter than either parent. It is important to note that the entities being recombined in these examples may range in size from single nucleotide substitutions – through DNA

motifs and protein domain-encoding sequences or whole genes – up to entire multi-partite virus genome segments or eukaryotic chromosomes. Moreover, horizontal gene transfer (reviewed by Lawrence & Hendrickson, 2003) may result in genetic recombination between organisms from distantly related – or even unrelated – taxa, in which case entirely new branches may arise on the evolutionary tree (Saunders & Stanley, 1999; Gibbs & Weiller, 1999). In this way, recombination might instantaneously bring together two genetic sequences that have evolved independently over millennia – essentially enabling "massively parallel" evolution of functional modules across vast numbers of genetically distinct lineages. Thus, the combinatorial property of recombination may result in much faster evolution rates than could be realised via mutation alone (Stemmer, 1994; Crameri *et al.*, 1998; Drummond, 2005).

As with so many biological phenomena, viruses have provided unique as well as general insights into recombination, and they offer outstanding prospects for testing a variety of theories experimentally (Elena & Sanjuan, 2007; de Visser & Elena, 2007). Moreover, while the relative importance of recombination in the evolution of cellular organisms and higher eukaryotes is apparently still under debate (Koonin, 2003; Shackelton & Holmes, 2004), there can be no doubt that recombination among viruses has led to profound evolutionary events, and is an important process driving the emergence of new diseases across a range of diverse hosts (Roossinck, 1997; Holland & Domingo, 1998; Worobey & Holmes, 1999). More specifically, there are numerous examples of promiscuous recombination within and among various groups of small ssDNA viruses including parvoviruses (Shackelton *et al.*, 2007), TT viruses (Worobey, 2000), circoviruses (Heath *et al.*, 2004; Ma *et al.*, 2007), and geminiviruses (Padidam *et al.*, 1999).

During the last decade, reports of geminivirus recombination detected by analyses of genome sequences sampled from nature have been published with increasing

frequency [‡], reflecting the exponential increase in geminivirus whole-genome sequences lodged in GenBank, improvements in the usability of software tools for detecting recombination (Martin & Rybicki, 2000; Martin *et al.*, 2005b), and presumably an increased awareness of the importance of recombination in geminivirus epidemiology.

Of course, the bipartite begomoviruses are also prone to reassortment – or "pseudorecombination" – in which the A component of one virus might adopt the B component of a distinct virus, and there are many published examples of viable pseudorecombinant geminiviruses discovered in nature (Torres-Pacheco *et al.*, 1993) and in the laboratory (Gilbertson *et al.*, 1993). Furthermore, homologous recombination between replicating geminivirus DNA and transgenic geminivirus sequences has been observed (Frischmuth & Stanley, 1998), indicating the need for caution in the dissemination of transgenic plants comprising pathogen-derived resistance mechanisms.

In addition to the numerous reports of inter-geminivirus homologous recombination, geminivirus DNA sequences have been found illegitimately integrated into host chromosomal DNA (Kenton *et al.*, 1995; Bejarano *et al.*, 1996), although I know of just one report of geminiviruses capturing host sequences (Stanley *et al.*, 1997). Further support for the importance of illegitimate recombination mechanisms in geminivirus biology can be found in the growing number of so-called "subgenomic" geminivirus DNA species that have been described [§]. Interestingly, recombination has relatively recently been shown to play

[‡] (Zhou *et al.*, 1997; Zhou *et al.*, 1998; Saunders & Stanley, 1999; Sanz *et al.*, 1999; Fondong *et al.*, 2000; Berrie *et al.*, 2001; Monci *et al.*, 2002; Willment *et al.*, 2002; Galvao *et al.*, 2003; Xie & Zhou, 2003; Jovel *et al.*, 2004; Kitamura *et al.*, 2004; Ndunguru *et al.*, 2005; Delatte *et al.*, 2005; Garcia-Andres *et al.*, 2006; Patil *et al.*, 2007; Lefeuvre *et al.*, 2007; Owor *et al.*, 2007a; Garcia-Andres *et al.*, 2007a; Duffy & Holmes, 2008; Shepherd *et al.*, 2008)

[§] (Nadeem *et al.*, ; Stanley & Townsend, 1985; MacDowell *et al.*, 1986; MacDonald *et al.*, 1988; Roberts *et al.*, 1988; Czosnek *et al.*, 1989; Stanley *et al.*, 1990; Frischmuth & Stanley, 1992; Stenger *et al.*, 1992; Frischmuth & Stanley, 1994; Stanley *et al.*, 1997; Liu *et al.*, 1998b; Mansoor *et al.*, 1999; Lotrakul & Valverde, 1999; Paximadis & Rey, 2001; Zhou *et al.*, 2001; Briddon *et al.*, 2003; Zhou

an important mechanistic role in geminivirus replication (Jeske *et al.*, 2001; Preiss & Jeske, 2003), in addition to the canonical rolling-circle model that was characterised earlier (Saunders *et al.*, 1991; Stenger *et al.*, 1991). Many people have recognised that geminivirus recombination represents a particularly worrisome threat to agriculture not only through the emergence of new diseases, but more specifically through the potentially rapid evolution of novel virus genotypes capable of overcoming both inbred (MSV-resistant maize, Damsteegt, 1983; Martin *et al.*, 1999), and transgenically derived host resistance (reviewed by Vanderschuren *et al.*, 2007; MSV-resistant maize, Shepherd *et al.*, 2007).

While bioinformatics tools for analysing recombination have been extremely valuable in establishing the importance and the consequences of geminivirus recombination, there are good reasons to pursue more specific experimental approaches. First, viruses isolated from nature have survived a relatively rigorous process of natural selection, leaving the vast majority of recombination events undetected. This selection process implies that the subset of recombination events that remain are unlikely to be representative of the whole; for example, both genome modularity and parental relatedness affect the viability of recombinant progeny (Martin *et al.*, 2005a). Second, bioinformatic interpretation of recombination data from naturally sampled sequences is partially confounded by the fact that signals of recombination in naturally obtained sequences are difficult to detect with confidence if the parental sequences are closely related, or if the exchanged regions are small (Posada & Crandall, 2001). Third, there is usually a great deal of uncertainty regarding the time-scales over which a given number of detected recombination events may have accrued in any given dataset of genomic sequences obtained from nature, making it difficult to gauge the frequency with which such events occur.

---

et al., 2003; Mansoor *et al.*, 2003a; Bull *et al.*, 2004; Stanley, 2004; Casado *et al.*, 2004; Patil *et al.*, 2007; Behjatnia *et al.*, 2007)

Recombination has been observed in mastreviruses in the course of genetic complementation-type experiments (Boulton *et al.*, 1989; Palmer & Rybicki, 2001) and in experiments manipulating geminivirus genome size (Gilbertson *et al.*, 2003), demonstrating that geminivirus recombinants are detectable within experimental time-frames under suitable selection conditions. Using heterodimeric infectious constructs, with the genomes fused ~150 bp downstream of the origin of virion-sense replication, Schnippenkoetter *et al.* (2001b) carried out an experimental investigation of recombination between diverse mastreviruses. Probably due to the specific design of these heterodimers, the recombinants they isolated comprised relatively small exchanged regions, each having a cross-over site at, or very near to, the origin of virion-sense replication. Recently, García-Andrés *et al.* (2007b) demonstrated recombination between isolates of the monopartite begomovirus species *Tomato yellow leaf curl Sardinia virus* (TYLCSV) and *Tomato yellow leaf curl virus* (TYLCV) during co-infection of tomato plants and characterised the recombinants, but I am not aware of any other attempts to develop systems for studying geminivirus recombination under controlled conditions in a laboratory. We had previously constructed and characterised a large number of MSV reciprocal chimaeras (Martin & Rybicki, 2002; Martin *et al.*, 2005a), and believed that these offered unique opportunities for establishing such a system. Two of these reciprocal chimaeras – MatMPCPVW and VWMPCPMat – comprise precise exchanges of both virion-sense ORFs between MSV-Mat (Martin *et al.*, 2001) and MSV-VW (Willment *et al.*, 2002) and were used in this study.

MSV-Mat is a representative of the maize-infecting MSV-A strain which is predominantly responsible for causing maize streak disease in Africa, and shares 89% nucleotide sequence identity with MSV-VW, a wheat-infecting isolate of the MSV-B strain that is considerably less pathogenic in maize (Martin *et al.*, 2001). MSV-Mat efficiently infects the moderately MSV-resistant maize cultivar PAN6099, but MSV-VW cannot, and both MatMPCPVW and VWMPCPMat are barely infectious in this

host (Martin & Rybicki, 2002). We utilised natural selection in PAN6099 plants to preclude parental chimaeric viruses, and to enrich recombinant progeny arising during co-infection with MatMPCPVW and VWMPCPMat. Here I describe full genome sequences of twelve recombinant viruses, and show that recombination is a powerful mechanism for rapidly reverting deleterious mutations and/or combining advantageous alleles during geminivirus infection. A collection of subgenomic virus DNA species are also described, which indicate an important role for illegitimate recombination in geminivirus molecular biology and evolution. I further provide evidence pointing to recombination hotspots in the MSV genome, and discuss the mechanisms that might give rise to such patterns of recombination in MSV.

Methods and materials

Viruses, plasmids, bacterial strains, maize genotypes

MSV-Mat (Martin *et al.*, 2001) and MSV-VW (Willment *et al.*, 2002) are two distinct strains of MSV isolated from maize and from wheat, respectively. Agroinfectious constructs of MatMPCPVW and VWMPCPMat were obtained from D.P. Martin (University of Cape Town, South Africa) and are described in Martin *et al.* (2002). The chimaeric virus MatMPCPVW consists of an MSV-MatA mutant in which the wild-type V1 and V2 ORFs (including intervening DNA) have been exchanged with those of MSV-VW. Similarly, VWMPCPMat consists of MSV-VW containing MSV-MatA V1 and V2 sequences. *A. tumefaciens* C58C1 [pMP90] (Koncz & Schell, 1986) was used to deliver viral DNA to maize seedlings by agroinoculation as described in Chapter 2. Seed of the hybrid maize genotype PAN6099 was obtained from D. Nowell (PANNAR Ltd., Greytown, South Africa).

A single maize plant that was obtained from a plantation in Greytown (South Africa) and that had been naturally infected with the MSV-A strain, was generously provided by Dr D. Shepherd, Molecular and Cell Biology Department, U.C.T.

Agroinoculation and screening of plants

Maize PAN6099 seedlings were germinated and agroinoculated as described in Chapter 2, except that inocula consisted of an equal mixture of MatMPCPVW- and VWMPCPMat-containing *A. tumefaciens* cultures. Batches of c.a. 100 seedlings were agroinoculated in each experiment. Seedlings were checked for symptom development weekly. Because the PAN6099 genotype is almost entirely resistant to infection with either MatMPCPVW or VWMPCPMat, symptoms were taken as an indication that recombination was likely to have taken place, and symptomatic plants were therefore selected for DNA isolation, at times ranging between forty and sixty days after inoculation.

Extraction, purification and cloning of viral DNA

This protocol is based on plasmid preparation protocols developed by others, and attempts to avoid or solve problems encountered in alkaline lysis/column purification protocols. Shearing of plant genomic DNA is a problem in methods based on pH-dependent denaturation-renaturation-precipitation of high molecular weight DNA. In order to liberate the maximum amount of viral DNA from plant tissue, extensive grinding of the material and subsequent shaking in extraction buffer is desirable. However, such treatments tend to shear the plant genomic DNA, ultimately making it very difficult or impossible to separate viral DNA from plant genomic fragments by denaturation-renaturation, ion-exchange column purification, or gel electrophoresis. However, in this protocol I used isopycnic gradient ultracentrifugation to separate supercoiled MSV dsDNA from relaxed, linear plant genomic dsDNA, based on their different buoyant densities. This approach has a number of advantages: i) it is unaffected by the degree of shearing of plant genomic DNA; ii) the method concentrates ds ccDNA, in addition to separating it from linear ds and ss DNA, making it efficient for obtaining virus DNA from relatively small samples or relatively poorly infected material; iii) the efficiency of the method is expected to be relatively independent of the size of the DNA species being isolated,

making it more reliable for approximating the relative ratios of subgenomic DNAs and full-length DNAs; iv) the method is relatively quick and simple to perform, and avoids the use of phenol.

Infected leaf material (0.5g to 2g) was frozen in liquid nitrogen, and ground very finely using a mortar and pestle before being transferred to a sterile disposable 15ml centrifuge tube. The ground plant material was simultaneously resuspended and thawed in 6ml extraction buffer (0.1M Tris-HCl, 0.1M NaCl, 0.1M EDTA, 1% (w/v) SDS, pH7) and shaken moderately for 10min. Debris was pelleted by centrifugation at 5000 x g for 10 minutes at 20 °C, and 5.5ml supernatant was removed to a sterile 15ml centrifuge tube containing 5.5g CsCl. The CsCl was dissolved by moderate shaking and the solution centrifuged at 5000 x g for 10min at 20 °C. A green floc formed on the surface, and 5ml solution was recovered from below the floc. Four hundred microlitres ethidium bromide solution (10mg/ml) and 1µg pBluescriptSK+ (pSK+; Stratagene, La Jolla, CA) plasmid DNA were added; the plasmid DNA co-separates with viral DNA during isopycnic centrifugation, providing a visual marker for recovering the appropriate region from the gradient. The mixture was placed in a 5ml Quick-Seal® (Beckman Coulter, Inc.) ultracentrifuge tube and centrifuged overnight in a Beckman Vti65.2 rotor at 55 000 rpm (~275 000 x g) at 20 °C. Two DNA bands were visualised using U.V. light at 310nm, and the thinner, lower band was collected by side-puncture using a wide-bore (18 gauge) hypodermic needle and a 2ml sterile disposable syringe. The collected fraction was transferred to a 2ml eppendorf tube, and the ethidium bromide extracted twice with equal volumes of salt-saturated isopropanol (isopropanol equilibrated with 5M NaCl in TE buffer). Two volumes of sterile ddH2O were added, and the total mixed thoroughly. One volume of isopropanol was added, the solution mixed and then left on ice for 10min. The DNA was pelleted by centrifugation at 13 000 rpm in a microfuge for 15 minutes, washed in 70% ethanol, dried briefly, and finally re-dissolved in 100ul of the appropriate restriction enzyme buffer.

Twenty units of restriction enzyme (*Eco*RI for MSV-Tas; *Bam*HI for all other viruses in this study) were added to 18ul pSK+/virus DNA and the mixture incubated at 37 °C for two hours. The DNA was purified from the restriction enzyme by ethanol precipitation (Sambrook *et al.*, 1989) and re-dissolved directly in 18 µl ligation buffer. Two units of ligase were added, ligation carried out overnight at 15 °C, and the resulting DNA used to transform competent *E. coli* DH5α according to standard methods (Sambrook *et al.*, 1989).

Standard blue/white selection on media containing IPTG and X-gal was used to identify bacterial colonies containing recombinant plasmids. White colonies were streaked onto Hybond™-N+ (G.E. Healthcare) membranes on Luria agar (10µg/ml ampicillin) and onto replica Luria agar (10µg/ml ampicillin) plates, which were then incubated overnight at 37 °C. Cells were lysed on the membranes and the released DNA bound according to the protocol entitled "Colony and plaque lifts" described in Hybond™-N+ Instructions (Amersham Biosciences UK Ltd., 2003).

DNA bound to membranes was detected using an equimolar mixture of MSV-Kom and MSV-Set DIG-labelled DNA probes. Probes were synthesised with the PCR DIG Probe Synthesis Kit (Roche Applied Science) using as templates agarose gel-purified MSV-Kom or Set-SV DNA released with *Bam*HI from pKNcoδ or pSNcoδ (described in Chapter 2). Probes of 1.3 kbp spanning C1, LIR and the 5' half of V1 were amplified and labelled using the degenerate forward primer MSV-DEG1 (5' – TTGGVCCGMVGATGTASAG – 3') and the degenerate reverse primer MSV-DEG2 (5' – CCAAKDTCAGCTCCTCCG – 3') (Willment *et al.*, 2001). Membranes were blocked and washed using the DIG Wash and Block Buffer Set (Roche Applied Science), and bound probe was visualised using a DIG Luminescent Detection Kit (Roche Applied Science) and a Syngene GeneGnome imaging system (Synoptics Ltd., Cambridge, U.K.).

Using universal M13 sequencing primers (M13-Fwd, 5' – CCCAGTCACGACG TTGTAAAACG – 3'; M13-Rev, 5' – AGCGGATAACAATTTCACACAGG – 3') plasmid inserts were amplified by PCR from all colonies (Gussow & Clackson, 1989) identified as containing virus DNA. Agarose gel electrophoresis was used to determine the sizes of the amplified products, and products of 2.7 kbp were identified as putative full-length mastrevirus genomes.

Screening and analysis of virus clones

Universal M13 sequencing primers (M13-Fwd, 5' – CCCAGTCACGACGTTG TAAAACG – 3'; M13-Rev, 5' – AGCGGATAACAATTTCACACAGG – 3') were used to amplify plasmid inserts by PCR (Gussow & Clackson, 1989) from all colonies identified as containing virus DNA. The sizes of the amplified products were determined by electrophoresis in 1% agarose gels. To identify recombinant virus genomes, inserts of 2.7kbp were digested with *Hpa*II, *Rsa*I or *Sau*IIIa and the resulting DNA fragments separated electrophoretically using 2% agarose gels. The resulting banding patterns were compared with those of MatMPCPVW and VWMPCPMat to determine whether recombination had occurred.

DNA sequencing and sequence analysis

DNA purification, sequencing and sequence analysis were performed as described in Chapter 3 (page 58).

Construction of infectious recombinant progeny virus clones and assessment of virulence

Agroinfectious constructs of selected recombinant progeny virus clones were assembled and the virulence of the viruses assessed as described previously (Martin *et al.*, 1999).

Results

We used two reciprocal chimaeric viruses, MatMPCPVW and VWMPCPMat (see Figure 4-1 & Figure 4-5, and Martin *et al.*, 2002), that differed at 11% of their

nucleotide positions to study recombination between co-infecting mastreviruses *in planta*. To improve our chances of isolating recombinant progeny, we used the partially MSV-resistant maize cultivar PAN6099 to distinguish between infections caused by individual parental chimaeric viruses – which are barely infectious in this host (Martin & Rybicki, 2002) – from infections comprising recombinant progeny viruses with (partially) restored virulence. We agroinoculated approximately 250 seedlings with mixed inocula of MatMPCPVW and VWMPCPMat, and identified thirty unambiguously symptomatic plants (data not shown). Based on experience gained while attempting to recover viral DNA from some very mildly infected plants in the course of the work presented in Chapter 3, and to preclude any doubts regarding mutations (Ge *et al.*, 2007) and/or recombination (Paabo *et al.*, 1990; Odelberg *et al.*, 1995; Zaphiropoulos, 1998) that might occur during PCR amplification, we developed and used a novel protocol for isolating and cloning MSV DNA directly from infected plants (described in *Methods*). This method yielded libraries from mildly infected plant tissue in which up to 80% of clones contained MSV DNA (determined by colony hybridisation; data not shown). Ultimately we isolated and cloned viral DNA from fifteen of these plants, and characterised between thirty and one hundred MSV DNA clones from each plant by colony PCR and RFLP analysis (data not shown). RFLP analysis revealed five plants that contained MatMPCPVW parental virus and one plant that contained VWMPCPMat parental virus. Candidates of each of these were sequenced to confirm that no recombination had occurred that had not been detected via RFLP analysis (data not shown).

Recombination patterns

We sequenced a selection of recombinant MSV genomic clones from each of six plants, based on unique RFLP patterns and available resources (Figure 4-1). Each clone presented here is identified by a two-component designation of the form x-yy, where x specifies the plant from which it was isolated, and yy is an arbitrary number

assigned to successive clones in each library. To gauge small-scale differences between clones and to get a feel for the power of our RFLP analysis to distinguish unique recombinants, we sequenced five genomes from plant 5 (two unique recombination patterns) and three genomes from plant 6 (two unique recombination patterns). In addition, we sequenced a single complete genome from each of plants A, 15, 23 and 24.

None of the breakpoints comprised *BamHI* restriction sites, so we are confident that the recombination patterns did not result from restriction and relegation during cloning of the viral DNA (data not shown). All plants appeared to contain virus populations exhibiting a single major recombination pattern, with minor variations between individuals comprising point mutations, small differences at individual recombination break-points, and/or minor additional recombination events (Figure 4-1). In general, two major classes of recombinants could be discerned. In the first and simplest of these, comprising between two and six cross-over events, progeny virus genomes essentially consisted of fusions of VWMPCPMat V-sense sequences and MatMPCPVW C-sense sequences with one invariable breakpoint at the V-sense origin of replication, and the other breakpoint(s) in the SIR and/or 3' region of V2 (clones A-01; 6-01, -10, and -46; 23-05; and 24-01). The second group of recombinant genomes contained between eight and eighteen cross-over regions, none of which occurred precisely at the V-sense origin of replication (clones from plants 5 and 15). All the breakpoints in this second group of genomes fell in the V-sense half of the genome. The total number of mutations seen in the twelve clones that we sequenced was fifteen. This translates to an average mutation frequency of 4.7×10^{-4} subs/site over the 60-day duration of the experiment, or 2.8×10^{-3} subs/site/year. We examined the sequences within regions containing cross-over sites, and the adjacent 20-bp regions on either side of each cross-over region, and found neither bias of GC content nor any other obvious distinguishing features (data not shown).

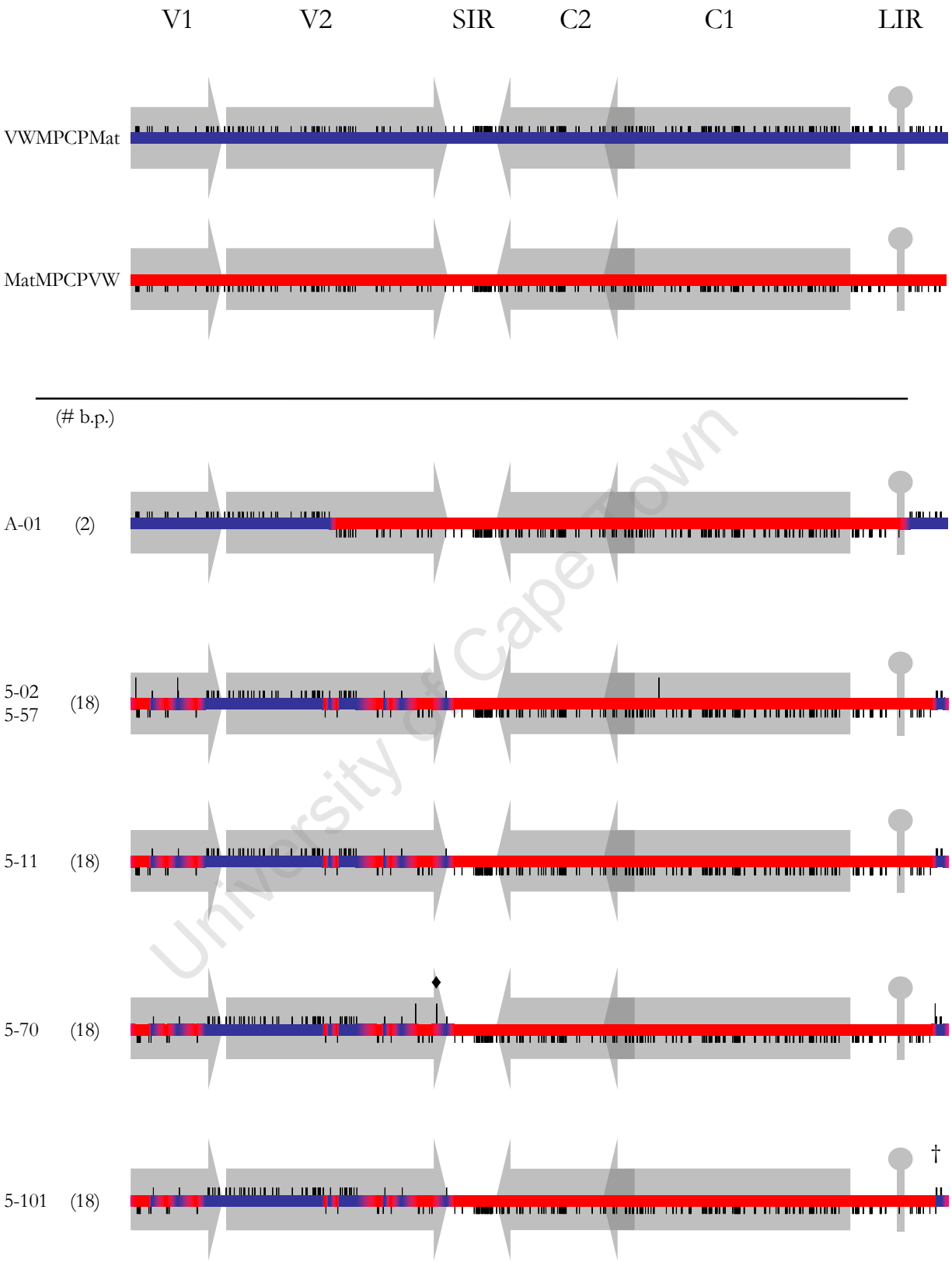


Figure 4-1~~Figure 4-1~~. ***(continued on next page)***

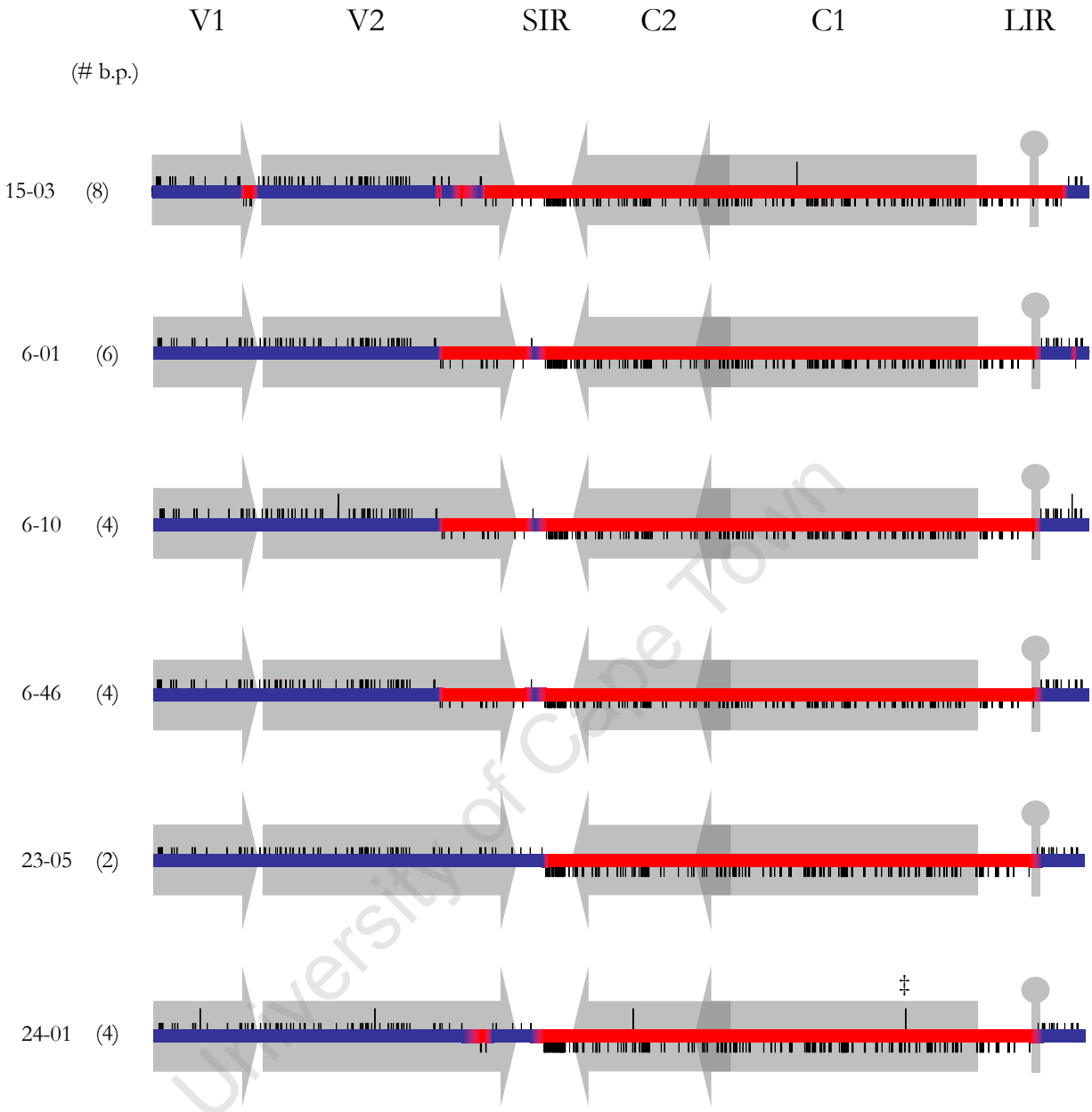


Figure 4-1. Recombination patterns and point mutations in full-length genomes.

The first two diagrams depict the chimaeric parental (input) viruses. Recombinant genomes are designated x-yy according to the plant from which they were isolated (x) and the number of the clone in the relevant library (yy). The number of break-points in each clone is given in parentheses. A red centre line indicates sequences inherited from MatMPCPVW, while blue indicates VWMPCPMat-derived sequences. Short vertical lines above or below the centre line indicate homology at informative sites to either MatMPCPVW or VWMPCPMat, respectively. Long vertical lines represent positions not homologous with either MSV-Mat or MSV-VW sequences. The positions of ORFs and the virion-sense origin of replication stem-loop sequence are indicated in shaded grey. The diagrams are to scale. † The breakpoint in this clone was slightly different to that in the two above. ‡ This mutation introduced a premature stop codon: AAG → TAG (C-sense). ♦ This single-base deletion resulted in a nonsense mutation.

Regional distribution of recombination breakpoints

Examination of our recombinant virus genomes (Figure 4-1) indicated that recombination cross-over events might have occurred more frequently within some genomic regions than in others. Figure 4-2 is a frequency distribution plot of breakpoints across the genome, and it indicates that more breakpoints occurred around the V-sense origin of replication and toward the 3' end of the V2 ORF than elsewhere.

Recombinant genomes were fitter than their parents

To test our hypothesis that the recombinant progeny we recovered would have had to be more competitive during the infection cycle than both parental input viruses, we quantified the virulence of selected recombinant daughter viruses isolated from plants 5 and 6. For comparative purposes, we also quantified the virulence of both parental chimaeras (MatMPCPVW and VWMPCPMat) as well as their constituent wild-type genomes (MSV-Mat and MSV-VW) in sweetcorn and in the PAN6099 cultivar. As described above, sweetcorn is a permissive host, whereas the PAN6099 cultivar is partially MSV-resistant and was the host that we used to generate the recombinant viruses. This also served to establish whether some of the genomes that we isolated were viable. We made dimeric constructs of clones 5-02, 5-11, 5-101, and 6-01, and obtained infections in sweetcorn following agroinoculation with each of them. To quantify virulence, the average percentage chlorotic leaf area was determined for successive leaves in plants that had been agroinoculated with each virus (Figure 4-3), which we interpreted as a surrogate measure of virus fitness (for rationale and methods, see Martin & Rybicki, 1998; Martin *et al.*, 1999; Martin *et al.*, 2001; Martin & Rybicki, 2002; Martin *et al.*, 2005a).

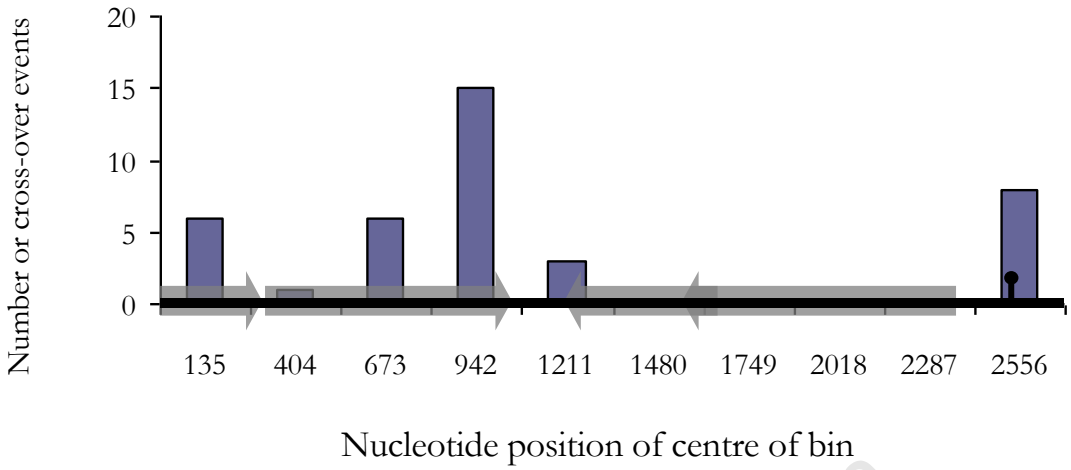


Figure 4-2. Genomic distribution of cross-over events.

Recombinant virus genomes were divided equally into ten regions of 269 bp, and independent cross-over events in full-length genomes were counted in each region. The approximate genomic position of each region is indicated by the cartoon on the x-axis depicting the ORFs (grey arrows) and origin of virion-sense replication (black stem-loop), while the nucleotide position corresponding to the centre of each of the ten regions is shown on the x-axis.

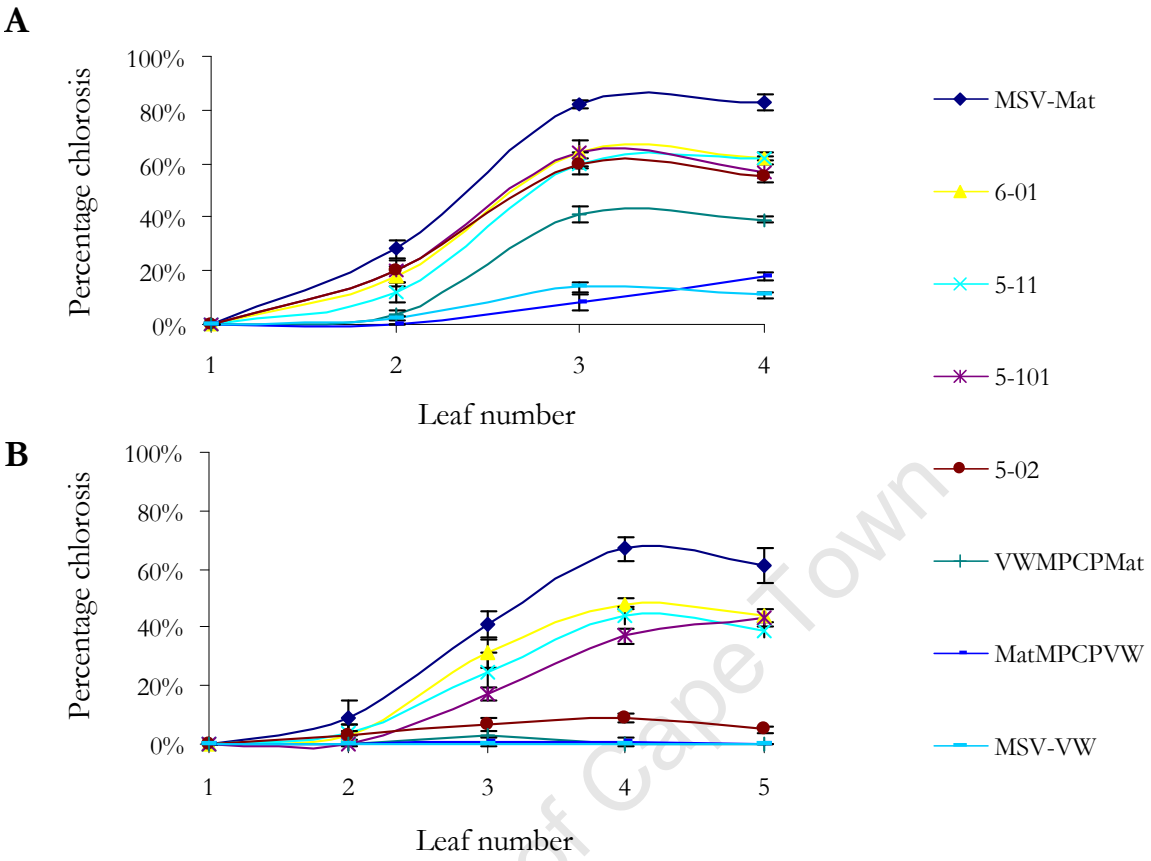


Figure 4-3. Symptom severity of recombinant viruses

The graphs show the percentage chlorotic area of successive leaves of infected plants following inoculation with either wild-type (MSV-Mat & MSV-VW), parental recombinant (MatMPCPVW & VWMPCPMat), or recombinant progeny viruses. **(A)** Symptom severity in sweetcorn, a relatively susceptible host. **(B)** Symptom severity in PAN6099, a relatively MSV-resistant cultivar, and the host used to generate the recombinant viruses in this study.

In both host cultivars, all of the recombinant daughter viruses were fitter than their parental chimaeric parents (MatMPCPVW and VWMPCPMat) and fitter than the wild-type MSV-VW. In descending order of virulence, which was approximately preserved in both hosts, the viruses were as follows: MSV-Mat; 5-11, 5-101 and 6-01; 5-02; VWMPCPMat; MatMPCPVW and MSV-VW. Note that the use of two hosts with different levels of infection susceptibility provided greater resolution in separating the viruses on the basis of virulence: some of the very pathogenic viruses

tended to bunch together at the top of the scale of virulence when analysed in sweetcorn, but were distinct from one another in terms of their virulence in PAN6099 plants. Conversely, very weak viruses were apparently incapable of infecting PAN6099, but could be rated effectively in sweetcorn.

Recombination resulted in MSV-Mat-like genotypes

Unsurprisingly, the recombinant progeny genomes that we recovered from plants co-inoculated with MatMPCPVW and VWMPCPMat were genotypically similar to MSV-Mat (Figure 4-4~~Figure 4-4~~). A neighbour-joining clustering dendrogram shows that all of the recombinant daughters were much more similar to MSV-Mat than they were to MSV-VW (Figure 4-4~~Figure 4-4~~-A). Furthermore, all of the recombinant progeny genomes were more akin to MSV-Mat than they were to MatMPCPVW. The accompanying bar graph illustrates the relative fitness of each virus with respect to MSV-Mat.

To visualise the relative individual contributions made by recombination and point mutation to the diversification of progeny viruses, a graph of the pairwise Hamming distances between all sequences was constructed using three dimensions (Figure 4-4~~Figure 4-4~~-C): co-ordinates on the x- and y-axes respectively represent distances from the wild-type reference sequences and from the chimaeric parental sequences due to recombination alone, while the areas of the circles representing each sequence are scaled according the distance of each sequence from parental sequences attributable to novel point mutations. In this graph, all of the progeny viruses lie well towards the MSV-Mat side (red area in the figure) of the line marking equidistance from the two wild-type sequences, indicating that recombination resulted in daughter sequences that were very similar to MSV-Mat. Conversely, the cluster of recombinant sequences straddles the line (dashed line in the figure) dividing sequences that are more similar to VWMPCPMat from those that are more

MatMPCPVW-like, indicating that there was little tendency to retain a majority of either parental chimaeric virus sequence.

To further illustrate the extent to which recombination had resulted in MSV-Mat-like daughter genotypes, and to show which regions of MSV-VW genomic sequence remained in the progeny, we analysed each genome to determine its regional identity to either MSV-Mat or MSV-VW wild-type ancestral sequences (Figure 4-5). The original junctions between MSV-Mat and MSV-VW genomic sequence from the parental chimaeric viruses were present in all of the clones except those from plant 5, in which a cross-over event between the parental chimaeras most likely occurred in the same region (i.e. between the same two informative sites) during co-infection. None of the recombinant viruses contained any MSV-VW sequences in the C-sense of their genomes. Most genomes retained small regions of MSV-VW sequence only between the V-sense origin of replication and the start of the V1 ORF, and in the 3' portion of the V2 ORF and the 5' portion of the SIR. Assuming that each cross-over event occurred in the middle of the area of uncertainty between the relevant informative sites, we estimated the proportions of each recombinant virus that comprised either MSV-Mat or MSV-VW genomic sequence (Figure 4-5). According to this estimate clone A-01 inherited the least MSV-Mat sequence (82%) and clones 15-03 and 23-05 inherited the most (91%).

Impact of recombination and mutation on proteins expressed from progeny genomes

In silico translation of the V1 ORFs of progeny viruses showed that recombinants from plant 5 expressed MPs that were very similar to the wild-type MSV-VW sequence, while the remaining recombinant progeny retained almost wild-type MSV-Mat MP sequences (Figure 4-6, and compare with Figure 4-5). The single point mutation in the V1 ORF of clone 24-01 (Figure 4-1) was silent, but both novel polymorphisms in this region of clones 5-02/57 resulted in amino acid substitutions: S6T, which is a fairly conservative change at a site that varies between MSV-Mat (A

at this position) and MSV-VW; and D52Y, a radical change from a small negatively charged residue to a large, aromatic, hydrophobic amino acid within the putative transmembrane domain.

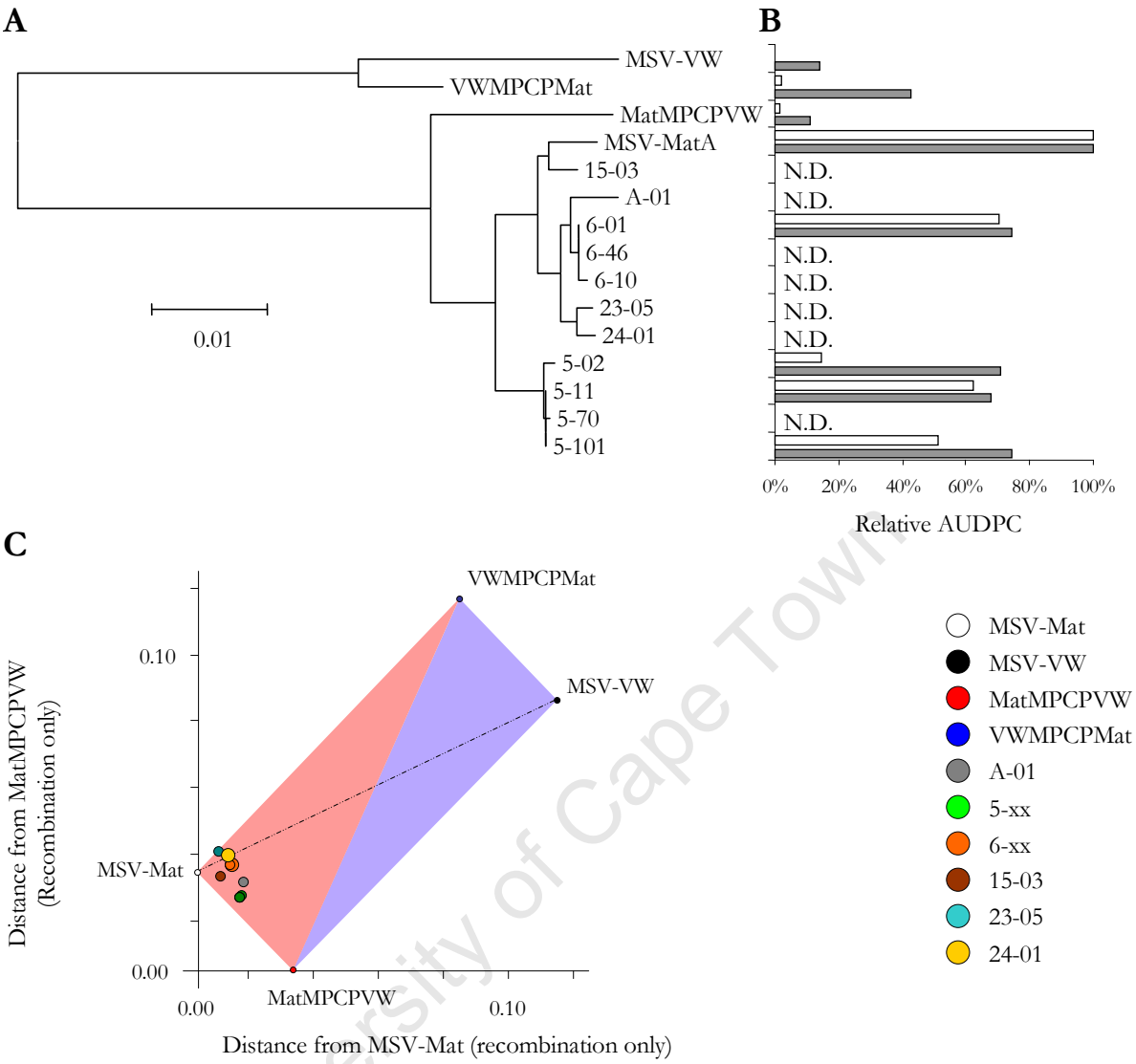


Figure 4-4. Generalised convergence of recombinant sequences on that of MSV-Mat.

(A) A neighbour joining clustering dendrogram constructed using pairwise Hamming distances (with pairwise deletion of gaps) between progeny recombinants, parental recombinants, MSV-Mat and MSV-VW. Rather than describing the evolutionary history of these sequences, this tree is intended as a graphical demonstration that recombination events tended to produce viruses that were more MSV-Mat-like. **(B)** The virulence (AUDPC of chlorotic leaf area) of recombinant progeny genomes relative to recombinant parental viruses, MSV-VW, and MSV-Mat, in MSV-sensitive (sweetcorn; grey bars) and MSV-resistant (PAN6099; white bars) maize genotypes. N.D., not determined. **(C)** A graph of the pair-wise Hamming distances between recombinant progeny viruses, parental recombinants, MSV-VW and MSV-Mat. Each point represents one genome sequence with the area of the point scaled according to the divergence from parental sequences due to novel point mutations (polymorphisms different to those observed in either MSV-Mat or MSV-VW; an arbitrary value of 0.002 was added to all data points to obtain non-zero areas for points representing parental sequences). Whereas the clustering of distances in the red area indicates a clear tendency for recombinants to become more MSV-Mat-like, the distances straddle the dashed line indicating that there is not a similar tendency for them to remain more like either one of the two parental recombinant sequences.

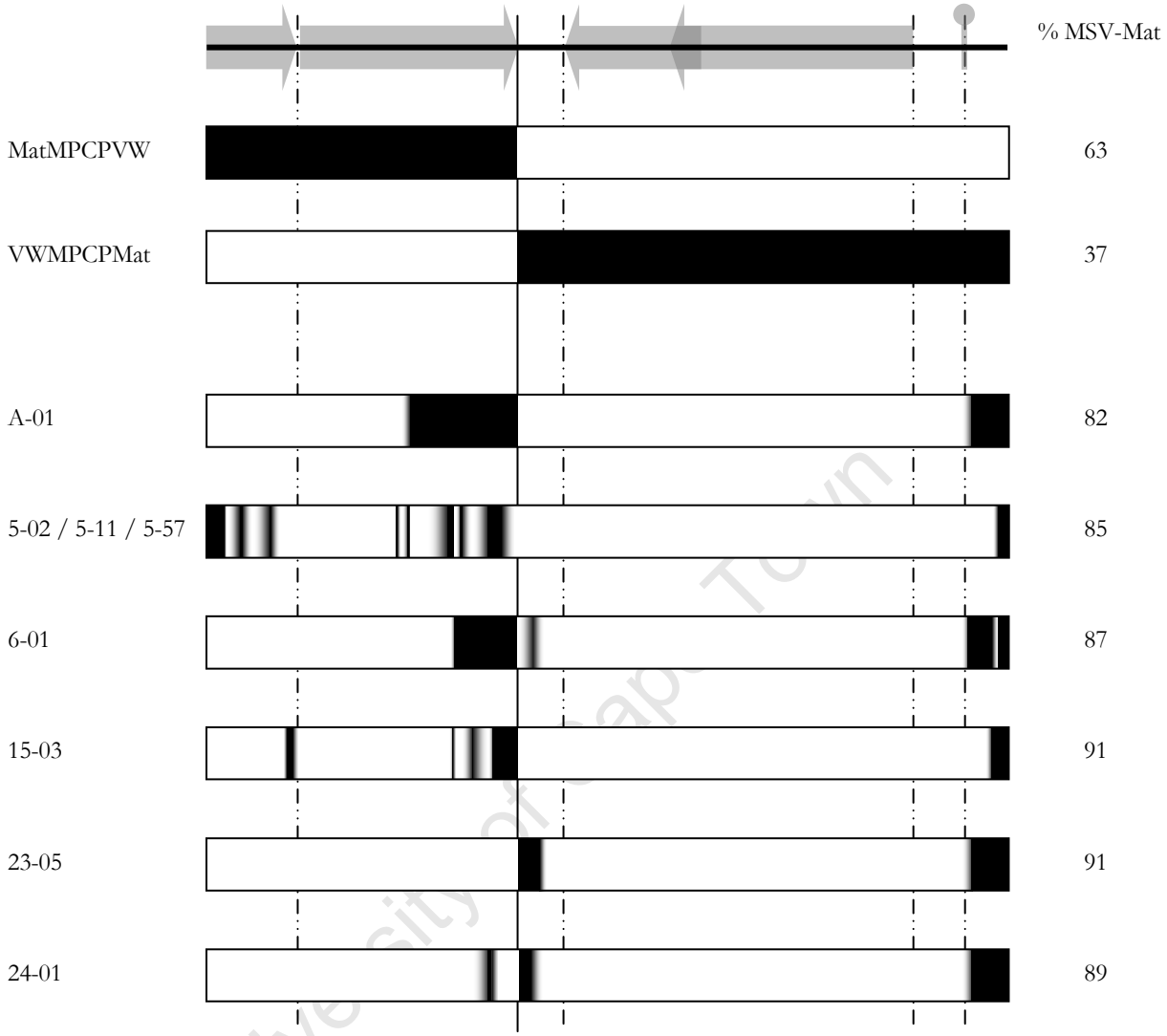


Figure 4-5. Wild-type virus composition of full-length recombinant genomes.

Parental (MatMPCPVW & VWMPCPMat) and recombinant genomes are depicted with white indicating MSV-Mat sequences and black denoting MSV-VW sequences. The estimated percentage of MSV-Mat sequence in each recombinant is shown in the column at the right of the figure. ORFs and the virion-sense origin of replication are shown in grey at the top. Vertical lines demarcate the intergenic regions and the position of the virion-sense origin of replication; the solid vertical line indicates the junction between MSV-Mat and MSV-VW sequences in the original parental virus constructs.

Because the MSV-Mat and -VW CP sequences are identical at all positions over the last ~144 C-terminal residues (Figure 4-7) all the recombinant clones containing MSV-VW sequences in the 3' portion of their V2 ORFs (Figure 4-5) nevertheless

encoded MSV-Mat-like CPs (Figure 4-7). The single point mutation in the V2 ORF of clone 6-10 (Figure 4-1) resulted in the change D72A in the corresponding CP sequence (Figure 4-7). The substitution in the V2 ORF of clone 24-01 was silent. Both mutations in the 3' portion of the V2 ORF of clone 5-70 affected the corresponding CP sequence: the 5'-proximal substitution resulted in V208I, while the single-base deletion nearer the 3' end of the ORF caused a frame-shift, resulting in completely altered protein sequence from residue 231 onwards, and loss of six amino acids from the C-terminal end due to premature termination.

Genomes 5-02/57 and 15-03 each had a single point mutation in the C1 ORF (Figure 4-1), and both mutations gave rise to substitutions in the respective RepA (Figure 4-8) and Rep (Figure 4-9) protein sequences: F213L in 5-02/57 RepA and Rep, which is relatively conservative; and S175F in 15-03 RepA and Rep, which constitutes a significant change from a very small polar residue to a much larger, aromatic, hydrophobic amino acid. The nonsense mutation near the 3' end of the C1 ORF in clone 24-01 noted in Figure 4-1 resulted in premature termination of both RepA and Rep proteins, but the theoretical translations of the remainder of the ORFs is provided nevertheless, because this data may be relevant to other individuals from that population that were not sequenced. With this in mind, the only other mutation affecting Rep in this genome resulted in K301Q, where MSV-VW had a methionine at this position.

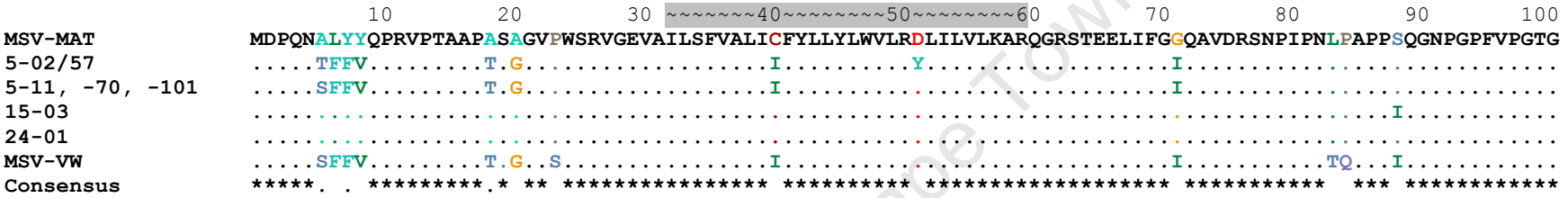


Figure 4-6. Multiple sequence alignment of translated recombinant and/or mutant V1 ORFs (MP).

Identity with MSV-Mat is depicted by full-stops. Variable residues are coloured according to functional similarity. The putative trans-membrane domain is demarcated in grey.

Figure 4-7. Multiple sequence alignment of translated recombinant and/or mutant V2 ORFs (CP).

Identity with MSV-Mat is shown with full-stops. Variable residues are coloured according to functional similarity. Features are as follows: grey, the nuclear localisation signal; and yellow, the general region thought to be involved in DNA binding.

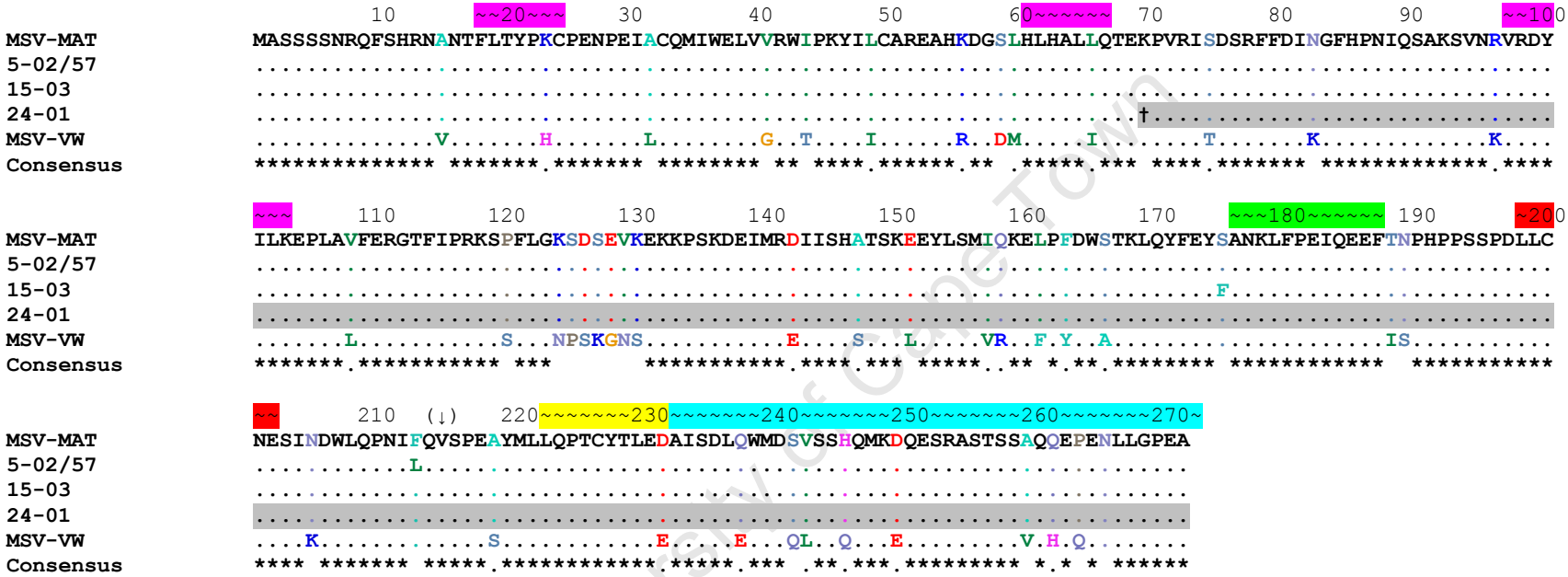


Figure 4-8. Multiple sequence alignment of translated recombinant and/or mutant C1 ORFs (RepA).

Identity with MSV-Mat is shown with full-stops. Variable residues are coloured according to functional similarity. † The sequence for clone 24-01 terminates prematurely due to a nonsense mutation. Features are as follows: pink, rolling-circle replication motifs; green, oligomerisation domain; red, Rb-binding motif; yellow, transactivation domain; and turquoise, putative GRAB-interaction domain. (↓) Indicates the corresponding position of the C1:C2 splice site in the *Rep* transcript; *RepA* and *Rep* are identical up to this point.

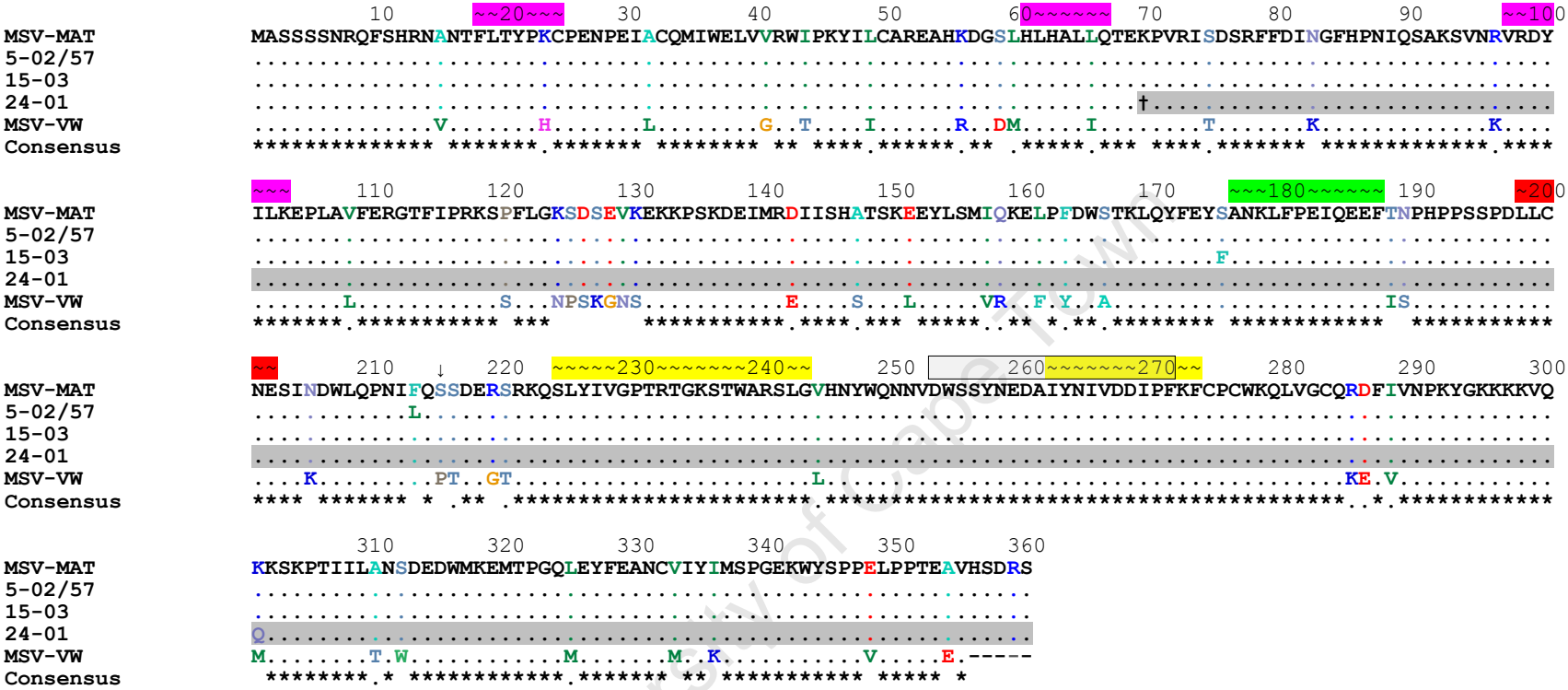


Figure 4-9. Multiple sequence alignment of translated recombinant and/or mutant spliced C1-C2 ORFs (Rep).

Identity with MSV-Mat is shown with full-stops. Variable residues are coloured according to functional similarity. † The sequence for clone 24-01 terminates prematurely due to a nonsense mutation. Features are as follows: pink, rolling-circle replication motifs; green, oligomerisation domain; red, non-functional Rb-binding motif; yellow, dNTP-binding motif; and shaded box, *myb*-like transactivation domain. ↓ Indicates the corresponding position of the C1:C2 splice site in the *Rep* transcript; *RepA* and *Rep* are identical up to this point.

Subgenomic and re-arranged sequences

PCR amplification of plasmid inserts from clones in our MSV genomic DNA libraries revealed a large proportion of DNA fragments that were smaller than a full-length MSV genome (2.7 kb; i.e. "subgenomic"), as well as some that were slightly larger (up to ~3 kb; data not shown). We attempted to sequence selected representatives of these clones, which revealed radical genomic rearrangements (Figure 4-10). We were unable to obtain complete sequences for many of the longer inserts and some of the shorter inserts, presumably because our internal, MSV-specific sequencing primers did not bind appropriately to rearranged and/or repeated DNA sequences (data not shown). Nevertheless, a diagrammatic representation of a selection of sequences that were successfully completed is presented here (Figure 4-10). Using the same methods, we subsequently isolated and sequenced DNA from a maize plant that had been naturally infected by the MSV-A strain, and obtained similar subgenomic DNA fragments, which are included in our analysis here.

These subgenomic fragments encompassed examples of large deletions, translocations, and inverse as well as direct repeats. Inspection of the sequences confirmed that the junctions between discontinuous segments of DNA did not comprise restriction sites, so we are confident that the genomic arrangements reported here represent *bona-fide* examples of MSV DNA in the infected plants that we studied. Moreover, three of the plasmid inserts contained substantial stretches of non-MSV DNA, all of which were matched with high confidence to maize DNA via BLAST search of the GenBank DNA sequence database (Table 4-1). Interestingly, one of these fragments matched mitochondrial DNA sequence, while the other two were chromosomally-derived.

All of the subgenomic DNA fragments that we analysed displayed recombination signatures compatible with the corresponding full-length genomes that we isolated

from the same plants (Figure 4-10). Short stretches of between one and four nucleotides of DNA of unconfirmed origin (shown in green in Figure 4-10) were found between some adjacent blocks of re-arranged MSV-sequence, but many junctions consisted of precise transitions from one perfectly matched sequence segment to another. Strikingly, several of these re-arranged sequences had breakpoints at very similar positions in the SIR, aside from breakpoints at various other positions in the genome. A multiple sequence alignment comparing these junctions with breakpoints in the same region in some of the full-length genomes revealed a very high density of breakpoints within a ~25 bp region near the 5' edge of the SIR (Figure 4-11).

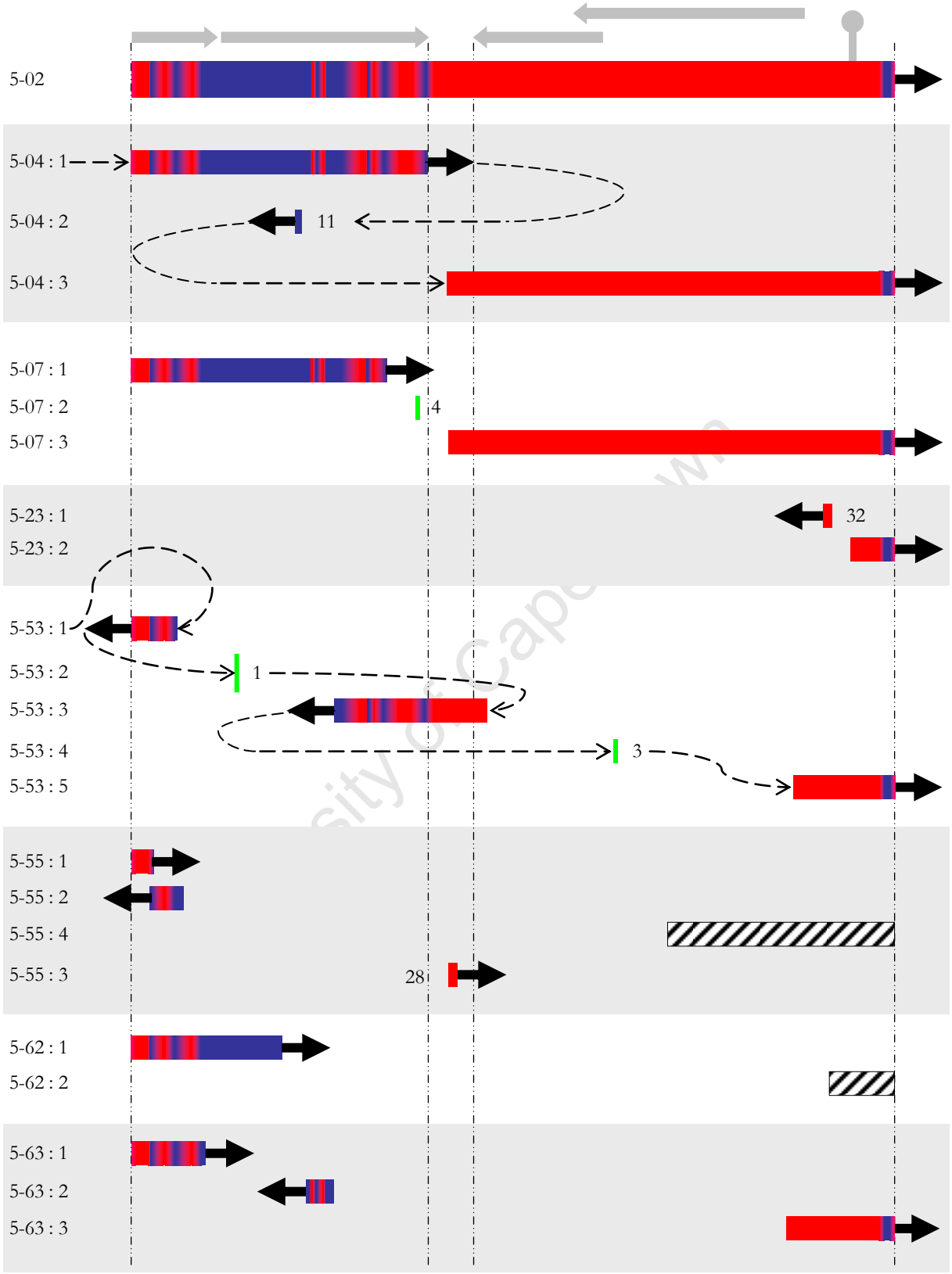


Figure 4-10~~Figure 4-10~~ ***(continued on next 2 pages…)***

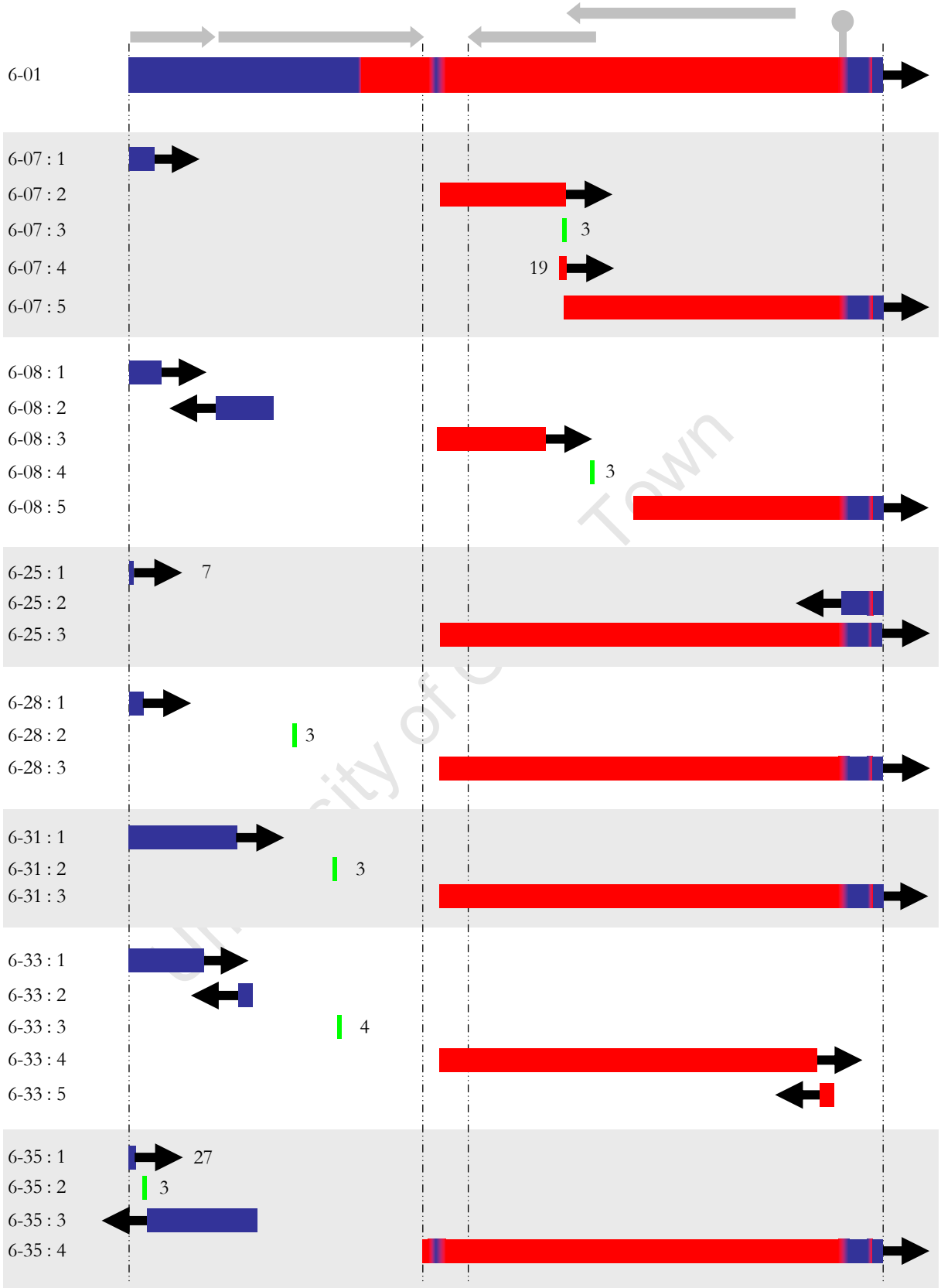


Figure 4-10~~Figure 4-10~~ ***(continued on next page…)***

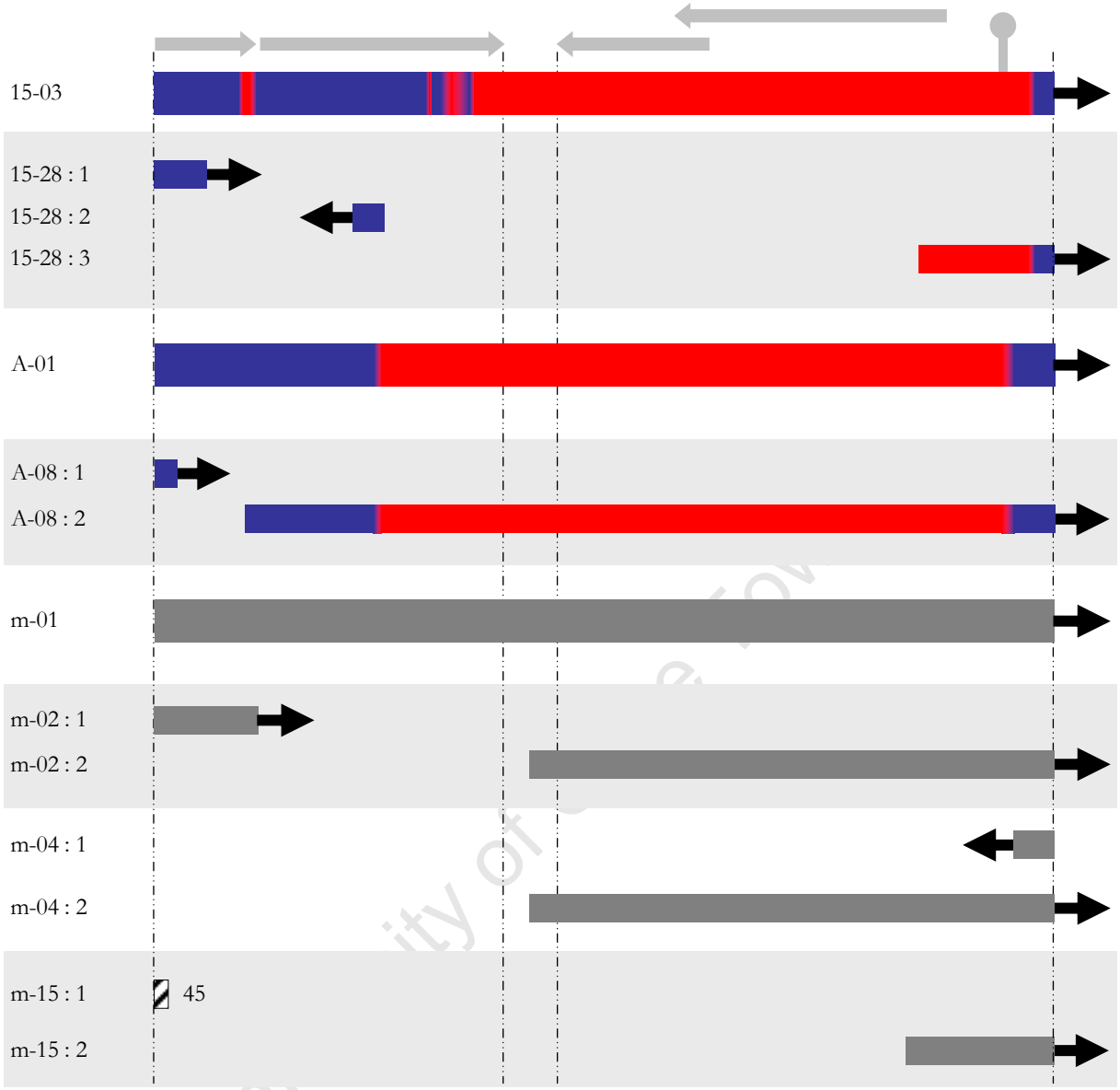


Figure 4-10. Genomic composition and arrangement of subgenomic genomes.

Representatives of full-length genomes isolated from each plant are shown above subgenomic sequences obtained from the same plant. Two subgenomic diagrams on the first page (5-04 and 5-53) explain how to read the cartoons (follow the dashed arrows): each clone is divided into its composite sequence elements which are numbered consecutively (5-04:1, 5-04:2, etc.) from 5' to 3', and represented on successive lines so that they are aligned with parental sequences where possible (i.e. overlapping sequence elements in a single clone represent either direct or inverted repeats). Thick black arrows indicate the direction of the corresponding sequence element in the clone (i.e. arrows pointing left indicate inverted sequences). MatMPCPVW sequences are red and VWMPCPMat seqences are blue; MSV isolated from a natural infection of maize is grey; sequence elements too short to identify are in green; non-MSV sequences are in white with black diagonal lines. The lengths of elements shorter than 50 bp are given in the diagram.

Table 4-1. Nucleotide BLAST results for non-MSV sequences identified in subgenomic sequences

Sequence (length)	Accession	Description	Max score	Total score	Query coverage	E value	Max ident
5-55 : 4 (802 bp)	DQ645539.1	*Zea mays* subsp. parviglumis mitochondrion Features in sequence: NADH dehydrogenase subunit 1 NADH dehydrogenase subunit 2	1447	1602	100%	0	100%
5-62 : 2 (239 bp)	DQ493649.1	*Zea mays* cultivar Coroico bz locus region	403	403	100%	10^{-109}	97%
m-15 : 1 (45 bp)	AF546188.1	Contiguous genomic DNA sequence comprising the 19-kDa-zein gene family from *Zea mays*	81.8	139	100%	10^{-13}	97%

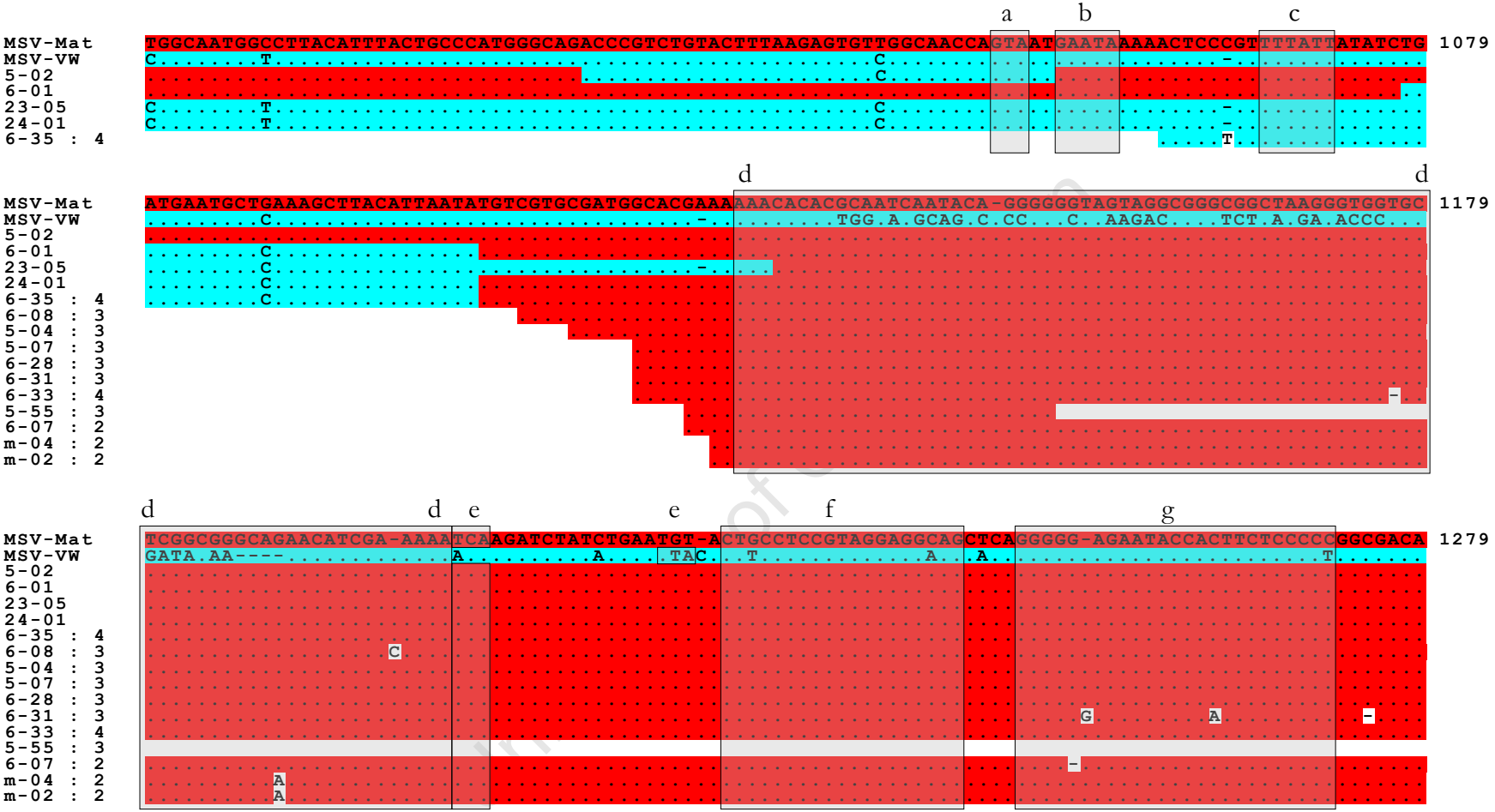


Figure 4-11. Multiple DNA sequence alignment of breakpoints in the SIR.

MSV-Mat sequences are highlighted in red; MSV-VW sequences are in blue. Identity with MSV-Mat is indicated with full-stops. Gaps are indicated with dashes. Positions are given in the MSV-Mat sequence, relative to the *Bam*HI site at the start of the V1 ORF. Indicated features are as follows: a, V2 stop codon; b, V-sense polyadenylation signals; c, C-sense polyadenylation signals; d, probable binding site of primer for C-sense replication; e, C2 stop codon; f & g, potential stem-loop sequences.

Discussion

We have shown experimentally that recombination can occur between distinct co-infecting species of MSV with high frequency, and that it represents an extremely powerful mechanism for rapid adaptation. The high resolution of our experimental system, arising from an absolute knowledge of the parental genomic sequences and from the high density of informative sites between parental viruses differing at 11% of their nucleotide positions, allowed us to document the occurrence of frequent cross-over events within individual genomes. One viable recombinant progeny genome comprised nine alternating sections of DNA sequences from distinct parents. Moreover, we were able to locate cross-over sites within relatively small regions of sequence (i.e. with relatively high precision). In contrast, recombination events typically detected in natural geminivirus populations are fewer per genome, and individual exchanged regions are considerably larger than the smallest exchanges that we report here (Martin *et al.*, 2001; Ndunguru *et al.*, 2005; Owor *et al.*, 2007a). This is probably due to a combination of factors which do not impinge on our experimental system: the limitations of the power of bioinformatic techniques to detect small recombination events between similar parental sequences (Posada & Crandall, 2001); incomplete sampling and therefore lack of knowledge of putative parental genomic sequences; and nucleotide substitutions becoming fixed in recombinant daughter sequences and thus obscuring their relationships to parental sequences. Our advantage here was in knowing exactly what the parental genomes were – which made our detection of recombination events very precise. In attempting to understand the biochemical processes that shape recombination patterns, one important factor complicates the interpretation of all recombinant genomes generated *in vivo* – natural selection.

Both chimaeric parental viruses that we used in this study were barely capable of infecting the PAN6099 cultivar that we used to select recombinant progeny. However, between them these chimaeric viruses comprised two complete and very

different wild-type MSV genomic sequences: MSV-VW, which cannot infect PAN6099 plants, and MSV-Mat, which is highly virulent in PAN6099. Within sixty days (but probably much more quickly), MSV-VW and MSV-Mat sequence segments in the chimaeric parents were re-shuffled and the recombinant progeny sorted so that ~90% of the MSV-Mat sequences were ultimately re-united with one another (Figure 4-5), and these recombinant daughters had reclaimed up to 75% of the fitness of MSV-Mat (Figure 4-3 and Figure 4-4~~Figure 4-4~~). Had the full complement of input sequences been maintained in the plants for a longer period, as might be the case in some natural co-infections (Jovel *et al.*, 2007) and in cases of constant re-occurrence of co-infection, there is good reason to believe that full recovery of wild-type MSV-Mat levels of fitness may have been achieved within relatively short periods of time.

Recombination patterns reflect selection as well as biochemical recombination mechanisms

The strong selection pressure that was undoubtedly responsible for driving such rapid adaptation is likely to have resulted in the loss of many recombinant genomes that were less fit than those that were ultimately recovered and characterised. The recombination patterns we observed were probably influenced both by the anatomy of the chimaeric input viruses, and by selection of daughter sequences; the simplest and most efficient process for obtaining MSV-Mat-like genomes from the input viruses would involve the transfer of the entire V1-V2 cassette from VWMPCPMat to MatMPCPVW via two cross-over events. Indeed, the majority of the recombinant daughter sequences appeared to have arisen through this type of exchange. Thus, when considering the frequency distribution of recombination breakpoints in our data-set, it is reasonable to expect a greater number of breakpoints near the original junctions between wild-type sequences in the original chimaeric parental viruses.

Nevertheless, the presence of significant proportions of MSV-VW DNA sequences in the V1 ORFs – which were reflected in the corresponding MPs – of the recombinant progeny from plant 5 indicates that the MP was not a crucial determinant of fitness (Figure 4-5 and Figure 4-6). We have shown previously that substitution of the MSV-Mat V1 sequence with corresponding regions from diverse viruses had the smallest negative impact on virulence, in comparison with exchanges of other coding regions (Martin & Rybicki, 2002; Martin *et al.*, 2005a). However, it is worth noting that the clones from plant 5 were the least virulent of the clones we tested (Figure 4-4~~Figure 4-4~~). Conversely, all of the recombinant virus genomes encoded CPs that were almost identical to that of MSV-Mat, despite some comprising significant proportions of MSV-VW DNA sequence in their V2 ORFs. This was at least partially due to the fact that MSV coat protein sequences are highly conserved, and the MSV-Mat and -VW CPs are consequently identical throughout the last two-thirds of their amino acid sequences, despite differences in their coding sequences. Perhaps the clearest evidence of selection for specific components of either wild-type ancestral virus was the complete absence of C-sense MSV-VW sequences in the recombinant daughter viruses (Figure 4-5). We previously showed that recombination with dissimilar homologous sequences generally has a substantial negative effect on the fitness of MSV-Mat (Martin *et al.*, 2005a), and the fact that all of the recombinant progeny viruses that we isolated here comprised more than 80% MSV-Mat sequences undoubtedly reflects this.

Notwithstanding the unquestionable role that selection played in determining the recombinant configurations of the genomes that we sampled, we found strong evidence of two mechanistically-determined breakpoint hotspots. Recombinant genomes isolated from multiple plants showed evidence of cross-over events that had occurred at the V-sense origin of replication. Similarly, many of the breakpoints that we observed in full-length genomes and in the rearranged subgenomic DNA fragments occurred within a ~25 bp region in the SIR, between the V2 stop codon

and the probable binding site of the primer for C-sense replication. While begomoviruses do not have a genomic region that is obviously equivalent to the mastrevirus SIR, it is relevant to note here that others have reported evidence of frequent cross-over events at the V-sense origin of replication and at the diametrically opposite junction between V-sense and C-sense genes in the A-components of diverse begomoviruses (Lefeuvre *et al.*, 2007; Garcia-Andres *et al.*, 2007b).

Recently, García-Andrés *et al.* (2007b) demonstrated recombination between the monopartite begomoviruses tomato yellow leaf curl Sardinia virus (TYLCSV) and tomato yellow leaf curl virus (TYLCV) during co-infection of tomato plants, and characterised the recombinants. They identified ten unique recombinant genomes, with each daughter comprising approximately half of either parental genome. In this study, all the recombinants had two cross-over sites, with one consistently at the origin of virion-sense replication. The second breakpoint was much less precisely and consistently defined, but most often occurred approximately at the diametrically opposite side of the circular genome, near the border between V- and C-sense ORFs. Similarly, Lefeuvre *et al.* (2007) identified recombination hotspots at the V-sense origin of replication and at the interface between terminating V- and C-sense genes in diverse begomovirus isolates sampled from natural populations in the islands of the south-west Indian Ocean. They proposed a sensible rationale for the existence of these recombination hotspots that accounts for both selection as well as biochemical factors.

First, they suggested that cross-over events that occur in intergenic regions are generally more likely to result in fit daughter viruses, because this avoids the disruption of intra-protein interactions that might result from breakpoints within coding regions. Second, they pointed out that any event which results in aborted V-sense synthesis during replication would result in an incomplete genome having a

defined 5'-end at the origin of replication. If such an incomplete genomic DNA fragment were repaired via homologous recombination mechanisms using a distinct, co-infecting genomic sequence as a substrate, then two cross-over events would ensue: one at the position at which the incomplete V-sense strand invaded the new, distinct template molecule, and the second at the origin of V-sense replication, when the 3'-end of the completed V-sense strand was re-ligated to its own tail by the Rep protein bound at its 5'-end. They conceived of a second possible mechanism to explain the fact that, according to their model, replication is most often aborted near the terminus of the C-sense coding regions: synthesis of the V-sense strand initially proceeds relatively unhindered in the same direction as transcription of the V-sense genes, but when synthesis continues into the C-sense region of the genome, the V-sense replication forks are more likely to be disrupted if they encounter C-sense transcription machinery proceeding in the opposite direction.

There is no direct evidence that this mechanism applies to geminiviruses, but replication-transcription collisions have been thoroughly documented, and have been shown to inhibit replication in both prokaryotes (French, 1992; Liu & Alberts, 1995; Elias-Arnanz & Salas, 1997; Elias-Arnanz & Salas, 1999) and eukaryotes (Haase *et al.*, 1994; Deshpande & Newlon, 1996; Aniskina *et al.*, 2003). Intriguingly, Olavarrieta and co-workers (2002) demonstrated that head-on replication-transcription collisions can result in DNA knotting, and proposed that such knotting may contribute to the deleterious effects that such collisions have on replication efficiency. The knotting and "knotted bubble" structures they described could help to explain some of the patterns of inverted and deleted segments seen in the subgenomic genomes that we sequenced. Perhaps of more direct relevance are a number of studies describing a phenomenon variously dubbed "transcription-dependent recombination" or "transcription-associated recombination" (TAR). These investigations have shown that transcription can interrupt replication – either via nascent RNA binding to the template DNA strand to form "R-loops" (Huertas & Aguilera, 2003; Yu *et al.*, 2003;

Huang *et al.*, 2006; Huang *et al.*, 2007), or via the direct collisions described above (French, 1992; Krasilnikova *et al.*, 1998; Takeuchi *et al.*, 2003) – and thereby cause recombination. The fact that the recombination breakpoint hotspot that we observed is so close to the 3'-end of the putative C-sense replication primer binding site (Figure 4-11) suggests that either synthesis of, or DNA extension from, this RNA primer may be additionally or alternatively responsible for interrupting replication at this position.

While the scenarios described above seem compellingly consistent with current thinking both on geminivirus molecular biology and on known mechanisms of recombination, there are other factors which may play important roles in geminivirus recombination. Indeed, any mechanism that results in pausing or blockage of replication is likely to result in DNA double-strand breaks and recombination (Wang *et al.*, 2004; Mirkin & Mirkin, 2007). It is well known that various mechanisms for avoiding replication-transcription collisions have apparently evolved in diverse organisms ranging from prokaryotes to eukaryotes, and it has been shown that in some cases even programmed replication pause sites (RPS), such as the *Ter* sites found at the *E. coli* replication terminus region, can result in recombination hot-spots (reviewed by Rothstein *et al.*, 2000). Thus, it is conceivable that geminivirus genomes may incorporate such replication pause sites to reduce the occurrence of replication-transcription collisions, and that recombination occurs more frequently at these RPSs. Another intriguing possibility is that high recombination rates in geminiviruses are related to a relatively recently discovered host-mediated systemic defence response characterised by increased somatic DNA recombination (Kovalchuk *et al.*, 2003), which is epigenetically heritable (Molinier *et al.*, 2006). Interestingly, a recent study found evidence that such defence response mediated recombination may be related to the regulation of transcription (Durrant *et al.*, 2007), which may implicate mechanisms related to TAR, as described above.

The origin and significance of subgenomic geminivirus genomes

There have been many reports of concatemeric and "subviral" or "subgenomic" forms of geminivirus genomic DNA, some of which date back some two decades to the earliest molecular studies of geminiviruses (Hamilton *et al.*, 1982; Stanley & Townsend, 1985; Stenger *et al.*, 1992). While concatemeric full-length genomes are generally assumed to result simply from imperfect resolution of unit-length genomes during RCR, a tidy understanding of the origins and roles of subgenomic forms has been more elusive. Such DNAs have frequently been encountered incidentally in the course of cloning, sequencing and/or characterising a wide variety of geminiviruses, but relatively few studies have set out to specifically investigate either the mechanisms by which geminivirus subgenomic DNAs are generated, or their significance in geminivirus biology. In addition to the numerous reports of sub-full-length and re-arranged genomes that have been isolated from infected plants, studies using 2-D gel electrophoresis have demonstrated that heterogenous populations of subgenomic virus DNA species are a ubiquitous feature of geminivirus replication (Saunders *et al.*, 1991; Jeske *et al.*, 2001; Preiss & Jeske, 2003; Jovel *et al.*, 2007).

Casado *et al.* (2004) purified half-geminate, isometric MSV particles and used PCR with LIR-specific abutting primers to amplify encapsidated subgenomic ssDNAs, which they cloned and sequenced. Although the processes of encapsidation, PCR, size-selection by gel electrophoresis, and cloning undoubtedly influenced the types of DNA species that they characterised, a striking proportion of them displayed breakpoints in, or adjacent to, the SIR. Moreover, a subset of their clones comprised the entire C-sense portion of the genome, with breakpoints corresponding to the V-sense proximal side of the putative C-sense replication primer binding site – apparently the same recombination and/or replication abortion hotspot that we describe here. While Casado *et al.* (2004) proposed encapsidation of nascent ssDNA in isometric (half-geminate) particles as the mechanism for this interrupted

replication, it would seem that the full-length recombinant genomes with breakpoints in this region that we report here argue against this.

In some cases, subgenomic DNAs have been shown to inhibit virus replication and/or attenuate symptoms, and have thus been legitimately identified as "defective interfering DNAs" (reviewed by Patil & Dasgupta, 2006). In much rarer cases, defined populations of sub-full-length forms have been shown to be replicated by one or more helper geminiviruses, and to be transmitted sustainably from one host plant to the next. These DNA species have therefore been termed "satellite viruses", and it has been documented that such satellite viruses may play significant roles in geminivirus evolution and can form important disease complexes (reviewed by Stanley, 2004).

In summary, the production of subgenomic DNAs is a standard, or at least frequent, feature of geminivirus replication, which suggests that they reflect an important mechanism of virus replication and/or a widespread host response to geminivirus infection. Furthermore, it is likely that the same processes lead both to the formation of subgenomic DNAs and to the generation of infectious full-length recombinant genomes. Subsets of these highly diverse populations of virus DNA, including both full-length and subgenomic molecules, are encapsidated and transmitted, permitting the frequent evolution of recombinant and satellite geminivirus genomes seen in nature. Here we have demonstrated the utility of a chimaera-based experimental system for investigating both mechanistic and evolutionary aspects of geminivirus recombination. We are currently extending and refining this approach through the construction of MSV chimaeras having multiple breakpoints evenly distributed throughout the genome.

CHAPTER 5: PERSPECTIVES AND CONCLUSIONS

Virology provides useful and unique perspectives on a range of fundamental biological fields of study including biochemistry, molecular biology, cell biology, genetics, taxonomy, and evolution. Because of their relative simplicity, viruses often allow supremely detailed and complete analyses of the genetic and biochemical processes that define them. Similarly, they may provide useful insights into more philosophical biological questions because they sometimes present opportunities to view such questions in a simpler context than might be possible when considering cellular life. Indeed, one might argue that viruses exist at the very boundary between the living and the non-living; by their nature they illustrate that useful definitions of life are elusive and that a distinction between living and non-living is less clear than many are wont to believe. Because they are obligate parasites viruses draw attention to the importance of their interactions with their hosts, and thereby accentuate the fact that no organism can be fully understood in isolation from its ecosystem. When considering evolution, viruses provide the clearest examples of a wide range of fundamental concepts such as mutation, recombination, positive selection, negative selection, genetic drift, and fitness landscapes.

It is therefore of general interest to develop viruses as biological tools that can be used to test existing hypotheses and to make observations that will inspire new theories, which in turn may be tested and refined. Apart from some specific contributions to a more complete understanding of geminiviruses in general and MSV in particular, much of the value of the work presented here lies in the successful development of unique experimental tools for studying mutation and recombination. While these tools will be most relevant to furthering our understanding of MSV and other geminiviruses, they could undoubtedly help to shed some light on more fundamental questions that are of broader interest.

As was mentioned in Chapter 1, the experiments described in this dissertation resulted not from the execution of a detailed plan, but arose out of opportunism and creativity over the course of time, and the strengths and weaknesses of the work presented here reflect that process. In hindsight it is clear that a more conservative approach and careful planning may have yielded more useful data, but on the other hand much of this experimental work was technically and conceptually novel, which made it difficult to anticipate the quantity or quality of the data that would be obtained. It should also be borne in mind that this work was carried out over a protracted period of time, with the result that some key tools that were either unavailable or prohibitively expensive at the outset are now relatively commonplace and affordable: the cost of DNA sequencing has plummeted, and high-fidelity rolling-circle amplification of geminivirus genomes from plant total DNA samples using Phi29 DNA polymerase has been thoroughly validated, and has recently revolutionised the cloning and sequencing of whole geminivirus genomes (Inoue-Nagata *et al.*, 2004; Haible *et al.*, 2006; Owor *et al.*, 2007b).

The construction of the initial MSV-Kom/-Set chimaeras described in Chapter 2 was a relatively risky first step that was taken more or less blindly: there were many reasons to believe that the chimaeric viruses were unlikely to be infectious, and it was not clear from the outset what specific insights infectious chimaeras might provide. Nevertheless, the gamble paid off with a spectacular demonstration of modularity in genome architecture (Martin *et al.*, 2005a), and provided useful biological corroboration of the relevance of specific MSV CP-MP interactions described subsequently by others (Liu *et al.*, 1997a; Liu *et al.*, 1999a; Liu *et al.*, 2001a). By demonstrating the feasibility of the approach, this work opened the way for colleagues to conduct a much larger study of MSV pathogenicity determinants in various maize cultivars employing a new set of chimaeric MSV constructs (Martin & Rybicki, 2002). Considering that MSV-Set represents a distinct MSV strain distantly related to the major maize pathogen strains, and that MSV-Set and MSV-Kom appear to be optimally adapted to different host species, it would be of interest to

extend the work from Chapter 2 to assess the pathogenicity of MSV-Kom/-Set chimaeras in a variety of hosts including maize, barley, wheat, and assorted wild grasses such as *Setaria* spp. with the aim of identifying determinants of host specificity.

The distinct symptomatologies displayed by the MSV chimaeras described in Chapter 2 raised the intriguing idea that they would be widely scattered across a theoretical "fitness landscape", within which the two wild-type parental viruses represent landmark "fitness peaks". In combination with some early indications that geminiviruses could be capable of extremely rapid evolution, it seemed possible that one might observe differences in the rates and/or trajectories of evolution of these chimaeras during a relatively short period of maintenance in host plants, and experiments exploring this idea were presented in Chapter 3. In addition to the chimaeric genomes passaged in maize, we made the serendipitous discovery of an MSV-infected sugarcane plant that had been maintained in isolation for a period of four years.

Unfortunately, the prohibitive cost of sequencing prevented us from performing many of the more detailed analyses that the realisation of our initial aims would have demanded: we were unable to perform site-by-site analysis of selection pressures, and could not distinguish between different chimaeras on the basis of the rates and/or trajectories of their evolution. Nevertheless, we were able to demonstrate the technical feasibility of the approach, and in the process we confirmed a very high basal rate of mutation for MSV – in the order of 7.4×10^{-4} subs/site/year for wild-type and chimaeric viruses in maize, and 1.6×10^{-3} subs/site/year for MSV in sugarcane.

Perhaps even more interesting was the observation of biased substitution rates, along with the implication that these substitution rate imbalances are not symmetrical with respect to the MSV virion- and complementary-sense DNA strands: while G→T

transversions in the virion-strand were massively over-represented, the frequency of complementary C→A transversions was found to be relatively low. This strongly implied that MSV incurs particular forms of strand-specific DNA damage as a result of the single-stranded nature of its genome during much of its life cycle in host plants.

It is now quite certain that geminiviruses, and probably many other small ssDNA viruses, exhibit very high mutation rates that are comparable to those of RNA viruses. This notion raises some intriguing paradoxes, and encourages us to re-assess the apparently reasonable and popular assumption that polymerase error rates are the primary determinant of mutation rates. While the evidence remains far from conclusive, there are tantalising indications from the data presented here and elsewhere that DNA damage may be a significant source of mutations in ssDNA viruses. Indeed, cellular organisms employ a vast cohort of DNA repair mechanisms to reverse the effects of DNA damage, and thereby maintain extremely low mutation rates. This raises the question: why do ssDNA viruses mutate so frequently when they replicate in the nuclei of such host cells, alongside host genetic material and using host replication machinery?

One likely explanation arises from the observation that geminivirus DNA does not appear to be methylated, which would preclude many of the DNA repair mechanisms employed by host cells. Another explanation is suggested by the biased, non-complementary mutation frequencies reported here and elsewhere: ssDNA is particularly prone to many forms of DNA damage, and geminivirus genomes spend much of their life-cycle in single-stranded form. Furthermore, it is possible that infected cells comprise a particularly damaging environment for DNA, either as a result of pathogenic effects, or perhaps due to host defence responses.

It is tempting to speculate about whether the high mutation rates of RNA viruses and small ssDNA viruses are essentially adaptive, or whether they are an unavoidable

burden. Since they are small and fecund, these viruses can "afford" high mutation rates; Drake (1991) has described a general, inverse correlation between genome size and mutation rate amongst a wide variety of DNA microbes. Certainly high mutation rates enable rapid adaptation, but they simultaneously carry a cost in terms of elevated numbers of deficient progeny. While the high mutation rates of RNA viruses may be interpreted as advantageous in the face of immune surveillance and other host defences, Drake (1998) disputed this interpretation and noted that two *E. coli*–infecting phages – the RNA phage Qβ and the DNA phage M13 – share similar life histories but display the expected differences in mutation rates according to their respective genetic materials.

Constituting the first reported attempts at fully controlled MSV evolution experiments, the work presented here has demonstrated that similar experiments are likely to yield extremely interesting insights into MSV evolution and molecular biology. For lack of data, small ssDNA viruses have been strikingly absent from studies comparing mutation rates across basic classes of organisms (Drake *et al.*, 1998; Drake, 1999), but there is now reason to expect that they might be included in similar surveys in the future. With the advent of affordable next-generation sequencing technologies permitting very deep genome sequencing of virus populations, experiments like these could be expected to provide detailed information about the occurrence and selection (positive or negative) of the thousands of possible mutations in the MSV genome. Because evolution can be thought of as an algorithm capable of computing solutions to the problem of fitness, this sequence information would not only illuminate the mechanisms of mutation, but could also reveal specific sites of positive or negative selection and could indicate epistatic interactions, all of which are relevant to deciphering molecular sequence/structure/function relationships.

While mutation is undoubtedly the ultimate source of genetic diversity, the experiments described in Chapter 4 focused on recombination – the powerful,

diversifying process of exchanging individual mutations (alleles) between genomes in a population. In the context of virus evolution and associated high mutation rates, recombination is often described as an important mechanism for escaping "Muller's ratchet" by purging deleterious alleles from the population, while facilitating the reassortment of beneficial alleles within distinct, individual progeny. Genetically, the process is akin to the various forms of mating seen in higher organisms. Therefore, although viruses might be considered by many to be "asexual" organisms, it is equally valid to describe many of them as promiscuously sexual.

The reciprocal chimaeric MSV constructs that formed the basis of the work presented in Chapters 2 and 3 are conceptually obvious tools for investigating recombination: while individual chimaeric viruses are relatively feeble in comparison to their parents, two reciprocal chimaeras together comprise the entire genetic sequences of both wild-type parents. Between them, they therefore carry the genetic potential to achieve wild-type fitness levels, through complementation and/or recombination. The results presented in Chapter 4 clearly demonstrated that, given the appropriate circumstances, recombination is a powerful adaptive mechanism of MSV. Furthermore, these experiments revealed likely recombination hot- and cold-spots which correspond closely to, and thereby validate, the results of computational analyses of recombination in mastreviruses. These recombination hot- and cold-spots can inform speculation about the mechanisms driving recombination in geminiviruses, and various possible scenarios were discussed in Chapter 4. Relatively recent advances can be expected to boost the power of similar experiments: it is now cost-effective to synthesise entire geminivirus genomes, making the construction of complicated reciprocal chimaeras trivial, while the use of Phi29 DNA polymerase for amplifying whole geminivirus genomes has greatly facilitated the process of cloning and sequencing of progeny viruses.

During the course of isolating, cloning and sequencing the full-length virus genomes presented in this dissertation, a large number of so-called subgenomic virus DNAs

were characterised. While subgenomic DNAs did not form a major focus of my work, they nevertheless fuelled much speculation and discussion between me and my colleagues over the years: such small, apparently "defective" geminivirus DNAs have been widely reported in the literature over the years and seem to be ubiquitously associated with geminivirus infections, but they remain poorly understood. Many of them are likely to originate from concatenated geminivirus genomic fragments making up the so-called "heterogenous DNA" species described by Jeske *et al.* (Jeske *et al.*, 2001; Preiss & Jeske, 2003), which is thought to be related to the processes of recombination-dependent replication (RDR) proposed by the same authors. Some of these subgenomic species are, however, apparently capable of autonomous or trans-replication, and may be encapsidated in and transmitted via half-geminate (icosahedral) virions. Indeed, the DNA-β satellites seem to represent such trans-replicated subgenomic DNAs which have survived and evolved over an extended period of time.

One can hardly help but wonder about the role played by subgenomic DNA species in relation to the parental geminiviruses from which they arise: do they represent mere debris resulting from some malfunction of virus replication or from host defense responses; do they constitute important and functionally efficient replication intermediates; do they play a role in regulating virus gene expression through unequal copy numbers of complementary- and virion-sense genes; are they harmless symbionts hitching a ride with their parental viruses; or are they active and detrimental parasites, so-called "defective interfering" molecules? While definitive answers to these questions are unlikely to be provided in the near future, concerted efforts to study geminivirus subgenomic DNAs might yield valuable insights into the molecular biology and ecology of these viruses.

When one considers these issues, it becomes apparent that geminiviruses may potentially serve as useful models for testing concepts that form the basis of much of ecology. It seems clear that viruses, satellite viruses, subgenomic DNAs, defective

interfering particles, and distinct, co-infecting virus strains may all represent examples of either co-operative, competitive, or parasitic relationships. This idea was recently eloquently posed by Elena and Sanjuàn (2007):

> *"Usually when talking about sociality, we tend to imagine the intricate relationships established between higher organisms (from social insects to the human society), but we hardly imagine viruses to be good candidates for testing theoretical predictions about cooperation, altruism, selfishness, or cheating. This appreciation is not correct because social interactions among viruses are quite common in nature, as, for example, synergistic symptoms among co-infecting plant virus (Malpica et al., 2006), predator-prey-like coevolutionary dynamics between some virus and their defective interfering particles (DePolo et al., 1987; Bangham & Kirkwood, 1993), or the existence of coviruses, that is, genome segments encapsidated into separate particles, which are required for a successful infection (Nee, 2000)."*

One final concept to ponder in respect to the work presented here is that of modularity in genetic architecture. As described in Chapter 2, the viability of the chimaeric viruses comprising sequences from such diverse parents was initially surprising in the light of the drastically detrimental effects of minor mutations in MSV clones reported elsewhere. However, if one appreciates the implicit modular nature of genetic sequences, then the paradox is resolved: exchanging a relatively self-contained functional genetic unit – represented in this case by the MSV V1-V2 ORF region – can be comparatively less disruptive than imposing a single, unfortunate, random mutation. In the light of modular genome architecture, recombination makes all the more sense: exploring novel combinations of modules that have already evolved to a state of functionality entails the exploration of spectacularly smaller (albeit still unimaginably vast) volumes of sequence space than does the alternative process of random mutation.

Over sufficiently long time periods, it can be said that the only certainty is change – the ecological niche of any given organism is bound to undergo a series of dramatic transformations. Any genetic material that is to survive such upheavals and be represented in organisms thereafter must therefore be part of a sufficiently changeable/evolvable system (i.e. population of organisms) – one that can adapt sufficiently rapidly to the changing environment. For example, the evolution of "mutator" races under conditions of environmental change is well documented. Similarly, it is safe to assume that the selective advantages of recombination must reflect its ubiquity amongst extant organisms. It therefore seems reasonable that any genetic feature which would enhance the efficacy of recombination for adaptation would be similarly selected during evolution, and modular genetic architecture would be an obvious example of such a feature. It therefore follows that recombination may be largely responsible for the selection of modularity in genomes.

Some organisms – like multicellular eukaryotes – apparently rely heavily on recombination for generating the requisite genetic diversity to enable adaptation within relatively short periods of time, with the contribution from mutation being comparatively small. Others, such as some RNA viruses, appear to derive a great deal of their useful genetic diversity via mutation, with recombination playing a relatively small role over the "short term". In the context of our theoretical understanding of evolution, it is clear that mutation and recombination are extremely complementary diversifying processes, and both appear to contribute substantially to the evolution of geminiviruses.

REFERENCES

Abouzid, A. M., Frischmuth, T., Jeske, H., 1988. A putative replicative form of abutilon mosaic virus (gemini group) in a chromatin-like structure. Molecular and General Genetics 212, 252-258.

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

Anderson, P. K., Morales, F. J., 1994. The emergence of new plant diseases.: The case of insect-transmitted plant viruses. Annals of the New York Academy of Sciences 740:181-194, 181-194.

Aniskina, I., Gornostaeva, S. I., Andreeva, L. E., Nenasheva, V. V., Tarantul, V. Z., Nikolaev, A. I., 2003. Viral promoter-enhancer element inhibits replication of autonomously replicating plasmid in mammalian cells. Molecular Biology (Moscowk) 37, 876-884.

Ao, P., 2005. Laws in Darwinian evolutionary theory. Physics of Life Reviews 2, 117-156.

Arguello-Astorga, G., Ascencio-Ibanez, J. T., Dallas, M. B., Orozco, B. M., Hanley-Bowdoin, L., 2007. High-frequency reversion of geminivirus replication protein mutants during infection. Journal of Virology 81, 11005-11015.

Arguello-Astorga, G. R., Ruiz-Medrano, R., 2001. An iteron-related domain is associated to Motif 1 in the replication proteins of geminiviruses: identification of potential interacting amino acid-base pairs by a comparative approach. Archives of Virology 146, 1465-1485.

Ascencio-Ibanez, J. T., Settlage, S. B., 2007. DNA abrasion onto plants is an effective method for geminivirus infection and virus-induced gene silencing. Journal of Virological Methods 142, 198-203.

Bangham, C. R., Kirkwood, T. B., 1993. Defective interfering particles and virus evolution. Trends in Microbiology 1, 260-264.

Barton, N. H., Charlesworth, B., 1998. Why sex and recombination? Science 281, 1986-1990.

Behjatnia, S. A. A., Dry, I. B., Rezaian, M. A., 2007. Characterization and transient replication of tomato leaf curl virus defective DNAs. Archives of Virology 152, 1127-1138.

Bejarano, E. R., Khashoggi, A., Witty, M., Lichtenstein, C., 1996. Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proceedings of the National Academy of Sciences U.S.A. 93, 759-764.

Bernard, H. U., 1994. Coevolution of papillomaviruses with human populations. Trends in Microbiology 2, 140-143.

Berrie, L. C., Rybicki, E. P., Rey, M. E., 2001. Complete nucleotide sequence and host range of South African cassava mosaic virus: further evidence for recombination amongst begomoviruses. Journal of General Virology 82, 53-58.

Biagini, P., 2004. Human circoviruses. Veterinary Microbiology 98, 95-101.

Bock, K. R., 1974. Maize streak virus. CMI/AAB Descriptions of Plant Viruses 133.

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

Bokanga, M., Lenne, J., Otim-Nape, W., Spence, N. D3. 1: Future control of infectious diseases in plants with emphasis on sub-Saharan Africa. 2006.
Ref Type: Report

Bosque-Perez, N. A., 2000. Eight decades of maize streak Virus Research. Virus Research 71, 107-121.

Boulton, M. I., 2002. Functions and interactions of mastrevirus gene products. Physiological and Molecular Plant Pathology 60, 243-255.

Boulton, M. I., King, D. I., Donson, J., 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., 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., Davies, J. W., 1993. Replication of maize streak virus mutants in maize protoplasts - evidence for a movement protein. Virology 192, 85-93.

Boulton, M. I., Steinkellner, H., Donson, J., Markham, P. G., King, D. I., Davies, J. W., 1989. Mutational analysis of the virion-sense genes of maize streak virus. Journal of General Virology 70, 2309-2323.

Briddon, R., Brown, J., Moriones, E., Stanley, J., Zerbini, M., Zhou, X., Fauquet, C., 2008. Recommendations for the classification and nomenclature of the DNA-beta satellites of begomoviruses. Archives of Virology 153, 763-781.

Briddon, R. W., Bedford, I. D., Tsai, J. H., Markham, P. G., 1996. 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., Bull, S. E., Amin, I., Idris, A. M., Mansoor, S., Bedford, I. D., Dhawan, P., Rishi, N., Siwatch, S. S., Abdel-Salam, A. M., Brown, J. K., Zafar, Y., Markham, P. G., 2003. Diversity of DNA beta, a satellite molecule associated with some monopartite begomoviruses. Virology 312, 106-121.

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

Briddon, R. W., Mansoor, S., Bedford, I. D., Pinner, M. S., Saunders, K., Stanley, J., Zafar, Y., Malik, K. A., Markham, P. G., 2001. Identification of dna components required for induction of cotton leaf curl disease. Virology 285, 234-243.

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

Briddon, R. W., Stanley, J., 2006. Subviral agents associated with plant single-stranded DNA viruses. Virology 344, 198-210.

Briddon, R. W., Bull, S. E., Amin, I., Mansoor, S., Bedford, I. D., Rishi, N., Siwatch, S. S., Zafar, Y., bdel-Salam, A. M., Markham, P. G., 2004. Diversity of DNA 1: a satellite-like molecule associated with monopartite begomovirus-DNA beta complexes. Virology 324, 462-474.

Brough, C. L., Gardiner, W. E., Inamdar, N. M., Zhang, X. Y., Ehrlich, M., Bisaro, D. M., 1992. DNA methylation inhibits propagation of tomato golden mosaic virus DNA in transfected protoplasts. Plant Molecular Biology 18, 703-712.

Brough, C. L., Hayes, R. J., Morgan, A. J., Coutts, R. H. A., Buck, K. W., 1988. Eeffects of mutagenesis in vitro on the ability of cloned tomato golden mosaic virus dna DNA to infect *Nnicotiana benthamiana* plants. Journal of General Virology 69, 503.

Bull, J. J., Badgett, M. R., Wichman, H. A., Huelsenbeck, J. P., Hillis, D. M., Gulati, A., Ho, C., Molineux, I. J., 1997. Eexceptional convergent evolution in a virus. Genetics 147, 1497-1507.

Bull, S. E., Tsai, W. S., Briddon, R. W., Markham, P. G., Stanley, J., Green, S. K., 2004. Diversity of begomovirus DNA beta satellites of non-malvaceous plants in east and south east Asia. Archives of Virology 149, 1193-1200.

Casado, C. G., Javier, O. G., Padron, E., Bean, S. J., McKenna, R., Agbandje-McKenna, M., Boulton, M. I., 2004. Isolation and characterization of subgenomic DNAs encapsidated in "single" T = 1 isometric particles of mMaize streak virus. Virology 323, 164-171.

Castellano, M. M., Sanz-Burgos, A. P., Gutierrez, C., 1999. Initiation of DNA replication in a eukaryotic rolling-circle replicon: identification of multiple DNA-protein complexes at the geminivirus origin. Journal of Molecular Biology 290, 639-652.

Caulfield, J. L., Wishnok, J. S., Tannenbaum, S. R., 1998. Nitric oxide-induced deamination of cytosine and guanine in deoxynucleosides and oligonucleotides. J Biol Chem. Journal of Biological Chemistry 273, 12689-12695.

Chao, L., 1990. Fitness of RNA virus decreased by Muller's ratchet. Nature 348, 454-455.

Chou, P. Y., Fasman, G. D., 1974. Prediction of protein conformation. Biochemistry 13, 222-245.

Christou, P., Twyman, R. M., 2004. The potential of genetically enhanced plants to address food insecurity. Nutrition Research Reviews 17, 23-42.

Collin, S., Fernandez-Lobato, M., Gooding, P. S., Mullineaux, P. M., 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.

Cooper, V. S., Lenski, R. E., 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. Nature 407, 689-690.

Corning, P. A., 2002. The re-emergence of "emergence": A venerable concept in search of a theory. Complexity 7, 18-30.

Covey, S. N., Al-Kaff, N. S., 2000. Plant DNA viruses and gene silencing. Plant Molecular Biology 43, 307-322.

Crameri, A., Raillard, S. A., Bermudez, E., Stemmer, W. P., 1998. DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature 391, 288-291.

Cromie, G. A., Connelly, J. C., Leach, D. R. F., 2001. Recombination at ~~Double~~double-~~Strand~~ strand ~~Breaks~~ breaks and DNA ~~Ends~~ends: ~~Conserved~~ conserved ~~Mechanisms~~ mechanisms from ~~Phage~~ phage to ~~Humans~~humans. Molecular Cell 8, 1163-1174.

Cserzo, M., Wallin, E., Simon, I., von, H. G., Elofsson, A., 1997. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. Protein Engineering 10, 673-676.

Cui, X., Li, G., Wang, D., Hu, D., Zhou, X., 2005. A ~~Begomovirus~~ begomovirus DNAbeta-encoded protein binds DNA, functions as a suppressor of RNA silencing, and targets the cell nucleus. Journal of Virology 79, 10764-10775.

Czosnek, H., Ber, R., Navot, N., Antignus, Y., Cohen, S., Zamir, D., 1989. ~~t~~Tomato yellow leaf curl virus ~~dna~~ DNA forms in the viral capsid~~e~~, in infected plants and in the insect vector. Journal of Phytopathology 124, 47-54.

Damsteegt, V. D., 1983. Maize streak virus: host range and vulnerability of maize germ plasm. Plant Disease 67, 734-7.

Davies, J. W., Stanley, J., 1989. Geminivirus genes and vectors. Trends in Genetics~~.~~ 5, 77-81.

Dayhoff, M. O., Schwartz, R. M., Orcutt, B. C., 1978. A model for evolutionary change in proteins. In: M. O. Dayhoff (Ed.), *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, pp. 345-352.

de Visser, J. A., Elena, S. F., 2007. The evolution of sex: empirical insights into the roles of epistasis and drift. ~~Nat.Rev.Genet.~~ Nature Reviews Genetics 8, 139-149.

de Visser, J. A., Lenski, R. E., 2002. Long-term experimental evolution in *Escherichia coli*. XI. Rejection of non-transitive interactions as cause of declining rate of adaptation. BMC Evolutionary ~~.~~Biology 2, 19.

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

Delatte, H., Martin, D. P., Naze, F., Goldbach, R., Reynaud, B., Peterschmitt, M., Lett, J. M., 2005. South West Indian Ocean islands tomato begomovirus populations represent a new major monopartite begomovirus group. Journal of General Virology 86, 1533-1542.

DePolo, N. J., Giachetti, C., Holland, J. J., 1987. Continuing coevolution of virus and defective interfering particles and of viral genome sequences during undiluted passages: virus mutants exhibiting nearly complete resistance to formerly dominant defective interfering particles. ~~J Virol~~Journal of Virology 61, 454-464.

Desbiez, C., David, C., Mettouchi, A., Laufs, J., Gronenborn, B., 1995. Rep protein of tomato yellow leaf curl geminivirus has an ATPase activity required for viral DNA replication. ~~Proc.Natl.Acad.Sci.~~ Proceedings of the National Academy of Sciences U.S.A. 92, 5640-5644.

Deshpande, A. M., Newlon, C. S., 1996. DNA replication fork pause sites dependent on transcription. Science 272, 1030-1033.

Dickinson, V. J., Halder, J., 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.

Domingo, E., Holland, J. J., 1997. RNA virus mutations and fitness for survival. ~~Annu.Rev.Microbiol.~~Annual Review of Microbiology 51, 151-178.

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

Drake, J. W., 1991. A constant rate of spontaneous mutation in DNA-based microbes. ~~Proc.Natl.Acad.Sci.U.S.A.~~ Proceedings of the National Academy of Sciences U.S.A. 88, 7160-7164.

Drake, J. W., 1999. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. ~~Ann.N.Y.Acad.Sci.~~ Annals of the New York Academy of Sciences 870, 100-107.

Drake, J. W., Charlesworth, B., Charlesworth, D., Crow, J. F., 1998. Rates of spontaneous mutation. Genetics 148, 1667-1686.

Drummond, D. A., 2005. On the conservative nature of intragenic recombination. Proceedings of the National Academy of Sciences U.S.A. 102, 5380-5385.

Duarte, E., Clarke, D., Moya, A., Domingo, E., Holland, J., 1992. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. ~~Proc.Natl.Acad.Sci.U.S.A.~~ Proceedings of the National Academy of Sciences U.S.A. 89, 6015-6019.

Duffy, S., Holmes, E. C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. Journal of Virology 82, 957-965.

Durrant, W. E., Wang, S., Dong, X., 2007. *Arabidopsis* SNI1 and RAD51D regulate both gene transcription and DNA recombination during the defense response. ~~Proc Natl Acad Sci U.S.A~~Proceedings of the National Academy of Sciences U.S.A. 104, 4223-4227.

Elena, S. F., Davila, M., Novella, I. S., Holland, J. J., Domingo, E., Moya, A., 1998. ~~evolutionary~~ Evolutionary dynamics of fitness recovery from the debilitating effects of ~~muller's~~ Muller's ratchet. Evolution 52, 309-314.

Elena, S. F., Lenski, R. E., 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. ~~Nat.Rev.Genet.~~ Nature Reviews Genetics 4, 457-469.

Elena, S. F., Sanjuan, R., 2007. Virus ~~Evolution~~evolution: ~~Insights~~ insights from an ~~Experimental~~ experimental ~~Approach~~approach. Annual Review of Ecology, Evolution, and Systematics 38, 27-52.

Elias-Arnanz, M., Salas, M., 1997. Bacteriophage phi29 DNA replication arrest caused by codirectional collisions with the transcription machinery. ~~EMBO J~~ EMBO Journal 16, 5775-5783.

Elias-Arnanz, M., Salas, M., 1999. Resolution of head-on collisions between the transcription machinery and bacteriophage phi29 DNA polymerase is dependent on RNA polymerase translocation. ~~EMBO J~~ EMBO Journal 18, 5675-5682.

Elmer, J. S., Brand, L., Sunter, G., Gardiner, W. E., Bisaro, D. M., 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.~~ Nucleic Acids Research 16, 7043-7060.

Etessami, P., Callis, R., Ellwood, S., Stanley, J., 1988. Delimitation of essential genes of cassava latent virus DNA 2. ~~Nucleic.Acids.Res.~~ Nucleic Acids Research 16, 4811-4829.

Fauquet, C. M., Bisaro, D. M., Briddon, R. W., Brown, J. K., Harrison, B. D., Rybicki, E. P., Stenger, D. C., Stanley, J., 2003. Revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. ~~Arch Virol~~Archives of Virology 148, 405-421.

Ferreira, P. T., Lemos, T. O., Nagata, T., Inoue-Nagata, A. K., 2008. One-step cloning approach for construction of agroinfectious begomovirus clones. ~~J Virol~~~~Journal of Virology~~ ~~Methods~~Journal of Virological Methods 147, 351-354.

Fondong, V. N., Pita, J. S., Rey, M. E., de Kochko, A., Beachy, R. N., Fauquet, C. M., 2000. Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. ~~J Gen Virol~~Journal of General Virology 81 Pt 1:287-97, 287-297.

Fontes, E. P., Santos, A. A., Luz, D. F., Waclawovsky, A. J., Chory, J., 2004. The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. ~~Genes Dev.~~Genes and Development 18, 2545-2556.

Francki, R. I. B., Hatta, T., Boccardo, G., Davies, J. W., 1980. The composition of ~~*chloris*~~ *Chloris* striate mosaic virus, a geminivirus. Virology 101, 233-241.

Frederico, L. A., Kunkel, T. A., Shaw, B. R., 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. Biochemistry 29, 2532-2537.

French, S., 1992. Consequences of replication fork movement through transcription units in vivo. Science 258, 1362-1365.

Frischmuth, T., Roberts, S., von Arnim, A., Stanley, J., 1993. Specificity of bipartite geminivirus movement proteins. Virology 196, 666-673.

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

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

Frischmuth, T., Stanley, J., 1998. Recombination between viral DNA and the transgenic coat protein gene of African cassava mosaic geminivirus. Journal of General Virology 79, 1265-1271.

Fuller, C., 1901. Mealie variegation. In: 1st Report of the Government Entomologist, Natal, 1899-1900.

Gafni, Y., Epel, B. L., 2002. The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. Physiological and Molecular Plant Pathology 60, 231-241.

Gallian, P., Biagini, P., Attoui, H., Cantaloube, J. F., Dussol, B., Berland, Y., de, M. P., de, L., X, 2002. High genetic diversity revealed by the study of TLMV infection in French hemodialysis patients. ~~J.Med.Virol.~~Journal of Medical Virology 67, 630-635.

Galvao, R. M., Mariano, A. C., Luz, D. F., Alfenas, P. F., Andrade, E. C., Zerbini, F. M., Almeida, M. R., Fontes, E. P., 2003. A naturally occurring recombinant DNA-A of a typical bipartite begomovirus does not require the cognate DNA-B to infect *Nicotiana benthamiana* systemically. ~~J Gen Virol~~Journal of General Virology 84, 715-726.

Garcia-Andres, S., Accotto, G. P., Navas-Castillo, J., Moriones, E., 2007a. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl disease in the Mediterranean basin. Virology 359, 302-312.

Garcia-Andres, S., Monci, F., Navas-Castillo, J., Moriones, E., 2006. Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: Evidence for the presence of a new virus species of recombinant nature. Virology 350, 433-442.

Garcia-Andres, S., Tomas, D. M., Sanchez-Campos, S., Navas-Castillo, J., Moriones, E., 2007b. Frequent occurrence of recombinants in mixed infections of tomato yellow leaf curl disease-associated begomoviruses. Virology 365, 210-219.

Garcia-Arenal, F., Fraile, A., Malpica, J. M., 2001. Variability and genetic structure of plant virus populations. ~~Annu.Rev.Phytopathol.~~Annual Review of Phytopathology 39, 157-186.

Garcia-Arenal, F., Fraile, A., Malpica, J. M., 2003. Variation and evolution of plant virus populations. ~~Int.Microbiol.~~International Microbiology 6, 225-232.

Garrido-Ramirez, E. R., Sudarshana, M. R., Lucas, W. J., Gilbertson, R. L., 2000. Bean dwarf mosaic virus BV1 protein is a determinant of the hypersensitive response and avirulence in *Phaseolus vulgaris*. Molecular Plant Microbe Interactions 13, 1184-1194.

Ge, L., Zhang, J., Zhou, X., Li, H., 2007. Genetic structure and population variability of tomato yellow leaf curl China virus. Journal of Virology 81, 5902-5907.

Gerrish, P., Lenski, R., 1998. The fate of competing beneficial mutations in an asexual population. Genetica 102-103, 127-144.

Ghanim, M., Czosnek, H., 2000. ~~tomato~~ Tomato yellow leaf curl geminivirus (TYLCV-is) is transmitted among whiteflies (~~*bemisia*~~ *Bemisia tabaci*) in a sex-related manner. Journal of Virology 74, 4738-4745.

Ghanim, M., Morin, S., Zeidan, M., 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, A. J., Keese, P. L., Gibbs, M. J., Garcia-Arenal, F., 1999. Plant virus evolution: past, present and future. In: E. Domingo, R. Webster, J. Holland (Eds.), Origin and Evolution of Viruses. Academic Press, New York, pp. 263-285.

Gibbs, M. J., Weiller, G. F., 1999. Evidence that a plant virus switched hosts to infect a vertebrate and then recombined with a vertebrate-infecting virus. Proceedings of the National Academy of Sciences 96, 8022.

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

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

Gilbertson, R. L., Sudarshana, M., Jiang, H., Rojas, M. R., Lucas, W. J., 2003. Limitations on geminivirus genome size imposed by plasmodesmata and virus-encoded movement protein: insights into DNA trafficking. Plant Cell 15, 2578-2591.

Gorbalenya, A. E., Koonin, E. V., 1989. Viral proteins containing the purine NTP-binding sequence pattern. Nucleic Acids Research 17, 8413-8440.

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

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

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

Gussow, D., Clackson, T., 1989. Direct clone characterization from plaques and colonies by the polymerase chain reaction. ~~Nucleic Acids Res.~~Nucleic Acids Research 17, 4000.

Gutierrez, C., 1999. Geminivirus DNA replication. ~~Cell Mol.Life Sci.~~Cellular and Molecular Life Sciences 56, 313-329.

Gutierrez, C., Ramirez-Parra, E., Mar, C. M., Sanz-Burgos, A. P., Luque, A., Missich, R., 2004. Geminivirus DNA replication and cell cycle interactions. ~~Vet.Microbiol.~~Veterinary Microbiology 98, 111-119.

Gutierrez, C., 2002. Strategies for geminivirus DNA replication and cell cycle interference. Physiological & Molecular Plant Pathology 60, 219.

Haase, S. B., Heinzel, S. S., Calos, M. P., 1994. Transcription inhibits the replication of autonomously replicating plasmids in human cells. ~~Mol Cell Biol~~Molecular and Cellular Biology 14, 2516-2524.

Haible, D., Kober, S., Jeske, H., 2006. Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. Journal of Virological Methods 135, 9-16.

Halley-Stott, R. P., Tanzer, F., Martin, D. P., Rybicki, E. P., 2007. The complete nucleotide sequence of a mild strain of b~~B~~ean yellow dwarf virus. Archives of Virology 152, 1237-1240.

Hamilton, W. D., Bisaro, D. M., 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.~~ Nucleic Acids Research 10, 4901-4912.

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

Hanley-Bowdoin, L., Settlage, S. B., Robertson, D., 2004. Reprogramming plant gene expression: a prerequisite to geminivirus DNA replication. Molecular Plant Pathology 5, 149-156.

Hanley-Bowdoin, L., Settlage, S. B., Orozco, B. M., Nagar, S., Robertson, D., 1999. Geminiviruses: Models for ~~Plant~~ plant DNA ~~Replication~~replication, ~~Transcription~~transcription, and ~~Cell~~ cell ~~Cycle~~ cycle ~~Regulation~~regulation. Critical Reviews in Plant Sciences 18, 71-106.

Hanson, S. F., Hoogstraten, R. A., Ahlquist, P., Gilbertson, R. L., Russell, D. R., 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.

Heath, L., Martin, D. P., Warburton, L., Perrin, M., Horsfield, W., Kingsley, C., Rybicki, E. P., Williamson, A. L., 2004. Evidence of unique genotypes of beak and feather disease virus in southern Africa. Journal of Virology 78, 9277-9284.

Hehnle, S., Wege, C., Jeske, H., 2004. Interaction of DNA with the movement proteins of geminiviruses revisited. Journal of Virology 78, 7698-7706.

Hey, J., 1998. Selfish genes, pleiotropy and the origin of recombination. Genetics 149, 2089-2097.

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

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

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

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

Hofgen, R., Willmitzer, L., 1988. Storage of competent cells for *Agrobacterium* transformation. ~~Nucleic Acids Res.~~Nucleic Acids Research 16, 9877.

Holland, J., Domingo, E., 1998. Origin and evolution of viruses. Virus Genes 16, 13-21.

Holmes, E. C., 2004. The phylogeography of human viruses. ~~Mol.Ecol.~~Molecular Ecology 13, 745-756.

Hong, Y., Stanley, J., van, W. R., 2003. Novel system for the simultaneous analysis of geminivirus DNA replication and plant interactions in *Nicotiana benthamiana*. ~~J Virol~~Journal of Virology 77, 13315-13322.

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

Hormuzdi, S. G., 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.

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

Hou, Y. M., Sanders, R., Ursin, V. M., Gilbertson, R. L., 2000. Transgenic plants expressing geminivirus movement proteins: abnormal phenotypes and delayed infection by Tomato mottle virus in transgenic tomatoes expressing the Bean dwarf mosaic virus BV1 or BC1 proteins [In Process Citation]. ~~Mol Plant Microbe.Interact.~~Molecular Plant Microbe Interactions 13, 297-308.

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

Huang, F. T., Yu, K., Balter, B. B., Selsing, E., Oruc, Z., Khamlichi, A. A., Hsieh, C. L., Lieber, M. R., 2007. Sequence dependence of chromosomal R-loops at the immunoglobulin heavy-chain Smu class switch region. ~~Mol Cell Biol~~Molecular and Cellular Biology 27, 5921-5932.

Huang, F. T., Yu, K., Hsieh, C. L., Lieber, M. R., 2006. Downstream boundary of chromosomal R-loops at murine switch regions: implications for the mechanism of class switch recombination. ~~Proc Natl Acad Sci U.S.A~~Proceedings of the National Academy of Sciences U.S.A. 103, 5030-5035.

Huertas, P., Aguilera, A., 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. ~~Mol Cell~~Molecular Cell 12, 711-721.

ICTVdB Management. 00.029.0.02. Curtovirus. Büchen-Osmond, C. ICTVdB - The Universal Virus Database, version 4 . 2006a. Columbia University.
Ref Type: Electronic Citation

ICTVdB Management. 00.029.0.02.001. Beet curly top virus. Büchen-Osmond, C. ICTVdB - The Universal Virus Database, version 4 . 2006b. Columbia University.
Ref Type: Electronic Citation

ICTVdB Management. 00.029.0.03. Begomovirus. Büchen-Osmond, C. ICTVdB - The Universal Virus Database, version 4 . 2006c. Columbia University.
Ref Type: Electronic Citation

ICTVdB Management. 00.029.0.04. Topocuvirus. Büchen-Osmond, C. ICTVdB - The Universal Virus Database, version 4 . 2006d. Columbia University.
Ref Type: Electronic Citation

Inamdar, N. M., Zhang, X. Y., Brough, C. L., Gardiner, W. E., Bisaro, D. M., Ehrlich, M., 1992. Transfection of heteroduplexes containing uracil.guanine or thymine.guanine mispairs into plant cells. ~~Plant Mol Biol~~Plant Molecular Biology 20, 123-131.

Inoue-Nagata, A. K., Albuquerque, L. C., Rocha, W. B., Nagata, T., 2004. A simple method for cloning the complete begomovirus genome using the bacteriophage phi29 DNA polymerase. ~~J Virol~~Journal of Virology ~~Methods~~Journal of Virological Methods 116, 209-211.

Isnard, M., Granier, M., Frutos, R., Reynaud, B., Peterschmitt, M., 1998. Quasispecies nature of three maize streak virus isolates obtained through different modes of selection from a population used to assess response to infection of maize cultivars. Journal of General Virology 79, 3091-3099.

Jenkins, G. M., Rambaut, A., Pybus, O. G., Holmes, E. C., 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. ~~J.Mol.Evol.~~ Journal of Molecular Evolution 54, 156-165.

Jeske, H., Lutgemeier, M., Preiss, W., 2001. DNA forms indicate rolling circle and recombination-dependent replication of *Abutilon* mosaic virus. ~~The EMBO Journal~~EMBO Journal 20, 6158-6167.

Jose, J., Usha, R., 2003. Bhendi Yellow Vein Mosaic Disease in India Is Caused by Association of a DNA-beta Satellite with a Begomovirus. Virology 305, 310-317.

Jovel, J., Preiss, W., Jeske, H., 2007. Characterization of DNA intermediates of an arising geminivirus. ~~Virus Res~~Virus Research 130, 63-70.

Jovel, J., Reski, G., Rothenstein, D., Ringel, M., Frischmuth, T., Jeske, H., 2004. *Sida micrantha* mosaic is associated with a complex infection of begomoviruses different from *Abutilon* mosaic virus. ~~Arch.Virol~~Archives of Virology 149, 829-841.

Kearney, C. M., Thomson, M. J., Roland, K. E., 1999. Genome evolution of tobacco mosaic virus populations during long-term passaging in a diverse range of hosts. ~~Arch.Virol~~Archives of Virology 144, 1513-1526.

Kenton, A., Khashoggi, A., Parokonny, A., Bennett, M. D., Lichtenstein, C., 1995. Chromosomal location of endogenous geminivirus-related DNA sequences in *Nicotiana tabacum* L. ~~Chromosome.Res.~~Chromosome Research 3, 346-350.

Kiraly, L., Barna, B., Kiraly, Z., 2007. Plant resistance to pathogen infection: forms and mechanisms of innate and acquired resistance. Journal of Phytopathology 155, 385-396.

Kitamura, K., Murayama, A., Ikegami, M., 2004. Evidence for recombination among isolates of Tobacco leaf curl Japan virus and Honeysuckle yellow vein mosaic virus. ~~Arch.Virol~~Archives of Virology 149, 1221-1229.

Koncz, C., 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. Molecular and General Genetics 204, 383-396.

Koonin, E. V., 2003. Horizontal gene transfer: the path to maturity. ~~Mol.Microbiol.~~Molecular Microbiology 50, 725-727.

Koonin, E. V., Ilyina, T. V., 1992. Geminivirus replication proteins are related to prokaryotic plasmid rolling circle DNA replication initiator proteins. Journal of General Virology 73, 2763-2766.

Koonin, E. V., Senkevich, T. G., Dolja, V. V., 2006. The ancient ~~Virus~~ virus ~~World~~ world and evolution of cells. Biology Direct. 1, 29.

Kosakovsky Pond, S. L., Frost, S. D. W., 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531-2533.

Kosakovsky Pond, S. L., Frost, S. D. W., 2005b. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution 22, 1208-1222.

Kotlizky, G., Boulton, M. I., Pitaksutheepong, C., Davies, J. W., 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.

Kovalchuk, I., Kovalchuk, O., Kalck, V., Boyko, V., Filkowski, J., Heinlein, M., Hohn, B., 2003. Pathogen-induced systemic plant signal triggers DNA rearrangements. Nature 423, 760-762.

Krasilnikova, M. M., Samadashwily, G. M., Krasilnikov, A. S., Mirkin, S. M., 1998. Transcription through a simple DNA repeat blocks replication elongation. ~~EMBO J~~ EMBO Journal 17, 5095-5102.

Kunik, T., Salomon, R., Zamir, D., Navot, N., Zeidan, M., Michelson, I., Gafni, Y., Czosnek, H., 1994. Transgenic tomato plants expressing the tomato yellow leaf curl virus capsid protein are resistant to the virus. Biotechnology (N.Y.) 12, 500-504.

Lapidot, M., Weil, G., Cohen, L., Segev, L., Gaba, V., 2007. Biolistic inoculation of plants with tomato yellow leaf curl virus DNA. ~~J Virol~~~~Journal of Virology Methods~~Journal of Virological Methods 144, 143-148.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., Gronenborn, B., 1995a. Identification of the nicking tyrosine of geminivirus Rep protein. ~~FEBS Lett.~~ FEBS Letters 377, 258-262.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J., Gronenborn, B., 1995b. *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.~~ Proceedings of the National Academy of Sciences U.S.A. 92, 3879-3883.

Lawrence, J. G., Hendrickson, H., 2003. Lateral gene transfer: when will adolescence end? ~~Mol.Microbiol.~~Molecular Microbiology 50, 739-749.

Lazarowitz, S. G., 1987. The Molecular Characterization of Geminiviruses. Plant Molecular Biology Reporter 4, 177-192.

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

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

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

Lefeuvre, P., Martin, D. P., Hoareau, M., Naze, F., Delatte, H., Thierry, M., Varsani, A., Becker, N., Reynaud, B., Lett, J. M., 2007. Begomovirus 'melting pot' in the south-west Indian Ocean islands: molecular diversity and evolution through recombination. Journal of General Virology 88, 3458-3468.

Lehman, N., 2003. A case for the extreme antiquity of recombination. ~~J Mol Evol~~Journal of Molecular Evolution 56, 770-777.

Lenski, R. E., Rose, M. R., Simpson, S. C., Tadler, S. C., 1991. Long-term experimental evolution in ~~escherichia~~ *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. The American Naturalist 138, 1315-1341.

Lenski, R. E., Travisano, M., 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proceedings of the National Academy of Sciences ~~of the United States of America~~U.S.A. 91, 6808-6814.

Liu, B., Alberts, B. M., 1995. Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. Science 267, 1131-1137.

Liu, H., Boulton, M. I., Davies, J. W., 1997a. Maize streak virus coat protein binds single- and double-stranded DNA in vitro. Journal of General Virology 78, 1265-1270.

Liu, H., Boulton, M. I., Oparka, K. J., Davies, J. W., 2001a. Interaction of the movement and coat proteins of Maize streak virus: implications for the transport of viral DNA. Journal of General Virology 82, 35-44.

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

Liu, H., Lucy, A. P., Davies, J. W., Boulton, M. I., 2001b. A single amino acid change in the coat protein of Maize streak virus abolishes systemic infection, but not interaction with viral DNA or movement protein. Molecular Plant Pathology 2, 223-228.

Liu, L., Davies, J. W., Stanley, J., 1998a. Mutational analysis of bean yellow dwarf virus, a geminivirus of the genus Mastrevirus that is adapted to dicotyledonous plants. ~~J Gen Virol~~Journal of General Virology 79, 2265-2274.

Liu, L., Pinner, M. S., Davies, J. W., 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~~Journal of General Virology 80, 501-506.

Liu, L., Saunders, K., Thomas, C. L., Davies, J. W., 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.

Liu, L., van, T. T., Pietersen, G., Davies, J. W., Stanley, J., 1997b. Molecular characterization of a subgroup I geminivirus from a legume in South Africa. ~~J Gen Virol~~Journal of General Virology 78, 2113-2117.

Liu, Y., Robinson, D. J., Harrison, B. D., 1998b. Defective forms of cotton leaf curl virus DNA-A that have different combinations of sequence deletion, duplication, inversion and rearrangement. Journal of General Virology 79, 1501-1508.

Lopez-Bueno, A., Villarreal, L. P., Almendral, J. M., 2006. Parvovirus variation for disease: a difference with RNA viruses? ~~Curr.Top.Microbiol.Immunol.~~Current Topics in Microbiology and Immunology 299, 349-370.

Lotrakul, P., Valverde, R. A., 1999. Cloning of a DNA-A-like genomic component of sweet potato leaf curl virus: nucleotide sequence and phylogenetic relationships. ~~Mol.Plant Pathol.~~Molecular Plant Pathology On-line publication/1999/0206lotrakul.

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

Ma, C. M., Hon, C. C., Lam, T. Y., Li, V. Y. Y., Wong, C. K. W., de Oliveira, T., Leung, F. C. C., 2007. Evidence for recombination in natural populations of porcine circovirus type 2 in Hong Kong and mainland China. Journal of General Virology 88, 1733.

MacDonald, H., Coutts, R. H. A., Buck, K. W., 1988. Characterization of a subgenomic DNA isolated from *Triticum aestivum* plants infected with wheat dwarf virus. Journal of General Virology 69, 1339-1334.

MacDowell, S. W., Coutts, R. H., Buck, K. W., 1986. Molecular characterisation of subgenomic single-stranded and double- stranded DNA forms isolated from plants infected with tomato golden mosaic virus. ~~Nucleic.Acids.Res.~~ Nucleic Acids Research 14, 7967-7984.

Malpica, J. M., Sacristan, S., Fraile, A., Garcia-Arenal, F., 2006. Association and host selectivity in multi-host pathogens. PLoS ONE 1, e41.

Mansoor, S., Briddon, R. W., Bull, S. E., Bedford, I. D., Bashir, A., Hussain, M., Saeed, M., Zafar, Y., Malik, K. A., Fauquet, C., Markham, P. G., 2003a. Cotton leaf curl disease is associated with multiple monopartite begomoviruses supported by single DNA beta. Arch.Virol Archives of Virology 148, 1969-1986.

Mansoor, S., Briddon, R. W., Zafar, Y., Stanley, J., 2003b. Geminivirus disease complexes: an emerging threat. Trends in Plant Science 8, 128-134.

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

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562-563.

Martin, D. P., Rybicki, E. P., 1998. Microcomputer-based quantification of maize streak virus symptoms in *Zea mays*. Phytopathology 88, 422-427.

Martin, D. P., Rybicki, E. P., 2002. Investigation of Maize streak virus pathogenicity determinants using chimaeric genomes. Virology 300, 180-188.

Martin, D. P., van der Walt, E., Posada, D., Rybicki, E. P., 2005a. The evolutionary value of recombination is constrained by genome modularity. PLoS Genetics 1, e51.

Martin, D. P., Williamson, C., Posada, D., 2005b. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21, 260-262.

Martin, D. P., Willment, J. A., Billharz, R., Velders, R., Odhiambo, B., Njuguna, J., James, D., Rybicki, E. P., 2001. Sequence diversity and virulence in *Zea mays* of Maize streak virus isolates. Virology 288, 247-255.

Martin, D. P., Willment, J. A., Rybicki, E. P., 1999. Evaluation of maize streak virus pathogenicity in differentially resistant *Zea mays* genotypes. Phytopathology 89, 695-700.

McGeoch, D. J., Gatherer, D., 2005. Integrating reptilian herpesviruses into the family herpesviridae. J Virol Journal of Virology 79, 725-731.

Mirkin, E. V., Mirkin, S. M., 2007. Replication fork stalling at natural impediments. Microbiol.Mol Biol Rev. Microbiology and Molecular Biology Reviews 71, 13-35.

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

Molinier, J., Ries, G., Zipfel, C., Hohn, B., 2006. Transgeneration memory of stress in plants. Nature 442, 1046-1049.

Monci, F., Sanchez-Campos, S., Navas-Castillo, J., Moriones, E., 2002. A natural recombinant between the geminiviruses Tomato yellow leaf curl Sardinia virus and Tomato yellow leaf curl virus

exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. Virology 303, 317-326.

Morales, F. J., Anderson, P. K., 2001. The emergence and dissemination of whitefly-transmitted geminiviruses in Latin America. Archives of Virology 146, 415-441.

Moriones, E., Navas-Castillo, J., 2000. Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. ~~Virus Res~~Virus Research. 71, 123-134.

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

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

Muller, H. J., 1964. The ~~Relation~~ relation of ~~Recombination~~ recombination to ~~Mutational~~ mutational ~~Advance~~advance. Mutat~~.~~ion Research~~.~~ 106, 2-9.

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

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

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

Nadeem, A., Weng, Z., Nelson, M. R., Cotton leaf crumple virus and cotton leaf curl virus are two distantly related geminiviruses. Molecular Plant Pathology On-line.

Nahid, N., Amin, I., Mansoor, S., Rybicki, E. P., van der Walt, E., Briddon, R. W., 2008. Two dicot-infecting mastreviruses (family *Geminiviridae*) occur in Pakistan. Archives of Virology In press.

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

Ndunguru, J., Legg, J. P., Aveling, T. A., Thompson, G., Fauquet, C. M., 2005. Molecular biodiversity of cassava begomoviruses in Tanzania: evolution of cassava geminiviruses in Africa and evidence for East Africa being a center of diversity of cassava geminiviruses. ~~Virol.J.~~Virology Journal 2, 21.

Nee, S., 2000. Mutualism, parasitism and competition in the evolution of coviruses. Philosophical Transactions of the Royal Society B: Biological~~Philos.Trans.R.Soc.Lond B Biol Sci~~ 355, 1607-1613.

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

Novella, I. S., Duarte, E. A., Elena, S. F., Moya, A., Domingo, E., Holland, J. J., 1995. Exponential increases of RNA virus fitness during large population transmissions. ~~Proc.Natl.Acad.Sci.U.S.A~~ Proceedings of the National Academy of Sciences U.S.A.92, 5841-5844.

Odelberg, S. J., Weiss, R. B., Hata, A., White, R., 1995. Template-switching during DNA synthesis by *Thermus aquaticus* DNA polymerase I. ~~Nucleic Acids Res.~~Nucleic Acids Research 23, 2049-2057.

Olavarrieta, L., Hernβndez, P., Krimer, D. B., Schvartzman, J. B., 2002. Dna knotting caused by head-on collision of transcription and replication. ~~J.Mol.Biol~~Journal of Molecular Biology 322, 1-6.

Owor, B. E., Martin, D. P., Shepherd, D. N., Edema, R., Monjane, A. L., Rybicki, E. P., Thomson, J. A., Varsani, A., 2007a. Genetic analysis of maize streak virus isolates from Uganda reveals widespread distribution of a recombinant variant. Journal of General Virology 88, 3154-3165.

Owor, B. E., Shepherd, D. N., Taylor, N. J., Edema, R., Monjane, A. L., Thomson, J. A., Martin, D. P., Varsani, A., 2007b. Successful application of FTA Classic Card technology and use of bacteriophage phi29 DNA polymerase for large-scale field sampling and cloning of complete maize streak virus genomes. Journal of Virological Methods 140, 100-105.

Paabo, S., Irwin, D. M., Wilson, A. C., 1990. DNA damage promotes jumping between templates during enzymatic amplification. Journal of Biological Chemistry 265, 4718-4721.

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

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

Palmer, K. E., Rybicki, E. P., 2001. Investigation of the potential of maize streak virus to act as an infectious gene vector in maize plants. ~~Arch.Virol~~Archives of Virology 146, 1089-1104.

Palmer, K. E., Schnippenkoetter, W. H., Rybicki, E. P., 1998. Geminivirus ~~Isolation~~ isolation and DNA extraction. In: G. Foster, S. Taylor (Eds.), Plant Virology Protocols: From Virus Isolation to Transgenic Resistance. Humana Press, Totowa, NJ, pp. 41-52.

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

Patil, B. L., Dasgupta, I., 2006. Defective interfering ~~d~~DNA~~na~~s of plant viruses. Critical Reviews in Plant Sciences 25, 47-64.

Patil, B. L., Dutt, N., Briddon, R. W., Bull, S. E., Rothenstein, D., Borah, B. K., Dasgupta, I., Stanley, J., Jeske, H., 2007. Deletion and recombination events between the DNA-A and DNA-B components of Indian cassava-infecting geminiviruses generate defective molecules in *Nicotiana benthamiana*. ~~Virus Res~~Virus Research. 124, 59-67.

Paximadis, M., Rey, M. E., 2001. Genome organization of Tobacco leaf curl Zimbabwe virus, a new, distinct monopartite begomovirus associated with subgenomic defective DNA molecules. ~~J Gen Virol~~Journal of General Virology 82, 3091-3097.

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

Pilartz, M., Jeske, H., 2003. Mapping of abutilon mosaic geminivirus minichromosomes. ~~J Virol~~Journal of Virology 77, 10808-10818.

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M., Tans, S. J., 2007. Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445, 383-386.

Posada, D., Crandall, K. A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. ~~Proc Natl Acad Sci U.S.A~~Proceedings of the National Academy of Sciences U.S.A. 98, 13757-13762.

Preiss, W., Jeske, H., 2003. Multitasking in replication is common among geminiviruses. ~~J Virol~~Journal of Virology 77, 2972-2980.

Redinbaugh, M. G., 2003. Transmission of Maize streak virus by vascular puncture inoculation with unit-length genomic DNA. Journal of Virological Methods 109, 95-98.

Roberts, E. J., Buck, K. W., Coutts, R. H., 1988. Characterization of potato yellow mosaic virus as a geminivirus with a bipartite genome. Intervirology 29, 162-169.

Rojas, M. R., Hagen, C., Lucas, W. J., Gilbertson, R. L., 2005. Exploiting chinks in the plant's armor: evolution and emergence of geminiviruses. ~~Annu.Rev.Phytopathol.~~Annual Review of Phytopathology 43, 361-394.

Roossinck, M. J., 1997. Mechanisms of plant virus evolution. ~~Annu.Rev.Phytopathol.~~Annual Review of Phytopathology 35, 191-209.

Rothstein, R., Michel, B., Gangloff, S., 2000. Replication fork pausing and recombination or "gimme a break". ~~Genes Dev.~~Genes and Development 14, 1-10.

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

Saeed, M., Behjatnia, S. A., Mansoor, S., Zafar, Y., Hasnain, S., Rezaian, M. A., 2005. A single complementary-sense transcript of a geminiviral DNA beta satellite is determinant of pathogenicity. Mol Plant Microbe Interactions 18, 7-14.

Saeed, M., Zafar, Y., Randles, J. W., Rezaian, M. A., 2007. A monopartite begomovirus-associated DNA beta satellite substitutes for the DNA B of a bipartite begomovirus to permit systemic infection. ~~J Gen Virol~~Journal of General Virology 88, 2881-2889.

Sambrook, J., Fritsch, E. F., Maniatis, T., 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanderfoot, A. A., 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., Lazarowitz, S. G., 1996. Getting it together in plant virus movement: Cooperative interactions between bipartite geminivirus movement proteins. Tr~~.~~ends in Cell Biology 6, 353-358.

Sanz, A. I., Fraile, A., Gallego, J. M., Malpica, J. M., Garcia-Arenal, F., 1999. Genetic variability of natural populations of cotton leaf curl geminivirus, a single-stranded DNA virus. ~~J.Mol.Evol.~~ Journal of Molecular Evolution 49, 672-681.

Sanz-Burgos, A. P., 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., Bedford, I. D., Briddon, R. W., Markham, P. G., Wong, S. M., Stanley, J., 2000. A unique virus complex causes *Ageratum* yellow vein disease. ~~Proc.Natl.Acad.Sci.U.S.A.~~ Proceedings of the National Academy of Sciences U.S.A. 97, 6890-6895.

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

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

Saunders, K., Salim, N., Mali, V. R., Malathi, V. G., Briddon, R., Markham, P. G., Stanley, J., 2002. Characterisation of Sri Lankan cassava mosaic virus and Indian cassava mosaic virus: evidence for acquisition of a DNA B component by a monopartite begomovirus. Virology 293, 63-74.

Saunders, K., 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., Yahara, T., Stanley, J., 2003. Aetiology: The earliest recorded plant virus disease. Nature 422, 831.

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

Schneider, M., Jarchow, E., Hohn, B., 1992. Mutational analysis of the 'conserved region' of maize streak virus suggests its involvement in replication. ~~Plant Mol Biol~~Plant Molecular Biology 19, 601-610.

Schneider, W. L., Roossinck, M. J., 2000. Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. Journal of Virology 74, 3130-3134.

Schneider, W. L., Roossinck, M. J., 2001. Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. Journal of Virology 75, 6566-6571.

Schnippenkoetter, W. H., Martin, D. P., Hughes, F. L., Fyvie, M., Willment, J. A., James, D., von Wechmar, M. B., Rybicki, E. P., 2001a. The relative infectivities and genomic characterisation of three distinct mastreviruses from South Africa. ~~Arch Virol~~Archives of Virology 146, 1075-1088.

Schnippenkoetter, W. H., Martin, D. P., Willment, J. A., Rybicki, E. P., 2001b. Forced recombination between distinct strains of Maize streak virus. Journal of General Virology 82, 3081-3090.

Shackelton, L. A., Hoelzer, K., Parrish, C. R., Holmes, E. C., 2007. Comparative analysis reveals frequent recombination in the parvoviruses. Journal of General Virology 88, 3294-3301.

Shackelton, L. A., Holmes, E. C., 2004. The evolution of large DNA viruses: combining genomic information of viruses and their hosts. Trends in Microbiology 12, 458-465.

Shackelton, L. A., Holmes, E. C., 2006. Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. Journal of Virology 80, 3666-3669.

Shackelton, L. A., Parrish, C. R., Truyen, U., Holmes, E. C., 2005. High rate of viral evolution associated with the emergence of carnivore parvovirus. ~~Proc.Natl.Acad.Sci.U.S.A~~ Proceedings of the National Academy of Sciences U.S.A. 102, 379-384.

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

Shepherd, D. N., Mangwende, T., Martin, D. P., Bezuidenhout, M., Kloppers, F. J., Carolissen, C. H., Monjane, A. L., Rybicki, E. P., Thomson, J. A., 2007. Maize streak virus-resistant transgenic maize: a first for Africa. Plant Biotechnology ~~-~~Journal 5, 759-767.

Shepherd, D. N., Martin, D. P., McGivern, D. R., Boulton, M. I., Thomson, J. A., Rybicki, E. P., 2005. A three-nucleotide mutation altering the Maize streak virus Rep pRBR-interaction motif reduces symptom severity in maize and partially reverts at high frequency without restoring pRBR-Rep binding. Journal of General Virology 86, 803-813.

Shepherd, D. N., Varsani, A., Windram, O. P., Lefeuvre, P., Monjane, A. L., Owor, B. E., Martin, D. P., 2008. Novel sugarcane streak and sugarcane streak Reunion mastreviruses from southern Africa and La Reunion. Archives of Virology 153, 605-9.~~-~~

Simons, J. N., Coe, D. M., 1958. Transmission of pseudo-curly top virus in Florida by a treehopper. Virology 6, 43-48.

Simpson, C. G., Clark, G., Davidson, D., Smith, P., Brown, J. W., 1996. Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. Plant Journal 9, 369-380.

Snijder, E. J., den Boon, J. A., Horzinek, M. C., Spaan, W. J., 1991. Comparison of the genome organization of toro-and coronaviruses: evidence for two nonhomologous RNA recombination events during Berne virus evolution. Virology 180, 448-452.

Stanley, J., 1983. Infectivity of the cloned geminivirus genome requires sequences from both DNAs. Nature 305, 643-645.

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

Stanley, J., 2004. Subviral DNAs associated with geminivirus disease complexes. ~~Vet.Microbiol.~~Veterinary Microbiology 98, 121-129.

Stanley, J., Bisaro, D. M., Briddon, R., Brown, J. K., Fauquet, C., Harrison, B. D., Rybicki, E., Stenger, D. C., 2005. *Geminiviridae*. In: C. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, L. A. Ball (Eds.), Virus Taxonomy, VIIIth Report of the ICTV. Elsevier / Academic Press, London, pp. 301-326.

Stanley, J., Frischmuth, T., Ellwood, S., 1990. Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. ~~Proc.Natl.Acad.Sci.~~ Proceedings of the National Academy of Sciences U.S.A. 87, 6291-6295.

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

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

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

Stanley, J., Saunders, K., Pinner, M. S., Wong, S. M., 1997. Novel defective interfering DNAs associated with ageratum yellow vein geminivirus infection of *Ageratum conyzoides*. Virology 239, 87-96.

Stanley, J., Townsend, R., 1985. Characterisation of DNA forms associated with cassava latent virus infection. ~~Nucleic.Acids.Res.~~ Nucleic Acids Research 13, 2189-2206.

Stasolla, C., Katahira, R., Thorpe, T. A., Ashihara, H., 2003. Purine and pyrimidine nucleotide metabolism in higher plants. Journal of Plant Physiology 160, 1271-1295.

Stemmer, W. P. C., 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389-391.

Stenger, D. C., Carbonaro, D., Duffus, J. E., 1990. Genomic characterization of phenotypic variants of beet curly top virus. Journal of General Virology 71, 2211-2215.

Stenger, D. C., McMahon, C. L., 1997. Genotypic diversity of beet curly top virus populations in the western United States. Phytopathology 87, 737-744.

Stenger, D. C., Revington, G. N., Stevenson, M. C., 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.~~ Proceedings of the National Academy of Sciences U.S.A. 88, 8029-8033.

Stenger, D. C., Stevenson, M. C., Hormuzdi, S. G., Bisaro, D. M., 1992. A number of subgenomic DNAs are produced following agroinoculation of plants with beet curly top virus. Journal of General Virology 73, 237-242.

Stenger, D. C., Seifers, D. L., French, R., 2002. Patterns of polymorphism in wheat streak mosaic virus: sequence space explored by a clade of closely related viral genotypes rivals that between the most divergent strains. Virology 302, 58.

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

Takeuchi, Y., Horiuchi, T., Kobayashi, T., 2003. Transcription-dependent recombination and the role of fork collision in yeast rDNA. ~~Genes Dev.~~Genes and Development 17, 1497-1506.

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

van Wezel, R., Dong, X., Blake, P., Stanley, J., Hong, Y., 2002. Differential roles of geminivirus Rep and AC 4(C 4) in the induction of necrosis in *Nicotiana benthamiana*. Molecular Plant Pathology 3, 461-471.

Vanderschuren, H., Stupak, M., Futterer, J., Gruissem, W., Zhang, P., 2007. Engineering resistance to geminiviruses--review and perspectives. Plant Biotechnology -Journal 5, 207-220.

Varma, A. N. U. P., Malathi, V. G., 2003. Emerging geminivirus problems: A serious threat to crop production. Annals of Applied Biology 142, 145-164.

Varsani, A., Oluwafemi, S., Windram, O. P., Shepherd, D. N., Monjane, A. L., Owor, B. E., Rybicki, E. P., Lefeuvre, P., Martin, D. P., 2008. *Panicum* streak virus diversity is similar to that observed for maize streak virus. Archives of Virology 153, 601-604.

von Arnim, A., Frischmuth, T., Stanley, J., 1993. Detection and possible functions of African cassava mosaic virus DNA B gene products. Virology 192, 264-272.

von Arnim, A., Stanley, J., 1992a. Determinants of tomato golden mosaic virus symptom development located on DNA B. Virology 186, 286-293.

von Arnim, A., Stanley, J., 1992b. Inhibition of African cassava mosaic virus systemic infection by a movement protein from the related geminivirus tomato golden mosaic virus. Virology 187, 555-564.

Wang, X., Ira, G., Tercero, J. A., Holmes, A. M., Diffley, J. F., Haber, J. E., 2004. Role of DNA replication proteins in double-strand break-induced recombination in *Saccharomyces cerevisiae*. Mol-ecular and Cellular -Biology 24, 6891-6899.

Ward, B. M., Medville, R., Lazarowitz, S. G., Turgeon, R., 1997. The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. Journal of Virology 71, 3726-3733.

Willment, J. A. Investigation of the molecular determinants of *Maize streak virus* replication. 1-207. 1999. University of Cape Town. 1999.
Ref Type: Thesis/Dissertation

Willment, J. A., Martin, D. P., Rybicki, E. P., 2001. Analysis of the diversity of African streak mastreviruses using PCR-generated RFLPs and partial sequence data. ~~J Virol~~Journal of Virologicaly~~t~~ Methods 93, 75-87.

Willment, J. A., Martin, D. P., van der Walt, E., Rybicki, E. P., 2002. Biological and genomic sequence characterization of *Maize streak virus* isolates from wheat. Phytopathology 92, 81-86.

Worobey, M., Holmes, E. C., 1999. Evolutionary aspects of recombination in RNA viruses. Journal of General Virology 80, 2535-2543.

Worobey, M., 2000. Extensive homologous recombination among widely divergent ~~tt~~ TT viruses. ~~The Journal of Virology~~ Journal of Virology 74, 7666-7670.

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

Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the 6th International Congress of Genetics 1, 356-366. 1932.
Ref Type: Conference Proceeding

Xia, X., Yuen, K. Y., 2005. Differential selection and mutation between dsDNA and ssDNA phages shape the evolution of their genomic AT percentage. BMC Genetics~~.~~ 6, 20.

Xie, Q., Sanz-Burgos, A. P., Guo, H., Garcia, J. A., Gutierrez, C., 1999. GRAB proteins, novel members of the NAC domain family, isolated by their interaction with a geminivirus protein. Plant Molecular Biology 39, 647-656.

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

Xie, Q., Suarez-Lopez, P., 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~~ EMBO Journal 14, 4073-4082.

Xie, Y., Zhou, X. P., 2003. Molecular characterization of squash leaf curl Yunnan virus, a new begomovirus and evidence for recombination. ~~Arch. Virol~~Archives of Virology 148, 2047-2054.

Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E., Lieber, M. R., 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. Nature ~~.~~Immunology~~.~~ 4, 442-451.

Zaphiropoulos, P. G., 1998. Non-homologous recombination mediated by *Thermus aquaticus* DNA polymerase I~~.~~: Evidence supporting a copy choice mechanism. Nucleic Acids Research 26, 2843-2848.

Zhan, X., Richardson, K. A., Haley, A., Morris, B. A., 1993. The activity of the coat protein promoter of *C~~.~~hloris* striate mosaic virus is enhanced by its own and C1-C2 gene products. Virology 193, 498-502.

Zhang, W., Olson, N. H., Baker, T. S., Faulkner, L., Agbandje-McKenna, M., Boulton, M. I., Davies, J. D., McKenna, R., 2001. Structure of the maize streak virus geminate particle. Virology 279, 471-477.

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

Zhou, X., Liu, Y., Robinson, D. J., 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~~Journal of General Virology 79, 915-923.

Zhou, X. P., Xie, Y., Zhang, Z. K., Qi, Y. J., Wu, J. J., 2001. Molecular characterization of a novel defective DNA isolated from tobacco tissues infected with tobacco leaf curl virus. Acta Virologica ~~Acta Virol~~ 45, 45-50.

Zhou, X., Xie, Y., Tao, X., Zhang, Z., Li, Z., Fauquet, C. M., 2003. Characterization of DNA ~~(~~beta~~)~~ associated with begomoviruses in China and evidence for co-evolution with their cognate viral DNA-A. Journal of General Virology 84, 237-247.