# STUDIES ON THE SUBGENOMIC DNA OF TOMATO GOLDEN MOSAIC VIRUS AND THE SEQUENCE OF WHEAT DWARF VIRUS DNA.

A thesis submitted for the degree of Doctor of Philosophy of the University of London and for the Diploma of Membership of the Imperial College.

by

Samuel Wallace MacDowell

September, 1986

Department of Pure and Applied Biology,
Imperial College,
London, S.W.7

**Abstract**

The single-stranded (ss) DNA form of a subgenomic species of the bipartite geminivirus tomato golden mosaic virus (TGMV) was shown by Southern blot and electron microscopic heteroduplex analyses to be derived from the B component by a highly specific deletion.

The covalently closed circular double-stranded (ds) form of the subgenomic DNA isolated from TGMV infected tobacco (*Nicotiana benthamiana*) plants was cloned into plasmid pEMBL 9 and the deletion end-points of 22 clones determined by dideoxy chain-termination sequencing. One of the clones was sequenced in its entirety.

Computer analysis of the sequence data revealed four types of clones that differed with respect to the location of the deletion. Comparison with the sequence of cassava latent virus (CLV) subgenomic DNA showed that equivalent regions of the genome were deleted in both cases but common features at the nucleotide level were not demonstrated.

Preliminary tobacco infectivity studies using the two TGMV genomic components in combination with cloned subgenomic DNA suggested that the latter replicates and delays symptom expression.

Restriction analysis, cloning and sequencing of virus-specific dsDNA isolated from wheat dwarf virus (WDV) infected wheat (*Triticum aestivum*) tissue, indicated that, similar to maize streak virus (MSV), the WDV genome apparently consists of a single DNA circle of 2749 nucleotides. Although WDV and MSV are serologically unrelated their putative capsid proteins showed 35% direct amino acid sequence homology and their DNAs 46% nucleotide sequence homology. The genomic organization of the two DNAs appears similar.

A comparison of the amino acid sequence of predicted proteins of WDV, MSV, TGMV and CLV revealed clear relationships between these viruses and suggested that the unipartite and bipartite geminiviruses have a common ancestral origin.

## Table of contents


| | | Page |
|---|---|---|
| Title | | 1 |
| Abstract | | 2 |
| Table of contents | | 3 |
| Index of figures | | 7 |
| Index of tables | | 11 |
| 1. | Introduction | 13 |
| 1.1 | Geminivirus history | 13 |
| 1.2 | Transmission | 19 |
| 1.3 | Geographical distribution and host range | 20 |
| 1.4 | Symptomatology and cytology | 21 |
| 1.5 | Relationships among geminiviruses | 23 |
| 1.6 | Virus structure and composition | 25 |
| 1.7 | Genome size and infectivity | 28 |
| 1.8 | Genome organization and comparison | 30 |
| 1.8.1 | Sequence determination | 30 |
| 1.8.2 | Intergenic regions | 31 |
| 1.8.3 | Coding regions | 32 |
| 1.9 | Transcription | 37 |
| 1.10 | Replication | 40 |
| 1.11 | Defective particles | 44 |
| 1.12 | Defective interfering viruses | 46 |
| 1.13 | Bacteriophage f1 | 48 |
| 1.14 | Adenoviruses | 49 |
| 1.15 | Parvoviruses | 50 |
| 1.16 | Papovaviruses | 51 |
| 1.17 | Herpesviruses | 53 |
| 1.18 | Defective RNA plant viruses | 54 |
| 1.19 | Mechanism for the generation of DI particles | 55 |
| 1.20 | Introduction to the present work | 59 |
| 2. | Materials and methods | 64 |
| 2.1 | Chemical supplies | 64 |
| 2.2 | Abbreviations | 66 |
| 2.3 | Safety considerations | 70 |

| | | |
|---|---|---|
| 2.4 | Common buffers | 71 |
| 2.5 | Miscellaneous | 71 |
| 2.6 | Phenol extraction and ethanol precipitation | 72 |
| 2.7 | Quantitation of nucleic acids | 73 |
| 2.7.1 | Spectrophotometric determination | 73 |
| 2.7.2 | Determination by UV induced fluorescence on agarose plates | 73 |
| 2.7.3 | Trichloroacetic acid (TCA) precipitation of labelled DNA | 73 |
| 2.8 | Sephadex chromatography | 74 |
| 2.8.1 | Sephadex G-50 column chromatography | 74 |
| 2.8.2 | "Spun-column" chromatography | 75 |
| 2.9 | Propagation of tomato golden mosaic virus | 75 |
| 2.9.1 | Growth of plants | 75 |
| 2.9.2 | Origin of tomato golden mosaic virus | 75 |
| 2.9.3 | Preparation of viral inoculum | 76 |
| 2.9.4 | Inoculation of plants | 76 |
| 2.10 | Extraction of DNA | 76 |
| 2.10.1 | Crude DNA preparations of tomato golden mosaic virus | 76 |
| 2.10.2 | Origin of wheat dwarf virus | 77 |
| 2.10.3 | Crude DNA preparations of wheat dwarf virus | 77 |
| 2.11 | Cesium chloride-ethidium bromide gradienting of DNA | 77 |
| 2.12 | Hydroxyl apatite chromatography | 78 |
| 2.13 | Agarose gel electrophoresis | 79 |
| 2.14 | Polyacrylamide gel electrophoresis | 80 |
| 2.15 | Methyl mercury gels | 80 |
| 2.16 | Recovery of DNA fragments from gels | 81 |
| 2.16.1 | Electroelution into dialysis bags | 81 |
| 2.16.2 | Electroelution onto DEAE paper | 81 |
| 2.16.3 | Electroelution with ISCO cups | 82 |
| 2.16.4 | Freeze squeeze | 83 |
| 2.17 | Purification of dsDNA with Elutip-d minicolumns | 84 |
| 2.18 | Nuclease reactions | 85 |
| 2.19 | Restriction endonuclease digestions | 85 |
| 2.20 | Reverse transcriptase reactions | 87 |

| | | |
|---|---|---|
| 2.21 | Nick translation | 88 |
| 2.22 | Preparation of ssDNA specific probes | 89 |
| 2.23 | DNA 'end-filling' reaction | 89 |
| 2.23.1 | Conversion of fragments with protruding 5'ends to blunt ends | 90 |
| 2.23.2 | Conversion of fragments with protruding 3'ends to blunt ends | 90 |
| 2.24 | DNA transfer and hybridization | 91 |
| 2.24.1 | Transfer to nitrocellulose paper | 91 |
| 2.24.2 | Transfer to GeneScreen | 92 |
| 2.24.3 | Transfer to GeneScreen Plus | 92 |
| 2.24.4 | Hybridization with nitrocellulose paper membranes | 93 |
| 2.24.5 | Hybridization with GeneScreen membranes | 94 |
| 2.24.6 | Hybridization with GeneScreen Plus membranes | 96 |
| 2.24.7 | Rehybridizations with GeneScreen Plus membranes | 97 |
| 2.25 | Colony lift hybridization | 97 |
| 2.26 | Large scale plasmid isolation | 98 |
| 2.27 | Small scale isolation of plasmid DNA | 99 |
| 2.28 | Electron microscopy | 100 |
| 2.29 | Cloning | 102 |
| 2.29.1 | Origin and storage of host strain bacteria and vector DNA | 102 |
| 2.29.2 | Media and antibiotics | 102 |
| 2.29.3 | Ligations | 103 |
| 2.29.4 | Preparation of competent JM101 cells | 104 |
| 2.29.5 | Transformation | 104 |
| 2.29.5.1 | Transformation with pEMBL vectors | 104 |
| 2.29.5.2 | Transformation with M13 vectors | 105 |
| 2.30 | Preparation of template DNA for DNA sequencing | 105 |
| 2.31 | Nucleotide sequencing | 106 |
| 2.31.1 | Stock solutions used for sequencing | 106 |
| 2.31.2 | Sequencing reactions | 108 |
| 2.31.3 | T-tracking | 109 |
| 2.31.4 | Sequencing gel electrophoresis | 110 |

| | | |
|---|---|---|
| 2.31.5 | Computer analysis | 111 |
| 3. | Results and discussion | 112 |
| 3.1 | Intracellular TGMV DNA species | 112 |
| 3.2 | Isolation of TGMV subgenomic ssDNA | 114 |
| 3.3 | Size estimate of TGMV subgenomic ssDNA | 114 |
| 3.3.1 | Size estimate by agarose gel electrophoresis | 114 |
| 3.3.2 | Size estimate by electron microscopy | 118 |
| 3.4 | Southern blot analysis of TGMV subgenomic ssDNA | 121 |
| 3.5 | Heteroduplex analysis of TGMV subgenomic ssDNA | 131 |
| 3.6 | Characterization of TGMV subgenomic dsDNA | 135 |
| 3.6.1 | Hydroxylapatite chromatography of TGMV crude DNA | 135 |
| 3.6.2 | Specificity of TGMV subgenomic dsDNA | 137 |
| 3.6.3 | Nuclease studies on TGMV subgenomic dsDNA | 144 |
| 3.7 | Molecular cloning of TGMV subgenomic dsDNA | 147 |
| 3.8 | Sequence of the deletion borders | 152 |
| 3.9 | Sequence of pB530 | 165 |
| 3.10 | Comparison to other geminiviruses | 174 |
| 3.11 | Vector construction and infectivity studies | 177 |
| 3.11.1 | Infection with cloned subgenomic DNA | 178 |
| 3.11.2 | Construction of pTBneo | 180 |
| 3.11.3 | Infection with pTBneo | 186 |
| 3.12 | Is encapsidated TGMV subgenomic DNA a DI particle? | 193 |
| 3.13 | Possible mechanisms for the formation of TGMV subgenomic DNA | 195 |
| 3.14 | Purification of WDV dsDNA | 205 |
| 3.15 | Characterisation of intracellular WDV DNA species | 207 |
| 3.16 | Restriction analysis of WDV oc dsDNA | 212 |
| 3.17 | Molecular cloning of WDV dsDNA | 214 |
| 3.18 | Sequencing strategy | 216 |
| 3.19 | Potential genes | 220 |
| 3.20 | Potential transcriptional control signals | 226 |
| 3.21 | Non-coding regions | 232 |
| 3.22 | Comparisons with other geminiviruses | 247 |

3.23 Relationships among the geminiviruses 266
Acknowledgements 270
References 271

## Index of Figures

1 Comparison of the geminiviral ORFs 32
2 Alignment of the highly conserved hairpin structures in the 'common regions' of TGMV, CLV and BGMV 33
3 CLV 2 showing the extent of the nucleotide deletions in the defective molecules 45
4 Replicative error model for the generation of DI particles containing simple deletions 57
5 Replicative error model for the generation of DI particles containing complementary termini 57
6 Replicative error model for the generation of 'snap-back' DI particles 57
7 Model for topo I mediated excisional recombination 57
8 'Slipped mispairing' model for the generation of deletions during DNA replication 60
9 Intracellular TGMV DNA species 113
10 Size estimation of TGMV subgenomic ssDNA 116
11 Size estimation of TGMV subgenomic ssDNA 117
12 Electron micrograph of gel purified TGMV subgenomic ssDNA 119
13 Electron micrograph of bacteriophage ØX174 ssDNA 120
14 Gel purification of TGMV subgenomic ssDNA 122
15 Examination of gel purified TGMV subgenomic ssDNA 124
16 Characterization of TGMV subgenomic ssDNA (Southern blot analysis) 126
17 Characterization of TGMV subgenomic ssDNA (Southern blot analysis) 127

| | | |
|---|---|---|
| 18 | Characterization of TGMV subgenomic ssDNA (Southern blot analysis) | 128 |
| 19 | Characterization of TGMV subgenomic ssDNA (Southern blot analysis) | 129 |
| 20 | Characterization of TGMV subgenomic ssDNA (heteroduplex analysis) | 132 |
| 21 | Characterization of TGMV subgenomic ssDNA (heteroduplex analysis) | 133 |
| 22 | Characterization of TGMV subgenomic ssDNA (heteroduplex analysis) | 134 |
| 23 | Hydroxylapatite chromatography (HAP) of TGMV crude DNA | 136 |
| 24 | Construction of recombinant plasmids containing specific fragments from TGMV DNAs A and B | 138 |
| 25 | Restriction analysis of recombinant plasmids containing specific fragments from TGMV DNAs A and B | 141 |
| 26 | Characterization of TGMV subgenomic dsDNA (specific probing) | 142 |
| 27 | Characterization of TGMV subgenomic dsDNA (nuclease sensitivity test) | 145 |
| 28 | Centrifugation of TGMV crude DNA in cesium chloride/ethidium bromide gradients | 146 |
| 29 | Restriction enzyme analysis of TGMV dsDNA | 148 |
| 30 | Size estimation of TGMV subgenomic dsDNA | 150 |
| 31 | Selection of recombinant plasmids containing TGMV subgenomic DNA | 151 |
| 32 | Restriction analysis of recombinant plasmids containing TGMV subgenomic DNA | 153 |
| 33 | Sequencing the nucleotides surrounding the deletion points in cloned subgenomic DNAs | 155 |
| 34 | Nucleotide sequence at the deletion points of 22 cloned subgenomic DNAs | 158 |
| 35 | Alignment of the predicted amino acid sequences encoded by TGMV B and subgenomic DNAs ORF BL1 | 163 |
| 36 | Sequencing pB530 | 167 |

| | | |
|---|---|---|
| 37 | Sequencing pB530 | 168 |
| 38 | Comparison of the TGMV B DNA and TGMV subgenomic DNA sequences | 170 |
| 39 | TGMV B DNA showing the region deleted in the subgenomic DNA | 173 |
| 40 | Intracellular viral DNA species in TGMV infected plants | 179 |
| 41 | Construction of pTBneo | 182 |
| 42 | Restriction analysis of pTBN | 184 |
| 43 | Restriction analysis of pTBneo | 185 |
| 44 | Infectivity studies | 189 |
| 45 | Infectivity studies | 190 |
| 46 | Infectivity studies | 191 |
| 47 | Direct repeats in the vicinity of the deletion points in the cloned subgenomic DNAs | 199 |
| 48 | Inverted repeats in the vicinity of the deletion points in the cloned subgenomic DNAs | 200 |
| 49 | Gel purification of WDV oc dsDNA | 206 |
| 50 | Intracellular WDV DNA species | 208 |
| 51 | Restriction analysis of WDV oc dsDNA | 209 |
| 52 | Centrifugation of WDV crude DNA in a cesium chloride/ethidium bromide gradient | 211 |
| 53 | DNA sequencing strategy | 217 |
| 54 | Restriction analysis of cloned WDV dsDNA | 218 |
| 55 | 'T-Tracks' of independently derived WDV DNA *Eco*RI subclones | 221 |
| 56 | Sequence of two independently derived *Eco*RI subclones | 222 |
| 57 | Nucleotide sequence of WDV | 223 |
| 58 | Open reading regions and open reading frames in the WDV DNA sequence | 227 |
| 59 | Potential coding regions in WDV DNA | 228 |
| 60 | Alignment of the AT-rich inverted repeat sequences containing putative polyadenylation signals in MSV-N DNA, TGMV DNAs A and B, BGMV DNAs 1 and 2 and CLV-K DNAs 1 and 2 | 235 |

| | | |
|---|---|---|
| 61 | Potential hairpin structures in WDV DNA | 238 |
| 62 | Comparisons of the major potential hairpin structures in WDV, MSV, BCTV, CLV, TGMV and BGMV DNAs | 239 |
| 63 | Comparison of the WDV and MSV-N DNA sequences | 241 |
| 64 | Sequence homology of WDV and MSV-N | 247 |
| 65 | Comparison of the genome organization of WDV, MSV and CLV | 251 |
| 66 | A) Alignment of the predicted amino acid sequences encoded by MSV ORF 10146 and WDV ORF 10906. B) Alignment of the predicted amino acid sequences encoded by MSV ORF 13011 and WDV ORF 14556 | 253 |
| 67 | Alignment of the predicted amino acid sequences of the coat proteins of WDV (29.4Kd), MSV-N (27.0Kd), TGMV (28.7Kd) and CLV-K (30.2Kd) | 254 |
| 68 | Alignment of the predicted amino acid sequences encoded by WDV ORF 30156 (30.2Kd) and OR 17292 (17.3Kd), MSV-N ORFs 31388 (31.3Kd) and 17768 (17.8Kd), TGMV ORF 40285 (40.3Kd) and CLV ORF 40300 (40.3Kd) | 256 |
| 69 | Alignment of the predicted amino acid sequences encoded by MSV ORF 21771 ORF and WDV OR2 | 258 |


## Index of Tables


| | | |
|---|---|---|
| 1 | Members of the geminivirus group | 15 |
| 2 | Alignment of the nucleotides surrounding the deletion points in 19 CLV subgenomic DNA clones | 47 |
| 3 | Nucleotide sequences at the deletion points of cloned TGMV subgenomic DNA and the corresponding regions of TGMV DNA B | 203 |
| 4 | Open reading frames in WDV DNA | 230 |

| | | |
|---|---|---|
| 5 | Nucleotide sequence at the predicted start sites for the WDV open reading frames | 230 |
| 6 | Most likely TATA boxes and polyadenylation signals for each ORF | 233 |
| 7 | Position of T/CCAAT sequences | 234 |
| 8 | Comparison of the amino acid sequences encoded by WDV ORFs with those of MSV, CLV and TGMV occupying similar relative positions in the respective DNAs | 249 |
| 9 | Potential splice sites for WDV ORF 30156 and OR 17292 | 263 |

To Paula and Pedro

# 1. Introduction

## 1.1. Geminivirus history

The geminiviruses comprise a group of plant-pathogenic viruses characterized by their twinned (geminate) particles and a genome of single-stranded (ss) DNA. The geminiviruses are endemic in tropical and warm temperate regions of the world where they cause some of the most serious diseases of crop plants. African cassava mosaic disease, for example, typically causes losses of about 70% in tuber yield of plants derived from infected cuttings (Bock, 1982). Epidemics of maize streak virus (MSV) can be devastating, causing almost total crop loss, especially when draught or crop maturity induces extensive migration of the leafhopper vector of the virus (Rose, 1978). Incidence of tomato yellow leaf curl virus (TYLCV) in autumn sown tomato crops in Jordan can reach 100%, causing crop losses of between 50 and 75% (Al-Musa, 1982).

Work on diseases now known to be caused by geminiviruses dates from the late 19$^{th}$ and early 20$^{th}$ century. Beet curly top disease was first reported in 1888 (cited in Larsen and Duffus, 1983). Warburg described cassava mosaic disease in East Africa in 1894 (cited in Storey and Nichols, 1938). Maize streak disease was first reported from South Africa in 1901 (Fuller, 1901) and tobacco leaf curl disease was known in Indonesia as early as 1912 (cited in Goodman, 1981 a).

However, it was not until the late 1970's that progress was made in virus purification and characterization, and the etiology of geminiviruses diseases better understood. Geminivirus particles were first clearly depicted by Bock *et al.* (1974) for MSV and by Mumford (1974) for beet curly top virus (BCTV). These authors reported an unusual capsid morphology for the virus consisting of small, geminate quasi-isometric units thereafter the name geminivirus was adopted (Harrison *et al.*, 1977). Evidence for a ssDNA genome was first presented by Goodman (1977 a,b) for bean golden mosaic virus (BGMV) and soon after by Harrison *et al.* (1977) for cassava latent virus (CLV) and MSV. In 1978 the geminiviruses were recognized by the International Committee on the Taxonomy of Viruses on the basis of their unique virion morphology and possession of a ssDNA genome (Matthews, 1979).

Interest in the molecular biology of the group, the realization of their potential as vectors for plant transformation and the availability of modern techniques for DNA manipulation and analysis resulted in rapid research progress. The number of both definitive and provisional members of the group has more than doubled over the last five years (Table I) and the complete nucleotide sequence of the genomes of four geminiviruses, tomato golden mosaic virus (TGMV), BGMV, MSV and CLV, are already known (Hamilton *et al.*, 1984; Howarth *et al.*, 1985; Mullineaux *et al.*, 1984; Stanley and Gay, 1983).

Definitive members of the group are those that are known to have geminate particles and possess circular ssDNA. At the present time seven viruses have been demonstrated as definitive members, namely BGMV (Galvez and Castano, 1976; Goodman 1977 a,b), CLV (Harrison *et al.*, 1977; Boch *et al.*, 1978), chloris striate mosaic virus (CSMV) (Francki *et al.*, 1979, 1980), MSV (Bock *et al.*, 1974; Harrison *et al.*, 1977) TGMV (Matyis *et al.*, 1975; Hamilton *et al.*, 1981) wheat dwarf virus (WDV) (Lindsten *et al.*, 1980; H. Macdonald, personnal communication) and potato yellow mosaic virus (PYMV) (Roberts *et al.*, 1986). Provisional members of the group have been identified on the basis of their particular morphology, cytopathology and a relationship to a definitive member of the group as indicated by serological or nucleic acid hybridization tests.

Bean golden mosaic disease was first reported from Brazil in 1965 (Costa, 1965). The disease is now known to be widespread in tropical and subtropical regions of the Americas (Bird and Maramorosh, 1975) and possibly in tropical Africa and Asia where similar diseases have been reported (Goodman and Bird, 1978). Transmission of the disease by *Bemisia Tabaci* (Genn) whiteflies has been known for some time (Costa, 1965) and sap transmission of the disease was reported some time later (Meiners *et al.*, 1973; Galvez and Castano 1976; Goodman *et al.*, 1977). The purification of a virus, BGMV, from infected plants in both Colombia and The United States and the demonstration that purified virus particles were infectious established BGMV as the causal agent of the disease (Galvez and Castano, 1976; Goodman *et al.*, 1977). The geminate morphology of the purified virus particles established BGMV as a geminivirus (Matthews, 1979). Further work demonstrated that BGMV had a ssDNA genome, the first report of a ssDNA containing plant virus (Goodman, 1977 a,b).

Table 1

## MEMBERS OF THE GEMINIVIRUS GROUP

| Virus[a] | Host range | Insect vector | Distribution | Ref. |
|---|---|---|---|---|
| | | Whitefly | | |
| Abutilon mosaic | Malvaceae | *Bemisia tabaci* | Tropical America | 1 |
| Bean golden mosaic | Leguminosae | *Bemisia tabaci* | Central & South America | 2 |
| Cassava latent | Euphorbiaceae | *Bemisia tabaci* | Africa & India | 3 |
| Cotton leaf crumple | Malvaceae | *Bemisia tabaci* | North America | 4 |
| Eupatorium yellow vein | Compositae | *Bemisia tabaci* | Japan | 5 |
| Euphorbia mosaic | Euphorbiaceae | *Bemisia tabaci* | America | 6 |
| Honeysuckle yellow vein mosaic | Capitoliaceae | *Bemisia loniceae* | Japan | 7 |
| Jatropha mosaic | Euphorbiaceae | *Bemisia tabaci* | Central America | 8 |
| Mungbean yellow mosaic | Leguminosae | *Bemisia tabaci* | South Asia | 9 |
| Potato yellow mosaic | Solanaceae | *Bemisia tabaci* | Venezuela | 10 |
| Solanum apical leaf curl | Solanaceae | | Peru | 11 |
| Soybean crinkle leaf | Leguminosae | *Bemisia tabaci* | Thailand | 12 |
| Squash leaf curl | Cucurbitaceae | *Bemisia tabaci* | Tropical America | 13 |
| Tobacco leaf curl | Solanaceae<br>Caprifoliaceae<br>Compositae | *Bemisia tabaci* | Pan Tropical | 14 |
| Tomato golden mosaic | Solanaceae | *Bemisia tabaci* | Tropical America | 15 |
| Tomato leaf curl & Tomato yellow leaf curl | Solanaceae | *Bemisia tabaci* | Mediterranean | 16,17 |
| Water melon curly mottle virus | Cucurbitaceae | *Bemisia tabaci* | North America | 18 |
| | | Leafhopper | | |
| Bean summer death | Leguminosae | *Orosius argentatus* | Australia | 19 |
| Beet curly top[b] | Chenopodiaceae<br>Compositae<br>Cruciferae<br>Leguminosae<br>Solanaceae | *Circulifer tenellus* | Mediterranean<br>North America | 20 |
| Chloris striate mosaic | Gramineae | *Nesoclutha obscura* | Australia | 21 |
| Miscanthus streak | Gramineae | | Japan | 22 |
| Maize streak | Gramineae | *Cicadulina mbila* | Africa & India | 23 |
| Tobacco yellow dwarf | Solanaceae | *Orosius argentatus* | Australia | 24 |
| Wheat dwarf | Gramineae | *Psammotettix alienus* | North Europe | 25 |
| | | Treehopper | | |
| Tomato pseudo curly top | Solanaceae | *Micrutallis malleifera* | North America | 26 |

[a] Provisional members, not yet recognized, are included.
[b] Only example host families are shown. The host range is much wider.

References for Table 1

1) Jeske and Werz (1980)
2) Goodman and Bird (1978)
3) Bock *et al*, (1978)
4) Brown and Nelson (1984, 1986)
5) Inouye and Osaki (1980)
6) Costa and Bennett (1950)
7) Osaki *et al*, (1979)
8) Bird *et al*, (1984)
9) Honda *et al*, (1983)
10) Roberts *et al*, (1986)
11) Hooker and Salazar (1983)
12) Iwaki *et al*, (1983)
13) Cohen *et al*, (1983)
14) Osaki and Inouye (1981)
15) Hamilton *et al*, (1981)
16) Cherif and Russo (1983)
17) Cohen and Nitzany (1966)
18) Brown and Nelson (1986)
19) Thomas and Bowyer (1980)
20) Thomas and Mink (1979)
21) Francki and Hatta (1980)
22) Yamashita (1983)
23) Bock (1974)
24) Thomas and Bowyer (1980)
25) Lindsten (1980)
26) Simons and Coe (1958)

Cassava mosaic disease was first described in East Africa in 1894 by Warburg (cited in Storey and Nichols, 1938). The disease is now known to occur in all parts of East, West, Central Africa and the adjacent islands (Walter, 1980). Diseases called cassava mosaic have also been reported from India, Java and South America but their etiology is uncertain (for references see Goodman, 1981a).

Cassava mosaic disease is transmitted by the whitefly *B. tabaci* (Storey and Nichols, 1938; Chart, 1958) Initial attempts to transmit the disease by sap inoculation proved difficult or almost impossible (Costa, 1976). It was later found that a virus could be transmitted from mosaic-affected cassava (*Manihot esculentum* Crontz) into several *Nicotiana* species (Bock *et al.*, 1978; Walter, 1980; Adejare and Coutts, 1982a; Sequeira and Harrison, 1982). The virus, named CLV because of its apparent absence from several mosaic-affected plants (Bock *et al.*, 1978), could not at first be returned to cassava by mechanical inoculation (Bock *et al.*, 1978, 1981), although bud-grafting was successful (Adejare and Coutts, 1982a). More recently, however, CLV was successfully transmitted from *Nicotiana spp* to susceptible Brazilian cassava cultivars in which it elicited typical mosaic symptoms, confirming CLV as the causal agent. Consequently, the virus was renamed African cassava mosaic virus (Bock and Woods, 1983). However, as much of the literature cited concerning the molecular biology of the virus still uses the old nomenclature, the virus will be referred to as CLV hereafter. The identification of geminate particles in purified virus preparations obtained from *N.benthamiana* tobacco plants infected with sap from mosaic-affected cassava (Bock *et al.*, 1978) and the demonstration of a ssDNA genome (Harrison *et al.*,1977) established CLV as a geminivirus.

Striate mosaic diseases of grasses and cereals in Australia were first described by Grylls (Grylls, 1963). The Australian disease was initially referred to as Australian wheat striate mosaic in view of its similarities to the North American wheat striate mosaic (Slykhuis, 1963). The latter was shown to be caused by a rhabdovirus, however, and the Australian disease was later referred to as chloris striate mosaic.

Chloris striate mosaic is transmitted by the cicadellid leafhopper *Nesoclutha pallida* (Grylls, 1963) but thus far has not been sap transmitted (Francki and Hatta, 1980). The viral etiology of the disease was demonstrated by feading virus-free leafhoppers with virus

purified from infected plants through a membrane and demonstrating that the leafhoppers subsequently became infective (Francki *et al.*, 1979). The geminate morphology of the purified virus particles and the demonstration of a ssDNA genome led to the inclusion of CSMV in the geminivirus group (Francki *et al.*, 1979, 1980).

Maize streak disease was first reported from South Africa in 1901 (Fuller, 1901) and has since been observed in many parts of Africa, Mauritius and Madagascar (Bock, 1974). The disease has also been reported from India and probably occurs elsewhere in South east Asia (Bock, 1974).

Maize streak is transmitted by the jassid leafhopper *Cicadulina mbila*, but is not sap transmissible (Storey, 1928). Virus particles were purified from infected maize, sugarcane and guinea grass (*Panicum maximum*) by Bock *et al.*, (1974). The demonstration that leafhoppers could be rendered infective by feeding through membranes on purified virus preparations confirmed MSV as the causal agent. Characterization of the virus by these workers revealed unusual geminate particles and the demonstration of a ssDNA genome by Harrison *et al.*, (1977) established MSV as a geminivirus.

Tomato golden mosaic was first reported from Brazil in 1975 (Matyis *et al.*, 1975). Similar diseases of tomato have been reported from Venezuela, Africa and Asia (Buck and Coutts, 1985). The disease is transmitted by the whitefly *B.tabaci* and is also sap transmissible (Costa 1976, Costa *et al.*, 1977). Virus particles were first purified from *N.glutinosa* tobacco plants mechanically inoculated with sap from tomato golden mosaic-affected tomatoes (*Lycopersicon esculentum*) (Matyis *et al.*, 1975). Demonstration that the virus is sap transmissible provided evidence that TGMV is the causal agent of the disease (Costa *et al.*, 1977). The identification of geminate particles in purified extracts from infected plants (Matyis *et al.*, 1975) and the demonstration of a ssDNA genome (Hamilton *et al.*, 1981) confirmed TGMV as a member of the geminivirus group.

Potato yellow mosaic disease (mosaico amarillo de la papa) was first reported from Venezuela in 1981 (Debrot, 1981). The disease is transmitted by the whitefly *B.tabaci* and has been sap transmitted to *Nicotiana* sp., *Petunia* sp. and *Lycopersicon* sp., but not to potatoes. Transmission to potatoes, however, has been achieved by grafting. The

demonstration of geminate particles (E. J. F. Roberts, personal communication) and the presence of a circular genome of ssDNA (Roberts *et al.*, 1986) established PYMV as a geminivirus.

Wheat dwarf disease was reported in 1961 from Czechoslovakia (Vacki, 1961) and later from Sweden (Lindsten *et al.*, 1970). The disease agent, which is transmitted by the cicadellid leafhopper *Psammotettix alienus*, was thought to be a virus although there were some indications that a mycoplasma could also be involved (Vacke, 1972). By infecting the leafhopper vector with wheat dwarf-affected plant sap after various centrifugation treatments and testing for infectivity, Lindsten (Lindsten, unpublished results, cited in Lindsten *et al.*, 1980) was able to show that not only mycoplasmas but also common viruses and viroids could be excluded as disease agents. The purification of unusually small virus particles from infected plants established WDV as the causal agent of the disease (Lindsten *et al.*, 1980). The predominantly geminate morphology of the purified virus particles and the demonstration that the viral genome consists of ssDNA (H. Macdonald, personal communication) led to the inclusion of WDV in the geminivirus group.

## 1.2. Transmission

The geminiviruses are transmitted in nature by insect vectors, either whiteflies (Aleyrodidae) or leafhoppers (Anchenorrhyncha, Cicadellideae). One exception is tomato pseudo-curly top virus (TPCTV) which is transmitted by a treehopper (Anchenorrhyncha, Membracidae) (Simons and Coe, 1958). The whitefly transmitted viruses all have the same vector species, *B.tabaci*, while the leafhopper-transmitted viruses each have a different vector species (Table I). None of the geminiviruses appears to be seed transmitted nor is there any evidence for natural spread by contact between infected and healthy plants.

Although not identical, vector relations of different geminiviruses show several features in common. Where these have been studied in detail, viruses were found to be cumulative, requiring a latent period in the vector of at least 4hr, and to persist in the vector for days and in some instances for life. Virus acquisition and inoculation generally occur in access periods of less than 1hr and inoculation efficiency increases with increasing access times (Harrison,

1985). With those vectors studied thus far the ability to transmit virus is retained through the moult from larva to adults but not *per os* and there is no evidence for virus replication in the vector.

Among the leafhopper transmitted geminiviruses, transmission of MSV by *C.mbila* has been studied in greatest detail (Storey 1928, 1938, 1939). The virus may be acquired in less than 1hr (minimum acquisition 15sec) and may be inoculated in 5min. The latent period in the vector is between 6 and 12hr at 30°C, during which time the virus moves from the gut lumen to the salivary glands where it becomes available for inoculation. Once the virus has been acquired, the vector retains the ability to transmit for life. The virus is retained during moulting but is not transmitted *per os*. The ability to transmit is inherited as a simple dominant sex-linked gene which apparently exerts its effect by controlling the permeability of the gut wall to passage of the virus. Broadly similar results were obtained for the transmission of BCTV by *Circulifer tenellus* (Bennett and Wallace, 1938; Severin, 1931) and for that of TPCTV by *Micrutalis malleifera* (Simons,1962).

Whitefly-transmitted geminiviruses require somewhat longer acquisition (15min to 1hr) and inoculation (10min to 2hr) access periods than leafhopper-transmitted geminiviruses, probably reflecting a difference in the feeding behaviour of whiteflies and leafhoppers. Transmission of euphorbia mosaic virus (EuMV) by *B.tabaci*, for example, requires minimum acquisition and inoculation access periods of 30 and 10min and a minimum latent period of 4hr (Costa and Bennett, 1950). Transmission is intermittent and may be retained for up to 20 days. The virus is retained during moulting but is not transovarially transmitted. As with leafhopper transmitted geminiviruses, there is no evidence for geminivirus replication in whitefly vectors.

## 1.3. Geographical distribution and host range

The geminiviruses occur predominantly in tropical and subtropical areas of the world but some are found in temperate regions, for example WDV. The geographical distribution of the geminivirus is generally more restricted than the distribution of their vectors, which is somewhat intriguing considering the persistence of these viruses in their vectors and the possibility of long distance flight by viruliferous

vectors (Goodman, 1981a). No epidemiological studies on the distribution of geminiviruses or the possibility of spread to new areas where the vector is already present have yet been made (Goodman, 1981a).

Although information on geminivirus host ranges is incomplete, it would appear that these viruses have narrower host ranges than other plant viruses (Goodman, 1981a). As a group, however, the geminiviruses infect a whole variety of plants including both monocotyledons and dicotyledons.

Within the geminiviruses, the leafhopper transmitted BCTV and MSV have very wide host ranges. BCTV is known to experimentally infect more than 300 species in 44 families of dicotyledonous plants (Bennett, 1971), especially in the Chenopodiacae, Compositae, Cruciferae, Leguminosae and Solanaceae. The host range of MSV, although confined to the Gramineae, is also very extensive. All cereal crops are susceptible to the virus and 54 grasses are symptomatic hosts (Damsteegt, 1983).

The host ranges of the whitefly transmitted geminiviruses are somewhat more restricted. BGMV, for example, infects only certain species in the legume genera *Phaseolus, Macroptilium, Vigna* and *Calopogonium* (Goodman and Bird, 1978). The host range of mungbean yellow mosaic virus (MYMV) is even narrower with only 7 of 26 species belonging to 6 families within the Euphorbiaceae being infected (Honda *et al.*, 1978).

Most whitefly-transmitted geminiviruses infect at least one genus in the Solanaceae, regardless of the family in which their natural hosts occur (Bock, 1982). Interestingly, within the Solanaceae, they all infect *Nicotiana* and *Datura* species, particularly *N.benthamiana* and *D.Stranmonium*. Because they are easy to grow and usually produce larger amounts of virus than their natural hosts, many of these species are used in preference to the natural hosts as propagation hosts. Thus, for example, *N.Benthamiana* is used to propagate both TGMV and CLV (Hamilton *et al.*,1981; Stanley, 1983).

## 1.4. Symptomatology and cytology

The geminiviruses can be divided into three broad groups on the basis of symptomological similarities (Goodman, 1981a). The first group comprises the leafhopper-transmitted geminiviruses which cause striate

mosaics in Graminae. These viruses include MSV, CSMV, and WDV. CSMV is found in all leaf tissues of *Chloris gayana* (Francki *et al.*, 1979) and MSV has been observed in the phloem, mesophyll, and guard cells of *Zea mays* (Markham *et al.*, cited in Stanley, 1986). WDV, on the other hand, appears to be limited to cells of the vascular tissue of *Triticum aestivum* (cited in Stanley, 1986).

The second group comprises the viruses that cause stunting, leaf curling and distortion in dicotyledons. These include the leafhopper-transmitted viruses BCTV and tobacco yellow dwarf virus (TYDV) and several whitefly-transmitted viruses such as tobacco leaf curl (TobLCV), tomato leaf curl (TomLCV) and squash leaf curl (SLCV) viruses. The majority of the viruses in this group, if not all, are limited to phloem cells of infected plants.

The third group comprises the whitefly-transmitted geminiviruses that cause striking veinal chlorosis and yellow and green mosaics in dicotyledons. Examples include BGMV, CLV, EuMV and TGMV. As with the previous groups, viruses in this group are mainly phloem-limited although some are known to infect other cell-types. CLV, for example, has been reported in cortex, mesophyll and epidermis of *N.benthamiana* (Sequeira and Harrison, 1982) and EuMV occurs in most cell types including the epidermis and mesophyll of *D. stramonium* (Kim and Fulton, 1984).

Some of the viruses mentioned above produce symptoms characteristic of more than one group. BGMV, TGMV and CLV, for example, cause leaf curl and distortion as well as typical mosaic symptoms. BCTV in tomatoes causes general yellowing symptoms and a strain of TobLCV in honeysuckle causes golden mosaic symptoms (Goodman, 1981a).

In all cases studied, geminiviruses appear to proliferate in the nuclei of infected cells where they form characteristic arrays of virus-like particles. Virus particles have been occasionally observed in the cytoplasm of CLV-infected *N.benthamiana* (Stanley, 1986) and TYLCV-infected *L.esculentum* (Cherif and Russo, 1983) leaf cells, but their presence in the cytoplasm is probably the result of nuclear breakdown prior to cell-to-cell spread. Virus particles have also been observed, although very rarely, in the cytoplasm and vacuoles of CSMV-infected *C.gayana* (Franck *et al.*, 1979.).

The cytopathological changes associated with geminivirus infections also occur predominantly in the nuclei of infected cells. The nuclear changes elicited by BGMV (Kim *et al.*, 1978) are typical of the majority of the geminiviruses infecting dicotyledonous plants. These changes include hypertrophy of the nucleoli, segregation of nucleolar material into granular and fibrillar regions, the appearance of electron dense fibrillar rings and the accumulation of virus-like particles in the nucleus often in the form of close-packed crystalline arrays (Kim *et al.*, 1978). Cytochemical studies showed that the segregated nucleoli consists mainly of ribonucleoprotein while the fibrillar rings are composed of deoxyribonucleoprotein (Kim *et al.*, 1978). These authors suggested that the fibrillar rings are the sites of viral DNA synthesis and virus assembly. This is in contrast to the Caulimoviruses, a group of double-stranded (ds) DNA containing plant viruses, where virus replication and virion assembly are thought to occur in characteristic vacuolated inclusion bodies found in the cytoplasm of infected cells (Shepherd, 1976). Electron dense inclusion bodies have been obeserved in the cytoplasm of EuMV infected *D. Stramonium*, however, although no virus particles were associated with these structures (Kim and Fulton, 1984). It is not known whether these cytoplasmic inclusions are virus coded or a host response to viral infection (Stanley, 1986).

The nuclear changes described for BGMV infections have not been reported for geminivirus-infected monocotyledonous-plants. In CSMV and MSV infected plants, cytopathological changes appear to be restricted to the presence of aggregates of virus-like particles in the nuclei of infected cells (Francki *et al.*, 1979, Bock *et al.*, 1974). In WDV-infected plants, however, tubes *ca.* 70mm long have been observed in the nuclei of infected root cells (Tonnenius and Oxelfelt, 1981). Although virus particles were present in most of the nucleoplasm of these cells, no particles were found in association with the tubes.

## 1.5. Relationships among geminiviruses

The geminiviruses are recognized by virtue of their geminate particles and the presence of circular ssDNA (Matthews, 1979). For many years, diagnosis of potential members of the group have relied on traditional methods, such as symptomatology and vector specificity.

Attempts to ascertain relationships among viruses on a basis of transmission vector, however, are probably of little use since completely unrelated viruses may be transmitted by a common vector (Goodman, 1981a). Moreover, although the presence of distinctive cytological changes in the nuclei of infected cells appears to be diagnostic for geminivirus infection, their absence is not considered to be so (Goodman, 1981a).

Serological tests are presently the most widely used way of establishing relationships among geminiviruses. Methods employed include gel double-diffusion tests, enzyme-linked immunosorbent assay (Elisa) and immunosorbent electron microscopy (ISEM).

Among the leafhopper-transmitted geminiviruses, no serological relationships have been detected between MSV, CSMV, and WDV (Bock *et al.*, 1977; Francki *et al.*, 1979; Lindsten *et al.*, 1980; Roberts *et al.*, 1984). A distant relationship has been detected between the leafhopper-transmitted geminiviruses BCTV and TYDV (Thomas and Boyer, 1980; Roberts *et al.*, 1984) but no relationships were detected between these viruses and CSMV, MSV or WDV (Roberts *et al.*, 1984).

Faint serological relationships have been detected between BCTV and the whitefly-transmitted viruses CLV and TGMV (Sequeira and Harrison., 1982; Stein *et al.*, 1983) but were not substantiated in a later investigation (Roberts *et al.*, 1984). No other serological relationships have been detected between the whitefly-transmitted and the leafhopper-transmitted geminiviruses (Sequeira and Harrison, 1982; Cohen *et al.*, 1983; Stein *et al.*, 1983; Roberts *et al.*, 1984).

Among the whitefly-transmitted geminiviruses, strong serological relationships have been detected between CLV, BGMV, EuMV, SLCV and TGMV (Stein *et al.*, 1983; Sequeira and Harrison, 1982; Roberts *et al.*, 1984). However, no serological relationships were detected between the whitefly-transmitted virus TobLCV and either CLV or BGMV (Osaki and Inoye, 1978).

Using spot hybridization techniques, imperfect nucleotide homologies were detected between probes for CLV DNA-1, which contains the coat protein (Stanley and Gay, 1983), and the DNA of the whitefly-transmitted geminiviruses BGMV, TGMV, TobLCV, TomLCV and TYLCV (Roberts *et al.*, 1984). No homologies were detected using the CLV DNA-2 probe. Moreover, no homologies were detected between either of the CLV

DNA probe and the leafhopper transmitted geminiviruses MSV, WDV and BCTV. Nucleotide sequence homologies were also detected between TGMV and EuMV (Buck and Coutts, 1985).

The results obtained from hybridization tests, with one exception in the case of TobLCV, closely parallel those obtained from serological assays. Because all the whitefly-transmitted geminiviruses have the same vector species, viz *B.tabaci*, whereas the leafhopper-transmitted geminiviruses each have a different vector, these results could be explained if their coat proteins have a key role in transmission by vectors (Roberts *et al.*, 1984).

Serology has not clearly defined geminivirus relationships. The results obtained so far suggest that within the geminiviruses there is a cluster of related viruses with whitefly vectors which are not closely related to the viruses with leafhopper vectors (Roberts *et al.*, 1984). This observation together with the demonstration that whitefly-transmitted geminiviruses have a bipartite genome while the leafhopper-transmitted geminiviruses appear to have a single component (see section 1.7), lend support to the view that geminiviruses should be subdivided into two groups on the basis of insect transmission vector. Relationships within the leafhopper-transmitted geminiviruses, however, can be as great as those between whitefly-transmitted and leafhopper-transmitted geminiviruses. As the geminiviruses appear to be moderately immunogenic, distant relationships may not be detected using the antisera presently available (Stein *et al.*, 1983). As more nucleotide sequences become available, more detailed information on geminiviral relationships may be obtained by direct nucleotide sequence comparison.

## 1.6. Virus structure and composition

The unique virion morphology of the geminiviruses sets these viruses apart from all other classes of viruses described thusfar. Electron micrographs of negatively-stained geminivirus preparations typically show twinned particles measuring 18-20nm in width and 30-36nm in length. Detailed ultrastrutural studies of freeze-dried and shadowed preparations of CSMV suggest that each particle consists of two incomplete icosahedra, each with 11 instead of 12 capsomeres, arranged in a T=1 surface lattice and joined at the discontinuities created by the missing capsomeres (Hatta and Francki, 1979).

Varying numbers of single units have been detected in negatively-stained preparations of BCTV, BGMV, CLV (synonym ACMV), CSMV MSV and TYDV (Larsen and Duffus, 1984; Goodman *et al.*, 1977; Adejare and Coutts, 1982a; Bock *et al.*,1974; Francki *et al.*,1979; Hatta and Francki, 1979; Thomas and Bowyer, 1980) The majority of single units are thought to be derived from the geminate particles during isolation or preparation for electron microscopy. This view is consistent with the demonstration that BCTV geminate particles, but not single units , are infective (Larsen and Duffus, 1984). Moreover, MSV geminate particles have been shown to contain intact viral ssDNA while single units contain predominantly degraded, heterodisperse ssDNA (Howell, 1984). However, purified preparations of CLV single particles are enriched in a minor population of circular ssDNA molecules of approximately half genome length suggesting that the circular molecules might be encapsidated in complete icosahedra (Stanley and Townsend, 1985).

Trimeric and/or tetrameric structures have been observed in preparations of BGMV (Goodman *et al.*, 1977), BCTV (Magyarosy, 1980; Larsen and Duffus, 1984), SLCV (Cohen *et al.*, 1983) and in the case of SALCV trimeric particles are the predominant type (Hooker and Salazar, 1983). Although the biological role of these unusual structures has not yet been determined, it has been suggested that encapsidated dimers of CLV genomic ssDNA could be accomodated and serve to stabilize tetrameric particles (Stanley and Townsend, 1985).

The geminate particles examined thus far contain a single major coat-protein sub-unit with Mr between 27,000 and 34,000 as estimated by SDS-polyacrylamide gel electrophoresis (Goodman *et al.*, 1980; Sequeira, 1982; Francki *et al.*, 1980; Thomas and Boyer, 1980). Additional minor protein components have been detected in purified preparations of BGMV, TGMV and TYDV but their origin and significance remains unknown (Goodman *et al.*, 1980; Adejare and Coutts, 1982a; Thomas and Boyer, 1980).

The only other identified component of geminivirus particles is ssDNA which consists predominantly of covalently closed circular molecules. A genome of ssDNA was first demonstrated by Goodman (1977b) for BGMV on the basis of nuclease and chemical sensitivity tests, hyperchromicity in the presence of formaldehyde and thermal denaturation profiles. Harrison *et al.* (1977) obtained similar results

for CLV and MSV and, in addition, demonstrated by electron microscopy that the ssDNA consisted predominantly of circular molecules. A genome of ss DNA has also been demonstrated for BGMV, CSMV, TGMV, WDV and PYMV (Reisman *et al.*, 1979; Francki *et al.*, 1980; Hamilton *et al.*, 1981; H. Macdonald, personal communication; Roberts *et al.*, 1986). The demonstration of a ssDNA genome sets the geminiviruses apart from the only other known group of plant DNA viruses, the caulimoviruses, which have a genome of dsDNA (reviewed in Shepherd and Lawson, 1981).

Size estimates of the circular ssDNA of the geminiviruses mentioned above ranging from 0.7 X $10^6$ to 0.8 X $10^6$D have been obtained by comparing the lengths of the circular ssDNA of the geminiviruses with the length of the circular ssDNA of the bacteriophage øX174. These values place the monopartite geminiviruses among the smallest known autonomously replicating viruses.

Geminiviruses DNA preparations also contain variable amounts of linear ssDNA which are resolved from the circular molecules by polyacrylamide gel electrophoresis (Goodman, 1981). It is generally accepted that the linear molecules arise from the circular molecules by breakage. This view is supported by the observation that circular ssDNA alone is infective and that the proportion of linear molecules in geminivirus DNA preparation increases with prolonged storage (Howarth and Goodman, 1982; Stanley, 1986).

The unusual capsid morphology and the small size of the circular ssDNA of the geminiviruses suggested to several investigators that each monomer in a pair might contain a distinct nucleic acid comprising the segments of a bipartite genome (Bock *et al.*, 1974; Galvez and Castano, 1986). Speculation concerning a bipartite genome was heightened by the evidence for two nucleic acid species associated with geminiviruses (Bock *et al.*, 1977). These are now known to be a linear and a circular molecule (Harrison *et al.*, 1977; Reisman *et al.*, 1979) and to have the same nucleotide sequences (Goodman, 1981). Moreover, analysis of the relative content of DNA and protein in geminiviruses has shown that there is only one DNA molecule per paired particle (Francki *et al.*, 1980; Goodman *et al.*, 1980). Francki *et al.* (1980) showed that gluteraldehyde-fixed particles of CSMV have a buoyant density in cesium-chloride of 1.35g/$cm^3$, indicating a DNA content of 19% rather than the 32% expected if there were two DNA molecules of 0.7 X $10^6$D in

each geminate particle. A similar conclusion was reached by Goodman *et al.* (1980) for BGMV, using both buoyant density measurements and chemical analysis of DNA and protein content.

## 1.7. Genome size and infectivity

The small size of the geminiviral DNA raised questions about genome size. Initial evidence for a bipartite genome came from an analysis of the fragments produced by restriction endonuclease digestion of both BGMV virion DNA and BGMV-specific dsDNA purified from infected plants (Haber *et al.*, 1981). Using restriction endonucleases *Hha*I and *Hae*III (both of which restrict ss as well as dsDNA) the gel-fractionated fragments of BGMV DNA revealed a total genome length of 5380 nucleotides, *ca.* twice that expected from electron-microscopic measurements of the DNA (2510 nucleotides) (Reisman *et al.*, 1979). Although these results could be explained by the presence of two related but distinct strains of the same virus, the idea of a bipartite genome was supported by analysis of the kinetics of infection of bean protoplasts with BGMV virion DNA (Haber *et al.*, 1981). The results illustrated a infectivity-dilution curve with a slope closer to two rather than to one which would be expected for the nucleic acid of a single component virus.

A bipartite genome was also suggested for TGMV on the basis of restriction endonuclease digestion of viral-specific dsDNA isolated from infected *N. benthamiana* leaves and subsequent molecular cloning of two distinct components into the *E.coli* plasmid pAT153 (Hamilton *et al.*,1982; Bisaro *et al.*,1982). Southern blot analysis revealed little sequence between the two components and showed both components to be present in approximately equal amounts in viral and virus-specific intracellular DNA forms.

Conclusive evidence for a bipartite genome was obtained in work with CLV. Exhaustive nucleotide sequence of CLV ssDNA revealed the presence of two distinct components of nearly identical sizes (2779 and 2724 nucleotides) (Stanley and Gay, 1983). From a knowledge of the nucleotide sequence, full length copies of the two DNA components were cloned in bacteriophage M13 vectors (Stanley, 1983). When excised from the cloning vectors, both components were found to be required to

infect *N.benthamiana* by mechanical inoculation, thereby confirming the bipartite nature of the CLV genome. Proof for a bipartite genome has now been presented for both TGMV and BGMV on the basis of sequencing and infectivity tests on the cloned DNA components (Hamilton *et al.*, 1984; Howarth *et al.*, 1985; Morinaga *et al.*, 1983).

A different result was obtained in work with both CSMV and MSV. Restriction enzyme analysis of CSMV ssDNA using restriction endonucleases *Hae*III and *Msp*I generated fragments whose cumulative size amounted to *ca.* 2,500 nucleotides (Marriott and Symons, 1982). These values compared reasonably well with a physical length of about 2,300 nucleotides estimated from electron micrographs of CSMV virion DNA (Francki *et al.*, 1980), indicating that only one DNA component is present in CSMV. Similar results were obtained from the restriction analysis of virus-specific cccDNA isolated from MSV-infected *Z.mays* (Mullineaux *et al.*, 1984).

Furthermore, exhaustive sequencing of both the Kenyan and the Nigerian isolates of MSV virion DNA revealed only one DNA component of 2681 and 2687 nucleotides respectively (Howell, 1984; Mullineaux *et al.*, 1984). Unequivocal proof for a single component is still lacking since attempts to infect plants by mechanical inoculation with either MSV virion DNA or cloned DNA have not been successful. The inability to mechanically transmit MSV, however, is thought to be due to difficulty in delivering the DNA to the site(s) where infection is initiated without incurring excessive tissue damage, rather than the absence of an undetected component of the genome (Mullineaux *et al.*, 1984).

These results are of interest in that three whitefly transmitted geminiviruses which infect dicotyledonous hosts, namely BGMV, CLV and TGMV, have a bipartite genome while two leafhopper transmitted geminiviruses which infect monocotyledonous hosts, namely MSV and CSMV, are thought to have a monopartite genome. A recent report that infective clones of BCTV, a leafhopper transmitted geminivirus which infects dicotyledonous plants, have been constructed and a single component unambiguosly demonstrated (J. Stanley, unpublished results; cited in Stanley, 1986) suggests a relationship between the number of DNA components and the nature of the insect vector.

## 1.8 Genome organisation and comparison

### 1.8.1 Sequence determination

The complete nucleotide sequences of TGMV (Hamilton *et al.*, 1984), BGMV (Howarth *et al.*, 1985), Kenyan and Nigerian isolates of CLV (CLV-K and CLV-N respectively) (Stanley and Gay, 1983; Stanley and Davies, 1985; Stanley *et al.*, 1985) and Kenyan and Nigerian isolates of MSV (MSV-K and MSV-N respectively) (Howell, 1984, 1985; Mullineaux *et al.*, 1984) have been determined. The sequences were established from cloned dsDNA for TGMV and BGMV and from virion DNA for both MSV isolates. In the first instance, the sequences of CLV DNAs were determined from virion DNA (Stanley and Gay, 1983; Stanley *et al.*, 1985) and later from infective clones (Stanley and Davies, 1985).

The whitefly-transmitted viruses TGMV, BGMV and CLV each contain two circular DNA components of similar size. These are denoted DNAs 1 and 2 in BGMV and CLV and A and B in TGMV. To facilitate comparison with other geminiviruses, DNAs A and B of TGMV will be referred to as DNAs 1 and 2, respectively. DNAs 1 and 2 of TGMV have 2588 and 2508 nucleotides (nt), the equivalent DNAs of BGMV have 2646 and 2587 nt, and those of CLV have 2779 and 2724 nt, respectively. In contrast, the Kenyan and the Nigerian strains of the leafhopper-transmitted virus MSV each have single circular DNA components of 2681 and 2687 nt, respectively.

Comparison of the sequences of the bipartite geminiviruses indicates that there is a closer relationship between TGMV and BGMV than there is between either of these viruses and CLV (Harrison, 1985). Also, a greater degree of nucleotide sequence homology was found between each DNA 1 (*ca.*60%) than between each DNA 2 (*ca.*40%), suggesting that there may be more rigid structural constraints on the DNA 1-encoded proteins (Hamilton *et al.*, 1984; Stanley, 1986). The DNA sequences of MSV-K and MSV-N are >99% homologous (Howell, 1984, 1985; Mullineaux *et al.*, 1984). No significant homologies were detected between the nucleotide sequence of MSV-N and those of TGMV, BGMV or CLV (Mullineaux *et al.*, 1984; Stanley, 1986).

### 1.8.2 Intergenic Regions

In each of TGMV, CLV and BGMV, the nucleotide sequences of the two DNA components have little sequence homology, except for a region of approximately 200 nt (the common region) which is almost identical in the two molecules. Indeed in BGMV the 'common region' is identical in both DNA components (Hamilton *et al.*, 1984; Stanley and Gay, 1983; Howarth *et al.*, 1985). In all three viruses, the 'common regions', the 5' ends of which have been arbitrarily designated as nucleotide 1, are situated almost exclusively within an intergenic region on both DNAs.

The 'common region' of each virus contains an inverted repeat sequence capable of forming a stable stem-loop structure consisting of a GC-rich stem and AT-rich loop. This structure is highly conserved in CLV, BGMV and TGMV (Fig.1) suggesting that it may contain recognition signals essential to one or more common functions such as replication, regulation of transcription and sites for capsid assembly. The remainder of the 'common region' is different in the three viruses. This lack of conservation contrasts with the relatively high degree of conservation observed between analogous putative coding regions, suggesting that sequences in the 'common region' may confer host specificity (Howarth *et al.*, 1985).

While most of the nucleotide sequence external to the 'common region' in DNA 1 appears to be involved in protein coding, only approximately 70% of DNA 2 in the case of BGMV and CLV and even less in the case of TGMV contain extended ORFs. The function of these apparently non-coding regions, if any, has not yet been determined.

In MSV there are two intergenic regions situated almost diammetrically opposed on the circular genome (Fig.2). In the case of MSV-K these regions have been referred to as the starting intergenic region (SIR) and terminal intergenic region (TIR) with regard to the proposed direction of transcription (Howell, 1984, 1985). There are more inverted repeat sequences capable of forming stable stem-loop structures in the MSV genome than in the genomes of the whitefly-transmitted geminiviruses. The most stable of these structures are located within the SIR, suggesting that they may have a regulatory role (Mullineaux *et al.*, 1984).

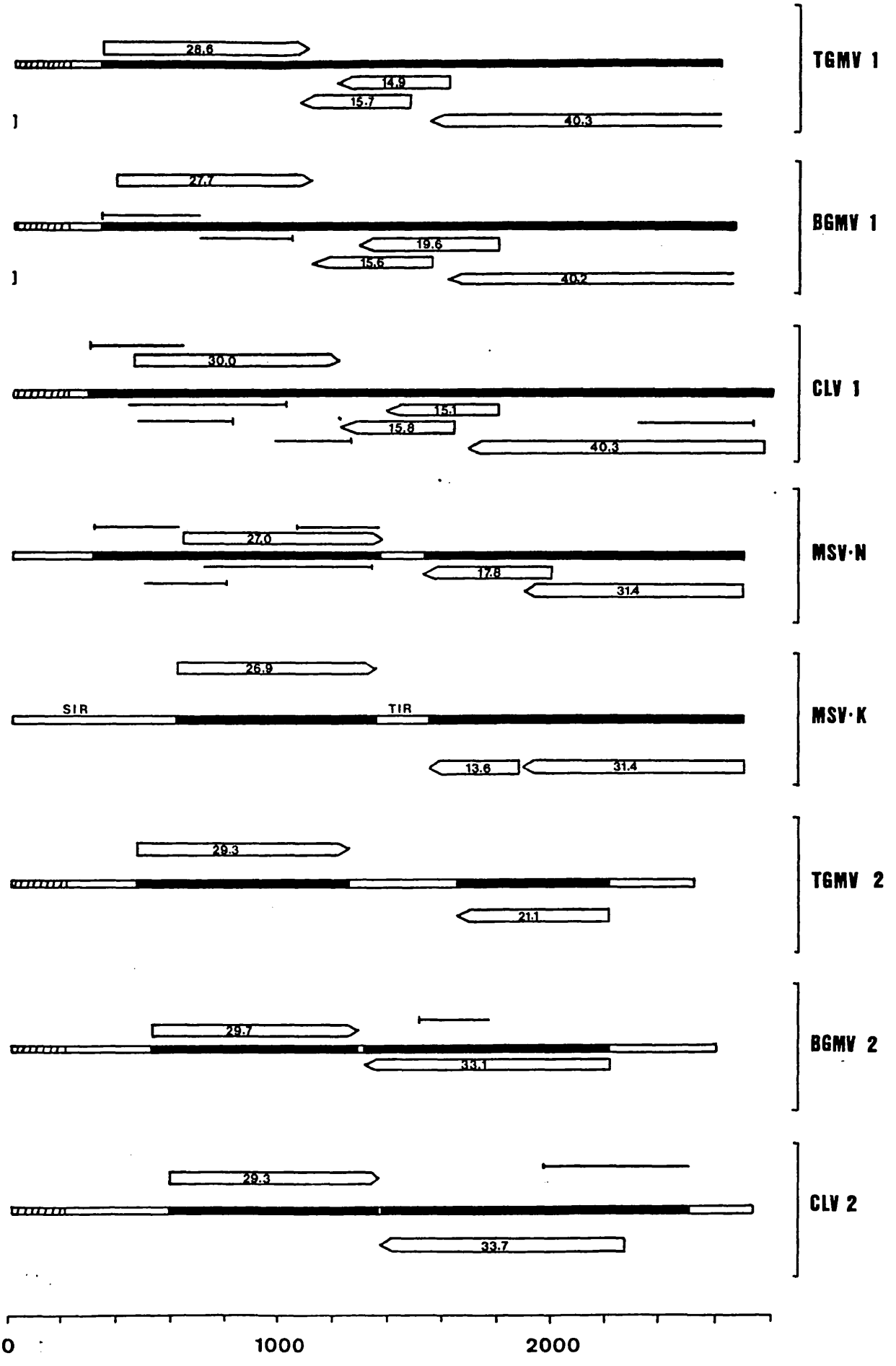
| | STEM | LOOP | STEM |
|---|---|---|---|
| TGMV 1 | GcGGCCAtCCG- | TtTAATATTACCGGa | TGGCCgC |
| TGMV 2 | GcGGCCAtCCGtTt | TAATATTACCGGa | TGGCCgC |
| CLV 1 | GgGGCCAaCCG- | TaTAATATTACCGGtT | GGCCcC |
| CLV 2 | GgGGCCAaCCG- | TaTAATATTACCGGtT | GGCCcC |
| BGMV 1 | gGcGGCCAtCCGaTa | TAATATTACCGGa | TGGCCgCc |
| BGMV 2 | gGcGGCCAtCCGaTa | TAATATTACCGGa | TGGCCgCc |

Fig.1. **Alignment of the highly conserved potential hairpin structures in the 'common regions' of TGMV, CLV and BGMV**

Nucleotides that are homologous in all three viruses are indicated with capital letters. Dashes are included to maximize sequence homologies. (Stanley, 1986).

Legend to Fig. 2. **Comparison of the geminiviral ORFs.**

The circular genomes are represented in a linear form. The genomes of TGMV, BGMV and CLV start at the begining of the 'common region'. The genomes of MSV-N and MSV-K start at the begining of the large intergenic region. The thick black lines show the coding regions, white boxes show the intergenic regions, and the hatched areas show the 'common regions' in the bipartite viruses. Conserved ORFs are indicated by the large open arrows containing the predicted protein Mr in KD. Non-conserved ORFs are drawn as thin solid arrows.
The information used in this figure was obtained from Hamilton *et al.* (1984), Howarth *et al.* (1985), Stanley and Davies (1985), Mullineaux *et al.* (1984) and Howell (1984).


28.6
14.9
15.7
40.3
TGMV 1
27.7
19.6
15.6
40.2
BGMV 1
30.0
15.1
15.8
40.3
CLV 1
27.0
17.8
31.4
MSV·N
26.9
SIR
TIR
13.6
31.4
MSV·K
29.3
21.1
TGMV 2
29.7
33.1
BGMV 2
29.3
33.7
CLV 2
0
1000
2000

### 1.8.3 Coding regions

Both genome components of TGMV, BGMV and CLV contain open reading frames (ORF) capable of coding for proteins of Mr greater than 10,000 when read from the first in-phase ATG triplet (Fig.2) (Hamilton *et al.*, 1984; Howarth *et al.*, 1985; Stanley and Gay, 1983; Stanley and Davies, 1985). Six such open reading frames are found in TGMV, four in DNA 1 and two in DNA 2, all of which have direct counterparts in both BGMV and CLV with respect to both their relative positions and their amino acid sequences. Most of these ORFs have similar coding capacities, except for the TGMV ORF analogous to the 33.7K ORF in CLV, which is truncated at the carboxy-terminus to give a coding capacity of 21.1KD and the BGMV ORF analogous to the 15.1KD CLV ORF, which is extended at the amino-terminus to give a coding capacity of 19.6KD. The high degree of amino acid sequence homology observed between the putative products of each of the six ORFs above suggests that they are functionally related. In addition, there are three other ORFs in BGMV and six in CLV that do not appear to be conserved. Whether these ORFs have arisen fortuitously or whether they code for proteins specific to the respective viruses is not known.

The genomes of all three viruses display a remarkable economy of genetic space. Both genome components in each of the three viruses have ORFs on the virion and the complementary strand which, together with the position of transcriptional regulatory sequences (see below), suggests that transcription is bidirectional. A number of the ORFs, including two of the conserved ones (The 15.1 and 15.8KD ORFs in CLV and their counterparts in TGMV and BGMV), overlap in different reading frames to a significant extent and in the complementary strand of DNA 1 all three reading frames are used.

Potential promoter regions conforming to the consensus TATA box sequence TATAA/TAA/T (Breathnach and Chambon, 1981) and polyadenylation signals AATAAA (Proudfoot and Brownlee, 1976) are available for most ORFs. Sequences surrounding the initiation codons of the putative ORFs do not conform to the consensus sequence CCA/GCCAUGG proposed by Kozak (1984) as a ribosomal recognition signal for the initiation of protein synthesis on eukaryotic mRNAs (Stanley, 1986). However,all the ORFs common to the three viruses, with the exception of the BGMV 27.7KD and 33.1KD ORFs and the TGMV 21.1KD ORF, have a preferred A or G (Kozak,

1984) as the third base preceding their putative initiation codons (Stanley, 1986). No common feature that might serve as a ribosomal recognition sequence has been identified in these viruses (Stanley, 1986).

The 30.2kD ORF in CLV and its counterparts in TGMV and BGMV (ORFs 28.6kD and 27.7kD respectively) have been identified as the coat protein gene on the basis of *in vitro* translation of mRNA (Townsend *et al.*, 1985), amino acid analysis (Townsend *et al.*, 1985; Stein, 1985) and size estimates by SDS-polyacrylamide gel electrophoresis of purified coat protein (Bock *et al.*, 1977; Goodman *et al.*, 1980; Hamilton *et al.*, 1981; Stanley *et al.*, 1985). The coat protein gene is the most highly conserved of all the putative coding sequences, reflecting the strong serological relationship observed between the three viruses (Harrison, 1985). No other virus-specific gene products, besides the coat protein, have been identified.

Comparison of the amino acid sequence derived from the CLV coat protein gene and that of ORF 29.9kD in CLV DNA 2 revealed a significant degree of sequence conservation (Kikuno *et al.*, 1984). As the two ORFs occupy equivalent positions in DNAs 1 and 2 it was concluded that the bipartite genome had evolved from a single common ancestral DNA. Stanley and Davies (1985) have suggested further that the product of the CLV 29.9kD ORF may represent a still as yet undetected component of the virus capsid.

Analysis of the MSV-N sequence revealed seven overlapping ORFs capable of coding for proteins of Mr>10,000, three in the virion sense and four in the complementary sense (Fig.2) (Mullineaux *et al.*, 1984). In contrast, only three large ORFs were found in the genome of MSV-K (Fig.2), one in the viral sense and two in the complementary sense (Howell, 1984, 1985), all of which have direct counterparts in MSV-N. As with the whitefly-transmitted geminiviruses, the distribution of the ORFs in relation to putative promoter and polyadenylation signals in the genomes of both MSV isolates suggests a bidirectional transcription strategy.

The only MSV-specific protein positively identified thus far is the coat protein. The MSV-N 27.0kD ORF and its MSV-K counterpart have been identified as the coat protein gene on the basis of *in vitro* translation of mRNA (Morris-Krsinich *et al.*, 1985), amino acid analysis and size estimates by SDS-polyacrylamide gel electrophoresis of purified

coat protein (Mullineaux *et al.*, 1984). The functions of the other ORFs are unknown, although the presence in MSV-K of a nonsense codon approximately in the middle of the sequence corresponding to the 21.8kD ORF in MSV-K suggests that this coding region may be fortuitous (Stanley, 1986).

In MSV-K, ORFs 13.6kD and 31.4kD are tandemly orientated in the same reading frame. The two putative proteins show nearly 40% amino acid sequence homology with the single product of the CLV 40.3 ORF, suggesting that the MSV-K 13.6kD and 31.4kD ORFs may be translated into a single protein by reading through the amber codon terminator in the 31.4kD ORF (Howell ,1984). In contrast, the analogous ORFs in MSV-N (ORFs 17.8kD and 31.4kD respectively) overlap in different reading frames (Mullineaux *et al.*, 1985).

## 1.9 Transcription

In all geminiviruses examined, the distribution of the ORFs together with the position of promoter and polyadenylation sequences suggest that transcription is bidirectional. This has been confirmed for both CLV (Towsend *et al.*, 1985) and MSV-N (Morris-Krsinich *et al.*, 1985) by the identification and mapping of polyadenylated transcripts isolated from infected plants.

In CLV, five major virus-specific polyadenylated transcripts have been mapped to the virion and the complementary sense of both DNAs 1 and 2 (Townsend *et al.*, 1985). Of the DNA 1-specific major transcripts, two (of 1.7Kb and 0.7Kb, respectively) were transcribed from the virion strand and one (of 1.0Kb) from the complementary strand. Of those for DNA 2, one (of 1.1Kb) was transcribed from the virion strand and one (of 0.9Kb) from the complementary strand. A minor transcript of 1.35Kb was mapped to the virion strand of DNA 2. None of the transcripts hybridized to both probes of full-length clones, suggesting that none of them is transcribed across the 'common region'.

The major transcripts identified in CLV are of suitable size and distribution to encode the products of the six ORFs conserved between CLV, TGMV and BGMV (Townsend *et al.*, 1985). Thus, the 1.0Kb transcript could encode the product of the 30.2Kb ORF in the virion strand of DNA 1. The 1.7Kb transcript could encode the product of the 40.3kD ORF in the complementary sense of DNA 1 and the 0.7Kb transcript could

encode the products of the 15.1 and/or 15.8kD overlapping ORFs. Hamilton *et al.* (1984) suggested that the expression of the corresponding two overlapping ORFs in TGMV would be possible by splicing, but no evidence for splicing has been found for any CLV transcript. The 0.9Kb and 1.1Kb transcripts could encode the products of the 29.3kD and 33.7kD ORFs of DNA 2 virion and complementary strand, respectively. All transcripts have a suitably positionned TATA box with the exception of the 1.1Kb transcript in the complementary strand of DNA 2, which has prompted the suggestion that the minor 1.35Kb might be a precursor of the 1.1Kb transcript (Townsend *et al.*, 1985).

*In vitro* translation of polyadenylated RNA from CLV-infected tissue showed a single virus-specific product of 30.0kD which was identified as the coat protein by immunoprecipitation with antiserum raised against purified CLV coat protein (Townsend *et al.*, 1985). The 1.0Kb transcript was identified as the coat protein mRNA by hybrid-arrested translation, confirming the identity of the 30.2kD ORF as the coat protein gene. This transcript has an untranslated leader of *ca.* one hundred and sixty nucleotides just downstream of a TATA box. The sequence GGCCAATC, conforming to the proposed consensus transcription modulator sequence (CAAT box) (Benoist *et al.*, 1980) is located sixty five nucleotides upstream of this TATA box. The same sequence is present in TGMV at approximately the same position relative to the proposed promoter of the coat protein gene (Hamilton *et al.*, 1984).

Nuclei isolated from TGMV-infected tobacco plants have been shown to direct the synthesis of virus-specific RNA when supplied with exogenous rNTPs (Coutts and Buck, 1982). Both strands of each component appeared to be transcribed, adding further support for bidirectional transcription in this virus. The effects of various salts, enzymes and known inhibitors of RNA polymerisation suggest that synthesis of viral-specific transcripts is catalysed by a DNA-dependent RNA polymerase type II enzyme (Coutts and Buck, 1985)

Three virus-specific transcripts have been detected in polyadenylated RNA from MSV-infected maize leaves (Morris-Krsinich *et al.*, 1985). Two of the transcripts, a major and a minor transcript of 0.9Kb and 1.05Kb respectively, were mapped on the virion sense DNA and the third, a minor transcript of 1.2Kb, was mapped on the complementary sense DNA.

The two virion sense transcripts are 3' coterminal but have different 5' termini. Of these transcripts, only the 1.05Kb transcript has a suitably positioned TATA box, suggesting that it may be a precursor of the 0.9Kb transcript (Morris-Krisinich *et al.*, 1985). It is possible, however, that the 0.9Kb transcript may be promoted independently from one of several T rich stretches of sequence located within sixty nucleotides of its 5' terminus (Morris-Krisinich *et al.*, 1985). Both transcripts could encode the product of the 27.0Kd ORF. In addition, the 1.05Kb transcript could also encode the product of the 10.9kD ORF. However, the location of the initiation codon for this ORF is just 4 nt downstream of the 5' terminus of the transcript, which argues against this possibility. Both transcripts terminate fifty seven nucleotides downstream of a possible polyadenylation signal AATAAA.

The 1.2Kb complementary sense transcript could encode the product of the 17.8kD ORF. The 5' end of this transcript appears to map slightly short of the first ATG for the putative 31.4 ORF. However, translation from another in-frame ATG at a position ninety nine nucleotides downstream would produce a 27.7kD protein (Morris-Krisinich *et al.*, 1985).

*In vitro* translation of MSV-specific polyadenylated RNA produced a 28.8kD protein which was identified as the coat protein by immunoprecipitation with antiserum raised against whole virus. The mRNA for this protein maps to the same region and has the same orientation as the virion sense transcripts of 1.05Kb and 0.9Kb, confirming the 27.0kD ORF as the coat protein gene. Unlike CLV and TGMV, where a consensus CAAT sequence precedes the coat protein transcript, there are no consensus CAAT sequences upstream of either of the virion sense MSV transcripts.

By analogy with simian virus 40 (SV40) (reviewed by Tooze, 1980), where bidirectional transcription segregates early and late functions on complementary strands, a model for the temporal regulation of gene expression based on the polarity of transcription has been proposed for the geminiviruses (Hamilton *et al.*, 1984; Howarth *et al.*, 1985). Since the gene for the coat protein (a product probably required late in the infection cycle) is located on the viral strand, it was suggested that late genes are expressed clockwise and early genes anti-clockwise. With whitefly-transmitted geminiviruses, the possibility exists of an additional intermolecular control between the two DNA components (Stanley, 1986)

The demonstration of bidirectional transcription implies that transcripts are synthesized from a dsDNA template. Intracellular dsDNA forms have indeed been found for a number of geminiviruses (see section 1.10). Sunter *et al* (1984) have suggested that virus-specific circular covalently closed (ccc) dsDNA isolated from TGMV-infected plants may adopt *in vivo* a transcriptionally active minichromosome structure analogous to that described for cauliflower mosaic virus (Olszewski *et al.*, 1982). The identification of specific nuclease S1-sensitive sites in the intergenic region of TGMV ccc dsDNA 2 prompted the suggestion that regions of possible secondary structure may have transcriptional regulatory function(s) (Sunter *et al.*, 1985).

## 1.10 Replication

Electron microscopy (Kim *et al.*, 1978) and studies with isolated nuclei (Coutts and Buck, 1985) suggest that geminiviral DNA synthesis takes place in the nuclei of infected host cells. Little is known about the mechanisms by which DNA replication is achieved, however, and most of what has been proposed remains speculative.

Intracellular dsDNA species of circular configuration have been identified for a number of geminiviruses, suggesting that these viruses replicate *via* circular dsDNA templates analogous to those of the ssDNA bacteriophages. Virus-specific unit-length ccc and, open circular (oc) dsDNA forms have been identified in extracts from TGMV-infected *N. benthamiana* (Hamilton *et al.*, 1982), unit-length ccc dsDNA has been isolated from *Z. mays* infected with MSV (Mullineaux *et al.*, 1984) and *N.benthamiana* infected with CLV (Stanley and Townsend, 1985), and unit-length oc dsDNA from *P. vulgaris* infected with BGMV or MYMV (Ikegami *et al.*, 1981, 1984). The demonstration that the circular dsDNA species isolated from both TGMV and MYMV-infected plants are infective strengthens the view that replication occurs via a circular dsDNA replicative form (Hamilton *et al.*, 1983; Ikegami *et al.*, 1984)..

The initial stage in the replication of geminiviruses upon infection of host cells is probably the conversion of virion ssDNA into a double-stranded form. Deproteinized virion ssDNAs from BGMV (Goodman, 1977a,b), TGMV (Hamilton *et al.*, 1981) and MYMV (Ikegami *et al.*, 1984), are infectious, which, if transcription is from dsDNA templates alone, implies that second strand synthesis relies entirely on host-coded

factors, as also suggested in studies with isolated nuclei (Coutts and Buck, 1985). As host factors might be expected to interact with similar recognition sequences on both components of a bipartite genome, it was suggested that the 'common regions' might contain the origins of replication for this process (Stanley and Gay, 1983; Hamilton *et al.*, 1984). The bacteriophage ØX174 ssDNA contains a stable hairpin structure, the primosome assembly site, which has been implicated in the host-directed synthesis of double-stranded replicative intermediates (Arai and Kornberg, 1981). By analogy, a similar function has been proposed for the highly conserved putative hairpin structures present in the 'common regions' of the bipartite geminiviruses (Hamilton *et al.*, 1984; Stanley and Davies, 1985).

Analysis of encapsidated MSV-N DNA revealed a population of small DNA molecules with ribonucleotides covalently linked to their fixed 5' deoxyribonucleotide ends and with variable 3' ends (Donson *et al.*, 1984). These molecules have a sequence of approximately eighty nucleotides which is complementary to a region immediately downstream of two stable hairpin structures in the virion ssDNA, almost exclusively within the terminal intergenic region. These molecules,which have also been reported for MSV-K (Howell, 1984), are thought to act as primers for second strand synthesis upon infection of host cells (Donson *et al.*, 1984; Howell, 1984). Conversely, the initiation of second-strand synthesis itself is thought to occur late in the infection cycle (Stanley, 1986) and probably involves RNA-primed DNA synthesis (Donson *et al.*, 1984; Howell, 1984). Donson *et al.* (1984) suggested that the ribonucleotides at the 5' end of the small molecules represent the remnants of a larger RNA primer that might have originally extended into the region of hairpin secondary structure, as has been reported for the ssDNA bacteriophage G4 (Sims and Dressler, 1978). In this context, the small DNA molecules should be viewed as intermediates in the synthesis of the second strand of DNA, which, by as yet unknown mechanisms, become selectively encapsidated (Donson *et al.*, 1984). Exhaustive searches failed to reveal a similar primer-like molecule associated with encapsidated CLV DNA, reflecting yet another fundamental difference between these viruses (Stanley and Townsend, 1985).

Virus-specific high molecular weight dsDNA species of greater than unit-length have been observed in TGMV and CLV infected tobacco plants (Hamilton *et al.*, 1982; Stanley and Townsend, 1985). For CLV, these molecules were shown to consist predominantly of covalently closed circular dimers of tandemly arranged viral molecules of the same DNA component (Stanley and Townsend, 1985). Hamilton *et al.* (1982) suggested that the high molecular weight species might be intermediates in the replication of the dsDNA replicative forms. Abnormal TGMV-infected plants containing unusually large amounts of the high molecular weight dsDNA species relative to genomic length ssDNA have been observed, prompting the suggestion that the former are genuine replicative intermediates which accumulated as a result of an incomplete infection cycle (Buck *et al.*, 1983; Hamilton *et al.*, 1983). The regular appearance of these molecules in plants infected with cloned TGMV and CLV DNA, in contrast to a subgenomic species which, in the case of CLV, was not detected even after multiple passaging of virus derived from the cloned DNA inoculum (see next section), is consistent with this view (Hamilton *et al.*, 1983; Stanley and Townsend, 1985).

Additionally, analysis of encapsidated virion ssDNA from both CLV and MSV revealed the presence of concatameric ssDNA molecules (Stanley and Townsend, 1985). These molecules consist predominantly of non-covalently bound dimers of genomic ssDNA together with true dimers of circular configuration. While the latter might have arisen by incomplete cleavage of nascent DNA strands during a rolling circle mechanism of DNA replication or by DNA replication from a dimeric dsDNA template (Stanley and Townsend, 1985; Stanley, 1986), the origin of the former is less clear. Since there are no major inverted repeats either within or between the ssDNA components, it was suggested that residual proteins might be responsible for the association between the DNA molecules in the non-covalently bound dimers (Stanley and Townsend, 1985). As with the dsDNA concatamers, both species of ssDNA dimers appear routinely in plants infected with cloned DNA.

Totipotent leaf mesophyll protoplasts of *N. plumbaginifolia* have been successfully infected with CLV and with full-length copies of CLV DNAs 1 and 2, opening a valuable new route for the investigation of geminiviral replication and genome function (Townsend *et al.*, 1986). Protoplasts infected with either virus or cloned DNA accumulated both ccc dsDNA and progeny virion ssDNA representative of each DNA

component. The appearence of virus-specific DNAs coincided with the onset of cell division, suggesting that CLV replication is dependent upon synthesis of host replication factors. Interestingly, protoplasts inoculated with DNA 1 alone, but not with DNA 2 alone, produced ccc dsDNA and virion ssDNA, demonstrating that DNA 1 encodes all the virus-specific functions necessary for replication of the genome. Since both components are required to initiate an infection in plants, it was suggested that DNA 2 might be responsible for systemic spread of the virus in the host (Townsend *et al.*, 1986).

Similar conclusions were drawn by Rogers *et al.* (1986). These workers used *Agrobacterium tumefaciens* Ti plasmid vectors to introduce either TGMV DNA 1 or DNA 2, as single or tandem inserts, into the chromosomal DNA of petunia cells. Regenerated plants were normal in morphology and fertile. Cross fertilization of plants carrying single copies of either DNA component produced phenotypically normal plants. In contrast, sexual crosses of plants containing tandem inserts of DNAs 1 and 2 produced progeny plants showing chlorotic lesions typical of virus infection. Analysis of total leaf DNA prepared from symptomatic progeny plants revealed the presence of free single and double-stranded viral DNA species identical to those obtained from TGMV-infected tobacco plants, indicating that replication of virus DNA occurred in these plants. Surprisingly, plants carrying tandem repeats of DNA 1 alone contained non-integrated single and double-stranded forms of DNA 1, again indicating that DNA component 1 contains all the virus-specific functions required for replication. No free DNA 2 was detected in plants carrying tandem repeats of component 2 alone. Since plants carrying tandem repeats of DNA 1 alone did not produce symptoms, it was concluded that DNA 2 encodes functions essential for symptom production (Rogers *et al.*, 1986). In contrast, results obtained from pseudorecombinant infectivity studies using heterologous combinations of the infective cloned DNA components of CLV-K and a Nigerian CLV isolate (CLV-N) demonstrated that DNA 1 contains at least some determinants of symptom induction (Stanley *et al.*, 1985). In conclusion, these results suggest that the products from both DNA components may be involved in symptom production.

## 1.11 Defective particles

Extracts of TGMV-infected tobacco plants contain a minor population of virus-specific subgenomic ssDNA, the size of which was estimated at *ca.* 2000 nt (Mr 0.6 X $10^6$) by agarose gel electrophoresis (Hamilton *et al.*, 1982, 1983). The subgenomic DNA species is encapsidated in virus particles and appears to be derived solely from component 2, as judged by Southern blot analysis using full-length DNA 1 and DNA 2 probes (Hamilton *et al.*, 1983). A weak hybridization signal with the DNA 1 probe was detected, which was attributed to the small amount of sequence homology between components 1 and 2. The subgenomic DNA species was not detected in plants infected with cloned DNA, suggesting that it is a defective molecule which arose through repeated passaging of the virus in *N. benthamiana* plants (Hamilton *et al.*, 1983).

CLV particles also contain a minor population of subgenomic ssDNA molecules of approximately half unit-length (*ca.*1200 nt) (Stanley and Townsend, 1985). These were shown to be circular by electron microscopy and appear to be encapsidated in single, possibly isometric particles. Southern blot analysis of CLV subgenomic ssDNA using DNA-specific probes demonstrated that it originates exclusively from DNA 2.

A ccc dsDNA form of the subgenomic ssDNA species was isolated from CLV-infected tobacco plants, suggesting that the latter replicates from its own dsDNA template (Stanley and Townsend, 1985). Full-length copies of the subgenomic dsDNA species were cloned and the deletion borders of 19 clones defined by nucleotide sequencing. Four clone types were obtained, which differed with respect to the extent of the deletion. In all cases, however, approximately half of the DNA 2 molecule, starting within 150 nt downstream of the 'common region', was deleted, completely removing the 29.3Kd ORF and the carboxy-terminus of the 33.7Kd ORF (Fig.3). One example was completely sequenced, showing that, with the exception of a few nucleotide substitutions and the insertion of a single T residue within the 'common region', no other major insertions or deletions had occurred. The presence of the 'common region' in all clones examined is consistent with its putative role in DNA replication and/or capsid assembly. Sequence analysis revealed the presence of small direct repeat sequences at the deletion borders

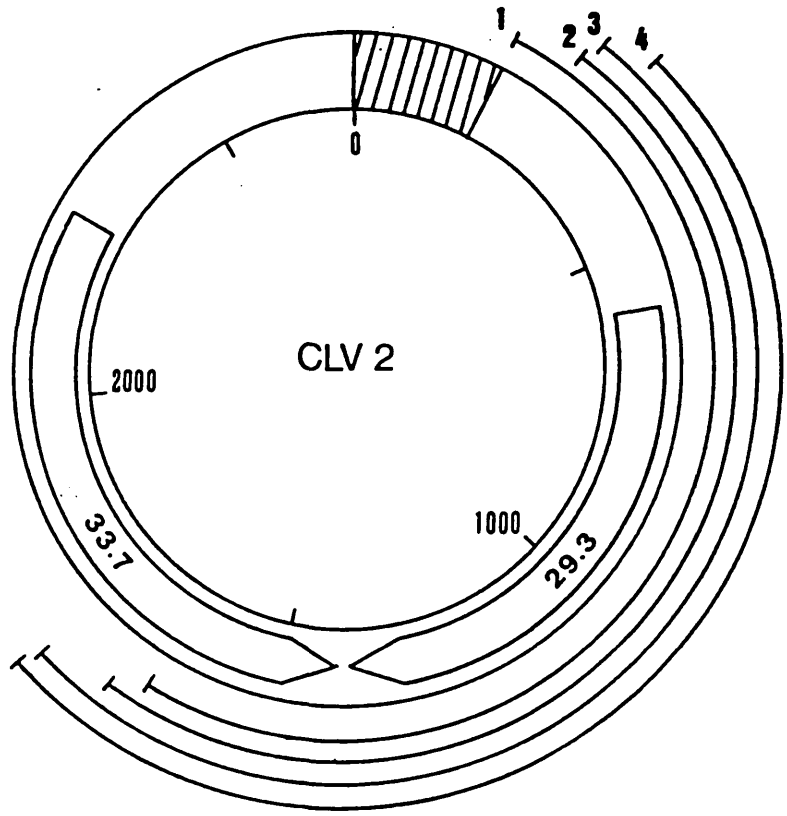



Fig. 3. **CLV 2 showing the extent of the nucleotide deletions in the defective molecules.**

The 'common region' is shown as the hatched area. The major ORFs are indicated as open arrows with the predicted protein Mr in Kd. Deletions 1 to 4 correspond to those shown in Table 2. (Stanley and Townsend, 1985).

(Table 2), the most abundant clone type having the largest repeat. Although the mechanism for the formation of the subgenomic DNA molecules was not determined, it was suggested that the repeated sequences may be responsible for the exact location of the deletions (Stanley and Townsend, 1985).

Infectivity studies in tobacco plants using cloned DNAs 1 and 2 in combination with cloned subgenomic DNA showed a direct correlation between the concentration of subgenomic DNA in the inoculum and both the time required for symptom expression and the number of inoculated plants that did not show symptoms (Stanley and Townsend, 1985). Inoculation with DNA 1 and varying amounts of subgenomic DNA did not elicit symptoms nor resulted in systemic spread of the virus. These observations suggest that CLV subgenomic DNA is similar to the helper-dependent defective interfering particles associated with animal virus infections (reviewed by Holland, 1985). As with TGMV, CLV subgenomic DNA was not detected in plants infected with the cloned DNA components alone, even after repeated passage.

## 1.12 Defective interfering viruses

Specific reference to DNA containing viruses will be given in this section, although some examples with RNA viruses are quoted and introduced.

Defective interfering (DI) viruses are viruses that lack one or more essential functions required for autonomous replication and therefore can only replicate in the presence of related helper virus. In addition, they inhibit the replication of helper virus by competing for helper virus-encoded products required for replication and/or encapsidation of the genome.

DI viruses were discovered by von Magnus in 1954. This worker showed that non-infectious 'incomplete particles' of influenza virus resulting from serial undiluted passages of influenza virus in embrionated eggs inhibited the growth of infectious influenza virus (von Magnus, 1954). The presence of DI viruses was subsequently reported for a number of RNA and DNA virus systems (reviewed by Holland, 1985), and in 1970 Huang and Baltimore proposed the name *defective interfering particles* for viral deletion mutants that contain normal viral structural proteins, require the presence of helper virus

| Type | Frequency | Sequence at points of deletion | DNA length |
|---|---|---|---|
| | | 190 200<br>CCCGCGCACTggttggcttc | |
| 1 | 1 | •• | 1315 |
| | | atctcatgCTTGTATAGCGC<br>1600 1610 | |
| | | 270 280<br>GCATATAGATGTTgtcagct | |
| 2 | 1 | ••••• | 1363 |
| | | tgtagcccATGTTTCTCCTG<br>1630 1640 | |
| | | 280 290<br>GATGTTGTCAGCtatgcata | |
| 3 | 1 | ••• | 1294 |
| | | ttttgaatgAGCCTTCGAAT<br>1710 1720 | |
| | | 340 350<br>GGGCCAATCGAATGacagct | |
| 4 | 16 | ••••••• | 1343 |
| | | tgagcctTCGAATGGGCTTA<br>1720 1730 | |

Table 2. **Alignment of the nucleotides surrounding the deletion points in 19 CLV subgenomic DNA clones.**

The small direct repeat sequences are indicated by a dot (•). In addition to the nucleotides shown in small register, one copy of each repeat is deleted. (Stanley and Townsend, 1985).

for replication and interfere specifically with the reproduction of homologous helper virus (Huang and Baltimore, 1970).

DI particles usually arise as a result of a deletion in an essential region of the parental virus genome. The deletion may be very small or may constitute over 90% of the genome. In all cases, however, DI particles must conserve or mimic parental recognition sequences involved in initiation of replication and encapsidation of the genome. In addition, packaging constraints may impose size limitations on the DI viral nucleic acid.

The degree of interference, as assessed by an inhibition of the absolute amount of helper virus produced in doubly infected cells compared to that produced in cells infected with helper virus alone, may vary from virtually 100%, as in the case of some DI particles of mouse adenovirus (Larsen, 1982), to only very slight, as in the case of poliovirus and other picornavirus DI particles (Cole and Baltimore, 1973). In some cases, DI particles of the same virus may interfere to different extents depending on the host cell line (Huang and Baltimore, 1970).

Because of their interfering properties, DI particles may play a role in the regulation of virus production and in the attenuation of effects of viral infections (Huang and Baltimore, 1970). DI particles have indeed been shown to protect animals from diseases and death caused by homologous virus and to prevent death of infected cell cultures (Holland *et al.*, 1980). DI particles have also been involved in the establishment of persistent infections and in virus and host evolution (Holland, 1985).

Amongst the DNA viruses, DI particles have been detected in bacteriophages, adenoviruses, parvoviruses, papovaviruses herpesviruses and geminiviruses. The remainder of this introduction focuses on a some representative examples of DI particles of DNA viruses as well as on possible mechanisms for their generation. Geminiviruses are omitted as they were discussed previously (section 1.11).

Bacteriophage f1

Multiple passages of bacteriophage f1 at high multiplicity of infection (m.o.i) leads to the accumulation of non-substituted deletion mutants lacking *ca.* 70% of the genome (Griffith and Kornberg, 1974;

Enea and Zinder, 1975). These mutants are unable to replicate independently yet outgrow the wild-type helper phage in a mixed infection. Restriction mapping and electron microscopic heteroduplex analysis indicate that these DI particles contain no intact cistrons but always retain part of two cistrons (the carboxyl-terminus of gene IV and the amino-terminus of gene II) and the intergenic region between them (Enea and Zinder, 1975; Enea *et al.*, 1977). The intergenic region has been shown to contain the origin of both viral and complementary strand synthesis (Horiuchi and Zinder, 1976).

Some DI particles contain tandem reiterations of part of the intergenic region. Enea *et al.* (1977) showed that DI particles having partial duplications grow better and interfere more strongly with the growth of helper phage than DI particles having no such duplications, suggesting that the presence of the reiteration confers a selective advantage over those mutants harbouring them. The ratio of RFII/RFI appears to be higher in DI particles containing partial duplications than in f1 (Enea *et al.*, 1977). Since RFII molecules, not RFIs, are those that are probably engaged in replication (Ray, 1976), it was suggested that the interference caused by these DI particles probably results from more efficient RF synthesis in the latter as compared to f1 (Enea *et al.*, 1977; Ravetch *et al.*, 1979). Sequence analysis of DI particles containing partial duplications indicates that the duplication contains the origin of complementary strand synthesis and possibly occurs at the origin of viral strand synthesis (Ravetch *et al.*, 1979). Since initiation of viral strand synthesis appears to be rate-limiting in f1, the selective advantage conferred by the partial duplication probably results from duplication of the origin of viral strand synthesis.

## 1.14 Adenoviruses

The adenoviruses are a group of animal viruses having icosahedral particles and a genome of linear dsDNA. These viruses were long thought not to produce DI particles. A number of defective deletion mutants had been observed which did not interfere with growth of the wild-type parental virus (see Daniell, 1976). True DI particles have now been isolated following serial passage of mouse adenovirus strain FL (AdFL) at high m.o.i. in BALB/c-3t3 cells (Larsen, 1982). These particles, which vary between 50 to 96% of the wild-type genome, become

apparent after 11 to 35 passages. They show a marked selective growth advantage over wild-type helper virus and often within a few passages after their appearence become the predominant species. In fact, they sometimes completely overgrow the helper virus, with the result that both viral DNA and cytopathic effects become undetectable.

None of the DI particles that have been analysed contain the same deletions. In all of them, however, at least 3% of each end of the AdFL genome is conserved, suggesting that sequences at the ends of the AdFL genome contain all the *cis* functions required for replication and encapsidation. The origins of replication of AdFL have in fact been located near each end of the AdFL genome (Temple *et al.*, 1981). Although the location of AdFL packaging sequence(s) is not known, an essential region for the packaging of viral DNA has been located within 400 base pairs (bp) of the left end of human adenovirus 16 (Ad16) (Hammaskjold and Winberg, 1980). Restriction analysis indicates that, in addition to the observed deletions, AdFL DI particles also contain cellular DNA substitutions (Larsen, 1982). The length of the substitutions vary between *ca.* 24% to 93% of the AdFL genome.

Another adenovirus defective particle exhibiting some interfering characteristics has been reported for human adenovirus type 12 (Ad12) (Deuring *et al*, 1981). This DI particle, which has the same size as wild-type helper virus, consists mainly of human KB cellular DNA flanked on either side by 700-1150 bp from the left terminus of Ad12. Electron microscopic observations of the DI particle after denaturation renaturation indicates that the molecule consists of a large inverted repeat of the type ABCDD'C'B'A'. This DI particle appears to be encapsidated more efficiently than wild-type helper virus, suggesting that, as in Ad16 (see above), the capsid assembly site(s) is (are) located in the left end of the molecule. In contrast to AdFL DI particles, the Ad12 defective has no apparent replicative advantage.

## 1.16 Parvoviruses

The parvoviruses are small icosahedral animal viruses with a linear ssDNA genome flanked at both ends by a short palindromic sequence. This sequence 'snaps back' to form a duplex structure which is used as a primer for second strand synthesis during replication. Among the parvoviruses, both adeno-associated virus and minute virus of

mouse have been shown to generate DI particles (de la Maza and Carter, 1980; Faust and Ward, 1979).

Adeno-associated virus (AAV) is a satellite virus, requiring the presence of either adenovirus or herpesvirus helper for the completion of its infection cycle. When AAV is coinfected with a helper adenoviruses, AAV DI particles with genome lengths ranging from nearly 100% to only 3% of that of standard virus are generated (de la Maza and Carter, 1980). These molecules have varying lengths of the internal region of the AAV genome deleted but retain the genome termini. Some of the DI particles contain 'snap back' DNA molecules. These consist of duplex hairpin molecules in which the complementary strands are covalently linked at variable sites distal from the normal genome termini. These molecules resemble AAV RF DNA, with the difference that in the latter the 'snap back' occurs at the terminal palindromic sequence (de la Maza and Carter, 1980).

DI particles have been observed after a single passage of minute virus of mouse (MVM) at high m.o.i. in EA cells (Faust and Ward, 1979). DNA extacted from incomplete MVM particles has several features in common with that from AAV DI particles, including the preservation of the genome termini and the heterogeneous length of the internal deletion. Approximately 50% of the MVM defective particles contain 'snap back' DNA molecules. Unlike those of AAV, the MVM 'snap back'molecules appear to be derived exclusively from the 5' end of the viral genome. Furthermore, these molecules gradually disappear after serial undiluted passage suggesting that they are unable to replicate (Faust and Ward, 1979).

## 1.17 Papovaviruses

The papovaviruses are non-enveloped icosahedral animal viruses with a dsDNA genome of circular nature. The generation of DI particles by SV40, a member of the group, has been amply documented and a number of these particles have been cloned and characterized (reviewed by Huang and Baltimore, 1977; Holland, 1985).

Serial passage of SV40 at high m.o.i. leads to the accumulation of defective virus particles with aberrant genomes. At early passages, defective particles usually contain duplications of genomic segments including the origin of DNA replication and compensating deletions in

other regions (Brockman *et al.*, 1975). After prolonged passage, all SV40 sequences not required for the proliferation of defective virus in the presence of wild-type helper appear to be lost and defective genomes consist predominantly of multiple copies of the origin of DNA replication, often interspersed with cellular DNA (Lee *et al*, 1975). Defective genome sizes are usually 10 to 20% shorter than that of wild-type virus, which are probably the smallest genome sizes that can be efficiently packaged.

Lee and Nathans (1979) demonstrated that reiterations of the replication origin confer a selective growth advantage to defective viruses over wild-type helper. These authors determined the relative rates of replication of defective particles containing different numbers of origins of replication and showed that the rate of replication is an exponential function of the number of origin segments per genome. In contrast, none of the defective genomes were encapsidated more efficiently than helper virus genomes, indicating that encapsidation efficiency is not proportional to the number of origins and is probably not a major factor in the evolution of defective SV40 particles.

O'Neill *et al.* (1982) have isolated two types of SV40 DI particles which can complement for infectivity. One type contains nearly all of the early SV40 sequences encircled by multiple viral origin and termination sequences. The second type contains the late SV40 sequences, also encircled by multiple origins and termini. Both defective DNAs were cloned into the bacterial plasmid pBR322 and shown to be infectious when coinoculated into BSC-1 cells, but not when applied separately. These DI particles therefore represent a new class of SV40 with a bipartite genome. The presence in these DI particles of multiple termination sequences adjacent to replication origins is thought to prevent transcription initiated at remote origins from interfering with normal transcription and/or replication of these particles, thereby allowing the selective growth advantage conferred by the multiple origins to be retained (Carroll and O'Neill, 1978; O'Neill *et al.*, 1982).

The nucleotide sequences at recombination sites in a number of SV40 DI particles have been determined (Johnson *et al.*, 1982 and references therein). The general conclusions to have emerged from analysis of these sequences are that recombination is not limited to a single specific nucleotide sequence, does not require extensive

homologies between parental strands, and may in some instances involve regions of patchy homology between recombining molecules as well as A-T rich sequences. In view of these findings, Johnson *et al* (1982) have ruled out a general homologous or sequence-specific recombination mechanism and suggested that repair and ligation of linear DNA molecules that result from the action of nonspecific cellular endonucleases are responsible for the formation of defective SV40 genomes.

## 1.17 Herpesviruses

The herpesviruses are enveloped icosahedral animal viruses containing a linear dsDNA genome. Several members of the herpesvirus group have been shown to generate DI particles upon repeated high-multiplicity passage, including herpes simplex virus (HSV), equine herpesvirus, cytomegalovirus, etc (reviewed by Frenkel, 1981). Of these, the HSV DI particles have been the best characterized.

The HSV genome is composed of two covalently linked components, large (S) and small (S), each of which contains a unique stretch of viral DNA sequences bracketed by inverted repeats (Roizman, 1979). DNA extracted from HSV DI particles consists of tandem reiterations of defined viral DNA sequences (repeat units) containing a segment from the end of the S component. Two classes of defective genomes have been identified which differ with respect to the additional sequences that they contain. Class I defective genomes contain additional DNA sequences from S, whereas in class II defective genomes the bulk of the repeat units is derived from DNA sequences from L.

Vlazny and Frenkel (1981) demonstrated that cotransfection of rabbit skin cells with standard helper virus DNA and monomeric repeat units (Mr 5.5 X $10^6$) derived from class I defective genomes of HSV-1 results in the regeneration of full-length (Mr 100 X $10^6$) defective genomes consisting of head to tail reiterations of the 'seed' unit. These authors concluded that the repeat units contain all the *cis* functions required for replication and packaging, and that the regenerated defective genomes arise by a rolling circle mechanism of replication in which a single repeat unit serves as the circularized template.

Defective HSV particles isolated from the nuclei of infected cells contain DNA molecules ranging from repeat unit size to full-length genome size (Vlazny *et al.*, 1982). These were shown to consist of different integral numbers (1 to 18) of repeat units. However, only DI particles containing full-length DNA molecules become enveloped and translocated to the cytoplasm. As all defective DNA molecules terminate at one end with the S-end sequences of standard virus DNA, Vlazny *et al.* (1982) concluded that these sequences contain a recognition signal for cleavage and packaging of viral DNA. Provided that they contain the cleavage/packaging signal, molecules significantly shorter than full-length standard virus can become encapsidated. However, the reason for the selective maturation and translocation to the cytoplasm of full-length DI particles is not clear.

## 1.18 Defective RNA plant viruses

In contrast to animal viruses, there is no conclusive evidence for DI particles of RNA plant viruses (Francki, 1985). Deletion mutations in RNA plant viruses have been reported for wound tumor virus (WTV) (Reddy and Black, 1974), pea enation mosaic virus (PEMV) (Adam *et al.*, 1979), soil-borne wheat mosaic virus (SBWMV) (Shirako and Brakke, 1984) and beet necrotic yellow vein virus (BNYVV) (Bouzoubaa *et al.*, 1985). In the case of WTV and PEMV, the deletions are associated with a loss in insect-transmission. In none of these examples, however, is there any evidence for loss or reduction of infectivity as a result of the deletions.

The satellite RNA plant viruses (reviewed by Francki, 1985) represent another group of plant viruses with properties similar to those of DI particles. Indeed, satellites are completely dependent for their replication on the helper-viruses with which they are associated and usually interfere with the replication of helper-virus and modify symptom expression on the host plants. They are different from DI particles, however, in that their genome has very little or no nucleotide sequence homology to the helper-virus genome, suggesting that they are not derived from the latter.

A typical example of satellitism in plant viruses is set by the interaction between tobacco necrosis virus (TNV) and its satellite (STNV). STNV RNA has negligible nucleotide sequence homology (<2%) to

TNV RNA (Shoulder *et al.*, 1974) and has been shown by *in vitro* translation (Klein *et al.* 1972) and amino acid sequence determination of the virus protein (Henriksson *et al.*, 1981) to code for its own coat protein . The interaction between STNV and TNV is highly specific in that the satellite virus can be activated by TNV only but not by other plant viruses (Kassanis and Nixon, 1961). This specificity is also manifested at the level of STNV and TNV strains. For instance, TNV strains capable of activating STNV-A or -B appear to be incapable of activating STNV-C, and *vice versa* (Kassanis and Phillips, 1970). STNV interferes with the replication of TNV. Addition of STNV and a progressive increase of its concentration in the inoculum result in a progressive suppression of the TNV reproduction, up to an extent where the presence of TNV in the inoculated plants becomes difficult to detect serologically (Kassanis, 1962). Interference can also occur between different STNV strains when one of them is present in the mixture in great excess. In this case the replication of the STNV strain which was in minority is depressed greatly (Kassanis and White, 1972). The mechanism responsible for the interference between STNV and TNV is not known but is thought to occur both at the level of translation and RNA replication (Jones and Reichmann, 1973).

## 1.19 Mechanisms for the generation of DI particles

The precise nature of the mechanisms for generating DI particles has not been determined for any virus as yet. The rearranged genomes of DI particles of DNA viruses could arise either by recombination events, including generalized (homologous), site-specific, or 'illegitimate' (nonhomologous) recombination, or by errors in viral replication. Analysis of DI particle genomes at the molecular level suggests the existence of multiple mechanisms for DI particle genesis and a number of hypothetical models have been proposed.

Lazzarini *et al.* (1981) have proposed a generalized scheme for the generation of DI particles of RNA viruses based on aberrant replication events. The scheme, which may apply to DI particles of ssDNA viruses, postulates that the replicase detaches from its template while still retaining the nascent daughter strand and resumes synthesis and elongation of the nascent strand at a different site. Depending on whether the new site is on another template or on the same template,

different types of DI particles could be generated. This is illustrated in Figs.4-6 for the generation of DI particles containing simple deletions, DI particles containing complementary termini, and 'snap back' DI particles. Resumption of replication on a particular template may reflect the relative concentration of this template in the vicinity of the polymerase (Lazzarini *et al.*, 1981). In addition, sequence specificity may determine the preference for a particular site. In the case of vesicular stomatitis virus (VSV), for example, resumption of RNA synthesis takes place preferentially at a site containing sequence similarities to the replication initiation site (Schubert *et al.*, 1979).

Another model based on errors in DNA replication (Efstratiadis *et al.*, 1980) could account for the generation of deletions in the genomes of some defective dsDNA viruses. The method was originally proposed for the generation of deletions in β-like globin genes and is illustrated in Fig.8. It involves mispairing between two short complementary sequences (R1' and R2) located on opposite strands of a duplex molecule during passage of a replication fork. As a result, a single-stranded loop is produced which could then be excised by DNA repair enzymes. Subsequent replication would generate a normal daughter duplex and a duplex containing a deletion comprising one of the two repeats involved in the mispairing (R1/R1' or R2/R2') and the sequences between them.

As mentioned above, rearranged DNA virus genomes may result from recombination events between parental DNA molecules. A deletion could arise if recombination occurs at different sites in the same DNA molecule. On the other hand, an insertion could result from unequal crossing-over between different DNA molecules. Both generalized and site-specific recombination involve reactions between homologous sequences. In generalized recombination, the enzymes responsible can use virtually any pair of homologous sequences as substrates, while only specific pairs of sequences are used in site-specific recombination. However, a number of DI particles that have been analysed at the molecular level lack any extensive homology ,symmetry, or sequence similarities around the recombinant joints. While some of these DI particles may be generated by one of the mechanisms involving errors in DNA replication, others are probably generated by 'illegitimate' recombination between parental DNA molecules. One such mechanism could

Legends to Figs. 4, 5, 6 and 7.

Fig. 4. **Replicative error model for the generation of DI particles containing simple deletions**

Virus replication terminates prematurely and then resumes at a more distant site towards the 5' end of the same template. The polymerase is shown as a black circle and the location of the deletion as an open triangle. (Reproduced from Lazzarini *et al.*, 1981).

Fig. 5. **Replicative error model for the generation of DI particles containing complementary termini.**

Virus replcation terminates prematurely and then resumes at a different site on the nascent strand. The polymerase is shown as a black circle and the location of the deletion as an open triangle. (Reproduced from Lazzarini *et al.*, 1981).

Fig. 6. **Replicative error model for the generation of 'snap-back' DI particles.**

A polymerase complex overtakes another, 'jumps' to the nascent strand in front and resumes synthesis. The polymerase is shown as a black circle and the location of the deletion as an open triangle. (Reproduced from Lazzarini *et al.*, 1981).

Fig. 7. **Model for topo I mediated excisional recombination.**

Topo I molecules (open circles) bind to sequences in the vicinity of the crossover points (rectangles) and each introduces a single-stranded break. The presence of additional breaks opposite to the topo I cleavage sites could result in a double-stranded break. Topo I-catalyzed formation of phosphodiester bonds and repair of the opposite strand would result in cyclisation of the excised DNA-containing SV40 sequences and ligation of the left (L) and right (R) chromosomal ends. (Reproduced from Bullock, 1985).

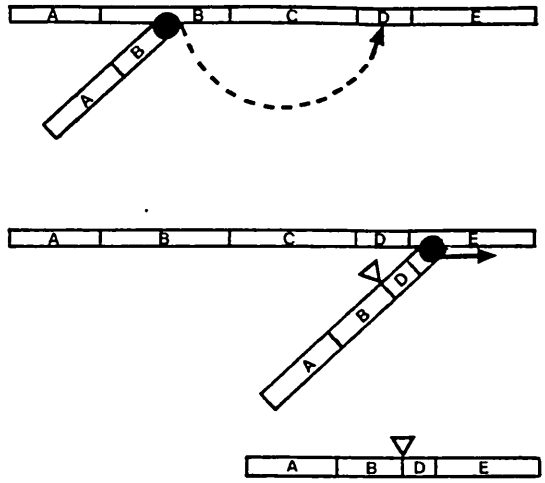



Fig. 4

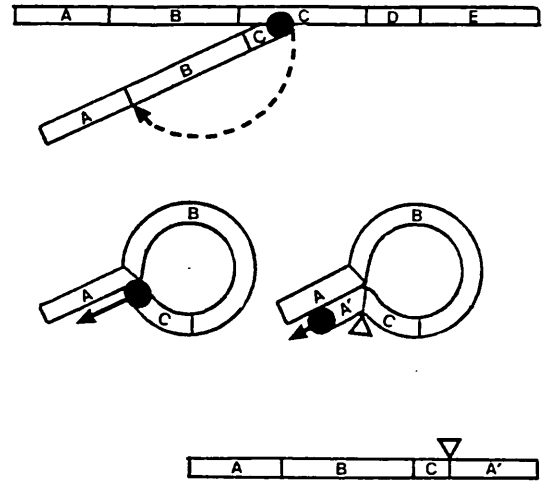



Fig. 5

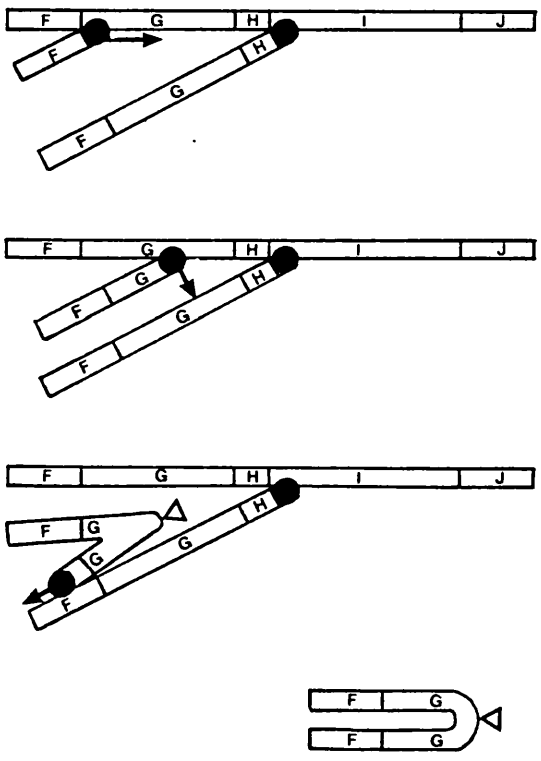



Fig. 6

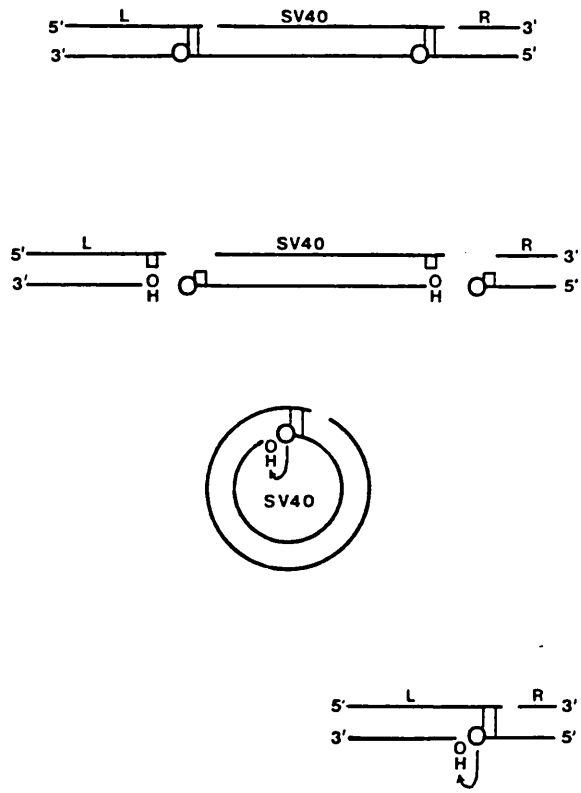



Fig. 7

involve cleavage at random locations by endogenous endonucleases followed by repair and ligation of the resultant linear fragments.

DNA topoisomerases may play a role in the mediation of 'illegitimate' recombination. These enzymes catalyze a variety of changes in the topological structure of DNA, including knotting, unknotting, catenation, decatenation, supercoiling and/or relaxation of DNA molecules (reviewed by Gellert, 1981). Topoisomerases are divided into two types, according to the mechanisms that they employ. Type I topoisomerases (topo I) break and reseal one strand of dsDNA at a time, whereas type II topoisomerases (topo II) act by introducing transient double-stranded breaks. Although these enzymes have no absolute sequence specificity, they show some degree of preference for particular sites. Thus, the majority of eukaryotic topo I cleavage sites occur immediately 3' of the sequences CTT or GTT (Been *et al.*, 1984). *Drosophila* topo II, on the other hand, has a preference for the consensus sequence 5'GT.A/TAY↓ATT.AT..G 3', where a dot means no preferred nucleotide, Y is for pyrimidine, and the arrow shows the point where cleavage occurs (Sander and Hsieh, 1985).

Bullock *et al.* (1985) showed that crossover points for chromosomal excision of SV40 from nontandem inserts are associated with eukaryotic topo I cleavage sites *in vitro*, which suggested a model for topo I-mediated excisional recombination. The model is illustrated in Fig.7, and could account for the generation of defective viral genomes containing deletions and/or insertions. Breakage of a DNA molecule at two different sites, followed by excision of the intervening fragment and resealing could generate a deletion. Alternatively, an insertion could be generated if resealing occurs between DNA fragments of different origin.

## 1.20 Introduction to the present work

One aim of the present work was to characterize the subgenomic species of TGMV. I was interested to know whether these species consisted of simple deletion mutants, or whether they contained other DNA rearrangements, such as reiterations, insertions, substitutions etc. Also, the molecular characterization of the subgenomic species might provide an insight into the mechanism responsible for its generation. For instance, the nucleotide sequences surrounding the recombinant

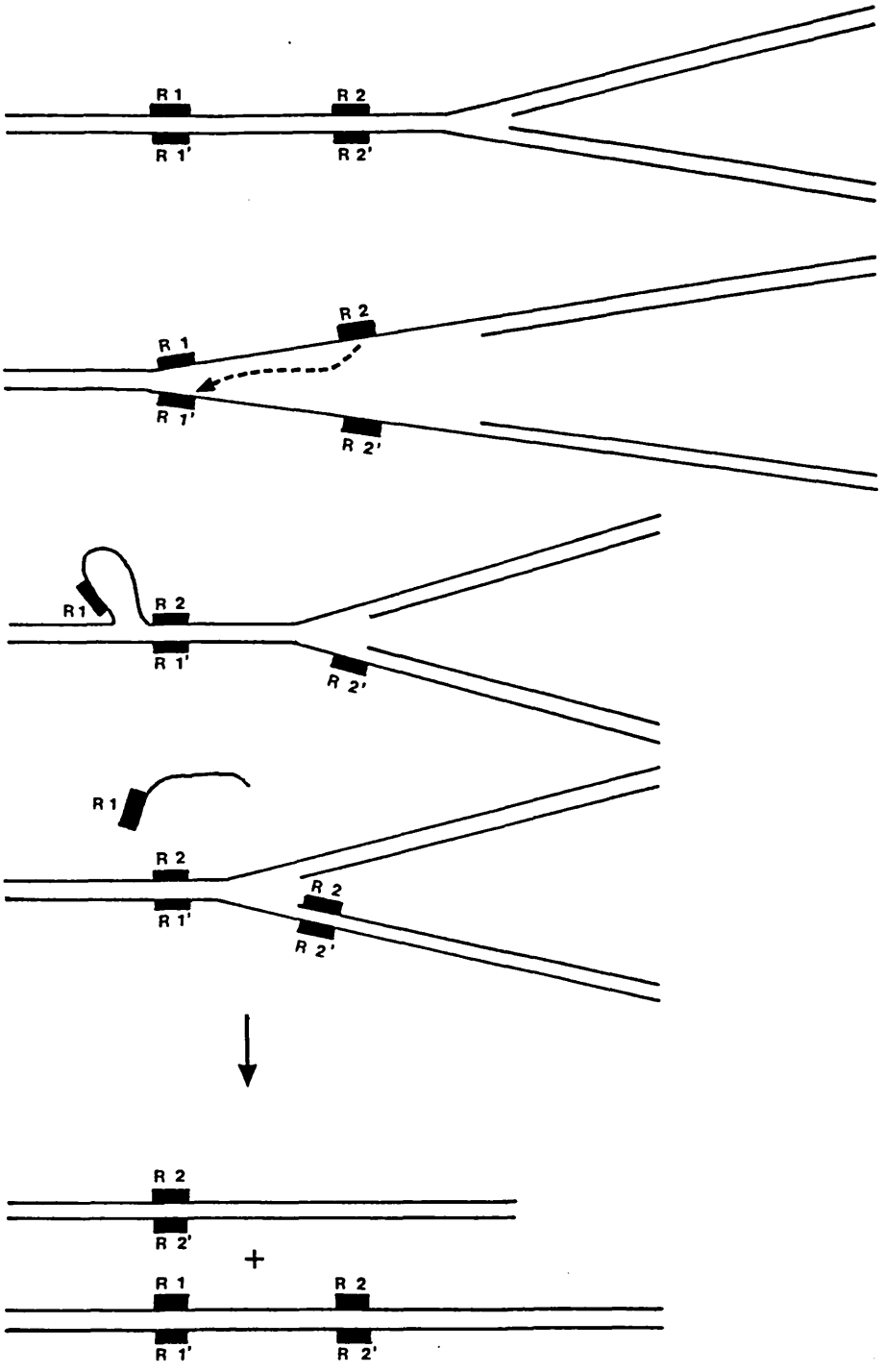



Fig. 8. **'Slipped mispairing' model for the generations of deletions during DNA replication.**

Mispairing of complementary sequences (R1' and R2') during passage of a replication fork results in a single-stranded loop. Excision of the loop by DNA repair enzymes followed by subsequent replication results in a normal daughter duplex and a duplex containing a deletion. Solid boxes represent short direct repeats (R1 and R2) and their complements (R1' and R2'). (Reproduced from Efstradiatis *et al.*, 1980).

joints might give some indications as to whether these molecules arise by homologous recombination or by any of the other mechanisms outlined in section 1.19. Characterization of the subgenomic species might also provide useful information with regard to the construction of plant vectors based on geminivirus replicons. TGMV subgenomic DNA consists of a population of naturally occurring, viable deletion mutants. Therefore, the location of the deletions in these molecules might indicate regions in the TGMV genome that may be replaced with foreign DNA without hampering the ability of the resulting chimeric DNA molecules to replicate and become encapsidated. Defective virus vector systems have been designed for a number of animal DNA viruses, including herpes virus (Spaete and Frenkel, 1982), SV40 (reviewed by Subramani and Southern, 1983) and adenovirus (Thummel *et al.*, 1981). These vector systems typically consist of two components: the defective virus DNA in which the foreign gene has been inserted, and the helper virus DNA which provides all the necessary *trans* functions for replication and packaging of the chimeric defective genome.

Rapid progress in the development of plant vectors based on the Ti plasmid of *A.tumefaciens* and in the techniques for the direct transfer of naked DNA into plant protoplasts are being used to good advantage to study gene expression and regulation in whole plants. Also, the parallel development of selectable antibiotic marker genes for plant transformation (Herrera-Estrella *et al.*, 1983) has proved invaluable in these studies and in the selection of transformed cells.

The natural ability of *A.tumefaciens* to transfer and integrate part of the Ti plasmid DNA (known as T-DNA) into the genome of host plants has been exploited for the construction of vectors capable of stably integrating foreign DNA into plant cell chromosomes. The oncogenic properties of the T-DNA have been circumvented by the development of 'disarmed' Ti-vectors lacking tumour-inducing genes (*onc* genes) (Zambryski *et al.*, 1983). In addition, problems associated with the large size of the Ti-plasmid have been largely obviated by the construction of binary vectors (Hoekema *et al.*, 1983). These consist of a two plasmids system: the vector and a helper plasmid. The vector replicates both in *A.tumefaciens* and *E.coli* and is easily mobilized between bacterial strains. It contains the T-DNA border sequences (required for chromosomal integration) flanking a selectable marker and suitable cloning sites. The helper plasmid consists of a Ti-plasmid

which has suffered a large deletion including the T-DNA but still retains the virulence (*vir*) region (required for T-DNA transfer). Following insertion of a foreign gene, the vector is mobilized from *E.coli* to *A.tumefaciens* containing the helper plasmid by conjugation. Complementation in *trans* then allows transfer and integration of the DNA segment between the border sequences of the vector into the plant nuclear genome.

Ti-plasmid - based vectors have a wide host range among dicotyledonous plants and apparently have no packaging constraints. Their main disadvantage is the inability so far to transform economically important monocotyledonous plants such as cereal crops and the long period of time required for the regeneration of transformed cells.

Direct transfer (without vectors) of foreign DNA into plant protoplasts has also been achieved. Paskowski *et al.* (1984) first reported the successful transformation of *N.tabacum* protoplasts by direct gene transfer. The method has now been extended to include cells from monocotyledonous as well as dicotyledonous species (Potrykus *et al.*, 1985; Lörz *et al.*, 1985). As with Ti-vectors, the foreign DNA is stably mantained and expressed over several generations. The main limitations include the present inability of plant regeneration from a number of agriculturally important plant species including cereal crops, and a relatively low and variable efficiency of transformation.

Thus far there has been little progress in the development of geminivirus-based vectors. Conceivably, these viruses might be used in the construction of high copy number transient expression vectors (Coutts *et al.*, 1986). Such vectors would allow the investigator to look at gene products, either RNA or protein, within hours of DNA uptake by plant cells. By introducing mutations in the DNA sequences flanking a particular gene, promoter function and other factors affecting plant gene expression could be rapidly analysed. This approach has indeed been successfully used to analyse the mechanisms of transcription and post-transcriptional processing in animal cells (reviewed by Rigby, 1982).

One of the problems associated with vectors based on viral replicons is the limit imposed on the amount of DNA that can be packaged into virus particles. For instance, insertions larger than 1.3Kb abolish infectivity of cauliflower mosaic virus (CaMV), another

potential vector for plant transformation (Walden and Howell, umpublished observations; cited in Howell, 1982). Because of the small size of the geminate particles, packaging constraints are likely to be particularly severe in the case of geminivirus-based vectors. This could be circumvented by the construction of geminivirus-based vectors that can replicate as episomes (Buck and Coutts, 1983). One possibility would be to use plant cells containing the viral genes required for geminivirus replication integrated into their chromosomes. Expression of these genes should then be able to support the sustained replication of plasmid vectors containing the geminiviral origin of replication (Coutts *et al.*, 1986). This host-vector system would be analogous to the SV-ORI plasmid/COS cell system developed by Mellon *et al.* (1981).

In contrast to Ti plasmids, geminiviruses may be used to construct vectors for the transformation of monocotyledonous plants. The nucleotide sequence of WDV, presented in this thesis, is a prerequisite for the construction of such vectors. In addition, comparison of the WDV sequence to those of other geminiviruses might help to clarify the evolutionary relationships among these viruses as well as increase our knowledge of their genome organisation and function.

## 2. Materials and methods

### 2.1 Chemical supplies

Materials used in the study were supplied by the following companies:

AMERSHAM : Amersham International PLC., White Lion Road, Amersham, Buckinghamshire, HP7 9LL.

ANGLIAN : Anglian Biotechnology Ltd., Unit 8, Hawkim Road, Colchester, Essex, CO2 8JX.

BDH : BDH Chemicals Ltd., Freshwater Road, Dagenham, Essex, RM8 1RZ.

BCL : Boehringer Corporation (London) Ltd., Boehringer Mannheim House, Bell Lane, Lewes, E. Sussex BN7 1LG.

BRL : Bethesda Research Laboratories (UK) Ltd., P.O.Box 145, Science Park, Cambridge CB4 4BE.

DIFCO : Difco Laboratories, P.O.Box 14B, Central Avenue, West Molesey, Surrey KT8 OSE.

FISONS : Fisons PLC, 41 Gatwick road, Crawley, Sussex RH10 2UL.

GILSON : Gilson Medical Electronics. Products supplied by Anachem, Anachem House, 20 Charles street, Luton, Bedfordshire LU2 OEB.

| | | |
|---|---|---|
| ISCO | : | Isco. Products supplied by Life Science Laboratories Ltd, Sarum Road, Luton, Bedfordshire LU3 2RA. |
| LKB | : | LKB Instruments Ltd., 232 Addington Road, Selsdon, South Croydon, Surrey CR2 8YD. |
| MILES | : | Miles Laboratories Ltd., P.O.Box 37, Stoke Poges, Slough, Bucks SL2 4LY. |
| MILLIPORE | : | Millipore (UK) Ltd., Abbey Road, London NW10 75P, UK. |
| NEB | : | New England Biolabs Inc., Products supplied by CP Laboratories Ltd., P.O.Box 22, Bishops Stortford, Hertfordshire CM22 7RQ. |
| NEN | : | New England Nuclear, 2 New Road, Southampton SO2 OAA. |
| OXOID | : | U.K. Order Department, Oxoid Ltd., Wade Road, Basingstoke, Hampshire, RG 24 OPW. |
| PACKARD | : | Packard Instruments Ltd., 13/17 Church Road, Caversham, Reading, Berkshire RG4 7AA. |
| P & S | : | P & S Biochemicals Ltd., 38 Queensland Street, Liverpool, L7 3JG. |
| PALL | : | Pall Process Filtration Ltd., Europe House, Havant Street, Portsmouth PO13 PD. |
| PHARMACIA P-L | : | P-L Biochemicals Ltd., P.O.Box 98, Northampton, NN3 1AW. |
| RAVEN | : | Raven Scientific Ltd., P.O.Box 2, Havehill, Suffolk, CB9 OES. |

ROSE : Rose Chemicals Ltd., 83 Darenth Road, London N16 6EB.

SCHLIECHER AND SCHüLL : Products supplied by Anderman and Co., Central Avenue, East Molesey, Surrey KT8 0Q2.

SIGMA : Sigma Chemical Co. Ltd., Fancy Road, Poole, Dorset BH1 77NH.

UNIVERSAL SCIENTIFIC : Universal Scientific Ltd., 231 Plashet Road, London.

VENTRON : Ventron GmbH. Products supplied by Lancaster Synthesis Ltd.,Eastgate, White Lund, Morecambe, Lancashire.

WACKER CHEMICALS : Wacker Chemicals Ltd., The Clock Tower, Mount Felix, Bridge Street, Walton-on-Thames, Surrey, KT12 1AS.

WHATMAN : Whatman Labsales, Springfield Mill, Maidstone, Kent, ME14 2LE.

WORTHINGTON : Worthington Diagnostic Systems Inc., Supplied by Flow Laboratories Ltd., P.O.Box 17, Second Avenue, Industrial Estate, Irvine, Scotland KA12 8NB.

## 2.2 Abbreviations

The following are non-standard abbreviations:

b : base(s)

bp : base pair(s)

BSA : Bovine serum albumin

cccDNA : Covalently closed circular DNA (supercoil)

cpm : counts per minute

D : Dalton

d- : deoxy, e.g. dNTP, deoxynucleoside triphophate

dd- : dideoxy, see above

$dH_2O$ : distilled water

$ddH_2O$ : double distilled water

ds : double stranded

DTT : dithiothreitol

EDTA : ethylene-diamine-tetra-acetic acid (di sodium salt)

g : relative centrifugal force

hr : hour

Kb : Kilobase(s)

kD : Kilodalton(s)

min : minute

Mr : relative molecular mass

ocDNA : open circular DNA

PEG : polyethylene glycol 6,000

RF : replicative form

rpm : revolutions per minute

SDS : Sodium dodecyl sulphate

sec : second

ss : single-stranded

TCA : trichloroacetic acid

TEMED : N, N'-tetramethylenediamine

Tris : trishydroxymethylaminoethane

u : unit

UV : ultra violet

**Amino acid codes**

| | | |
|---|---|---|
| A | Ala | Alanine |
| B | Asp, Asn | Aspartic, Asparagine |
| C | Cys | Cysteine |
| D | Asp | Aspartic |
| E | Glu | Glutamic |
| F | Phe | Phenylalanine |
| G | Gly | Glycine |
| H | His | Histidine |
| I | Ile | IsoLeucine |
| K | Lys | Lysine |
| L | Leu | Leucine |
| M | Met | Methionine |
| N | Asn | Asparagine |

| | | |
|---|---|---|
| P | Pro | Proline |
| Q | Gln | Glutamine |
| R | Arg | Arginine |
| S | Ser | Serine |
| T | Thr | Threonine |
| V | Val | Valine |
| W | Trp | Tryptophan |
| X | Xxx | Unknown |
| Y | Tyr | Tyrosine |
| Z | Glu,Gln | Glutamic, Glutamine |
| * | End | Terminator |

**Virus names**

| | |
|---|---|
| AMV | avian myoblastis virus |
| AAV | adeno-associated virus |
| Ad12 | adenovirus type 12 |
| Ad16 | adenovirus type 16 |
| AdFL | adenovirus strain FL |
| BCTV | beet curly top virus |
| BGMV | bean golden mosaic virus |
| BNYVV | beet necrotic yellow vein virus |
| CaMV | cauliflower mosaic virus |
| CLV | cassava latent virus |
| CSMV | chloris striate mosaic virus |
| EuMV | euphorbia mosaic virus |
| HSV | herpes simplex virus |
| MSV | maize streak virus |
| MVM | minute virus of mouse |
| MYMV | mungbean yellow mosaic virus |
| PEMV | pea enation mosaic virus |
| PYMV | potato yellow mosaic virus |
| SBWMV | soil-borne wheat mosaic virus |
| SLCV | squash leaf curl virus |
| STNV | satellite tobacco necrosis virus |
| SV40 | simian virus 40 |
| TGMV | tomato golden mosaic virus |

| | |
|---|---|
| TNV | tobacco necrosis virus |
| TobLCV | tobacco leaf curl virus |
| TomLCV | tomato leaf curl virus |
| TPCTV | tomato pseudo-curly top virus |
| TYDV | tobacco yellow dwarf virus |
| TYLCV | tomato yellow leaf curl virus |
| WDV | wheat dwarf virus |
| WTV | wound tumor virus |

## 2.3 Safety considerations

A protective laboratory coat was worn at all times and protective gloves were worn when handling hazardous chemicals, radioactive isotopes and carrying out DNA recombinant work. Glasses were worn when working with radioactive isotopes, and experiments were performed behind a protective perspex screen. Gëiger monitors were used to check potential radioactive contamination. All radioactive waste was disposed of in accordance with the Imperial College Safety Regulations and a record of disposals maintained.

Registration with the Genetic Manipulation Advisory Group (GMAG) together with the required medical examination were completed for work with recombinant DNA. An Envair class II microbiological safety cabinet was used for recombinant DNA work. All materials for disposal were autoclaved and spillages were cleaned with concentrated hypochlorite followed by industrial methylated spirit.

Work on TGMV and WDV DNA was classified by the Department of Pure and Applied Biology Microbiological Safety sub-committee as good microbiological practice, and was carried out under the destructive pests and diseases of plants Act with licences HH 1234A/73 and PHF 29A/143 respectively which were issued by the Ministery of Agriculture, Food and Fisheries (MAFF) and held by Dr. R.H.A. Coutts.

## 2.4 Common buffers

NaPB : 50mM $Na_2H_2PO_4/Na_2HPO_4$ (pH 7.0)
prepared as a 20 X stock solution.

SSC : 150mM Nacl, 15mM sodium citrate
prepared as a 20 X stock solution.

TAE : 40mM Tris-Hcl (pH 8.2) 5mM acetic acid,
1mM EDTA.
prepared as a 25 X stock solution.

TE : 10mM Tris-Hcl (pH 8.0) 1mM EDTA

## 2.5 Miscellaneous

All chemicals used were supplied by BDH unless otherwise stated and were of analytical grade.

All solutions were autoclaved where possible. Those which were not autoclavable were filter sterilised (Millipore, 0.2µm) before use. All glassware and plasticware for storage of any chemical materials or for use in recombinant DNA work were autoclaved or purchased in a sterile state. Autoclaving was carried out for 20min. at 121°C, at 15lb/in$^2$ pressure.

A number of items of glassware and plasticware were siliconized by soaking in a 5% solution of dichlorodimethyl silane (Sigma) in chloroform for 30min. in a fume-cupboard, followed by extensive rinsing in several changes of $dH_2O$ prior to baking at 180°C for 2hr

Phenol was redistilled and stored in aliquots at -70°C. At least one day before use, frozen phenol was melted at 68°C and 8-Hydroxyquinoline (Sigma) added to a final concentration of 0.1% (w/v). This solution was equilibrated with 0.1M Tris (pH 8.0) followed by TE (pH 8.0) and then stored at 4°C.

Dialysis tubing was cut into pieces.of convenient length (10-20cm) and boiled for 10min in a large volume of 10mM EDTA. The tubing was then rinsed in $dH_2O$ and boiled for a further 10 min in

1mM EDTA. After cooling, tubing was stored at 4°C in 1mM EDTA. Prior to use the tubing was thoroughly washed in $dH_2O$.

Formamide was deionised by mixing with 5g/100 ml of Amberlite MB-1 "monobed" resin (BDH) continuously for 30min at room temperature. Deionized formamide was filtered through filter paper and stored at -20°C.

Centrifugations were performed in MSE Hi-spin 18 or 21 centrifuges for low speed spins and in Beckman L8-55 or Spinco L-1 centrifuge for high speed spins. A Hettick mikrofuge was used to centrifuge samples contained in Eppendorf tubes.

Gilson "pipetman" automatic pipettors fitted with disposable tips were used to measure volumes from 1 to 1,000µl.

Where samples in Eppendorf tubes were to be heated, this was performed in a multi-block heater (Lab-line instruments) which contained paraffin.

## 2.6 Phenol extraction and ethanol precipitation

Nucleic acid solutions were mixed with equal volumes of buffer saturated phenol in either Eppendorf tubes or larger polypropylene tubes fitted with plastic caps. The contents were mixed by vortexing or inversion until an emulsion formed. The phases were separated by centrifugation for 5min in a microfuge or 10min at 10,000rpm in either the MSE HS18 or HS21 centrifuges. The upper aqueous phases were transferred to fresh tubes and extracted further, first with a 1:1 mixture of phenol and sevag (24:1 mixture of chloroform and isoamyl alcohol) and then with sevag alone. The nucleic acids were recovered by ethanol precipitation as follows.

Nucleic acid solutions were mixed with 2 vol ethanol and kept at -20°C overnight. Low ionic strength solutions were adjusted to 0.25M Sodium Acetate (from a 2.5M sodium acetate (pH 5.2) stock solution) prior to ethanol addition. Precipitated nucleic acids were pelleted by centrifugation (5 to 10min in a microfuge or 10 to 30min in either the MSE HS21 or HS18 centrifuges), the pellets were washed in 70% (v/v) ethanol, dried under vacuum and then resuspended in a buffer of choice.

## 2.7 Quantitation of nucleic acids

### 2.7.1 Spectrophotometric determination

The concentration of various nucleic acid preparations was determined by UV absorption spectroscopy using a Beckman Model 35 spectrophotometer with quartz cells of path length 1cm. Readings were taken at wavelengths of 260nm and 280nm. The concentration of nucleic acid in the samples was calculated from the reading at 260nm using the following relationships:

1 $O.D._{260}$ = 50µg/ml dsDNA
1 $O.D._{260}$ = 40µg/ml ssDNA

The ratio between the readings at 260nm and 280nm ($O.D._{260}/O.D._{280}$) provided an estimate of the purity of the nucleic acid preparation. Pure preparations of DNA were assumed to have a ratio of $\geq 1.8$.

### 2.7.2 Determination by UV induced fluorescence on agarose plates

3 to 5 ml of a 1% (w/v) agarose solution containing 0.5 µg/ml of ethidium bromide were poured into Petri dishes and allowed to set. The DNA samples, along with a series of DNA solutions of known concentration (0.5-2.0µg/ml) were spotted onto the surface of the agarose and allowed to dry. Plates were observed under UV illumination and the concentration of DNA estimated by comparing the intensity of fluorescence in the sample with that of the solutions of known concentration.

### 2.7.3 Trichloroacetic acid (TCA) precipitation of $[^{32}P]$ labelled DNA

Samples (1-2µl) for assay were spotted onto the centres of whatman 3MM filter discs. The discs were then washed with 50ml of 5% (w/v) TCA containing 10mM sodium pyrophosphate followed by 20ml of 85% ethanol, dried under an infra-red lamp and placed into

scintillation vials containing 10ml of a scintillant prepared as follows:

| | |
|---|---|
| 666ml | toluene |
| 333ml | 96% ethanol |
| 4g | PPO (2, 5-diphenyloxazole; Sigma) |
| 0.1g | POPOP (1, 4-bis(5-phenyloxazole) benzene; Packard) |

The radioactivity on the discs was then counted in a Packard Tri-carb scintillation counter.

## 2.8 Sephadex chromatography

### 2.8.1 Sephadex G-50 column chromatography

Sephadex G-50 (30g) was added to 10mM NaPB (pH 7.0), 0.2% (w/v) SDS (250ml) and autoclaved for 15min at 15lb/in$^2$. The Sephadex was allowed to cool at room temperature and poured into a 10ml glass pipette plugged with siliconized glass wool until the column was three-quarters full with packed Sephadex. The column was washed with several volumes of 10mM NaPB (pH 7.0), 0.2% (w/v) SDS. Before use, the buffer in the column was allowed to level with the Sephadex and the DNA sample (in a total volume of 0.1ml) loaded directly above the Sephadex. The DNA was eluted by continuous addition of 10mM NaPB (pH 7.0), 0.2% (W/v) SDS to the column. Fractions (0.5ml) off the column were collected with a Gilson microcol TDC 80 fraction collector. Progress of the incorporated and unincorporated [$^{32}$P] dNTPs through the column was monitored with a hand-held Gëiger-counter. Following collection of the leading-peak labelled material, the column was disconnected from the fraction collector and unincorporated dNTPS eluted into a wash beaker. The column was then clamped off for further use. The radioactivity in each individual fraction was measured in a scintillation counter by Cerenkov counting. Fractions containing labelled material were pooled and stored at -20°C for further use.

### 2.8.2 "Spun-column" chromatography

Sephadex G-75 (30g) was added to 250ml of TE buffer (pH 8.0), autoclaved for 15min at 15lb/in$^2$ and cooled to room temperature. Columns were prepared by filling to the brim 2ml disposable syringes (plugged with discs of whatman 3MM paper) with the Sephadex. Columns were inserted into 15ml Corex tubes and centrifuged in a bench top centrifuge at 1,000rpm for 2min to pack the Sephadex. The eluted buffer was discarded and decapped Eppendorf tubes fitted to the base of the corex tubes, under the columns. Samples of DNA (in a total volume of 0.1ml) were applied to the columns and the assemblies centrifuged at 1,000rpm for 1min. Eppendorf tubes containing eluant from the syringes were stored and fresh Eppendorf tubes replaced onto the assemblies. The procedure was repeated 3 to 4 times after application of further TE (pH 8.0) (0.1ml) to the columns. The radioactivity in each of the eppendorf tubes was measured in a scintillation counter as previously (Section 2.8.1). After use the spun-columns were discarded.

## 2.9 Propagation of Tomato Golden Mosaic Virus

### 2.9.1 Growth of plants

Seeds of *Nicotiana benthamiana* (supplied by Dr. R.H.A. Coutts) were germinated on acid-washed sand moistened with $dH_2O$. Seedlings were potted in John Innes nº 2 compost in 4" pots and kept under polythene sheets for the first two weeks of growth to prevent dessication. Plants were grown in an insect-free green house at a temperature of 25-28°C, 70-78% relative humidity, under mercury vapour or incandescent strip lights of *ca.* 15,000 lux with a 16 hr photoperiod.

### 2.9.2 Origin of tomato golden mosaic virus

The virus was originally supplied by Dr. A.S.Costa from the Instituto Agronomico, Estado de Sao Paulo, Brazil and propagated by Dr. W.D.O. Hamilton (Hamilton *et al.*, 1981).

### 2.9.3 Preparation of viral inoculum

TGMV inoculum was prepared according to the procedure of Stein *et al.* (1983).

TGMV-infected *N. benthamiana* plants were homogenized in a polytron with ice-cold homogenization buffer (0.1M trisodium citrate, 0.75% (w/v) sodium sulphite, 5mM disodium EDTA, 1% (v/v) 2-mercaptoethanol, 0.325% (w/v) L-ascorbic acid (Fisons)). Homogenization buffer was used at a rate of 2ml/g of plant tissue. The homogenate was made 2.5% (v/v), with respect to Triton X-100 and stirred for 16hr at 4°C. The homogenate was then filtered through muslin and centrifuged (10,000rpm, 15min) at 4°C. The supernatant was further centrifuged at 40,000rpm for 2hr in a Beckman type 65 rotor or at 27,000rpm for 3hr in a Beckman type 30 rotor at 4°C. The resulting pellets were resuspended in 1XTAE buffer and stored in aliquots at -20°C.

### 2.9.4 Inoculation of plants

*N. benthamiana* plants were grown to the 6-8 leaf stage before inoculation. The lower leaves were removed leaving the upper 3 leaves which were then dusted with carburundum (500 mesh) Viral inoculum (10-15µl) was applied to each leaf and rubbed with a small pad of cotton wool. Inoculations with cloned TGMV DNA and other DNAs in combination with virus were carried out in a similar fashion, and used known concentrations of the cloned DNAs (10-15µg).

## 2.10 Extraction of viral DNA

### 2.10.1 Crude DNA preparations of tomato golden mosaic virus

Viral inoculum prepared as described in section 2.9.3. was made 0.1% (w/v) with respect to SDS and extracted once with phenol, once with phenol/Sevag and once with Sevag alone. The DNA was precipitated with 2 vol of ethanol at -20°C overnight and recovered by microcentrifugation (10min). After washing with 70% (v/v) ethanol, the pellets were dried in a vacuum dessicator, resuspended in 1XTAE buffer and stored in aliquots at 4°C or -20°C.

2.10.2 **Origin of wheat dwarf virus**

WDV infected wheat leaves (*Triticum triticalis*) were supplied by Dr.Klas Lindsten, Department of Plant and Forest Protection, Swedish University of Agricultural Sciences, 5-750 07 Uppsala, Sweden. The infected leaves were stored at -20°C prior to processing.

2.10.3 **Crude DNA preparations of wheat dwarf virus**

This procedure was adapted from Ikegami *et al.* (1982).

Infected wheat leaves were ground with a pestle and mortar in 0.1M Tris-Hcl (pH 7.0), 0.1M NaCl, 0.01M EDTA, 1% (w/v) SDS, 1% (v/v) B-mercaptoethanol, 0.325% (w/v) ascorbic acid, 100µg/ml Proteinase K (Sigma; auto digested at 60°C for 1hr). Homogenization buffer was used at a rate of 2ml/g of plant tissue. After incubation at room temperature for 30min, the homogenate was filtered through muslin, extracted once with a 4:1 mixture of phenol/chloroform and once with chloroform alone. The DNA was precipitated with 2 vol. ethanol at -20°C overnight and recovered by centrifugation (10,000rpm, 10min) at 4°C. The pellets were dried in a vacuum dessicator and resuspended in 1XTAE. The DNA solution was then extracted once with phenol, once with phenol/sevag, once with sevag alone and reprecipitated with 2 vol ethanol at -20°C overnight. The DNA was recovered by microfugation (10min), the pellets dried, resuspended as above and stored in aliquots at 4°C or -20°C.

2.11 **Cesium chloride-ethidium bromide gradienting of DNA**

Centrifugation of TGMV or WDV crude DNA extracts in cesium chloride-ethidium bromide gradients was performed as previously (Sunter *et al.* 1984). Solid cesium chloride (1g) was added to each ml of the crude DNA solutions. The contents were mixed until all of the salts had dissolved and 80µl of ethidium bromide (10mg/ml in $dH_2O$) were added for every ml of the final solution, and that the final density of the solution was *ca.* 1.55g/ml and the concentration of ethidium bromide *ca.* 600µg/ml. The solution was centrifuged (32,000 rpm, 65hr at 20°C) in a Beckman type 50 rotor. The tubes

were then punctured at the base and 0.5ml fractions collected using a Gilson microcol TDK 80 fraction collector. Ethidium bromide was removed by exhaustive extraction with 1-butanol saturated with water. Cesium chloride was removed by dialysis against TE buffer (pH 8.0) at 4°C overnight. The DNA was precipitated with 2 vol of ethanol at -20°C overnight and recovered by microcentrifugation (10min).The pellets were dried under vacuum, resuspended in TE (pH 8.0) and the fractions analysed by agarose gel electrophoresis.

## 2.12 Hydroxyl apatite chromatography

Hydroxyl apatite (HAP) chromatography was performed according to the procedures of Colman *et al.* (1978). This technique was used to prepare crude DNA extracts of TGMV infected plants that were enriched in the viral specific dsDNA forms.

The HAP was equilibrated with 8M urea, 0.24M NaPB (pH 6.8). After removing the fines, the HAP was poured into 2ml syringes plugged with siliconized glass wool to give a final column volume of 1ml and washed with 2 vol. of 8M urea, 0.24M NaPB (pH 6.8). Negative pressure was provided throughout the procedure by a peristaltic pump.

The crude DNA samples from TGMV infected plants (110µl containing up to 500µg DNA) were denatured at 95°C for 5min in a paraffin bath and immediately diluted with 890µl of 9M urea, 0.27M NaPB (pH 6.8) containing 0.9% (w/v) SDS. Samples were then applied to the columns and washed with 4ml of 8M urea, 0.24M NaPB (pH 6.8) to elute RNA and ssDNA. Fractions(0.5ml) were collected as before (Sect. 2.8.1.) Columns were washed with 6ml of 0.01M NaPB (pH 6.8) to remove residual urea. dsDNA was eluted with 3ml of 0.3M NaPB (pH 6.8) followed by 3ml of 0.4M NaPB (pH 6.8). The eluant was added in small amounts to allow for equilibration and fractions (0.5ml) were collected as above. 10µl of each fraction or each alternate fraction were spotted onto agarose-ethidium bromide plates (Section 2.7.2) prior to examination under UV light. Those fractions within each wash that contained nucleic acids were pooled and dialysed overnight at 4°C against 0.5XTAE buffer. The dialysed fractions were then mixed with 2 vol ethanol and kept at -20°C overnight. The nucleic

acids were recovered by centrifugation (10,000rpm, 30min), the pellets dried under vacuum, resuspended in TE (pH 8.0) and stored at 4°C.

## 2.13 Agarose gel electrophoresis

Several sizes of horizontal submarine slab gels were used during the investigation. These included: 'full-size' gels (22.5 X 15.5 X 0.4cm); 'half-size' gels (11.0 X 15.5 X 0.4cm); wide 'mini-gels' (6.5 X 10.0 x 0.4cm and 15.0 X 10.0 X 0.4cm); and 'mini-gels' (10.0 X 6.0 X 0.4cm). Electrophoresis tanks utilized were large home-made tanks, a wide 'mini-sub' DNA cell (Biorad) and a 'mini-sub DNA cell' (Biorad). Several comb sizes for casting gel slots were also used which varied in width from 0.2 to 0.8cm.

Agarose was used at several different concentrations ranging from 1 to 1.5% (w/v) Agarose (Sigma; Seakem, type II, low-endo-osmosis) was prepared by solubilizing the required amount in 1XTAE buffer using a magnetic stirrer with a heater facility. Ethidium bromide (Sigma) was then added to a final concentration of 0.5µg/ml from a 10mg/ml stock solution (stored in the dark at 4°C). Gels were allowed to cool to 60°C before adding ethidium bromide. Samples were mixed with 0.1 vol of loading buffer (50% (v/v) glycerol, 10XTAE, 0.4% (v/v) bromophenol blue) prior to loading into gel slots. Sample volumes were usually 10-30µl for full and half size gels and 5-15µl for 'mini-gels'.

Gels were electrophoresed in 1XTAE buffer at 200mA for 2-4hr in the large and wide 'mini-sub' tanks and at 100mA for 30-60min in the 'mini-sub' tanks. In most cases electrophoresis was terminated when the bromophenol blue dye front had migrated *ca.* three-quarters the gel length. Electrophoresis was carried out using either a LKB Biochrom 2103 or a Pharmacia EPS 500/400 power pack. Agarose gels electrophoresed without ethidium bromide were stained with the dye afterwards at a concentration of 0.5µg/ml for 30min and destained with $dH_2O$ for 10min. Bands were visualized on a Chromato-Vue Transilluminator C-63 UV light box. Photographs were taken with an Exa 1 camera fitted with two Y(K2) yellow filters (Hoya) and one UV(0) filter (Hoya). Either Ilford Hp5 or PanF 35mm films were used which were developed in Promicol ultrafine grain developer.

## 2.14 Polyacrylamide gel electrophoresis

Polyacrylamide gels were used to purify TGMV subgenomic ssDNA species and WDV oc dsDNA.

The gels were cast in siliconized glass tubes (8.0 X 0.6cm internal diameter). The gel solution contained 4% (w/v) acrylamide (prepared as a 40% (w/v) stock solution consisting of 39.6% (w/v) acrylamide (Sigma), 0.4% (w/v) N, N'-methylene bisacrylamide (Sigma)), 1XTAE, 0.05% (w/v) TEMED and 0.075% (w/v) ammonium persulphate. The acrylamide was allowed to polymerise under an overlay of $dH_2O$ for at least 1 hr at room temperature. Samples were mixed with 0.1 vol loading buffer (50% (w/v) glycerol, 10XTAE, 0.4% (w/v) bromophenol blue) and layered onto gels in volumes as large as 50µl. Gels were electrophoresed in 1XTAE buffer at 10mA/tube for 2-3 hr using a Shandon tube gel electrophoresis tank with a discontinuous buffer system. Electrophoresis was carried out using either an LKB Biochrom or a Pharmacia EPS 500/400 power pack. Following electrophoresis gels were stained with 0.5µg/ml ethidium bromide for 10-15 min, rinsed in $dH_2O$ and bands visualized and photographed as described in section 2.13.

## 2.15 Methyl mercury gels

Methyl mercury (Me-Hg) gels (Bailey and Davidson, 1976) were used to estimate the sizes of TGMV subgenomic ssDNAs using *Hae*III ØX174 RF DNA restriction fragments as size markers.

Gels were prepared by adding 250µl of 1M Me-Hg (Ventron) to 50ml of a 1% (w/v) agarose solution prepared in 1XBE (40mM sodium borate, 1mM EDTA (pH 8.2); prepared as a 10 X stock solution (pH 7.38)). The agarose was cooled to 50°C before addition of Me-Hg. Samples of DNA (17µl) were mixed with 8µl of loading buffer prior to loading into the gel slots. The loading buffer was prepared as follows:

| | |
|---|---|
| 50µl | 10X BE (400mM borate, 10mM EDTA, (pH 7.38) |
| 50µl | glycerol |
| 45µl | 0.5% (w/v) bromophenol blue |
| 5µl | 1M Me-Hg (Ventron) |

Gels were electropheresed in BE at 80V for 2hr using an LKB Biochrom 2103 power pack. The gels were then stained for 30min in the following solution: 250ml $dH_2O$, 175µl β-mercaptoethanol, 12.5µl 10mg/ml ethidium bromide. The gels were visualized and photographed as previously (section 2.13). All operations concerning Me-Hg were performed in a fume hood.

## 2.16 Recovery of DNA fragments from gels

DNA samples of interest in gels were localized with UV light. Gel slices containing the DNA bands were excised using a sharp sterile blade and the DNA recovered by one of the following methods (see results section for details).

### 2.16.1 Electroelution into dialysis bags

This method was communicated by Dr. R.H.A. Coutts.

A bag of dialysis tubing filled with 1XTAE was fitted onto the tip of a glass column (cut from a 10ml pipette) plugged with siliconized glass wool. The column was carefully filled with 1XTAE to avoid any trapped air bubbles, and the whole placed in a tube gel apparatus. The gel slice(s) containing the band(s) of interest were loaded into the column and the DNA electrophoresed into the dialysis bag at 160V for 3-4hr. After reversed electrophoresis for *ca.* 2min to release any DNA bound to the dialysis membrane, the dialysis bag was removed and the residual buffer transferred to eppendorf tubes. This solution was then pheol extracted and the DNA recovered and subsequently stored in 1XTAE as previously (section 2.6)

### 2.16.2 Electroelution onto DEAE paper

This method was an adaptation of that described by Dretzen *et al* (1981)

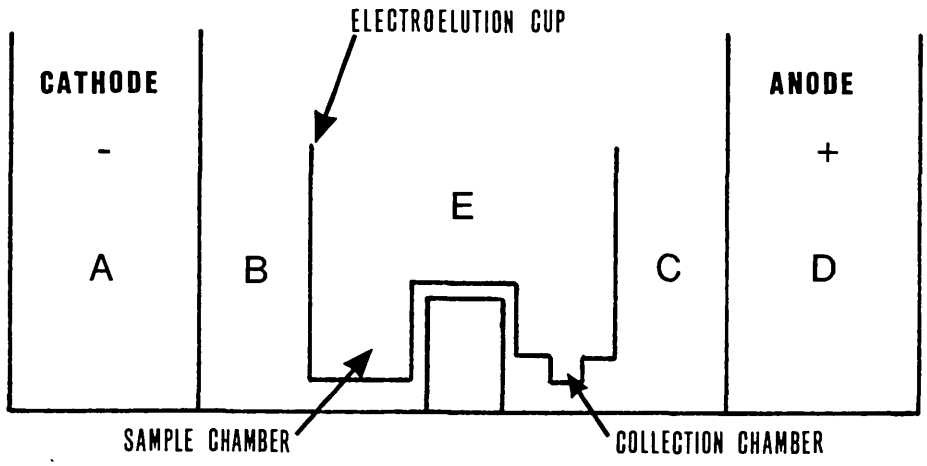
Cylindrical glass tubes were sealed at one end with a disc of Whatman DE 81 DEAE-cellulose paper backed by a dialysis membrane and a rubber ring employed to hold the membrane in place. The DEAE-cellulose discs were soaked in 2.5M Nacl for several hours, washed in several changes of $dH_2O$ and stored in 1mM EDTA at 4°C prior to

use. Tubes were filled with 1XTAE and placed in a tube gel apparatus. The gel slice(s) containing the band(s) of interest were loaded into the tubes and the DNA electroeluted onto the DEAE cellulose discs at 100V for 2-3hr. Discs were then transferred to siliconized glass vials, washed with cold $dH_2O$, drained and placed in Eppendorf tubes containing 0.5ml of 20mM Tris-Hcl (pH 7.5), 1mM EDTA, 1.5M Nacl. Discs were shredded by vortexing and then incubated at 37°C for 2hr with occasional agitation. Mixtures were transferred to small Eppendorf tubes (0.5ml), previously punctured at the base and plugged with siliconized glass wool. Eluates were microfuged through the glass wool into larger Eppendorf tubes (1.5ml) with detached caps. Pooled eluates were extracted and the DNA recovered and subsequently stored in 1XTAE as previously (section 2.6).

2.16.3 **Electroelution with ISCO cups**

This method was performed according to Zassenhaus *et al.* (1982).

The electroelution apparatus is shown diagramatically below :



It consisted of an electroelution cup (ISCO) and an electrophoresis tank with separate cathodic and anodic reservoirs. The cathodic and anodic reservoirs were themselves divided by a semipermeable dialysis membrane thus creating compartments A, B, C and D. Similarly, dialysis membranes held in place by plastic compression rings were present at the bottom of the sample and collection chambers of the electroelution cup. Compartments A, C and D were filled with high ionic strength buffer (3M sodium acetate + 1XTEA (1XTEA= 50mM Tris-acetate (pH 7.8), 0.1mM EDTA)) and compartment B was filled with a lower ionic strength buffer (2XTEA). The gel slice(s) containing the DNA band(s) of interest were placed in the sample chamber in the electroelution cup which in turn was filled with 10ml of very low ionic strength buffer (0.005XTEA). The cups were then placed in the tank and the DNA electroeluted into the collection chamber at 100 Volts for 1 hr. Diffusion of the high ionic strength buffer in compartment C into the collection chamber formed a high-salt barrier which prevented the electroeluting DNA from either sticking to or penetrating the membrane. The electroeluate was recovered by transferring the buffer in the cup (except for the fluid in the collection chamber) to a Sterilin tube with a pasteur pipette. The fluid in the collection chamber was then removed with a Gilson automatic pipetor filled with a disposable tip and transferred to Eppendorf tubes. The buffer was then returned to the cup and electroelution repeated for 1hr more. After removal of the fluid for a second time, the collection chamber was rinsed with 100µl of 0.15M sodium acetate, pooled eluates were extracted and the DNA recovered and subsequently stored in TE (pH 8.0) as previously (section 2.6)

This method worked equally well for agarose and polyacrylamide gels. Yields were usually > 80-90% and the DNA recovered amenable to a number of *in vitro* reactions.

### 2.16.4 Freeze squeeze

This method was a modification of that described by Tautz *et al.* (1983) and was only used for elution of small DNA fragments (<2.0kb) from agarose gels for subcloning into M13 DNA vectors.

The tip of a 1ml disposable syringe was cut off, sealed with a hot scalpel and punctured with a hot needle. The gel slice(s) containing the DNA band(s) of interest were crushed by passage through the syringe and collected in Eppendorf tubes. Three vol/slice of 1M Nacl, 10mM Tris-Hcl (pH 8.0), 1mM EDTA were added and the mixtures incubated at room temperature for 15-30min in the dark. Tubes were then immersed in liquid nitrogen or dry ice, or kept at -70°C until the mixtures were completely frozen. After thawing at room temperature, the mixtures were tranferred to small eppendorf tubes (0.5ml) previously punctured at the bottom and plugged with siliconized glass wool. The tubes were nested into larger Eppendorf tubes (1.5ml) and microfuged. The pooled eluates were extracted and the DNA recovered and subsequently stored in TE (pH 8.0) as previously (section 2.6).

## 2.17 Purification of dsDNA with Elutip-d minicolumns

The electroeluted inserts of PBH 401 and PBH 604 DNA (Bisaro *et al.*, 1982) were purified using Elutip-d minicolumns (Schleicher & Schüll) according to the manufacturers instructions.

The stock solutions used were:

Buffer solution I (0.2M Nacl, 20mM Tris-Hcl (pH 7.4), 1.0mM EDTA)

Buffer solution II (1.0M Nacl, 20mM Tris-Hcl (pH 7.4), 1.0mM EDTA)

After removal of the protective cap and excision of the tip, columns were attached to 2ml syringes loaded with buffer solution II (2ml). This solution was injected into the columns to pre-wash the matrix. Using 10ml syringes, the columns were then injected with buffer solution I (5ml) to remove buffer solution II and prime them. The columns were then fitted to S&S Disposable Filter Holders FB 030/20 "white rim" and the DNA samples (resuspended in buffer solution I) slowly forced through the filters and columns using 2ml syringes. The columns and filters were then washed with buffer solution I (2ml) using the same syringes to remove residual gel contaminants. The columns were removed from the disposable filter holders and attached to 1ml syringes loaded with buffer solution II (800µl). The DNA was eluted into Eppendorf tubes, recovered by ethanol precipitation.and resuspended in buffer as previously (section 2.6)

## 2.18 Nuclease reactions

DNAse I (Sigma, type III) was prepared as a 1mg/ml stock solution and stored at -20°C. Digestions were carried out in 10mM Tris-Hcl (pH 7.4), 10mM $MgCl_2$ (prepared as a 10 X stock solution and stored at -20°C) for 1-2hr at 37°C. Unless otherwise stated, 1µg of enzyme was used per reaction.

RNAse I (Sigma, type XI-A) was prepared as a 1mg/ml stock solution in $dH_2O$, boiled for 10min to remove contaminating DNAse activity and stored at -20°C. Digestions were carried out for 1-2hr at 37°C with no specific buffer requirement. Unless otherwise stated, 1µg of enzyme was used per reaction.

Nuclease S1 (Sigma, type III) was diluted to 9u/µl in storage buffer (25% (v/v) glycerol, 10mM sodium acetate, 5mM $MgSO_4$, 0.5mM $ZnSO_4$, (pH 5.0) and stored at -20°C. Digestions were performed in 10mM sodium acetate, 5mM $MgSO_4$, 0.5mM $ZnSO_4$, (pH 5.0), (prepared as a 10 X stock solution, stored at -20°C) for 1-2hr at 37°C. Unless otherwise stated, 4.5 to 9.0u of enzyme were used per reaction.

Exonuclease III was purchased from BCL at 160µ/µl and stored at -20°C. Digestions were carried out in 20mM Tris-HCl (pH 7.5), 7mM $MgCl_2$, 100mM KCl, 100µg/ml gelatin, 1mM DTT (prepared as a 10 X stock solution, stored at 4°C) for 1-2hr at 37°C. Unless otherwise stated, the enzyme was diluted 1:4 in 1 X reaction buffer and 40 to 80u of enzyme used per reaction.

Before gel electrophoresis of nuclease reaction mixtures, they were usually extracted once with phenol/Sevag, once with Sevag and precipitated with 2 vol of ethanol at -20°C overnight. The nucleic acids were pelleted, dried and resuspended in TE (pH 8.0) as previously (section 2.6)

## 2.19 Restriction endonuclease digestions

Restriction endonucleases were purchased from Amersham, Anglian, Bcl, Biolabs, Pharmacia and P&S Biochemicals. Storage was at -20°C. All restriction digestions were carried out either in 'Breaks-all' buffer (Hamilton *et al.*, 1982) or in TA buffer

(O'Farrel *et al.*, 1980). The latter buffer was preferable for most restriction enzyme digests and was used almost exclusively. The buffers were prepared as 10 X stock solutions (see below), filter sterilized (millipore 0.2μm pore size) and stored in aliquots at 4°C and -20°C.

10 X 'Breaks-all'

200mM Tris-HCl (pH 7.5)
70mM $MgCl_2$
1M KCl
10mM DTT
1mg/ml gelatine

10 X TA buffer

330mM Tris-acetate (pH 7.9)
660mM Potassium acetate
100mM Magnesium acetate
50mM DTT
1mg/ml BSA
500mM NaCl

Reactions typically contained 0.2 to 1μg of DNA in a volume of 20μl with 5 to 10u of enzyme. Enzymes in double digests were added simultaneously except where these required different temperatures of incubation. Reactions were incubated at the temperature recommended by the manufacturers, for periods of time varying between one to several hours, depending on the amount of DNA to be digested and the enzyme used. Gel purified DNA was usually digested overnight in the presence of spermidine (5mM). When large quantities of DNA were to be cleaved, reaction volumes were 'scaled up' so that the final concentration of DNA remained <0.5μg/μl.

If the restricted DNA was to be directly analysed by gel electrophoresis, digests were mixed with the appropriate volume of loading buffer and loaded into gel slots. Otherwise the DNA was extracted and subsequently resuspended in the desired buffer as described previously (section 2.6)

## 2.20 Reverse transcriptase reactions

Complementary DNA to TGMV subgenomic ssDNA labelled with [$^{32}$P] was prepared by a modification (Gould and Symons, 1977) to the method of Taylor *et al.*, (1976) and used Avian Myoblastis virus (AMV) reverse transcriptase (Pharmacia).

The primer for the reaction was prepared by digesting salmon sperm DNA with DNAseI so as to generate a large number of random oligodeoxynucleotide fragments. The reaction was 'set up' as follows:

| | |
|---|---|
| 20µl | 0.1M Tris-HCl (pH 7.4), 0.1M $MgCl_2$ |
| 14µl | 1mg/ml DNAseI (Sigma) |
| 166µl | 6mg/ml salmon sperm DNA (Sigma) |

Reactions were incubated for 2hr at 37°C and then heated to 120°C for 10min in a paraffin bath to denature the DNAse and the primer DNA. The primer was stored at -20°C for later use.

The reverse transcriptase reaction was 'set up' as follows:

| | |
|---|---|
| 5µl | 0.5M Tris-HCl (pH 8.3) and 80mM $MgCl_2$ |
| 5µl | 80mM DTT |
| 2µl | each of 15mM dATP, dGTP, TTP |
| 10µCi | [$\alpha$-$^{32}$P] dCTP (Amersham, 410mCi/mmole) vacuum dried before use |
| 25µl | 5mg/ml DNA primer |
| xµl | TGMV subgenomic ssDNA (up to 1 µg) |

with $dH_2O$ to a final volume of 50µl

The contents were mixed by vortexing and 1µl aliquots spotted onto two discs of Whatmam 3MM filter paper to assess input and background levels of radioactivity. Between 10 to 15u of AMV reverse transcriptase were then added and the reaction incubated at 37°C. After 90min, another 1µl aliquot was spotted to measure the amount of label incorporated into the cDNA and the reaction terminated by addition of 0.5M EDTA (1µl). Labeled DNA was separated from unincorporated dNTPs either by fractionation on a sephadex G-50 column or by 'spun-column' chromatography (section 2.8). The discs

were washed in TCA (except the one used to measure the input levels) and counted in a Scintillation counter (section 2.7.3).

## 2.21 **Nick translation**

Nick translations were performed according to the procedures of Rigby *et al.* (1977).

The reaction was 'set up' as follows:

| | |
|---|---|
| 20µl | 5 X reaction buffer (se below) |
| 5µl | each of 300µm dATP, dGTP, TTP |
| 10-50µci | [$\alpha^{-32}$P] dCTP (Amersham) vacuum dried before use |
| 1µl | 3mg/ml BSA |
| xµl | DNA template (up to 1µg) |
| with $dH_2O$ to a final volume of 100µl | |

The contents were mixed by vortexing and aliquots (2µl) were spotted onto two discs of Whatman 3MM filter paper to assess input and background levels of radioactivity.

DNAseI (2µl of a 200ng/ml solution) and DNA polymerase I (1µl of a 3.5u/µl stock solution; Amersham) were then added and the reaction incubated at 16°C. The DNAseI solution was prepared fresh each time by addition of 1mg/ml DNAseI (2µl) to $dH_2O$ (10ml). After 1-2 hr, another aliquot (2µl) was spotted to measure the amount of label incorporated into the probe and the reaction terminated by addition of 1N NaOH (10µl). The labelled DNA was separated from unincorporated dNTPs either by fractionation on a Sephadex G-50 column or by 'Spun-column' chromatography (section 2.8). The discs were washed in TCA (except the one used to measure the input level) and counted in a scintillation counter as described in section 2.7.3.

The contents of the 5 X reaction buffer were as follows:

| | |
|---|---|
| 500mM | Tris-HCl (pH 7.4) |
| 50mM | $MgCl_2$ |
| 100mM | β-mercaptoethanol |
| 250µg/ml | BSA |

## 2.22 Preparation of ssDNA specific probes

Probes complementary to the TGMV virion sense B DNA component were prepared according to the procedures of Hu *et al.* (1982).

The primer (M13 hybridization primer, Amersham) was diluted to 1µg/ml in $dH_2O$ and stored at -20°C. The primer was boiled for 2min and 'snap-cooled' before use. The annealing reaction was 'set up' as follows:

1µl primer
1µl 10 X *Hin*cII buffer (see below)
1µl 0.1M DTT
1µl ssM13 clone of TGMV B(-) at 1µg/µl concentration
6µl $dH_2O$

The contents were mixed by vortexing and incubated for 5min at 65°C followed by 10min at room temperature. The contents were transferred to eppendorf tubes containing 10µCi [$\alpha$-$^{32}$P] dCTP (Amersham; dried under vacuum before use). 1µl dATG mix (500mM each dATP, dGTP, TTP) and 1µl Klenow fragment of DNA polymerase I (Amersham; 4u/µl) were then added and the reaction incubated at 15°C for 1hr. The reaction was terminated by addition of 85µl TE (pH 8.0) and 0.5M EDTA (5µl). Labelled DNA was separated from unincorporated dNTPs by 'spun-column' chromatography.

The constituents of the 10 X *Hin*cII buffer were as follows:

100mM Tris-HCl (pH 7.9)
600mM NaCl
66mM $MgCl_2$

## 2.23 DNA 'end-filling' reactions

The conversion of DNA fragments with cohesive ends to blunt ends was carried out according to Maniatis *et al.*, (1982).

2.23.1 Conversion of fragments with protruding 5'ends to blunt ends

The reactions were 'set-up' as follows:

2.5µl 10 X nick-translation buffer (see below)
1µl dNTP mix (A solution of all four dNTPs each at 2mM)
xµl restriction fragment (up to 1µg DNA)
with $dH_2O$ to a final volume of 25µl

The Klenow fragment of DNA polymerase I (1µl) (Amersham, 4u/µl) was added and the mixture incubated for 15min at 22°C. The DNA was then extracted, recovered and subsequently resuspended in TE (pH 8.0) as previously (section 2.6.).

The constituents of the 10 X nick-translation buffer were as follows:

0.5M Tris-HCl (pH 7.2)
0.1M $MgSO_4$
10mM DTT
500µg/ml BSA (Sigma)

2.23.2 Conversion of fragments with protruding 3'ends to blunt ends

The reactions were 'set up' as follows:

2µl 10 X T4 polymerase buffer (see below)
1µl dNTP mix (A solution of all four dNTPs each at 2mM)
xµl Restriction fragment (up to 1µg DNA)
with $dH_2O$ to a final volume of 20µl

T4 DNA polymerase (1µl) (Anglian; 5u/µl) was added and the mixture incubated at 37°C for 5min. 1µl of 0.5M EDTA were then added and the DNA extracted and finally resuspended in TE (pH 8.0) as previously (section 2.6)

The constituents of the 10 X T4 polymerase buffer were as follows:

0.33M Tris acetate (pH 7.9)
0.66M Potassium acetate
0.10M Magnesium acetate
50mM DTT
1mg/ml BSA

## 2.24 DNA transfer and hybridization

### 2.24.1 Transfer to nitrocellulose paper

Samples of DNA fractionated on agarose gels were transferred to nitrocellulose paper (Schleicher & Schüll) according to the method of Southern (1975).

After electrophoresis was completed, gels were stained with 0.5µg/ml ethidium bromide if electrophoresis had been performed in the absence of the dye and photographed as previously, but this was optional.

Gels were then prepared for transfer as follows: The DNA was depurinated by soaking the gel in 0.25M HCl (250ml) for 15min at room temperature with occasional shaking. The gel was washed briefly in $dH_2O$ and then soaked twice in 0.5M NaOH containing 1M NaCl (250ml) for 15min at room temperature to denature the DNA. The gel was briefly washed in $dH_2O$ and then neutralized by soaking in 0.5M Tris-HCl (pH 7.4) containing 3M NaCl (250ml) for 10min at room temperature with occasional shaking.

Transfer was then performed as follows. Two pieces of Whatman 3MM paper (longer and wider than the gel) were placed on an elevated glass plate (fitted into a pirex dish) so that the ends formed wicks. The pieces of filter paper were wetted with 20XSSC buffer and the pyrex dish filled with *ca.* 500ml of the same buffer. The gel was placed on top of the filter paper and strips of parafilm placed along each side of the gel. A piece of nitrocellulose paper (cut about 1-2mm larger than the gel in both dimensions) preiviously soaked in 20XSSC for *ca.* 1hr was then placed on top of the gel. Care was taken not to trap air bubbles between the gel and the nitrocellulose paper. A wedge of paper consisting of 6-8 pieces of

dry 3MM filter paper and 2-3" of absorbent paper towels (Kimwipe) cut to the same size as the gel was then placed on top of the membrane. A weight, supported by a plastic plate, was placed on top of the stack and the whole sealed with 'cling-film' to prevent evaporation of buffer. Transfer was performed overnight (16-24hr) at room temperature. After transfer was complete, the nitrocellulose paper was lifted away from the gel, soaked in 2XSSC for 10-20min, dried under a infra-red lamp and baked at 80°C for 2hr.

2.24.2 **Transfer to GeneScreen**

Transfer of DNA fragments fractionated on agarose gels to GeneScreen membranes was carried out according to the manufacturers instructions.

After electrophoresis was complete and photographs taken, the gel was treated for transfer as follows:

The DNA was depurinated by incubating the gel in 0.25N HCl (200ml) for 15min at room temperature with occasional agitation and then denatured by soaking the gel in 0.2N NaOH containing 0.6M NaCl (200ml) for 30min at room temperature with occasional agitation. The gel was then neutralized in 1XSSC at room temperature for 1hr with three changes of buffer (200ml each change). The membrane was soaked in 1XSSC for 15 to 30min prior to transfer, which was performed as described for nitrocellulose paper using 1XSSC. After transfer the membrane was washed in 1XSSC to remove residual agarose, placed with transferred DNA uppermost on a piece of filter paper and dried at room temperature. The DNA was baked onto the membrane at 80°C for 2hr.

2.24.3 **Transfer to GeneScreen Plus**

Transfer of DNA fragments fractionated on agarose gels to GeneScreen Plus membranes was carried out according to the manufacturers instructions.

After electrophoresis was completed and photographs taken (optional), the gel was treated for transfer as follows:

The DNA was depurinated by incubating the gel in 0.25N HCl (250ml) for 15min at room temperature with gentle agitation and then

denatured by soaking the gel in 0.4N NaOH containing 0.6M NaCl (250ml) for 30min at room temperature with gentle agitation. The gel was then neutralized by incubating in 0.5M Tris-HCl (pH 7.5) containing 1.5M NaCl (250ml) for 30min at room temperature with gentle agitation. The membrane was wetted in $dH_2O$ and soaked in 10XSSC for 15min prior to transfer. Transfer was performed as described for nitrocellulose paper using 10XSSC, with the concave side of the membrane (side B) in contact with the gel. After transfer, the membrane was immersed in an excess of 0.4N NaOH for 30-60sec to ensure complete denaturation of the immobilized DNA and then neutralized in an excess of 0.2M Tris-HCl (pH 7.5), 2XSSC. The membrane was placed transferred DNA uppermost on a piece of filter paper and dried at room temperature. There was no necessity to fix DNA onto GeneScreen Plus by baking.

2.24.4 **Hybridization with nitrocellulose paper membranes**

This procedure was performed according to the method of Wahl *et al.*, (1979).

The baked nitrocellulose paper was placed in a heat-sealable bag and 1ml of prehybridization solution (see below) added for each $10cm^2$ of nitrocellulose paper. As much air as possible was removed from the bag before sealing with a plastic bag making apparatus. The sealed bag was then incubated in a shaking water-bath for at least 4hr at 42°C. The bag was then opened and the prehybridization solution removed and replaced with 1ml of hybridization solution (see below) for each $10cm^2$ of nitrocellulose paper. The bag was sealed as before and incubated in a shaking water bath for 16-24hr at 42°C. The nitrocellulose paper was then removed from the bag and washed three times in 2XSSC, 0.1% (w/v) SDS (250ml) for 5min each wash at room temperature and then twice in 0.1XSSC, 0.1% (w/v) SDS (250ml) for 15min each wash at 42°C. The nitrocellulose paper was then dried at room temperature over a sheet of filter paper, wrapped in Saran Wrap and exposed to Kodak X-Omat S film for several hours to several days at room temperature or at -70°C with or without the use of intensifying screens. Autoradiograms were developed using a Kodak X-Omat automatic processor.

The constituents of the various solutions used above were as follows:

a) Prehybridization solution

50% (v/v) Formamide (deionized)
5 X SSC
50mM NaPB
0.5 mg/ml Salmon sperm DNA (sheared and denatured)
1% glycine
5 X Denhardt's solution (see below)

b) Hybridization solution:

50% (v/v) Formamide (deionized)
5 X SSC
20mM NaPB
100µg/ml Salmon sperm DNA (sheared and denatured)
1 X Denhardt's solution (see below)
10% Na dextran Sulfate (Mr. 500,000)
[$\alpha$-$^{32}$P] labelled DNA probe (with the exception of ssDNA specific probes, all probes were denatured at 100°C for 5min and 'snap-cooled' in an ethanol/ice bath prior to hybridization)

c) 50 X Denhardt's solution:

1% (w/v) Ficoll
1% (w/v) Polyvinylpyrrolidone
1% (w/v) BSA.

2.24.5 **Hybridization with GeneScreen membranes**

The procedure was carried out according to the manufacturers instructions.

The membrane was incubated in a sealed plastic bag with prehybridization solution (1ml/10cm$^2$ membrane) (see below) for at least 6 hr at 42°C with constant agitation, 2-3ml of hybridization

solution (see below) were then added to the bag, The bag was sealed and incubated for 16-24hr at 42°C with constant agitation. After hybridization the membrane was removed from the bag and washed twice in 100ml of a solution containing 0.3M Nacl, 0.06M Tris-HCl (pH 8.0), 0.002M EDTA at room temperature for 5min with constant agitation, and twice in 100ml of a solution containing 0.3M NaCl, 0.06M Tris-Hcl (pH 8.0), 0.002M EDTA, 0.5% (w/v) SDS at 60°C for 30min with constant agitation and twice in 100ml of 0.003M Tris base at room temperature for 30min again with constant agitation. The membrane was then dried and exposed to X-ray film as described earlier (section 2.24.4)

The constituents of the various solutions used above were as follows:

a) Prehybridization solution:

50% (w/v) Formamide (deionized)
10 X Denhardt's solution
0.05M Tris-HCl (pH 7.5)
1M NaCl
0.1% sodium pyrophosphate
0.1% (w/v) SDS
100µg/ml Salmon sperm DNA (sheared and denatured)
10% sodium dextran sul te (Mr 500,000)

b) Hybridization solution

50% (v/v) Formamide (deionized)
10 X Denhardt's solution
0.05M Tris-HCl (pH 7.5)
0.1% sodium pyrophosphate
0.1% (w/v) SDS
100µg/ml Salmon sperm DNA (sheared and denatured)
[$\alpha^{-32}$P] labelled DNA probe treated as in (section 2.24.4)

2.24.6 **Hybridization with GeneScreen Plus membranes**

This procedure was carried out as described in Maniatis *et al.*, (1982).

The membrane was incubated in a sealed plastic bag with prehybridization solution (1ml/10cm$^2$ membrane) (see below) for at least 6hr at 42°C with constant agitation. Up to 0.5ml of the probe was then added to the bag (all probes, with the exception of ssDNA specific probes, were denatured and 'snap-cooled' as previously prior to hybridization) The bag was sealed and incubated for 16-24hr at 42°C with constant agitation. The membrane was then removed from the bag and washed three times in 2XSSC containing 1% (w/v) SDS (200ml) for 20min each wash at 70°C with gentle agitation. The membrane was then dried and exposed to X-ray film as described earlier (section 2.24.4).

The constituents of the various solutions used above were as follows:

a) Prehybridization solution

50% (v/v) Formamide (deionized)
5 X SSPE
1 X Denhardt's solution
100µg/ml Salmon sperm DNA (sheared and denatured)
10% (v/v) SDS

b) 20 X SSPE; consisted of:

2.4M NaCl
0.3M NaCitrate
0.26M $NaH_2PO_4H_2O$
0.02M EDTA
SSPE was adjusted to pH 6.5 with NaOH.

2.24.7 Rehybridizations with GeneScreen Plus membranes

To avoid irreversible binding of the probe, these membranes were never allowed to dry completely during the first hybridization and subsequent exposure to X-ray film.

The probe was removed according to the manufacturers instructions as follows. The membrane was incubated in 0.4N NaOH (200ml) at 42°C for 30min with gentle agitation and then in 200ml of 0.1 X SSC containing 0.1% (w/v) SDS and 0.2M Tris-Hcl (pH 7.5) at 42°C for 30mim with gentle agitation. The membrane was blotted dry between paper towels and autoradiographed for an appropriate period of time to determine if all of the initial radioactivity had been removed. If this was not the case, the procedure was repeated and the membrane re-checked. Prehybridization and hybridization of these stripped membranes were then performed as before (section 2.24.6.).

2.25 Colony lift hybridization

This procedure was used to identify colonies harbouring TGMV specific subgenomic DNA cloned into Pembl9 vectors.

Transformed cells producing white colonies on selective agar plates (see section 2.29.5) were transferred with sterile tooth picks to two grided agar plates (bottom agar containing 100µg/ml Ampicillin) so that each individual colony was represented in each of the plates in the same relative location on the grid. The plates were incubated at 37°C overnight to allow the colonies to grow. The plate to be used as a master plate was then stored at 4°C. Colonies in the other plate were lifted with a pall biodyne A nylon membrane (Pall) according to the manufacturers instructions. The membrane was marked in three asymmetric locations with a pencil and then carefully placed onto the surface of the plate in direct contact with bacterial colonies for 10-20sec. The plate was marked in the same locations with a marker pen. The membrane was then transferred onto a piece of filter paper saturated with denaturing solution (2.5M NaCl, 0.5M NaOH) for 60sec and then to another piece of filter paper saturated with a neutralizing solution (3M Na Acetate (pH 5.5)) for a further 60sec. The membrane was then dried onto a piece of filter paper at room temperature and baked at 80°C for 2hr.

Prehybridization and hybridization were then carried out according to the procedure described for GeneScreen membranes (section 2.24.5)

## 2.26 Large scale plasmid isolation

This method is a 'scaled-up' version of that described by Ish-Horowicz and Burke, (1981) and includes a LiCl precipitation step to remove the bulk of contaminating RNA (Bisaro *et al.*, 1982).

A starter culture was prepared by innoculation of 2TY broth (10ml) containing 100µg/ml ampicillin with a single bacterial colony and incubated with shaking overnight at 37°C. 200ml of 2TY broth containing 100µg/ml Ampicillin was then inoculated with the starter culture and incubated at 37°C overnight with shaking. Prior to harvest, the cells were chilled on ice for 20min, and then transferred to 250ml MSE bottles and centrifuged at 4°C (7,000 rpm, 10min). Pelleted cells were resuspended in 10ml of 10mM EDTA, 25mM Tris-HCl (pH 8.0), transferred to 50ml MSE polypropylene tubes and centrifuged at 4°C (7,000rpm, 10min). The resultant pellet was resuspended in 3ml of 50mM glucose, 10mM EDTA, 25mM Tris-HCl (pH 8.0) containing 12mg/ml lysozyme (Sigma) and incubated for 30min on ice. To complete cell lysis, 0.2M NaOH containing 1% (w/v) SDS (6ml) was added to each tube. The tubes were capped and the contents mixed by inversion prior to incubating on ice for 10min. For each tube 4.5ml of 3M Potassium acetate (pH 4.8) was added, the contents mixed by inversion and the tubes incubated on ice for a further 30min. The resultant precipitate was pelleted by centrifugation at 4°C (13,000rpm, 15min). The supernatant was transferred to a fresh 50ml tube, and 30ml of ethanol added and the mixture held at -20°C for 30min. The resultant DNA precipitate was pelleted by centrifugation at 20°C (10,000rpm, 10min), the pellets dried briefly <u>in vacuo</u> and finally resuspended in 3ml of 100mM NaCl, 50mM Tris-HCl (pH 7.8). A filter sterilised solution of 8M LiCl (1ml) was then added to each tube and the mixtures held on ice for 30min. The RNA was pelleted by centrifugation at 4°C (10,000rpm,10min). The supernatant was transferred to a Corex tube (15ml capacity), mixed with 2 vol. ethanol and held at -20°C for 30min. The DNA precipitate was pelleted by centrifugation at 4°C (10,000rpm, 10min). The pellet was washed with 70% ethanol, dried briefly *in vacuo*, resuspended in

0.5ml 50mM Tris-HCl (pH 7.5), 100mM NaCl, 10mM EDTA and transferred to an eppendorf tube. RNAse A (25µl of a 1mg/ml stock) was then added to the tube and the contents incubated at 37°C for 1hr. The solution was extracted once with phenol, once with Sevag and the DNA precipitated by the addition of 1.5vol of isopropanol at -20°C for at least 1hr. Plasmid DNA was recovered by microcentrifugation (5min). The DNA pellet was washed in 70% ethanol, vacuum dried, resuspended in TE (pH 8.0) and stored at -20°C.

## 2.27 Small scale isolation of plasmid DNA

This procedure is a modification of the method of Birnboim and Doly, (1979) as described in Maniatis *et al.*, (1982).

2TY broth (1.5ml) containing 100µg/ml Ampicillin was inoculated with a single bacterial colony and incubated with shaking at 37°C overnight. The culture was then poured into an Eppendorf tube, microfuged (1min) and the medium removed by aspiration, leaving the bacterial pellet as dry as possible. The pellet was resuspended by vortexing in 100µl of an ice-cold solution of 50mM glucose, 10mM EDTA, 25mM Tris-HCl (pH 8.0). A freshly prepared solution of 0.2N NaOH containing 1% (w/v) SDS (200µl) was then added. The contents were mixed by inverting the tube rapidly two or three times and the tube incubated on ice (5min). 150µl of ice-cold solution of 3M potassium acetate (pH 4.8) were then added and the contents mixed by gently vortexing of the tube in an inverted position for 10sec. The tube was incubated on ice for 5min and then microfuged at 4°C (5min). The supernatant was transferred to a fresh eppendorf tube and extracted with an equal volume of Phenol/Sevag. After microcentrifugation (5min), the supernatant was transferred to a fresh eppendorf tube and mixed with 2vol of ethanol at room temperature. After standing for 2min at room temperature the tube was microfuged at room temperature (5min). The pellet was washed with 1ml of 70% ethanol (1ml), dried briefly *in vacuo* and resuspended in 50µl of TE (pH 8.0) containing 20µg/ml RNAse A. The DNA was then characterized by gel electrophoresis or restriction endonuclease digestion followed by gel electrophoresis. Storage was at 4°C.

## 2.28 Electron microscopy

Heteroduplex analysis and DNA spreading of TGMV subgenomic ssDNA were performed according to the procedures of Davies *et al.*, (1971).

Heteroduplexes were formed by the alkali denaturation method and reactions were 'set up' as follows:

| | | |
|---|---|---|
| 12.5µl | denaturation solution (0.2N NaOH, 40mM EDTA) | |
| xµl | TGMV subgenomic ssDNA | equimolar amounts of |
| yµl | Purified viral DNA inserts of either pBH401 or pBH407 recombinant plasmids | each DNA to achieve a final concentration of *ca.* 4-5µg/ml |

with Tri-$dH_2O$ to a final volume of25µl

Denaturation was allowed to proceed for 10min at room temperature and then 5min at 50°C. The solution was neutralized by addition of 5µl neutralization buffer (2M Tris-HCl (pH 7.2). Recrystallized formamide (30µl) was then added and renaturation allowed to proceed for 2hr at 3°C. The samples were then spread by the formamide/cytochrome C technique described below.

The hypophase and hyperphase solutions were prepared immediately prior to use and were used within 15-30min. The hypophase solution contained 17% (v/v) formamide, 10mM Tris-HCl (pH 8.7), 0.1mM EDTA prepared in $ddH_2O$.

The hyperphase solution contained:

45µl Formamide
10µl 1M Tris-HCl (pH 8.5)
4µl 2mg/ml Cytochrome C (Sigma, type V)
xµl DNA sample (final concentration: 0.1-0.5µg/ml)
with tri-$dH_2O$ to a final volume of 100µl.

Formamide used in the hyperphase solution had distilled in a rotary flash evaporator under vacuum and recrystallized prior to use. Cytochrome C was purified of contaminants by incubating a stock solution (20mg/ml in $ddH_2O$) overnight at 4°C and microcentrifugation

(5min). The solution free of pelleted contaminants was millipore filtered and diluted in dd$H_2O$ to a 2mg/ml concentration.

Spreadings were performed in a Teflon trough (9cm in diameter) using a glass microscope slide (cleaned in chromic acid) as a ramp. The trough was filled with the hypophase to give a convex meniscus and the microscope slide was supported by a teflon bar to achieve an angle of 30-45°C across the trough. After having cleaned the hypophase by aspiration with tissue (Kleenex), a small amount of fine talcum powder was sprinkled over the hypophase with a camel hair brush and a small volume (25µl) of the hypophase solution was slowly (over 15-40sec) delivered to the slide (about 0.3-1cm above the meniscus) with a drawn out capillary tube fitted with a mouth suction adaptor. Care was taken to mantain a continuous flow of liquid between the delivery tube and the hypophase. Formation of the protein monolayer was visualized by a movement of the talcum powder grains away from the slides towards the edges of the trough. The monolayer was picked up with a Collodium coated EM grid 60-75 sec after spreading. Only one grid was taken per spreading. After each spread the ramp was removed, rinsed with 90% ethanol and air dried before replacement. The surface of the hypophase was cleaned by aspiration with medical tissue and topped up with more solution if required.

Grids were stained by immersion in freshly prepared 90% (v/v) ethanol solution containing 50µM uranyl acetate (from a 50mM stock solution in 50mM HCl, stored at 4°C) for 30sec, rinsed in 90% (v/v) ethanol for 10sec and dried under a lamp. The grids were then shadowed with a platinum:palladium alloy (80:20) by vacuum evaporation in an Edwards E306 shadower. Between 14 and 18 turns of the alloy, wound on normal gauge tungsten wire, were used to shadow the grids. Shadowing was performed at 8° for 5-8sec from one direction followed by rotary shadowing until the filament was exausted. The grids were examined in a Siemens 1A electron microscope operated at 50KV. Molecules were photographed on Kodak 4489 sheet film. Contour lengths were measured with a Hewlet-packard 9821A calculator coupled to a 9864A digitizer. ØX174 ssDNA was used as a ssDNA size marker. Magnifications were calibrated with a replica plate grid (Ladd, 28,600 lines/in.).

## 2.29 Cloning

### 2.29.1 Origin and storage of host strain bacteria and vector DNA

pEMBL9 vector DNA (Dente *et al.*, 1983) was a gift from Dr. P. Wise. M13mp8 and mp9 vector DNA (Messing and Vieira, 1982) were purchased from Biolabs. JM101 cells (Δ*lac pro, thi, sup* E, F' *tra* D36, *pro* AB, *lac* $I^q$ ΔM15) were a gift from Dr. W.D.O. Hamilton. All vector DNAs were stored at -20°C. JM101 cells were kept in sealed minimal agar plates (section 2.29.2) at 4°C and re-streaked every 1-2 months.

### 2.29.2 Media and Antibiotics

The constituents of the normal media additives and antibiotics are listed below:

a) Minimal agar

500ml 3% (w/v) Bacto-Agar (Difco)
500ml M9 salts
1ml 0.1M $CaCl_2$
1ml 1M $MgSO_4$
0.5ml 1% (w/v) Vitamin B1
1ml 20% (w/v) glucose

M9 salts consisted of:

6g $Na_2HPO_4$
3g $KH_2PO_4$
0.5g NaCl
1g $NH_4Cl$
$dH_2O$ to 1 litre.
All components were autoclaved except vitamin B1 which was filter sterilized.

b) 2TY broth

16g Bacto-tryptone (Difco)
10g Bacto yeast extract (Difco)
5g NaCl
$dH_2O$ to 1 litre

c) Bottom agar

2TY containing 1.5% (w/v) Bacto-agar (Difco)

d) H top agar (Miller, 1972)

10g Bacto-tryptone
8g NaCl
8g Bacto-agar
$dH_2O$ to 1 litre

All broth and agar was autoclaved for 15min at 15 $lb/in^2$

X-gal (5-bromo-4-chloro-3-indoyl-beta-galactosidase; Anglian) was prepared as a 20mg/ml stock solution in dimethyl formamide and stored at 4°C.

IPTG (isopropyl-beta-D-thiogalactopyranoside; Sigma) was prepared as a 24mg/ml stock solution in $dH_2O$ and stored at 4°C.

Ampicillin (Sigma, Sodium salt) was prepared as a 25mg/ml stock solution in $dH_2O$ and stored at 4°C for up to 2 weeks. Ampicillin was normally used at a working concentration of 100µg/ml.

Selective agar plates for pEMBL cloning were prepared by spreading 30µl of 20mg/ml X-Gal and 20µl of 24mg/ml IPTG on the surface of bottom agar plates containing 100µg/ml Ampicillin.

2.29.3 **Ligations**

pEMBL9 DNA or either of the M13mp phage DNAs were digested with the appropriate restriction enzyme for 1-2hr as described (section 2.19.). Restricted vector DNA was extracted once with phenol/sevag, once with sevag and precipitated with 2vol of ethanol at-20°C overnight. The DNA was recovered by microcentrifugation

(10min), dried *in vacuo* and resuspended in TE (pH 8.0) at a concentration of 10ng/µl and stored at -20°C or 4°C.

DNA for cloning was prepared as described later (see results).

About 50ng of pEMBL or 20ng of M13 vector DNA were used per reaction with a 2-3 fold molar excess of the DNA to be cloned. All ligations were performed in a 15µl vol of 66mM Tris-Hcl (pH 7.6), 10mM $MgCl_2$, 15mM DTT, 1mM spermidine, 1mM ATP at 6°C overnight or at room temperature for 5-6hr. For 'blunt-end' ligations 1µl of undiluted T4 ligase (10u/ul; Anglian) was used, while for cohesive-end ligations the enzyme was diluted 1:5 in 20mM NaPB, 60mM KCl, 5mM DTT, 1mM EDTA, 50% glycerol (pH 7.6) and 1µl (*ca.* 2u) used. In most instances, the vector DNA and the DNA to be cloned were phenol extracted and ethanol precipitated in the same tube and resuspended directly in ligation buffer.

### 2.29.4 Preparation of competent JM101 cells

10ml of 2TY broth were inoculated with a single colony of JM101 picked from a minimal agar plate, and incubated at 37°C overnight with shaking.

A Sterilin tube containing 10ml of 2TY broth was then inoculated with 10µl of the overnight culture and incubated at 37°C for 4hr with shaking. The cells were cooled on ice for 10min, centrifuged at 4°C (4,000g, 5min) and resuspended in 5ml of 50mM $CaCl_2$. The cells were then kept on ice for 20min, recentrifuged as above and resuspended in 1ml of 50mM $CaCl_2$. The cells were kept at 4°C for 12-24hr before use.

### 2.29.5 Transformation

#### 2.29.5.1 Transformation with pEMBL vectors

The ligation mix was added to 200µl of JM101 competent cells. The cells were incubated on ice for (40min) and then 'heat-shocked' at 42°C for 3min. 2TY broth (800µl) was then added and the cells incubated at 37°C for 15min. The cells were then spread onto selective agar plates (100µl of cells/plate). The plates were allowed to dry briefly, inverted and incubated at 37°C overnight.

Cells giving rise to white colonies were assumed to carry recombinant plasmid DNA and were picked up for further characterization. Blue colonies were assumed to harbour 'wild-type' pEMBL plasmid only.

2.29.5.2 **Transformation with M13 vectors**

A portion of the ligation mix (≈3.75µl) was added to 200µl of competent JM101 cells. The cells were incubated on ice for 40min and then 'heat-shocked' as above. These cells were then added to 3ml H-top agar at 42°C containing 30µl of 20mg/ml X-Gal, 20µl of 24mg/ml IPTG and 50µl of an overnight culture of JM101 cells. The H-top agar was gently vortexed and layered onto previously poured bottom agar plates. The plates were allowed to dry for 10-20min in a Lamina Flow before being inverted and incubated at 37°C overnight. Plaques carrying recombinant phage were selected on the basis of colour as described for pEMBL vectors.

12.30 **Preparation of template DNA for DNA sequencing**

Template DNA was prepared according to Sanger *et al.*, (1980).

A Sterilin tube containing 1ml of 2TY broth and 100µl of an overnight culture of JM101 cells was inoculated with a single plaque (or 5µl of frozen inoculum) and incubated for 5-6hr at 37°C with shaking. The contents were transferred to an Eppendorf tube and microfuged (2min). 800µl of supernatant was transferred to a fresh eppendorf tube and phage precipitated by the addition of 200µl of 2.5M NaCl, 20% (w/v) PEG 6000 for 15min at room temperature. The remaining supernatant was transferred to another Eppendorf tube and stored at -70°C to be used as a source of inoculum for the preparation of further template DNA. The precipitated phages were pelleted by microcentrifugation (5min). The supernatant was discarded and the tube recentrifuged for a few seconds. The remaining supernatant was then removed with a Gilson automatic pipettor fitted with a disposable tip. The pellet was resuspended in 100µl of 10mM Tris-HCl (pH 7.5), 0.1mM EDTA and extracted with an equal volume of phenol. The aqueous phase was then transferred to a fresh Eppendorf tube and template DNA precipitated by the addition

of 0.1vol 2.5M sodium acetate (pH 5.2) and 2vols ethanol at -20°C overnight. The DNA was pelleted by microcentrifugation (10min) dried *in vacuo*, resuspended in 50µl of TE (pH 8.0) and stored at -20°C.

## 2.31 Nucleotide sequencing

### 2.31.1 Stock solutions used for sequencing are listed below:

a) 10 X TBE (pH 8.9)

| | |
|---|---|
| Tris | 162g |
| Sodium borate | 27.5g |
| EDTA | 9.5g |
| $dH_2O$ to 1 litre | |

b) 6% (w/v) acrylamide/7M urea

| | |
|---|---|
| Urea | 288g |
| Acrylamide (Sigma) | 34.2g |
| N; N'-methylenebisacrylamide (Sigma) | 1.8g |
| 10 X TBE | 60ml |
| $H_2O$ to 600ml | |

The acrylamide solution was filtered through Whatman nº1 filter paper and stored at 4°C in 35ml aliquots.

c) Deoxynucleotide triphosphates (dNTP) mixes

dNTP mixes were prepared according to the table below:

| | T-mix | C-mix | G-mix | A-mix |
|---|---|---|---|---|
| 0.5mM dTTP | 1µl | 20 | 20 | 20 |
| 0.5mM dCTP | 20 | 1 | 20 | 20 |
| 0.5mM dGTP | 20 | 20 | 1 | 20 |
| 50mM Tris-HCl pH 8.0, 1mM EDTA | 5 | 5 | 5 | 5 |

dNTPs were prepared as 25mM stock solutions in $dH_2O$. 0.5mM dNTPs solutions were prepared by diluting the stocks in $dH_2O$. All dNTPs were stored at -20°C.

d) Di-deoxynucleotide triphosphates (ddNTPs)

ddNTPs were made up as 10mM stock solutions in 50mM Tris-HCl (pH 8.0), 1mM EDTA. Working stength solutions, given below, were prepared by $dH_2O$ dilution of the stocks.

ddATP → 0.04mM
ddCTP → 0.175mM
ddGTP → 0.30mM
ddTTP → 0.375mM

All stocks and working strength solutions were stored at -20°C.

e) Termination mixes

Termination mixes were prepared by adding equal volumes of the appropriate dNTP mixes and ddNTP working strenth solutions. (eg. T termination mix = 1:1 volume of T-mix and ddTTP). Termination mixes were stored at -20°C.

f) Formamide dye mix

0.03% (w/v) xylene cyanol FF (Sigma)
0.03% (w/v) bromophenol blue (Sigma)
20mM EDTA
This solution was prepared in deionised formamide.

g) 10 X annealing buffer

100mM Tris-H l (pH 8.5)
100mM $MgCl_2$

h) Primer

The primer used during sequencing was the M13 specific synthetic 15-mer (5'-TCCCAGTCACGACGT-3') produced by New England Biolabs. It was diluted to 1ng/ml and stored in small aliquots at -20°C.

i) 'Sticky' Silane solution

| | |
|---|---|
| 25ml | Absolute ethanol |
| 750µl | 10% Acetic acid |
| 75µl | γ-methacryl oxypropyl-trimethoxy silane (Wacker chemicals) |

j) [α-$^{35}$S] dATP (>600 Ci/mmole) was purchased from Amersham as an aqueous solution containing approximately 1mCi/120µl. The solution was aliquoted in 20µl volumes and stored at -20°C.

2.31.2 **Sequencing Reactions**

Sequencing reactions were performed as described by Sanger *et al.*, (1977) and modified by Sanger *et al.*, (1980;1982).
Annealing of the primer to the DNA template was 'set up' as follows:

| | |
|---|---|
| 5µl | DNA template |
| 1µl | 10 X annealing buffer |
| 1µl | Primer |
| 3µl | $dH_2O$ |

The contents were mixed, incubated at 65°C for 5min and cooled to room temperature for 30min.
The polymerization reaction was carried out as follows:
1µl $^{35}$S dATP and 1µl Klenow fragment of DNA polymerase I (Amersham; diluted to 1u/µl before use in 50mM $KPO_4$ (pH 7.0), 0.25mM DTT, 50% (v/v) glycerol) were added to the primer/DNA template mixture. The contents were mixed by vortexing and aliquoted into four separate eppendorf tubes (2.5µl per tube) labelled A, T, C, and G respectively. 2µl of the respective termination mix were then added

to the sides of each of the four tubes. The reactions were simultaneously started by a quick spin in the microfuge and allowed to proceed for 20min at room temperature. 2µl of 0.5mM dATP "chase" were then added to each of the tubes as described above and incubation at room temperature was carried out for a further 15min. The reactions were terminated by simultaneous addition of 4µl of formamide dye mix to each of the tubes. The mixtures were boiled at 100°C for 5min to denature the DNA strands and 2µl of each mixture was then immediately loaded onto gels.

2.31.3 **T-tracking**

The following reaction mixes and volumes were used to analyze 20 clones at one time.

Annealing of the primer to the DNA templates was 'set up' as follows:

| | |
|---|---|
| 8µl | Primer |
| 12µl | 10 X annealing buffer |
| 24µl | $dH_2O$ |

The contents were mixed and 2µl added to each of 20 eppendorf tubes containing 2µl of DNA template. The tubes were incubated at 65°C for 5min and cooled down to room temperature for 30min. The polymerization reaction was 'set up' as follows:

| | |
|---|---|
| 32µl | T-termination mix |
| 4µl | $dH_2O$ |
| 1µl | Klenow fragment (4µ/µl) |
| 6µl | $^{35}S$ dATP |

The contents were mixed and 2µl added to the sides of each of the 20 tubes containing the primer/DNA template mixes. The reactions were simultaneously started by a quick spin in a microfuge and allowed to proceed for 20min at room temperature. 2µl of 0.5mM dATP chase were then added to the tubes as described above and incubation at room temperature carried out for a further 15min. The reactions were

terminated and prepared for loading as described for the sequencing reaction using 4µl of each reaction mix for loading onto gels.

2.31.4 **Sequencing gel electrophoresis**

The reaction products were electrophoresed on 6% (w/v) acrylamide/7M UREA gels as described by Sanger and Coulson (1978) with some modifications (Garoff and Ansorge, 1981).

Glass plates measuring 200 X 350 X 4mm were used to cast the gels. The spacers and combs were cut from 0.35mm thick plasticard (Raven). Plates were washed thoroughly in Decon 90 followed by ethanol and then acetone. The notched plates were treated with ≈20ml of a 2% solution of dimethyl dichlorosilane in 1,1,1-trichloroethane (BDH, Repelcote) and the un-notched plates with 25ml of 'sticky-silane'. After rinsing once more with ethanol, the plates and spacers were clamped together and the bottom and side edges sealed with electrophoresis tape (Universal Scientific). 210µl of a 10% (w/v) ammonium persulphate solution and 28µl of TEMED were added to 35ml of acrylamide solution prior to pouring the gel. After pouring the gel and inserting the comb, the plates were wrapped in cling film and the gel allowed to set for at least 2hr (usually overnight) before electrophoresis. Samples were loaded using a drawn out capillary attached to a home-made screw-advance micro-pipettor. Gels were run in 1 X TBE buffer at 40W (with the voltage and current limited at 1600V and 35mA respectively) using a Raven Sequencing gel apparatus and an LKB Biochrom 2103 power pack. Usually, each sequencing reaction was run in two separate gels, one for a 'short-run', the other for a 'long-run'. 'Short-run' gels (and T-tracking gels) were stopped immediately after the bromophenol blue dye front had migrated off the bottom of the gel (1.5-2.0hr electrophoresis). 'Long-run' gels were stopped 2hr after the termination of 'short-run' gels. After electrophoresis was completed, gels were allowed to cool, removed from the gel apparatus and the plates separated. The un-notched plates with the gel still attached were soaked in 2l 10% acetic acid for 20-30min to fix the DNA. They were then removed, blotted dry with Kim wipe and placed in an 80°C oven for at least 2hr. Gels were then exposed to Kodak X-Omat S X-ray film (18 X 43cm) for 48hr and developed using the Kodak X-Omat automatic processor.

2.31.5 **Computer analysis**

DNA sequences obtained as well as any other DNA and RNA sequences used for comparison purposes were stored on the ERC/VMX VAX computer of the University of Edinburgh connected to Imperial College via the JANET system. Sequences were analysed using the Program Library of the University of Wisconsin Genetics Computer Group (Devereaux *et al.*, 1984). Individual programmes when they were used for analysis are described in the results section.

## 3. Results and discussion

### 3.1 Intracellular TGMV DNA species

The major viral-specific DNA species present in DNA extracts obtained from TGMV-infected *N. benthamiana* plants (TGMV crude DNA), as well as their sensitivities to DNAse I, RNAse A, Exonuclease III and S1 nuclease, are shown in Fig. 9A. The viral DNA species include unit-length oc dsDNA, unit-length linear dsDNA, unit-length ccc dsDNA, unit-length ssDNA and subgenomic ssDNA (Hamilton *et al.*, 1982). The DNA in the gel shown in Fig.9A was transferred to a hybridization membrane (GeneScreen plus) and probed with a mixture of clones pBH401 (TGMV 1 cloned into pAT153 at the unique *Eco*RI site) and pBH604 (TGMV 2 cloned into pAT153 at the unique *Cla*I site) (Hamilton *et al.*, 1983) labelled with [$^{32}$P] by the nick-translation procedure (Rigby *et al.*, 1977) to show additional intracellular viral DNA species (Fig.9B). Besides the previously described dsDNA species (Hamilton *et al.*, 1982), three discrete fast-migrating bands can be visualized in the autoradiograph (bands a, b and c in lane 5). These were shown to correspond to dsDNA forms of the TGMV subgenomic DNA species (Section 3.6). The results of the endonuclease reactions have been previously discussed by Hamilton *et al.* (1982) and only those results concerning the subgenomic ssDNA species will be discussed here.

Hamilton *et al.* (1982) demonstrated the single-stranded nature of the predominant form of TGMV subgenomic DNA on the basis of its strong sensitivity to S1 nuclease (an enzyme that degrades single-stranded nucleic acids (Vogt, 1973)) and DNAse I and its resistance to Exonuclease III (an enzyme which degrades dsDNA with 3'-OH termini (Weiss, 1976)) and RNAse A. The same reactions were performed under essentially the same conditions as employed by Hamilton *et al.* (1982) and the same results were obtained, thereby confirming the previous findings (Figs.9A, 9B). Although S1 nuclease preferentially degrades ssDNA, it can also cleave ccc dsDNA, as has been demonstrated for TGMV ccc dsDNA (Sunter *et al.*, 1984). In addition, dsDNA may be completely digested by S1 nuclease if the enzyme is present at very high concentrations (Maniatis *et al.*,

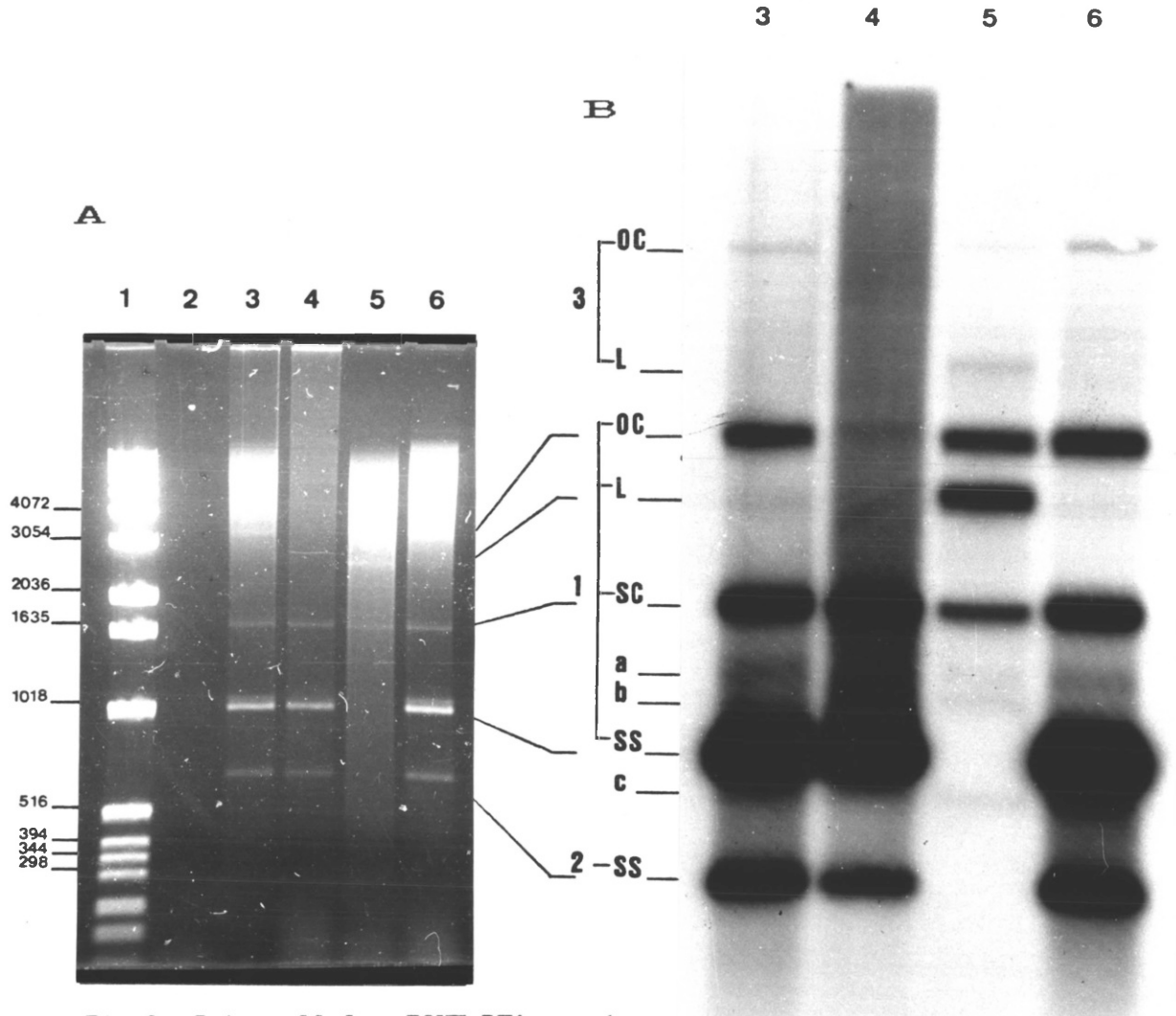



Fig.9. **Intracellular TGMV DNA species.**

A) Electrophoresis in a 1.2% agarose gel containing 0.5µg/ml ethidium bromide.
(1) Kb marker (fragment sizes are indicated in base pairs); (6) TGMV crude DNA (see section 3.1 for definition); (2), (3), (4) and (5), TGMV crude DNA digested with DNAse I, RNAse A, Exonuclease III and S1 nuclease, respectively.

B) Autoradiogram of a blot of the gel shown in Fig.9A probed with a mixture of nick-translated TGMV 1 (pBH401) and 2 (pBH604) DNAs. Only lanes 3 to 6 (as above) are shown.

oc=open circular dsDNA; L=linear dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1, 2 and 3 refer to genomic length, subgenomic and dimeric (M. Slomka; personal communication) viral DNA, respectively. Bands a, b, and c are previously uncharacterized viral DNA species.

1982). However, the fact that the TGMV subgenomic species were completely digested under conditions that had little effect on the viral dsDNA species (this consisted primarily of the linearization of circular dsDNA molecules) leaves no doubt as to its single-stranded nature.

## 3.2 Isolation of TGMV subgenomic ssDNA

The ssDNA form of TGMV subgenomic DNA was isolated by gel elution techniques. TGMV crude DNA, prepared as described in section 2.10.1, was used as source material. Fractionation of the DNA samples was performed by agarose gel (usually 1% agarose containing 0.5µg/ml ethidium bromide) electrophoresis. The bands corresponding to TGMV subgenomic ssDNA were visualized by UV fluorescence (several bands were processed at one time to increase yields) and gel slices containing the bands were rapidly excised to minimize UV damage to the DNA molecules.

Several elution procedures were attempted at different times (see section 2.16 for details). The yields of recovered material varied considerably with each technique. However, the major and more persistent problem associated with the majority of the techniques employed was the presence of agarose contaminants (probably sulphate residues) in the eluate, which considerably inhibited subsequent enzymatic reactions, particularly reverse transcription. Repeated cycles of phenol extraction and ethanol precipitation achieved little towards improving the purity of the recovered DNA. The only method that proved satisfactory was the electroelution procedure of Zassenhaus *et al.* (1982) (section 2.16.3). The DNA was efficiently recovered (usually >80% recovery) and, for practical purposes, free of agarose contaminants. For these reasons, this method was used throughout most of the work described.

## 3.3 Size estimate of TGMV subgenomic ssDNA

### 3.3.1 Size estimate by agarose gel electrophoresis

The size of TGMV subgenomic DNA was estimated from its mobility in agarose gels relative to those of DNA fragments of known

size. Size estimates were obtained both from denaturing (1% agarose containing 0.5 mM methyl-mercury) (Fig.10) and nondenaturing gels (1% agarose containing 0.5µg/ml ethidium bromide) (Fig.11). The former contained phage ØX174 RF *Hae*III fragments as internal markers and the latter the 1Kb ladder and/or phage λ *Cla*I fragments. Sizes were estimated by the local 2 X 3 point method of Southern (1979) using the following formula:.

$$L = K_1/(M-M_0)+K_2$$

where:

$$M_0 = \frac{M_3-M_1((L_1-L_2)/(L_2-L_3)X(M_3-M_2)/(M_2-M_1))}{1-((L_1-L_2)/(L_2-L_3)X(M_3-M_2)/(M_2-M_1))}$$

$$K_1 = \frac{L_1-L_2}{1/(M_1-M_0)-1/(M_2-M_0)}$$

$$K_2 = L_1-K_1/(M_1-M_0)$$

$M_0$ is an empirical correction factor determined by imposing the condition that the three lines joining the three points (L1, L2 and L3) all have the same slope. $L_1$, $L_2$ and $L_3$ are the three molecular weight (or nucleotide length) standards with mobilities $M_1$, $M_2$ and $M_3$ which most closely span the unknown molecular weight (or nucleotide length) L of mobility M. The estimation is performed twice for each value. The first calculation uses two markers that are larger, and one smaller, than the unknown; the second uses two markers that are smaller, and one larger. The final result is the average of these two values.

A size estimate of 1246nt (averaged from three independent measurements) was obtained for TGMV subgenomic ssDNA electrophoresed in methyl-mercury gels (section 2.15). A significantly higher value of 1335nt (averaged from five independent measurements) was obtained for TGMV subgenomic ssDNA electrophoresed in non-denaturing gels. As secondary structures in the ssDNA that might influence its mobility in gels are disrupted under denaturing conditions, it is thought that size estimates obtained from methyl-mercury gels more closely reflect the real value than those obtained under nondenaturing

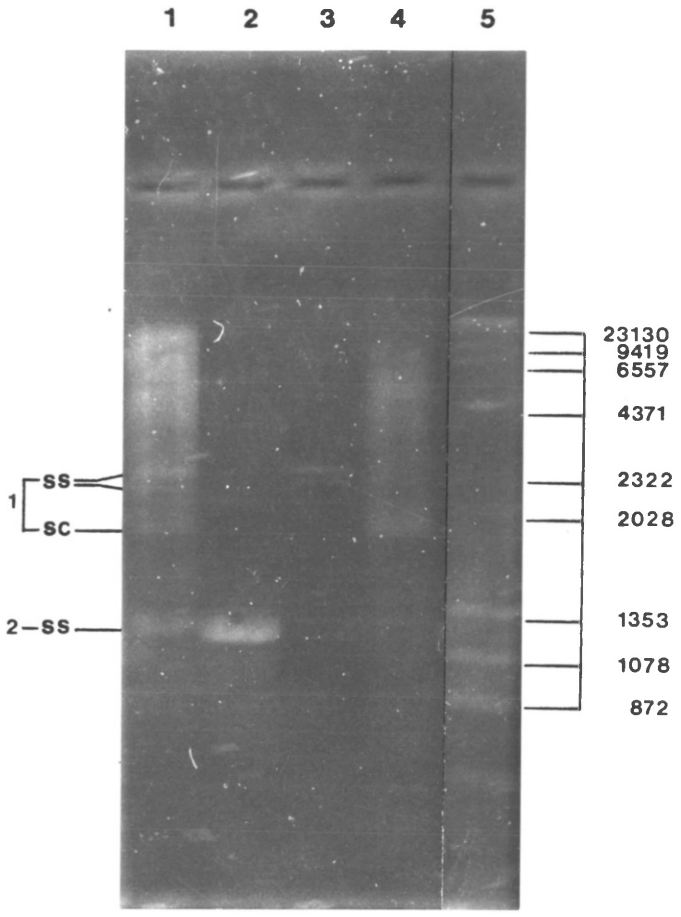



Fig.10. **Size estimation of TGMV subgenomic ssDNA.**

Electrophoresis in a methyl-mercury agarose gel (1% agarose). (1) TGMV crude DNA; (2) gel purified TGMV subgenomic ssDNA; (3) gel purified TGMV genomic-length ssDNA; (4) gel purified TGMV genomic-length supercoiled DNA; (5) ØX174 RF dsDNA digested with restriction enzyme *Hae*III (fragment sizes are indicated in base pairs).

sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic-length and subgenomic viral DNA respectively.

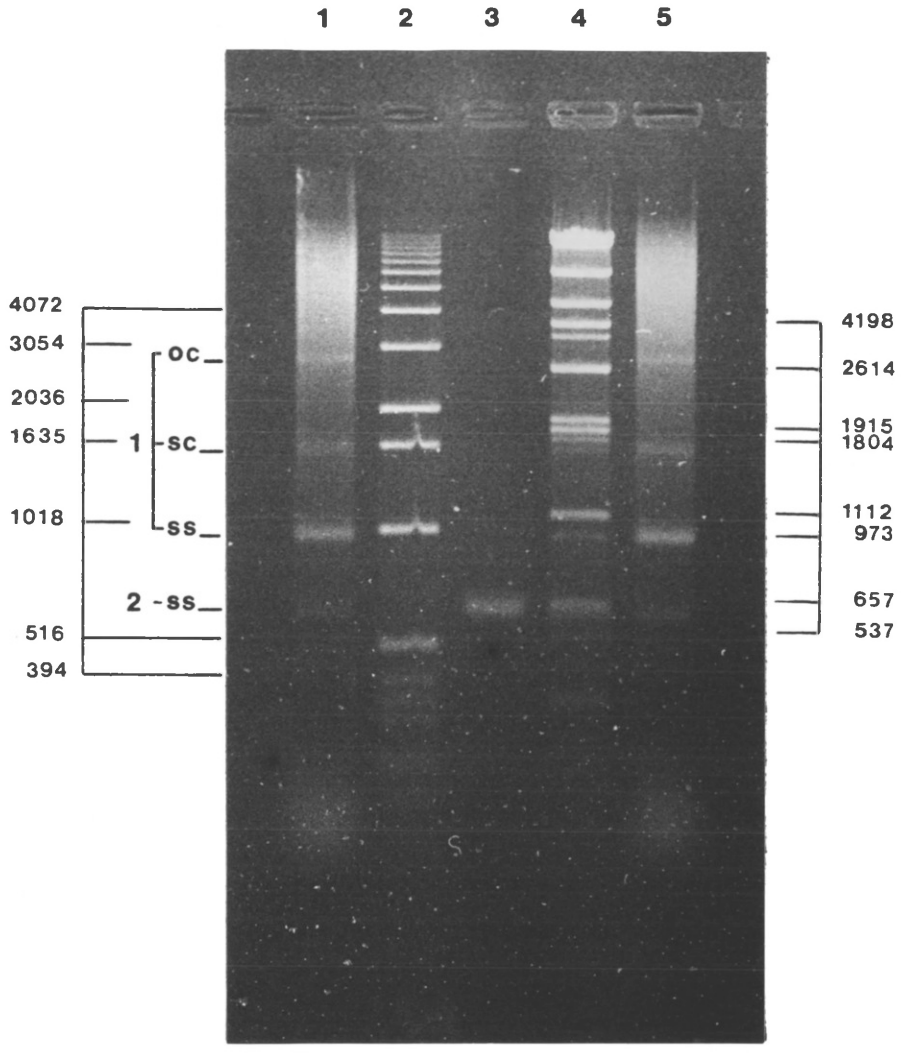



Fig.11. **Size estimation of TGMV subgenomic ssDNA.**

Electrophoresis in a 1.2% agarose gel containing 0.5µg/ml ethidium bromide.
(1) TGMV crude DNA; (2) Kb marker (fragment sizes are indicated in base pairs); (3) gel purified TGMV subgenomic ssDNA; (4) λ-phage DNA digested with restriction enzyme *Cla*I (fragment sizes are indicated in base pairs); (5) TGMV crude DNA.

oc=open circular dsDNA; L=linear dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic length and subgenomic viral DNA respectively.

conditions. That this is indeed the case is demonstrated by the size estimates obtained for unit-length viral ssDNA under each condition. In methyl-mercury gels, viral ssDNA was resolved as two bands (Fig.10, lane 2) with size estimates of 2499 and 2252nt respectively (averaged from two independent measurements). Conversely, a size estimate of 2008nt (averaged from five independent measurements) was obtained for the single viral ssDNA band in nondenaturing gels. While the identity of the two viral ssDNA bands observed in methyl-mercury gels has not yet been determined (i.e. whether they correspond to components I and 2 or to the linear and circular forms of viral ssDNA), their estimated sizes more closely match the real values (2558 and 2508nt for components I and 2 respectively (Hamilton *et al.*, 1984)) than do those obtained under nondenaturing conditions.

In contrast, Hamilton *et al.* (1982, 1983) reported a size estimate of *ca.* 1900nt for subgenomic ssDNA. This value was an extrapolation from the mobilities of ØX174 ssDNA and TGMV genomic ssDNA (of which the size of the latter had been determined by electron microscopy and by electrophoresis of glyoxalated DNA and glyoxalated DNA standards; Hamilton *et al.*, 1981) in non-denaturing gels, and therefore at most very approximate.

### 3.3.2 Size estimate by electron microscopy

TGMV subgenomic ssDNA was eluted from agarose gels by the 'dialysis bag' method (section 2.16.1) and examined by electron microscopy. As can be seen in the electron micrograph shown in Fig.12, samples of gel purified TGMV subgenomic ssDNA contain a mixture of circular and linear molecules. Hamilton *et al.* (1981) demonstrated that TGMV genomic ssDNA consists of a mixture of circular and linear components and suggested that the linear molecules were derived by breakage of the circular molecules. The same is likely to be true in the case of the subgenomic ssDNA species. However, as subgenomic ssDNA (present in TGMV crude DNA) comigrates with host DNA in agarose gels (Fig.11), it is likely that some of the linear molecules observed in the electron micrographs also consist of co-purified host DNA.

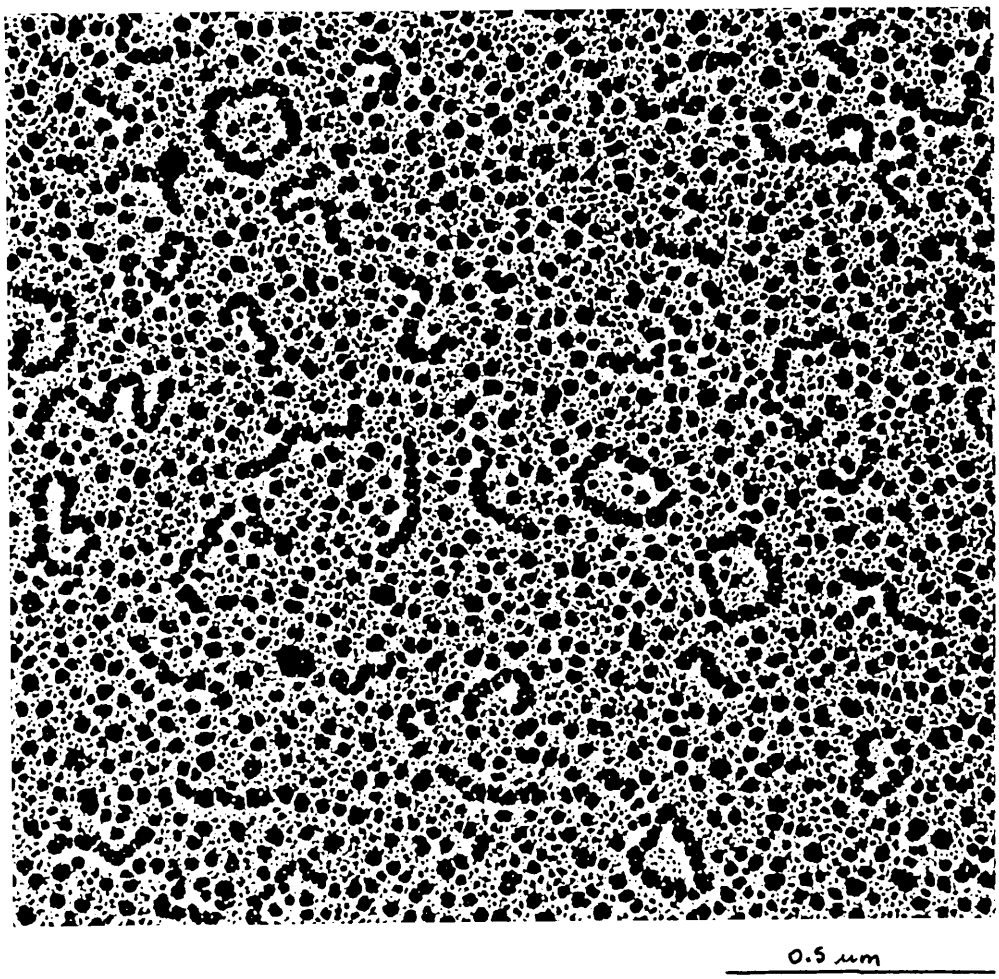



Fig. 12. Electron micrograph of gel purified TGMV subgenomic ssDNA.

Magnification X 71000.

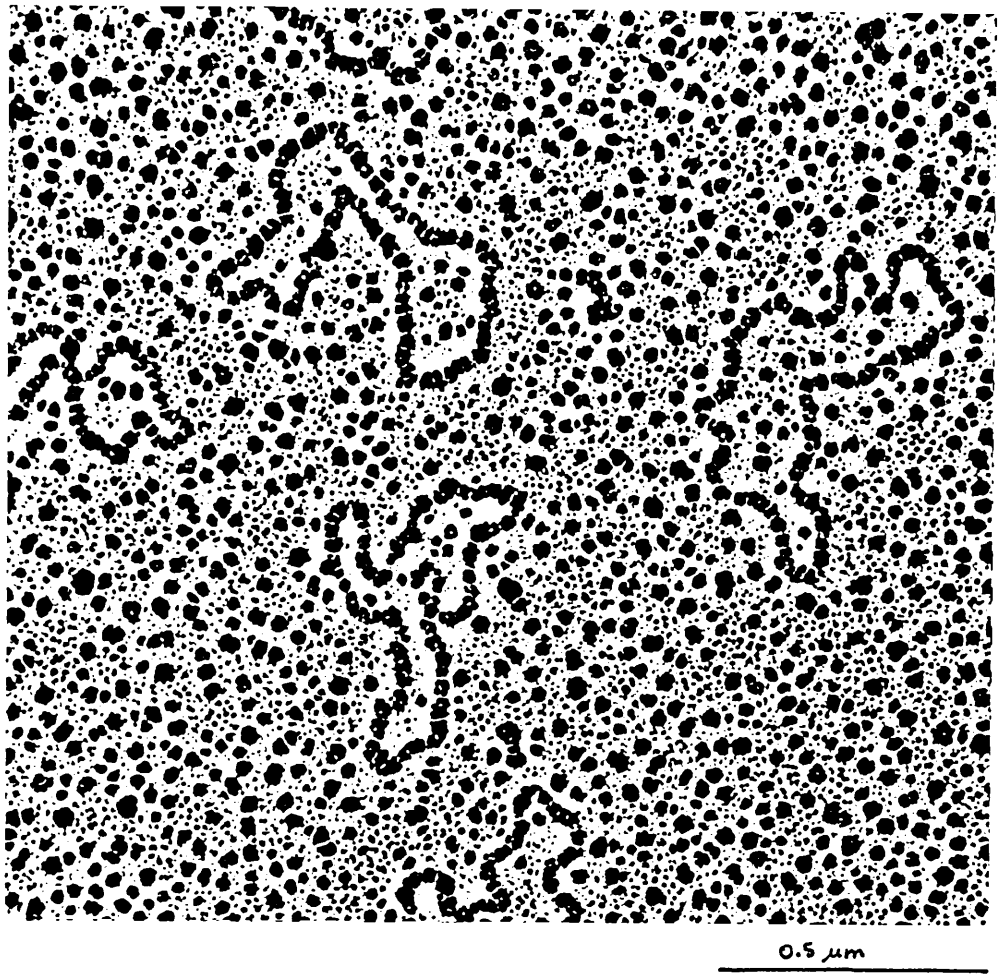



Fig. 13. Electron micrograph of bacteriophage ØX174 ssDNA.

Magnification X 71000.

The size of TGMV subgenomic ssDNA was estimated from electron micrographs by comparing its mean contour length to that of circular phage ØX174 ssDNA (used as an external standard, Fig.13). Contour lengths were determined for 70 subgenomic and 30 phage ssDNA molecules. For each circular molecule, two measurements were taken (of the internal and external circumference, respectively) and the average of the two values was used to calculate the mean value. TGMV subgenomic ssDNA had a mean contour length of 1.83 ± 0.16cm and phage ØX174 ssDNA had a mean contour length of 7.93 ± 0.43cm. Since ØX174 DNA contains 5386 nucleotides (Sanger *et al.*, 1978), the size of TGMV subgenomic ssDNA was calculated as 1242nt (1.83 ÷ 7.93 X 5386). This value is virtually identical to that obtained from methyl-mercury gels (1246nt) and probably represents an accurate average size estimate of the population of subgenomic DNA molecules.

### 3.4 Southern blot analysis of TGMV subgenomic ssDNA

One of the strategies used to investigate the nature of TGMV subgenomic ssDNA consisted of hybridizing a [$^{32}$P]-labelled cDNA probe made to gel purified subgenomic ssDNA to immobilised restriction fragments derived from TGMV 1 and TGMV 2. Failure of any of the fragments to hybridize to the probe would indicate regions of the TGMV genome that are deleted in the subgenomic species.

As TGMV subgenomic ssDNA recovered from agarose gels was invariably contaminated with co-migrating host DNA background and, in some preparations, with a small amount of unit-length viral ssDNA (Fig.14A), it was necessary to further purify the subgenomic ssDNA before preparing the cDNA probe, as contaminating DNA might reduce the specificity of the probe.

Circular and linear DNA molecules of the same Mr have different electrophoretic mobilities in polyacrylamide gels. Hamilton *et al.* (1981) demonstrated by electron microscopy of the eluted bands that the circular form of TGMV virion ssDNA molecules has a slower mobility than the linear form in polyacrylamide gels. Also, dsDNA migrates faster on polyacrylamide gels than ssDNA of similar Mr. These properties of polyacrylamide gels were therefore exploited to separate the circular subgenomic DNA molecules from

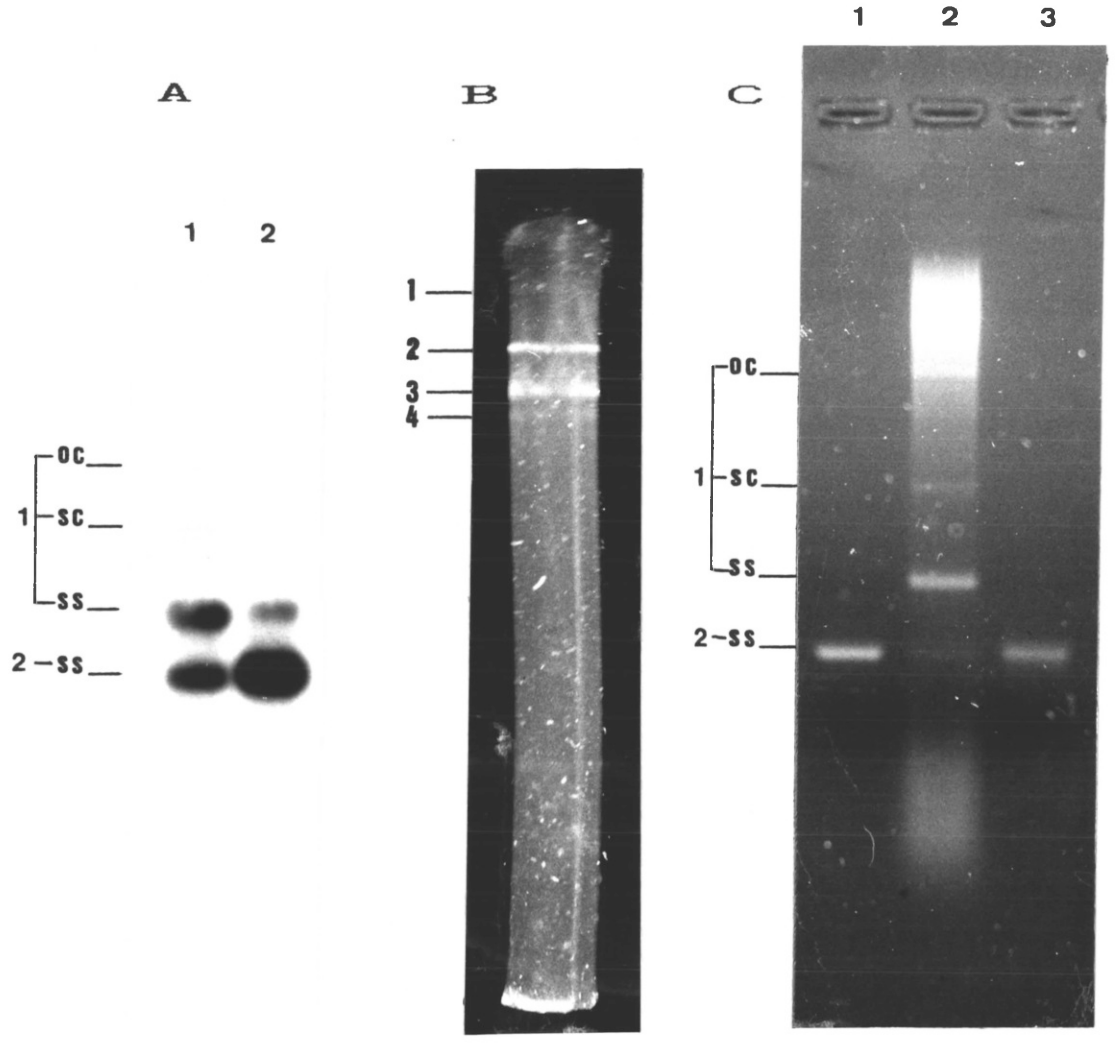



Fig.14. **Gel purification of TGMV subgenomic ssDNA.**

A) Autoradiogram of a blot (from a 1.2% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated TGMV DNA B (pBH 604).
(1) TGMV crude DNA; (2) agarose gel purified TGMV subgenomic ssDNA.

B) Polyacrylamide (4%) tube gel electrophoresis of agarose gel purified TGMV subgenomic ssDNA. See section 3.4 for explanation.

C) Electrophoresis in a 1.2% agarose gel containing 0.5µg/ml ethidium bromide.
(1) TGMV subgenomic circular ssDNA (band 2 in Fig.14B) after sequential purification from agarose gel and polyacrylamide gel respectively; (2) TGMV crude DNA; (3) TGMV subgenomic linear ssDNA (band 3 in Fig.14B) after sequential purification from agarose and polyacrylamide gels respectively.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic length and subgenomic viral DNA respectively.

contaminating DNA molecules (the latter including both the linear form of TGMV subgenomic ssDNA as well as comigrating host DNA).

When TGMV subgenomic ssDNA recovered from agarose gels was fractionated on 4% polyacrylamide tube gels, two major bands (Fig.14B, bands 2 and 3), corresponding to the circular (band 2) and the linear (band 3) forms of TGMV ssDNA, were routinely detected. Some preparations also contained two additional minor bands (Fig.14B, bands 1 and 4). The identity of these was not investigated, but, as they were only occasionally detected, it was thought unlikely that they represented different subgenomic species. The slower migrating band (band 1) probably corresponded to unit-length circular viral ssDNA, while the faster migrating band (band 4) probably consisted of host DNA or one of the double-stranded forms of the subgenomic DNA (see section.3.6). The two major bands (bands 2 and 3) were separately eluted and examined by agarose gel electrophoresis (Fig.14C). Both bands had the same electrophoretic mobility in non-denaturing agarose gels and, as in polyacrylamide tube gels (Fig.14B), circular subgenomic ssDNA produced a tighter band than the linear form. Analysis of the circular subgenomic ssDNA species by Southern blot analysis revealed an electrophoretically pure band (Fig.15B). The purified circular subgenomic ssDNA was therefore used to probe the TGMV DNA 1 and TGMV DNA 2 restriction fragments as follows.

Full-length TGMV DNA components 1 and 2, excised from recombinant plasmids pBH401 and pBH604 (Hamilton *et al.*, 1983) with restriction endonucleases *Eco*RI and *Cla*I respectively, were eluted from agarose gels and further purified using Elutip-d columns (Schleicher and Schüll) to remove residual agarose contamination. The *Eco*RI linearized TGMV DNA 1 insert was digested with restriction enzymes *Hgi*AI; *Apa*I; *Bam*HI; *Sca*I; *Xho*I; and *Pvu*I. The *Cla*I linearized TGMV DNA 2 insert was digested with *Pvu*I; *Eco*RI; *Pst*I; *Bam*HI/*Pst*I; *Bam*HI; *Tth*III1; *Sca*I/*Eco*RV; *Sca*I/*Bgl*II; and *Bcl*I/*Bgl*II. Restriction enzyme reactions were performed overnight to ensure maximum digestion of the DNA samples. The fragments were then resolved by agarose gel electrophoresis (1.5% agarose containing 0.5µg/ml ethidium bromide) and visualized by UV fluorescence. All the expected fragments (with the exception of a 76bp *Pst*I/*Cla*I fragment from TGMV DNA 2, which is too small to be detected by UV

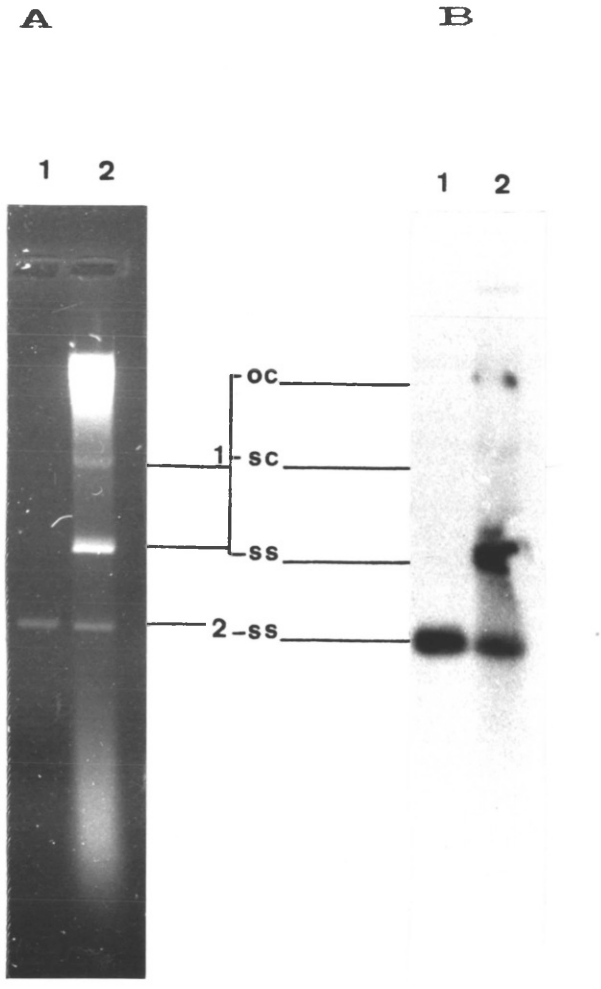



Fig.15. **Examination of gel purified TGMV subgenomic ssDNA.**

A) Electrophoresis in a 1.2% agarose gel containing 0.5µg/ml ethidium bromide.
(1) TGMV subgenomic circular ssDNA (band 2 in Fig.14B) sequentially purified from agarose and polyacrylamide gels respectively; (2) TGMV crude DNA.

B) Autoradiogram of a blot of the gel shown in Fig.15A probed with TGMV subgenomic cDNA.
(1) and (2) as above.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic length and subgenomic viral DNA respectively.

illumination), as predicted from the nucleotide sequences of DNAs 1 and 2 (Hamilton *et al.*, 1984), were detected (Fig.16). The fragments were then transferred to GeneScreen Plus menbrane, hybridized with [$^{32}$P]-labelled cDNA made to the subgenomic ssDNA using reverse transcriptase (section 2.20) and autoradiographed (Fig.17).

No hybridization was detected between the subgenomic probe and TGMV DNA 1 derived restriction fragments which lacked the 200 base 'common region'. A weak hybridization, however, was detected with DNA 1 derived restriction fragments which contained the 'common region' (Fig.17, lanes 2 to 7). In contrast, a strong hybridization was detected between the probe and many TGMV DNA 2 derived fragments, including some which did not contain the 'common region' (Fig.17, lanes 9 to 17). DNA 2 derived restriction fragments located between nucleotides 246 (Pst I site) and 1569 (Sca I site), however, did not hybridize to the probe. This corresponds to a stretch of DNA *ca.*1.3Kb long. When a similar blot was probed with a mixture of [$^{32}$P]-labelled TGMV 1 (pBH401) and TGMV 2 (pBH604), all the restriction fragments that had been visualized by UV illumination were detected (Fig.18).

The results above indicate that the subgenomic DNA is specific to DNA 2. As shown diagrammatically in Fig.19B, *ca.*1.2Kb of the DNA 2 genome hybridized to the subgenomic DNA. This is approximately the same size as that estimated for the subgenomic DNA molecule by electrophoresis through denaturing gels and by electron microscopy (sections 3.2. and 3.3.). Therefore, within the limits of resolution of the method used (±50 bases), the subgenomic ssDNA appears to be a homogeneous population of molecules formed by the deletion of approximately half the DNA 2 molecule. The deleted region, defined by restriction fragments which did or did not hybridize with the subgenomic probe, comprises all of ORF BR1 and the C-terminal region of ORF BL1 and is located approximately between nucleotides 246 and 1569 (Fig.19B). As the 'common region' is retained in the subgenomic ssDNA, the weak hybridization detected with some of the DNA 1 derived restriction fragments can be accounted for entirely by the presence of the 'common region' in these fragments (Fig.19A).

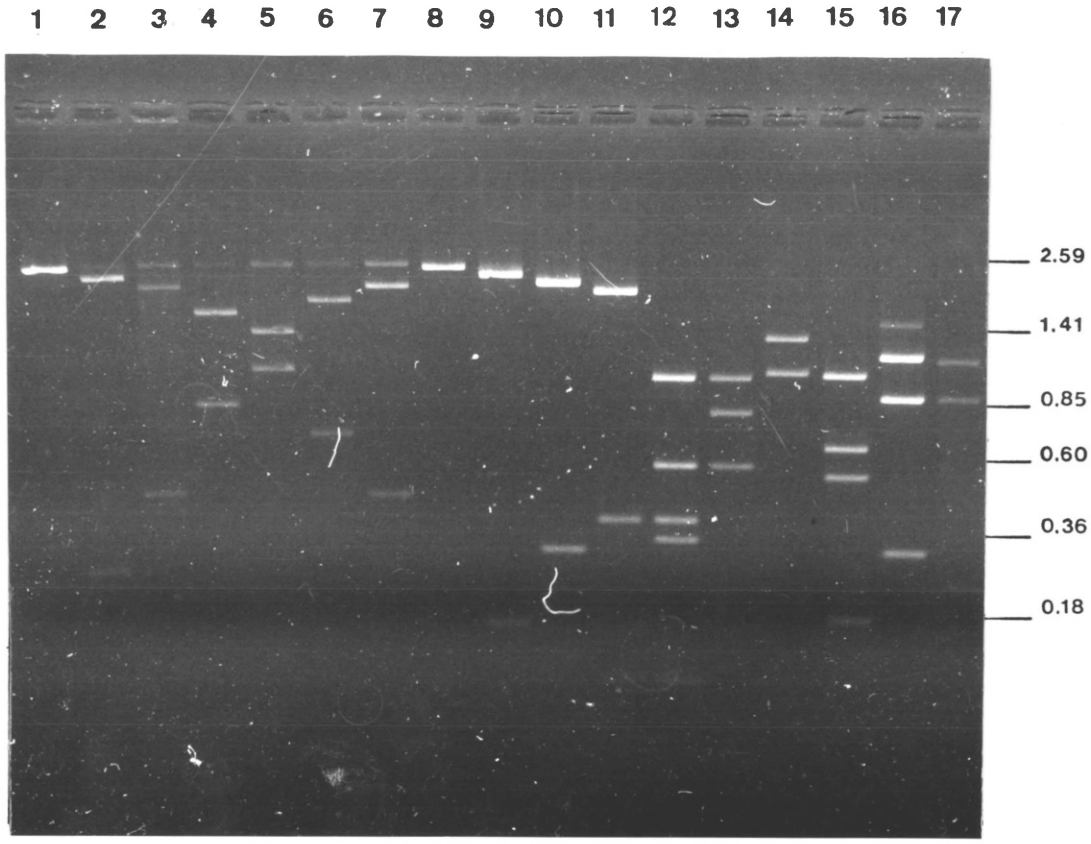



Fig.16. **Characterization of TGMV subgenomic ssDNA (southern blot analysis).**

Electrophoresis in a 1.5% agarose gel containing 0.5µg/ml ethidium bromide.
(1) gel purified TGMV DNA 1 (excised from pBH401 with restriction enzyme *Eco*RI); TGMV DNA 1 (*Eco*RI linearized) digested with (2) *Hgi*AI, (3) *Apa*I, (4) *Bam*HI, (5) *Sca*I, (6) *Xho*I, (7) *Pvu*I; (8) gel purified TGMV DNA 2 (excised from pBH604 with restriction enzyme *Cla*I); TGMV DNA 2 (*Cla*I linearized) digested with (9) *Pvu*I, (10) *Eco*RI, (11) *Pst*I, (12) *Bam*HI + *Pst*I (13) *Bam*HI, (14) *Tth*III1, (15) *Sca*I + *Eco*RV, (16) *Sca*I + *Bgl*II, (17) *Bcl*I + *Bgl*II.
Sizes of DNA fragments are given in Kb.

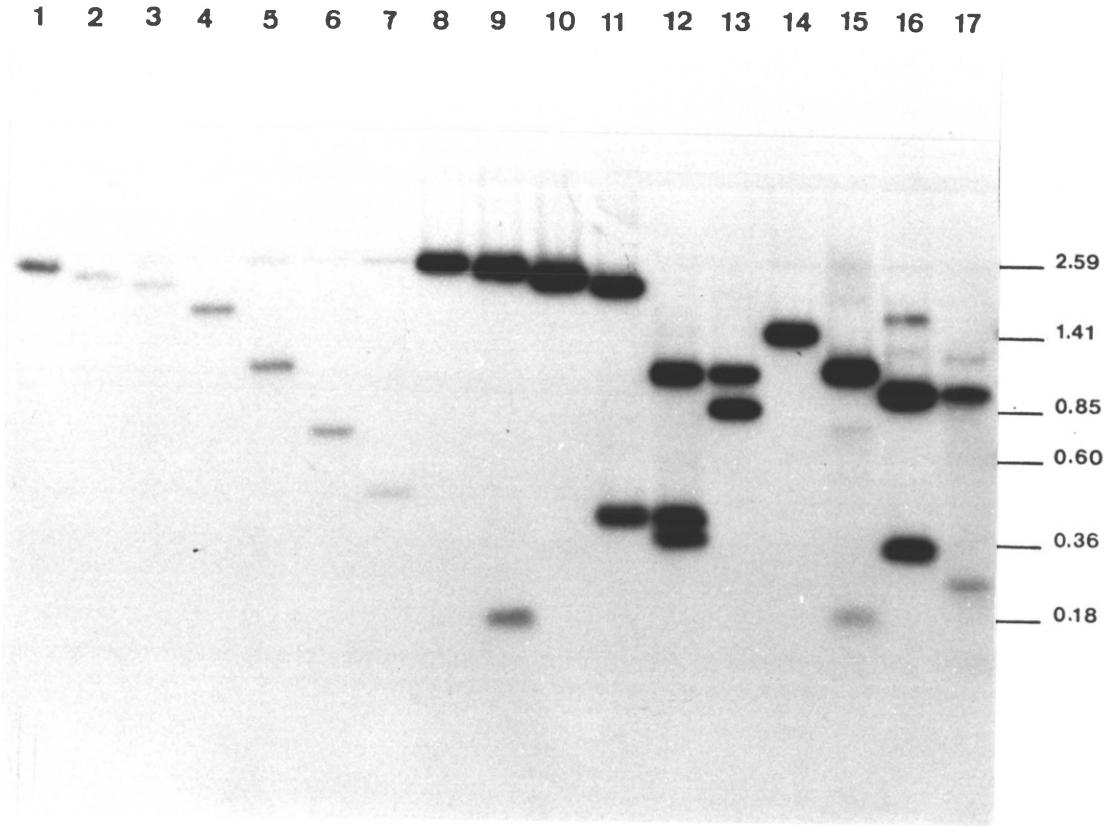



Fig.17. **Characterization of TGMV subgenomic ssDNA (Southern blot analysis).**

Autoradiogram of a blot of the gel shown in Fig.16 probed with TGMV subgenomic cDNA.
(1) to (17) as in Fig.16.
Sizes of DNA fragments are given in Kb.

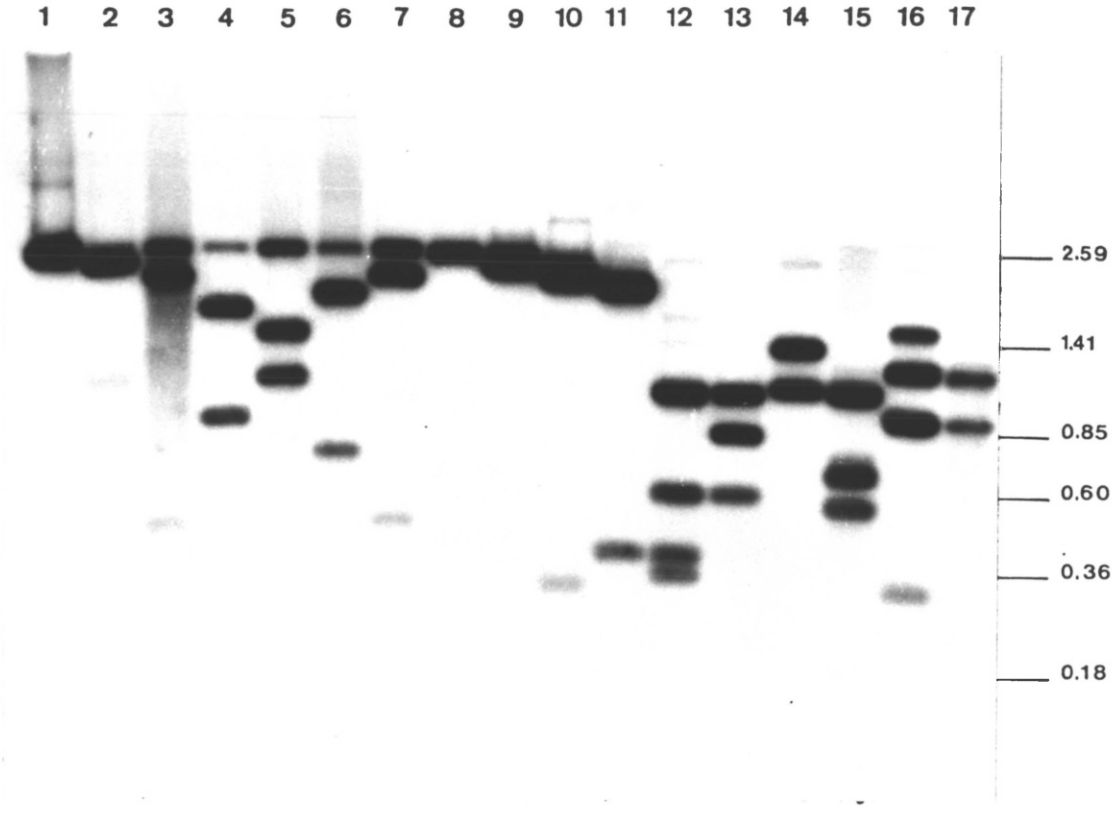



Fig.18. **Characterization of TGMV subgenomic ssDNA (Southern blot analysis).**

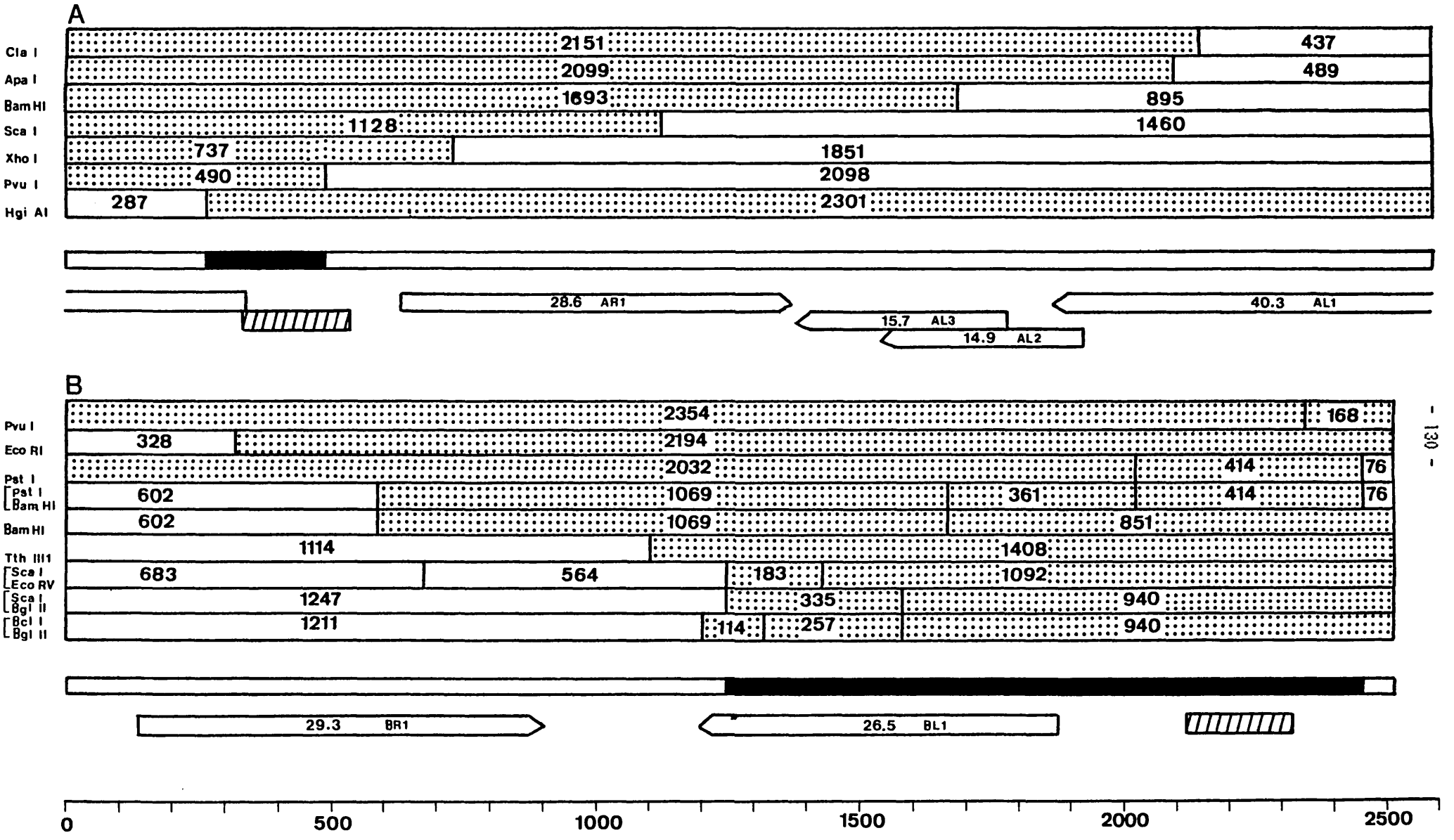
Autoradiogram of a blot of a gel similar to that shown in Fig.16 probed with a mixture of nick-translated TGMV 1 (pBH401) and 2 (pBH604) DNAs.
(1) to (17) as in Fig.16.
Sizes of DNA fragments are given in Kb.

Legend to Fig.19. **Characterization of TGMV subgenomic ssDNA (Southern blot analysis).**

A) Restriction map of *Eco*RI linearized TGMV DNA 1 (pBH401 insert).

B) Restriction map of *Cla*I linearized TGMV DNA 2 (pBH604 insert).

Only the location of the cleavage sites for the enzymes used to characterize TGMV subgenomic ssDNA are shown. The sizes of the restriction fragments generated by these enzymes are given in base pairs. Dotted areas indicate the restriction fragments that hybridize to TGMV subgenomic cDNA. The genomes of TGMV and are represented as open boxes and the black boxes within these show the regions of DNAs 1 and 2 that are retained in (or hybridize to) TGMV subgenomic DNA, as defined by those fragments that did or did not hybridize to the subgenomic probe. The major ORFs in DNAs 1 and 2 are indicated as open arrows and the sizes of the corresponding proteins are given in Kd. The hatched boxes indicate the position of the 200bp 'common region'. The scale is in base pairs.


A
Cla I
Apa I
BamHI
Sca I
Xho I
Pvu I
Hgi AI
2151
437
2099
489
1693
895
1128
1460
737
1851
490
2098
287
2301
28.6 AR1
40.3 AL1
15.7 AL3
14.9 AL2
B
Pvu I
Eco RI
Pst I
Pst I
Bam HI
BamHI
Tth III1
Sca I
Eco RV
Sca I
Bgl II
Bcl I
Bgl II
2354
168
328
2194
2032
414
76
602
1069
361
414
76
602
1069
851
1114
1408
683
564
183
1092
1247
335
940
1211
114
257
940
29.3 BR1
26.5 BL1
0
500
1000
1500
2000
2500

## 3.5 Heteroduplex analysis of TGMV subgenomic ssDNA

A second approach used to characterise TGMV subgenomic ssDNA consisted in the electron microscopic examination of heteroduplex formed between the subgenomic DNA and either TGMV DNA 1 or TGMV DNA

Cloned full-length TGMV DNA components 1 and 2 were excised from recombinant plasmids pBH401 and pBH407 (the latter a non-infective DNA 2 molecule cloned into pAT153 at its unique *Eco*RI site; Bisaro *et al.*, 1982) respectively with restriction enzyme Eco RI and gel purified (section 2.16.2). Each TGMV insert was then denatured and hybridized (separately) to gel purified (section 2.16.1) subgenomic ssDNA as described in section 2.28. The samples were then spread by the formamide/cytochrome c technique (section 2.28.) and examined under the electron microscope.

Very few putative heteroduplexes between the subgenomic ssDNA and TGMV DNA 1 were found. In these only a small region of each molecule appeared to be involved in the formation of the heteroduplexes (Fig.20). The size of this region (*ca.*0.2Kb) and its position close to one extremity of the *Eco*RI linearized TGMV DNA 1 component (see Fig.19A for the postion of the 'common region' relative to the Eco RI site) are consistent with it being the 'common region'. This is in good agreement with the results obtained by Southern blot analysis (section 3.4), adding further support to the view that the subgenomic ssDNA is specific of DNA 2. The difficulty in finding heteroduplexes between the subgenomic ssDNA and TGMV DNA 1 probably results from the short length of the complementary region in these molecules. Indeed, the frequency in which heteroduplexes are formed depends on the percentage of homology between the strands. Hence, heteroduplexes would be expected to be only a small fraction of the total number of single-stranded and renatured molecules if the homologous regions are short.

In contrast, heteroduplexes between the subgenomic DNA and TGMV DNA 2 were frequently observed. These appeared as double-stranded circles with two single-stranded tails (Fig.21) the mean lengths of which were calculated as 0.73 ± 0.06cm and 1.68 ± 0.18cm respectively from 24 independent measurements. Taking the length of

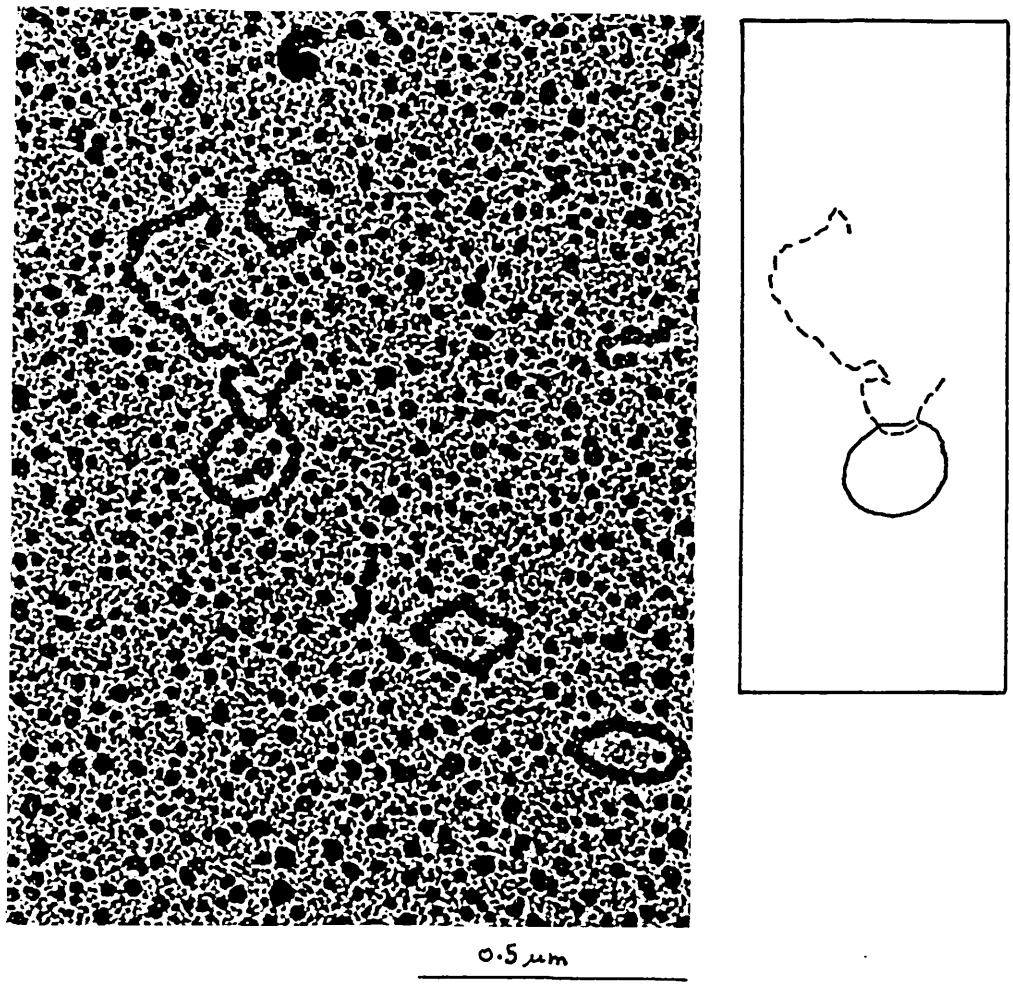



Fig.20. **Characterization of TGMV subgenomic ssDNA (heteroduplex analysis).**

Electron micrograph of a (putative) heteroduplex formed between gel purified TGMV subgenomic ssDNA and TGMV DNA I (gel purified full-length insert excised from pBH401 with restriction enzyme *Eco*RI). The diagram shows an interpretation of the heteroduplex with TGMV subgenomic ssDNA drawn as a solid line and TGMV DNA I as a broken line. Magnification X 71000.

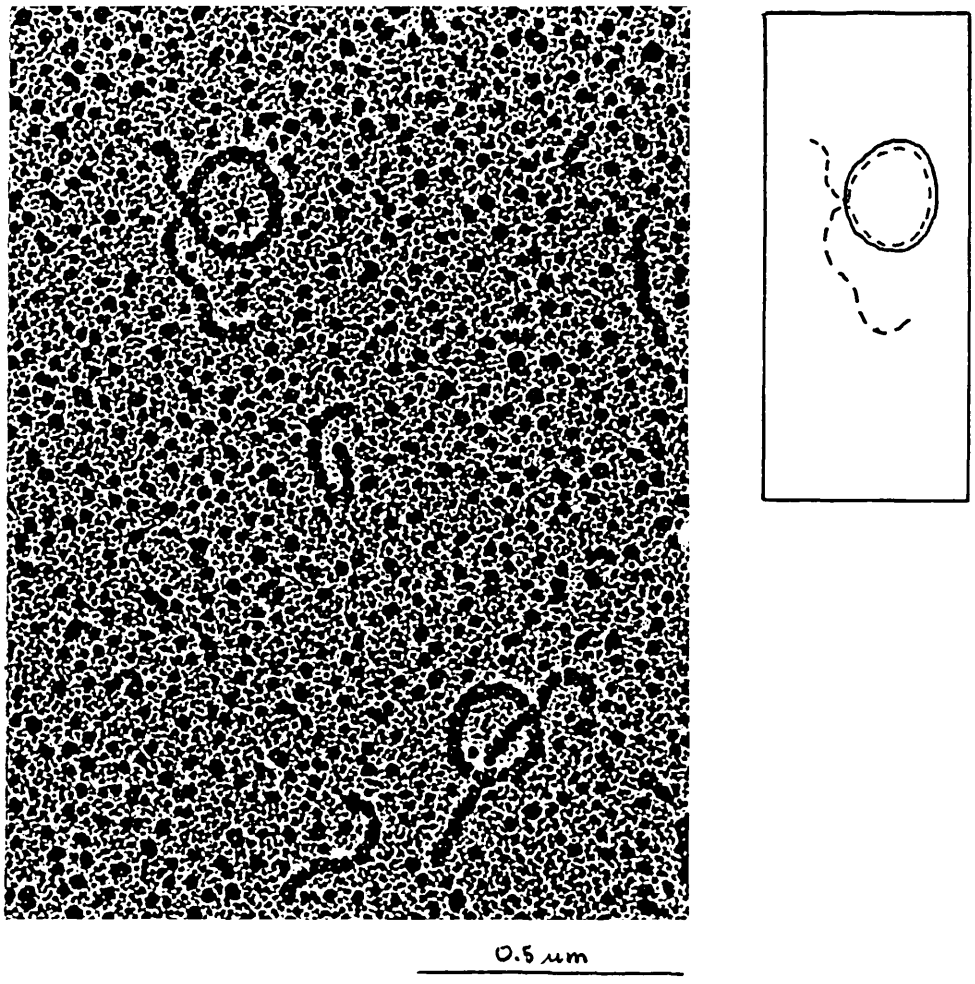



Fig.21. **Characterization of TGMV subgenomic ssDNA (heteroduplex analysis).**

Electron micrograph of a heteroduplex formed between gel purified TGMV subgenomic ssDNA and TGMV DNA 2 (gel purified full-length insert excised from pBH407 with restriction enzyme *Eco*RI). The diagram shows an interpretation of the heteroduplex with TGMV subgenomic ssDNA drawn as a solid line and TGMV DNA 2 as a broken line. Magnification X 71000.

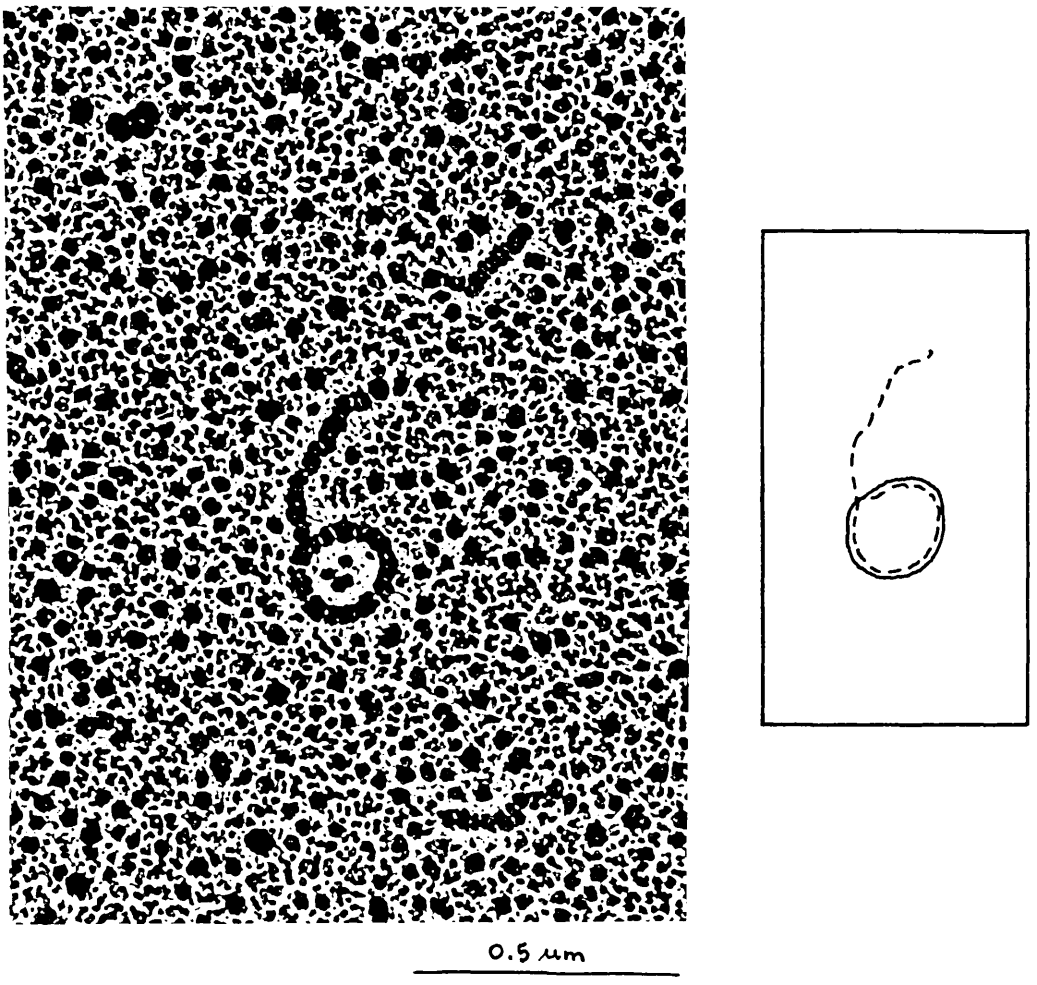



Fig.22. **Characterization of TGMV subgenomic ssDNA (heteroduplex analysis).**

Electron micrograph of a heteroduplex formed between gel purified TGMV subgenomic ssDNA and *Cla*I digested TGMV DNA 2 (gel purified full-length insert excised from pBH407 with restriction enzyme *Eco*RI). The diagram shows an interpretation of the heteroduplex with TGMV subgenomic ssDNA drawn as a solid line and *Cla*I digested TGMV DNA 2 as a broken line. Magnification X 71000.

the subgenomic ssDNA as 1240nt (sections 3.2 and 3.3), the added length of the two tails should be approximately equal to 1280nt (the length of TGMV DNA 2 minus that of the subgenomic DNA). The length of the single-stranded tails was therefore calculated as *ca.*390nt (0.73 X 1280 ÷ (0.73+1.68)) and 890nt (1.68 X 1280 ÷ (0.73+1.68)) respectively.

These tails correspond to the region of DNA 2 that is missing in the subgenomic DNA (see diagram in Fig.21). In order to determine the position of the deletion relative to full-length TGMV 2, *Eco*RI linearized DNA 2 was digested with restriction enzyme *Cla*I (322) before the formation of heteroduplexes. This resulted in heteroduplexes lacking the shorter tail (Fig.22), indicating that the deletion in the subgenomic DNA lies approximately between nucleotides 260 and 1560, in good agreement with the position estimated by Southern blot analysis (section 3.4).

## 3.6 Characterization of TGMV subgenomic dsDNA

### 3.6.1 Hydroxylapatite chromatography of TGMV crude DNA

The small size of bands a, b and c, shown in Fig.9B (lane 5), and their resistance to S1 nuclease, suggested that they might correspond to dsDNA forms of the subgenomic species. As these DNA species were present in very small amounts, it was necessary to enrich them for further characterization. This was achieved by hydroxylapatite chromatography (HAP) of TGMV crude DNA. Although the method used had been originally described for the removal of contaminating chromosomal DNA in bacterial ccc dsDNA purification (Colman *et al.*, 1978), it proved quite effective in enriching TGMV crude DNA for viral dsDNA species.

After heat denaturation at 100°C for 5min, TGMV crude DNA was subjected to HAP chromatography as described in section 2.12. Single-stranded and double-stranded fractions were collected, separately pooled and fractionated by agarose gel electrophoresis (Fig.23, lanes 2 and 3). Most of the host DNA and the viral ssDNA species were eluted in the single-stranded fraction (Fig.23, lane 3). The double-stranded fraction (TGMV crude dsDNA), on the other hand, contained a number of viral dsDNA bands, including the unit-

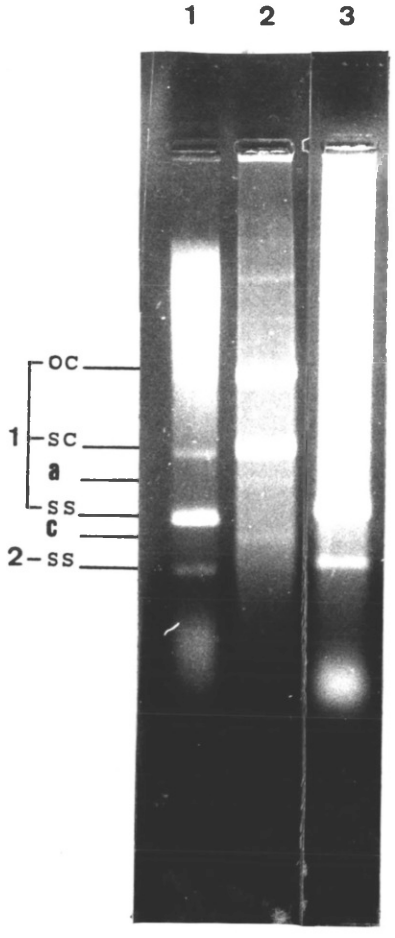



Fig.23. **Hydroxylapatite chromatography (HAP) of TGMV crude DNA.**

Electrophoresis in a 1.2% agarose gel containing 0.5µg/ml ethidium bromide.
(1) TGMV crude DNA; (2) double-stranded and (3) single-stranded hydroxylapatite fractions containing TGMV crude DNA.
oc=open circular ssDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic-length and subgenomic viral DNA respectively. bands a and c are previously uncharacterized viral species.

length and high molecular weight dsDNA species described previously (Hamilton *et al.*, 1982) as well as host DNA. In addition, bands a and c were now clearly visible by UV illumination. Band b, however, did not appear to be present in the double-stranded fraction. This could either be due to it having been eluted in the single-stranded fraction or to it not having resolved from band a in the gel.

As the DNA had been denatured prior to being loaded onto the HAP column, only ccc dsDNA was expected to be eluted in the double-stranded fraction. It was therefore surprising to find in this fraction a number of other viral DNA bands as well as host DNA. Although the presence of viral oc dsDNA in the double-stranded fraction could be explained by 'nicking' of the ccc dsDNA during the procedure, the presence of host DNA suggests that incomplete denaturation or partial renaturation of the DNA had occurred.

### 3.6.2 Specificity of TGMV subgenomic dsDNA

To determine the specificity of bands a and c, TGMV crude dsDNA was fractionated by agarose gel electrophoresis, transferred to GeneScreen Plus membrane and hybridized to specific TGMV DNA restriction fragments cloned in pEMBL19. These consisted of a 1408bp fragment of DNA 2 spanning restriction sites *Tth*III1 (1436) to *Cla*I (322), a 1114bp fragment of DNA 2 spanning *Cla*I (323) to *Tth*III1 (1436) and a 1851bp fragment of DNA 1 spanning *Xho*I (400) to *Eco*RI (2251) (see Fig.19 A and B for restriction maps). The procedure used to clone these fragments is illustrated in Fig.24. To clone the DNA 2 fragments, pBH604 was digested with *Tth*III1 and the resulting cohesive ends were converted to blunt-ends with the Klenow fragment of DNA polymerase I (section 2.23.1). The same plasmid was then digested with *Cla*I. The TGMV DNA 2 derived *Cla*I/*Tth*III1 fragments were then gel purified and cloned into pEMBL19 double-digested with *Acc*I and *Sma*I. To clone the TGMV DNA 1 derived 1851bp *Xho*I/*Eco*RI restriction fragment, pBH401 was double-digested with *Xho*I and *Eco*RI. The TGMV DNA 1 derived 1851bp *Xho*I/*Eco*RI restriction fragment was then gel purified and cloned into pEMBL19 double-digested with *Sal*I and *Eco*RI. The resultant recombinant plasmids

Legend to Fig.24. **Construction of recombinant plasmids containing specific fragments from TGMV DNAs 1 and 2.**

A) Construction of recombinant plasmids containing TGMV DNA 2 (pBH604) derived *Cla*I/*Tth*III1 fragments.

B) Construction of a recombinant plasmid containing the TGMV DNA 1 (pBH401) derived 1851bp *Xho*I/*Eco*RI fragment.

See section 3.6.2 for details of construction
TGMV sequences are indicated by dots. The 200bp 'common region' is indicated as a black box. Sizes of restriction fragments are given in base pairs. A=*Acc*I; C=*Cla*I; E=*Eco*RI; H=*Hin*dIII; K=*Kpn*I; S=*Sma*I; S=*Sal*I; T=*Tth*III1; X=*Xho*I.

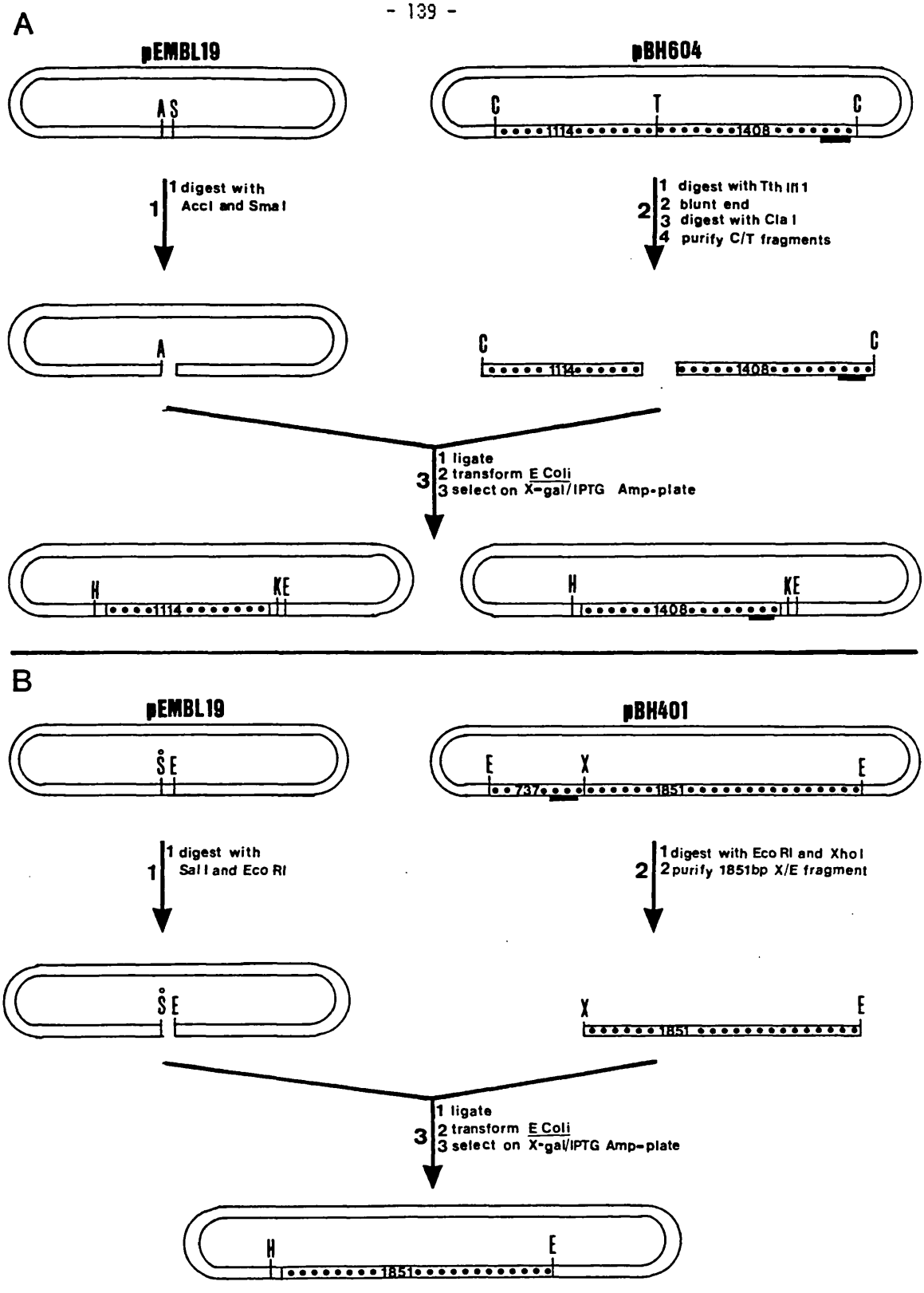

A
pEMBL19
A S
pBH604
C T C
1114
1408
1 digest with
Acc I and Sma I
2 1 digest with Tth III 1
2 blunt end
3 digest with Cla I
4 purify C/T fragments
A
C C
1114
1408
3 1 ligate
2 transform E Coli
3 select on X-gal/IPTG Amp-plate
H KE
1114
H KE
1408
B
pEMBL19
S E
pBH401
E X E
737
1851
1 digest with
Sal I and Eco RI
2 1 digest with Eco RI and Xho I
2 purify 1851bp X/E fragment
S E
X E
1851
3 1 ligate
2 transform E Coli
3 select on X-gal/IPTG Amp-plate
H E
1851

Fig. 24

were isolated as described in section 2.27 and analysed by restriction digestion and agarose gel electrophoresis to check that the right fragments had been cloned (Fig.25). Each plasmid was then separately labelled with [$^{32}$P] by nick-translation (section 2.21) and used to probe separate identical blots (Fig.26).

Similar to the subgenomic ssDNA (Fig. 26A, lane 1; Fig.26B, lane 1), both bands a and c hybridized to the 1408bp *Tth*III1/*Cla*I fragment from DNA 2 (Fig.26A, lanes 2 and 3) but not to the 1114 *Cla*I/*Tth*III1 DNA 2 fragment (Fig.26B, lanes 2 and 3). The latter fragment, but not the former, is located in the region of DNA 2 that is deleted from the subgenomic ssDNA (section 3.4; Fig.19B). Also similar to the subgenomic ssDNA (Fig 26C, lane 1), neither band hybridized to the 1851bp *Xho*I/*Eco*RI fragment of DNA 1 (Fig.26C, lanes 2 and 3) which lacked the 'common region'. Actually, a very weak hybridization signal could be seen in the regions corresponding to the subgenomic ssDNA and band a in lanes 1 and 2 of Figs.26 B and C. As bands a and c are different forms of the same DNA species (see section 3.6.3), it would be expected that both bands would hybridize if there were any homology between these bands and the probe. Also, no hybridization is seen with band a in lane 3 of the same Figs. It is therefore proposed that the detected signal is artifactual rather than genuine. Lane 4 in each blot contains a mixture of pBH604 (TGMV DNA 2) digested with *Tth*III1 and *Cla*I and pBH401 (TGMV DNA 1) digested with *Xho*I and *Eco*RI. These digests were included to determine the amount of cross-hybridization, if any, between the different fragments. As expected, the 1408bp *Tth*III1/*Cla*I fragment from DNA 2 hybridized to the 737bp *Xho*I/*Eco*RI fragment of DNA 1, as both contain the 'common region' (Fig.26A, lane 4). There was no detectable cross-hybridization, however, between any of the fragments used as probes. The two bands migrating in front of the 1851bp fragment of DNA 1 in lane 4 of Fig.26C do not correspond to any of the fragments used as probes and are possibly products of a weak Eco RI star activity on pBH401. As there is no detectable cross-hybridization between the fragments used as probes, the very weak hybridization signal observed in the regions corresponding to band a and the subgenomic ssDNA in lanes 1 and 2 of Figs.26 B and C is thought to be due to a stacking effect created by the high

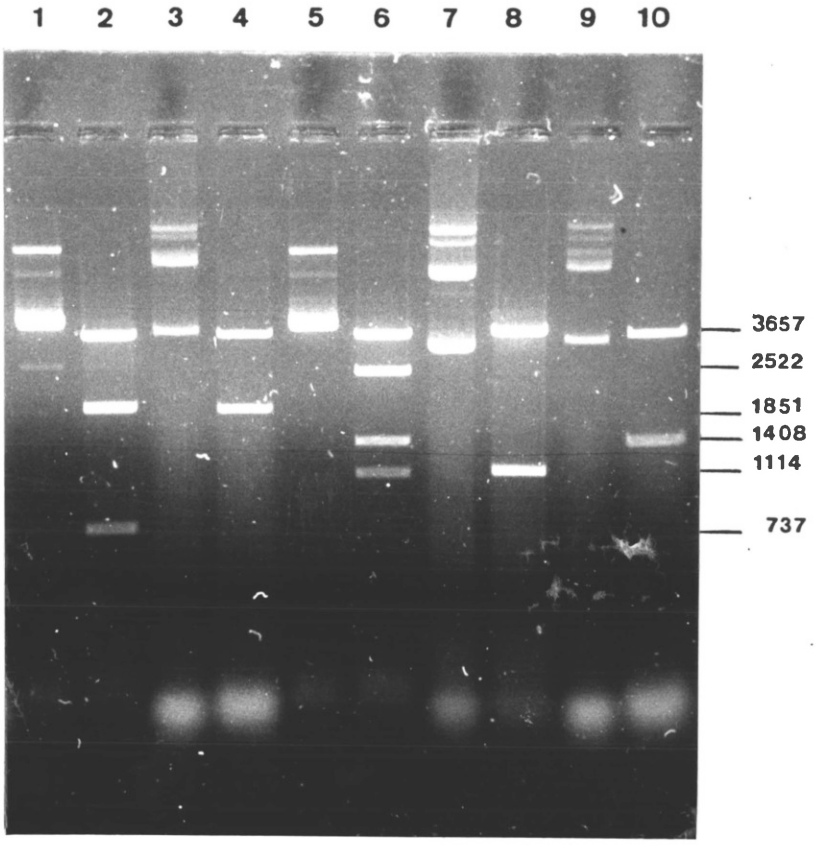



Fig.25. **Restriction analysis of recombinant plasmids containing specific fragments from TGMV DNAs 1 and 2.**

Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.
(1) pBH401; (2) pBH401 double-digested with *Xho*I and *Eco*RI; (3) pEMBL19 containing the TGMV DNA 1 derived 1851bp *Xho*I/*Eco*RI fragment; (4) pEMBL19 containing the TGMV DNA 1 derived 1851bp *Xho*I/*Eco*RI fragment double-digested with *Hin*dIII and *Eco*RI; (5) pBH604; (6) pBH604 double-digested with *Cla*I and *Tth*III1; (7) pEMBL19 containing the TGMV DNA 2 derived 1114bp *Cla*I/*Tth*III1 fragment; (8) pEMBL19 containing the TGMV DNA 2 derived 1114bp *Cla*I/*Tth*III1 fragment double-digested with *Hin*dIII and *Kpn*I; (9) pEMBL19 containing the TGMV DNA 2 derived 1408bp *Tth*III1/*Cla*I fragment; (10) pEMBL19 containing the TGMV DNA 2 derived 1408bp *Tth*III1/*Cla*I fragment double-digested with *Hin*dIII and *Eco*RI. (see figure 24 for the location of restriction enzyme sites).
Sizes of DNA fragments are given in base pairs.

Legend to Fig.26. **Characterisation of TGMV subgenomic dsDNA (specific probing).**

A) Autoradiogram of a blot (from a 1.2% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated cloned 1408bp *Cla*I/*Tth*III1 fragment from TGMV DNA 2.
(1) TGMV crude DNA; (2) TGMV crude dsDNA (double-stranded fraction from HAP column) preparation I; (3) TGMV crude dsDNA (double-stranded fraction from hydroxylapatite column) preparation II; (4) a mixture of plasmid pBH604 double-digested with *Cla*I and *Tth*III1 and plasmid pBH401 double-digested with *Eco*RI and *Xho*I.

B) Autoradiogram of a blot (from a 1.2% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated cloned 1114bp *Cla*I/*Tth*III1 fragment of TGMV DNA 2.
(1) to (4) as in Fig.26A.

C) Autoradiogram of a blot (from a 1.2% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated cloned 1851bp *Xho*I/*Eco*RI fragment of TGMV DNA 1.
(1) to (4) as in Fig.26A.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic-length and subgenomic viral DNA. Bands a and c are previously uncharacterized viral DNA species. Fragment sizes are given in base pairs.

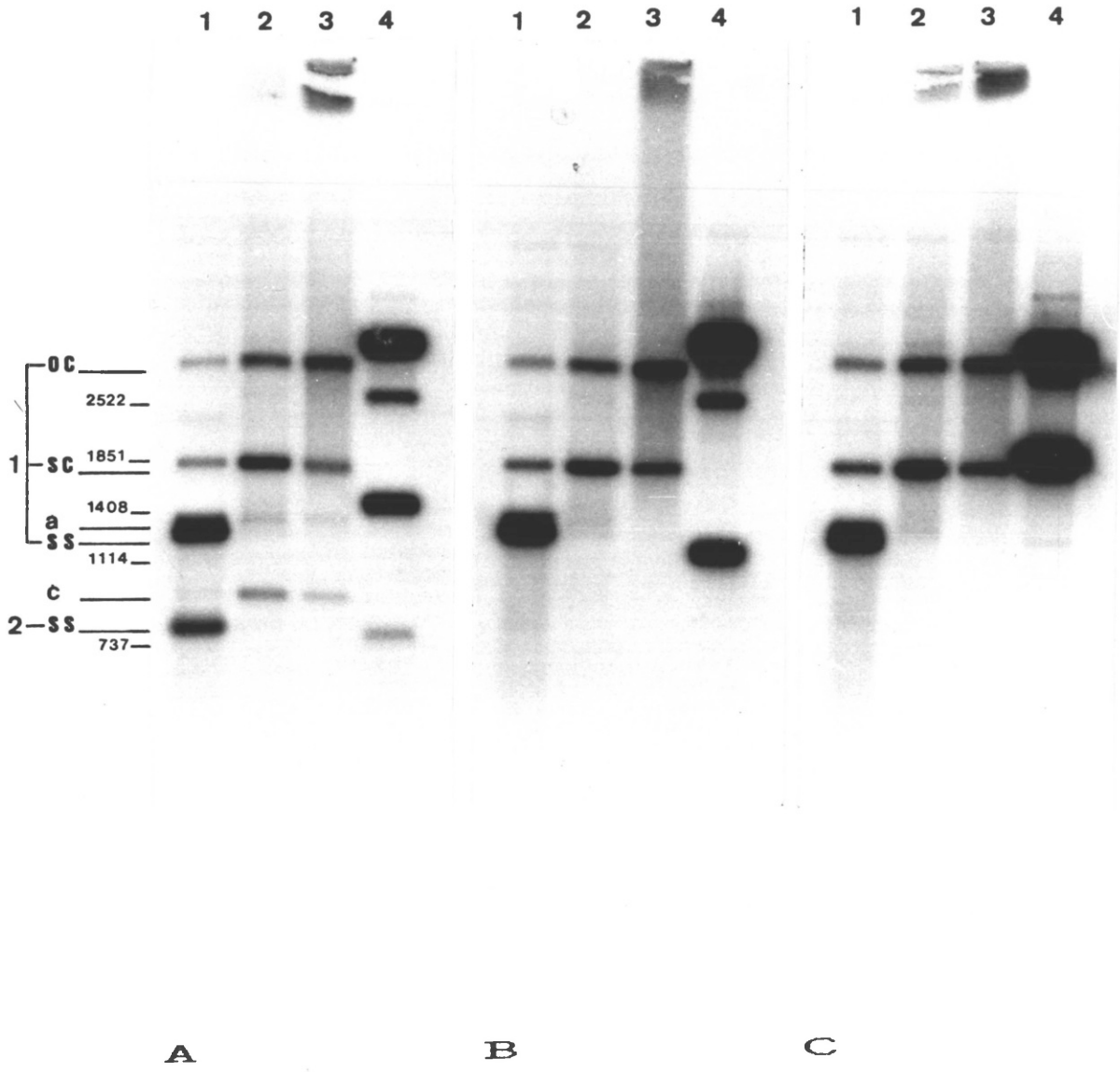



Fig. 26

concentration of these DNA species in these areas, which causes an increase in the background. These results indicate that bands a and c have the same specificity as the subgenomic ssDNA and probably correspond to dsDNA forms of the latter.

### 3.6.3 Nuclease studies on TGMV subgenomic dsDNA

In order to determine the conformation of the two subgenomic dsDNA species (bands a and c), TGMV crude dsDNA (section 3.6.2) was digested (separately) with DNAse I, RNAse A, S1 nuclease and exonuclease III (section 2.18), fractionated by agarose gel electrophoresis, transferred to GeneScreen Plus and hybridized to [$^{32}$P]-labelled nick-translated TGMV 2 (pBH604) (Fig.27).

As expected, both species were completely degraded by DNAse I and resistant to RNAse A (Fig.27, lanes 4 and 5), confirming their DNA nature .Also, both species were resistant to S1 nuclease under conditions were ssDNA was degraded (Fig.27, lane 6; Fig.9 lane 5), confirming that they are both double-stranded molecules. This was indeed suggested by the fact that they had been eluted in the HAP double-stranded fraction. Only band a, however, was sensitive to exonuclease III (Fig.27, lane 7), an enzyme which specifically hydrolyses dsDNA molecules carrying a 3'-OH terminus (Weiss, 1978). During this reaction a species with the mobility of the subgenomic ssDNA was produced (Fig.27, lane 7), suggesting that band a is an oc dsDNA molecule with a 'nick' in one of the strands only. Digestion of such a molecule with exonuclease III would indeed be expected to degrade the 'nicked' strand while leaving the circular strand intact. Band c, on the other hand, has the characteristics of a ccc dsDNA molecule. This is consistent with the behaviour of this molecule in cesium chloride-ethidium bromide gradients (Fig.28). As can be seen in lane 4 of Fig 28A, band c appears to have a density similar to that of unit-length ccc dsDNA, as judged by their relative concentrations in each fraction. When gel-purified, band c was usually contaminated with small amounts of the oc dsDNA molecule (band a) (Fig.29, lane 10), suggesting that the latter arises by 'nicking' of the former. These results, together with a size estimate of *ca.*1272bp for the linearized form (section 3.7), indicate that bands a and c correspond to the oc and ccc dsDNA forms

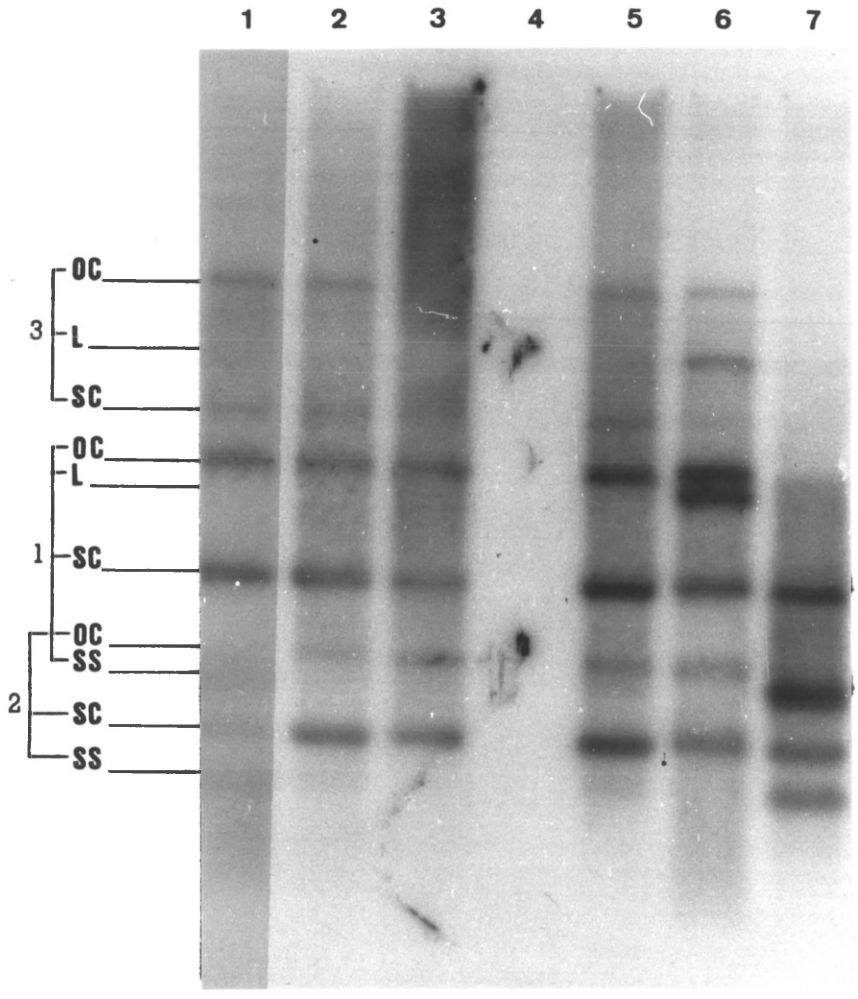



Fig.27. **Characterization of TGMV subgenomic dsDNA (nuclease sensitivity test).**

Autoradiogram of a blot (from a 1.0% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated TGMV DNA 2 (pBH604).
(1) TGMV crude DNA; (2) TGMV crude dsDNA (double-stranded fraction from HAP column) preparation I; (3) TGMV crude dsDNA (double-stranded fraction from hydroxylapatite column) preparation II; (4), (5), (6) and (7) TGMV crude dsDNA (double-stranded fraction from hydroxylapatite column) preparation I digested with DNAse I, RNAse A, S1 nuclease and Exonuclease III, respectively.

oc=open circular dsDNA; L=linear dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1, 2 and 3 refer to genomic length, subgenomic and dimeric (M. Slomka; personal communication) viral DNA, respectively.

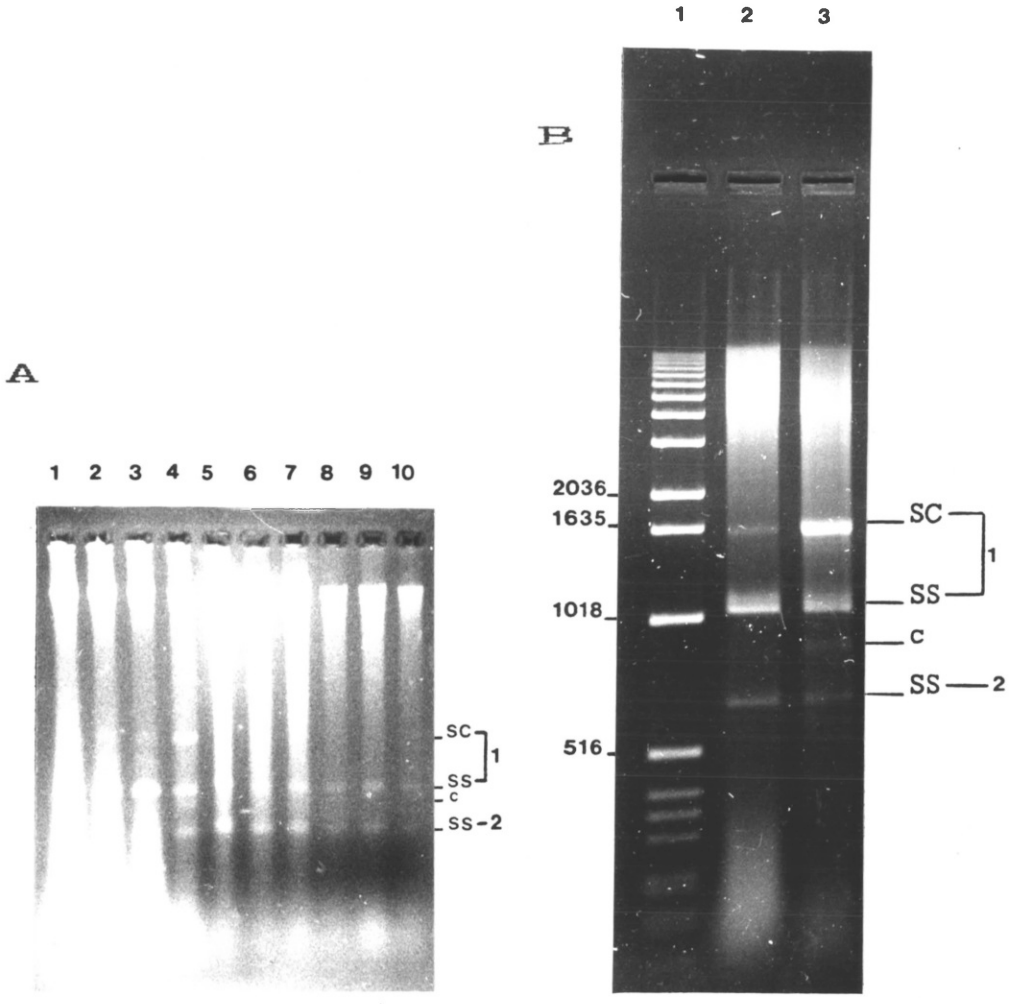



Fig.28. **Centrifugation of TGMV crude DNA in cesium chloride/ethidium bromide gradients.**

A) TGMV crude DNA was centrifuged in cesium chloride/ethidium bromide gradients and fractionated as described in section 2.11 and fractions 1 to 10 (from bottom to top of the gradient were fractionated by electrophoresis on a 1% agarose gel containing 0.5µg/ml ethidium bromide.
(1) to (10) correspond to fractions 1 to 10.

B) Electrophoresis in a 1.4% agarose gel containing 0.5µg/ml ethidium bromide.
(1) Kb marker (fragment sizes are given in base pairs); (2) TGMV crude DNA; (3) same as (4) in Fig.28A.

sc=supercoiled dsDNA (covalently closed circular dsDNA) and ss=single-stranded DNA. 1 and 2 refer to genomic length andsubgenomic viral DNA respectively. Band c is the ccc dsDNA form of TGMV subgenomic DNA.

of the subgenomic DNA derived from DNA 2. Subgenomic dsDNA species have also been demonstrated for the geminiviruses CLV (Stanley and Townsend, 1985) and WDV (H.Macdonald, personal communication).

The identification of viral dsDNA species in geminivirus-infected plants suggested that these viruses replicate *via* dsDNA intermediates (Ikegami *et al.*, 1981; Hamilton *et al.*, 1982; Stanley and Townsend, 1985). Similarly, the identification of intracellular dsDNA forms of TGMV, CLV and WDV subgenomic DNAs suggests that the latter replicate from their own dsDNA templates.

### 3.7 Molecular cloning of TGMV subgenomic dsDNA

In order to define the exact location of the deletion points in the subgenomic DNA, it was decided to clone the subgenomic dsDNA and sequence across the deletion borders. To find single restriction sites which could be used to clone full-length subgenomic molecules, gel purified covalently closed circular subgenomic DNA was digested with restriction nucleases *Eco*RV, *Hpa*I, *Bam*HI and *Bgl*II, fractionated by agarose gel electrophoresis, transferred to GeneScreen Plus and hybridized to [$^{32}$P]-labelled nick-translated TGMV DNA 2 (pBH604) (Fig.29 lanes 6 to 10). The enzymes above were chosen as they were expected to cleave the subgenomic DNA at single sites, as deduced from the TGMV DNA 2 sequence (Hamilton *et al.*, 1984) and the location of the deletion in the subgenomic DNA.

Digestion with restriction enzymes *Eco*RV and *Bam*HI appeared to linearize most of the subgenomic DNA (Fig.29, lanes 6 and 8). Approximately half of the subgenomic DNA molecules appeared to be linearized by *Hpa*I (Fig.29, lane 7) and two fragments were produced when the subgenomic DNA was digested with *Bgl*II (Fig.29, lane 9). The generation of two fragments by *Bgl*II indicated the presence of two sites for this enzyme in the subgenomic DNA. This was unexpected, since only one site for this enzyme is present in the cloned full-length TGMV DNA 2 molecule (Hamilton *et al.*, 1984). However, digestion of gel purified unit-length intracellular viral ccc dsDNA with *Bgl*II also produced two fragments, the smaller of which had the same electrophoretic mobility as the smaller *Bgl*II fragment derived from the subgenomic DNA. This indicates some degree

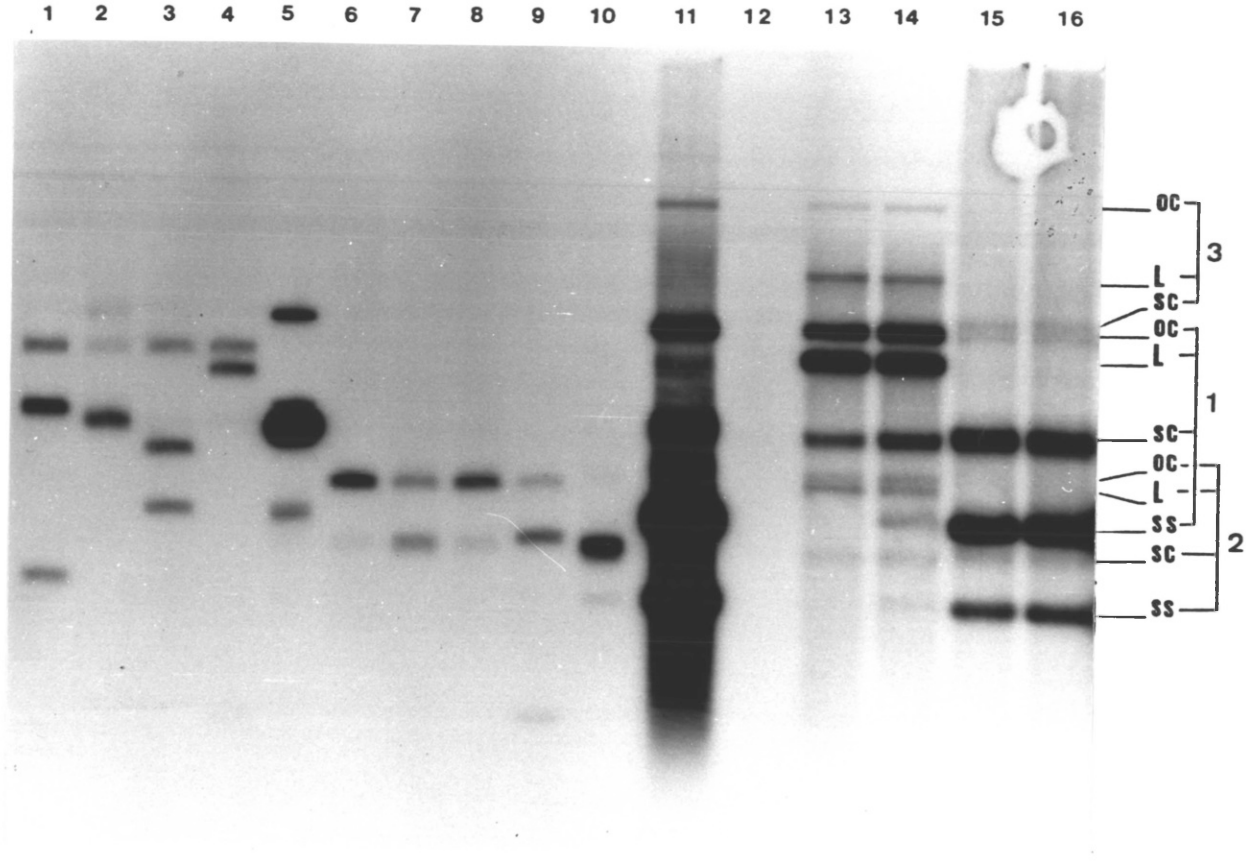



Fig.29. **Restriction enzyme analysis of TGMV dsDNA.**

Autoradiogram of a blot (from a 1.2% agarose gel containing 0.5µg/ml ethidium bromide) probed with a ssDNA probe specific to TGMV DNA 2 virion sense..
Gel purified TGMV genomic-length ccc dsDNA digested with (1) *Eco*RV, (2) *Hpa*I, (3) *Bam*HI, (4) *Bgl*II; (5) gel purified TGMV genomic-length ccc dsDNA (the faster migrating band is contaminating ssDNA); gel purified TGMV subgenomic ccc dsDNA digested with (6) *Eco*RV, (7) *Hpa*I, (8) *Bam*HI, (9) *Bgl*II; (10) gel purified TGMV subgenomic ccc dsDNA (the faster migrating band is contaminating ssDNA); (11) TGMV crude DNA; (12) TGMV crude DNA digested with DNAse I; (13) and (14) TGMV crude DNA digested with 9 and 4.5 units of S1 nuclease respectively; (15) and (16) TGMV crude DNA digested with 40 and 80 units of Exonuclease III respectively; (16) TGMV crude digested with RNAse A.

oc=open circular dsDNA; L=linear dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1, 2 and 3 refer to genomic length, subgenomic and dimeric (M. Slomka; personal communication) viral DNA, respectively.

of polymorphism in the TGMV DNA 2 sequence, as has also been found for MSV (Howell, 1984) and CLV (Stanley and Gay, 1983). As with the subgenomic DNA, digestion with *Hpa*I failed to linearize a large proportion of the genomic-length ccc dsDNA. This could be due to partial digestion or to further heterogeneity in the DNA 2 sequence. Digestion of genomic-length ccc dsDNA with *Eco*RV and *Bam*HI produced the expected number of bands (two in each case) (Fig.29, lanes 1 and 3).

The results above indicate the presence of a single site for *Bam*HI and *Eco*RV in most of the subgenomic DNA molecules. As can be seen in lane 13 of Fig.29, the previously observed band b (Fig.9A, lane 5) had a similar electrophoretic mobility as *Bam*HI and *Eco*RV linearized subgenomic DNA, indicating that it corresponds to the linear form of the subgenomic dsDNA. A size estimate of *ca.*1270bp was obtained for *Bam*HI linearized gel purified subgenomic ccc dsDNA (Fig.30, lane 3),which is in good agreement with the estimated size (*ca.*1240nt) for the corresponding ssDNA molecule.

Since the cloning vector pEMBL9 (Dente *et al.*, 1983) has a unique *Bam*HI site in the polylinker within the gene β-galactosidase, it was decided to clone the subgenomic DNA at this site. Gel purified subgenomic ccc dsDNA was therefore linearized at its unique *Bam*HI site and ligated to *Bam*HI digested pEMBL9 (section 2.29.3). The ligation mixture was used to transform *E.coli* JM101 cells (2.29.5) and bacterial colonies harbouring recombinant plasmids were selected on selective agar plates (section 2.29.2) by virtue of their white colour (section 2.29.3). As gel-purified subgenomic ccc dsDNA was contaminated with host DNA (as a result of it being purified from agarose gel fractionated TGMV crude dsDNA. Fig.30, lane 4), a further selection proved necessary to descriminate recombinant plasmids carrying the subgenomic DNA from those carrying host DNA inserts. The white bacterial colonies were transferred to gridded agar plates, grown overnight at 37°C, lifted with Pall biodyne A nylon menbranes (section 2.25) and hybridized to [$^{32}$P]-labelled nick-translated TGMV 2 (gel purified full-length inserts of pBH604 excised with restriction enzyme *Cla*I) (Fig.31). Recombinant plasmids from colonies showing strong hybridization signal (as opposed to the weak background signal shown by negative colonies) were isolated (section 2.27) and analysed by agarose gel

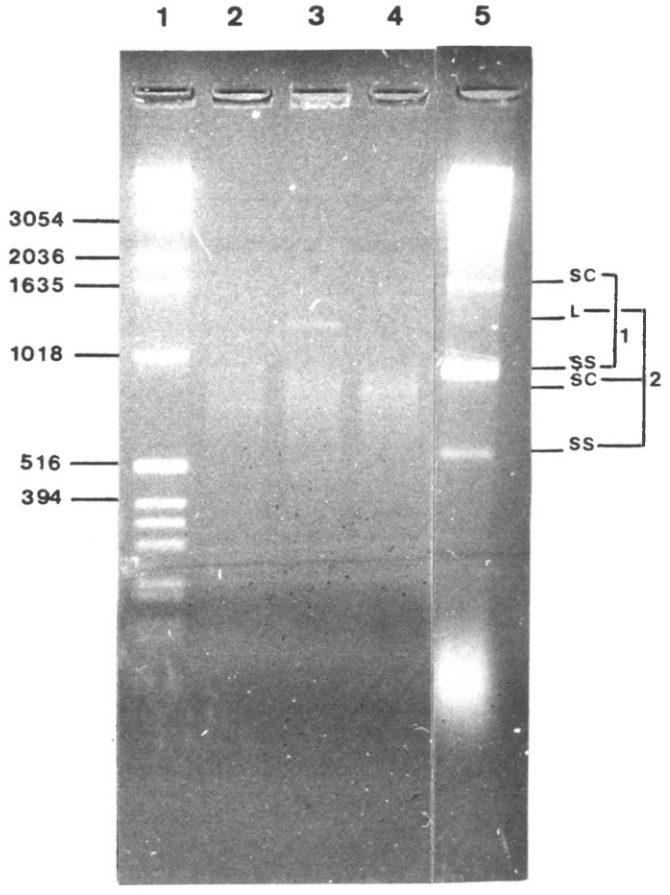



Fig.30. **Size estimation of TGMV subgenomic dsDNA.**

Electrophoresis in a 1.2% agarose gel containing 0.5µg/ml ethidium bromide.
(1) Kb marker (fragment sizes are given in base pairs); (2) and (3) gel purified TGMV subgenomic ccc dsDNA digested with Hpa II (underloaded) and Bam HI respectively; (4) gel purified TGMV subgenomic ccc dsDNA; (5) TGMV crude DNA.

L=linear dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic length and subgenomic viral DNA respectively.

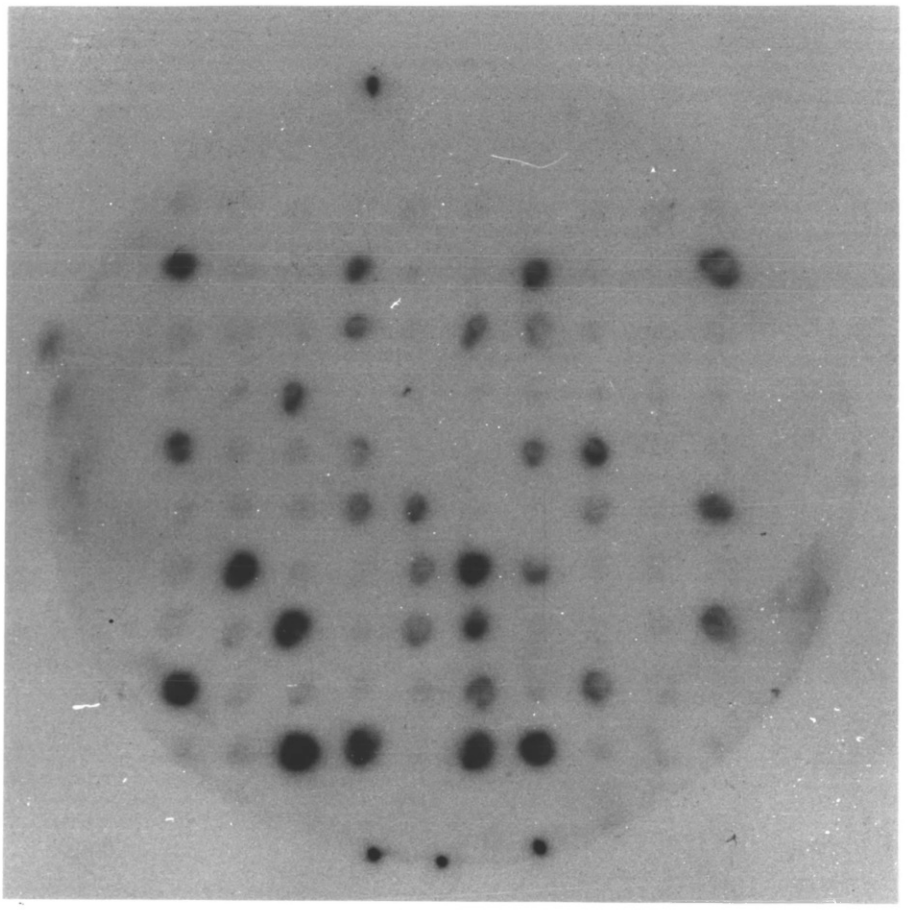


Fig.31. **Selection of recombinant plasmids containing TGMV subgenomic DNA.**

Bacteria (*E.coli* JM101) harbouring recombinant plasmids were transferred to Pall Biodyne A nylon membranes and probed with nick-translated TGMV DNA 2 (pBH604).

electrophoresis before and after digestion with restriction nuclease *Bam*HI (Fig.32A, lanes 3 to 15). All the recombinant plasmids that were analysed (thirty six, six of which are shown in Fig.32A) had the same size insert with an electrophoretic mobility intermediate to that of genomic-length viral ccc dsDNA and ssDNA (Fig.32A, lane 1, 4, 6, 8, 10, 12 and 14). This is indeed the same mobility as linearized subgenomic dsDNA (Fig.30, lane 3). Recombinant plasmids carrying the subgenomic DNA were named pB5 (pB501 to pB536) as the subgenomic ssDNA is commonly referred to as band 5 (Hamilton *et al.*, 1982).

## 3.8 Sequence of the deletion borders

On the basis of the results obtained from Southern blot analysis of the subgenomic ssDNA (section 3.4), the 5' deletion point had been located in the vicinity of a *Pst*I (246) site (Fig.19B). If that site had not been deleted, the subgenomic DNA should contain two *Pst*I sites (246 and 2354). Double-digestion of the pB5 clones with *Bam*HI and *Pst*I should therefore generate three subgenomic DNA derived fragments (two *Bam*HI/*Pst*I fragments and one *Pst*I/*Pst*I fragment) of *ca.* 0.48Kb, 0.41Kb and 0.36Kb, with the deletion border located in the larger fragment. Conversely, only two *Bam*HI/*Pst*I fragments of *ca.* 0.89Kb (0.48Kb + 0.41Kb) and 0.36Kb should be generated if the deletion included the *Pst*I (246) site. When the pB5 clones were double-digested with *Bam*HI and *Pst*I, the majority produced three fragments, indicating that the *Pst*I (246) site was present in these clones (Fig.32B, lanes 1,2,5 and 6, bands m, n and o). However, only two fragments were produced in reactions involving clones pB503, pB504, pB523 and pB536 (Fig.33B, lanes 3 and 4, bands p and m). While at first site this could be taken to mean that the *Pst*I (246) site had been deleted in these clones, further examination showed that this was not the case. Indeed, the larger fragment (band p) produced in these reactions corresponded to the smaller two (bands n and o) of the three fragments expected if the *Pst*I (246) site were present, rather than to the two larger ones (bands m and n), which would be expected if that site were part of the deletion (compare lanes 2 and 3 of Fig.32B). It was concluded that the *Pst*I (2354) site was absent from these clones as a result

Legend to Fig.32. **Restriction analysis of recombinant plasmids containing TGMV subgenomic DNA.**

A) Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.
(1) TGMV crude DNA; (2) pEMBL9 digested with *Bam*HI; (3) pEMBL9; (4), (6), (8), (10), (12) and (14) clones pB501 to pB506 digested with *Bam*HI; (5), (7), (9), (11), (13) and (15) clones pB501 to pB506.

B) Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.
(1) to (6) clones pB501 to pB506 double-digested with *Bam*HI and Pst I; (7) TGMV crude DNA.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic length and subgenomic viral DNA respectively. m, n, o and p are restriction fragments derived from TGMV subgenomic DNA (see section 3.7). The small DNA fragment in lane 10 (Fig.32A) and lane 3 (fig.32B) is probably a *host DNA Bam*HI fragment which was cloned with the subgenomic DNA.

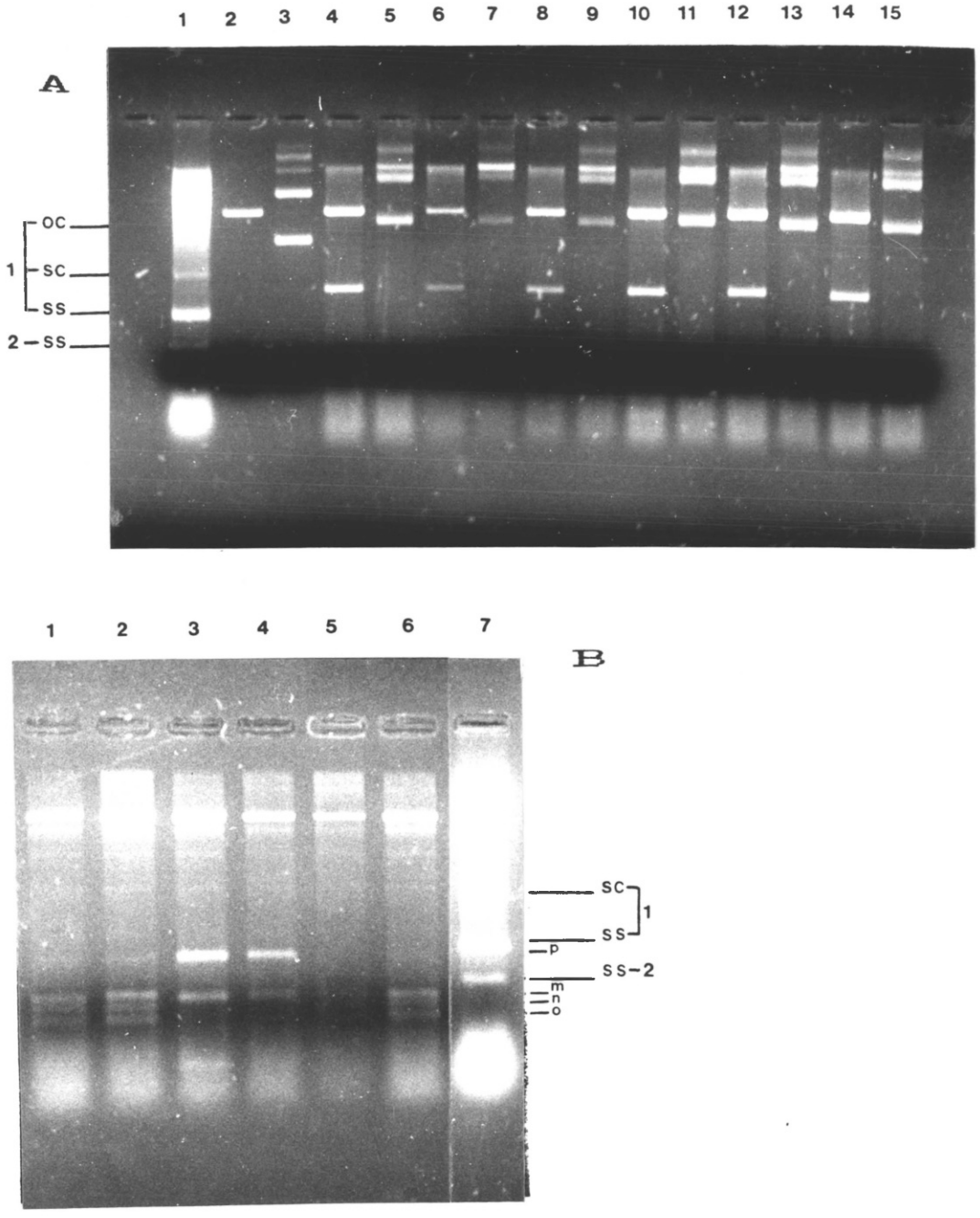
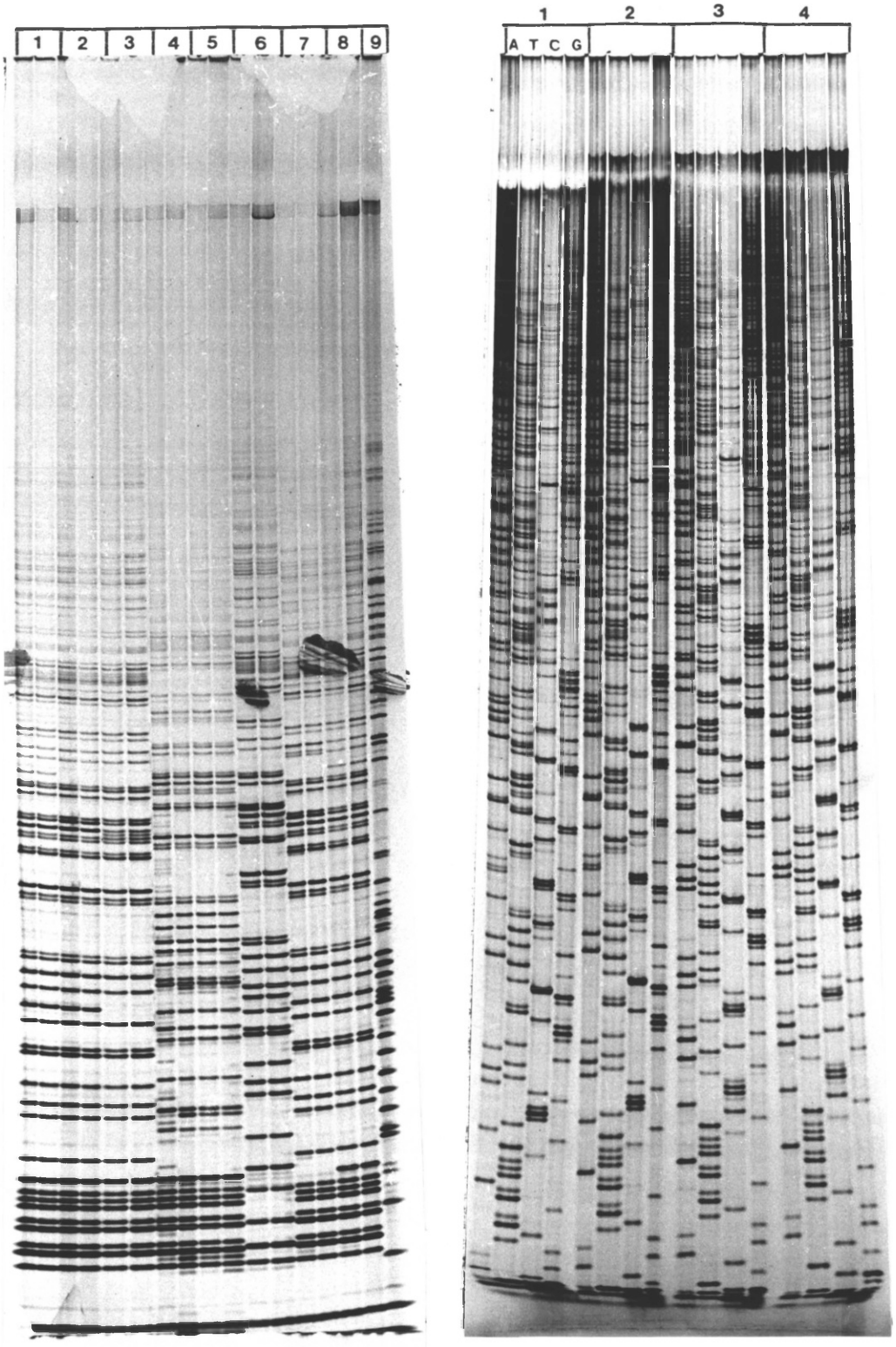

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
A
OC
1
SC
SS
2 – SS
1 2 3 4 5 6 7
B
SC
1
SS
P
SS-2
m
n
o

Fig. 32

Legend to Fig.33. **Sequencing the nucleotides surrounding the deletion points in cloned subgenomic DNAs.**

A) 'T-Tracks' of the *Bam*HI/*Pst*I fragment containing the deletion joint. Fragments were derived from subgenomic DNA clones (1) pB512, (2) pB513, (3) pB514, (4) pB519; (5) pB522, (6) pB523, (7) pB526 and (8) pB527; (9) M13mp8.

B) Sequencing gel of the *Bam*HI/*Pst*I fragment containing the deletion border. Fragments were derived from subgenomic DNA clones (1) pB530, (2) pB531, (3) pB532 and (4) pB534. A, T, C and G reactions are as indicated in (1).




A B

Fig. 33

of sequence heterogeneity and that the deletion occurred downstream of the *Pst*I (246) site in all the examined clones.

For each of twenty two clones of subgenomic DNA, the 0.48Kb *Bam*HI/*Pst*I fragment containing the deletion borders was gel purified (section 2.16.4) and ligated (section 2.29.3) to *Bam*HI/*Pst*I double-digested M13mp8. In M13mp8 (Messing and Vieira, 1982), the *Pst*I site is positioned upstream of the *Bam*HI site relative to the sequencing primer, ensuring that sequencing of the subcloned *Bam*HI/*Pst*I fragments by the dideoxy chain termination method (Sanger *et al.*, 1977) would proceed in the *Pst*I→*Bam*HI direction. The ligation mixtures were used to transform competent JM101 cells and bacterial plaques carrying recombinant phage DNA selected on the basis of colour (section 2.29.5). For each clone of subgenomic DNA, at least two white plaques were picked and recombinant DNA template for sequencing prepared as described in section 2.30. To determine that they contained the right inserts, the recombinant DNA templates were 'T-tracked' (section 2.31.3) (Fig.33A) and the resulting patterns of 'Ts' compared to that of the TGMV DNA 2 sequence. The first 130-140 bases of the *Bam*HI/*Pst*I fragment of each of the twenty two cloned subgenomic DNAs were then sequenced by the dideoxy chain termination method (section 2.31.2) (Fig.33B).

In all twenty two clones, the deletion borders were located within bases downstream of the *Pst*I (246) site (Fig.34). Four different types of deletion borders were obtained, derived from four different 5' borders and three 3' borders. The location of the deletion borders was highly specific, with the 5' deletion borders being located within a region of ten nucleotides (255-265) and the 3' deletion borders within a region of eleven nucleotides (1529-1540). Of the twenty two clones examined, two had deletion 1 (256-1528), three had deletion 2 (262-1537), ten had deletion 3 (264-1539) and seven had deletion 4 (266-1528) (Fig.34). The region deleted in the cloned subgenomic dsDNA comprises the whole of ORF BR1 and the C-terminal region of ORF BL1 and is clearly similar to that deleted in the subgenomic ssDNA (sections 3.4 and 3.5) (Fig.39). Assuming that no other deletions or insertions had occurred, the predicted nucleotide lengths of the four clone types would be 1249nt (type 1), 1246nt (type 2), 1246nt (type 3) and 1259nt (type 4). These lengths

Legend to Fig.34. **Nucleotide sequence at the deletion points of 22 cloned subgenomic DNAs.**

Sequences are shown in the virion sense. TGMV 2 (1) is the published sequence (Hamilton *et al.*, 1984) and TGMV 2 (2) is the revised sequence. Types 1, 2, 3 and 4 clones are separated by gaps (in that order). Polymorphisms in the subgenomic DNAs are indicated in boxes.

```
          242       250         260         1529        1540        1550
TGMV 2(1) CTGCAGAGTCTTGTTATTTCTGCA.....CCCCTGATCAGATCCCGCATATGATGG
TGMV 2(2) CTGCAGAGTCTTGTTATTTCTGCA.....CCCCTGATCAGATCCCGCATATGATGG

pB504     CTGCAGAGTgTgGT                CCCCTGATCAGATCCCGCATATGATGG
pB523     CTGCAGAGTgTgGT                CCCCTGATCAGATCCCGCATATGATGG

pB501     CTGCAGAGTCTTGTTATTTC                AGATCCCGCATATGATGG
pB526     CTGCAGAGTCTTGTTATTTC                AGATCCCGCATATGATGG
pB528     CTGCAGAGTCTTGTTATTTC                AGATCCCGCATATGATGG

pB503     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB513     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB514     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB511     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB529     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB531     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB507     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB512     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB527     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG
pB530     CTGCAGAGTCTTGTTATTTCTG                ATCCCGCATATGATGG

pB502     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
pB508     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
pB519     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
pB520     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
pB522     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
pB532     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
pB534     CTGCAGAGTCTTGTTATTTCTGCA      CCCCTGATCAGATCCCGCATATGATGG
```

```
                1560       1570       1580       1590       1600       1610
TGMV 2(1)    TGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCCGAGGCCTTAATTCTATT
TGMV 2(2) GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT

pB504     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB523     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT

pB501     GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCa AGGCCTTAATTCTATT
pB526     GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCa AGGCCTTAATTCTATT
pB528     GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCa AGGCCTTAATTCTATT

pB503     GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB513     GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGaCCTTAATTCTATT
pB514     GCTGGGCCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTtTATT
pB511     GCTGGGCCCTAcAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTtTATT
pB529     GCTGGGCCCTgTAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB531     GCTGGGCCCTgTAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB507     GCTGGGCCCTgTAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB512     GCTGGGCCCTgTAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB527     GCTGGGCCCTgTAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB530     GCTGGGCCCTgTAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT

pB502     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB508     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB519     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB520     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB522     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB532     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
pB534     GCTGGGtCCTATAGTACTCCTTGAGGCCCAAGATTCTCC AGGCCTTAATTCTATT
```

```
                  1620       1630       1640
TGMV 2(1) TATTGGGCCTGTTAG            ATGTG
TGMV 2(2) TATTGGGCCTGTTAGCCCAATAGTAGATGTG

pB504     TATTGGGCCTGTTAGCCCAATAGTAGATGTG
pB523     TATTGGGCCTGTTAGCCCAATAGTAGATGTG

pB501     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB526     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB528     TATTGGGCCTGTTAGCCCAATtGTAGATGTG

pB503     TATTGGGCCTGTTAGCCCAATAGTAGATGTG
pB513     TATTGGGCCTGTTAGCCCAATAGTAGATGTG
pB514     TATTGGGCCTGTTAGCCCAgTAGTAGATGTG
pB511     TATTGGGCCTGTTAGCCCAgTAGTAGATGTG
pB529     TATTGGGCCTGTTAGCCCAATAGTAGATGTG
pB531     TATTGGGCCTGTTAGCCCAATAGTAGATGTG
pB507     TATTGGGCCTGTTAGtCCAATAGTAGATGTG
pB512     TATTGGGCCTGTTAGtCCAATAGTAGATGTG
pB527     TATTGGGCCTGTTAGtCCAATAGTAGATGTG
pB530     TATTGGGCCTGTTAGtCCAATAGTAGATGTG

pB502     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB508     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB519     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB520     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB522     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB532     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
pB534     TATTGGGCCTGTTAGCCCAATtGTAGATGTG
```

correspond to just under half that of the revised TGMV 2 sequence (2522nt; see below and section 3.9).

Comparison of the nucleotide sequences obtained for the *Bam*HI/*Pst*I fragments of the twenty two ✓ clones of subgenomic DNA to that of TGMV DNA 2 (Hamilton *et al.*, 1984; Fig.34, TGMV 2(1)) revealed a number of differences other than the major deletion, including point mutations and small insertions and deletions (Fig.34). On reexamination of the TGMV 2 sequencing gels (Hamilton W.D.O. and Stein V.), however, all the observed insertions and deletions in this limited region were found to be the result of sequence-reading errors. As a result of the corrections made to the TGMV 2 sequence, the C-terminus of ORF BL1 was extended to nucleotide 1502 in the revised sequence (Fig.38), giving a new coding capacity of 26,466d (234 amino acids; Fig.35). This is still smaller than those of the corresponding ORFs in DNAs 2 of CLV and BGMV (33.1Kd and 33.7Kd respectively). However, exhaustive searches failed to reveal further reading-errors which might be responsible for the truncation of the TGMV ORF relative to those of CLV and BGMV.

As shown in Fig.34, there are a number of point mutations, including transitions and transversions, between the revised TGMV 2 sequence (TGMV 2 (2)) and the twenty two ✓ subgenomic sequences. These probably represent polymorphisms in the populations of both the subgenomic DNA and DNA 2, as had been previously observed (sections 3.7 and 3.8; I.T.D. Petty, personal communication). Interestingly, the polymorphisms within the twenty two ✓ examined clones appeared to be specific to each of the four deletion types. Thus, type 1, 2 and 4 clones each had their own characteristic pattern of polymorphisms, suggesting that each of these types were derived from a single original event. In type 3 clones, however, the situation appeared more complex, as these showed intratypical polymorphisms. From the limited data available, type 3 clones could be subdivided into four groups in which group 1 consisted of clones pB503 and pB513, group 2 of clones pB514 and pB511, group 3 of clones pB529 and pB531 and group 4 of clones pB507, pB512, pB527 and pB530. It is not known however, whether each of these groups originated separately or whether they are the result of genetic drift in a population of molecules originated from a single event.

Legend to Fig.35. **Alignment of the predicted amino acid sequences encoded by TGMV B and subgenomic DNAs ORF BL1.**

The amino acid sequences were determined with the programme 'translate' and aligned with the programme PRETTY (Deveraux *et al.*, 1984). For types 1 to 4 cloned subgenomic DNA, the sequences were determined using the TGMV 2 DNA sequence. Amino acids which differ from those in the TGMV 2 ORF BL1 are shown in small case.

```
          1                                                            50
tgmvb.bl1 MDSQLACPPN AFNYIESNRD EYQLSHDLTE IILQFPSTAS QLSARLSRSC
pb530.bl1 MDSQLACPPN AFNYIESNRD EYQLSHHLTE IILQFPSTAS QLSAkLSRSC
type1.bl1 MDSQLACPPN AFNYIESNRD EYQLSHDLTE IILQFPSTAS QLSARLSRSC
type2.bl1 MDSQLACPPN AFNYIESNRD EYQLSHDLTE IILQFPSTAS QLSARLSRSC
type3.bl1 MDSQLACPPN AFNYIESNRD EYQLSHDLTE IILQFPSTAS QLSARLSRSC
type4.bl1 MDSQLACPPN AFNYIESNRD EYQLSHDLTE IILQFPSTAS QLSARLSRSC

          51                                                          100
tgmvb.bl1 MKIDHCVIEF RQQVPINATG SVVVEIHDKR MTDNESLQAS WTFPVRCNID
pb530.bl1 MKIDH*VIEF RQQVPINATG SVVVEIHDKR MTDNESLQAS WTFPVRCNID
type1.bl1 MKIDHCVIEF RQQVPINATG SVVVEIHDKR MTDNESLQAS WTFPVRCNID
type2.bl1 MKIDHCVIEF RQQVPINATG SVVVEIHDKR MTDNESLQAS WTFPVRCNID
type3.bl1 MKIDHCVIEF RQQVPINATG SVVVEIHDKR MTDNESLQAS WTFPVRCNID
type4.bl1 MKIDHCVIEF RQQVPINATG SVVVEIHDKR MTDNESLQAS WTFPVRCNID

          101                                                         150
tgmvb.bl1 LHYFSSSFFS LKDPIPWKLY YKVCDSNVHQ RTHFAKFKGK LKLSTAKHSV
pb530.bl1 LHYFSSSFFS LKDPIP1KLY YKVCDSNVHQ RTHFAKFKGK LKLSTAKHSV
type1.bl1 LHYFSSSFFS LKDPIPWKLY YKVCDSNVHQ RTHFAKFKGK LKLSTAKHSV
type2.bl1 LHYFSSSFFS LKDPIPWKLY YKVCDSNVHQ RTHFAKFKGK LKLSTAKHSV
type3.bl1 LHYFSSSFFS LKDPIPWKLY YKVCDSNVHQ RTHFAKFKGK LKLSTAKHSV
type4.bl1 LHYFSSSFFS LKDPIPWKLY YKVCDSNVHQ RTHFAKFKGK LKLSTAKHSV

          151                                                         200
tgmvb.bl1 DIPFRAPTVK ILSKQFTDKD VDFSHVGYGK WDRKMIRSAS TSTIGLTGPI
pb530.bl1 DIPFRAPTVK ILSKQiTvKD VDFSHVGYGK WDRKMIRSAS TSTIGLTGPI
type1.bl1 DIPFRAPTVK ILSKQFTDKD VDFSHVGYGK WDRKMIRSAS TSTIGLTGPI
type2.bl1 DIPFRAPTVK ILSKQFTDKD VDFSHVGYGK WDRKMIRSAS TSTIGLTGPI
type3.bl1 DIPFRAPTVK ILSKQFTDKD VDFSHVGYGK WDRKMIRSAS TSTIGLTGPI
type4.bl1 DIPFRAPTVK ILSKQFTDKD VDFSHVGYGK WDRKMIRSAS TSTIGLTGPI

          201                                                         250
tgmvb.bl1 ELRPGESWAS RSTIGPSPSY AGSDQGDAMH RIST*..... ..........
pb530.bl1 ELRPGESWAS RSTtGPSPSY AGSeitrlcs lnmig*.... ..........
type1.bl1 ELRPGESWAS RSTIGPSPSY AGSDQGtrlc slnmig*... ..........
type2.bl1 ELRPGESWAS RSTIGPSPSY AGSeitrlcs lnmig*.... ..........
type3.bl1 ELRPGESWAS RSTIGPSPSY AGSeitrlcs lnmig*.... ..........
type4.bl1 ELRPGESWAS RSTIGPSPSY AGSDQGcrnn ktlqlkydwl tqgdhtkwpt

          251                                  283
tgmvb.bl1 .......... .......... .......... ...
pb530.bl1 .......... .......... .......... ...
type1.bl1 .......... .......... .......... ...
type2.bl1 .......... .......... .......... ...
type3.bl1 .......... .......... .......... ...
type4.bl1 mrrnsrtrvg rtiarpsgni ktdgrhvplr nf*
```

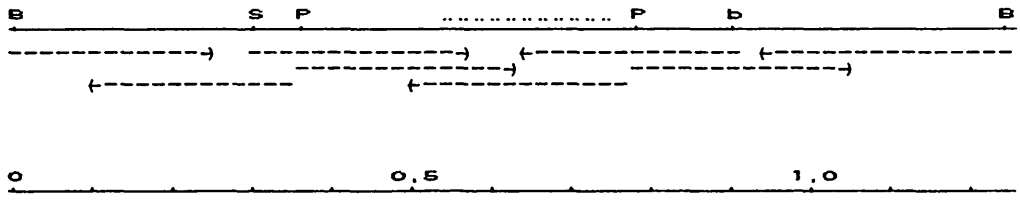
Fig. 35

Type 1 clones had two polymorphisms close to the deletion border, at nucleotides 251 and 253 (Fig.35). The sequence surrounding the deletion joint in these clones was therefore GTGTGGTCCCCT. Interestingly, the same sequence is present 42nt upstream, at nucleotides 207-218 in the TGMV 2 sequence, and a similar sequence GTGGTTCCC is present 42nt upstream of the right hand deletion border, at nucleotides 1487-1495. The significance of this, if any, is not known.

## 3.9 Sequence of pB530

To determine whether other insertions or deletions had occurred, it was decided to sequence one clone of subgenomic DNA in its entirety. Clone pB530, a representative of the most abundant deletion type (type 3) was chosen as candidate.

The sequencing strategy is shown below,




where the continuous line represents the cloned subgenomic DNA, the dotted line shows the location of the 'common region and the broken arrows indicate the direction and extent of sequence obtained for each clone. Restriction enzyme sites for *Bam*HI (B), *Sph*I (S), *Pst*I (P) and *Bcl*I (b) are indicated, and the scale indicates the distances in Kb from the *Bam*HI site. The strategy consisted in subcloning fragments generated by *Bam*HI/*Pst*I and *Sph*I/*Sca*I double-digests into M13mp18 and mp19 (digested with *Pst*I and double-digested with *Bam*HI and *Pst*I in the first instance and double-digested with *Sph*I and *Sma*I in the second). As M13mp18 and mp19 have the cloning sites in opposite orientation relative to the sequencing

primer, each fragment could be sequenced in both directions, allowing >50% of the subgenomic sequence to be read in both orientations. To minimize the number of fragments to be screened, each fragment was separately eluted from agarose gels (section 2.16.4) and cloned (section 2.29). Recombinant template DNA was prepared as described (section 2.30), screened by "T-track" (section 2.31.3) (Fig.36) and sequenced by the dideoxy chain termination method (section 2.31) (Fig.37).

A comparison of the pB530 (1246nt) and the revised TGMV 2 (2522nt) DNA sequences is shown in Fig.38. In addition to the major deletion, the subgenomic DNA sequence revealed twenty six mutations, including seven transitions, seventeentransversions, one deletion and one insertion, compared to the corresponding region of the sequence of DNA 2. Interestingly, the deletion (a G residue at position 129 in the TGMV 2 sequence) occurred at the position in the stem just 5' to the loop of the hairpin structure which is highly conserved in the bipartite geminiviruses (Hamilton *etal.*, 1984; Stanley and Gay, 1983; Howarth *et al.*, 1985), resulting in an increase of one nucleotide in the loop. The insertion (of a G residue), on the other hand, occurred between nucleotides 2222 and 2223 in the TGMV 2 sequence, 16nt upstream of the start of ORF BL1 in the complementary sense. As commented previously (section 3.8), these differences probably represent polymorphisms in the populations of both the subgenomic DNA and DNA 2. As the subgenomic DNA does not contain any complete ORFs and, even in the event of it being expressed *in vivo*, its gene products would probably not be required for its proliferation, the only regions of its DNA sequence to be under selective pressure are probably the origins of replication and packaging. Therefore, mutations in the population of subgenomic DNA would be expected to accumulate at a higher rate than in the DNA 2 population. In this context, it is noteworthy that a mutation at nucleotide 762 in clone pB530 generates an in phase stop codon at nucleotides 764-762 (corresponding to nucleotides 2041-2039 in the TGMV 2 sequence) in ORF BL1, such that it could produce a protein of only 55 amino acids (Fig.35). Initiation from a second in phase ATG triplet (at nucleotides 690-692), however, could produce a protein of 155 amino acids (Fig.35). If the mutation causing the in phase stop codon had not been present, translation of clone pB530 ORF BL1

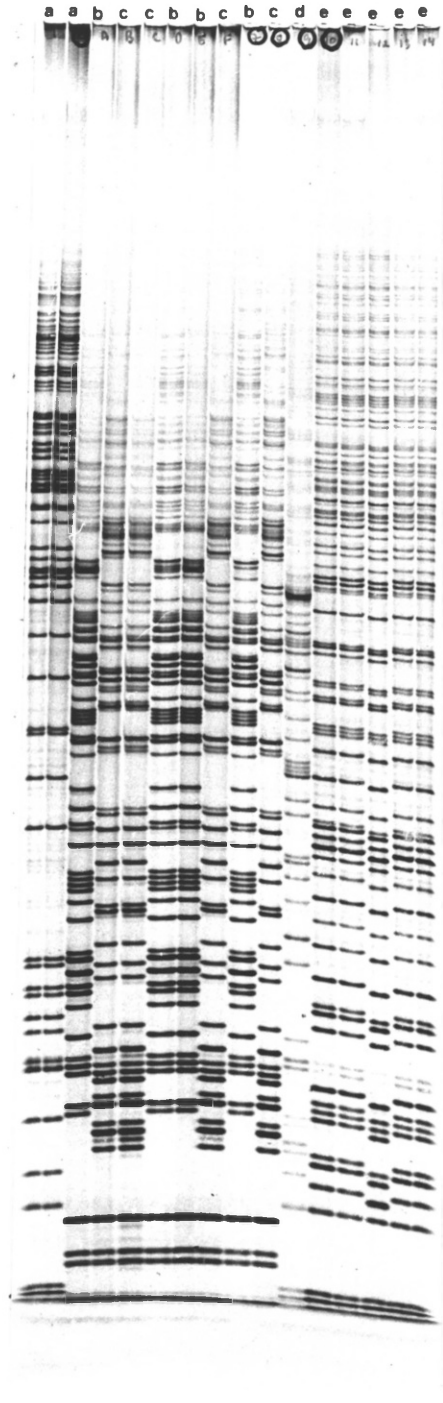



Fig.36. **Sequencing pB530.**

'T-tracks' of recombinant M13 clones containing restriction fragments derived from double-digestion of pB530 with *Bam*HI and *Pst*I.

(a) Fragment n cloned in M13mp9; (b) fragment o clomed in M13mp8; (c) fragment m cloned in M13 mp8; (d) fragment m cloned in M13mp9, (e) fragment o cloned in M13mp9. (see section 3.9 and Fig.32B for details).

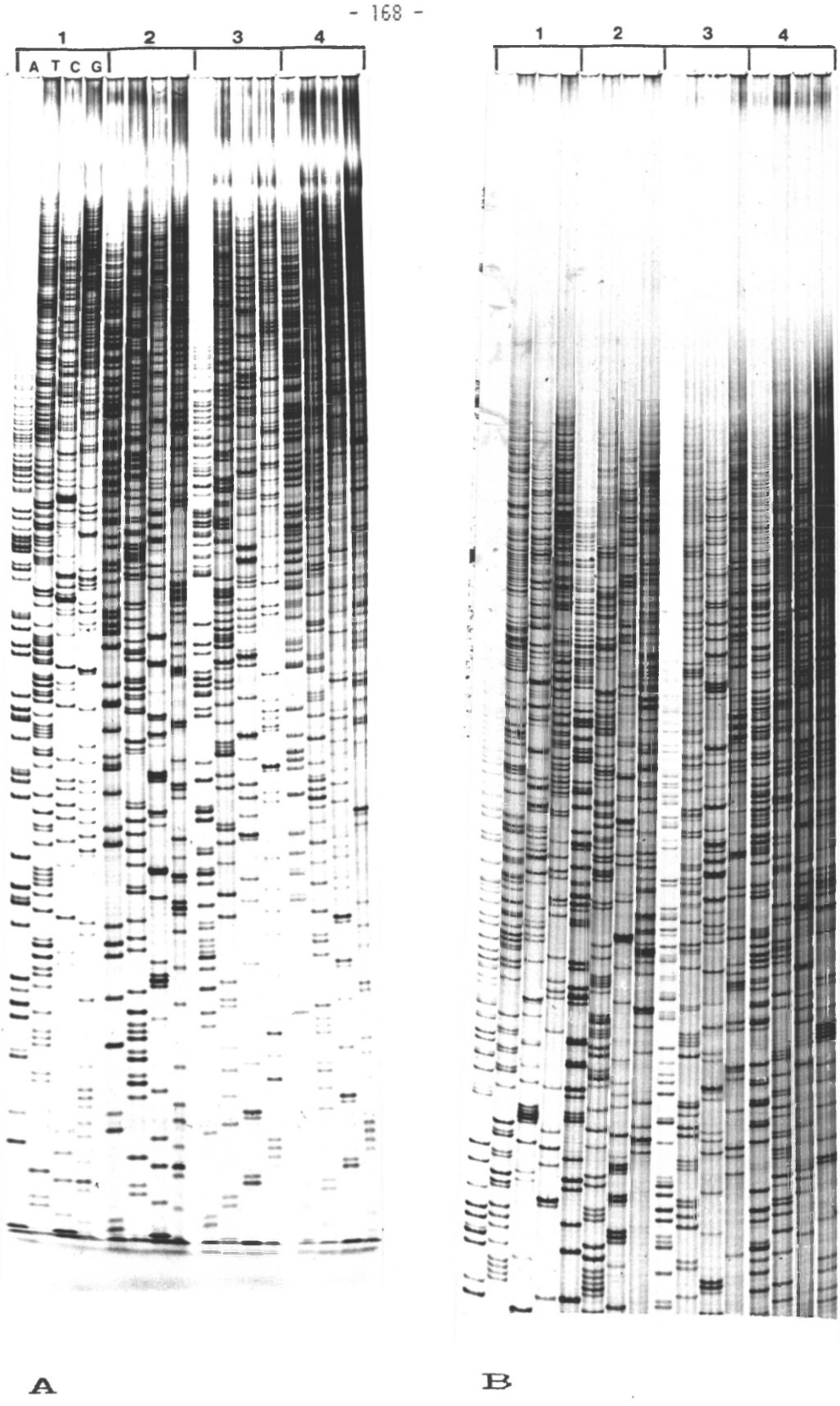



Fig.37. **Sequencing pB530.**

Sequencing gels of pB530 derived *Bam*HI/*Pst*I fragments (1) Fragment o cloned in M13mp8; (2) fragment m cloned in M13mp8; (3) fragment m cloned in M13mp9; (2) fragment o cloned in M13mp9. A, T, C and G reactions are as indicated in (1). A) 2hr run, B) 4hr run. (see section 3.9 and Fig.32B for details).

could produce a protein of 235 amino acids, i.e. one more than the TGMV 2 ORF BL1 (Fig.35). This is because the next in phase termination codon occurs downstream (in the complementary sense) of the deletion joint, at nucleotides 224-222 in the subgenomic sequence. If such protein were produced, the last twelve amino acids at its C-terminus would be different from those of the protein coded by the TGMV 2 ORF BL1 (Fig.35). Also, this protein would contain four additional amino acid changes as a result of point mutations in the sequence of clone pB530 compared to that of TGMV 2 (Fig.35).

Most eukaryotic mRNAs have a sequence of polyadenylic acid at their 3' end. This is added by the enzyme poly-A polymerase. The signal for the addition of the adenosine residues is carried in the transcript in the form of a polyadenylation signal G/ATAA (Messing *et al.*, 1983) which is usually about twenty nucleotides from the translation termination codon. However, the putative polyadenylation signal for TGMV ORF BL1 (at nucleotides 1331-1326; Hamilton *et al.*, 1984) is deleted in the subgenomic DNA. The next possible polyadenylation signal conforming to the consensus sequence G/ATAA (Messing *et al.*, 1983) occurs at nucleotides 1196-1192 (GATAA) in the subgenomic DNA sequence, 257nt downstream of the translation termination codon of clone pB530 ORF BL1. To contain this polyadenylation signal, the transcript for the subgenomic DNA ORF BL1 would have to encompass the 200 base 'common region'. Whether the putative hairpin structure present in this region would present an obstacle to transcription is not known. However, since polyadenylation does not appear to be absolutely essential for translation, the subgenomic DNA ORF BL1 could possibly be expressed in the absence of a poly(A) tail.

For all four subgenomic DNA clone types, translation of ORF BL1 could produce a protein larger than that produced by full-length DNA B assuming no further mutations to produce stop codons or frameshifts (Fig.35). In the case of type 4 clones, that protein could have an additional fourty nine amino acids. Whether these proteins are expressed and what effect they might have on the virus infection cycle, however, remains to be investigated.

Legend to Fig.38. **Comparison of the TGMV 2 DNA and TGMV subgenomic DNA sequences.**

The sequences were aligned with the program 'zip' (Deveraux *et al.*, 1984). Gaps were inserted by the program to increase the similarity. The TGMV 2 DNA sequence was determined from pBH604 (Hamilton *et al.*, 1984) and subsequently revised in this thesis. The subgenomic DNA sequence was determined from pB530.

Comparison of TGMV 2 (pBH604) and TGMV subgenomic DNA (pB530)

```
   1 GAGGTGATGGCATTTTGGTAATTAGAAGGGTTACTACCATTTGGTTTGGG 50
     !!!!! !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
   1 GAGGTAATGGCATTTTGGTAATTAGAAGGGTTACTACCATTTGGTTTGGG 50

  51 GCTACAAAAGTTATATGAATTGGTAGTAAGGTAGCTCTTATATATTAGAA 100
     !!! !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
  51 GCTTCAAAAGTTATATGAATTGGTAGTAAGGTAGCTCTTATATATTAGAA 100

 101 GTTCCTAAGGGGCACGTGGCGGCCATCCGTTTTAATATTACCGGATGGCC 150
     !!!!!!!!!!!!!!!!!!!!!!!!!!!  !!!!!!!!!!!!!!!!!!!
 101 GTTCCTAAGGGGCACGTGGCGGCCATCC.ATTTAATATTACCGGATGGCC 149

 151 GCGCGATCGTCCTCCCGACCCGTGTCCGTGAATTGCGCCGCATTGTCGGC 200
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 150 GCGCGATCGTCCTCCCGACCCGTGTCCGTGAATTGCGCCGCATTGTCGGC 199

 201 CACTTGGTGTGGTCCCCTTGTGTTAACCAATCATATTTAAGCTGCAGAGT 250
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 200 CACTTGGTGTGGTCCCCTTGTGTTAACCAATCATATTTAAGCTGCAGAGT 249

 251 CTTGTTATTTCTGCACTCATTAACTGGTCCCTTTAATTTGAAATATCTTT 300
     !!!!!!!!!!!!
 250 CTTGTTATTTCTG..................................... 262

                                 .
                                 .
                                 .

1490 GTTCCCAACCTATTAAGTGCTTATACGATGCATTGCATCCCCCTGATCAG 1539

 262 ..................................................  262

1540 ATCCCGCATATGATGGGCTTGGCCCTATAGTACTCCTTGAGGCCCAAGAT 1589
     !!!!!!!!!!!!!!!!!!!!!!!!!!! !!!!!!!!!!!!!!!!!!!!!
 263 ATCCCGCATATGATGGGCTTGGCCCTGTAGTACTCCTTGAGGCCCAAGAT 312

1590 TCTCCAGGCCTTAATTCTATTGGGCCTGTTAGCCCAATAGTAGATGTGGA 1639
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 313 TCTCCAGGCCTTAATTCTATTGGGCCTGTTAGTCCAATAGTAGATGTGGA 362

1640 GGCGGATCTGATCATTTTTCTATCCCATTTCCCATATCCCACGTGGCTGA 1689
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 363 GGCGGATCTGATCATTTTTCTATCCCATTTCCCATATCCCACGTGGCTGA 412

1690 AATCGACATCTTTATCTGTAAATTGTTTGGACAATATTTTGACAGTGGGT 1739
     !!!!!!!!!!!!!! !!!!!! !!!!!!!!!!!!!!!!!!!!!!!!!!!
 413 AATCGACATCTTTAACTGTAATTTGTTTGGACAATATTTTGACAGTGGGT 462

1740 GCCCGGAAAGGGATATCAACGGAGTGTTTAGCTGTCGATAATTTCAGCTT 1789
     !!!!!!!!!!!!!!!!!!!!!!! !!!!!!!!!!!!!!!!!!!!!!!!!
 463 GCCCGGAAAGGGATATCAACGGAATGTTTAGCTGTCGATAATTTCAGCTT 512

1790 CCCTTTGAATTTCGCAAAATGAGTCCTCTGGTGAACATTAGAGTCGCAAA 1839
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 513 CCCTTTGAATTTCGCAAAATGAGTCCTCTGGTGAACATTAGAGTCGCAAA 562
```

```
1840 CTTTGTAATATAGTTTCCATGGGATTGGGTCTTTGAGCGAGAAGAACGAC 1889
     !!!!!!!!!!!!!!!!! !!!!!!!!!!!!!!!!!!!!!!! !!!!!!!!!
 563 CTTTGTAATATAGTTTCAATGGGATTGGGTCTTTGAGCGAAAAGAACGAC 612

1890 GATGAGAAATAGTGGAGATCTATATTGCATCTCACCGGAAAAGTCCATGA 1939
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 613 GATGAGAAATAGTGGAGATCTATATTGCATCTCACCGGAAAAGTCCATGA 662

1940 CGCTTGTAAGGATTCATTGTCAGTCATTCTTTTGTCATGAATCTCCACGA 1989
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 663 CGCTTGTAAGGATTCATTGTCAGTCATTCTTTTGTCATGAATCTCCACGA 712

1990 CCACGGATCCTGTTGCGTTTATCGGAACCTGTTGCCTGAACTCAATCACA 2039
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 713 CCACGGATCCTGTTGCGTTTATCGGAACCTGTTGCCTGAACTCAATCACT 762

2040 CAGTGGTCTATCTTCATACAGCTACGGCTCAGTCTGGCGCTTAATTGAGA 2089
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! !!!!!!!!!!!!!!!!
 763 CAGTGGTCTATCTTCATACAGCTACGGCTCAGTTTGGCGCTTAATTGAGA 812

2090 AGCTGTTGACGGAAACTGCAAGATTATCTCAGTCAAGTCATGAGATAACT 2139
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 813 AGCTGTTGACGGAAACTGCAAGATTATCTCAGTCAAGTGATGAGATAACT 862

2140 GATATTCATCTCGGTTTGATTCAATGTAATTGAATGCATTTGGGGGACAA 2189
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 863 GATATTCATCTCGGTTTGATTCAATGTAATTGAATGCATTTGGGGGACAA 912

2190 GCTAACTGAGAATCCATATATTATGAAGACCT.GCCTCGCAGAGGCAGCG 2238
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! ! !! !!!!!!!!!!!!!!!!!
 913 GCTAACTGAGAATCCATATATTATGAAAATCTGGCCTCGCAGAGGCAGCG 962

2239 TTTCACTGAAAATAATAAGCCAAGAGAATAGCTATGAAATTCAAGCCTTG 2288
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
 963 TTTCACTGAAAATAATAAGCCAAGAGAATAGCTATTAAAATCAAGCATTG 1012

2289 CTGCCGGCAGCAACGAACTGAAAATATTAGCTCAAGAGAATAGCTATGAA 2338
     !!!!  !!!!!!!!!! !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1013 CTGCATGCAGCAACGACCTGAAAATATTAGCTCAAGAGAATATCTATGAA 1062

2339 ATTCAACCCTCGCTGCAGGCAATGAGGAACTGAAATACTAACAGAAAATA 2388
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1063 ATTCAACCCTCGCTGCAGGCAATGAGGAACTGAAATACTAACAGAAAATA 1112

2389 ATCGTTCAGGAAAAATAAAAGAAGATATTAAGCCTAATAATTTAGTAGCC 2438
     !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1113 ATCGTTCAGGAAAAATAAAAGAAGATATTAAGCCTAATAATTTAGTAGCC 1162

2439 ACATAGCTAAGAAACTTGTCAAGAGATAATTATCATATGTCGGCGTAGAA 2488
     !!!!!!!!!!!!!!!!!!! !!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
1163 ACATAGCTAAGAAACTTGTTAAGAGATAATTATCATATGTCGGCGTAGAA 1212

2489 CTGGAAATGGGTAGCATATATATAAAACCCTAAT 2522
     !!!!!!!!!!!!!!!!!!! !!!!!!!! !!!!!
1213 CTGGAAATGGGTAGCATATTTATAAAACACTAAT 1246
```

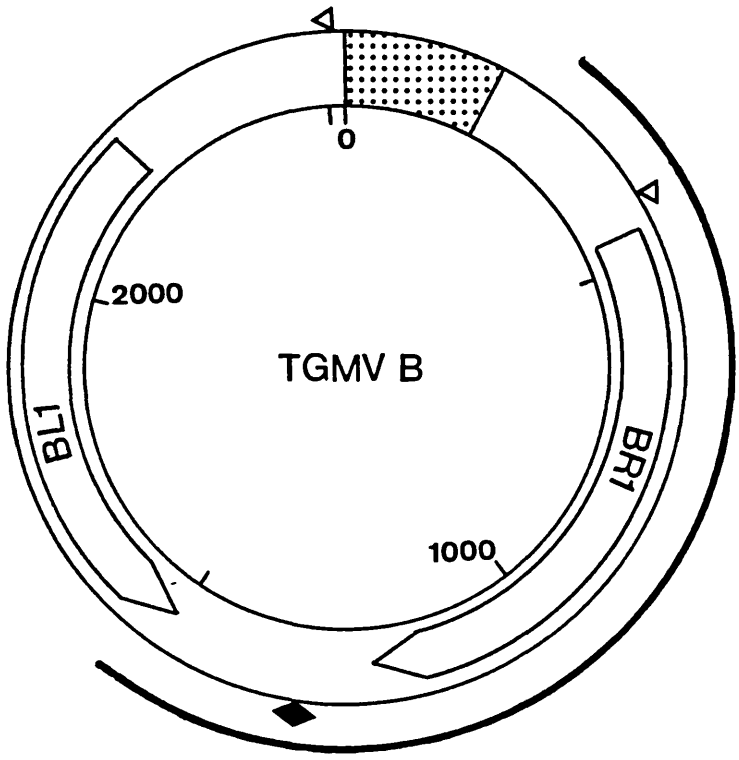



Fig.39. **TGMV 2 DNA showing the region deleted in the subgenomic DNAs.**

The major ORFs are shown as open arrows. Open triangles indicate the position of TATA boxes and shaded triangles the position of putative polyadenylation signals. The stippled area shows the location of the 200bp 'common region' and the heavy arc shows the location of the deletion.

## 3.10 Comparison to other geminiviruses

The subgenomic DNA species of TGMV and CLV (Stanley and Townsend, 1985) share a number of features in common. In each case, the single-stranded form of the subgenomic DNA is encapsidated and intracellular double-stranded forms have been identified. These have been shown to be derived exclusively from DNA 2 by a deletion of approximately half the genome length. For both viruses, sequencing of cloned subgenomic DNA revealed four different types of deletion, all of which occurred in a similar region of DNA 2, completely deleting TGMV ORF BR1 and its CLV counterpart (ORF 29.3) and removing the C-terminus of TGMV ORF BL1 and of the corresponding CLV ORF (ORF 33.7).

While in CLV sixteen of the nineteen clones examined contained the same deletion, in TGMV the number of clones representing each deletion type appeared to be more unevenly distributed. Unlike TGMV (see section 3.13), CLV contains small direct repeats at the deletion points of all four characterized defective types (Table 2). Stanley and Townsend (1985) suggested that these repeats may contribute to the exact location of the recombination event, which could account for the greater predominance of the most abundant clone type in CLV, since it contains the largest repeat. On the other hand, the absence of repeats in TGMV subgenomic DNA could explain why the number of representatives of each clone type is more evenly distributed than in CLV.

There appears to be a greater degree of specificity in TGMV than in CLV with respect to both the location of the deletion borders and the size of the subgenomic DNA molecules. In TGMV, the left hand and right hand deletion borders were located within regions of 10 and 11 nucleotides respectively, giving rise to a narrow size range (1246-1259 nucleotides) (section 3.7) for the population of clones examined. In CLV, on the other hand, the left hand and right hand deletion borders were located within regions of 195 and 116 nucleotides respectively, giving rise to a comparatively wider size range (1294-1343 nucleotides).

The narrow size range of both TGMV and CLV subgenomic DNA suggests that a stringent size selection is imposed on the subgenomic molecules. As the size of CLV subgenomic DNA is

approximately half that of the genomic DNAs, Stanley and Townsend (1985) suggested that individual subgenomic molecules may be encapsidated within half size, quasi isometric particles seen in electron micrographs of CLV preparations. This is supported by the observation that preparations of CLV isometric particles are enriched in subgenomic DNA (J.C.Sequeira, cited in Stanley and Townsend, 1985). Since half size, quasi isometric particles have also been observed in TGMV preparations (Matyis *et al.*, 1975), it is possible that TGMV subgenomic DNA may also be encapsidated in such particles. Analysis of the DNA content in CSMV and BGMV virions indicated the presence of a single genomic DNA molecule per geminate particle (Francki *et al.*, 1980; Goodman *et al.*, 1980). Presumably, the half size, quasi isometric particles could accommodate about half the amount of DNA as the geminate ones. The apparently stringent size selection on the subgenomic DNA could therefore be the result of packaging constraints imposed by it being encapsidated in isometric particles. The narrower size range of TGMV subgenomic DNA compared to CLV subgenomic DNA could reflect more rigorous packaging constraints on the former.

The formation of both TGMV and CLV subgenomic DNAs by deletion of specific regions of DNA 2 is probably a consequence of the need to retain sequences required for their proliferation. Since subgenomic ssDNA molecules are encapsidated, any sequences required for packaging must be retained within the subgenomic DNA. As mentioned previously, the identification of dsDNA forms of the subgenomic species suggests that the latter replicates from its own template. If this is the case, then sequences required for replication must also be retained in the subgenomic molecules. In this respect, it may be significant that the 'common region', which has been postulated to contain the origin(s) of DNA replication (Stanley and Gay, 1983; Hamilton *et al.*, 1984), is retained in the subgenomic DNA of both TGMV and CLV.

Since the deletion is limited to one half of DNA 2, it would appear that at least two regions of this molecule, one on either side of the deletion, must be retained in the subgenomic DNA. Presumably, the 'common region' is one of these regions. The 'common region' contains an inverted repeat sequence capable of

forming a stable hairpin structure (Stanley and Gay, 1983; Hamilton *et al.*, 1984). It has been proposed, by analogy to the primosome assembly site in the bacteriophage ØX174 ssDNA (Arai and Kornberg, 1981), that this structure might function as a start signal for the conversion of ssDNA into dsDNA replicative intermediates (Hamilton *et al.*, 1984; Stanley and Davies, 1985). Primer-like DNA molecules have been found in association with encapsidated MSV (Donson *et al.*, 1984; Howell, 1984) and WDV (H.Macdonald, personal communication) ssDNA. These are complementary to a region on the virion DNA located almost diametrically opposite to a hairpin structure analogous to that found in the 'common regions' of TGMV and CLV (section 3.21). Although in ØX174 the formation of the primosome takes place at the primosome assembly site, priming of DNA synthesis can occur at many different locations. This is thought to be due to movement of the primosome along the ssDNA to the sites at which priming actually occurs (Arai and Kornberg, 1981). A similar situation could be envisaged for the geminiviruses. It is possible therefore, that priming of second-strand synthesis in TGMV and CLV might occur at regions other than the 'common region', even if the hairpin structure in the 'common region' of these viruses functions as a 'primosome' assembly site. The fact that priming of second-strand synthesis in MSV and WDV appears to take place at a single site suggests that sequences around that site may specify its exact location. It may be significant that in both viruses the primer-like molecules occur downstream (in the complementary sense) of inverted repeat sequences (Donson *et al.*, 1984; H.Macdonald, personal communication). Sequence analysis of a clone of WDV subgenomic DNA revealed a deletion of 1211nt. The deletion is located *ca.*80nt downstream of a hairpin structure analogous to that found in the 'common regions' of TGMV and CLV (section 3.21) and just overlaps the 5' end of the priming site at the other extremity (H.Macdonald, personal communication). If priming of second-strand synthesis in DNA 2 of TGMV and CLV took place at a site diametrically opposite to the 'common region', the specificity of the location of the deletion in these viruses could be accounted for by the need to retain the priming site as well as the 'common region'. The difference in

the specificity of the location of the deletion points in TGMV and CLV could be a consequence of the distance between the priming site and the 'common region' in these viruses, in combination with the stringent size selection. However, unlike WDV, no significant inverted repeat sequences, which may be indicative of a primer site, are found immediately downstream of the deletion in either TGMV or CLV.

### 3.11 Vector construction and infectivity studies

The potential of geminiviruses as vectors for the transfer of genes to whole plants as well as isolated plant cells is now beginning to be considered ( Buck and Coutts, 1983; Hull and Davies, 1983; Coutts *et al.*, 1985). One approach to vector construction would be to design vectors based on the subgenomic DNA. The present work indicates that the region of DNA 2 deleted in TGMV subgenomic DNA is not required for replication of the latter in the presence of the two genomic components. It is possible, therefore, that this region may be replaced with a foreign gene without interfering with the ability of the resulting molecule to replicate in the presence of components A and B. A similar approach has been used in the development of early SV40 vectors (Ganem *et al.*, 1976). These workers used defective SV40 DNA molecules consisting of a series of reiterated copies of the origin of DNA replication interspersed with host DNA to propagate a segment of bacteriophage-λ in monkey cells in the presence of wild-type helper virus. This strategy, however, has now been superseded by the use of replacement vectors (reviewed by Subramani and Southern, 1983). In this system, either the late or early region is replaced with foreign DNA and the recombinant genome propagated by complementation with a conditionally defective, temperature-sensitive, helper virus at the nonpermissive temperature. The recombinant provides early functions and the helper virus provides late functions (or *vice versa*) ensuring that the hybrid genome is not outgrown by the helper virus.

### 3.11.1 Infection with cloned subgenomic DNA

The first step towards constructing a plant vector based on the subgenomic DNA was to demonstrate that cloned TGMV subgenomic DNA could replicate in tobacco plants in the presence of both genomic components. When *N. benthamiana* plants were infected with cloned TGMV DNA components 1 and 2, no subgenomic DNA could be detected in the virus progeny formed, nor in intracellular DNA extracts (Hamilton *et al.*, 1983; Fig.40, lane 1). Similarly, no subgenomic DNA could be detected in *N. benthamiana* plants infected with cloned CLV DNAs 1 and 2, even after five passages of virus derived from the cloned DNA inoculum (Stanley and Townsend, 1985). The detection of the subgenomic species in plants infected with an inoculum consisting of cloned subgenomic DNA in combination with both cloned DNA components should therefore provide good indication for replication of the cloned subgenomic DNA in infected plants.

As mentioned above, DNA extracted from *N. benthamiana* plants infected with cloned TGMV DNA components 1 and 2 did not show any detectable subgenomic DNA. Since the infectivity of inocula derived from plants infected with components 1 and 2 (first passage inocula) is considerably higher than that of cloned DNA, it was decided to see whether infection with cloned subgenomic DNA in combination with first passage inoculum would result in replication of the former. As a preliminary test, two *N. benthamiana* plants were inoculated with first passage inoculum (sections 2.9.3 and 2.9.4) in combination with cloned subgenomic DNA (pB530) which had been excised from the vector by digestion with restriction enzyme *Bam*HI. Approximately 25µg of cloned DNA (including the vector) were used for each plant. As a control, two more plants were infected with first passage inoculum alone. Only one of the four plants in this preliminary experiment showed symptoms, fortunately one that had been inoculated with the subgenomic DNA. DNA was extracted from this plant (section 2.10.1), fractionated by agarose gel electrophoresis and transferred to GeneScreen Plus menbrane. Probing with [$^{32}$P]-labelled nick-translated TGMV (pBH 604) revealed the presence of a band comigrating with subgenomic ssDNA (Fig.40, lanes 2 and

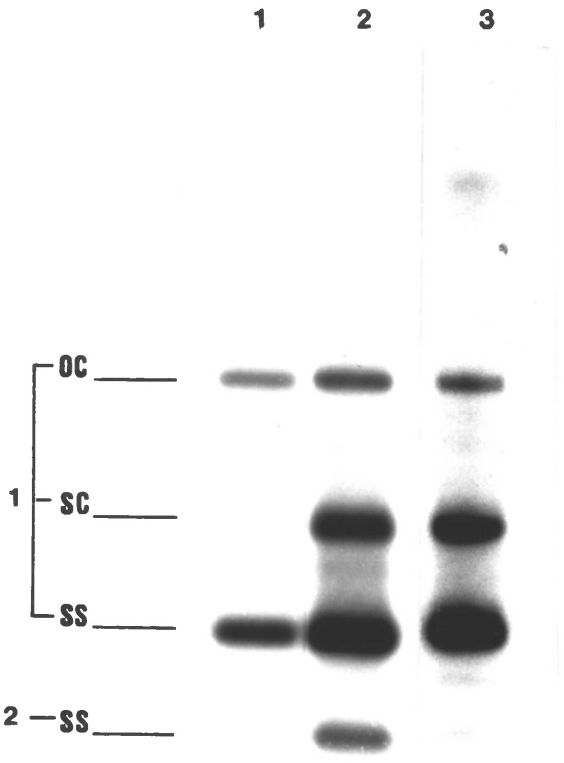



Fig.40. **Intracellular viral DNA species in TGMV infected plants.**

Autoradiogram of a blot (from a 1.2% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated TGMV 2 DNA.
(1) DNA extracts from *N. benthamiana* plants infected with cloned DNAs A and B; (2) DNA extracts from *N. benthamiana* plants infected with native TGMV; (3) DNA extracts from a *N. benthamiana* plant infected with first passage inoculum and cloned subgenomic DNA.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic-length and subgenomic viral DNA.

3). In view of the observation that clone-infected plants did not contain detectable levels of subgenomic DNA, even after multiple passages in the case of CLV, it was assumed that infection with cloned subgenomic DNA in the presence of first passage inoculum did result in replication of the former. Further infectivity tests were therefore conducted with first passage inoculum.

Since the subgenomic DNA used in the experiment above contained a nonsense codon in ORF BL1, the only extended ORF in this molecule, it would appear that expression of the subgenomic DNA is not required for its replication. Recent studies have shown that DNA 1 contains all the viral coded genes required for replication and suggested that DNA 2 is required for systemic spread of virus in the plant as well as sympton production (Townsend *et al.*, 1986; Rogers *et al.*, 1986; section 1.10) Since TGMV subgenomic DNA lacks all of ORF BR1 and the C-terminus of ORF BL1, it is unlikely that inoculation with the subgenomic DNA in combination with DNA 1 alone would result in infection. This has indeed been demonstrated for CLV subgenomic DNA (Stanley and Townsend, 1985).

### 3.11.2 Construction of pTBneo

Characterisation of TGMV subgenomic DNA showed that the region of DNA B between nucleotides 255 and 1529 does not contain sequences essential for DNA replication and packaging. It was therefore decided to replace this region with a foreign gene and see if inoculation of tobacco plants with the resulting hybrid DNA molecule in combination with components 1 and 2 would result in replication of the former and expression of the foreign gene.

Since it is not known which sequences, other than the 'common region', are required in *trans* for DNA replication and packaging, it was decided to replace the 1114bp *ClaI/TthIII* fragment of pBH604 (Fig.19B) with the foreign gene. This fragment comprises *ca.*87% of the region deleted in the subgenomic DNA and does not extend beyond the deletion borders. There was no risk, therefore, of essential sequences being lost. This fragment also contains all of ORF BR1 and its polyadenylation signal. Another possibility that was considered was to insert the foreign

gene directly into a cloned subgenomic DNA molecule. However, for the reason stated above it was thought that this could cause disruption of an essential region.

The construction of the hybrid DNA molecule (pTBneo) is illustrated in Fig.41. The foreign gene used consisted of the neomycin phosphotransferase II (npt II) coding region of transposon Tn5 coupled to the Ti plasmid-nopaline synthase (NOS) promotor and polyadenylation signal (a gift from M.W.Bevan; Bevan *et al.*, 1983). This chimeric gene (NOS-npt II) confers resistance to the antibiotics kanamycin and G418 to which plants are normally sensitive and its activity in plant extracts can be easily assayed (Herrera-Estrella *et al.*, 1983; Bevan *et al.*, 1983). The plasmid pNeo6 consisting of the NOS-npt II gene cloned into pUC9 was digested with restriction enzyme *Eco*RI and the resulting cohesive ends were converted to blunt ends with the Klenow fragment of *E.coli* DNA polymerase I (section 2.23.1). The same plasmid was then digested with *Hin*dIII and the excised NOS-npt II gene purified by gel elution (2.16.4). Similarly, pBH604 was digested with *Tth*III1 and the resulting cohesive ends converted to blunt ends as above. The same plasmid was then digested with *Hin*dIII to excise the TGMV DNA $\lambda$ 1114bp *Cla*I/*Tth*III1 fragment, and the 5.06Kbp fragment of pBH604 consisting of pAT153 and the TGMV DNA $\lambda$ 1408bp *Cla*I/*Tth*III fragment was gel purified as above. This fragment was then ligated (section 2.29.3) to the NOS-npt II gene and the ligation mixture used to transform competent *E.coli* JM101 cells (section 2.29.5). Bacteria containing recombinant DNA were selected by their resistance to ampicillin. Recombinant DNA was isolated as in section 2.27 and examined by agarose gel electrophoresis. One plasmid (pTBN) which appeared to be of the right size for a plasmid in which the TGMV DNA $\lambda$ 1114bp *Cla*I/*Tth*III1 fragment had been replaced by the NOS-npt II gene was further analyzed by restriction enzyme digestion followed by agarose gel electrophoresis (Fig.42). The restriction fragment patterns produced were indeed those expected for such a plasmid. Because the restriction sites at each end of the insert (the 1408bp *Cla*I/*Tth*III1 fragment of DNA $\lambda$ coupled to the NOS-npt II gene) in pTBN are different (*Hin*dIII and *Cla*I), it would be difficult for

Legend to Fig.41. **Construction of pTBneo.**

See section 3.11.2 for details of the construction.
TGMV 2 sequences are shown in (.....); the neomycin phosphotransferase II (npt II) coding region is shown in (••••••); the Nos promoter and termination sequences are shown in (------). Approximate sizes are given in Kb. ORFs are shown as open arrows and the position of the 200bp 'common region' is indicated by a shaded box. Open triangles indicate the position of TATA boxes and shaded triangles indicate the position of polyadenylation signals. Restriction enzyme sites shown are: C=*Cla*I, H=*Hin*dIII, P=*Pst*I, S=*Sal*I and T=*Tth*III1.

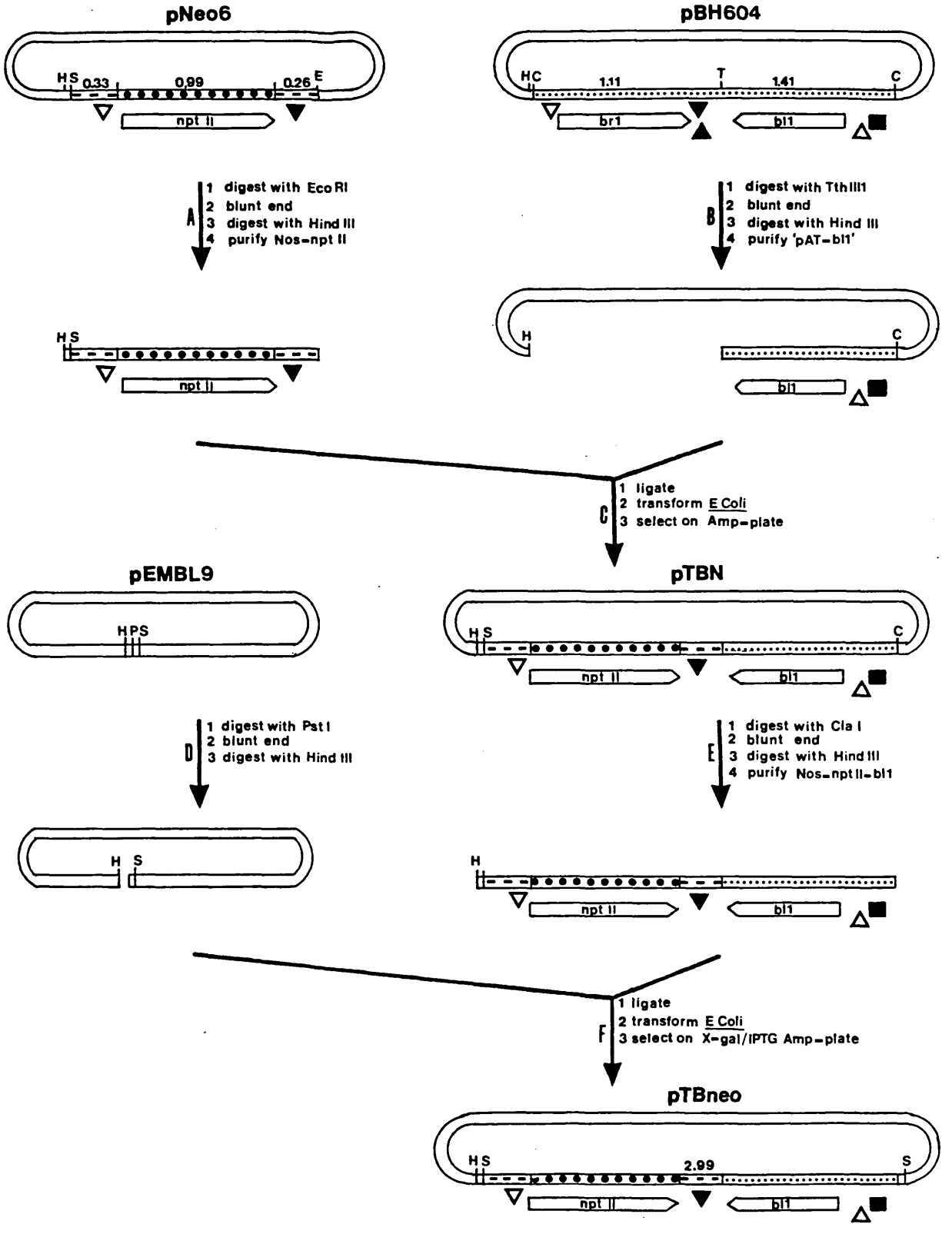

pNeo6
HS 0.33 0.99 0.26 E
npt II
pBH604
HC 1.11 T 1.41 C
br1
bl1
A
1 digest with EcoRI
2 blunt end
3 digest with Hind III
4 purify Nos-npt II
B
1 digest with TthIII1
2 blunt end
3 digest with Hind III
4 purify 'pAT-bl1'
HS
npt II
H C
bl1
C
1 ligate
2 transform E Coli
3 select on Amp-plate
pEMBL9
HPS
pTBN
HS C
npt II
bl1
D
1 digest with Pst I
2 blunt end
3 digest with Hind III
E
1 digest with Cla I
2 blunt end
3 digest with Hind III
4 purify Nos-npt II-bl1
H S
H
npt II
bl1
F
1 ligate
2 transform E Coli
3 select on X-gal/IPTG Amp-plate
pTBneo
HS 2.99 S
npt II
bl1

Fig. 41

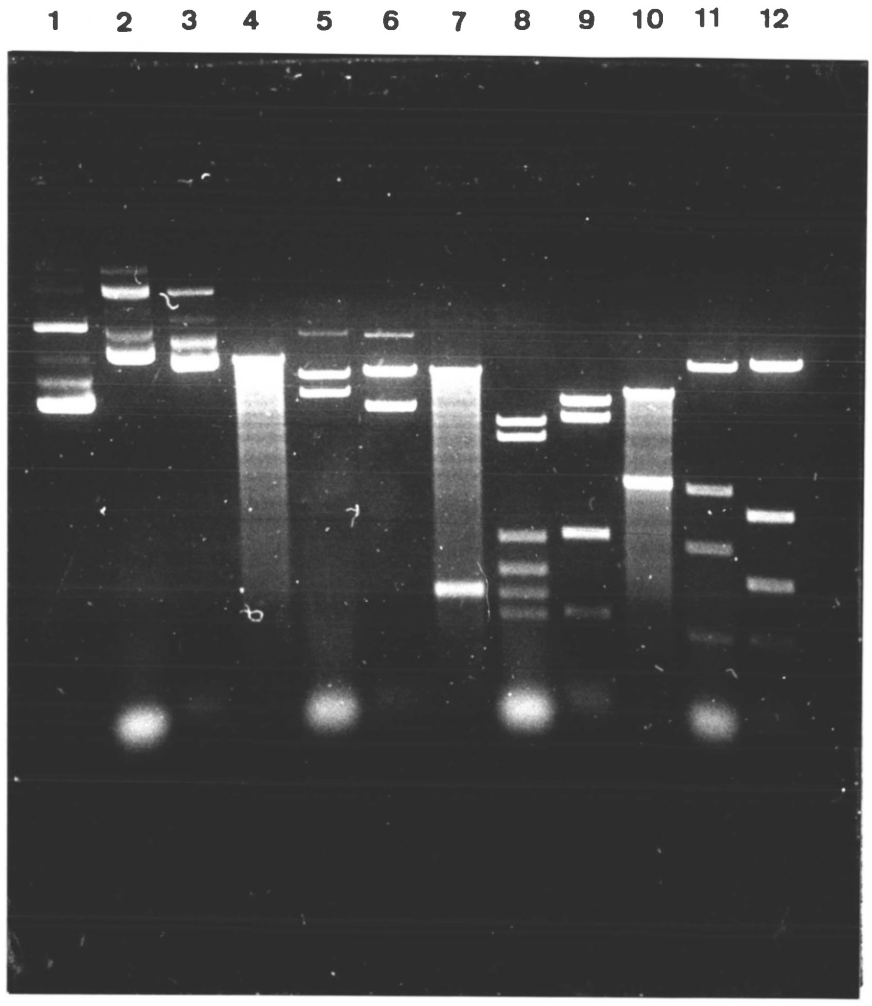



Fig.42. **Restriction analysis of pTBN.**

Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.
(1) pNeo6; (2) pTBN; (3) pBH604; (4) pNeo6 double-digested with HindIII and *Cla*I; (5) pTBN double-digested with *Hin*dIII and *Cla*I; (6) pBH604 double-digested with *Hin*dIII and *Cla*I; (7) pNeo6 double-digested with *Sal*I and *Pst*I; (8) pTBN double-digested with *Sal*I and *Pst*I; (9) pBH604 double-digested with *Sal*I and *Pst*I; (10) pNeo6 double-digested with *Hin*dIII and *Bam*HI; (11) pTBN double-digested with *Hin*dIII and *Bam*HI; (12) pBH604 double-digested with *Hin*dIII and *Bam*HI;

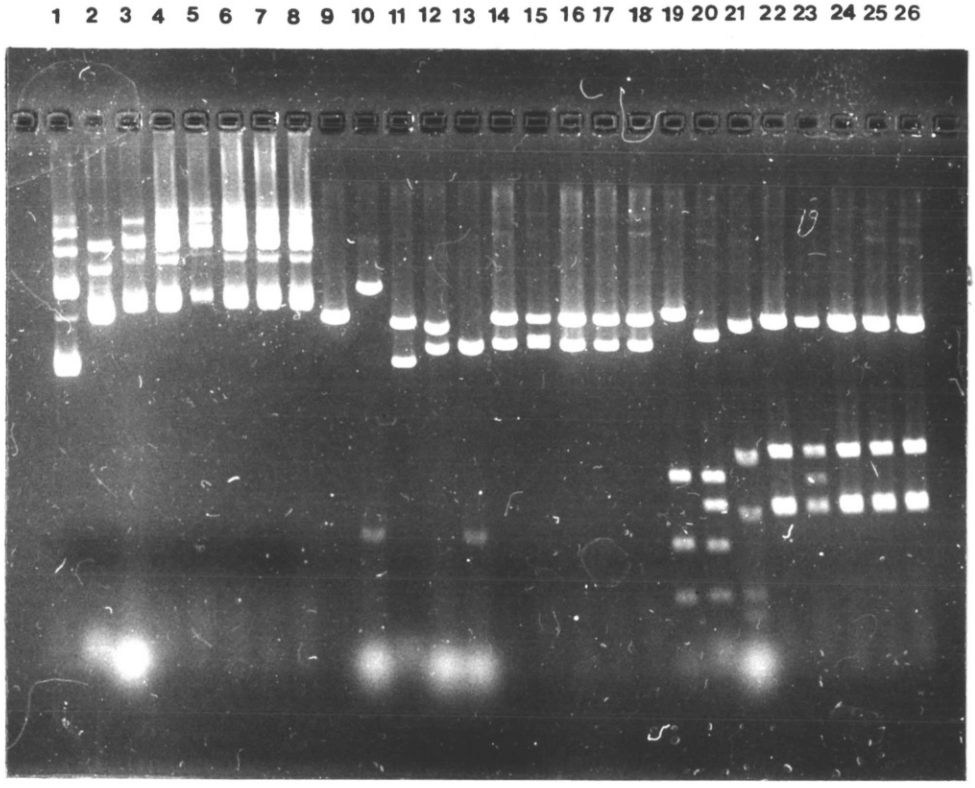



Fig.43. **Restriction analysis of pTBneo.**

Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.
(1) pEMBL19; (2) pBH604; (3) pTBN; (4) to (8) pTBneo 1 to 5; (9) pEMBL19 digested with *Sal*I; (10) pTBN digested with *Sal*I; (11) pBH604 digested with *Cla*I; (12) pTBN double-digested with *Sal*I and *Cla*I; (13) pTBN double-digested digested with *Hin*dIII and *Cla*I; (14) to (18) pTBneo 1 to 5 digested with *Sal*I; (19) pBH604 double-digested with *Cla*I and *Bam*HI; (20) pBH604 double-digested with *Hin*dIII and *Bam*HI; (21) pTBN double-digested with *Sal*I and *Bam*HI; (22) to (26) pTBneo 1 to 5 double-digested with *Sal*I and *Bam*HI.

plant cells to circularize the excised insert upon infection. A procedure was therefore devised to introduce the same restriction site on each side of the insert. To achieve this pTBN was digested with *Cla*I and the resulting cohesive ends converted to blunt ends as above. The same plasmid was then digested with *Hin*dIII and the excised insert purified by gel elution as above. Similarly, pEMBL 9 was digested with *Pst*I and the resulting cohesive ends converted to blunt ends with T4 polymerase (2.23.2). This plasmid was then digested with *Hin*dIII and ligated to the purified insert from pTBN. The ligation mixture was then used to transform competent *E.coli* JM101 cells and bacteria containing recombinant DNA selected by virtue of their white colour on X-gal/IPTG selective plates (section 2.29). Five recombinant clones were analyzed by restriction enzyme digestion followed by agarose gel electrophoresis (Fig.43). Of these, four contained the right insert, which could be excised by digestion with *Sal*I. One of these (pTBneo 1) was selected for further experiments.

### 3.11.3 Infection with pTBneo

Plasmids pB530 (cloned TGMV subgenomic DNA) and pTBneo (section 3.11.2) were isolated by the 'large scale' procedure (section 2.26) and full-length inserts were excised with restriction enzymes *Bam*HI and *Sal*I respectively. Three groups of four *N.benthamiana* plants each were then inoculated as follows. Plants of group 1 were inoculated with 45µl of first passage inoculum each, plants of group 2 were inoculated with 45µl of first passage inoculum and 25µg of *Bam*HI-digested pB530 each, and plants of group 3 with 45µl of first passage inoculum and 25µg of *Sal*I-digested pTBneo each. Inoculations were performed as described in section 2.9.4.

Plants were allowed to grow for about four weeks after inoculation. Plants inoculated with first passage inoculum alone developed symptoms between 8 to 11 days post-inoculation (8, 8, 11 and 11 days, respectively). Those inoculated with first passage inoculum in combination with the pTBneo insert developed symptoms between 8 and 14 days post-inoculation (8, 11, 12 and 14

days, respectively), and those infected with first passage inoculum in combination with cloned subgenomic DNA developed symptoms between 12 and 15 days post-inoculation (12, 12, 12 and 15 days, respectively). No differences in symptoms were detected in any of the plants.

For each plant, crude DNA extracts were obtained (section 2.10.1) and analyzed by agarose gel electrophoresis (Fig.44, lanes 2 to 13). Neither subgenomic DNA nor the pTBneo insert could be detected by UV illumination. The DNA in the gel was then transferred to GeneScreen Plus membrane and hybridized to a [$^{32}$P]-labelled ssDNA probe complementary to TGMV 2 virion sense DNA component (section 2.22). The results are shown in Fig.45.

DNA extracts from plants inoculated with first passage inoculum alone contained a small amount of DNA with the electrophoretic mobility of subgenomic ssDNA (Fig.45, lanes 2 to 5). This is in contrast to experiments with CLV, for which five passages of virus derived from cloned DNA inoculum failed to produce subgenomic DNA (Stanley and Townsend, 1985). The presence of subgenomic DNA in all four plants inoculated with first passage inoculum alone indicates that the formation of subgenomic DNA occurs relatively frequently. While it is likely that subgenomic DNA arose independently in each of these plants, it is possible that undetectable levels of subgenomic DNA might have already been present in the inoculum which subsequently became enriched upon second passage. In this respect, it is noteworthy that infection of tobacco plants with the cloned DNA components of PYMV, a closely related bipartite geminivirus, resulted in the formation of detectable levels of subgenomic DNA (E.J.F. Roberts, personal communication). It would be interesting, therefore, to sequence across the deletion borders of some of the subgenomic DNA molecules present in plants infected with TGMV first passage inoculum to see whether more than one type of subgenomic DNA has been formed and whether these differ between different plants.

DNA extracted from plants infected with first passage inoculum in combination with cloned subgenomic DNA (Fig.45, lanes 6 to 9) also contained small amounts of subgenomic DNA. Although it is difficult to estimate the relative amount of subgenomic DNA in each of the lanes, since some lanes were loaded with more DNA

than others, it appears that a larger amount of subgenomic DNA was formed in plants inoculated with a mixture of first passage inoculum and cloned subgenomic DNA than in those inoculated with first passage inoculum alone, suggesting that replication of cloned subgenomic DNA occurred in the former. Interestingly, a DNA species with higher electrophoretic mobility than the subgenomic DNA was formed in one of the plants inoculated with first passage inoculum and cloned subgenomic DNA (Fig.45, lane 8). DNAs with the mobility of subgenomic DNA, and in some experiments smaller subgenomic DNAs (Fig.45, lanes 11 and 12), were also formed when plants were inoculated with first passage inoculum and the pTBneo insert (Fig.45, lanes 10 to 13). The smaller subgenomic DNA formed in one of these plants (Fig.45, lane 12) appeared to have the same mobility as that formed in the plant inoculated with first passage inoculum and cloned subgenomic DNA (Fig.45, lane 8). While in the latter a DNA with the mobility of subgenomic DNA had also been formed, no such DNA could be detected in the former.

The fact that smaller subgenomic DNAs have not been previously detected in plants inoculated with native TGMV and were not detected in plants inoculated with first passage inoculum alone suggests that the smaller subgenomic DNA species are derived from the clones used in this experiment. It is not known whether the smaller subgenomic molecules were formed by an extension of the deletion present in the native subgenomic DNA or whether they were formed by internal deletions in other regions of the genome. The fact that these molecules are present in detectable levels suggests that they are able to replicate. Indeed, a DNA with similar electrophoretic mobility to subgenomic oc dsDNA can be seen in the respective lanes (Fig.45, lanes 8 and 12, between ss1 and sc1). If this assumption is correct, then these molecules must have retained sequences required for DNA replication. As discussed previously, the stringent size selection observed for native subgenomic DNA was attributed to rigid packaging constraints due to encapsidation in single isometric particles. It will be interesting, therefore, to see whether the smaller subgenomic molecules are encapsidated.

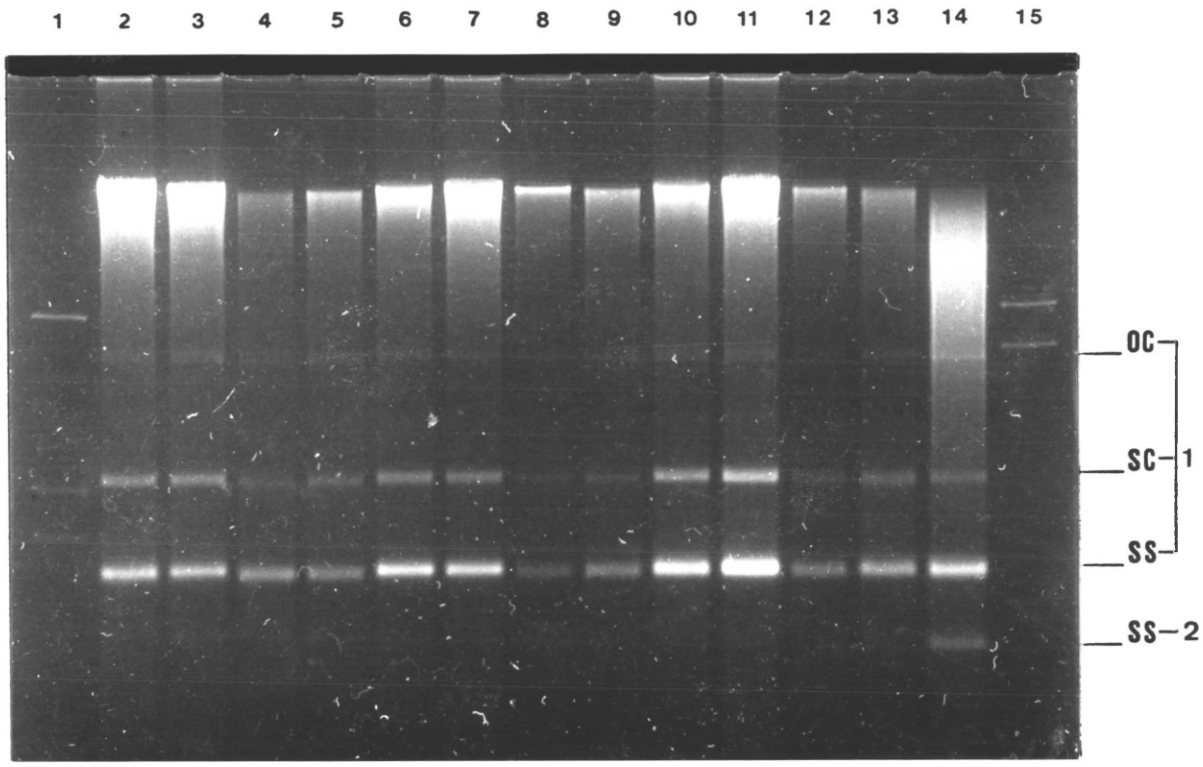



Fig.44. **Infectivity studies.**

Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.
(1) pBH604 double-digested with *Cla*I and *Tth*III1; (2) to (5) DNA extracts from *N. benthamiana* plants infected with first passage inoculum; (6) to (9) DNA extracts from *N. benthamiana* plants infected with first passage inoculum and cloned subgenomic DNA (*Bam*HI digested pB530); (10) to (13) DNA extracts from *N. Benthamiana* plants infected with first passage inoculum and *Sal*I digested pTBneo; (14) DNA extracts obtained from *N. benthamiana* plants infected with TGMV (TGMV crude DNA); (15) pTBneo digested with *Sal*I.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic-length and subgenomic viral DNA.

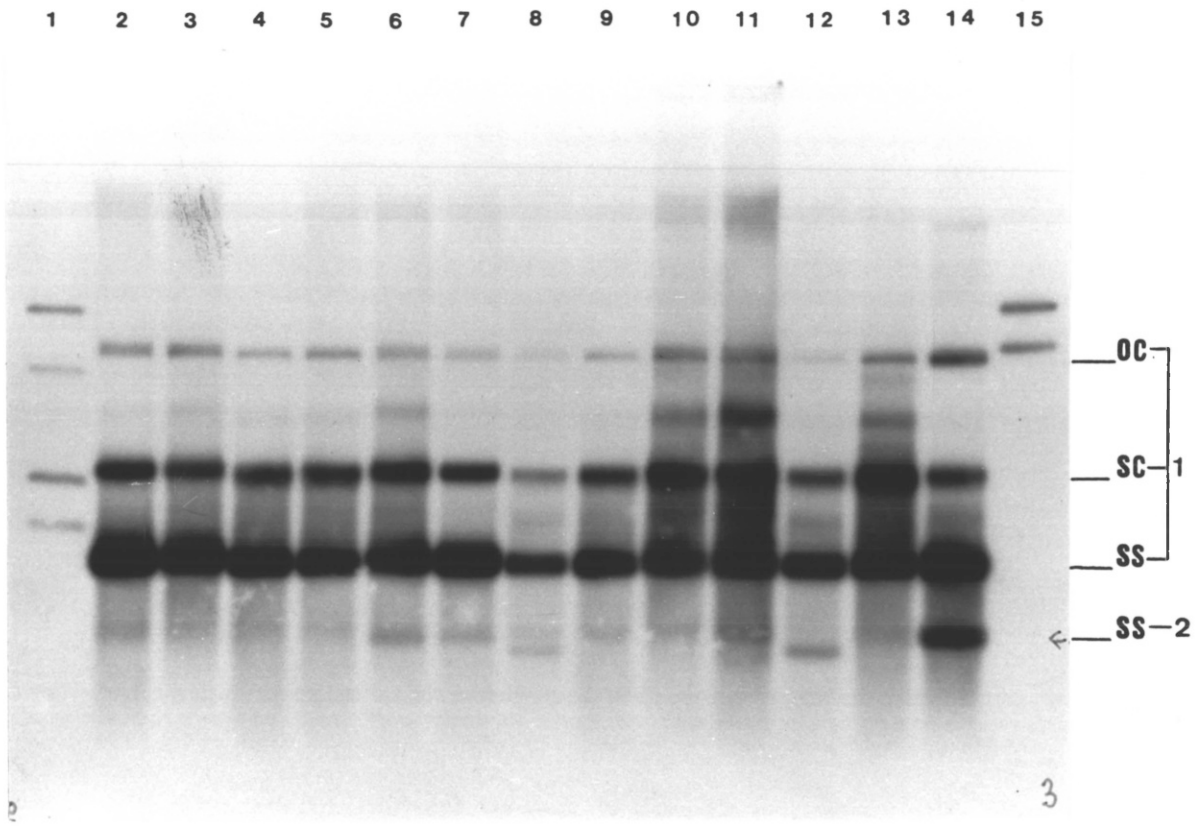



Fig.45. **Infectivity studies.**

Autoradiogram of a blot of the gel shown in Fig.44 probed with a ssDNA specific probe complementary to virion sense TGNV DNA 2. (1) to (15) as in Fig.44.

oc=open circular dsDNA; sc=supercoiled dsDNA (covalently closed circular dsDNA); ss=single-stranded DNA. 1 and 2 refer to genomic-length and subgenomic viral DNA.

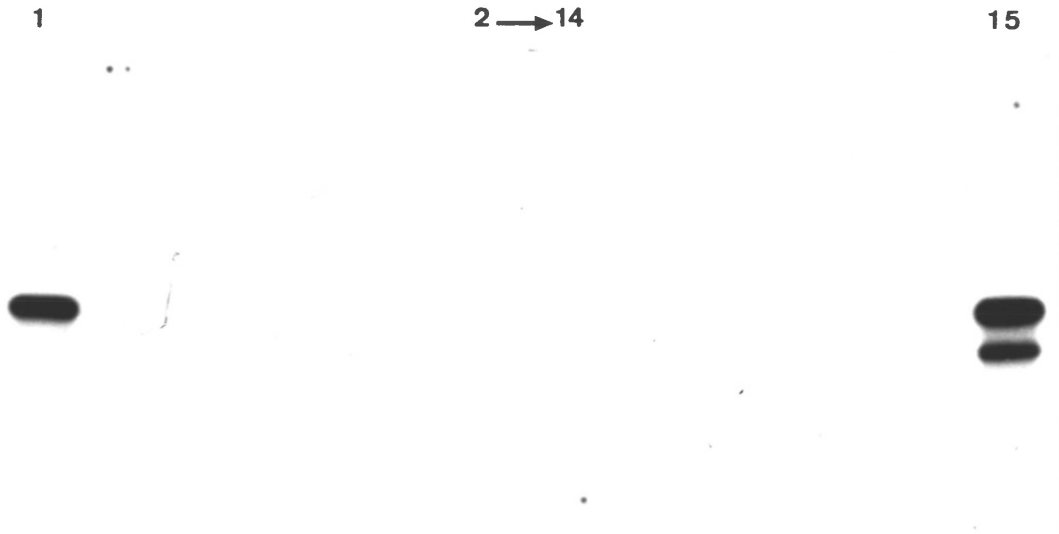



Fig.46. **Infectivity studies.**

Autoradiogram of a blot of the gel shown in Fig.44 probed with nick-translated pNeo6.
(1) to (15) as in Fig.44.

Because the pTBneo insert (*ca.*2.99Kbp) is larger than genomic length DNA (2522bp), its mobility on agarose gels is expected to be lower than that of DNA 2. When DNA extracts obtained from plants inoculated with first passage inoculum and the pTBneo insert were probed with [$^{32}$P]-labelled TGMV DNA 2, no DNA bands with the expected mobility for the pTBneo insert could be detected (Fig.45, lanes 10 to 13). The possibility existed, however, that the pTBneo insert might have suffered internal deletions, resulting in a mobility similar to that of DNA 2. To verify this assumption, the blot shown in Fig.45 was washed to remove the probe (section 2.24.7) and rehybridized to [$^{32}$P]-labelled pNeo6 (section 3.11.2). The results (Fig.46) show conclusively that the pTBneo construct, or any derivative containing the Nos-npt II gene or part of it, failed to replicate in the infected plants. The fact that subgenomic DNA, including the smaller species, was formed in these plants suggests that the Nos-npt II insert may have been excised by the plant. This may be partly due to the fact that the pTBneo insert was too large to be packaged. Experiments with a TGMV mutant in which the gene coding for the coat protein had been disrupted indicated that packaging is neither required for DNA replication nor for systemic spread of viral DNA (C.E. Brough, personal communication). However, packaging could increase the efficiency with which the virus is spread in the plant. The fact that TGMV subgenomic DNA is able to increase its numbers in the virus population over successive passages indicates that it must have a selective advantage over the genomic length components. If the rate of DNA elongation is a limiting factor in viral DNA replication, then the small size of the subgenomic DNA could be the reason for this selective advantage. By the same argument, the large size of the pTBneo insert could confer a replicative disadvantage on these molecules over the genomic DNAs, which could result in overgrowth of the former by the latter.

A possible way of creating selection for the pTBneo insert would be to inoculate plants with this insert in combination with TGMV DNA component 1 and a clone of DNA 2 in which ORF BL1 has been inactivated by insertion mutagenesis but retains ORF BR1

intact. If ORF BL1 is required for infection, complementation between the pTBneo insert, which retains a complete ORF BL1, and the defective DNA B will be required for infection, thus creating selective pressure for the maintenance of the pTBneo insert. Since the pTBneo insert lacks the polyadenylation signal for ORF BL1, further manipulations may be required to restore this signal. This vector system would be analogous to the SV40 replacement vectors (section 3.11). Due to packaging constraints, the size of the fragments that can be inserted into SV40 replacement vectors cannot be larger than the size of viral DNA being replaced. Whether a similar situation occurs in TGMV remains to be investigated. If this turns out to be the case, this vector system may be more useful for studies in plant protoplasts where encapsidation of the vector DNA would not be essential.

Still another problem associated with the use of complementation vectors is the possibility of rescue by recombination. Indeed, in all instances where pairwise inoculations of plants with combinations of cloned CaMV DNAs containing non-overlapping mutations that destroyed infectivity resulted in infection, rescue was due to CaMV genomes being restored to their normal form by recombination (Howell *et al.*, 1981). No experiments have yet been conducted to determine the possibility of recombination between defective TGMV genomes. Obviously these will be required if the successful development of TGMV complementation vectors is to take place.

## 3.12 Is encapsidated TGMV subgenomic DNA a DI particle?

Inoculation of tobacco plants with cloned CLV DNAs 1 and 2 in combination with increasing concentrations of cloned CLV subgenomic DNA resulted in progressive increases in the time required for symptom expression and the number of plants that did not show symptoms (Stanley and Townsend, 1985). These results indicate that CLV subgenomic DNA interferes with virus multiplication or spread in the host and is therefore analogous to DI particles of animal viruses (Holland, 1985). Similarly, inoculation of tobacco plants with first passage TGMV inoculum in combination with cloned TGMV subgenomic DNA appeared to delay

symptom expression by 1 to 7 days (with an average of about 3 days) compared to plants inoculated with first passage inoculum alone (section 3.11.3). These results suggest that, similar to CLV subgenomic DNA, TGMV subgenomic DNA may also have interfering properties.

TGMV subgenomic DNA could interfere with multiplication of the genomic components by competing for viral coded products required for DNA replication and/or encapsidation. Another possibility would involve expression of the subgenomic DNA to produce altered forms of the protein encoded by ORF BL1 in DNA 2 (section 3.9). Experiments in which tandem copies of DNA 1 and DNA 2 have been inserted into the genome of *Petunia* plants *via* Ti plasmid vectors demonstrated that DNA 1 encodes all the virus coded gene products required for DNA replication and suggested that DNA 2 is required for systemic spread of the virus in the host as well as symptom development. Altered forms of the protein encoded by ORF BL1 might therefore interfere with cell to cell transmission of the virus and/or symptom expression, possibly by competing for the binding sites of the wild-type protein, therefore delaying the time for symptom development.

Unlike some DI particles of animal DNA viruses, such as those of SV40 and HSV (reviewed by Holland, 1985), TGMV subgenomic DNA does not contain reiterated copies of the origin of DNA replication which are thought to confer a replicative advantage to those molecules harbouring them. Selection for the subgenomic DNA, however, may be achieved in other ways. As mentioned previously (section 3.11.3) the small size of the subgenomic DNA molecules may result in faster rounds of replication compared to the genomic DNAs or, possibly, in more efficient encapsidation. Alternatively, TGMV subgenomic DNA could be selected by virtue of its interfering properties, by preventing the virus from killing the host. In the case of animal viruses, DI particles have been shown to protect their hosts from death caused by the homologous virus (Holland *et al.*, 1980). The effects of TGMV subgenomic DNA on symptom expression, however, appear to be very mild. Also, there is no evidence that infection in the absence of the subgenomic DNA results in death of the host.

Admittedly, the results obtained for TGMV subgenomic DNA are only preliminary and further experiments will be required to confirm whether TGMV subgenomic DNA does indeed interfere with symptom development and replication. To avoid misinterpretation of the results, it will be important to use cloned DNAs 1 and 2 as inoculum rather than first passage inoculum, since the latter may already contain undetectable amounts of subgenomic DNA. A slight delay in symptom expression was also observed when plants were inoculated with first passage inoculum in combination with the pTBneo insert (section 3.11.2). While it is possible that the pTBneo insert, or rather subgenomic molecules derived from it, may have interfered with symptom expression, it may also be that contaminants within the DNA preparations or even a slight dilution of the inoculum may be responsible for delaying symptom development.

## 3.13 Possible mechanisms for the formation of TGMV subgenomic DNA

Analysis of TGMV subgenomic DNA has shown that (i) it is derived exclusively from DNA 2, (ii) it is approximately half the size of DNA 2 and (iii) it is formed by deletions in specific regions of DNA 2. While the possibility exists that these characteristics may be a direct consequence of the mechanism responsible for the generation of the deletion, it is more likely that they result from subsequent selection of subgenomic molecules meeting specific requirements. As discussed in section 3.10, the stringent size selection and the specific location of the deletion points probably reflect packaging constraints imposed by the subgenomic DNA being encapsidated in half-size particles and the need to retain essential sequences for DNA replication and encapsidation, respectively. The absence of subgenomic DNAs derived from DNA 1 is less clear. One possibility is that subgenomic DNAs derived from DNA 1 may interfere with the infection process to such an extent that they are strongly selected against. DNA 1 encodes protein(s) required for DNA replication (Rogers *et al.*, 1986) as well as the coat protein (Hamilton *et al.*, 1984). The high degree of amino acid sequence conservation shown by the proteins encoded by ORF AL1 in DNA 1

and equivalent ORFs in both mono and bipartite geminiviruses (see section 3.22) suggests a role in DNA replication for these proteins. Thus, a truncated form of the protein encoded by ORF AL1 might interfere with DNA replication by competing with the wild-type protein for binding sites in the viral DNA. Similarly, a truncated coat protein might interfere with DNA packaging. Sequence analysis of a cloned subgenomic molecule derived from WDV, a geminivirus more closely related to TGMV 1 than to TGMV 2, indicated that the deletion comprises the whole of the coat protein gene and ORF 10146 (H.Macdonald, personal communication; section 3.19) leaving no truncated ORFs. Whether TGMV DNA 1 derived subgenomic DNAs with truncated ORFs are selected or not, there must be other factors preventing the formation of DNA 1 subgenomics with no truncated ORFs. It is possible that the distribution of sequences required for DNA replication and encapsidation in DNA 1 is such that the generation of a subgenomic DNA by a single deletion that would fulfill the stringent size requirement for packaging is precluded. Alternatively, the formation of subgenomic DNA from only DNA 2 may be a consequence of differences in the secondary or tertiary structures of DNAs 1 and 2. The much greater S1 nuclease-sensitivity of DNA 1 compared to DNA 2 indicates significant structural differences between the supercoiled forms of these DNAs (Sunter *et al.*, 1985). In this respect it may be significant that the 5' deletion points in the subgenomic DNAs are located within or immediately downstream of an inverted repeat structure (243 5'-TGCAGAGAGTCTTGT↓TATTTC↓TG↓CA↓-3' 265) in close proximity to an S1-sensitive site in DNA B (Sunter *et al.*, 1985).

A number of mechanisms have been proposed for the generation of deletions in DNA molecules (see section 1.19). Some of these involve errors in DNA replication, others involve recombination (either homologous or illegitimate) between the same (intramolecular) or different (intermolecular) DNA molecules. Insight into the mechanisms by which deletions are formed may be gained by analysis of the DNA sequences flanking the deletion endpoints. Indeed, several mechanisms have been postulated from the results of such analyses. Thus, the presence of short direct repeats at the deletion borders in β-like globin genes has led to

the proposal of the 'slipped mispairing' model (Efstratiadis *et al.*, 1980; section 1.19). It was expected, therefore, that examination of the sequences surrounding the deletion points in TGMV subgenomic DNA might give some indication as to what mechanisms are more likely to explain its formation.

Computer assisted analysis of the TGMV DNA 2 sequence did not reveal any extensive perfect homology (>8bp) between the sequences flanking the 5' deletion points and those flanking the 3' deletion points in the cloned subgenomic DNA molecules. No significant homologies were detected even when mismatches were allowed. Searches for homologies were performed with the computer program BESTFIT (Deveraux *et al.*, 1984), which finds the best region(s) of homology between two sequences, and involved regions comprising one hundred and fifty nucleotides on each side of the deletion points. Intramolecular homologous recombination in a dsDNA form of DNA 2 was therefore ruled out as a possible mechanism for the formation of subgenomic DNA.

Examination of the nucleotide sequences surrounding deletion sites in the *lac I* gene of *E.coli* (Farabaugh *et al.*, 1978), in *lac I-Z* fusion strains of *E.coli* (Albertini *et al.*, 1982), in the human β-globin gene family (Efstratiadis *et al.*, 1980), and in ORF II of the CaMV genome (Howarth *et al.*, 1981) strongly suggested that short direct repeat sequences of as few as 2-8bp may be involved in the generation of these deletions. In the majority of cases, the deletions were formed by eliminating DNA segments comprised between the repeats together with one copy of the repeats, the other copy remaining after the deletion. A similar pattern has been found for CLV subgenomic DNA (Stanley and Townsend, 1985). In all nineteen clones of subgenomic DNA that were analysed, the deletions occurred between short direct repeat sequences (2-8nt), one copy of which was entirely removed (Table 2). A possible mechanism by which these deletions could be formed (Efstratiadis *et al.*, 1980) has been described previously (section 1.19, Fig.8). The possibility of enzymes which recognize short homologous sequences and promote recombination has also been invoked (Albertini *et al.*, 1982).

All direct repeat sequences ≥3nt located in the vicinity of the deletion points in TGMV subgenomic DNA are shown in Fig.47.

Although a number of direct repeats were found, the largest of which were 5nt long, there were no repeats ≥3nt at the deletion points. In type 2 clones, however, the dinucleotide TC is located immediately upstream of both the 5' and 3' deletion points. The presence of this very short direct repeat at the deletion points of one of the least abundant clones is probably not significant. It seems unlikely, however, that the presence of direct repeats at the deletion points in CLV subgenomic DNA may be fortuitous, suggesting that TGMV and CLV subgenomic DNAs may be formed by different mechanisms. Similar to TGMV subgenomic DNA, no direct repeats were found at the deletion points in a cloned WDV subgenomic DNA (H. Macdonald, personal communication).

The presence of inverted repeat sequences near the deletion points may help bring these points together in the ssDNA form of DNA 2 by basepairing. Double-stranded cleavage (or two single-stranded cleavages) at the site of pairing, possibly by host endonucleases, followed by re-ligation could generate a subgenomic DNA molecule. There are no known eukaryotic ligases that can use ssDNA as a substrate. If the subgenomic DNA is formed by cleavage of viral ssDNA, then re-ligation would probably have to be catalyzed by a virus-coded product. Inverted repeats ≥3nt in the vicinity of the deletion points are shown in Fig. 48. In type 2 clones, the deletion points may be located within a 4nt inverted repeat (due to the presence of a two nucleotides direct repeat at the deletion points in these clones (see above) it is impossible to ascertain the exact location where the deletion occurred). However, no inverted repeats were found at the deletion points of other clones.

A noteworthy feature of the nucleotide sequences at the deletion points is the presence of two copies of the palindromic sequence TGCA in both border sequences, each of which is part of larger repeat sequences (Table 3). It is possible that subgenomic DNA may be formed by site-specific recombination mediated by the repeats TGCA. However, only three deletion points occur at this sequence and, of these, only one occurs immediately 3' to this sequence. To account for the formation of all types of deletions, cleavage at this site would have to be accompanied by further exonucleolytic degradation prior to religation. Although the

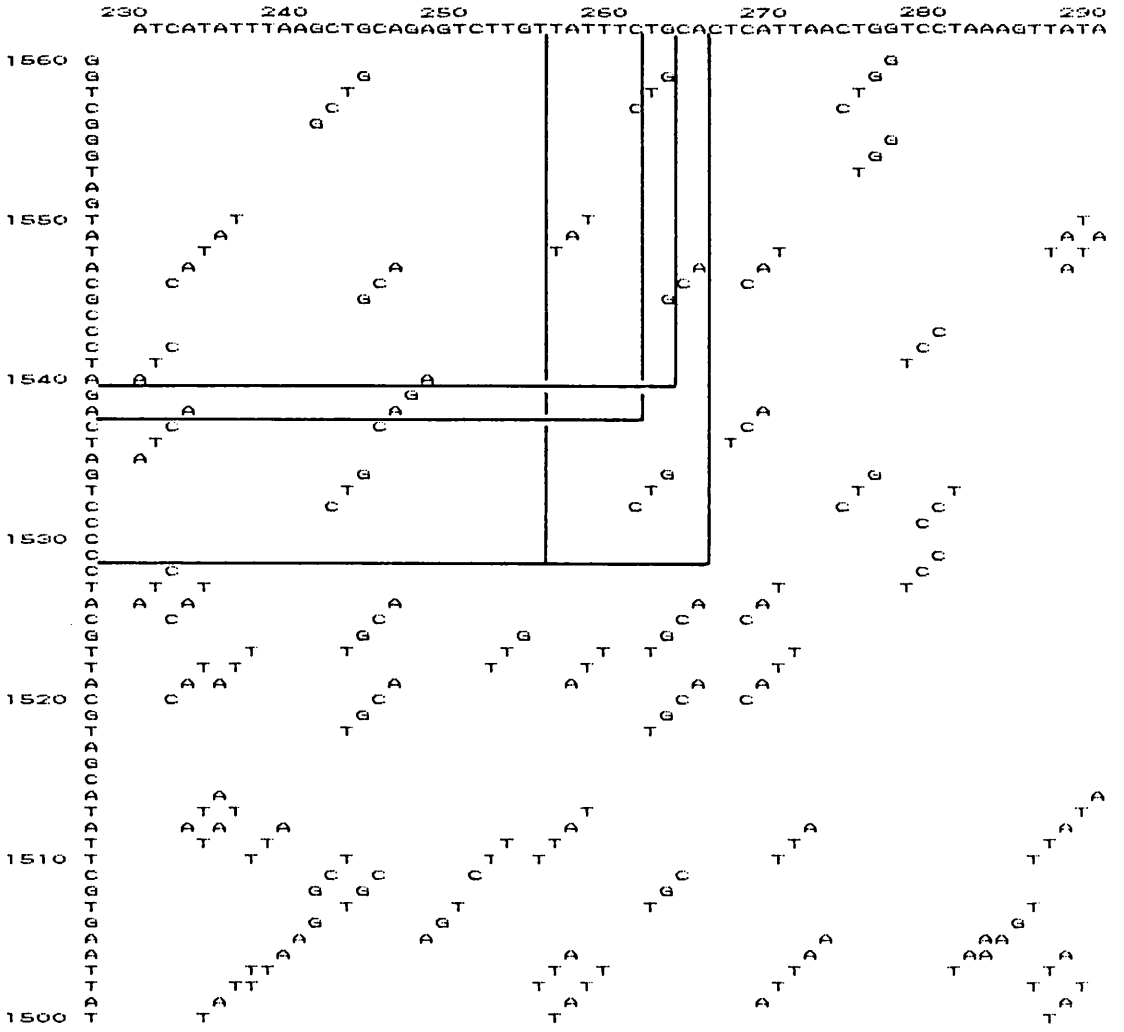



Fig.47. **Direct repeats in the vicinity of the deletion points in the cloned subgenomic DNAs.**

The TGMV 2L DNA sequence flanking the 5' deletion points is indicated horizontally and the sequence flanking the 3' deletions is indicated vertically. The deletion joints are indicated by the solid lines. All direct repeats ≥3nt are shown.

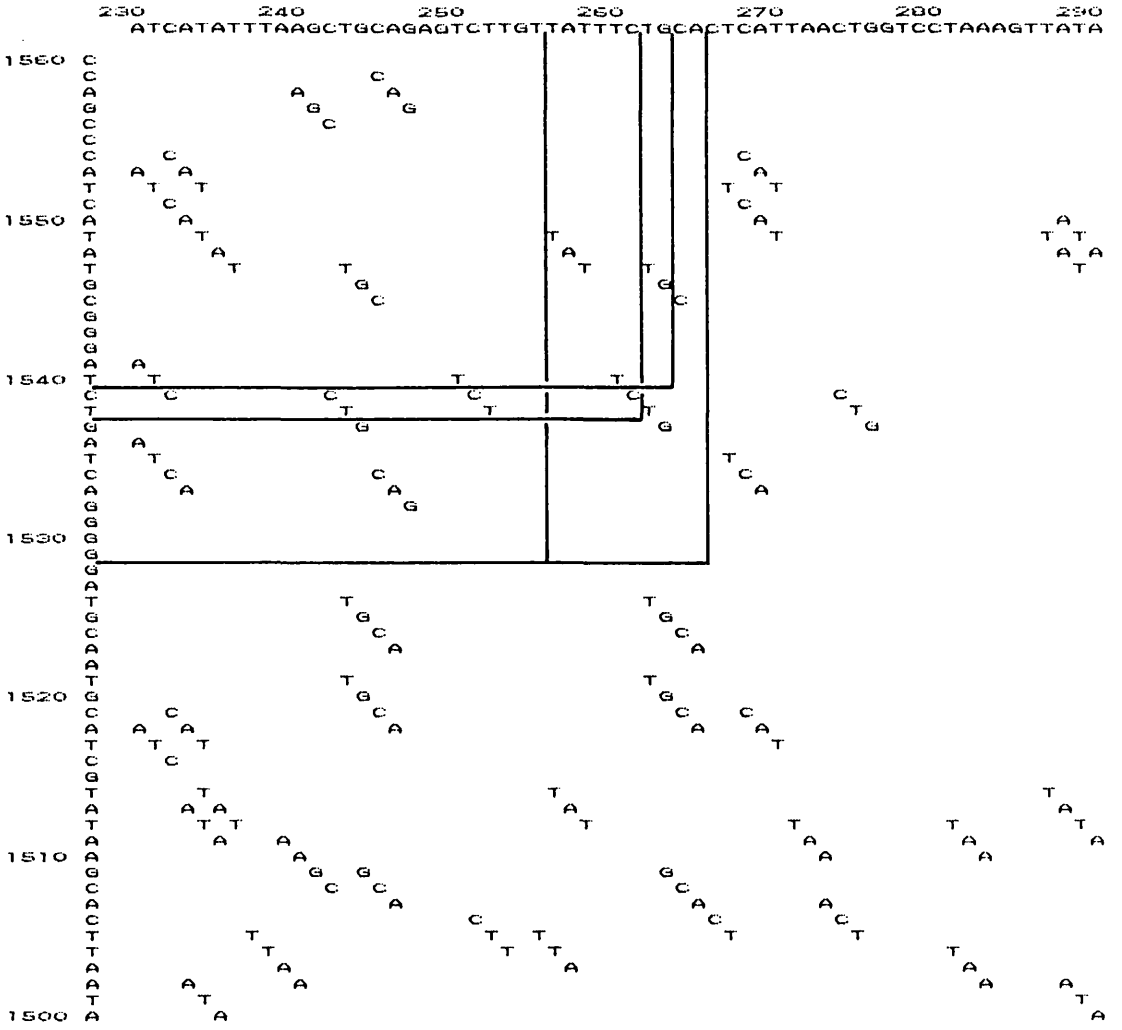



Fig.48. **Inverted repeats in the vicinity of the deletion points in the cloned subgenomic DNAs.**

The TGMV 2 DNA sequence flanking the 5' deletion points is indicated horizontally and the revert-complement of the sequence flanking the 3' deletions is indicated vertically. The deletion joints are indicated by the solid lines. All inverted repeats ≥3nt are shown.

sequence TGCA occurs at three other places in DNA 2, deletions involving these sequences would probably not meet the criteria for selection. The sequence TGCA occurs only once in DNA 1.

The absence of significant homologies at the deletion points suggests that TGMV subgenomic DNA may be formed either by 'illegitimate' recombination or by errors in DNA replication.

The simplest way by which 'illegitimate' recombination could take place involves random cleavage of DNA 2 by endonucleases followed by recircularisation of the linearised fragments. Although a number of different subgenomic DNAs could be formed by this mechanism, only those fulfilling specific requirements would be selected. A similar mechanism has been proposed for the formation of SV40 DI particles (Johnson *et al.*, 1982). Cleavage of DNA 2 could also occur during breakdown of geminate particles to single particles. Depending on whether or not ssDNA is arranged in a sequence specific way in geminate particles, cleavage may or may not be sequence specific. In the latter case, only those molecules able to replicate and be encapsidated would be maintained.

'Illegitimate' recombination could also be mediated by DNA topoisomerases types I (topo I) or II (topo II). These enzymes catalyze a variety of changes in the topological structure of DNA and are thought to play a role in DNA replication, transcription and recombination (Gellert, 1981).

Eukaryotic topo I acts by introducing transient single-stranded breaks in both ssDNA and dsDNA molecules. The nicking/closing reaction can be uncoupled by the addition of denaturing agents, thereby allowing observation of the products of the nicking step. Thus, it has been demonstrated that topo I generated nicks contain 5' termini ending in hydroxyl groups and topo I molecules covalently bound to the 3' phosphates (Champoux, 1976, 1977). Binding of the enzyme to the 3' end is thought to store the energy required for re-synthesis of the phosphodiester bonds during the closing step. Analysis of the nucleotide sequence at topo I breakage sites has shown that the majority of sites occur immediately 3' of the sequences CTT or GTT (Been *et al.*, 1984). Bullock *et al.* (1985) showed that crossover points for chromosomal excision of SV40 from nontandem inserts are

associated with eukaryotic topo I cleavage sites *in vitro*, and suggested a model for topo I-mediated excision recombination (Fig.7). The observation that topo I can catalyze the ligation *in vitro* of heterologous DNA fragments showing no sequence complementarity lends further support to a role for this enzyme in 'illegitimate' recombination (Edwards *et al.*, 1982).

The positions of the sequences CTT and GTT near the deletion points in TGMV subgenomic DNA are shown in Table 3. None of the deletion points occurred immediately 3' of either of these sequences. Furthermore, in only one case did a deletion point occur within one of these sequences (GTT). In CLV, on the other hand, two deletion points may have occurred immediately 5' of the sequences CTT and GTT and a third deletion point may have occurred within a CTT sequence. In both TGMV and CLV, CTT and GTT sequences are present close (within 40nt) to the deletion points. Since the 5' ends of topo I generated nicks are not protected by bound topo I molecules, their initial position may be altered by exonucleolytic degradation. This could obscure the nature of the primary sequences involved and possibly account for the presence of deletion points which are not immediately 3' of the sequences CTT or GTT. However, since the specificity for these sequences is not absolute, topo I-mediated recombination could also occur at other sequences. It may be significant that all 3' deletion points in TGMV are located immediately upstream or downstream of the trinucleotide GAT (Table 3). Similarly, the 3' deletion points in CLV may also be located immediately upstream or downstream of another trinucleotide, in this case CAT, depending on the exact location of these points. Both trinucleotides match the consensus sequence 5'-A/T G/C T/A T-3' determined for rat liver and wheat germ topo I.

Topo II introduces transient double-stranded breaks in duplex DNA and therefore may play a role in 'illegitimate' recombination. *Drosophila* topo II cleavage sites occur preferentially in non-transcribed spacer segments and at the 5' and 3' ends of genes (Udvardy *et al.*, 1985). In contrast, topo I cleavage sites appear to occur predominantly in transcribed regions (Gilmour *et al.*, 1986). The 5' deletion points in TGMV and CLV occur in an intergenic region, whereas the 3' deletion

| Type | Subgenomic DNA clones | | N° of clones | length (nt) |
|---|---|---|---|---|
| 1 | AGCTGCAGAGTgTgGT | CCCCTGATCAGATCCCGCATATGATG | 2 | 1249 |
| 2 | AGCTGCAGAGTCTTGTTATTTC | AGATCCCGCATATGATG | 3 | 1246 |
| 3 | AGCTGCAGAGTCTTGTTATTTCTG | ATCCCGCATATGATG | 10 | 1246 |
| 4 | AGCTGCAGAGTCTTGTTATTTCTGCA | CCCCTGATCAGATCCCGCATATGATG | 7 | 1259 |

DNA B clone (pBH604)

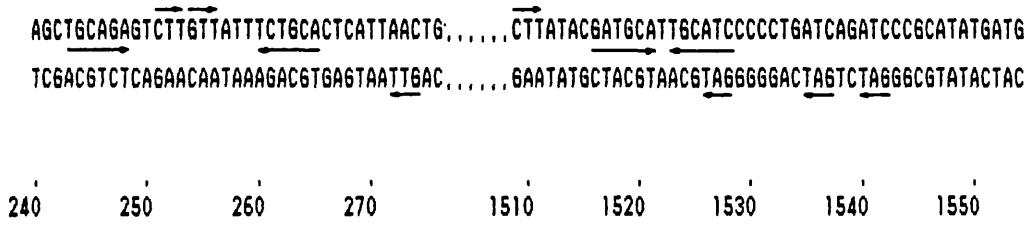



Table 3. **Nucleotide sequences at the deletion points of cloned TGMV subgenomic DNA and the corresponding regions of TGMV DNA 2.**

For the subgenomic DNA the sequence of the strand encapsidated in virions is shown. For DNA 2 the sequence of both strands is shown and arrows indicate the positions of the sequences CTT, GTT, GAT and inverted repeat sequences containing the palindromic sequence TGCA.

points occur near the C-terminus of the corresponding ORFs BL1 and 33.7. The location of deletion points in TGMV and CLV may therefore be more consistent with a topo II, rather than a topo I, mediated recombination. *Drosophila* topo II has been shown to have a specificity, although not absolute, for the consensus sequence 5'-GT.A/TAY↓ATT.AT..G-3', where a dot means no preferred nucleotide, Y is for pyrimidine and the arrow shows the point where cleavage occurs (Sander and Hsieh, 1985). The sequences at the deletion points in both TGMV and CLV do not match this consensus sequence. It is possible, however, that plant topo II may have different consensus sequences.

As mentioned above, TGMV subgenomic DNA could be formed by an error in DNA replication. Very little is known, however, about the mechanism of geminivirus DNA replication. On the available evidence (see section 1.10) it would appear that viral ssDNA is first converted into a circular dsDNA replicative form which then acts as a template for further synthesis of ssDNA, possibly by a rolling circle mechanism.

The subgenomic DNA could be formed during the conversion of ssDNA to dsDNA as a result of the replicase complex detaching from one site (the 3' deletion points) on the template and resuming synthesis at a more distant site (the 5' deletion points). Although the distance between these points would have to be relatively large in terms of the number of nucleotides (*ca.*1250nt) between them, the actual physical distance might be considerably reduced by the secondary structure adopted by the ssDNA (or partially dsDNA) molecule. A similar mechanism was originally proposed for the formation of DI particles of animal RNA viruses (Leppert *et al.*, 1977; Huang, 1977). Alternatively, subgenomic DNA could be formed during synthesis of ssDNA from a dsDNA template by a rolling circle mechanism. A deletion could be generated by cleavage of the nascent strand at a pseudo recognition site by an endonuclease-ligase enzyme analogous to the ØX174 gene A protein. However, there are no significant homologies at the 5' and 3' deletion points which would be required for this mechanism.

## 3.14 Purification of WDV dsDNA

Total DNA from WDV-infected wheat leaves (*Triticum aestivum* L.cv.Diamant) was extracted as described in section 2.10.3. The extraction procedure was adapted from that used to isolate BGMV oc dsDNA from infected bean leaves (Ikegami *et al.*1981). It included 2-mercaptoethanol and L-ascorbic acid in the extraction buffer to prevent oxidation of the nucleic acid extracts and all steps following the first ethanol precipitation in the original procedure were omitted. After the first ethanol precipitation the nucleic acid extracts were further extracted with phenol to remove residual proteins, re-precipitated with ethanol and resuspended in 1 X TAE buffer.

Total DNA from WDV-infected leaves (WDV crude DNA) is shown in Fig.49 (lane 2). Since the extraction procedure used did not include any steps to selectively enrich the DNA preparation in viral DNA, the amount of the latter was very small compared to that of host DNA. Only one DNA band (band c in Fig.49 A) could be routinely observed on agarose gels by UV illumination. This band had an electrophoretic mobility slightly lower than that of TGMV oc dsDNA (Fig.49 A, lane 1) and was thought to represent the equivalent WDV DNA species. In the absence of a healthy control (only infected plants were supplied), however, it was not possible at this stage to ascertain whether this band was of viral or host origin.

To isolate the putative WDV oc dsDNA species, agarose gel slices containing 'band c' were excised and the DNA electroeluted (section 2.16.3). Because of the low abundance of 'band c' relative to the host DNA background, the eluted DNA consisted primarily of the latter (Fig.49 A, lane 3; Fig.49 B). As discussed in section 3.4, polyacrylamide gels can be used to separate circular from linear DNA of similar Mr. If, as mentioned above, 'band c' consisted of circular viral DNA, fractionation of the eluted DNA by polyacrylamide electrophoresis should result in two bands, with the top band consisting of the circular viral DNA and the bottom band of linear host and viral DNA. Indeed, when the eluted DNA was fractionated on 4% polyacrylamide tube gels,

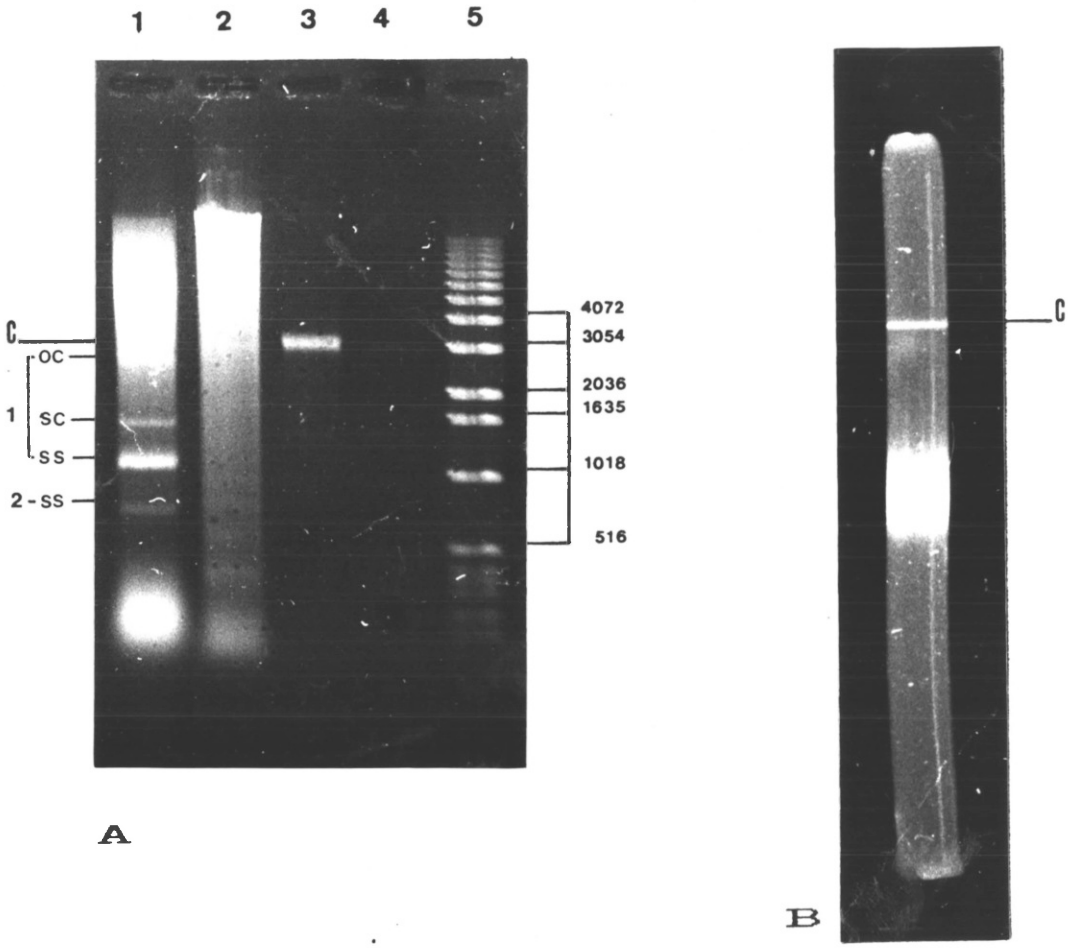



Fig. 49. **Gel purification of WDV oc dsDNA.**

A) Electrophoresis in a 1% agarose gel containing 0.5µg/ml ethidium bromide.
(1) TGMV crude DNA; (2) Total DNA from WDV-infected wheat leaves (WDV crude DNA); (3) agarose gel purified WDV oc dsDNA (band c); (4) WDV oc dsDNA after sequential purification from agarose gel and polyacrylamide gel respectively (underloaded); (5) Kb marker (fragment sizes are indicated in base pairs).
The bands labelled oc (open circle dsDNA), sc (supecoiled dsDNA), ss (single-stranded DNA) refer to those present in TGMV crude DNA. 1 and 2 refer to genomic and subgenomic-length respectively. Band c is the putative oc form of WDV dsDNA.

B) Polyacrylamide (4%) tube gel electrophoresis of agarose gel purified WDV oc dsDNA. See section 3.14.

two bands were obtained (Fig.49 B). The slower migrating band, consisting of the putative viral oc dsDNA, was eluted for further characterization.

## 3.15 Characterisation of intracellular WDV DNA species

Aliquots of WDV crude DNA were digested separately with Exonuclease III, S1 nuclease, RNAse A and DNAse I (section 2.18). The reaction products were fractionated on a 1.5% agarose gel, transferred to GeneScreen membrane and hybridized to the purified putative viral oc dsDNA labelled with [$^{32}$P] by nick-translation. The results are shown in Fig.50. The pattern of bands was remarkably similar to that obtained from similar TGMV crude DNA reaction products (see Fig.9 B), indicating that the eluted band was probably of viral origin. Conclusive evidence for a viral origin was obtained by probing a similar blot with [$^{32}$P]-labelled cDNA made to virion ssDNA extracted from purified virus particles (a gift from Dr. R.H.A.Coutts) (Fig.51, lanes 1 to 5). All the major bands that hybridized to the putative WDV oc dsDNA also hybridized to virion ssDNA. The apparent absence of some of the minor bands in the blot shown in Fig.51 is probably due to the lower percentage (1.0%) of agarose used to fractionate the reaction products.

The viral bands observed in Fig.50 were tentatively identified as genomic-length oc, linear and ccc dsDNA (bands c, d and e, respectively); genomic-length ssDNA (band g); dimeric oc and linear dsDNA (bands a and b, respectively) and subgenomic ccc dsDNA (band j), on the basis of their differential sensitivities to the various nucleases used in the experiment. Thus, ssDNA was identified by its sensitivity to S1 nuclease and resistance to Exonuclease III, ccc dsDNA was identified by its relative resistance to both S1 nuclease and Exonuclease III (see below), and oc and linear dsDNA were identified by their resistance to S1 nuclease and sensitivity to Exonuclease III. The oc and linear dsDNA forms could be distinguished between them by their relative amounts in S1 nuclease reactions. This enzyme attacks single-stranded regions in duplex molecules, such as nicks in oc dsDNA and unpaired regions in ccc dsDNA molecules, which results in an

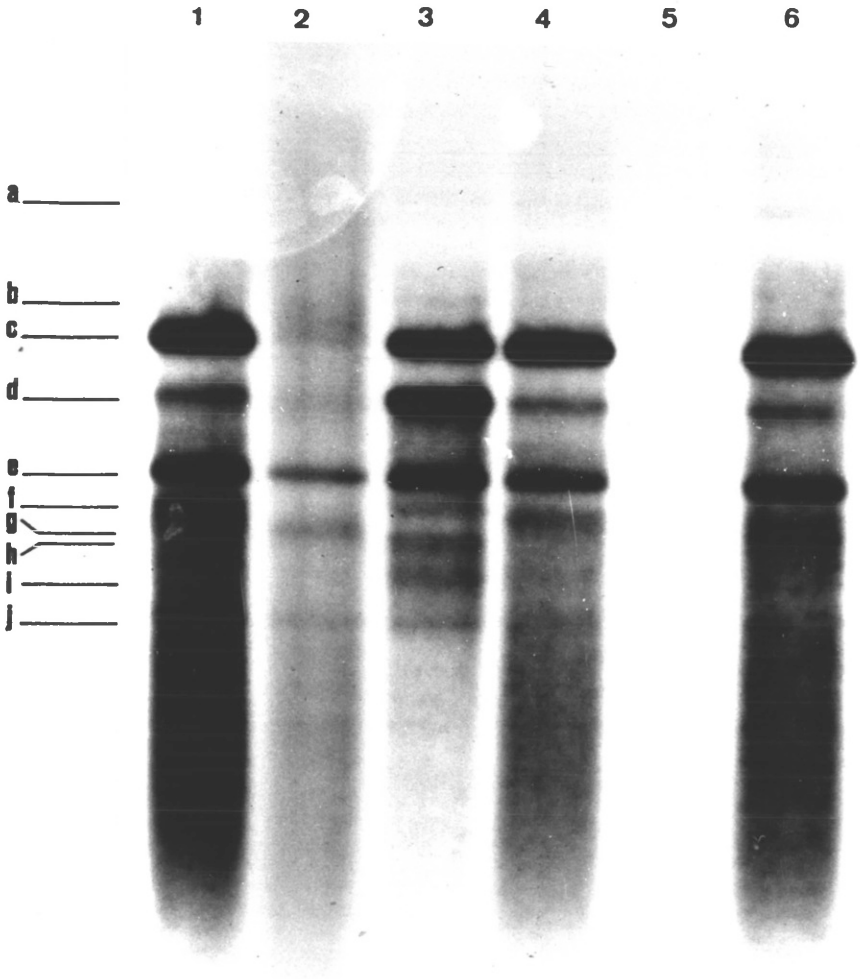



Fig. 50. **Intracellular WDV DNA species.**

Autoradiogram of a blot (from a 1.4% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated gel purified WDV oc dsDNA.
(1) WDV crude DNA; WDV crude DNA digested with (2) Exonuclease III, (3) S1 nuclease, (4) RNAse A, (5) DNAse I; (6) WDV crude DNA.
See section 3.15 for proposed identities of bands a to j.

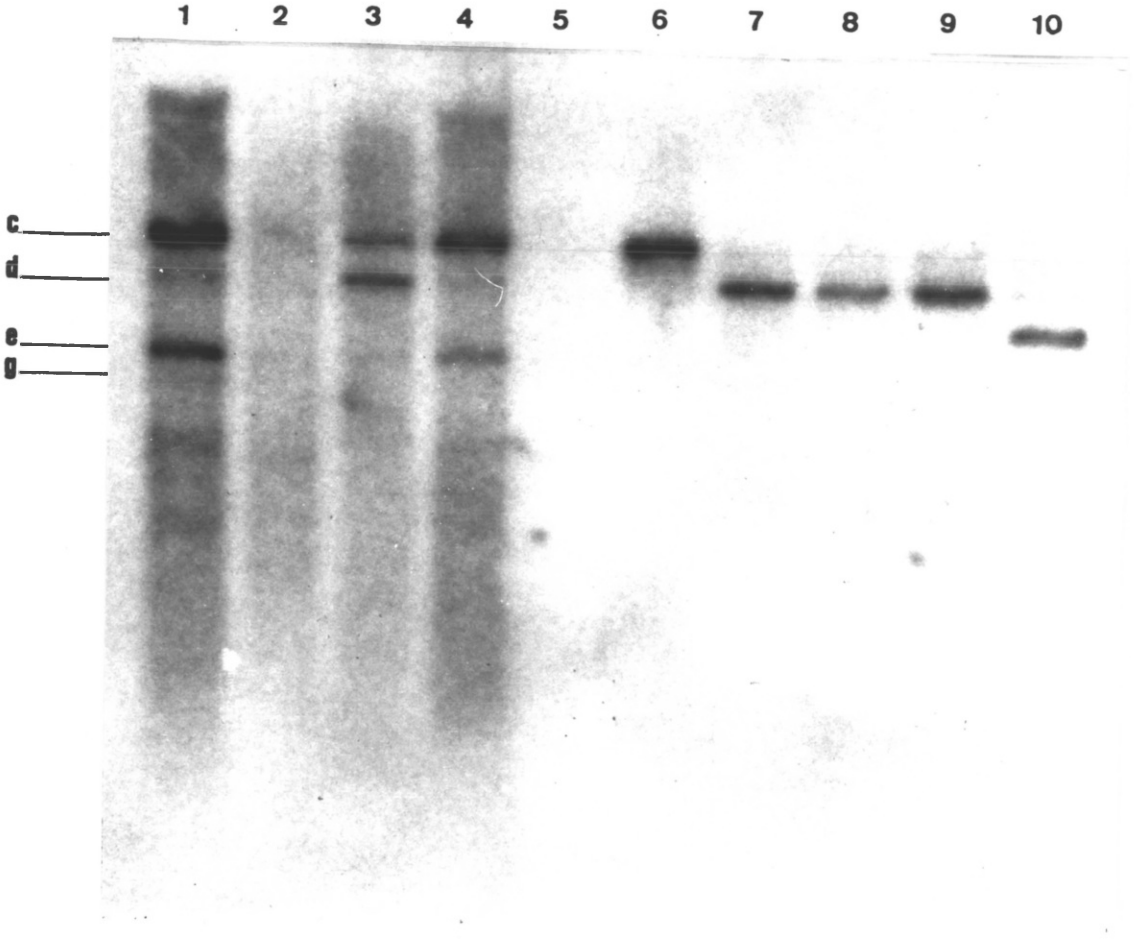



Fig. 51. **Restriction analysis of WDV oc dsDNA.**

Autoradiogram of a blot (from a 1.0% agarose gel) probed with cDNA made to purified WDV virion ssDNA.
(1) to (5) as in Fig. 50; (6) agarose gel purified WDV oc dsDNA; WDV oc dsDNA digested with restriction endonucleases (7) *Hin*dIII, (8) *Sma*I, (9) *Cla*I, (10) *Eco*RI.
See section 3.15 for proposed identities of bands c, d, e and g.

increase in the amount of the linear dsDNA relative to that of the circular dsDNA of the same species. Other subgenomic dsDNA species (bands f, h and i), possibly oc and linear forms, were also detected (Fig.50, lane 3) but their identity was difficult to ascertain. On the basis of the results shown in Fig.50, it would appear that there are at least two populations of differently-sized subgenomic DNA molecules. Indeed, when encapsidated WDV DNA was analysed by Southern blot-hybridization experiments, a number of discrete bands of less than genomic-length were detected (H.Macdonald, personal communication).

One of the problems in identifying 'band e' as WDV ccc dsDNA was its apparent sensitivity to Exonuclease III (Figs. 50 and 51, lane 2). In these and in other experiments, the amount of this DNA species was reduced in the presence of Exonuclease III. As mentioned previously (section 3.6.3), this enzyme specifically hydrolyses dsDNA molecules having 3'-OH termini and therefore should have no effect on ccc dsDNA. However, there is some circumstantial evidence to suggest that this band may indeed be the ccc form of genomic-length WDV dsDNA. Firstly, 'band c' and 'band d', which are probably the oc and linear forms of WDV dsDNA, were virtually completely digested by Exonuclease III, whereas 'band e' was only partly digested (Fig.50, lane 2). Secondly, the sedimentation properties of this DNA species in cesium chloride/ethidium bromide gradients are consistent with that of ccc dsDNA. Fractions collected from a cesium chloride/ethidium bromide gradient containing WDV crude DNA are shown in Fig.52 (lanes 2 to 16). Although the positions of 'band e' and 'band c' in the gradient overlap to some extent (Fig.52, lanes 6 to 9), the amount of 'band e' relative to that of 'band c' is higher towards the bottom of the gradient (Fig.52, lanes 5 and 6) and lower towards the top (Fig.52, lanes 9 to 16). These results indicate that 'band e' has a greater density than 'band c' (WDV oc dsDNA) which is consistent with the former being ccc dsDNA. A second DNA species (band j) co-sedimenting with 'band e', but with higher electrophoretic mobility, is probably a subgenomic ccc dsDNA species (band j in Fig. 50). The DNA at the bottom of the gradient (Fig.52, lanes 2 to 4), on the other hand, is probably ssDNA, which has a greater density than dsDNA.

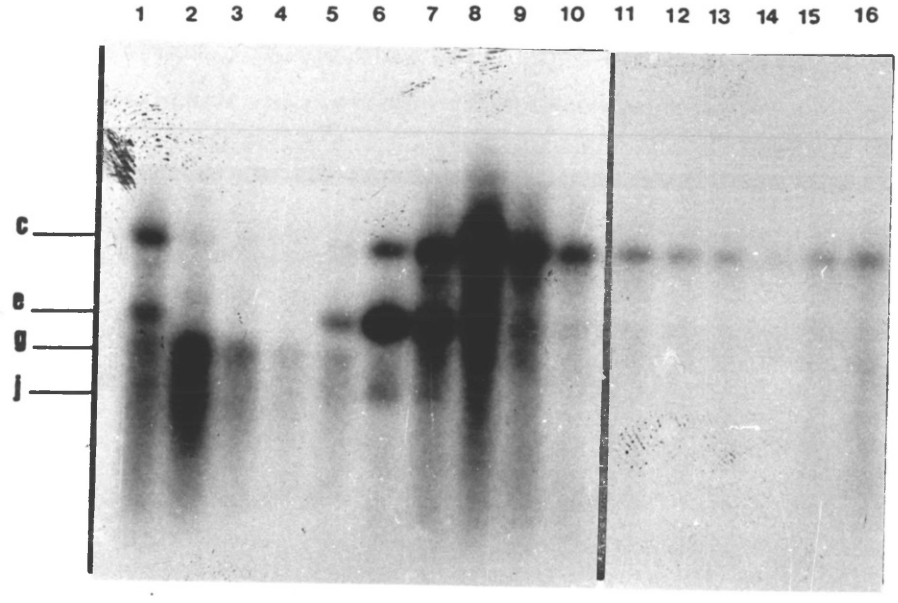



Fig. 52. Centrifugation of WDV crude DNA in a cesium chloride/ethidium bromide gradient.

Autoradiogram of a blot (from a 1.0% agarose gel containing 0.5µg/ml ethidium bromide) probed with nick-translated gel purified WDV oc dsDNA.
(1) WDV crude DNA; (2) to (16) fractions 1 to 15 from bottom to top of the gradient.
See section 3.15 for proposed identities of bands c, e, g, and j.

Thirdly, the electrophoretic mobility of 'band e' in relation to the other viral bands is similar to that of TGMV ccc dsDNA which is also consistent with it being the ccc form of WDV dsDNA. The apparent sensitivity of 'band e' to Exonuclease III is probably best explained by the presence of small amounts of endonuclease activity in the enzyme preparation.

The detection of intracellular WDV genomic-length and dimeric dsDNA species similar to those described for TGMV (Hamilton *et al.*, 1982), CLV and MSV (Stanley and Townsend, 1985) provides further evidence for a role for these species in geminiviral DNA replication. A noteworthy difference between DNA extracts obtained from either WDV or TGMV infected plants is the much lower abundance of viral ssDNA relative to dsDNA in the former. It would be interesting to see whether this difference arises as a result of the different DNA extraction procedures used or whether it reflects a difference in the infection cycle of these viruses. In my experience, the proportion of viral ssDNA to dsDNA in DNA extracts obtained by either method from plants infected with TGMV is not greatly affected, suggesting that the observed differences may not be due to the extraction procedures.

### 3.16 Restriction analysis of WDV oc dsDNA

To find restriction enzymes that cleaved WDV $^{ds}$DNA at single sites, agarose gel purified WDV oc dsDNA was separately digested with restriction enzymes *Hin*dIII, *Sma*I, *Cla*I and *Eco*RI. The reaction products were fractionated on a 1.0% agarose gel containing 0.5µg/ml ethidium bromide, transferred to nitrocellulose membrane and probed with [$^{32}$P]-labelled cDNA made to purified viral ssDNA extracted from virus particles. The results are shown in Fig.51 (lanes 6 to 10). *Hin*dIII, *Sma*I and *Cla*I each produced a single DNA fragment (Fig.51, lanes 7 to 9) of similar electrophoretic mobility to that of viral linear dsDNA (Fig.51, band d), suggesting that each of these enzymes cleaved WDV DNA at a single site. The possibility existed, however, that one or more of these enzymes might have cleaved at two (or more) very closely situated sites, which might not have been detected within the limits of resolution of the gel used. Restriction

enzyme *Eco*RI, on the other hand, produced a fragment smaller than viral linear dsDNA (Fig.51, lane 10), indicating more than one site for this enzyme in WDV DNA. A smaller fragment(s), however, could not be detected. This could be explained either by poor DNA transfer to the membrane or, alternatively, by the presence of several closely situated *Eco*RI sites. In the latter case, if the fragments produced were too small, they might have migrated off the bottom of the gel. Small amounts of a fragment of similar size to viral linear dsDNA were also produced in the *Eco*RI reaction. This fragment could have been a partial product or, alternatively, represented a minor population of a DNA species with a single *Eco*RI site.

Thus far, all the leafhopper-transmitted geminiviruses that have been characterized at the molecular level, namely MSV (Mullineaux *et al.*, 1984; Howell, 1984) and BCTV (Stanley *et al.*, 1986), appear to have a single component genome. In the case of BCTV, this has been unambiguously demonstrated by the construction of infective clones from virus-specific dsDNA isolated from infected *Beta vulgaris* (Stanley *et al.*, 1986). Conclusive evidence for a single MSV DNA component is still lacking, however, due to inability to infect plants by mechanical inoculation with virion or with cloned DNA. This is thought to be due to difficulty in the monocotyledonous host plant in delivering the DNA to the site(s) where infection is initiated without incurring excessive cell damage (Mullineaux *et al.*, 1984). The pattern of fragments produced by digestion of WDV oc dsDNA with restriction enzymes *Hin*dIII, *Cla*I, *Sma*I and *Eco*RI also suggested the presence of a single DNA component for this virus. However, the possibility of two different DNA components with the same number of sites for the enzymes used could not be discounted. The small amount of linear DNA produced by digestion with *Eco*RI could be due to a second component having a single site for this enzyme. Also, a second component may have been present in levels too low to be detected by Southern blotting. It was expected that subsequent cloning and sequencing of WDV DNA would help to answer some of the questions raised by these results.

## 3.17 Molecular cloning of WDV dsDNA

Restriction enzyme analysis of WDV oc dsDNA indicated the (probable) presence of single sites for *Hin*dIII, *Sma*I and *Cla*I. In principle any of these sites could be used to clone full-length WDV DNA into the vector pEMBL9. This vector contains single sites for both *Hin*dIII and *Sma*I in the cloning region within the gene encoding β-galactosidase. However, unlike *Hin*dIII, which generates DNA molecules with cohesive ends, digestion with *Sma*I generates molecules with blunt ends, which are more difficult to ligate. On the other hand, the single *Cla*I site in pEMBL9 lies outside the β-galactosidase gene and therefore is not very useful for cloning, since insertion inactivation of β-galactosidase is used to select for recombinant plasmids. Although it would be possible to clone *Cla*I linearised WDV dsDNA into the unique *Acc*I site in the cloning region of pEMBL9, since both enzymes produce compatible termini, insertion into this site would result in a hybrid site which is not recognized by either of the parental enzymes, making it impossible to excise the full-length insert if required. For the above reasons it was decided to clone WDV dsDNA at its *Hin*dIII site.

Two attempts to clone gel-purified WDV oc dsDNA digested with *Hin*dIII were unsuccessful, probably due to the presence of contaminants in the DNA preparation. To avoid the elution step, it was decided to use viral ccc dsDNA purified by cesium chloride/ethidium bromide gradient centrifugation as source of DNA for cloning. WDV crude DNA was centrifuged in cesium chloride/ethidium bromide gradients and 0.5ml fractions were collected and processed as described in section 2.11. The DNA from each fraction was then fractionated by agarose gel electrophoresis, transferred to nitrocellulose membrane and probed with [$^{32}$P]-labelled nick-translated WDV oc dsDNA (Fig.52, lanes 2 to 16). Fractions 4 and 5 (Fig.52, lanes 5 and 6), which appeared to consist predominantly of WDV ccc dsDNA (see section 3.15), were pooled and digested with restriction enzyme *Hin*dIII. The reaction products were extracted with phenol/sevag, ethanol precipitated, resuspended in $dH_2O$ and ligated to *Hin*dIII

linearised pEMBL9 .The ligation mixture was used to transform competent *E.coli* JM101 cells and plated on bottom-agar plates containing ampicillin, X-Gal and IPTG (section 2.29; W.D.O. Hamilton, personal communication). Cells giving rise to white colonies were harvested and plasmid DNA extracted as described in section 2.27. Recombinant plasmids carrying WDV DNA inserts were selected by dot-blot analysis using [$^{32}$P]-labelled nick-translated WDV oc dsDNA as a probe and analysed by restriction enzyme digestion and agarose gel electrophoresis. All recombinant plasmids (named pWD) that hybridized to the WDV DNA probe, except one (pWD2), contained the same insert with respect to both size and the position of restriction sites (W.D.O. Hamilton, personal communication). The plasmid pWDV2 was subsequently demonstrated by restriction mapping and sequencing (H.Macdonald, personal communication) to be a clone of a *sub*genomic species of the cloned genomic-length DNA component. One of the recombinant clones containing apparently full-length WDV DNA, pWD1, was chosen as source of DNA for sequencing.

The fact that a single DNA component was cloned in this experiment and, subsequently, in another one using *Eco*RI digested WDV oc dsDNA (see section 3.18), strengthens the notion that WDV is a monopartite geminivirus. However, the existence of a minor second component, possibly of non-genomic-length, cannot yet be discounted.Similar to MSV, conclusive evidence for a single WDV DNA component will have to await the successful infection of host plants with cloned DNA. A possible way in which this could be achieved has been recently described by Grimsley *et al.* (1986). These workers have systemically infected turnip (*Brassica rapa*) plants with cloned CaMV DNA by inoculating wounded leaves with *A. tumefaciens* in which tandemly duplicated CaMV genomes had been inserted into the T-DNA of the Ti plasmid. Transfer of the T-DNA to host cells followed by escape of the viral genome, either by intramolecular recombination or *via* production of a 35s transcript, resulted in systemic infection. There is some evidence that the T-DNA of Ti-plasmids can be transferred to *Narcissus*, a monocotyledonous plant (Hooykaas-Van Slogteren *et al.*, 1984). There is, therefore, a possibility that 'Agroinfection' of monocotyledonous plants might be feasible.

Ti plasmid vectors have also been used to infect petunia plants with cloned TGMV DNA (Rogers *et al.*, 1986; see section 1.10). In this study, petunia leaf discs were transformed with either TGMV DNA 1 or DNA 2, as single or tandem inserts. Regenerated plants were phenotipically normal. Cross fertilization of plants carrying tandem inserts, but not single inserts, of either DNA component produced progeny plants showing symptoms typical of virus infection. Analysis of leaf DNA prepared from symptomatic progeny plants revealed the presence of free viral ss and dsDNA indicating that escape of the viral DNA components had occurred. This method, however, requires the regeneration of whole plants from single cells which, in the case of the gramineae, has yet to be achieved reproducibly.

## 3.18 Sequencing strategy

The strategy used to sequence cloned WDV DNA consisted of subcloning DNA fragments generated by restriction enzyme digestion of gel purified WDV DNA (full-length insert excised from pWD1 with restriction enzyme *Hin*dIII) into M13mp8 or mp9 vectors (Messing and Vieira, 1982; section 2.29.6) and to sequence these by the di-deoxy chain termination method (section 2.31). In a few cases, subclones were generated by digestion with Bal 31 as described in Maniatis *et al.* (1982) (a gift from Dr.W.D.O. Hamilton). Template DNA for sequencing was prepared as described in section 2.30.

The sequencing strategy is shown in Fig.53. Initially, restriction fragments were subcloned randomly and the different subclones obtained identified by 'T-Tracking' (section 2.31.3). It was already known that WDV DNA had recognition sites for *Hin*dIII, *Cla*I, *Sma*I and *Eco*RI (section 3.16). To find further restriction enzymes that could be used for subcloning, gel purified WDV DNA was digested (separately) with restriction enzymes *Hpa*II, *Sau*3A, *Acc*I, *Hae*III, *Pst*I, *Bam*HI, *Xho*I, *Taq*I, *Bgl*II and *Sal*I, all of which generate termini compatible with those generated by cleavage at the M13 cloning site (Fig.54). Only *Hpa*II, *Sau*3A, *Hae*III, *Taq*I and *Pst*I cleaved WDV DNA. Of these, all except *Pst*I recognize tetranucleotide sequences. It

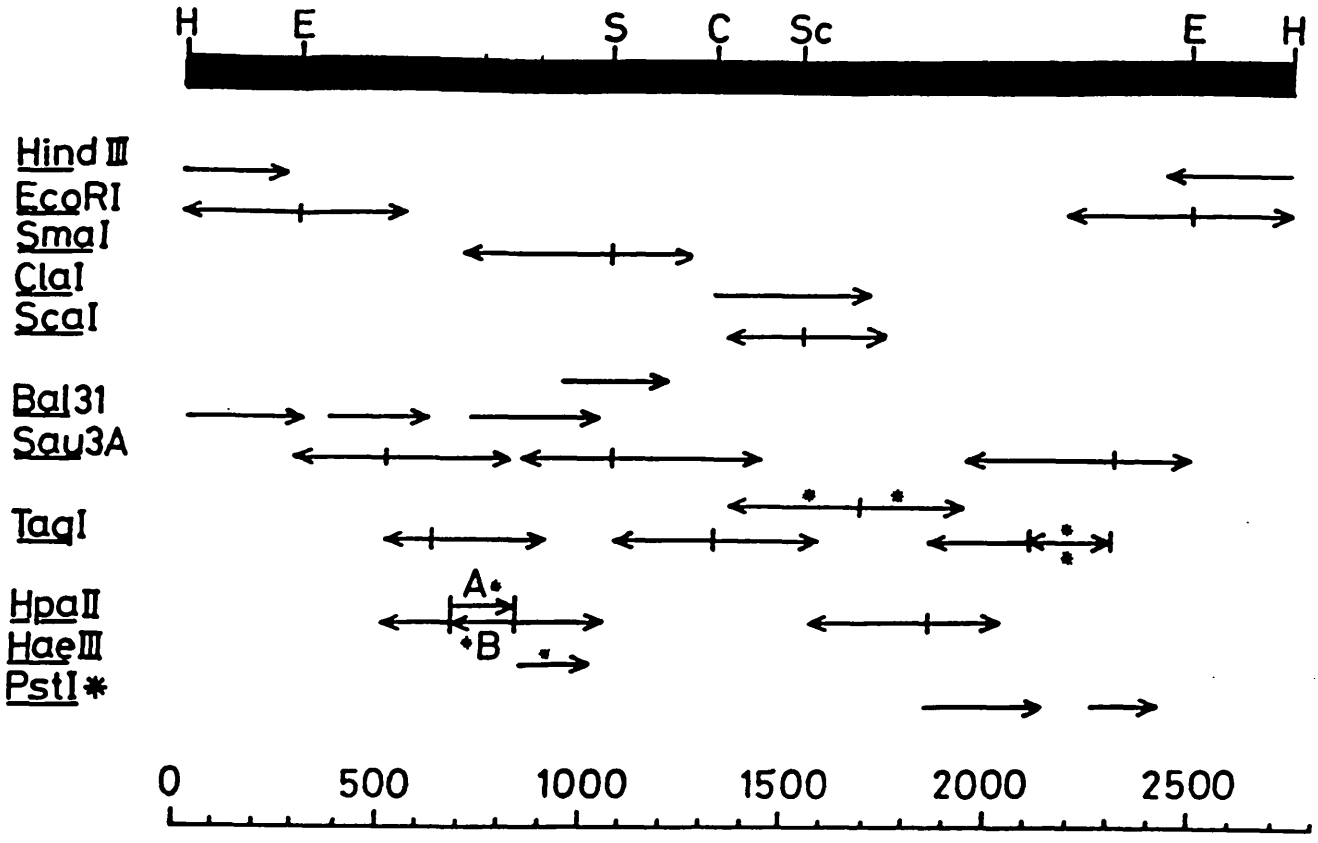



Fig. 53. DNA sequencing strategy.

The cloned WDV DNA is represented as the thick black line. Selected restriction sites are shown. H = *Hin*dIII, E = *Eco*RI, S = *Sma*I, C = *Cla*I, SC = *Sca*I. The arrows indicate the extent and direction of sequence obtained from each clone. The restriction enzyme used to generate the sequencing start-point is indicated at the left of each series of clones. The *Pst*I clones were generated by four base-specific cleavages by the enzyme *Pst*I. The scale indicates the distance in base pairs from the *Hin*dIII site. The labels A and B refer to clones used as templates for cDNA probes described in section 3.18. Clones labelled with an asterisk were sequenced by H. Macdonald.

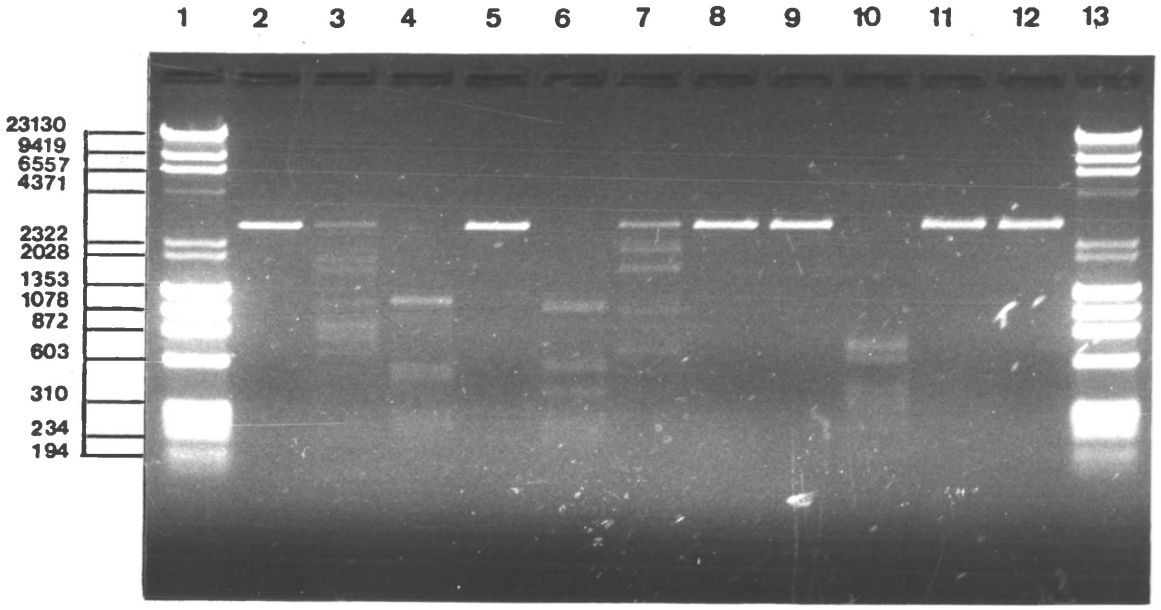



Fig. 54. **Restriction analysis of cloned WDV dsDNA.**

Electrophoresis in a 1.0% agarose gel containing 0.5µg/ml ethidium bromide.

(1) λ-phage DNA digested with restriction enzyme *Hin*dIII + ØX174 bacteriophage DNA digested with restriction enzyme *Hae*III; (2) gel purified WDV DNA (excised from pWD1 with restriction enzyme *Hin*dIII),; WDV DNA (*Hin*dIII linearized) digested with (3) *Hpa*II, (4) *Sau*3A, (5) *Acc*I, (6) *Hae*III, (7) *Pst*I, (8) *Bam*HI, (9) *Xho*I, (10) *Taq*I, (11) *Bgl*II, (12) *Sal*I; (13) as in (1).

was subsequently found that the *Pst*I fragments had been unintentionally generated by a star activity and that there were no *Pst*I sites in the WDV sequence. As more information about the sequence was obtained, it became possible to gel purify and subclone specific restriction fragments, thereby accelerating the sequencing procedure. Thus, most of the *Hpa*II, *Sau*3A and *Taq*I subclones were generated from the WDV DNA derived 2.2Kb *Eco*RI fragment. A number of fragments were cloned in both M13mp8 and mp9, thereby allowing their sequences to be determined in both orientations. Sequencing also revealed the presence of the unique ScaI recognition site which was used to generate two subclones (see Fig.53).

The sequence was obtained by noting the most intense band at each level on the sequencing gel. In some instances, however, regions of ambiguity were encountered where interpretation of the sequencing data was extremely difficult. These included the presence of a band of similar intensity across all the tracks in a particular region, variations in band spacing (compressions) possibly due to local secondary structure and the appearence of each band as a doublet or triplet. To aid in the interpretation of such data the following rules were applied (Davies, 1982; Hindley, 1983):

- Upper C is always more intense than lower C.
- Upper G is often more intense than lower G (particularly when the double G is preceded by a T).
- Upper A is often less intense than lower A

When interpreting the sequence data, the spaces between between bands were also taken into account, to ensure that the proposed sequence fitted into the space available. In some instances, however, unambiguous interpretation of the sequence data was only possible by sequencing the same region in both directions.

The sequences derived from the various subclones were assembled visually by searching for overlapping regions. More than 99% of the WDV genome was sequenced in both orientations, leaving no ambiguous section. The assembled sequence was stored in a VAX computer for subsequent analysis using the Program Library of the University of Wisconsin Genetics Computer Group (Deveraux *et al.*, 1984).

The WDV DNA sequence is shown in Fig.57 in the virion sense and consists of 2749 nucleotides. The virion strand was identified by hybridization of [$^{32}$P]-labelled cDNA from two *Hpa*II subclones (labelled A and B in Fig.53) to virion ssDNA. Only cDNA to clone B hybridized and thus the sequence obtained from this clone and all other clones in this orientation were complementary to the virion DNA (H.Macdonald, personal communication).

Analysis of the WDV DNA sequence with the MAPSORT progam (Deveraux *et al.*, 1984) revealed the presence of two *Eco*RI sites, both within sequencing distance from the *Hin*dIII (cloning) site. To sequence across the *Hin*dIII site, gel purified WDV oc dsDNA (see section 3.14) was digested with *Eco*RI and the reaction products ligated to *Eco*RI digested M13mp9. Recombinant phage DNA were identified by 'T-Tracking' (Fig.55). Four types of subclones were obtained. These corresponded to the two *Eco*RI fragments, each in both orientations, expected from similar DNA molecules as pWD1. Sequencing of one *Eco*RI subclone across the *Hin*dIII site demonstrated the full-length nature of the cloned WDV DNA.

## 3.19 Potential genes

To investigate the potential coding capacity of the genome, the WDV DNA sequence was screened for open coding regions (i.e. regions that are not interrupted by one of the nonsense codons TAA, TGA or TAG) and open reading frames (i.e. open coding regions starting with an ATG triplet) in all three reading frames on both the viral (+) and complementary strands (-) using the MAP program (Deveraux *et al.*, 1984) (Fig.58 A and B). Twelve open coding regions (OR) longer than 200nt were found (Fig.58 A), two in reading frame +1 (nucleotides 222-438 and 1144-1407), one in reading frame +2 (nucleotides 245 to 526), three in reading frame +3 (nucleotides 498-1280, 1380-1604 and 2043-2079), two in reading frame -1 (nucleotides 711-472 and 2178-1960), one in reading frame -2 (nucleotides 2620-1799) and three in reading frame -3 (nucleotides 777-357, 1292-852 and 1904-1452). Four of these ORs comprise open reading frames (ORF) longer than 200nt (Fig.58B) and, assuming that translation initiates at the

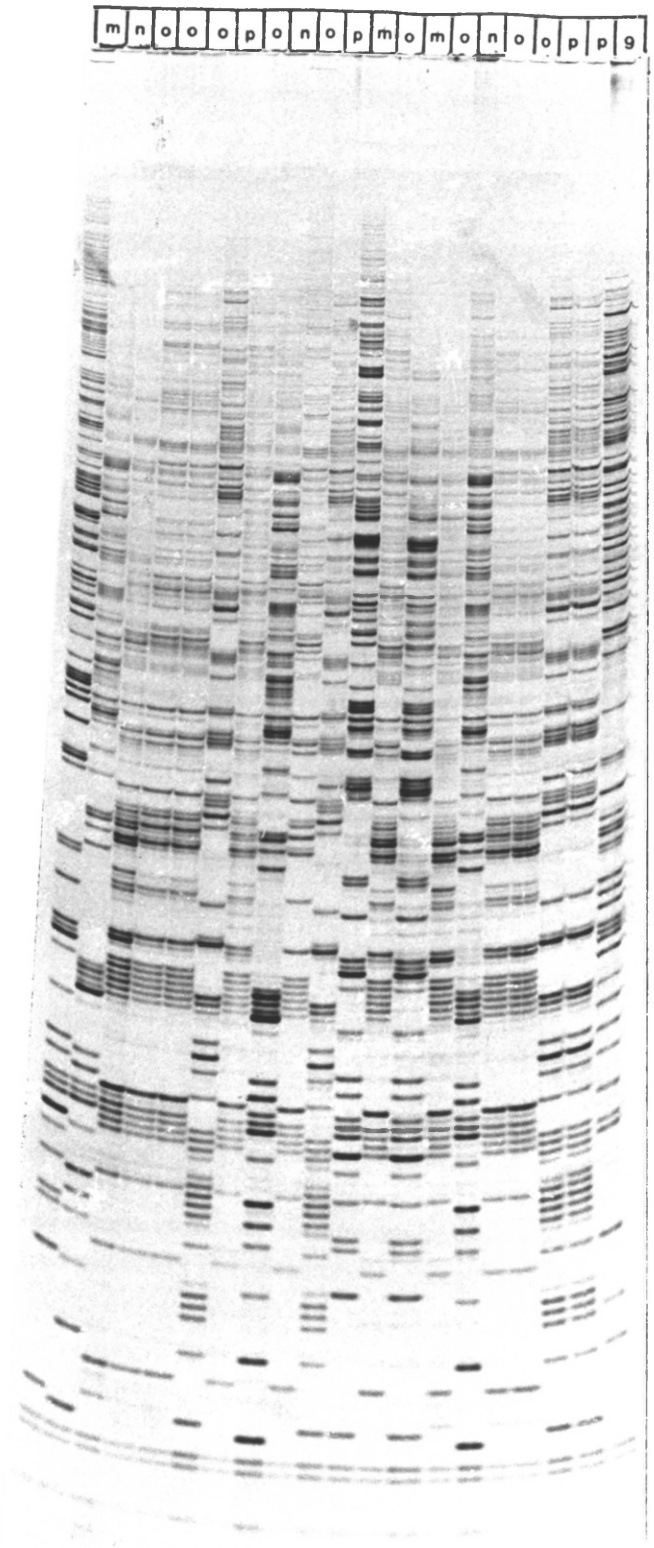



Fig. 55. **'T-Tracks' of independently derived WDV DNA *Eco*RI subclones.**

The four types of subclones m, n, o and p correspond to the two fragments, each in both orientations, expected from DNA molecules having the same sequence as the pWD1 insert. 9 = M13mp9.

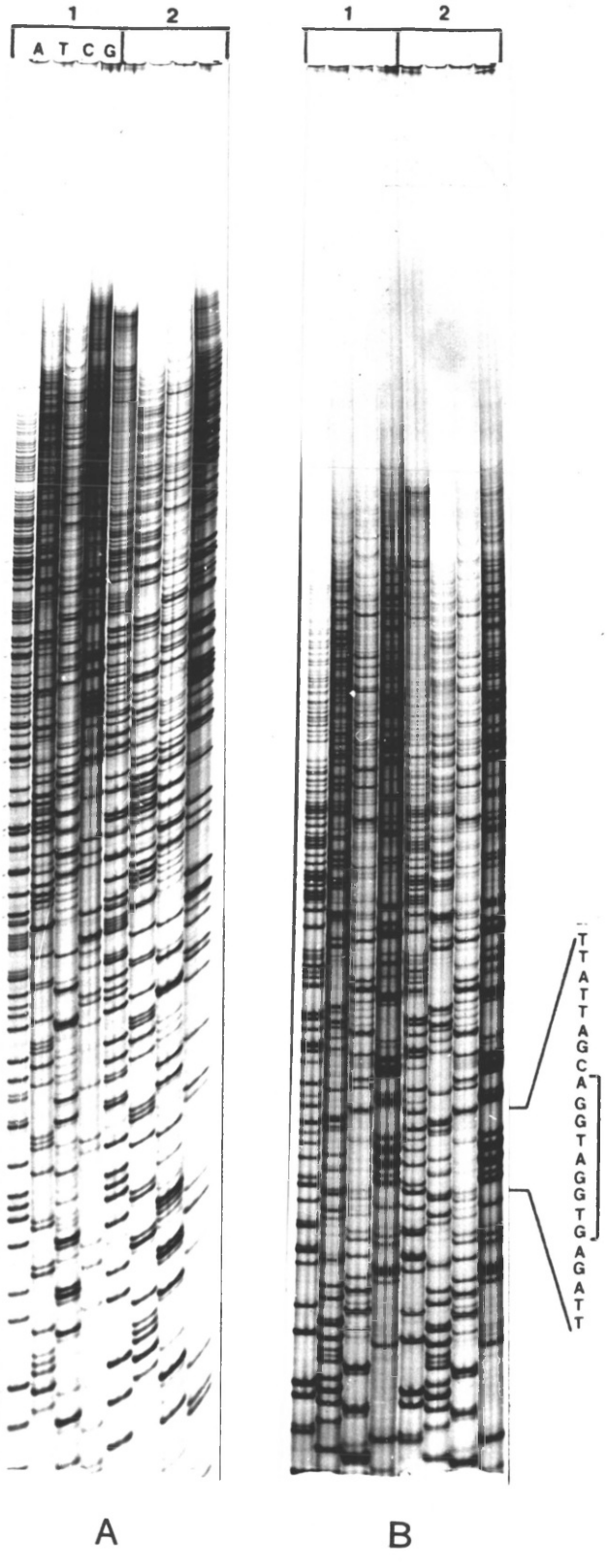



Fig. 56. **Sequence of two independently derived *Eco*RI subclones.**

Sequencing gel of two *Eco*RI subclones obtained obtained from *Eco*RI digested WDV oc dsDNA as described in section 3.18. The sequence obtained from these subclones (type p, see Fig.55) is the same as that from pWD1 starting at nucleotide 2075 in the complementary sense. The sequence GTG GAT GGA (see section 3.22) is indicated by a vertical line. A, T, C and G reactions are as indicated in (1). A) 2hr run, B) 4hr run.

Legend to Fig. 57. **Nucleotide sequence of WDV.**

The virion DNA strand is shown starting within the intergenic region containing the major potential hairpin structures (see section 3.21).

Sequence of wheat dwarf virus DNA Length: 2749 ..

```
   1  CTAGTGTTCC CACGGTAGCG TAGCGAATCT TGTGGGCCCT GTTCGGTGTG
  51  CGGTCGGGGG GCCTCCACGC GGGTTATAAT ATTACCCCGC GTGGTGGCCC
 101  CCGACGCGCA CTCGGCTTTT CGTGAGTGCG CGGAGGCTTT TGGACCACAT
 151  CTTTTCTGAT CACTTTCGTG GAAGATGTTG ATTTATCACA CTTTTGACGG
 201  GGAAATCTGT GCCATGCCTT AGCTTATAAG GAAGTGCGTG GTAGCCCATC
 251  TCGATGGAGC AGGCAATAGC CCCCCCGCTT CCTATACGGG ACTATCAATA
 301  CCAGACCCCT TCCATTCCCG GTTCCTCCGA CTACGCCTGG CGAACATTTG
 351  TGTTCGTTAC CTTTGGTTTG CTAATAGCCG TAGGCGTTGC TTGGCTTGCT
 401  TACACTCTGT TTCTGAAAGA TTTAATTTTA GTGTGTAAGG CGAAGAAGCA
 451  AAGGAGGACC GAGGAAATTG GTTACGGGAA TACACCGGCC AGATTAAATG
 501  GTGACCAACA AGGACTCCCG AGGTAAGGGG AAGCGGAAGA TGGAAGAAGG
 551  TGAATCCAGC GGAAGGTGGA AGGGGGCTGT GTATAAGCGG CGTAAACAGG
 601  CGTATAAGGT AGTACCTGTG AAGCCCCCAG CTCTCTGCGT ATTCCGCTAC
 651  AACTGGTTGA ATAGCGACAG GACCAATATT GTTGTGGGTA ATACACCCCG
 701  GGTCGATCTG ATTACCTGTT TTGCTCAGGG TAAGGCCGAT AATAATCGGC
 751  ATACAAATCA GACCGTCCTA TACAAATTTA ACATACAGGG TACCTGCTAT
 801  ATGTCCGATG CATCAGCTCC GTTCATCGGT CCAGTCCGCC TCTACCACTG
 851  GTTAGTCTAT GATGCAGAGC CGAAACAGGC TATGCCAGAC GCCACTGACA
 901  TCTTTACGAT GCCTTGGAAT CTGCTGCCGA GTACGTGGAC TGTGCAACGT
 951  GCTTGGTCGC ATCGATTCGT GGTGAAAAGG AAGTGGACCG TGAACCTTGT
1001  TACTGATGGA CGGAAGGTCG GGTCTAAGAC CGTTGACCAG CGCTACAACT
1051  GGGTAGTCGG CAAGAATATC GTTGACGCAA ATAAGTTCTT CAAAGGTTTG
1101  CGTGTCACGA CGGAGTGGAT GAACACGGGT GACGGCAAGA TAGGCGACAT
1151  TAAGAAGGGA GCACTGTATC TTATTAGCAG TACTCGTGGT GGTGTTACTG
1201  GTGATAGTGC CTCTACGGCG TTTGATGTTG TATGTGCCTA TACGCACGCG
1251  TGTTATTTCA AAGCCATCGG CATTCAGTAA TAAAATAATA TTTTATTTAT
1301  CTCATGTCAT TCGATTACAG AGGCTCGGCT ACGAGCAAAG ACAAACCAAA
1351  TATAACAAAC AACAACCCTT ACACAATGAC ATCGGAAAAC GAAATACAAC
```

1401 ACCCTGAGAT ATTACATTTA TAGAAACTGT ACGCCGTCCG CGCTAGGACA

1451 GTCACTGCGA AGCAGTGACG TCTTCGCCGG AGGCGAACGA GTAGTTGATG

1501 AACGTCTCGC CTTCATACAT GTAGTGAACA ACAGTGTTAG AGTACATGTA

1551 ATCCGACTGT TCGGGAGTCA TATCCTTGAG CCAATCTTCG TCTGGATTAA

1601 CTAAAATGAT GCAAGGTATT CCACCCCGTA TGACCTTTCG CTTACCATAT

1651 TTTGGATTGA CCGTGAAGTC ACGCTGAGCC CCGACGAAGC ACTTCCAGTT

1701 GGGTGTGAAC TTGAATGGAA TGTCGTCGAT GATATTATAC TTGGCGTTGA

1751 CGTCATATGT TGTGAAATCA ACTAGACTGT TATAATAATT GTGTGTCCCT

1801 AGAGACCTTG CCCAGGAAGT CTTTCCTGTT CTGGTTGGCC CGCAGATGTA

1851 GATGGACTTA TGCCTCCCCG GTGACTCCTG GAATAATCGT CCATCCACTC

1901 TAAGTCAGAT TGCGCTTGAT CCGCAGGAGT GGAAGTACAA AGGATATAGG

1951 ATTCGAGGCT TACGGAGTAG AGATGTTCAT TTTTCCAGCT TTCAATGGTC

2001 TCATGGCAAA TGAGTGATTC GGTTGGAAAC TCAGGTGTGT AAGTGGCAAC

2051 TGGGTCAGGA AATAGATGGC GTGCCGTGTA CTCGAAGTCT TTGAGACGGA

2101 TAGACCATTC AAACGGAAAA CGATTGCAAA CCATGCTGAG GAATTCCTCG

2151 CGAGAGGAAC TAGATTCAAT GATCTGTTTC ATATCCGCAT CACGGTCTTT

2201 ACGACCTGGA GTTGAAACAG CCACGAATGT TCCCCACTCA GCTGTGTTTA

2251 CATCGGAGTC AACCTCCTTC GTGATGTAAT CACGAACTTG GTTGCAGTCT

2301 TTGGCAGCTT GTATATTTGG ATGGAATATG GAGAATGGAG ATGTATCCAT

2351 ACGGAGGTTT AAGGCATTGG GATTGGTGAT GGAAGCACGA AGCTTGTTCT

2401 GCACGAGAAC GTGCAGATGT GGTGATCCAT CTTCGTGGAG CTCTCTAACA

2451 GCAGCGATGT AGAGGGGCTC ATATTTGTTC AAGAGAGTGC GAAGTGAATC

2501 CAAGGCGTAC TGTGGCTCAA GGGTACATTC AGGATATGTT AGAAAGAGGT

2551 ACTTGGAATA GACACGGAAC CTGGGTGCAG ATGAAGAGGC CATGGTAGTG

2601 AACAGAAGTC CGGCAGGTCC TTAGCGAAAA AACGGGGTGT GCCAGAAAAC

2651 TCTATCCTCT ACCCTGCGTG GAGGTGTGAA TTCTGCACAC TGCAAATGCA

2701 ATGTGTCCAA TGCTTTATAT AGGGCAGGTT TTGGCGGGAG AACAGGGCC

proximal 5' ATG triplet (see below), all four ORFs have the potential to code for proteins with mol.wts. ≥10.000. The exact location of these ORFs and the number of amino acids encoded by each of them are shown in Table 4. Thus, when read in the virion sense the DNA sequence could encode two proteins of mol.wt. 10146 and 29407. In the complementary sense the DNA could encode two proteins of mol.wt. 14556 and 30156. These are shown in relation to the circular DNA genome in Fig.59. The possibility that some of the ORs that do not comprise extensive ORFs may be expressed by splicing or other post-transcriptional processing mechanism will be discussed in section 3.22.

In over 90% of mRNAs whose sequences are known, protein synthesis initiates at the first AUG triplet. This observation has led to the 'scanning' hypothesis, according to which the eukaryotic ribosome moves down the mRNA from the 5' end until it encounters the first AUG codon, whereupon translation is initiated (Kozak, 1983). In a few cases, however, initiation has been found to occur at AUG codons located downstream of the first AUG triplet, presumably because the latter is embedded in a sequence that poorly matches the consensus sequence CCA/GCCAUGG for the start of translation proposed by Kozak (1984) (Hunt, 1985). The predicted start site for ORF 30156 corresponds closely to the consensus sequence for the start of translation (Table 5). There is also some similarity between this consensus sequence and the start site for ORF 10146. In contrast, the start sites for ORFs 29407 and 14556 do not conform to this consensus. However, ORF 29407 may use an alternative start codon 42 nucleotides downstream of the first ATG. There is a preferred purine (A) three nucleotides upstream of this second ATG codon (Table 5).

## 3.20 Potential transcriptional control signals

Eukaryotic mRNA is synthesized by RNA polymerase II in the cell nucleus. RNA polymerase binds to a region in the 5' flanking sequence of the gene, known as the promoter region, initiates a mRNA chain and then copies through the gene until it reaches a termination signal. Almost all mRNA transcripts are further

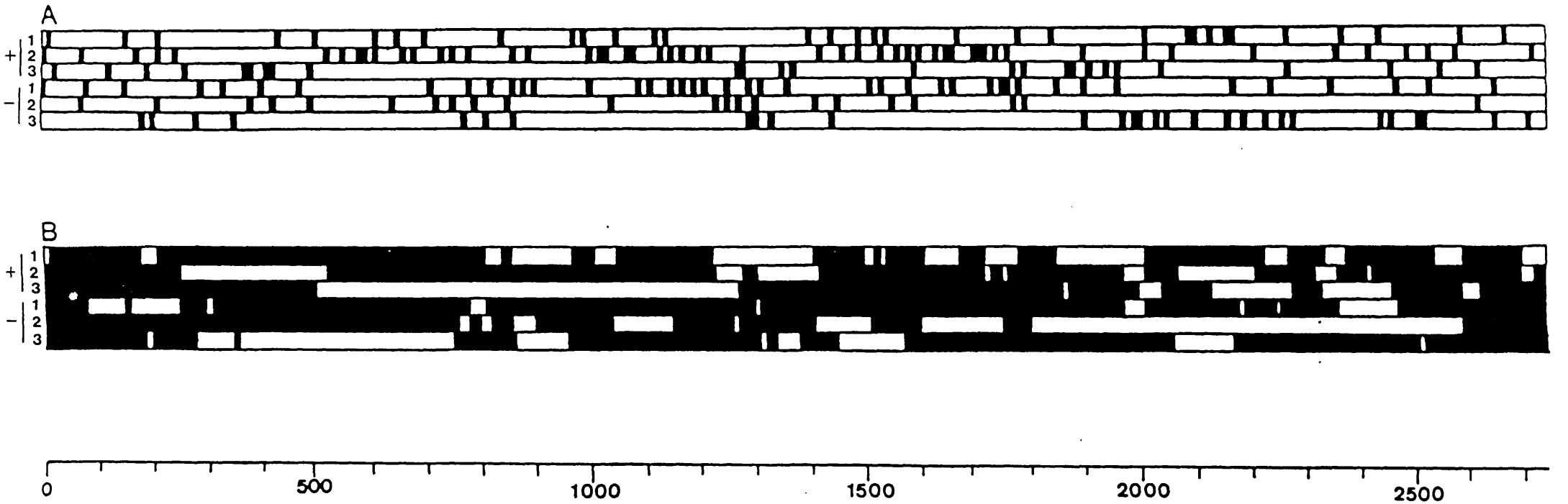



Fig. 58. Open reading regions and open reading frames in the WDV DNA sequence.

A) Open reading regions in the WDV DNA sequence. The open reading regions found for the virion (+) and complementary (-) strands of WDV DNA are shown. Each reading frame was divided into blocks of ten and shaded if it contained the second base of a stop codon.

B) Open reading frames in the WDV DNA sequence. The open reading frames starting with the 5' proximal ATG codon are shown as open rectangles in all six reading frames. The scale is in base pairs.

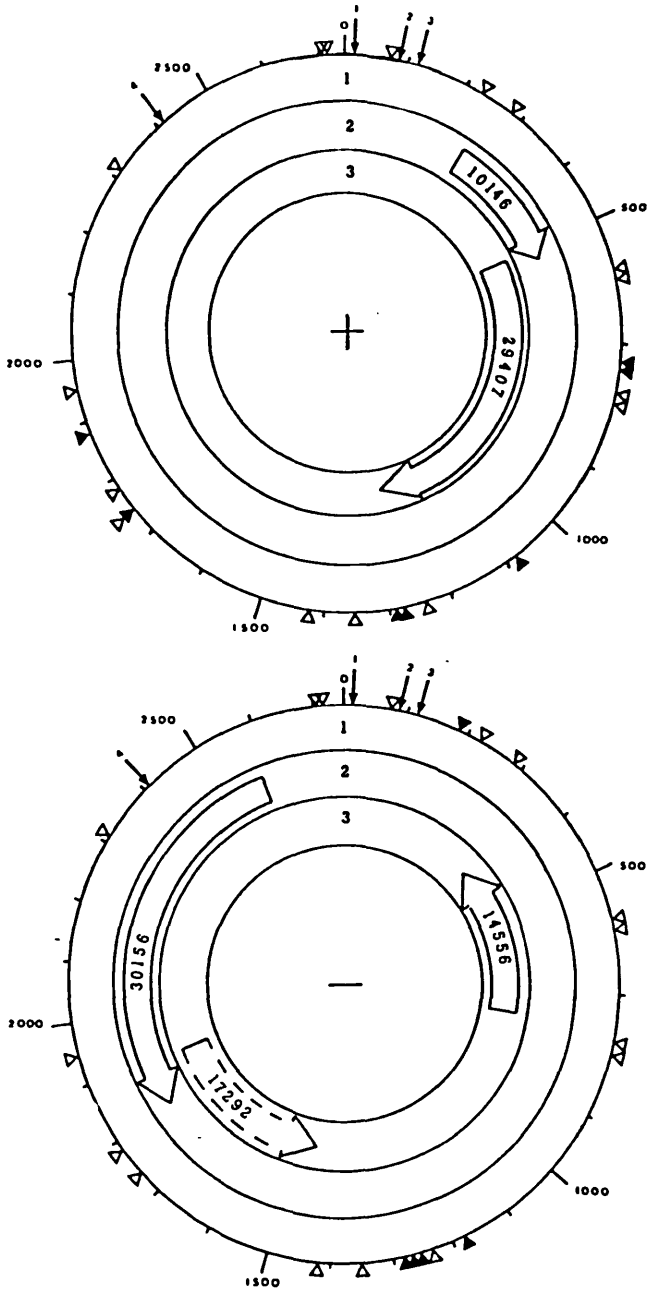



Fig. 59. **Potential coding regions in WDV DNA.**

Open reading frames starting with an ATG triplet and coding for a polypeptide with mol.wt.≥10000 are shown as open arrows. The mol.wt. of the potential product is given within each arrow. The stippled arrow represents the potential coding region on the sequence which may be spliced onto ORF 30156 (see section 3.22). Open triangles indicate the position of TATA boxes and shaded triangles potential polyadenylation signals (G/AATAA). The arrows numbered 1 to 4 show the position of the potential hairpin structures detailed in Fig.61.

processed to form the mature mRNA molecules which are then transported to the cell cytoplasm where translation takes place.

Comparison of the nucleotide sequences flanking a vast number of eukaryotic genes has revealed short regions of homology which have been implicated in the initiation of transcription, translation and in processing of the primary transcript. Although most of the information available has been obtained from animal genes, present evidence suggests that plant genes have similar regulatory sequences to those found in animal genes (Messing *et al.*, 1983).

An almost ubiquitous sequence, located 20-30 bases upstream of the transcriptional initiation site (cap site), is thought to be involved in positioning RNAse polymerase II for correct initiation. This sequence, known as the TATA or Hogness-box, has the consensus sequence TATAA/TA/T(Breathnach and Chambon, 1981). Experiments *in vivo* and *in vitro* have shown that this sequence is required for accurate and efficient initiation of transcription (Grosveld *et al.*, 1981; McKnight and Kingsbury, 1982).

In addition to the TATA box, a sequence located 80-100 nucleotides upstream of the initiation point of the mRNA may also be involved in regulation of transcription of some animal genes. This sequence, known as the CAAT box, has the consensus sequence GGC/TCAATCT (Benoist *et al*, 1980). This sequence is required for efficient transcription of some animal genes *in vivo*, e.g. the rabbit β-globin gene (Grosveld *et al.*, 1982) but not for others, e.g. the thymidine kinase gene (McKnight and Kingsbury, 1982). Deletion of the CAAT box has no effect on transcription *in vitro*. Sequences with limited homology to the CAAT box have also been found in plant plant genes and the consensus sequence $C/TA_{2-5}G/TNGA_{2-4}C/TC/T$ has been proposed for the plant counterpart (the AGGA box) to the animal CAAT box (Messing *et al.*, 1983).

Most eukaryotic mRNAs have a sequence of polyadenylic acid (poly (A) tail) at the 3' end, the significance of which is not clearly understood. The poly (A) tail is added by the enzyme poly (A) polymerase to the RNA in the nucleus after transcription. Most polyadenylated animal mRNAs, but not of yeast mRNAs, have the sequence AATAAA located about 10-30 bases from the poly (A) tail (Proudfoot and Brownlee, 1976). Studies *in vivo*

| Protein mol. wt. | Reading frame | Start | Stop | Amino Acids |
|---|---|---|---|---|
| 10146 | +2 | 254 | 526 | 90 |
| 29407 | +3 | 498 | 1280 | 260 |
| 14556 | -3 | 752 | 357 | 131 |
| 30156 | -3 | 2593 | 1799 | 264 |
| 17292* | -3 | 1904 | 1452 | 150 |

Table 4. **Open reading frames in WDV DNA.**

*Open reading region which may be spliced onto ORF 30156 (see section 3.22).

| ORF | start site |
|---|---|
| 10146 | 249: tCtCgATGG |
| 29407 | 493: attaaATGG or<br>535: ggAagATGG |
| 14556 | 759: tttgtATGc |
| 30156 | 2598: CtACCATGG |
| consensus* | $CC^{A}_{G}CCATGG$ |

Table 5. **Nucleotide sequence at the predicted start sites for the WDV open reading frames.**

An alternative start site for ORF 29407 is shown (see section 3.19)
*Consensus sequence for the start of translation (Kozak, 1984).
Homologies to the consensus sequence for the start of translation are shown in upper case

using SV40 late genes suggest that this sequence is required for polyadenylation and determines the position of the poly (A) tail (Fitzgerald and Shenk, 1981). Similar sequences are also present in the majority of the plant genes examined and the consensus sequence G/ATAA$_{1-3}$ has been proposed for the polyadenylation signal in plants (Messing *et al.*, 1983). Unlike animal genes, most plant genes have two putative polyadenylation signals and these are located closer to the stop codon than to the poly (A) tail (Messing *et al.*, 1983).

To locate potential transcriptional control signals associated with the ORFs, the WDV DNA sequence was screened for eukaryotic promoter regions and polyadenylation signals using the FIND program (Deveraux *et al.*, 1984). The positions of potential promoter regions and polyadenylation signals in the WDV genome are shown in Fig.59.

The proposed complete consensus sequence for plant gene promoters is TC/GTATAT/A$_{1-3}$C/TA (Messing *et al.*, 1983). However, since this consensus was derived from very few (12) sequences including only three monocotyledonous genes, the search for potential promoters in the WDV genome was limited to the core sequence (TATA). All the potential genes have at least one TATA box within 250bp upstream of the initiation codon. The position of the most likely TATA boxes for each ORF and their homology to the consensus sequence are indicated in Table 6. For all ORFs, with the exception of ORF 29407, there is a potential promoter TATA box with >60% homology to the consensus sequence for plant promoters. Although sequences resembling the animal CAAT box consensus sequence were found in the WDV DNA sequence (Table 7), none appeared to be suitably positioned relative to the most likely TATA boxes for each of the ORFs. None of the ORFs were found to have AGGA boxes.

All ORFs with the exception of ORF 30156 possess a G/ATAA sequence within 250bp of the termination codon. The positions of the most likely polyadenylation signals for each ORF are indicated in Table 6. In the case of ORF 29407, the nearest polyadenylation signal overlaps the TAA termination codon and is part of an AT-rich region. The nearest potential polyadenylation signals for ORF 30156 in the opposite orientation are also

contained within this AT-rich region (Fig.60). Interestingly a similar situation occurs in the DNAs of all the geminiviruses that have been sequenced (Fig.60). In all cases the pairs of polyadenylation signals in opposite orientations form, or are part of inverted repeats. The significance, if any, of this structure is not known. Its preservation in the DNAs of all the geminiviruses that have been sequenced suggests that it may have a function. Transcriptional termination signals (terminator) in bacteria contain inverted repeats just prior to the point of termination (Adhya and Gottesman, 1978). By analogy, it is possible the the AT-rich region in the geminiviral DNAs may act as a terminator. The location of the 3' end of the CLV coat protein transcript immediately downstream of this region (Townsend *et al.*, 1985) and of that of the MSV coat protein only about 40 bases downstream (Morris-Krsinisch *et al.*, 1985) may be consistent with this hypothesis.

The distribution of the major ORFs and associated TATA and G/ATAA boxes in the WDV genome suggests that, like MSV, CLV, TGMV, BGMV and BCTV, transcription is bidirectional. This has been confirmed for both CLV (Townsend *et al.*, 1985) and MSV-N (Morris-Krsinich *et al.*, 1985) by the identification and mapping of polyadenylated transcripts isolated from infected plants and will undoubtedly be confirmed for the other geminiviruses.

## 3.21 Non-coding regions

There are two apparently non-coding regions situated almost diammetrically opposite on the WDV circular genome (Fig.59). Assuming that all the major ORFs are involved in protein coding, the first intergenic region would comprise nucleotides 2594 to 253, and the second nucleotides 1281 to 1788. However, if the ORF between nucleotides 1904 and 1452 (OR 17292) in the complementary strand is involved in protein coding, as is suggested in section 3.22, the second intergenic region would be between nucleotides 1281 and 1451.

The genomes of leafhopper transmitted geminiviruses MSV-N and MSV-K also contain two diametrically opposed intergenic regions (Mullineaux *et al.*;, 1984; Howell, 1984). In the case of MSV-K these have been

| ORF | TATA | | G/ATAA | |
|---|---|---|---|---|
| 10146 | 73: | gtTATAATATt (-181) | 738: | GATAAtA (+212) |
| | 223: | ctTATAAggAa ( -31) | 741: | AATAAtc (+215) |
| 29407 | 223: | ctTATAAggAa (-275) | 1279: | AATAAAA ( +1) |
| | 281: | cCTATAcgggA (-217) | 1284: | AATAAtA ( -4) |
| 14556 | 774: | TGTATAggACg ( -22) | 187: | GATAAAt (+170) |
| | 804: | caTATAgcAgg ( -23) | | |
| 30156 | 2721: | taTATAAAgCA (-128) | 1297: | AATAAAA (+502) |
| | 2723: | cCTATATAAAg (-130) | 1301: | GATAAAt (+498) |
| 17292[a] | 2721: | taTATAAAgCA | 1297: | AATAAAA (+155) |
| | 2723: | cCTATATAAAg | 1301: | GATAAAt (+151) |
| consensus[b] | | $T^{C}_{G}TATA^{T}_{A}A_{1-3}{}^{T}_{C}A$ | | $^{G}_{A}ATAA_{1-3}$ |

Table.6. **Most likely TATA boxes and polyadenylation signals for each ORF.**

For each TATA box and polyadenylation signal the position in the DNA sequence and homology to the consensus sequences (upper case) are indicated. The distance between each TATA box and the first ATG of each ORF and between each polyadenylation signal and the stop codon of each ORF or OR are indicated in brackets.
aOpen reading region which may be spliced onto ORF 30156 (see section 3.22).
[b] Promoter and polyadenylation consensus sequences for plant genes (Messing *et al.*, 1983).

| DNA sense | | CAAT |
|---|---|---|
| virion | 293: | taTCAATac |
| | 671: | GaCCAATgT |
| | 1579: | aGCCAATCg |
| | 1990: | ttTCAATgg |
| | 2164: | atTCAATga |
| | 2705: | GtCCAATgc |
| complementary | 473: | aaCCAATtT |
| | 1684: | GGTCAATCc |
| | 2374: | tcCCAATgc |
| | 2380: | caCCAATCc |
| | Consensus[b]: | GG$^{C}_{T}$CAATCT |

Table.7. **Position of T/CCAAT sequences.**

The positions of T/CCAAT sequences in the virion and complementary DNA sense, and homology of surrounding nucleotides to the animal consensus sequence (upper case) are indicated.
[b]CAAT box consensus sequence for animal genes (Benoist *et al.*, 1980).

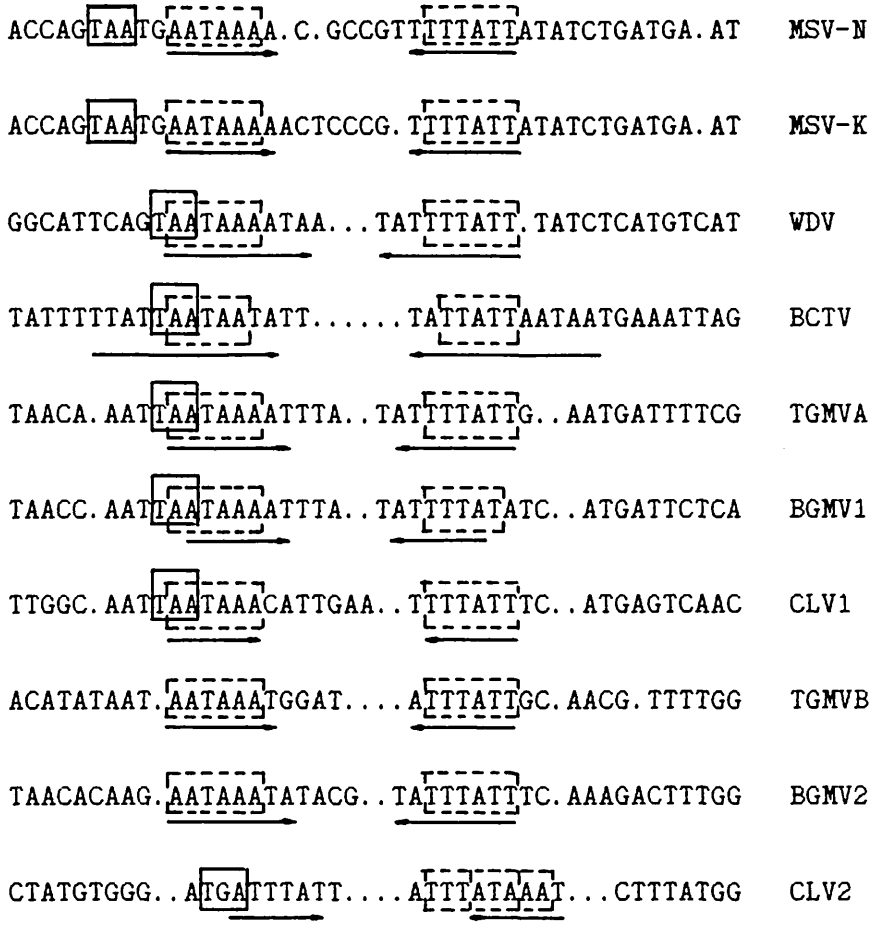



Fig.60. **Alignment of the AT-rich inverted repeat sequences containing putative polyadenylation signals in MSV-N DNA, MSV-K DNA, WDV DNA, BCTV DNA, TGMV DNAs A and B, BGMV DNAs 1 and 2 and CLV-K DNAs 1 and 2.**

Sequences are aligned to show the relative positions of the two polyadenylation signals (broken boxes). Termination codons are boxed and the arrows show the extent of the inverted repeats.

referred to as the starting intergenic region (SIR) and terminal intergenic region (TIR) with regard to the proposed direction of transcription (Howell, 1984). Nine inverted repeat sequences potentially capable of forming stem-loop structures with a $\Delta G \geq -14$Kcal/mol (the free energy of the primosome assembly site of ØX174; Arai and Kornberg, 1981) have been identified on the MSV-N DNA sequence. Three of the five most stable ($\Delta G \geq -20$Kcal/mol) potential hairpin structures are located in the large intergenic region, suggesting that they may be involved in regulatory functions (Mullineaux *et al.*, 1985). Two potential hairpin structures are also located in the large intergenic region of MSV-K (Howell, 1984). A single intergenic region is present in the DNA of the leafhopper-transmitted geminivirus BCTV and it also contains a stable hairpin structure (Stanley *et al.*, 1986)

Intergenic regions are also present in the genomes of the whitefly transmitted geminiviruses TGMV, BGMV and CLV (Hamilton *et al.*, 1984; Howarth *et al.*, 1985; Stanley and Gay, 1983). In each of these viruses, the nucleotide sequences of the two DNA components have little sequence homology, except for a 200 base 'common region' which is situated almost exclusively within an intergenic region region on both DNAs, analogous in position to the large intergenic region of the MSV isolates (see Fig.2). With the exception of TGMV DNA 2, there are no significant intergenic regions in the DNAs of the whitefly-transmitted geminiviruses other the one comprising the 'common region'. The 'common region' of each virus contains a highly conserved inverted repeat sequence capable of forming a stable stem-loop structure (Fig.1). This structure has been implicated in functions such as replication, regulation of transcription or capsid assembly.

A search for hairpin structures using the STEMLOOP program (Deveraux *et al.*, 1984) revealed four hairpin structures with $\Delta G \geq -14$Kcal/mol, as calculated by the rules of Tinoco *et al.* (1973) (Fig.61; H.Macdonald, personal communication). The three most stable of these potential hairpin structures are located within the first intergenic region (Figs.59 and 61; hairpin numbers 1, 2 and 3) and the fourth within ORF 30156. These hairpins may be involved in functions such as replication, regulation of transcription and sites for capsid assembly. It is

noteworthy that the sequence TAATATTAC present in the loop of one hairpin (Fig.61, hairpin number 2) is conserved in similar putative structures in all the other geminivirus sequences examined (Fig.62). This sequence might function as a recognition sequence for a protein involved in the synthesis of ssDNA from a dsDNA template (see section 3.22).

Analysis of encapsidated MSV ssDNA revealed a population of small DNA molecules (*ca.*80nt) with ribonucleotides covalently linked to their 5' ends (Donson *et al.*, 1984; Howell, 1984,1985). These molecules, which are thought to act as primers for second strand synthesis upon infection of host cells, are complementary to a region immediately downstream of stable hairpin structures, almost exclusively within the small intergenic region. Similar molecules have been detected in WDV and are complementary to a region in the second intergenic region, approximately between nucleotides 1370 and 1450 (H.Macdonald, personal communication). An inverted repeat sequence capable of forming a hairpin structure with $\Delta G = -10.0$ is located immediately upstream of the priming site, between nucleotides 1451 and 1469.

## 3.22 Comparisons with other geminiviruses

The genome organisation of WDV is remarkably similar to that of MSV (Fig.64B). To find whether the two genomes are related at the nucleotide sequence level, comparisons were made using the ZIP and COMPARE/DOT PLOT programs (Deveraux *et al.*, 1984).

A comparison of the MSV-N (Mullineaux *et al.*, 1985) and WDV DNA sequences using the ZIP program (gap weight 2.5; gap length weight 0.25) is shown in Fig.63. The two sequences were aligned starting from the first base of the conserved nanonucleotide sequence TAATATTAC in the loop of the most stable hairpin structure in the large intergenic region (or first intergenic region) of these viruses (see section 3.21). A total of 1538 matches were obtained between the two sequences, corresponding to an overall sequence homology of 57.2% with respect to MSV-N or 48.4% if the percentage is calculated with respect to the length of the gapped sequences (3177 nucleotides). In either case the overall sequence

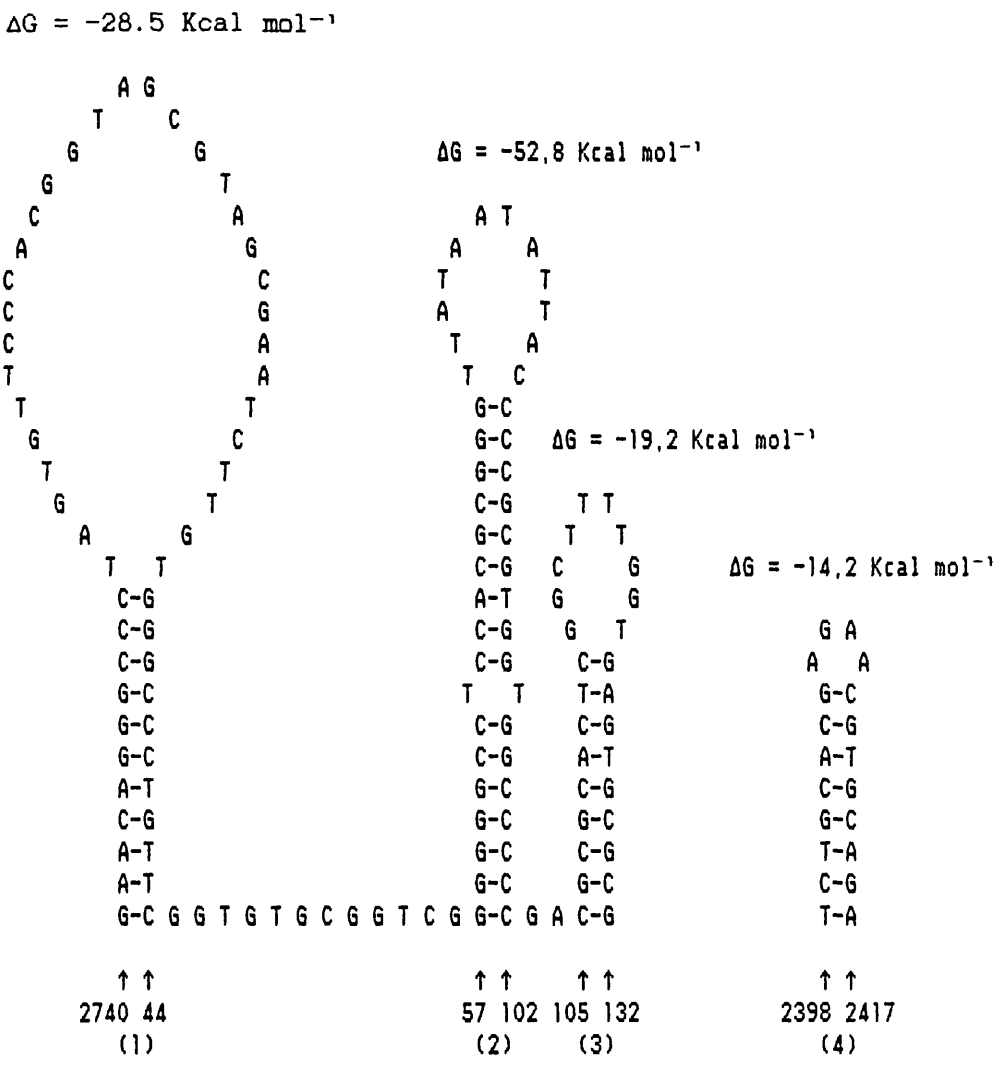



Fig.61. **Potential hairpin structures in WDV DNA.**

The putative hairpin structures shown have a free energy (ΔG) of ≥-14 kcal/mol (H.Macdonald, personal communication), as calculated by the rules of Tinoco *et al.* (1973). The structures are numbered according to the position of the centre of the loop (see Fig.59). The coordinates of the first and last bases of each stem are shown. The bases separating stems 1, 2 and 3 are shown. Potential hairpin structures in the WDV DNA sequence were found with the STEMLOOP program (Deveraux *et al.*, 1984).

GTCGGGGGGCCTCCACGCGGGTTA**TAATATTAC**CCCGCGTGGTGGCCCCCGAC WDV

GGCCCAGCAGGAAAAGAAGGCGCGCAC**TAATATTAC**CGCGCCTTCTTTTCCTGCGAGGGCC MSV-N

GGGCCGGCTAGCAAGAAAAGAAGGCGCGCAC**TAATATTAC**CGCGCCTTCTTTTCCTGCGGTAGGGCCC MSV-K

GGGCCATCCGTTA**TAATATTAC**CGGATGGCCC BCTV

GCGGCCATCCGTT.**TAATATTAC**CGGATGGCCGC TGMVA

GCGGCCATCCGTTT**TAATATTAC**CGGATGGCCGC TGMVB

GCGGCCATCCGATA**TAATATTAC**CGGATGGCCGC BGMV1

GCGGCCATCCGATA**TAATATTAC**CGGATGGCCGC BGMV2

GGGGCCAACCGTA.**TAATATTAC**CGGTTGGCCCC CLV1

GGGGCCAACCGTA.**TAATATTAC**CGGTTGGCCCC CLV2

Fig.62. **Comparisons of the major potential hairpin structures in WDV, MSV, BCTV, CLV, TGMV and BGMV DNAs.**

The sequences have been aligned to show the conserved nanonucleotide TAATATTAC (bold) in the loop of the potential hairpin structures. The stem of the potential hairpin structres are underlined.

homology is greater than the random expectation of 25%, indicating that the two sequences are significantly homologous. An overall sequence homology of 46% with respect to MSV-N was obtained with the same program under more stringent conditions (gap weight 4; gap length weight 0.1).

A plot of regions of >60% homology between the sequences of WDV and MSV DNA obtained with the COMPARE/DOT PLOT programs (window size 21; stringency 14) is shown in Fig.64A. A diagonal line can be clearly discerned among the background dots confirming the relationship between the two sequences. In contrast, there was insufficient nucleotide sequence homology between the DNAs of WDV and those of TGMV and CLV-K to align the sequences.

All the ORFs in WDV have a counterpart in MSV-N with respect to both size and relative position in the genome (Fig.64A). To find whether the predicted amino acid sequences encoded by the ORFs in the two viruses are related, comparisons were made using the GAP program (gap weight 1; gap length weight 0.1). Only the amino acid sequences of ORFs occupying similar relative positions in the two viral genomes were compared. The results are given in Table 8. Direct amino acid sequence similarities between predicted proteins ranged from 25.0 to 44.4%.

It is generally believed that there may be functionally similar amino acids that can be substituted without causing radical changes in the function of the protein. It is therefore useful to treat some amino acids as equivalent in protein sequence comparisons. Using the SIMPLIFY programme (Deveraux *et al.*, 1984) which simplifies a peptide sequence into broad amino acid groups (P, A, G, S, T = neutral, weakly hydrophobic; Q, N, E, D = hydrophilic, acid amine; H, K, R = hydrophilic, basic; L, I, V, M = hydrophobic; F, Y, M = hydrophobic aromatic; C = cross-link forming) and the GAP program (gap weight 1; gap length weight 0.1) to align the simplified sequences, conserved similarities between the WDV and MSV predicted proteins ranging from 54.6 to 67.7% were obtained (Table 8).

Comparisons were also made between the predicted amino acid sequences encoded by WDV ORFs and those encoded by TGMV and CLV

Legend to Fig. 63. **Comparison of the WDV and MSV-N DNA sequences.**

The sequences were aligned with the ZIP program (gap weight 2.5; gap length weight 0.25; Deveraux *et al.*, 1964). The WDV and the MSV-N (Mullineaux *et al.*, 1964) DNA sequences were aligned starting from the first base of the conserved nanonucleotide TAATATTAC in the loop of the major potential hairpin structure in the starting intergenic region of the two viruses.

COMPARISON OF THE MSV AND WDV DNA SEQUENCES

```
     1                                                            50
MSV  TAATATTACC gCGCcTt... .......... .........C TTTTCcTgcG
WDV  TAATATTACC cCGCgTggtg gcccccgacg cgcactcggC TTTTCgT..G

     51                                                          100
MSV  AGgGCcCGG. .....TaGGg CCcgAgCgaT TTgaTGtaaA gTTTgGtcct
WDV  AGtGCgCGGa ggcttTtGGa CCacAtC..T TTtcTGatcA cTTTcG....


     101                                                         150
MSV  gcttTGtA.. ....TGATTT ATCtaaagca gcCcaTTcTa AaG....AAT
WDV  ....TGgAag atgtTGATTT ATC....... .aCacTTtTg AcGgggaAAT

     151                                                         200
MSV  CcGgtCCcgG tCactaTAaa TTgccTAA.c AAGTGCG... .AttCatTC.
WDV  CtGtgCCatG cCt...TAgc TT..aTAAgg AAGTGCGtgg tAgcCcaTCt

     201                                                         250
MSV  ..ATGGAtC. ..CAcAgAaC gCCCtgtaTT atcAgcCGcG ggTAcCcAcA
WDV  cgATGGAgCa ggCA.AtAgC cCCCccgcTT cctAtaCGgG acTAtCaAtA

     251                                                         300
MSV  gCAG....CT cCgAcatCCG GaggagTgCC gtgGAgT.CG CgTaGGCGAg
WDV  cCAGacccCT tCcAttcCCG G.....TtCC tccGAcTaCG CcT.GGCGAa

     301                                                         350
MSV  gtagctatTT TGaGcTTtGT TgCaT.TGaT TTGCTttTAc C.........
WDV  ......caTT TGtG.TTcGT TaCcTtTGgT TTGCTaaTAg Ccgtaggcgt

     351                                                         400
MSV  TGCT...... ..TTACctTt gGgTgCTGAg AGAccTtATc TTAGTtctgA
WDV  TGCTtggctt gcTTACacTc tGtTtCTGAa AGAttTaATt TTAGTgtgtA

     401                                                         450
MSV  AGGctCGAca AgGCAgatcc AcGGAGGAgC tgatatttgG tGGAcAagct
WDV  AGG..CGAag AaGCA..... AaGGAGGAcC .........G aGGAaA....

     451                                                         500
MSV  gTGGaTAgGa Gcaaccctat ccctAATAtA CCaGC..... .........A
WDV  tTGGtTAcGg G......... ....AATAcA CCgGCcagat taaatggtgA

     501                                                         550
MSV  CCAcCAAgtc aggGcAaTCC CGggccattt gttccAGGcA cGGGataagc
WDV  CCAaCAA... ...GgAcTCC CG........ .....AGGtA aGGG......

     551                                                         600
MSV  attcagccat gtccacgtcc AAGaGGAAGc gGGgAGA... TGAtTC....
WDV  .......... .........g AAGcGGAAGa tGGaAGAagg TGAaTCcagc

     601                                                         650
MSV  .GAAttGGaG tAAGcGGGtg aCTaaGaAgA AGCcttcttc agctGGgcTg
WDV  gGAA..GGtG gAAGgGGG.. .CTgtGtAtA AGC....... ....GGcgTa

     651                                                         700
MSV  AAgAGGgcT. ...GGcAG.. .....cAAGg CCgataggCC A..TCcCTGC
WDV  AAcAGGcgTa taaGGtAGta cctgtgAAGc CC......CC AgcTCtCTGC
```

```
     701                                                        750
MSV  aaATcCagaC aCTcCAgCac GcTG..... ....GGACCA ccAT......
WDV  gtATtC...C gCTaCAaCtg GtTGaatagc gacaGGACCA atATtgttgt

     751                                                        800
MSV  .GaTAAcggT cCcCtCCGGa ggagtatGT. GAcCTcATcA aCaccTaTGC
WDV  gGgTAA...T aCaCcCCGG. .......GTc GAtCTgATtA cCtgtTtTGC

     801                                                        850
MSV  cCgaGGatct GaCGAgggcA AcCGcCAcAC cAgcgAGACt cTgacgTACA
WDV  tCagGGtaag GcCGAtaatA AtCGgCAtAC aAatcAGACc gTcctaTACA

     851                                                        900
MSV  AgaTcgcCAT cgAC....TA CC.aCTtcgT tgCCGAcGCg gCAGCctgTC
WDV  AatTtaaCAT ..ACagggTA CCtgCTataT gtCCGAtGCa tCAGC...TC

     901                                                        950
MSV  gcTaC..... TCCA..aCaC CggTACCggt gtaatgTGGc TgGTgTATGA
WDV  cgTtCatcgg TCCAgtcCgC CtcTACC... ....acTGGt TaGTcTATGA

     951                                                       1000
MSV  caccacTcCc G.GCgG..AC AaGCT....C cGACcCCgCa aACtATaTTT
WDV  ......TgCa GaGCcGaaAC AgGCTatgcC aGACgCCaCt gAC.ATcTTT

     1001                                                      1050
MSV  gCctaccctG AcaCgcTGaA AgC.GtgGCC GgccACaTGG AaaGTGagcC
WDV  aC.......G AtgCctTGgA AtCtGctGCC GagtACgTGG ActGTGcaaC

     1051                                                      1100
MSV  GgGagctGTg tCATCGcTTC GTGGTGAAAc GGcgaTGGtt gtTcAACaTg
WDV  GtGcttgGTc gCATCGaTTC GTGGTGAAAa GGaagTGGac cgTgAACcTt

     1101                                                      1150
MSV  GagACcGAcG GgCGcAttgG tTCGGaTaTt cctCCcTcGA ......aTgC
WDV  GttACtGAtG GaCGgA..aG gTCGGgTcTa agaCCgTtGA ccagcgcTaC

     1151                                                      1200
MSV  AAgTtGGaAG cCttGCAAGc gcAAcATC.. .tACttccAc AAGTTCacgA
WDV  AAcTgGGtAG tC.gGCAAG. ..AAtATCgt tgACgcaaAt AAGTTCttcA

     1201                                                      1250
MSV  gtGGgTTGgG aGTgAgaACG cAGTGGAaGA AtgtaACG.. .GACGG.AgG
WDV  aaGGtTTGcG tGTcAcgACG gAGTGGAtGA A...cACGgg tGACGGcAaG

     1251                                                      1300
MSV  AgTtGGtGcC ATccAGAgaG GAGCgCTGTA ...cATggtC AtTgCcCccG
WDV  A.TaGGcGaC ATtaAGAagG GAGCaCTGTA tcttATtagC AgTaCtCgtG

     1301                                                      1350
MSV  GcaaTGGccT TACT...... ......TtTA C......... .........T
WDV  G...TGGtgT TACTggtgat agtgccTcTA Cggcgtttga tgttgtatgT

     1351                                                      1400
MSV  GCCcATggGC AgaccCGtcT GTacTTTaAg AG.tgTtGGC AacCAGTAAT
WDV  GCCtATacGC A....CGcgT GTtaTTTcAa AGccaTcGGC AttCAGTAAT
```

```
     1401                                                        1450
MSV  .gAATAAaAc gccgtTTTTA TTaTATCTgA TG.aATgCtg aaaGcTTACA
WDV  aaAATAAtA. .....TTTTA TT.TATCTcA TGtcATtC.. ...GaTTACA

     1451                                                        1500
MSV  .........t TAatatgtCG tGCgAtGgC. .ACgAAA... AACAcAC.AC
WDV  gaggctcggc TA......CG aGCaAaGaCa aACcAAAtat AACAaACaAC

     1501                                                        1550
MSV  AAtCaaTACA ggggggtagt cggcgggcgg ctAAgGgtgg tgctcggcgg
WDV  AAcCctTACA .......... .......... .cAAtG.... ..........

     1551                                                        1600
MSV  gcagaACATC GaAAAAtcA. .......... ...AGATcT. ..ATcT...G
WDV  .....ACATC GgAAAAcgAa atacaacacc ctgAGATaTt acATtTataG

     1601                                                        1650
MSV  AA..TGTACt gCcTCCG... TAGGAggCAG cTCAggGgGA gaatAcCAtT
WDV  AAacTGTACg cCgTCCGcgc TAGGA..CAG .TCActGcGA ....AgCAgT

     1651                                                        1700
MSV  ....TCTcCc CC...GGCGA catAatGTAa aTGATGcA.G T.TtGCCTcg
WDV  gacgTCTtCg CCggaGGCGA acgA..GTAg tTGATGaAcG TcTcGCCTtc

     1701                                                        1750
MSV  AaA..... .......... .......... ....TAcTCC agCTGccCtG
WDV  AtAcatgtag tgaacaacag tgttagagta catgTAaTCC gaCTGttCgG

     1751                                                        1800
MSV  GAGTCATtTC CTTcAtCCAA TCTTCaTCcG agTTggCgAg gATtATtgtA
WDV  GAGTCATaTC CTTgAgCCAA TCTTCgTCtG gaTTaaCtAa aATgATgcaA

     1801                                                        1850
MSV  GG.cTTagAC ttCtTcTGcA CCTTTttCtt tTTACCATAc TTgGGgTTtA
WDV  GGtaTTccAC ccCgTaTG.A CCTTTcgC.. .TTACCATAt TTtGGaTTgA

     1851                                                        1900
MSV  CaaTGAAaTC cCtCTGAcag CCaACtAAct gtTTCCAaca aGGacaGAAt
WDV  CcgTGAAgTC aCgCTGAgcc CCgACgAAgc acTTCCAgtt gGGtgtGAAc

     1901                                                        1950
MSV  TTaAAcGGAA TaTCaTCtAc GATgTTATAg aTtGCGTctt CGTtgTATGa
WDV  TTgAAtGGAA TgTCgTCgAt GATaTTATAc tTgGCGTtga CGTcaTATGt

     1951                                                        2000
MSV  aGaccAATCA ACattAtTtT gccAgTAATT aTGaacCCCT AGgcttCTgG
WDV  tGtgaAATCA ACtagAcTgT tatAaTAATT gTGtgtCCCT AGagacCTtG

     2001                                                        2050
MSV  CCCAaGtAGa tTTTCCgGTT CTtGTTGGgC CGacGATGTA GAgG...cTc
WDV  CCCAgGaAGt cTTTCCtGTT CTgGTTGGcC CGcaGATGTA GAtGgactTa

     2051                                                        2100
MSV  TGCtTtCttg atctttcatC tGaTGA...C TGGAtacAga aTCCATCCAt
WDV  TGCcTcC... .........C cGgTGActcC TGGAataAtc gTCCATCCAc
```

```
     2101                                                2150
MSV  TggAgGTCAG aaATTGC... ..ATCCtCga GgGTataaca gGtAgGTtgA
WDV  TctAaGTCAG ..ATTGCgct tgATCCgCag GaGT...... .GgAaGTacA

     2151                                                2200
MSV  AggAGcATgT AaGcTTCGgG aCTaACctGG A....AGATG TT........
WDV  A..AGgATaT AgGaTTCGaG gCTtAC..GG AgtagAGATG TTcatttttc

     2201                                                2250
MSV  .AGgctggag cCaaTCAttg ATtGaCTCAT taCAAAgtAa atcagGTGAT
WDV  cAG....... .CttTCA... ATgGtCTCAT ggCAAAtgA. .....GTGAT

     2251                                                2300
MSV  gaGGgTGGA. ..TgAGGatT G...GTG..A ACTcttcctg aatcTCAGGA
WDV  tcGGtTGGAa acTcAGGtgT GtaaGTGgcA ACT....... .gggTCAGGA

     2301                                                2350
MSV  AAaAGcTtat tTGCaGaGTA tTCaAAaTac tgcaatTTTG .......TgG
WDV  AAtAGaTggc gTGCcGtGTA cTCgAAgT.. .....cTTTG agacggaTaG

     2351                                                2400
MSV  ACCAaTCAAA gGGAAgctct TTctggAtCA TGgaGAGGtA cTCtTCtttg
WDV  ACCAtTCAAA cGGAAaacga TTgcaaAcCA TGctGAGGaA tTCcTC....

     2401                                                2450
MSV  gaagtaGCGt GtGaAAtaat gtCTcGcATT attTcATCTt Tagaaggctt
WDV  ......GCGa GaGgAA.... ..CTaG.ATT caaTgATCTg T.........

     2451                                                2500
MSV  tttttccTTt AccTCtGaAT CA..GatTTT cCtA...GGA aggGggACtt
WDV  .......TTc AtaTCcGcAT CAcgGtcTTT aCgAcctGGA gttGaaAC..

     2501                                                2550
MSV  cctAGgaAtG AAaGTaCCtC tCTCA..... ...aACA... ....CAgCCa
WDV  ...AGccAcG AAtGTtCCcC aCTCAgctgt gtttACAtcg gagtCAaCC.

     2551                                                2600
MSV  gaggtTCCTT gagaATGTAA TCcCtcAcCc TGtTTaCtGa C.TTGGCAct
WDV  .....TCCTT cgtgATGTAA TCaC.gAaCt TGgTTgCaGt CtTTGGCAgc

     2601                                                2650
MSV  cTGaATATTT GGgTGaAacc catttATATc aaAGAAccTt GAGtcAgaTA
WDV  tTGtATATTT GGaTGgA... .....ATAT. ggAGAA..Tg GAG..AtgTA

     2651                                                2700
MSV  TCCtTAC... .......... ..cGGctTct cTGtctGAAG CA..AtGCaT
WDV  TCCaTACgga ggtttaaggc attGGgaTtg gTGatgGAAG CAcgAaGCtT

     2701                                                2750
MSV  GT........ ...AaaTGCA aA......ct TCCATCTTta TGtgcCTCTC
WDV  GTtctgcacg agaAcgTGCA gAtgtggtga TCCATCTTcg TGgagCTCTC

     2751                                                2800
MSV  gggCA.CA.. ....TAGAat G.....TATT T......... ........Gg
WDV  taaCAgCAgc gatgTAGAgg GgctcaTATT Tgttcaagag agtgcgaaGt
```

```
     2801                                                      2850
MSV  GAATCCAA.. CGaACaacGa GCTCccaGaT catctgACAg gcgatTTCAG
WDV  GAATCCAAgg CGtACtgtG. GCTCaagGgT ......ACA. .....TTCAG

     2851                                                      2900
MSV  GATtTtcTgG A......cAC TTtGgATAGg ttaGGAACgT Gttagcgttc
WDV  GATaTgtTaG AaagaggtAC TTgGaATAGa cacGGAACcT G.........

     2901                                                      2950
MSV  cggtgtgaga actgacGGTt gGatGAgGAg GAGGCCATtG ccGacgACgG
WDV  .......... ......GGT. .GcaGAtGAa GAGGCCATgG taGtgaACaG

     2951                                                      3000
MSV  A......GCt tgAGG..CTg AGgGAtggcA gACtGGGaGc tCC...AAAC
WDV  AagtccgGC. ..AGGtcCTt AGcGA..aaA aACgGGGtGt gCCagaAAAC

     3001                                                      3050
MSV  TCTATagTaT ACCC...... ...GTGcGcc TTC....... ..gAAATcCg
WDV  TCTATccTcT ACCCtgcgtg gagGTGtGaa TTCtgcacac tgcAAATgCa

     3051                                                      3100
MSV  ccG.cTCCct TGtcTTATAg tGGttGCA.. aaTgGGCcGG ...ACcGGGC
WDV  atGtgTCCaa TGctTTATAt aGG..GCAgg ttTtGGCgGG agaACaGGGC

     3101                                                      3150
MSV  C......ggC CCA..GcAG. GaAaaGAA.. .....GGC.. ..........
WDV  CctagtgttC CCAcgGtAGc GtAgcGAAtc ttgtgGGCcc tgttcggtgt

     3151                      3177
MSV  GCG....... ......CAC. .......
WDV  GCGgtcgggg ggcctcCACg cgggtta
```

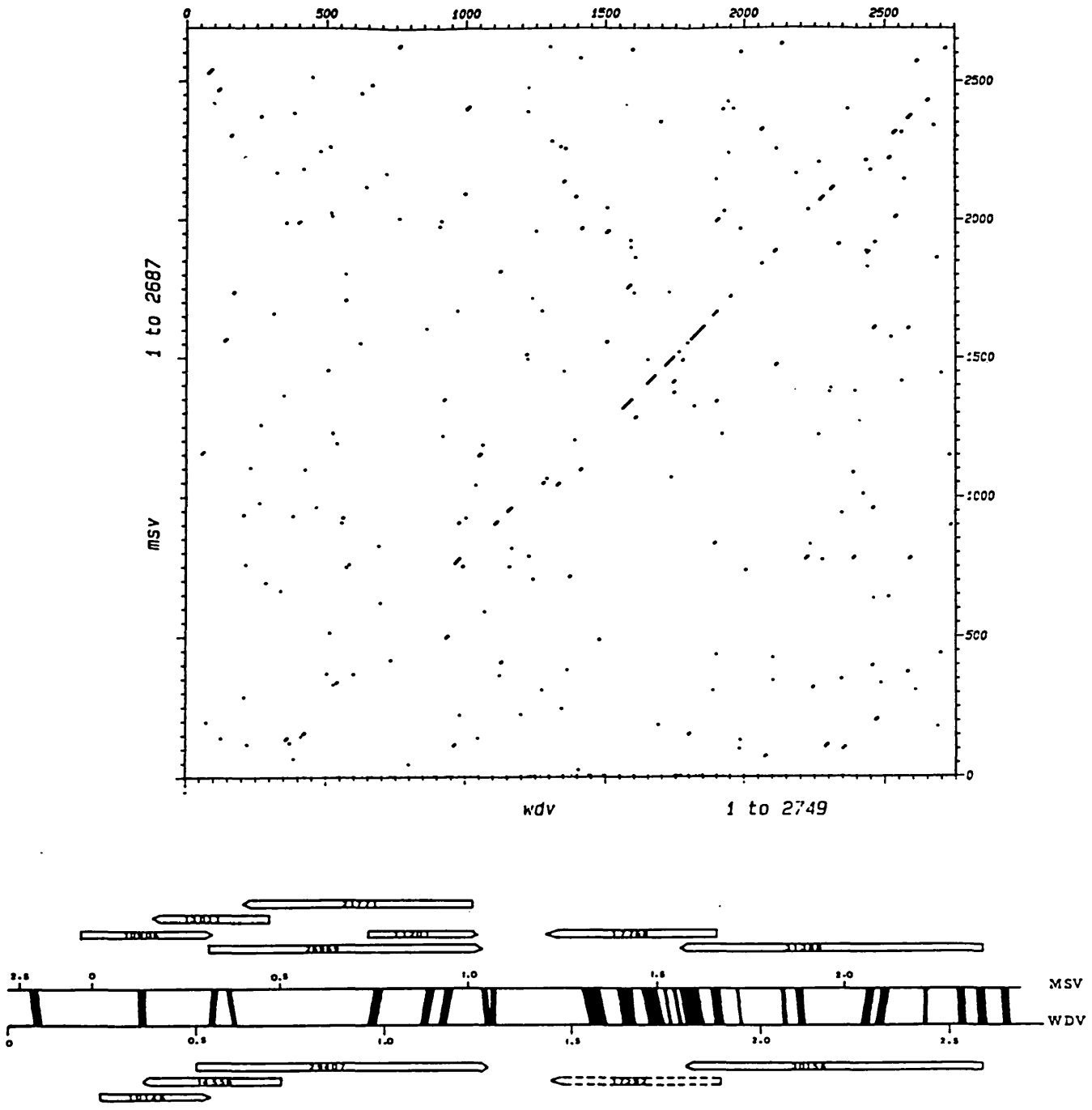



Fig. 64. **Sequence homology of WDV and MSV-N.**

A) Plot of regions of >60% homology between the sequences of WDV and MSV-N (Mullineaux *et al.*, 1984) DNAs obtained using the COMPARE/DOT PLOT programs (Deveraux *et al.*, 1984) with a window size = 21 and stringency = 14.

B) Linear representation of the WDV and MSV DNA sequences aligned to give a maximum sequence homology. The MSV sequence is numbered as in Mullineaux *et al.* (1984). The shaded areas indicate regions of at least 21 bases with >60% homology. The potential coding regions on both strands of each virus are indicated by open arrows above (MSV) or below (WDV) the sequence alignment. The box drawn with stippled lines below the WDV sequence represents the open reading region which may be spliced onto ORF 30156 (see section 3.22).

ORFs occupying similar relative positions in the respective DNAs (see Fig.65). The results are given in Table 8. Amino acid sequence similarities between TGMV and WDV predicted proteins ranged from 24.7 to 35.6% (direct) and 51.7 to 63.2% (conserved). Those between CLV-K and WDV predicted proteins ranged from 22.6 to 37.2% (direct) and 50.5 to 64.3% (conserved).

One of the problems in interpreting the results obtained from amino acid sequence comparisons is that of determining the level of significance of the values obtained. It is frequently assumed that the average percentage of similarity between two random amino acid sequences is equal to 5%. However, this value does not take into account the usually widely different relative frequencies with which each amino acid occurs in a specific sequence and between different sequences. It also does not take into account the introduction of gaps by the program to increase the similarity between the sequences. To test the significance of the values obtained, comparisons were made between a number of random sequences. The sequences were randomized with the SHUFFLE programme (Deveraux *et al.*, 1984) and compared with the GAP programme using the same gap weight and gap length weight as above. The sequences which were compared included identical, related and unrelated sequences. The percentages of similarity obtained were found to be relatively independent of the degree of relatedness between the original (i.e. prior to randomization) sequences. The mean percentages of similarity obtained for >70 random comparisons were 25.77 ± 3.52% (direct) and 55.49 ± 4.92% (conserved).

Assuming that the threshold of homology is equal to the mean percentage of similarity plus the standard deviation for random comparisons (29.29 and 60.41% for direct and conserved similarities, respectively), then all WDV ORFs, with the exception of ORF 14556, have homologous counterparts in MSV-N. Thus WDV ORF 10145 is related to MSV-N ORF 10906 , WDV ORF 29407 is related to MSV-N ORF 26969 and WDV ORF 30156 is related to MSV-N ORF 31388. Alignment of the WDV ORFs to the MSV ORFs are shown in Figs.66 to 68.

In contrast, only two CLV-K ORFs and one TGMV ORF have percentages of similarity to WDV ORFs above the threshold of

Table 8. **Comparison of the amino acid sequences encoded by WDV ORFs with those of MSV, CLV and TGMV occupying similar relative positions in the respective DNAs.**

Comparisons were made with the GAP programme(gap weight 1; gap length weight 0.1; Deveraux *et al.*, 1984). To obtain the percentage of similarity for conserved amino acids the sequences were simplified with the SIMPLIFY programme(Deveraux *et al.*, 1984) prior to aligning with the GAP programme.The groups used by the SIMPLIFY program consist of: A = P, A, G, S, T (neutral, weakly hydrophilic); D = Q, N, E, D (hydrophilic, acid amine); H = H, K, R (hydrophilic, basic); I = L, I, V, M (hydrophobic); F = F, Y, W (hydrophobic aromatic; C = C (cross-link forming). Percentages of similarity were calculated with respect to the shortest sequence.
See Fig.65. for the position of the ORFs.

| ORF | | | | | Percentage similarity | |
|---|---|---|---|---|---|---|
| WDV | MSV | CLV | TGMV | Amino acid difference | Direct | Conserved |
| 10146 | 10906 | | | -11 | 34.8 | 60.7 |
| " | | 13100 | | -23 | 22.6 | 50.5 |
| 14556 | 13011 | | | +29 | 25.0 | 54.6 |
| " | | 13500 | | +10 | 24.3 | 57.8 |
| " | | 21000 | | -110 | 27.2 | 59.0 |
| 29407 | 11201 | | | +162 | 26.2 | 63.6 |
| " | 26969 | | | +16 | 37.6 | 62.5 |
| " | | 29300 | | +4 | 23.8 | 57.4 |
| " | | 30200 | | +2 | 31.5 | 56.7 |
| " | | | 28651 | +13 | 28.4 | 55.7 |
| " | | | 29341 | +4 | 24.7 | 51.7 |
| 30156 | 31388 | | | -8 | 44.4 | 67.7 |
| " | | 15600 | | +124 | 26.1 | 58.3 |
| " | | 33700 | | -34 | 26.0 | 54.0 |
| " | | 40300 | | -94 | 37.2 | 64.3 |
| " | | | 26466 | +30 | 25.9 | 57.7 |
| " | | | 40285 | -88 | 35.6 | 63.2 |
| 17292 | 17768 | | | -3 | 53.0 | 76.3 |
| " | | 15100 | | +15 | 20.9 | 54.9 |
| " | | 15800 | | +16 | 25.6 | 48.0 |
| " | | 33700 | | -148 | 27.3 | 59.6 |
| " | | 40300 | | -208 | 36.4 | 64.2 |
| " | | | 14871 | +21 | 21.8 | 53.7 |
| " | | | 15677 | +18 | 21.2 | 53.6 |
| " | | | 26466 | -84 | 31.5 | 47.9 |
| " | | | 40285 | -202 | 49.2 | 65.4 |
| OR2 | 21771 | | | -47 | 30.2 | 60.7 |

Table 8

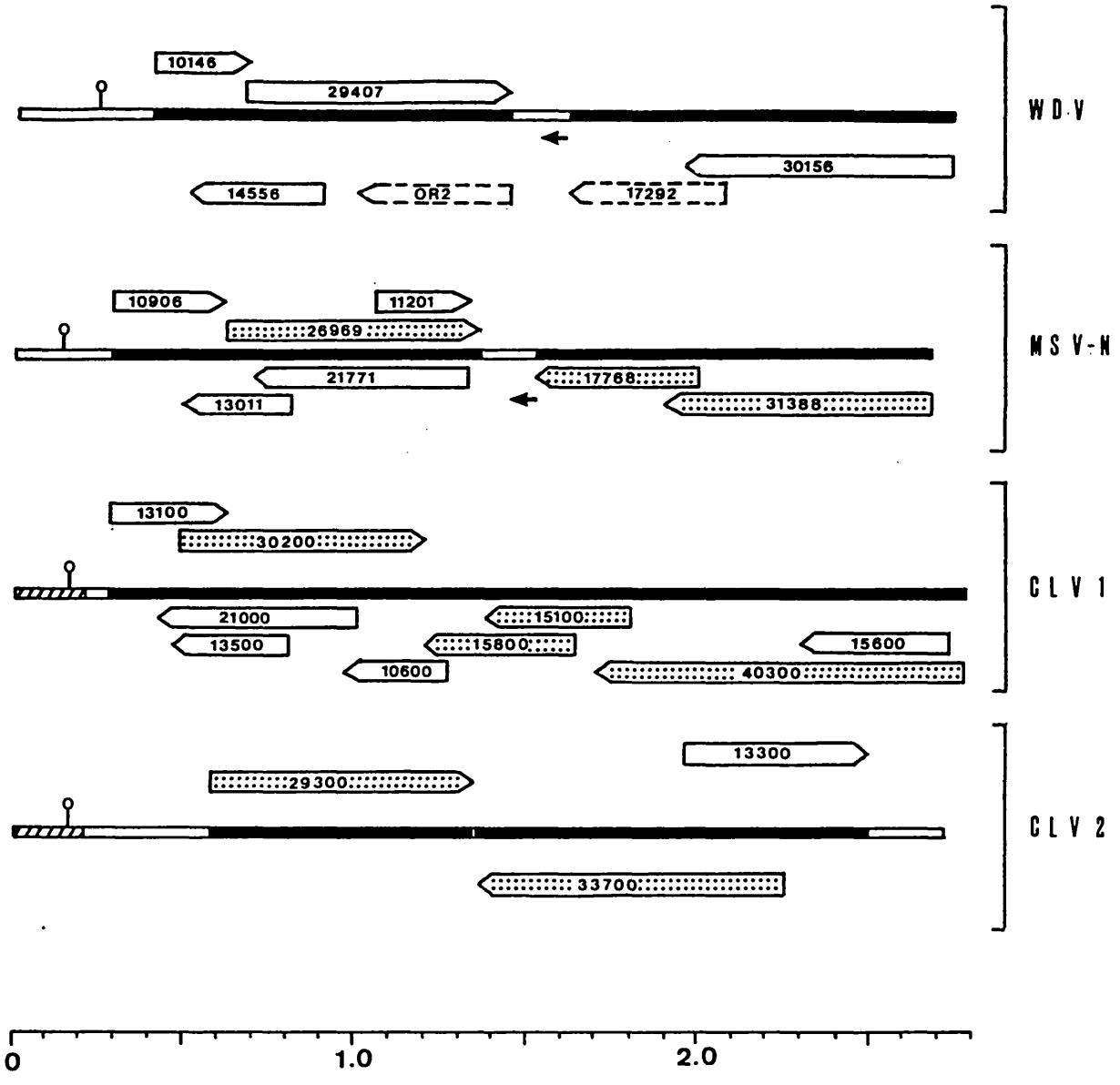



Fig. 65. **Comparison of the genome organization of WDV, MSV and CLV.**

The circular genomes are represented in a linear form. The genomes of WDV and MSV-N start at the begining of the starting intergenic region. The DNA components of CLV-K start at the beginning of the 'common region'. The thick black lines show the coding regions, white boxes show the intergenic regions and hatched areas show the 'common region' of CLV DNAs 1 and 2. The stem-loop structures containing the conserved nanonucleotide TAATATTAC are indicated as (♀). ORFs are indicated by large arrows. The shaded (stippled) arrows in the MSV-N genome are conserved in MSV-K and those in the CLV-K genome are conserved in all the geminiviruses which infect dicotyledonous plants. The predicted mol.wt. of the encoded proteins are given in daltons. The small closed arrows show the location of the putative primer in WDV and MSV-N DNA (H.Macdonald, personal communication; Donson *et al.*, 1984). The scale is in Kb.

homology. Thus, both CLV ORF 40300 and TGMV ORF 40285 are related to WDV ORF 30156 and the CLV coat protein (ORF 30200) is related to WDV ORF 29407. The percentage of similarity of the TGMV coat protein gene (ORF 28651) to WDV ORF 29407, on the other hand, is just under the threshold of homology (Table 8).

The size and amino acid content of the MSV-N ORF 26969 suggested that it might code for the capsid protein (Mullineaux *et al.*, 1984). This was confirmed by *in vitro* translation of MSV-specific polyadenylated mRNA (Morris-Krsinich *et al.*, 1985). The homology of this ORF to the WDV ORF 29407 suggests that the latter codes for the WDV coat protein. Indeed, purified WDV preparations contain a protein of this mol.wt. (Dr.R.H.A.Coutts, personal communication).

The relationships between the putative coat protein of WDV with those of MSV-N, CLV-K and TGMV-K (Table.8) are distant when compared with the homologies between the coat proteins of CLV-N and TGMV (71.3% direct and 84.6% conserved homologies; Hamilton *et al.*; 1984). Weak homologies between the MSV-N coat protein and those of CLV-K and TGMV have also been reported by other authors (Mullineaux *et al.*, 1985). Recently, the sequence of the leafhopper transmitted geminivirus BCTV has been determined (Stanley *et al.*, 1986). The sequence of the putative coat protein of BCTV was found to be more closely (although weakly) related to those of the leafhopper-transmitted MSV and WDV than to those of the whitefly-transmitted geminiviruses BGMV, TGMV and CLV. These observed homologies are in good agreement with the serological relationships among these viruses. Indeed, strong serological relationships have been detected between CLV and TGMV but not between these viruses and MSV or WDV (Stein *et al.*, 1983; Roberts *et al.*, 1984). Although weak serological relationships were detected between BCTV and both CLV and TGMV (Sequeira and Harrison, 1982; Stein *et al.*, 1983) these were not substantiated in a later investigation (Roberts *et al.*, 1984). No serological relationships have been detected between BCTV, MSV and WDV (Roberts *et al.*, 1984). The correlation between serological relationship, now extended to amino acid sequence homology between the coat proteins, and transmission by insect vectors has led to the suggestion that the coat protein plays a key role in

**A**

```
            1                                                  50
                    •   o   ••   •  o ••••  o •    ••      •    ••
MSV 10146   MEQAIAPPL. PIRDYQYQTP .SIPGSSDYA WRTFVFVTFG LLIAVGVAWL
WDV 10906   .....MDPQN ALY.YQPRVP TAAPTSGGVP WSRVGEVAIL SFVALICFYL

            51                                                 100
             • •• ••••  ••• •• ••  • ••• •  • ••  •••  o o     ••
MSV 10146   AYTLF.LKDL ILVCKAKKQ. RRTEE.IGYG NTPARLN..G DQ.....QGL
WDV 10906   LY.LWVLRDL ILVLKA.RQG RSTEELIFGG QAVDRSNPIP NIPAPPSQGN

            101   109
            •
MSV 10146   PR*.......
WDV 10906   PGPFVPGTG*
```

**B**

```
            1                                                  50
                          o •••   o     •   •• • o    ••  •   ••
MSV 13011   .......... ....MVVPAC WSVMICRDGL SALLPALFSP AE.EG...FF
WDV 14556   MPIIIGLTLS KTGNQIDPGC ..ITHNNIGP VAIQPVVAEY AESWGLHRYY

            51                                                 100
            •• ••   •    •• •        o   ••  o    •  •   •  • ••
MSV 13011   LVTRL...LQ FESS.PRFLL DVDMAECLSR AWNKWPGIAL TWWCWYIRDR
WDV 14556   LI.RLFTPLI H.SPLPPSAG FTFFHLPLPL TSGVLVGHHL .IWPVYSRNQ

            101                                  136
             o     •   • • •  •           •   • •
MSV 13011   VAPIHSLSTK YQL.LRGS.. .......ALS SLQN*..
WDV 14556   FPRSSFASSP YTLKLNLSET ECKQAKQRLR LLANQR*
```

Fig.66. **A) Alignment of the predicted amino acid sequences encoded by MSV ORF 10146 and WDV ORF 10906. B) Alignment of the predicted amino acid sequences encoded by MSV ORF 13011 and WDV ORF 14556.**

The alignments were done with the GAP program (gap weight 1; gap length weight 0.1). The gapped sequences were then simplified with the SIMPLIFY (Devereaux *et al.*, 1984) programmeand further gaps introduced to increase the number of matches between conserved amino acids (i.e. those in the same group). The groups used by the program SIMPLIFY consist of: A = P, A, G, S, T (neutral, weakly hydrophobic); D = Q, N, E, D (hydrophilic, acid amine); H = H, K, R (hydrophilic,basic); F = F, Y, W (hydrophobic, aromatic) and C = C (cross-link forming). An open circle indicates conserved amino acids. A closed circle indicates identity of amino acids. Gaps were inserted to increase the similarity. Asterisks denote stop codons.

Legend to Fig.67. **Alignment of the predicted amino acid sequences of the coat proteins of WDV (29.4Kd), MSV-N (27.0Kd), TGMV (28.7Kd) and CLV-K (30.2kd).**

The alignment was done for each pair of sequences individually using the GAP program (gap weight 1; gap length weight 0.1) and the four sequences were then aligned by eye. The gapped sequences were then simplified with the SIMPLIFY program (Deveraux *et al.*, 1984) and further gaps introduced to increase the number of matches between conserved amino acids (i.e. those in the same group). The groups used by the program SIMPLIFY consist of: A = P, A, G, S, T (neutral, weakly hydrophobic); D = Q, N, E, D (hydrophilic, acid amine); H = H, K, R (hydrophilic, basic); F = F, Y, W (hydrophobic, aromatic) and C = C (cross-link forming). Conserved amino acids common to WDV and any of the remaining three viruses are shown in upper case. An open circle indicates identity of amino acids between WDV and MSV-N. A closed circle indicates identity of amino acids for all four viruses. An open circle with a dot indicates identity of amino acids in the coat proteins of all four viruses and also in the predicted amino acid sequences of TGMV ORF BR1 (29 341) and CLV-K ORF 2R1 (29 200). Gaps were inserted to increase the similarity. Asterisks denote stop codons.

```
          1                                                  50

                                                o o  o  oo      ••      o
WDV  29.4  .......... .......... ......MvTn KdSRG....k GKRkmeegeS
MSV  27.0  .......... .......... ......MsTs KrkRGddsnw SKRvtkkkpS
TGMV 28.7  mpkrdapwrl magtskvsrs an....ysPr gslp..... .KRdawvnrP
CLV  30.2  mskrpgdiii stpgskvrrr lnfdspyrnr atAptvhvtn rKR.awvnrP

51                                                           110
                                                        ⊙     ⊙
o   • o   o    o         • o             o  o o       ooo   o •
SgRwKGAVYK rrKqAyKvvP vkPPALcVfR yNwlnSdrTn IvVGnTPRV DLITcFAqGkA
SaglKrAgsK ....AdR... ...PSLqIqt lQhagT..Tm ItVPsGGvc DLIntYArGsd
myR.KPrIYR slRgPdvpkG ceGPc....K vQ...Syeqr hdISlvGKV c.ISdvTrGnG
myR.KPTMYR myRsPdiprG ceGPc....K vQ...Sfeqr ddVkhlGic kVISdvTrGpG

111                                                          170
                                                  ⊙  ⊙          ⊙
  o•oo  o   oo  •       oo o           oo ooo  •   o o o    oo
DnNRHTnQTV LYKfnIQGtc YMSDASA.pF igPvlryhWL VYDAePk..Q AMPdAtdIFT
EgNRHTsETL tYKiaID.yh FVADAAAcrY snTgtgvmWL VYDTtPgg.Q A.PtPqtIFA
ithRvGkrfc VksvyIlGki WMdEnik.lk nhTnsvmfWL VrDrrPygt. PMdfGq.VFn
lthRvGkrfc IksiyIlGki WLdETik.kq nhTnnvifYL LrDrrPygna PqdfGq.IFn

171                                                          230
                      ⊙
 o  o   •  •• o o   o •••••o o   o  oooo  o oo       o    o
MPwnL..lPS TWTVqRawsH RFVVKRKWtv NLVTDGRkVG SkTVdqRYnW VvgKniVDaN
yPdtLkawPA TWkVsRelcH RFVVKRRWlf NMeTDGR.IG SdippsnasW kpcKrnIyfh
Mf...dnePS TaTVkndlrd RFqViHRFha kVtGgqyasn eqALvrRFwk VnnnvvyNhQ
Mf...dnePS TaTIkndlrd RFqVlRKFha tVVGgpygMk eqALvkRFyr LnhHvtyNhQ

231                                                          290

oo  oo o    o o o  oo   o o  ooo o       o  o     oo        o   •
KFfKGLrVT. .TEWMNtGDG KIGDIkKGAL YLIsSTrgGV TgdsasTAfd vvcAyThacY
KFtsGLgVr. .TQWkNvTDG gVGaIqRGAL YMViAPgnGL Tf....TAh. ...GqTrl.Y
eagKyenhTe nAllLymAcT HaSNpvyATL kIri..... ..........  .........Y
eagKyenhTe nAllLymAcT HaSNpvyATL kIri..... ..........  .........Y

291  298

• o  o o
F.KAIGiQ*
F.KSVGnQ*
FydSITn*
FydSIGn*
```

Fig. 67

Legend to Fig.68. **Alignment of the predicted amino acid sequences encoded by WDV ORF 30156 (30.2Kd) and OR 17292 (17.3Kd), MSV-N ORFs 31388 (31.3Kd) and 17768 (17.8Kd), TGMV ORF 40285 (40.3Kd) and CLV ORF 40300 (40.3Kd).**

The alignment was done for each pair of sequences individually using the GAP program (gap weight 1; gap length weight 0.1) and the six sequences were then aligned by eye. The gapped sequences were then simplified with the SIMPLIFY program[mv2] (Deveraux *et al.*, 1984) and further gaps introduced to increase the number of matches between conserved amino acids (i.e. those in the same group). The groups used by the program[mv] SIMPLIFY consist of: A = P, A, G, S, T (neutral, weakly hydrophobic); D = Q, N, E, D (hydrophilic, acid amine); H = H, K, R (hydrophilic, basic); F = F, Y, W (hydrophobic, aromatic) and C = C (cross-link forming). Conserved amino acids common to WDV and any of the remaining three viruses are shown in upper case. An open circle indicates identity of amino acids between WDV and MSV-N. A closed circle indicates identity of amino acids for all four viruses. Gaps were inserted to increase the similarity. Asterisks denote stop codons.

```
                  1                                                  50
                  •••••  • •        ••• •• •   o   o      •      o ••
WDV  30.2, 17.3   MASSSPRR.F RVySKyLFLT YPECTLEPQy ALDsLrTLLn KYEPlYIaav
MSV  31.3, 17.8   MASSSSnRqF shrnantFLT YPkCPeNPEi AcQmIweLVv RWiPkYIlca
TGMV 40.3         M.PShPkR.F qInAKnyFLT YPQCSLskEe SLsqLqALnt piNkkFIkic
CLV  40.3         M...rTPR.F RIgAKnVFLT YPkCSIpkEh lLsfIqTLsl qsNPkFIkic

     51                                                            110
     •• • ••• • •• • •       o           o   o o   •••••• • •  o •••••
     RELHEDGSPH LHVLVQnklR aSITNpNaLn lrMDTSpFSI ..FHPNIQAA KdcNQVRDYI
     REaHkDGSlH LHaLLQtekp vrIS..... ...DSrfFdI ngFHPNIQSA KsvNrVRDYI
     RELHEDGqPH LHVLIQfegK yccqN.Qrff dlVsPTrsAh ..FHPNIQrA KsssDVktYI
     RELHQNGePH LHaLIQfegK iTItN.NrLf dcVhPScsTs ..FHPNIQGA KsssDVksYL

     111                                                           170
      ••    •    o •••       o  •  o           •  o  ••    • •• •••
     tKEvDSdVnt aEWGTF.V.A vSTP..GRkD .........R DAdMKQIIes SSSREEFLSM
     lKEplA.Vf. .ErGTF.I.P rkSPflGKsD sevkekkpsK DeiMRDIIsh ATSKEEYLSM
     dKDgDTlV.. ..WGeFqVdG rSAr..GgcQ t......s.n DAaaeaLN.. ASSKEEaLqI
     dKDgDT.V.. .EWGqFqIdG rSAr..GgqQ .......san DAyaKaLN.. SGSKsEaLnV

     171                                                           230
          •• ••   o  ••• o   •••       o  o  •   o o • o   •       ••
     VcNRfPFEWS IRLKDFEYTA rHLFPDPvAT YTPeFPTeS. .LIChETIEs WkNehLYsVS
     IqkelPFDWS tKLqyFEYSA nKLFPEiqee FTnphPPsSp dLLCnESINd WlQpnIFqVS
     IrEKiPekYl fqfHNlNsnl dRIFdkTpeP WlPpFhvsSf tnVpdE.Mrq W.........
     IrElvPkDFv LqfHNlNsnl dRIFqEPpAP YvSpFPcsSf dqVpvE.IEe Wv........

     231                                                           290
      o  o  o   o       o  oo  o ooo  o  o  o          ooo  o  o  o  o  o
     lESYIL..cT stPaDQAqSD LEWMDdySrS H..R.ggisP STSAGQpEqE rLpGqgl* *
     pEAYML1qpT cyTlEDAiSD LQWMDsvS.S HqmKdqesrA STSSAQqEpE nLlGpea* *
     .......... .......... .......... .......... .......... ........ .
     .......... .......... .......... .......... .......... ........ .

     291                                                           350
        oo   o  •     •  •  oo •••••• •  •••••• ooo    oo    o    o
     lrVDGrlfQE S...PGRHkS IYIcGPTRTG KTSWARSLGT HNYYNSlVDF TTYDvN..Ak
     ppMDGfciQs Sde.rSRKqS LYIvGPTRTG KSTWARSLGv HNYWQnnVDW SSYNeD..Ai
     ..aEnyfgks SaarPeRpiS IiIeGdSRTG KTmWARSLGP HNYlsGhLDl nSrvysnkve
     ..aDn..vrD SaarPwRpnS IvIeGdSRTG KTiWARSLGP gNYlcGhLDl SPkvfnnaAw

     351                                                           410
     ••• •••••• o  o  oo  o    • •••• o •••••• oo       • oo • • oooo
     YNIIDDIPFK FtPnWKcFV. ..GAQRDFtV NPKYGK.RKV .IRGGIPcII LvNPDEDWL.
     YNIVDDIPFK FcPcWKqlV. ..GcQRDFiV NPKYGKkKKV q.KkSkPtII LaNSDEDWM.
     YNVIDDVTpq Y.lklKhWke liGAQRDWqt NcKYGK..pV qIKGGIPsIV LcNPgEgasy
     YNVIDDVdph Y...lKhFke fmGSQRDWqs NTKYGK..pV qIKGGIPtIf LcNPgptssy

     411                                                      463
     •  ••• o   o   o      o  oo    • o              o
     KD.MTPEQsD YMySNTVVhy MYeGETfinY SfASgedvtA .....SQ*
     KE.MTPgQlE YfeANcIIyi MspGE..kwY SpPelppteP vhsdrS*
     KvfLdkEEnt pLknwTfhna kFvflnsplY qsSTqss*
     KEfLAeEkqE aLkAwALkna IFiTlTeplY SgSnqs.... .hsqtSQeas hpa*
```

Fig. 68

```
            1                                                   50
                                          ∘∘  • ∘ ∘∘∘ ∘  ∘∘ • •∘ ∘
MSV 21771   ..... ..... .....MGS KSKAIAGGND HVQRSSLDGT
WDV OR2     NIILLLNADG FEITRVRIGT YNIKRRRGTI TSNTTTSTAN KIQCSLLNVA

            51                                                  100
              ∘•  ∘ •• •  ∘•∘  •• ∘ • •∘ ••   •  ∘   •  ∘∘∘∘ • ∘∘
MSV 21771   NSSVRYILPL RSHSQPTREL VEV.DVALAR LPTCIRGRN. IRTNAP.VGL
WDV OR2     YLAVTRVHPL RRDTQTFEEL ICVNDI.LAD YP.VVALVNG LRPDLPSISN

            101                                                 150
             •  • •••• •∘•   ∘ •  •  •∘• •   •  •  ∘        •∘• •
MSV 21771   HVEQPSPFHH EAMTQLPAHF PCGRPRFQRV RVGKYSLRGR SLSAGSGVIH
WDV OR2     KVHGPLPFHH ESMRPSTLHS P..RTRQQIP RHRK.DV... .....SGVWH

            151                                                 200
                     ∘ ∘   • ∘
MSV 21771   QPHYTGTGVG VATGCRVGNE VVVDGDLVRQ SLAGVAVALV RSSGIGVDEV
WDV OR2     .........S LFRLCIID*. ..........  ..........  ..........

            201                                 236

MSV 21771   TYSSGGDRYH GGPSVLECLD LQGWPIGLAA SPLQPS*
WDV OR2     .......... ..........  ..........  .......
```

Fig.69. **Alignment of the predicted amino acid sequences encoded by MSV ORF 21771 ORF and WDV OR2.**

The two sequences were first aligned with the GAP program (gap weight 1; gap length weight 0.1). They were then simplified with the SIMPLIFY programme (Deveraux *et al.*, 1984) and further gaps introduced to increase the number of matches between conserved amino acids (i.e. those in the same group). The groups used by the SIMPLIFY program consist of: A = P, A, G, S, T (neutral, weakly hydrophobic); D = Q, N, E, D (hydrophilic, acid amine); H = H, K, R (hydrophilic, basic); F = F, Y, W (hydrophobic, aromatic) and C = C (cross-link forming). The closed circles indicate identical amino acids and the open circles indicate conserved amino acids. Asterix denote stop codons. Gaps were inserted to increase the similarity.

insect transmission ( see section 1.5; Roberts *et al.*, 1984; Stanley *et al.*, 1986).

Alignment of the predicted amino acid sequences of the coat proteins of WDV, MSV-N, CLV-K and TGMV is shown in Fig.67. There are 55 conserved amino acids (22.1% conserved homology), 20 of which are directly homologous in all four coat proteins (7.9% direct homology) as compared to a random expectation of 22.0% (conserved) and 2.5% (direct). There is therefore a significant level of direct homology between the four sequences, suggesting a functional role in the secondary, tertiary or quaternary structure of the proteins. Comparison of the coat proteins of MSV and WDV to the coat protein of satellite tobacco necrosis virus, an RNA plant virus, revealed a significant degree of amino acid sequence homology between these proteins, which was attributed to the icosahedral structure of their particles (Mohana Rao, 1986). Similarities between the coat proteins of WDV, MSV, CLV and TGMV may also be imposed by the unusual geminate morphology of these viruses. Six of the identical amino acids in the same positions in WDV, MSV, CLV and TGMV are lysine and arginine residues. Such residues may interact with and neutralize phosphate groups in the DNA sequence as has been suggested in the case of tobacco mosaic virus (TMV) coat protein (Butler, 1984).

The most highly conserved region between the WDV and MSV DNA sequences lies between nucleotides 1500 and 2000 in WDV and 1300 and 1800 in MSV-N (Fig.64). In MSV-N, this region corresponds mainly to ORF 17768. In contrast, no significant ORF starting with an ATG triplet was found in the corresponding section of the WDV sequence (Fig.58 B). However, when read in the complementary sense, there is an OR starting at nucleotide 1904 and ending at nucleotide 1454 (Fig.58 A). This region has the potential to code for a protein of mol.wt. 17292 containing 150 amino acids. When aligned with the predicted amino acid sequence of MSV-N ORF 17768 using the GAP program (gap weight 1; gap length weight 0.1), the two sequences showed 76.3% conserved homology and 53.0% direct homology, suggesting that OR 17292 may be functional in WDV.

The first three amino acids of the gene product of MSV-N ORF 17768 (including the methionine start codon) are coded by the

sequence ATG GAT GGT. A similar sequence is present at the 5' end of OR 17292 in WDV DNA except that the ATG codon in MSV is replaced by a GTG codon in WDV. It was possible, therefore, that the clone from which the sequence was derived contained an A→G transition in the start codon of the ORF. However, the presence of a GTG codon was confirmed in two independently derived *Eco*RI subclones (see section 3.18; Fig.56). Prokaryotic ribosomes can occasionally initiate at GUG codons but eukaryotic ribosomes seem to initiate exclusively at AUG codons (Kozak, 1983). The latter has been confirmed by *in vitro* studies using mouse ascites and wheat germ ribosomes (Brown and Smith, 1970; Kozak, 1981). Thus it seems unlikely that GUG serves as an initiator codon for OR 17292 in WDV. It is more probable that WDV ORF 30156 and OR 17292 are transcribed on a single mRNA which encodes a composite protein (see below).

The ORFs 31388 and 17768 of MSV-N correspond to ORFs P1a and P1b of MSV-K, respectively, except that P1b is 43 amino acids shorter than ORF 768 at its amino terminus. In MSV-K, ORF P1a and P1b are tandemly arranged in the same reading frame. The two proteins encoded by P1a and P1b show nearly 40% amino acid identity with the single product of ORF 40300 in the CLV-K genome, suggesting that ORFs P1a and P1b may be translated into a single composite protein by reading through the amber codon terminating ORF P1a (Howell, 1984). When the predicted amino acid sequence of WDV OR 17292 was compared to that of CLV-K ORF 40300 using the GAP program (gap weight 1; gap length weight 0.1), 36.4% direct and 64.2% conserved homology were obtained compared to a random expectation of 29.29% (direct) and 60.41% (conserved). Comparison of WDV OR 17292 to TGMV ORF 40285 (the counterpart of CLV-K ORF 40300) showed 42.3% direct and 65.4% conserved homology (Table 8). As previously mentioned, comparison of WDV ORF 30156, the counterpart of MSV-N and MSV-K ORFs 31388 and P1a, respectively, to CLV-K ORF 40300 and TGMV ORF 40285 also showed significant levels of homology (Table 8). An alignment of ORF 30156 and OR 17292 of WDV, ORFs 31388 and 17768 of MSV-N, ORF 40300 of CLV-K and ORF 40285 of TGMV is shown in Fig.68. There are significant regions of homology between WDV ORF 30156 and MSV-N ORF 31388 and the N-terminal portions of TGMV ORF 40285 and

CLV-K ORF 40300, and between WDV OR 17292 and MSV-N ORF 17768 and the C-terminal portions of TGMV ORF 40285 and CLV-K ORF 40300. Several regions contain sequences of four to six amino acids which are identical in the four viruses. One region contains a sequence of 12 amino acids of which ten are invariant in all four viruses. Overall, between the four sets of proteins, i.e. WDV (ORF 30156 + OR 17292), MSV-N (ORFs 31388 + 17768), TGMV ORF 40285 and CLV-K ORF 40300, there are 129 conserved amino acids of which 82 are identical in in all four viruses. Based on the average number of amino acids (355) encoded by CLV ORF 40300 and TGMV ORF 40285 this corresponds to 23.0% direct and 36.3% conserved homology as compared to a random expectation of 2.5% (direct) and 22.0% (conserved). Thus it is likely that WDV ORF 30156 and OR 17292 are transcribed on a single mRNA to encode a composite protein corresponding to the single gene products of CLV-K ORF 40300 and TGMV ORF 40285. The observation that the nearest polyadenylation signal for ORF 30156 is downstream of the termination codon of OR 17292 is consistent with this hypothesis. MSV-N ORFs 31388 and 17768 might also encode a composite protein, even though ORF 17768 starts with an ATG, and therefore may be independently expressed.

Two possible mechanisms can be envisaged for the production of a composite protein from WDV ORF 30156 and OR 17292 (or their MSV-N counterparts). Firstly, a precursor transcript could be spliced to produce a mRNA encoding the composite protein. The consensus sequence for splice junctions in plant genes is T/AC/ANT/AG↓*GTAA/TGTA/TT/AT/AA/T* for the donor (left) boundary and *T/ATTT/AT/AT/AAA/TT/ATT/AG/AT/CAG*↓GT/A for the acceptor (right) boundary, where the arrows show the sites where splicing occurs and intron sequences are indicated in italics (Lycett *et al.*, 1984). Of these nucleotides, only the first and the last two bases of the introns, GT and AG respectively, are conserved in all plant genes. Potential splice sites in WDV ORF 30156 and OR 17292 are shown in Table 9. There are seven AG dinucleotides in OR 17292 which, if involved in splicing, would preserve the highly homologous region starting at position 318 in Fig.68. A comparison of the nucleotide sequences flanking these dinucleotides to the acceptor boundary consensus sequence is

shown in Table 9. Four of the (acceptor) sequences have >40% homology to the consensus sequence (Table 9, positions 1889, 1892, 1912 and 1914). Of these, the sequence starting at position 1892 has the higher (>58%) homology to the consensus sequence. A similar (acceptor) sequence, although with only *ca.* 41% homology to the consensus, is conserved in MSV-N (nucleotides 1655-1642). There are twenty four GT dinucleotides in WDV ORF 30156 downstream of the DNA sequence (in the complementary sense) coding for the highly homologous region ending at position 101 in Fig.68. A comparison of the nucleotide sequences flanking these dinucleotides to the donor boundary consensus sequence is shown in Table 9. Homologies to the consensus sequence ranged from *ca.* 33% to 73%. Two (donor) sequences with *ca.* 73% homology to the consensus (Table 9, positions 2136 and 2292) could, in combination with the (acceptor) sequence at position 1892, join the two reading frames to give a single (in phase) coding region. However, the composite proteins thus formed would be significantly shorter than the TGMV and CLV counterparts. Alternatively, if splicing occurred between the (donor) sequence starting at position 1968 (see Table 9) and the (acceptor) sequence at position 1892, an in phase composite protein of similar size as the TGMV and CLV counterparts could be produced. The (donor) sequence at position 1968 shows 60% homology to the consensus and, unlike the two (donor) sequences at positions 2136 and 2292, is conserved in MSV-N (nucleotides 1732-1723). The MSV-N (donor) sequence at position 1732 and the (acceptor) sequence at position 1655 (see above) could join the two reading frames to give a single (in phase) coding region.

Analysis of MSV-specific polyadenylated mRNA revealed a single minor 1.3Kb complementary sense transcript (Morris-Krsinich *et al.*, 1985). The 5' and 3' ends of this transcript were positioned near nucleotides 2360 and 1160, respectively, by S1-nuclease mapping. The 5' end of this transcript, therefore, appeared to map slightly short of the first ATG (nucleotides 2374-2372) for ORF 31388. However, if translation started from another in-frame ATG at a position ninety nine nucleotides downstream, a 27.7Kd protein could be produced (Morris-Krsinich *et al.*, 1985). This transcript could also encode the putative product of ORF 17768. It appears, therefore, that MSV-N ORFs 31388 and 17292 may be

Table 9. **Potential splice sites for WDV ORF 30156 and OR 17292.**

Only potential splice sites that would preserve the regions of significant homology between WDV (ORF 30156 + OR 17292), MSV (ORFs 31388 + 17768), CLV-K ORF 40300 and TGMV ORF 40285 and that conform to the GT/AG boundary rule of Breathnach *et al.* (1978) are shown. Homology to the plant consensus sequence (Lycett *et al.*, 1984) is shown in upper case. 'h' refers to the number of nucleotide matches to the consensus sequence. The position of each potential splice site in the WDV DNA sequence is indicated on the left. 'p' refers to the position in the amino acid sequences shown in Fig. 68 where splicing would occur. Any combination of donor and acceptor splice sites for which the sum of 'n' is equal to 3 would produce an in phase composite protein. Potential splice sites for MSV-N ORFs 31388 and 17768 are shown in italics (see section 3.22).

ACCEPTOR BOUNDARY

| | INTRON | EXON | n | h | p |
|---|---|---|---|---|---|
| | T/A T T T/A T/A T/A A A/T T/A T T/A G/A T/C A G ↓ | G T/A | | 17 | |
| 1845: | c T g c g g g c c a A c C A G | a A | 1 | 6 | 318 |
| 1871: | c c g g g g A g g c A t a A G | t c | 0 | 4 | 309/310 |
| 1878: | g g a g T c A c c g g G g A G | G c | 1 | 6 | 307 |
| 1889: | c g a T T A t T c c A G g A G | t c | 0 | 8 | 300/301 |
| 1892: | g g a c g A t T A T T c C A G | G A | 0 | 10 | 299/300 |
| *1655:* | *g g a T T c t g T a T c C A G* | *t c* | *0* | *7* | *299/300* |
| 1912: | c a a T c T g A c T T A g A G | t g | 2 | 8 | 293 |
| 1914: | c g c A A T c T g a c t T A G | a g | 1 | 7 | 292 |

DONOR BOUNDARY

| | EXON | INTRON | n | h | p |
|---|---|---|---|---|---|
| | T C<br>A A N T<br>A G ↓ G T A A<br>T G T A T T A<br>T A A T | | | 15 | |
| 1862: | c A T A a | G T c c a T c T A c | 1 | 8 | 269 |
| 1880: | c A G g a | G T c A c c g g g g | 1 | 5 | 260 |
| 1903: | T t A g a | G T g g a T g A c g | 2 | 6 | 252 |
| 1943: | c C T T c | G T A c t T c c A c | 1 | 8 | 239 |
| 1968: | A C T c c | G T A A G c c T c g | 3 | 9 | 228/229 |
| *1732:* | *T C C A G* | *G T t A G T c c c g* | *3* | *10* | *228/229* |
| 2035: | c C T g a | G T t T c c A A A c | 2 | 8 | 204 |
| 2055: | A C C c a | G T t g c c A c T T | 3 | 8 | 197/198 |
| 2086: | T t C g a | G T A c a c g g c A | 2 | 6 | 187 |
| 2102: | T A T c c | G T c T c a A A g A | 1 | 9 | 182 |
| 2110: | g A A T G | G T c T a T c c g T | 2 | 9 | 179 |
| 2119: | T t T c c | G T t T G a A T g g | 2 | 8 | 176 |
| 2126: | c A A T c | G T t T t c c g T T | 1 | 8 | 174 |
| 2136: | g C A T G | G T t T G c A A T c | 3 | 11 | 170/171 |
| 2165: | A t C T a | G T t c c T c T c g | 1 | 7 | 162 |
| 2198: | A g A c c | G T g A t g c g g A | 1 | 6 | 150 |
| 2207: | A g G T c | G T A A a g A c c g | 1 | 8 | 138 |
| 2210: | T C C A G | G T c g t a A A g A | 1 | 10 | 137 |
| 2223: | T g G c t | G T t T c a A c T c | 3 | 7 | 128/130 |
| 2229: | c A T T c | G T g g c T g T T T | 3 | 9 | 126/128 |
| 2241: | g C T g a | G T g g G g A A c A | 2 | 8 | 122 |
| 2256: | c C G A t | G T A A a c A c A g | 3 | 9 | 117/118 |
| 2268: | A g G A G | G T t g a c T c c g | 3 | 7 | 113/114 |
| 2288: | A g T T c | G T g A t T A c A T | 1 | 10 | 107 |
| 2292: | A C C A a | G T t c G T g A T T | 3 | 11 | 105/106 |

transcribed on a single mRNA. However, there is as yet no evidence for splicing of this or any other transcript in MSV-N.

If splicing does not occur, a composite protein between WDV ORF 30156 and OR 17292 (or their MSV-N counterparts) could be formed by translational frameshift prior to the termination codon of ORF 30156 (or MSV-N ORF 31388) as has been suggested for the expression of the gag-pol polyprotein of Rous sarcoma virus and the two major ORFs of yeast *Ty* elements (reviewed by Varmus, 1985). Very little is known, however, about the mechanisms by which translational frameshift occurs. Comparison of the DNA sequences of the two major ORFs of two Ty elements (Ty*912* and Ty*917*) showed no significant homology in the overlapping region, suggesting that the primary structure of this region is not important in promoting frameshifting (Clare and Farabaugh, 1950). It has been proposed that frameshifting could be caused by an offset anticodon-codon pairing involving the conserved U33 nucleotide immediately 5' to the normal anticodon in so-called 'shifty' tRNAs (Weiss, 1984).

The demonstration that TGMV DNA 1 and CLV DNA 1 are capable of independent replication indicated that these DNAs encode all the virus-coded proteins required for viral DNA replication (Rogers *et al.*, 1986; Townsend *et al.*, 1986). A significant degree of homology (61% direct and 75% conserved homology) has been detected between the 40300 ORF encoded by CLV-K DNA 1 and the 40.9 ORF encoded by the single DNA component of the geminivirus BCTV (Stanley *et al.*, 1986). Since both CLV-K DNA 1 and BCTV DNA are capable of independent replication, the high degree of conservation between CLV-K ORF 40300 and BCTV ORF 40.9 and, by extension, TGMV ORF 40285, WDV ORF 30156 + OR 17292, MSV-N ORFs 31388 + 17768 and MSV-K ORFs P1a + P1b (Hamilton *et al.*, 1984; Mullineaux *et al.*, 1985; Howell, 1984; Rogers *et al.*, 1986) suggests that the proteins encoded by these ORFs may be involved in DNA replication. The presence of extensive regions of amino acid sequence homology might reflect conserved functional domain (s) in these proteins. Some of these regions may be involved in the interaction between DNA and protein. The conserved nanonucleotide TAATATTAC in the loop of analogous hairpin structures in the DNAs of all geminiviruses which have been

sequenced is a potential candidate for interacting with conserved regions in the ORFs above. Interestingly, this nanonucleotide sequence is highly homologous to the recognition and cleavage site for the bacteriophage ØX174 gene A protein (Rogers *et al.*, 1986), which cleaves the nascent strand during rolling-circle synthesis of ssDNA from a dsDNA template (Kornberg, 1982). It is possible that the geminiviral ORFs above may have a similar function as the ØX174 gene A protein.

There is at least one more example in WDV where joining of ORs of the DNA could produce a composite protein. The OR between nucleotides 1292 and 852 in reading frame -3 (WDV OR2) could code for a polypeptide of 146 amino acids. Comparison of the predicted amino acid sequences encoded by WDV OR2 and MSV-N ORF 21771 using the GAP program (gap weight 1; gap length weight 0.1) showed 30.2% direct and 60.7% conserved homology. The alignment in Fig.69 shows significant regions of homology between the two sequences (e.g. that between positions 105-113). There is a potential TATA box (TCTATAAAT) upstream of WDV OR2, at nucleotides 1424-1416, conforming to the consensus sequence for plant promoters TCTATAT/AA$_{1-3}$T/CA (Messing *et al.*, 1983). Three ATG codons are located between this TATA box and the beginning of OR2 (Fig.58B). The nearest polyadenylation site conforming to the consensus sequence G/AATAA (Messing *et al.*, 1983), however, is located 665nt downstream of the termination codon of OR2. A counterpart of MSV-N ORF 21771 is not present in MSV-K, suggesting that this ORF may not be functional. Indeed, no complementary sense transcripts mapping to that region were detected in MSV-N specific polyadenylated mRNA (Morris-Krsinich *et al.*, 1985). It is possible, therefore, that the conservation of amino acid sequences in those open reading regions may be due to the fact that they overlap the more conserved coat protein genes of these viruses in the opposite sense.

### 3.23 Relationships among the geminiviruses

On the basis of amino acid sequence conservation in putative proteins encoded by the leafhopper-transmitted WDV, MSV-N and MSV-K (section 3.22), it is suggested that, in addition

to the coat protein genes, at least two more ORFs (ORF 30156 and OR 17292 in WDV, ORFs 31388 and 17768 in MSV-N and ORFs P1a and P1b in MSV-K) are involved in protein coding in these viruses. Three viral-specific transcripts isolated from MSV-N infected maize leaves have been mapped, two to the viral and one to the complementary sense of MSV DNA (Morris-Krsinich *et al.*, 1985). Both viral sense transcripts could encode the coat protein and the complementary sense transcript could encode the product of ORF 17768 and possibly that of ORF 31388 (Morris-Krsinich *et al.*, 1985). These results are therefore consistent with expression of the three ORFs above. Although there is significant homology between the amino acid sequences encoded by WDV ORF 10146 and MSV-N ORF 10906, the absence of a corresponding protein in MSV-K suggests that they may be fortuitous. Indeed, although the larger (1.05Kb) of the two viral sense MSV-N specific transcripts could encode the product of ORF 10906, the location of the initiation codon for this ORF is just four nucleotides downstream of the 5' terminus of the transcript which argues against this possibility. However, a number of minor MSV-N transcripts have not yet been mapped (Morris-Krsinich *et al.*, 1985) and these may account for expression of this and/or other ORFs. Further work, including mapping of WDV-specific transcripts and identification of virus-encoded proteins, will be required to define the number of real protein coding regions of WDV and MSV.

Amino acid sequence comparisons presented in this thesis (section 3.22) and in other studies (Mullineaux *et al.*, 1985; Rogers *et al.*, 1986; Stanley *et al.*, 1986) also suggest conserved functional domains in proteins encoded by the leafhopper-transmitted geminiviruses with monopartite genomes and those encoded by the whitefly-transmitted geminiviruses with bipartite genomes. Further similarities between the viruses in the two subgroups above include their overal genome organisation, including a bidirectional transcription strategy, and similarities in non-coding regions of the DNA such as the presence of an AT-rich region containing inverted repeats composed solely of A and T residues and the conserved nanonucleotide TAATATTAC present in the loop of similar putative hairpin structures. These results provide strong evidence for an

evolutionary relationship between these two types of geminiviruses. On the basis of limited sequence homology between the coat protein gene of CLV-K DNA 1 (ORF 30200) and a gene of similar size in a corresponding position in CLV-K DNA 2 (ORF 29200), Kikuno *et al.* (1984) suggested that the bipartite genome evolved from a single common ancestral DNA. Comparison of the coat proteins genes for WDV, MSV-N, TGMV and CLV-K to CLV-K ORF 29200 and a corresponding protein in TGMV DNA 2 (ORF 29341) showed that six of the 20 amino acids (i.e. 30%) which are identical in the coat protein of the four viruses are also identical at the same positions in CLV ORF 29200 and TGMV ORF 29341 (Fig.67), compared with a random expectation of 5 X 0.34 = 1.7% (CLV ORF 29200 and TGMV ORF 29341 are 34% directly homologous) (Dr.K.Buck, personal communication). These results therefore extend the suggestions of Kikuno *et al.* (1984) to provide evidence for a common ancestral origin of the monopartite and bipartite geminivirus genomes (Dr.K.Buck, personal communication). On the basis of similar comparisons between the coat proteins of MSV-N, CLV-K and TGMV and the products of CLV-K ORF 30200 and TGMV ORF 29341, Mohana Rao (1985) has further suggested that the DNA of MSV is ancestral and the bipartite genomes of CLV and TGMV evolved from it.

Although direct proof for the monopartite nature of WDV and MSV is still lacking, the demonstration that cloned BCTV DNA is infectious indicates that a single DNA component has the potential to accommodate all the viral functions required for infectivity (Stanley *et al.*, 1986). This raises questions as to why a second component has evolved in the whitefly-transmitted geminiviruses. DNA 1 of the whitefly-transmitted geminiviruses CLV and TGMV encodes all the virus-specific replicative functions (Townsend *et al.*, 1986; Rogers *et al.*, 1986) as well as the coat protein. Since both DNA components are required to initiate an infection in plants, it has been suggested that DNA 2 may be responsible for systemic spread of the virus in the host (Townsend *et al.*, 1986). On the basis that CLV is readily mechanically transmissible to *N. benthamiana* while BCTV has so far only been introduced into this host using its leafhopper vector, Stanley *et al.* (1986) have suggested that

the putative spread factor encoded by DNA 2 may only be required for spread of the virus from peripheral leaf cells to phloem cells in which it proliferates. This is consistent with the fact that BCTV is apparently phloem-limited while CLV has been observed in the cortex, mesophyll and epidermis (Stanley *et al.*, 1986). Therefore, if this suggestion is correct, introduction of CLV DNA 1 directly into the phloem using methods similar to those for mechanical transmission of BCTV to its natural host *B.vulgaris* may result in a systemic-infection (Stanley *et al.*, 1986). It is not clear, however, why such a function would have evolved in nature, where the virus appears to be transmitted exclusively by insects. One possibility is that the feeding mechanism of whiteflies may be less efficient than that of leafhoppers in delivering the virus to phloem cells. It may be, therefore, that the leafhopper transmitted geminiviruses, which may rely entirely on their vectors to reach the phloem cells, do not encode a spread factor. This could explain why mechanical transmission of these viruses to their natural hosts is difficult and in some cases (e.g. MSV) has not been achieved. However, the relative difficulties in mechanically transmitting some of these viruses may also reflect the different responses of different host plants to the wounding effects of mechanical transmission. In addition to the putative function involved in systemic spread of the virus, DNA 2 of the whitefly-transmitted geminiviruses may encode an insect transmission factor. Indeed, there is a correlation between the number of DNA components and the insect vector. All the whitefly-transmitted geminiviruses that have been examined have two components while the leafhopper-transmitted geminiviruses appear to have a single component. On the basis of serological relationships among geminiviruses and amino acid similarities between the coat proteins, it has been suggested that the coat protein may play a key role in insect transmission (Roberts *et al.*, 1984; Stanley *et al.*, 1986). A DNA 2 product could interact with the coat protein to specify transmission by whiteflies. In the absence of a DNA 2 component, the coat protein may be the sole virus-coded protein required for the interaction of virus with leafhopper vector (Stanley *et al.*,1986).

**Acknowledgements**

I am very grateful to my supervisor Dr. R.H.A. Coutts for his help and advice throughout the course and to Dr. K.W. Buck for valuable discussion during the preparation of this manuscript.

I would like to thank Nino Villaroel for his help with DNA spreading and electron microscopy, and all the people in the lab with whom I have worked.

I would also like to thanks CAPES, Brazil, for having given me a studenship and Dr. Klein for his immense support.

## References

Adam, G., Sander, E., Shepherd, R. J. (1979) Structural differences between pea enation mosaic virus strains affecting transmissibility by *Acrythosiphon pisum* (Harris). *Virology*, **92**: 1-14.

Adejare, G. O., Coutts, R. H. A. (1982a) The isolation and characterization of a virus from Nigerian cassava plants affected by the cassava mosaic disease, and attempted transmission of the disease. *Phytopathologische Zeitschrift*, **103**: 193-210.

Adhya, S., Gottesman, M. (1978) Control of transcription termination. *Annual Review of Biochemistry*, **47**: 967-996.

Albertini, A. M., Hofer, M., Calos, M. P., Miller, J. H. (1982) On the formation of spontaneous deletions: The Importance of Short Sequence Homologies in the Generation of Large Deletions. *Cell*, **29**: 319-328.

Al-Musa, A. (1982). Incidence, economic importance, and control of tomato yellow leaf curl in Jordan. *Plant Disease*, **66**: 561-563.

Arai, K-I., Kornberg, A. (1981) Unique primed start of phage ØX174 DNA replication and mobility of the primosome in a direction opposite chain synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **78**: 69-73.

Bailey, J. M., Davidson, N. (1976) Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. *Analytical Biochemistry*, **70**: 75-85.

Been, M. D., Burgess, R. R., Champoux, J. J. (1984) Nucleotide sequence preference at rat liver and wheat germ type 1 DNA topoisomerase breakage sites in duplex SV40 DNA. *Nucleic Acids Research* **12**: 3097-3113.

Bennett, C. W. (1971) The curly top disease of sugar beet and other plants. *American Phytopathological Society*. Monograph, Nº 7, St. Paul, Minnesota.

Bennett, C. W., Wallace, G. (1938). Relation of the curly top virus to the vector, *Eutettix tenellus. Journal of Agricultural Research*, **56**: 31-51.

Benoist, C., O'Hare, K., Breathnach, R., Chambon, P. (1980) The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Research*, **8**: 127-142.

Bevan, M. W., Flavel, R. B., Chilton, M. (1983) A chimaeric antibiotic resistance gene as a selectable marker for plant cell transformation. *Nature*, **304**: 184-187.

Bird, J. and Maramorosch, K. (1978) Virus and virus diseases associated with whiteflies. *Advances in Virus Research*, **22**: 55-110.

Bird, J., Kim, K. S., Rodriguez, R. L., Martin. E. M., Escudero, J. (1984). Transmission and cytopathology of Jatropha mosaic disease caused by a whitefly-transmitted geminivirus. *American Phytopathologycal Society- Canadian Phytopathologycal Society*. Annual meeting. August. Guelph. Ontario. Abstract A 554.

Birnboim, H. C., Siegel, A. (1982) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research*, **7**: 1513-1523.

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

Bock, K. R. (1974) Maize streak virus. *CMI/AAB Description of plant Viruses*, Nº 133, Commonwealth Mycologycal Institute. Kew, England.

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

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

Bock, K. R., Guthrie, E. J., Meredith, G., Barher, M. (1977) RNA and protein components of maize streak and cassava latent viruses. *Annals of Applied Biology*, **85**: 305-308.

Bock, K. R., Guthrie, E. J., Meredith,G. (1978) Distribution, host range, properties and purification of cassava latent virus, a geminivirus. *Annals of Applied Biology*, **90**: 361-367.

Bock, K. R., Guthrie, E. J., Figueiredo, G. (1981) A strain of cassava latent virus occuring in coastal districts of Kenya. *Annals of Applied Biology*, **99**: 151-159.

Bouzoubaa, S., Guilley, H., Jonard, G., Richards, K., Putz, C. (1985) Nucleotide sequence analysis of RNA-3 and RNA-4 of beet necrotic yellow vein virus, isolates F2 and G1. *Journal of general virology*, **66**: 1553-1564.

Breathnach, R., Chambon, P. (1981) Organisation and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry*, **50**: 349-383.

Brockman, W. W., Lee, T. N. H., Nathans, D. (1975) Evolutionary variants of simian virus 40. Characterization of cloned complementing variants. *Virology*, **66**: 36-52.

Brown, J. K., Nelson, M. R. (1984) Geminate particles associated with cotton leaf crumple disease in Arizona. *Phytopathology*, **74**: 987-990.

Brown, J. K., Nelson, M. R. (1986) Whitefly-borne virus of melons and lettuces. *Phytopathology*, **76**: 239.

Brown, J. L., Smith, A. E. (1970) Initiator codons in eukaryotes. *Nature*, **226**: 610-612.

Buck, K. W., Coutts, R. H. A. (1983) The potential of geminiviruses in the construction of plant host vector systems. *Plant Molecular Biology*, 2: 351-354.

Buck, K. W., Coutts, R. H. A. (1985) Tomato golden mosaic virus. *CMI/ABB Description of Plant Viruses*, Nº 303, Commonwealth Mycologycal Institute. Kew England.

Bullock, P., Champoux, J., Botcham, M. (1985) Association of crossover points with topoisomerase I cleavage sites: a model for nonhomologous recombination. *Science*, **230**: 954-958.

Butler, P. J. G. (1984) The current picture of the structure and assembly of tobacco mosaic virus. *Journal of General Virology*, 65: 253-279.

Carroll, D., O'Neill, F. J. (1978) Genome maps of simian virus 40 defectives propagated in human glioblastoma cells. *Virology*, **87**: 120-129.

Champoux, J. J. (1976) Evidence for an intermediate with a single-stranded break in the reaction catalyzed by the DNA untwisting enzyme. *Proceedings of the National Academy of Sciences of the United States of America*, **73**: 3488-3491.

Champoux, J. J. (1977) Strand breakage by the DNA untwisting enzyme results in covalent attachment of the enzyme to DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **74**: 3800-3804.

Chant, S. R. (1958) Studies on the transmission of cassava mosaic virus by *Bemisia* species (Aleyrodidae). *Annals of Applied Biology*, **46**: 210-215

Cherif, C. and Russo, M. (1983) Cytologycal evidence of the association of a geminivirus with the tomato yellow leaf curl disease in Tunisia. *Phytopathologishe Zeitschrift*, **108**: 221-225.

Clare, J., Farabaugh P. (1985) Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proceedings of the National Academy of Science of the United States of America*, **82**: 2829-2833.

Cohen, S., Nitzany, F. E. (1966) Transmission and host range of tomato yellow leaf curl virus. *Phytopathology*, **56**: 1127-1131.

Cohen, S., Duffus, J. E., Larsen, R. C.,Hiu, H. Y., Flock, R. A. (1983) Purification, serology and vector relationships of squash leaf curl virus, a whitefly-transmitted geminivirus. *Phytopathology*, **73**: 1669-1673.

Cole, C. N., Baltimore, D. (1973) Defective interfering particles of poliovirus. IV. Mechanisms of enrichment. *Journal of Virology*, **12**: 1414-1426.

Colman, A., Byers, M. J., Primrose, S. B., Lyons, A. (1978) Rapid purification of plasmid DNAs by hydroxyapatite chromatography. *European Journal of Biochemistry*, **91**: 303-310.

Costa, A. S. (1965) Three whitefly-transmitted virus diseases of beans in Sao Paulo, Brazil. *Food and Agriculture Organization, Plant Protection Bulletin*, **13**: 121-130.

Costa, A. S. (1976) Whitefly transmitted plant diseases. *Annual Review of Phytopathology*, **14**: 429-449.

Costa, A. S., Bennett, C. W. (1950) Whitefly transmitted mosaic disease of *Euphorbia prunifolia*. *Phytopathology*, **40**: 266-283.

Costa, A. S., Oliveira, A. R., Silva, D. N. (1977). *Summa Phytopathologica*, **3**: 194-200.

Coutts, R. H. A., Buck, K. W. (1985) DNA and RNA polymerase activities of nuclei and hypotonic extracts of nuclei isolated from tomato golden mosaic virus infected tobacco leaves. *Nucleic Acids Research* **13**: 7881-7897.

Coutts, R. H. A., Buck, K. W., Roberts, E. J. F., Brough, C. L., Hayes, R. J., Macdonald, H., MacDowell, S. W., Petty, I. T. D., Slomka, M. J., Hamilton, W. D. O., Bevan, M. W. (1986) Studies on the replication of tomato golden mosaic virus and the construction of gene vectors. In: *UCLA symposia series*, (Liss, A. R., ed.) New York. In press.

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

Daniell, E. (1976) Genome structure of imcomplete particles of adenovirus. *Journal of Virology*, **19**: 685-708.

Davis, R. W., Simon, M., Davidson, N. (1971) Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *Methods in Enzymology*, **21**: 413-428.

Davies, R. W. (1982) DNA sequencing. In: *Gel Electrophoresis of Nucleic Acids: A Practical Approach*, (D. Rickwood and D. D. Hames eds.) pp 117-146.

Debrot, E. A. (1981) Natural infection of potatoes in Venezuela with the whitefly transmitted mosaico amarillo tomato virus. *International Society of Plant Pathology:- Association of Applied Biologists.* Abstracts of papers given at a workshop on Pathogens Transmitted by Whiteflies in Kable College, Oxford, July 31$^{st}$ 1981.

Dente, L., Cesarini, G., Cortese, R. (1983) pEMBL: a new family of single stranded plasmids. *Nucleic Acids Research*, **11**: 1645-1655.

Deuring, R., Klotz, G., Doerfler, W. (1981) An unusual symmetric recombinant between adenovirus type 12 DNA and human cell DNA. *Proceedings of the National Academy of Science of the United States of America*, **78**: 3142-3146.

Deveraux, J., Haerberli, P., Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research*, **12**: 387-395.

Donson, J., Morris-Krsinish, 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. *The EMBO Journal*, **3**: 3069-3073.

Dretzen, G., Bellard, M., Sassone-Corsi, P., Chambon, P. (1981) A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Analytical Biochemistry*, **112**: 295-298.

Edwards, K. A., Halligan, B. D., Davis, J. L., Nivera, N. L., Liu, L. F. (1982) Recognition sites of eukaryotic DNA topoisomerase I: DNA nucleotide sequencing analysis of topo I cleavage sites on SV40 DNA. *Nucleic Acids Research*, **10**: 2565-2576.

Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., Proudfoot, N. J. (1980) The structure and evolution of the human β-globin gene family. *Cell*, **21**: 653-668.

Enea, V., Zinder, N. D. (1975) A deletion mutant of bacteriophage f1 containing no cistrons. *Virology*, **68**: 106-114.

Enea, V., Horiuchi, K., Turgeon, B. G., Zinder, N. D. (1977) Physical map of defective interfering particles of bacteriophage f1. *Journal of Molecular Biology*, **111**: 395-414.

Farabaugh, P. J., Schmeissner§, U., Hofer, M., Miller, J. H. (1978) Genetic studies of the *lac* repressor on the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *Journal of Molecular Biology*, **126**: 847-863.

Faust, E. A., Ward, D. C. (1979) Incomplete genomes of the paravirus minute virus of mice: selective conservation of genome termini, including the origin for DNA replication. *Journal of Virology*, **32**: 276-292.

Fitzgerald, M., Shenk, T. (1981) The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell*, **24**: 251-260.

Francki, R. I. B. (1985) Plant virus satellites. *Annual Review of Microbiology*, **39**: 151-174.

Francki, R. I. B., Matta, T. (1980) Chloris striate mosaic virus. *CMI/AAB Description of Plant Viruses*, Nº 221, Commonwealth Mycological Institute. Kew England.

Francki, R. I. B., Hatta, T., Grylls, N. E., Grivell, C. J. (1979) The particle morphology and some other properties of chloris striate mosaic virus. *Annals of Applied Biology*, **91**: 51-59.

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

Frenkel, N. (1981) Defective interfering herpesviruses. In : *The Human Herpesviruses-An Interdisciplinary Prospective*, (Nahmias, A. J., Dowdle, W. R., Schinazy, R. S., eds.) pp. 91-120. Elsevier/North-Holland, Amsterdam.

Fuller, C. (1901) Mealie variegation. *1st report on the Government Entomologist*. Natal. 1899-1900. pp. 17-19.

Galvez, G. E. and Castano, M. (1976) Purification of the whitefly-transmitted bean golden mosaic virus. *Turrialba*, **26**: 205-207

Ganem, D., Nussbaum, A. L., Davoli, D., Fareed, G. C. (1976) Propagation of a segment of bacteriophage λ-DNA in monkey cells after covalent linkage o a defective simian virus 40 genome. *Cell*, **7**: 349-359.

Garoff, H., Ansorge, W. (1981) Improvements of DNA sequencing gels. *Analytical Biochemistry*, **15**: 450-457.

Gellert, M. (1981) DNA topoisomerases. *Annual Review of Biochemistry*, **50**: 879-910.

Gilmour, D. S., Pflugfelder, G., Wang, J. C., Lis, J. T. (1986) Topoisomerase I interacts with transcribed regions in drosophila cells. *Cell*, **44**: 401-407.

Goodman, R. M. (1977a) Infectious DNA from a whitefly-transmitted virus of *Phaseolus vulgaris*. *Nature*, **266**: 54-55.

Goodman, R. M. (1977b) Single-stranded DNA genome in a whitefly-transmitted plant virus. *Virology*, **83**: 171-179.

Goodman, R.M. (1981a) Geminiviruses: the single-stranded DNA viruses of plants. in: *Handbook of Plant Virus Infections and Comparative Diagnosis* (Kurstak, E., ed.) pp. 883-910. Biomedical Press, Elsevier/North-Holland, Amsterdam.

Goodman, R. M. (1981b) Geminiviruses. *Journal of general virology*, **54**: 9-21.

Goodman, R. M. and Bird, J. (1978) Bean golden mosaic virus. *CMI/AAB Descriptions of Plant Viruses*, Nº 192, Commonwealth Mycological Institute, Kew, England.

Goodman, R. M., Bird, J., Thongmeearkom, P. (1977) An unusual virus-like particle associated with golden yellow mosaic of beans. *Phytopathology*, **67**: 37-42.

Goodman, R. M., Shock, T. L., Haber, S., Browning, K. S., Bowers, G. R. (1980) The composition of bean golden mosaic virus and its single-stranded DNA genome. *Virology*, **106**: 168-172.

Gould, A. R., Symons, R.H. (1977) Determination of the sequence homology between the four RNA species of cucumber mosaic virus by hybridization analysis with complementary DNA. *Nucleic Acids Research*, **4**: 3787-3808.

Griffith, J., Kornberg, A. (1974) Mini M13 bacteriophage: circular fragments of M13 DNA are replicated and packaged during normal infections. *Virology*, **59**: 139-152.

Grimsley, N., Hohn, B., Hohn, T., Walden, R. (1986) "Agroinfection", an alternative route for viral infection of plants by using the Ti plasmid. *Proceedings of the National Academy of Sciences of the United States of America*, **83**: 3282-3286.

Grosveld, G. C., Shewmaker, C. K., Jat, P., Flavell, R.A. (1981) Localization of DNA sequences necessary for transcription of the rabbit β-globin gene *in vitro*. *Cell*, **25**: 215-226.

Grosveld, G. C., de Boer, E., Shewmaker, C. K., Flavell, R. A. (1982) DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. *Nature*, **295**: 120-126.

Grylls, N. E. (1963) A striate mosaic virus disease of grasses and cereals in Australia, transmitted by the cicadellia *Nesoclutha obscura*. *Australian Journal of Agriculture*, **14**: 143-153.

Haber, S., Ikegami, M., Bajet, N. B., Goodman, R. M. (1981) Evidence for a divided genome in bean golden mosaic virus, a geminivirus. *Nature*, **289**: 324-326.

Hamilton, W. D. O., Sanders, R. C., Coutts, R. H. A., Buck, K. W. (1981) Characterisation of tomato golden mosaic virus as a geminivirus. *FEMS Microbiology Letters*, **11**: 263-267.

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

Hamilton, W. D. O., Bisaro, D. M., Coutts, R. H. A., Buck, K. W. (1983) Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. *Nucleic Acids Research*, **11**: 7387-7396.

Hamilton, W. D. O., Stein V. E., Coutts, R. H. A., Buck, K. W. (1984) Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: Potential coding regions and regulatory sequences. *The EMBO Journal*, **3**: 2197-2205.

Hammarskjold, M. L., Winberg, G. (1980) Encapsidation of adenovirus 16 DNA is directed by a small DNA sequence at the left end of the genome. *Cell*, **20**: 787-795.

Harrison, B. D. (1985) Advances in geminivirus research. *Annual Review of Phytopathology*, **23**: 55-85.

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

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

Henriksson, D., Tanis, R. J., Tashian, R. E., Nyman, P. O. (1981) Amino acid sequence of the coat protein subunit in satellite tobacco necrosis virus. *Journal of Molecular Biology*, **152**: 171-179.

Herrera-Estrella, L., de Block, M., Messens, E., Hernalsteens. J.-P., van Montagu, M., Schell, J. (1983) Chimeric genes as dominant selectable markers in plant cells. *The EMBO Journal*, **2**: 987-995.

Hindley, J. (1983) DNA sequencing. In: *Laboratory techniques in biochemistry and molecular biology*, Vol 10, (T.S. Work, Burton R. S., eds.) pp. 200-206. Elsevier Biomedical Press.

Holland, J. J. (1985) Generation and replication of defective viral genomes. In: *Virology*. (B. N. Fields *et al.* eds.) Chapter 6, pp. 77-99. Raven Press, New York.

Holland, J. J., Kennedy, S. I. T., Semler, B. L., Jones, C. L., Roux, L., Grabau, E. A. (1980) Defective interfering RNA viruses and the host cell response. In: *Comprehensive Virology*, Vol. 16, (Frankel-Conrat, H., Wagner, R. R., eds.) pp. 137-192. Plenum Press, New York.

Honda, Y., Iwaki, M., Saito, Y. (1983) Mechanical transmission, purification and some properties of whitefly-borne mung bean yellow mosaic virus in Thailand. *Plant Disease*, **67**: 801-804.

Honda, Y., Iwaki, M., Saito, Y., Thongmeearkom, P., Kittisak, K., Deema, N. (1983) Mechanical transmission, purification, and some properties of whitefly-borne mung bean yellow mosaic virus in Thailand. *Plant Disease*, **67**: 801-804.

Hoekema, A., Hairsch, P. R., Hooykas, P. J. J., Schilperoot, R. A. (1983) A binary plant vector strategy based on separation of VIR and T-region of the *Agrobacterium tumefaciens* Ti plasmid. *Nature*, **303**: 179-181.

Hooker, W. J., Salazar, L. F. (1983) A new plant virus from the high jungle of the Eastern Andes, Solanum apical leaf curling virus (SALCV). *Annals of Applied Biology*, **103**: 449-454.

Hooykaas-van Slogteren, G. M. S., Hooykaas, P. J. J., Schilperoot, R. A. (1984) Expression of Ti plasmid genes in monocotyledonous plants infected with *Agrobacterium tumefaciens*. *Nature*, **311**: 763-764.

Horiuchi, K., Zinder, N. D. (1976) Origin and direction of synthesis of bacteriophage f1 DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **73**: 2341-2345.

Howarth, A. J., Gardner, R. C., Messing, J., Shepherd, R. J. (1981) Nucleotide sequence of naturally occuring deletion mutants of cauliflower mosaic virus. *Virology*, **112**: 678-685.

Howarth, A. J., Goodman, R. M. (1982) Plant viruses with genomes of single-stranded DNA. *Trends in Biochemical Sciences*, **7**: 180-182.

Howarth, A. J., Caton, J., Bossert, M., Goodman, R. M. (1985) Nucleotide sequence of bean golden mosaic virus and a model for gene regulation in geminiviruses. *Proceedings of the National Academy of Sciences of the United States of America*, **82**: 3572-3576.

Howell, S. H. (1982) Plant molecular vehicles: Potential vectors for introducing foreign DNA into plants. *Annual Review of Plant Physiology*, **33**: 609-650.

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

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

Howell, S. H., Walker, L. L., Walden, R. M. (1981) Rescue of *in vitro* generated mutants of cloned cauliflower mosaic vius genome in infected plants. *Nature*, **293**: 483-486.

Hu, N., Messing, J. (1982) The making of strand-specific M13 probes. *Gene*, **17**: 271-277.

Huang, A. S. (1973) Defective interfering viruses. *Annual Review of Microbiology*, **27**: 101-117.

Huang, A. S. (1977) Viral pathogenesis and molecular biology. *Bacteriological Reviews*, **41**: 811-821.

Huang, A. S., Baltimore, D. (1970) Defective viral particles and viral disease processes. *Nature*, **226**: 325-327.

Hull, R., Davies, J. W. (1983) Genetic engineering with plant viruses, and their potential as vectors. *Advances in virus research*, **28**: 1-33.

Hunt, T. (1985) False starts in translational control of gene expression. *Nature*, **316**: 580-581.

Ikegami, M., Harber, S., Goodman, R. (1981) Isolation and characterization of virus-specific double-stranded DNA from tissues infected by bean golden mosaic virus. *Proceedings of the National Academy of Sciences of the United States of America*, **78**: 4102-4106.

Ikegami, M., Morinaga, T., Miura, K. (1984) Infectivity of virus-specific double-stranded DNA from tissue infected by mung bean yellow mosaic virus. *Virus Research*, **1**: 507-512.

Inouye, T., Osaki, T. (1980) The first record in the literature of the possible plant virus disease that appeared in "Manyoshu" a Japanese classic anthology, as far back as the time of the 8$^{th}$ century. *Annual Phytopathological Society of Japan*, **46**: 49-50.

Ish-Horowicz, R., Burke, J. F., (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Research*, **9**: 1989-2998.

Iwaki, M., Thongmeearkom, P., Honda, Y., Deema, N. (1983) Soybean crinkle leaf: A new whitefly-borne disease of soybean. *Plant Disease*, **67**: 546-548.

Jeske, H., Werz, G. (1980) Ultrastructural and biochemical investigations on the whitefly transmitted abutilon mosaic virus (AbMV). *Phytopatholische Zeitschrift*, **97**: 43-55.

Jones, I. M., Reichmann, M. E. (1973) The proteins synthesized in tobacco leaves infected with tobacco necrosis virus and satellite tobacco necrosis virus. *Virology*, **52**: 46-56.

Johson, A. D., Barkan, A., Mertz, J. E. (1982) Nucleotide sequence analysis of the recombinant joints in 16 naturally arising deletion mutants of simian virus 40. *Virology*, **123**: 464-469.

Kassanis, B. (1962) Properties and behaviour of a virus depending for its multiplication on another. *Journal of General Microbiology*, **27**: 477-488.

Kassanis, B., Nixon, H. L. (1961) Activation of one virus by another. *Journal of General Microbiology*, **25**: 459-471.

Kassanis, B., Phillips, M. P. (1970) Serological relationship of strains of tobacco necrosis virus and their ability to activate strains of satellite virus. *Journal of General Virology*, **90**: 119-126.

Kassanis, B., White, R. F. (1972) Interference between two satellite viruses of tobacco necrosis virus. *Journal of General Virology*, **17**: 177-183.

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

Kim, K. S., Fulton, R. W. (1984) Ultrastructure of *Datena stramonium* infected with an Euphorbia virus suggestive of a whitefly-transmitted geminivirus. *Phytopathology*, **74**: 236-241.

Kim, K. S., Shock, T.L., Goodman, R.M. (1978). Infection of *Phaseolus vulgaris* by bean golden mosaic virus: ultrastructural aspects. *Virology*, **89**: 22-23.

Kim, K. S., Bird, J., Rodriguez, R. L., Martin, E. M., Escudero, J. (1986) Ultrastructural studies of *Jatropha gossypifolia* infected with jatropha mosaic virus, a whitefly-transmitted geminivirus. *Phytopathology*, **76**: 80-85.

Klein, W. H., Nolan, C., Lazar, J. M., Clark, J. M. (1972) Translation of satellite tobacco necrosis virus ribonucleic acid. I. Characterization of *in vitro* procariotic and eucariotic translation products. *Biochemistry*, **11**: 2009.

Kornberg, A. (1980) *DNA Replication.* Freeman, San Francisco.

Kozak, M. (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Research*, **9**: 5233-5252.

Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiological Reviews*, **47**: 1-45.

Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Research*, **12**: 857-871.

Larsen,R. C., and Duffus, J. E. (1984) A simplified procedure for the purification of curly top virus and the isolation of its monomer and dimer particles. *Phytopathology*, **73**: 114-118

Larsen, S. H. (1982) Evolutionary variants of mouse adenovirus containing cellular DNA sequences. *Virology*, **116**: 573-580.

Lazzarini ,R. A., Keene J. D., Schubert, M. (1981) The origins of defective interfering particles of the negative-strand RNA viruses. *Cell*, **26**: 145-154.

Lee, T. N. H., Nathans, D. (1979) Evolutionary variants of simian virus 40: replication and encapsidation of variant DNA'. *Virology*, **92**: 291-298.

Lee, T. N. H., Brockman, W. W., Nathans, D. (1975) Evolutionary variants of simian virus 40: Cloned substituted variants containing multiple initiation sites for DNA replication. *Virology*, **66**: 53-69.

Leppert, M., Kort, L., Kolakofsky, D. (1977) Further characterization of Sendai virus DI-RNAs: A model for their generation. *Cell*, **12**: 539-552.

Lindsten, K., Vacke, J., Gerhardson, B. (1970) A preliminary report on three cereal virus diseases new to Sweden spread by *Macrosteles* and *Psammotettix* leafhoppers. *National Swedish Institute for Plant Protection Contributions*, **14**: 285-297.

Lindsten, K., Lindsten, B., Abdelmoeti, M., Junti, N. (1980) Purification and some properties of wheat dwarf virus. *Proceedings on the 3rd Conference on Virus Diseases of the Graminae*, pp. 27-33, Rothamsted, England.

Lörz, H., Baker, B., Schell, J. (1985) Gene transfer to cereal cells mediated by protoplast transformation. *Molecular and General Genetics*, **199**: 178-182.

Lycett, W. G., Croy, R. R. D., Shirsat, A. H., Boulter, D. (1984) The complete nucleotide sequence of a leguminin gene from pea (*Pisum sativum* L.). *Nucleic Acids Research*, **12**: 4493-4507.

von Magnus, P. (1954) Incomplete forms of influenza virus. *Advances in Virus Research*, **2**: 59-79.

Magyarosy, A. C. (1980) Beet curly top virus transmission by artificially fed and infected beet leafhoppers *(circulifer tenellus)*. *Annals of Applied Biology*, **96**: 301-305

Maniatis, T., Fritsch, E. F., Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbour Press, New York.

Marriott, T. W., Symons, R. M. (1982) Determination of the genome structure of chloris striate mosaic virus, a geminivirus. *Abstracts of the IVth International Conference on Comparative Virology, Control of Virus Diseases*. Banff, Canada.

Matthews, R. E. F. (1979) Classification and Nomenclature of viruses. Third report of the International Committee on the Taxonomy of Viruses. *Intervirology*, **12**: 129-296

Matyis, J. C.,Silvar, D. M., Oliveira, A. R., Costa, A. S. (1975) Purificacao e morfologia do virus do mosaico dourado do tomateiro. *Summa Phytopathologica*, **1**: 267-274

de la Maza, L. M., Carter, B. J. (1980) Molecular structure of adeno-associated virus variant DNA. *The Journal of Biological Chemistry*, **255**: 3194-3203.

McKnight, S. L., Kingsbury, R. (1982) Transcriptional control signals of a eukaryotic protein coding gene. *Science*, **217**: 316-324.

Meiners, J. P., Lawson, R. H., Smith, F. F., Diaz, A. J. (1973). Mechanical transmission of a whitefly-borne disease agent of beans in El Salvador. Abstracts of the 30$^{th}$ Annual Meeting of the Potomac Division of the American Phytopathological Society, *Phytopathology*, **63**: 803-804

Mellon, P., Parker, V., Gluzman, Y., Maniatis, T. (1981) Identification of DNA sequences required for transcription of the human α1-globin gene in a new SV40 host-vector system. *Cell*, **27**: 279-288.

Messing, J., Vieira, J. (1982) A new pair of M13 vectors for selecting either double-digest restriction fragments. *Gene*, **19**: 269-276.

Messing, J., Geraghty, D., Heidecker, G., Hu, N.-T., Kridl, J., Rubenstein, I. (1983) Plant gene structure. In: *Genetic Engineering of Plants*, (Kasuge,T., Meredith, C. P., Hollaender, A., eds.) Plenum Press, New York, pp. 211-217.

Mohana Rao, J. K. (1985) Sequence homology among the coat proteins og geminiviruses. *Biochemical and Biophysical Research Communications*, **31**: 892-896.

Mohana Rao, J. K. (1986) Sequence homology between the coat proteins od DNA and RNA plant viruses. *Biochemical and Biophysical Research Communications*, **134**: 372-377.

Morinaga, T., Ikegami, M., Miura, K. (1983) Infectivity of the cloned DNAs from multiple genome components of bean golden mosaic virus. *Proceedings of the Japanese Academy B*, **59**: 363-366

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

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

Mullineaux, P. M., Donson, J., Stanley, J., Boulton, M. I., Morris-Krsinich, B. A. M., Markham, P. G., Davies, J. W. (1985) Computer analysis identifies sequence homologies between potential gene products of maize streak virus and those of cassava latent virus and tomato golden mosaic virus. *Plant Molecular Biology*, **5**: 125-131.

Mumford, D. L. (1974) Purification of curly top virus. *Phytopathology*, **64**: 136-139

O'Farrel, P. H., Kutter, E., Nalcanishi, M. (1980) A restriction map of the bacteriophage T4 genome. *Molecular and General Genetics*, **179**: 421-435.

Olszewski, N., Hagen, G. Guilfoyle, T. J. (1982) A transcriptionally active, covalently closed minichromosome of cauliflower mosaic virus DNA isolated from infected turnip leaves. *Cell*, **29**: 395-402.

O'Neill, F. J., Maryon, E. B., Carroll, D. (1982) Isolation and characterization of defective simian virus 40 genomes which complement for infectivity. *Journal of Virology*, **43**: 18-25.

Osaki, T., Inoye, T., (1981). Tobacco leaf curl virus. *CMI/AAB. Description of Plant Viruses*, Nº 232, Commonwealth Mycological Institute. Kew, England.

Osaki, T., Kobatake, M., Inouye, T. (1979) Mosaic of honeysuckle (*Lonicera Japonica Thunb.*) a disease caused by tobacco leaf curl virus in Japan. *Annals of the Phytopathologycal Society of Japan*, **45**: 62-69

Paszkowski, J., Shillito, R. D., Saul, M., Mandak, V., Hohn, T., Hohn, B., Potrykus, I. (1984) Direct gene transfer to plants. *The EMBO Journal*, **3**: 2717-2722.

Potrykus, I., Saul, M. W., Petruska, J., Paskowski, J., Shillito, R. (1985) Direct gene transfer to cells of a graminaceous monocot. *Molecular and General Genetics*, **199**: 183-188.

Proudfoot, N. J., Brownlee, G. G. (1976) 3' non-coding region sequences in eukaryotic messenger RNA. *Nature*, **263**: 211-214.

Ravetch, J. V., Horiuchi, K., Zinder, N. D. (1979) DNA sequence analysis of the defective interfering particles of bacteriophage f1. *Journal of Molecular Biology*, **128**: 305-318.

Ray, D. S. (1976) Replication of filamentous bacteriophages.In: *Comprehensive Virology*, Vol 7 (Fraenkel-Conrat, H., Wagner, R., eds.) pp. 105-166. Plenum Press, New York.

Reddy, D. V. R., Black, L. M. (1974) Deletion mutations of the genome segments of wound tumor virus. *Virology*, **61**: 458-473.

Reisman, D., Ricciardi, R. P., Goodman, R. M. (1979) Size and topology of single-stranded DNA from bean golden mosaic virus. *Virology*, **97**: 388-395

Rigby, P. W. J. (1982) Expression of cloned genes in eukaryotic cells using vector systems derived from viral replicons. In: *Genetic Engineering* Vol.3, (Williamson, R. ed.) pp.84-134. Academic Press, London.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., Berg, P. (1977) Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *Journal of Molecular Biology*, **113**: 237-251.

Roberts, E. J. F., Buck, K. W., Coutts, R. H. A. (1986) A new geminivirus infecting potatoes. *Plant Disease*, **70**: 603.

Rogers, S. G., Bisaro, D. M., Horsch, R. B., Fraley, R. T., Hoffmann, N. L., Brand, L., Elmer, J. S., Lloyd, A. M. (1986) Transgenic plants containing integrated tomato golden mosaic virus genomes: Independent replication of the A component and symptom production in hybrid progenny. *Cell*, **45**: 593-600.

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

Roizman, B. (1979) The structure and isomerization of herpes simplex virus. *Cell*, **16**: 481-494.

Rose, D. J. W. (1978) Epidemiology of maize streaak virus. *Annual Review of Entomology*, **23**: 259-282

Sander, M., Hsieh, T. (1985) *Drosophila* topoisomerase II double-stranded DNA cleavage: analysis of DNA sequence homology at the cleavage site. *Nucleic Acids Research*, **13**: 1057-1073.

Sanger, F., Coulson, A. R. (1978) The use of thin acrylamide gels for DNA sequencing. *FEBS Letters*, **87**: 107-110.

Sanger, F., Nicklen, S., Coulson, A. R. (1977) DNA Sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **74**: 5463-5467.

Sanger, F., Coulson, A. R., Friedman, T., Air, G. M., Barrel., B. G., Brown, N. L., Fiddes, J. C., Hutchinson, C. A., Slocombe, P. M., Smith, M. (1978) *Journal of MOlecular Biology*, **125**: 225-246.

Sanger, F., Coulson, A. R., Barrel, B. G., Smith, A. J. H., Roe, B. A. (1982) Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *Journal of Molecular Biology*, **143**: 161-178.

Sanger, F. Coulson, A. R., Hong, G. F., Hill, D. F., Petersen, G. B. (1982) Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, **162**: 729-773.

Schubert, M., Keene, J. D., Lazzarini, R. A., Emerson, S. U. (1978) The complete sequence of a unique RNA species synthesized by a DI particle of VSV. *Cell*, **15**: 103-112.

Sequeira, J. C. (1982) Purification, properties and relationships of cassava latent virus. *PhD thesis.* Dundee University. Scotland. Cited in Harrison, 1985.

Sequeira, J. C., Harrison, J. D. (1982) Serological studies on cassava latent virus. *Annals of Applied Biology*, **101**: 33-42

Severin, H. H. P. (1931) Modes of curly-top transmission by the beet leafhoppers, *Euterrix tenellus* (Baker). *Helgardia*, **6**: 253-276

Shepherd, R. J. (1976) DNA viruses of higher plants. *Advances in Virus Research*, **20**: 305-339

Shepherd, R. J., Lawson, R. H. (1981) Caulimoviruses. In: *Handbook of Plant Virus Infection and Comparative Diagnosis*, (Kurstak, E., ed.) chap.25. Biomedical Press, Elsevier/North-Holland, Amsterdam.

Shirakko, Y., Brakke, M. K. (1984) Spontaneous deletion mutation of soil-borne wheat mosaic virus RNA II. *Journal of General Virology*, **65**: 855-858.

Shoulder, A., Darby, G., Minson, T. (1974) RNA-RNA hybridization using $^{125}$I-labelled RNA from tobacco necrosis virus and its satellite. *Nature*, **251**: 733-735.

Simons, J. N. (1962) The pseudo-curly top disease in South Florida. *Journal of Economical Entomology*, **55**: 358-363

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

Sims, J., Dressler, D. (1978) Site-specific initiation of a DNA fragment: nucleotide sequence of the bacteriophage G4 negative-strand initiation site. *Proceedings of the National Academy of Sciences of the United States of America*, **75**: 3094-3098.

Slykhuis, J. T. (1963) Vector and host relations of north American wheat striate mosaic virus. *Canadian Journal of Botany*, **41**: 1171-1185

Southern, E. M. (1975) Detection of a specific sequence among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, **98**: 503-517.

Southern, E. M. (1979) DNA length measurement by gel electrophoresis. *Analytical Biochemistry*, **100**: 319.

Spaete, R. R., Frenkei, N. (1982) The herpes simplex virus amplicon: a new eucaryotic defective-virus cloning-amplifying vector. *Cell*, **30**: 295-304.

Stanley, J. (1983) Infectivity of the cloned geminivirus genome requires sequences from both DNAs. *Nature*, **305**: 260-262

Stanley, J. (1986) The molecular biology of geminiviruses. *Advances in Virus Research*, **30**: 139-177.

Stanley, J., Davies, J. W. (1985) Structure and function of the DNA genome of geminiviruses. In: *Plant Molecular Virology* Vol II (J.W.Davies,ed.) pp. 191-218. CRC press, Boca Raton, Florida.

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

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

Stanley, J., Townsend, R., Curson, S. J. (1985) Pseudorecombinants between cloned DNAs of two isolates of cassava latent virus. *Journal of General Virology*, **66**: 1055-1061.

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

Stein, V. E. (1985) Nucleotide sequence analysis of infectious cloned DNA of tomato golden mosaic virus. *Phd thesis*, University of London.

Stein, V. E., Coutts, R. H. A., Buck, K.W. (1983) Serological studies on tomato golden mosaic virus, a geminivirus. *Journal of General Virology*, **64**: 2493-2498

Storey, H. H. (1928) Transmission studies on maize streak disease. *Annals of Applied Biology*, **15**: 1-25

Storey, H. H. (1938) Investigation of the mechanism of plant viruses by insect vectors. II. The part played by puncture in transmission . *Proceedings of the Royal Society. London Series B* ,**125**: 455-477

Storey, H. H. (1939) Transmission of plant viruses by insects. *Botany Review*, **5**: 240-272.

Storey, H. H., Nichols R. F. W. (1938) Studies on the mosaic diseases of cassava. *Annals of Applied Biology*, **98**: 261-276

Subramani, S., Southern, P. (1983) Analysis of gene expression using simian virus 40 vectors. *Analytical Biochemistry*, **135**: 1-15.

Sunter, G., Coutts, R.H.A., Buck, K.W. (1984) Negatively supercoiled DNA from plants infected with a single-stranded DNA virus. *Biochemical and Biophysical Research Communications*, **118**: 747-752.

Sunter, G., Buck, K. W., Coutts, R. H. A. (1985) S1-sensitive sites in the supercoiled double-stranded form of the B DNA component of tomato golden mosaic virus: Identification of regions of potential alternative secondary structure and regulatory function. *Nucleic Acids Research*, **13**: 4645-4659.

Tautz D., Renz, M. (1983) An optimized freeze-squeeze method for the recovery of DNA fragments from agarose gels. *Analytical Biochemistry*, **132**: 14-19.

Taylor, J. M., Illmensee, R., Summers, J. (1976) Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. *Biochimica et Biophysica Acta*, **473**: 57-71.

Temple, M., Antoine, G., Delius, H., Stahl, S., Winnacker, E.-L. (1981) Replication of mouse adenovirus strain AdFL DNA. *Virology*, **109**: 1-12.

Thomas, J. E. and Bonyer, J. W. (1980) Properties of tobacco yellow dwarf and bean summer death viruses. *Phytopathology*, **70**: 214-217

Thomas, P. E., Mink, G. I. (1979) Beet curly top virus. *CMI/AAB Description of Plant viruses*, Nº 210. Commonwealth Mycological Institute. Kew, England

Thummel, C, Tijan, R., Grodzicker, T. (1981) Expression of SV40 T antigen under control of adenovirus promoters. *Cell*, **23**: 825-836.

Tinoco, I., Borer, P. N., Dengler, B., Levine, M. D. (1973) Improved estimation of secondary structure in ribonucleic acids. *Nature*, **246**: 40-41.

Tomenius, K., Oxelfelt, P. (1981) Preliminary observations of virus-like particles in nuclei in cells of wheat infected with the wheat dwarf disease. *Phytopathologische Zertschrift*, **101**: 163-167.

Tooze, J. (1980) *Molecular Biology of Tumor Viruses Part 2, DNA Tumor Viruses*. Cold Spring Harbor Laboratory Press, New York.

Townsend R., Watts, J., Stanley, J. (1986) Synthesis of viral DNA in *Nicotiana plumbaginifolia* protoplasts inoculated with cassava latent virus (CLV): evidence for the independent replication of one component of the CLV genome. *Nucleic Acids Research* **14**: 1253-1265.

Townsend, R., Stanley, J., Curson, S. J., Short, M. N. (1985) Major polyadenylated transcripts of cassava latent virus and location of the gene encoding coat protein. *The EMBO Journal*, **4**: 33-37.

Townsend R., Watts, J., Stanley, J. (1986) Synthesis of viral DNA in *Nicotiana plumbaginifolia* protoplasts inoculated with cassava latent virus (CLV): evidence for the independent replication of one component of the CLV genome. *Nucleic Acids Research*, **14**: 1253-1265.

Udvardy, A., Schedl, P., Sander, M., Hsieh, T. S. (1985) Novel partitioning of DNA cleavage sites for drosophila topoisomerase II. *Cell*, **40**: 933-941.

Vacke, J. (1961) Wheat dwarf virus disease. *Biologia Plantarum*, 3: 228-233.

Vacke, J. (1972) Host plants range and symptoms of wheat dwarf virus. *Vedecke Prace Vyzk. Ustavu Rostl. Vyroby, Praha-Ruzyne*, **17**: 151-162.

Varmus, H. E. (1985) Reverse transcriptase rides again. *Nature*, **314**: 583-584.

Vlazny, D. A., Frenkel, N. (1981) Replication of herpes simplex virus DNA: localization of replication recognition signals within defective virus genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **78**: 742-746.

Vlazny, D. A., Kwong, A., Frenkel, N. (1982) Site-specific cleavage/packaging of herpes simplex virus DNA and the selective maturation of nucleocapsids containing full-length viral DNA. *Proceedings of the National Academy of Sciences of the United States of America*, **79**: 1423-1427.

Vogt, V. M. (1973) Purification and further properties of single strand specific nuclease from *Aspergillus oryzae*. *European Journal of Biochemistry*, **33**: 192-200.

Wahl, G. M., Stern, M., Stark, G. R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulphate. *Proceedings of the National Academy of Sciences of the United States of America*, **76**: 3683-3687.

Walter, B. (1980) Isolation and purification of a virus transmitted from mosaic-disease cassava in the Ivory Coast. *Plant disease*, **64**: 1040-1042

Weiss, B. (1976) Endonuclease II of *Escherichia coli* is exonuclease III. *Journal of Biological Chemistry*, **251**: 1896-1901.

Weiss, R. B. (1984) Molecular model of ribosome frameshifting. *Proceedings of the National Academy of Sciences of the United States of America*, **81**: 5797-5801.

Yamashita, S. (1983) Miscanthus strak virus. In: *Handbook of Plant Viruses in Japan*. (Yora, K., Saito, S., Do, Y., Inovye, T., Tomaru, K., eds.) pp.372-373. Tokyo: Asakura Shoten.

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Van Montagu, M., Schell, J. (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *The EMBO Journal*, **2**: 2143-2150.

Zassenhaus, H. P., Butow, R. A., Hannon, Y. P. (1982) Rapid electroelution of nucleic acids from agarose and acrylamide gels. *Analytical Biochemistry*, **125**: 125-130.