2.
PS

MOLECULAR INVESTIGATIONS OF SUBGROUP I GEMINIVIRUSES

by

FIONA LESLEY HUGHES

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Faculty of Science, University of Cape Town.

December 1990

CERTIFICATION BY SUPERVISOR

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

Signed by candidate

Senior Lecturer in Virology

Department of Microbiology

University of Cape Town.

30th November, 1990

ACKNOWLEDGEMENTS

I would like to express my thanks to the following:

My supervisor, Dr. Ed Rybicki, for his enthusiasm and encouragement on this project, over and above his duties as a supervisor. Thanks also for pushing new techniques, Ed !

The Foundation for Research Development for their generous financial support of this project.

Professor Ralph Kirby (currently HOD, Department of Microbiology, Rhodes University, Grahamstown, Cape) for his freely given ideas and advice when I hit problems, and for a sympathetic ear.

Professor von Wechmar for her considerable assistance in procuring specimens, for transferring streak isolates between hosts, and plant maintenance. My thanks also to the plant room staff for plant maintenance.

Katherine Harborne and Dr Jean-Claude Autrey for streak-diseased sugarcane specimens.

Dr Trevor Sewell and his assistants (Department of Biochemistry), and Mr Kassie Kasdorf (Plant Protection Institute, Stellenbosch, Cape), for assistance with electron microscopy.

Michelle Babaya and Erik Edge for the use of restriction map data.

My research colleagues and the technical staff of the department for providing support, and the camaraderie that makes working in a laboratory worthwhile.

Finally, I wish to thank my family, both immediate and more distant, for their love and oft-needed moral support.

A special word of appreciation is due to my husband, Guy, for listening patiently about the inevitable daily setbacks and projected long term gloom (and for so ably picking up the technical jargon in order to allow me to do so without having to constantly explain what I am talking about), and for fiercely championing my cause, whether or not it was actually necessary.

CONTENTS

MOLECULAR INVESTIGATIONS OF SUBGROUP I GEMINIVIRUSES

By

Fiona Lesley Hughes

Department of Microbiology, University of Cape Town, Rondebosch 7700, Cape, Republic of South Africa.

ABSTRACT

The diversity of Subgroup I geminiviruses causing streak disease in maize, sugarcane, and indigenous wild grasses was investigated. The virus isolates studied originated from maize (several southern African isolates), two sugarcane cultivars (from Natal province, South Africa, and from Mauritius), wheat, and three grasses (*Panicum*, *Setaria*, and *Eleusine* spp. from South Africa). The following methods were used: analysis of restriction fragment length polymorphisms (RFLPs) between viral genomes in individual infected plants; DNA cross-hybridization between virus isolates; restriction endonuclease mapping of whole virus genomes; and nucleic acid sequencing. The complete genome of the Natal sugarcane streak virus isolate was sequenced. Partial sequences were obtained for other isolates, either by sequencing the ends of cloned viral genomes, or by sequencing a 250 base pair fragment of a highly conserved open reading frame that had been amplified using the polymerase chain reaction technique. The viruses being studied were compared both among themselves and with other Subgroup I geminiviruses of known DNA sequence, on the basis of sequence (nucleotide and amino acid) and restriction map data. Distance matrix methods were used to infer phylogenetic relationships between Subgroup I geminiviruses from restriction map and sequence data. Phylogenies deduced from sequence data were considered to be more accurate than those deduced from map data. Regardless of the method of analysis used, however, the relationships between the Subgroup I geminiviruses studied here remained constant. Thus, three strains of MSV (maize, *Setaria*, and *Eleusine* strains) were distinguished. Streak viruses distinct from MSV were also identified: panicum streak virus (PanSV), and two distantly related strains (Natal and Mauritius) of sugarcane streak virus (SSV). Restriction mapping of different geographical isolates of the maize strain of MSV demonstrated

that variation existed within a single strain of virus. RFLP analysis indicated that minor variation existed between virus genomes within single diseased plants.

Methods used to type Subgroup I geminiviruses were evaluated, and discrepancies in the serological typing of geminiviruses from Subgroups I and III were pointed out. A unified scheme was proposed for distinguishing between distinct Subgroup I geminiviruses and strains of geminiviruses. The origins of maize and sugarcane streak viruses were speculated upon.

ABBREVIATIONS

1. Viruses.

ACMV	African cassava mosaic virus
BCTV	beet curly top virus
BGMV	bean golden mosaic virus
CSMV	chloris striate mosaic virus
DSV	digitaria streak virus
MSV	maize streak virus
MiSV	miscanthus streak virus
PSMV	paspalum striate mosaic virus
SSV	sugarcane streak virus
TGMV	tomato golden mosaic virus
TLCV	tobacco leaf curl virus
WDV	wheat dwarf virus

2. Nucleic Acid Terms.

A, C,	adenosine, cytosine
G, T, U	guanosine, thymidine, uridine
kb	kilobase
bp	base pair(s)
ccc	covalently closed circular
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
(d)dNTP	(di)deoxy nucleotide triphosphate
ds	double stranded
IR	intergenic region
N	nucleotide
ORF	open reading frame
RF	replicative form
RFLP	restriction fragment length polymorphism
RNase	ribonuclease
(t)RNA	(transfer) ribonucleic acid
ss	single stranded

3. Serological Terms.

(M)Ab	(monoclonal) antibody
Ag	antigen
DAS-ELISA	double antibody sandwich-ELISA
ELISA	enzyme-linked immunosorbent assay
ISEM	immunosorbent electron microscopy
SDI	serological differentiation index

4. Units.

k, m, μ, n	kilo-, milli-, micro-, nano- (prefixes)
m, l, g	metre, litre, gram
Ci	Curie(s)
°C	degrees Centigrade
Da	Dalton(s)
mol	mole
M	molar

5. Chemicals.

DMSO	dimethylsulphoxide
EDTA	ethylenediaminetetraacetic acid
EthBr	ethidium bromide
EtOH	ethanol
KOAc	potassium acetate
MOPS	3-[N-morpholino]propane sulphonic acid
NaOAc	sodium acetate
NBT	Nitro Blue tetrazolium
SDS	sodium dodecyl sulphate
SSC	salt-sodium citrate buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
X-gal	5-bromo-4-chloro-3-indolyl galactosidase
X-phosphate	5-bromo-4-chloro-3-indolyl phosphate

6. Miscellaneous.

cv.	cultivar
h	hour(s)
min	minute(s)
M_r	relative molecular mass
MW	molecular weight
OD_{600}	optical density at 600 nm wavelength
PAGE	polyacrylamide gel electrophoresis
rpm	revolutions per minute
sp(p)	species (plural)
Tm	melting temperature
UV	ultra violet
vol	volume
v/v	volume per volume ratio
w/v	weight per volume ratio

CHAPTER I

GENERAL INTRODUCTION.

CONTENTS

CHAPTER I.
GENERAL INTRODUCTION.

1.1. Historical Background.

Maize streak disease was first described as "mealie variegation" by Fuller (1901) in Natal, South Africa. The causal agent was discovered to be a virus by Storey (1924, 1932a), who termed it maize streak virus (MSV). The virus was found to be obligately transmitted by the leafhopper *Cicadulina mbila* (Naude) (Storey, 1924, 1925a). Subsequently, several other *Cicadulina* species were reported as being able to transmit MSV (see review by Rose, 1978; Dabrowski, 1987). In 1978, MSV was designated the type virus of the newly described group taxon Geminivirus (Harrison *et al.*, 1977; Matthews, 1979).

Early studies indicated that there were several distinctly different African streak viruses adapted to different host ranges (Storey & McClean, 1930; McClean, 1947). These studies were based on the transmission of virus isolates between different host species and symptomatology.

In a subsequent study of streak virus transmission between maize, sugarcane, and *Panicum maximum*, the relatively new technique of immunodiffusion was employed, using antiserum to the maize isolate. From the results it was concluded that the maize, sugarcane, and *Panicum* isolates were strains of the same virus, MSV (Bock *et al.*, 1974). The maize isolate was given as the type strain.

Since that date other transmission tests (Pinner *et al.*, 1988) and more sophisticated serological assays (Pinner *et al.*, 1988; Dekker *et al.*, 1988; Pinner & Markham, 1990) have been performed on a wide range of streak isolates from different hosts and locales, and it has been claimed that all forms of streak disease in the Gramineae in Africa are caused by strains of the same virus, MSV (Damsteegt, 1983; Pinner *et al.*, 1988; Dekker *et al.*, 1988; Pinner & Markham, 1990b).

In southern Africa there is a wide range of naturally occurring MSV isolates. Consequently, there is no necessity for strict quarantine in this part of the world when working with MSV isolates, unlike the situation pertaining to Europe or America. These conditions make this location ideal for working with a wide range of streak isolates and potential host species.

Accordingly, this laboratory has been involved for some time in establishing the extent of variation in MSV in the southern African subcontinent. It was hoped

that in so doing, some insight into the evolutionary pattern of the virus would be gained (Clarke, 1987; Kirby *et al.*, 1989). In the process of studying MSV variation, nucleic acid techniques have been explored for use in geminivirus typing. These originally included DNA cross-hybridization (Southern- and dot-blotting), restriction fragment length polymorphism comparisons, and restriction endonuclease mapping. (Clarke *et al.*, 1989; Rybicki *et al.*, 1989; Hughes *et al.*, 1990a).

1.2. Project Aims.

The initial aims of the project were to continue and expand on the work initiated in this laboratory in cloning and mapping maize isolates of MSV from different regions of southern Africa, and to continue investigating the use of restriction endonuclease mapping as a reliable means of typing MSV isolates.

During the early stages of this study, an isolate of streak diseased sugarcane was obtained from Natal, South Africa. The DNA hybridization and restriction map properties of this isolate differed substantially from those of the maize isolates of MSV collected thus far. Consequently, the project was altered to include a comparison of streak isolates from grasses with those from maize. Finally, the overall objective came to be to determine whether all African grass streak isolates are indeed strains of MSV, or whether "MSV" is in fact a loose description for a cluster of related but distinct viruses.

1.3. Geminiviruses.

1.3.1. Geminivirus Taxonomy.

The Geminiviruses, formally recognized in 1978 by the International Committee on the Taxonomy of Viruses (ICTV), are one of the most recently described taxonomic groups of plant pathogenic viruses (Harrison *et al.*, 1977; Matthews, 1979). These viruses are distinguished by their unique virion morphology and nucleic acid characteristics. Geminiviruses have twinned quasi-icosahedral capsids (hence "geminate") measuring approximately 20 nm x 30 nm, and their genomes consist of either one or two circular molecules of single-stranded DNA of between 2500 and 2800 bases in length (Lazarowitz, 1987). Members of this group that have monopartite genomes (see below) are the smallest known autonomously

replicating viruses (Howarth & Goodman, 1982; Lazarowitz *et al.*, 1989). Geminiviruses are transmitted by insect vectors (Harrison et al., 1977).

Until recently, only two geminivirus subgroups were recognized formally (Matthews, 1979; Lazarowitz, 1987). Subgroup I included all monopartite genome geminiviruses and Subgroup II included the bipartite genome geminiviruses.

Proposals have been made that three subgroups should be recognized (eg. Stanley, 1985; Howarth & Vandemark, 1979), and a new classification scheme for the geminiviruses has just been approved by the ICTV (E. P. Rybicki, pers. comm.). The geminiviruses are classified as follows:

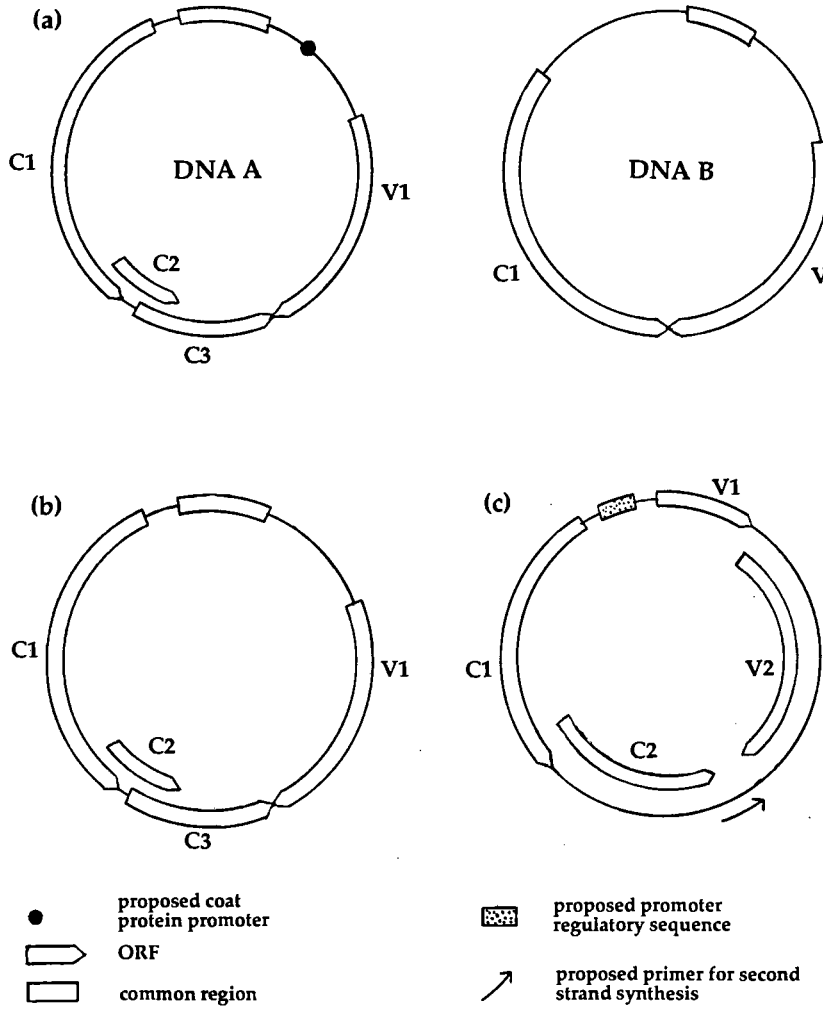
Subgroup I - those viruses causing streaking and stunting symptoms in monocotyledonous plants (in the Gramineae) and being obligately transmitted by leafhopper vectors. Members of this subgroup have monopartite genomes (see Fig. 1.1). The known members are MSV, digitaria streak virus (DSV, Dollet *et al.*, 1986), wheat dwarf virus (WDV, Vacke, 1972; Lindsten *et al.*, 1980), chloris striate mosaic virus (CSMV, Grylls, 1963; Francki *et al.*, 1980) and miscanthus streak virus (MiSV, Yamashita *et al.*, 1985).

Subgroup II - those viruses having monopartite genomes and being obligately transmitted by leafhoppers, but infecting dicotyledonous hosts. Their genome structure is closer to that of Subgroup III geminiviruses than to that of the Subgroup I viruses (see Fig. 1.1). Some examples from this Subgroup are beet curly top virus (BCTV, Mumford, 1974) and tobacco yellow dwarf virus (TYDV, Osaki & Inouye, 1978).

Subgroup III - those viruses causing mosaic or leaf curl symptoms in dicotyledonous plants and being transmitted by the whitefly, *Bemisia tabaci*. Viruses in this subgroup have bipartite genomes (see Fig. 1.1). Some examples are tomato golden mosaic virus (TGMV, Hamilton *et al.*, 1981), bean golden mosaic virus (BGMV, Goodman, 1977), and African cassava mosaic virus (ACMV, Bock *et al.*, 1978).

In this thesis, I will be concerned only with the Subgroup I Geminiviruses. These are also referred to in the body of the thesis as "grass" or "cereal geminiviruses". Accordingly, the rest of this literature survey will be concerned almost solely with properties of the Subgroup I Geminiviruses.

Figure 1.1. Genome organization in geminiviruses.



Legend: (a) Subgroup III geminivirus bipartite genome. (b) Subgroup II geminivirus monopartite genome. (c) Subgroup I geminivirus monopartite genome. V = virion sense open reading frame. C = complementary sense open reading frame.

1.3.2. Geographical Occurrence of Cereal Geminiviruses.

The grass geminiviruses are widely distributed in the Old World, occurring in roughly three different geographical regions: Australia and the Far East; Africa, nearby islands and possibly the Indian subcontinent; and northern Europe.

In the Australasian region, CSMV and related viruses occur in Australia (Grylls, 1963; Greber, 1989) and MiSV occurs in Japan (Yamashita *et al.*, 1985). DSV was discovered on the Pacific island Vanuatu, near Australia (Dollet *et al.*, 1986), but appears to be more closely related to MSV than to CSMV (Donson *et al.*, 1987; Andersen *et al.*, 1988; Hughes *et al.*, 1990b).

WDV is the only known European grass geminivirus to date. It has been reported from Sweden (Lindsten *et al.*, 1980), Czechoslovakia (Woolston *et al.*, 1988), and from Hungary (Bisztray & Gaborjanyi, 1989).

A number of viruses at present identified as MSV have been found in Africa, ranging from Egypt to South Africa (Storey & McClean, 1930; Bock *et al.*, 1974; Ammar *et al.*, 1982; Pinner *et al.*, 1988), as well as from Mauritius (Autrey & Ricaud, 1983). A streak virus, presumed to be the *Pennisetum* strain of MSV, has been reported on *Pennisetum* and wheat in India (Seth et al., 1972a, 1972b). Whether the virus is more closely related to MSV or to the Australasian viruses is uncertain. Indeed, there is no conclusive proof as yet that the Indian virus is in fact a geminivirus (Soto & Buddenhagen, 1982).

1.3.3. Transmission of Cereal Geminiviruses.

Cereal geminiviruses can be transmitted neither mechanically nor by seed, their only natural mode of transmission being by leafhoppers (Storey, 1928; Harrison, 1985). Unlike the dicot-infecting geminiviruses, which are all transmitted by a single whitefly species, grass geminiviruses are transmitted by a wide range of genera and species in the Cicadellideae. MSV is transmitted by several *Cicadulina* spp. (see refs. below). DSV is transmitted by *Nesoclutha declivata* (Linnavuori) (Julia & Dollet, 1989) and cannot be transmitted by *Cicadulina* spp. (Julia & Dollet, 1989; Pinner *et al.*, 1989). WDV is transmitted by *Psammotettix alienus* (Dhalb.) (Lindsten *et al.*, 1980). CSMV and related viruses are transmitted by *Nesoclutha pallida* (Evans) (Grylls, 1963), and the vector of MiSV is unknown (Ikegami *et al.*, 1989).

The best-studied vector/virus system for this geminivirus subgroup is that of MSV, investigated in great detail by Storey (1928, 1932).

In southern and central Africa, MSV is transmitted chiefly by *Cicadulina mbila* (Storey, 1924, 1925). In West Africa, *C. triangula* Ruppel is the major vector (Soto & Buddenhagen, 1982; Okoth *et al.*, 1987), while in Egypt the main streak vector is *C. bipunctella-zeae* China (Ammar *et al.*, 1982). Other species of *Cicadulina* capable of transmitting MSV are *C. bipunctata bipunctata* (Melichar) (*C. zeae*) (Storey, 1936), *C. latens* Fennah, *C. parazeae* Ghauri, *C. storeyi* China, (Bock et al., 1974), *C. ghaurii* (Dabrowski, 1987a), and *C. hartmansi* Dabrowski (Okoth *et al.*, 1987).

Not all races of a vector species are active, ie. capable of virus transmission (Storey, 1928). A bioassay of *C. ghaurii* showed that an average of 10% of males and 30% of females were capable of transmission after a 4h acquisition feed (Dabrowski, 1987b). The ability to transmit virus seems to be controlled by a dominant sex-linked gene (Storey, 1928, 1932b).

MSV can be acquired by active races of *C. mbila* after an acquisition probe into the mesophyll of less than 1 hour (minimum acquisition time 15 seconds) and inoculation can occur in as brief a probe as 5 minutes, during which time the insect's stylets are thought to reach the phloem (Storey, 1938). However, the frequency of transmission by, and persistence of MSV in, individual insects increases with the virus concentration in the plant, the length of the acquisition access period, and on the duration of inoculation (Okoth *et al.*, 1987). Similar findings have been made for CSMV and its vector, *N. pallida* (Grylls, 1963).

MSV undergoes a latent period in the vector before transmission can occur. The length of this period is temperature dependent. At 30°C, the minimum latent period is 6 - 12 hours (Storey, 1928), the medium period being 16 - 20 hours (Okoth *et al.*, 1987). At 16°C, the minimum latent period is 85 hours (Storey, 1928). The latent period represents the time required for the virus to pass from the insect's gut lumen to the haemocoel and thence to the salivary glands, where it becomes available for inoculation (Storey, 1932b).

The gene that controls the insect's virus transmitting ability appears to act by controlling the passage of virus to the haemolymph. This so, since inactive leafhoppers can be rendered active by puncturing the gut with a needle (Storey, 1932b). This evidence has been taken to imply that a single vector protein regulates virus passage through the gut wall (Harrison, 1985). However, since individual insects transmit virus with differing degrees of efficiency, it has been assumed that other autosomal genes modify the effect of the major gene controlling transmission (Storey, 1932b). Indeed, Medina *et al.* (1990) have shown that there are two barriers to infection; transport of virus across the insect's gut lining, and virus entry into the salivary glands. They demonstrated that MSV particles are taken up into cells of the filter chamber in active vectors only. Virus-vector specificity is consistent with receptor-mediated endocytosis, followed by envelopment of aggregated virus by the endoplasmic reticulum. A rapid transcytotic pathway was proposed for virus in

cells of the ventriculus, with inclusions probably breaking up in the haemolymph. From there, virions enter the salivary glands via specialized secretory cells (Medina *et al.*, 1990). It appears that a protease inhibitor motif exists in the virion coat protein sequence, which may be responsible for resistance to proteolysis in the insect gut (P. G. Markham, pers. comm. to E. P. Rybicki).

MSV can be acquired and transmitted by all 5 nymphal instars of *C. mbila* and is retained through ecdysis, although the virus does not appear to be capable of replication in the vector (Boulton & Markham, 1986). The virus cannot be transmitted to the following insect generation through the egg (Storey, 1928).

For CSMV at least, the proportion of insects transmitting virus, the length of the latent period, and the duration of transmission are dependent on the specific acquisition source (Greber, 1989). Thus, for field-infected *Chloris gayana*, 10% of feeding insects could transmit virus over 3 - 10 days after a latent period of 3 - 7 days. Using freshly infected *C. gayana* or wheat as acquisition sources, 50% of feeding insects could transmit virus for up to 9 weeks after a latent period of only 12 hours.

Little is known definitively about whether specific viruses or strains of a virus are preferentially adapted to particular species or races of vector. It has been shown recently though that paspalum striate mosaic virus (PSMV), which is related to CSMV, is transmitted by only one specific biotype of *N. pallida* (Greber, 1989).

1.3.4. Host Ranges of Cereal Geminiviruses.

The host range of Subgroup I geminiviruses is confined exclusively to the Gramineae. Unlike the whitefly-transmitted geminiviruses, however, grass geminiviruses generally have wide, often overlapping, host ranges. Both annual and perennial grasses as well as cereal crops are infected. In part this reflects the wide feeding range of their insect vectors (Rose, 1978; Damsteegt, 1983). For example, abundant oviposition and nymphal development of *C. mbila* is supported on several grass spp. in the genera *Aegilops, Andropogon, Avena, Bothriochloa, Digitaria, Echinochloa, Eleusine, Hyparrhenia, Panicum, Schizachrium, Sorghastrum, Sorghum, Trichachne, Trichloris,* and *Zea* (Damsteegt, 1983).

DSV was originally isolated from *Digitaria sanguinalis* (*D. setigera*; see Pinner & Markham, 1990b) (Dollet *et al.*, 1986). MiSV has been isolated from *Miscanthus*

saccifloris (Yamashita *et al.*, 1985). WDV infects a variety of grasses and most cereal crops including *Triticum, Avena,* and *Hordeum* (Vacke, 1972). However, the host ranges of DSV, MiSV, and WDV have not yet been fully explored.

The most detailed host range studies of Subgroup I geminiviruses have been performed on MSV and CSMV and illustrate the extensive host ranges to be found for Subgroup I geminiviruses.

In southern Africa, streak disease has been recorded in wild and cultivated grass species in the following tribes: Andropogoneae (*Cymbopogon, Imperata, Rottboelia*); Eragrosteae (*Dactyloctenium, Diplachne, Eleusine, Eragrostis, Leptochloa, Setaria*); Paniceae (*Digitaria, Panicum, Paspalum*); Sporoboleae (*Sporobolus*); Zoysieae (*Tragus*); Maydeae (*Zea, Euchlaena*); Hordeae (*Hordeum*) and Aveneae (*Avena*) (Storey, 1925b; Storey & McClean, 1930; McClean, 1947). Each grass variety studied tended to have a strain of MSV that was specialized to it ("host-adapted") and either non-virulent or only weakly virulent to other plant species (Storey & McClean, 1930; Bock, 1982).

While for most of these hosts, the identity of the infecting virus or strain of virus has not been rigorously tested, McClean (1947) distinguished between five different viruses. These comprised a severe ("A-type") and a mild ("B-type") strain of MSV, a sugarcane streak virus, a *Sporobolus* virus, and a virus that was apparently common to *Eleusine indica* and *Paspalum notatum.*

Damsteegt (1983) found that for the A-type MSV alone, 54 of 138 grass accessions from around the world were susceptible upon experimental transmission using *C. mbila.* The susceptible hosts (a total of 33 genera and 54 species) came from the tribes Andropogoneae, Aveneae, Chlorideae, Festuceae, Glycerieae, Hordeae, and Paniceae. Within each genus tested, not all species were susceptible and some of the susceptible species showed a low infection incidence. This was possibly due to plant heterogeneity and/or vector feeding preferences (Damsteegt, 1983). In addition, MSV type-A infected all cereal crops tested (barley, rye, wheat, oats, maize, and rice), as well as several subspecies of *Zea mays* and various *Tripsacum* accessions (Damsteegt, 1983).

In Mauritius, natural streak infections have been recorded on *Coix, Cenchrus, Brachiaria, Panicum, Paspalum,* and *Digitaria* spp. as well as on *Z. mays* and *Saccharum* hybrids (Autrey & Ricaud, 1983). Distinct, host-adapted strains were obtained from

these grasses, and Autrey & Ricaud (1983) considered that only the viruses found on *Coix lachryma-jobi, Cenchrus echinatus, Brachiaria reptens,* and *B. eruciformis* to be epidemiologically relevant to MSV infection of cultivated maize in Mauritius.

A study has been made of the host ranges of CSMV and related Australian grass geminiviruses (Greber, 1989). Of 25 grasses and cultivated cereals tested, the type strain of CSMV alone infected species from 15 genera, either naturally or experimentally. Oats, barley, wheat, maize, and the grasses *Dactyloctenium aegypticum* and *Leptochloa filiformis* were infected by all of the viruses examined (Greber, 1989). Host range and ease of host infection varied greatly between the viruses tested. *Microlaena stipoides* was an exclusive host for CSMV-M, a strain of CSMV. *Bromus catharticus* was exclusive for the *B. catharticus* geminivirus, while *Paspalum* spp. were readily infected only by isolates of paspalum striate mosaic virus (PSMV) (Greber, 1989).

Even cereal geminiviruses recognized as being distinct from one another and coming from widely different parts of the world often show overlapping host ranges. Thus, WDV, MSV, and CSMV all naturally infect *Hordeum vulgare* and *Avena sativa,* the latter being recognized as the principal host of WDV (Lindsten *et al.,* 1980; Damsteegt, 1983; Greber, 1989). Likewise, CSMV, PSMV and MSV all naturally infect *Zea mays,* the principal host of MSV (Greber, 1989; Damsteegt, 1983). Besides crop plant hosts, CSMV and MSV share a number of wild grass hosts, e.g. *Chloris gayana, Eleusine indica, Lolium multiflorum, Paspalum conjugatum,* and *Setaria italica* (Greber, 1989; McClean, 1947; Damsteegt, 1983; Pinner *et al.,* 1988).

Recently, an artificial inoculation method termed "agroinfection" has been developed (Grimsley *et al.,* 1987). Using this technique, whole geminivirus genomes can be cloned into the Ti plasmid of *Agrobacterium tumefaciens* and inoculated into meristematic host tissue. Apparently normal infection ensues, and normal virus particles can be recovered in quantity (Grimsley et al., 1988). By this means DSV was successfully agroinoculated into maize and *A. sativa,* as well as into *D. setigera* (initially identified incorrectly as *D. sanguinalis*) (Donson *et al.,* 1988), thereby apparently extending its potential host range.

1.3.5. Epidemiology and Economic Importance of Cereal Geminiviruses.

MSV and WDV cause serious disease in maize and wheat crops (Van Rensburg & Kuhn, 1977; Rose, 1978; Mzira, 1984; Kim *et al.*, 1989; Lindsten *et al.*, 1980). Streak disease affects barley, wheat, oats, sugarcane, pearl millet (*Pennisetum typhoides*) and finger millet (*Eleusine coracana*) crops in southern and East Africa, but is of minor economic importance (Soto & Buddenhagen, 1982). The other members of Subgroup I mainly affect wild grasses and have little economic impact (Greber, 1989). Again, MSV is epidemiologically the best studied virus of this subgroup.

Yield loss in maize due to maize streak disease is directly related to time of infection - infected seedlings produce no yield or are killed, while plants infected at the 2nd, 4th, 6th, 8th, and 10th leaf stages lose approximately 55%, 45%, 40%, 33%, and 25% in grain weight loss, respectively (Bock, 1982).

MSV epidemics are more severe in tropical regions (Rose, 1978; Autrey & Ricaud, 1983). Virus spread between crops is facilitated by successive cropping and the presence of wild grasses as reservoirs of both virus and vectors (Autrey & Ricaud, 1983), as well as the wide variety of plant species that leafhoppers feed on, the ability of the vectors to transmit MSV persistently, and the insects' often considerable migration distances (Rose, 1978). In the warm wet season, the fecund, long-bodied form of *C. mbila* is produced. This morph flies less than 10 metres, and only isolated pockets of disease develop. However, with the onset of crop maturity or under drought conditions - which cause the food plants of leafhoppers to dry out - the stronger-flying, short-bodied form of *C. mbila* is produced. Extensive migration into irrigated crops occurs, spreading disease over great distances and resulting in widespread crop failure (Rose, 1978).

Disease avoidance can be practiced by adjusting planting dates to avoid migrating leafhoppers landing on young plants (Rose, 1978). The vector can be controlled by applying systemic insecticides (eg. Carbofuran) to the planting furrow during maize planting (Mzira, 1984).

However, the development and use of streak-resistant cultivars is probably the most effective and economically viable means of preventing streak epidemics. Naturally occurring resistance to MSV has been found in maize on Reunion Island (where MSV has long been endemic). The resistance appeared to be simply inherited

and was rapidly fixed in breeding, being easily transferred to other maize lines (Soto & Buddenhagen, 1982). The incorporated resistance resulted in lowered disease incidence and reduced disease severity (Soto & Buddenhagen, 1982; Bock, 1982; Damsteegt, 1983). More recently, it has been shown that resistance to MSV in maize is quantitatively inherited, with relatively small numbers of genes involved (Kim *et al.*, 1989). Approximately five pairs of different alleles, both dominant and recessive, confer MSV resistance in maize. (G. J. M. A. Gorter, pers. comm. to E. Rybicki). Thus, it is envisaged that simple recurrent selection or modified back-cross breeding methods could be used to breed for MSV resistance in Africa (Kim *et al.*, 1989).

1.3.6. Disease Symptoms and Cytology.

Cereal geminiviruses all cause similar symptoms in the Gramineae, comprising leaf streaking and (where disease is severe) stunted shoot and root growth (Bock, 1982).

Maize streak disease first manifests as minute pale circular spots on the lowest exposed portion of the leaf. Only newly formed leaf tissue develops symptoms, leaves below the point of infection remaining healthy. The spots develop into streaks up to several millimetres in length along the leaf veins, primary veins being less affected than secondary or tertiary veins. The streaks often fuse laterally to give narrow, broken, chlorotic stripes, which may extend over the entire length of fully affected leaves (Storey, 1936). The chlorosis is caused by failure of chloroplasts to develop in the tissue surrounding the vascular bundles (McClean, 1947). This results in reduced photosynthesis and increased respiration, leading to a reduction in leaf length and plant height (McClean, 1947). Thus, maize plants infected at an early stage become severely stunted, producing undersized, misshapen cobs or giving no yield at all (Rose, 1978).

In the MSV group, depending on the isolate infecting the host plant, fully developed symptoms vary from severe chlorotic striping to mild streaks to lesions consisting of only a few sparse flecks on the leaves (McClean, 1947; Pinner *et al.*, 1988). Also, lesion colour varies from white to yellow with incident light, the white lesions being translucent in transmitted light (Pinner *et al.*, 1988). Some virus strains give red pigmentation on maize leaves and abnormal shoot and flower bunching in grasses (Pinner *et al.*, 1988; M. B. von Wechmar, pers. comm.). One virus strain may

react differently in different hosts. Thus, MSV type-A gives severe symptoms in maize but gives only a mild diffuse chlorotic mosaic on oats and low frequency mild streaking in certain other grasses (Damsteegt, 1983).

Leaf-curling (Greber, 1989) and development of pronounced vein enations have been reported in association with streak disease, but this has been attributed to leafhopper feeding damage (Pinner *et al.*, 1988).

Cytologically, virus-like particles have been shown to accumulate in the nucleus of infected cells. Whereas the Subgroup III geminiviruses are largely phloem-limited, Subgroup I geminiviruses can infect almost all leaf cell types of their hosts (Harrison, 1985). Thus, MSV infects cells in all leaf tissues, while CSMV infects all leaf cells except epidermal cells (Bock *et al.*, 1974; Hatta & Francki, 1979; Francki *et al.*, 1979). DSV occurs in linear aggregates or large crystalline arrays of geminate particles in the nuclei of companion and phloem parenchyma cells, (Dollet *et al.*, 1986). Fibrillar rings in cell nuclei and segregation of the nucleolus were observed in CSMV infected tissue (Francki *et al.*, 1979). More unusually for geminiviruses, aggregates of virus-like particles have been observed in the cytoplasm of intact cells; crystalline arrays of DSV (Dollet *et al.*, 1986), extensive sheets of PSMV (Greber, 1989), and random aggregates of CSMV and the *Bromus catharticus* and *Digitaria didactyla* viruses (Francki *et al.*, 1979; Greber, 1989). Interestingly, those viruses serologically related to MSV show typical small blocks of crystalline inclusions regardless of plant host (Pinner *et al.*, 1990). CSMV and the geminiviruses causing streak in sugarcane and *Panicum* exhibit non-crystalline inclusions (Pinner *et al.*, 1990). It is possible that coat protein determinants of serological relatedness are involved in virus aggregation in the plant cell.

1.3.7. Virus Particle Morphology and Virion Properties.

Like geminiviruses in other subgroups, the cereal geminiviruses have twinned capsids measuring approximately 18 nm x 30 nm, consisting of two incomplete icosahedra with T = 1 surface lattice with a total of 22 capsomers, as shown for CSMV (Hatta & Francki, 1979). The capsid is made up from a single species of polypeptide, of MW 28 - 34 x 10^3, and forms an efficient immunogen (Bock, 1982). No lipid or carbohydrate has been reported for this group. Subgroup I geminiviruses have an S_{20W} (sedimentation coefficient at 20°C in water) of 70 - 80 S

(Svedberg units) for particle pairs. Each virion contains one molecule of circular single-stranded DNA, 2670 - 2850 base pairs (Bock *et al.*, 1974; Francki *et al.*, 1979; MacDowell *et al.*, 1985; Dollet *et al.*, 1986; Greber, 1989; Ikegami *et al.*, 1989).

1.3.8. Genomic Organization.

This topic will be covered more comprehensively in Chapter 3, and so only a brief outline will be given here, along with a brief comparison between the genomic organization of the Subgroup I geminiviruses and that of the other two subgroups (see also reviews by Lazarowitz, 1987; Davies & Stanley, 1989) .

Subgroup I geminiviruses have four functional open reading frames (ORFs) which are bidirectionally transcribed (see Fig. 1.1).

On the virion strand, the V1 ORF encodes a non-structural protein (Mullineaux *et al.*, 1988) which is essential to virus spread within the plant, but non-essential to virus replication (Boulton *et al.*, 1989; Lazarowitz *et al.*, 1989). The V1 of product of MSV appears to control streak width in infected maize (Boulton *et al.*, 1990). Immediately downstream of V1 lies V2, which encodes the viral capsid protein (generally 28 - 30 kDa). A 1.2 kb transcript which includes both the V1 and V2 ORFs is the most abundant viral transcript. Thus it is likely that these ORFs are expressed from strong promoters. In addition, smaller quantities of a 0.9 kb transcript are also produced. While an authentic coat protein messenger, it has no consensus TATA or CAAT sequences immediately upstream (Morris-Krsinich *et al.*, 1985; Mullineaux *et al.*, 1988). As in all geminiviruses, the coat protein has an important role in vector transmission and appears to define the specificity of the transmitting insect (Briddon *et al.*, 1990a, 1990b).

Two ORFs, C1 and C2, exist on the complementary strand. Evidence exists that these may be spliced to give an approximately 41 kDa replication associated protein (Schalk *et al.*, 1989; Mullineaux *et al.*, 1990). The host cell appears to exert a negative control over the splicing process (Mullineaux *et al.*, 1990). It is not known for certain whether these two ORFs each encode a functional polypeptide in addition to the "spliced" protein. However, with the exception of MSV, the C2 ORF in Subgroup I geminivirus genomes has no AUG initiation codon (see Chapter 3).

The intergenic region lying between the V1 and C1 transcription initiation codons contains transcriptional control elements and a highly conserved stem-loop

structure speculated to function in virus replication (Fenoll *et al.*, 1990). In MSV, host range, specific infectivity, and severity of chlorosis are collectively determined by a single base change in this intergenic region (Boulton *et al.*, 1990).

The smaller intergenic region lying between the transcription termination codons of V2 and C2 contains a binding site on the virion strand for an endogenous DNA primer required in complementary strand synthesis. This primer is approximately 80 bases long, with a few ribonucleotides at the 5' terminus.(Mullineaux *et al*, 1985; see Chapter 3).

The Subgroup III geminiviruses have bipartite genomes, the two DNA molecules being termed DNAs A and B (see Fig. 1.1). A and B share little homology except for the strongly conserved intergenic region (the "common region") equivalent to the larger intergenic region of the Subgroup I viruses. DNA A contains four ORFs and DNA B contains two. Again, genes are bidirectionally transcribed (Lazarowitz, 1987; Davies & Stanley, 1989).

The V1 ORF of DNA A encodes the capsid protein. This protein is not required (in TGMV, at least) for systemic spread or symptom development. In addition to the coat protein gene, DNA A contains the genes for virus replication. Only the C1 product has been directly implicated in DNA replication. AC2 is required for accumulation of ssDNA and is needed for systemic infection of plants (Lehto *et al.*, 1990). The function of AC3 remains obscure.

While DNA A but not DNA B of African cassava mosaic virus (ACMV) can replicate autonomously in *Nicotiana plumbaginifolia* protoplasts (Townsend *et al.*, 1986), both viral components are required for systemic infection in intact plants. Thus DNA B must encode polypeptides involved in cell to cell (local) and/or vascular (long distance) spread (Davies & Stanley, 1989). It appears that determinants of symptom development, and hence symptom type are encoded on DNA B of TGMV (Arnim *et al.*, 1990).

The common (intergenic) regions of DNAs A and B are believed to contain one or more origins of DNA replication, and control elements for bidirectional transcription (Lehto *et al.*, 1990). Subgroup III geminivirus genomes do not appear to be associated with a primer molecule. It has been speculated that second strand synthesis may instead be completely host-directed (Davies & Stanley, 1989).

The Subgroup II geminiviruses, as epitomised by BCTV, are intermediate between the subgroups I and III in genomic organization. They have a single component genome, but the organization of this molecule is equivalent to that of DNA A of the Subgroup III geminiviruses (Stanley *et al.*, 1986; see Fig. 1.1).

1.3.9. Geminivirus Replication.

Cytological evidence indicates that geminivirus DNA replicates, and virions accumulate, in the nucleus of infected cells (Sequiera & Harrison, 1982; Stanley, 1983). The evidence from infection of *Phaseolus vulgaris* protoplasts with BGMV suggests a sequence of virus movements from the rough endoplasmic reticulum to the nuclear membrane, the highest virus concentration occurring finally inside the nucleus (Steffensen *et al.*, 1987).

While details of the precise mechanism of replication are unavailable, it is likely that the viral DNA forms double-stranded replicative intermediates, since covalently closed circular, double-stranded geminivirus DNA is found in infected plants (Townsend *et al.*, 1986; Dollet *et al.*, 1986; Lazarowitz, 1988; Ikegami *et al.*, 1989). It is also known that Subgroup I geminiviruses each have an endogenous DNA primer for complementary strand synthesis (1.3.7). It has been speculated that the highly conserved stem-loop structure in the large intergenic region acts as an origin of replication (Fenoll *et al.*, 1990).

In the Subgroup I geminiviruses, the extent of viral replication is affected negatively by host control over splicing of the C1-C2 intron (Mullineaux et al., 1990), which thereby controls the amount of viral replication-associated protein that is produced.

1.4. An Introduction to Methods Used in Typing Cereal Geminiviruses.

1.4.1. Biological Techniques.

Virus inoculation into indicator plants by insect vector is a commonly used technique in geminivirus typing (Pinner *et al.*, 1988; von Wechmar & Hughes, 1990). The method is useful for establishing whether two virus isolates share similar transmission characteristics, host range, and symptoms, and viruses may be

grouped accordingly. In addition, a panel of differential hosts (different genotypes of a single host species) can be constructed to differentiate between strains of one virus. This is of wide application in agriculture, both epidemiologically and in plant breeding.

However, these techniques are labour-intensive and demanding in both time and space when carried out on even a medium scale (Hill, 1984). In addition, to the taxonomist, such information as may be obtained is at best preliminary to the classification of viruses.

1.4.2. Serological Techniques.

Geminiviruses are suited to serological study, as the capsid protein is a good immunogen (Bock, 1982), eliciting antisera of useful quality. A variety of serological techniques can therefore be usefully applied to studying geminiviruses.

In any serological reaction, homologous serum titres are usually higher than heterologous ones. The number of two-fold dilution steps separating the homologous and heterologous titres is the Serological Differentiation Index (SDI), which may serve as a measure of the degree of serological relatedness between two viruses (van Regenmortel, 1986).

Antigenic relatedness between viruses may be assessed from antigen-antibody precipitation reactions, resulting from antigen-antibody diffusion towards one another in an agarose gel, termed immunodiffusion (reviewed by Matthews, 1970). Gel immunodiffusion was the first serological technique to be used in typing geminiviruses (Bock *et al.*, 1974) and is still of limited use in approximating geminivirus relationships (Pinner *et al.*, 1988). However, the method is insensitive and many artifacts and interpretational difficulties with precipitation patterns can arise if the test conditions are not rigorously controlled (Matthews, 1970).

In recent years immunodiffusion techniques have been largely superseded by enzyme-linked immunosorbent assay (ELISA; reviewed by Clark & Bar-Joseph, 1984) and immunoelectroblotting (Rybicki, 1984; Rybicki & von Wechmar, 1982), using either polyclonal antiserum or monoclonal antibodies. These methods are far more sensitive than precipitin assays. In addition, less material is required and a large number of samples can be handled at one time.

A number of variations on the ELISA technique exist. A good comparative account is given by Rybicki & von Wechmar (1985). The double-antibody sandwich method (DAS-ELISA) is extremely strain-specific and is useful in differentiating between closely related viruses. However, by the same token, the method has only limited use in detecting distantly related viruses. Indirect ELISA is less specific than DAS-ELISA and has been used to estimate the degree of serological cross-reactivity between viruses. An advantage over DAS-ELISA is that separate antibody-enzyme conjugates for each distinct virus or serotype tested are not required (Rybicki & von Wechmar, 1981; Jaegle & van Regenmortel, 1985; Clark & Barbara, 1987).

However, the use of polyclonal antisera in either ELISA or in immunodiffusion may give problems. Impure virus preparations elicit antisera that cross-react with plant proteins. Different bleedings give different antibody titres, as may animals from different stocks in different laboratories. Thus, the SDIs obtained from these different sources may vary considerably, and it may be difficult to compare reliably the results of serological tests made in different laboratories.

Monoclonal antibodies (MAbs), which each recognize a single epitope, are an improvement over polyclonal antisera in that they provide an opportunity to select antibodies of high affinity. This in turn permits increased reaction specificity, giving reliable results even when using impure virus preparations (van Regenmortel, 1986). A further advantage is that the use of MAbs allows the standardization of serological tests in different laboratories, since antibody-producing hybridomas can be transferred between laboratories (van Regenmortel, 1986).

The disadvantages of MAbs are, however, that they are laborious and expensive to produce, and that a comprehensive panel must be used in serological assays. Otherwise, for example, it could be concluded erroneously that two viruses are closely related when a MAb reacts with what is perhaps the only shared epitope between the two viruses. Likewise, it may be concluded that two closely related viruses are serologically different where a MAb against the only epitope that differs between the two viruses is used to assess their serological relatedness (van Regenmortel, 1986). A further disadvantage of MAbs is that not all MAbs are suited to all kinds of serological assay (van Regenmortel, 1986).

A disadvantage of ELISA techniques, whether monoclonal or polyclonal antisera are used, is that viral antigens are identified by only one property - their

serological reactivity. The only measure of this is an indirect colorimetric assay. Thus, particularly where polyclonal antisera are made use of, apparent serological cross-reactions due to undetected contamination of the virus inoculum used to inject animals can lead to false positives (Rybicki, 1984).

Immunoelectroblotting (IEB), also termed Western blotting, allows the antigen being studied to be sized in addition to discriminating between the cross-reactivity due to the antigen and that due to other contaminants (Rybicki, 1984; Rybicki & von Wechmar, 1982).

To conclude, the main advantage of serological typing of geminiviruses is that the range of techniques available make it a versatile approach to typing. However, since less than 5% of the genome may confer the antigenicity of a virus (Matthews, 1970; Hull, 1986), viruses are being typed on the basis of only a portion of the entire information contained in the viral genome. More importantly, serological data comprise pairwise comparisons and not absolute data. Thus, for example, a comparison between A and B and another between C and D will not yield any information on the relationship between A and C.

However, given these drawbacks, a wide variety of African and Australian cereal geminiviruses have been serologically typed (Dekker *et al.*, 1988; Pinner & Markham, 1990a, 1990b), using indirect ELISA with polyclonal antisera to a number of the viruses being typed and MAbs to a reference isolate of MSV. Pinner & Markham (1990b) compared a number of ELISA methods and the use of sap extracts over purified virus extracts as antigen source. It was shown that the SDI obtained varied depending on the virus source and the specific test used. The most accurate serotyping was obtained with indirect ELISA, and plant sap extracts were the most reliable antigen source (Pinner & Markham, 1990b).

1.4.3. Nucleic Acid Techniques.

1.4.3.1. DNA Cross-hybridization Between Virus Isolates.

This technique is based on the ability of complementary single-stranded nucleic acid sequences to hybridize to one another. Hybridization also occurs between sequences which are not fully complementary and, therefore, the technique

can be used to gain an approximate assessment of sequence similarity between two nucleic acids.

When double-stranded nucleic acids are heated, the bonds linking the two strands melt and the strands dissociate (the nucleic acid denatures). The temperature at which 50% of the sequences are denatured is termed the melting temperature (Tm). If the complementary strands are incubated together at temperatures below the Tm and under appropriate conditions, the strands reanneal (see review by Hull, 1986). Both the rate of hybridization and, more importantly in this context, the stability of the resultant nucleic acid hybrid are affected by the factors that determine the Tm of a nucleic acid: salt concentration, pH, presence of helix destabilizing agents (eg. formamide), and the size and base composition of the nucleic acids in question (Meinkoth & Wahl, 1983). A heterologous nucleic acid duplex has a lower Tm than a homologous one and so is less stable.

Whereas hybridization parameters were originally determined for nucleic acids in solution, the general principles apply to mixed phase reactions (Meinkoth & Wahl, 1983), where one of the nucleic acids is immobilized on a solid matrix and the other is a labelled probe strand added in solution (Denhardt, 1966), making the technique more amenable to common laboratory use.

For probe DNAs of greater than 50 nucleotides, an empirical formula for Tm has been derived (Meinkoth & Wahl, 1983):

$$Tm = 81.5^{\circ}C + 16.6 \log M + 0.41 (\%G + \%C) - 550/n$$

where:

M is the ionic strength (moles/litre) of the hybridization solution.

n is the length of the shortest chain in the duplex.

The nucleic acid of most plant viruses has a G+C content in the range of 40 - 50%. For DNA of 40% G+C in 0.3M NaCl the Tm is 88.2°C and for 50% G+C it is 93.3°C (Hull, 1986).

Because both temperature and salt concentration affect the melting of double-stranded nucleic acid, reaction conditions (stringency) can be selected for distinguishing between nucleic acids of different degrees of complementarity. Low stringency conditions will allow hybridization between nucleic acids showing less complementarity.

Commonly, hybridization is carried out under relatively low stringency conditions (65°C to 72°C) to allow a high rate of hybridization. A series of post-hybridization washes of increasing stringency (higher temperature and/or lower ionic strength, using sodium dodecyl sulphate, SDS, to assist in the process) are then performed. The higher stringency conditions no longer permit the probe and target DNAs to form a stable hybrid, where sequence mismatch occurs between them. The greater the mismatch, the lower the wash-stringency needed to remove the mismatched probe.

In an early application of DNA hybridization techniques, Boulton & Markham (1986), using the dot-blot technique, could distinguish between grass geminivirus isolates from Africa and Mauritius.

1.4.3.2. Restriction Endonuclease Site Mapping of Virus Genomic DNA.

Restriction endonucleases cleave double-stranded DNA molecules at specific nucleotide sequences (recognition sites) that are usually palindromic and four to six nucleotides in length (Nei & Li, 1979). Different enzymes will recognize different sites. Thus, by a number of available methods (see Maniatis *et al.*, 1982), it is possible to construct for any DNA sequence a "map" specifying the relative positions of the cleavage sites of a number of different restriction endonucleases.

The degree of sequence relatedness between two or more DNAs can be estimated from restriction map comparisons, as follows.

Once restriction sites are determined for two or more different but related DNA sequences, for example viral DNAs, the proportion of sites shared by them can be computed, and is given by:

$$S = \frac{2\,n_{xy}}{dn_x + dn_y}$$ (Aoki *et al.*, 1981)

where:

S = the average proportion of shared sites between two isolates.

n_x = the total number of restriction sites observed in isolate x.

n_y = the total number of restriction sites observed in isolate y.

n_{xy} = the number of sites shared by both isolates.

The proportion of sites shared between two DNA sequences is expected to decline as the sequences become increasingly divergent (Nei & Li, 1979; Aoki *et al.*, 1981). Thus, nucleotide sequence diversity between genomes may be estimated from restriction map data, as follows:

$$\pi = \frac{(-\ln S)}{r} \qquad \text{(Nei \& Li, 1979)}$$

where:

π = nucleotide diversity between 2 DNA sequences (estimated as the average number of nucleotide differences per site between the two sequences).

S = the average proportion of shared sites between two isolates.

r = the number of base pairs in the restriction site.

The theory behind this equation depends on the assumption that all nucleotides are distributed at random over a DNA sequence with a given base composition. Although this assumption is not always satisfied, only an extremely non-random base distribution would affect the estimate (Nei & Li, 1979).

Such estimates of nucleotide diversity may be used to compute the evolutionary distance between related genomes (Gibbs & Fenner, 1984), defined as the number of base substitutions per homologous site that have occurred since the divergence of the two genomes (Aoki *et al.*, 1981). Such substitutions are assumed to be equally probable at all nucleotide sites, and the substitution rate is assumed to be constant and identical for all directions of base change.

1.4.3.3. Nucleotide Sequence Comparisons.

Sambrook *et al.* (1989) have reviewed the development of DNA sequencing methods and give practical details for the use of current methods. The two rapid sequencing techniques commonly used at present are the enzymatic method (dideoxy-mediated chain termination) of Sanger *et al.* (1977) and the chemical degradation method of Maxam & Gilbert (1977). Both methods generate populations of radiolabelled oligonucleotides that begin from a fixed point and terminate randomly at a fixed residue. These oligonucleotide populations are

electrophoretically resolved such that single nucleotide differences in DNA lengths can be discriminated. The order of nucleotides along the DNA is read from an autoradiographic image of the gel, thereby yielding the DNA sequence. Thereafter, either manually or by using computer programmes, potential open reading frames (ORFs) in the DNA sequence can be identified and the amino acid sequence encoded therein may be deduced.

Molecular sequences provide precisely comparable characters, observed at or near gene level. This data can be used for direct comparisons between organisms, such as viruses, and differences between them can be precisely quantified. Moreover, such data are absolute. Thus, if we have sequence data for A and B and have deduced how closely they are related, then sequence data obtained for C will establish its relationship with both A and B - and with any other entity subsequently sequenced. This property is also true of restriction map data, but not of serological data (see section 1.4.2).

A further use of sequence data is in the inference of the phylogenetic relationship between organisms, in this case Subgroup I geminiviruses, within a taxonomic group (see below, section 1.5).

1.5. Methods of Phylogenetic Analysis.

From observable genetic differences between individuals (such as DNA sequence data), inferences on their evolutionary relationships may be made. The central model of molecular evolution used in phylogenetic inference is one of random evolutionary changes, occurring at a stochastically constant rate (the "molecular clock"), such that the rate of change is the same in all lineages (Felsenstein, 1988). However, analysis can often proceed without the assumption of a molecular clock.

A number of methods for computing phylogenetic relationships (which are commonly depicted as trees) are currently in use (Felsenstein, 1988). The methods usually assume simple base substitutions to be the evolutionary mechanism. This may not accurately reflect the true details of genetic change, and so any tree produced is of necessity only an approximation of the true phylogeny (Bishop *et al.*, 1987).

The three major families of methods currently used to infer phylogenies are (1) the Maximum Likelihood methods, (2) the Parsimony methods, and (3) the Distance Matrix methods (Felsenstein, 1988).

(1). Maximum Likelihood Methods.

With a model (M) and data (D), the likelihood of a tree (T) is the probability (P) of observing the data, given the tree and model, ie. P(D;T,M). The probability of all possible sets of data = 1. However, when D is constant and T is varied, the different values of P(D;T,M) need not add up to one, and are rather called likelihoods. The Maximum Likelihood method simply chooses that tree which maximizes the probability that the observed data could have occurred, given the model, M (Bishop *et al.*, 1987; Felsenstein, 1988). Although their statistical properties are sound, these methods are computationally complex and slow to perform (Felsenstein, 1988), even using computer programs. Consequently, most biologists currently prefer to use parsimony or distance matrix methods (Felsenstein, 1988; Feng & Doolittle, 1987).

(2). Parsimony Methods.

These methods find the evolutionary tree that invokes the fewest base substitutions necessary to explain the evolution of the observed data on that phylogeny. Any base may be substituted for any other, thereby removing all information on the root of the tree. The resulting unrooted tree represents an equivalence class of rooted trees, all of which have equal numbers of substitutions (Felsenstein, 1988; Bishop *et al.*, 1987).

In practice, species are added one at a time to the tree and repeated local rearrangements of the tree are done, accepting those which improve the tree. The statistical soundness of these methods appears variable, depending on the data and the precise analysis used (Felsenstein, 1988). Simulation studies suggest that parsimony methods work best when the sequences in question are little diverged from one another (Felsenstein, 1988).

(3). Distance Matrix Methods.

These fit a tree to a matrix of pairwise distances (eg. percentage DNA sequence difference) between the species. The phylogeny predicts the distance for each pair as the sum of branch lengths in the path from one species to another through the tree. The preferred phylogeny is that which minimizes the discrepancy between the observed and expected distances, as evaluated by a measure of the goodness of fit. A molecular clock may or may not be assumed (Felsenstein, 1988).

In this thesis, use is made of two Distance Matrix methods, the Fitch & Margoliash (1967) least squares method, and Saitou & Nei's (1987) "neighbour joining" method.

Fitch & Margoliash (1967) introduced the first distance matrix method. The measure of lack of fit between observed (D_{ij}) and expected (d_{ij}) distances computed from the tree was given by a weighted least squares measure:

$$\sum_i \sum_j W_{ij} (D_{ij} - d_{ij})^2$$

where the weights W_{ij} were $1/D_{ij}^2$.

In the Phylogeny Inference package (PHYLIP, J. Felsenstein, Department of Genetics, University of Washington, Seattle), the programme FITCH uses the Fitch & Margoliash (1967) algorithm to construct trees. In addition, a global rearrangement option may be invoked, giving a better chance of finding the best available tree. With this option, the position of every species is reconsidered after the last species has been added to the tree, by removing each group and re-adding it in all possible ways.

Saitou & Nei's (1987) "neighbour joining" algorithm estimates an additive tree from a distance matrix, under the principle of minimum evolution. The unique final tree has a parsimonious topology, and branch lengths are given. Starting with a starlike tree, the algorithm finds pairs of neighbouring species that minimize the total branch length at each subsequent stage of clustering of species. The algorithm yields an exact result when there is a tree that perfectly fits the data (Felsenstein, 1988).

The field of phylogeny inference is still in an incomplete state (Felsenstein, 1988). Even the assessment of the reliability of estimated phylogenies is problematic.

Only fragments of methods exist, each with many properties unknown (Felsenstein, 1988). Felsenstein (1988) reviews the statistical tests and resampling methods currently in use. In the studies detailed in this thesis, phylogenies were checked by using different types of data (restriction site data; DNA and amino acid sequence data), and by using two different methods to analyse the data. In addition, for any one analysis and set of data, reordered data sets and subsets of data were analysed separately.

CHAPTER 2.

CHARACTERIZATION OF THE SUGARCANE STREAK AGENT AS A DISTINCT GEMINIVIRUS.

CONTENTS

CHAPTER 2.
CHARACTERIZATION OF THE SUGARCANE STREAK AGENT AS A DISTINCT GEMINIVIRUS.

SUMMARY.

The relationship between maize streak virus and the geminivirus causing streak disease in sugarcane was investigated. The sugarcane streak virus did not cross-protect maize against superinfection by maize streak virus. The DNA of sugarcane streak virus did not cross-hybridize significantly with that of maize streak virus and *vice versa*. Native replicative form (RF) viral DNA (genome size approximately 2.8 kb) and cloned RF DNA were subjected to restriction endonuclease mapping and limited nucleotide sequencing. DNA sequence divergence between the sugarcane streak virus and other cereal geminiviruses was estimated from aligned restriction maps. The results of the studies made show sugarcane streak virus to be as unrelated to maize streak virus and digitaria streak virus as these are different from each other. The virus is only distantly related to wheat dwarf virus and chloris striate mosaic virus. Based on these results, it is proposed that the agent causing sugarcane streak is a distinct geminivirus.

2.1. INTRODUCTION

Streak disease of sugarcane (*Saccharum oficinarum*) was first described on cultivar Uba in Natal, South Africa, in 1914 (Storey, 1925), but was already widespread at that time. The disease had considerable impact on sugar production in Natal in the 1920's and 1930's. By 1935, more than 60% of Natal's sugarcane (virtually all Uba) was infected, with an estimated total crop loss of 7% in the 1934/1935 season alone (McClean & Halse, 1936). After 1945, by which time Uba had been replaced almost entirely by resistant varieties, sugarcane streak became uncommon and of minor economic importance. It is now seen only in susceptible clones in the breeding programme of the SASAES (South African Sugar Association Experimental Station) at Mt. Edgecombe, Natal, where new clones are tested for streak resistance by interplanting with infected Uba (Bock & Bailey, 1989).

The streaking symptoms caused by this virus are finer and far more sparse than those caused by typical maize strains of MSV. The sugarcane streak virus is

transmissable to maize by *C. mbila*, inducing persistent but mild streak symptoms (Storey & McClean, 1930; McClean, 1947; Bock *et al.*, 1974). However, depending on the sugarcane cultivar being infected, the maize strain of MSV causes either no infection, transient infection, or permanent infection with severe symptoms (Storey & McClean, 1930; McClean, 1947; Bock *et al.*, 1974; Pinner et al., 1988).

Wild grass hosts of the sugarcane streak virus in Africa have not yet been identified. In Mauritius, however, natural infection of the local strain of sugarcane streak occurs in *Cenchrus echinatus*, which probably constitutes the natural reservoir of the virus on the island (Bock & Bailey, 1989).

There is evidence that the virus varies geographically. The Mauritian isolate differs serologically from the Natal isolate (Dekker *et al.*, 1988), and could not be transmitted experimentally to cultivar Uba (Ricaud & Felix, 1978). The relationship of the Natal and Mauritian sugarcane streak viruses to the streak virus in Egyptian sugarcane, transmitted by *C. bipunctella-zeae* (Ammar *et al.*, 1982), is unclear. Streak disease was also reported in the 1920's in India and Burma (refs. cited by Storey & McClean, 1930), but the relationship of the causal virus(es) to other sugarcane streak viruses is likewise unclear.

McClean (1947) considered the sugarcane streak agent to be distinct from maize streak virus on the basis of biological studies, including cross-protection, host range, and symptom differences. However, the sugarcane streak virus is currently accepted as a strain of MSV on the basis of results obtained from serological studies, which show the two viruses to be related, albeit distantly (Bock *et al.*, 1974; Pinner *et al.*, 1988; Dekker *et al.*, 1988; Pinner & Markham, 1990b).

The exact relationship of the sugarcane streak virus (here termed sugarcane streak virus, SSV) to MSV is still not altogether clear. In particular, a distant serological relationship to MSV need not imply that the virus in question is a strain of MSV. Digitaria streak virus (DSV), for example, was originally identified as a strain of MSV on the basis that it is serologically related to MSV (Dollet *et al.*, 1986). However, as for SSV, the relationship is a distant one (Dekker et al., 1988), and following DNA sequence comparisons (Donson *et al.*, 1988), DSV is now accepted as a distinct geminivirus (Julia & Dollet, 1989; Accotto *et al.*, 1989; G. P. Martelli, pers. comm. to E. P. Rybicki).

Restriction enzyme site mapping of MSV has hitherto been used successfully in this laboratory (Clarke, 1987; Clarke *et al.*, 1989; Kirby *et al.*, 1989) to estimate

sequence differences between MSV isolates from southern Africa. In the present study, restriction mapping and DNA cross-hybridization were used to investigate genomic differences between MSV, SSV, and DSV, as well as to test the cross-protection of maize by SSV against MSV superinfection. Results obtained indicate that the classification of the causal agent of streak disease in sugarcane should be reconsidered and it is suggested that it is a distinct virus, SSV.

2.2. MATERIALS AND METHODS

2.2.1. Virus Isolates.

The sources and geographical origins of streak-diseased sugarcane, *Setaria*, *Digitaria*, maize, and wheat isolates are shown in Table 2.1. DSV infected *Digitaria setigera* was originally obtained from G. Boccardo, Torino with the permission of the Directorate of Plant and Seed Control, Department of Agriculture, permit no. 14/14/A/16/16. Streak infected sugarcane cv. Uba and *D. setigera* were propagated vegetatively (Appendix A.1).

Table 2.1. Origins of streak-infected plants used in the studies reported in Chapter 2.

Isolate	Source plant	Geographical origin	Obtained from
MSV-CT	maize	Potchefstroom, OFS	M B von Wechmar[a]
MSV-G	maize	George, E. Cape	M B von Wechmar[a]
MSV-Koe	maize	Koedoespoort, N. Tvl	M B von Wechmar[a]
MSV-Kom	maize	Komatipoort, E. Tvl	M B von Wechmar[a]
MSV-PE	maize	Port Elizabeth, E. Cape	M B von Wechmar[a]
MSV-RSE	maize	Riviersonderend, Cape	J Laubscher[a]
MSV-SW	maize	Otjiwarongo, Namibia	H Hoffsman[a]
SetSV	*Setaria* sp.	Mt. Edgecombe, Natal	K Harborne[b]
MSV-W	wheat	Wilderness, E. Cape	M B von Wechmar[a]
SSV	Sugarcane	Mt. Edgecombe, Natal	K Harborne[b]
DSV	*Digitaria setigera*	Vanuatu, New Hebrides	G Boccardo[c]

Key: Tvl = Transvaal. OFS = Orange Free State (the OFS, Transvaal, Cape, and Natal are provinces of South Africa). a = University of Cape Town. b = South African Sugar Association Experimental Station, Mt. Edgecombe. c = IFA del CNR, Torino, Italy.

2.2.2. Isolation of replicative form (RF) DNA.

Total DNA was extracted from diseased sugarcane, maize, *D. setigera*, and the *Setaria* sp. as detailed in Appendix A.3. Where necessary, viral double-stranded RF-DNA was separated from the bulk of the host chromosomal DNA by CsCl-EthBr density centifugation (Maniatis *et al.*, 1982).

2.2.3. Molecular cloning of viral RF-DNA.

In preliminary experiments, SSV RF-DNA was digested with a number of restriction enzymes (A.5). *Pst* 1 cleavage produced full-length linear RF-DNA and this was ligated into the *Pst* 1 site of the plasmid vector Bluescript SK (as in A.8). The ligation mix was used to transform *E. coli* LKIII (A.9, A.10). After restriction endonuclease cleavage with restriction enzymes shown previously to cut SSV RF-DNA, and restriction enzyme digestion to size the recombinant plasmids' inserts, four recombinant plasmids having correctly sized inserts and the same restriction site profiles as uncloned SSV RF-DNA were selected. Of these, one plasmid, pSS100, was labelled with digoxygenin (A.13.c) and probed back to Southern-blotted (A.13.b) total DNA extracts of both infected and uninfected sugarcane to confirm the viral origin of the insert (A.13.d).

2.2.4. DNA cross-hybridization between viral genomes from different host sources.

The viral inserts of CsCl-purified cloned SSV (A.11.d) and cloned MSV-SW (Chapter 4; Clarke *et al.*, 1989) were excised from their vectors using *Pst* I and *Bgl* II, respectively, and gel-purified (A.7). These insert DNAs were labelled for use as probes by random primed incorporation of digoxygenin-labeled dUTP, as detailed in Appendix A.13.c.

Total plant DNA extracts were electrophoresed on duplicate agarose gels (A.6) and blotted onto Hybond-N (as in Appendix A.13.b), but with the DNA being cross-linked onto the membrane by UV irradiation (450 mW cm^{-2} at 254nm). One blot was probed with the MSV-SW probe and the other with the SSV probe, as described in Appendix A.13.d, using 50 ng probe per ml of hybridization fluid.

2.2.5. Coinfection of SSV and MSV in Maize.

SSV was transmitted from infected sugarcane cv. Uba to maize (sweetcorn cv. Melody) by *C. mbila*, as described in Appendix A.1. From Melody sweetcorn, SSV was transmitted to two maize cultivars, Kalahari Early Pearl ("Witplat") and Melody sweetcorn. After SSV symptoms were well established, the severe MSV isolate, MSV-Koe (Babaya, 1990; von Wechmar & Hughes, 1990; see Table 2.1) was transmitted in the same manner to the SSV-infected maize plants, using fresh leafhoppers.

Two weeks after superinfection with MSV-Koe, some plants were used for DNA extractions, while others were left to grow for longer in order to observe symptom development. Total DNA extracts were made (A.3, A.4) from SSV-infected sugarcane and sweetcorn, from maize infected with MSV-Koe, from the SSV and MSV coinfected maize cultivars, and from uninfected maize and sugarcane. Samples of all of these extracts were electrophoresed through duplicate agarose gels (A.6) and blotted onto Hybond-N^+ (A.13.b). One blot was probed with the MSV-SW probe and the other with the SSV probe, using 40 ng probe per ml of hybridization fluid, as described above (2.2.5).

2.2.6. Restriction enzyme mapping of SSV.

Using the methods given in Appendices A.5 and A.6, and following the strategy outlined in A.14, a restriction endonuclease cleavage map for SSV was constructed. The following restriction enzymes were used: *Apa* I, *Bam* HI, *Bgl* I, *Bgl* II, *Cla* I, *Eco* RI, *Hind* III, *Kpn* I, *Pst* I, *Pvu* II, *Sac* I, *Sal* I, and *Xho* I. Both RF-DNA and cloned DNA of SSV were used to map the genome.

The first 300bp of pSS100 was sequenced in the forward direction (Appendix B.2). The sequence was aligned with already published MSV sequences using GENEPRO version 4.0 software (Riverside Scientific, Seattle, Wa., USA). This comparison was used to orientate the SSV restriction map with known MSV maps and maps of DSV, CSMV, and WDV (see Figure 2.1).

2.2.7. Estimation of DNA sequence divergence between geminiviruses and construction of a phylogenetic tree.

Using GENEPRO version 4.2 software, restriction maps were generated from published sequence data for DSV (Donson *et al.*, 1987), CSMV (Andersen *et al.*, 1988), WDV (MacDowell *et al.*, 1985), MSV-K (Kenyan isolate, Howell, 1984; 1985), MSV-N (Nigerian isolate, Mullineaux *et al.*, 1984), and MSV-S (South African isolate, Lazarowitz, 1988).

The number of restriction sites shared between isolates was obtained by making pairwise comparisons of the aligned restriction maps of SSV, DSV, CSMV, WDV, and the MSV isolates. Sites were regarded as common or shared if their mapped positions coincided within an error of 2% of total genome length.

DNA sequence divergence and evolutionary distances between SSV and other Subgroup I geminiviruses were then estimated, using the mathematical models of Aoki *et al.* (1981) and Nei & Li (1979) (see Chapter 1, section 1.4.3.2).

The average proportion of shared sites between isolates (S) was calculated as:

$S = 2\, n_{xy} / dn_x + dn_y$ (Aoki et al., 1981).

where:

n_x and n_y = total number of restriction sites observed in isolates x and y, respectively.

n_{xy} = number of sites shared by both isolates.

From the S values so calculated, a table of estimated distance (p) values was drawn up, calculated using the equation:

$\pi = (-\ln S) / r$ (Nei & Li, 1979).

where:

π = nucleotide diversity between the 2 DNA sequences (estimated as the average no. of nucleotide differences per site between the 2 sequences).

r = the number of bases per restriction site.

The matrix of π values was inserted into the Fitch and Margoliash (1967) algorithm, using the FITCH programme (global rearrangement (G) and jumble (J) options) from the Phylogeny Inference Package (PHYLIP) Version 3.1. (J. Felsenstein, Department of Genetics, University of Washington, Seattle, USA).

FITCH gives an unrooted phylogenetic tree. A rooted tree was derived by outgroup analysis, using SSV, WDV, and CSMV as outgroups to find the root position of the tree. As a check, the π values were also used in the programme NJTREE, which uses the neighbour joining method of Saitou & Nei (1987), as modified by Studier & Keppler (1988).

2.2.8. Serology.

2.2.8.1. Sources of Antisera.

Antiserum to sucrose-gradient-purified MSV-CT (see Table 2.1) was raised in rabbits. IgG was purified from the antiserum and conjugated with alkaline phosphatase (Clarke *et al*, 1989). Antiserum to SSV was a gift from P. Jones, Rothamsted, to K. Harborne, SASAES.

2.2.8.2. Virus Extraction.

MSV-George (Table 2.1) and SSV virus particles were isolated from infected maize and sugarcane, respectively, as described in Appendix A.2.

2.2.8.3. Immunosorbent electron microscopy.

Immunosorbent electron microscopy (ISEM), using antiserum to SSV to trap and decorate SSV particles, was performed as described (A.15). Antiserum was diluted 1:1000 in 100 µM Na acetate buffer pH 7.5 to trap, and 1:10 to decorate, the virus particles.

2.2.8.4. Immunoelectroblotting.

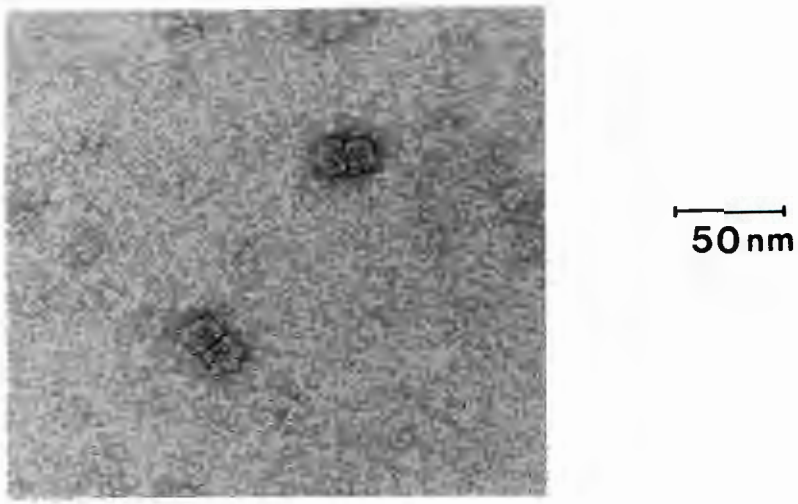
SDS-PAGE was performed as described in Appendix A.16. Gels were either stained with PAGE blue 83 (A.16) or electroblotted onto nitrocellulose membranes ("Western" blots, A.17). As described in Appendix A.18, the nitrocellulose blots were probed with host-absorbed antiserum to MSV-CT and the colour reaction developed.

2.3. RESULTS.

2.3.1. Serology.

ISEM showed the virus particles in streak-infected sugarcane to be typically geminate, with approximate particle dimensions of 20 nm x 30 nm (Fig. 2.1).

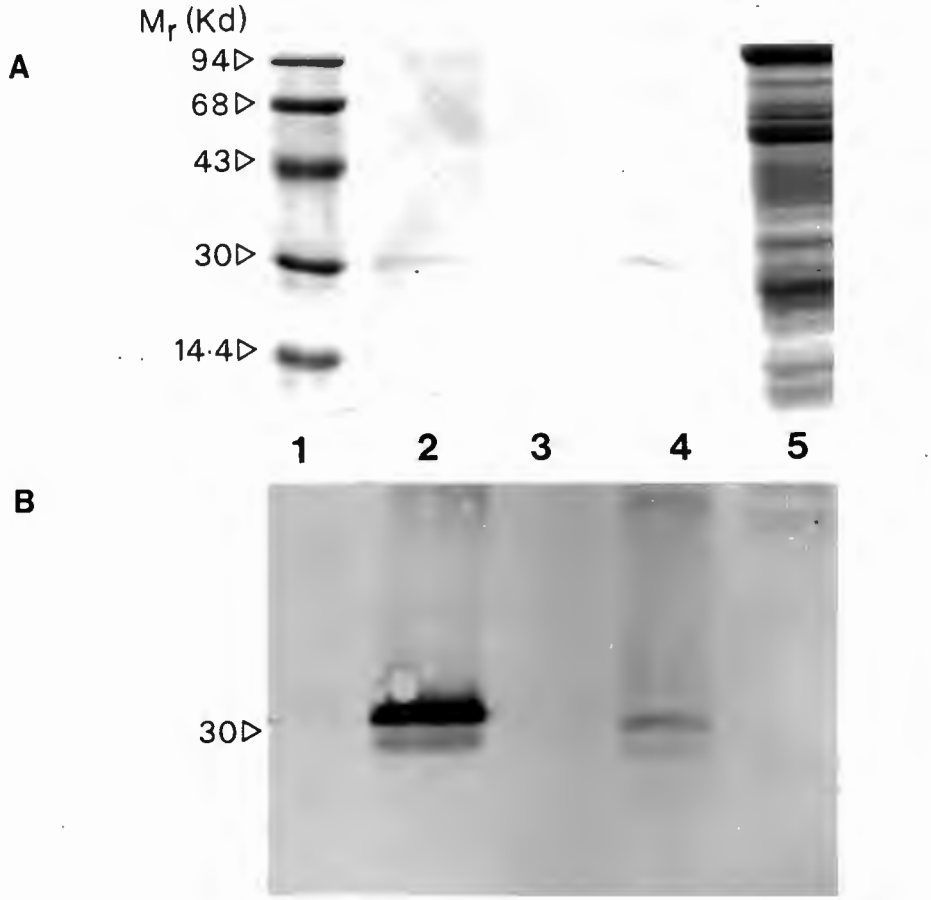
Figure 2.1. Electron micrograph of sugarcane streak virus particles.



Virus particles stained with 1% uranyl acetate. Size bar represents 50 nm.

Although the anti-MSV antiserum used was raised against the MSV-CT isolate, an MSV-G virus preparation was used in these experiments, as prior testing had shown that MSV-G and MSV-CT behaved identically towards the anti-MSV antiserum (E. P. Rybicki, pers. comm.). Western blots showed that the capsid protein of SSV reacted poorly with antiserum to MSV, the colour reaction developing far more slowly and less intensely than that shown by an approximately equal amount of MSV capsid protein (Fig. 2.2). Both the stained gels and the Western blots showed that the MSV-G and SSV capsid proteins both had an apparent M_r of 30 kd (Fig. 2.2). The SSV antiserum was of too low a titre - at a dilution of 1 in 50 - to reveal cross-reaction in Western blots, but reacted homologously (not shown).

Figure 2.2. SDS-PAGE and Western blot comparisons of SSV and MSV.

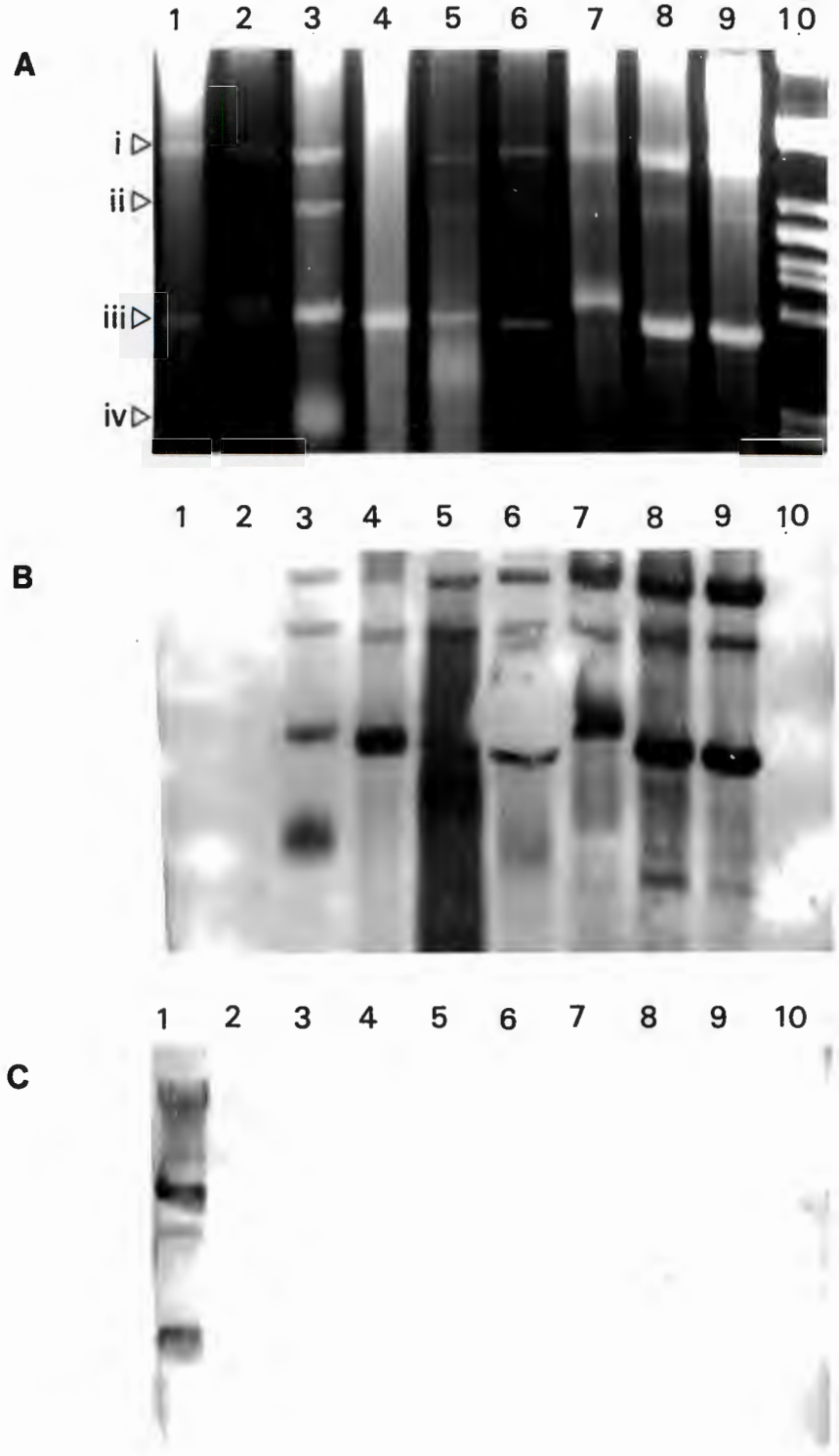


Legend: (a). PAGE-Blue 83 stained gel. (b). Electroblot of (a) probed with antiserum to MSV-CT. lane 1; protein molecular weight markers (Pharmacia LMW). Lanes 2-5; MSV-G, uninfected maize sap, SSV, uninfected sugarcane sap, respectively. All virus and sap preparations were purified, except the uninfected sugarcane sap.

2.3.2. DNA cross-hybridization.

Cloned MSV-SW hybridized strongly to all MSV RF-DNA isolates tested, albeit to a slightly lesser extent to the *Setaria* isolate (Fig. 2.3b). However, MSV-SW hybridized only marginally to DSV and SSV RF-DNA under the stringency conditions used. Only barely visible staining was detectable after prolonged reaction of the blot with substrate (Fig. 2.3b). Under the same hybridization and washing conditions, cloned SSV hybridized strongly to SSV RF-DNA but not at all to any MSV isolate or to DSV, even after prolonged development of the blot (Fig. 2.3c). The same results were obtained using ^{32}P-dCTP labeled probes (not shown).

Figure 2.3. DNA cross-hybridization between SSV, DSV, and MSV isolates.

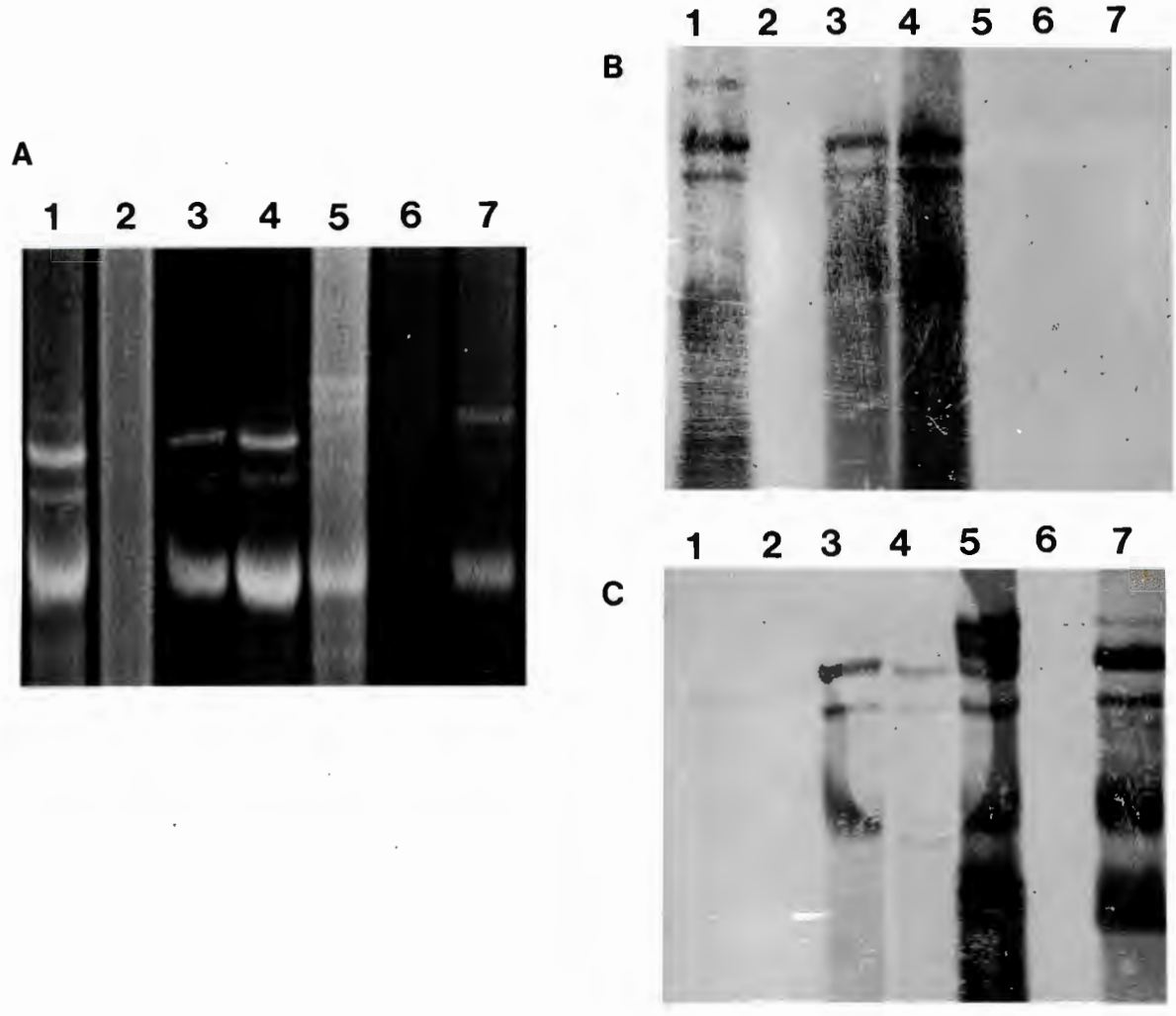


(a) 0,8% agarose gel stained with EthBr. (b) Southern blot of (a), probed with cloned MSV-SW DNA. (c) Southern blot of (a), probed with cloned SSV DNA. Lane 1; SSV. Lane 2; DSV. Lane 3; SetSV. Lane 4; MSV-W. Lanes 5-9; maize isolates of MSV as follows; MSV-SW, MSV-RSE, MSV-Kom, MSV-G, MSV-CT, respectively. Lane 10; *Pst* 1-restricted Lambda DNA molecular weight markers. i = open circular dsDNA; ii = linear dsDNA; iii = ccc dsDNA; iv = ssDNA, as described in Hughes *et al.* (1991).

2.3.3. Coinfection of SSV and MSV in maize.

SSV could be transmitted to maize by *C. mbila*, but less efficiently than MSV could be transmitted under the same conditions (M. B. von Wechmar, pers. comm.). The acquisition period for SSV was longer (7 days, using first instar nymphs) than for MSV isolates (1 to 2 days), and symptoms only became apparent after 10 to 12 days at 24°C, compared with 6 days for MSV isolates. SSV produced similar symptoms in both maize and sugarcane, that is, fine, narrow streaks which did not coalesce.

Figure 2.4. Molecular detection of MSV and SSV in SSV-infected maize superinfected with MSV-Koe.



Legend: (a). EthBr-stained 0.8% agarose. (b). Southern blot of (a), probed with cloned MSV-SW DNA. (c). Southern blot of (a), probed with cloned SSV DNA. Lane 1; MSV-Koe-infected maize. Lane 2; uninfected maize. Lane 3; MSV-Koe / SSV-infected maize, cv. Melody. Lane 4; MSV-Koe / SSV-infected maize, cv. Kalahari Early Pearl. Lane 5; SSV-infected maize. Lane 6; uninfected sugarcane. Lane 7; SSV-infected sugarcane.

MSV-Koe produces severe streaking in maize, generally with coalescence of individual lesions into broad stripes along the length of the leaf (Babaya, 1989; von Wechmar et al., 1990). MSV-Koe superinfection of maize plants already infected with SSV resulted in the development of the characteristic streak symptoms of normal MSV infections after 6 to 7 days. These streaks increased rapidly in number and size, masking the finer streaks of the SSV infection with each new emergent leaf until eventually the infection resembled a typical MSV infection.

In Southern blots of SSV and MSV coinfected maize, using equal amounts of probe DNA, MSV-SW hybridized much more strongly to the coinfected maize samples than did SSV (Fig. 2.4). In addition, while MSV staining intensity was approximately the same for both the Witplat and Melody coinfections, SSV gave a more intense reaction with the Melody sample than with the Witplat sample, and in both samples less staining was observed than for the infection of SSV alone in Melody sweetcorn (Fig. 2.4).

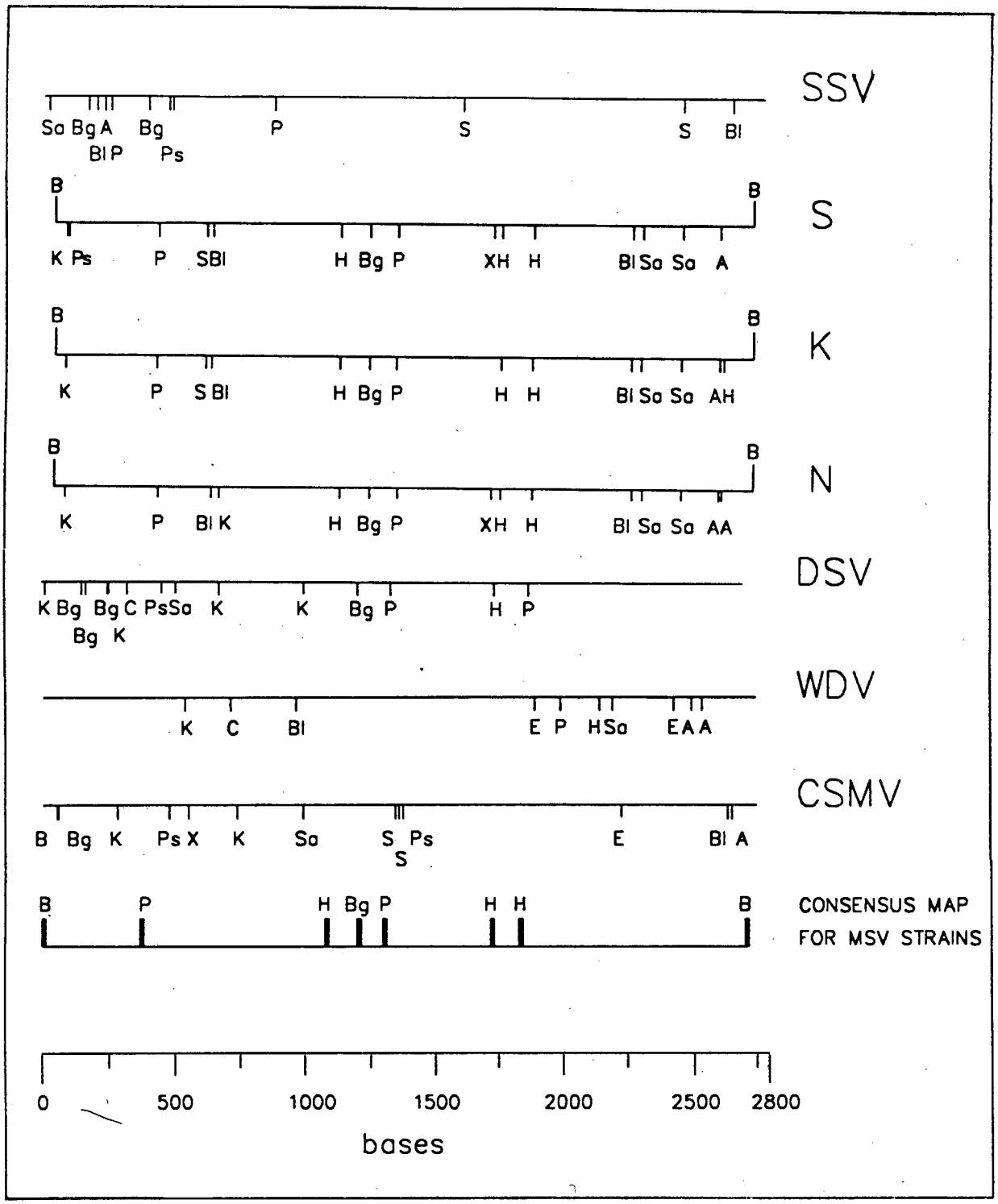
2.3.4. Restriction Endonuclease Cleavage Mapping of SSV.

Restriction mapping of SSV RF-DNA gave an estimated genome size of 2.8 kbp. Investigation of native and cloned DNA gave no evidence for the existence of more than one genome component. Preliminary sequencing of SSV established the orientation of the restriction map so that it could be directly compared with maps of maize isolates of MSV (MSV-S, MSV-K, and MSV-N) and the related geminiviruses DSV, WDV, and CSMV (Fig. 2.5).

The restriction map of SSV is entirely different to those of WDV and MSV isolates. SSV apparently shared only one restriction site with CSMV, and two restriction sites with DSV (Fig. 2.5).

These results were quantified to give estimated DNA sequence divergences (Table 2.2), and an unrooted phylogenetic tree for the geminiviruses was constructed, as described above (2.2.7), from which a rooted tree was derived by outgroup analysis (Fig. 2.6). The tree shown in Figure 2.6 was the best of 95 trees examined by FITCH, having an average standard deviation of 7.5%.

Figure 2.5. Aligned restriction maps of Subgroup I geminiviruses.



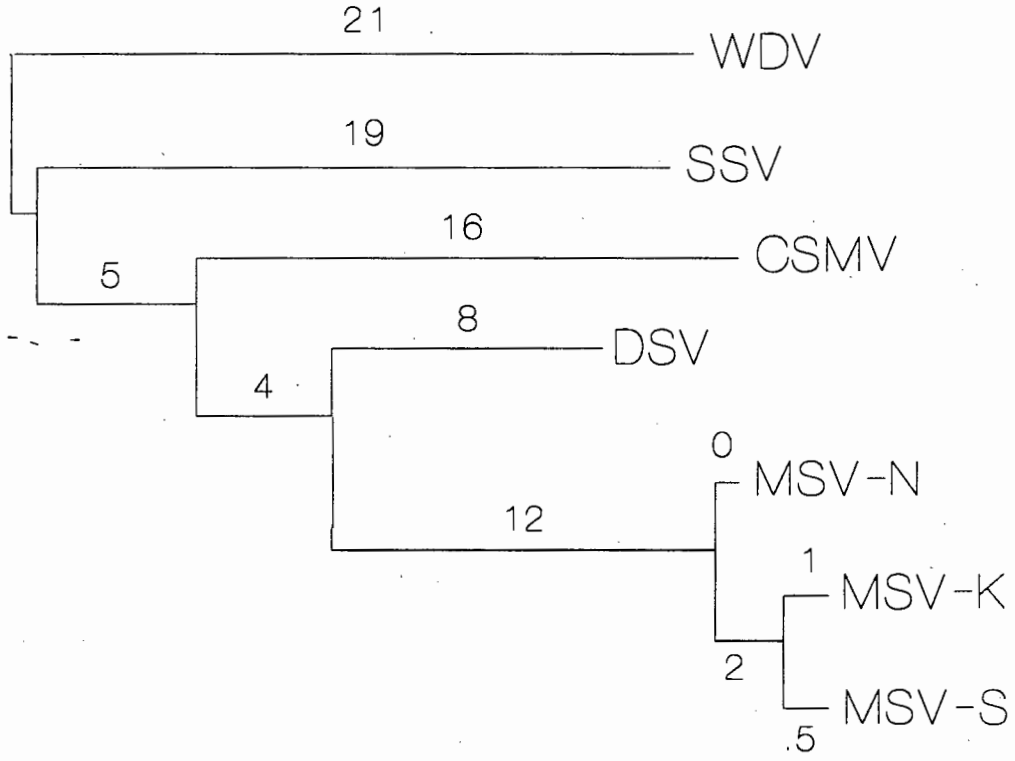
Legend: SSV = sugarcane streak virus. S = South African isolate of MSV (Lazarowitz, 1988). K = Kenyan isolate of MSV (Howell, 1984). N = Nigerian isolate of MSV (Mullineaux *et al.*, 1984). DSV = digitaria streak virus (Donson *et al.*, 1987). WDV = wheat dwarf virus (MacDowell *et al.*, 1985). CSMV = chloris striate mosaic virus (Andersen *et al.*, 1988). The composite "consensus map" at the bottom shows those restriction sites conserved in 5 out of 6 MSV isolates (Clarke *et al.*, 1989). Abbreviations for the restriction enzyme sites are as follows: A = *Apa* I; B = *Bam* HI; Bl = *Bgl* I; Bg = *Bgl* II; C = *Cla* I; E = *Eco* RI; H = *Hind* III; K = *Kpn* I; P = *Pvu* II; Ps = *Pst* I; S = *Sal* I; Sa = *Sac* I; X = *Xho* I.

Table 2.2. Matrix showing sequence variation between Subgroup I geminiviruses estimated from shared restriction endonuclease sites.

	MSV-S	MSV-K	MSV-N	DSV	SSV	CSMV	WDV
MSV-S	**16**	**14**	**14**	**3**	**0**	**2**	**1**
MSV-K	0.0170	**15**	**13**	**4**	**0**	**2**	**1**
MSV-N	0.0222	0.0292	**16**	**5**	**0**	**2**	**1**
DSV	0.2682	0.2145	0.1832	**14**	**2**	**2**	**0**
SSV	0.4408	0.4339	0.4408	0.3118	**12**	**1**	**0**
CSMV	0.3300	0.3241	0.3300	0.3184	0.4209	**13**	**0**
WDV	0.4273	0.4209	0.4273	0.4148	0.3994	0.4069	**10**

Key: The upper right of the table shows the number of shared restriction sites between pairs of viruses (in bold). Sequence divergence values are shown in the lower left of the table. The divergence values along the diagonal of identity = 0.0000. Isolate designations are as given in Figure 2.5.

Figure 2.6. Rooted phylogenetic tree of SSV and other Subgroup I geminiviruses. The estimated sequence divergences between viruses and branch populations are shown.



2.4. DISCUSSION

2.4.1. Serology.

It was not the purpose of this thesis to carry out detailed serological studies, but rather to characterize grass streak isolates by comparing their nucleic acid properties. The serological relationship between MSV and SSV has been studied in detail elsewhere (Pinner *et al.*, 1988; Dekker *et al.*, 1988; Pinner & Markham, 1990a, 1990b) by means of immunodiffusion and ELISA techniques, and using the same sugarcane source material as in this present study (clonally propagated sugarcane cv. Uba obtained from K. Harborne, Mt. Edgecombe, Natal). For this reason it was felt that a detailed serological study here would be superfluous.

Accordingly, although an ELISA test (DAS-ELISA) was carried out (following the method described by Rybicki & von Wechmar, 1982a), it was only to confirm that the viruses used in this study were similar to those used by the above-mentioned groups. The results (not shown) tallied with those already reported (Dekker *et al.*, 1988).

Since no comparison of MSV and SSV using immuno-electroblotting (Western blotting) has been reported by other groups, it was considered useful to carry out such a test. The dissimilarity between MSV and SSV seen in Western blots, where similar amounts of viral coat proteins reacted very differently towards antiserum to MSV indicated a distant serological relatedship between SSV and MSV. The results are in agreement with the findings of other serological studies ((Pinner *et al.*, 1988; Dekker *et al.*, 1988; Pinner & Markham, 1990b).

In a detailed serological study of grass geminiviruses, Dekker *et al.* (1988) reported that their sugarcane isolate from South Africa, S(SA)S, was serologically distinct from maize MSV isolates, with a serological differentiation index (SDI) of 3.6 against antisera raised to a Nigerian isolate of MSV (M(N)M). Further, S(SA)S reacted to only one of five groups of monoclonal antibodies (McAbs) raised to M(N)M, showing lower binding than that observed with M(N)M, indicating that SSV shares relatively few epitopes with MSV. Likewise, SSV and DSV were also very distantly related serologically to one another (SDI of 6.2). By contrast, DSV, recognized as a distinct geminivirus (Accotto *et al.*, 1989; Julia & Dollet, 1989), had

an SDI of 2.8 to M(N)M and reacted as strongly as M(N)M with four out of the five groups of monoclonal antibodies to M(N)M (Dekker *et al.*, 1988).

2.4.2. DNA cross-hybridization.

DNA cross-hybridization between virus isolates (Fig. 2.3) was indicative of the limited extent of sequence similarity between MSV, DSV and SSV DNAs. The insignificant hybridization of MSV DNA to DSV DNA in the present study reflects the overall sequence similarity of 64% between DSV and MSV (Donson *et al.*, 1987). The equally insignificant hybridization of MSV DNA to that of SSV, and *vice versa*, indicates at least as low a sequence similarity between SSV and MSV as that between DSV and MSV. These results also show that SSV is at least as distantly related to DSV as to MSV, since it cross-hybridizes with neither virus.

It is to be noted that the relationship established between MSV, DSV and SSV using DNA cross-hybridization data is in agreement with the distant relationship indicated by serological data (Dekker *et al.*, 1988).

2.4.3. Coinfection of SSV and MSV in Maize.

In this study SSV was transmitted to maize plants after lengthy acquisition feeds by first instar nymphs of *C. mbila*, and symptoms took longer to develop than for MSV isolates. Two factors could cause the inefficient transmission. *C. mbila* prefers certain plant species over others as feeding hosts (van Rensberg, 1982; Damsteegt, 1983). Thus our leafhopper colony, being adapted to feeding on maize, was perhaps feeding less efficiently on sugarcane and in so doing taking up reduced amounts of virus over a relatively long period of time.

This does not, however, explain the longer latent period observed for SSV in maize. McClean (1947) reported that even under optimized conditions, SSV had a minimum latent period in maize of 6 days compared with a minimum latent period of 3 days for MSV. Lowered replication rates of SSV in maize could account for this, perhaps indicating that SSV is less well adapted to maize than is MSV - possibly constituting evidence of a distant relationship between the two viruses.

The DNA cross-hybridization results for SSV/MSV coinfection show that SSV does not cross-protect maize against superinfection with MSV. Cross-protection by a virus strain is specific, affording no protection against infection by an unrelated

virus (McClean, 1947). The implication is that SSV and MSV are not closely related. This experiment supports early findings, where the mild form of MSV in maize (MSV-B) cross-protected against superinfection by the severe form, MSV-A, but SSV transmitted into maize cross-protected against neither MSV-A nor MSV-B (McClean, 1947).

The difference in staining intensity between SSV and MSV in the coinfected samples showed that there was less SSV DNA than MSV-Koe DNA present in these samples. Considering that the plants were infected first with SSV, this implies that MSV was replicating faster than SSV, evidence that MSV is better adapted to maize than is SSV. The development of symptoms towards the typical MSV pattern also suggests that MSV out-competes SSV in maize. Moreover, SSV replication appears to be better supported in a very susceptible maize cultivar, since SSV in coinfected sweetcorn stained more intensely than SSV in coinfected Witplat while MSV-Koe showed no significant difference in staining intensity (and hence replication efficiency) in either cultivar.

2.4.4. Restriction Mapping and Phylogenetic Analysis.

It has been found in this laboratory (Clarke *et al.*, 1989; Kirby *et al.*, 1989) that restriction mapping provides a useful indication of the degree of genomic similarity between MSV isolates, in that positions of certain restriction sites tend to be conserved between strains. As shown in Figure 2.5, "commonly conserved sites" denotes those restriction endonuclease sites conserved in five out of six virus isolates (Clarke *et al.*, 1989). SSV retains none of the seven commonly conserved sites found in the mapped MSV isolates. The marked difference between the restriction map of SSV and the maps of known MSV strains from widely-spaced locations in Africa (Fig. 2.5) argues against a close relationship between these two viruses. Furthermore, as can be seen from Figure 2.5, the SSV map differs from those of MSV, DSV, WDV and CSMV as much as the maps of these four viruses differ from each other.

The phylogenetic tree calculated from the restriction site conservation data (Table 2.2, Fig. 2.6) was rooted using WDV as an outgroup. SSV is separated from WDV, CSMV, DSV, and the MSV isolates, and has a calculated branch length of 19%. It should be noted however that comparing restriction maps of viruses having

few or no sites in common gives estimated sequence divergences that are beyond the method's accurate area of estimation (Nei & Li, 1979; Gibbs & Fenner, 1984). Thus, while the tree indicates that SSV is not closely enough related to MSV to be considered a strain of MSV, the exact relationship between these viruses cannot be deduced. The same applies to CSMV, DSV, and WDV.

The method used here gives a serviceable indirect estimate of sequence divergence between geminiviruses (compare trees in Chapters 4 and 5) but does not eliminate the need for sequence data to definitively establish the ancestral branching patterns in the phylogenetic tree.

In conclusion, it is proposed that SSV is a distinct geminivirus. This conclusion is supported by the lack of cross-protection afforded by SSV against MSV, the marginal hybridization of its DNA to that of MSV, the lack of conserved restriction sites between MSV and SSV, as well as the distant serological relationship between SSV and MSV and the estimated sequence divergence of greater than 30% between SSV and the MSV isolates, compared with the estimated sequence divergence of 5% between MSV strains. Moreover, SSV is distinct from DSV by restriction map and DNA cross-hybridization criteria, and from WDV and CSMV by restriction map criteria.

CHAPTER 3

COMPLETE NUCLEOTIDE SEQUENCE OF THE SUGARCANE STREAK GEMINIVIRUS.

CONTENTS

CHAPTER 3.
COMPLETE NUCLEOTIDE SEQUENCE OF THE SUGARCANE STREAK GEMINIVIRUS.

SUMMARY

The complete nucleotide sequence of the genome of the geminivirus sugarcane streak virus (SSV) was determined from cloned replicative-form cccDNA. The genome is contained in one DNA circle of 2758 nucleotides. The sequence has four open reading frames with the potential to encode proteins of MW >10 kDa: two in the viral (+) sense and two in the complementary (-) sense. Each open reading frame has a counterpart among the open reading frames reported for other Subgroup I geminiviruses. A potential binding site for a DNA replication primer and potential transcriptional control sequences were identified on the (+) strand. Nucleotide and derived amino acid sequence comparisons were made between SSV and other Subgroup I geminiviruses. Phylogenetic trees were constructed from aligned coat protein and replication-associated protein sequences of SSV and other grass geminiviruses. The results confirm that SSV is a distinct geminivirus and indicate its phylogenetic relationship to other Subgroup I geminiviruses.

3.1. INTRODUCTION.

The *Cicadulina mbila*-transmitted geminivirus isolated from diseased sugarcane from Natal (Chapter 2) has been claimed for some time to be a strain of maize streak virus (MSV) (Pinner *et al.*, 1988), since it can infect maize, is transmitted by the same vector, and is serologically related to MSV. It has been proposed, however, that the sugarcane virus is a distinct virus (termed sugarcane streak virus, SSV) rather than a strain of MSV (Hughes *et al.*, 1990; Chapter 2) on the basis that it has been shown to be very different from MSV serologically (Dekker *et al.*, 1988; Pinner & Markham, 1990a, 1990b), it does not cross-protect maize against MSV infection (McClean, 1947; Chapter 2), the nucleic acid acids of the two viruses do not cross-hybridize, and the SSV restriction endonuclease map is entirely different from those of MSV isolates (Hughes *et al.*, 1990; Chapter 2).

Obtaining the genomic DNA sequence of SSV would allow sequence comparisons to be performed between SSV and other already sequenced Subgroup I

geminiviruses at both the nucleotide and deduced amino acid level. Such comparisons would be expected to give more definitive information on the taxonomic status of SSV and its exact relationship with other members of its subgroup.

An examination of percentage DNA sequence similarities gives some indication of the level of sequence similarity expected between Subgroup I geminiviruses that are considered to be isolates of the same virus and the level of similarity shown between distinct Subgroup I geminiviruses.

The three sequenced isolates of MSV available to date - a Nigerian isolate (MSV-N, Mullineaux *et al.*, 1984), a Kenyan isolate (MSV-K, Howell, 1984, 1985), and a South African isolate (MSV-S, Lazarowitz, 1988) - are approximately 98% identical in nucleotide sequence (Lazarowitz, 1988). DSV, the most closely related geminivirus to MSV, shares only 64% nucleotide sequence identity with MSV (Donson *et al.*, 1987). CSMV and WDV differ from each other and from MSV and DSV by about 50% in direct sequence comparisons (Andersen *et al.*, 1988; MacDowell *et al.*, 1985). However, the two sequenced WDV isolates, WDV-S (from Sweden, MacDowell *et al.*, 1985) and WDV-CJI (from Czechoslovakia, Woolston *et al.*, 1988) share an overall nucleotide sequence homology of 98.3% (Woolston *et al.*, 1988).

Thus, for SSV to be considered a distinct virus rather than a strain of MSV, a minimum expectancy based on sequence similarity is that it should share no more than about 65% overall DNA sequence similarity with MSV.

In this chapter, the complete nucleotide sequence of SSV is presented and compared at both the nucleotide and derived amino acid level with other sequenced Subgroup I geminiviruses. The deduced phylogenetic relationship of SSV to other Subgroup I geminiviruses is given. In addition, comparative inferences are drawn on determinants of transcriptional specificity and host specificity in cereal geminiviruses.

The results confirm that SSV is a distinct geminivirus, and as distantly related to its nearest neighbours, MSV and DSV, as these two viruses are distant from one another.

3.2. MATERIALS AND METHODS.

3.2.1. DNA Cloning and Sequencing.

Streak-diseased sugarcane cv. Uba from Mt. Edgecombe, Natal (see Chapter 2) was used as the source of DNA for cloning and sequencing SSV.

Total plant DNA containing SSV viral DNA was extracted as detailed in Appendix A.3.

Following the methods detailed in Appendix A (A.8 - A.10), double-stranded replicative-form DNA of SSV was cloned in both orientations as an approximately full-length *Pst* I fragment, and as two separate *Bgl* II fragments, into the plasmid vector Bluescript SK (Stratagene, California) restricted with *Pst* I and with *Bam* HI, respectively.

The smaller of the two *Bgl* II fragments was sequenced entirely in both directions (see Appendix B.2), using pUC sequencing primers (Bethesda Res. Labs.).

Sets of ordered deletions were constructed (Appendix B.1) for the *Pst* I clones and for the clones containing the larger of the two *Bgl* II fragments, via exonuclease III / S1 nuclease shortening, according to Henikoff (1984). Transformants were sized initially by electrophoresis of uncut plasmid preparations (rapid method; A.11.a). Transformants containing plasmids of apparently differing size were again subjected to plasmid preparation ("miniprep"; A.11.c). This miniprep DNA was digested with *Pvu* II. The restriction patterns of the electrophoresed digests were compared with patterns expected for *Pvu* II-digested recombinants containing progressively shortened viral insert DNA. Appropriate clones were selected such that the insert of each clone was never more than 300 bp shorter than that of the next longest clone, in order to obtain overlapping sequence.

DNA was sequenced by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977), using the Sequenase sequencing kit (see Appendix B.2) and the pUC "forward" sequencing primer. Sequence ambiguities due to band compressions were eliminated by the addition of dimethylsulphoxide (DMSO) into the labelling and termination mixes to a final concentration of 10%, together with incubating the labeling reactions at 15°C and the termination reactions at 45°C (see B.2).

A synthetic SSV primer was used to sequence across a region of the virion strand for which no appropriately shortened clone could be obtained. The primer

was designed by consideration of the virion strand sequence proximal to and upstream of the gap in the sequencing direction. The primer sequence chosen (5'-TTGAAGCGACGGCTAGG-3') was checked for unique binding to the designated primer region and fulfilled the requirement for a greater than 50% G-C content, necessary for efficient primer binding. The primer was purchased from Beckman Instruments (SA) (Pty) Ltd., Cape Town.

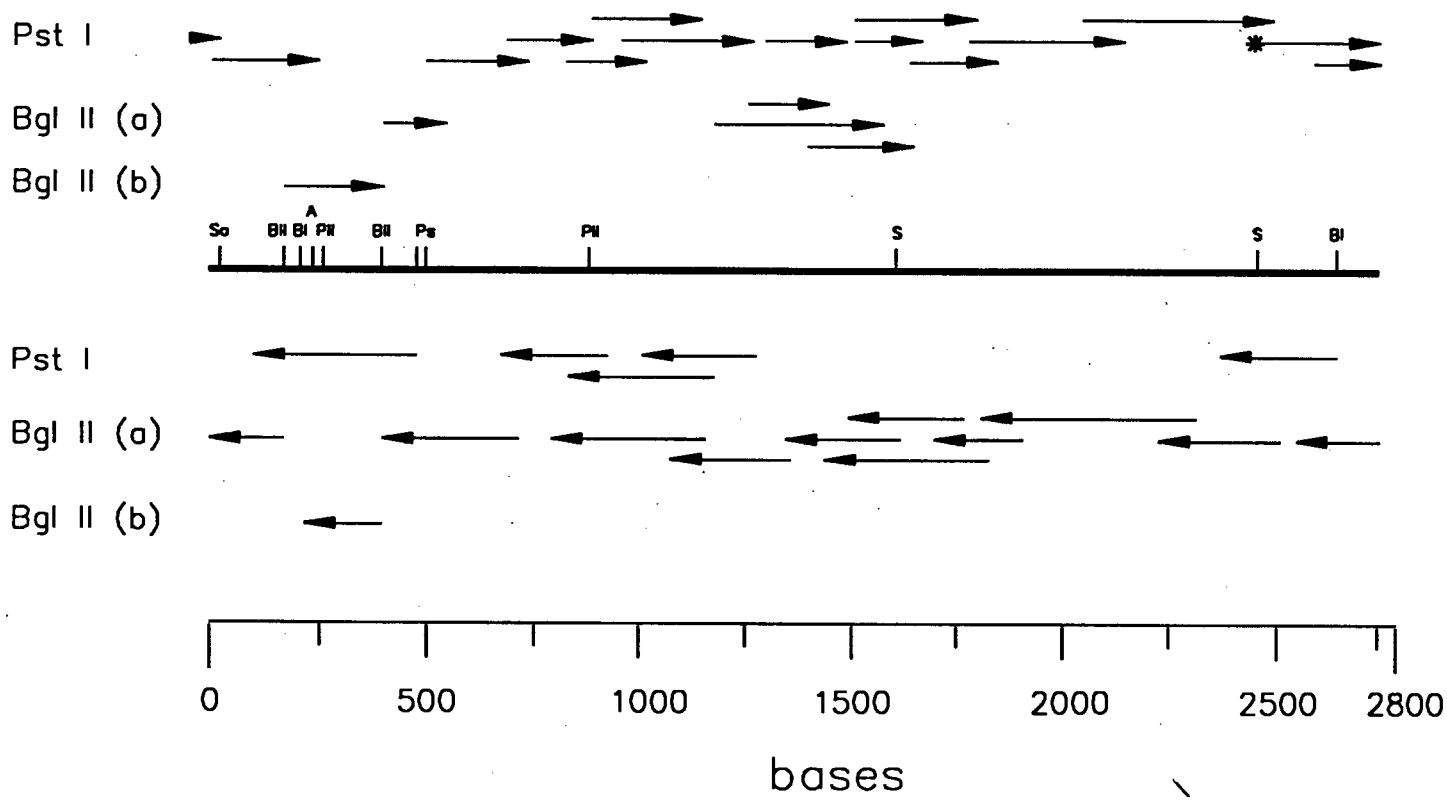
3.2.2. Sequence Analysis and Comparisons with Other Geminiviruses

SSV sequence information was stored, assembled and analysed on an IBM-compatible personal computer using GENEPRO Ver. 4.20 software (Riverside Scientific 18332. 57th Ave. N.E. Seattle WA98155), and the UWGCG package, Version 6.1 (Devereux *et al.*, 1984) on a DEC/VAX 6000-330 mainframe computer. The UWGCG programmes GAP and BESTFIT were used for pairwise sequence similarity comparisons between nucleotide and derived amino acid sequences of SSV and those of other Subgroup I geminiviruses. The sequences of MSV-N, MSV-K, MSV-S, CSMV, WDV, and DSV were obtained from the GENBANK database. Since MSV-S, MSV-N, and MSV-K gave nearly identical results in comparisons made between MSV and other geminiviruses, only those comparisons made using MSV-S are shown.

3.2.3. Phylogenetic analysis.

The UWGCG programme LINEUP was used for multiple alignment of predicted coat- and replication-associated protein sequences of SSV, MSV-S, DSV, CSMV, and WDV, aligned pairwise using GAP, as above. The UWGCG programme DISTANCES was then used for calculation of matrices of pairwise sequence similarity values between aligned sequences. These values were converted to difference values by subtraction from 1.0000, and used as data matrices (shown in Table 4) for the programme FITCH in PHYLIP version 3.1 (phylogeny inference package; Felsenstein, 1988), as used by Howarth & Vandemark (1989). FITCH uses the algorithm of Fitch & Margoliash (1967) to construct an unrooted phylogenetic tree. The options G (global) and J (jumble) were used for each data set to ensure that the shortest possible tree was generated.

Figure 3.1. DNA sequencing strategy.



Legend: The central line represents the SSV DNA sequence and shows mapped restriction sites (see Chapter 2): A = *Apa* I; BI = *Bgl* I; Bg = *Bgl* II; Ps = *Pst* I; P = *Pvu* II; Sa = *Sac* I; S = *Sal* I. Arrows denote the position, extent, and direction of each piece of sequence obtained. *Bgl* II (a) = large cloned *Bgl* II fragment. *Bgl* II (b) = small cloned *Bgl* II fragment. * = location of the binding site of the synthetic oligonucleotide primer (5'-TTGAAGCGACGGCTAGG-3').

The same data sets were also run with the programmes KITSCH from PHYLIP, and with NJTREE version 2.0. The former uses the same method as FITCH but assumes a molecular clock; the latter generates phylogenetic trees by the neighbour-joining method of Saitou and Nei (1987) as modified by Studier and Keppler (1988).

3.3. RESULTS AND DISCUSSION.

3.3.1. Nucleotide Sequence of SSV.

The complete nucleotide sequence of the SSV genome was determined in both orientations (Fig. 3.1). More than 99% readable sequence was obtained in both senses and an average of six readings per nucleotide sequenced was made.

Restriction mapping (Hughes *et al.*, 1990) and sequence data were consistent with the genome being contained in a single DNA circle of 2758 nucleotides. Sequence orientation was deduced by comparison with the published sequence of MSV-S. The virion strand sequence of SSV is shown in Figure 3.2. The ATG start codon of the 11.16 kDa open reading frame (ORF) is defined as position 1 of the sequence, in conformity with the published sequence of DSV (Donson *et al.*, 1987). Structural features of the genome are shown in Figure 3.3 and summarized in Table 3.1.

Two nucleotide transversions were encountered between clones. An A/T change at position 281 (second base of codon) alters Asn to Ile in the 11.16 kDa ORF - but occurs in a region of the polypeptide that is not well conserved between the viruses being compared in this study. The second transversion, T to G at position 1782 (third base of codon), does not result in an amino acid change in the complementary sense 34.94 kDa ORF.

3.3.2. Potential Coding Regions.

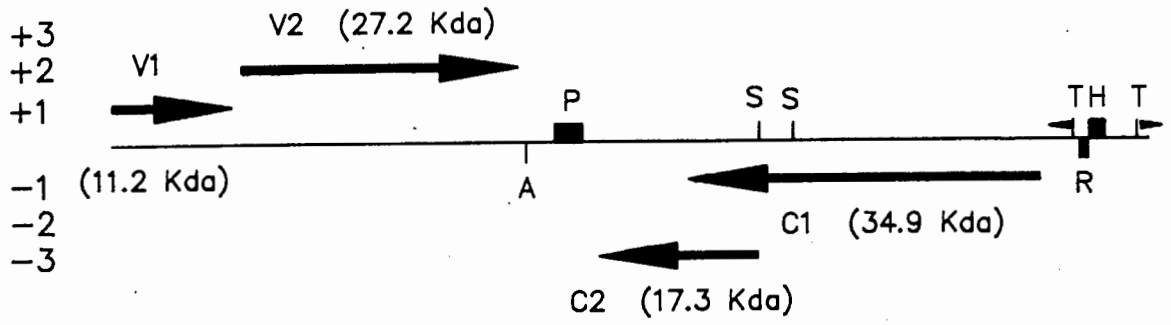
The SSV nucleotide sequence was screened in all three reading frames in both the virion (+) and complementary (-) senses for coding regions with the potential to code for proteins with molecular weights >10 kDa. Open reading frames (ORFs) were identified initially by computer inspection and by comparison with ORFs of other sequenced cereal geminiviruses.

Figure 3.2. Nucleotide sequence of SSV DNA.

```
ATGGATAGCT TTGGCAGAGC TCCTCCTCTT TGGCCGCAGT CAGCTTTGCC TCGGGTCCCT    60
GGTGCTGCTC CGTCGTCCTC TGGACTGCCG TGGAGCCGCG TCGGTGAGAT AGCTATCTTT   120
ACCTTTGTTG CAGTATTGGC GCTTTACCTG CTCTGGAGTT GGGTGGGTAG AGATCTTCTC   180
TTGGTTCTTA AGGCTCGCCG CGGAGGCACT ACGGAGGAGC TGACATTTGG GCCCAGGGAG   240
CGCCACAGTT TACCAGCTGT AGCTGTTGCT CGGGTGGAAA ATCCTCCGTG TCCCAGTGGT   300
TCAGTGGAAG CTAGACCGTT CACCGGTTAG CTGTGCCCTC AGCATGCCTC TCTCTGGAAT   360
GAAGCGTAAG AGGTCCGATG AGACGGGTCG CCGTAAGAGA TCTTCTGGGG TGAAGCAGGG   420
GCGTACATCT GCCGCCCGTG CTGGTTCTGC TGTTCGCAGG ACTCGTCCCT CTCTGCAGAT   480
CCAGACGCTG CAGGCGGCTG GTACTTCAAT GATCGAAGTG CCATCCGGTG GCGTTTGTGA   540
CCTTCTTGGT TCCTTCTCCC GTGGCTCTGA CGAGGGGAAC AGGCACACCA ACGAGACTGT   600
CATCTACAAG GTGGCACTGG ACTACCACTT CATTGCTACC GCTGCTGCCT GCAAGTATTC   660
ATCTATTGGC ACTGGCGTGG TGTGGTTGGT GTATGACGCA CAGCCTTCGG GGAATCCCCC   720
GACAGTGAAG GATATTTTCC CTCACCCTGA TACTTTGACA GCGTTCCCGT ACACCTGGAA   780
GGTTGGTCGT GAAGTGTGTC ATCGGTTCGT GGTGAAGCGG AGGTGGACTT TCACGATGGA   840
AGTTAACGGG CGTATCGGTT CCGATATACC TCCTAGTACC AGCTGTTGGC CCCCGTGCCG   900
TAAGAACATT TACTTCCACA AGTTTGTCAC CGGCTTAGGT GTTAAAACGG AGTGGAAGAA   960
CACTACCGGC GGTGAAGTCG GTGACATTAA GAAGGGTGCC CTGTATATTG TAATTGCCCC  1020
TGGCAATGGG TTAGATTTTA CGGTGCATGG TAATGCCCGT TTGTATTTCA AGTCAGTCGG  1080
GAACCAGTAG GTCCTGAGTT TAATAAAATT CCATTTTATT TCATGATAAG TCTAGCCTTG  1140
TCTCTGACAA GGTGGGTTAC ACAGGCAGCC GCTTGGCTGA CAAAAACACT AAACTGATAC  1200
AAACGGCGGC AGATCGTAGG CGGCTAAGGG ATAGGAATAT GAAAAAAAAA ACACATCCAA  1260
ACTGACATTA TTGAAATTGG TCGGCGCTCA GACTTCGCTA GAAGGATGGG CTGTAGCTGA  1320
CACCGCTGGC AGGCTGAAGA ACCGCTCTCC CGGTTCTAGC ATGTATGTCT CACAGTTTGC  1380
ATTGAAATAG TCCTGCTGGG CAGGCGTCAT GTCTCTCAAC CAGTCCTCAT CCTCATTGGC  1440
TAGCACTATA GTGGGGATGG ATTTACTAGC TACTTTCTTA CGCTTGCCGT ACTTTGGATT  1500
TACAATATAA TCCTTCTGGC AGCCAATCAG TTGCTTCCAG CATGGACAGT ACTTGAAGGG  1560
TATGTCGTCG ATGATGTTGT ATTCTGCGTC CTCGTCGTAG GAGGACCAGT CGACGTTGTT  1620
CTGCCAGTAA TTATGTCTTC CCAGGCTCCT GGCCCAGGAG GACTTGCCTG TTCTTGTTGG  1680
CCCGAGGATG TACAGGCTGA GCTTCCTGTT GGAATTAGAG TTCTGGATGT TTCATGCATC  1740
CATATCAGAT CAGAAGTTGC TTCTTCCAGG GTTAAGCAAC TGGGATTAGC TAGCATGTAT  1800
GCAGAAGGAC TAACCTGGTA GAGGTTGTTC TCAAGCCAGG TATTTATTGC CGTAGGATCG  1860
AGGAGTGCTG GTGTGCTTTG AGGGAAAGGA CTGGTGTATA TTTCCTGGAT GTCAGGAAAC  1920
AGTTTATTTG CCGAATACTC AAAGTATTGT AGTTTGGTGG CCCACTCGTA AGGTAGAGCT  1980
TTCTGTATCA TGGAGAGGTA CTCTTGCTTG TTGGTGGAAT GTTCTATGAT ATCTCTTACG  2040
ATATCATCCT TGGATTGCTT CTTGGAATTT GGCTCAGTCG AATCTCCTAG GAAGGAGGTC  2100
TTCCTAGGTA TGAACGTACC TCTCTCCCAG TATGTAACCG GGTTCTTCAT GGCGTATGCT  2160
CTGACCTTGT TGGCACTCTT TGCACTCTGA ATGTTGGGAT GGAAGCCATC AATGTCGAAG  2220
AAGCCGGAGT CAGTTGTGTA GACAGGTTTT GCACTCTGAG CGAGTACGTG GATGTGATAC  2280
CCTCCGTCTT CATGTGTCTC CCTGACAGAT AGGACATACA CTGGTGTCCA GTGTGCAATG  2340
AGGCTCCATA TGTGTAGACC TACAGCTTCA GGTTCTAATG GACAGCGAGA GTAGGTGAGG  2400
AACGTGTTCA CGTTGCGGTG TTTGAAGCGA CGGCTAGGAG TGGAAGACAC AGTTGATCCT  2460
ACGGTCGACA TCTCTTTGGG CAGTGTCTGA TTTGGAGCCT CAACCAACCT CCTACAGGTT  2520
TGCGATTTAG AATCACGCAC TCCCCCTGCC CTCTTATAGG AGTGCTAAAT GGGCTGGGCC  2580
GAATTTTTTC TTGGGCCGCA ATGAAAAAGC GCAAACCAAT AATATTACAG CGCTTTTTCA  2640
TGCGAGGCCC ACCAGGCCGA CGGGTCCCGG TTCTGGGTGC GTTTTGTTCT TTATTGCCGG  2700
CTGCTTTGAA ACGCCGGCTT TGGATCGCAC TATAAGGGGG TTGTGTCATG TGTTCGAC     2758
```

Legend: Sequence of the virion (+) strand is shown, beginning with the first nucleotide of the ATG start codon for the 11.16 kDa ORF (V1), in conformity with the nucleotide sequence of DSV (Donson *et al.*, 1987). Arrows indicate the conserved hairpin structure in the large intergenic region.

Figure 3.3. Map of the SSV genome derived from sequence data.



Legend: Large arrows indicate the direction and extent of open reading frames. Numbers at left indicate reading frames with respect to virion sequence. The designation and predicted M_r of each potential product is given.
A = potential polyadenylation signal AATAAA
H = conserved hairpin loop
P = probable binding site of endogenous DNA primer
R = *rpe* 1 equivalent
S = probable pre-mRNA splice positions.
T = TATA box most likely to be promoter-associated (small arrowheads indicate probable direction of transcription)

Table 3.1. Features of the SSV genome.

ORF[a]	frame	product (kDa)	TATA[b]	start	stop	AATAA[f]	core UAS
V1	1+	11.16	2731	1[c]	328[d]	1102	2575-2598
V2	2+	27.20	2731	344[c]	1088[d]	1102	2575-2598
C1	3-	34.94	2555	2471[c]	1529[e]	1120	
C2	1-	17.28	2555	1731	1290[e]	1120	

Key: a = open reading frame. b = CTATAA - most likely TATA box matching the consensus promoter sequence for plant genes, $T^{C}/_{G}TATA^{T}/_{A}A_{1\text{-}3}$ (Messing *et al.*, 1983). c = AUG. d = UAG. e = UGA. f = most likely polyadenylation signal matching the consensus $^{A}/_{G}ATAA$ for plant genes (Messing *et al.*, 1983).

A table of codon usage among the sequenced Subgroup I geminiviruses was compiled from the coat protein and replication-associated protein ORFs, using the programme CODONFREQUENCY (UWGCG). This table was used in the programme FRAMES (UWGCG) to indicate the frequency of rare codons in potential ORFs of SSV, and as such was used as a test of ORF validity for SSV (not shown).

Two ORFs were found in the virion sense sequence. These were termed V1 and V2 after Boulton *et al.* (1989). V1 potentially encodes a polypeptide of 11.16 kDa, corresponding to the 10.9 kDa protein of MSV which is essential to systemic spread of the virus in the host plant (Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989). V2 encodes a 27.2 kDa polypeptide corresponding to the 26.8 kDa ORF of MSV, which encodes the capsid protein (Morris-Krsinich *et al.*, 1985). The capsid protein both determines vector specificity (Briddon *et al.*, 1990) and is essential for systemic spread in the host (Lazarowitz *et al.*, 1989; Boulton *et al.*, 1989).

Two partly overlapping ORFs (C1 and C2) were found in the complementary sense. C1 has the coding potential for a polypeptide of 34.94 kDa and corresponds to the 31.3 kDa ORF of MSV-S. C2 potentially encodes a polypeptide of 17.28 kDa (corresponding to the 17.2 kDa ORF of MSV-S), but lacks an ATG start codon. The corresponding ORFs of CSMV, DSV and WDV also lack an initial ATG codon.

Two other ORFs were found in the (-) sense with the potential to encode proteins of > 10 kDa. However, neither had an upstream TATA sequence, one contained a number of rare codons, and the other included a hairpin structure that is highly conserved in the common intergenic region of all geminiviruses (see below). Therefore it seemed unlikely that either of these ORFs were functional.

It has been shown for WDV and DSV (Schalk *et al.*, 1989; Accotto *et al.*, 1989) and deduced for MSV and CSMV (Mullineaux *et al.*, 1984; Lazarowitz, 1988; Schalk *et al.*, 1989) that the ORFs C1 and C2 on the (-) strand are transcribed as a single unit. An intronic sequence is spliced out of the pre-mRNA transcript to yield a fusion product of the two ORFs (here termed C'), showing maximal amino acid homology to the ~40 kDa replication-associated protein of the bipartite genome geminiviruses.

The C1 and C2 ORFs of SSV were examined for presence of an intronic sequence, according to Schalk *et al.* (1989). The potential intronic region of SSV (positions 1723 to 1814) is shown in Figure 3.4.

Figure 3.4. Intronic region of the overlapping C1 and C2 open reading frames (ORFs) showing splice sites.

(a).

```
SSV    TACCAG GT TAGTCCTTCTGCATACATGCTAGCTAATCCCAGTTGCTTAACCCT
MSV    TTCCAG GT TAGTCCCGAAGCTTACATGCTCCTTCAACCTACCTGTTATACCCT
DSV    TTTCAG GT CAGTCGTGAAGCTTATTGCTTGTTGAACCCTACCTGTTATACCCG
WDV    TACTCC GT AAGCCTCGAATCCTATATCCTT......TGTACTTCCACTCCTGC
CSMV    GT CCTCAAGCCCTCAGTCTGCATGCAGGTATTTCGGAGGAACAAGCACGTATT

SSV    GGAAGAAGCAACTT CTGATC TGATATGGATGCATGAAACATCC AG AACT
MSV    CGAGGATGCAATTT CTGAC CTCCAATGGATGGATTCTGTATCC AG TCAT
DSV    CGAGGAAGCTATCT CTGAC TTACAATGGATGTCTGATTACACT AG GTCA
WDV    GGATCAAGCGCAAT CTGAC TTAGAGTGGATGGACGATTATTCC AG GAGT
CSMV   GACCTCCAATGGATGT CTGAC CTAACCAGGTCTAGAGCCCTGG AG TCCG
```

(b).

```
SSV    SSPDLLCNESINDWLQPNIFQ  VSPEAY.....
MSV    SSPDLLCNESINDWLQPNIFQ  VSPEAY.....
DSV    SEPDLLCQETITDWLQNDLFQ  VSAEAY.....
WDV    SLICH.ETIES..WKNEHLYS  VSIESY.....
CSMV   SR.DMSDHPVIGEWLQQELYTV  SPQAL.....

SSV        NIQ  NSN.SNRKLSLYILGPTRTGKSSWARSLG...
MSV    MDGFCIQ  SSDERSRKQSLYIVGPTRTGKSTWARSLG...
DSV             VTDGV.RKRSLYILGPTRTGKSTWARSLG...
WDV    LRVDGRLFQ  ESPGRHKSIYICGPTRTGKTSWARSLG...
CSMV           SPGV  IRRRSLYICGPTRTGKTSWARSLG...
```

Legend: (a) DNA sequences aligned as shown in Schalk *et al.* (1989). Splice sites and the lariat sequence are displayed in bold type. (b). Aligned amino acid sequences of the spliced region of the resultant C'(~40 kDa) protein. Sequences aligned as in Schalk *et al.* (1989). Exon sequences are displayed in bold type.

The region is unusual in that the highly conserved "lariat" sequence CTGAC found in all other Subgroup I geminiviruses is represented unambiguously in SSV as CTGATC (bases 1748 - 1753).

Experimental evidence suggests that the host may exert a negative control over intron splicing in DSV (Mullineaux *et al.*, 1990). This control appears to be partly operative in *Digitaria* and permanently expressed in tobacco.

Yet not all heterologous systems exert negative control on DSV intron splicing - a HeLa cell splicing system was found to splice the DSV intron (Mullineaux *et al.*, 1990). This splicing control phenomenon, if general, may explain why Subgroup I geminiviruses cannot infect dicotyledonous plants, and why the Subgroup III geminiviruses, which infect dicotyledonous plants, have a single large ORF encoding the replication-associated protein rather than the spliceable C1-C2 system operative in Subgroup I geminiviruses. It has been speculated that the bipartite geminivirus genome evolved from an ancestral monopartite genome (Howarth & Goodman, 1986). Possibly, the acquisition of a "permanently spliced" ORF was intrinsic to the geminivirus host transition from monocots to dicots.

3.3.3. Noncoding Regions and Potential Control Sequences.

There are two apparent intergenic regions in SSV, located on opposite sides of the genome (Fig. 3.3), as reported for other Subgroup I geminiviruses (Lazarowitz, 1987).

The larger intergenic region of SSV (positions 2472-2758) features a hairpin structure (positions 2601-2641, see Figs. 3.2 and 3.3) that is highly conserved in the equivalent region of all other geminiviruses (Lazarowitz, 1987). The SSV hairpin has a free energy of formation (ΔG) of -19.2 kcal/mol, calculated according to Salser (1977). As in all other geminiviruses, the conserved sequence TAATATTAC is present in the loop of the SSV hairpin, while the stem sequences of the structure are dissimilar to those of other geminiviruses.

Potential transcription signals (promoter and polyadenylation sequences) were searched for. Based on their positions, those that seemed most likely to be functional are detailed in Table 3.1.

In MSV, a GC-rich "upstream activating sequence" (UAS) immediately upstream of the conserved hairpin structure in the large intergenic region is

required for rightward transcription of V1 and V2 (Fenoll *et al.*, 1988). The crucial region of the UAS is the rightward (or core) promoter element (*rpe* 1), comprising two adjacent "GC-boxes" which potentially form a hairpin structure (Fenoll *et al.*, 1988; Fenoll *et al.*, 1990). Maize nuclear factors are proposed to bind to rpe 1 G residues located on opposite faces, separated by one and a half turns, of the DNA helix of double stranded replicative-form virus DNA.

Figure 3.5. Aligned "Upstream Activating Sequences" of SSV, MSV, DSV, WDV, and CSMV, showing core sequences.

```
          *****                          *****
MSV-N     TGGGCCGGACC..........GGGCCGGCCC

MSV-S     TGGGCCGGACC..........GGTCCGGCCCA

MSV-K     TGGGCCGGACC..........GGGCCGGCTA

DSV       TGGGCCAGAGGCCCAATACCATGGGCCGCG

SSV       TGGGCCGAATTTTTTCT....TGGGCCGCA

WDV       TGGGCCCTGTTCGGTGTGCGGTCG

CSMV      TGGGCCATAACAGGCCGGCTCTTT
```

Legend: Upstream activating sequence *rpe* 1 equivalents, aligned to give maximum homology between core sequences (indicated in bold type). Each sequence shown extends to the nucleotide preceding the 5' terminus of the conserved stem-loop structure of the large intergenic region. MSV-N = Nigerian isolate of MSV, (Mullineaux *et al.*, 1984). MSV-K = Kenyan isolate of MSV (Howell, 1985). Asterisks indicate G residues implicated in nuclear factor binding (Fenoll *et al.*, 1990).

A sequence corresponding to *rpe* 1 of MSV was located at positions 2575 to 2597 in the large intergenic region of SSV, immediately upstream of the conserved hairpin structure (Fig. 3.5, Table 3.1). The *rpe* 1 equivalent in SSV has two direct repeats of the core sequence 5'-TGGGCCG-3' (Fenoll *et al.*, 1990) separated by an intervening AT-rich stretch of 10 nucleotides (Fig. 3.5). The AT-rich segment spaces the GC core sequences in SSV such that a hairpin structure cannot form, and the sites critical to nuclear factor binding on the DNA helix would not be topologically equivalent to those of MSV. The GC cores of the DSV UAS equivalent are separated by 15 nucleotides, while WDV and CSMV appear to have only one GC core

sequence each (Fig. 3.5). Thus, the presence of two adjacent GC-boxes in the UAS of MSV isolates is probably peculiar to MSV, and their topological positioning on the DNA helix may be fortuitous.

It could be speculated that the number of core sequences per virus may relate to transcription requirements of the host plant and could contribute, at least in part, to restriction of a virus to certain hosts. Variations in the GC-rich region of the UAS of the bipartite genome geminivirus, squash leaf curl virus, are associated with differences in host range (Lazarowitz, 1987). It may be speculated that the duplicated GC core sequences of SSV, DSV, and MSV facilitate their infection of maize, but that the spacing of the GC boxes of MSV make it transcriptionally more active than SSV or DSV in maize and hence more virulent.

3.3.4. Potential DNA Primer Binding Site.

Whereas the presence of a putative primer molecule for SSV was not determined experimentally in this study, the existence of a primer binding site (PBS) on the (+) strand was inferred by comparing the DNA sequence of the smaller of the two intergenic regions of SSV with corresponding sequences from MSV, DSV, WDV, and CSMV (Donson *et al.*, 1984 ; Donson *et al.*, 1987 ; Hayes *et al.*, 1988 ; Andersen *et al.*, 1988).

Figure 3.6. DNA sequence of the binding site on the SSV virion strand for the endogenous DNA complementary strand primer, aligned with primer binding sites of MSV and DSV.

	5' terminus	3' terminus
SSV	TTGGCT.	**GACAAAAACACTAAACTGA**.....**AATATGAAAAAAAAAAC** AC..AT
DSV	GCGGC..	**AACTAAAACACACATGACC**.....**TAAAATTAAAAAAAAAC** ACAGAT
MSV-S	ATGGCAC	GA.**AAAAACACACGCAATC**.....**CAAAACATCGAAAAATC A.AGAT**
Consensus	..GGC..	GACAAAAACACACA.....AAAA.AAAAAAAAAAC ACAGAT

Legend: Primer binding site sequences are shown in bold type. Consensus sequence = two out of three nucleotides in agreement.

The 3' and 5' PBS termini are fairly well conserved between MSV and DSV (Donson *et al.*, 1987), and SSV sequences showing homology to these termini were

found. The potential PBS of SSV extended from position 1179 to position 1251 (Fig. 3.6), giving a putative primer length of 72 bases. This compares with 79 bases for WDV, 80 bases for DSV, 82 bases for MSV, and 90 bases for CSMV. Other than at the 3' and 5' termini, the PBS sequence is not well conserved between Subgroup I geminiviruses.

3.3.5. Sequence Comparisons with Other Geminiviruses.

The entire nucleotide sequence of SSV was compared directly with those of MSV-S, DSV, WDV, and CSMV (Table 3.2). Acceptable comparisons were obtained between MSV, DSV, and SSV, but GAP and BESTFIT both gave unreliable results for SSV-WDV and SSV-CSMV comparisons. Accordingly, it was decided that the DNA sequences of individual ORFs of SSV should be compared (using GAP) with the corresponding ORFs of the other cereal geminiviruses (Table 3.2). To clarify the relationship of SSV to its nearest relatives, as indicated by overall percentage sequence identity, comparisons were also made between MSV and DSV.

Table 3.2. Percentage nucleotide similarity between whole genomes and open reading frames (ORFs) of SSV, MSV-S, DSV, and CSMV.

	total DNA		(+) sense ORFs		(-) sense ORFs	
Comparison	GAP	BESTFIT	V1	V2	C1	C2
MSV-DSV	62.1	67.3	64.00	63.90	66.70	70.80
SSV-DSV	59.1	64.2	55.16	63.00	60.50	67.70
SSV-MSV	61.4	64.8	60.20	67.00	65.23	66.70
SSV-CSMV	49.2[a]	58.4[b]	42.00[e]	51.70	52.50	57.80
SSV-WDV	45.0[c]	60.8[d]	40.50	46.74	48.60	56.20

Key: Total DNA sequences were aligned using the UWGCG programmes, GAP and BESTFIT. ORF DNA sequences were aligned with GAP. V1, V2, C1, and C2 correspond to the 11.16 kDa, 27.2 kDa, 34.94 kDa and 17.28 kDa ORFs, respectively, of SSV. A percentage similarity of 37% was taken to imply that sequences were unrelated (MSV-S V1 DNA aligned with CSMV C1 DNA gives a 37% similarity). a = alignment of 2854 bases and gaps. b = alignment of 1818 bases. c = alignment of 2863 bases and gaps. d = alignment of 991 bases. e = alignment of V1 of SSV with the truncated version of V1 of CSMV (Andersen *et al.*, 1988).

In terms of overall sequence homology, SSV, DSV and MSV are approximately equally closely related (Table 3.2). CSMV and WDV differ from each other and from MSV and DSV by about 50% in direct sequence comparisons (Andersen et al., 1988; MacDowell et al., 1985). In the comparisons made here, SSV showed the least homology with WDV and CSMV (Tables 3.2 and 3.3).

Comparing nucleotide sequences of individual ORFs (using GAP), SSV ORFs showed slightly greater homology to those of MSV than to those of DSV. However, MSV ORFs showed a greater homology to DSV than to SSV ORFs, with the exception of V2.

The nucleotide sequences of the ORFs shown in Table 3.2 were translated using GENEPRO and the deduced amino acid sequences were compared using GAP (Table 3.3).

Table 3.3. Percentage similarities between amino acid sequences deduced from nucleotide sequences of open reading frames of SSV, MSV-S, DSV, and CSMV.

	V1		V2		C1		C2		C'[a]	
Comparison	dir	tot	dir	tot	dir	tot	dir	tot	dir	tot
MSV-DSV	38.8	70.6	61.7	72.1	65.4	76.8	65.7	72.0	67.0	76.2
SSV-MSV	47.7	67.0	65.7	77.5	58.1	73.9	68.0	75.8	63.1	73.9
SSV-DSV	41.7	62.6	66.8	76.6	54.3	71.5	69.4	72.6	61.3	72.3
SSV-CSMV	20.0[b]	34.2[b]	46.2	54.4	36.9	55.5	54.4	58.0	41.4	55.1
SSV-WDV	27.0	53.0	33.5	44.4	36.6	59.4	51.4	59.2	42.4	56.1

Key: ORF designations are as shown in Table 3.2. C'is a fusion product of C1 and C2, after Schalk *et al.*, (1989). Percentage similarities for direct (dir) and total (tot) amino acid matches were made using the UWGCG programme, GAP. * = alignment of SSV V1 amino acid sequence with the truncated version (11.7 kDa) of the V1 (14.5 KDa) protein sequence of CSMV (Andersen *et al.*, 1988).

All ORFs showed similar degrees of overall amino acid conservation between MSV, DSV, and SSV, but in terms of direct sequence matches between viruses, the V1 ORF consistently showed the least amino acid conservation. It could be

speculated that the V1 protein is not as tightly constrained evolutionarily as the coat and replication-associated proteins. However, considering the function of the V1 protein in viral spread within the plant (Boulton *et al.*, 1989, 1990), together with the low level of direct amino acid sequence conservation of this protein between viruses (Table 3.3), the V1 protein could perhaps be one of the viral determinants of host specificity.

The highest direct amino acid sequence homology for the V1 ORF was that shared by SSV and MSV. The derived product of ORF V2 of SSV showed approximately equal amino acid similarity to both MSV and DSV V2 gene products, the homology being slightly higher than that between the MSV and DSV products. The deduced C' fusion proteins of MSV and DSV were slightly more similar to one another than the C' of SSV was to that of either DSV or MSV.

It should be noted that a very high degree of genomic conservation is expected between Subgroup I geminiviruses, since they have very small DNA genomes with few genes, some overlapping, and apparently the same mode of replication in the plant cell nucleus (Lazarowitz, 1987). Hence, while the nucleotide and derived amino acid sequence homologies shown here appear high, they are in fact significantly low for this group of viruses. This point will be reiterated in Chapters 4 and 5, where comparisons between a number of grass geminiviruses will be made.

3.3.5. Phylogenetic Analysis.

Amino acid sequences were used in the phylogenetic analysis because this data is more conservative than DNA data, allowing distant evolutionary relationships to be demonstrated. The deduced amino acid sequences of the replication-associated C1-C2 fusion protein (C') and the coat protein were chosen for phylogenetic analysis because these two proteins are common to all geminiviruses, have apparently tightly conserved functions, and have previously been used in phylogenetic analysis of geminiviruses (Howarth and Vandemark, 1989). Amino acid sequence alignments for WDV, CSMV, MSV-S, DSV and SSV were used to construct data matrices as described earlier (see Table 3.4).

The shortest phylogenetic trees that could be calculated using FITCH are shown in Figure 3.7, and agree well with the relevant parts of the C' and coat protein

trees generated by Howarth & Vandemark (1989). All FITCH runs gave the same results and the trees shown are topologically identical to trees obtained using NJTREE ver. 2.0 (not shown).

The two trees shown in Figure 3.7 differ in their relative placement of SSV, DSV and MSV-S. SSV and DSV are shown as the most closely related viruses in the coat protein tree (Fig. 3.7.a), whereas in the C' tree (Fig. 3.7.b) it is DSV and MSV-S which appear closest. The coat protein is intimately associated with vector transmission, and its evolution would thus be expected to be constrained to accommodate vector transmissibility (Howarth & Vandemark, 1989). In theory, therefore, one would expect the coat protein of SSV to be phylogenetically closer to that of MSV than to that of DSV, since both MSV and SSV are transmitted by *C. mbila* (Storey & McClean, 1930), whilst DSV is transmitted by *Nesoclutha declivata* (Julia & Dollet, 1989).

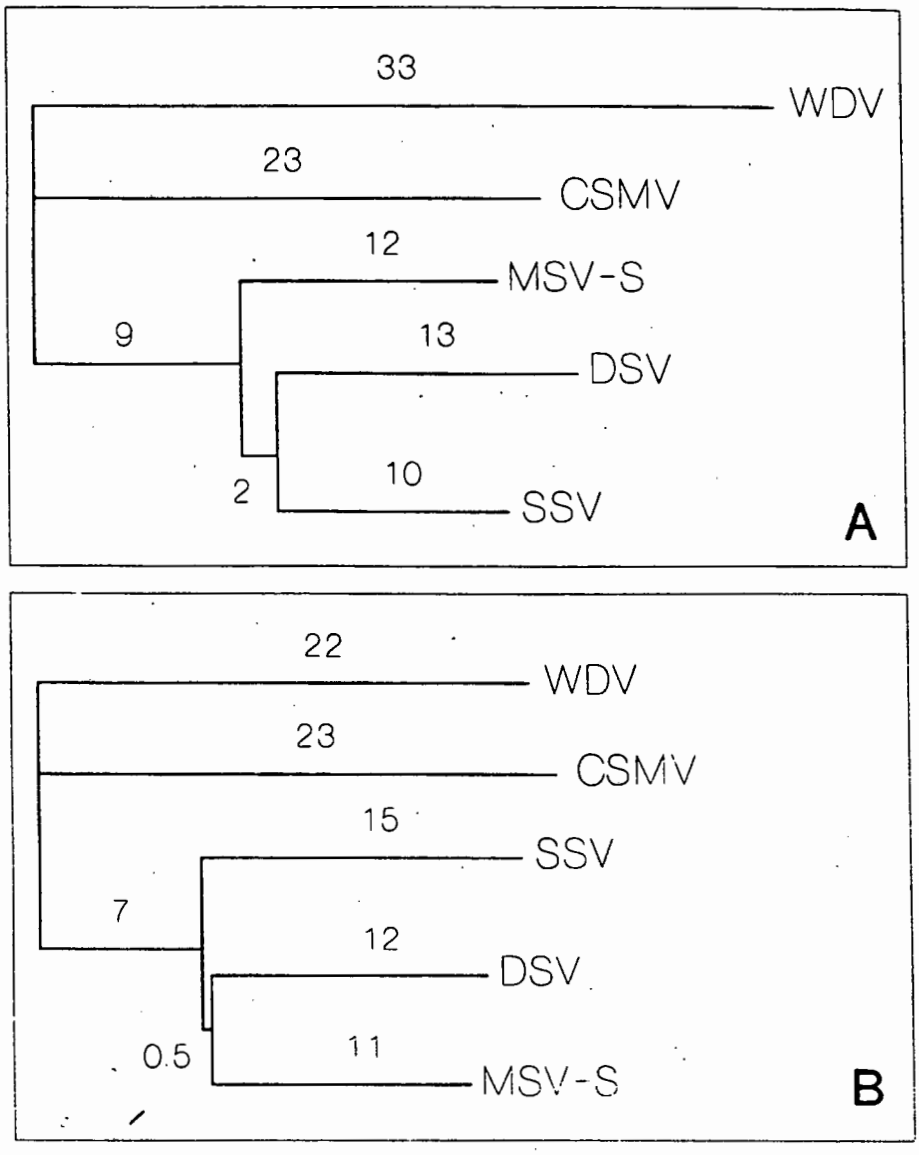
Table 3.4. Pairwise sequence difference matrices for the coat proteins and C' proteins of cereal geminiviruses.

	SSV	MSV-S	DSV	CSMV	WDV
SSV	0.0000	0.2254	0.2336	0.4564	0.5565
MSV-S	0.2611	0.0000	0.2787	0.4315	0.5492
DSV	0.2773	0.2381	0.0000	0.4564	0.5533
CSMV	0.4494	0.4242	0.4298	0.0000	0.5602
WDV	0.4387	0.4160	0.4217	0.4530	0.0000

Key: Upper right-hand side of diagonal = coat protein pairwise distance matrix. Lower left-hand side of diagonal = replication-associated C1-C2 fusion protein (C') pairwise distance matrix. Values shown were calculated using the UWGCG programme DISTANCES after alignment of sequences using LINEUP, and converted from similarity values to differences by subtraction from 1.0000.

However, this discrepancy may be an artifact of the methods used to produce the trees rather than a reflection of the true phylogenies of the individual genes.

Figure 3.7. Unrooted phylogenetic trees for coat and replication-associated proteins of SSV, MSV-S, DSV, CSMV and WDV, calculated using FITCH.



Legend: (A) Coat protein tree. (B) C' (putative C1-C2 fusion) protein tree. Horizontal branch lengths are proportional to genetic distance; vertical lengths are arbitrary. Figures along horizontal branches are internodal percent sequence divergences calculated by the programme. The tree in (A) is the best of 23 examined, with an average standard deviation of branch length of 3.05%. The tree in (B) is the best of 27, with an average standard deviation of 0.63%.

The same data sets used in KITSCH gave a coat protein tree in which SSV and MSV-S were closest relatives (not shown), with a similar standard deviation as in the FITCH tree for the coat protein (Fig. 3.7.a). Thus, the only safe conclusions to be drawn are that SSV, MSV-S and DSV form a group of viruses which are equally distantly related to one another, and quite distinct from WDV and CSMV.

The relative positions of CSMV and SSV in the trees shown in Figure 3.7 differ from their positioning in a tree previously derived for the same viruses from restriction endonuclease cleavage maps (Hughes et al., 1990; Chapter 2). This is because the use of restriction maps in generating phylogenetic trees is not very reliable for sequences differing by more than 20% since the number of shared sites is then too small for accurate phylogenetic analysis (Gibbs and Fenner, 1984; see also Chapter 2). The sequence-derived phylogenies are therefore more likely to reflect actual phylogenies.

3.4. CONCLUSIONS.

Nucleotide sequencing of SSV has clarified the taxonomic status of SSV with repect to other Subgroup I geminiviruses. The three sequenced isolates of MSV available to date (MSV-N, MSV-K, and MSV-S) are approximately 98% identical in nucleotide sequence (Lazarowitz, 1988). DSV is the most closely related geminivirus to MSV, and is now generally recognized as a distinct Subgroup I geminivirus (Accotto *et al.*, 1989; G. P. Martelli, pers. comm. to E. P. Rybicki). SSV has a lower overall percentage nucleotide sequence identity with MSV than has DSV. Thus, the data presented here confirm that SSV is a distinct geminivirus, as suggested by McClean (1947) and Hughes *et al.* (1990) (see also Chapter 2), and not a strain of MSV. Further, nucleotide and derived amino acid sequence comparisons indicate that SSV, DSV, and MSV probably form a group of equidistantly related viruses, only distantly related to WDV and CSMV.

FOOTNOTE : The nucleotide sequence data reported in this chapter have been submitted to the GenBank nucleotide sequence database and have been assigned the accession number M33829.

CHAPTER 4

NUCLEIC ACID TYPING OF STREAK VIRUS ISOLATES.

CONTENTS

CHAPTER 4.
NUCLEIC ACID TYPING OF STREAK VIRUS ISOLATES.

SUMMARY

Streak virus isolates from a variety of graminaceous hosts were compared by restriction fragment length polymorphism (RFLP) analysis, restriction endonuclease mapping, and by the extent of DNA cross-hybridization between virus isolates.

Viral RF-DNA was multiply cloned from a single streak-diseased maize plant and a single streak-diseased *Setaria* plant. RFLPs were generated for these clones, using restriction enzymes with 4-base recognition sites. In each infected plant, minor sequence variation between individual virus genomes was found.

In DNA cross-hybridization experiments, using both cloned and uncloned virus DNA as target, the greatest degree of cross-hybridization was shown between maize MSV, the *Setaria* streak virus isolate, and an *Eleusine* sp. streak isolate (also found in wheat), followed by that between the Natal and Mauritius sugarcane streak isolates. The *Panicum* sp. streak virus isolate hybridized with MSV and with the *Setaria* isolate - but only marginally. SSV-N, DSV, and MSV were mutually non-cross-hybridizing.

Restriction maps were constructed for four maize isolates of maize streak virus, three grass streak isolates (from *Panicum* sp., *Setaria* sp., and *Eleusine* sp.), and a Mauritian sugarcane streak isolate. Maps of maize isolates of MSV were all highly similar to one another. The maps of the *Setaria* and wheat/*Eleusine* isolates bore a limited resemblance to each other and to maize MSV maps. Maps of the two sugarcane streak isolates and the *Panicum* streak isolate were highly dissimilar to each other and to all other mapped Subgroup I geminiviruses. Restriction maps constructed in this study and maps derived from published sequences of Subgroup I geminiviruses were used to construct a phylogenetic tree for this subgroup.

Results from hybridization studies, restriction mapping, and from phylogenetic analysis indicated that the *Setaria* and wheat/*Eleusine* streak isolates could be considered to be strains of MSV. The Mauritian sugarcane streak isolate appeared to be a strain of SSV, but was very distantly related to the Natal strain. The *Panicum* streak isolate appeared to be a distinct virus.

4.1. INTRODUCTION.

The early literature on maize streak virus indicated the existence of a number of geminiviruses distinct from MSV, as well as of isolates that were considered to be strains of MSV (Storey & McClean, 1930; McClean, 1947). Later, these different viruses were designated "host-adapted" strains of MSV (Bock *et al.*, 1974). Once established, this classification trend has continued to the present (Pinner *et al.*, 1988; Pinner & Markham, 1990b), so that any geminivirus now found in an African grass is automatically designated a strain of MSV, even when found to be only distantly related serologically to the maize strain of MSV or to "strains" of MSV from other grasses.

Having found that SSV constitutes a distinct Subgroup I geminivirus (Hughes *et al.*, 1990a; 1990b; Chapters 2 and 3), whereas MSV isolates from widely separated geographical regions have remarkably similar nucleic acid characteristics (Hughes *et al.*, 1990c; Chapter 2), it was considered worthwhile to characterize by the same means a selection of streak virus isolates from different host species, particularly where the streak symptoms differed from the typical severe streak symptoms of MSV as found in maize.

In this Chapter is reported the molecular typing of streak virus isolates by means of DNA cross-hybridization, restriction endonuclease mapping, and partial DNA sequencing. A phylogenetic tree of Subgroup I geminiviruses was constructed from restriction map data. In addition, a survey was made of virus genome variability within naturally infected single plants of a *Setaria* sp. and maize, using RFLP analysis.

4.2. MATERIALS AND METHODS.

4.2.1. Streak Virus Isolates Used And Their Origins.

The origins of isolates used in this study are shown in Table 4.1. Streak-infected *Digitaria setigera* was admitted into South Africa with the permission of the Directorate of Plant and Seed Control, Department of Agriculture (permit no. 14/14/A/16/16). Streak-infected sugarcane, cultivar H 443098, was similarly admitted from Mauritius (permit no. 14/2/2/1(9/19/22)).

Table 4.1. Origins of streak-infected plants used in the studies reported in Chapter 4.

Isolate	Source plant	Geographical origin	Obtained from
MSV-Koe	maize	Koedoespoort, N. Tvl	M B von Wechmar[a]
MSV-Kom	maize	Komatipoort, E. Tvl	M B von Wechmar[a]
MSV-PE	maize	Port Elizabeth, E. Cape	M B von Wechmar[a]
MSV-RSE	maize	Riviersonderend, Cape	J Laubscher[a]
MSV-SW	maize	Otjiwarongo, Namibia	H Hoffsman[a]
SetSV	*Setaria* sp.	Mt. Edgecombe, Natal	K Harborne[b]
WESV	wheat	Ventersdorp, Tvl	M B von Wechmar[a]
WESV	*Eleusine* sp.	Howick, Natal	M B von Wechmar[a]
PanSV	*Panicum* sp.	Karino, N. E. Tvl	M B von Wechmar[a]
SSV-N	sugarcane[1]	Mt. Edgecombe, Natal	K Harborne[b]
SSV-M	sugarcane[2]	Reduit, Mauritius	L J-C Autrey[c]
DSV	*Digitaria setigera*	Vanuatu, New Hebrides	G Boccardo[d]

Key: Tvl = Transvaal. OFS = Orange Free State (the OFS, Transvaal, Cape, and Natal are provinces of South Africa).
1 = cv. Uba. 2 = cv. H443098. a = University of Cape Town.
b = South African Sugar Association Experimental Station, Mt. Edgecombe.
c = Mauritius Sugar Industry Research Institute, Reduit, Mauritius.
d = IFA del CNR, Torino, Italy.

The streak virus isolated out of wheat from Ventersdorp, Transvaal proved to be identical to the virus isolated previously out of an *Eleusine* sp. from Howick, Natal (unpublished data). In these studies, the wheat isolate was used for extraction of viral DNA, since far more viral DNA was present in the wheat sample than in *Eleusine*. With the exception of the *Setaria* sp. and Riviersonderend maize isolates, streak-infected isolates were maintained as detailed in A.1.

4.2.2. Cloning of Virus RF-DNA.

Virus DNAs of the following isolates were cloned: MSV-SW, MSV-RSE, SetSV, WESV, PanSV, DSV, SSV-N, and SSV-M (see table 4.1). Total plant DNA was extracted from streak-infected plants (A.3) and where the concentration of viral RF-DNA was low, the RF-DNA was purified from the total plant DNA (A.4, A.11.b). RF-DNA from each virus was cleaved with a panel of restriction enzymes (C.6) in

single-enzyme digests (A.5) to identify those enzymes having unique cleavage sites on the viral genome. Such an enzyme was then used to linearize the viral DNA to clone into a suitable plasmid vector (see Table 4.2).

Table 4.2. Construction details of cloned streak virus isolates used in studies detailed in Chapter 4.

Isolate	linearized using	cloning vector	cloning site used	clone designation
MSV-Koe	*Bam* HI	pUC 19	*Bam* HI	pKoe100
MSV-Kom	*Bam* HI	pUC 19	*Bam* HI	pKom100
MSV-PE	*Bgl* II	pEcoR251	*Bgl* II	pBC200
MSV-RSE	*Bam* HI	pUC 18	*Bam* HI	pRSE100
MSV-SW	*Bgl* II	pEcoR251	*Bgl* II	pSW100
SetSV	*Bam* HI	pUC 18	*Bam* HI	pSET100
WESV	*Bgl* I*	pUC 19	*Sma* I	pWES100
PanSV	*Bam* HI	pUC 18	*Bam* HI	pPS100
DSV	*Hind* III	pUC 18	*Hind* III	pDSV100
SSV-N	*Pst* I	Bluescript SK	*Pst* I	pSS100
SSV-M	*Sal* I	pUC 18	*Sal* I	pSSM100
SSV-M	*Bgl* II	pUC 18	*Bam* HI	pSSM200

Key: pBC200 - cloned by B A Clarke (MSc. thesis, 1987).
pKoe100 - cloned and mapped by M Babaya (M. Sc. thesis, 1989).
pKom100 - cloned and mapped by E Edge (Hons. thesis, 1989).
*** cut ends filled in using klenow enzyme and dNTPs to create blunt ends, as described in Appendix B.1.**

Cloning and transformation methods were as detailed in A.8 - A.10. Recombinants containing viral inserts were detected by probing (A.13.d) colony blots (A.13.a) with gel-purified insert DNA of cloned MSV-SW (A.7) labeled with digoxygenin-dUTP (A.13.c). For viruses showing little or no hybridization with MSV-SW, virus inserts were identified by restriction with suitable enzymes, checking the digestion products to see if the cloned insert was of the expected full-length genome size, with the same number and type of restriction sites as the uncloned virus. Recombinants with putative complete genome viral inserts were digoxygenin-labelled (A.13.c) and used to probe (A.13.d) Southern-blotted total

DNA extracts of infected and healthy plants (A.13.b) to confirm the viral origin of the inserts.

4.2.3. Clonal Analysis of Virus from a Single Plant.

Field-infected isolates - maize from Riviersonderend and a *Setaria* sp. from Mt. Edgecombe (Table 4.1) - were used. For each isolate, total plant DNA was isolated (A.3), digested with *Bam* HI and ligated into *Bam* HI-linearized pUC 18 (A.8). The ligation mixes were used to transform *E. coli* LK112 (A.9, A.10). Transformants with inserts (white colonies on X-gal plates, C.3) were colony-blotted (A.13.a) and probed with cloned MSV-SW insert DNA, as detailed in 4.2.2. Minipreps were made from positively staining colonies (A.11.c).

Each MSV-RSE miniprep was digested with *Hae* III and with *Rsa* I, restriction enzymes with a 4-base recognition site. The *Setaria* virus clones were digested not only with *Hae* III and *Rsa* I, but also with *Hind* III and *Bgl* I (6-base recognition sites). Digestions and electrophoresis were performed as in A.5 and A.6. Agarose gels of electrophoretically separated DNA digests were photographed and clones with identical restriction patterns grouped together. Clones giving unusual restriction patterns were re-cut to eliminate the possibility of incomplete digestion as the cause of the pattern difference. Representatives from each pattern grouping were subjected to further restriction digestion to check which inserts comprised complete viral genomes. A full restriction map was made of one clone from the most common restriction pattern grouping in each plant isolate (see 4.2.5).

4.2.4. DNA Cross-Hybridization Between Virus Isolates.

4.2.4.1. Preparation of Probes.

Cross-hybridization was performed using the following labelled recombinant plasmids as probes: pRSE100, pSET100, pPS100, pDSV100, pSS100, and pSSM100 (see Table 4.2). Since the viral insert could not be reliably separated from vector DNA, pUC 19 and Bluescript SK were used as probe controls. pUC 19 and Bluescript SK target controls were intrinsic to the blots of the cloned virus isolates. One µg of DNA was used for each probe-labelling reaction. Plasmids were digested to completion (A.5) with *Bam* HI (for pRSE100, pSET100, and pPS100), *Hind* III (for pDSV100, pUC 19, and SK), *Pst* I (for pSS100), or with *Kpn* I and *Pst* I (for pWES100)

to liberate the viral insert from the vector as well as to linearize the DNA for labelling. pSSM100 was cut with *Kpn* I and *Bgl* I, which left a portion of vector DNA (~300 base pairs) attached to the viral insert. Probe DNAs were labelled and stored as described in A.13.c.

4.2.4.2. Southern Blotting of Cloned and Uncloned Viral DNAs.

In the original experiment performed, eight replicates were made of each blot detailed below. The following uncloned isolates were used: MSV-RSE, SetSV, PanSV, DSV, SSV-N, and SSV-M (see Table 4.1).

For each replicate, samples of extracts of total plant DNA containing uncloned virus were separated electrophoretically (A.6) and Southern blotted (A.13.b). Since quantitation of viral DNA against the plant DNA background was impossible, trial gels were run to adjust the quantity loaded of each preparation, such that an approximately equal amount of viral RF-DNA of each isolate was used.

In a separate experiment, a sample of uncloned WESV (Table 4.1) as well as the samples already mentioned above, were subjected to gel electrophoresis (A.6) and Southern-blotting (A.13.b). In addition, samples of uncloned WESV were electrophoresed separately (8 lanes), Southern-blotted as above, and cut into strips (one lane of DNA per strip).

With the exception of pWES100, the cloned isolates used as target DNAs were the same ones used for making probe DNAs. Since virus and vector DNAs were approximately the same size, each cloned isolate was digested in such a manner that upon electrophoretic separation, bands of viral DNA could not be confused with vector bands. Accordingly, the following digests were made (A.5): pRSE100 - *Bam* HI/*Xho* I, pSET100 - *Bam* HI/*Eco* RI, pPS100 - *Bam* HI/*Bgl* I, pDSV100 - *Hind* III / *Pst* I, pSS100 - *Sal* I/*Pst* I, pSSM100 - *Kpn* I/*Sal* I. DNAs used were quantitated prior to digestion, as in A.11.d. A single "bulk" digest was made for each DNA being digested, containing enough DNA to load replicate lanes (200 ng DNA per lane) in six gels. After electrophoresis (A.6), the six replicate gels were Southern-blotted (A.13.b).

4.2.4.3. Hybridization and Detection.

Hybridization and detection were carried out as described in A.13.d. Blots were hybridized overnight with 40 ng of labelled probe DNA per ml of hybridization fluid used.

In the first experiment, each DNA probe (4.2.6.1), excepting pWES100, was used to probe one replicate blot containing uncloned target DNA and one blot containing cloned target DNA (4.2.6.2).

In the second experiment, the pWES100 probe was used against the blot containing all uncloned target DNAs, including WESV. The strip-blots containing uncloned WESV were probed individually with one of pRSE100, pDSV100, pSET100, pPS100, pSS100, pSSM100, or the SK and pUC control probes.

4.2.5. Restriction Endonuclease Mapping of Cloned Virus Isolates and Partial Sequencing of Cloned Viruses.

This was performed as described in A.5 and A.6, with an example of the method and logic used in map construction being shown in Appendix A.14.

The following clones (see Table 4.2) were mapped: pSW100, pRSE100, pWES100, pSET100, pPS100, pSS100, pSSM100. The recombinant pBC200, previously cloned and mapped by B. Clarke (Clarke, 1987; Clarke *et al.*, 1989) was remapped. The following enzymes were used in mapping each isolate; *Apa* I, *Bam* HI, *Bgl* I, *Bgl* II, *Cla* I, *Eco* RI, *Hind* III, *Kpn* I, *Pst* I, *Pvu* II, *Sac* I, *Sal* I, and *Xho* I.

Complete nucleotide sequences of the following Subgroup I geminiviruses were obtained from the Genbank database: MSV-K (Howell, 1984; 1985), MSV-N (Mullineaux *et al.*, 1985), MSV-S (Lazarowitz, 1988), WDV (MacDonald *et al.*, 1985), CSMV (Andersen *et al.*, 1988), and DSV (Donson *et al.*, 1987). Restriction maps were generated from these published sequences using GENEPRO version 4.2 software (Riverside Scientific, Seattle, Wa, USA).

Maps of sugarcane and grass isolates were oriented with respect to maps of maize isolates by sequencing from 100 to 400 bases of the viral insert of the recombinant in the forward and/or reverse directions (as described in B.2), sequencing inwards from the ends of the viral insert DNA. The sequences so obtained were aligned with the equivalent sequences of MSV, DSV, and SSV-N, and

percentage sequence identities were calculated, using GENEPRO ver. 4.2 (Riverside Scientific, Seattle, USA).

4.2.6. Construction of Phylogenetic Trees from Restriction Map Data.

DNA sequence divergence and evolutionary distances were estimated from aligned restriction maps, using the mathematical models of Aoki *et al.* (1981), and Nei & Li (1979) (see Chapter 1, section 1.4.3.2).

Restriction sites were regarded as common or shared sites if their mapped positions coincided within an error of 2% of total genome length.

The average proportion of shared sites between isolates (S) was calculated as:

$$S = 2\, n_{xy} / dn_x + dn_y$$ (Aoki et al., 1981; see Chapters 1 and 2).

From the S values so calculated, a table of estimated distance (p) values was drawn up, calculated using the equation:

$$\pi = (-\ln S) / r$$ (Nei & Li, 1979; see Chapters 1 and 2).

The π values obtained were inserted into the Fitch and Margoliash (1967) algorithm to produce a rooted phylogenetic tree using the global rearrangement (G) option of the FITCH programme in PHYLIP (Phylogeny Inference Package) version 3.1 (J. Felsenstein, Department of Genetics, University of Washington, Seattle).

As checks, the same calculated values were used in the programme NJTREE, which uses the neighbour joining method of Saitou & Nei (1987), as modified by Studier & Keppler (1988). In addition, subsets of the data were used in both NJTREE and FITCH. The global rearrangement (G), jumble (J), shuffle (S), and outgroup (O) options of the FITCH programme were used on the data subsets.

Figure 4.1. Diagrammatic representation of the *Hae* III and *Rsa* I restriction patterns obtained with MSV-RSE and SetSV clones.

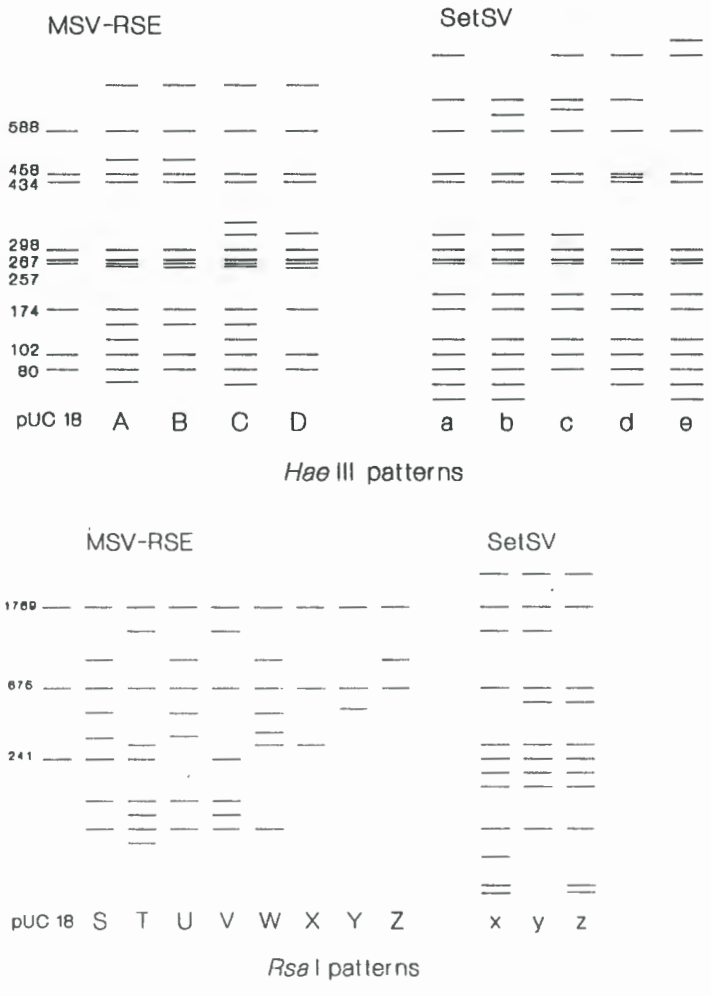
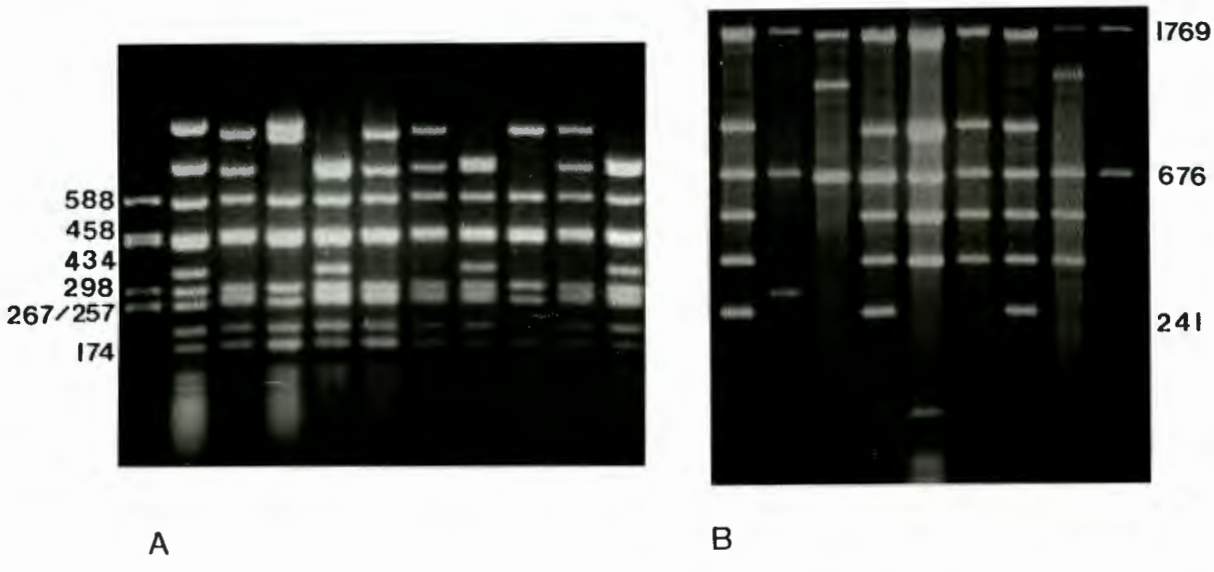


Figure 4.2. Examples of the restriction fragments observed in *Rsa* I and *Hae* III digests of MSV-RSE and SetSV clones.



Legend: EthBr-stained 1.5% TBE agarose gels. (a). MSV-RSE clones digested with *Hae* III. (b). SetSV clones digested with *Rsa* I.

4.3. RESULTS.

4.3.1. Clonal Analysis of Virus from a Single Plant.

4.3.1.1. Maize isolate from Riviersonderend.

Seventy-eight recombinants hybridizing positively with the MSV-SW probe were chosen. Of these, 46 contained whole viral genome inserts. Recombinants containing sub-genomic inserts were not used in the analysis. The *Rsa* I and *Hae* III restriction patterns obtained are depicted in Figure 4.1 and the data summarized in Table 4.3. An example of an agarose gel with *Hae* III-digested miniprep DNA is shown in Figure 4.2.

Table 4.3. Summary of the restriction pattern groupings obtained from *Hae* III and *Rsa* I digests of individually-cloned MSV-RSE genomes.

Hae III group	Rsa I group	No. of clones
A	U	13
A	V	3
A	W	2
A	Z	1
B	S	5
B	V	4
B	X	2
B	Y	1
C	T	12
D	T	1

The predominant genome variant was obtained in both orientations, being cloned more frequently in one orientation - that corresponding to pRSE100 - (32 clones - *Hae* III groups A and B; *Rsa* I groups T, U, V, W, X, Y, Z,) than the other (13 clones - *Hae* III groups C and D; *Rsa* I group S).

There was only minor variation between clones. In the less frequent orientation, all clones gave the same *Rsa* I cutting patterns, and only one clone showed variation from the predominant *Hae* III pattern. In the more frequent orientation, two *Hae* III patterns and 7 *Rsa* I patterns were encountered. However, since both *Hae* III pattern groups shared members of *Rsa* I group V, it was deduced that *Hae* III groups A and B belonged to the same major genome variant. In addition,

of the 7 *Rsa* I pattern groups, there appeared to be a loss of an internal *Rsa* I site in the vector in 5 of the groups (U, W, X, Y, Z), totalling 19 of the 46 clones analysed.

4.3.1.2. *Setaria* sp. isolate from Mt. Edgecombe.

The MSV-SW probe hybridized positively with 42 recombinants. Of these, 30 contained whole viral genome inserts. An example of *Rsa* I-digested miniprep DNA is shown in Figure 4.2. The *Rsa* I, and *Hae* III restriction patterns obtained are shown diagrammatically in Figure 4.1 and the data summarized in Table 4.4.

Table 4.4. Summary of the restriction pattern groupings obtained from *Hae* III and *Rsa* I digests of individually-cloned SetSV genomes.

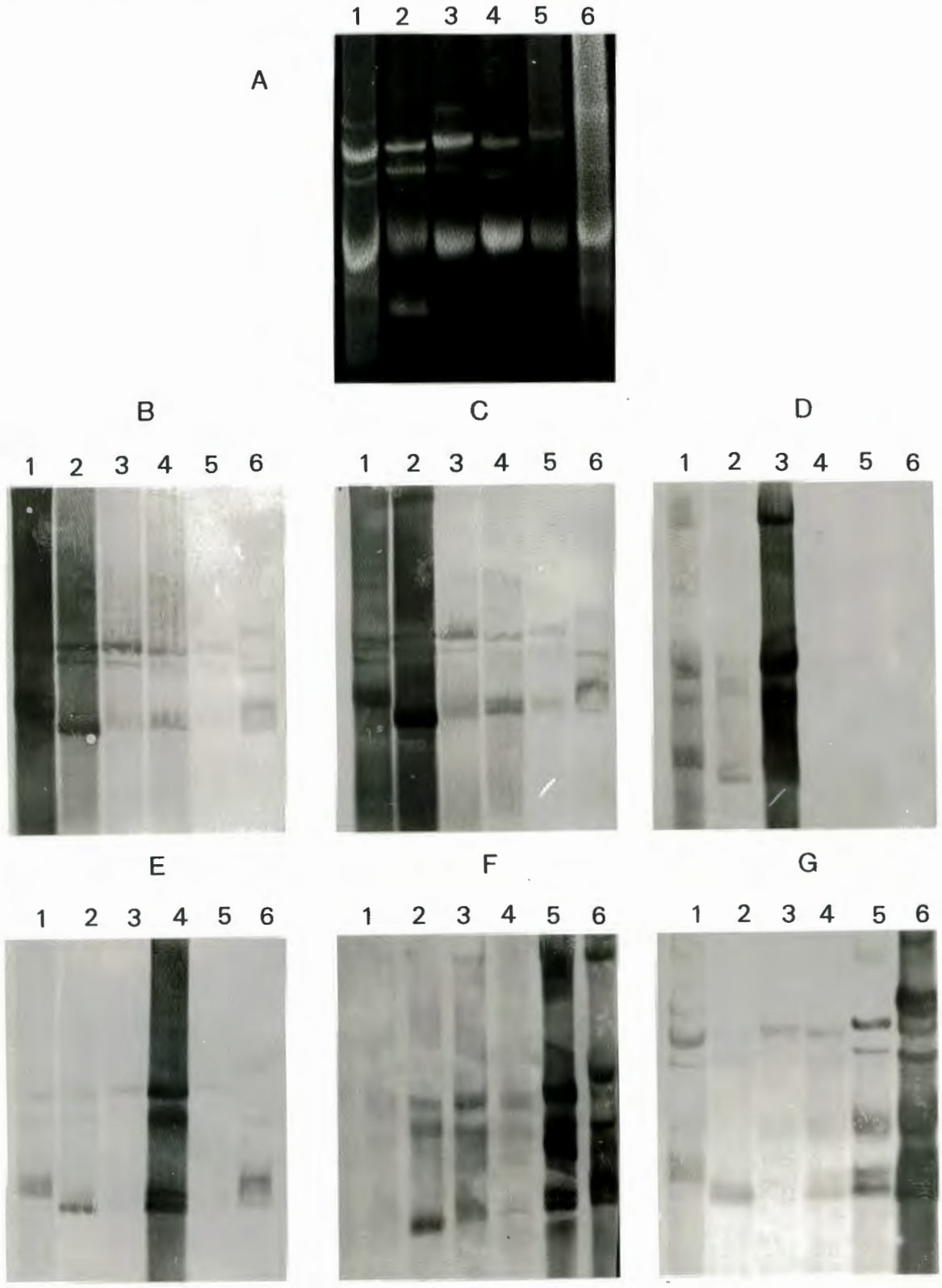
Hind III / Bgl I group	*Hae* III group	*Rsa* I group	Number of clones
A	a	x	6
A	b	x	10
A	c	y	1
A'	d	z	11
A'	e	z	1

Of the 30 complete genomes cloned, 29 belonged to the same major variant, as deduced from the *Hind* III and *Bgl* I restriction patterns obtained. This variant was obtained in both orientations: orientation A - 17 clones, corresponding to pSET100, and orientation A' - 12 clones.

In orientation A, there were 2 dominant *Hae* III pattern variations (a, b) and one dominant *Rsa* I variation (group x). One clone in this grouping gave unique restriction patterns with both *Hae* III and *Rsa* I. In orientation A', only one *Rsa* I pattern was obtained (group z), and 2 *Hae* III pattern variations (d, e), with *Hae* III pattern 'e' being seen in one clone only. The variation appeared to be due to the loss of a *Hae* III site.

One clone obtained during this survey had an insert which included a whole genome plus an additional 1000 bp of viral DNA. This was determined by restriction analysis combined with hybridization of restricted DNA from this clone with labeled MSV-SW probe DNA. This clone was not analysed further.

Figure 4.3. Extent of DNA cross-hybridization between Subgroup I geminivirus isolates, using cloned viral genome probes against total plant DNA containing uncloned viral DNA.



Legend: (a) EthBr-stained 0.8% TBE agarose gel. Lane 1; MSV-RSE. lane 2; SetSV. lane 3; PanSV. lane 4; DSV. lane 5; SSV-N. lane 6; SSV-M. marker lane not shown. (b) - (g) Southern blots of gels, lanes as in (a), probed with digoxygenin-labelled probe DNAs, as follows: (b) MSV-RSE; (c) SetSV; (d) PanSV; (e) DSV; (f) SSV-N; (g) SSV-M.

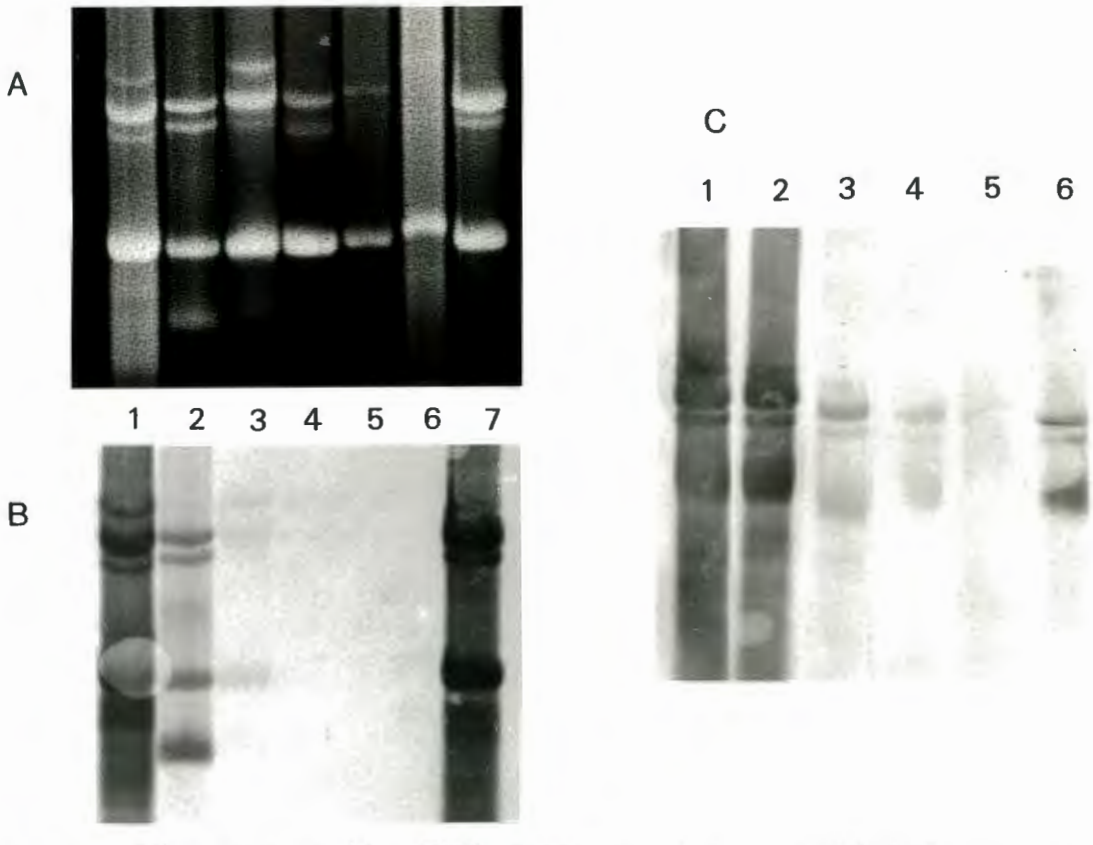
4.3.2. DNA Cross-Hybridization Between Virus Isolates.

4.3.2.1. Uncloned virus DNA.

The extent of DNA cross-hybridization between virus isolates is shown in Figures 4.3 and 4.4, and summarized in Table 4.5.

Each viral DNA isolate hybridized best to itself, as judged by staining intensity on the blots. The greatest extent of cross-hybridization between isolates was seen between MSV-RSE and the SetSV and WESV isolates, and between SSV-N and SSV-M (Figs. 4.3, 4.4). The PanSV isolate showed limited hybridization with MSV-RSE, but very much less than that shown between MSV-RSE and SetSV or WESV. Under the conditions used, MSV-RSE did not cross-hybridize with DSV or SSV-N, and cross-hybridized with SSV-M only marginally. DSV did not cross-hybridize with either MSV-RSE or the SSVs.

Figure 4.4. Extent of DNA cross-hybridization between WESV and other Subgroup I geminivirus isolates, using cloned viral genome probes against total plant DNA containing uncloned viral DNA.



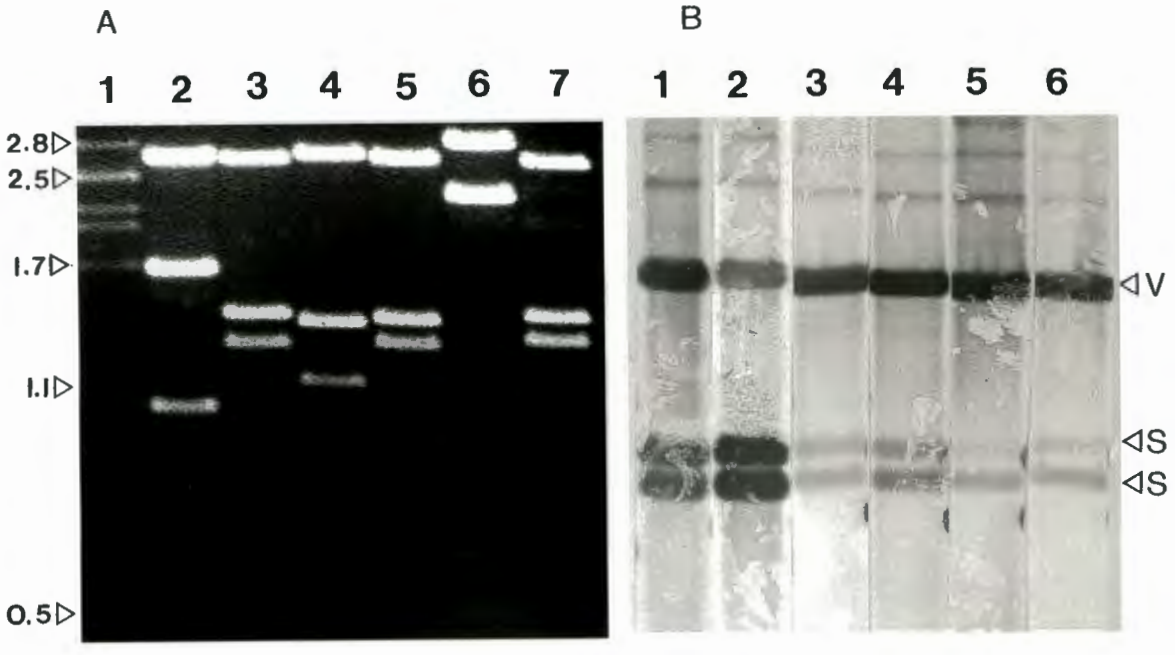
Legend: (a) EthBr-stained 0.8% TBE agarose gel. Lane 1; MSV-RSE. lane 2; SetSV. lane 3; PanSV. lane 4; DSV. lane 5; SSV-N. lane 6; SSV-M. lane 7; WESV. *Pst* I—digested lambda DNA markers not shown. (b) Southern blot of (a), probed with digoxygenin-labelled pWES100. (c) Southern blots of WESV DNA, probed as follows: lane 1, MSV-RSE; lane 2, SetSV; lane 3, PanSV; lane 4, DSV; lane 5, SSV-N; lane 6, SSV-M.

4.3.2.2. Cloned virus DNA.

The extent of DNA cross-hybridization between cloned virus isolates is summarized in Table 4.6.

An example of such cross-hybridization is shown in Figure 4.5. The gels were somewhat overloaded for the purposes of this experiment, but the results reflected those obtained for the uncloned isolates, indicating that the levels of cross-hybridization shown between virus isolates were not due to infection of the source plant by two or more different viruses.

Figure 4.5. Extent of DNA cross-hybridization between Subgroup I geminivirus isolates, using cloned viral genome probes against cloned target viral DNA.



Legend: (a) EthBr-stained 0.8% TBE agarose gel. lane 1; *Pst* I-digested lambda DNA markers. Lane 2; pRSE100. lane 3; pSET100. lane 4; pPS100. lane 5; pDSV100. lane 6; pSS100. lane 7; pSSM100. DNA in lanes 1 - 6 was digested as described in the text. (b). Lane 3 of (a), Southern blotted and probed with digoxygenin-labelled probe DNAs, as follows: lane 1; MSV-RSE. lane 2; SetSV. Lane 3; PanSV. Lane 4; DSV. Lane 5; SSV-N. Lane 6; SSV-M. V = vector DNA. S = cloned SetSV insert DNA.

Table 4.5. Extent of DNA cross-hybridization observed between digoxygenin-labelled viral DNA probes and uncloned target viral DNAs.

		Target DNA					
Probe DNA x	MSV-RSE	SetSV	PanSV	DSV	SSV-N	SSV-M	WESV
MSV-RSE	5	4	3	3	1	2	4
SetSV	4	5	2	2	1-2	2	3-4
PanSV	3	2	5	0-1	1	1	2
DSV	2	2	1	5	1	2	1
SSV-N	1-2	3	3	2	5	4	0-1
SSV-M	2	1-2	1	2	3-4	5	1-2
WESV	4	3	1-2	0-1	0	0-1	5

Key: Degree of reaction; 5 = most intense reaction (homologous DNA). 4 = almost as intense as 5. 3 = intermediate intensity. 2 = faint reaction. 1 = barely visible reaction. 0 = no observed reaction.

Table 4.6. Extent of DNA cross-hybridization observed between digoxygenin-labelled viral DNA probes and cloned target viral DNAs.

		Probe DNA				
Target DNA	MSV-RSE	SetSV	PanSV	DSV	SSV-N	SSV-M
MSV-RSE upper band	5	4	2	2	1	2
lowerband	5	4	1	1	1	0-1
SetSV upper band	4	5	2	2-3	1	1-2
lowerband	4	5	2	2-3	2	2
PanSV single band	3	3	5	2	2-3	3
DSV upper band	0	1	0	5	0	0
lowerband	2	2	1	5	1-2	2
SSV-N upper band	0	1-2	1-2	1-2	5	4
lowerband	0	0	2	1	5	4
SSV-M upper band	1	1	1	1	4	5
lowerband	3	2-3	3-4	4	4	5

Key: Degree of reaction; 5 = most intense reaction (homologous DNA). 4 = almost as intense as 5. 3 = intermediate intensity. 2 = faint reaction. 1 = barely visible reaction. 0 = no observed reaction.

4.3.3. Restriction Endonuclease Maps of Cloned Virus Isolates.

Aligned restriction maps are shown in Figures 4.6 and 4.7. The maize isolate maps were all markedly similar. A "consensus MSV" map, showing sites that are conserved between five out of six isolates, is shown in Figures 4.6 and 4.7. As well as the overall consensus, however, each maize isolate has restriction features that distinguish it from other isolates. Thus, MSV-S has a unique *Pst* I site; MSV-K lacks an *Xho* I site, but has four *Hind* III sites; MSV-N has two *Kpn* I sites and no *Sal* I; MSV-SW lacks *Sal* I and *Xho* I; MSV-Kom has four *Sac* I sites; and MSV-RSE lacks a *Kpn* I site.

Of all the non-maize virus isolates, the maps of the *Setaria* and wheat/*Eleusine* virus isolates show the greatest similarity to the consensus MSV map, although they show only limited similarity to one another (Fig. 4.7). The maps of the other isolates show no correspondence with the consensus MSV map and little or no correspondence with one another. Thus, SSV-N and SSV-M apparently share only 4 sites with one another, and 2 sites with DSV. PanSV and WDV have no sites in common with any other virus, and there are only 2 shared sites between DSV and CSMV (see Fig. 4.7).

To summarize, all maize isolates show similar restriction map patterns, whereas maps of the non-maize isolates are dissimilar to the consensus MSV map and to one another, with the exception of the *Setaria* and wheat/*Eleusine* isolates which conform with the consensus MSV map to a large extent, though they each bear only a limited resemblance to other isolates.

Figure 4.6. (shown on p. 79). Aligned restriction maps of MSV isolates from maize and from selected grasses.

Figure 4.7. (shown on p. 80). Aligned restriction maps of selected Subgroup I geminiviruses.

Combined Legend to Figures 4.6 and 4.7: MSV-S = South African isolate (Lazarowitz, 1988). MSV-K = Kenyan isolate (Howell, 1984). MSV-N = Nigerian isolate (Mullineaux *et al.*, 1984). WDV = wheat dwarf virus (MacDowell *et al.*, 1985); CSMV = chloris striate mosaic virus (Andersen *et al.*, 1988); DSV = digitaria streak virus (Donson *et al.*, 1987). Other map designations are as shown in Table 4.1. The composite "consensus map" at the bottom shows those restriction sites conserved in 5 out of 6 MSV isolates. Abbreviations for restriction enzyme sites are as follows: A = *Apa* I; B = *Bam* HI; Bl = *Bgl* I; Bg = *Bgl* II; C = *Cla* I; E = *Eco* RI; H = *Hind* III; K = *Kpn* I; P = *Pvu* II; Ps = *Pst* I; S = *Sal* I; Sa = *Sac* I; X = *Xho* I.

Figure 4.6.

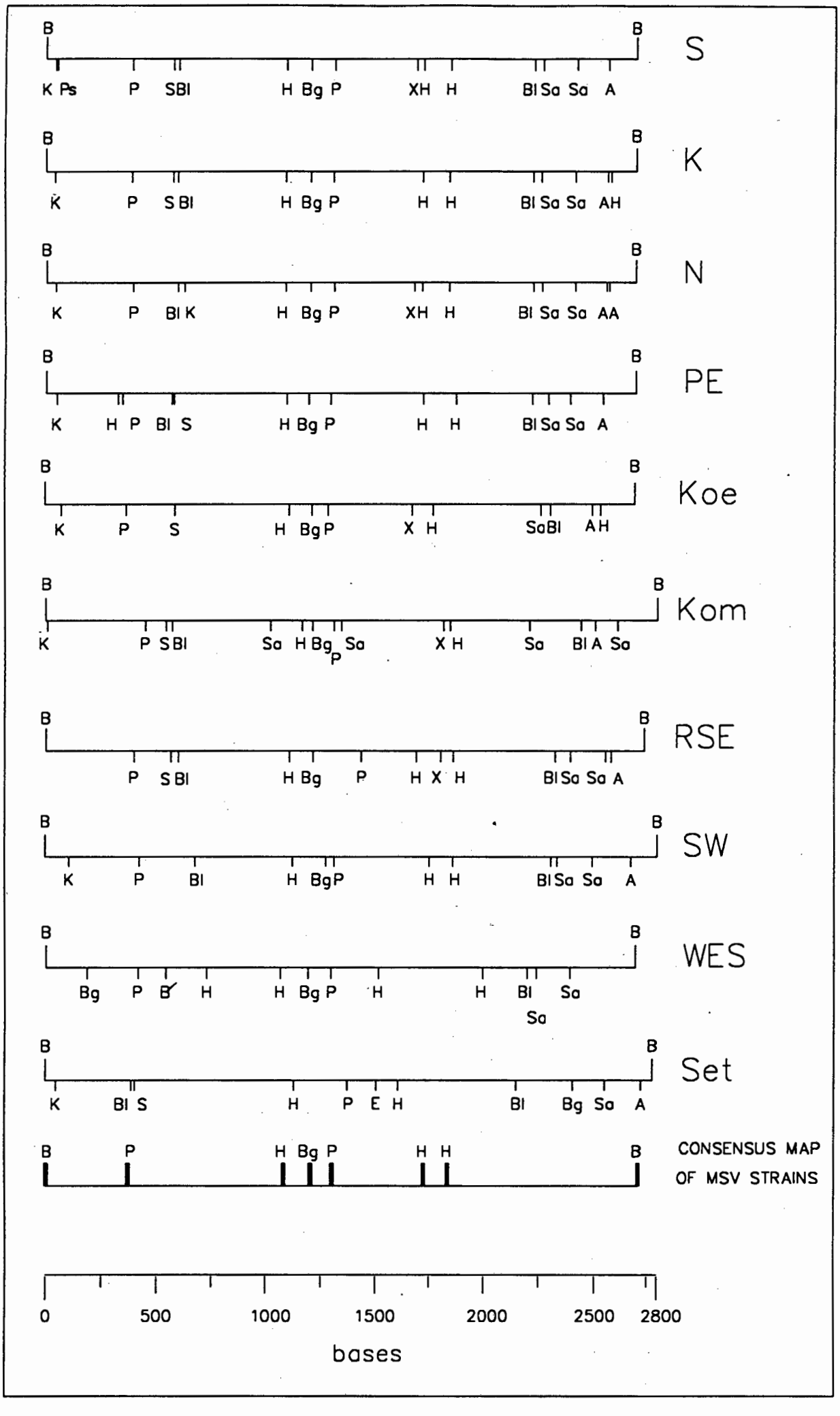
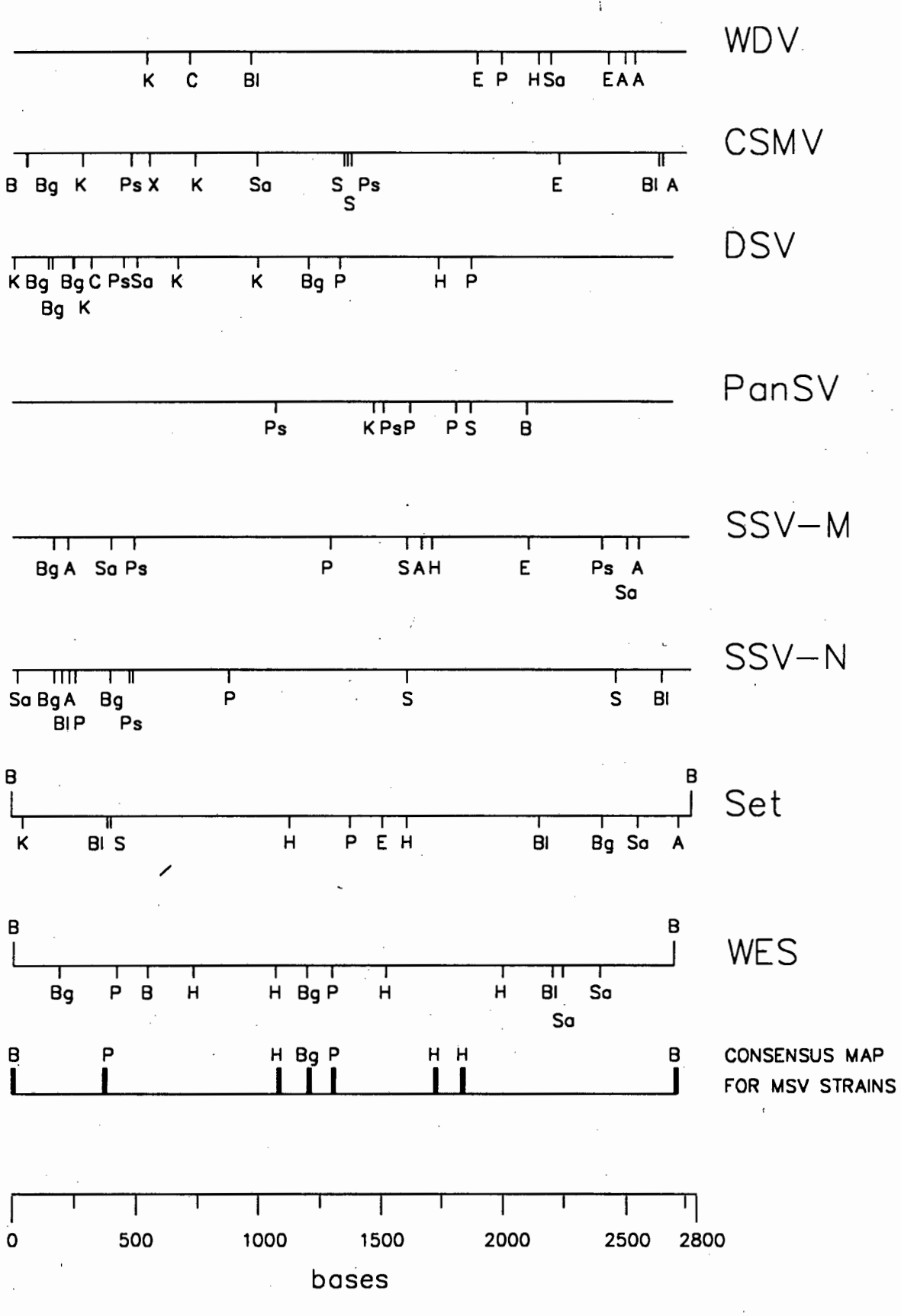

B
B
S
K Ps P SBI H Bg P XH H Bl Sa Sa A
K
K P S Bl H Bg P H H Bl Sa Sa AH
N
K P Bl K H Bg P XH H Bl Sa Sa AA
PE
K H P Bl S H Bg P H H Bl Sa Sa A
Koe
K P S H Bg P X H Sa Bl A H
Kom
K P S Bl Sa H Bg P Sa X H Sa Bl A Sa
RSE
P S Bl H Bg P H X H Bl Sa Sa A
SW
K P Bl H Bg P H H Bl Sa Sa A
WES
Bg P B H H Bg P H H Bl Sa Sa
Set
K Bl S H P E H Bl Bg Sa A
B P H Bg P H H B
CONSENSUS MAP
OF MSV STRAINS
0 500 1000 1500 2000 2500 2800
bases

Figure 4.7.


WDV
K C Bl E P H Sa E A A
CSMV
B Bg K Ps X K Sa S Ps S E Bl A
DSV
K Bg Bg C Ps Sa K K Bg P H P Bg K
PanSV
Ps K Ps P P S B
SSV–M
Bg A Sa Ps P S A H E Ps A Sa
SSV–N
Sa Bg A Bg P S S Bl Bl P Ps
Set
B B K Bl S H P E H Bl Bg Sa A
WES
B B Bg P B H H Bg P H H Bl Sa Sa
B P H Bg P H H B
CONSENSUS MAP FOR MSV STRAINS
0 500 1000 1500 2000 2500 2800
bases

4.3.4. Partial Sequencing of Cloned Virus Isolates.

Because the viral DNA insert of each clone was sequenced from either the forward or the reverse primer site on the vector, sequence was obtained for one strand only for each end of the insert that was sequenced. The partial sequences of each clone that was end sequenced are shown in Figure 4.8. Percentage sequence identities between the partial sequences obtained and the equivalent sequences of MSV-S, SSV-N, and DSV, calculated using GENEPRO, are shown in Table 4.7.

Table 4.7. Percentage DNA sequence homology shown between partial sequences obtained for cloned streak virus isolates and the corresponding sequence from MSV-S, DSV, and SSV-N.

Clone	genome region to which sequence corresponds	% sequence homology to MSV-S	SSV-N	DSV
pSET100 [176.R]	LIR (inverted)	82	43	44
pSET100 [369.F]	V1 ORF	78	56	59
pWES100 [250.R]	C1 ORF	81	61	55
pWES100 [273.F]	C1 ORF (inverted)	89	57	35
pWES100 [273.F]	LIR	78	-	-
pPS100 [218.F]	C1 ORF	61	64	67
pSSM100 [164.F]	C1 ORF (inverted)	72	72	67
pSSM200 [182.F]	V1 ORF (inverted)	63	70	58
pSSM200 [114.R]	V1 ORF	60	56	53

Key: Contents of square brackets give the number of nucleotides sequenced and the sequencing direction; F = forward, R = reverse. ORF = open reading frame - designations are as given by Boulton *et al.* (1989; see Chapter 3). LIR = large intergenic region (see Chapter 3).

Figure 4.8. (shown on p. 82). Partial DNA sequence obtained from selected clones of Subgroup I geminiviruses.

Legend: Clone designations are as detailed in Table 4.2. F = sequence obtained using pUC universal (forward) sequencing primer. R = sequence obtained using pUC reverse sequencing primer.

Figure 4.8.

pSET100 [R]
GGATCCATGG CTGAATCGCA CTTGTGAGAC AATTTATAGT
CGCGTCCCAC GGCTTCTTTA AACAAGGCAG CTTTGATATT
TCACGACAAA AGCAAATCTG AGCTTTAAAT CAAAGACGCT
CGGTCCCTAC CGGGCCCTCG CAGGTAAAGG GGGCGAGGTA
ATATTATTGC TCGCCC

pSET100 [F]
AGAGTGCTAT CTATCTCTCC CACGGGTACC CACAGCAGCT
CCGACCACCG GAGGGTTGTC GTGGAGTCAC GTCGGCGAGG
TAGCTATACT GACTTTGTTG CTTTGATTTG CATTTATCTG
CTTTACCTTT GGGTGTTGAG AGACCTTATC TTAGTCTTGA
AGGCAAGACG CGGGAGGTCC ACGGAGGAGC TGATATTTGG
ATCTGAAGCT GTGGATAGGA GGCACCCTAT CCCTAATACT
TTGGAACCTA CAGCTCCGGT GCATCCGGAC CGTTCGTTCC
AGGTCAGGGA TAAGCAGTAA GCCATGTCGA CGTCCAAGAG
GAAGCGTGCC AGTAGCGGCA ATGGAATAGC GTCCACCAAG
AAGAAAGG

pWES100 [R]
CGACTAGAGG ATCCCCGGGC ACATATATAT ACTTTGGGGT
CCAACGACCA ACGAGCTCCC AAATCATCTG ACAGAGGATT
TCTGGATTTT CTGGACAGTG TGGATATGTT AAGAACGTGT
TAATGTTCCT GTGTGAGAAC TGACGGTTGG ATGAGGAGGA
GGCCATATCG GACGACTCCG AGCAGCTTGC GGGATGGCAG
GATGGGAGCT CCAAACTCTA TATCAACCGG TTGCGCCTTC
GAAATCCGCC GCTCCCCCTT TTATAGTGGT TGTTTATGGC
CGGACCGG

pWES100 [F]
GGAGATATGC ATTTACATGC ATTAATCCAG ACAGAGAAAC
CGGTAAAGGG TTCCACCCAA ACATTCAGAG TGCAAAATCA
GTAAACAAGG TCAGGGATTA CATTCTCAAG GAACCTCTGG
CCTTGTTTGA GAGAGGTACT TTCATTCCTA GGAAGTCAAG
CTTCCTAGGA AATCCTTCCA AGGGAAATTC TGAAAAGAAA
CCTTCCAAAG ATGAAATAAT GAGAGAGATT ATTTCACACT
CCACTTCAAA

pPS100 [F]
TCTAGAGGAT CCCAAGGACA AGTGGGAGAA AGGTACTTAC
ATACCAAGGA AGAAGAGCTT CGTTCCTCCA GGTAAGGAAA
ACTCTGAGAA GAAGCCTTCT AAGGATGAGG TCATGAAGGA
GATCATGACA CATGCCACTA GCAGAGCAGA GTACCTTTCA
CTAGTGCAGA CTTCACTGCC TTACGACTGG GCCACAAAAC
TGTCATACTT TGAGTACTCA GCTAGTC

pSSM100 [F]
GTCGACGTTG TTTTGCCAGT AATTATGTCT TCCCAGGCTA
CGGGCCCAAC TAGATTTTCC AGTTCTTGTT GGGCCGAGGA
TGTAGAGGCT AAGCTTTCTT GTACCTGTAT TGTTCTGGAT
GTAGAAGACA TCCATTCTAG GTCAGAGGAT GCCTCTTGAG
GCTT

pSSM200 [F]
GGTACGTTTT ATCTCTCAGC ACCCAAACCC AAAGTAAGTA
AAGCGCTACG ACTGCAACAA AGGTAAAGAT AGCTATTTCA
CCAACGCGGC TCCACGGCAA CGCAGGAGCG GACGGAGCTG
CCCGGGGGAC ACGAGGCAAA GCTACTGCGG TGCAAAGCAG
GGGGACACGT CCNTTAAGGC AT

pSSM200 [R]
TCTTGTCTTC GTGTTGAAGG CTCGACGAGG CGGATCTACG
GAGGAGCTGC ATTTTGGGCC CAGGGAGCGG GAGTCTGTCG
CTTCTGCCGA CAGTCTCGTC CAGTCGCTGT TCCG

Tentative relationships between the viruses indicated in Table 4.7 could be established, given that between Subgroup I geminivirus members, some parts of the genome are intrinsically more conserved than others. The sequences obtained for pSET100 and pWES100 showed greater similarity to MSV-S sequence, and were largely dissimilar to SSV-N and DSV sequences. While pSET100 showed equally poor homology to DSV and SSV sequences, pWES100 showed slightly better sequence homology to SSV than to DSV. The pPS100 sequence was as poorly related to those of MSV, SSV, and DSV as the corresponding sequences of these latter viruses were among themselves (see Chapter 3). Sequences obtained for pSSM100 and pSSM200 appeared to be almost equally similar to corresponding sequences from both MSV-S and SSV-N, and only slightly less similar to DSV sequences.

4.3.5. Phylogenetic Analysis from Restriction Map Data.

The divergence values shown in the lower left of Table 4.8 were used in the FITCH programme to generate the phylogenetic tree shown in Figure 4.9. Using the entire data set, FITCH examined 1444 trees, with an average standard deviation of 13.8%. The tree generated using NJTREE on the entire data set (not shown) was topologically similar overall to that produced using FITCH.

As checks, subsets of the data shown in Table 4.6 were also analysed using NJTREE and FITCH. Thus, the maize isolates and SetSV and WESV ("MSV" subset) were analysed separately, as were all grass isolates and a single MSV isolate, MSV-S ("grasses" subset). The global rearrangement (G), jumble (J), shuffle (S), and outgroup (O) options of FITCH were used.

The trees generated using the grasses data subset agreed well with the appropriate portion of the tree generated using the entire data set, and gave a comparable standard deviation (12 to 12.5%).

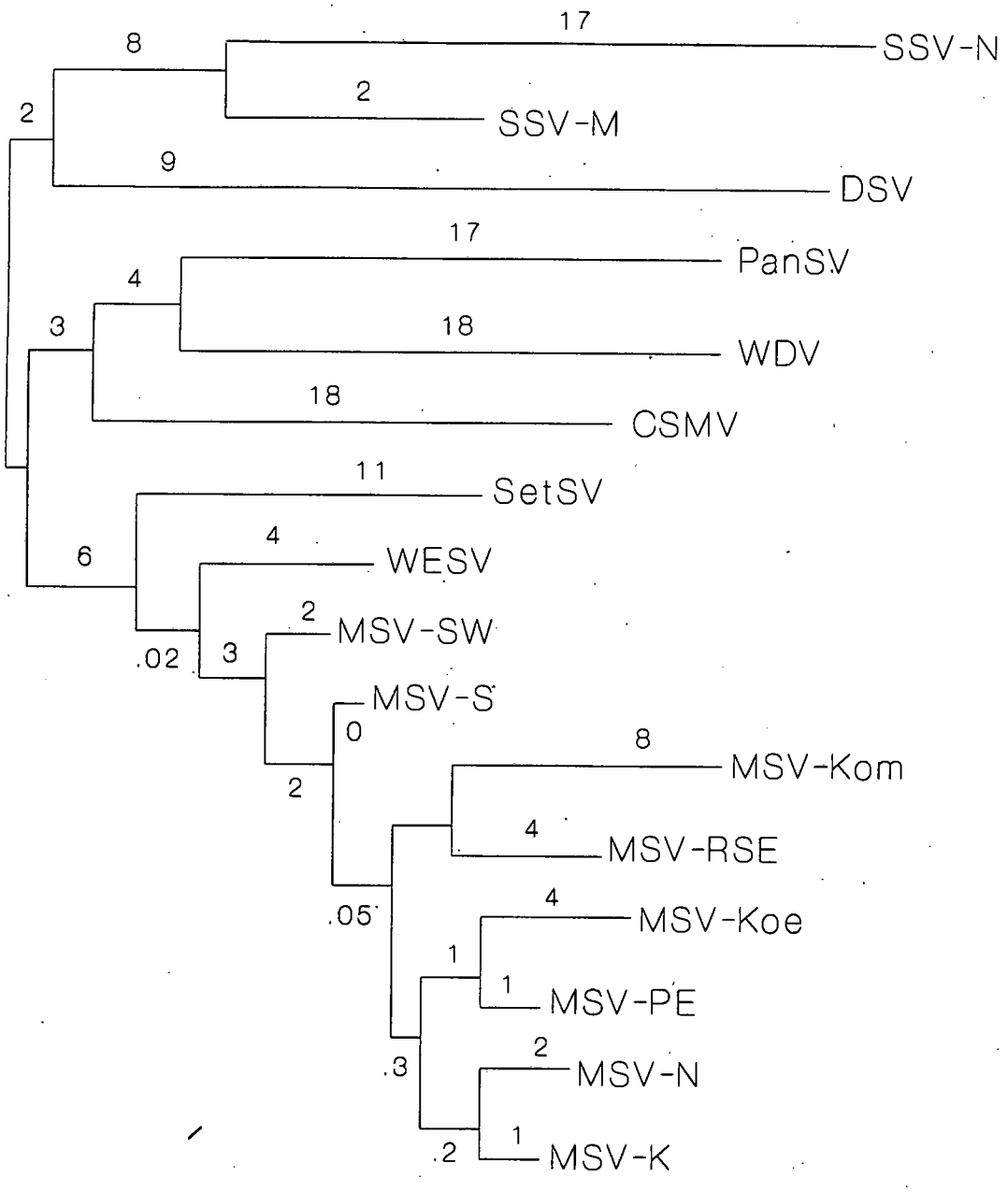
In the initial analyses made, the precise placement of the maize MSV isolates with respect to one another differed between trees generated, depending on whether the whole data set was used, or only the MSV isolates, and whether FITCH or NJTREE was being run. However, from four runs of the MSV data subset, using the J and S options, the same optimum tree was obtained.

Table 4.8. Matrix showing sequence variation between Subgroup I geminiviruses estimated from shared restriction endonuclease sites.

MSV-S	**16**	**14**	**14**	**13**	**10**	**9**	**11**	**12**	**8**	**5**	**0**	**0**	**4**	**3**	**2**	**1**
MSV-K	0.017	**15**	**13**	**12**	**9**	**8**	**10**	**9**	**8**	**4**	**0**	**0**	**4**	**4**	**2**	**1**
MSV-N	0.022	0.029	**16**	**12**	**8**	**7**	**9**	**9**	**8**	**4**	**0**	**0**	**4**	**5**	**2**	**1**
MSV-PE	0.029	0.037	0.043	**15**	**10**	**8**	**9**	**9**	**6**	**4**	**0**	**0**	**4**	**4**	**1**	**1**
MSV-Koe	0.062	0.074	0.099	0.056	**13**	**5**	**6**	**8**	**6**	**4**	**0**	**0**	**2**	**4**	**1**	**1**
MSV-Kom	0.096	0.110	0.137	0.110	0.177	**16**	**7**	**5**	**5**	**4**	**0**	**0**	**2**	**4**	**2**	**2**
MSV-RSE	0.052	0.062	0.085	0.079	0.135	0.127	**14**	**8**	**5**	**3**	**0**	**0**	**3**	**2**	**1**	**1**
MSV-SW	0.031	0.074	0.079	0.074	0.081	0.177	0.087	**13**	**8**	**5**	**0**	**0**	**3**	**3**	**2**	**0**
WESV	0.099	0.093	0.099	0.141	0.129	0.177	0.166	0.081	**13**	**5**	**0**	**1**	**3**	**3**	**1**	**0**
SetSV	0.172	0.203	0.209	0.203	0.190	0.209	0.244	0.153	0.153	**12**	**0**	**0**	**0**	**2**	**2**	**0**
PanSV	0.407	0.399	0.407	0.399	0.384	0.407	0.392	0.384	0.384	0.376	**7**	**0**	**0**	**1**	**0**	**0**
SSV-N	0.441	0.434	0.441	0.434	0.421	0.441	0.427	0.421	0.421	0.415	0.376	**12**	**4**	**2**	**1**	**0**
SSV-M	0.208	0.203	0.207	0.203	0.305	0.324	0.244	0.238	0.238	0.415	0.376	0.183	**12**	**4**	**1**	**1**
DSV	0.268	0.214	0.183	0.214	0.203	0.220	0.324	0.251	0.251	0.312	0.392	0.312	0.196	**14**	**2**	**0**
CSMV	0.330	0.324	0.330	0.441	0.427	0.330	0.434	0.312	0.427	0.305	0.384	0.421	0.421	0.318	**13**	**0**
WDV	0.427	0.421	0.427	0.421	0.407	0.312	0.415	0.407	0.407	0.399	0.356	0.399	0.399	0.415	0.407	**10**

Key: The upper right of the table shows the number of shared restriction sites between pairs of viruses (in bold). Sequence divergence values are shown in the lower left of the same table. The divergence values along the diagonal (MSV-S vs MSV-S, etc) = 0.0000. MSV-S = South African (Potchefstroom, Transvaal) isolate (Lazarowitz, 1988). MSV-K = Kenyan isolate (Howell, 1984). MSV-N = Nigerian isolate (Mullineaux *et al.*, 1984). Other map designations are as shown in Table 4.1.

Figure 4.9. Consensus unrooted phylogenetic tree of the Subgroup I geminiviruses derived from sequence divergences as estimated from restriction map data.



Legend: Isolate designations as given in Figs. 4.6 and 4.7. Vertical distances are arbitrary. The numbers shown above the horizontal branches (not drawn to scale) represent mutational distances (expressed as % divergence from nodal branch populations). Average percent standard deviation of branch length was 13.0%.

The O option rooted the tree on SetSV, which agreed best with the "MSV" portion of the overall tree. The "sub-tree" so generated was incorporated into the consensus tree containing all of the viruses in the analysis (Fig. 4.9).

The maize group of viruses formed a closely-related cluster, but with some evidence of diversity within the group (Fig. 4.9). In all trees generated, the SetSV and WESV isolates are the most closely related grass viruses to the maize isolates of MSV, with WESV being more closely related to maize isolates than SetSV. The relationship between SetSV and WESV is more distant than that between any of the maize isolates of MSV.

The two sugarcane viruses branched from the same node of the tree, but were distantly related to one another. SSV-N was was more closely related to DSV than to MSV, but DSV and SSV-M were equally closely related to the MSV isolates and to one another. The PanSV isolate grouped with WDV. This was presumably because PanSV shared no restriction sites with any other isolate, and only a highly distant and approximate placement could be made. For the same reason, PanSV, CSMV, and WDV show the greatest divergence from one other and from all other virus isolates.

4.4. DISCUSSION.

4.4.1. Clonal Analysis of Virus from a Single Plant.

A drawback in this study was that the methods used had a direct influence on the results achieved, and so only limited conclusions could be drawn. At the time that this study was performed, only maize isolates of MSV had been studied and it was assumed that all streak isolates from African grasses contained a single *Bam* HI site and that all isolates would cross-hybridize with MSV. As is evident from the results above (4.3.2., 4.3.3) this is not always so.

Using *Bam* HI to restrict total plant DNA for mass cloning is not ideal since there is a possibility of non-specific endonuclease activity ("starring") with this enzyme (Maniatis *et al.*, 1982). Consequently, it can not be easily determined whether a recombinant with a less than full-length viral insert contains a sub-genomic piece of DNA formed *in vivo*, as has been found for WDV (1.470 kbp; MacDonald *et al.*, 1988) and MiSV (1.4 $\pm$ 0.24 kbp; Ikegami *et al.*, 1989), or whether

the insert represents a cutting artifact. Because of this uncertainty, clones with less than full-length inserts were not analysed further. In the survey reported here, "sub-genomic" inserts were obtained at a relatively high frequency (39.6%). It is possible that this percentage is due to a cloning artifact, whereby incomplete fragments of viral DNA are selected over full-length genomes.

The study was also narrow in that if more than one unrelated virus were to infect a single plant, as has been shown for MSV-Koe and SSV-N (Chapter 2), the probe used would only detect MSV-related virus DNA and not, for example, SSV-related virus DNA. To a certain extent, this possibility would appear to be discounted by the DNA cross-hybridization results (4.3.2), unless one virus were present at marginal levels. There is evidence that in a dual infection between very different viruses one virus is eliminated by being out-competed by the other during the course of infection (McClean, 1947). The infections investigated in this study were well advanced and so were likely to be single infections by the time of harvesting - but the possibility of dual infection by unrelated viruses cannot be completely ruled out.

What can be concluded from this study, however, is that there appears to be relatively little sequence variation based on restriction site variability between individual whole virus genomes in a single infection (at least as determined with the enzymes used in this study). This finding becomes an important consideration when determining geminivirus variation based on percentage nucleic acid homology between isolates (see Chapters 3 and 5). Comparisons of genomic variation between virus isolates from geographically separated regions or from different host species are only valid if the variation is greater than that found within a single plant.

However, virus variation such as that found between clones in this survey, does point to the generation of limited genomic variation within a host plant, which could perhaps be of evolutionary significance in the long term.

4.4.2. DNA Cross-Hybridization Between Virus Genomes.

Under a defined set of reaction conditions, the extent of DNA cross-hybridization between two viral DNAs is determined by the number of regions of shared sequence similarity and the extent of sequence similarity within those regions (see Chapter 1, section 1.4.3.1). Thus, stable hybridization between viral

DNAs having a particular level of overall similarity to be to one another can be obtained by manipulating the stringency of the hybridization and post-hybridization washing conditions used (Chapter 1, 1.4.3.1).

While this technique is not strictly quantitative, the results correlate well with known levels of sequence similarity between geminiviruses. Under the stringency conditions used here, DSV, MSV, and SSV-N are mutually non-cross-hybridizing. This correlates directly with the extent of sequence homology between these viruses. Overall, they share only approximately 60% total DNA sequence homology, with relatively low nucleotide sequence homology among their V1 ORFs and almost no shared sequence in their intergenic regions (Chapter 3; Hughes *et al.*, 1990b).

In this study, the PanSV isolate cross-hybridized only with MSV-RSE, and the WESV and SetSV isolates, and then only to a minor extent. Thus, it could be deduced that the *Panicum* virus is distantly related to MSV, a deduction borne out by the partial sequence analysis made (4.3.4).

The Mauritian sugarcane streak virus DNA cross-hybridized best with SSV-Natal. Unlike SSV-N however, the Mauritian isolate also hybridized with all other virus isolates tested, albeit marginally. The WESV and SetSV isolates cross-hybridized best with one another and with MSV-RSE, indicating a close relationship between these three isolates. Again, these results correlated well with partial sequence analyses (4.3.4).

4.4.3. Restriction Mapping and Partial Sequencing of Cloned Viral Isolates.

Restriction endonuclease mapping is typically an inexact method, with an unavoidable experimental error of up to 10% in site placement (Kaplan, 1983). In this study, the SSV-N genome was the most complex one to map, having a number of sites clustered in a small region (10 sites in approximately 800 bp). However, comparison of the restriction map obtained experimentally with that derived from the SSV-N nucleotide sequence (Chapter 3), showed that all sites were placed correctly in relation to one another, with an average error of 32 bases (between 1% and 2% of the genome length). Thus, restriction mapping, as performed in this laboratory, gives accurate enough results for typing Subgroup I geminiviruses and for the use of map data in constructing phylogenetic trees.

The map differences between MSV, DSV and SSV at least, can be correlated with DNA sequence differences between MSV, DSV and SSV (Donson *et al.*, 1987; Hughes *et al.*, 1990b; Chapter 3). These map differences are in agreement with DNA cross-hybridization results, where DSV, SSV and MSV were mutually non-cross-hybridizing (4.3.2), and reflect serological findings elsewhere (Dekker *et al.*, 1988).

The PanSV isolate, with its distinctive restriction map, would appear to be a distinct geminivirus, given the above correlation between restriction map differences and genomic differences. This proposition is supported by the lack of extensive DNA cross-hybridization between the PanSV isolate and any of the other viruses studied here, and the limited DNA sequence similarity between it and MSV-S, SSV-N, and DSV, as established by partial sequencing.

Of all the non-maize isolates, the maps of the WESV and SetSV isolates show the greatest similarity to the maps of maize isolates of MSV. This ties in with DNA cross-hybridization results, where DNA of the SetSV and WESV isolates cross-hybridized strongly with MSV DNA, although to a lesser extent than that shown between maize isolates (Chapter 2). In addition, these two isolates showed a high percentage DNA sequence similarity to MSV-S sequences. Taken together, these results indicate that the SetSV and WESV isolates are most likely grass strains of MSV.

4.4.4. Phylogenetic Analysis.

Since the average experimental error in restriction mapping was estimated at 2% of the genome length, as determined for pSS100 (4.4.3), a 2% error was chosen for restriction site alignment. That is, sites were considered as being shared if they fell within 2% of the genome length (about 50 bp) of one another.

Since the SetSV and WESV maps share a number of sites with maize MSV isolates, the phylogenetic relationship between MSV-Set and other MSVs is fairly accurately represented on the tree shown in Figure 4.9. Their positions on the phylogenetic tree confirm the designation of the SetSV and WESV isolates as strains of MSV, hence MSV-Set and MSV-WES, respectively.

From their phylogenetic placement, DSV, PanSV and SSV-N would be classified as distinct viruses. From its placement on the tree, the Mauritian SSV isolate (SSV-M) appears to be a strain of SSV, but is distantly related to SSV-N.

PanSV is grouped with WDV on the tree - probably artificially. It should be noted here that the estimated sequence divergences of WDV, CSMV, DSV, PanSV and SSV are beyond the accurate area of estimation of this method (Nei & Li, 1979; Gibbs & Fenner, 1985; see Chapter 1). Any apparent sharing of sites between these more distantly related viruses could be due to stochastic errors, since the overall number of shared sites is very small. Thus, while they are certainly different, the exact relationship of SSV and PanSV to one another and to other viruses, MSV and DSV in particular, cannot be deduced accurately from Fig. 4.9.

4.5. CONCLUSIONS.

All of the typing methods used in this study give rise to consistent conclusions about the relationships between the virus isolates studied here. That is, that all maize isolates of MSV are markedly similar; that MSV-Set and MSV-WES are the most closely related grass streak isolates to the maize isolates of MSV; that PanSV and SSV are distinct from each other, from DSV, and from MSV maize and grass isolates; and that SSV-M is only distantly related to SSV-N.

Serological comparisons between Subgroup I geminiviruses (Dekker *et al.*, 1988; Pinner & Markham, 1990b) have been used to derive phylogenetic relationships (Rybicki, 1990) which agree well with those derived in the present study. Thus, nucleic acid typing methods may be used to type a streak virus isolate unequivocally, as either a strain of an existing virus or as a distinct virus.

In practice, a field isolate of a potentially new virus would first be used to probe a reference panel of viral DNAs that are mutually non-cross-hybridizing, to determine its broad relationship with existing viruses. Thereafter, a detailed map of the new isolate could be drawn up for comparison with a reference panel of virus maps. In this way, absolute data can be accumulated for typing African grass geminiviruses and it is hoped that eventually the full range of variability of these viruses can be covered.

CHAPTER 5

TYPING OF GRASS GEMINIVIRUSES BY POLYMERASE CHAIN REACTION AMPLIFICATION AND SEQUENCING OF A CONSERVED NUCLEOTIDE SEQUENCE.

CONTENTS

CHAPTER 5.
TYPING OF GRASS GEMINIVIRUSES BY POLYMERASE CHAIN REACTION AMPLIFICATION AND SEQUENCING OF A CONSERVED NUCLEOTIDE SEQUENCE.

SUMMARY

The application of the polymerase chain reaction (PCR) DNA amplification technique to the typing of isolates of Subgroup I geminiviruses is described. Oligonucleotide (17-mer) primers containing degeneracies were used to amplify an approximately 250 base pair fragment from the highly conserved C2 open reading frame. Target DNA was amplified directly from total DNA extracts of streak-infected plants. PCR-amplified DNA fragments were obtained for MSV-RSE, SetSV, SSV-N, SSV-M, PanSV, and WESV, the latter being amplified from both wheat and an *Eleusine* sp. Amplified DNA from all of these isolates, excepting the PanSV isolate, was sequenced, either directly or after cloning into the plasmid Bluescript SK.

Sequence generated from PCR-amplified fragments was used in a phylogenetic analysis of the above isolates and already sequenced Subgroup I geminiviruses. Comparisons were made at both the nucleotide and derived amino acid sequence levels. Analyses indicated that the *Setaria* and *Eleusine* viruses were strains of MSV, and that SSV-M was a strain of SSV, only distantly related to SSV-N. The application of PCR amplification of geminivirus genomic fragments to phylogenetic analysis and taxonomy of Subgroup I geminiviruses is discussed.

5.1. INTRODUCTION.

Investigations of maize and grass geminivirus isolates (Chapter 4) indicated the existence of a range of genomic diversity, with some isolates appearing to be strains of MSV and others appearing to be distinct viruses. It was pointed out that phylogenetic analysis of restriction endonuclease mapping data (Chapter 4) could not always give a precise phylogenetic relationship between different viruses, particularly where few or no restriction sites were shared between viruses.

Relationships between viruses can be deduced unequivocally by sequencing entire viral genomes. This has been done for MSV isolates (Howell, 1984, 1985;

Mullineaux *et al.*, 1985; Lazarowitz, 1988) and SSV-Natal (Hughes *et al*, 1990b; Chapter 3), for example. However, it is clearly not practical in terms of time and materials to sequence every streak virus in its entirety in order to ascertain its precise relationship with other streak viruses.

By use of the polymerase chain reaction (PCR) DNA amplification technique (Saiki *et al.*, 1988), nucleotide sequence from a defined portion of the viral genome can be obtained, in theory, for any Subgroup I geminivirus isolate (Rybicki & Hughes, 1990). By this method, the equivalent genome region from different virus isolates can be compared rapidly and directly. A further advantage is that comparisons can be based on a region of the genome which may not be amenable to serological assay, thereby extending the range of useful data for virus typing.

A study of published geminivirus genomic sequences indicated that a region of approximately 250 base pairs in the C2 open reading frame (which potentially encodes an approximately 17 kDa polypeptide) is well conserved between Subgroup I geminiviruses (Rybicki & Hughes, 1990). The primers designed for amplifying this region comprised the longest possible oligonucleotide sequences containing the fewest differences that bounded the area to be amplified (Rybicki & Hughes, 1990; Fig. 5.1). PCR amplification of this conserved C2 region has been used successfully in this laboratory to type isolates of MSV and related grass geminiviruses (Rybicki & Hughes, 1990). Accordingly, this region was chosen for amplification from streak virus infected plants in the present study, using the conditions already defined by E. P. Rybicki (Rybicki & Hughes, 1990).

In this chapter is reported the PCR-amplification of a conserved 250 bp region from several grass geminiviruses, and the nucleotide sequencing of the amplified products. This enabled comparisons to be made at both the DNA and derived amino acid sequence levels between fully-sequenced geminiviruses and streak virus isolates subjected to PCR-amplification. Phylogenetic analyses were made on the basis of these comparisons, and the results were compared with those of analyses performed on whole genomic sequences.

5.2. MATERIALS AND METHODS.

5.2.1. Sources of Viral DNA.

The following virus isolates were used; MSV-RSE, SetSV, WESV (from both wheat and *Eleusine*), PanSV, DSV, SSV-N, and SSV-M. The origins of these viruses are shown in Table 4.1, Chapter 4. Total DNA extractions from infected plants (A.3) and RF-DNA purifications (A.4 and A.11.b) were made.

5.2.2. Oligonucleotide Primers.

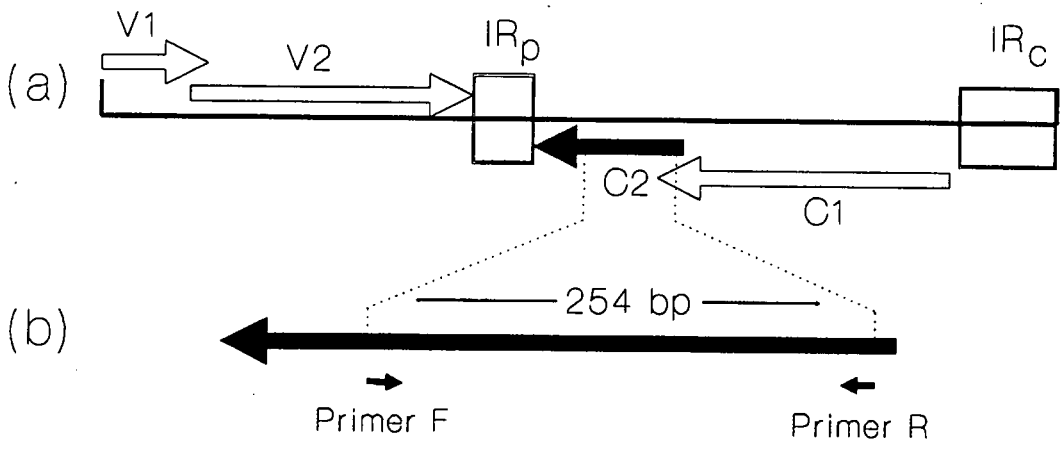
Primers were designed by E. Rybicki (Rybicki & Hughes, 1990) and purchased from Beckman Instruments (SA) (Pty) Ltd, Cape Town. The primers were synthesized as 17-mer oligonucleotides having 5 degeneracies in primer F and 4 in primer R, as shown in Figure 5.1. All four possible bases were used at each degenerate position.

5.2.3. Polymerase Chain Reaction.

Reactions and reaction conditions were as detailed by Rybicki & Hughes (1990). The GeneAmp^{TM} kit from Perkin Elmer-Cetus was used. Thermostable *Taq* polymerase, nucleotides and buffer were all included in the kit. Reaction mixes were made up according to the GeneAmp kit instructions.

In 500 µl microfuge tubes, sample mixes were made up with dH_2O and primers (F and R, 5 µl of 5 µM stock) to a volume of 18 µl. A stock reaction mix was made up in a separate tube. Reagents were added in the order: double distilled water, 10x buffer, nucleotides, and enzyme. The reaction mix volume was calculated to allow for 30 µl mix per sample tube and 20 µl extra "wastage". Only after dispensing reaction mix to the sample tubes were the DNA samples (total plant DNA extract diluted $^1/_5$ in dH_2O) added. This minimized contamination. The negative control sample (dH_2O) was added last.

Figure 5.1. Considerations guiding the choice of amplification primers specific for Subgroup I geminiviruses.



Legend:

(a). Genomic organization in a typical cereal geminivirus. Positions of genomic features are scaled to MSV. Large arrows indicate open reading frames (ORFs); boxes IR_c, IR_p = intergenic regions containing transcriptional control elements and endogenous primer binding sites respectively (Lazarowitz *et al.*, 1989). V1 and V2 = genomic-sense ORFs. C1 and C2 = complementary-sense ORFs (Boulton *et al.*, 1989).

(b). Detail of (a) showing ORF C2, the location of the two sequence blocks showing maximum similarity between viruses (indicated by small arrows) chosen for use as primer sequences, and the target region for amplification lying between the primers.

(c). Sequences of the two amplification primers F (forward) and R (reverse): degenerate positions are indicated by *.

F: 5'-T**A*CCA*TCTTC*TC-3'

R: 5'-GGAAA**CT*C*TGGGC-3'

The final reaction volume was 50 µl, and all reaction mixes were covered with 50 µl of mineral oil to prevent evaporation.

Thermal cycling was done in a custom-made programmable apparatus consisting of a graphite block with contoured wells fitting 14 x 500 µl microfuge tubes (L. Purves and J. Thornthwaite, Dept Chemical Pathology, Univ. Cape Town). The apparatus was heated by low-voltage direct current, with solenoid valve-controlled water cooling. Both heating and cooling rates were about 2°C/sec. Temperature was monitored by a thermocouple probe in a reaction vial containing 100 µl of glycerol.

Thermal cycling started with 2 min at 94°C for initial denaturation, followed by 30 - 36 cycles of 1 min denaturation at 94°C, 2 min annealing at 42°C, and 2 min elongation at 72°C.

For routine detection of amplified DNA, 20 µl samples were mixed 1:1 with loading buffer (Appendix C.8), and electrophoresed in 2% TBE agarose gels containing 50 ng/ml ethidium bromide. The DNA molecular weight markers VI from Boehringer-Mannheim (154-2176 bp) were used.

5.2.4. Cloning and Sequencing of PCR Fragments.

5.2.4.1. Direct Sequencing of PCR products.

This method was used for PCR-amplified DNA from the wheat and *Eleusine* sp. isolates of WESV. PCR-amplified DNA was purified from PCR primers and excess nucleotides in a spin dialysis cartridge (Ultrafree-MC 30 000 NMWL Filter unit, Millipore Corp., Bedford, MA 01730, USA). The PCR reaction mix (after DNA amplification) and 360 µl dH_2O were loaded into a cartridge, and centrifuged at 2000 x g for 4 min at room temperature. Two washes were made, by adding 400 µl dH_2O per wash to the cartridge and centrifuging as already described. The remaining liquid (containing the purified DNA) was removed from above the filter of the cartridge and freeze-dried. Purified, freeze-dried DNA was resuspended in 10 µl dH_2O, and 5 µl was used per sequencing reaction.

Template denaturation and primer annealing were combined in a single step. Reaction mixes were made up, each containing 5 µl DNA, 1 µl F or R primer (0.5 pmol), 2 µl 5x Sequenase buffer (Sequenase kit version 2.0, Sequenase Corp., United States Biochemicals, Ohio, USA), 1 µl of 5% Tween-20, and 1 µl of 5% Nonidet P-40.

Mixes were incubated at 95°C for 3 min and then flash-cooled in an ice-NaCl mixture.

Extension and termination sequencing reactions were carried out as detailed in Appendix B.2, with the following modifications: the labelling mix was used at a dilution of 1 in 20; 1 µl of manganese buffer (as supplied in the Sequenase kit, ver. 2.0) was added to each extension reaction. This reduced secondary structure formation, and allowed bases near to the primer to be read; DMSO in the extension reactions was replaced with 0.75 µl each of 5% Tween-20 and 5% Nonidet P-40; and a stock of each termination mix was made, containing 2.5 µl each of 5% Tween-20 and 5% Nonidet P-40 per 20 µl termination mix. In the termination reactions, 3.1 µl of this mix was used per reaction.

5.2.4.2. Sequencing of cloned PCR products.

PCR-amplified fragments from the SSV-M and SetSV isolates were blunt-end ligated without purification into *Sma* I-digested Bluescript SK (A.8), and used to transform *E. coli* LK112 (A.9 and A.10). Recombinant plasmids were cleaved with *Bgl* I, and clones containing single full-length inserts were size-selected from the electrophoresed restriction products. Selected clones were fully sequenced in both directions (requiring short and medium runs) by the Sanger dideoxy sequencing method (Appendix B.2). One SSV-M clone and two SetSV clones were sequenced.

5.2.5. Sequence-Handling Software.

All sequence searches, matches and comparisons were performed on genomic sequences obtained from the GenBank database or entered by hand using GenePro™ software (Riverside Scientific, Ca. USA) on an IBM XT-compatible 640K microcomputer, or the UWGCG Sequence Analysis Software Package, version 6.1 (University of Wisconsin Genetics Computer Group; Devereux *et al.*, 1984) on a DEC 6000-330 VAX mainframe. The GCG programmes GAP, LINEUP and DISTANCES were used for sequence alignments, aligned sequence presentation, and calculation of pairwise sequence similarities respectively.

5.2.6. Phylogenetic Analyses.

Pairwise sequence similarities were generated from aligned nucleotide sequences using the programme DISTANCES. The DNA similarity data were converted to distances by subtraction of each value from 1.000. An unrooted phylogenetic tree was constructed from the distance data using the programme FITCH in PHYLIP (phylogeny inference package) version 3.1 (J. Felsenstein, Department of Genetics, University of Seattle, Oregon, USA).

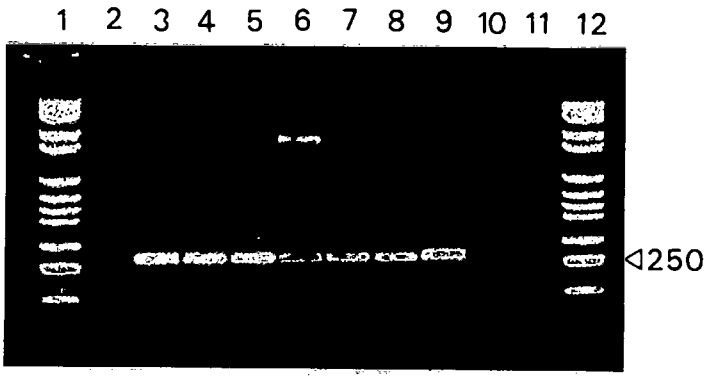
The phylogenetic analysis was repeated using the NJTREE version 2.0 neighbour-joining programme of Saitou & Nei (1987).

5.3. RESULTS.

5.3.1. PCR Amplification.

All geminivirus samples tested amplified positively. A single ethidium-staining product of 250 bp was amplified from all isolates, except for PanSV (Fig. 5.2).

Figure 5.2. PCR amplification of a conserved genomic region from a number of Subgroup I geminiviruses, obtained using degenerate primers and total genomic DNA isolated from streak-infected plants.



Legend: EthBr-stained 2% TBE agarose gel. Lanes 1 & 12: Boehringer-Mannheim MW markers VI. Lane 2: DNA extract from uninfected maize. Lane 3: MSV-RSE. Lane 4: SetSV. Lane 5: WESV. Lane 6: PanSV. Lane 7: DSV. Lane 8: SSV-N. Lane 9: SSV-M. Lane 10: blank PCR reaction containing no input DNA. Lane 11: CETUS kit control DNA. Arrow indicates 250 bp.

In addition to the 250 bp band, higher molecular weight bands were produced with the PanSV isolate - presumably products of non-specific priming. No amplification was observed with C2 primers and the GeneAmp kit control DNA, with primers and DNA extracts of uninfected plants, or in samples containing primers and no template DNA (Fig. 5.2).

5.3.2. Sequencing of PCR Fragments.

PCR-amplified fragments of SSV-M, SetSV, and WESV were fully sequenced in both directions. The MSV-RSE fragment was not sequenced, since other maize isolates of MSV show no significant sequence differences in this region (Rybicki & Hughes, 1990; this Chapter, Fig. 5.3). In addition, 198 bases in the large intergenic region, the most variable region in the Subgroup I geminivirus genome (see Chapter 1), had been sequenced previously from a clone of MSV-RSE (originating from the study reported in 4.2.3) picked at random. This sequence showed 97% sequence identity to the corresponding sequence from MSV-S.

The sequences obtained from the SSV-M, SetSV, and WESV fragments were aligned with the equivalent region from entire genomic sequences of Subgroup I geminiviruses (Fig. 5.3). Since C2 is a complementary sense ORF, complementary sense sequence is shown in Figure 5.3.a. The nucleotide sequences obtained for both the wheat and the *Eleusine* sp isolates of WESV were identical, and so the WESV sequence is given only once. Sequence obtained for the SSV-M fragment was identical to that of the analogous region of a whole-genome clone of SSV-M (pSSM100; see Fig. 4.6), confirming that the amplified fragment originated from SSV-M, and not from a contaminant geminivirus.

The translation product of the single long open reading frame (comprising most of the C2 gene) found in each sequence is shown in Figure 5.3.b. A percentage similarity matrix of nucleotide and amino acid sequences is given in Table 5.1. Conservation of the protein sequence coded for by these genome fragments is evident from Fig. 5.3.b and Table 5.1.

Figure 5.3. Alignment of PCR-amplified C2 gene fragment sequences from Subgroup I geminiviruses.

```
                                                                      60
WESV    GCTTAGGGCT TCATAATTAT TGGCAAAATA ATGTTGATTG GT...CTTCG TACAACGAAG
SetSV   GCTTAGGGCG TCATAATTAC TGGCAAAATA ATGTTGATTG GT:..CTTCC TACAACGAAG
SSV-M   GCCTGGGAAG ACATAATTAC TGGCAAAACA ACGTCGACTG GT...CCTGC TACGACGAGG
SSV-N   GCCTGGGAAG ACATAATTAC TGGCAGAACA ACGTCGACTG GT...CCTCC TACGACGAGG
DSV     GGCTAGGTAG ACATAATTAC TGGCAAAATA ATGTAGACTG GG...CTTCT TATGACGAGG
MSV-K   GCCTAGGGGT TCATAATTAC TGGCAAAATA ATGTTGATTG GT...CTTCA TACAACGAAC
MSV-N   GCCTAGGGGT TCATAATTAC TGGCAAAATA ATGTTGATTG GT...CTTCA TACAACGAAG
MSV-S   GCCTAGGGGT TCATAATTAC TGGCAAAATA ATGTTGATTG GT...CTTCA TACAACGAAG
CSMV    GTCTTGGAAC TCATCACTAC TGGCAGCACT CAGTGAACTT CCTAGAGGAA TGGAACTGCC
WDV     CTCTAGGGAC ACACAATTAT TATAACAGTC TAGTTGATTT CA...CAACA TATGACGTCA

                                                                     120
WESV    ACGCAATCTA CAACATCGTA GATGATATTC CTTTTAAATT CTGTCCTTGT TGGAAACAGT
SetSV   ACACAATCTA CAACATCGTA GATGATATCC CCTTTAAATA TTGTCCTTGC TGGAAACAGT
SSV-M   ACGCAGTCTA CAACGTCATT GACGACATCC CCTTCAAGTT CTGTCCTTGC TGGAAGCAAC
SSV-N   ACGCAGAATA CAACATCATC GACGACATAC CCTTCAAGTA CTGTCCATGC TGGAAGCAAC
DSV     AAGCCCAGTT CAATGTCATT GATGACATAC CATTCAAGTT CTGTCCTTGT TGGAAACAGT
MSV-K   ACGCAATCTA CAACATCGTA GATGATATTC CGTTTAAATT CTGTCCTTGT TGGAAACAGT
MSV-N   ACGCAATCTA TAACATCGTA GATGATATTC CGTTTAAATT CTGTCCTTGT TGGAAACAGT
MSV-S   ACGCAATCTA CAACATCGTA GATGATATTC CGTTTAAATT CTGTCCTTGT TGGAAACAGT
CSMV    AGGCCCAGTT CAACATCATT GATGACATCC CGTTCAAGTT CGTCCCTTGT TGGAAGGGAC
WDV     ACGCCAAGTA TAATATCATC GACGACATTC CATTCAAGTT CACACCCAAC TGGAAGTGCT

                                                                     180
WESV    TAGTTGGCTG TCAGAAAGAG TTCGTGGTAA ATCCAAAATA TGGCAAGAAG AAGAAAGTTC
SetSV   TGGTTGGCTG TCAAAAAGAA TTCGTTGTAA ATCCTAAATA CGGGAAGAAG AAGAAGGTTC
SSV-M   TGATTGGTTG CCAAGAGAAC TACGTCGTTA ATCCCAAGTA TGGGAAGAAA CGGAGAGTAG
SSV-N   TGATTGGCTG CCAGAAGGAT TATATTGTAA ATCCAAAGTA CGGCAAGCGT AAGAAAGTAG
DSV     TGATTGGTTG TCAGAAAGAA TACGTCGTCA ACCCTAAGTA TGGTAAAAAG AAGCNTGTTG
MSV-K   TAGTTGGCTG TCAGAGGGAT TTCATTGTAA ACCCCAAATA TGGTAAGAAG AAGAAGGTGC
MSV-N   TAGTTGGCTG TCAGAGGGAT TTCATTGTAA ACCCCAAGTA TGGTAAAAAG AAAAAGGTGC
MSV-S   TAGTTGGCTG TCAGAGGGAT TTCATTGTAA ACCCCAAGTA TGGTAAGAAG AAAAAGGTGC
CSMV    TCGTCGGCAG CCAGTATGAC CTGACGGTAA ACCCCAAGTA CGGGAAGAAG AAAAGAATCC
WDV     TCGTCGGGGC TCAGCGTGAC TTCACGGTCA ATCCAAAATA TG...GTAAG CGAAAGGTCA

                                                             230
WESV    AGATGAAGTC TAAGCCTACA ATCATCCTCG CCAACTGGGA TGAAGACTGG
SetSV   AGATGAAGTC TAAACCTACT ATCATTCTCG CCAACTCGGA TGAAGACTGG
SSV-M   CCAAGAAAAG CATCTCCACA ATTGTTCTTG CCAACGAGGA TGAAGACTGG
SSV-N   CTAGTAAATC CATCCCCACT ATAGTGCTAG CCAATGAGGA TGAGGACTGG
DSV     CTTCTAAATC CATACCATCA ATAATCCTCA CCAATCCGGA TGAAGATTGG
MSV-K   ACAAGAAGTC TAAGCCTACA ATAATCCTCG CCAACTCGGA TGAAGATTGG
MSV-N   AGAAGAAGTC TAAGCCTACA ATAATCCTCG CCAACTCGGA TGAAGATTGG
MSV-S   AGAAGAAGTC TAAGCCTACA ATAATCCTCG CCAACTCGGA TGAAGATTGG
CSMV    CCA...ACGG GATTCCATGT ATTATCCTCG TAAACGAGGA TGAAGACTGG
WDV     TACGGGGTGG AATACCTTGC ATCATTTTAG TTAATCCAGA CGAAGATTGG
```

(a). Alignment of nucleotide sequences. The sequences are shown from 5' to 3', but inverted with respect to the genomic strand, and with respect to the sequences shown in Fig. 5.1. The sequences from MSV-S, MSV-K, MSV-N, DSV, CSMV, and WDV were obtained from the Genbank Database (see Chapter 3).

Figure 5.3. (continued). Alignment of PCR-amplified C2 gene fragment sequences from Subgroup I geminiviruses.

```
          1                                                    44
WESV      LGLHNYWQNNVDWS.SYNEDAIYNIVDDIPFKFCPCWKQLVGCQ
SetSV     LGRHNYWQNNVDWS.SYNEDTIYNIVDDIPFKYCPCWKQLVGCQ
SSV-M     LGRHNYWQNNVDWS.CYDEDAVYNVIDDIPFKFCPCWKQLIGCQ
SSV-N     LGRHNYWQNNVDWS.SYDEDAEYNIIDDIPFKYCPCWKQLIGCQ
DSV       LGRHNYWQNNVDWA.SYDEEAQFNVIDDIPFKFCPCWKQLIGCQ
MSV-K     LGVHNYWQNNVDWS.SYNEHAIYNIVDDIPFKFCPCWKQLVGCQ
MSV-N     LGVHNYWQNNVDWS.SYNEDAIYNIVDDIPFKFCPCWKQLVGCQ
MSV-S     LGVHNYWQNNVDWS.SYNEDAIYNIVDDIPFKFCPCWKQLVGCQ
CSMV      LGTHHYWQHSVNFLEEWNCQAQFNIIDDIPFKFVPCWKGLVGSQ
WDV       LGTHNYYNSLVDFT.TYDVNAKYNIIDDIPFKFTPNWKCFVGAQ

                                                    75
WESV      KEFVVNPKYGKKKKVQMKSKPTIILANWDED
SetSV     KEFVVNPKYGKKKKVQMKSKPTIILANSDED
SSV-M     ENYVVNPKYGKKRRVAKKSISTIVLANEDED
SSV-N     KDYIVNPKYGKRKKVASKSIPTIVLANEDED
DSV       KEYVVNPKYGKKKXVASKSIPSIILTNPDED
MSV-K     RDFIVNPKYGKKKKVHKKSKPTIILANSDED
MSV-N     RDFIVNPKYGKKKKVQKKSKPTIILANSDED
MSV-S     RDFIVNPKYGKKKKVQKKSKPTIILANSDED
CSMV      YDLTVNPKYGKKKRIPN.GIPCIILVNEDED
WDV       RDFTVNPKYGKRK.VIRGGIPCIILVNPDED
```

(b). Alignment of predicted amino acid sequences of the single large ORF in all sequences in (a), starting from position 3.

Table 5.1. Pairwise Percent Sequence Similarities Between the C2 Gene PCR Fragments of Subgroup I Geminiviruses.

	SSV-M*	SSV-N*	DSV	MSV-K	MSV-N	MSV-S	SETSV*	WESV*	CSMV	WDV
SSV-M	100	83	73	70	69	70	72	70	63	54
SSV-N	95	100	74	70	70	71	72	70	62	54
DSV	86	93	100	74	74	74	73	73	61	58
MSV-K	91	91	85	100	97	98	87	91	61	58
MSV-N	92	92	86	99	100	99	86	90	60	59
MSV-S	92	92	86	99	100	100	87	91	61	58
SetSV	89	92	86	95	96	96	100	91	58	56
WESV	89	92	86	96	96	97	96	100	59	58
CSMV	74	77	76	74	73	74	73	74	100	58
WDV	73	74	75	73	71	74	71	73	75	100

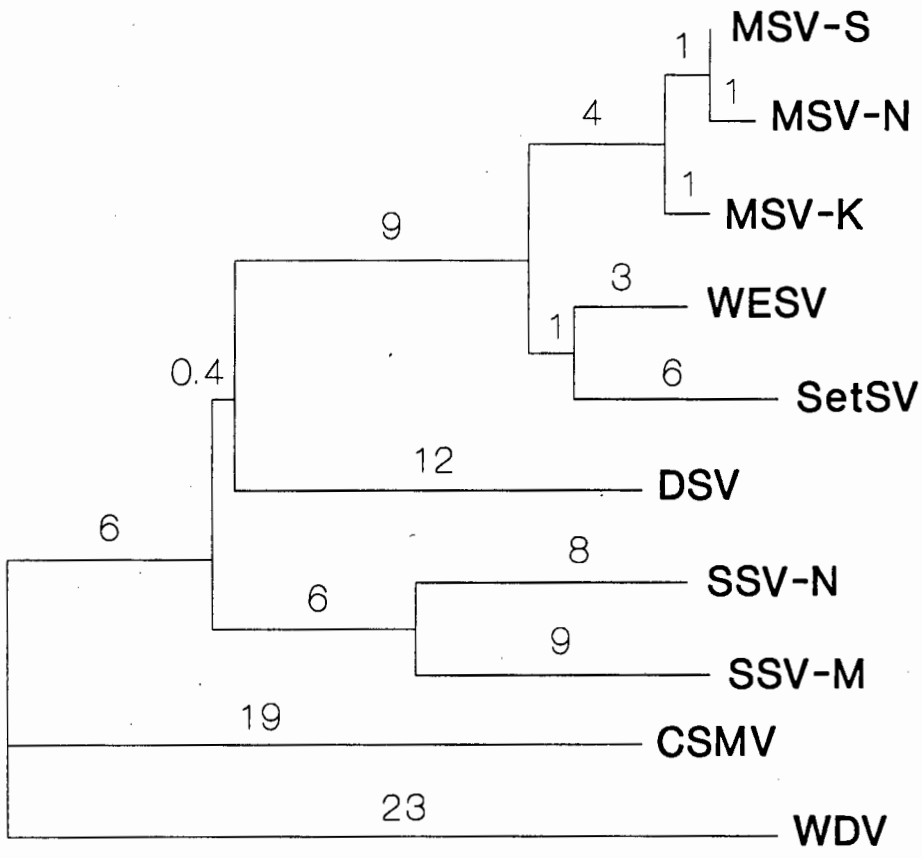
Key: The upper right of the table shows percent direct nucleotide similarities, rounded off to the nearest whole number; the lower left of the table shows percent predicted amino acid sequence similarities. All values were calculated using the UWGCG programme DISTANCES from multiple alignments of sequences performed using LINEUP.

The SetSV and WESV nucleotide sequences were related most closely to one another and to the MSV isolates. The nucleotide sequence of SSV-M was related most closely to that of SSV-N. The homology between SSV-M and SSV-N was lower than that shown between SetSV, WESV and the maize isolates of MSV.

5.3.3. Phylogenetic Analysis.

The unrooted tree generated using FITCH is shown in Figure 5.4. Analysis of the same data set with the species in different orders gave the same result, as did use of the jumble (J) option and several different random number seeds (Felsenstein, 1984; Howarth & Vandemark, 1989; Rybicki & Hughes, 1990). The phylogenetic analysis repeated using the NJTREE version 2.0 neighbour-joining programme (Saitou and Nei, 1987) yielded a topologically identical tree with very similar branch lengths (not shown).

Figure 5.4. **Phylogenetic tree calculated from pairwise sequence distance data for C2 ORF PCR fragments.**



Legend: Shortest tree (= one with smallest sum of squares) calculated using FITCH from nucleic acid percent sequence difference matrix derived from data in Table 5.1 by subtraction of each value from 100. Vertical distances are arbitrary; horizontal branch lengths are proportional to mutational distances (shown above each branch expressed as % divergence from node). Sum of squares was 0.139; average percent standard deviation of branch length was 4.0%; the programme examined 236 trees.

In the tree shown in Figure 5.4, the WESV and SetSV isolates arise from the same branch as the maize isolates of MSV, but are clearly separated from the maize isolates. The two SSV isolates, while fairly distantly related to one another, lie on the same branch of the tree. The SSVs are slightly more distantly related to the MSVs than is DSV, which forms its own branch. CSMV and WDV are the most distantly related viruses, both with respect to one another and to all of the other viruses shown in Figure 5.4.

5.4. DISCUSSION.

The polymerase chain reaction (PCR) oligonucleotide primers, as designed, are theoretically capable of the specific and sensitive amplification of a 250 bp sequence in the C2 ORF from a number of Subgroup I geminiviruses, including WDV, CSMV (Rybicki & Hughes, 1990), and MiSV (E. P. Rybicki, pers. comm.). As expected, streak virus DNA could be amplified from streak-diseased *Panicum, Setaria, Eleusine* and wheat, from two different sugarcane streak virus isolates, from DSV, and from maize isolates of MSV. The PanSV isolate, however, continually gave more than one band upon electrophoresis of PCR products, even after a second cycle of amplification using gel-purified DNA (A.7) of the correct size, obtained from the first amplification reaction. This multiple banding was also obtained when cloned PanSV (pPS100) was used as source DNA. Such multiple banding was felt to be indicative of non-specific priming.

As performed in this study, it was found that sequencing from cloned PCR-amplified products gave cleaner, more easily readable sequence than did direct sequencing of uncloned PCR-amplified products.

Repeated attempts were made to obtain sequence from the PanSV isolate. These attempts included the use of direct sequencing of PCR-amplified PanSV DNA that had been generated from a second round of amplification using gel-purified DNA, as described above; use of the F and R PCR primers to prime the whole-genome clone of PanSV (pPS100, see Chapter 4) for sequencing; and repeated attempts at cloning PCR-amplification products.

No sequence at all was obtained using the first two methods described above. Using the third method, out of many clones tested, only two clones containing $\pm$ 250 base pair inserts were found. These were partly sequenced.

Of the initial sequence obtained, one clone showed no sequence homology to any sequenced Subgroup I geminivirus - though this could have been due to the sequence being from the large intergenic region, in which case it is possible that no homology would be detected. However, the terminal sequence of either strand of this same clone was that of only one of the primers, indicating the occurrence of random priming.

Only 58 bases, in one direction, were sequenced from the second clone (pPS200). While the sequence was from the C2 region, one end of the insert lay

168 bp into the sequence potentially targeted by PCR primers. From the size of the insert, the other end presumably lay well outside the PCR target sequence. The degenerate F primer sequence, 5'-TTCAGCCAGTCTTCGTC-3', was found at the start of the fragment sequence.

From all of the above, it was concluded that the designed primers could not specifically amplify the target C2 DNA of PanSV, under the reaction conditions used in this study. Likewise, until the entire C2 region of the PanSV genome has been sequenced, the reason for the failure of the primers to hybridize specifically to the target sequence cannot be determined.

Despite this, an indication of the level of DNA sequence similarity between PanSV and other Subgroup I geminiviruses can be obtained. The limited nucleotide sequence obtained from pPS200 was compared with the DNA sequences shown in Figure 5.3.a. PanSV showed the following percentage DNA sequence similarities: MSV-S 72%; SetSV 69%; WESV 69%; DSV 65.5%; SSV-N 77.5%; SSV-M 71%. These results indicate that although all of the sequence similarities are poor for this part of the genome, PanSV is most closely related to SSV-N, equally distantly related to MSV and SSV-M, and most distantly related to DSV. It should be noted, however, that these results are based on comparisons using a small number of nucleotides and should be treated with caution. Comparisons made earlier between MSV-S, DSV, and SSV-N sequences and the C1 ORF partial sequence obtained from pPS100 (see Chapter 4, Fig. 4.6, Table 4.5) showed PanSV to share the greatest sequence similarity with DSV (67%, compared with 64% with SSV-N and 61% similarity with MSV-S).

Thus, the precise relationship between PanSV and MSV, SSV, and DSV cannot be determined with the sequence comparisons at hand. However, considering that the C1 and C2 ORFs are the most highly conserved parts of the genome in this Subgroup (Chapter 3; Rybicki & Hughes, 1990), the percentage sequence similarities obtained over 276 bases are low enough that, taken with its distinctive restriction map and lack of extensive DNA hybridization to other viruses studied (Chapter 4), PanSV could be considered a distinct Subgroup I geminivirus. It is proposed that the virus should be termed panicum streak virus (PanSV).

The phylogenetic tree shown in Figure 5.4 indicated that the *Setaria* sp. virus and the wheat/*Eleusine* sp. virus isolates are best considered as strains of MSV,

termed MSV-Set and MSV-WES, respectively. MSV-Set appeared to be more distantly related to the (type) maize strain than MSV-WES.

Although SSV-M could be grouped with SSV-N, it was very nearly as distantly related evolutionarily to this virus as to DSV or to the MSV strains. Rybicki & Hughes (1990) have proposed that SSV-M could be considered a distinct Subgroup I geminivirus, on the basis that the distinct subgroup III geminiviruses bean golden and tomato golden mosaic viruses (see Chapter 1, section 1.3.1) share 79% DNA similarity and 87% direct protein similarity in the same region (compare with Table 5.2).

The two sequenced isolates of WDV available (WDV-S, MacDonald *et al.*, 1985; and WDV-CJI, Woolston *et al.*, 1989) showed no sequence differences in the region of the genome corresponding to that amplified by the PCR primers used in the present study. Hence only WDV-S is shown in Figures 5.3 and 5.4. In passing, it is worth mentioning that WDV-S and WDV-CJI are probably isolates of the same strain of WDV, since they share 98.3% DNA sequence identity overall (Woolston *et al.*, 1988). This is comparable with the Nigerian, Kenyan and South African isolates of the maize strain of MSV, which also share approximately 98% DNA sequence identity (Lazarowitz, 1988).

Of all of the viruses shown in Figure 5.4, CSMV and WDV were the most distantly related, both with respect to the other viruses shown in Figure 5.4, and to one another.

The C2 PCR-based phylogeny presented here is very similar to that derived by Howarth & Vandemark (1989), also using FITCH and distance data derived from replication-associated protein sequence. This similarity would be expected, since the C2 ORF DNA sequence used here encodes the most conserved region of the replication-associated protein(s) (Rybicki & Hughes, 1990). The PCR approach has the advantage, however, that only a short DNA sequence is required to give essentially the same result.

The present study indicates that "MSV" in fact exists as a loose "streak virus cluster" which includes MSV (maize, *Setaria*, and *Eleusine* strains), DSV, PanSV, and SSV (Natal and Mauritius strains). It is probably necessary, however, to sequence SSV-M in its entirety in order to determine whether it really is a strain of SSV, like SSV-N, or a distinct Subgroup I geminivirus. Finally, it is possible that further study could bring to light other members of the cluster.

CHAPTER 6

GENERAL DISCUSSION.

CONTENTS

CHAPTER 6.
GENERAL DISCUSSION

6.1. Evaluation of Methods Commonly Used to Type Cereal Geminiviruses.

6.1.1. Shared Host Range.

Early host range studies using field isolates of virus, hosts and vectors indicated that some streak isolates could not be transmitted between certain host species, and it seemed plausible that streak viruses could be differentiated by host range alone (Storey & McClean, 1930; McClean, 1947).

Pinner *et al.* (1988) have used this criterion to suggest that because a wide range of streak isolates from maize and wild grasses could all infect seedlings of Golden Bantam sweetcorn at the one to two leaf stage, they were all isolates of MSV. However, this reasoning is not altogether correct.

Firstly, Golden Bantam, being a sweetcorn cultivar, is one of the most highly susceptible maize cultivars to any grass streak virus (Damsteegt, 1983). Sweetcorn is used as a "catch-all" in our laboratory to transfer grass viruses that would otherwise be lost if transfer was attempted to other maize cultivars (M. B. von Wechmar, pers. comm.). Further, the one to two leaf seedling represents maize at its most susceptible to any streak virus (Damsteegt, 1983). Therefore, it is to be expected that most viruses carried by *C. mbila* will infect maize under these conditions, even should such a virus normally infect maize with difficulty, if at all.

More importantly, transmission to maize does not in itself prove that two grass geminiviruses are the same. CSMV, a distinctly different virus to MSV (Andersen *et al.*, 1988), also infects maize naturally, and both viruses share several natural and experimental hosts (Damsteegt, 1983; Greber, 1989; see Chapter 1, section 1.3.4). Thus, since the grass geminiviruses infect so many different species in so many families of the Gramineae - introduced cereal crop plants often being particularly susceptible - it would be as big a mistake to say that every streak virus infecting maize is MSV as it would be to say that if MSV were found in *Paspalum*, for example, the virus so isolated is automatically paspalum streak virus.

Because Subgroup I geminiviruses have overlapping host ranges, factors such as ease of infection (acquisition access period and latent period) and incidence of

infection (the number of plants infected *versus* the number of plants exposed to infection) are probably as important considerations as host range, *per se*, in this Subgroup. It would appear, then, that host range is at its most useful when taken in conjunction with the findings of other (molecular) methods used in differentiating between these viruses.

6.1.2. Shared Vector.

It has been implied that because *C. mbila* transmits a range of streak viruses between members of the same and different host species, all those viruses are automatically strains of MSV (eg. Bock *et al.*, 1974; Pinner *et al.*, 1988). However, it is not impossible that *C. mbila* could carry a range of different streak viruses between a wide range of hosts. After all, a single species of whitefly (*Bemisia tabaci*) is the sole vector of all the known Subgroup III geminiviruses - a geographically widespread group having a wide spectrum of dicotyledonous hosts (Goodman, 1981). Besides which, several other species of *Cicadulina* also transmit streak viruses (see Chapter 1, 1.3.3). It is not yet clear whether all *Cicadulina* species are equally capable of transmitting all African streak viruses. Certainly, different *Cicadulina* spp. transmit the maize strain of MSV with differing degrees of efficiency (Okoth *et al.*, 1977; Dabrowski, 1977). No comparative studies have been made regarding the efficiency of different *Cicadulina* species in transmitting other African grass geminiviruses. Of the Australian grass geminiviruses, however, it has been shown that of two strains of CSMV transmitted by *Nesoclutha pallida*, the *Microlaena stipoides* strain is transmitted exclusively by only one biotype of *N. pallida* (Greber, 1989). Further research is needed to show whether one species of *Cicadulina* is better adapted than another for transmitting different African grass geminiviruses.

6.1.3. Serological Methods.

The serological differentiation index (SDI) between two viruses depends not only on the antiserum and source of virus used, but also on whether the virus preparation largely contains intact or disrupted virus particles (Pinner & Markham, 1990b). This so, since disrupted particles have both internal and external epitopes exposed. Thus, shared internal epitopes may contribute to an apparently lower SDI between two viruses than if intact virions were used as antigen source. In addition,

the method chosen for typing will also affect the SDI obtained (Pinner & Markham, 1990b), with indirect ELISA giving the most accurate typing results.

Using indirect ELISA and cross-absorbed antisera, Pinner & Markham (1990b) distinguished four MSV strains, maize (type), *D. setigera* (DSV), *Panicum maximum*, and sugarcane (SSV-N and SSV-M). Admittedly, the study was considered more appropriate to the differentiation of field isolates than to studying geminivirus phylogeny (Pinner & Markham, 1990b), but DSV is currently accepted as a distinct geminivirus, rather than a strain of MSV (G. P. Martelli, pers. comm. to E. P. Rybicki).

Serological comparisons of the same exactitude as those performed recently for African grass geminiviruses (Dekker *et al.*, 1988; Pinner & Markham, 1990b) have not yet been performed among geminiviruses in Subgroups II and III. Nevertheless, it is apparent that Subgroup III geminiviruses are also closely related serologically.

In gel double diffusion tests using antiserum to ACMV, Stein *et al.* (1983) obtained an SDI of 1.0 between ACMV and BGMV, and an SDI of 3.0 between ACMV and TGMV. Using antiserum to ACMV, an SDI of 0.0 between between ACMV and TGMV was obtained.

Roberts *et al.* (1984) also found TGMV, ACMV, and BGMV to be serologically related. Relationships were assessed by immunosorbent electron microscopy (ISEM), whereby grids coated with homologous antiserum entrapped at least 100-fold more virus particles than uncoated grids. Amongst ACMV, BGMV, and TGMV, antisera gave similar increase factors with heterologous viruses to those obtained with homologous viruses.

Of course, only approximate SDI values can be obtained using immunodiffusion or ISEM. However, the serological relationship between ACMV, BGMV, and TGMV is close enough to be detected even when using DAS-ELISA (Thomas *et al.*, 1986).

Perhaps it is to be expected that geminiviruses transmitted by similar vector species are closely related serologically, since the capsid protein confers vector specificity (Briddon *et al.*, 1990a) and has a special function in virus transport within the vector (Briddon *et al.*, 1990a; Medina *et al.*, 1990). For example, DSV (from Vanuatu) is serologically related (SDI of 3.0) to an African streak virus isolate from *Panicum maximum* (Pinner & Markham, 1990a). Three Australian striate mosaic

viruses (CSMV, bromus striate mosaic virus, and digitaria striate mosaic virus) showed a similar degree of serological relationship (SDIs of 3.0 to 4.7) to the panicum virus isolate (Pinner & Markham, 1990a). While the panicum virus is transmitted by *Cicadulina mbila* (Pinner & Markham, 1990b), DSV and the Australian viruses are all transmitted by leafhopper species of the same genus, *Nesoclutha* (Greber, 1989; Julia & Dollet, 1989).

6.1.4. Nucleic Acid Methods.

Undoubtedly, complete nucleic acid sequencing is the most precise means of typing geminiviruses (see Chapters 3 and 5), but it is impractical to clone and sequence every streak virus isolate. Instead, an approximation of virus relationships can be obtained relatively quickly and easily by determining the extent of DNA cross-hybridization between two viruses, by comparing restriction maps, and particularly by the variety of comparisons that can be made using PCR-amplified portions of viral genomes (Rybicki & Hughes, 1990; Chapter 5).

The extent of DNA cross-hybridization shown between two virus isolates gives a rapid indication of the degree of relatedness between them. An approximate distinction can be made between virus strains (one isolate shows reduced hybridization to the other) and distinct viruses (no cross-hybridization between isolates), as was demonstrated in Chapter 4. However, differentiation between closely related strains is difficult, and no differentiation between isolates of one strain can be made.

Restriction mapping can also be used to determine degrees of relationship between grass geminiviruses (see Chapters 2 and 4). This method is particularly useful in distinguishing between closely related strains or isolates of a single strain. Reliable phylogenetic analyses can be made from map data where a large proportion of restriction sites are shared between isolates (as in Chapter 4). However, estimating phylogenetic relationships from restriction maps becomes inaccurate when too few sites are shared between isolates (see Chapters 2 and 4). Nonetheless, we can deduce from DNA sequence data (Chapters 3, 4, and 5) that the number of shared sites becomes low once the overall DNA sequence similarity has dropped to 65% or below. Thus, with accumulated data, it could be predicted that a Subgroup I geminivirus with a distinctive restriction map is likely to be a distinct virus, while

two geminiviruses showing an overall similarity in their restriction maps are likely to be strains of the same virus, or if the maps are nearly identical, isolates of the same strain.

PCR, when carefully designed primers are used, can be a rapid and powerful method for typing geminiviruses (Rybicki & Hughes, 1990; Chapter 5). A comparison of published sequences (Rybicki & Hughes, 1990) would indicate the parts of the genome showing greater or lesser DNA homology between viruses. By making use of different sets of primers, different portions of the genome can be examined, and the viruses typed accordingly. Viral relationships can be graded by a number of criteria: with which set of primers amplification is achieved; the magnitude of amplification achieved; or the presence of non-specific bands on gels, indicating lowered primer-binding specificity for a poorly related virus. In addition, PCR amplification products may be probed with labelled viral DNA (Rybicki & Hughes, 1990). Even more information can be obtained by comparing nucleotide sequences of amplified fragments and, where potential ORFs are amplified, their deduced amino acid sequences (Rybicki & Hughes, 1990; Chapter 5). Thus, using PCR amplification, potentially the entire genome can be used to type grass geminiviruses, as opposed to the approximately 5% genome coding capacity made use of in serological typing (Matthews, 1970; Hull, 1986).

Thus, DNA techniques offer greater versitility and precision in typing geminiviruses than other methods discussed above; the greatest power of DNA techniques (other than DNA hybridization) being their ability to give rise to absolute, rather than comparative, data.

6.2. Criteria Defining Strains versus Distinct Viruses in the Subgroup I Geminiviruses.

In theory, a virus should receive the same taxonomic placement, regardless of the approach used in classifying it. In practice, different taxonomic placements of Subgroup I geminiviruses have been made, depending on the typing method used. It has been concluded from serological studies and some host range studies (Bock *et al.*, 1974; Pinner *et al.*, 1988; Dekker *et al.*, 1988; Pinner & Markham, 1990b) that all African grass geminiviruses are strains of MSV (see above, 6.1.3). On the other hand, from other host range studies (Storey & McClean, 1930; McClean, 1947) and from

nucleic acid studies (Hughes *et al.*, 1990a, 1990b; Rybicki & Hughes, 1990; this thesis), it has been concluded that the African grass geminiviruses studied so far comprise strains of MSV and other distinct viruses (eg. Chapters 4 and 5).

Yet the serological and nucleic acid data obtained to date are not inconsistent with one another, or with host range and symptom expression data. The only disagreement lies in how the data are interpreted (see sections 6.1.1 and 6.1.2). For example, Dekker *et al.* (1988) considered S(SA)S (SSV-N) to be a strain of MSV despite it being very distantly related serologically to maize isolates of MSV, whereas nucleic acid data indicated SSV-N to be a distinct geminivirus (Hughes *et al.*, 1990a, 1990b; this thesis, Chapters 2 and 3). A similar situation pertains with DSV (Donson *et al.*, 1987; Pinner *et al.*, 1988; Accotto *et al.*, 1989).

In fact, direct correlations can be made between nucleic acid data presented in this thesis (DNA cross-hybridization, restriction mapping, PCR amplification) and published serological data (Pinner *et al.*, 1988; Dekker *et al.*, 1988). Rybicki (1990) constructed phylogenetic trees from the SDI data generated by Dekker *et al.* (1988). The trees agreed well with analyses of sequence and restriction map data (Howarth & Vandemark, 1989; Rybicki & Hughes, 1990; this thesis), and the same taxonomic conclusions could be drawn.

From such correlations, a unified scheme is proposed for differentiating, on a molecular basis, between Subgroup I geminiviruses that are distinct viruses and those that are strains of the same virus. The scheme is illustrated overleaf, in Table 6.1. It allows for relatively unsophisticated typing (immunodiffusion and DNA cross-hybridization) applicable to less well-equipped laboratories, as well as the more sophisticated serological and nucleotide sequencing techniques and computer-derived comparisons.

Table 6.1. Unified scheme for differentiating between Subgroup I geminiviruses.

Typing method used	Taxonomic Inference: strain of virus	Taxonomic Inference: distinct virus
Restriction maps	most sites shared	few/no sites shared
DNA cross-hybridization	at least 50%	30% or less
DNA sequence similarity	80% or greater	65% or less
deduced amino acid sequence similarity	90% similarity or more.	75% similarity or less.
gel immunodiffusion	homologous reaction	spur formation
Indirect ELISA : using		
Polyclonal antiserum	SDI lower than 3.0	SDI = 3.0 or higher
Monoclonal antibodies	most epitopes shared	few epitopes shared

6.3. Discrepancies in the Serological Classification of Geminiviruses in Subgroups I and III.

As detailed above (section 6.1.3), Subgroup III geminiviruses are closely related serologically. While their serological interrelationships have not been as extensively tested as those of Subgroup I geminiviruses, the degree of serological relatedness between ACMV, BGMV, and TGMV, for example (Stein *et al.*, 1983), is as high as that between MSV, SSV, DSV, and PanSV, for example. This correlates well with similarities between the deduced coat protein amino acid sequences of viruses within each subgroup: eg. 88% between TGMV and BGMV (Harrison, 1985), 72% between MSV and DSV (Chapter 3). Yet, while TGMV and BGMV are recognized as being distinct, DSV is still sometimes referred to as a strain of MSV. The same applies to other African grass geminiviruses (eg. Pinner *et al.*, 1988).

A perusal of the literature provides an insight on how the confusion over serotyping of all African Subgroup I geminiviruses as strains of MSV may have arisen historically.

The original discovery of a serological relationship between the maize, sugarcane, and *Panicum* streak viruses by the imprecise gel diffusion technique

(Bock *et al.*, 1974) appeared to contradict earlier conclusions that these viruses were distinct (McClean, 1947).

Subsequently, it was determined by ISEM that Subgroup III (whitefly-transmitted) geminiviruses are closely related to one another serologically, but that they shared no serological relationship with either MSV or BCTV. Moreover, BCTV and MSV (both leafhopper-transmitted) were serologically unrelated to one another (Roberts *et al.*, 1984). In the light of subsequent molecular studies, it would no longer be expected that MSV (Subgroup I) and BCTV (Subgroup II) should be serologically related. However, at the time, this study gave rise to the present misperception that all bipartite genome geminiviruses are closely related serologically, while all Subgroup I geminiviruses are serologically unrelated (see also Harrison, 1985). The implication is, therefore, that the finding of a serological relationship between two Subgroup I geminiviruses is definitive proof that they are strains of the same virus.

The misperception was compounded during the discovery and characterization of DSV. When DSV was first isolated, Dollet *et al.* (1986) were properly cautious, in the absence of other data, that DSV could well be a strain of MSV, given the weak serological reaction between the two viruses demonstrated in a gel diffusion test. They also proposed that should the vector prove to be different from the MSV vector and/or the nucleotide sequence prove to be significantly different from that of MSV, then the two viruses could be accepted as being distinct. These provisos have been met (Donson *et al.*, 1987; Julia & Dollet, 1989), and DSV is generally regarded as a distinct Subgroup I geminivirus. Yet, even now, DSV is still sometimes assigned as a strain of MSV (eg. Pinner & Markham, 1989b) - something that would not be tolerated had it been a Subgroup III geminivirus (see above, section 6.1.3).

Of course, it could be argued that the Subgroup III geminiviruses have been incorrectly classified, and that ACMV, BGMV, and TGMV, for example, should be reclassified as strains of a single virus. However, apart from throwing the literature into confusion, evidence exists that ACMV and TGMV, at least, are distinct viruses. Defective subgenomic viral DNA B of ACMV used to transform *Nicotiana benthamiana* acted as a cross-protectant against infection challenge with ACMV, but not against challenge by TGMV, which could not amplify the ACMV subgenomic DNA to trigger the interference phenomenon (Stanley *et al.*, 1990).

The dichotomy between serotyping of Subgroup I and Subgroup III geminiviruses can be resolved by accepting that within either subgroup, distinct viruses may be serologically closely related, though not necessarily so. This would contribute towards a more uniform taxonomic treatment of geminiviruses in general.

6.4. Putative Origins of Maize and Sugarcane Streak Viruses.

From phylogenetic analyses made in this thesis, it is apparent that the viruses found in the *Setaria* and *Eleusine* grass spp. are the most closely related streak viruses to the maize (type) strain of MSV. Other African grass streak viruses, as determined from investigations at the molecular level, are related only very distantly to MSV (eg. a streak isolate from *Coix lachryma-jobi*, Dekker *et al.*, 1988; Rybicki, 1990; SSV and PanSV, this thesis). This is perhaps not entirely fortuitous. Long before the introduction of maize into Africa, *Eleusine coracana* (finger millet) and *Setaria italica* (Italian millet) were grown as staple food crops. Finger millet is susceptible to streak infection (Soto & Buddenhagen, 1982). It is likely that maize, when first introduced, was grown mainly as a subsistence crop to supplement the traditional crops. Our experience of the epidemiology of streak disease in wild grasses has been that such disease occurs in small, scattered loci (M. B. von Wechmar, pers. comm.). It is therefore possible that millet crops provided a high, localized concentration of virus inoculum for spread into maize.

Single base changes affecting host range and specific infectivity, such as have been reported by Boulton *et al.* (1990), may have facilitated the transition to maize as an established host of "ancestral MSV". Boulton *et al.* (1990) also noted that a single base change in the V1 ORF of MSV changed the broad streaks typical of severe MSV infection of maize to a narrow streak (mild) symptom, typical of infection of maize by grass isolates. It has been postulated that streak viruses giving severe symptoms on wild plants are naturally self-eliminating (Rose, 1978). It can be speculated, therefore, that "ancestral MSV" most likely gave mild symptoms in maize at first. The mutation to an aggressive phenotype (severe streaking in maize) possibly occurred later. The V1 mutation observed by Boulton *et al.* (1990) may represent the equivalent of a reverse mutation to an ancestral mild symptom type.

In this thesis, only a limited number of grass streak isolates were studied. It is quite possible that other relatives of MSV (maize), more closely related than MSV-Set or MSV-WES, may yet be found. For example, *Pennisetum typhoides* (pearl millet) is also a traditional crop in Africa, is a favourite host of *C. mbila* (van Rensburg, 1982), and is susceptible to streak infection (Soto & Buddenhagen, 1982).

With regard to the sugarcane streak viruses, *Cenchrus echinatus* is thought to be the natural reservoir of SSV-M in Mauritius (Bock & Bailey, 1989). The origin of SSV-N, which is only distantly related to SSV-M, is less clear. As yet, no strain of virus closely related to SSV-N has been found in local wild grasses. This is not to say that such a strain does not exist, but the possibility should be borne in mind that SSV-N may have other geographical origins.

Available records state that cv. Uba became infected with streak only after its introduction into Natal (Storey & McClean, 1930). During the late 19th Century, large numbers of Asian immigrants entered Natal to work in the sugarcane industry. It has been speculated (R. Kirby, M. B. von Wechmar, pers. comm.) that immigrants often brought ratoons of their own into the country, planting them in small plots along with their household crops. Thus, the possibility that SSV-N may have originated from elsewhere than the African continent is open to speculation.

Ironically, unravelling the origins and interrelationships of the African Subgroup I geminiviruses is complicated by the presence of MSV itself. In the same way that streak can be transmitted from grasses to maize, so streak can be transmitted from maize to wild grasses. In South Africa at least, due to the widespread growth of maize and wheat, the overlapping wheat and maize growing seasons, and the prevalence of streak disease in these crops, it is likely that the maize strain of MSV could swamp other streak virus populations, leading to their eradication by gradual dilution with the more virulent maize virus. It would be of interest to catalogue the variety in the African Subgroup I geminiviruses before this can occur.

APPENDIX A

GENERAL METHODS.

CONTENTS

APPENDIX A
GENERAL METHODS.

A.1. Maintenance and propagation of streak isolates.

Maize isolates were transferred to laboratory grown maize, as described in detail in Clarke (1987) and von Wechmar & Hughes (1990). Briefly, non-viruliferous leafhoppers were allowed to feed on infected field samples for 5 - 10 hours. Uninfected two-week old maize seedlings were then exposed to viruliferous leafhoppers for several days, before being sprayed with systemic insecticide containing either Diazinon or oxydementomethyl to kill leafhoppers. Infected plants were incubated in a plant growth room under the following conditions: 14 hour day / 10 hour night cycle, 24°C day / 21°C night temperature cycle, approximately 70% humidity. MSV-RSE was not transferred to laboratory maize, as only one DNA extraction was made from the original field sample.

The *Panicum* sp., *Digitaria setigera*, and sugarcane isolates were maintained by vegetative propagation. The *Setaria* sp. isolate could not be propagated and so only one DNA isolation was made from the original sample.

A.2. Virus Extraction.

Virus particles were isolated from infected plant material as described by Clarke (1987).

Freshly harvested leaves were homogenized at room temperature with an equal volume/weight of 0.1 M Na acetate buffer pH 4.8 (Miller & Golder, 1950). The homogenate was filtered through a single layer of cheesecloth and the pH of the filtrate immediately adjusted back to pH 4.8 with 10% (w/v) glacial acetic acid. After centrifugation for 10 min at 10 000 rpm in a Sorvall SS34 rotor to remove plant debris, the supernatant was ultracentrifuged at 30 000 rpm in a 60Ti rotor for 180 min at 4°C. The virus pellet was resuspended in 0.1 M Na acetate pH 4.8 and the centrifugation cycle was repeated. The purified virus suspension was stored at 4°C.

A.3. Extraction of double-stranded replicative-form viral DNA from virus-infected plants.

Total DNA was isolated from healthy and infected plants by a modification of the method of Ikegami *et al.* (1981).

Plant material was cut into small pieces, frozen in liquid nitrogen for 5 minutes, and ground into a fine powder. An equal volume of extraction buffer (0.1 M Tris-HCl pH 7.0, 0.1 M NaCl, 0.1 M EDTA, 1% [w/v] SDS) was added and the mixture was stirred for 5 minutes at room temperature. The homogenate was filtered through a single layer of cheesecloth, and the filtrate was extracted once with half its volume of a 1:1 mix (v/v) of phenol:chloroform. Phenol was prepared as described in Maniatis *et al.* (1982). This mixture was centrifuged for 10 minutes at 10 000 rpm in a Sorvall SS34 rotor, and the aqueous phase removed to a fresh SS34 tube. The aqueous phase was mixed with an equal volume of propan-2-ol (isopropanol), and centrifuged immediately for 15 minutes at 12 000 rpm in an SS34 rotor. The crude DNA pellet was resuspended in 4 ml TE buffer (Appendix C.4) and reprecipitated, using 0.1 vol 4 M LiCl and 2 vol 96% EtOH. The pellet was resuspended in 1 ml TE.

A.4. Purification of viral RF-DNA.

A 10 ml column was packed with Sephadex-G100 suspended in TE buffer and equilibrated with TE. 1 ml (1/10 column vol) of total DNA extract was loaded onto the column and eluted with TE in 400 µl aliquots. Each aliquot was scanned in a UV spectrophotometer between 310 nm and 230 nm. Those aliquots giving a typical DNA UV absorption profile were pooled, EtOH-precipitated as above, and the DNA pellet resuspended in TE. This process served to rid the samples of residual protein and plant pigments not removed by phenol/chloroform extraction, giving DNA pure enough for restriction endonuclease digestion.

DNA samples could be further purified if necessary by passage through a Qiagen column (see A.11.b).

A.5. Restriction endonuclease digestion.

The procedures followed were essentially those described in Maniatis *et al.* (1982). Generally, 4 units of restriction enzyme were used to digest 1 µg of DNA in a volume of 20 ul, using the appropriate restriction buffer and reaction temperature for that enzyme (see appendices C.5 and C.6). A special buffer (C.5) was used for *Sma* I, as specified in Maniatis *et al.* (1982). Digests were usually incubated for a maximum of 2 hours. Double- and triple-enzyme digests were carried out simultaneously if the salt and temperature requirements of the enzymes were compatible. If not, sequential digestions were made, using the enzyme with the lowest salt optimum first. The salt concentration was then adjusted for the second enzyme. If the digestion products were to be analysed electrophoretically, the reaction was terminated by addition of gel loading buffer (C.8). If the digestion products were to be used in subsequent enzyme reactions, the digest was either EtOH-precipitated directly or Genecleaned (see A.12), depending on the stringency of the requirements of the subsequent manipulation.

A.6. DNA electrophoresis in agarose gels.

Electrophoresis of DNA was performed using horizontal slab-gel systems, essentially as described in Maniatis *et al.* (1982). TBE buffer (see C.7) was used as gel and tank buffer. Agarose concentration varied between 0.7% and 1.5% depending on the sizes of the DNA fragments being analysed. Ethidium bromide (EthBr) (1.0 mg/ml aqueous stock solution) was added to the gel mix before pouring, to a concentration of 20 ng/ml. DNA samples were mixed with 1/6 vol of loading buffer (C.8) and electrophoresed either at 1.5 V/cm overnight or 3.75 V/cm for 1 - 5 hours. DNA bands were visualized using a 305 nm wavelength (UV) transilluminator and photographed with Polaroid 667 or 665 film. DNA fragments sizes were calculated from standard curves constructed from the migration distances of *Pst* I-digested lambda DNA fragments of known size.

A.7. Gel Purification of DNA Fragments.

DNA restriction products (A.4) were loaded into adjacent lanes in a TAE (40 mM Tris-acetate pH 7.5, 2 mM EDTA) agarose gel and separated by electrophoresis (A.6). One lane in the gel was stained with 100 ng/ml EthBr for 10

min in the dark, viewed on a UV transilluminator (345 nm) and the position of the band of interest marked by notching the gel. The stained portion was matched to the unstained part of the gel and bands corresponding to the notched band were excised. The rest of the gel was stained and viewed to ensure that the correct band had been excised.

DNA was removed from the excised bands by the Geneclean method (A.11), with the following modification: after addition of a 3x vol of the NaI solution, the tube containing the agarose slice was heated in a water bath at 55°C until the agarose melted (approximately 5 - 10 min). Thereafter "Glassmilk" was added and the rest of the procedure carried out as usual (see A.11).

A.8. Construction of recombinant plasmids.

Appropriately digested vector and insert DNAs (A.5), were ligated in a 50 µl ligation mix at an optimal DNA concentration of 0.25 pM. Sterile distilled water and DNAs to be ligated were added to a sterile microfuge tube on ice, to a volume of 45 ul. To this was added 5 µl of 10x ligation buffer (0.5 M Tris-HCl pH 7.6, 0.1 M $MgCl_2$, 10 mM ATP, 10 mM dithiothreitol), thawed from -70°C on ice, followed by 0.25 units of T4 DNA ligase. For ligation of blunt-ended DNA species, 2 units of enzyme were used. Ligation was allowed to proceed at room temperature for 1 hour to overnight. Cleaning up the ligation mix before transformation was not considered necessary and was omitted. Ligation mixes were either used immediately or were stored at -70°C.

A.9. Preparation of *Escherichia coli* LK112 competent cells.

E. coli strain LK112 is a Rec A^- derivative of *E. coli* strain LK111, which has the following genotype:

F^-, thi 1, thr 1, leu B6, lac Y1, ton A21, sup E44, lambda$^-$, rK^-, mK^+, lac I, lac Z, delta M15.

Competent cells were prepared following the method of Chung & Miller (1988).

A 1:100 dilution of an overnight culture of LK112 in Luria-Bertani (LB) broth (C.1) was made into 25 ml pre-warmed LB. The culture was incubated at 37°C with constant agitation until an OD_{600} of 0.30 to 0.45 was obtained. The culture flask was

rapidly chilled in ice-water, and the culture pelleted by centrifugation at 4000 rpm at 4°C for 6 min in a pre-cooled Sorvall SS34 rotor. The bacterial pellet was resuspended in 1/10th volume cold transformation/storage buffer (TSB) (LB pH 6.1, 10% PEG-4000, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 5% DMSO) and the suspension kept on ice for 10 min to allow cells to develop competence. Competent cells could be stored in TSB at -70°C, if required, with an approximately 10-fold drop in competence.

A.10. Transformation of competent cells.

The method of Chung & Miller (1988) was followed.

Five µl ligation mix (A.8) was dispensed to a sterile microfuge tube and pre-chilled on ice. To this was added 100 µl of the competent cell suspension (A.9), and the tubes were kept on ice for 5 to 30 min. Thereafter, 900 µl TSB containing 20 mM glucose was added and the transformation mix was incubated at 37°C for 1 hour to allow the transformed cells to express the plasmid-encoded ampicillin-resistance gene, B-lactamase. Transformation mixes could be plated immediately, or stored for up to 1 week at 4°C, or stored indefinitely at -70°C. Cells were plated (100 ul/plate) on Luria agar (LA) plates (C.1), containing 100 ug/ml ampicillin (C.2). Where pUC or Bluescript vectors were used, X-gal plates (C.3) were used. Plates were incubated overnight at 37°C.

A.11. Plasmid preparative methods.

A.11.a. Rapid method for plasmid size screening.

A plasmid-containing bacterial colony was inoculated into 800 µl LB (Amp50, C.2) in a microfuge tube, and incubated with shaking at 37°C for 3 hours. The culture was centrifuged ("microfuged") for 30 sec in an Eppendorf-type centrifuge for use with 1.5 ml microfuge tubes. The supernatant was discarded. To the bacterial pellet was added 30 µl STE (0.1 M NaCl, 20 mM Tris-HCl pH 7.5, 10 mM EDTA) and 30 µl of a 1:1 mix of phenol/chloroform, and the whole was vortexed for 15 sec. The tube was microfuged for 3 min. and the aqueous phase transferred to a fresh tube containing 30 µl loading buffer (C.8, containing 1 mg/ml DNase-free RNase A). A 10 µl aliquot of this mixture was loaded onto a 0.7% agarose gel.

A.11.b. The Qiagen column method.

The supplier's protocol for the use of Qiagen-tip 20 columns (Diagen Inc., West Germany) was followed. Buffers P1, P2, P3, QB, QC, and QF were supplied in the kit.

Three ml of an overnight bacterial culture in LB (Amp50) was pelleted in a microfuge and the pellet resuspended in 300 µl of buffer P1 (50 mM Tris-HCl, 10 mM EDTA, 100 µg RNase A / ml, pH 8.0). After mixing in 300 µl of buffer P2 (0.2 M NaOH, 1% SDS), the tube was kept at room temperature for 5 min. Then 300 µl of buffer P3 (2.55 M KOAc pH 4.8) was mixed in and the tube microfuged at 4°C for 15 min. The supernatant was loaded onto a Qiagen-tip 20 column previously equilibrated with 1 ml buffer QB (0.75 M NaCl, 50 mM MOPS pH 7.0, 15% EtOH). The column was washed with 2 x 1 ml of buffer QC (1 M NaCl, 50 mM MOPS pH 7.0, 15% EtOH). The DNA was eluted with 800 µl of buffer QF (1.2 M NaCl, 50 mM MOPS pH 8.0, 15% EtOH), precipitated with 0.5 vol of isopropanol, microfuged at 4°C for 30 min, washed with 70% EtOH, and resuspended in 50 µl TE.

The columns were also used to purify RF-DNA from total plant DNA extracts, after precipitating a suitable quantity of total plant DNA extract (determined empirically, depending on the relative amounts of viral and plant DNA in the extract) and resuspending the pellet in buffer P1. Thereafter, the solution was treated as a plasmid preparation, as above.

A.11.c. Small scale plasmid preparation ("miniprep").

An adaptation of the method of Ish-Horowicz & Burke (1981) was followed.

A 1.5 ml aliquot of a 5 ml overnight bacterial culture in LB (Amp50) was microfuged for 40 sec, all excess supernatant removed, the cell pellet resuspended in 100 µl solution I (25 mM Tris-HCl pH 8.0, 10 mM EDTA, 50 mM glucose), and the tube placed on ice. Immediately, 200 µl solution II (0.2 M NaOH, 1% [w/v] SDS) was added, mixed, and the tube was held on ice. After 4 min, 150 µl cold (4°C) solution III (5 M KOAc pH 4.8) was added. The tube was vortexed briefly and kept 5 min on ice before microfuging for 4 min at 4°C. Nucleic acid in the supernatant was precipitated with 2 vol EtOH and resuspended in 400 µl TE. The solution was

reprecipitated (40 µl 4 M LiCl, 800 µl EtOH) and the pellet finally resuspended in 100 µl TE.

A.11.d. Large scale plasmid preparation ("maxiprep").

The method of Ish-Horowicz & Burke (1981) was followed, (solutions as in A.11.c). A 100 - 200 ml overnight culture, grown in LB (Amp50), was harvested by centrifugation in a Sorvall SS34 rotor at 4000 rpm for 7 min. The cell pellet was resuspended in 2 ml of solution I, and kept 5 min at room temperature before addition of 4 ml of solution II. After shaking to mix, the tube was placed on ice for 5 min. Thereafter, 1.5 ml of Cold solution III was added and the tube was shaken briefly and kept a further 5 min on ice before centrifugation at 10 000 rpm (SS34 rotor) for 10 min at 4°C. Nucleic acid in the supernatant was precipitated with an equal volume of isopropanol, and pelleted by centrifugation at 12000 rpm (SS34 rotor) for 15 min. The pellet was resuspended in 4 ml TE. CsCl (1.2 g per ml solution) and EthBr (200 µl of a 10 mg/ml stock) were added, and the solution's refractive index was adjusted to 1.396. After a clearance spin (16 000 rpm for 10 min, SS34 rotor) the preparation was sealed in a 5 ml Quick-seal tube, and spun overnight at 55000 rpm in a VTi 65 rotor in a Beckman L8-70 ultracentrifuge at 15°C. The plasmid band was recovered from the side of the tube as described in Maniatis *et al.* (1982). EthBr was removed by exhaustive extraction with TE/NaCl-saturated isopropanol. The DNA solution was precipitated by addition of 2 vol TE and 1 resultant vol isopropanol, pelleting the DNA in a microfuge for 5 min. The DNA pellet was washed with 70% EtOH and resuspended in TE.

DNA concentration was assayed spectrophotometrically, with the assumption that one absorbance unit at 260 nm is equivalent to 50 µg DNA per ml (Maniatis *et al.*, 1982).

A.12. Purifying DNA with Geneclean.

The Geneclean kit (BIO 101 Inc., California) was used, following supplier's instructions. In brief; 2 to 3 vol of 6 M NaI solution (supplied) was added to the DNA to be cleaned. 5 µl "Glassmilk" (aqueous silica matrix) was added (for more than 5 µg DNA, an extra 1 µl glassmilk was added per extra 0.5 µg DNA). The mixture was vortexed briefly and kept 5 min at room temperature. The glassmilk

and bound DNA was pelleted in a microfuge (5 sec) and the pellet was washed 3 times with 200 - 700 µl of cold (-20°C) "New wash" (Tris-buffered solution, pH 7.0 - 8.5, containing 50% EtOH, made up as directed from supplied concentrate). The pellet was resuspended in the required volume of eluting fluid (dH_2O or low salt buffer), incubated at 45°C - 55°C for 3 min, microfuged for 30 sec, and the DNA-containing supernatant was removed to a fresh tube. Generally, at least an 80% recovery was achieved.

A.13. Membrane hybridization of DNA.

A.13.a. Colony blotting.

This was performed essentially as described in the Hybond-N^+ manual (Amersham plc., Britain). Master and replica plates were prepared by stabbing out colonies in a grid pattern on LA (Amp100) plates and incubating overnight at 37°C. A Hybond-N^+ (charged nylon) membrane, marked for later orientation with the plate, was placed onto the agar surface for 1 min. The membrane was carefully peeled off and placed, colonies face-up, on a pad of filter paper soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 7 min. This step also served to fix the DNA to the membrane. The membrane was then transferred, colony side up, to a filter pad soaked in neutralizing solution (0.5 M Tris-HCl pH 7.2, 1 mM EDTA, 1.5 M NaCl) for 3 min. This step was repeated with a fresh pad of neutralizing solution. Membranes were washed in 2x SSC (C.9) + 0.5% SDS at 65°C overnight, to rid all cell debris and residual agar. Membranes were either stored damp, heat-sealed in plastic, or hybridized immediately, as in A.13.d, using probe DNA prepared as in A.13.c.

A.13.b. "Southern" transfer of electrophoresed DNA from an agarose gel.

The method of Southern (1975), as adapted by Amersham for use with Hybond-N^+ membranes (Hybond-N^+ manual), was followed.

After electrophoresis, DNA in the gel was nicked by brief exposure to UV radiation at 305 nm. The gel was given two 15 min washes in denaturing solution (0.5 M NaOH, 1.5 M NaCl) and two 30 min washes in neutralizing solution (1 M

ammonium acetate, 20 mM NaOH), with shaking. The gel was placed on a glass plate and overlaid (avoiding air bubbles) with Hybond-N^+, cut to fit and wetted in solution II, followed by 3 sheets of Whatmans 3MM filter paper similarly wetted. A stack of absorbent paper towels was placed onto this assemblage and a 1 kg weight put on top of this. Transfer was allowed to proceed from 1 - 16 hours. After blotting, well positions were marked by pencil, the membrane was washed briefly in 2 x SSC (C.9) and either stored damp, heat-sealed in plastic, or hybridized immediately, as in A.13.d, using probe DNA prepared as in A.13.c.

A.13.c. Digoxygenin labelling of DNA probes.

The DNA nonradioactive labelling and detection kit (Boehringer Mannheim) was used both for labelling DNA probes by random primed incorporation of digoxygenin-labelled dUTP, and for the detection of DNA hybrids by immunoassay (A.13.d). The supplier's instructions for use of the kit were followed.

Briefly, linearized Genecleaned DNA, made up to 15 µl with dH_2O, was heat-denatured at 95°C for 10 min and chilled quickly on ice/NaCl. Two µl hexanucleotide mixture, 2 µl dNTP labeling mixture, and 1 µl Klenow enzyme, as supplied, were added and the reaction mix was incubated at 37°C for 1 - 16 hours. The labeled DNA was precipitated at -70°C for 30 min with 20 µg yeast tRNA, 1/10th vol 4 M LiCl, and 3 vol cold EtOH; centrifuged in an Eppendorf-type microfuge at 4°C for 15 min; the pellet rinsed with 70% EtOH and resuspended in 50 µl TE containing 0.01% SDS. Probes were stored at -20°C, and were stable for at least 1 year.

A.13.d. DNA hybridization and immunological detection of bound probe DNA.

The supplier's instructions for use of the DNA nonradioactive labelling and detection kit (see A.13.c) were followed.

Nylon membranes with bound DNA (A.13.a, A.13.b) were pre-hybridized for a minimum of 1 hour in hybridization solution (5x SSC, 0.1% (w/v) N-lauroylsarcosine (Na salt), 0.02% SDS, 0.5% (w/v) blocking reagent (supplied) added to the made-up solution and dissolved at 65°C). Membranes were then sealed in plastic bags with 3 ml per 100 cm^2 membrane of hybridization solution containing

10 - 50 ng freshly heat-denatured (95°C) labelled DNA per ml solution. Membranes were hybridized 6 - 16 hours at 65°C, and then given 2 low stringency washes (2x SSC, 0.1% (w/v) SDS) of 5 min each at room temperature and 2 high stringency washes (0.1x SSC, 0.1% (w/v) SDS) of 15 min each at 65°C. The membranes were either stored air-dried or used directly for detection of hybridized DNA.

Detection procedures were carried out at room temperature. Volumes given are for 100 cm^2 of membrane. Membranes were rinsed in buffer I (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl), blocked for 30 min in 100 ml solution II (0.5% (w/v) blocking reagent (supplied) in buffer I, dissolved at 65°C) and incubated for 30 min in 20 ml of diluted antibody-conjugate solution (antibody-conjugate, as supplied, diluted 1:5000 in buffer I). After two 15 min washes with 100 ml buffer I, followed by a 2 min equilibration in buffer III (0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 50 mM $MgCl_2$) the membranes were incubated in 10 ml colour solution (45 µl NBT solution, 35 µl X-phosphate solution, as supplied, in 10 ml buffer III) in the dark. The colour precipitate started to form between 5 - 30 min, and the reaction was complete between 1 - 16 hours. Membranes were rinsed in TE to stop the reaction and were photographed wet, or air-dried and stored in the dark at room temperature.

A.14. Strategy Followed in the Construction of Restriction Endonuclease Maps.

Initially, single enzyme digests (A.5) were made of the DNA to be mapped. Inspection of the electrophoretically separated (A.6) digestion products showed which enzymes had cleavage sites in the DNA being mapped and, where cloned viral DNA was being mapped, how many sites were present in the DNA insert. If the DNA to be mapped was not cleaved with a particular restriction enzyme, that digest was repeated with the addition of 100 ng of phage Lambda DNA into the digest. If the Lambda DNA was digested but not the DNA being mapped, it was concluded that the viral DNA in question did not include a recognition site for that enzyme. Often however, when using cloned DNA, internal digestion controls were provided by restriction enzyme recognition sites already present in the vector.

In some cases, information on site positions could be gained from single digests (eg. *Sal* I digest in example given below).

Recognition sites were fixed first for those enzymes having only one site in the insert. Single and double enzyme digests were performed (A.5), making judicious use of external (vector) sites to place each site unambiguously (see example given below). Thereafter, making use of vector sites and fixed (single) insert sites, the sites of those enzymes with two recognition sites present in the insert were mapped. Finally, using vector sites and already mapped insert sites (single sites by preference), the sites of those enzymes with three or more recognition sites in the insert were mapped, working always with the smaller number of sites first. Site positions were confirmed by comparing predicted restriction fragment sizes with experimentally obtained fragment sizes in chosen double or triple enzyme digests (see below).

Table A.I. Size of Restriction Fragments Generated for pSET100 (expected total size of 5400 base pairs).

enzyme digests	fragment sizes (bp)	total size (bp)
Bgl II	+5000.	+5000
Bgl II / *Bam* HI	2700 2350 360.	5410
Bam HI	2700 (doublet).	5400
Bgl II / *Pst* I	3000 2350.	5350
Pst I	+5000.	+5000
Sal I	+5000 380.	+5380
Bgl II / *Sal* I	3050 2000 380.	5430

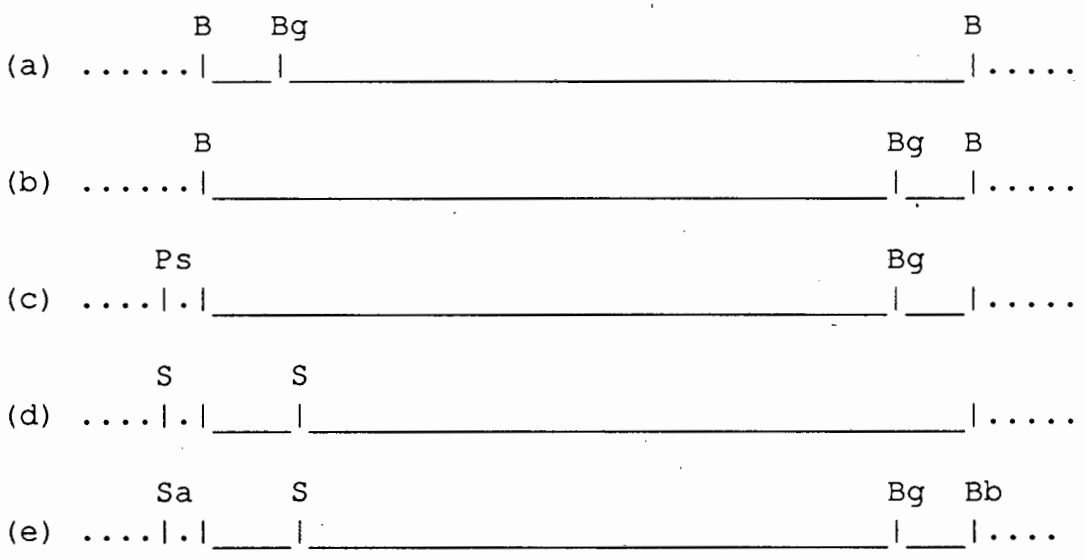
The strategy used is illustrated as follows: the example shown is part of the mapping experiment performed on pSET100 (that is, the complete *Setaria* streak virus genome [approx. 2.7 kbp] linearized using *Bam* HI and cloned into the *Bam* HI-cut plasmid vector, pUC 18 [2686 bp] - see Chapter 4).

From Table A.I, it can be seen that *Bgl* II gives a single large fragment of over 5 kbp (fragment sizes above 5 kbp and below 150 bp cannot be accurately sized from a 0.8% agarose gel). Since pUC 18 has no *Bgl* II site, this implies that the insert has a single *Bgl* II site, but no information is given on its position.

The *Bgl* II / *Bam* HI fragments correspond to the *Bam* HI-excised vector (2700 bp) and two viral insert *Bgl* II / *Bam* HI fragments. Thus, it can be deduced that the *Bgl* II site is 360 bp from the *Bam* HI site (Fig. A.I a, A.I b; see overleaf) but no information is given as to which end of the insert the *Bgl* II site is located.

The *Pst* I digest showed that the viral insert has no *Pst* I site. The site cut was that of pUC 18, which has a *Pst* I site 12 bp from the *Bam* HI site in the vector's multiple cloning site. The *Pst* I site is therefore a useful one for mapping. From the *Bgl* II / *Pst* I digest made (Table A.I; Fig. A.I c), it can be deduced that *Bgl* II is orientated with respect to *Bam* HI as shown in Fig. A.I b (the orientation shown in Fig. A.I a would have given fragments of approximately 400 bp and 4800 bp).

Figure A.I. Deduced positions of the *Bgl* II and *Sal* I restriction sites of the pSET100 viral insert DNA.



Key: The fragments are not drawn to scale.
B = *Bam* HI; Bg = *Bgl* II; Ps = *Pst* I; S = *Sal* I.
. . . . = vector DNA; _____ = insert DNA; | = restriction site.

Only a single *Sal* I digest was necessary to position the *Sal* I site of the viral insert DNA. pUC 18 has only one *Sal* I site - approx. 50 bp from the *Bam* HI site in the vector's multiple cloning site. Thus there was only one feasible position for the *Sal* I site in the viral insert, shown in Fig. A.I d.

The *Bgl* II / *Sal* I digest served to confirm the positioning of the *Bgl* II and *Sal* I restriction sites in the viral insert DNA (Table A.I; Fig. A.I e). The restriction

fragment sizes expected from such a digest were obtained. The smallest fragment (380 bp) corresponded with the expected *Sal* I/*Sal* I fragment. The large fragment (3050 bp) comprised vector DNA (from S^a to B^b) and the adjoining small *Bam* HI - *Bgl* II (B^b-Bg) portion of the insert DNA. The 2000 bp fragment corresponds with the expected size of the internal *Sal* I - *Bgl* II fragment.

A.15. Immunosorbent electron microscopy.

Immunosorbent electron microscopy (ISEM) was performed essentially as described by von Wechmar & Milne (1983).

Antiserum diluted to 1:1000 in 0.1 M Na acetate buffer pH 5.0 was adsorbed to carbon-coated grids. Purified virus preparation (A.2) was subsequently adsorbed to the grids, followed by decoration with antiserum diluted 1:10 in the same buffer. Grids were washed thoroughly with buffer between adsorption steps. The grids were stained with 2% (w/v) uranyl acetate made up in NaOAc buffer pH 5.0 and viewed in a Zeiss EM109 transmission electron microscope at 50000 x magnification.

A.16. SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE).

SDS-PAGE was performed essentially as described by Laemmli (1970).

Gel slabs (1.5 mm thick) comprising a 12% resolving gel and a 4% stacking gel were poured in a Hoefer SE-600 vertical slab gel apparatus (Hoefer Scientific Instruments).

Samples to be electrophoresed were disrupted by being mixed with an equal volume of protein dissociation mix (10% [w/v] SDS, 10% 2-mercaptoethanol, 15% glycerol, 0.01% bromophenol blue, 0.125 M Tris-HCl pH 6.8) and heated at 100°C for 10 min.

Electrophoresis was carried out with a constant current of 30 mA until the tracking dye front reached the gel base.

Gels were stained with 0.2% (w/v) PAGE blue 83 (BDH chemicals) dissolved in a 45:45:10% (v/v) mixture of methanol, water and glacial acetic acid, and destained with the same mixture without the PAGE blue 83.

A.17. Electroblotting.

Gels were electroblotted onto nitrocellulose membranes (Schleicher & Schuell, BA85, 0.45 mm pore) essentially as described by Rybicki & von Wechmar (1982), but using a "semi-dry" or horizontal electrode assembly.

The electroblotting apparatus was assembled as follows: a 1 cm thick layer of nappy liners (Johnson & Johnson) was laid on a horizontal 1 cm thick carbon electrode (the anode), followed by 2 sheets of 3MM filter paper, a sheet of nitrocellulose, the polyacrylamide gel, another layer of nappy liners, and lastly the second carbon electrode (the cathode). All components of this assembly were pre-soaked in transfer buffer (0.375 M Tris-HCl pH 8.3, 0.192 M glycine, 20% methanol). Transfer was allowed to proceed for 1 hour with a current of approximately 1 amp.

A.18. Indirect Immunoassay.

The method followed was essentially that of Williamson *et al.* (1988). All steps were carried out at room temperature, with constant agitation to ensure even reaction of reagents with blots.

Nitrocellulose blots were blocked for 2 hours at 37°C in blocking buffer (10 mM Tris-HCl pH 7.4, 0.15 M NaCl, containing 5% non-fat milk powder (Protea, RSA) and 0.05% Tween-20). Blots were then incubated for 2 hours at room temperature with host-absorbed antiserum diluted 1:500 in blocking buffer. After three 5 min washes in 10 mM Tris-HCl, 0.15 M NaCl, 0.05% Tween-20, pH 7.4, the blots were incubated for 2 hours at room temperature in goat-anti-rabbit-alkaline phosphatase conjugate (Seravac Labs, Cape Town) diluted 1:5000 in blocking buffer. After a further three washes (as above), phosphatase activity was assayed using the substrates Nitro Blue tetrazolium (Sigma, 100 µl of a 75 mg/ml stock in 70% dimethylformamide) and 5-bromo-4-chloro-3 indolyl phosphate (Sigma, 100 µl of a 50 mg/ml stock in 100% formamide) in substrate buffer (100 mM Tris-Cl buffer pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$). Blots were not shaken during this step.

APPENDIX B

SEQUENCING METHODS.

CONTENTS

APPENDIX B
SEQUENCING METHODS.

B.1. Progressive shortening of cloned DNA in pUC or Bluescript vectors.

The method of Henikoff (1984) was followed. Buffers used are given below.

Twelve µg DNA was digested with appropriate enzymes (A.5) to obtain cleavage of two restriction sites in the vector polylinker sequence, such that one 5' overhanging end (from which to shorten the insert) and one 3' overhang (to protect the vector from being shortened) were obtained, the 5'overhang being nearest the insert. It was necessary for both sites to be situated at the same end of the insert, with neither site occurring in the insert DNA being shortened.

The digested DNA was Genecleaned (A.12) and resuspended in 100 µl exonuclease III buffer and equilibrated at 37°C. A 9 µl sample was removed to one of 11 tubes kept on ice, each containing 25 µl S1 mix. 300 units of exonuclease III were added to the equilibrated DNA mix at 37°C and 9 µl samples were removed to the remaining S1 tubes at 20 sec intervals. Faster or slower rates of shortening were achieved by manipulating reaction temperature, amount of enzyme used, and incubation time between removal of aliquots. S1 mixes were raised to room temperature for 30 min. The S1 reaction was stopped by addition of 3.5 µl S1 stop solution and holding at 70°C for 10 min. Six µl of every fraction was checked for shortening on a 0.8% agarose gel. Klenow mix (3.4 µl) and 4 units Klenow enzyme were added to each tube and left at room temperature for 3 min. A dNTP mix (0.125 mM per µg DNA) was added and the tubes left a further 5 min. To each tube was added 120 µl ligation mix, and ligation was allowed to proceed overnight at room temperature. A 10 µl aliquot of ligation mix was used to transform 100 µl of LK112 competent cell suspension (A.10).

Solutions for Exonuclease shortening:

10 x S1 buffer

1.1 ml 3 M NaOAc (pH 4.6)
5.0 ml 5 M NaCl
5.0 ml glycerol
30 mg $ZnSO_4$
(autoclave)

Exonuclease III buffer

660 µl 1 M Tris-HCl (pH 8.0)
66.4 µl 0.1 M $MgCl_2$
9.27 ml dH_2O
(autoclave)

S1 mix (for 12 tubes)

41 µl 10 x S1 buffer
259 µl sterile dH_2O
60 units S1 nuclease

S1 stop solution

0.30 M Tris base
0.05 M EDTA
(do not pH since a high pH is needed to inactivate the S1 nuclease)

Klenow mix

20 mM Tris-HCl (pH 8.0)
7 mM $MgCl_2$
(autoclave)
(make just before use)

Ligase mix (for 12 tubes)

180 µl 10x ligation buffer (A.8)
24 µl T4 ligase (1 unit / µl)
1.236 ml sterile dH_2O

B.2. DNA Sequencing Protocol.

Template DNA was prepared for sequencing as in A.11.d., then digested with DNase-free RNase A (1 µl of a 10 mg/ml aqueous stock solution) for 30 min on ice, followed by one phenol/chloroform extraction and EtOH precipitation.

The Sequenase kit (United States Biochemical Corporation) was used for preparing templates for sequencing. The protocol followed was modified from the Sequenase protocol for double-stranded DNA sequencing. In all steps, complete mixing of reagents was essential. Reaction volumes given below are for a single template, but scaled-up reaction mixes could prepared for dispensing to individual templates so that up to 8 sequencing reactions could be performed in any one session.

B.2.1. Alkaline Denaturation of dsDNA templates.

Five µg of template DNA was diluted to 20 µl with dH_2O; 2 µl of 2 N NaOH was added, and the tube was kept at room temperature for 5 min. The denatured DNA was precipitated by addition of 80 µl neutralizing/precipitating solution (5% 2 M ammonium acetate pH 7.5, 95% EtOH [96%]) with incubation on ice for 10 min, followed by centrifugation for 20 min at 4°C. The pellet was gently rinsed with 70% EtOH and dried.

B.2.2. Priming Reaction.

To the dried denatured template was added 7 µl dH_2O, 2 µl Sequenase buffer (supplied), and 1 µl of appropriate primer (1 pmol/µl). The mixture was placed in a 65°C water bath for 2 min and allowed to cool slowly to room temperature (at least 30 min).

B.2.3. Labelling and Termination Reactions.

Before starting, the termination reaction tubes were prepared. 2.5 µl termination mix (ddNTP/dNTP, as supplied) and 0.25 µl DMSO was added to each of 4 tubes, such that the ddATP, ddCTP, ddGTP, ddTTP termination mixes were added to tubes labelled A, C, G, and T, respectively. The termination mixes were prewarmed at 45°C, 30 sec before addition of labelling mix.

To the template/primer complex were added in order; 1 µl 0.1 M dithiothreitol (supplied), 2 µl of a 1/12 dilution in TE of the supplied dGTP labeling mix, 0.5 µl ^{35}S-dATP (Amersham, 400 mCi/mol), 2 µl of a 1/8 dilution in TE of the Sequenase enzyme (modified T7 DNA polymerase), and 1.7 µl DMSO. The labelling mix was incubated for 4 min in a water bath at 18°C. Thereafter, 4.3 µl of labelling mix was added to each of the prewarmed termination mixes and incubated at 45°C for 5 min. The reaction was stopped by addition of 5 µl stop buffer (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol; supplied).

Reactions were either electrophoresed immediately on a sequencing gel, or stored at -20°C for up to 1 week.

B.2.4. Gel Preparation and Gel Pouring.

Gel apparatus, including gel plate assemblies and electrophoresis equipment, was supplied by Omeg Scientific, Cape Town.

Glass plates were not silanised as it was found that if plates were washed well with Contrad (Merck S.A. Pty Ltd) and rinsed thoroughly with water immediately after use, the upper glass plate could be lifted from the gel surface without sticking.

Back (34 cm wide x 42 cm long x cm thick) and front (as per back plate, but with a 2.5 cm deep x 29 cm wide cut-away at the top edge) glass gel plates were assembled with plastic spacer strips (1 cm x 42 cm) between them at each side of the plates (giving a gel thickness of 0.2 mm). A small spacer (1 cm x 1 cm x 0.1 mm) was inserted between the gel plates at the centre of the lower edge. The whole assembly was taped securely along all edges, excepting the top.

Seventy ml of sequencing gel mix (C.11), 60 µl of 50% ammonium persulphate (freshly prepared weekly), and 60 µl TEMED were mixed in a beaker, avoiding bubbles, the TEMED being added just before pouring. Using a 60 ml syringe, the gel mix was poured smoothly and continuously into the plate assembly which was held at a 60° angle, taking care not to introduce bubbles. The assembly was then laid flat and the straight-edged slot-former (rubbed with a little TEMED) was inserted 4 mm into the top of the assembly and clamped in position. After gel polymerization (usually 30 min), the gel assembly was unclamped and untaped and clamped onto the gel electrophoresis stand. The top and bottom reservoirs of the stand were filled with 1 x TBE (C.7). The slot-former was removed and the "sharkstooth" comb inserted with its teeth just piercing the gel surface.

B.2.5. Electrophoresis of DNA Sequencing Reactions.

TBE heated to approximately 65°C was added to the top and bottom reservoirs of the gel apparatus. This eliminated the necessity of pre-running the gel at 30 mA for 30 min to heat it to the running temperature. The TBE cooled to 55°C during pouring, and cooled gradually to 50°C during sample loading.

Templates, prepared as in B.2.1 - B.2.3, were heat-denatured at 80°C in a water bath for 2 min immediately prior to loading onto the gel. Samples were loaded in groups of 4 (A, C, G, and T of one template, in that order) with the power off, and

then immediately run into the gel at 900 - 1200 V. This both stopped the samples from diffusing out of the wells and the gel from cooling significantly. Generally, 2.5 µl of sample was loaded per well.

The samples were electrophoresed at a constant power of 96 watts - though this was adjusted to 98 or 100 watts if the gel cooled down, or to 94 watts if the gel heated up. Samples were electrophoresed either until the first loading dye front (bromophenol blue) reached the base of the gel (a "short" run - generally 45 min), or until the second dye front (xylene cyanol) reached the gel base (a "medium" run - generally 90 min). Thereafter, a fresh aliquot of tracking dye could be loaded into the sample well and the sample electrophoresed until the new bromophenol blue front reached the base of the gel (a "long" run - approximately 130 min). The gel temperature was held at 45° - 50°C for short runs, and at 40°C for medium and long runs.

After electrophoresis, the gel assembly was dismantled from the stand and laid on a flat surface. The upper plate was carefully lifted from the gel resting on the lower plate. A sheet of Whatmans 3MM paper was laid over the gel and gently smoothed in order to attach the gel to the paper. The paper with attached gel was lifted from the lower plate and the gel surface covered with clingfilm (3M Corporation). The gel was dried under vacuum at 80°C.

B.2.6. Autoradiography of DNA Sequencing Gels.

The dried paper-backed sequencing gel (B.2.5) was exposed to Curix X-ray film (Agfa-Gevaert) in an autoradiography cassette overnight, or longer if necessary. X-ray film was developed with by immersion in X-ray developing fluid (Agfa, Agfatech, R.S.A.) for 3 min, a brief washing in water to rinse off excess developer, and a 3 min immersion in X-ray fixative (Agfa, Agfatech, R.S.A.) followed by a thorough wash in water and air drying. DNA sequence was read from the autoradiograph and entered into an IBM-compatible computer.

APPENDIX C

STOCK SOLUTIONS AND MEDIA.

CONTENTS

APPENDIX C.
STOCK SOLUTIONS AND MEDIA.

C.1. Luria-Bertani Broth (LB) and Agar (LA).

10 g tryptone

5 g yeast extract

5 g NaCl

15 g agar

Make up to 1 l with deionized water. Autoclave.

If desired, add 10 ml/l of 20% filter-sterilized glucose after autoclaving.

C.2. Ampicillin.

25 mg/ml aqueous stock of the sodium salt.

Filter-sterilize. Store at -20°C.

Used at 50 µg/ml in LB. Add after autoclaving.

Used at 100 µg/ml in LA. Allow autoclaved LA to cool to 55°C before adding ampicillin.

C.3. X-gal Plates.

To the surface of a LA (Amp100) plate, spread 35 µl X-gal (20 mg/ml in dimethylformamide. store at -70°C) and 7 µl IPTG (23.8 mg/ml in dH_2O. store at -20°C) with 100 µl TE. The plates must be spread with X-gal/IPTG at least 30 min before use, to avoid toxicity to bacterial cells.

X-gal = 5-bromo-4-chloro-3-indolyl-B- galactosidase.

IPTG = isopropyl-B-D-thio-galactopyranoside.

C.4. Tris-HCl / EDTA Buffer (TE).

10 mM Tris-HCl pH 7.5 or 8.0.

1 mM EDTA pH 8.0

(autoclave)

C.5. Restriction Endonuclease Buffers.

RE-0, -50, -100, -150 made up according to New England Biolabs (1988/1989 catalogue). *Sma* I buffer made up according to Maniatis *et al.* (1982).

All restriction buffers were filter-sterilized.

RE-0.

10 mM Tris-HCl pH 7.5
10 mM $MgCl_2$
100 µg/ml BSA*

RE-50.

10 mM Tris-HCl pH 7.5
10 mM $MgCl_2$
100 µg/ml BSA
50 mM NaCl

RE-100.

10 mM Tris-HCl pH 7.5
10 mM $MgCl_2$
100 µg/ml BSA
100 mM NaCl

RE-150.

10 mM Tris-HCl pH 7.5
10 mM $MgCl_2$
100 µg/ml BSA
150 mM NaCl

(* BSA = bovine serum albumin)

Sma I Buffer.

20 mM KCl
10 mM Tris-HCl pH 8.0
10 mM $MgCl_2$
1 mM dithiothreitol

C.6. Salt and Temperature Preferences of Restriction Endonucleases used in studies detailed in this thesis.

From specifications listed in the New England Biolabs 1988/1989 catalogue.

Enzyme	Buffer Used	Usable Buffer Range*
Apa I	RE-0	RE-0 - RE-50
Bam HI	RE-150	RE-100 - RE-150
Bgl I	RE-100	RE-50 - RE-150
Bgl II	RE-100	RE-50 - RE-150
Cla I	RE-50	RE-0 - RE-100
Eco RI	RE-50	RE-50 - RE-150
Hae III	RE-50	RE-0 - RE-150
Hind III	RE-50	RE-50 - RE-100
Kpn I	RE-0	RE-0
Pst I	RE-100	RE-0 - RE-150
Pvu II	RE-100	RE-0 - RE-150
Rsa I	RE-50	RE-0 - RE-150
Sac I	RE-0	RE-0
Sal I	RE-150	RE-150
Sma I	Sma I buffer	
Xho I	RE-150	RE-50 - RE-150

* = 30 - 100% digestion in 60 min at optimum temperature.

Optimum reaction temperature for *Apa* I = 30°C.

Optimum temperature for *Sma* I = 25°C.

Optimum temperature for all other restriction enzymes shown above = 37°C.

C.7. 5x Tris-Borate-EDTA Buffer (TBE).

54 g Tris base

27.5 g boric acid

20 ml EDTA (0.5 M pH 8.0)

Make up to 1 litre with deionized water. When diluted to 1x, pH = 7.8.

C.8. Agarose Gel Loading Buffer.

1x TBE

0.25% bromophenol blue filter-sterilize.

40% (w/v) sucrose

C.9. 20x SSC (salt sodium citrate buffer).

3.0 M NaCl

0.3 M Na_3 citrate pH 7.0

C.10. 40% Acrylamide Stock.

38 g acrylamide monomer

2 g methylene bisacrylamide

dH_2O to 100 ml

C.11. Sequencing Gel Mix.

74 ml acrylamide stock (40%)

239 g urea

100 ml 5x TBE

dH_2O to 500 ml

Filter through Whatmans no. 1 paper. Store at 4°C or at room temperature in the dark. Prepare weekly, but the mix can last up to 3 weeks if necessary.

REFERENCES

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 J.* **8**, 1033-1039.

AMMAR, E.D., KIRA, M.T., & ABUL-ATA, A.E. (1989). Natural occurence of streak and mosaic diseases on sugarcane cultivars at Upper Egypt and transmission of sugarcane streak by *Cicadulina bipunctella zeae* China. *Ann. Virol.* **133**, 183-185.

ARNIM, A.V., ELLWOOD, S., & STANLEY, J. (1990). Determinants of tomato golden mosaic virus symptom development are encoded on DNA B. In: *VIIIth International Congress of Virology, Berlin, Abstracts*, p 466.

ANDERSEN, M.T., RICHARDSON, K.A., HARBISON, S-A., & MORRIS, B.A.M. (1988). Nucleotide sequence of the geminivirus chloris striate mosiac virus. *Virology* **164**: 443-449.

AOKI, K., TATENO, Y., & TAKAHATA, N. (1981). Estimating evolutionary distances from restriction maps of mitochondrial DNA with arbitrary G+C content. *J. Molec. Evol.* **18**, 1-8.

AUTREY, L.J-C., & RICAUD, C. (1983). The comparitive epidemiology of two diseases of maize caused by leafhopper-borne viruses in Mauritius. pp. 277-285. In: *Plant Virus Epidemiology* (Eds., R.T. Plumb, J.M. Thresh). Blackwell, Oxford.

BABAYA, M. (1990). *Studies on transgenic tobacco plants containing geminiviral DNA.* M. Sc. thesis, Dept. Microbiology, Univ. Cape Town, S. Africa.

BISHOP, M.J., FRIDAY, A.E., & THOMPSON, E.A. (1987). Inference of evolutionary relationships. pp. 359-385. In: *Nucleic Acid and Protein Sequence Analysis, A Practical Approach.* (Eds. M.J. Bishop & C.J. Rawlings). IRL Press, Eynsham, Oxford, U.K.

BISZTRAY, G., & GABORJANYI, R. (1989). Isolation and characterization of wheat dwarf virus found for the first time in Hungary. *Z. Pflanzenkr. Pflanzenschutz.* **96**, 449-454.

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

BOCK, K.R., & BAILEY, R.A. (1989). Streak. pp. 322-329. In: *Diseases of Sugarcane; Major Diseases* (Eds., C. Ricaud, B.T. Egan, A.G. Gillespie Jr., & C.G. Hughes). Elsevier, Amsterdam, Oxford, New York, Tokyo.

BOCK, K.R., GUTHRIE, E.J., & MEREDITH, G. (1978). Distribution, host range, properties and purification of cassava latent virus, a geminivirus. *Ann. Appl. Biol.* **90**, 361-367.

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*. *Ann. Appl. Biol.* **77**, 289-296.

BOULTON, M.I., KING, D.I., & DAVIES, J.W. (1990). A single nucleotide change affects the pathogenicity of maize streak virus. In: *VIIIth International Congress of Virology, Berlin, Abstracts*, p126.

BOULTON, M.I., & MARKHAM, P.G. (1986). The use of squash-blotting to detect plant pathogens in insect vectors. pp. 55-69. In: *Developments in Applied Biology 1. Developments and Applications in Virus Testing* (Eds., R.A.C. Jones & L. Torrance). Assoc. Appl. Biologists, Wellesbourne, U. K.

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. *J. Gen. Virol.* **70**, 2309-2323.

BRIDDON, R.W., PINNER, M.S., STANLEY, J., & MARKHAM, P.G. (1990a). Geminivirus coat protein gene replacement alters insect specificity. *Virology* **177**, 85-94.

BRIDDON, R.W., STANLEY, J., & MARKHAM, P.G. (1990b). Investigation of the specificity of geminivirus insect transmission using chimeric viruses. In: *VIIIth International Congress of Virology, Berlin, Abstracts*, p126.

CHATANI, M., MATSUMOTO, Y., MIZUTA, H., IKEGAMI, M., BOULTON, M.I., & DAVIES, J. (1990). Nucleotide sequences of infectious clones of the geminivirus miscanthus streak virus. In: *VIIIth International Congress of Virology, Berlin, Abstracts*, p466.

CHUNG, C.T., & MILLER, R.H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucl. Acids Res.* **16**, 3580.

CLARK, M.F., & BARBARA, D.J. (1987). A method for the quantitative analysis of ELISA data. *J. Virol. Methods* **52**, 173-176.

CLARK, M.F., & BAR-JOSEPH, M. (1984). ELISA in plant virology. *Methods in Virology* 7, 51-85.

CLARKE, B.A. (1987). *Molecular biological studies of southern African isolates of maize streak virus*. M. Sc. thesis, Dept. Microbiology, Univ. Cape Town, S. Africa.

CLARKE, B.A., RYBICKI, E.P., HUGHES, F.L., KIRBY, R., & VON WECHMAR, M.B. (1989). Characterization of southern African isolates of maize streak virus : typing of three isolates by restriction mapping. *Intervirology* **30**, 86-95.

DABROWSKI, Z.T. (1987a). Two new species of *Cicadulina* China Hemiptera Euscelidae from West Africa. *Bull. Entomol. Res.* **77**, 53-56.

DABROWSKI, Z.T. (1987b). *Cicadulina-ghauri* Hemiptera Euscelidae. Distribution, biology and maize streak virus (MSV) transmission. *J. Appl. Entomol.* **103**, 489-496.

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

DAVIES, J.W., & STANLEY, J. (1989). Geminivirus genes and vectors. *Trends Genet.* **5**, 77-81.

DEKKER, E.L., PINNER, M.S., MARKHAM, P.G., & VAN REGENMORTEL, M.H.V. (1988). Characterisation of maize streak virus isolates from different plant species by polyclonal and monoclonal antibodies. *J. Gen. Virol.* **69**, 983-990.

DEVEREUX, J., HAEBERLI, P., & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.* **12**, 387-395.

DOLLET, M., ACCOTTO, G.-P., LISA, V., MENISSIER, J., & BOCCARDO, G. (1986). A geminivirus, serologically related to maize streak virus, from *Digitaria sanguinalis* from Vanuatu. *J. Gen. Virol.* **67**, 933-937.

DONSON, J., ACCOTTO, G.-P., BOULTON, M.I., MULLINEAUX, P.M., & DAVIES, J.W. (1987). The nucleotide sequence of a geminivirus from *Digitaria sanguinalis*. *Virology* **161**, 160-169.

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 J.* **3**, 3069-3073.

DONSON, J., GUNN, H.V., WOOLSTON, C.J., PINNER, M.S., BOULTON, M.I., MULLINEAUX, P.M., & DAVIES, J.W. (1988). *Agrobacterium*-mediated infectivity of cloned digitaria streak virus DNA. *Virology* **162**, 248-250.

EDGE, E. (1989). *Molecular studies on two isolates of maize streak virus in South Africa.* Hons. thesis, Dept. Microbiology, Univ. Cape Town, S. Africa.

FELSENSTEIN, J. (1984). Distance methods for inferring phylogenies: a justification. *Evolution* **38**, 16-24.

FELSENSTEIN, J. (1988). Phylogenies from molecular sequences: Inference and reliability. *Annu. Rev. Genet.* **22**, 251-265.

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