ASPECTS OF LUTEOVIRUS MOLECULAR BIOLOGY IN RELATION TO THE INTERACTION BETWEEN BYDV-PAV AND THE Yd2 RESISTANCE GENE OF BARLEY

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Adelaide

by

John Paul Rathjen (B.Sc. Hons)

March, 1995

Errata

p40 Section 2.2.12. Ribonucleotides (ATP, CTP and GTP) were used in *in vitro* transcription reactions, not deoxynucleotides as indicated.

p42 Section 3.1. The acronym for subterranean clover red leaf virus is SCRLV, not SCLRV.

Fig. 6.1 Plasmid pC8 contains a recognition site for *Stu*I site between the 35S promoter and terminator as indicated for pC7.1.

Figs 6.6 and 6.7 The extremities of individual RACE products are indicated by numbers in circles.

TABLE OF CONTENTS

Chapter Six **Full-length cDNA Clones of BYDV-RPV and BYDV-PAV Genomic RNAs for Agroinfection**

SUMMARY

The general aim of this thesis was to examine the interaction between the PAV strain of barley yellow dwarf luteovirus (BYDV-PAV), and a naturally occurring resistance gene from barley known as Yd2. Barley yellow dwarf disease is caused by at least five different strains of BYDV. These strains fall into either of the two luteoviral subgroups. Subgroup I BYDV strains are sensitive to Yd2 (*e.g.* BYDV-PAV), whereas subgroup II strains (*e.g.* BYDV-RPV) are not. Knowledge of the genomic region controlling interaction of BYDV-PAV with the Yd2 resistance gene offers the potential for understanding the function of the gene, and possibly to facilitate its isolation.

The partial nucleotide sequence of soybean dwarf luteovirus (SDV) was completed in order to gain a better understanding of the organisation and evolution of the luteoviral genome. The SDV genome is 5861 nucleotides (nt) in length and encodes five major reading frames possessing conservation of sequence and organisation with known luteovirus sequences. Comparative analyses of the genome structure revealed that SDV shares sequence homology and features of genome organisation with BYDV-PAV in the 5'-half of the genome, yet is more closely related to subgroup II luteoviruses in its 3' coding regions. In addition, SDV lacks the small 3' open reading frame (ORF 6) unique to subgroup I luteoviruses. The SDV genome is thus a chimaera most likely to have been formed by RNA recombination between members of different luteoviral subgroups.

These data were used to derive a model of luteovirus genome organisation based on the association of similar gene activities to modules. ORFs in the 5' half of the genome control replication, while those in the 3' half control viral movement, including cell-cell, long-range, and plant-plant movement. The functional organisation of the luteovirus genome proposed here allows the design of module-swapping experiments between the genomes of BYDV-PAV and -RPV. Change in the sensitivity of

recombinant viruses to Yd2 will allow definition of the genomic region controlling the interaction with the resistance gene.

Infectious cDNA clones of BYDV genomic RNAs are required before manipulation of the viral genome is possible. Luteoviruses are not mechanically transmissible, so adaptation of the agroinfection technique for the infection of intact plants was proposed. The first step in this procedure was to construct a plasmid vector (pCass) containing the transcriptional promoter and terminator elements from the cauliflower mosaic virus (CaMV) 35S gene. The promoter was modified such that viral sequences can be inserted at the first transcribed nucleotide. cDNA clones of the three cucumber mosaic cucumovirus genomic RNAs were cloned into pCass to test its utility in the construction of infectious clones. A generalised strategy for cloning viral cDNA sequences into pCass was devised. The CMV cDNAs cloned into pCass were infectious when inoculated together onto susceptible host plants. The infectivity of the constructs was substantially enhanced when the transcriptional cassette, containing the CaMV 35S transcriptional elements fused to the viral cDNA, was excised by restriction digestion prior to inoculation.

The second part of the strategy was to place full-length cDNA clones of BYDV-RPV and -PAV genomic RNAs under control of the pCass transcriptional elements, and to assemble the entire construct in an *Agrobacterium* binary vector suitable for agroinfection. A full-length cDNA clone of an unsequenced isolate of BYDV-RPV was constructed using a polymerase chain reaction (PCR) mediated approach. Restriction sites were engineered into the BYDV-RPV genomic RNA sequence using silent mutagenesis in the PCR primer sites, allowing construction of the full-length clone. Sequence analysis of the 5' and 3' genomic ends of the BYDV-RPV isolate revealed structures not previously reported for this virus. In particular, the 5' terminal nucleotide sequence shows conservation with those of other subgroup II luteoviruses, in contrast to a published report. A form of this sequence was present at the 3' genomic terminus, albeit in the reverse orientation and complementary to the 5' sequence. This has not been previously demonstrated for any other luteovirus. Full-length cDNAs of

the BYDV-RPV clone assembled here and a clone of BYDV-PAV constructed by other workers were subcloned into the binary vector under control of the CaMV 35S transcriptional sequences. The full-length BYDV cDNA clones were not infectious in preliminary agroinfection experiments.

STATEMENT

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference has been made in the text.

I consent to this thesis being made available for photocopying and loan.

John P Rathjen

March, 1995

ACKNOWLEDGMENTS

An undertaking as large and as prolonged as this Ph.D has proved to be inevitably relies on the help and support of a number of people. It is difficult to acknowledge the full extent of each contribution in this format without appearing glib. However, I owe thanks to the following people:

- my supervisor, Professor Bob Symons, for guidance, encouragement and support.

- the Grains Research and Development Corporation for generous support.

- Chris Ford, for too many things to mention here, but including general supervision, satirical humour and thesis correction.

- Shou-Wei Ding, for a lot of scientific (and other) guidance, and more thesis corrections.

- Nich Collins - we're both still alive after four years.

- Masoud Shams-bakhsh, for patience and friendship.

- members of the Symons' lab and the Waite community past and present, including (in no order) Jim McInnes, Alex Jeffries, Dave Warrilow, Nick Paltridge, Neil Shirley, Ali Afsharifa, Yan Wan Chow Wah, Mandy Jolly, Helen Healy, Dingo, Dags, Molesworth, Dave Tatts and Kyles, Bujun Shi, Wendy Winnall, Bob Staysis, Ian Dry, John Randles, Peter Langridge, Geoff Fincher, Tony Rathjen, Abdolreza Bagheri, Angelo Karakousis, Jan Nield, Wan-Xiang Li, Ruth Ellickson, Barbara Engel.

- Drs Paul Keese and Peter Waterhouse, CSIRO Division of Plant Industry, Canberra.

- my extended family for continued interest and support.

I wish to pay special tribute to my parents for their support, commitment and understanding, and for providing a first-rate environment during the large proportion of these studies.

Finally, I thank Tina for her love and dedication, which I hope to repay. No thanks to Rachel who impeded the process.

CHAPTER ONE

REVIEW OF THE LITERATURE

1.1 Introduction

Barley yellow dwarf disease is caused by a suite of plant luteoviruses, all of which share the name barley yellow dwarf virus (BYDV) albeit with specific strain designations. As members of the luteovirus group, the BYDVs are phloem-limited, spherical viruses containing a single molecule of positive-sense RNA. The virus particles can be transmitted only by aphid vectors. BYDV is limited to the phloem tissue of the host plant with the consequence that the titre of the virus is very low, making purification of virus particles difficult. Each BYDV strain causes similar yellowing symptoms in susceptible plants; the name luteovirus (of which BYDV-MAV is the type member) is derived from the Latin *luteus*, which means yellow. BYDV has a wide host range, infecting most members of the graminae including the major monocotyledonous crop plants. Because the biological characteristics (symptomatology, aphid transmission, host range, phloem limitation) of each BYDV strain are similar, they will be referred to in the singular throughout this thesis except where specific strains are discussed.

Barley yellow dwarf was first reported as a disease of viral aetiology and aphid transmission by Oswald and Houston in California in 1951. Prior to this, symptoms similar to those caused by BYDV had been recorded, and in some cases an association with aphid transmission had been made (Burnett, 1984, and references therein). Subsequent to its discovery, BYDV isolates differing in specificity of aphid transmission and serological relationships were described. The application of techniques including cytopathology, double-stranded RNA analysis, and latterly molecular biology, has led to the recognition that barley yellow dwarf disease is caused by a number of viruses, all of which are luteoviruses but otherwise fall into two subgroups. The molecular differences between these two subgroups are substantial despite their similar biology and are reflected in the other members of the luteovirus group.

1.2 Biology of BYDV

1.2.1 Symptoms of BYDV infection

Symptoms caused by infection with BYDV are highly variable and depend on the following factors; the strain of the virus, the virulence of the strain and the dosage of infection, the plant species that is infected, the age of the plant, and environmental conditions, especially low temperatures and high light intensity. The age of the plant at infection is critical for all species in the development of symptoms. Crop plants infected at the seedling stage are most severely affected and become badly stunted, with partial or complete sterility of the heads/florets. In addition, the root systems of infected plants are poorly developed. The colour changes to the leaf associated with BYDV infection are complex but may include the following: In wheat and barley, leaves exhibit diffuse or blotchy yellowing near leaf tips, which then extends down the leaf leaving a strip of green along the side of the midrib. In oats more dramatic discolouration can be seen, leaves turning yellowish brown or pale orange, and becoming a striking red or purple in cool conditions. An additional symptom seen in oats is stiffening of the leaves, which are shorter and more erect than in non-infected plants. Rye and triticale show few symptoms. All crop species can be infected without showing symptoms, especially if infected at the post-seedling stage, however a yield loss is still associated with this condition. The flag leaves of plants infected when mature may turn yellow or red, especially in wheat and barley. Overall, however, the most common symptom in infected plants is stunting (Mathre, 1982; Burnett, 1984; Wiese, 1987; Paliwal and Comeau, 1987).

1.2.2 Location of luteoviruses in the plant

Luteoviruses appear to be confined to the phloem (carbohydrate conducting) tissue of the plant. Virus particles have been visualised in the phloem parenchyma, companion cells and sieve tube elements (Esau and Hoefert, 1972, Gill and Chong, 1975, 1976, 1979a, 1979b), and also in seeds (Eweida *et al.*, 1988), although seed transmission does not occur. Infection causes degeneration and blockage of sieve tube

elements which may be responsible for the typical luteovirus symptoms (stunting, discolouration and leaf rolling; Esau, 1957; Jensen, 1969). Plant tissues other than phloem can be infected under laboratory conditions, including mesophyll, epidermis, xylem (Gill and Chong, 1981), and also undifferentiated cells (Miller, 1994). Additionally, Barker (1987) reported that potato leafroll luteovirus (PLRV) could invade non-phloem tissue in *Nicotiana clevelandii* when the plants were secondarily infected with potato Y potyvirus (PYV). Taken together, these results suggest that luteoviruses are not confined to the phloem because they are deficient in some factor necessary for replication in non-phloem tissues, but lack the ability to escape from phloem tissue (where they are deposited by the aphid). Presumably PYV is able to complement this missing function.

1.2.3 Host range and transmission of BYDV

All major monocotyledonous crop plants, including wheat, barley, oats, rye, and maize are hosts of BYDV. Rice is infected by a close relative of BYDV strain RPV (1.3.3) called rice giallume virus (RGV; Osler, 1984, cited by Plumb, 1990). Nearly all members of the graminae appear to be hosts including the common pasture grasses *Festuca* spp. and *Lolium* spp., and serve as a reservoir of infection for crop species (Conti *et al.*, 1990, Henry *et al.*, 1992).

BYDV can only be transmitted by aphids and is not mechanically or seed transmissible. The aphid ingests the virus when feeding on the sieve tube elements of infected phloem tissue (Fig. 1.1). The virus passes from the insect's foregut to the hindgut, where it is transported in membrane-bound vesicles to the haemocoel, or body cavity (Gildow, 1985). The hindgut is the point of acquisition of the virus; if it is not absorbed at this point, then it passes through the aphid and is excreted in the honeydew (Gildow, 1990, 1993). Acquired virus becomes suspended in the haemolymph (equivalent to blood) and circulates throughout the insect. The virus particles are infectious at this stage and as such the aphid becomes a reservoir for the virus. Transmission of the virus is dependent on virus particles crossing firstly from the

haemolymph into the accessory salivary gland (Gildow and Rochow, 1980), and then from the gland cells into the salivary duct (Gildow, 1990). Healthy plants are infected when the virus is expelled with salival secretions as the aphid feeds . The virus does not replicate in the aphid; however the aphid retains the ability to infect plants at least 2-3 weeks after acquisition and even after moulting. This type of transmission is called circulative, non-propagative and persistent.

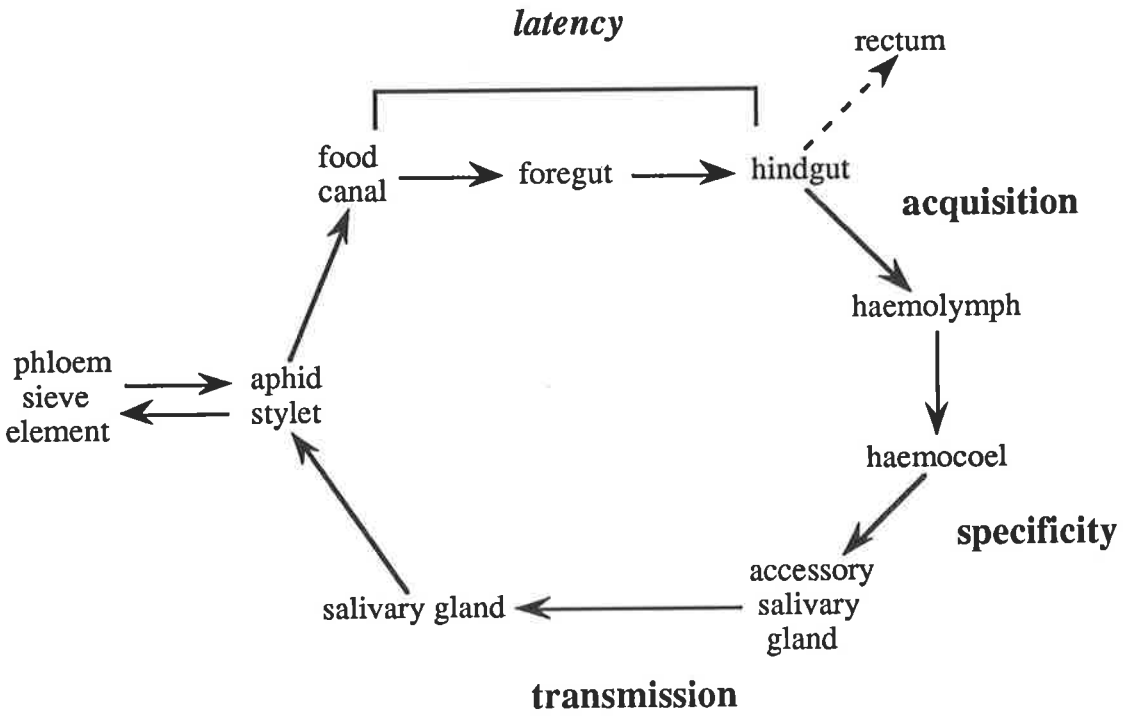


Figure 1.1. Passage of BYDV through the aphid vector (redrawn from Gildow, 1990). Arrows show the path of virus movement from the plant, through the aphid, and back to the plant. Important points in the passage of the virus through the aphid (latency as part of the gut contents, acquisition by transfer to the haemolymph, specific transport to the accessory salivary gland, and transmission to the plant) are shown. Refer to text for details.

1.2.4 Significance and epidemiology of BYDV

BYDV has been reported from nearly all countries where cereals are grown and is regarded as the most important viral pathogen of cereal crops worldwide (Burnett, 1984). Virus spread is entirely dependent on movement of viruliferous aphids (Mathre, 1982; Paliwal and Comeau, 1987). Therefore the probability of an epiphytotic developing depends upon many factors, but especially on the ability of the aphid to transmit the viral isolate present, the numbers of aphids, and the proportion of aphids which carry the virus (Halbert *et al.*, 1992). The most severe infections of crops occur when viruliferous aphids colonise a crop early in the growth season. The aphids may be local, regional, or migrate long distances; evidence suggests that crop infections do not necessarily correspond to the viral isolates predominant in the local area, which implies that long distance migration is an important factor (Irwin and Thresh, 1990). Severe outbreaks of BYDV are then dependent on sufficiently cool and damp conditions to allow the build-up of aphid numbers, leading to secondary infection of neighbouring crops (Wiese, 1987). If such conditions do not occur then the infection will be limited to the original 'foci' (equivalent to where a viruliferous aphid has landed in the crop and started a colony) and crop losses will be minimal (Plumb, 1990).

Yield losses from BYDV thus vary considerably from year to year and also in the country (and hence climate) in which the crop is grown. For example, in countries with severe summers (such as Australia) or winters (such as Canada) where no crops are planted between the major growing seasons, the reservoir of viruliferous aphids can be considerably diminished. This means that a severe outbreak of BYDV is less likely. Generally speaking, losses attributable to BYDV occur at a low level (1-3%) between exceptional years where conditions for aphid growth are ideal. Losses in such years can be a considerable proportion of total production; in some cases entire crops can be lost (Conti *et al.*, 1990).

1.3 Taxonomy of BYDV and the luteoviruses

1.3.1 Members of the luteovirus group

A list of luteovirus members, their synonyms, abbreviations and genomic RNA sequence information is given in Table 1.1. These data will be referred to throughout this thesis.

1.3.2 Particle morphology and composition in the luteovirus group

BYDV belongs to the luteovirus group, of which BYDV-MAV is the type member. The main characteristics of luteovirus particles are as follows (Waterhouse *et al.*, 1988): Isometric particles of 25-30 nM in diameter which sediment at ~104-127 *S* with a buoyant density in CsCl of 1.38-1.40 g/cm^3. The virus particles are composed of 180 subunits of coat protein monomer (M_r 2.2-2.4 x $10^{3)}$ in a T=3 arrangement (Miller, 1994). The viral particle contains a single molecule of RNA which is positive-sense and single-stranded (M_r ~2.0 x 10^6, about 6 kilobases (kb)). A viral protein linked to the 5' end of the genomic RNA (VPg) has been reported for BYDV-RPV and PLRV but not for other members of the group (Mayo *et al.*, 1982; Murphy *et al.*, 1989). Luteovirus particles have A_{260}/A_{280} ratios of 1.6-1.9. The virus particles are moderately stable with a thermal inactivation point of 45-75°C, and are strongly immunogenic in rabbits. Most members of the luteovirus group are serologically interrelated, with some clustering of viruses (Waterhouse *et al.*, 1988). A satellite RNA is sometimes associated with the RPV isolate of BYDV (Miller *et al.*, 1991).

1.3.3 Separation of BYDV into five strains

BYDV was first divided into strains when Rochow (1969) and Johnson and Rochow (1972) defined five isolates of BYDV based on their specific transmission by different aphid species (Table 1.2; also see Gill, 1967). The specificity of transmission appears to act at the point where virus particles are transported from the haemolymph to the salivary gland of the aphid (Fig. 1.1; Gildow and Rochow, 1980). Transmission

Table 1.1 General data for luteovirus group members

Name (Synonym)	Abbreviation	Sequenced strain	Sequence length (nt)
barley yellow dwarf	**BYDV**		
BYDV-MAV		BYDV-MAV-PS1	5273*[1]
BYDV-PAV		BYDV-PAV-Vic	5677[2]
		BYDV-PAV-P	5179*[1]
BYDV-RMV		BYDV-RMV-(IL, MN, NY)	partial[3]
BYDV-RPV		BYDV-RPV-NY	5600*[4]
BYDV-SGV		BYDV-SGV-NY	partial[5]
bean leafroll	**BLRV**	BLRV-(German isolate)	partial[6]
legume yellows			
Michigan alfalfa			
pea leafroll	PeLRV		
beet western yellows	**BWYV**	BWYV-(FL1, GB1)	5641[7]
beet mild yellowing			
malva yellows			
turnip mild yellows			
carrot red leaf	**CaRLV**		ND†
cucurbit aphid-borne yellows	**CABYV**	CABYV-N	5669[8]
groundnut rosette assistor	**GRAV**		ND
Indonesian soybean dwarf	**ISDV**		ND
potato leafroll	**PLRV**	PLRV-A	5882[9]
		PLRV-C	5883[9]
		PLRV-N	5882[10]
		PLRV-S	5987‡[11]
solanum yellows	**SYV**		ND
tomato yellow top	**ToYTV**		ND
soybean dwarf	**SDV**	SDV-Y	partial[12]
subterranean clover	SCLRV		
redleaf virus			
strawberry mild yellow edge			
tobacco necrotic dwarf	**TNDV**		ND

*Sequence probably incomplete; †Not Determined; ‡Contains a non-viral 5' extension (Mayo and Jolly, 1991); [1]Ueng *et al.*, 1992; [2]Miller *et al.*, 1988a; [3]Domier *et al.*, 1994; [4]Vincent *et al.*, 1991; [5]GenBank accession no. UO6865; [6]Prill *et al.*, 1990; [7]Veidt *et al.*, 1988; [8]Guilley *et al.*, 1994; [9]Keese *et al.*, 1990; [10]van der Wilk *et al.*, 1989; [11]Mayo *et al.*, 1989; [12]Smith *et al.*, 1993.

specificity is not absolute, and at least 23 aphid species able to transmit at least one strain of BYDV are known (Plumb, 1990). These include the important cereal pathogens *Metopolophium dirhodum*, which is able to transmit BYDV-MAV and -PAV, and *Rhopalosiphum rufiabdominalis*, which can transmit BYDV-PAV, -RMV and -RPV (Gildow, 1990). Serological studies have affirmed the division of BYDV into five strains (1.3.4.2).

Table 1.2. Definition of BYDV strains by the specificity of aphid transmission[a,b]

BYDV strain	transmission	aphid species
-RPV	specific	*Rhopalosiphum padi* (Linnaeus)
-RMV	specific	*R. maidis* (Fitch)
-MAV	specific	*Sitobion (=Macrosiphum) avenae* (Fabricus)
-SGV	specific	*Schizaphis graminum* (Rodani)
-PAV	non-specific	*R. padi* and *S. avenae*

[a]Rochow (1969); [b]Johnson and Rochow (1972).

1.3.4 Separation of BYDV into two subgroups

Other than the classification of BYDV into strains of different aphid transmission specificity, the strains can be clustered into two subgroups. This section briefly reviews the serological, biological, cytopathological and double-stranded RNA data that led to the formation of the subgroups. Ultimately, the most convincing data for this division is based on the genome organisation of the strains, presented in section 1.4.1.

1.3.4.1 Cross-protection studies

The relatedness of two viruses can be determined by the reaction of the host plant to simultaneous infection with both viruses. Closely related viruses will show 'cross-protection' in the host, that is, the plant will be less severely affected than in single infections with either of the two viruses. Conversely, viruses that are not related may induce symptoms equal to or greater than that either virus gives in a single infection ('synergism'). This phenomenon has been used to establish the relationship between BYDV strains of different aphid transmission specificity. Smith (1963) showed that BYDV isolates probably corresponding to BYDV-MAV and BYDV-PAV conferred cross-protection in oats, thus establishing a close relationship between these two strains. These results were confirmed by later workers (Jedlinski and Brown, 1965; Aapola and Rochow, 1971; Halstead and Gill, 1971). However, synergism existed between BYDV-RPV and BYDV-MAV, and also between BYDV-RPV and BYDV-PAV, thus demonstrating the distance between these viruses (Aapola and Rochow, 1971; Halstead and Gill, 1971). Finally, Gill and Comeau (1977) showed that BYDV-PAV and BYDV-RMV interacted synergistically and are therefore unrelated.

1.3.4.2 Serology

Antisera raised against BYDV-MAV recognised both BYDV-MAV and BYDV-PAV particles (Aapola and Rochow, 1971; Lister and Rochow, 1979). However, BYDV-RPV particles did not react with BYDV-MAV antiserum (Lister and Rochow, 1979). These studies were extended by Rochow and Carmichael (1979) who raised further antisera, against the PAV isolate of BYDV and also against BYDV-RMV. In heterologous reactions conducted between the five isolates of BYDV (BYDV-MAV, -PAV, -RPV, -RMV and -SGV; Rochow, 1969; Johnson and Rochow, 1972), the following relationships were discovered: The RPV and RMV strains shared common antigens, but were not related to BYDV-MAV or BYDV-PAV. Similarly, the antiserum raised against BYDV-MAV reacted with BYDV-PAV particles; the opposite reaction (BYDV-PAV antiserum, BYDV-MAV particles) also occurred. Finally, the

BYDV-PAV antiserum showed a faint reaction with BYDV-SGV. Subsequent immunological tests by Rochow (1979) found that all of 181 field isolates of BYDV reacted homologously in ELISAs with one of the five antisera raised against the type strains of BYDV.

1.3.4.3 Cytopathology

Gill and Chong (1975, 1976, 1979a, 1979b) examined oat tissue for cytopathological effects induced by infection with different strains of BYDV. Their major findings were as follows (summarised in Table 1.3): For BYDV-MAV, -PAV and -SGV strains, distortions in the cytoplasm included the appearance of single-membraned vesicles containing densely staining fibrils. Densely staining filaments accumulated in the cytoplasm and nucleus. Changes to the nucleus included distortion of the nuclear outline, the aggregation and accumulation of densely staining material, and clumping of the heterochromatin late in the infection cycle. Progeny virus particles were first observed in the cytoplasm, amongst the filaments. For BYDV-RPV and -RMV, fibril containing vesicles were also observed in the cytoplasm but were bound by double membranes. No filament clusters accumulated in the cytoplasm as for the BYDV-MAV type strains. A further unique feature was the formation of tubular aggregates in the cytoplasm. The nucleus remained normal in outline, with progressive disintegration of the heterochromatin. Progeny virus particles were first observed in the nucleus (in the nucleolus) and ultimately in the cytoplasm.

The five strains thus fell naturally into two groups based on changes to the nucleus, the site in the cell where progeny virus particles first appeared, and the type of fibril-containing membranous structures in the cytoplasm (Gill and Chong, 1979b). Of further interest was the similarity in cytopathological symptoms between BYDV-RPV and -RMV and beet western yellows luteovirus (BWYV; Esau and Hoefert, 1972), a relationship reflected by a positive serological reaction between BYDV-RPV and BWYV. Gill and Chong (1979b) proposed the division of BYDV into two subgroups (Table 1.4), subgroup I containing BYDV-MAV, -PAV and -SGV, and subgroup II

Table 1.3 Summary of cytopathological features for the five strains of barley yellow dwarf virus[a]

cytopathological inclusions and alterations	Subgroup I			Subgroup II	
	MAV	PAV	SGV	RPV	RMV
alterations in nucleus	distortion of nuclear outline, then aggregation and accumulation of persistent, densely staining material			nucleus +/- normal in outline; heterochromatin slowly disintegrates	
first occurrence of virus progeny	in cytoplasm			around nucleus at intermediate[1] stage	
	early[1] or late[1] phase	late phase	early or intermediate		
membranous inclusions in cytoplasm	single-membraned vesicles containing fibrils			double-membraned vesicles, containing fibrils	
proliferated tubules	absent		present		
filaments	in cytoplasm and nucleus			only in cytoplasmic voids	late phase in cytoplasm

[a]Redrawn and simplified from Gill and Chong (1979b); [1]phases are defined as follows; early - alteration first in the cytoplasm; middle - then alteration in the nucleus; late - subsequent alterations in nucleus and cytoplasm.

containing BYDV-RPV and -RMV, on the basis of their cytopathological data, and also considering the serological and cross protection studies done by others (1.3.4.1; 1.3.4.2).

Table 1.4. Subgroups of BYDV defined by serological, biological and cytopathological analyses[a]

Subgroup I	Subgroup II
BYDV-MAV	BYDV-RPV
BYDV-PAV	BYDV-RMV
BYDV-SGV	

[a]Gill and Chong (1979b)

1.3.4.4 Double-stranded RNA analysis

Double-stranded viral RNAs (dsRNA) are extracted from whole plant tissue, rather than viral particles, and as such represent all RNAs involved in replication and expression of the viral genome. This is as opposed to RNAs that are packaged in the virions, which are usually only those required for infection (genomic RNA(s)). RNAs shorter than the genomic RNA usually represent subgenomic (or messenger RNAs), therefore differences in the dsRNA profiles of two RNA viruses represent differences in the organisation and expression of the respective viral genomes. This in turn represents dissimilarity in the relationship between the viruses.

Gildow *et al.* (1983) examined dsRNA profiles of all five strains of BYDV to establish points of similarity or difference between them. BYDV-PAV, -MAV and -SGV possessed five bands of dsRNA after electrophoresis on polyacrylamide gels, of

M_r 3.6x10^6 (genomic RNA), 2.0x10^6, 1.2x10^6, 0.55x10^6, and 0.50x10^6. In contrast, BYDV-RPV and -RMV possessed only four bands, of M_r 3.8x10^6 (genomic RNA), 1.6x10^6, 1.2x10^6, and 0.55x10^6. Therefore, the strains of BYDV group I associate on the basis of number and size of dsRNAs, and are distinct from the strains of subgroup II, which are similar to each other in these respects.

1.4 Molecular biology of the luteoviruses and BYDV

The genomic RNAs of many luteoviruses have been sequenced (Table 1.1). From these data, the luteoviruses can be divided into two subgroups which correspond to subgroup I and II of the barley yellow dwarf viruses (Table 1.4; Veidt *et al.*, 1988; Miller, 1994). Furthermore, inferences about some of the luteoviral strategies for gene expression have been made from the sequence data, and in some cases proven by further work. This section provides a review of the current knowledge of the molecular biology of the luteoviruses, with the assumption that knowledge in one viral system will be applicable to all members of that luteoviral subgroup, and is thus directly relevant to studies of the barley yellow dwarf viruses.

1.4.1 Description of the luteovirus genomes

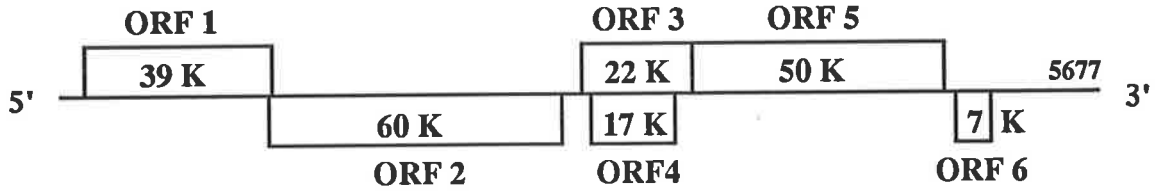
Comparisons in this section arbitrarily refer to the genomes of BYDV-PAV-Vic (Miller *et al.*, 1988a) as representative of subgroup I luteoviruses, and PLRV-N (van der Wilk *et al.*, 1989) as representative of subgroup II luteoviruses. The genomes of both subgroups are represented in Fig. 1.2. This thesis follows the ORF naming convention of Martin *et al.* (1990) and Miller *et al.* (1994); thus the six ORFs of subgroup I are numbered 1-6 while those of subgroup II are numbered 0-5.

The full-length sequences of luteovirus genomes are generally between 5600 and 5900 nucleotides (nt; Table 1.1). Although the genomes of both subgroups specify six ORFs, the genome types are differently organised. There are two major blocks of coding sequence in both subgroups, however subgroup II has an extra ORF in the 5' coding block (ORF 0) which is absent in subgroup I, whereas subgroup I has a unique

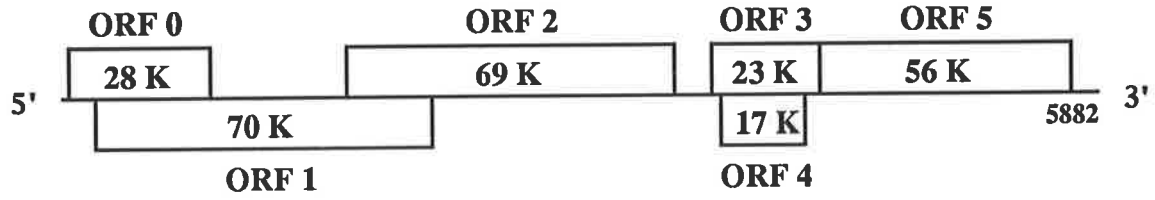
Fig. 1.2. Genome organisation of luteoviruses and related viruses. (A) Genome organisation of representatives of the two luteoviral subgroups. The virus genomes depicted are those of BYDV-PAV-Vic (subgroup I; Miller *et al.*, 1988a) and PLRV-N (subgroup II; van der Wilk *et al.*, 1989). Open boxes represent open reading frames (ORFs). The numbers within the boxes refer to the relative molecular mass of the putative protein deduced from the nucleotide sequence of the ORF. Approximate sizes of the genomes in kilobases (kb) are shown. The ORF numbering scheme follows that of Martin *et al.* (1990); ORF 2 of each genome encodes the **GDD** motif associated with RNA-dependent RNA polymerase activity, except in (B) where it occurs in ORF 3. (B) Genome organisation of the RNA associated with BWYV strain ST9 (Chin *et al.*, 1993). (C) Genome organisation of the two genomic RNAs of PEMV (Demler and de Zoeten, 1991; Demler *et al.*, 1993). Luteovirus abbreviations are given in Table 1.1.

A

Subgroup I

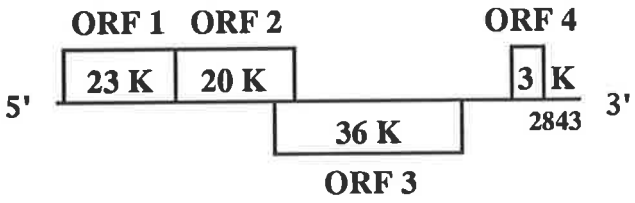


Subgroup II



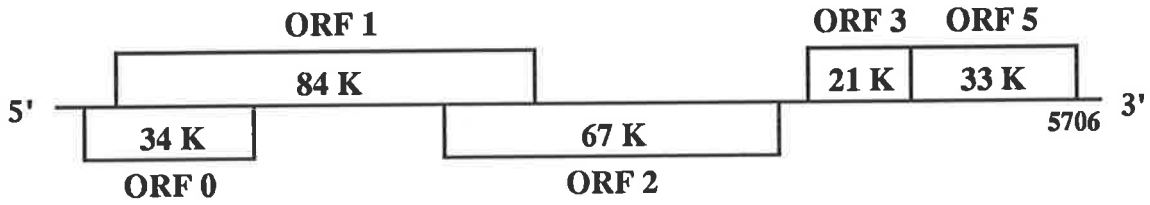
B

BWYV-ST9 associated RNA



C

PEMV RNA 1



PEMV RNA 2

ORF (ORF 6) downstream of the 3' coding block. The two subgroups have the same organisation and approximate sizes of ORFs 3, 4 and 5; these ORFs also show high levels of sequence homology. Thus the major differences between the subgroups is in the organisation and size of the 5' coding block.

Subgroup I has two genes in the 5' coding region. ORF 1 follows a 5' untranslated region (UTR) of 141 nt and specifies a protein of M_r 39 K (all sizes given are for BYDV-PAV-Vic; Miller *et al.*, 1988a). ORF 1 overlaps ORF 2 by 13 nt. The complete coding sequence of ORF 2, which contains the Gly-Asp-Asp (**GDD**) amino acid motif specific to RNA-dependent RNA polymerases (1.4.2.4), specifies a protein of M_r 60 K although the first initiation (AUG) codon occurs significantly into the reading frame. On the basis of this gene organisation and considering other evidence reviewed below, Miller *et al.* (1988a) hypothesised that the ORF 2 product would be expressed as a frameshift fusion protein with the product of ORF 1; strong evidence now exists for such a translational mechanism (1.4.4.2; Brault *et al.*, 1992; Di *et al.*, 1993) A non-coding or intergenic sequence of 116 nt separates ORF 2 from the second coding block, which commences with ORF 3.

In contrast to subgroup I, the subgroup II genome contains three overlapping genes at the 5' end. The first of these, ORF 0, specifies a protein of M_r 28 K (all sizes given are for PLRV-N; van der Wilk *et al.*, 1989), and overlaps ORF 1 by 608 nt. The putative protein product of ORF 1 is M_r 70 K. ORF 1 overlaps the 5' end of ORF 2 by 580 nt. ORF 2 also specifies a protein product of M_r 70 K, which like BYDV-PAV contains the **GDD** RNA-dependent RNA polymerase motif (1.4.2.4). van der Wilk *et al.* (1989) proposed that ORF 2 would be expressed as a frameshift fusion with the product of ORF 1 on the basis of similarity of organisation of ORFs 1 and 2 with those of BYDV-PAV. Evidence now suggests that this is indeed the case (1.4.4.2; Prüfer *et al.*, 1992; Garcia *et al.*, 1993; Kujawa *et al.*, 1993). The 5' coding block (ORFs 0, 1 and 2) is followed by an intergenic region of 197 nt, after which the second coding block starts with ORF 3 (as in subgroup I).

The organisation of ORFs 3, 4 and 5 is similar between the subgroups. ORF 3, which encodes the viral coat protein (1.4.2.5), specifies a protein of M_r ~22 K. Nested within ORF 3 (but in a different reading frame) is ORF 4, which encodes a protein of M_r 17-21 K. ORF 5 is contiguous and in-frame with ORF 3, and is separated from ORF 3 by an amber (UAG) stop codon. It has coding potential for a protein of M_r ~50 K. The location of ORF 5 relative to ORF 3 led many workers to propose that ORF 5 may be expressed as a readthrough or fusion protein with the product of ORF 3 (Miller *et al.*, 1988a; Veidt *et al.*, 1988; van der Wilk *et al.*, 1989). Evidence now exists to support this hypothesis (1.4.4.6).

ORF 5 is the last coding sequence in subgroup II luteoviruses, and is followed by a non-coding region of 143 nt in PLRV-N. Subgroup I luteoviruses have a second intergenic region of 106 nt after which ORF 6 initiates, specifying a protein (in BYDV-PAV-Vic) of M_r 6.7 K. ORF 6 is followed by 3' non-coding region of 564 nt.

1.4.2 Functions encoded by luteoviral ORFs

1.4.2.1 ORF 0

ORF 0 is only present in subgroup II luteoviruses (1.4.1) and has the most poorly conserved nucleotide sequence of any luteoviral ORF (Guilley *et al.*, 1994). The amino acid sequences derived from the coding sequence of ORFs 1 of the various subgroup II luteoviruses are hydrophobic, which suggests that they may be associated with membranes (Mayo *et al.*, 1989). Veidt *et al.* (1992) deleted ORF 0 from a cloned cDNA of BWYV but the mutated virus was still able to replicate in plant protoplasts. The authors suggested that ORF 0 may have a role in determination of host range, based on the poor amino acid conservation and dispensability of the ORF.

1.4.2.2 ORF 1: (a) Helicase

RNA helicases are thought to responsible for unwinding duplex RNA during replication and transcription. Habili and Symons (1989) found motifs conserved in the putative RNA helicases of the predicted amino acid sequences of plant viruses to be

present in ORFs 1 and 2 of both luteoviral subgroups. However, Gorbalenya and Koonin (1989) were unable to find helicase motifs in any member of the luteovirus group using curtailed motif formulae in a general database search. Similarly, Martin *et al.* (1990) noted that the putative nucleotide-binding motif **GXGK(T/S)** that is essential for helicase activity (Gorbalenya and Koonin, 1993) was 'absent or poorly conserved' in the sequences of BYDV-PAV, BWYV and PLRV. While this discrepancy has not been resolved, there is evidence that ORF 1 of subgroup II encodes a protease (see below; Demler and de Zoeten, 1991; Koonin and Dolja, 1993; Miller *et al.*, 1994), which would potentially exclude the possibility that the ORF 1 product also has helicase activity.

1.4.2.3 ORF 1: (b) Protease

Several workers have found an amino acid motif diagnostic of picornavirus-like proteases in the deduced amino acid sequence of ORF I of subgroup II luteoviruses, and also in the homologous ORF of the related RNA 1 of pea enation mosaic virus (PEMV; Demler and de Zoeten, 1991; Koonin and Dolja, 1993; Miller *et al.*, 1994). The motif, **H**($X_{\sim 25}$)**[D/E]**($X_{70\text{-}80}$)**T[R/K]XGXSG**, is fully conserved in these luteoviruses except for the basic (**[R/K]**) amino acid (Miller *et al.*, 1994). While this constitutes strong evidence for the existence of a protease in subgroup II, so far no direct evidence for proteolytic cleavage of subgroup II proteins by a virally encoded protease has been published. The product of ORF 5 does appear to be cleaved, however this event occurs in viruses of both subgroups so is not specific to subgroup II (1.4.4.6; Bahner *et al.*, 1990; Filichkin *et al.*, 1994). The conserved motif does not appear in the amino acid sequence of any subgroup I ORF (Miller *et al.*, 1994).

1.4.2.4 ORF 2: RNA-dependent RNA polymerase

The RNA-dependent RNA polymerase gene of both subgroups is thought to be encoded by ORF 2. This is because of the presence of the diagnostic amino acid motif **GXXXTXXXN**($X_{25\text{-}40}$)**GDD** which is located approximately three-quarters of the way

from the N terminus of the putative protein (Miller *et al.*, 1994). Although this motif is conserved in almost all known RNA-dependent RNA polymerases (Kamer and Argos, 1984), actual polymerase activity has not yet been demonstrated for any member of the luteoviruses. Several workers have shown that ORF 2 is expressed as a fusion with the product of ORF 1 in both subgroups (1.4.4.2; Brault and Miller, 1992; Prüfer *et al.*, 1992; Di *et al.*, 1993; Garcia *et al.*, 1993; Kujawa *et al.*, 1993), which implies that the product of ORF 1 also has a role in RNA-dependent RNA polymerase activity. However, there is no evidence thus far that the ORFs 1 of each subgroup have any structure or function in common.

Despite the unity of putative function of the luteoviral ORFs 2, the amino acid sequences in each subgroup show diverse evolutionary origins (Miller *et al.*, 1988a; Veidt *et al.*, 1988). Thus the ORF 2 of subgroup I is more closely related to those of members of the carmovirus group (type member carnation mottle virus (CarMV)) than it is to that of subgroup II. In turn, ORF 2 of luteovirus subgroup II is most closely related to the RNA-dependent RNA polymerase ORF of members of the sobemovirus group (type member southern bean mosaic virus (SBMV)). This dissimilarity between subgroups I and II is a major factor (other than genome organisation) for the taxonomic division of the luteoviruses (Habili and Symons, 1989).

1.4.2.5 ORF 3: Coat protein

The coat protein is encoded by ORF 3. This has been demonstrated for a number of luteoviruses, usually by recognition of expressed recombinant protein by antisera raised against viral particles (Miller *et al.*, 1988b; Veidt *et al.*, 1988; Kawchuk *et al.*, 1989; Smith and Harris, 1990; Vincent *et al.*, 1991; Smith *et al.*, 1993), but also by comparison of deduced and actual coat protein amino acid sequences (Miller *et al.*, 1988b). The coat protein gene is not necessary for the replication of BWYV RNA in plant protoplasts (Reutenauer *et al.*, 1993).

1.4.2.6 ORF 4: (a) Genome linked viral protein

A genome linked viral protein (or VPg) has been reported for two subgroup II luteoviruses. Mayo *et al.* (1982) reported a protein of M_r 7 K linked to the 5' genomic terminus of PLRV-S, while Murphy *et al.* (1989) found that a M_r 17 K protein was linked to the 5' terminus of BYDV-RPV. When the first luteovirus sequence was published (BYDV-PAV; Miller *et al.*, 1988a), the authors proposed that ORF 4 encoded the VPg based on the similarity between the size of the VPg reported for BYDV-RPV (M_r 17 K) and the coding capacity of ORF 4 (also M_r 17 K). van der Wilk *et al.* (1989) proposed the same role for ORF 4 of PLRV-N although the M_r 17 K protein would require processing to reach the required weight of M_r 7 K.

Miller *et al.* (1994) have questioned whether subgroup I luteoviruses possess a VPg. Their argument is based on the premises that (1) the VPg is involved in viral RNA replication (as for poliovirus; Kuhn and Wimmer, 1987), (2) the VPg must therefore interact specifically with the viral replicase, and (3) the polymerase gene (ORF 2) of subgroup I luteoviruses is similar to that of members of the carmovirus group, members of which are known not to possess a VPg. This is in contrast to subgroup II luteoviruses where the polymerase gene is similar to that of the sobemoviruses. The type member of this group, southern bean mosaic virus (SBMV) does possess a VPg (Mang *et al.*, 1982).

If ORF 0 does in fact encode a membrane-bound protein (Mayo *et al.*, 1989), and the protease designation for subgroup II ORF 1 is correct (Demler and de Zoeten, 1991; Miller *et al.*, 1994), then the following argument can be made (Demler and de Zoeten, 1991): The subgroup II arrangement of membrane anchor-protease-polymerase closely parallels the picornavirus--like virus arrangement of membrane anchor-VPg-protease-polymerase (Domier *et al.*, 1987). Taken together with the large size of subgroup II ORF 1 (M_r ~70 K) versus that of subgroup I (M_r ~39 K), it follows that ORF 1 of subgroup II may encode the VPg (Miller *et al.*, 1994), which would be released from the ORF 1 product by proteolytic cleavage. This argument is supported by the evolutionary theory of Koonin and Dolja (1993), which states that conservation

of distinct arrays of genes is one of the most important rules governing evolution of positive-strand RNA viruses. Note that this proposition significantly decreases the likelihood of the existence of a VPg in subgroup I, also as argued by Miller *et al.* (1994).

In support of the argument that ORF 4 does not encode the VPg, Reutenauer *et al.* (1993) found that ORF 4 was not necessary for infection of BWYV in plant protoplasts, thus precluding an essential role for this gene in replication. Moreover, RNA 1 of PEMV (which is highly related to the genomic RNA of subgroup II luteoviruses) lacks a homologue to ORF 4, yet is linked to a VPg (Reismen and de Zoeten, 1982; Demler and de Zoeten, 1991). Finally, Tacke *et al.* (1993) have proposed that ORF 4 of PLRV encodes a cell-cell movement protein (see below).

1.4.2.7 ORF 4 (b) Cell-cell movement protein

The cell-cell movement protein of the luteoviruses has not been positively identified. However, Tacke *et al.* (1993) have proposed that ORF 4 encodes this function based on biochemical properties of the recombinant protein. Firstly, recombinant PLRV ORF 4 protein bound non-specifically to single-stranded nucleic acids, a property which was conditioned by basic sequences at the C-terminus of the protein (Tacke *et al.*, 1991). A second domain located in the N-terminal portion of the protein directed formation of homodimers of the protein (Tacke *et al.*, 1993). The authors postulated that these features of the ORF 4 protein would allow it to package viral nucleic acids into ribonucleoprotein complexes. Such a structure is consistent with current models of plant virus RNA movement across cellular membranes (Citovsky and Zambryski, 1991; Fujiwara *et al.*, 1993). Furthermore, the ORF 4 protein was phosphorylated *in planta*, as is the movement protein of tobacco mosaic virus (TMV; Citovsky *et al.*, 1993). However, subcellular localisation of the protein by differential centrifugation found that the majority of the ORF 4 product was not localised to the cellular or membranous fractions (Tacke *et al.*, 1993), as might be expected for a cell-cell movement protein (Citovsky and Zambryski, 1991).

1.4.2.8 ORF 5: Aphid transmission

It has been widely speculated that the product of ORF 5 is involved in aphid transmission (*e.g.* Bahner *et al.*, 1990; Martin *et al.*, 1990). This is because the ORF is expressed as a readthrough fusion with the coat protein (1.4.4.6), and thus would be an external component of the virus particle, where it could interact with cellular receptors in the aphid's hindgut and accessory salivary gland (1.2.3; Gildow and Rochow, 1980; Gildow, 1985). A large molecular weight protein corresponding to the fusion protein has been detected as a component of most luteovirus particles (Martin *et al.*, 1990). However, no direct evidence to support the role of ORF 5 in aphid transmission has been published. Young *et al.* (1991) found that ORF 5 was necessary for replication of BYDV-PAV in protoplasts, but this has since been refuted by Dinesh-Kumar (1993, cited by Miller *et al.*, 1994). Similarly, Reutenauer *et al.* (1993) showed that ORF 5 was unnecessary for the replication of BWYV in protoplasts.

1.4.2.9 ORF 6

ORF 6 is present in all subgroup I luteoviruses sequenced to date (Miller *et al.*, 1988a; Ueng *et al.*, 1992; Kelly *et al.*, 1994; Chalhoub *et al.*, 1994). No function has yet been ascribed to the putative product of this ORF. ORF 6 varies considerably in size in different isolates and strains of subgroup I, potentially encoding a protein of between M_r 4.3 K and 6.7 K. The amino acid sequence derived from ORF 6 is conserved in the amino-terminal ~20 residues, but thereafter the sequence is the most variable in the entire subgroup I genome (Miller *et al.*, 1988a; Chalhoub *et al.*, 1994). Truncation of the ORF 6 sequence such that the coding region was reduced to M_r ~3.3 K abolished the replication of an infectious BYDV-PAV clone in oat protoplasts (Young *et al.*, 1991), despite the natural variation in length and poor conservation of sequence in the C-terminal region of the putative protein (Chalhoub *et al.*, 1994). Further evidence for the expression of ORF 6 includes (1) the existence of a subgenomic mRNA that should allow translation of ORF 6 by positioning of the ORF close to the 5' end of the message (Kelly *et al.*, 1994) and (2) the pattern of

nucleotide variability in the 5' part of the ORF; most changes occur in the third position of the codon, thus minimising changes to the putative amino acid sequence of the protein (Chalhoub *et al.*, 1994).

1.4.3 Genomes related to the luteoviruses

1.4.3.1 BWYV-ST9 associated RNA

Virion preparations of BWYV strain ST9 contain two major RNA species; the larger (~6 kb; Table 1.1) is the genomic RNA, but the smaller (~2.9 kb) is a novel species with some characteristics of a satellite RNA (the BWYV-ST9 associated RNA, or aRNA; Chin *et al.*, 1993). Sequence analysis of the aRNA revealed four ORFs, arranged as a block of three ORFs at the 5' end of the genome, and a small solitary ORF downstream of the major coding block (Fig. 1.2; Chin *et al.*, 1993). Analysis of the deduced amino acid sequences of the ORFs revealed the following features: Open reading frames 2 and 3 showed homology to putative RNA-dependent RNA polymerase genes of the carmoviruses, and also to that of luteovirus subgroup I. The **GDD** motif characteristic of RNA-dependent RNA polymerases (Argos and Kamer, 1984) occurred in ORF 3. No significant relationships could be found for ORFs 1 or 4. The BWYV-ST9 aRNA was able to replicate autonomously in plant protoplasts and inoculated leaves, but was dependent on the BWYV genomic RNA for encapsidation and cell-cell movement (Passmore *et al.*, 1993). Interestingly, plants infected with BWYV-ST9 contain approximately 10 times more virions per gram of tissue than do plants infected with BWYV isolates that lack the ST9 aRNA (Falk and Duffus, 1984), a phenomenon that may be related to the synergistic interactions between subgroup I and II barley yellow dwarf viruses (1.3.4.1).

1.4.3.2 Pea enation mosaic virus (PEMV)

Pea enation mosaic virus (PEMV) has some characteristics in common with luteoviruses. It is a spherical virus which can be transmitted in a persistent, circular manner by aphids, although unlike the luteoviruses mechanical transmission is also

possible. The interaction of the virus with the aphid appears to be similar to that of the luteoviruses at the subcellular level (Harris and Bath, 1972; Harris *et al.*, 1975). Similarly, the cytopathology of PEMV infection resembles that of subgroup II luteoviruses (Demler and de Zoeten, 1991). PEMV shows a strong affiliation with the phloem tissue of infected plants, although it is not phloem limited. However, PEMV differs most markedly from members of the luteovirus group in that it possesses a bipartite genome, each member of which is able to replicate in the absence of the other RNA species (Demler *et al.*, 1993; Demler *et al.*, 1994).

RNA 1 of PEMV has an ORF organisation markedly similar to the genome of subgroup II luteoviruses (Fig. 1.2; Demler and de Zoeten, 1991). There are five significant ORFs with varying amounts of homology to those of members of the luteovirus group. ORF 2 contains the **GDD** RNA-dependent RNA polymerase motif and is highly related to ORF 2 of subgroup II luteoviruses and the cognate ORF of the sobemoviruses. Similarly, ORFs 3 (coat protein) and 5 (possible aphid transmission factor) are also related to the corresponding ORFs of subgroup II, while ORF 1 shows a weaker relationship. ORF 0 does not share homology with ORF 0 of subgroup II and the deduced amino acid sequence is less hydrophobic, although it does contain a region capable of encoding a membrane-spanning protein (Demler and de Zoeten, 1991). The most notable feature of organisation of RNA 1 *vis-à-vis* the subgroup II genome is the absence of ORF 4. Although RNA 1 of PEMV can replicate autonomously in infected plants, it is not able to move from the site of infection. This is evidence for the proposed movement protein nature of the luteoviral ORF 4 product (1.4.2.7).

RNA 2 of PEMV also encodes a protein containing the **GDD** motif, however in this case the ORF (ORF 2) is most closely related to that of luteovirus subgroup I and the putative RNA-dependent RNA polymerase gene of the carmoviruses (Fig. 1.2; Demler and de Zoeten, 1993). ORF 2 is overlapped by ORF 1, an organisation which is similar to that of subgroup I luteoviruses. However, analysis of the deduced amino acid sequence of ORF 1 failed to identify any homology with the corresponding ORFs of subgroup I, or with the carmoviruses and related viruses. PEMV RNA 2 encodes three

further ORFs downstream of ORF 2; these are thought not to have a structural role (*i.e.* as components of the virions) but are possibly involved in cell-cell movement of the virus. Unlike RNA 1, infections involving RNA 2 only are able to infect the plant systemically (Demler *et al.*, 1994), therefore the movement function is likely to reside on RNA 2. There is some doubt as to whether the 3' terminal M_r 15 K ORF is significant (Demler *et al.*, 1993).

The most striking feature of PEMV is the marked similarity of RNA 1 to the genome of subgroup II luteoviruses, in contrast with RNA 2 which has a polymerase gene closely related to that of subgroup I luteoviruses (Demler and de Zoeten, 1991; Demler *et al.*, 1993). Infections with both RNAs 1 and 2 were necessary to reproduce wildtype PEMV symptoms, although RNA 1 directed the synthesis of virus-like particles and recreated typical PEMV cytopathology in the absence of RNA 2 (Demler *et al.*, 1994). RNA 2 conditioned systemic movement of the virus (Demler *et al.*, 1994). This second result should perhaps be viewed with caution, because all experiments relied on mechanical inoculation of viral RNAs. Delivery of RNA 1 to phloem tissue may allow systemic phloem-limited infection, by analogy to subgroup II of the luteoviruses.

PEMV is therefore a complex of two unrelated RNAs, which self-replicate but are otherwise interdependent for encapsidation and movement functions (Demler *et al.*, 1994). This relationship further demonstrates the synergism that seems to exist between luteoviral subgroups I and II, as reviewed above for strains of BYDV (1.3.4.1) and postulated for BWYV-ST9 and its aRNA. RNA 2 of PEMV is directly analogous to the BWYV-ST9 aRNA, because of its autonomous replication but dependency on RNA 1 for encapsidation and (presumably) aphid transmission. However, the relationship between RNAs 1 and 2 of PEMV is more complex than the interaction between independent luteoviruses, because of the evolution to interdependence.

1.4.4 Strategies for gene expression in the luteoviruses

Fig. 1.3 contains a diagrammatic representation of the strategies for gene expression in the luteoviruses, and should be referred to throughout this section.

1.4.4.1 Translation from the genomic RNA

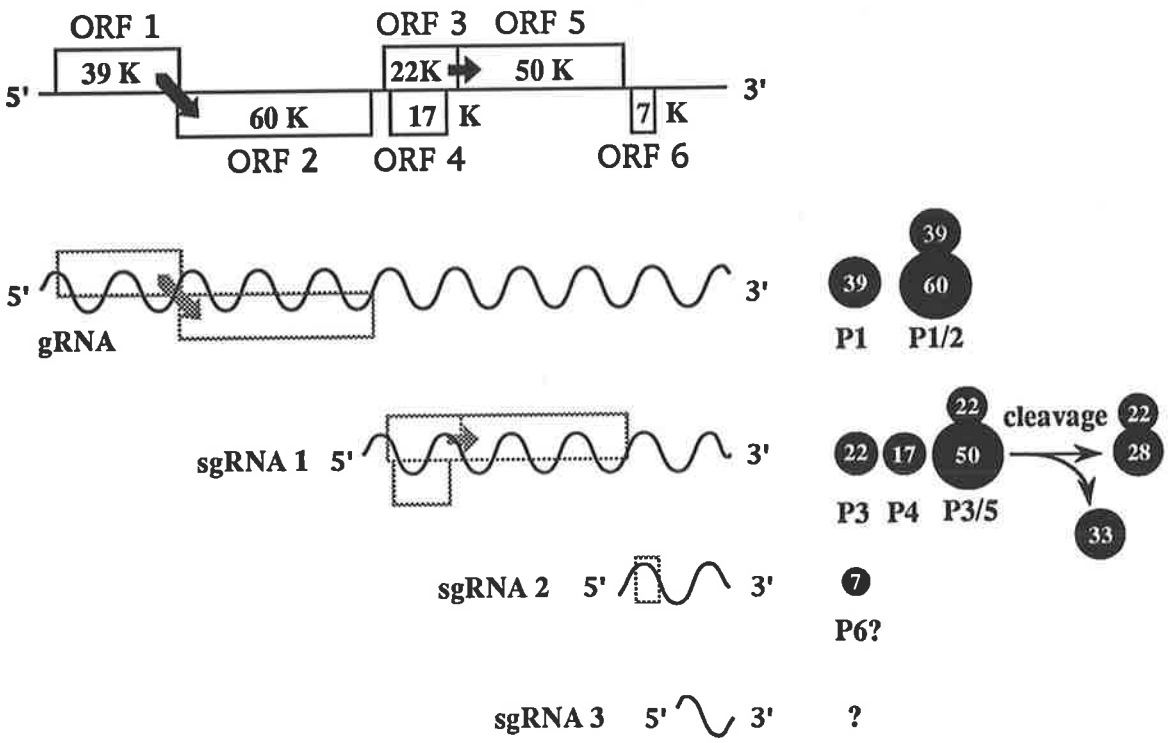
ORF 1 of subgroup 1, and ORFs 0 and 1 of subgroup 2, are believed to be translated directly from the respective genomic RNAs. *In vitro* translation of genomic BYDV-PAV RNA gave a major protein product of M_r 39 K, corresponding to the predicted size of the ORF 1 product (Young *et al.*, 1991). Other minor products were also seen. In subgroup II luteoviruses, *in vitro* translation of genomic RNAs gave major products of M_r 28 K and 70 K for PLRV (Mayo *et al.*, 1989), or M_r 25 K and 66 K for BWYV (Veidt *et al.*, 1992). The smaller product corresponds to the predicted size of ORF 0, and the larger to the predicted size of ORF 1. Mutation of the 5' proximal AUG resulted in the loss of the M_r 26 K protein in *in vitro* translations of BWYV RNA, thus establishing the relationship between the 26 K product and ORF 0 (Veidt *et al.*, 1992). The AUG start codons of ORFs 0 and 1 are the first such codons in the subgroup II genome (Miller *et al.*, 1994), and are separated by 133 nt in PLRV (Mayo *et al.*, 1989) or 142 nt in BWYV (Veidt *et al.*, 1988). Therefore, it appears that a proportion of ribosomes are able to scan past the first AUG codon to initiate translation at the second (ORF 1) AUG. No other luteoviral ORF is believed to be translated from the genomic RNA by direct initiation of translation at its start codon.

1.4.4.2 Expression of ORF 2 by -1 frameshift from ORF 1

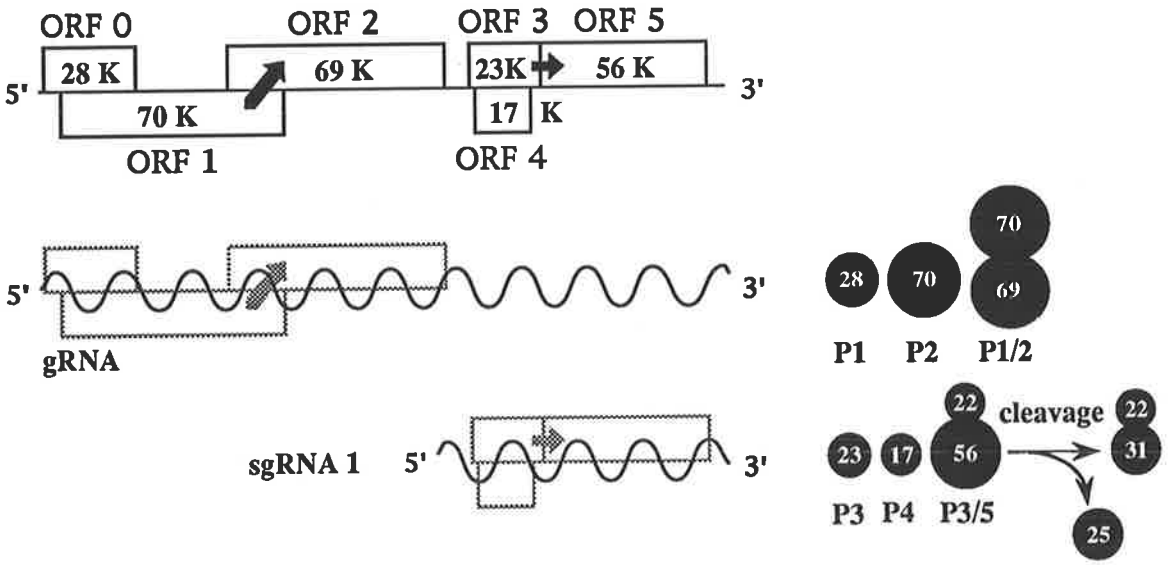
ORF 2 is expressed as a fusion with the ORF 1 product by -1 ribosomal frameshifting in both luteoviral subgroups (Miller *et al.*, 1994). The structures required for -1 frameshifting are different in each subgroup, and possibly also between isolates of PLRV. However, the basic requirements are similar and appear to be a 'shifty heptanucleotide' sequence normally consisting of three A, U, or G residues, followed by either UUUA, UUUU, AAAC or AAAU (Miller *et al.*, 1994), as well as secondary

Fig. 1.3. Expression strategies of luteovirus ORFs. Depiction of subgroup I and II genomes are as for Fig. 1.2. Wavy lines represent genomic and subgenomic RNAs from which proteins are translated. Grey boxes superimposed on RNAs represent ORFs translated from the RNAs. Filled black circles represent protein products with relative molecular masses given in K (000's). Refer to the text (1.4.4) for details. (A) Expression strategies for subgroup I luteoviruses (BYDV-PAV). Large arrows in the genome diagram refer to -1 frameshifting and stop codon readthrough events respectively. ORFs 1 and 2 are translated from the genomic RNA, ORF 2 as a frameshift fusion with the product of ORF 1. Protein product sizes (in K) are given at the right of the diagram. ORFs 3, 4 and 5 are translated from sgRNA 1. ORF 5 is translated as a readthrough fusion with the product of ORF 3, after which it is processed to give M_r 50 K and 33 K products. sgRNA 2 is sufficient for translation of ORF 6 but it is not known if this event actually occurs. The role of sgRNA 3 in gene expression (if any) is unknown. (B) Expression strategies for subgroup II luteoviruses (PLRV). Large arrows in the genome diagram refer to -1 frameshifting and stop codon readthrough events respectively. ORFs 0, 1 and 2 are translated from the genomic RNA, ORF 2 as a frameshift fusion with the product of ORF 1. Protein product sizes (in K) are given at the right of the diagram. ORFs 3, 4 and 5 are translated from sgRNA 1. ORF 5 is translated as a readthrough fusion with the product of ORF 3, after which it is processed to give a M_r 53 K product which is packaged into virions. The M_r 25 K product has not been observed but is depicted by analogy to subgroup I. The molecular weight for the M_r 25 K product was derived by subtraction of the M_r 53 K product from the full-length ORF 3-ORF 5 fusion protein (deduced size M_r 78 K).

A Subgroup I (BYDV-PAV)



B Subgroup II (PLRV)



structure(s) in the RNA template. Each subgroup will be treated separately here because of the variable requirements for frameshifting.

Brault and Miller (1992) detected -1 frameshift translation of ORF 2 equal to ~1% of ORF 1 translation in BYDV-PAV using a reporter gene construct in carrot cells. The ORF 1 stop codon was absolutely required for frameshifting. The authors noted a potential shifty heptanucleotide sequence (GGGUUUU) in the 13 nt overlap between ORFs 1 and 2, as well as potential RNA secondary structures that could form either side of the nominated frameshift site. While the role(s) of these structures was not investigated, it is interesting to note the possible pseudoknot structure downstream of the shifty heptanucleotide sequence that is analogous to a similar structure proposed to be involved in frameshifting in subgroup II luteoviruses (see below; Kujawa *et al.*, 1993; Garcia *et al.*, 1993). The proposed shifty heptanucleotide frameshift site was subsequently confirmed by amino acid sequencing of the putative readthrough product (Di *et al.*, 1993). Furthermore, a M_r 99 K protein corresponding to the predicted size of the transframe (ORF 1/2) protein was precipitated by an antiserum raised against the translation product of ORF 2, thus identifying the frameshift protein (Di *et al.*, 1993). Surprisingly, the terminal 600 nt of the BYDV-PAV genome was required for efficient frameshifting in wheat germ extracts (Di, 1992, cited by Miller *et al.*, 1994).

Translational frameshifting between ORFs 1 and 2 of subgroup II luteoviruses occurs at a similarly low level to that in subgroup I (~1%; Prüfer *et al.*, 1992; Garcia *et al.*, 1993). However, the frameshift site of subgroup II is different from subgroup I in the following ways. Firstly, the region of overlap between ORFs 1 and 2 is much larger (582 nt in PLRV). The shifty heptanucleotide sequence, a non-canonical UUUAAAU in PLRV (Prüfer *et al.*, 1992; Kujawa *et al.*, 1993) or GGGAAAC in BWYV (Garcia *et al.*, 1993), occurs roughly 25% of the distance into the overlap. There is some controversy over the RNA secondary structures required to support frameshifting. Prüfer *et al.* (1992) found that a stem-loop structure 3' of the proposed frameshifting site in PLRV (German isolate) was necessary for frameshifting. An alternative weak pseudoknot structure that could be drawn from the sequence at that site

did not form. This was refuted by Kujawa *et al.* (1993), who found that mutations which abolished the pseudoknot structure of PLRV (Polish isolate) reduced the efficiency of frameshifting. Disruption of the proposed stem-loop structure by deletion of its 3' end of did not affect frameshifting efficiency. The few nucleotide differences between the sequences of the different PLRV isolates in the frameshifting region may account for the conflicting results of the two groups (Kujawa *et al.*, 1993). However, frameshifting in BWYV was also dependent on a homologous weak pseudoknot 3' of the frameshifting site, supporting the role of the pseudoknot in frameshifting in subgroup II luteoviruses. Additionally, BWYV does not have the potential to form a stem-loop structure at that point in the genome (Garcia *et al.*, 1993), despite its close relationship to PLRV.

1.4.4.3 Translation from subgenomic RNAs: ORFs 3, 4 and 5

Both luteoviral subgroups transcribe a major subgenomic RNA (sgRNA 1) of 2.5-3.0 kb from the genomic RNA, the 5' end of which maps upstream of ORF 3. ORFs 3, 4 and 5 are believed to be translated from sgRNA 1, albeit by different mechanisms. This section discusses the mapping of the 5' end of sgRNA 1 for different luteoviruses, and expression strategies for the ORFs that reside on it.

1.4.4.4 Mapping of the 5' end of sgRNA 1

The 5' end of sgRNA 1 has been mapped for different isolates of both BYDV-PAV and PLRV. In BYDV-PAV, the 5' end of sgRNA 1 has been localised in different viral isolates to nucleotide 2769 (Dinesh-Kumar *et al.*, 1992) or nucleotide 2670 (Kelly *et al.*, 1994) of the Vic isolate. Both groups used northern blot, RNase protection and primer extension techniques to obtain their data. The conflicting results may be due to differences between the two isolates of BYDV-PAV, but comparative analyses suggests that Kelly *et al.* (1994) are more likely to be correct (Miller *et al.*, 1994). This is because the sequence at the 5' end of sgRNA 1 as determined by Kelly *et al.* (1994) closely matched the sequence at the 5' end of the genomic RNA. Such a relationship

occurs for the genomic and subgenomic RNAs of many different unrelated viruses. Furthermore, Kelly *et al.* (1994) mapped two further RNAs (sgRNA 2 and sgRNA 3, possibly corresponding to the two small dsRNA species described by Gildow *et al.*, (1983)), one of which (sgRNA 2) also contained the conserved sequence at its 5' end. Localisation of the 5' end of sgRNA 1 to nucleotide 2670 of BYDV-PAV-Vic as determined by Kelly *et al.* (1994) would give a leader sequence of 188 nt before the AUG initiation codon of ORF 3, and a total subgenomic RNA length of ~3 kb.

A similar controversy exists over the starting point of sgRNA 1 in PLRV. Tacke *et al.* (1990) estimated from primer extension experiments that the 5' end of sgRNA 1 of a field isolate of PLRV lay 40 nt upstream of the AUG translational start site of ORF 3. This result gave a predicted size of ~2.3 kb for sgRNA 1, which is significantly less than the value of ~2.6 kb obtained by the same authors using northern analyses. Miller and Mayo (1991) mapped the 5' end of sgRNA 1 of PLRV-S (Table 1.1) to a position corresponding to nucleotide 3376 of PLRV-N. This predicts a subgenomic size of ~2.5 kb, in good agreement with the data of Tacke *et al.* (1990), but some 200 nt less than their own estimates from northern blots. However, the sequence at the 5' end of sgRNA 1 as determined by Miller and Mayo (1991) closely matched the 5' terminal genomic sequences of other PLRV isolates. As for BYDV-PAV, this is considered strong evidence that the location of the 5' of sgRNA 1 by Miller and Mayo (1991) is correct (Miller *et al.*, 1994). The AUG translational initiation codon of PLRV-N occurs at nucleotide 3588; thus the leader sequence of sgRNA 1 in this virus consists of 212 nt. Note that the 5' extremity of sgRNA 1 in both luteoviral subgroups maps to the C-terminus of the coding region of ORF 2.

1.4.4.5 Translation of ORFs 3 and 4 from sgRNA 1

Mayo *et al.* (1982) was unable to synthesise coat protein in *in vitro* translations of PLRV genomic RNA. Other workers have demonstrated that ORFs 3 (coat protein) and 4 are translated from sgRNA 1. *In vitro* translation of BYDV-RPV or BYDV-PAV RNA molecules similar to sgRNA 1 (Vincent *et al.*, 1991; Dinesh-Kumar *et al.*, 1992)

showed that the putative ORF 4 (M_r 17 K) product accumulated to an equal or greater amount than the putative ORF 3 product. This is despite the internal location of ORF 4 relative to ORF 3. This result was confirmed *in vivo* by Tacke *et al.* (1990) who found seven times greater ß-glucuronidase (GUS) activity in protoplasts transformed with PLRV ORF 4-GUS fusions than ORF 3-GUS fusions. Similar experiments conducted by Dinesh-Kumar and Miller (1993) with BYDV-PAV found a ratio of ORF 4/ORF 3 expression of ~2:1. Therefore, the ORF 4 translational initiation codon appears to sequester ribosomes at the expense of the ORF 3 AUG. The greater translational efficiency of the ORF 4 AUG relative to that of ORF 3 was mainly due to the sequence context in which the AUG codon occurred (Dinesh-Kumar and Miller, 1993). The translation of overlapping ORFs from a single subgenomic RNA is reminiscent of the translation of ORFs 0 and 1 from the genomic RNA in subgroup II luteoviruses as described above (1.4.4.1).

1.4.4.6 Expression of ORF 5 as a readthrough fusion with the coat protein

The organisation of ORF 5 relative to ORF 3, as well as the sequence context around the ORF 3 stop codon (which is highly conserved in all luteoviruses), led early workers to propose that ORF 5 is translated after ribosomal 'readthrough' of the ORF 3 stop codon (Miller *et al.*, 1988a; Veidt *et al.*, 1988). Thus the ORF 5 product would be expressed as a fusion protein with the ORF 3 (coat protein) product. *In vitro* translation of sgRNA 1-like RNA molecules results in the formation of three major protein products. The smaller of these correspond to the products of ORFs 3 and 4, as discussed above. The third protein product is equal in size to the predicted translational product of ORF 3 added to ORF 5 (Veidt *et al.*, 1988; Dinesh-Kumar *et al.*, 1992). The readthrough product has been identified by western blot analysis in plant protoplasts and infected tissue using antibodies raised against the product of ORF 5 (Bahner *et al.*, 1990 (PLRV); Reutenauer *et al.*, 1993 (BWYV); Cheng *et al.*, 1994a; Filichkin *et al.*, 1994 (BYDV-PAV)). The fusion protein is cleaved at the carboxyl terminus to give products of M_r ~50 K and ~33 K in BYDV-PAV, the larger of which is a component of

virus particles (Filichkin *et al.*, 1994). A similar cleavage event may also occur in PLRV infections (Bahner *et al.*, 1990). The efficiency of readthrough of the ORF 3 stop codon varied from ~1%, measured by reporter gene fusions in plant protoplasts (Tacke *et al.*, 1990), to 7-15% in *in vitro* translations (Dinesh-Kumar *et al.*, 1992).

1.4.4.7 Other subgenomic RNAs of BYDV-PAV

Subgroup II luteoviruses posses a single subgenomic RNA (sgRNA 1) which is homologous and functionally equivalent to sgRNA 1 of subgroup I. However, subgroup I luteoviruses appear to transcribe a further two subgenomic RNAs from the genomic 3' region (Kelly *et al.*, 1994), although their role (if any) in expression of ORF(s) has not been characterised.

sgRNA 2 of BYDV-PAV is approximately 850 nt in size, and maps to nucleotide 4809 of the Vic isolate (Kelly *et al.*, 1994). This gives a leader sequence of 111 nt before the AUG translational start codon of ORF 6. Synthetic sgRNA 2 directs the translation of a protein corresponding to the size of ORF 6 *in vitro*, although it is not known if this ORF is expressed *in vivo*. sgRNA 3 is the most abundantly expressed viral RNA in BYDV-PAV infection. It is approximately 350 nt in size, and maps to nucleotide 5348 of the BYDV-PAV-Vic genome. There are no conserved ORFs 3' of the sgRNA 3 start site, and synthetic sgRNA 3 was unable to direct the synthesis of any small proteins *in vitro* (Chalhoub *et al.*, 1994; Kelly *et al.*, 1994).

The 5' terminal sequences of BYDV-PAV sgRNA 2 and the genomic RNA are similar, and match the sequence at the 5' end of one determination of sgRNA 1 (Kelly *et al.*, 1994). This sequence is not matched by that at the 5' end of sgRNA 3, although such a sequence does occur immediately upstream and contiguous with the proposed start site of sgRNA 3 (Miller *et al.*, 1994). Such homology is evidence of the functionality of the sgRNAs, rather than (say) merely representing breakdown products from other viral RNAs. Despite this, no function has yet been ascribed to either sgRNA 2 or sgRNA 3. However, Miller *et al.* (1994) reported that (1) a region between nucleotides 4513 and 5009 (encompassing the 3' end of ORF 5 and the 5' terminal half

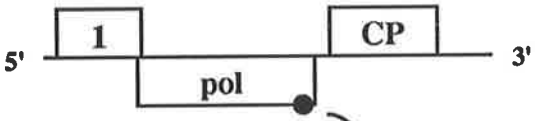
of ORF 6) was necessary for translation of uncapped transcripts in wheat germ extracts and (2) the 3' terminal 600 nt of the BYDV-PAV genome was required for efficient frameshifting between ORFs 1 and 2. It is therefore possible that sgRNAs 2 and 3 could have some role(s) in mediating these effects, although it is not clear if this is due to expression of ORFs or through a structural role of the RNAs themselves.

1.5 Evolution of the luteoviruses

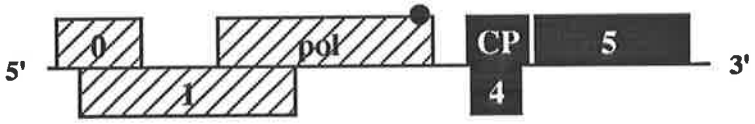
The luteoviruses fall into two subgroups based on comparisons between their putative RNA-dependent RNA polymerase genes (ORF 2; Habili and Symons, 1989). Subgroup I luteoviruses are most closely related to the carmoviruses in this region, whereas subgroup II luteoviruses share more homology with the sobemoviruses. However, the subgroups are closely related in their 3' cluster of genes (ORFs 3, 4 and 5). Such a relationship clearly arose by RNA-RNA recombination between unrelated plant viruses, and may represent the most recent example of such recombination in the positive-strand RNA viruses (Koonin and Dolja, 1993). However, it is not known if subgroup I arose from subgroup II luteoviruses or vice versa, or if they arose independently by repeated recombination between the donor of ORFs 3, 4 and 5 with successive polymerase donors. Miller *et al.* (1994) have proposed (Fig. 1.4) that the subgroup I genome was formed after a strand-switching RNA recombinational event between a diantho-like virus (possessing similar organisation of ORFs 1 and 2 to subgroup I luteoviruses) and a subgroup II luteovirus (donating ORFs 3, 4 and 5). Recombination is proposed to have occurred at small conserved sequences (5'-ACAAA-3') at the respective putative sgRNA 1 promoters. A second RNA recombination event would be necessary for the prototypical subgroup I genome to obtain the long 3' UTR (with or without ORF 6) that is unique to subgroup I luteoviruses.

Fig. 1.4. Recombinational model for the formation of the subgroup I luteovirus genome (redrawn and modified from Miller *et al.*, 1994). Open boxes represent open reading frames. Unshaded boxes are of carmovirus-like origin, diagonally shaded boxes of subgroup II (sobemo-like) origin, filled black boxes of luteovirus origin. Circles (filled and unfilled) represent position of putative subgenomic RNA promoters at which recombination is proposed to have occurred. Arrows show indicate RNA recombinational events and show the direction of evolution.

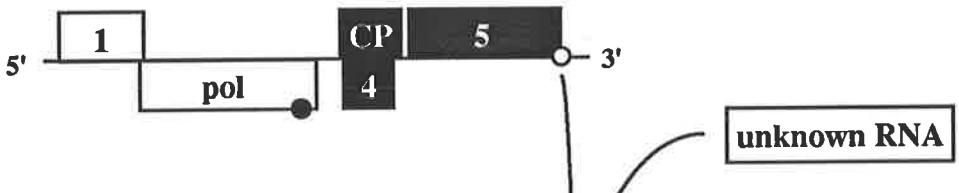
Dianthovirus



Subgroup II



Proto-subgroup I



Subgroup I

1.6 Control of BYDV infection

Losses due to BYDV infection have been controlled in the following ways: (1) Spraying of aphids with organophosphate insecticides after primary infection to prevent the spread of aphids from their original foci of colonisation (2) manipulation of sowing date so that plants have passed the vulnerable 2-3 leaf stage by the time viruliferous aphids migrate into the crops and (3) exploitation of cereal varieties that are resistant to the virus (Johnstone *et al.*, 1990). Of these options, cultivar resistance is the most attractive because of its inexpensiveness, the ecological advantage in circumventing pesticide usage, and also the flexibility it allows in agricultural practice. Useable resistances have been found for all of the major Triticae crops, however these may be linked to undesirable agronomic traits, and the genetics of resistance to BYDV is generally poorly understood (Comeau and Jedlinski, 1990). This thesis therefore confines discussion to the better studied resistances, with particular emphasis on the Yd2 gene of barley.

1.6.1 Resistance to BYDV conferred by the Yd2 gene

Resistance to BYDV was first observed in the commercial barley cultivar Rojo, and was subsequently shown to be conferred by a single recessive gene now known as Yd1 (Suneson, 1955). Stronger resistance to BYDV was found in four varieties of barley by Rasmusson and Schaller (1959), and shown to segregate as a single, incompletely dominant gene that was designated Yd2. Screening of 6689 accessions of the world barley collection for reaction to BYDV infection resulted in the identification of 117 further resistant varieties (Schaller *et al.*. 1963). 113 of these originated from Ethiopia, while three of the remaining four varieties were hybrids with Ethiopian varieties in their parentage. The final accession originated from China. Genetic studies of 16 BYDV resistant barley varieties of Ethiopian origin showed that all possessed the same gene (Yd2) for resistance (Damsteegt and Bruehl, 1964; Schaller *et al.*, 1964). The gene was localised to chromosome three of the barley genome by co-segregation

with morphological markers in genetic crosses (Schaller *et al.*, 1964). Yd2 has now been incorporated into commercial cultivars of barley world wide (Schaller, 1984).

Catherall *et al.* (1970) reported that the Yd2 gene existed in allelic forms that differed in their effectiveness against BYDV. Alleles that provided high, intermediate or low levels of 'tolerance' to BYDV retained their relative effectiveness when crossed into a new genetic background. Similarly, the Yd2 allele of the commercial variety Shannon was less effective against a BYDV-PAV/BYDV-RPV mixed isolate than those present in two Ethiopian barley lines (Larkin *et al.*, 1991). Other workers have found that the phenotype of Yd2 can vary from recessive to incompletely dominant depending on environmental conditions, although the arbitrary nature of symptom classification may obscure the true genetic relationship (Schaller, 1984). Yd2 was more effective in fast-growing barley cultivars than in those that were slower growing (Catherall *et al.*, 1970; Jones and Catherall, 1970a); segregation of these traits was not attempted so it is not clear if there is a primary relationship between growth rates and BYDV resistance.

Importantly, several authors have shown that the Yd2 gene is active against BYDV-PAV and BYDV-MAV (both subgroup I luteoviruses), but not BYDV-RPV (subgroup II; Jones and Catherall, 1970b; Baltenberger *et al.*, 1987; Herrera and Plumb, 1989; Larkin *et al.*, 1991). One of three BYDV-RPV isolates appeared to be more susceptible to Yd2 resistance than the others (Banks *et al.*, 1992), however it is possible that the resistance apparently observed in this case was in fact due to the suppression of virus replication by a small satellite RNA that is found with some isolates of BYDV-RPV (Dr P Waterhouse, CSIRO Division of Plant Industry, Canberra, Australia, personal communication; Miller *et al.*, 1991). The Yd2 gene also provides some resistance against mixed BYDV-PAV/BYDV-RPV infections, which normally devastate the host plant (Baltenberger *et al.*, 1987; Larkin *et al.*, 1991). The BYDV resistance conferred by Yd2 appears not to be active in plant protoplasts (Larkin *et al.*, 1991).

1.6.2 Other major gene resistance to BYDV

Resistance to BYDV has been observed in the wheat grass *Thinopyrum intermedium* (Xin *et al.*, 1988, and references therein). This is important because the resistance gene can potentially be transferred by standard cytogenetic procedures to wheat, for which no major gene conferring resistance to BYDV has been described. The *Th. intermedium* resistance resides on the ß arm of the group 7 chromosome, and appears to reduce the concentration of both BYDV-PAV and BYDV-RPV coat protein antigens in plants infected with these viruses (Brettell *et al.*, 1988). Incorporation of the gene into elite commercial cultivars by genetic manipulation will however be a lengthy process. In Italy, resistance to rice giallume virus (probably a subgroup II luteovirus) conferred by a single incompletely dominant gene has been observed in rice (Baldi *et al.*, 1990). The gene is now being exploited in rice breeding programs (Baldi *et al.*, 1991).

1.7 Aims

The general aim of the project in this laboratory is to characterise the molecular events in the interaction of BYDV-PAV with the Yd2 gene of barley leading to expression of the resistance phenotype. Within this framework, the specific aims of this thesis are:

(1) To complete the nucleotide sequence of soybean dwarf luteovirus, already known to be closely related to BYDV-PAV (Habili and Symons, 1989).

(2) From the results from this work, establish a strategy for the investigation of the viral ORF conditioning the interaction of BYDV-PAV with the Yd2 gene.

CHAPTER TWO

GENERAL MATERIALS AND METHODS

2.1 Materials

All general laboratory reagents were at least analytical grade in standard. Suppliers are listed only where alternate sources might affect performance or quality of reagents. Solutions were prepared under sterile conditions with ultra-pure water, and autoclaved where appropriate.

2.1.1 Synthetic oligodeoxyribonucleotides

Synthetic oligodeoxyribonucleotides were prepared on an Applied Biosystems (USA) Model 380B DNA synthesiser by Dr Neil Shirley in the Department of Plant Science, University of Adelaide. Oligonucleotides were purified by ion exchange HPLC using a MonoQ column (Pharmacia, USA). Sequences of oligonucleotides are given in the text of this thesis.

2.1.2 Nucleotides and radionucleotides

Ultrapure nucleotide triphosphates (NTPs) and deoxynucleotide triphosphates (dNTPs) were obtained from Pharmacia. α-^{32}P-dATP (10 mCi/ml), α-^{32}P-UTP (10 mCi/ml) and α-^{35}S-dATP (12.5 mCi/ml) were obtained from Bresatec (Australia).

2.1.3 Bacterial strains, growth media and cloning vectors

Escherichia coli strain DH5α (*supE*44 Δ*lac*U169 (Ø80 *lacZ*ΔM15) *hsdR*17 *recA*1 *endA*1 *gyrA*96 *thi*-1 *relA*1; BRL, USA) was used for all routine cloning work in this thesis. The dcm$^-$ dam$^-$ strain JM110 (*dam dcm supE*44 *hsdR*17 *thi leu rpsL lacY galK galT ara tonA thr tsx* Δ(*lac-proAB*) F' [*traD*36 *proAB*$^+$ *lacI*^{q} *lacZ*ΔM15]; Yanisch-Perron *et al.*, 1985) was used where necessary. Bacteria were grown in LB broth (1% ($^w/_v$) bacto-tryptone, 0.5% ($^w/_v$) yeast extract, 1% ($^w/_v$) NaCl, pH 7.0) or 2YT (1.6% ($^w/_v$) bacto-tryptone, 1% ($^w/_v$) yeast extract, 0.5% ($^w/_v$) NaCl, pH 7.0). Antibiotics were added to growth media where appropriate in the following concentrations; ampicillin 50-100 µg/ml; kanamycin 50 µg/ml; rifampicin 25 µg/ml. Bacteria were plated out on solid media composed of LB broth containing 1.5% ($^w/_v$)

bacteriological agar (DIFCO, USA). Routine cloning was carried out using the vector pBluescript SK+ (Stratagene, USA), a phagemid carrying ampicillin resistance and with promoter sequences for T3 and T7 RNA polymerases flanking the polylinker. Other vectors were used as indicated in the text of this thesis.

2.2 Methods

Methods were carried out according to standard procedures (*e.g.* Sambrook *et al.*, 1989) or using manufacturers specifications except where indicated. Routine methods used throughout this thesis are recorded here with listing of suppliers and solution components where appropriate. Specific methods are listed in each chapter as necessary.

2.2.1 Purification of vector DNAs

2.2.1.1 Small scale preparations of plasmid DNA

The following procedure was used to purify small amounts of plasmid DNA for routine manipulations. Plasmid DNA was isolated from 1.5 ml of a stationary phase, plasmid-containing bacterial culture.

Bacteria grown in the appropriate medium containing antibiotics as necessary were pelleted by centrifugation at full speed at room temperature (RT) in a bench micro-centrifuge (Eppendorf, Germany). The supernatant was discarded and the pellet resuspended in 100 µl of GET buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl pH 8.0). Two hundred µl of 0.2 M NaOH, 1% SDS (prepared fresh) was added, mixed gently and left at RT for ~1 min. One hundred and fifty µl of KAcF (3 M potassium acetate, 1.8 M formic acid) was added and mixed by gentle inversion of the tube. The mixture was centrifuged for 5 min at RT in a bench micro-centrifuge, and 350 µl of the supernatant removed and placed in a fresh tube. DNA was precipitated from the supernatant by the addition of 400 µl ice-cold 2-propanol, and pelleted by centrifugation for 5 min as previously. The supernatant was discarded and the pellet washed by vortex mixing in 400 µl ice-cold 70% ethanol. The DNA was pelleted by

centrifugation as previously described and the supernatant discarded, after which the pellet was dried either *in vacuo* or by evaporation at RT for 15 min. The pellet was resuspended in 20 µl of TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA) containing 20 µg/ml DNase-free RNase A.

2.2.1.2 Large scale preparations of plasmid DNA

The following method was used to purify large amounts (≥150 µg) of plasmid DNA. A plasmid-containing bacterial culture was grown overnight to stationary phase (~16 h) in 400 ml of 2YT containing appropriate antibiotic(s), in a baffled 2 litre flask at 37°C. Cells were sedimented by centrifugation (5,000 rpm, 15 min, 4°C, Sorvall GSA rotor) and washed in STE buffer (50 mM NaCl, 1 mM EDTA, 25 mM Tris-HCl pH 8.0). The cells were pelleted by centrifugation as before then resuspended in 4 ml of GET buffer containing 1 mg/ml lysozyme. The resuspended cells were incubated on ice for 10 min, before the addition of 8 ml of freshly prepared 0.2 M NaOH, 1% SDS, followed by gentle mixing. After incubation on ice for 10 min the mixture was centrifuged as previously except that centrifugation was at 10,000 rpm for 10 min. The supernatant was removed and strained through four layers of cheesecloth, before precipitation of nucleic acids by the addition of 12 ml of ice-cold 2-propanol. The pellet was washed with ice-cold 70% ethanol, after which it was dried in air at RT before resuspension in 3 ml TE. LiCl (10 M) was added to a final concentration of 2.5 M to precipitate RNA, and the solution placed on ice for 10 min. RNA was pelleted by centrifugation in a Sorvall HB4 rotor at 10,000 rpm for 15 min at 4°C. The supernatant was removed and DNA precipitated by addition of an equal volume of ice-cold 2-propanol, followed by centrifugation as previously to pellet DNA. The pellet was washed in ice-cold 70% ethanol and dried in air at RT. DNA was resuspended in 400 µl TE and transferred to a micro-centrifuge tube. RNase A was added to 20 µg/ml and the mixture incubated at 37°C for 1 h. The solution was extracted twice with phenol:chloroform (2.2.2), and once with chloroform to remove proteins. Plasmid DNA was precipitated from solution by the addition of an equal volume of 13% PEG

8000, 1.6 M NaCl, followed by incubation at RT for 5 min. DNA was recovered by centrifugation at full speed in a bench micro-centrifuge for 5 min at RT, and the pellet resuspended in 400 µl TE. DNA was again precipitated from solution by addition of 3M sodium acetate pH 4.6 to a concentration of 0.3 M, and 2.5 volumes of ice-cold ethanol. The DNA was pelleted by centrifugation as previously, and the pellet washed in 400 µl ice-cold 70% ethanol before re-centrifugation. The pellet was dried and resuspended in 400 µl TE.

2.2.1.3 Purification of M13 single-stranded DNA

Bacterial cultures infected with recombinant bacteriophage M13mp18 or M13mp19 were grown for 5 h. Bacteria were pelleted from 1.5 ml of each culture by centrifugation in a bench micro-centrifuge for 15 min at RT. One ml of the supernatant was removed and transferred to a new micro-centrifuge tube, and phage particles precipitated by the addition of 200 µl of 20% ($^{w}/_{v}$) PEG 8000, 2.5 M NaCl. The solution was incubated at RT for 5 min, then placed on ice for a further 15 min. Phage were pelleted by centrifugation for 10 min as previously, and the supernatant removed. The phage pellet was resuspended in 120 µl of 50 mM Tris-HCl pH 7.5, 5 mM EDTA, 0.5% ($^{w}/_{v}$) SDS. Phage DNA was recovered after phenol extraction and ethanol precipitation.

2.2.2 Phenol:chloroform extraction and ethanol precipitation of DNA

DNA solutions were vortexed thoroughly with one volume of phenol:chloroform (containing one volume of redistilled phenol (BDH, Australia) equilibrated in 50 mM Tris-HCl pH 8.0, and one volume of chloroform) and centrifuged for 10 min at room temperature (full speed in an Eppendorf micro-centrifuge for small quantities, or 10,000 rpm in a Sorvall HB4 rotor for larger solutions). The aqueous phase was recovered and the extraction repeated as necessary.

DNA was routinely precipitated from solutions with ethanol. Briefly, $^{1}/_{10}{}^{th}$ volume of 3M sodium acetate (pH 4.6) was added followed by 2.5 volumes of ice-cold

ethanol. The solutions was incubated on ice for 15 min, followed by centrifugation at high speed at RT for 15 min in an Eppendorf micro-centrifuge for small volumes, or at 10,000 rpm at 4°C for 15 min in a Sorvall HB4 rotor for larger volumes. Pellets were washed in 70% ethanol prior to drying *in vacuo* or on the bench at RT.

2.2.3 Restriction digestion of DNA

DNA was digested with restriction endonucleases supplied by Boehringer Mannheim (Germany), Promega (USA), New England Biolabs (USA) or Bresatec (Australia), using buffer systems recommended or supplied by the manufacturers. Restriction digests employing two enzymes were conducted using buffer conditions as close as possible to that recommended for each enzyme alone.

2.2.4 Gel electrophoresis

2.2.4.1 Agarose gel electrophoresis

Agarose minigels were prepared from 0.7-2.0% ($^{w}/_{v}$) solutions of SeaKem GTG agarose (FMC, USA) in 1xTBE (89 mM Tris-borate pH 8.3, 2 mM EDTA). Ten ml of the molten agarose solution was poured onto a 7.5 x 5.0 cm glass microscope slide after positioning of an appropriate well comb. One half volume of urea loading buffer (3x concentration is 2 M urea, 50 mM Tris-HCl, 20% ($^{w}/_{v}$) sucrose, 10 mM EDTA, 0.06% ($^{w}/_{v}$) xylene cyanol, 0.06% ($^{w}/_{v}$) bromophenol blue) was added to DNA samples before loading of the wells. Preparations of phage λ DNA digested with *Eco*RI, or phage SPP-1 DNA digested with *Eco*RI, or pUC19 DNA digested with *Hpa*II (Bresatec, Australia), were used as high, medium and low range molecular weight markers respectively. Gels were electrophoresed in 1xTBE running buffer at 80-120 mA. DNA was visualised by staining gels with ethidium bromide (10 μg/ml ($^{w}/_{v}$) in water). Gels were destained in water before photography under short wavelength UV light.

2.2.4.2 Polyacrylamide (sequencing) gel electrophoresis

Denaturing polyacrylamide gels were prepared from 50 ml solutions containing 6% ($^{w}/_{v}$) acrylamide, 0.3% ($^{w}/_{v}$) bisacrylamide, 7 M urea, 1xTBE. Polymerisation was initiated by the addition of 400 µl freshly prepared 10% ($^{w}/_{v}$) ammonium persulfate and 40 µl of TEMED. The polymerising solution was poured into gels of 20 x 40 x 0.04 cm, with well formation by shark's tooth combs. Gels were allowed to set for at least 60 min, then pre-electrophoresed at 50 W until gel temperature was approximately 50°C. Gels were electrophoresed at 50°C at constant power after loading and denaturing of samples in formamide loading solution (95% ($^{v}/_{v}$) formamide, 0.1% bromophenol blue, 0.1% xylene cyanol, 10 mM EDTA).

2.2.5 Purification of DNA from agarose gel slices (Geneclean)

Slices of agarose containing DNA fragments of interest were excised from agarose gels after detection with ethidium bromide and long wavelength UV light (2.2.4.1). DNA was extracted from the gel by the Geneclean procedure, using kits supplied by Bio101 (USA) or Bresatec (Australia). Extraction followed manufacturer's directions, with use of TBE modifier to allow purification of DNA from gels containing 1xTBE.

2.2.6 First-strand cDNA synthesis

RNA was denatured in the presence of 50 ng of specific first-strand oligonucleotide primer in TE buffer by heating to 80°C for 2 min, followed by cooling at RT for 5 min. The annealed RNA-primer mixture was then subjected to reverse transcription under the following buffer conditions; 10 mM Tris-HCl pH 8.5, 50 mM KCl, 5 mM $MgCl_2$, 10 mM DTT, 1 mM each dNTP, 1 U/µl RNasin (Promega, USA), and 8 U of AMV reverse transcriptase (Promega) in a final volume of 20 µl. The reaction was incubated at 45°C for 30 min, then stopped by heating at 80°C for 5 min. cDNAs were occasionally purified by phenol:chloroform extraction and ethanol precipitation (2.2.2).

2.2.7 Polymerase chain reaction (PCR)

Conditions for PCR varied depending on the DNA polymerase used to catalyse the reaction. Reactions using Taq polymerase utilised buffer conditions containing 50 mM KCl, 10 mM Tris-HCl pH 9.0, 0.1% ($^{v}/_{v}$) Triton X-100, 1.5-3.0 mM $MgCl_2$ and 1-5% of the first-strand cDNA reaction (2.2.6). Approximately 0.3 μM of each primer and 200 μM of each dNTP was used in each reaction. Reactions employing the high-fidelity Vent DNA polymerase (New England Biolabs, USA) were carried out using recommended buffer conditions (10 mM KCl, 20 mM Tris-HCl pH 8.8, 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 0.1% ($^{v}/_{v}$) Triton X-100), except that the concentrations of DNA primers and Mg^{2+} ions (present as $MgSO_4$) were optimised for each reaction. PCR reactions utilising Vent DNA polymerase contained dNTPs at a concentration of 500 μM each. PCR reactions were carried out on automated machines (DNA Thermal Sequencer) supplied by Corbett Research (Australia). These machines accepted either 0.5 ml microfuge tubes in a block configuration, or 30 μl capillary tubes in a circular formation. Heating and cooling in each case was by a fan-assisted Peltier effect mechanism.

2.2.8 End-filling using Klenow

Endfilling of double-stranded DNA fragments with 3' recessed ends for cloning or radioactive labelling using the large fragment of *E. coli* DNA polymerase I (Klenow fragment) was performed in a reaction containing 50 mM NaCl, 6 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, and 100 μM each dNTP. Individual dNTPs were substituted with radioactively labelled species where appropriate. The reaction was incubated at 37°C for 15 min, then terminated by incubation at 70°C for 10 min. DNAs were purified by phenol:chloroform extraction and ethanol precipitation (2.2.2). This method was also used for creation of blunt ends from 3' overhangs, with the substitution of T4 DNA polymerase for Klenow enzyme. In this case, the concentration of individual dNTPs was increased to 200 μM.

2.2.9 Ligation of vector and insert DNAs

Restricted, dephosphorylated vector (20-50 ng) was ligated with the DNA fragment to be cloned in molar ratios of 3:1, 2:1 and 1:1 (insert:vector respectively). The ligation was carried out in a volume of 20 µl containing 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.5 mM ATP and T4 DNA ligase (Bresatec, Australia). One fifth of a unit of T4 DNA ligase was used for sticky-end ligations, or 1 U for blunt-end ligations. The mixture was incubated for 4 h at RT, then diluted 1:5 in pure H_2O before use in transformations (2.2.10).

2.2.10 Transformation of *E. coli* with plasmids

Competent cells were prepared in bulk and stored at -80°C according to the method of Hanahan (1983). Briefly, 100 ml of LB containing 10 mM $MgSO_4$ and 10 mM $MgCl_2$ was inoculated with 1 ml of a fresh *E. coli* DH5α overnight culture in a 1 litre flat bottomed flask. Cultures were grown to an OD_{600} of 0.45 - 0.55, then centrifuged at 4000 rpm in a Sorvall HB4 rotor at 4°C for 5 min. The bacterial pellets were drained, then resuspended in 32 ml of FSB (100 mM KCl, 45 mM $MnCl_2.4H_2O$, 10 mM potassium acetate, 10 mM $CaCl_2.2H_2O$, 10% ($^v/_v$) glycerol, 3 mM hexammine cobalt chloride). The bacteria were pelleted by centrifugation as previously after incubation on ice for 15 min. Pellets were resuspended in 8 ml FSB and placed on ice. Fresh DMSO (280 µl) was added to the bacterial suspension, and mixed in by swirling the tube gently. The bacteria were incubated for a further 5 min on ice, after which the DMSO step was repeated. Cells were divided into 400 µl aliquots and snap frozen in liquid nitrogen, then stored at -80°C. For transformations, individual aliquots were thawed on ice. Approximately one quarter of each ligation mixture diluted in water (2.2.9) was mixed with a 100 µl aliquot of competent cells, then incubated on ice for 20-30 min. Cells were heat shocked at 42°C for 90 s, then allowed to recover by incubation at 37°C for 60 min after the addition of 0.9 ml 2YT. Approximately one third of each transformation suspension was plated out onto solid media containing

antibiotics as appropriate. Bacterial cells prepared and transformed according to this method had a competence of 10^7-10^8 cfu/μg DNA.

2.2.11 DNA sequencing

The dideoxynucleotide chain termination sequencing method (Sanger *et al.*, 1980) was used to determine DNA sequence. DNA sequencing was performed using the Klenow fragment of *E. coli* DNA polymerase I on M13 single-stranded DNA templates, using kits supplied by Bresatec (Australia), or using T7 DNA polymerase on double-stranded plasmid DNAs, using kits supplied by Pharmacia. Double-stranded DNAs were purified and denatured prior to sequencing according to the following method: Plasmid DNAs purified on a small-scale as described (2.2.1.1) were denatured by the addition of 4 μl of 2M NaOH to 16 μl of plasmid DNA. Denatured DNAs were purified from solution components by passage through micro-spin columns containing Pharmacia CL-6B resin, followed by elution in 20 μl TE. Ten μl of the purified denatured DNA was used in each sequencing reaction.

2.2.12 *In vitro* transcription of plasmid clones

In vitro transcription was used routinely to generate radioactive probes. Plasmids to be transcribed were linearised by digestion with the appropriate restriction endonuclease at the terminus of the sequence of interest. Linearised DNAs were purified by Geneclean (2.2.5) or phenol:chloroform extraction followed by ethanol precipitation (2.2.2). Transcription mixtures were set up as follows: 1-2 μg of linearised DNA was transcribed in a mixture containing 40 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT, 0.5 U/μl RNasin (Promega), 0.5 mM each of dATP, dCTP and dGTP, 12 μM UTP, 50-100 μCi α-^{32}P-UTP, and 20-40 U of T3 or T7 RNA polymerase (Promega, USA) as appropriate. Transcription reactions were incubated at 37°C for 90 min. The DNA template was destroyed by addition of 1U of RNase-free DNase (Promega) and incubation at 37°C for 15 min. Transcripts were purified either by polyacrylamide gel electrophoresis (2.2.4.2)

followed by elution and phenol:chloroform extraction coupled with ethanol precipitation (2.2.2), or phenol:chloroform extraction (2.2.2) followed by repeated precipitations in 2.5 M ammonium acetate and 2.5 volumes of ethanol to remove unincorporated radioactive label. Probes were stored in TE containing 5 mM β-mercaptoethanol.

2.2.13 RNA dot-blots

RNA dot-blots were used for routine indexing of plants for viral infection. Small-scale extractions of total RNA from 0.5-1.0 g of plant tissue were performed according to the method of Verwoerd *et al.* (1989) and resuspended in 20 µl TE. One µl of the RNA was denatured at 85°C for 5 min in a solution containing 50% ($^{v}/_{v}$) deionised formamide, 10 mM EDTA in a volume of 10 µl. The denatured RNA solution was brought to 4xSSC (1xSSC is 150 mM NaCl, 15 mM sodium citrate pH 7.0), and 4 µl of the denatured RNA was spotted onto a nylon membrane (Hybond N+, Amersham, UK). The RNA was fixed to the membrane by contact with a pad of absorbent paper soaked in 50 mM NaOH for 5 min, then washed in 2xSSC for 5 min. The filter was prehybridised in 5 ml of a solution containing 5xSSC, 5xDenhardts solution (50xDenhardts solution is 1% ($^{w}/_{v}$) Ficoll 400 (Pharmacia), 1% ($^{w}/_{v}$) polyvinylpyrrolidone, 1% ($^{w}/_{v}$) bovine serum albumin), 0.5% SDS and 20 µg/ml sheared and denatured salmon sperm DNA, in a hybridisation bottle at 65°C for 30 min in a rotating hybridisation oven. 1-5x10^5 cpm of radioactive RNA probe (2.2.12) was added to the prehybridisation solution and incubation continued at 70°C overnight. The hybridisation solution was discarded and the filter washed as follows: Two times in 2xSSC, 0.1% SDS at RT for 10 min; once in 1xSSC, 0.1% SDS at 70°C for 15 min; and once in 0.1xSSC, 0.1% SDS at 70°C for 20 min. The filter was blotted dry and autoradiographed to detect radioactive signal.

CHAPTER THREE

DETERMINATION OF THE COMPLETE NUCLEOTIDE SEQUENCE OF SOYBEAN DWARF VIRUS ISOLATE TAS1

3.1 Introduction

Soybean dwarf virus (SDV) is a luteovirus with a wide host range, predominantly infecting members of the Fabacae (Leguminosae) but not restricted to this family (Damsteegt *et al.*, 1990). As with other luteoviruses it is transmitted by aphids, the common vectors being *Aulacorthum solani* and *Acyrthosiphon pisum.* SDV is economically important in Japan where it causes losses to soybean production, however in the USA and Australia it predominantly infects forage or pasture legumes. Symptoms of SDV infection may include yellowing, leafrolling and dwarfing in peas and beans, while infection of subterranean clover causes a characteristic reddening of the leaves (hence the Australasian synonym subterranean clover red leaf virus (SCLRV; Ashby and Johnstone, 1985)).

Serological relationships have been found between SDV particles and those of most other luteoviruses (D'Arcy, 1986), however the reactions between SDV and BWYV, and SDV and PLRV, are probably the most significant (D'Arcy *et al.*, 1989). Analysis of the double-stranded RNAs (dsRNA) extracted from SDV-infected plants revealed two species (Smith *et al.*, 1991). The larger of these, which probably corresponds to the genomic RNA, was of M_r 3.4 x 10^6 (~5 kb), while the smaller species, probably equivalent to sgRNA 1 of BYDV-PAV, was of M_r 1.9 x 10^6 (~2.8 kb). This compares with the 4-5 species of dsRNA associated with BYDV infection (Gildow *et al.*, 1983).

The aim of the current work was to complete the partial sequence of a Tasmanian strain of SDV (SDV-Tas1) that had already been obtained in this laboratory. Initial data analysis had suggested that SDV was a subgroup I luteovirus (Habili and Symons, 1989), which contrasts with the serological data relating SDV to the subgroup II luteoviruses BWYV and PLRV (D'Arcy *et al.*, 1989). Therefore, by completion of the genomic nucleotide sequence, comparisons of ORF organisation and sequences could be made with other luteoviruses. This would help to establish the relationship between SDV and other luteoviruses, and possibly also to understand the function of the luteoviral genome. At the time this work commenced, BYDV-PAV was the only

subgroup I luteovirus for which the entire genomic RNA sequence had been obtained, which provided further interest in the nucleotide sequence of the SDV genome.

3.2 Materials and Methods

3.2.1 Acknowledgments

This project was formed as a collaboration with Drs Peter Waterhouse and Wayne Gerlach, CSIRO Division of Plant Industry, Canberra, Australia. The majority of SDV-Tas1 genomic cDNA clones (designated pSD; Fig. 3.1) were generated by these workers. Sequencing of the pSD clones was performed in this laboratory, and was done in collaboration with Ms Litsa E Karageorgos, Mr Timothy Hercus and Dr Nuredin Habili.

3.2.2 Purification of SDV-Tas1 genomic RNA

SDV isolate Tas1 (Helms *et al.*, 1983) was transmitted to pea (*Pisum sativum* cv. Sugar Snap) by viruliferous *Aulacorthum solani* aphids. Infected plant material was harvested three weeks post infection and virus purified by a modification of the method of Waterhouse and Helms (1984). Briefly, 50 g of infected tissue was ground to a powder in liquid nitrogen, then placed in a Waring blender with 150 ml of 0.1 M sodium citrate pH 6.0, 0.2% ß-mercaptoethanol, 2% Celluclast (Novo, Denmark), and blended until homogenous. The mixture was incubated overnight at 30°C with shaking, followed by extraction with chloroform:butanol. Virus was precipitated from the supernatant by addition of PEG 6000 to 8% ($^{w}/_{v}$) and NaCl to 0.4 M. The mixture was centrifuged to pellet virus, and the pellet resuspended overnight in 6 ml of sodium phosphate pH 7.6. Virus suspensions were ultracentrifuged at 44,000 rpm in a Beckman Ti-50 rotor for 3 h at 15°C, and the virion pellet resuspended overnight in 6 ml 0.01 M sodium phosphate pH 7.6. Virions were purified by ultracentrifugation as previously through a 20% sucrose cushion, then disrupted by resuspension in 0.5 ml of TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA) containing 2% ($^{w}/_{v}$) SDS, followed by incubation at 60°C for 10 min. RNA was purified by successive phenol:chloroform

extractions of the disrupted virions, followed by precipitation in 0.3 M sodium acetate pH 5.2 and 2.5 volumes of ice-cold ethanol. The RNA pellet was washed in ice-cold 70% ethanol and dried under vacuum. Purity of the RNA was checked by agarose gel electrophoresis (not shown).

3.2.3 Construction of cDNA clones from SDV-Tas1 genomic RNA

cDNA was produced from SDV-Tas1 genomic RNA (3.2.2) using an adaptation of the method of Gubler and Hoffman (1983). Briefly, 5 µg of the purified RNA was annealed to 5 µg of random sequence hexanucleotides, then treated with AMV reverse transcriptase (Promega) according to manufacturer's specifications. First strand cDNA was purified by phenol:chloroform extraction followed by ethanol precipitation, then resuspended in 20 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM $(NH_4)_2SO_4$, 5 mM $MgCl_2$, 1 mM DTT, 0.25 mM dNTPs, 0.1 mg/ml BSA. RNase H was added to 0.02 U/µl, and *E. coli* DNA polymerase I to 0.3 U/µl in a final volume of 100 µl, and the reaction incubated at 15°C for 60 min. Second-strand cDNA was purified by phenol:chloroform extraction followed by ethanol precipitation, then treated with T4 DNA polymerase in the presence of dNTPs to repair ragged ends. The cDNA was purified as before, then ligated to *Bam*HI linkers (New England Biolabs, USA), followed by digestion with *Bam*HI and insertion into pUC9 also digested with this enzyme. Recombinant colonies carrying SDV sequences were selected by colony hybridisation (Waterhouse *et al.*, 1986) using ^{32}P labelled SDV RNA as a probe.

3.2.4 Sequence analysis of pSD clones

pUC9 clones containing SDV cDNA inserts generated above (3.2.3) were subcloned by digestion with *Bam*HI and insertion into M13mp18 or M13mp19 vectors. Single-stranded M13 DNA was prepared prior to sequencing, which was carried out using kits supplied by Bresatec (Australia). Clones too large to sequence by single-pass sequencing were subcloned further to create smaller inserts, or alternatively synthetic oligonucleotides were synthesised to viral sequences previously identified to enable

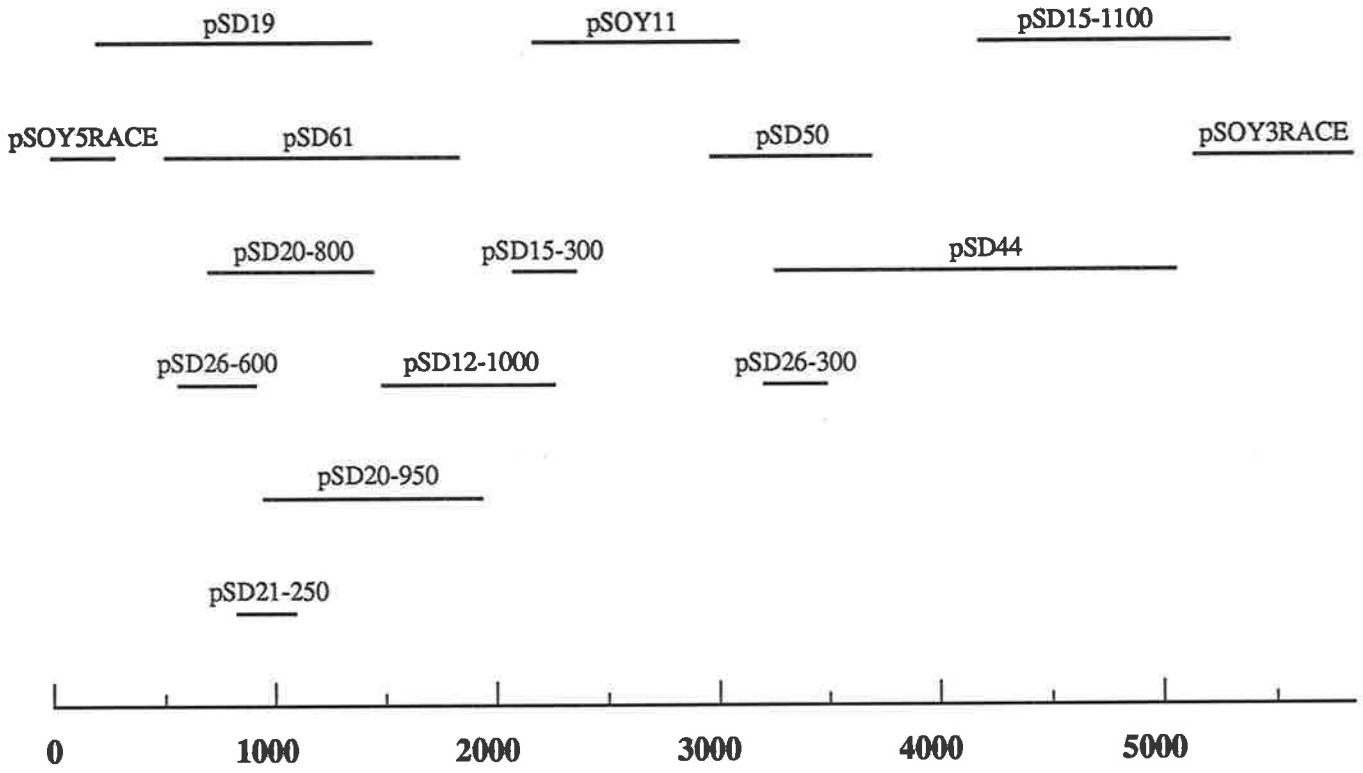
sequencing further along the clone. Both strands of all clones were sequenced. Sequences were compiled and analysed using the computer package of Staden (1980).

3.2.5 Cloning of SDV-Tas1 genomic fragments not represented in the initial cDNA clone population

Cloning of central region. A 1.2 kb DNA fragment covering the central region of the SDV genome that was not represented in the initial cDNA cloning experiment was amplified using standard reverse transcriptase-PCR (RT-PCR) conditions. Briefly, reverse transcription using AMV reverse transcriptase was primed from positive-strand viral RNA using the oligodeoxynucleotide primer SDV 3089 (5'-CTCTCGTAGGGCAGCAAGAC-3'; complementary to residues 3070-3089 of the SDV-Tas1 genome) in a reaction volume of 20 µl. One microlitre of the cDNA product was amplified using *Taq* polymerase and enzymically phosphorylated primers in a PCR reaction employing primer SDV 1853 (5'-ATAGCCAATAAATGGTCCAA-3'; homologous to residues 1853-1872 of the SDV-Tas1 genome) as the second strand primer. Thermocycling for PCR was [94°C/1 min; 50°C/1 min; 72°C/90 s]$_{30}$ and was performed on a DNA Thermal Sequencer (Corbett, Australia). The major PCR product of ~1.2 kb was resolved by agarose gel electrophoresis and the band of interest purified using Geneclean, then blunt-end cloned into the *Sma*I site of M13mp18 to create pSOY01. This clone was restricted with *Sac*I to release a 900 nt fragment which was cloned into the *Sac*I site of pGEM1 (Promega) to give pSOY11. Sequencing of the clone was completed using synthetic oligonucleotides and double-stranded DNA templates.

RACE cloning of the 5' end of SDV-Tas1 genomic RNA. This was performed largely as described by Frohman (1990; see Fig. 3.2 for a description of the RACE procedure). First-strand cDNA was synthesised from the oligodeoxynucleotide SDV 621 (5'-CCTCCTTCTTCTGAATGA-3'; complementary to residues 604-621 of the SDV-Tas1 genome) and purified from the primer and reaction components using a Qiagen TIP-5 column. The cDNA was tailed with dATP using terminal

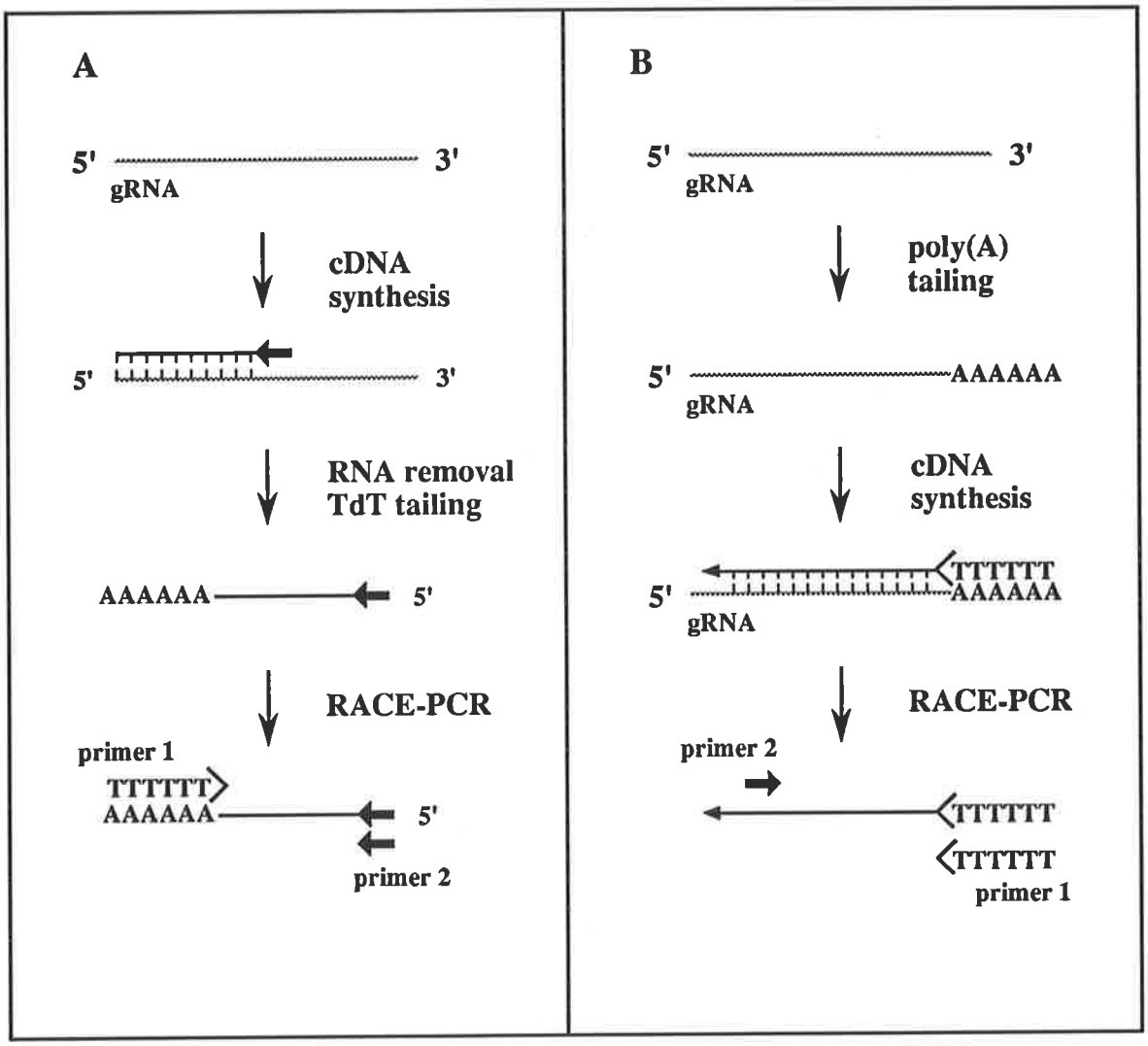
Fig. 3.1. cDNA clones used to determine the sequence of SDV-Tas1. All clones used to derive the nucleotide sequence presented in Fig. 3.3 are shown. The scale indicates the size of the cloned inserts in nucleotides. pSD clones were generated by random priming on SDV-Tas1 genomic RNA, while pSOY clones were constructed using PCR. More than one clone of pSOY5RACE and pSOY3RACE were sequenced (see text).


pSD19
pSOY11
pSD15-1100
pSOY5RACE
pSD61
pSD50
pSOY3RACE
pSD20-800
pSD15-300
pSD44
pSD26-600
pSD12-1000
pSD26-300
pSD20-950
pSD21-250
0
1000
2000
3000
4000
5000

deoxynucleotidyl transferase, then heated at 70°C for 15 min to denature the enzyme. The reaction was diluted to 200 µl with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and 1 µl of this solution was used in the PCR. PCR amplification of the cDNA employed primer SDV 282 (5'-GTGCAGCAAACACGCCTTGGAG-3'; complementary to residues 261-282) as the specific primer and the adaptor-primer ARACE5 (5'-GACTCGAGATCGA[T]$_{17}$-3'). The thermocycling profile for the reaction was [94°C/5 s; 55°C/5 s; 72°C/30 s]$_{45}$, using Vent DNA polymerase in a capillary DNA Thermal Sequencer (Corbett). The single major reaction product of 300 nt was resolved on a 2% agarose minigel and blunt-end cloned into the *Sma*I site of pBluescript SK+ (Stratagene) to create pSOY5RACE. Confirmation of 5' end clones was by sequence analysis to identify overlap with other SDV-Tas1 genomic RNA clones. The entire nucleotide sequence of eight independent clones was determined from both strands using double-stranded DNA templates.

RACE cloning of the 3' end of SDV-Tas1 genomic RNA. Total RNA was isolated from SDV-Tas1 infected pea essentially as described by Dunsmuir *et al.* (1988). One microgram of the RNA was treated with poly(A) polymerase, then reverse transcribed by AMV reverse transcriptase using ARACE5 as the primer. PCR was performed on the first-strand cDNA using primers ARACE5 and SDV 5178 (5'-GGGCATATATCGATGGTTTA-3'; homologous to residues 5178-5197) and Vent DNA polymerase. The reaction profile was [94°C/5 s; 51°C/5 s; 72°C/45 s]$_{40}$ and was carried out on a capillary DNA Thermal Sequencer. Reaction products were cloned into pBluescript as described above to create the plasmid pSOY3RACE, and overlap with other SDV-Tas1 genomic clones confirmed by sequence analysis. Southern analysis was used to confirm that only the ~700 nt product of the RACE reaction contained sequences homologous to SDV-Tas1 genomic RNA (data not shown). Two independent clones corresponding to the SDV-Tas1 3' 700 nt were sequenced after subcloning to reduce the size of the inserts. Sequencing was completed using double-stranded DNA templates.

Fig. 3.2. Rapid Amplification of cDNA Ends (RACE) -PCR for the amplification of termini of RNA molecules. (A) 5' end determination. cDNA synthesis is primed from the genomic RNA (gRNA; stippled line) with a sequence specific primer (thick black arrow). The cDNA is purified from the RNA template and excess first-strand primer, then tailed with terminal deoxynucleotidyl transferase (TdT) and dATP. The tailed cDNA can then be used as a template for PCR, using a non-specific $d(T)_n$ primer, and either the first-strand primer or an internal sequence specific primer. (B) 3' end determination. Luteovirus RNAs are not polyadenylated, so this must be done *in vitro* before cDNA synthesis (using a non-specific $d(T)_n$ primer) can be performed. Thereafter PCR is carried out with an internal sequence-specific primer and the non-specific first-strand primer.


A
5'
3'
gRNA
cDNA
synthesis
5'
3'
RNA removal
TdT tailing
AAAAAA
5'
RACE-PCR
primer 1
TTTTTT
AAAAAA
5'
primer 2
B
5'
3'
gRNA
poly(A)
tailing
5'
AAAAAA
gRNA
cDNA
synthesis
TTTTTT
5'
AAAAAA
gRNA
RACE-PCR
primer 2
TTTTTT
TTTTTT
primer 1

3.2.6 Computer analysis of the nucleotide sequence of SDV-Tas1

Open reading frames were detected using the computer program DNA Strider version 1.1. Alignment of nucleotide sequences was performed using the UWGCG program GAP (Devereux *et al.*, 1994) using default values of 3.00 for gap weight and 0.10 for gap length weight.

3.2.7 Cloning of the 3' end of the SDV-AP1 genome

Total RNA was extracted from a whole young subterranean clover plant infected with SDV isolate AP1 (a kind gift of Dr G Johnstone, Department of Primary Industries, Tasmania, Australia) according to the method of Maes and Messens (1992). One hundred nanograms of the RNA was reverse transcribed by AMV reverse transcriptase using SDV3TERM (5'-GGGGCAGGTGGACACAAAG-3'; complementary to residues 5843-5861 of the SDV-Tas1 genome) as the first strand primer in a reaction volume of 20 µl. One microlitre of the cDNA was PCR amplified by Vent polymerase in a 20 µl reaction employing SDV 5178 as the second strand primer. The reaction profile was [94°C/5 s; 49°C/5 s; 72°C/30 s]$_{40}$, and was performed on a capillary DNA Thermal Sequencer. The major reaction product of 680 nt was gel purified and cloned into pBluescript as described above to create pSAP-700. The insert was further subcloned and sequenced in its entirety using double-stranded DNA templates.

3.3 Results

3.3.1 Cloning and sequencing of the SDV-Tas1 genome

Clones covering the full-length nucleotide sequence of the SDV-Tas1 genome were generated here by a mixture of random and directed approaches (Fig. 3.1). The initial cloning procedure used random hexanucleotides to prime first-strand cDNA synthesis. Twelve independent clones covering the large proportion of the genome were generated, leaving gaps in the centre of the genome and at the genomic termini. Regions of the genome not represented in the pool of random cDNA clones were

obtained by PCR, with either single- (genomic termini; Fig. 3.2) or double-sided (central genomic region) specificity. Approximately 80% of the genome was covered by more than one cDNA clone (Fig. 3.1), including the genomic termini where multiple clones (eight at the 5' end, two at the 3' end) were sequenced to ensure fidelity of the sequence.

Sequencing of both strands of all clones, including multiple clones at the 5' and 3' genomic termini, revealed a total nucleotide sequence of 5861 nt (Fig. 3.3). The total length of the sequence is dependent on the length of clones of the 5' and 3' genomic termini, which varied in length by one nucleotide (3.3.2). Therefore, the longest sequence present amongst the clones sequenced was taken to represent the full-length sequence. Analysis of the terminal sequences presented in Chapter Four provides evidence that this determination of the 5' and 3' genomic ends is correct. No nucleotide variation between clones was observed, despite the fact that many overlapping clones were sequenced.

3.3.2 Determination of the 5' and 3' terminal sequences of the SDV-Tas1 genome

The PCR-based technique Rapid Amplification of cDNA Ends (RACE; Frohman, 1990) was used to determine the nucleotide sequences of both the 5' and 3' genomic termini of SDV-Tas1 RNA. This technique employs PCR amplification of first-strand cDNA with only one sequence-specific primer (Fig. 3.2), which means that the end of any RNA molecule can be amplified so long as sequence information exists internally of the terminus in question. Thus, first-strand cDNA synthesised from a sequence-specific primer at the 5' end of the genomic RNA must first be tailed with poly(dATP), before synthesis of the second cDNA strand and subsequent PCR amplification using a non-specific $d(T)_n$ oligonucleotide as the second-strand primer. Similarly, for amplification of the 3' terminus, the virus genomic RNA must be tailed with poly(ATP) using *E. coli* poly(A) polymerase prior to synthesis of the first-strand cDNA, which is primed non-specifically with a $d(T)_n$ oligonucleotide. PCR is then carried out using the $d(T)_n$ primer and an internal sequence specific primer.

Fig. 3.3. Nucleotide sequence of SDV-Tas1 genomic RNA. Numbers refer to nucleotide positions in the genome. Translations of the five major ORFs defined in Fig. 3.4A are given, with amino acid sequences as single letter abbreviations. The number of each ORF is indicated on the left hand side of the diagram.

```
   1  AGUAAAGUUGACACCUUUACAGAAGUGGUCUUACUUGUAAGAGUUAACUCAUCAAGAGUUAAUAUAAGAUCCACCUCCCG   80
             10                  30                  50                  70

                                                                  ORF1  M   F   N   F   D   S
  81  GCACCUUCGUAUCGUGUUUGAGGUAUCUCUAGUGUUUGGUUGUUUAAAUACUAGCUAAUAUUCAUGUUUAAUUUCGAUAG   160
             90                  110                 130                 150

ORF1  L   V   S   A   T   A   K   V   V   K   D   F   I   H   F   C   Y   N   R   A   R   H   V   Y   Y   A   L
 161  UUUAGUGUCGGCCACCGCCAAGGUGGUCAAAGAUUUUAUUCAUUUUUGUUAUAAUAGGGCGAGGCACGUAUAUUAUGCCC   240
             170                 190                 210                 230

ORF1   K   R   W   L   W   E   L   Q   G   V   F   A   A   H   D   A   F   V   D   M   C   Y   D   A   M   Y
 241  UCAAACGCUGGCUUUGGGAACUCCAAGGCGUGUUUGCUGCACAUGAUGCCUUUGUGGACAUGUGCUACGACGCCAUGUAU   320
             250                 270                 290                 310

ORF1  G   V   E   E   F   E   W   E   L   Q   K   Q   F   S   S   A   E   H   D   V   L   I   A   K   H   E   F
 321  GGCGUCGAGGAGUUUGAGUGGGAGUUGCAAAAGCAAUUCUCCAGUGCCGAACAUGAUGUGCUCAUCGCCAAGCACGAAUU   400
             330                 350                 370                 390

ORF1  E   R   L   L   K   D   G   A   P   L   R   T   W   P   Q   P   C   A   P   L   G   S   F   R   S   S   D
 401  UGAGCGCCUAUUGAAAGAUGGCGCCCCAUUGAGGACAUGGCCACAACCAUGUGCUCCUUUGGGUAGUUUCCGGUCGUCUG   480
             410                 430                 450                 470

ORF1   D   F   Q   E   A   A   R   E   V   K   T   V   L   D   G   P   E   P   S   L   I   K   G   S   G   D
 481  ACGACUUCCAAGAAGCUGCCAGGGAAGUGAAAACUGUCCUUGAUGGACCUGAACCCUCCUUGAUUAAAGGGUCAGGAGAU   560
             490                 510                 530                 550

ORF1  Y   S   L   D   N   P   N   R   I   E   K   F   I   N   L   I   Q   K   K   E   V   L   S   A   T   E   R
 561  UACUCACUUGACAAUCCUAACCGGAUUGAAAAGUUUAUCAACCUCAUUCAGAAGAAGGAGGUACUUUCCGCCACCGAGCG   640
             570                 590                 610                 630

ORF1   M   I   K   H   A   Y   E   E   H   I   G   E   A   P   F   G   K   W   F   N   T   L   P   S   R   M   D
 641  AAUGAUCAAACAUGCUUAUGAAGAGCAUAUCGGUGAGGCACCAUUCGGAAAGUGGUUCAACACUUUGCCCUCUCGUAUGG   720
             650                 670                 690                 710

ORF1   Y   I   K   R   A   A   S   K   R   A   K   A   A   K   R   S   N   S   I   R   Q   M   V   E   E   V
 721  ACUACAUCAAGAGGGCCGCUUCAAAGAGAGCCAAGGCGGCUAAAAGAUCCAACUCUAUCCGCCAAAUGGUAGAAGAGGUA   800
             730                 750                 770                 790

ORF1  N   V   I   P   D   F   I   S   I   C   D   V   V   Q   V   D   T   G   E   K   L   P   P   K   K   D   K
 801  AAUGUCAUUCCUGACUUUAUCUCAAUAUGUGAUGUUGUCCAGGUGGACACAGGUGAGAAACUCCCCCCCAAAGAAAGACAA   880
             810                 830                 850                 870

ORF1   D   G   E   P   M   E   P   E   P   K   L   K   M   V   R   R   V   R   F   E   H   Y   G   D   A   R   K
 881  AGAUGGGGAGCCAAUGGAACCUGAACCCAAACUAAAAAAUGGUGAGAAGGGUUAGGUUCGAGCAUUAUGGUGAUGCUCGUA   960
             890                 910                 930                 950

ORF1   Y   I   R   Q   H   I   R   N   N   N   M   R   L   T   D   G   S   D   V   S   H   A   T   I   N   R
 961  AGUACAUAAGACAGCAUAUUCGCAACAACAACAUGCGUCUUACUGACGGCUCUGAUGUUAGUCAUGCUACCAUCAACCGC   1040
             970                 990                 1010                1030

ORF1  Y   A   L   K   F   C   E   D   L   E   L   D   M   T   S   T   C   Y   A   G   D   Y   A   M   T   M   V
1041  UAUGCUCUUAAGUUUUGCGAAGACUUGGAGCUUGAUAUGACCAGUACUUGCUAUGCUGGUGAUUAUGCUAUGACAAUGGU   1120
            1050                1070                1090                1110

ORF1   P   I   P   L   K   N   D   I   E   R   A   K   I   V   H   S   P   A   A   R   Q   I   R   Q   E   L   G
1121  GCCCAUUCCCCUCAAGAAUGACAUUGAAAGGGCGAAGAUUGUCCAUUCUCCUGCAGCCAGGCAAAUCAGGCAGGAGUUGG   1200
            1130                1150                1170                1190

ORF1   V   L   N   A   E   V   F
ORF2               ORF2  G   F   L   E   G   L   C   S   D   S   G   F   E   S   P   F   S   I   L   G   L   P
1201  GCGUUCUCAACGCUGAGGUUUUUUAGAGGGGCUCUGCUCCGACUCUGGUUUUGAAUCCCCGUUUUCUAUUUUGGGGUUAC   1280
            1210                1230                1250                1270

ORF2   E   I   V   V   R   S   G   A   A   P   R   K   S   R   S   V   I   S   F   L   S   Q   F   T   L   G
1281  CAGAGAUCGUGGUUCGCAGUGGAGCUGCACCUAGGAAGAGUCGUAGUGUAAUUAGUUUUUUAUCGCAGUUUACCUUAGGU   1360
            1290                1310                1330                1350

ORF2  L   D   Y   Q   C   P   N   P   S   L   H   N   A   L   V   A   V   E   R   R   V   F   T   V   G   K   G
1361  CUAGAUUAUCAAUGCCCAAAUCCCAGUUUACACAAUGCAUUGGUGGCGGUUGAACGACGUGUGUUCACCGUUGGGAAAGG   1440
            1370                1390                1410                1430

ORF2   N   E   V   V   L   P   Y   K   N   K   P   G   I   F   S   N   L   D   Y   F   R   D   S   I   V   N   K
1441  AAAUGAGGUAGUGCUACCUUACAAGAACAAACCAGGAAUAUUUUCCAAUCUUGAUUACUUUAGAGACUCAAUUGUCAACA   1520
            1450                1470                1490                1510

ORF2   V   G   C   P   R   T   H   T   P   E   E   L   A   A   T   Y   H   S   G   K   R   S   L   Y   N   A
1521  AAGUUGGCUGUCCGAGGACCCACACUCCUGAGGAACUUGCUGCAACGUACCACUCUGGAAAGAGAAGUUUGUAUAAUGCU   1600
            1530                1550                1570                1590

ORF2  A   V   Q   S   L   K   K   K   A   V   E   R   S   D   A   N   V   T   A   F   L   K   M   E   K   H   L
1601  GCAGUUCAAAGCCUCAAAAAGAAGGCAGUCGAAAGGAGUGAUGCCAAUGUGACAGCUUUUCUCAAGAUGGAAAAACAUUU   1680
            1610                1630                1650                1670

ORF2   M   S   K   K   I   A   P   R   L   I   C   P   R   N   K   R   Y   N   V   E   L   G   R   R   L   K   F
1681  AAUGAGUAAGAAGAUAGCACCCAGGUUGAUAUGUCCCCGCAACAAACGGUAUAAUGUUGAAUUGGGACGUCGAUUGAAGU   1760
            1690                1710                1730                1750

ORF2   N   E   K   K   F   M   H   A   I   D   S   T   F   D   S   P   T   V   L   S   G   Y   D   S   F   R
1761  UCAAUGAGAAGAAAUUUAUGCAUGCAAUCGACUCAACCUUUGAUUCCCCAACUGUUCUUAGUGGAUAUGACAGUUUCAGA   1840
            1770                1790                1810                1830

ORF2  V   G   K   I   I   A   N   K   W   S   K   F   K   R   P   V   A   I   G   V   D   A   S   R   F   D   Q
1841  GUUGGGAAGAUAAUAGCCAAUAAAUGGUCCAAAUUCAAGAGACCAGUUGCAAUAGGUGUUGAUGCGAGCAGAUUUGAUCA   1920
            1850                1870                1890                1910

ORF2   H   V   G   V   E   A   L   Q   W   E   H   S   I   Y   N   G   A   F   K   D   P   I   L   K   E   L   L
1921  ACAUGUGGGGGUAGAAGCACUCCAAUGGGAGCACUCAAUUUACAACGGUGCAUUCAAAGAUCCCAUUCUUAAGGAGUUGC   2000
            1930                1950                1970                1990
```

```
 ORF2   H  W  Q  T  E  N  R  I  M  L  F  V  E  D  K  I  L  K  F  K  V  K  G  H  R  M
2001  UACACUGGCAAACAGAGAAUAGAAUUAUGCUGUUUGUUGAAGAUAAAAUCCUCAAGUUCAAGGUCAAAGGACAUAGAAUG  2080
           2010                 2030                 2050                 2070

 ORF2 S  G  D  I  N  T  S  S  G  N  K  L  I  M  C  G  M  M  H  Y  Y  F  K  T  L  G  V
2081  UCCGGCGACAUUAACACCUCUUCUGGCAACAAAUUAAUCAUGUGUGGUAUGAUGCACUACUACUUCAAAACUCUUGGAGU  2160
           2090                 2110                 2130                 2150

 ORF2   K  A  E  L  C  N  N  G  D  D  C  V  I  I  C  E  R  K  D  E  N  K  F  Q  H  M  H
2161  CAAAGCCGAGCUCUGCAAUAACGGCGAUGAUUGUGUUAUCAUAUGCGAGCGGAAAGAUGAGAACAAAUUCCAACACAUGC  2240
           2170                 2190                 2210                 2230

 ORF2   S  W  F  K  D  Y  G  F  D  M  Q  I  E  T  P  V  Y  K  I  G  Q  I  E  F  C  Q
2241  ACAGCUGGUUUAAAGACUAUGGGUUUGACAUGCAGAUUGAGACUCCUGUCUACAAGAUUGGACAGAUAGAGUUUUGUCAA  2320
           2250                 2270                 2290                 2310

 ORF2 S  K  P  V  K  I  N  G  Y  Y  R  M  V  R  K  P  E  S  I  S  K  D  A  H  S  L  I
2321  AGUAAACCAGUUAAAAAUUAAUGGCUAUUAUAGGAUGGUGCGUAAACCAGAGAGCAUAUCCAAAGACGCCCAUUCCCUCAU  2400
           2330                 2350                 2370                 2390

 ORF2   S  M  A  S  A  E  D  V  K  T  F  M  S  A  T  A  Q  C  G  M  I  L  N  S  G  V  P
2401  UUCGAUGGCAUCAGCUGAAGAUGUCAAAACUUUCAUGAGUGCAACCGCCCAGUGCGGUAUGAUCUUGAAUAGUGGUGUAC  2480
           2410                 2430                 2450                 2470

 ORF2   V  L  D  A  F  H  K  C  L  F  K  A  S  G  Y  K  K  V  S  Q  E  A  I  E  R  I
2481  CUGUUCUUGAUGCUUUUCACAAAUGCCUAUUUAAAGCAUCGGGGUAUAAGAAAGUAUCGCAAGAAGCUAUCGAAAGGAUA  2560
           2490                 2510                 2530                 2550

 ORF2 V  S  F  G  T  Q  D  K  L  G  L  R  K  E  R  V  E  E  P  I  T  M  D  N  R  L  S
2561  GUCUCUUUUGGAACACAGGACAAACUCGGGCUCAGAAAAGAGCGAGUCGAAGAACCAAUAACAAUGGAUAACAGGUUGAG  2640
           2570                 2590                 2610                 2630

 ORF2   Y  W  E  S  S  G  V  D  P  Q  T  Q  V  L  V  E  R  Y  F  D  N  L  T  V  H  I  E
2641  UUAUUGGGAGUCCAGUGGGGUUGACCCUCAGACACAGGUCCUUGUCGAACGAUAUUUUGACAAUCUGACGGUCCAUAUCG  2720
           2650                 2670                 2690                 2710

 ORF2   P  R  G  V  K  R  L  T  P  L  L  D  K  T  L  L  S  I  A  S  V  A  R  K  S  V
2721  AACCCCGCGGUGUAAAGAGAUUGACGCCUUUACUAGAUAAAACUUUGCUUUCAAUUGCGAGUGUAGCACGAAAAUCUGUG  2800
           2730                 2750                 2770                 2790

 ORF2 S  L  P  I  L  S  K
2801  UCGCUACCAAUACUUUCAAAAUAGAAGCACUCAUAGUAACCAUAGUAAUAAUAUUAGUAGUAGUAAUUAGUUUAUUAGAU  2880
           2810                 2830                 2850                 2870

2881  UACGCUAGUUCGUACAGAUAAAUUGCAACGGAUUAUAAAUUCCUAUCAGGUUUUGCUUCUGGAUUUAUUUCCAGCAUCCC  2960
           2890                 2910                 2930                 2950

                                                                        ORF3  M  V  A
2961  UGUAUCAAUAUUGGCAAUUUAUUUUGUUUACCUAAAAAUUUCAAAGAAUCUGCGCGAGAUAAUCAACGAGUAUGGUCGCG  3040
           2970                 2990                 3010                 3030

 ORF3 V  S  N  V  A  I  Q  R  R  R  S  R  R  A  A  R  R  A  P  R  V  Q  L  M  A  V  P
 ORF4       M  S  Q  Y  N  D  D  A  L  V  G  Q  Q  D  A  L  Q  E  F  S  S  W  L  F  Q
3041  GUUAGCAAUGUCGCAAUACAACGACGACGCUCUCGUAGGGCAGCAAGACGCGCUCCAAGAGUUCAGCUCAUGGCUGUUCC  3120
           3050                 3070                 3090                 3110

 ORF3  T  A  T  S  R  P  Q  R  R  G  R  Q  R  R  R  R  R  N  N  R  G  G  S  F  V  S  G
 ORF4   R  P  P  A  D  H  N  A  E  D  D  N  D  D  E  G  E  I  I  E  E  E  A  L  F  P
3121  AACGGCCACCAGCCGACCACAACGCAGAGGACGACAACGACGACGAAGGAGAAAUAAUAGAGGAGGAAGCUUUGUUUCCG  3200
           3130                 3150                 3170                 3190

 ORF3   G  S  G  K  A  H  T  F  V  F  S  K  D  G  I  N  G  S  S  K  G  S  I  T  F  G
 ORF4 E  D  Q  A  R  L  T  H  S  C  F  Q  R  T  A  S  M  V  V  P  R  E  V  S  L  S  G
3201  GAGGAUCAGGCAAGGCUCACACAUUCGUGUUUUCAAAGGACGGCAUCAAUGGUAGUUCCAAGGGAAGUAUCACUUUCGGG  3280
           3210                 3230                 3250                 3270

 ORF3 P  S  L  S  E  C  K  P  F  S  D  G  I  L  K  A  Y  H  E  Y  K  I  T  S  V  L  L
 ORF4  R  L  Y  Q  N  A  S  H  S  L  M  E  Y  S  R  P  T  M  N  I  R  S  R  V  S  Y  Y
3281  CCGUCUUUAUCAGAAUGCAAGCCAUUCUCUGAUGGAAUACUCAAGGCCUACCAUGAAUAUAAGAUCACGAGUGUCUUAUU  3360
           3290                 3310                 3330                 3350

 ORF3   Q  F  I  T  E  A  S  S  T  S  S  G  S  I  A  Y  E  L  D  P  H  C  K  Y  S  E  I
 ORF4   S  S  S  P  R  P  L  P  P  R  Q  V  P  S  L  M  N  L  T  H  T  A  S  T  P  K
3361  ACAGUUCAUCACCGAGGCCUCUUCCACCUCGUCAGGUUCCAUCGCUUAUGAACUUGACCCACACUGCAAGUACUCCGAAA  3440
           3370                 3390                 3410                 3430

 ORF3   Q  S  L  L  N  K  F  S  I  T  K  S  G  S  K  R  F  P  T  R  A  I  N  G  L  E
 ORF4 F  N  R  Y  S  I  N  S  V  S  Q  R  A  V  R  N  V  S  Q  P  E  L  S  M  A  S  N
3441  UUCAAUCGCUACUCAAUAAAUUCAGUAUCACAAAGAGCGGUUCGAAACGUUUCCCAACCAGAGCUAUCAAUGGCCUCGAA  3520
           3450                 3470                 3490                 3510

 ORF3 W  H  D  T  S  E  D  Q  F  K  I  H  Y  K  G  N  G  E  S  K  I  A  G  S  F  K  I
 ORF4  G  M  I  P  V  R  I  N  S  R  S  T  I  K  G  T  E  S  P  R  S  Q  A  P  S  R  S
3521  UGGCAUGAUACCAGUGAGGAUCAAUUCAAGAUCCACUAUAAAGGGAACGGAGAGUCCAAGAUCGCAGGCUCCUUCAAGAU  3600
           3530                 3550                 3570                 3590

 ORF3   S  I  N  V  L  T  Q  N  A  K  *  V  D  G  E  P  G  P  K  P  G  P  D  P  A  P  Q
 ORF4   R  S  M  S
3601  CUCGAUCAAUGUCCUAACGCAGAAUGCUAAAUAGGUAGACGGAGAACCGGGUCCGAAACCAGGUCCAGACCCAGCACCCC  3680
           3610                 3630                 3650                 3670

 ORF5   P  T  P  T  P  K  P  T  P  A  K  H  E  R  F  I  A  Y  T  G  T  L  S  T  L  I
3681  AACCAACACCAACACCUAAACCAACGCCAGCCAAACACGAGAGGUUUAUCGCGUACACCGGUACGUUGUCUACACUCAUU  3760
           3690                 3710                 3730                 3750

 ORF5 S  A  R  Q  S  S  D  S  I  S  L  Y  S  I  R  N  Q  R  I  R  Y  I  E  D  E  N  S
3761  AGUGCUAGGCAGUCUUCUGAUAGCAUCUCCUUAUAUUCCAUUCGGAAUCAGAGGAUCAGAUACAUUGAAGAUGAAAACUC  3840
           3770                 3790                 3810                 3830
```

```
ORF5  S  W  T  N  I  D  A  K  W  Y  S  Q  N  S  V  E  A  I  P  M  F  V  Y  P  V  P  E
3841  CAGUUGGACAAACAUAGAUGCAAAAUGGUACUCACAGAACUCUGUUGAGGCUAUUCCAAUGUUUGUUUAUCCAGUACCUG  3920
            3850                3870                3890                3910

ORF5   G  T  W  S  I  E  I  S  C  E  G  Y  Q  A  A  S  S  T  S  D  P  H  R  G  K  C
3921  AAGGUACUUGGUCAAUUGAGAUUUCUUGUGAAGGCUACCAGGCAGCGUCUAGCACUUCAGACCCGCAUCGUGGAAAAUGC  4000
            3930                3950                3970                3990

ORF5  D  G  M  I  A  Y  D  D  D  S  S  K  V  W  N  V  G  Q  Q  N  N  V  T  I  T  N  N
4001  GAUGGCAUGAUUGCUUAUGACGACGAUUCAUCCAAGGUGUGGAAUGUUGGCCAGCAGAAUAAUGUAACCAUAACCAACAA  4080
            4010                4030                4050                4070

ORF5  K  A  D  N  D  W  K  Y  G  H  P  D  P  L  D  L  M  I  N  G  D  R  F  D  Q  N  Q
4081  CAAGGCCGAUAAUGAUUGGAAGUAUGGCCACCCAGAUCCUCUAGAUCUGAUGAUCAAUGGUGACAGAUUCGAUCAAAAUC  4160
            4090                4110                4130                4150

ORF5   V  V  E  K  D  G  I  I  S  F  H  L  V  T  T  G  P  N  A  S  F  F  L  V  A  P
4161  AAGUAGUCGAGAAAGAUGGAAUUAUCUCAUUCCAUUUAGUAACUACUGGCCCCAAUGCGAGUUUCUUUCUUGUCGCUCCA  4240
            4170                4190                4210                4230

ORF5  A  V  K  K  T  A  K  Y  N  F  C  V  S  Y  G  D  W  T  D  R  D  M  E  F  G  M  V
4241  GCUGUUAAGAAAACAGCCAAAUACAACUUUUGUGUUUCUUACGGGGAUUGGACAGAUAGAGAUAUGGAAUUUGGUAUGGU  4320
            4250                4270                4290                4310

ORF5  S  V  V  L  D  E  H  L  E  G  A  R  S  S  Q  Y  V  R  K  S  P  R  P  G  H  F  G
4321  AUCGGUGGUGCUUGAUGAGCACUUAGAAGGUGCAAGGAGUUCUCAGUAUGUUAGAAAAUCGCCAAGACCAGGUCAUUUUG  4400
            4330                4350                4370                4390

ORF5   V  N  R  S  H  R  L  Q  D  S  F  T  P  V  E  Y  V  S  D  D  D  S  S  S  S  S
4401  GCGUCAAUCGCUCUCACCGAUUGCAAGAUAGUUUCACUCCUGUGGAAUAUGUCAGUGAUGAUGAUAGUAGUAGUAGU  4480
            4410                4430                4450                4470

ORF5  S  I  V  S  N  R  P  S  T  P  D  N  D  S  D  I  Q  F  A  N  S  L  K  G  K  L  P
4481  AGUAUAGUUAGCAAUCGACCAUCAACUCCGGACAAUGAUAGUGAUAUCCAGUUCGCGAAUUCGUUGAAAGGUAAGCUUCC  4560
            4490                4510                4530                4550

ORF5  S  Q  T  K  L  P  P  K  G  F  Q  S  R  L  S  A  R  E  K  E  E  I  S  K  S  K  P
4561  GUCUCAGACGAAACUUCCUCCCAAAGGAUUCCAAUCACGGUUAAGCGCAAGAGAAAAAGAAGAGAUUUCAAAAUCAAAGC  4640
            4570                4590                4610                4630

ORF5   S  N  V  E  R  Q  V  G  P  L  V  D  A  Y  G  Y  P  S  Q  T  G  V  Y  D  A  A
4641  CUUCAAAUGUUGAACGUCAAGUUGGUCCUCUAGUUGAUGCAUAUGGAUACCCUUCCCAGACGGGUGUUUACGAUGCAGCU  4720
            4650                4670                4690                4710

ORF5  R  E  I  L  Q  S  K  E  A  A  E  N  L  A  E  L  E  K  D  L  K  E  I  N  K  L  E
4721  AGGGAGAUUCUGCAGUCUAAAGAGGCUGCAGAAAACUUGGCAGAAUUAGAGAAGGAUCUAAAAGAAAUCAACAAGCUGGA  4800
            4730                4750                4770                4790

ORF5   P  P  D  T  I  E  Q  E  E  I  P  D  F  V  A  P  S  E  R  V  I  A  E  D  D  R  D
4801  ACCGCCGGAUACAAUCGAACAGGAAGAAAUCCCUGAUUUCGUAGCCCCAUCGGAGAGAGUGAUUGCGGAAGAUGAUCGCG  4880
            4810                4830                4850                4870

ORF5   Y  V  P  S  I  W  R  N  A  D  Q  A  V  V  I  S  S  F  E  P  T  D  W  S  R  P
4881  AUUACGUUCCUUCAAUUUGGCGUAACGCCGACCAAGCUGUGGUAAUCUCAUCUUUUGAGCCUACAGAUUGGAGUCGGCCA  4960
            4890                4910                4930                4950

ORF5  A  Y  E  S  G  D  P  P  K  K  A  G  L  L  K  G  T  L  S  K  L  G  G  S  L  R  S
4961  GCUUAUGAGUCAGGUGAUCCUCCUAAGAAGGCAGGUUUGCUCAAAGGCACCCUCUCAAAAUUGGGAGGGUCGCUUAGAAG  5040
            4970                4990                5010                5030

ORF5  G  E  S  S  L  R  G  N  L  R  K  T  Q  D  Q  S  D  L  E  Y  K  L  S  R  L  N  V
5041  UGGCGAGUCGUCUCUUAGAGGUAACCUCCGCAAGACGCAAGAUCAAAGUGAUCUAGAGUACAAGUUGUCUAGACUCAAUG  5120
            5050                5070                5090                5110

ORF5   P  Q  R  S  Q  Y  Q  R  I  L  A  N  L  G  K  V  R  A  R  A  Y  I  D  G  L  D
5121  UUCCGCAGCGAUCCCAAUACCAACGGAUCUUGGCGAACCUUGGGAAAGUGCGUGCGCGGGCAUAUAUCGAUGGUUUAGAU  5200
            5130                5150                5170                5190

ORF5  L  V
5201  UUAGUUUAAUAUAAAUAGAAGUAAGUAUGCGCUUUCCAUCACAUUUAGUUGGAGCCAACCUUGGUGAGUUGCUGUCGUGG  5280
            5210                5230                5250                5270

5281  GAUCCUGGGAAACAGGUUCGGUGAACAAAAAGCCGGUAAUCUCGGGCGUUUGUCAGACGCCGUCGCAUCAACACAGUAUG  5360
            5290                5310                5330                5350

5361  CAUAUCCAUCAAUAACCCAACGUUCUUGUGACUAAAAAUUAAAGUUAGUUUCCGCCUCUCUUAUAAACCACAUAUAAGAU  5440
            5370                5390                5410                5430

5441  GCGAGCUUGUGGAGAGUACCGUCCCUUGAAUAAGGAUGGUUGAAAGUUUCUGGUUUCAAUCCAUUGAUCAAAAUUGACAC  5520
            5450                5470                5490                5510

5521  UGCUUCUGGUGACUACACUGCCGCUGGCGAUACAGCAUAAACAAUUUCCACUACUUGUUCGAUGAAAUGAGAACAAUUAG  5600
            5530                5550                5570                5590

5601  UAAAGUCUUGCCGGUCACAUGAUUAUUUGGCGGUGCUACCGUUCGCCCCUUUGAUGGCGUUAAAUGAAAGCGGAGUAAAC  5680
            5610                5630                5650                5670

5681  CGACAUCACCUCACCAGGUUGGGUUAACCUGGGGAAGCCGGGAAAAUGCCCGGAACACUUGGUUUAGUGGAUUUUAAACU  5760
            5690                5710                5730                5750

5761  CGAGGGCUCUUCCUCGCUAAAAAGGGGGUAGCCAGGUUGAAAACUGGCAUUAGGGAGUGAUCCCUCUUUGUUGUUAAGGU  5840
            5770                5790                5810                5830

5841  GGCUUUGUGUCCACCUGCCCC  5861
            5850
```

Selection of RACE products for cloning differed for the 5' and 3' ends. RACE of the 5' end of SDV-Tas1 generated only one band, of similar size (~300 nt) to that predicted by comparison of the 5' sequence of SDV-Tas1 and BYDV-PAV-Vic (data not shown). This band was cloned and found to overlap the existing sequence at the 5' end by the predicted 73 nt. RACE of the 3' end generated multiple bands between ~700 and ~1000 nt. Southern analysis of the RACE products using the cloned 700 nt RACE cDNA as probe showed that only this band contained SDV-Tas1 3' genomic sequences (data not shown). Sequencing of the cloned 700 nt band revealed the expected overlap of the existing sequence by 100 nt at the 3' end. Multiple clones obtained from the 5' and 3' RACE reactions were sequenced in order to determine the precise termini of the viral genome. Of these, sequence variation occurred predominantly as the presence or absence of the terminal nucleotide in both the 5' and 3' RACE reactions (data not shown). The longest sequence detected in RACE-PCR was assumed to represent the true terminus of the genomic RNA. About 50% of the clones used to determine the 5' and 3' ends contained the complete terminal sequence defined in this way. However, it cannot be excluded that the 5' terminal nucleotide is a U (or a run of U's), or that the 3' terminal nucleotide is an A (or a run of A's), because of the RACE strategy used to obtain these results.

3.3.3 Genome organisation of SDV-Tas1

Translation of the positive-sense strand of the SDV-Tas1 genomic RNA in all three reading frames reveals five major ORFs arranged in two groups (Fig. 3.4A). This genome organisation is similar to that of subgroup I of the luteoviruses (Figs 3.4B and 4C), except that SDV lacks an ORF corresponding to ORF 6 of the subgroup I genome (see below). ORF 1 of SDV-Tas1 begins after a 5' leader sequence of 143 nt, potentially encoding a protein of M_r 40 K (Fig. 3.4B). There is no homologue to ORF 0 of luteovirus subgroup II in the SDV-Tas1 genome (Fig. 3.4C). ORF 2 is overlapped in a different reading frame by ORF 1 over 7 nt. The coding sequence of ORF 2 specifies a protein product of M_r 59 K between successive in-frame stop codons,

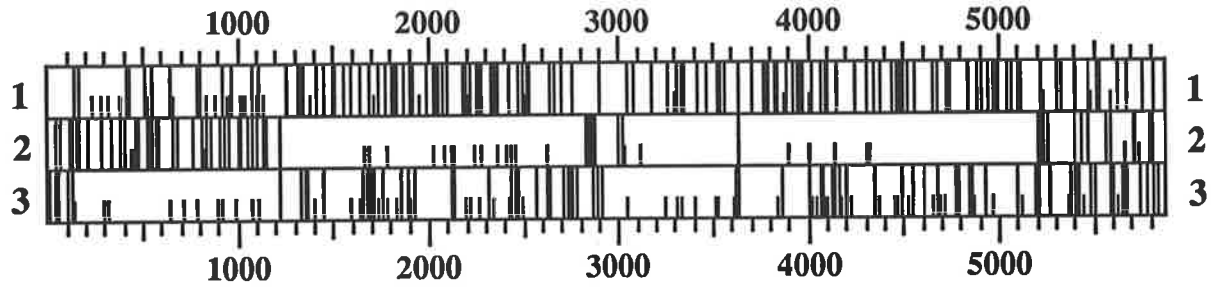
although by analogy to BYDV-PAV, ORF 2 is likely to be expressed as a M_r ~99 K frameshift product of the first and second reading frames, rather than as a separate entity by internal initiation of translation on the genomic RNA. A non-coding intergenic sequence of 210 nt separates ORFs 1 and 2 from the second block of coding sequence. The characteristic arrangement of the 3' block of three genes common to all luteoviruses (ORFs 3, 4 and 5; Martin *et al.*, 1990) is conserved in SDV (Figs 3.4B and 4C). ORF 3, which is known to encode the coat protein in the luteoviruses, is the first ORF to initiate after the non-coding sequence and potentially encodes a M_r 22 K protein. ORF 4 is completely contained within the coat protein gene, extending for 567 nt thus encoding a protein of M_r 21 K. ORF 5 is contiguous and in-frame with ORF 3, but separated by a UAG (amber) stop codon. The reading frame specifies a protein product of M_r 48 K when calculated from the first methionine residue, although it is likely that ORF 5 is expressed as a readthrough protein from ORF 3, as demonstrated for other members of the luteovirus group (Tacke *et al.*, 1990; Bahner *et al.*, 1990; Dinesh-Kumar *et al.*, 1992). Such an expression strategy would give a protein product of M_r ~80 K before potential proteolysis as has been shown for members of both luteoviral subgroups (Bahner *et al.*, 1990; Filichkin *et al.*, 1994). The size of the potential ORF 5 protein encoded by SDV-Tas1 (M_r 58 K) is significantly larger than that of BYDV-PAV-Vic (M_r 50 K) and accounts for most of the difference in length between the two genomes. A large 3' untranslated region (UTR) of 654 nt follows the UAG stop codon of ORF 5. There is no ORF in the SDV-Tas1 genome that corresponds to ORF 6 of subgroup I luteoviruses.

3.3.4 Analysis of the coding potential of the 3' UTR of SDV genomic RNA

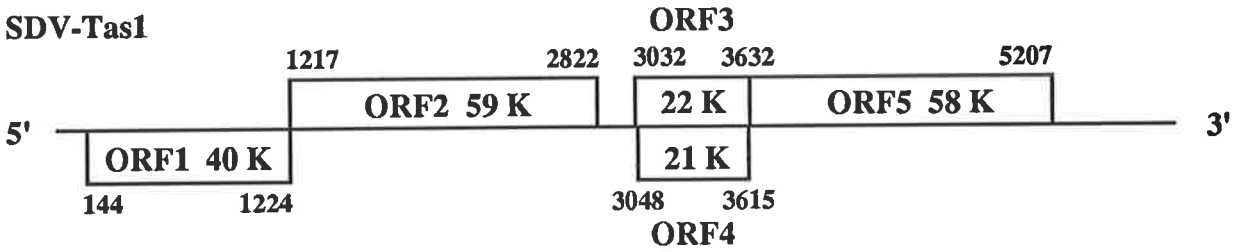
The 3' untranslated region of SDV-Tas1 (654 nt) is comparatively long given that there are apparently no significant ORFs in this region (Fig. 3.4A). To gain insight into the conservation of sequence and coding potential of this region, the sequence of the 3' end of a second SDV isolate (SDV-AP1) was determined. Comparison between the nucleotide sequences of SDV-Tas1 and SDV-AP1 downstream of ORF 5 reveals

Fig. 3.4. Open reading frames encoded in the nucleotide sequence of SDV-Tas1. (A) Three-phase translation of the SDV-Tas1 genomic sequence. Each phase is represented by a horizontal box numbered 1, 2 or 3 at the left side of the diagram, referring to phases 0, +1 and -1 respectively. Translational start codons (AUG) are represented by vertical lines reaching half the height of each box, while stop codons (UGA, UAG and UAA) are represented by vertical lines fully crossing the box. Other numbers refer to nucleotide positions in the genome. (B) Schematic representation of the ORFs of SDV-Tas1. Open boxes represent open reading frames (ORFs). ORFs above the line are in the +1 reading frame, those below are in the -1 reading frame. The nucleotide positions of initiation and termination of the ORFs are shown (small print), as are the potential protein sizes deduced by conceptual translation of each ORF. (C) Genome organisation of representatives of subgroup I and II luteoviruses (drawn to scale with Fig. 3.4B). Open boxes represent open reading frames. The virus genomes depicted are those of BYDV-PAV-Vic and PLRV-N. Numbers refer to the size of proteins potentially encoded by each ORF. The length of each genome in nucleotides is given at the 3' end of the diagram.

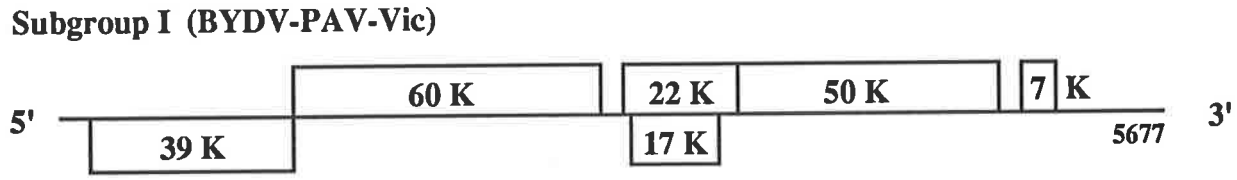
A


1000
2000
3000
4000
5000
1
2
3
1
2
3
1000
2000
3000
4000
5000

B


SDV-Tas1
ORF3
1217
2822
3032
3632
5207
ORF2 59 K
22 K
ORF5 58 K
5'
3'
ORF1 40 K
21 K
144
1224
3048
3615
ORF4

C


Subgroup I (BYDV-PAV-Vic)
60 K
22 K
50 K
7 K
5'
3'
39 K
17 K
5677


Subgroup II (PLRV-N)
28 K
69 K
23 K
56 K
5'
3'
70 K
17 K
5882

thirty-seven nucleotide changes, including five deletions and one insertion (Fig. 3.5A). Translation of the SDV-AP1 sequence in three reading frames (Fig. 3.5B) revealed only one ORF of appreciable size, extending from bases 5654 to 5788 of the SDV-Tas1 genome and capable of encoding a protein of M_r 5 K. However, this ORF is not conserved in the 3' genomic sequence of SDV-Tas1, which makes its significance doubtful. There is a lack of correlation in the incidence and positioning of small ORFs between the SDV-Tas1 and SDV-AP1 sequences which strongly suggests that this region has no coding function in the SDV genome. Further analysis of the 3' UTR of SDV was conducted by aligning the region 3' of ORF 5 in SDV-Tas1 with that of BYDV-PAV-Vic (data not shown). Only two regions of significant homology were detected, spanning but not including the region encoding ORF 6 of BYDV-PAV (data not shown; Chalhoub *et al.*, 1994). No homology was detected between the sequences of SDV-Tas1 and BYDV-PAV-Vic in the region of the putative ORF located in the 3' genomic sequence of SDV-AP1. Therefore, on the basis of conservation of nucleotide sequence between SDV strains and BYDV-PAV, it is unlikely that SDV contains a significant ORF 3' of ORF 5.

3.4 Discussion

The complete nucleotide sequence of SDV strain Tas1 was determined by sequencing both strands of cDNA clones that cover the entire RNA genome. Two strategies were used in the construction of cDNA clones. Firstly, cDNA synthesis was primed from purified SDV-Tas1 genomic RNA using hexanucleotides of random sequence. Viral sequences not represented in the initial population of cDNAs were isolated using PCR, either with two specific primers, or using a RACE protocol. The sequence determined in this way consists of 5861 nt which is significantly larger than that of BYDV-PAV (5677 nt; Miller *et al.*, 1988a) and BWYV (5641 nt; Veidt *et al.*, 1988) but similar to that of PLRV (5882 nt; van der Wilk *et al.*, 1989; Keese *et al.*, 1990).

The most significant usage of PCR in the cloning procedure was in determination of the 5' and 3' terminal genomic sequences of SDV-Tas1. The termini of viral genomic RNAs can be sequenced directly (by degradative enzymic sequencing, or primer extension/dideoxy sequencing on the RNA template) only if the viral RNA is abundant and easily purified. As luteoviral RNA is difficult to purify in large amounts, the alternative protocol of RACE was used. While this is relatively simple in 5' end analysis, the 3' end of the genomic RNA must first be polyadenylated to provide a priming site for synthesis of the first-strand cDNA. Thereafter the protocols for amplification of the 5' and 3' ends are the same.

There are two prime considerations when using RACE to determine the end of a molecule. Firstly, the full-length sequence may be truncated either by failure of the reverse transcriptase to continue transcription to the end of the molecule, or by internal binding of the non-specific $d(T)_n$ primer during amplification. Failure of reverse transcription is a potential problem only with 5' end determination, because in 3' end determination the reverse transcriptase extends towards the sequence-specific primer site, so prematurely terminated transcripts will not be amplified. The second major consideration in using RACE is the error rate of the polymerase used in PCR, which is compounded by the exponential nature of PCR amplification. In the application described here only truncation of the sequence was observed.

Several 5' and 3' end clones were sequenced to establish the extent of nucleotide variation at the termini of the viral genome. Selection of sequences for cloning after RACE amplification was either by size after agarose gel electrophoresis, or by Southern blot with a specific SDV-Tas1 probe. Sequence variation was observed only as the presence or absence of the terminal nucleotide. Therefore, not only is the genomic RNA homogenous in length, but the RACE process faithfully reproduces the complete ends of the viral sequence. Mapping of the genomic termini in this way relies on the assumption that the longest sequence recovered from the RACE reaction represents the true terminus of the RNA. Also, RACE does not exclude that the 5' terminal nucleotide is a U, or that the 3' terminal nucleotide is an A, because of the necessity for tailing the

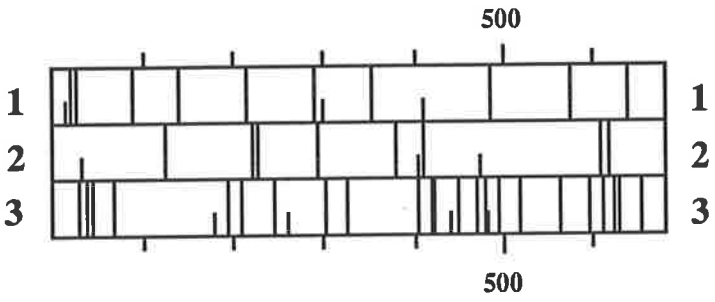
Fig. 3.5. Nucleotide variation between the genomes of SDV-Tas1 and SDV-AP1 at the 3' end of the viral genome. (A) Nucleotide changes in the SDV-AP1 genome relative to SDV-Tas1. The full sequence is the 3' end of the SDV-Tas1 genome (downstream of ORF 5); changes to this sequence in SDV-AP1 are indicated in the line above. Δ indicates a nucleotide deletion, +(N) an insertion. Other changes are represented by nucleotide abbreviations (capital letters). The primer sites used in the amplification of the SDV-AP1 sequence are underlined. (B) Three-phase translation of the sequences presented in Fig. 3.5A. Interpretation of the diagram is as for Fig. 3.4A.

A

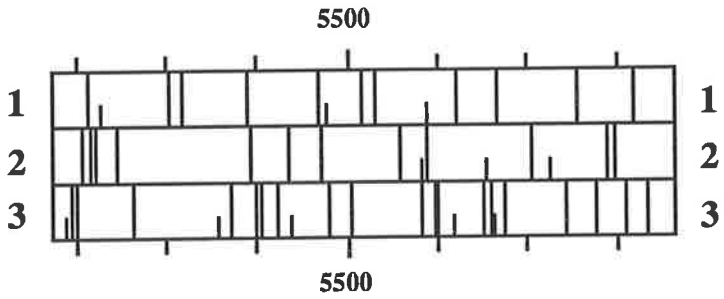
```
AP1                                          GC              G
5178 GGGCAUAUAUCGAUGGUUUAGAUUUAGUUUAAUAUAAAUAGAAGUAAGUAUGCGCUUUCCAUCACAUUUAGUUGGAGCCA 5257
          5190                  5210                  5230                  5250
AP1                                                      Δ
5258 ACCUUGGUGAGUUGCUGUCGUGGGAUCCUGGGAAACAGGUUCGGUGAACAAAAAGCCGGUAAUCUCGGGCGUUUGUCAGA 5337
          5270                  5290                  5310                  5330
AP1   U     U                                      Δ              GUC     +A
5338 CGCCGUCGCAUCAACACAGUAUGCAUAUCCAUCAAUAACCCAACGUUCUUGUGACUAAAAAUUAAAGUUAGUUUCCGCCU 5417
          5350                  5370                  5390                  5410
AP1       G          C                        A                                    C
5418 CUCUUAUAAACCACAUAUAAGAUGCGAGCUUGUGGAGAGUACCGUCCCUUGAAUAAGGAUGGUUGAAAGUUUCUGGUUUC 5497
          5430                  5450                  5470                  5490
AP1                    G                        A        G          U   UU
5498 AAUCCAUUGAUCAAAAUUGACACUGCUUCUGGUGACUACACUGCCGCUGGCGAUACAGCAUAAACAAUUUCCACUACUUG 5577
          5510                  5530                  5550                  5570
AP1                    C                 U       UC      A        Δ         ΔΔ
5578 UUCGAUGAAAUGAGAACAAUUAGUAAAGUCUUGCCGGUCACAUGAUUAUUUGGCGGUGCUACCGUUCGCCCCUUUGAUGG 5657
          5590                  5610                  5630                  5650
AP1                                      U       A   C                      C
5658 CGUUAAAUGAAAGCGGAGUAAACCGACAUCACCUCACCAGGUUGGGUUAACCUGGGGAAGCCGGGAAAAUGCCCGGAACA 5737
          5670                  5690                  5710                  5730
AP1                                            U             U          A   C
5738 CUUGGUUUAGUGGAUUUUAAACUCGAGGGCUCUUCCUCGCUAAAAAGGGGGUAGCCAGGUUGAAAACUGGCAUUAGGGAG 5817
          5750                  5770                  5790                  5810
AP1
5818 UGAUCCCUCUUUGUUGUUAAGGUGGCUUUGUGUCCACCUGCCCC  5861
          5830                  5850
```

B

SDV-AP1



SDV-Tas1



template nucleic acid prior to amplification. Tailing of the 5' cDNA product with an alternative nucleotide (followed by RACE-PCR using the complementary polynucleotide as the non-specific primer) would allow exact specification of the SDV Tas1 5' genomic sequence.

RNA viruses are believed to exist as mixtures of related but variant genomes known as quasi-species (Domingo, 1992). The sequence variants probably arise from the high mutation rate associated with viruses with RNA genomes (usually 10^{-4} - 10^{-5} per base; Holland *et al.*, 1992), which in turn seems to be a function of the lack of a proof-reading activity in the viral replicase (Steinhauer *et al.*, 1992). Sequence 'drift' is contained because most sequence variants possess a low fitness relative to the dominant sequence in the population. Hence populations of replicating RNA genomes appear to consist primarily of a dominant or master sequence, with small percentages of sequence variants of below average fitness (Domingo and Holland, 1988). Despite this, no sequence variants were observed when sequencing SDV-Tas1. This is perhaps not surprising, given that (1) each cDNA clone is a random sampling of a single genome in the population and (2) each genome probably only differs from consensus by 1-2 nt (Domingo *et al.*, 1978). Therefore, given an average clone size of roughly 400 nt, and the limited extent of the SDV-Tas1 genome covered by more than two clones, the probability of detecting sequence variants is likely very low. Similarly low levels of sequence variability were found in overlapping clones of BYDV-PAV-Vic (1/2883 nt; Miller *et al.*, 1988a) and BYDV-PAV-P (6/5179 nt; Ueng *et al.*, 1992). Moreover, the significance of such sequence variants is doubtful; often they cause no amino acid change from the wildtype sequence, or specify conservative (similar chemical properties) amino acid substitutions. Finally, it is difficult to distinguish between mutations derived from genuine variants in the viral population, and those artefactually generated in the cloning process.

The genome organisation of SDV-Tas1 resembles that of subgroup I more closely than subgroup II luteoviruses (Fig. 3.4). While SDV shares most of the familiar features of subgroup I genome organisation, including the lack of ORF 0, and an ORF 1

of M_r ~40 K slightly overlapped by an ORF 2 of M_r ~60 K, there are also aspects that are unique or in unique combination in SDV. The total length of the sequence is greater than might be expected for a nominal subgroup I luteovirus. This is largely due to the size of ORF 5 (M_r 58 K, versus 50 K for BYDV-PAV-Vic; Miller *et al.*, 1988a). In addition, the M_r 21 K coding potential of ORF 4 of SDV-Tas1 is relatively large (the sequence of BYDV-PAV, PLRV and BWYV vary from M_r 17-19.5 K), which may be evolutionally important given the confinement of ORF 4 within the boundaries of the M_r 22 K coat protein ORF. The size of the intergenic region is similar in SDV-Tas1 and the subgroup II luteoviruses (~200 nt), but significantly shorter in subgroup I (~110 nt). Perhaps the most major difference is the lack of ORF 6 of subgroup I in the SDV genome, which does not exist in two distinct isolates of SDV.

The absence of ORF 6 in the SDV genome is a major point of dissimilarity to the subgroup I luteoviruses, *viz.* BYDV-PAV and -MAV. While it is not certain that ORF 6 is expressed, the following lines of evidence point to its importance (reviewed in 1.4.2.9): The 5' end of a small subgenomic RNA (sgRNA 2) maps close to the putative initiation codon of ORF 6 in BYDV-PAV (Kelly *et al.*, 1994), and is therefore potentially responsible for its expression. Crucially, this subgenomic RNA appears to be absent in SDV, which possesses a single large subgenomic RNA corresponding to the sgRNA 1 of both luteovirus subgroups (Smith *et al.*, 1991). Also, the pattern of nucleotide and amino acid sequence conservation in ORF 6, as well as the lethal effects of its truncation (Young *et al.*, 1991), imply that this ORF is translated and that the gene product is necessary for BYDV-PAV replication. Therefore the absence of this ORF in SDV is a major distinction between it and the subgroup I luteoviruses. Another variation is the role of the SDV 3' UTR in comparison with that of subgroup I. The 3' UTR of BYDV-PAV appears to affect both ribosomal frameshifting and cap-independent translation (Miller *et al.*, 1994). While the mechanism of this function is obscure, it follows that the two small subgenomic RNAs of BYDV-PAV-Vic (sgRNAs 2 and 3) might mediate the effect, acting in *trans*. This is especially pertinent if the coding potential of the BYDV-PAV 3' UTR is discounted. If the 3' UTR of SDV is

also necessary for these functions, as may be inferred on the basis of homeology, then it must act in *cis* as sgRNAs 2 and 3 appear to be absent in SDV (Smith *et al.*, 1991). Despite the poor nucleotide sequence conservation in the 3' UTRs of SDV and BYDV-PAV it is unlikely that this region has no function in either virus, otherwise its relatively extreme length would not be preserved.

The organisational similarities of SDV relative to other luteoviruses implies that it is likely to share strategies for gene expression. In particular, it is likely that ORF 2 is expressed as a frameshift fusion with the product of ORF 1, as is the case for BYDV-PAV (Brault and Miller, 1992; Di *et al.*, 1993). This proposition is based on the obvious similarity of gene organisation shared by the two viruses in this region, as well as the lack of an initiation codon in SDV ORF 2. Likewise, ORFs 3, 4 and 5 are probably expressed from the major subgenomic RNA described by Smith *et al.* (1991) as for other luteoviruses. ORF 5 is likely to be expressed as a readthrough product of ORF 3, and ORF 4 synthesised by internal initiation from the coat protein messenger RNA. These features of expression appear to be common to all luteoviruses described so far (Bahner *et al.*, 1990; Tacke *et al.*, 1990; Dinesh-Kumar *et al.*, 1992). Sequence comparisons between SDV and other viruses are presented in Chapter Four.

CHAPTER FOUR

STRUCTURE AND EVOLUTION OF THE SDV-TAS1 GENOME

4.1 Introduction

The previous Chapter showed that SDV has a genomic structure similar to that of the group I luteoviruses other than the lack of ORF 6 in the 3' portion of the viral genome. This Chapter extends this study to include comparisons of the deduced amino acid sequences of the viral ORFs between SDV and other luteoviruses, and also the nucleotide sequences of the non-coding regions of the genome. This comparison has two purposes; to establish the relatedness of SDV to the other luteoviruses, and to gain further understanding of the evolutionary events leading to the generation of the two luteoviral subgroups. The general aim of these studies within the context of this thesis is to enable the design of experiments for the investigation of the interaction of BYDV-PAV with the Yd2 barley resistance gene.

A further goal of comparative studies is to refine the methods for taxonomic classification of plant viruses. The more traditional methods of classification rely on readily observable characteristics such as host range of the virus, transmission characteristics, and the symptoms caused in susceptible hosts. These have been strengthened by the addition of biochemical data, such as the size of coat protein components, dsRNA species involved in replication, and the presence or absence of a VPg, and also immunological studies, *i.e.* the serological relationships between virions. While these have generally provided a successful scheme for taxonomic classification, the advent of molecular biology has further advanced the resources available for description of plant viruses. Classification using molecular data has largely reproduced that generated with classical techniques. However, amino acid or nucleotide sequences allow comparisons at different points in the genome, using larger data sets than the measurement of a single physical characteristic. Thus greater confidence in the inference of relationships is possible. Moreover, this type of analysis has developed the idea of the viral RNA genome as a collection of genetic elements that evolve at different rates, rather than as a single entity that evolves uniformly. This last concept has been reinforced by the recognition of the role of genetic recombination between RNA genomes in their evolution (Lai, 1992).

A further contribution of comparative sequence analysis is to the understanding of gene function. In particular, the recognition of amino acid sequence motifs associated with particular molecular activities, either inferred or proven, is a powerful method for elucidating the function of viral ORFs. Similarly, conserved motifs in the nucleotide sequence define elements such as those involved in viral RNA replication, initiation of subgenomic RNA synthesis, and especially in the case of luteoviruses, aberrant translation events. Such inferential data naturally takes second preference to empirical experimental results, but in many cases may set the direction of subsequent investigation.

4.2 Materials and Methods

4.2.1 Virus abbreviations

Virus abbreviations and references to sequence data used in this Chapter are as follows: BaYMV - barley yellow mosaic bymovirus (Kashiwazaki *et al.*, 1990, 1991); BLRV - bean leafroll luteovirus (Table 1.1); BWYV- beet western yellows luteovirus (Table 1.1); BWYV-ST9 aRNA - independently replicating RNA associated with BWYV strain ST-9 (Chin *et al.*, 1993); BYDV - barley yellow dwarf luteovirus (Table 1.1); CABYV - cucurbit aphid-borne yellows luteovirus (Table 1.1); CarMV - carnation mottle carmovirus (Guilley *et al.*, 1985); CMV - cucumber mosaic cucumovirus (Rezaian *et al.*, 1984); LTSV - lucerne transient streak sobemovirus (Drs A.C. Jeffries and R.H. Symons, University of Adelaide, Australia, personal communication); MBV - mushroom bacilliform mycovirus (Revill *et al.*, 1994); PEMV - pea enation mosaic penamovirus (Demler and de Zoeten, 1991; Demler *et al.*, 1993); PLRV - potato leafroll luteovirus (Table 1.1); PVX - potato virus X potexvirus (Kraev *et al.*, 1988, cited by Koonin and Dolja, 1993); RCNMV - red clover necrotic mosaic dianthovirus (Xiong and Lommel, 1989); SBMV - southern bean mosaic sobemovirus (Wu *et al.*, 1987); SDV - soybean dwarf luteovirus (Chapter Three).

4.2.2 Computer analyses

Amino acid sequences were aligned using the UWGCG programs GAP for pairwise comparisons or PileUp for multiple sequence alignments (Devereux *et al.*, 1984). Both programs use the algorithm of Needleman and Wunsch (1970) to create pairwise sequence alignments. Default values of 3.00 for gap weight and 0.10 for gap length weight were used in all comparisons except for multiple alignment of motif sequences (4.3.1), where a gap weight value of 10.00 and a gap length weight of 1.00 were used to suppress gap insertion. Multiple alignment using PileUp proceeds with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster can then be aligned to the next most related sequence or cluster of aligned sequences. The final alignment is achieved by a series of progressive, pairwise alignments that include increasingly dissimilar sequences and clusters, until all sequences have been included in the final pairwise alignment.

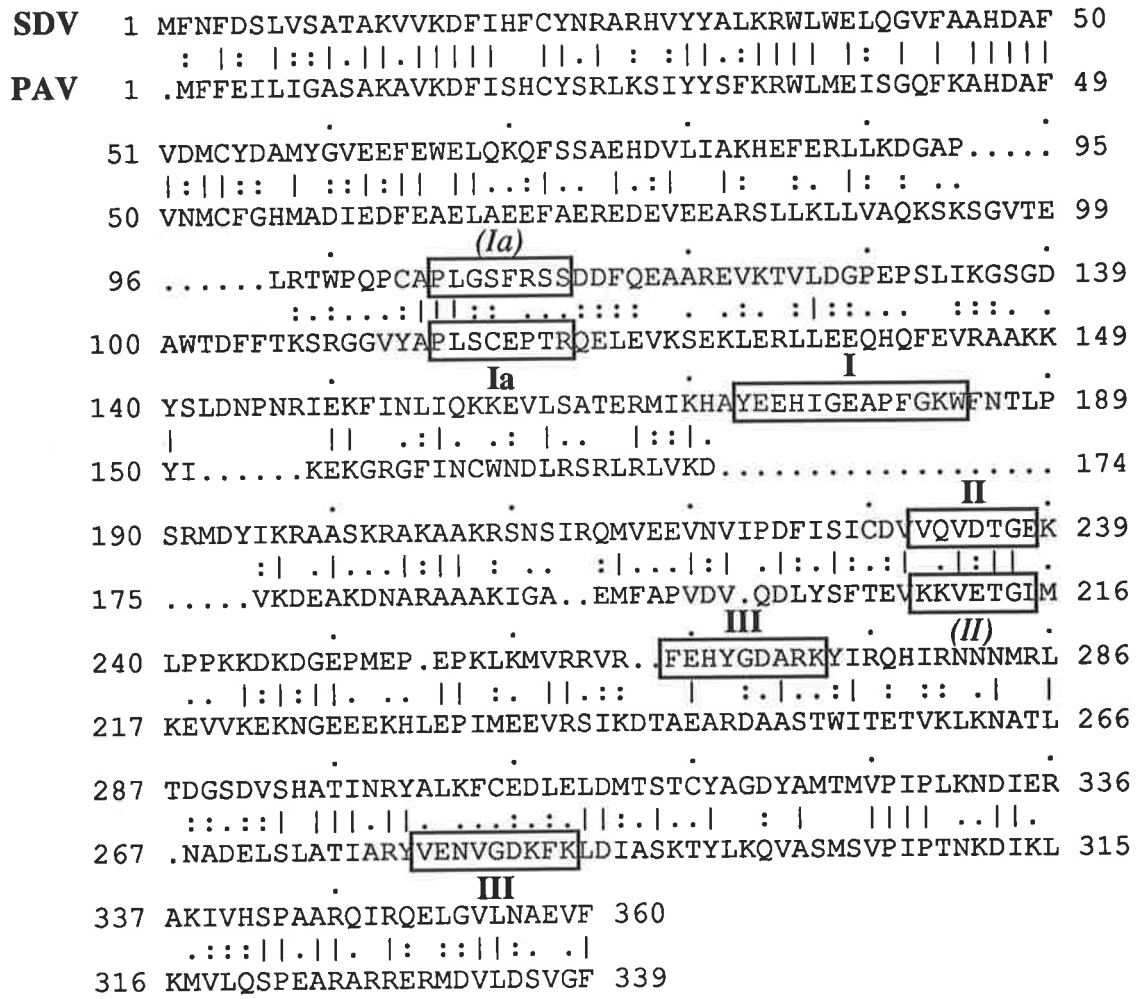
PileUp scores the similarity between every possible pair of sequences. These similarity scores are used to create a clustering order that can be represented as a dendrogram. The clustering strategy represented by the dendrogram is called UPGMA, which stands for unweighted pair-group method using arithmetic averages. For a pairwise alignment of clusters of sequences, the comparison score between any two positions in those clusters is the arithmetic average of the scores for all possible symbol comparisons at those positions.

4.3 Results

4.3.1 Analysis of the coding potential of ORFs 1 and 2

Putative helicase. The 5' half of the SDV genome encompassing ORFs 1 and 2 of SDV shares similar organisation to that of the subgroup I luteoviruses (see Chapter Three). This similarity extends to homology between the deduced amino acid sequences of the respective ORFs. ORF 1 of SDV is larger than that of BYDV-PAV, potentially encoding 360 amino acids as against 339 for BYDV-PAV. The deduced amino acid sequences of the SDV and BYDV-PAV ORFs 1 are 32% identical after the

Fig. 4.1. Alignment of deduced amino acid sequences of ORFs 1 of SDV and BYDV-PAV. Sequences were aligned by the program GAP using default values as indicated (4.2.2). The upper sequence belongs to SDV. Numbers refer to amino acid positions in the deduced sequence. Putative helicase motifs I, Ia, II and III are boxed. Motifs not nominated by Habili and Symons (1989) but inferred by homology are boxed and indicated by bracketed italics.


SDV
PAV
(Ia)
Ia
I
II
III
(II)
III

addition of gaps, with the greater percentage of the identical amino acids at the amino and carboxyl termini of the deduced proteins (Fig. 4.1). The deduced sequence of the ORF 1 protein shows no significant homology to other sequences in electronic databases. Similarly, no experimental evidence exists for the role of the ORF 1 product *in vivo*, although Habili and Symons (1989) have presented comparative data suggesting that a helicase activity is encoded in the deduced sequences of ORF 1 and the N-terminal sequence of ORF 2. However, this proposal is challenged here by a re-examination of the data.

Nucleic acid helicases (referred to hereafter as 'helicases') are polynucleotide-dependent nucleoside triphosphate (NTP) phosphatases which possess the ability to unwind double-stranded nucleic acid complexes. They are believed to be involved in basic genetic processes such as genome replication, transcription, recombination and repair. Numerous putative helicases from many classes of organisms have been identified on the basis of seven conserved amino acid sequences or motifs (Gorbalenya and Koonin, 1993). The varying levels of sequence conservation of the seven motifs has allowed grouping of the putative helicases into at least three superfamilies. All putative helicases share a homologous form of the 'Walker' A and B sites (Walker *et al.*, 1982; corresponding to motifs I and II) believed to form the active component of the NTP hydrolysing activity of the helicase. However these sites are also present in NTP-hydrolysing enzymes that are not helicases. The remainder of the motifs (Ia, III, IV, V and VI) thus specify the nature of the enzyme, and also allow classification of the putative enzyme into the relevant superfamily. Habili and Symons (1989) nominated sequences present in the deduced amino acid sequences of ORF 1 of either SDV or BYDV-PAV (or both) as representative of motifs I, Ia, II and III. Sequences corresponding to motifs IV and VI (but not V) were found in the deduced amino acid sequences of ORF 2 of both viruses. The authors concluded from the sequence alignments that the translation product of ORF 1, necessarily fused by frameshift translation to the N-terminal portion of ORF 2, encodes a putative helicase.

Alignment of the seven helicase motifs for representative proteins of helicase superfamily 1 (SF1), which contains the helicases or putative helicases of positive-strand RNA viruses (Gorbalenya and Koonin, 1993), is presented in Fig. 4.2. The deduced sequences nominated by Habili and Symons to correspond to these motifs in SDV and BYDV-PAV are shown underneath the SF1 alignment. No sequence was proposed for SDV corresponding to motif Ia, likewise there is no proposed motif II for BYDV-PAV. Where these proposed motifs are absent in one or the other of the two viruses, homologous sequences have been traced using the program GAP (Figs 4.1 and 4.2). This was not possible for motif I, where no BYDV-PAV sequence was nominated (see below), or motif V, where no sequence was proposed for either virus.

Habili and Symons (1989) nominated a potential motif I sequence for SDV but not BYDV-PAV. Motif I is the highly conserved 'Walker A' NTP binding motif **GxGK[S/T]**, which is necessary for helicase activity (Gorbalenya and Koonin, 1993). The principle that the motif strategy is based upon is the strong conservation of particular sequences in homologous proteins of closely related organisms. Consensus sequences defined in this way can be used to identify similar sequences in more distantly related proteins. Therefore, it is a strict requirement when comparing distantly related sequences using a motif strategy that potential motifs to be used in the alignment are conserved between closely related entities such as SDV and BYDV-PAV. Alignment of the deduced amino acid sequences of ORF 1 of these viruses revealed no homology at the position of the proposed motif I in the genome; in fact, BYDV-PAV appears to contain a deletion relative to SDV in the position of the proposed motif (Fig. 4.1). Additionally, the proposed motif I of SDV contains only the sequence **GK** in common with the consensus alignment, so it is unlikely that it represents a functional NTP-binding domain. Therefore, the sequence proposed by Habili and Symons (1989) for SDV does not fulfil the criteria to represent helicase motif I.

The proposed Ia motif of Habili and Symons (1989) is present in BYDV-PAV but not SDV; similarly, the putative motif II (containing the 'Walker B' motif) is

Fig. 4.2. Alignment of proposed helicase motifs for SDV and BYDV-PAV ORFs 1 (Habili and Symons, 1989) with selected proteins of helicase superfamily 1 (SF1; Gorbalenya and Koonin, 1993). Motifs are numbered above the sequences. Distances between the motifs and protein termini are indicated. Consensus residues are highlighted by bold uppercase typing except for deviating residues. Asterisks indicate positions of identity in the consensus sequence. Luteovirus motifs have been slightly modified from that published where necessary to fit with more recent consensus sequences (not shown). Luteovirus motifs not nominated by Habili and Symons (1989) have been inferred by alignment of ORF 1 sequences using GAP (Fig. 4.1, and see text) and are represented in italics. RUBV, rubella virus; PVX, potato virus X (encodes two helicases), HSV-1, herpes simplex virus.

		I		Ia		II		III	
'HEL'	RUBV	IRVWNmAA**GAGKT**	10	DLYVCPTN	36	IYI**DE**AF	17	VICV**GD**RD**Q**	19
p166	PVX	ACVIHGAg**GSGKS**	15	ITVVLPTN	32	VIF**Dd**YS	19	VILT**GD**SR**Q**	27
p25	PVX	PLVVHAVA**GAGKS**	15	HTLGVPDK	19	AIL**DE**YT	9	QALF**aD**PY**Q**	9
UL5	HSV-1	VYLITGNA**GSGKS**	11	CVVTGATR	121	IVI**DE**AG	33	LVCV**Gs**PT**Q**	43
UvrD	E.coli	NLLVLAGA**GSGKT**	20	IMAVTFTN	152	ILV**DE**FQ	19	VMIV**GD**DD**Q**	25
		* ***				**		*	
SDV	ORF1	YEEHIGEA**pFGKw**	*-60*	*PLGSFRSS*	*120*	VQV**Dt**GE	27	FEHY**GD**AR**k**	148
PAV	ORF1			PLSCEPTR	*87*	*KKV**et**GL*	*65*	VENV**GD**KF**k**	112

		IV		V		VI	
'HEL'	RUBV	ERSRHTW**R**	55	AY**T**VREA**QG**MSVGT**a**	21	**Vs**L**TR**ASD	?
p166	PVX	YYLNATH**R**	51	TF**T**YAGC**QG**LTKPK**V**	17	**tALSR**ATD	502
p25	PVX	FYLETSF**R**	57	FV**k**PCQV**tG**LEFKV**V**	17	**nAITR**SK-	9
UL5	HSV-1	AIFINNK**R**	461	AM**T**ITRS**QG**LSLDK**V**	15	**VAMSR**TTS	36
UvrD	E.coli	IRLEQNY**R**	271	LM**T**LHSA**KG**LEFPQ**V**	30	**VGVTR**AMQ	112
				* *		*	
SDV	ORF2	VAVERRV**f**			104	**VeLgR**RLK	354
PAV	ORF1	VAVERRV**f**			105	**iiLgt**RLK	347

represented only in SDV. Sequences homologous to proposed motifs have been traced where missing using GAP (Figs 4.1 and 4.2). In both cases the pattern of conservation between the deduced amino acid sequences of the two viruses at the site of the proposed helicase motif does not reflect the consensus sequence of the true motif (Fig. 4.2). Two further points concerning the positioning of the proposed motifs in the sequence of ORF 1 are relevant. Firstly, in SDV, the sequence of proposed and deduced motifs is Ia-I-II, rather than I-Ia-II as in all other putative helicases (Fig. 4.1). Although this could conceivably reflect rearrangement of the viral genome, it is more likely that the assignation of the motifs in the deduced amino acid sequence of SDV ORF 1 is incorrect, given the poor relationship between the proposed motif sequences and the consensus. Secondly, while the sequences of SDV and BYDV-PAV proposed to correspond to motif III show nominal conservation to each other and to the motif III consensus, alignment of the deduced ORF 1 protein sequences of SDV and BYDV-PAV places the respective proposed motif III sites at different (non-homologous) points in the genome (Fig. 4.1). Thus the apparent conservation of the proposed luteoviral motif III appears to be artefactual.

The remainder of the proposed helicase motifs fall into the N-terminal region of the deduced ORF 2 protein sequence. The sequence proposed to correspond to motif IV is highly conserved between SDV and BYDV-PAV, but shows no homology with the consensus sequence. No candidate sequence for motif V was nominated. Finally, the proposed motif VI sequence is poorly conserved between the two viruses, and shows little homology to the consensus sequence. The higher level of conservation of proposed motif sequences in ORF 2 is possibly a function of the high homology between the deduced protein sequence of SDV and BYDV-PAV in this ORF, rather than because of specific conservation of the proposed helicase motifs.

While none of the proposed helicase motifs show convincing relationships to the true consensus sequences, the critical example is motif I. This is because it is the most extensively conserved of the seven motifs, and also because its defined role in the function of the helicase as a NTPase means that it is indispensable. Its absence in the

deduced amino acid sequence of either SDV or BYDV-PAV as argued here is the major factor in the preclusion of a helicase role for the transframe product of ORFs 1 and 2. However, there is also insufficient evidence that the proposed luteovirus sequences correspond to the seven putative helicase motifs. The absence of a helicase in the subgroup I luteovirus genome is not unexpected because no such enzyme (or putative enzyme) has been identified in any single-stranded RNA virus with a genome smaller than 6 kb (Koonin and Dolja, 1993).

RNA-dependent RNA polymerase. The **GDD** amino acid motif associated with RNA-dependent RNA polymerase (RdRp) activity (Kamer and Argos, 1984) is encoded in the nucleotide sequence of SDV ORF 2. The deduced amino acid sequence of this ORF shows high homology to that of BYDV-PAV, with 61% identical residues

Table 4.1 Amino acid sequence comparisons of the putative RNA-dependent RNA polymerases of selected luteoviruses, carnation mottle virus and southern bean mosaic virus

	CarMV	PLRV	SBMV	SDV
BYDV-PAV	34.0[a]	17.8	20.9	60.8
CarMV	-	16.1	16.9	34.4
PLRV	-	-	31.1	15.6
SBMV	-	-	-	21.4

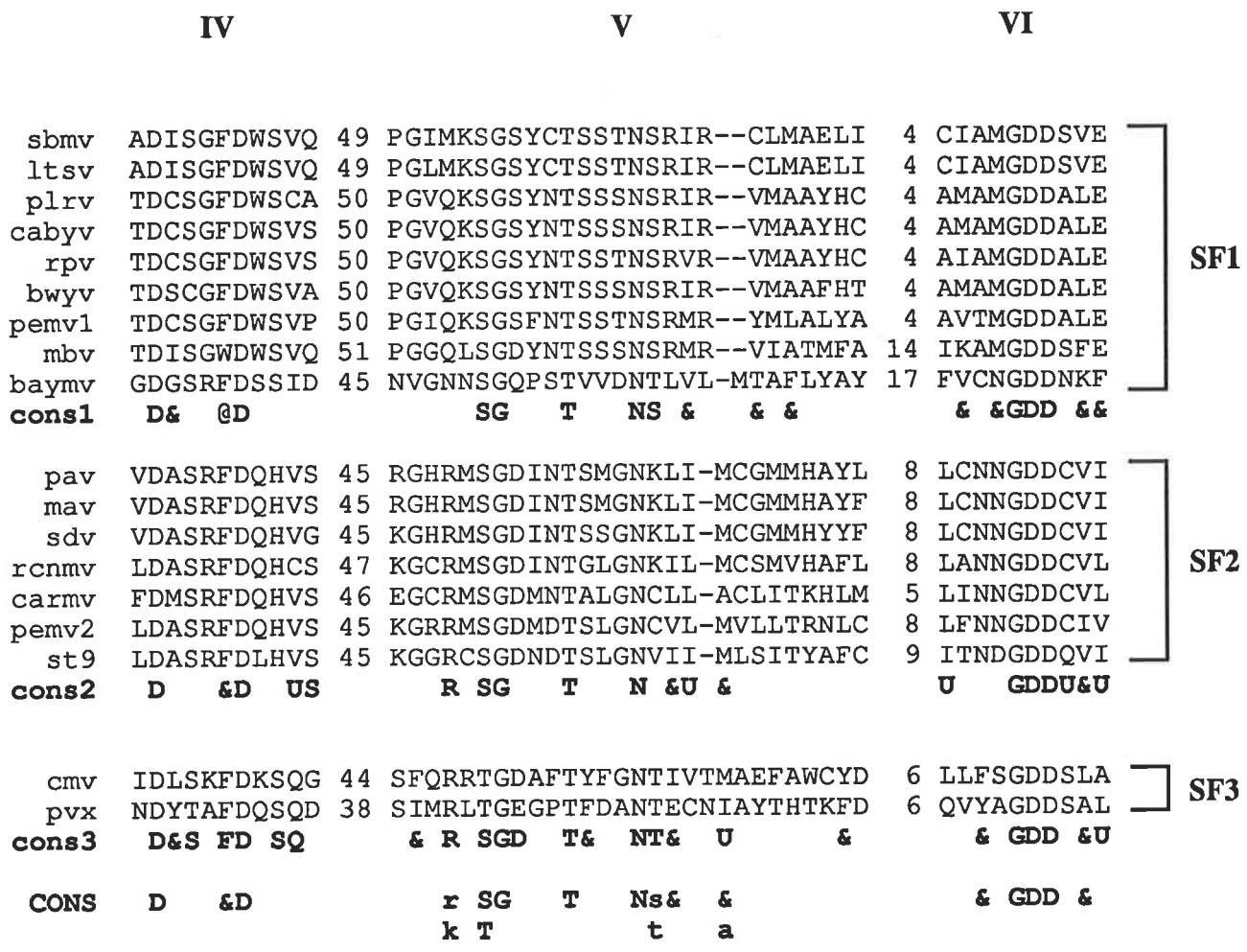
[a]Numbers are percentage of amino acids that are identical between the sequences and were derived using the UWGCG program GAP (Devereux *et al.*, 1984).

after the addition of gaps (Table 4.1). Subgroup I luteoviruses have a carmovirus-like polymerase gene; this relationship extends to the deduced amino acid sequence of SDV ORF 2 which is 34% identical to the polymerase gene of CarMV. Furthermore, SDV ORF 2 shows little homology to either of the polymerase genes of SBMV (21% identity) or PLRV (16% identity). SDV therefore possesses a carmovirus-like polymerase gene, rather than the sobemovirus-like protein of the subgroup II luteoviruses. Thus the subgroup I-like organisation of SDV is reflected in the sequence relationships of ORFs 1 and 2.

RdRps contain universal sequence motifs conserved in all known sequences of positive-strand RNA viruses. Two implications arise from this. Firstly, all RdRps have descended from a common ancestor (albeit with subsequent diversion into lineages), and secondly RdRps of divergent lineages may be compared through the sequences of shared motifs. Eight such motifs have been defined and have been used to infer phylogenies of the enzyme (Koonin, 1991), and by association taxonomic relationships between the viruses containing the respective RdRps (Koonin and Dolja, 1993). The RdRp genes fall into three 'supergroups' with separate consensus sequences for each motif. Subgroup I luteoviruses and CarMV fall into supergroup II, while subgroup II luteoviruses and SBMV are classified as supergroup I (Koonin and Dolja, 1993). Of the eight Koonin motifs only three (IV, V and VI) are conserved in all putative RdRps of single-stranded plant RNA viruses. Therefore, in this thesis only the sequences represented by these motifs were used in comparisons between the putative RdRps of SDV and other plant RNA viruses.

Investigation of the relationship of the SDV RdRp to those of other viruses was approached by alignment of the deduced amino acid sequence of SDV ORF 2 with the three universally conserved RdRp motifs. RdRp sequences used for comparison (for abbreviations see 4.2.1) were obtained from subgroup I and II luteoviruses, CarMV, the sobemoviruses SBMV and lucerne transient streak virus (LTSV), PEMV RNAs 1 and 2 (supergroups II and I respectively), red clover necrotic mosaic dianthovirus (RCNMV; supergroup II), BWYV-ST9 aRNA (supergroup II), as well as the more distantly related

Fig. 4.3. Conserved sequence motifs in the RNA-dependent RNA polymerases of luteovirus and luteovirus-like plant RNA viruses. The comparison includes all known luteovirus RdRp sequences, as well as sequences known to be closely related to luteovirus RdRps and more diverse sequences (BaYMV, CMV and PVX). Only three of the eight motifs of Koonin (1991) were used in the comparison, and were chosen because of their universal conservation in the sequences of positive-strand RNA viruses. Numbers above the sequences refer to motif identity. Distances between motifs are shown. The sequences are arranged according to membership of each superfamily (SF1, SF2 or SF3; right hand scale) and consensus sequences for each superfamily are given (Koonin and Dolja, 1993). Residues conserved in at least 75% of total sequences (not presented) are shown. In the general consensus, residues conserved in all three superfamilies (upper case) or in two patterns (lower case) are shown. In the consensus lines, U represents a bulky aliphatic residue (I, L, V, M), @ represents an aromatic residue (F, Y, W), and & designates a bulky hydrophobic residue (either aliphatic or aromatic). Abbreviations for the viral origin of each RdRp sequence is given in the text (4.2.1).

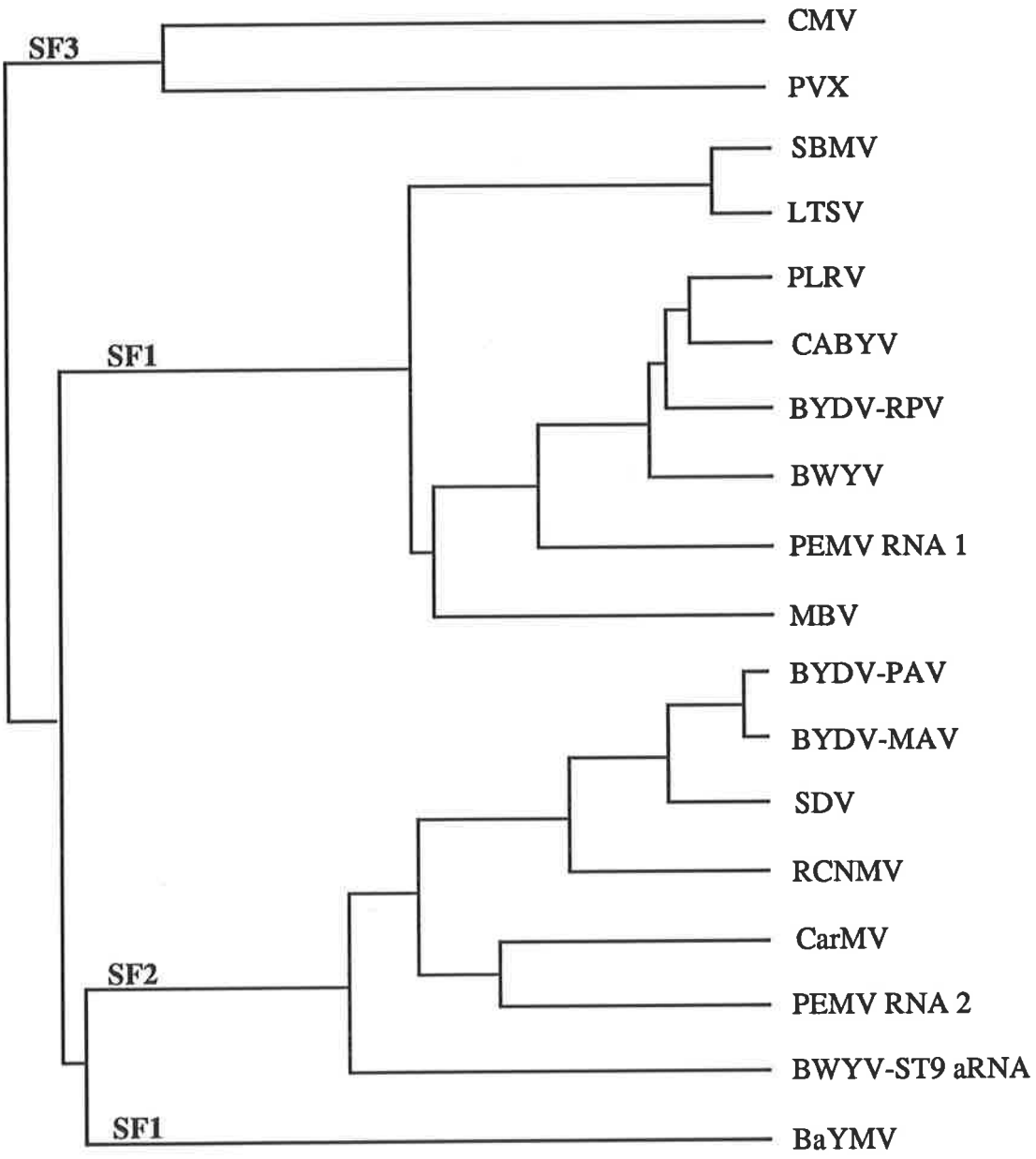

IV
V
VI
sbmv ADISGFDWSVQ 49 PGIMKSGSYCTSSTNSRIR--CLMAELI 4 CIAMGDDSVE
ltsv ADISGFDWSVQ 49 PGLMKSGSYCTSSTNSRIR--CLMAELI 4 CIAMGDDSVE
plrv TDCSGFDWSCA 50 PGVQKSGSYNTSSSNSRIR--VMAAYHC 4 AMAMGDDALE
cabyv TDCSGFDWSVS 50 PGVQKSGSYNTSSSNSRIR--VMAAYHC 4 AMAMGDDALE
rpv TDCSGFDWSVS 50 PGVQKSGSYNTSSTNSRVR--VMAAYHC 4 AIAMGDDALE
bwyv TDSCGFDWSVA 50 PGVQKSGSYNTSSSNSRIR--VMAAFHT 4 AMAMGDDALE
pemv1 TDCSGFDWSVP 50 PGIQKSGSFNTSSTNSRMR--YMLALYA 4 AVTMGDDALE
mbv TDISGWDWSVQ 51 PGGQLSGDYNTSSSNSRMR--VIATMFA 14 IKAMGDDSFE
baymv GDGSRFDSSID 45 NVGNNSGQPSTVVDNTLVL-MTAFLYAY 17 FVCNGDDNKF
cons1 D& @D SG T NS & & & & &GDD &&
SF1
pav VDASRFDQHVS 45 RGHRMSGDINTSMGNKLI-MCGMMHAYL 8 LCNNGDDCVI
mav VDASRFDQHVS 45 RGHRMSGDINTSMGNKLI-MCGMMHAYF 8 LCNNGDDCVI
sdv VDASRFDQHVG 45 KGHRMSGDINTSSGNKLI-MCGMMHYYF 8 LCNNGDDCVI
rcnmv LDASRFDQHCS 47 KGCRMSGDINTGLGNKIL-MCSMVHAFL 8 LANNGDDCVL
carmv FDMSRFDQHVS 46 EGCRMSGDMNTALGNCLL-ACLITKHLM 5 LINNGDDCVL
pemv2 LDASRFDQHVS 45 KGRRMSGDMDTSLGNCVL-MVLLTRNLC 8 LFNNGDDCIV
st9 LDASRFDLHVS 45 KGGRCSGDNDTSLGNVII-MLSITYAFC 9 ITNDGDDQVI
cons2 D &D US R SG T N &U & U GDDU&U
SF2
cmv IDLSKFDKSQG 44 SFQRRTGDAFTYFGNTIVTMAEFAWCYD 6 LLFSGDDSLA
pvx NDYTAFDQSQD 38 SIMRLTGEGPTFDANTECNIAYTHTKFD 6 QVYAGDDSAL
cons3 D&S FD SQ & R SGD T& NT& U & & GDD &U
SF3
CONS D &D r SG T Ns& & & GDD &
k T t a

barley yellow mosaic bymovirus (BaYMV; supergroup I) and supergroup III viruses cucumber mosaic cucumovirus (CMV) and potato virus X potexvirus (PVX; see 4.2.1 for abbreviations and references). A further divergent RdRp sequence added to the comparison was that of mushroom bacilliform mycovirus (MBV; unclassified), which possesses striking organisational and sequence similarities to subgroup II luteoviruses (Revill *et al.*, 1994).

Alignment of the viral RdRp sequences with the three RdRp motifs is given in Fig. 4.3. As expected, the RdRp of SDV is most similar to those of supergroup II enzymes and in particular to the members of luteovirus subgroup I. The phylogenetic tree derived from this alignment using the UWGCG program PileUp shows the following relationships (Fig. 4.4). Firstly, the 18 sequences form three clusters, with some doubt as to whether the BaYMV sequence is clustered with the subgroup I type luteovirus sequences. However, as Koonin and Dolja (1993) placed BaYMV with the subgroup II luteoviruses in supergroup I, it seems likely that in this analysis, BaYMV represents an outgroup with respect to the other sequences. Thereafter the sequences group as predicted by the analysis of Koonin and Dolja (1993). The subgroup I luteoviruses and related sequences of supergroup II form one cluster, as do the subgroup II luteoviruses and related sequences of supergroup I. Also, the supergroup III sequences of CMV and PVX form a loose cluster that is distinct from the remaining sequences.

Of major interest to this thesis is the relationships within the luteoviral subgroup I and II clusters. The subgroup I cluster tentatively describes the following relationships: There are three main lineages, composed of the BWYV ST9-aRNA, the cluster formed by CarMV and PEMV RNA 2, and the cluster formed by RCNMV and the luteoviruses SDV, BYDV-PAV and BYDV-MAV. The SDV RdRp is closely related to those of BYDV-PAV and -MAV as predicted by pairwise sequence comparisons (Table 4.1). The close relationship between RCNMV and the subgroup I luteoviruses has already been noted by Miller *et al.* (1994), who proposed that the subgroup I luteovirus genome was created by recombination between an ancestor of

Fig. 4.4. Phylogeny of selected RNA-dependent RNA polymerase motifs. The dendrogram was generated from the data presented in Fig. 4.3. using the UWGCG program PileUp. The three motifs were treated as a single sequence with modification of default variables to suppress misalignment by insertion of gaps (4.2.2). The superfamily (SF) identity of each branch as defined by Koonin and Dolja (1993) is shown. Virus acronyms are given in 4.2.1.

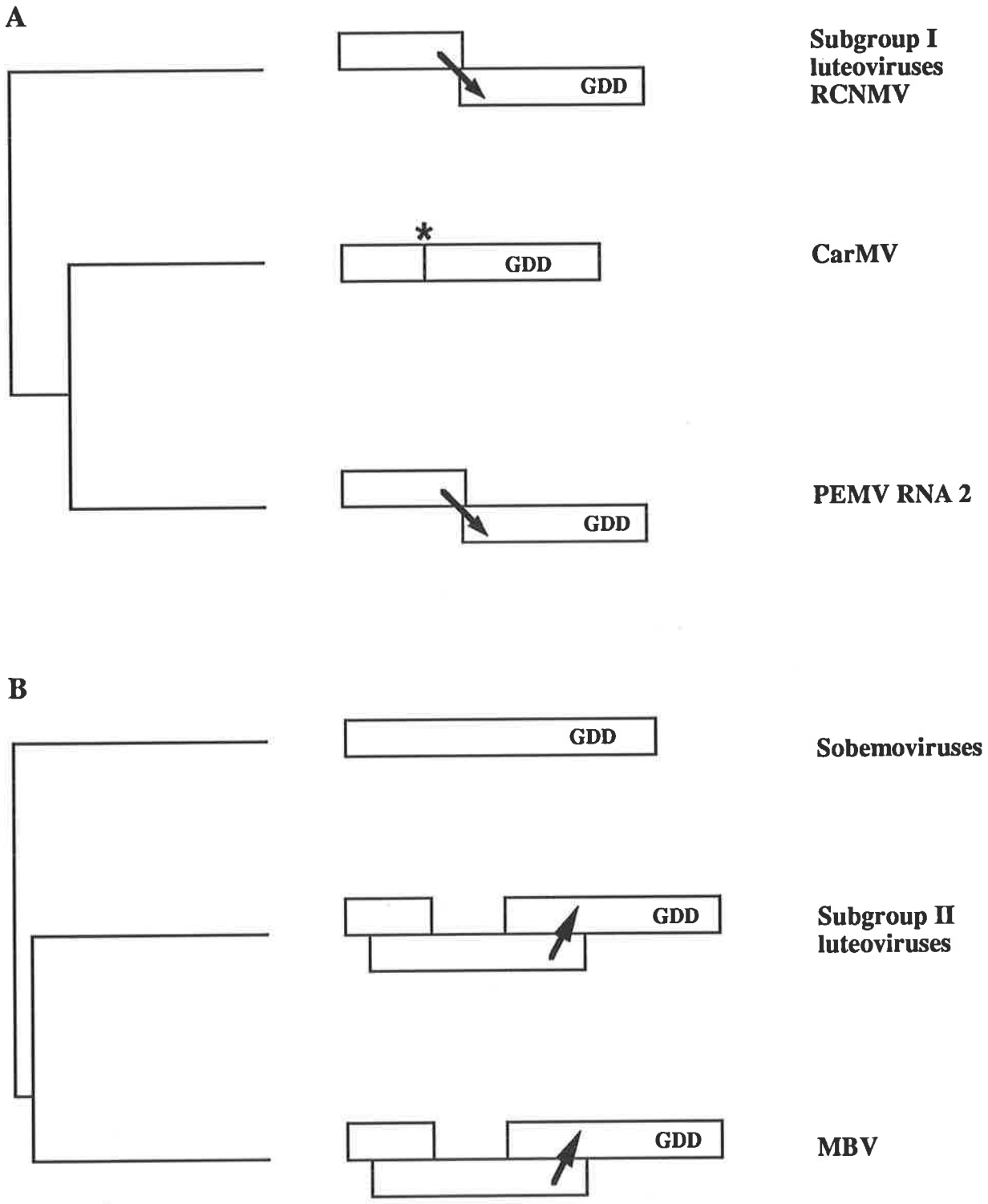

SF3
CMV
PVX
SF1
SBMV
LTSV
PLRV
CABYV
BYDV-RPV
BWYV
PEMV RNA 1
MBV
SF2
BYDV-PAV
BYDV-MAV
SDV
RCNMV
CarMV
PEMV RNA 2
BWYV-ST9 aRNA
SF1
BaYMV

RCNMV and a subgroup II luteovirus (possibly BYDV-RPV; 1.5). The most surprising result of the analysis presented here is the clustering of the CarMV and PEMV RNA 2 putative RdRps, given the differences in gene organisation and probable mode of gene expression of these two viruses. The putative RdRp of CarMV appears to be expressed after suppression (readthrough) of a stop codon from the preceding gene, whereas for PEMV RNA 2, expression of the putative RdRp appears to be by frameshift translation from the previous gene, by analogy to the luteoviruses and RCNMV (Demler *et al.*, 1993; Xiong *et al.*, 1993).

The supergroup I-like cluster consisting of the subgroup II luteoviruses and related viruses also appears to contain three lineages. These are composed of the sobemoviruses (SBMV and LTSV), the subgroup II luteoviruses and PEMV RNA 1, and MBV. The association of the subgroup II luteoviruses with the sobemoviruses and PEMV RNA 1 is unsurprising. However, it is noteworthy that the putative RdRp of MBV, which has a luteoviral subgroup II type organisation of its 5' three ORFs, is approximately equally similar to the putative RdRps of the sobemoviruses than it is to the subgroup II luteovirus cluster. This is despite the fact that the MBV RdRp (ORF 3) is likely expressed by frameshift from ORF 2 (Revill *et al.*, 1994), whereas the coding regions of SBMV are quite differently organised and frameshift expression of the putative RdRp domain of ORF 2 has not been proposed.

In conclusion, SDV acts as an orthodox subgroup I luteovirus in this comparison. This is in agreement with aspects of the genome organisation of SDV discussed above and in Chapter Three. The viral sequences compared here largely segregated according to the supergroup classification of Koonin and Dolja (1993), apart from BaYMV which appeared to form an outlier to the other sequences. The most striking point of the comparison is the association of the CarMV and PEMV RNA 2 RdRps, and the approximately equal relationship between MBV, the sobemoviruses and the subgroup II luteoviruses. Thus a paradigm appears to exist in both of the clusters that contain luteovirus sequences whereby sequence homology between the RdRps of the different viruses is not supported by the expression strategy for the gene (Fig. 4.5).

Fig. 4.5. Relationships between sequence homology and ORF organisation of RdRp genes of luteovirus and luteovirus-like RNAs. Dendrograms were derived from Fig. 4.4 and are drawn in proportion. Organisation of contiguous ORFs encoding replicase components are shown. Open boxes represent open reading frames and are drawn to rough scale. **GDD** represents the RdRp amino acid motif (Kamer and Argos, 1984) and is drawn in the approximate position of the motif in the ORF. Arrows indicate a frameshifting event, asterisks indicate a readthrough event. The sobemovirus RdRp is possibly processed after translation by protease cleavage. Abbreviations are given in the text (4.2.1).


A
GDD
Subgroup I
luteoviruses
RCNMV
*
GDD
CarMV
GDD
PEMV RNA 2
B
GDD
Sobemoviruses
GDD
Subgroup II
luteoviruses
GDD
MBV

The RdRp tree (Fig. 4.4) was generated using only three of the eight motifs as raw data, however comparisons based on all motifs, or using complete or partial sequences of the RdRp genes, largely reproduced the results presented here (data not shown).

4.3.2 Analysis of ORFs 3, 4 and 5

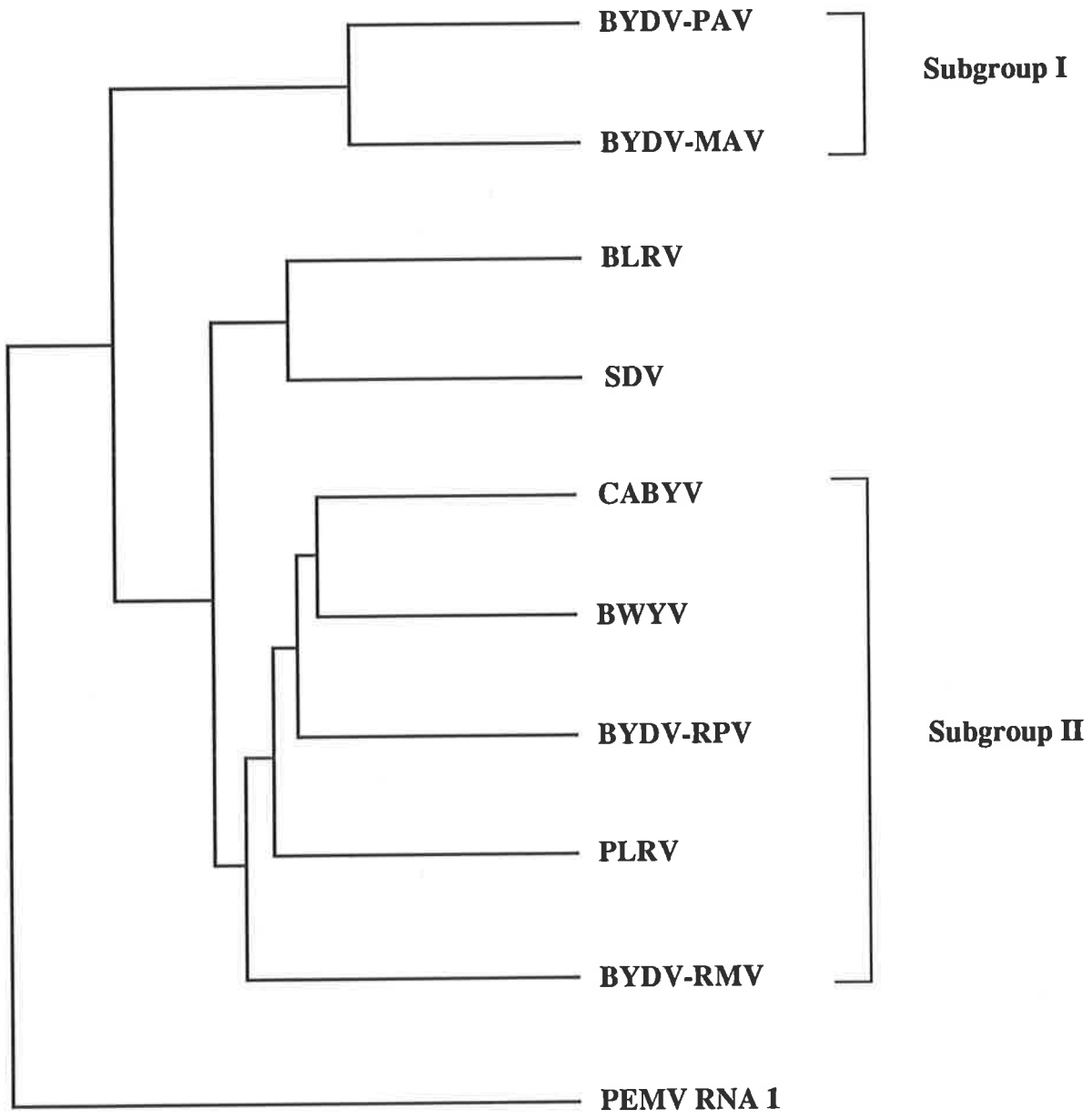
ORF 3. The coat protein of all luteoviruses is encoded by ORF 3, the first ORF to initiate after the intergenic sequence. This is also the case for SDV-Tas1, as the deduced amino acid sequence of ORF 3 can be aligned with partial sequences obtained from the capsid protein (data not shown; Dr A. A. Kortt, CSIRO Division of Biotechnology, Melbourne, Australia, unpublished data), and because of the high homology between the deduced amino acid sequence of SDV ORF 3 and those of the other luteovirus ORFs 3 (see below). There are a large number of deduced sequences for the coat proteins of the various luteoviruses available, which makes inference of a phylogeny possible. The luteovirus coat protein genes are sufficiently conserved that a motif strategy for sequence alignment is unnecessary. Alignment of the sequences and generation of a dendrogram using PileUp reveals the following relationships (Fig. 4.6): The coat proteins form two major clusters, with the coat protein of PEMV RNA 1 behaving as an outlier. The smaller cluster consists of the subgroup I luteoviruses BYDV-PAV and BYDV-MAV, while the larger consists of the subgroup II luteoviruses as well as the unclassified luteoviruses BYDV-RMV, SDV and BLRV. BYDV-RMV probably constitutes an orthodox group II luteovirus based on the similarity of its deduced coat protein sequence to the other members of group II, and the cytopathological symptoms caused by this virus upon infection (Gill and Chong, 1979b). However, SDV and BLRV behave quite differently in the comparison, forming a cluster distinct from that of the subgroup II luteoviruses. The following conclusions can be drawn from the data. Firstly, the coat proteins of SDV and BLRV are most closely related to each other and may form a subgroup of their own. Secondly, the SDV-like coat proteins are more closely related to those of the subgroup II than the subgroup I luteoviruses. This last point is highly significant because of the subgroup I

nature of the SDV 5' ORFs, and shows that SDV possesses a chimaeric genome containing elements of both subgroup I and II luteoviruses (4.4.1).

ORF 4. ORF 4 of the luteoviruses overlaps the coat protein gene (ORF 3). Various roles have been proposed for the protein product of ORF 4, including function as a VPg or cell-cell movement protein (reviewed in Chapter One). In SDV, ORF 4 encodes a protein of 189 amino acids (M_r 21 K) which is larger than that of BYDV-PAV or PLRV (M_r 17 K). Examination of the data for other luteovirus sequences reveals that ORF 4 varies in length from 141 amino acids (M_r 15.5 K) for BLRV to 192 amino acids (M_r 21 K) for CABYV (Table 4.2). This compares with the limited variation in length of the coat protein gene (ORF 3) of 196-208 amino acids (M_r 22-23 K). Comparisons between the deduced amino acid sequences of the luteoviral ORFs 4 reveal similar relationships to those found by coat protein sequence comparisons (data not shown). However, the overall level of sequence conservation is significantly less for ORF 4 than for ORF 3. As for the coat protein comparison, the deduced amino acid sequence of SDV ORF 4 is most similar to that of BLRV, despite the size difference of the ORFs. BLRV appears to contain a severe truncation at the amino terminus of the deduced ORF 4 protein relative to other luteoviruses (Table 4.2). Alignment of deduced protein sequences of SDV and BLRV in this region reveals that there are sequences upstream of the BLRV ORF 4 AUG putative initiation codon that show homology to sequences at the amino terminus of the deduced SDV ORF 4 protein (data not shown). Therefore it is possible that translation of BLRV ORF 4 initiates upstream of the first methionine codon in the sequence, albeit with initiation of translation at some codon other than a canonical AUG. On average, the large percentage of ORF 4 length variation occurs at the carboxyl-terminus of the deduced proteins; in some cases (CABYV and BYDV-RMV) the ORF 4 reading frame extends into the coding sequence of ORF 5.

ORF 5. ORF 5 of SDV potentially encodes a protein of 58 K. This would result in a protein of ~90 K after readthrough of the coat protein stop codon, as appears to occur in other luteoviruses (Bahner *et al.*, 1990; Reutenauer *et al.*, 1993; Cheng *et al.*,

Fig. 4.6. Phylogeny of coat protein sequences of luteoviruses and pea enation mosaic virus. The dendrogram was obtained by alignment of complete protein sequences using PileUp (4.2.2). Subgroup membership of the luteovirus sequences is indicated where known.


BYDV-PAV
BYDV-MAV
Subgroup I
BLRV
SDV
CABYV
BWYV
BYDV-RPV
PLRV
BYDV-RMV
Subgroup II
PEMV RNA 1

1994a). The deduced amino acid sequence of SDV ORF 5 shows a similar pattern of conservation to those of ORFs 3 and 4, with closest homology to the subgroup II luteoviruses (data not shown). The sequences of luteoviral ORFs 5 are conserved in the amino half of the deduced proteins, but highly variable at the carboxyl half (Mayo *et al.*, 1989). This type of sequence conservation extends to the deduced amino acid sequence of SDV ORF 5 (data not shown). The block of amino acids located in the centre of the variable carboxyl half of the deduced amino acid sequences of ORFs 5 that is conserved between BWYV, CABYV and PLRV (Guilley *et al.*, 1994) is not present in the SDV sequence (data not shown).

4.3.3 Putative transcriptional motifs in the nucleotide sequence of SDV-Tas1

The 5' terminal nucleotide sequence of SDV-Tas1 matches that of BYDV-PAV closely (31/45 identical nucleotides after the addition of gaps; Fig. 4.7). This sequence is different from that defined as consensus for the subgroup II luteoviruses (Keese *et al.*, 1990); thus there is a correlation between the 5' ORF organisation and the 5' terminal nucleotide sequence in the luteovirus group. Conserved sequences at the 5' genomic termini of positive-strand RNA viruses are considered to encode recognition sites for the viral RdRp, albeit active in the minus-sense RNA strand. The 3' terminus of the SDV-Tas1 genomic RNA encodes the 5'-CCC-3' motif which is also present at the 3' terminus of plant RNA viruses encoding a carmovirus-like RdRp ORF; CarMV (Guilley *et al.*, 1985); turnip crinkle virus (Carrington *et al.*, 1989); BWYV ST9-aRNA (Chin *et al.*, 1993); PEMV RNA 2 (Demler *et al.*, 1993); and RNAs 1 and 2 of RCNMV (Lommel *et al.*, 1988; Xiong and Lommel, 1989). A lack of homology (or complementarity) between the 5' and 3' genomic sequences exists in many plant RNA virus groups, including the subgroup II luteoviruses (Miller *et al.*, 1994). The similarity between BYDV-PAV and SDV-Tas1 at their respective genomic RNA termini strengthens the relationship established by the homology of their ORFs 1 and 2.

In addition to the homologies at the 5' genomic termini that occur within diverse plant RNA virus groups, internal genomic sequences may also show homology to the 5'

Table 4.2. Length variation (relative to PLRV ORF 4) at the C- and N- termini of luteovirus ORF 4 deduced amino acid sequences after multiple sequence alignment

Virus	Length (aa)	deduced M_r (K)	Initiation[a] rel. to PLRV	Stop[b] rel. to PLRV
BLRV	141	15.5[c]	+56	+33
BWYV	175	19.3	+9	+23
BYDV-MAV	154	16.9	+13	+8
BYDV-PAV	153	16.8	+13	+8
CABYV	191	21.0	+6	+41
PLRV	156	17.2	0	0
BYDV-RMV	192	21.1	+8	+41
BYDV-RPV	153	16.8	+3	-3
SDV	189	20.8	+6	+34

[a]Start point of coding region relative to PLRV (number of amino acid residues) after alignment using PileUp.

[b]Finishing point of coding region relative to PLRV (number of amino acid residues) after alignment using PileUp.

[c]Assuming an average base weight of M_r 110.

genomic ends. In many cases these have been proven to define the 5' terminus of viral subgenomic or messenger RNAs, by implication forming the negative strand promoter for synthesis of the subgenomic RNA (reviewed for the luteoviruses in Chapter One). A sequence with strong similarity to the 5' terminal 23 nt occurs at nucleotide 2731 of the SDV-Tas1 genome. Two lines of evidence suggest that this sequence represents the 5' end of a subgenomic RNA. There is close homology between the internal genomic sequence and the 5' end of the genomic RNA (21/27 nt after the addition of gaps; Fig. 4.7B). Secondly, the SDV internal genomic sequence and the 5' end of BYDV-PAV sgRNA 1 map to the same amino acid residue (**VK**) of the deduced ORF 2 protein (Fig. 4.7C; Kelly *et al.*, 1994). Initiation of SDV sgRNA 1 from nucleotide 2731 would give an RNA product of ~3100 nt, which is close to the value of 2800 nt derived from the molecular mass of sgRNA 1 dsRNA estimated by Smith *et al.* (1991).

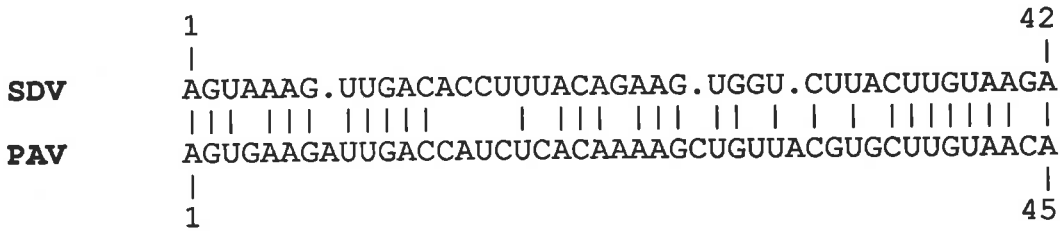
Only one subgenomic RNA has been observed in SDV infected tissue (Smith *et al.*, 1991). However, a search of the genome with the query sequence 5'-GUAAAG-3', which is completely conserved between the 5' genomic terminus and the putative sgRNA 1 5' terminus, revealed a homologous sequence at nucleotide 5600 of the SDV-Tas1 genome. This sequence shares 10/12 and 9/13 identical nucleotides with the 5' genomic and putative sgRNA 1 respectively after addition of gaps (data not shown). Transcription of a sgRNA from this point in the minus-strand genomic RNA would result in product of 261 nt. Such a sgRNA would be analogous to sgRNA 3 of BYDV-PAV (Kelly *et al.*, 1994) because of its small size (261 nt versus 329 nt for BYDV-PAV sgRNA 3) and lack of evident coding capacity.

4.3.4 Sequences for control of translation in the SDV genome

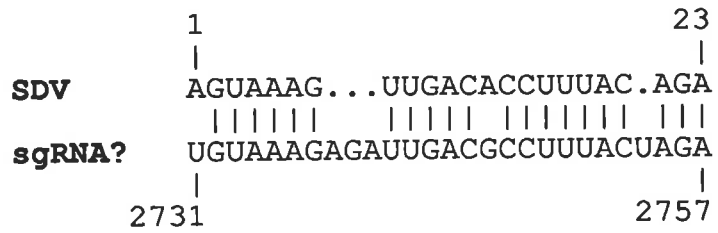
Other than internal initiation of translation for expression of ORFs 1 and 4, luteovirus RNAs are known to undergo two aberrant translational events. The first of these is -1 frameshifting from ORF 1 to express ORF 2 in BYDV-PAV (Brault and Miller, 1992). Frameshifting occurs at a 13 nt overlap between the ORFs at the slippery sequence 5'-GGGUUUU-3' (see Chapter One for a review). The organisation

Fig. 4.7. Putative transcriptional motifs in the sequence of SDV-Tas1. Sequence identity (SDV or BYDV-PAV) is indicated in bold on the left. Numbers in (A) and (B) correspond to nucleotide position in the genome. Alignment of sequences was by eye except for (C) where GAP was used. Insertion of gaps to maximise the alignment where necessary is indicated by dots. (A) Alignment of 5' terminal genomic sequences of SDV and BYDV-PAV. (B) Identification of a possible subgenomic RNA initiation site in the SDV genome by alignment with the 5'-terminal nucleotide sequence. (C) Alignment of the carboxyl-termini of the deduced protein sequences of SDV and BYDV-PAV ORFs 2. Amino acid residues corresponding to the known subgenomic RNA 1 start site of BYDV-PAV-Vic (Kelly *et al.*, 1994), and that proposed above for SDV-Tas1 in (B) are indicated in bold and underlined. Numbers refer to amino acid position in the deduced amino acid sequence of each ORF 2; asterisk indicates a stop codon.

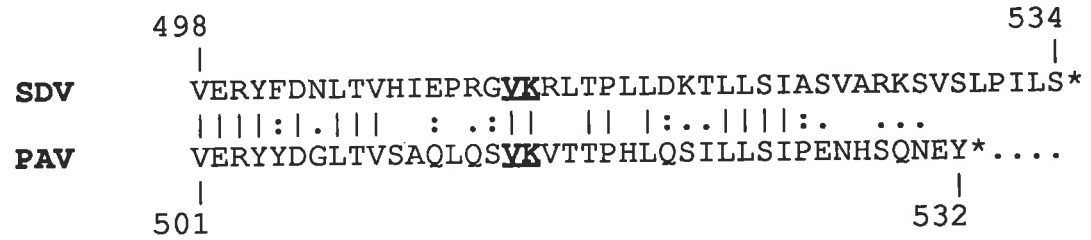
A


1
42
SDV
AGUAAAG.UUGACACCUUUACAGAAG.UGGU.CUUACUUGUAAGA
PAV
AGUGAAGAUUGACCAUCUCACAAAAGCUGUUACGUGCUUGUAACA
1
45

B


1
23
SDV
AGUAAAG...UUGACACCUUUAC.AGA
sgRNA?
UGUAAAGAGAUUGACGCCUUUACUAGA
2731
2757

C


498
534
SDV
VERYFDNLTVHIEPRGVKRLTPLLDKTLLSIASVARKSVSLPILS*
PAV
VERYYDGLTVSAQLQSVKVTTPHLQSILLSIPENHSQNEY*....
501
532

of the SDV genome is very similar to BYDV-PAV at the ORF 1/2 overlap, although it extends for 7 nt in SDV rather than 13 nt. SDV encodes a potential shifty sequence 5'-AGGUUUU-3' which is closely related to that of BYDV-PAV, so probably has the same function. Thus it is likely that ORF 2 of SDV is expressed by -1 frameshifting after translation of ORF 1.

The second aberrant translational event in luteovirus gene expression is readthrough of the ORF 3 stop codon to express ORF 5 (see Chapter One for review). Sequence requirements for readthrough are not known, however the following evidence suggests that SDV ORF 5 is also expressed by stop codon suppression. Firstly, ORF 5 is positioned contiguous and in-frame with ORF 3. The ORFs are separated by a single stop codon. Lastly, the sequence surrounding the stop codon is 5'-AAAUAGGUAGA-3' which is identical in all luteoviruses sequenced to date and possibly has a role in translational readthrough.

4.4 Discussion

4.4.1 Evolution of SDV

The genome of SDV-Tas1 contains familiar features of luteovirus genome organisation, but analysis of nucleotide and deduced amino acid sequences reveals diverse origins of the 5' and 3' coding blocks. Thus the 5' half, including the 5' noncoding region and ORFs 1 and 2, is closely related to the subgroup I luteoviruses and other carmo-like viruses, whereas the 3' coding region, which encodes the three genes common to all luteoviruses, is more closely related to subgroup II. Such a chimaeric form most likely arose by RNA recombination, although it is not clear if the SDV genome was formed as a result of recombination between the pre-existing luteoviral subgroups, or as a reiteration of the original event proposed by Miller *et al.* (1994) leading to the formation of the subgroup I from the subgroup II luteoviruses.

The sequences of SDV-Tas1 and the subgroup II luteoviruses BWYV and PLRV contain areas of homology throughout the length of the intergenic region (data not shown). In addition, SDV and the subgroup II luteoviruses have an intergenic

region of ~200 nt, while that of subgroup I is ~120 nt. Thus there is an association between the subgroup homology of the 3' coding block and the size of the internal noncoding sequence. Significant homology exists between the deduced amino acid sequences of ORFs 2 of SDV and BYDV-PAV towards the carboxyl terminus of the putative protein (Fig. 4.7C). Therefore, the recombination event that paired the two halves of the SDV genome most likely occurred between the 3' end of the polymerase gene (ORF 2) and the 5' end of the intergenic region of a subgroup II luteovirus. This is a similar scenario to that proposed to occur in the formation of the subgroup I luteovirus genome, with a RCNMV-like ancestor as the donor of the 5' ORFs (1.5; Miller *et al.*, 1994).

There is some evidence to suggest that the SDV genome arose as a consequence of recombination between subgroup I and II luteoviruses. Firstly, the relationship between ORFs 2 of SDV and BYDV-PLRV is approximately as close as that between the various ORFs 2 of the subgroup II luteoviruses (Fig. 4.4). This implies that the SDV RdRp donor was a subgroup I luteovirus rather than an ancestral virus, which would presumably be not as closely related to BYDV-PAV as is SDV. Further evidence to strengthen this hypothesis comes from the position and coding nature of the putative subgenomic promoter sequences in ORF 2. These occur at the same point in ORF 2 of both BYDV-PAV and SDV, and encode the same amino acid sequence (Fig. 4.7C). Interestingly, and perhaps significantly, the putative subgenomic promoter of RCNMV RNA 1 also maps to the same position in the RdRp ORF, as approximated by alignment of the deduced amino acid sequences using GAP (data not shown). However, neither the nucleotide or amino acid sequence of the putative subgenomic promoter is conserved between RCNMV and BYDV-PAV (or SDV), despite the high homology of the deduced protein sequences (39% identity between BYDV-PAV (or SDV) and RCNMV after the addition of gaps). Thus the poor conservation of the putative subgenomic promoter (which is nonetheless conserved with the 5' terminal genomic sequence in all three viruses) is evidence of the malleability of this sequence;

conversely its conservation between BYDV-PAV and SDV is evidence of recent divergence.

A similar argument can be made for the relationship between luteoviral coat protein sequences. The close relationship between SDV and the subgroup II luteoviruses relative to subgroup I at this point in the genome implies that the donor of the SDV 3' coding region was a subgroup II luteovirus rather than an ancestral virus. In any case it is likely that the luteovirus group evolved with the creation of ORF 4 (see below), so by this reasoning the donor of the genes had to be a luteovirus. The close relationship between the deduced coat protein (and ORF 4) sequences of SDV and BLRV is provocative. BLRV could be a close relative of SDV thus strengthening the contention that SDV possesses the third variant genome type in the luteovirus group (see below). Alternatively, if BLRV is an orthodox subgroup II luteovirus, then it is the obvious candidate as the donor of ORFs 3, 4 and 5 to SDV. Further characterisation of the BLRV genome is necessary before its full relationship to SDV is known.

There are further features which distinguish the SDV and subgroup I genomes in the region 3' of the ORF 5 coding sequence. All isolates of BYDV-PAV and -MAV encode a small 3' ORF (ORF 6; Miller *et al.*, 1988a; Ueng *et al.*, 1992; Chalhoub *et al.*, 1994) which is not contained in the genome of two distinct SDV isolates (Chapter Three). Evidence for the expression of ORF 6 is given by the existence of a small subgenomic RNA capable of expressing the ORF (see review in Chapter One; Young *et al.*, 1991; Kelly *et al.*, 1994). An additional small subgenomic RNA (sgRNA 3) is expressed in Australian strains of BYDV-PAV (Kelly *et al.*, 1994). Studies of RNAs expressed in SDV infection using Northern analysis and a probe corresponding to the 3' end of the SDV genome revealed only the genomic RNA and the large sgRNA 1 common to all luteoviruses (Smith *et al.*, 1991). Therefore, the coding and probable expression of ORF 6 is a major difference between the SDV and subgroup I genomes. Sequence analyses presented here suggest the existence of a possible subgenomic promoter at nucleotide 5600 of the SDV Tas-1 genome that would direct transcription of an RNA corresponding to BYDV-PAV sgRNA 3 (Kelly *et al.*, 1994). The small size

of such an RNA (~260 nt) may explain why it was not observed in the study of Smith *et al.* (1991).

Differences in genome organisation, expression and evolution distinguish SDV from the subgroup I luteoviruses. SDV therefore appears to possess the third variant genome type in the luteovirus group. Further appreciation of the relationship between SDV and subgroup I will be allowed by greater understanding of the role of the genomic region 3' of ORF 5 in both genome types. Characterisation of additional SDV-like genomes will also aid the classification of luteovirus genomes. In particular, the completion of the partial sequence of BLRV will reveal if it is an SDV homologue, or an orthodox subgroup II luteovirus. The existence of further SDV-like luteoviruses will strengthen the case for classification of the SDV genome as a third subtype.

4.4.2 Genetics of the luteoviruses; development of a strategy for investigation of the interaction of BYDV-PAV with the Yd2 resistance gene of barley

Evolution of plant viruses proceeds in part by recombinatorial exchange of functional gene clusters known as modules (Gibbs, 1987; Zimmern, 1988). This phenomenon is most clearly demonstrated in the luteovirus group, where subgroups I and II possess RdRp modules (ORFs 0-2) from diverse evolutionary lineages (Habili and Symons, 1989). Moreover, the 3' gene cluster (ORFs 3-5) also behaves as such a module, demonstrated by the evolution of the SDV genome (4.4.1). Localisation of the majority of ORFs to two clusters in this way (excluding ORF 6 of BYDV-PAV and -MAV) allows a rudimentary genetic analysis of the luteovirus genome. The characteristic luteovirus biological properties (aphid transmission, phloem limitation) are likely to be encoded by the 3' module, as these are the only ORFs common to both subgroups. While luteoviruses create typical symptoms in their hosts (yellowing, leafrolling) it is unlikely that symptomatology is specifically controlled by the virus. However, given that the symptoms of luteoviral infection may be a result of phloem necrosis (Jensen, 1969), luteoviral symptomatology can be tentatively assigned to the phloem limitation function which appears to reside in the 3' module as argued above.

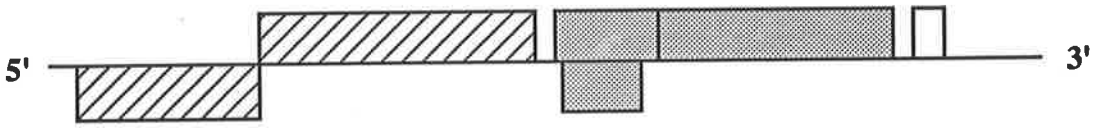
Phloem limitation may be specified by ORF 4, given that ORF 3 is known to encode the coat protein and ORF 5 likely encodes a structural protein controlling aphid transmission (Filichkin *et al.*, 1994), and the proposed role for ORF 4 as a (phloem-specific) movement protein (Tacke *et al.*, 1993).

This genetic analysis of luteoviral ORF function nominally precludes a VPg role for the product of ORF 4. Localisation of the ORFs to modules implies that the gene products of each module act independently of those of the other module. As the VPg is believed to play a role in replication and possibly transcription of the viral genome, by this analysis it should be encoded by the 5' module along with the putative RdRp gene (ORF 2). Such a placement agrees with the proposal by Miller *et al.* (1994) that the VPg is encoded by ORF 1 of the subgroup II genome. ORF 6 of the subgroup I genome is encoded separately from either module and as such does not conform to the analysis presented here. It may have evolved subsequent to the formation of the subgroup I from the subgroup II luteoviruses, thus distinguishing the new subgroup, similar to that proposed for the small overlapping ORF 2b which is present in the cucumoviruses but not the other groups of the family Bromoviridae (Ding *et al.*, 1994). ORF 6 of subgroup I is possibly required for replication as truncation of the ORF causes the virus to become non-infectious in plant protoplasts (Young *et al.*, 1991).

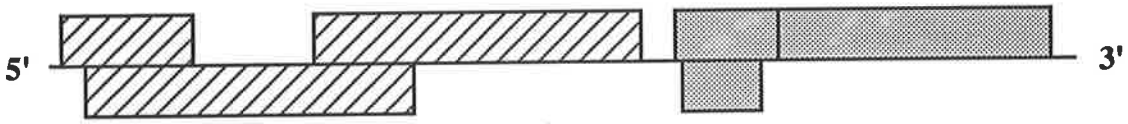
This thesis proposes a model for the functional organisation of the luteoviral genome (Fig. 4.8). In this model, the two ORF modules of the luteovirus represent the two major functions of the virus, replication and movement. 'Movement' here refers to cell-cell movement (presumably governed by ORF 4), long-distance movement within the plant (presumably governed by the viral particle (ORFs 3 and 5)), and transmission from plant to plant via aphids (presumably governed by ORF 5). 'Replication' refers to genome replication and transcription, which must be governed by ORF 2, and by implication ORF 1, as ORF 2 is expressed as a frameshift protein with the product of ORF 1 in both subgroups (Brault and Miller, 1992; Prüfer *et al.*, 1992). ORF 0 is also likely to be involved in replication because of the association of this ORF with ORFs 1

Fig. 4.8. Functional organisation of the luteoviruses. Genomes representative of luteoviral subgroups I (BYDV-PAV) and II (PLRV) as well as soybean dwarf virus are shown. Genes implicated in viral replication (diagonal lines; ORFs 0-2) or movement (includes cell-cell, long distance, and plant-aphid-plant movement; stippled boxes; ORFs 3-5) are distinguished.

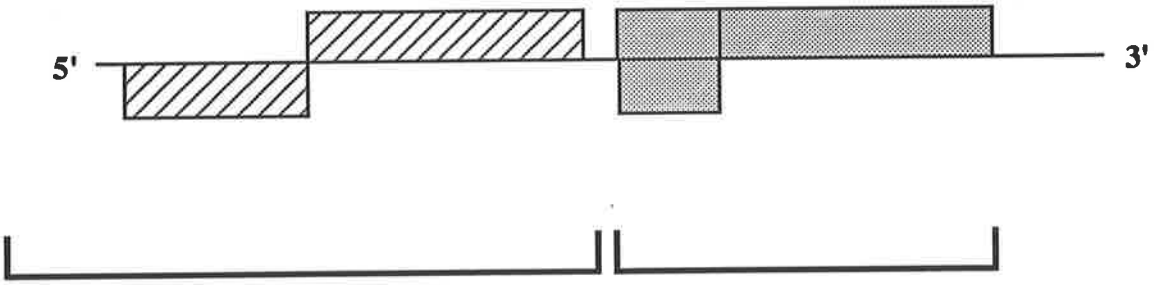
Subgroup I



Subgroup II



SDV



and 2 in the subgroup II luteoviruses and PEMV RNA 1, as well as the mycovirus MBV (4.3.1), thus conforming to the module concept.

The functional organisation of the luteoviruses as proposed above gives rise to a basis for the investigation of the interaction of BYDV-PAV with the Yd2 resistance gene of barley. Studies on the mode of action of disparate viral resistance genes show that they appear to act either to contain virus multiplication (≈replication) or localise infection to the site of inoculation (≈movement; Fraser, 1987). This is convenient, because not only does it reflect the two major functions of the virus, but also the functional organisation of the luteovirus genome. It follows that 'domain-swapping' (or in this case, module swapping) experiments can be designed for BYDV-PAV and BYDV-RPV to detect the viral ORF controlling the interaction with the Yd2 gene. BYDV-PAV, but not BYDV-RPV, is restricted by Yd2. Therefore, reciprocal exchange of the movement module between the viruses has the potential to change the interaction of the parent virus with the Yd2 gene. If there is no change in the interaction with the resistance gene, then resistance must act at the level of replication. The opposite case (change of the interaction between the parent virus and Yd2) would identify movement as the target of resistance. Preliminary results obtained in this way could be extended to identify the ORF responsible for Yd2 interaction.

4.4.3 Evolution of the luteoviruses

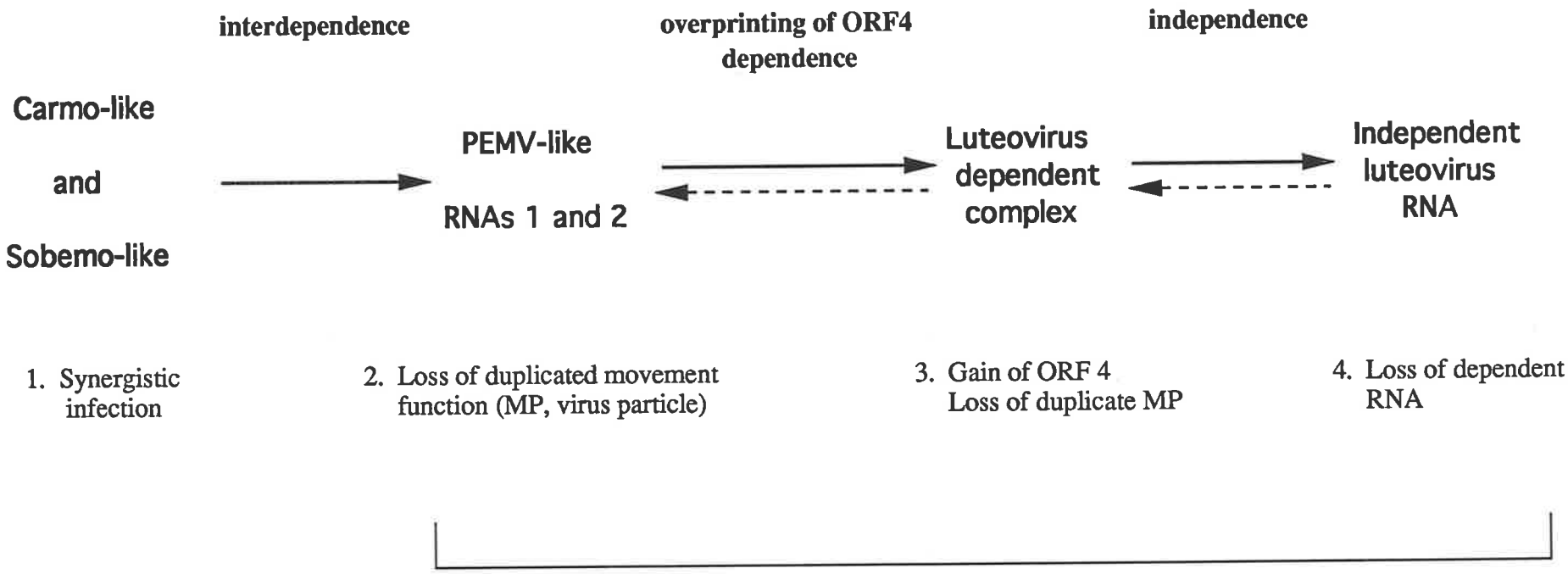
Certain luteoviruses form synergistic relationships with other luteo or luteo-like viruses across subgroup boundaries. These include strains of BYDV, which cause more severe symptoms when two viruses of each subgroup infect the same host than infection with either virus alone (reviewed in Chapter One). Furthermore, luteoviruses form mixed infections with dependent viruses that are known as persistent complexes (Murant, 1993). Such complexes include a competent luteovirus and a second virus or viral RNA that is dependent on the luteovirus for aphid transmission and probably encapsidation. Dependent viruses (proposed to be members of the new 'umbravirus' group; Murant, 1993) can only be transmitted mechanically when they exist in single

infections. Where known, persistent complexes appear to be composed of a subgroup II luteovirus and a dependent virus with affinities to subgroup I luteoviruses. The only such dependent viral RNA genome that has been sequenced is the BWYV-ST9 aRNA (Passmore *et al.*, 1993), which has affinities to subgroup I luteoviruses. The BWYV-ST9 aRNA possesses a putative RdRp gene but is dependent on BWYV for efficient movement in plants and for aphid transmission (Passmore *et al.*, 1993). Double infection with the BWYV-ST9 aRNA increases the yield of BWYV virus particles by ~10-fold (Chin *et al.*, 1993).

The other major example of luteovirus-like RNAs involved in a dependent (or interdependent) relationship is PEMV. In this example (reviewed in Chapter One), RNA 1 is subgroup I-like but lacks the ability to move from cell to cell. RNA 2 is subgroup II-like but is dependent on RNA 1 for encapsidation and aphid transmission, while providing the cell-cell movement function for both RNAs. Both RNAs carry genes for their independent replication. PEMV is thus analogous to the luteovirus persistent complexes in that both RNAs replicate autonomously, but is different to the extent that the RNAs are interdependent for some aspect of viral movement.

Luteovirus-like sequences thus form an evolutionary continuum, the common feature being the representation of carmo-like and sobemo-like RdRp modules at all evolutionary points. A direction of evolution can be inferred by hypothesising that the starting point of luteovirus evolution was the synergistic double infection of a common host by carmo-like and sobemo-like viruses (Fig. 4.9). Continued replication of these viruses in the presence of each other could lead to the development of interdependence similar to that exhibited by PEMV RNAs 1 and 2. Formation of such a relationship can be explained as loss of duplicated factors; in the case of PEMV, viral coat protein and cell-cell movement function. The event which separated the luteoviruses from other virus groups was most likely the overprinting of ORF 4 in the coat protein reading frame (Keese and Gibbs, 1992). In the example of PEMV, creation of ORF 4 in the RNA 1 coat protein reading frame would abolish the dependence of RNA 1 on RNA 2 for cell-cell movement. That ORF 4 encodes a phloem-specific movement gene is

Fig. 4.9. Possible pathway for the evolution of the luteoviruses from a synergistic infection of carmo-like and sobemo-like viruses. See text for details.


interdependence
overprinting of ORF4 dependence
independence
Carmo-like
and
Sobemo-like
PEMV-like
RNAs 1 and 2
Luteovirus dependent complex
Independent luteovirus RNA
1. Synergistic infection
2. Loss of duplicated movement function (MP, virus particle)
3. Gain of ORF 4
Loss of duplicate MP
4. Loss of dependent RNA
Recombination event involving replication module

central to this thesis; in addition to the evidence for this presented in Chapters one and four, it follows that a new cell-cell movement function encoded in the sequence of a coat protein/readthrough structure that potentiates aphid transmission should limit virus movement to that tissue targeted by the aphid.

Gain of movement function to the prototypical luteovirus genome would allow two further directions of evolution. The first is obviously separation of the RNAs and emergence of the new luteo sequence as an independent virus (this may also have occurred prior to the creation of ORF 4). The second is formation of the luteovirus persistent complexes. Using the BWYV-ST9 aRNA as an example, continued association of a progenitor luteovirus sequence with the second RNA could lead to loss of the cell-cell movement function of the second RNA, using the previous argument of loss of duplicated function. However, the RNAs would continue to exist in complex because of the synergistic effects of the divergent RdRp genes (*e.g.* Passmore *et al.*, 1993). The luteovirus sequence could exit (or re-enter) this relationship at any time.

While this model places the hypothesised progenitor carmo- and sobemo-like viruses at one end of the luteoviral evolutionary spectrum, and the luteoviruses at the other, it does not follow that each step in the pathway is unidirectional. In particular, it is possible that ORF 4 of the luteoviruses could be lost as the luteovirus entered an interdependent relationship with another RNA (as could explain the evolution of PEMV), or at other stages gain of function via RNA recombination is possible. The continued association of carmo and sobemo-like RdRp modules provides ample opportunity for the recombination event leading to the formation of luteovirus subgroup I.

4.4.4 Molecular taxonomy of plant RNA viruses

One of the tenets of current molecular taxonomy is the central importance of the RdRp in viral evolution. Koonin and Dolja (1993) have written that

"...the view of a virus as a relatively stable, slowly evolving "core" of the replicative genes accompanied by a much more flexible "shell" of the genes coding for

virion components and "accessory" proteins appears to be a strongly preferable, and in a sense the "correct" one. This concept implies that the derived phylogeny of the "core" gene complex should constitute the basis of phylogenetic taxonomy of positive-strand RNA viruses, at least when higher taxa are considered."...

This view derives from the fact that only the RdRp gene is universally represented in the genomes of positive-strand RNA viruses, and that all known RdRps share conserved sequence domains and thus evolved from a common ancestor (Altschul *et al.*, 1990; Koonin and Dolja, 1993). However, the example of genome evolution exhibited by the luteovirus group demonstrates that the RdRp should not be regarded as the basic unit of evolution. These enzymes appear to be encoded as functional modules which are interchangeable between divergent genomes. The nucleotide sequences recognised by individual RdRps in replication and transcription of the viral genome appear to be relatively simple (*e.g.* Fig. 4.5), so are easily obtained in a context of high mutation rates. Phylogenies based upon the sequence of this enzyme therefore merely represent the phylogeny of the enzyme itself, rather than of the virus genome encoding the RdRp.

The use of phylogenetic analysis of viral proteins as a means of establishing whole virus phylogeny must be questioned. Such analyses rely on a narrow view of evolution predicated on random nucleotide change as the sole mechanism for sequence evolution. However, RNA recombination, both homologous (approximated by the evolution of the SDV genome from subgroup I and II ancestors) and non-homologous (demonstrated by the evolution of the two luteoviral subgroups) is a major force in the genome evolution of positive-strand RNA viruses (Lai, 1992). A phylogenetic approach to taxonomic classification is limited in this context. Hierarchical evolutionary structures can not represent the horizontal transfer of genetic information mediated by RNA recombination.

CHAPTER FIVE

CONSTRUCTION OF A PLASMID VECTOR FOR *IN PLANTA* TRANSCRIPTION OF PLANT VIRUS cDNAs

5.1 Introduction

The strategy for investigation of the interaction between BYDV-PAV and the Yd2 resistance gene of barley outlined in Chapter Four (4.4.2) is based upon manipulation of the viral genome. BYDV possesses an RNA genome so it is necessary to construct full-length infectious cDNA clones of BYDV-PAV and -RPV before manipulation of sequences is possible. Such a clone exists for BYDV-PAV, and is composed of a full-length BYDV-PAV cDNA fused at the 5' end of the viral genome to a bacteriophage T7 RNA polymerase promoter (Young *et al.*, 1991). Linearisation of the clone at the 3' end of the viral sequence allows the *in vitro* transcription of positive-sense RNA copies directed by T7 RNA polymerase. Synthetic RNA molecules made in this way produce infectious replicating virus particles when electroporated into *Triticum monococcum* protoplasts (Young *et al.*, 1991).

A protocol based on electroporation of protoplasts with synthetic RNAs is unsuitable for the purposes of this thesis. This is because the Yd2 gene is apparently inactive in plant protoplasts (Larkin *et al.*, 1991), so assay of recombinant viruses with respect to the Yd2 gene must take place in intact plants. This necessitates the use of aphids to transfer progeny virus from protoplasts to the plant. The use of synthetic RNA and protoplasts has the drawbacks of the technical difficulties in maintaining and inoculating protoplasts, as well as feeding aphids on the infected protoplasts, with the associated potential for environmental release of the recombinant virus. Further drawbacks to the use of *in vitro* transcribed clones are the expense and technical intricacy of the transcription process (Boyer and Haenni, 1994).

A simpler technique for the inoculation of plants with cloned virus nucleic acids is agroinfection (Grimsley *et al.*, 1986). Tandemly repeated copies of the viral genome are cloned into a plasmid vector able to replicate in both *Agrobacterium* and *E. coli* known as a binary vector. The binary vector contains the repeat elements necessary for transfer of the *Agrobacterium* T-DNA to the nucleus. The viral sequences are cloned between the T-DNA repeats, then transferred to *Agrobacterium* containing a helper plasmid encoding the functions necessary for nuclear transfer of the T-DNA. Plants are

inoculated with dense suspensions of *Agrobacterium* containing both the recombinant binary vector and the helper plasmid. Although the exact process of viral infection is not known, it is believed that the viral cDNA circularises by homologous recombination after excision and nuclear transport of the T-DNA (Grimsley, 1990). Agroinfection has been used successfully for the infection of monocotyledonous plants with phloem limited viruses, in which case the inoculation must be targeted to meristematic tissue (Grimsley and Bisaro, 1987; Lazarowitz, 1988).

Adaptation of agroinfection for the use of RNA viruses requires the addition of a plant RNA polymerase II promoter sequence to the T-DNA (Turpen *et al.*, 1993). This is necessary to direct the *in planta* transcription of positive-sense RNA molecules from the viral cDNA, which could be subsequently replicated by the virally encoded RdRp. This Chapter describes the construction of a plasmid vector suitable for synthesis of viral cRNA containing plant RNA polymerase promoter and terminator sequences. The promoter is modified such that viral sequences can be inserted at the first nucleotide of the transcribed sequence. This is important because non-viral nucleotides at the 5' end of the transcript can severely decrease or abolish infectivity (Boyer and Haenni, 1994). The utility of this vector is demonstrated in the construction of full-length infectious cDNA clones of cucumber mosaic cucumovirus.

5.2 Materials and Methods

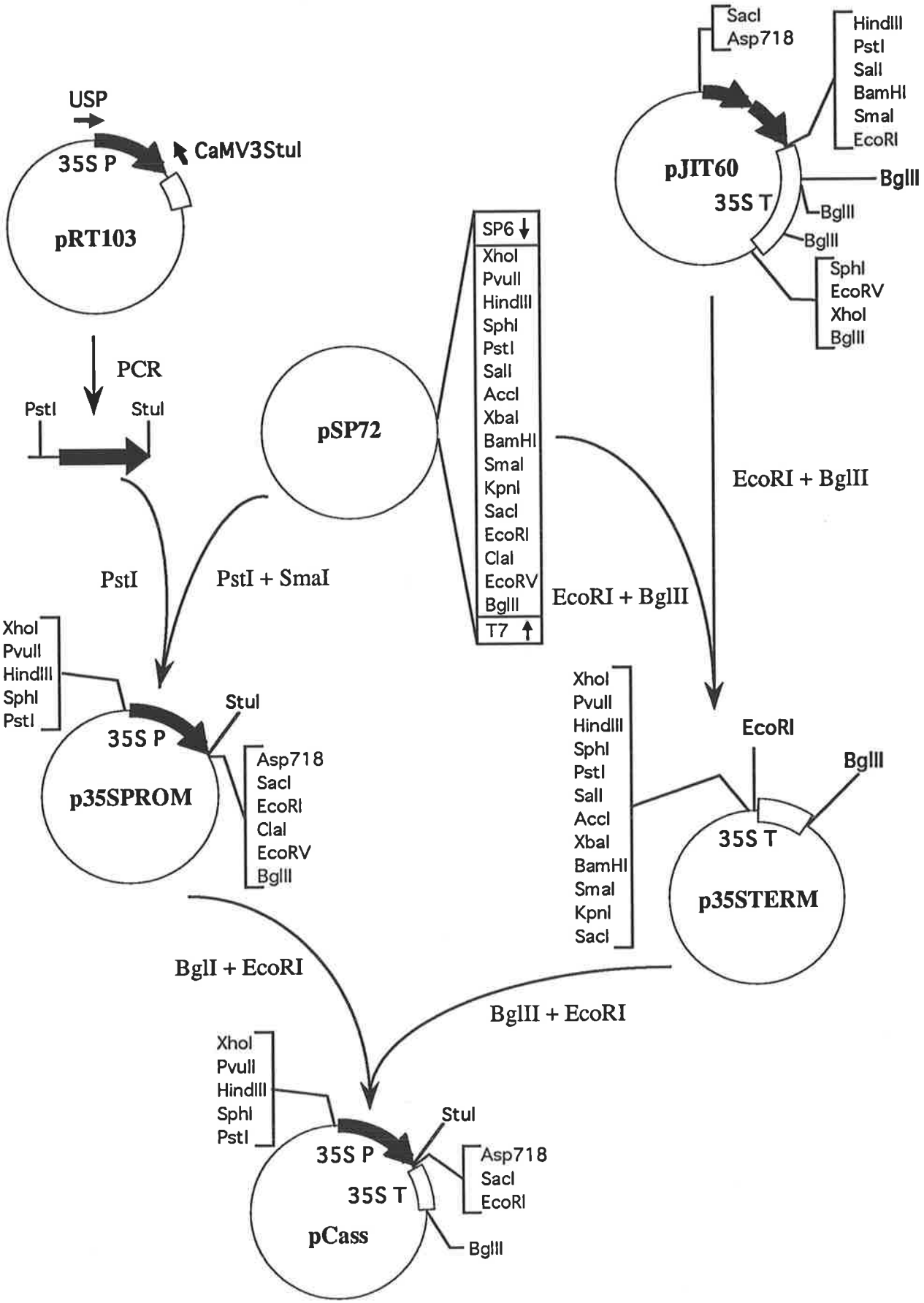
5.2.1 Acknowledgments

The work presented here involving construction of full-length cDNA clones of cucumber mosaic virus genomic RNAs into pCass, and infection studies conducted using these clones, was carried out in collaboration with Dr Shou-Wei Ding and Ms Wan-Xiang Li.

5.2.2 Construction of a vector for *in planta* transcription of viral cDNAs

Mutagenesis of the CaMV 35S promoter. The promoter that drives transcription of the 35S RNA of cauliflower mosaic virus (CaMV) was obtained as a cloned

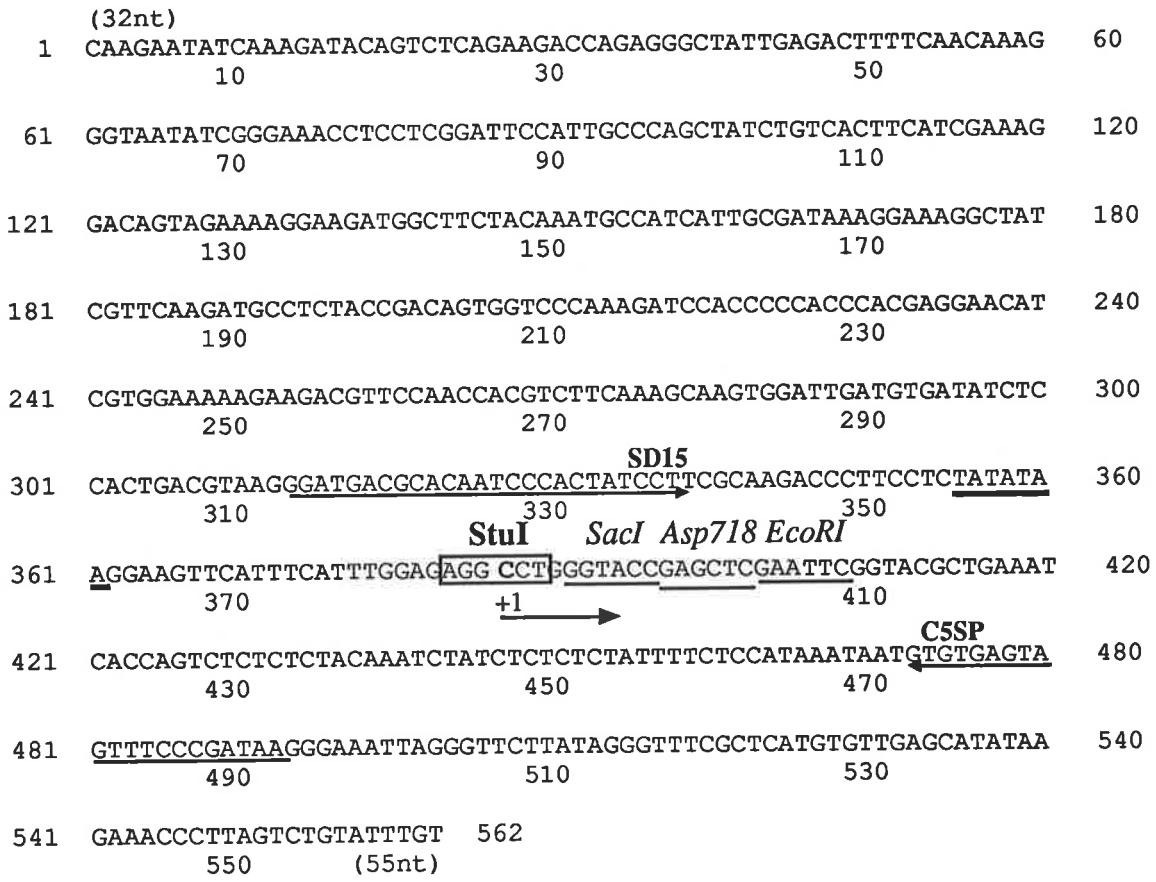
Fig. 5.1. Assembly of pCass. Structures of plasmids and restriction sites are given. Arrows indicate the direction of recombinant manipulations. The CaMV 35S promoter is represented as a thick black arrow. Small filled arrows represent PCR primers. A detailed restriction map of pCass is shown.


USP
35S P
CaMV3StuI
pRT103
PCR
PstI
StuI
pSP72
SP6
XhoI
PvuII
HindIII
SphI
PstI
SalI
AccI
XbaI
BamHI
SmaI
KpnI
SacI
EcoRI
ClaI
EcoRV
BglII
T7
SacI
Asp718
HindIII
PstI
SalI
BamHI
SmaI
EcoRI
pJIT60
35S T
BglII
BglII
BglII
SphI
EcoRV
XhoI
BglII
EcoRI + BglII
PstI
PstI + SmaI
EcoRI + BglII
XhoI
PvuII
HindIII
SphI
PstI
StuI
35S P
p35SPROM
Asp718
SacI
EcoRI
ClaI
EcoRV
BglII
XhoI
PvuII
HindIII
SphI
PstI
SalI
AccI
XbaI
BamHI
SmaI
KpnI
SacI
EcoRI
BglII
35S T
p35STERM
BglI + EcoRI
BglII + EcoRI
XhoI
PvuII
HindIII
SphI
PstI
StuI
35S P
35S T
pCass
Asp718
SacI
EcoRI
BglII

fragment in the plasmid pRT103 (Töpfer *et al.*, 1987). The sequence corresponding to the transcriptional start site of the CaMV 35S promoter was modified to contain a recognition site for the restriction endonuclease *Stu*I using oligonucleotide-directed PCR mutagenesis (Fig. 5.1; Higuchi *et al.*, 1988). PCR was performed using the M13 universal sequencing primer (USP) as the upstream primer and oligonucleotide CaMV3StuI (5'-AGGCCTCTCCAAATGAAATGAAC-3'; Yamaya *et al.*, 1988) containing the *Stu*I modification (underlined bases) as the downstream primer. PCR was performed on 0.1 ng of pRT103 template DNA, using Vent DNA polymerase according to the manufacturer's specifications except that the concentration of $MgSO_4$ in the reaction mixture was 3 mM. The reaction profile was [94°C/5 s; 47°C/5 s; 72°C/30 s]$_{25}$; 72°C/5 min; 25°C/5 min and was carried out on a capillary DNA Thermal Sequencer (Corbett, Australia). Reaction products were electrophoresed on a 2% agarose 1xTBE minigel and the major DNA species of ~480 nt excised after ethidium bromide/UV visualisation. The DNA was extracted from the gel slice using Geneclean, digested with *Pst*I (cuts at the 5' end of the promoter sequence corresponding to the pRT103 polylinker), and ligated into *Pst*I-*Sma*I digested pSP72 (Promega) to create the clone p35SPROM. Faithful incorporation of the mutation was verified by dideoxy sequencing (data not shown; Fig. 5.2).

Completion of the vector. The CaMV 35S transcriptional termination signal necessary to complement the modified promoter was obtained from the plasmid pJIT60 (Fig. 5.1), a pUC-based plasmid containing the 35S RNA transcriptional promoter and terminator sequences of CaMV (a gift of Dr I.B. Dry, CSIRO Division of Horticulture, Glen Osmond, South Australia). The plasmid was digested with *Eco*RI and *Bgl*II to release a 240 nt fragment of the transcription terminator, which was purified and ligated into pSP72 also restricted with *Eco*RI and *Bgl*II to create p35STERM. The terminator sequence was subcloned from p35STERM by excision with *Eco*RI and *Bgl*II and insertion into *Eco*RI-*Bgl*II restricted p35SPROM to create pCass. Thus the CaMV 35S terminator and modified promoter flank a short polylinker of four restriction enzyme sites (Fig. 5.1). The completed vector was partially sequenced to confirm the CaMV

Fig. 5.2. Partial sequence of the transcriptional cassette of pCass. Approximately 32 nt of the 35S promoter 5' sequence, and 55 nt of the 35S terminator 3' sequence, are missing by comparison to the sequence of pRT103 (data not shown). The *Stu*I restriction site is boxed, and the sites for *Sac*I, *Asp*718 and *Eco*RI shown. The position of the +1 nucleotide of transcription is indicated (arrow). Primer binding sites for sequencing of cloned inserts are underlined (thin arrows); primer SD15 allows 5' sequences to be read while primer C5SP enables sequencing of 3' sequences. The TATA box of the 35S promoter is underlined (thick line).


(32nt)
1 CAAGAATATCAAAGATACAGTCTCAGAAGACCAGAGGGCTATTGAGACTTTTCAACAAAG 60
10 30 50
61 GGTAATATCGGGAAACCTCCTCGGATTCCATTGCCCAGCTATCTGTCACTTCATCGAAAG 120
70 90 110
121 GACAGTAGAAAAGGAAGATGGCTTCTACAAATGCCATCATTGCGATAAAGGAAAGGCTAT 180
130 150 170
181 CGTTCAAGATGCCTCTACCGACAGTGGTCCCAAAGATCCACCCCCACCCACGAGGAACAT 240
190 210 230
241 CGTGGAAAAAGAAGACGTTCCAACCACGTCTTCAAAGCAAGTGGATTGATGTGATATCTC 300
250 270 290
SD15
301 CACTGACGTAAGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTCTATATA 360
310 330 350
StuI SacI Asp718 EcoRI
361 AGGAAGTTCATTTCATTTGGAGAGG CCTGGGTACCGAGCTCGAATTCGGTACGCTGAAAT 420
370 +1 410
C5SP
421 CACCAGTCTCTCTCTACAAATCTATCTCTCTCTATTTTCTCCATAAATAATGTGTGAGTA 480
430 450 470
481 GTTTCCCGATAAGGGAAATTAGGGTTCTTATAGGGTTTCGCTCATGTGTTGAGCATATAA 540
490 510 530
541 GAAACCCTTAGTCTGTATTTGT 562
550 (55nt)

35S promoter and terminator sequences and the restriction sites flanked by these elements (Fig. 5.2).

5.2.3 cDNA clones of cucumber mosaic cucumovirus genomic RNAs

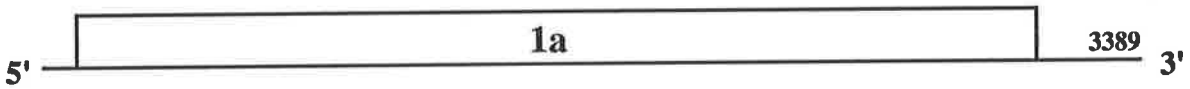
The tripartite virus cucumber mosaic cucumovirus (CMV) strain Q (Q-CMV; Francki *et al.*, 1966) was used in this work (Fig. 5.3). Full length cDNA clones of each genomic RNA designated pQCR1, 2 and 3 (corresponding to RNAs 1, 2 and 3 respectively) were generated by reverse transcription and PCR and cloned into the *Bam*HI and *Sma*I sites of pUC19 for RNAs 1 and 3, or pUC18 for RNA 2 (W.-X. Li and S.-W. Ding, unpublished data). The complete nucleotide sequences of all three Q-CMV genomic RNAs have been published (Rezaian *et al.*, 1984; Rezaian *et al.*, 1985; Davies and Symons, 1988).

5.2.4 Cloning a full-length cDNA of CMV RNA 2 into pCass

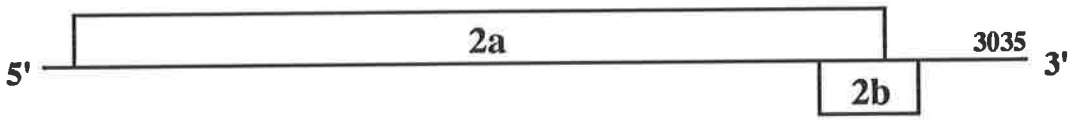
The full-length cDNA copy of RNA 2 of CMV strain Q cloned into the *Bam*HI and *Sma*I sites of pUC18 (pQCR2; 5.2.3) was the starting point of this work. pQCR2 was digested with *Fsp*I (cuts at nucleotide 738 of Q-CMV RNA 2) and *Eco*RI (cuts 5' of nucleotide 1 in the pUC18 polylinker; Fig. 5.4). The blunt-sticky ended fragment of ~750 nt was cloned into *Eco*RI-*Sma*I digested pBluescript SK+ to create pSKR25. The precise 5' terminus, as well as the remaining 738 nt of Q-CMV and some plasmid polylinker sequences, was amplified from pSKR25 by PCR using oligonucleotides SD9 (5'-GTTTATTCTCAAGAGCGTATGG-3'; homologous to nucleotides 1-22 of RNA 2) and the T3 primer which is complementary to the bacteriophage T3 promoter sequence of pBluescript. The PCR was performed with Vent DNA polymerase to ensure blunt ends, using a thermocycling profile of [94°C/1 min; 50°C/1 min; 72°C/2 min]$_{30}$ on a DNA Thermal Sequencer (Corbett). The single reaction product was resolved on a 1.5% agarose/1xTBE minigel, excised and isolated from the gel slice using Geneclean, then digested with *Sac*I which cuts in the pBluescript polylinker 3' of the viral sequence. The *Sac*I-blunt ended PCR product was cloned into *Stu*I-*Sac*I digested pCass

Fig. 5.3. Genome organisation of cucumber mosaic cucumovirus (strain Q). The three genomic RNAs (RNAs 1, 2 and 3) are shown. Open boxes represent open reading frames with the following functions; ORF 1a and 2a are components of the viral RdRp; ORFs 2b and 3a control viral movement, and CP represent the viral coat protein gene. The size of each RNA in nucleotides is given at the 3' end of the molecule.

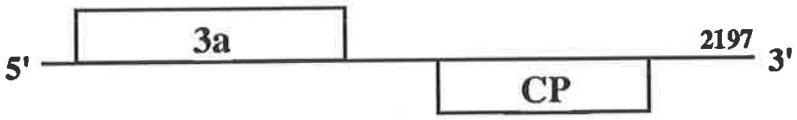
RNA 1



RNA 2



RNA 3

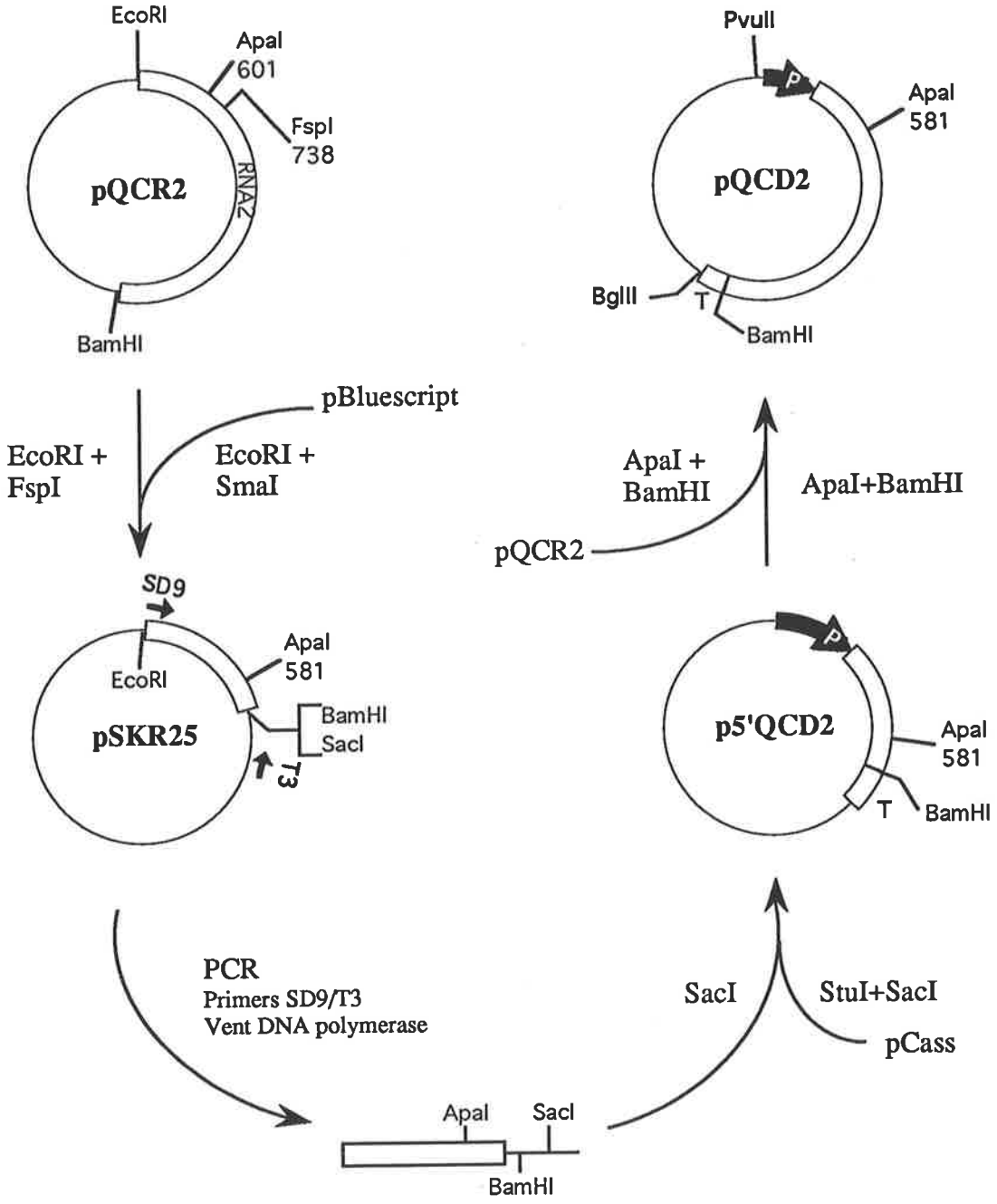


to create p5'QCD2. Correct fusion of the 35S promoter and 5' Q-CMV sequence was confirmed by dideoxy nucleotide sequencing (data not shown) using primer SD15 (Fig. 5.2). p5'QCD2 was restricted with *Apa*I, which cuts internal to the viral sequence at nucleotide 581, and *Bam*HI, which was carried over in the pBluescript polylinker and cuts 3' of the viral sequence. The remaining Q-CMV RNA 2 sequence was obtained by digesting pQCR2 with *Apa*I and *Bam*HI, and was inserted into the *Apa*I-*Bam*HI restricted p5'QCD2 vector to create pQCD2. The sequence integrity of the recombinant junctions was verified by sequencing of the DNA (data not shown).

5.2.5 Cloning a full-length cDNA of CMV RNA 3 into pCass

The starting point for this work was a full-length cDNA of Q-CMV RNA 3 (pQCR3) cloned into the *Sma*I and *Bam*HI sites of pUC19 (5.2.3). pQCR3 was digested with *Xba*I to release a fragment containing the 5' 1003 nt of Q-CMV RNA 3 and a few nucleotides derived from the plasmid polylinker fused to the 5' end of the viral sequence (Fig. 5.5). The fragment was cloned into *Xba*I-digested pBluescript SK+ to create pSKR35. The precise 5' viral terminus, as well as the complete viral sequence of pSKR35 and a portion of the pBluescript polylinker, was amplified by PCR using the oligonucleotide SD10 (5'-GTAATCTTACCACTTTCTTTCACG-3'; homologous to nucleotides 1-23 of Q-CMV RNA 3) and the T3 primer as above (5.2.4). The single reaction product of ~1000 nt was resolved on 1% agarose/1xTBE, excised and purified from the gel and digested with *Eco*RI (cuts in the pBluescript polylinker downstream of the viral sequence) to create a blunt-sticky ended fragment. This was ligated into *Stu*I-*Eco*RI digested pCass to create p5'QCD3, and the 35S promoter/5' viral sequence fusion was verified by sequencing (data not shown). The remaining RNA 3 sequences were excised from pQCR3 by digestion with *Sac*I (cuts at viral nucleotide 492) and *Bam*HI (cuts 3' of the viral sequence in the polylinker) and cloned into *Sac*I-*Bam*HI digested p5'QCD3. The sequence integrity of the recombinant junctions was verified by sequencing of the DNA (data not shown).

Fig. 5.4. Cloning of a full-length cDNA corresponding to CMV-Q RNA 2 into pCass. Interpretation of the diagram is as for Fig. 5.1. Restriction sites for release of the 35S transcriptional cassette containing the viral cDNA in pQCD2 are shown.


EcoRI
ApaI
601
FspI
738
pQCR2
RNAZ
BamHI
PvuII
P
ApaI
581
pQCD2
BglII
T
BamHI
pBluescript
EcoRI +
FspI
EcoRI +
SmaI
ApaI +
BamHI
ApaI+BamHI
pQCR2
SD9
EcoRI
ApaI
581
BamHI
SacI
pSKR25
T3
P
p5'QCD2
ApaI
581
T
BamHI
PCR
Primers SD9/T3
Vent DNA polymerase
SacI
StuI+SacI
pCass
ApaI
SacI
BamHI

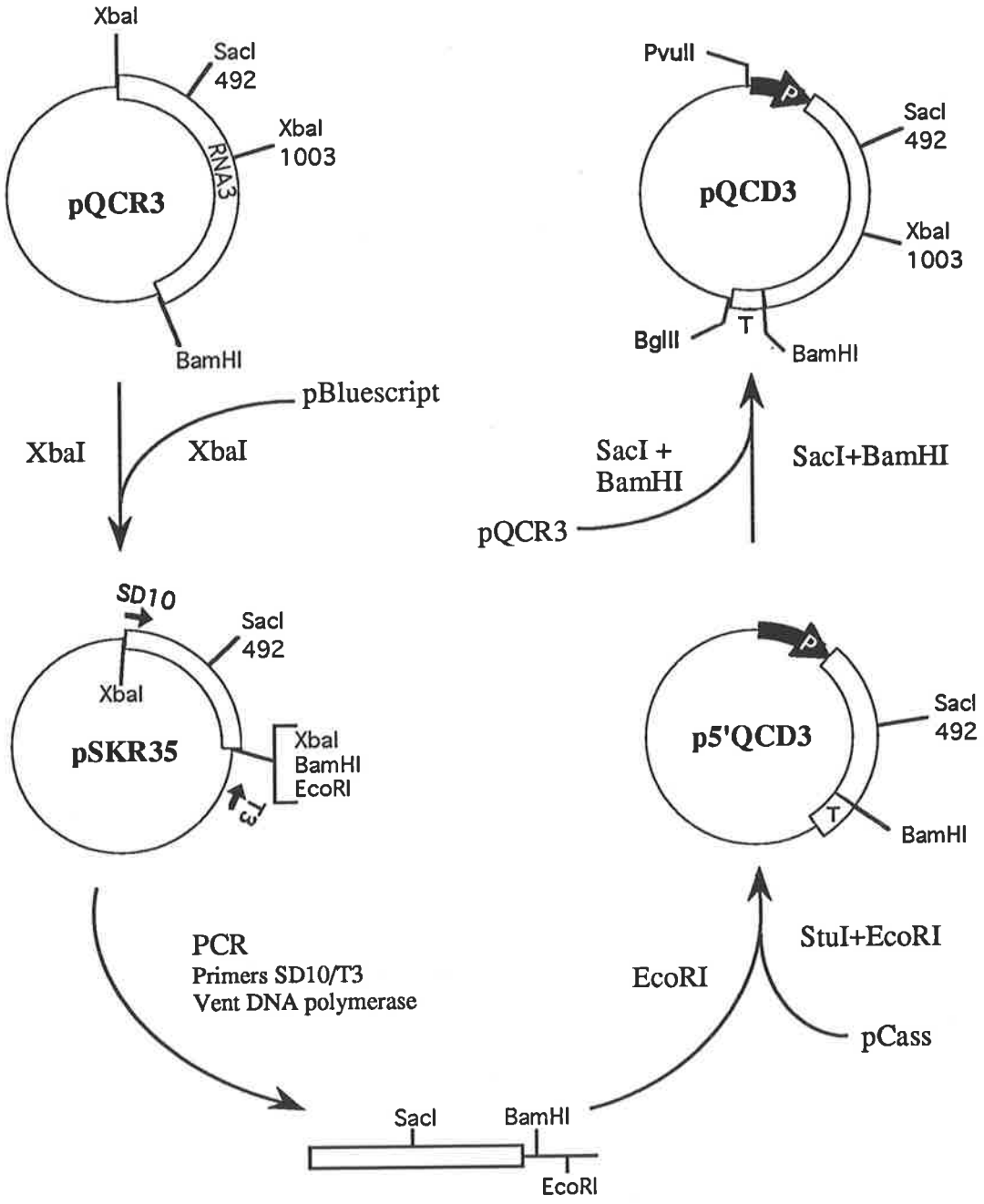
5.2.6 Cloning a full-length cDNA of CMV RNA 1 into pCass

A full-length cDNA of Q-CMV RNA 1 cloned into the *Bam*HI and *Sma*I sites of pUC19 (pQCR1) formed the basis of this work (5.2.3). pQCR1 was digested with *Eco*RI and *Fsp*I (cuts at nucleotide 539 of RNA 1) to release a 5' fragment with some vector sequences fused to the 5' terminal viral sequence (Fig. 5.6). The fragment was cloned into *Eco*RI-*Sma*I digested pBluescript SK+ to create pSKR15. The 5' viral sequence was amplified as previously (5.2.4) using oligonucleotide SD8 (5'-GTTTTATTTACAAGAGCGTACGG-3'; homologous to nucleotides 1-23 of RNA 1) and the T3 primer. The PCR product of ~550 nt was purified as above and cleaved with *Sac*I 3' of the viral sequence, then cloned into *Stu*I-*Sac*I restricted pCass to create p5'QCD1. The promoter-viral 5' end fusion was confirmed by dideoxy sequencing (data not shown). Repeated efforts to assemble the remaining RNA 1 sequences in p5'QCD1 failed, so an alternative strategy of cloning the 35S promoter/5' viral fusion and the 35S terminator into pQCR1 was employed. p5'QCD1 was digested with *Hind*III and the restriction site made blunt by end-filling with the Klenow fragment of *E. coli* DNA polymerase I and dNTPs. The promoter/5' viral fusion was then released from the linearised vector by digestion with *Nco*I. pQCR1 was digested with *Eco*RI upstream of the viral 5' end, and the restriction site made blunt by end-filling using Klenow fragment and dNTPs. The linearised vector was then treated with *Nco*I to release the 5' RNA 1 sequences and allow insertion of the promoter/5' viral cDNA fusion from p5'QCD1 (above) using a blunt-sticky end strategy to create pQCD1-T. The 35S terminator was released from pQCD3 (5.2.5 and Fig. 5.5) by digestion with *Bam*HI and *Bgl*II, and cloned into the *Bam*HI site of pQCDT-1 to create pQCD1. The sequence integrity of recombinant junctions was verified by sequencing of the DNA (data not shown).

5.2.7 Plant infections with cDNA clones of Q-CMV RNAs in pCass

Nicotiana glutinosa or cucumber (*Cucumis sativus* cv. Green Gem) plants were grown to the four-leaf or cotyledon stage respectively under natural lighting conditions,

Fig. 5.5. Cloning of a full-length cDNA corresponding to CMV-Q RNA 3 into pCass. Interpretation of the diagram is as for Fig. 5.1. Restriction sites for release of the 35S transcriptional cassette containing the viral cDNA in pQCD3 are shown.

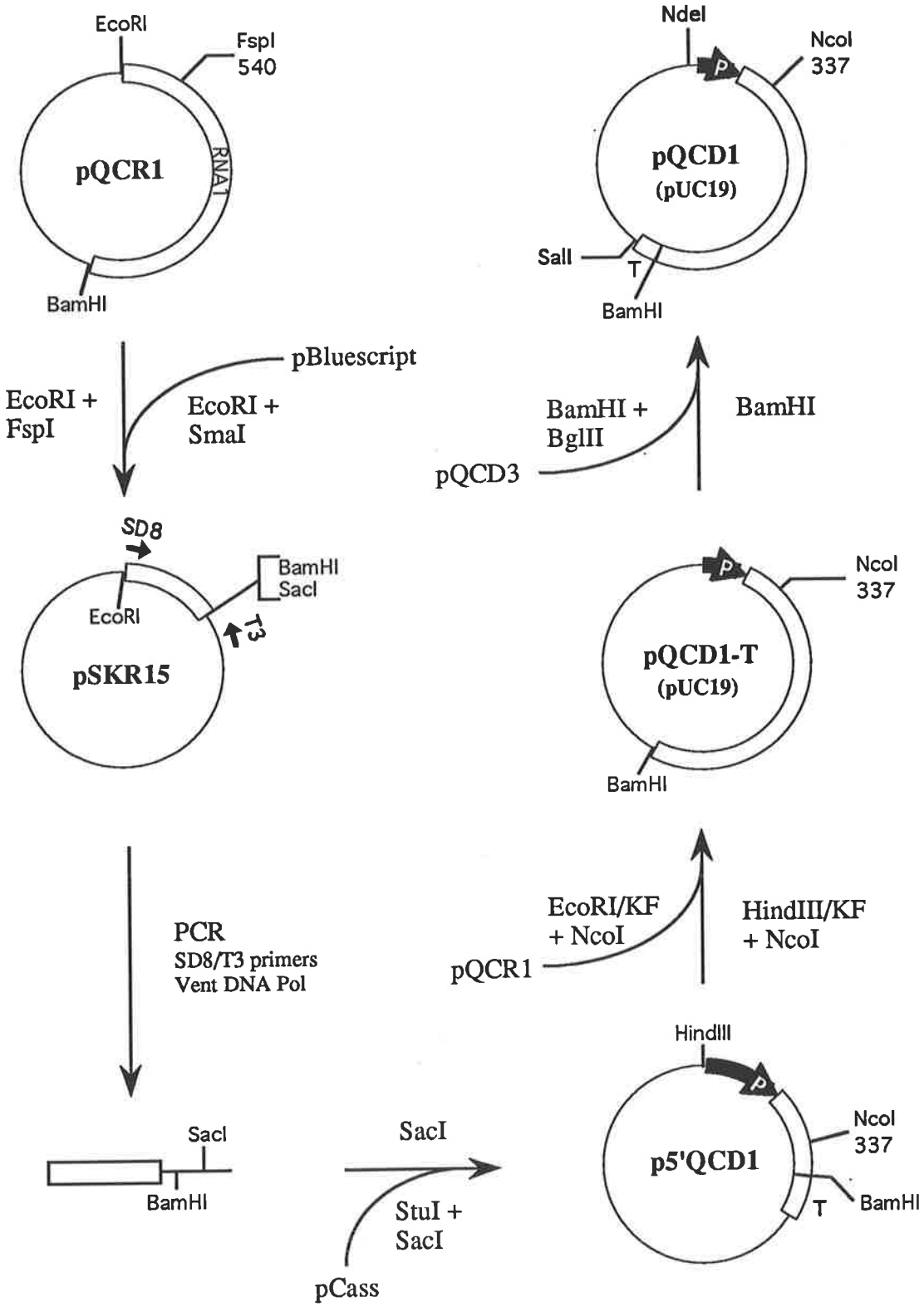

XbaI
SacI
492
XbaI
1003
RNA3
pQCR3
BamHI
pBluescript
XbaI
XbaI
PvuII
P
SacI
492
pQCD3
XbaI
1003
BglII
T
BamHI
SacI +
BamHI
pQCR3
SacI+BamHI
SD10
SacI
492
XbaI
pSKR35
XbaI
BamHI
EcoRI
T3
P
SacI
492
p5'QCD3
T
BamHI
PCR
Primers SD10/T3
Vent DNA polymerase
StuI+EcoRI
EcoRI
pCass
SacI
BamHI
EcoRI

then placed in the dark for 24 hours prior to inoculation. Plants were infected by rubbing equal amounts ($^{w}/_{w}$; 5 µg each plasmid in a final volume of 10 µl) of plasmids pQCD1, pQCD2 and pQCD3 onto the fourth leaf (for *N. glutinosa*) or cotyledons (for cucumber) dusted with carborundum powder. Infected tissue was harvested 10-14 days post inoculation (dpi) for northern analysis. Where restriction digestion was required to release the viral cassette (consisting of the full-length CMV cDNA flanked by the 35S promoter and terminator), restriction was performed with the following endonucleases; *Nde*I and *Sal*I for pQCD1; *Pvu*II and *Bgl*II for pQCD2 and pQCD3 (Figs 5.4, 5.5 and 5.6). Restriction digests were purified by phenol/chloroform extraction and ethanol precipitation prior to inoculation. Control inoculations were performed in the same manner as for plasmids except that the inoculum consisted of pure Q-CMV virions at a concentration of 100 µg/ml. All inoculations involving recombinant DNAs were conducted under containment (C1) conditions.

5.2.8 Northern analysis of CMV-infected plants

Total RNA was prepared from ~1 g of infected plant tissue using the small-scale method of Verwoerd *et al.* (1989). Two and a half µg of the RNA per sample was denatured by incubation at 65°C for 10 min in 1xMOPS running buffer (0.1 M MOPS pH 7.0, 40 mM sodium acetate, 5 mM EDTA), 6% ($^{v}/_{v}$) formaldehyde, 50% ($^{v}/_{v}$) formamide, then loaded onto a 1.2% agarose 1% formaldehye 1xMOPS buffer submarine gel in 1xMOPS running buffer. After electrophoresis RNA was transferred from the gel to a nylon filter (Hybond N+, Amersham) by capillary transfer, and the RNA fixed to the filter by UV crosslinking. The filter was transferred to a hybridisation bottle on fine mesh, then prehybridised for 4 hr at 65°C in 10 ml of 5xSSC (1xSSC is 150 mM NaCl, 15 mM sodium citrate pH 7.0), 5xDenhardts solution (50xDenhardts solution is 1% ($^{w}/_{v}$) Ficoll 400 (Pharmacia, USA), 1% ($^{w}/_{v}$) polyvinylpyrrolidone, 1% ($^{w}/_{v}$) bovine serum albumin), 50% ($^{v}/_{v}$) deionised formamide, 20 mM sodium phosphate buffer pH 6.8, 1% ($^{w}/_{v}$) SDS, 0.1 mg/ml denatured salmon sperm DNA, 0.1 mg/ml denatured *E. coli* tRNA. A ^{32}P labelled

Fig. 5.6. Cloning of a full-length cDNA corresponding to CMV-Q RNA 1 into pCass. Interpretation of the diagram is as for Fig. 5.1. Restriction sites for release of the 35S transcriptional cassette containing the viral cDNA in pQCD3 are shown.


EcoRI
FspI
540
pQCR1
RNA1
BamHI
NdeI
NcoI
337
P
pQCD1
(pUC19)
SalI
T
BamHI
pBluescript
EcoRI +
FspI
EcoRI +
SmaI
BamHI +
BglII
BamHI
pQCD3
SD8
BamHI
SacI
EcoRI
T3
pSKR15
NcoI
337
P
pQCD1-T
(pUC19)
BamHI
PCR
SD8/T3 primers
Vent DNA Pol
EcoRI/KF
+ NcoI
HindIII/KF
+ NcoI
pQCR1
HindIII
SacI
BamHI
SacI
StuI +
SacI
pCass
P
NcoI
337
p5'QCD1
T
BamHI

riboprobe was prepared by *in vitro* transcription (2.2.12) from clone pCMVall+ (Dr S. -W. Ding, unpublished data), corresponding to the 3'-terminal ~300 nt of Q-CMV RNA 2, and complementary to all CMV genomic and subgenomic RNAs (Symons, 1979). Approximately 10^5 cpm of the probe was denatured by incubation at 85°C for 2 min, then added directly to the prehybridisation solution and incubated overnight at 65°C. After incubation the probe was discarded, and the filter washed twice in 2xSSC, 0.1% ($^w/_v$) SDS at 50°C for 15 min, then once in 0.2xSSC, 0.1% ($^w/_v$) SDS at 75°C for 45 min. The filter was blotted dry and radioactive species detected by autoradiography at room temperature.

5.3 Results

5.3.1 Synthesis of a cloning vector for plant RNA viruses

The aim of this work was to develop a suitable vector for the transcription of cloned viral cDNAs within the plant cell. The promoter for the 35S transcript of CaMV was chosen for this purpose, firstly because it expresses strongly in different plant species and tissues, and also because it does not require intronic sequences for maximal promoter activity (McElroy and Brettell, 1994). Transcription of viral cDNAs containing the native 5' end sequence from the CaMV 35S promoter is therefore possible by insertion of the sequence of interest at the +1 nucleotide of transcription. This thesis follows the strategy of Yamaya *et al.* (1988) who modified the CaMV 35S transcriptional start site from 5'-AGGACA-3' (underlined nucleotide indicates +1 nucleotide) to 5'-AGGCCT-3' (changed nucleotides are underlined). The mutated sequence contains a recognition sequence for the restriction endonuclease *Stu*I, which cleaves centrally leaving blunt ends. This allows insertion of the viral cDNA at the +1 transcription site.

Modification of the 35S promoter was achieved by PCR mutagenesis (Fig. 5.1). The modified promoter was cloned into the plasmid pSP72, which was chosen because it lacks α-complementation, so there is no bacterial promoter driving expression across the polylinker. This is important because viral sequences can be toxic to the bacterial

cell (*e.g.* MacFarlane *et al.*, 1991), an effect which is presumably mediated by transcription. A small fragment of the 35S transcriptional terminator competent for termination (Dr I.B. Dry, personal communication) was obtained by digestion of pJIT60 with *Eco*RI and *Bgl*II. Cloning of the truncated 35S transcriptional terminator downstream of the modified promoter created a small polylinker of four endonuclease recognition sites that facilitates the cloning of viral sequences (Figs 5.1 and 5.2). The structure of the vector has allowed design of a standard procedure for cloning viral cDNAs into the vector (5.3.2).

5.3.2 Cloning CMV genomic cDNAs into pCass

Full-length cDNAs of Q-CMV genomic RNAs were cloned into pCass using a generalised strategy. This involved subcloning of the 5' terminal portions of Q-CMV cDNAs 1, 2 and 3 into pBluescript SK+ prior to PCR amplification of the viral sequences (Figs 5.4, 5.5 and 5.6). The genomic 5' end of the viral cDNA was precisely determined by the sequence of the appropriate PCR primer. Cloning of the viral 5' cDNA into pBluescript prior to PCR amplification allowed exploitation of the restriction sites in the pBluescript polylinker that were co-amplified with the viral sequence. The short polylinker of three restriction sites downstream of the *Stu*I recognition sequence in pCass allowed directional cloning of the viral sequence, placing the 5' genomic nucleotide at the starting point of the transcript. While this strategy was successful in cloning cDNAs of Q-CMV RNAs 2 and 3, a minor variation was required for RNA 1. This was because repeated attempts to assemble full-length cDNAs of RNA 1 cloned into pCass on the pSP72 backbone failed, implying that the full-length construct is toxic to *E. coli*. Therefore, as the RNA 1 cDNA existed prior to this work as a full-length (non-infectious) clone in pUC19, the infectious construct consisting of the modified promoter, full length RNA 1 cDNA and transcriptional terminator was also assembled in pUC19. This construct proved to be stable when transformed into *E. coli* strain DH5α.

5.3.3 Infection of *N. glutinosa* with the pQCD clones

To test the infectivity of the CMV genomic cDNAs cloned into pCass, purified circular plasmid DNAs pQCD1, 2 and 3 were mixed in equal amounts and inoculated onto plants which had been kept in the dark for 24 hours. The mixture of pQCD1, pQCD2, and pQCD3 systemically infected *N. glutinosa* at a variable frequency of 20-40% (Table 5.1). The cloned cDNAs produced dark and light green mosaic symptoms typical of CMV infection in *N. glutinosa*. Symptoms generated by cDNA infections in cucumber were indistinguishable from those produced by purified Q-CMV virions (Fig. 5.7). Total RNAs were purified from healthy and systemically infected leaves and subjected to northern blot analysis using a ^{32}P-labelled RNA probe with sequence complementary to the 3' terminal 300 nt of RNA 2. This RNA 2-derived probe is capable of detecting all Q-CMV RNAs because the 3' terminal sequences (about 300 nt) of these RNAs are highly homologous (Symons, 1979). Viral genomic RNAs (1, 2 and 3) and subgenomic RNAs (4 and 4A) were detected both in cDNA-inoculated and virion-inoculated *N. glutinosa* plants (Fig. 5.8). Western blot analysis also demonstrated the accumulation of Q-CMV coat protein in the cDNA-inoculated plants (data not shown).

Restriction enzyme digestion of the cDNA clones prior to inoculation was investigated in an attempt to improve the frequency of infection obtained with the cloned CMV cDNAs. Linearisation of the plasmids did not significantly improve the efficiency of infection relative to that achieved with closed circular DNA (Table 5.1). However, if the viral expression cassettes (consisting of full-length viral cDNA flanked by the 35S promoter and terminator) were excised from the vector by restriction enzyme digestion, marked increases in infectivity were obtained. In three independent experiments comprising inoculation of five *N. glutinosa* plants, 14/15 plants became infected after double digestion of the plasmid DNAs (Table 5.1). These data demonstrate the infectivity of Q-CMV genomic cDNAs cloned into pCass.

Fig. 5.7. Cucumber plants infected with CMV. Aerial view of young cucumber plants 14 days post inoculation. The plant on the left was inoculated with wildtype Q-CMV virions, while that on the right was inoculated with the three pQCD cDNA clones. The plant in the middle is a healthy control. Apparent spots on the leaves of the healthy control are water marks.

Fig. 5.8. Northern blot analysis of fractionated total RNAs extracted from *N. glutinosa* plants. Q-CMV RNAs detected by northern analysis from plants inoculated with Q-CMV virions (lane 3) or by Q-CMV genomic cDNA clones pQCD1, 2 and 3 (lane 4). RNA isolated from healthy *N. glutinosa* plants was loaded in lane 1, while pure Q-CMV RNAs isolated from virions were loaded in lane 2. CMV genomic RNAs (1, 2, and 3) and subgenomic RNAs (4, 4A, and 5) are indicated on the left margin. Sizes of genomic RNAs are given in Fig. 5.3.

Figure 5.7

Figure 5.8

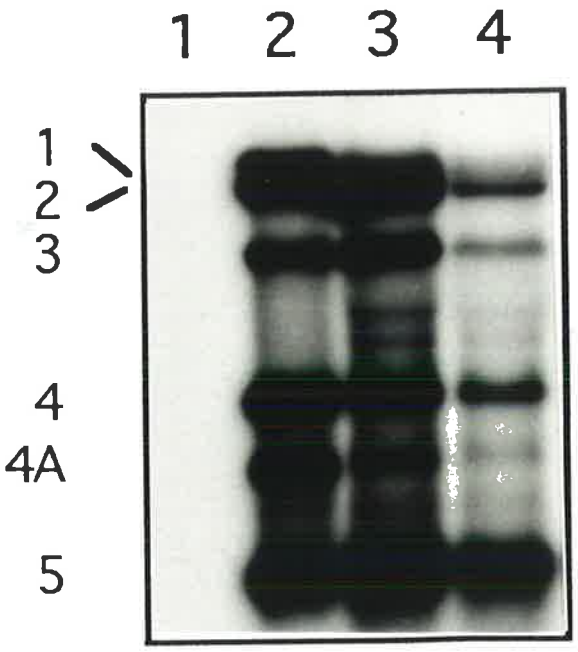


Table 5.1. Infectivity of cloned CMV-Q cDNAs in pCass

Inoculation	Circular	Linearised	Released insert
1	1/5[a]	2/5	4/5
2	1/5	2/5	5/5
3	2/5	-	5/5
4	2/5	-	-
Total	6/20	4/10	14/15

[a]Number of plants that became infected with CMV-Q over the total number of plants inoculated.

5.4 Discussion

This Chapter describes the synthesis of a universal cloning vector for the expression of plant RNA virus cDNAs *in planta*. The well characterised promoter of the 35S RNA of CaMV was chosen to drive transcription in the vector. Advantages of this promoter include its activity in a wide variety of plant species including both dicots and monocots. In addition, the vector should also be able to direct transcription in plant protoplasts. The ultimate usage of the vector within this thesis is to express BYDV cDNAs in cereal tissues, in particular the vegetative meristem of oat and barley plants. Therefore, a promoter with strong expression lacking tissue specific expression was required; to date the CaMV 35S promoter best fits these criteria (McElroy and Brettell, 1994). Other promoters with higher levels of expression in monocot plant tissues are available, but these often require intronic sequences for optimal activity, or are hybrid promoters with uncharacterised transcriptional initiation sites. The second major advantage of the CaMV 35S promoter is that the position from which the first nucleotide is transcribed is known. Such information is necessary because the addition

of non-viral nucleotides to the 5' end of synthetic RNA genomic transcripts can seriously diminish or abolish infectivity. Therefore, knowledge of the transcriptional start site allows cloning of the viral sequence such that the first transcribed nucleotide corresponds to the first nucleotide of the viral sequence.

The majority of cloned infectious plant virus cDNAs require *in vitro* transcription before the infectious RNA transcripts are used to infect whole plants or plant protoplasts (Boyer and Haenni, 1994). There are a number of problems associated with this approach that are circumvented by the use of cDNAs which are truly infectious and do not require *in vitro* transcription prior to inoculation. The first point derives from the promoter used to drive transcription of the viral cDNA *in vitro*, which is most commonly the bacteriophage T7 promoter. The T7 RNA polymerase has certain sequence requirements at the 3' end of the promoter that are necessary for optimal expression. Fusion of the plant virus genomic sequence to the transcriptional start site may disrupt the sequence and significantly reduce the yields of cRNA (Boyer and Haenni, 1994). Further disadvantages of an infection system based on *in vitro* transcription include the potential for degradation of RNA transcripts, especially in the infection process where it is impossible to exclude the presence of RNases. Other drawbacks include the cost of the 7-methyl guanosine cap analogue, and the general intricacy of the procedure.

The vector described here was based on the plasmid pSP72. This is a pUC-based plasmid designed for *in vitro* transcription, and hence contains promoter sites for T7 and SP6 bacteriophage RNA polymerases. pSP72 was chosen because it contains no endogenous bacterial promoter at the polylinker, and thus viral cDNAs cloned into the vector are unlikely to be toxic to the cell. The T7 and SP6 promoters are recognised in a highly specific manner by the cognate bacteriophage DNA-directed RNA polymerase, and are unlikely to be transcribed in the absence of these enzymes. Despite this reasoning, repeated attempts to assemble full-length cDNA corresponding to Q-CMV RNA 1 in pCass failed, suggesting that the construct is toxic to the cell. The complete construct comprising the modified 35S promoter/viral cDNA fusion and the

35S terminator was subsequently assembled in a pUC19 background. Paradoxically, pUC19 does possess a bacterial promoter driving expression across the polylinker, although this promoter should be repressed in cells not induced by galactose or its analogues. The toxic effect of RNA 1 cDNA exhibited in pCass could possibly be due to the effect of cryptic or unrecognised promoters.

A strategy for cloning viral cDNAs into pCass applicable to any positive-strand plant RNA virus genome was evolved. This involves cleavage of the vector at the transcription initiation site with *Stu*I, followed by precise PCR mediated blunt-end fusion with the 5' genomic terminus. Cloning of the remaining genomic sequences is facilitated by restriction sites in the viral sequence and those derived from the PCR or present in pCass. This strategy was demonstrated here in the cloning of cDNAs corresponding to the three genomic RNAs of CMV strain Q. The three clones were infectious when inoculated together onto *N. glutinosa* or cucumber seedlings. This was demonstrated here by the development of typical CMV mosaic symptoms in infected cucumber plants. Symptoms produced by infection with the cDNA clones were identical to those produced by infection with native Q-CMV virus particles. In addition, the accumulation of viral genomic and subgenomic RNAs in plants inoculated with the cDNA clones was demonstrated. Infectious virus particles also could be recovered from plants infected by the three pQCD clones (data not shown).

The infectivity of the pQCD clones was markedly increased by excision of the transcription cassette prior to inoculation. Low infectivity of closed circular viral cDNA clones driven by the CaMV 35S promoter has also been reported for brome mosaic bromovirus (Mori *et al.*, 1991) and tomato mosaic tobamovirus (Weber *et al.*, 1992). The increase in infectivity obtained by digestion of the infectious clones to release the transcriptional cassette described here is similar to the results of Neeleman *et al.* (1993) who found that cDNA clones of alfalfa mosaic virus were only infectious after release of the viral inserts flanked by the transcriptional controls. While it is not clear why such restriction should increase infectivity, it is possible that the excised DNA is transported more efficiently to the nucleus of the plant cell, resulting in greater

transcription of the viral sequence. Other scenarios such as differential susceptibility to cellular nucleases and efficiencies of transcription (due possibly to supercoiling effects) are also possible. This result is relevant to the proposed use of the vector as an intermediate in the construction of potentially agroinfectious clones in the T-DNA of a binary vector (5.1). Excision of the T-DNA from the binary vector would mimic the *in vitro* excision of the transcriptional cassette described here, although the resulting linear DNA species would be much larger than the CMV constructs. Additionally, the T-DNA would be expected to be preferentially targeted to the nucleus as this is part of the *Agrobacterium* strategy for plant transformation.

CHAPTER SIX

FULL-LENGTH cDNA CLONES OF BYDV-RPV AND BYDV-PAV GENOMIC RNAS FOR AGROINFECTION

6.1 Introduction

The previous Chapter showed how the expression vector pCass can be used to create infectious cDNA clones of a systemically-infecting, mechanically transmissible plant RNA virus. Application of this system to the luteoviruses is more difficult because members of the group are not mechanically transmissible, so direct inoculation of leaves with viral cDNAs cloned into a transcription vector (as for CMV) is unlikely to cause infection (*e.g.* Leiser *et al.*, 1992). Therefore, adaptation of the agroinfection technique for the mechanical inoculation of cloned luteovirus sequences was proposed (Chapter Five). Agroinfection has largely been used to introduce cloned geminivirus DNAs to monocotyledonous plants (Grimsley, 1990). Geminiviruses are mechanically non-transmissible and possess circular single-stranded DNA genomes that are either mono- or bipartite. The viral DNA is cloned as a greater-than-unit-length fragment in the T-DNA of a suitable *Agrobacterium* binary vector, which is transferred to certain strains of *A. tumefaciens* or *A. rhizogenes* (Marks *et al.*, 1989). Thick suspensions of the *Agrobacterium* containing the plasmid are used to inoculate the vegetative meristem of the plant. Infection appears to occur as a consequence of recombination between repeated sequences in the cloned viral DNA, releasing the infectious circular form of the viral DNA (Grimsley *et al.*, 1986). Thus agroinfection provides a means of circumventing the mechanical non-transmissibility of the geminiviruses, and potentially also of the luteoviruses. However, luteovirus sequences must be converted to RNA before they are infectious; thus it is necessary to place the luteovirus cDNAs under the control of transcriptional signals prior to cloning into a binary vector. The aim of this work was therefore to use the functional elements of the expression vector pCass to construct transcriptionally competent full-length clones of BYDV-PAV and -RPV suitable for agroinfection. Successful agroinfection of BWYV using a strategy similar to that described here was reported subsequent to the commencement of this work (Leiser *et al.*, 1992).

While a full-length clone of BYDV-PAV exists that can be conveniently adapted for agroinfection (Young *et al.*, 1991), no such clone exists for BYDV-RPV.

Prior to the initiation of this work no sequence data was available for BYDV-RPV, although partial sequence of the closely related Chinese isolate BYDV-GPV (Cheng *et al.*, 1994b) was available. Therefore the first step towards constructing a full-length clones was to define the genomic termini of an Australian isolate of BYDV-RPV (BYDV-RPV-Vic; Waterhouse *et al.*, 1986). The near-complete sequence of a New York isolate of BYDV-RPV (BYDV-RPV-NY; Vincent *et al.*, 1991) was published shortly after the sequencing of the BYDV-RPV-Vic genomic termini, facilitating construction of the full-length clone. This Chapter describes construction of a full-length genomic cDNA of BYDV-RPV-Vic using a PCR based approach, and the cloning of full-length BYDV-PAV and RPV genomic cDNAs under the control of the CaMV 35S promoter into the binary vector pBIN19. Preliminary attempts to use these clones in agroinfection are described.

6.2 Materials and Methods

6.2.1 Generation of plant material infected with BYDV-RPV-Vic

An Australian isolate of BYDV-RPV, termed BYDV-RPV-Vic (Waterhouse *et al.*, 1986), was obtained from Dr Peter Waterhouse, CSIRO Division of Plant Industry, Canberra, Australia. Virus was maintained by serial transmission in oat (*Avena sativa* cv. New Zealand Cape) using viruliferous *Rhopalosiphum padi* aphids. For generation of large amounts of infected plant material, non-viruliferous *R. padi* aphids were placed on BYDV-RPV-Vic infected oat tissue for an acquisition period of 48 h. Viruliferous aphids were transferred to week-old oat (cv. New Zealand Cape) seedlings densely planted in 12" pots and grown in insect-proof cages under natural light conditions. After a virus transmission period of 48 h the aphids were killed by application of pyrethrum insecticide (AgChem, Australia). The infected seedlings were grown for a further three weeks after which leaf tissue was harvested, then used immediately for dsRNA extraction or stored until needed at -80°C.

6.2.2 Purification of BYDV-RPV double-stranded RNA (dsRNA)

For purification of BYDV-RPV-Vic dsRNA, 30 g of infected oat tissue (6.2.1) was frozen in liquid N_2 and ground to a fine powder in a mortar and pestle. The grounds were transferred to a sterile 100 ml beaker and mixed with 40 ml 2xSTE (1xSTE is 100 mM NaCl, 50 mM Tris-HCl pH 7.0, 1 mM EDTA) and 6 ml 10% SDS. Fifty ml of Tris-HCl saturated phenol (pH ~8.0) was added and the mixture stirred at room temperature (RT) for 30 min. The extract was transferred to centrifuge pots and the phases separated by centrifugation at 10,000 rpm for 15 min at 4°C in a Sorvall GSA rotor. The aqueous layer was removed to a new sterile beaker and an equal volume of chloroform:isoamyl alcohol (24:1) added. The mixture was stirred for a further 30 min at RT, and the centrifugation step repeated. The aqueous phase was removed and adjusted to 16% ($^{v}/_{v}$) ethanol before the addition of 5 g Whatman CF-11 cellulose. The dsRNA was allowed to bind to the CF-11 over 2 h with stirring at RT. The cellulose was repeatedly washed by pelleting by centrifugation at 10,000 rpm for 10 min at 4°C in a GSA rotor followed by resuspension in ~50 ml 2xSTE 16% ($^{v}/_{v}$) ethanol. Washing was continued until the cellulose appeared white and fluffy, after which it was suspended in ~50 ml 2xSTE 16% ethanol. The cellulose was transferred to a sterile RNase-free glass column (2x20 cm) and allowed to settle, after which it was washed under gravity flow with 350 ml of 1xSTE 16% ($^{v}/_{v}$) ethanol at RT over ~4 h to remove single-stranded RNAs (ssRNA). dsRNA was eluted from the column with 36 ml of 1xSTE and precipitated from solution by the addition of $^{1}/_{10}^{th}$ volume 3 M sodium acetate pH 5.2 and 2.5 volumes of ethanol. The precipitated nucleic acids were pelleted by centrifugation in Corex tubes at 10,000 rpm at 4°C for 30 min in a Sorvall HB4 rotor and the supernatant discarded. Nucleic acid pellets were washed with 70% ethanol at -20°C and centrifuged as before, after which the supernatant was discarded. The pelleted nucleic acids were dried *in vacuo* and resuspended in 400 µl TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA), and stored at -20°C.

6.2.3 PCR verification of BYDV-RPV dsRNA

One µl of dsRNA (6.2.2) was denatured by addition of methyl mercuric hydroxide (MeHgOH) to a concentration of 6 mM in a final volume of 20 µl. After incubation for 15 min at RT the MeHgOH was deactivated by the addition of 2 µl 350 mM β-mercaptoethanol ($C_f \approx 30$ mM) and incubation at RT for 5 min. First-strand cDNA synthesis was carried out using 2 µl of the denatured RNA and was primed either by oligonucleotide RPV1.oli (5'-CTTCGCCGACATCTACACCTGGG-3'; homologous to nucleotides 2025-2047 of the BYDV-RPV-NY genome) to prime cDNA synthesis from minus-strand RNA, or RPV2.oli (5'-GGAGCTTCAAGGGCATCATCGCCC-3'; complementary to residues 3186-3209 of the BYDV-RPV-NY genome) to prime cDNA synthesis from plus-strand RNA. Reverse transcription was performed using AMV reverse transcriptase (Promega, USA) according to the manufacturer's specifications in a final volume of 10 µl. One µl of the cDNA was amplified in a PCR reaction employing RPV1.oli and RPV2.oli as the PCR primers and was catalysed by Vent DNA polymerase (New England Biolabs) essentially as recommended by the supplier. The profile for thermocycling was $[94^oC/1\ min; 60^oC/1\ min; 72^oC/1\ min]_{30}$; 72°C/5 min; 25°C/5 min and was carried out on a DNA Thermal Sequencer (Corbett, Australia). Reaction products were electrophoresed on a 1.5% agarose/1xTBE minigel and visualised by ethidium bromide/UV.

6.2.4 RACE-PCR of the BYDV-RPV-Vic 5' and 3' genomic termini

One µl of BYDV-RPV-Vic dsRNA (6.2.2) was denatured as previously (6.2.3), and 2 µl of the denatured RNA reverse transcribed using the oligonucleotide RPVRACE5 (5'-ATGAATTCTGTAGATCCAACTCGTTATA-3'; complementary to nucleotides 766-788 of the BYDV-RPV-NY genome) for 5' end determination, or RPVRACE3 (5'-ATGAATTCGAAAACTTCGGTATACAAAC-3'; homologous to nucleotides 5094-5114 of the BYDV-RPV-NY genome) for 3' end determination. The reaction was carried out using AMV reverse transcriptase (Promega) according to the

manufacturer's specifications in a volume of 20 µl, and was stopped by incubation of the mixture at 80°C for 3 min. The volume of each cDNA solution was increase to 200 µl with TE, then RNase A added to a final concentration of 20 µg/ml. The mixture was incubated at 37°C for 1 h to degrade the RNA, then extracted once with phenol:chloroform. The supernatant was recovered and precipitated by the addition of ammonium acetate to 2.5 M and one volume of ice-cold 2-propanol. Precipitated nucleic acids were recovered by centrifugation in a benchtop microfuge at full speed (14,000 rpm), then washed in ice-cold 70% ethanol. The nucleic acid pellet was resuspended in 200 µl TE and the precipitation repeated. The nucleic acid pellet was dried *in vacuo* after the final 70% ethanol wash, then resuspended in 10 µl TE. The purified cDNAs were poly(dA) tailed from the 3'OH group with dATP and terminal deoxynucleotidyl transferase (Promega) according to the manufacturer's specifications. The tailed cDNAs were incubated at 80°C for 5 min to denature the enzyme, then diluted to 200 µl with TE and stored at -20°C. RACE-PCR was performed on 1 µl of the tailed cDNA using RPVRACE5 or RPVRACE3 as the sequence-specific primer, and ARACE5 (5'-GACTCGAGATCGA$[T]_{17}$-3') as the non-specific primer. The PCR reaction was catalysed by Vent DNA polymerase according to the manufacturer's directions, with a thermal cycle profile of 94°C/3 min; $[55°C/1$ min; 72°C/1 min 30 s; $94°C/40$ s$]_{40}$; 72°C/5 min; 25°C/5 min on a DNA Thermal Sequencer (Corbett). Reaction products were resolved on 1.5% agarose 1xTBE minigels and visualised with ethidium bromide/UV. The major reaction product of ~800 nt (5' end) or ~680 nt (3' end) was excised from the gel and purified using Geneclean before blunt-end cloning into the *Sma*I site of pBluescript SK+ (Stratagene). Clones were identified by dideoxy sequencing on double-stranded templates.

6.2.5 PCR amplification and cloning of cDNA segments covering the BYDV-RPV-Vic genome

Segment 1. Amplification of a BYDV-RPV-Vic genomic cDNA corresponding to the 5' ~2 kb was carried out using cDNA from the 5' RACE reaction (6.2.4) as

template. PCR was performed using Vent DNA polymerase and primers RPV 2082 (5'-AAAGCCTGGGATCTCTTGTT-3'; complementary to residues 2063-2082 of the BYDV-RPV-NY genome) and RPV5TERM (5'-ACAACGAAAGAAGCTTAGGA-3'; homologous to residues 1-20 of the BYDV-RPV-Vic genome) each at 4.25 μM. The thermal cycle for amplification was [94°C/15 s; 51°C/5 s; 72°C/2 min]$_{45}$; 72°C/5 min; 25°C/5 min and was carried out on a capillary DNA Thermal Sequencer (Corbett). Reaction products were resolved on 1.0 % agarose 1xTBE minigels and visualised with ethidium bromide/UV. The major product of ~2 kb was excised from the gel and purified using Geneclean. Purified DNA was cloned into the *Sma*I site of pBluescript SK+ to create the clone pRPVseg1, and the cDNA termini verified by dideoxy sequencing from dsDNA templates.

Segment 2. A BYDV-RPV-Vic internal genomic fragment of ~2 kb (corresponding to nucleotides 2025-4003 of the BYDV-RPV-NY genome) was amplified after synthesis of cDNA from minus-strand RNA primed by oligonucleotide RPV1.oli (6.2.3). Denaturation of dsRNA and conditions for cDNA synthesis were as described previously (6.2.3). PCR was performed using oligonucleotides RPV1.oli and RPV-1 (5'-TCATGGTAGGCCTTGAGTATTCCAT-3'; complementary to nucleotides 3979-4003 of the BYDV-RPV-NY genome), both at a concentration of 0.3 μM. The reaction was catalysed by Vent DNA polymerase, with a thermocycle profile of [94°C/10 s; 55°C/5 s; 72°C/70 s]$_{40}$; 72°C/5 min; 25°C/5 min, and was carried out on a capillary DNA Thermal Sequencer (Corbett). Gel electrophoresis and visualisation of the major product of ~2 kb was as described above. The purified DNA was cloned into the *Sma*I site of pBluescript SK+, and was designated pRPVseg2 after verification of the terminal sequences by dideoxy sequencing from dsDNA templates.

Segment 3. Amplification of a second internal genomic fragment of BYDV-RPV-Vic (corresponding to nucleotides 3876-5192 of the BYDV-RPV-NY genome) was carried out using cDNA generated for 3' genomic RACE (6.2.4). PCR was performed using oligonucleotides RPV 3876 (5'-AGCCGTGGCGAGACATTCGT-3'; homologous to nucleotides 3876-3895 of the BYDV-RPV-NY genome) and RPV 5192

(5'-GATCGTCTTCTGACTCCGAAT-3'; complementary to nucleotides 5173-5192 of the BYDV-RPV-NY genome) each at a concentration of 3.8 μM. The PCR reaction employed Vent DNA polymerase with a thermal cycle of [94°C/10 s; 55°C/5 s; 72°C/80 s]$_{40}$; 72°C/5 min; 25°C/5 min, and was carried out on a capillary DNA Thermal Sequencer (Corbett). PCR products were resolved on 1.5% agarose 1xTBE minigels and visualised with ethidium bromide/UV. The major product of ~1.3 kb was excised from the gel and purified with Geneclean, before cloning into the *Sma*I site of pBluescript to create pRPVseg3. The terminal sequences of the cDNA were verified by dideoxy sequencing of the double-stranded clone.

Segment 4. Amplification of a BYDV-RPV-Vic genomic cDNA corresponding to the 3' ~700 nt was carried out using cDNA generated for 3' RACE (6.2.4) as template. PCR was carried out using oligonucleotides RPVRACE3 (6.2.4) and RPV3TERM (5'-ACAAAAGCTTCTTAGAGATC-3'; complementary to the 3'-terminal 20 nt of BYDV-RPV-Vic), each at a concentration of 2.5 μM. The reaction was catalysed by Vent DNA polymerase using a thermocycle profile of [94°C/40 s; 49°C/1 min; 72°C/90 s]$_{40}$; 72°C/5 min; 25°C/5 min, and was carried out on a capillary DNA Thermal Sequencer (Corbett). PCR products were resolved on 1.5% agarose 1xTBE minigels and visualised with ethidium bromide/UV. The major DNA product of ~700 nt was excised from the gel, purified using Geneclean, and cloned into the *Sma*I site of pBluescript. The clone was designated pRPVseg4 after verification of the terminal nucleotide sequences by dideoxy sequencing from dsDNA templates.

6.2.6 Restriction analysis of PCR segments 1-4

cDNAs corresponding to segments 1-4 were amplified as described (6.2.5) and purified by gel electrophoresis followed by Geneclean of the excised fragments. Purified DNAs were subjected to digestion in separate reactions with two or more of the following restriction endonucleases, using buffers recommended by the manufacturer (Boehringer Mannheim); *Bam*HI, *Bcl*I, *Bgl*II, *Cla*I, *Eco*RI, *Pst*I *Sal*I, *Sph*I, *Xba*I, *Xho*I. Restricted DNA solutions were brought to 100 mM NaCl where

necessary before endfilling with Klenow fragment and the appropriate ^{32}P-labelled deoxynucleotide. Radioactive fragments were electrophoresed on 10% polyacrylamide 7 M urea 1xTBE gels (40x20 cm) under denaturing conditions to detect small fragments (≥20 nt), and on 1.5% agarose 1xTBE minigels (non-denaturing) to detect larger fragments (≥300 nt). The gels were dried onto Whatman 3MM paper, and DNA fragments visualised by autoradiography.

6.2.7 Overlapping cDNA segments of the BYDV-RPV-Vic genome mutated to contain restriction sites in the overlaps

Segment 1M. Synthesis of a DNA fragment approximately corresponding to segment 1 was performed as previously (6.2.5) apart from the following details: The complementary primer was RPV 2289M (5'-GGAACG*CTCGAG*CCAGC-3'; complementary to residues 2255-2270 of the BYDV-RPV-NY genome but modified at the italicised nucleotide to contain a recognition site for *Xho*I) rather than RPV 2082. The cDNA corresponding to segment 1M cloned into the *Sma*I site of pBluescript was designated pRPVseg1M, and verified as before by partial nucleotide sequencing.

Segment 2M. A DNA fragment approximately corresponding to segment 2 was synthesised as previously (6.2.5) apart from the following details: cDNA synthesis was primed with oligonucleotide RPV 4194M (5'-GGTGCCAC*TCTAGA*CCGTTG-3'; complementary to residues 4175-4194 of the BYDV-RPV-NY genome except for modification at the italicised nucleotides to contain a recognition site for *Xba*I). The oligonucleotides for amplification were RPV 4194M and RPV 2253M (5'-TCGCTGG*CTCGAG*CGTTC-3'; homologous to residues 2253-2270 of the BYDV-RPV-NY genome except for modification at the italicised nucleotide to create a recognition site for *Xho*I). The cDNA corresponding to segment 2M cloned into the *Sma*I site of pBluescript was designated pRPVseg2M, and verified as before by partial nucleotide sequencing.

Segment 3M. A DNA fragment approximately corresponding to segment 3 was synthesised as previously (6.2.5) apart from the following details: In the PCR, the

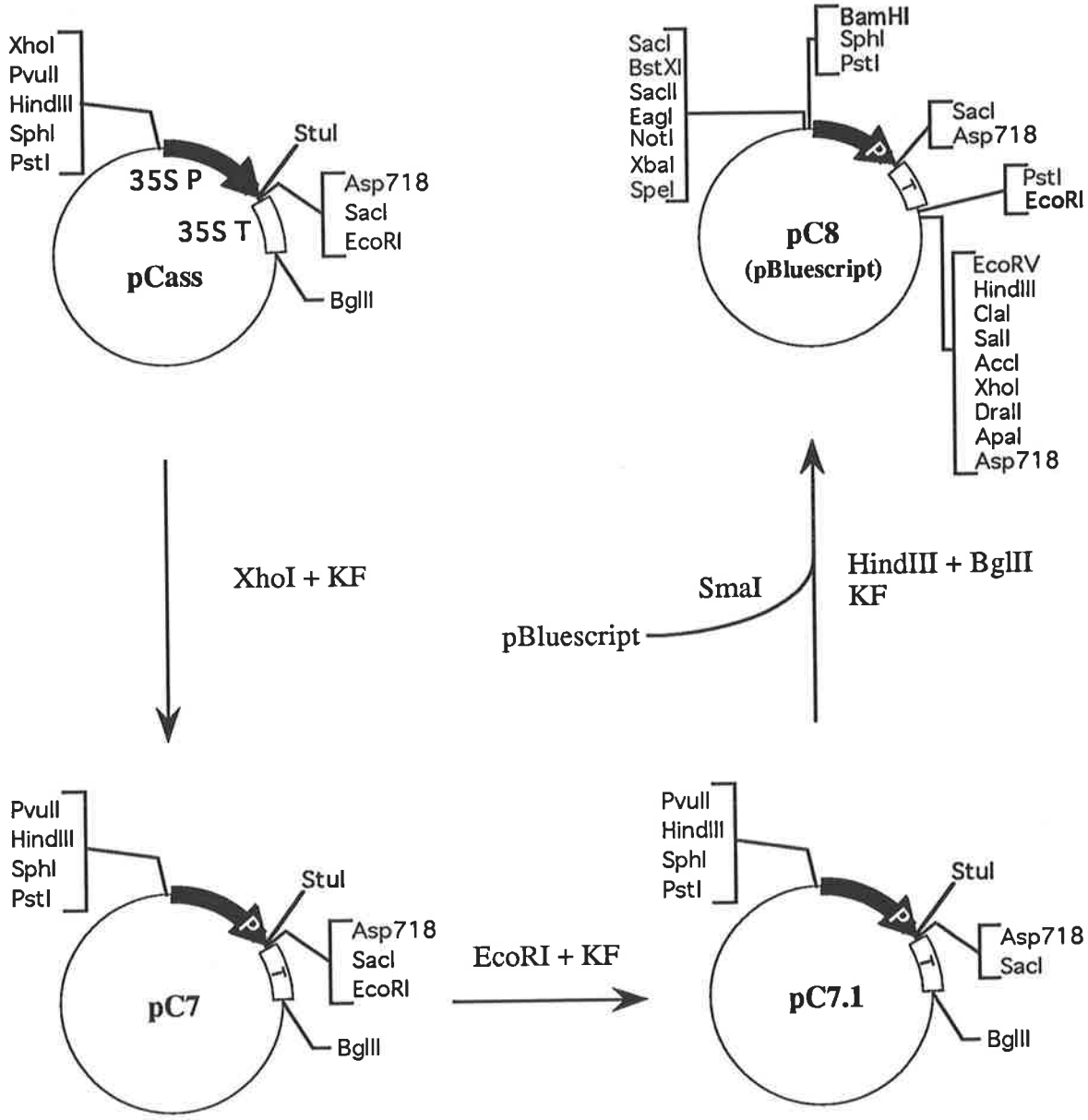
complementary primer was oligonucleotide RPV 5024M (5'-AGTGGGGGATCCCGAACTA-3'; complementary to residues 5006-5024 of the BYDV-RPV-NY genome and containing a modification at the italicised nucleotide to insert a *Bam*HI recognition site into the sequence) and the homologous primer was RPV 4168M (5'-AGATGATCAACGGTCTAGAGT-3'; homologous to residues 4168-4188 of the BYDV-RPV-NY genome and containing modifications at the italicised nucleotides to insert a recognition site for *Xba*I). The cDNA corresponding to segment 3M cloned into the *Sma*I site of pBluescript was designated pRPVseg3M, and verified as before by partial nucleotide sequencing.

Segment 4M. A DNA fragment approximately corresponding to segment 4 was synthesised as previously (6.2.5) with the following difference: The homologous primer RPVRACE3 was substituted for RPV 5001M (5'-AAGCGTAGTTCGGGATCC-3'; homologous to residues 5001-5018 of the BYDV-RPV-NY genome and mutated at the italicised residue to contain a recognition site for *Bam*HI). The cDNA corresponding to segment 4M cloned into the *Sma*I site of pBluescript was designated pRPVseg4M, and verified as before by partial nucleotide sequencing.

6.2.8 Assembly of a full-length cDNA of BYDV-RPV-Vic in pCass

Modification of pCass. Modification of pCass started from a pCass derivative (pC7; 6.2.11) in which the *Xho*I site at the 5' end of the polylinker had been deleted (Fig. 6.1). pC7 was digested with *Eco*RI, then endfilled with Klenow fragment and dNTPs to destroy the *Eco*RI site, before religation to create pC7.1. The 'cassette' fragment containing the CaMV 35S promoter and terminator flanking three restriction sites (*Stu*I, *Asp*718, *Sac*I) was excised from pC7.1 by restriction digestion with *Hind*III and *Bgl*II. The fragment was purified from the vector by gel electrophoresis/Geneclean, then treated with Klenow fragment and dNTPs to create blunt ends. The blunted fragment was cloned into the *Sma*I site of pBluescript to create pC8. pC8 contains the transcriptional elements of pCass cloned between the *Bam*HI

Fig. 6.1. Derivatives of pCass used in the construction of BYDV cDNA clones for agroinfection. Large dark arrow denotes CaMV 35S promoter (P), 35S terminator sequence is indicated (T). KF indicates endfilling of restriction fragments with dNTPs and the Klenow fragment of *E. coli* DNA polymerase I.

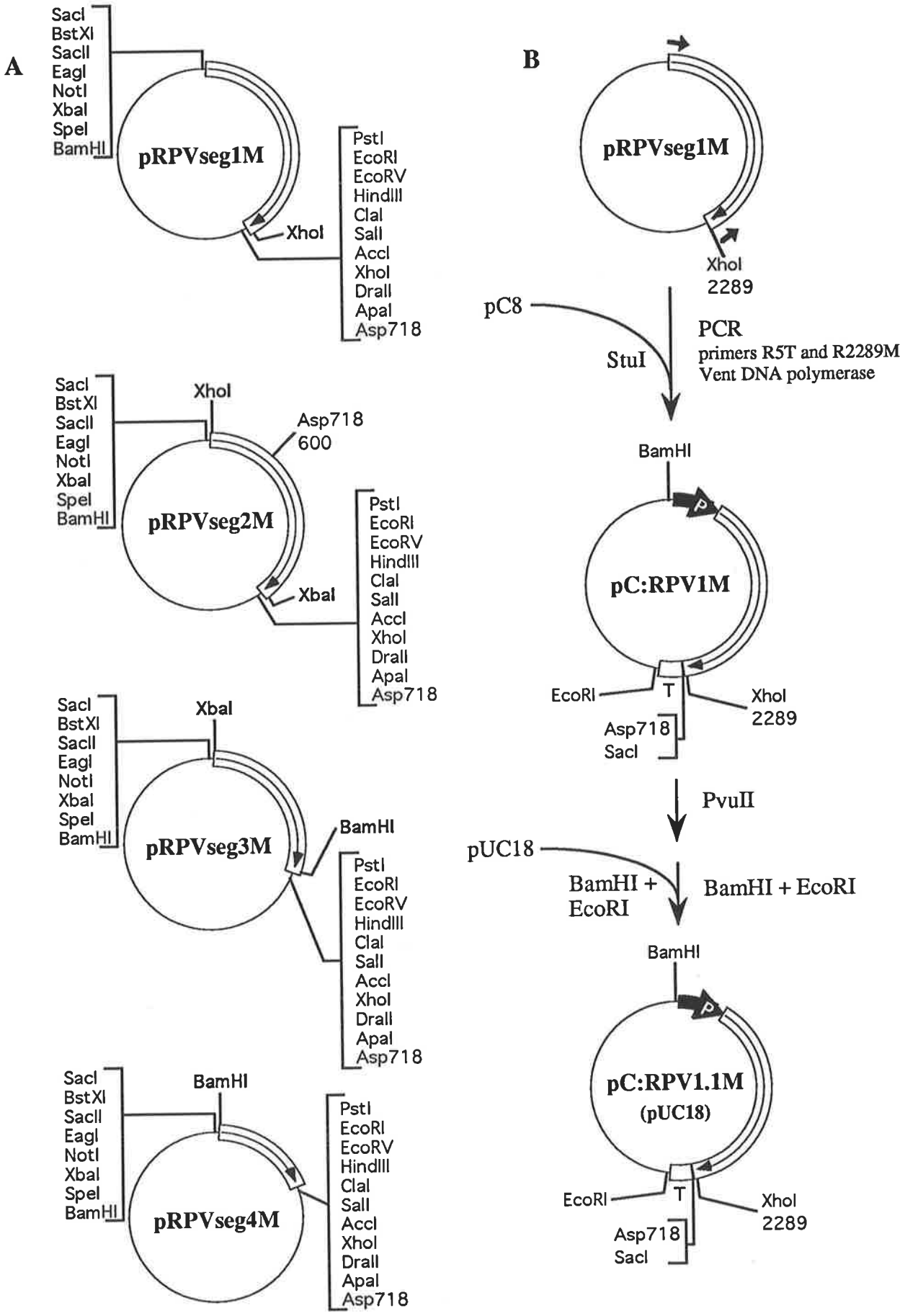

XhoI
PvuII
HindIII
SphI
PstI
35S P
35S T
pCass
StuI
Asp718
SacI
EcoRI
BglII
XhoI + KF
SacI
BstXI
SacII
EagI
NotI
XbaI
SpeI
BamHI
SphI
PstI
SacI
Asp718
PstI
EcoRI
pC8
(pBluescript)
EcoRV
HindIII
ClaI
SalI
AccI
XhoI
DraII
ApaI
Asp718
pBluescript
SmaI
HindIII + BglII
KF
PvuII
HindIII
SphI
PstI
StuI
Asp718
SacI
EcoRI
pC7
BglII
EcoRI + KF
PvuII
HindIII
SphI
PstI
StuI
Asp718
SacI
pC7.1
BglII

and *Eco*RI sites of the pBluescript polylinker, interspersed with recognition sites for *Sph*I and *Pst*I at the 5' end, and *Pst*I at the 3' end.

Assembly of segment 1M into pCass. Segment 1M was PCR-amplified using the oligonucleotide primers RPV5TERM (6.2.5) and RPV 2289M (6.2.7) each at 2.5 μM, and 200 ng of pRPVseg1M (Fig. 6.2A) as template. The reaction was catalysed by Vent DNA polymerase using conditions suggested by the manufacturer, except that the final concentration of $MgSO_4$ was 3 mM. The reaction profile was 94°C/1 min; [94°C/10 s; 50°C/5 s; 72°C/90 s]$_{10}$; 72°C/5 min; 25°C/5 min, performed on a capillary DNA Thermal Sequencer. The PCR product was purified from a 0.7% agarose 1xTBE minigel using Geneclean, and cloned into the *Stu*I site of pC8 (Fig. 6.2B). Clones of the correct orientation were selected by digestion with *Xho*I, and the sequence at the termini of the BYDV-RPV cDNA checked by dideoxy sequencing. The clone derived from this work was designated pC:RPV1M. The complete insert comprising the plasmid polylinker as well as the pCass and BYDV-RPV cDNA sequences was excised from pC:RPV1M by digestion with *Pvu*II, then purified and digested further with *Bam*HI and *Eco*RI. The resulting fragment, containing the BYDV-RPV segment 1M sequence as well as the CaMV 35S transcriptional signals, was cloned into pUC18 also digested with *Bam*HI and *Eco*RI to create pC:RPV1.1M.

Assembly of segments 2M, 3M and 4M. The *Xho*I site 3' of the insert of pRPVseg4M (Fig. 6.2A) was destroyed by cleavage with this enzyme, followed by endfilling and religation to create pRPV4M1.1 (Fig. 6.2C). The viral cDNA was excised from pRPV4M1.1 with *Bam*HI and *Asp*718, and cloned into pSP72 also digested with these enzymes to create pRPV4M1.2. Segments 2M and 3M were cloned into this plasmid in the following manner: Segment 2M cDNA was released from pRPVseg2M (Fig. 6.2A) by digestion with *Xho*I and *Xba*I, while segment 3M cDNA was released from pRPVseg3M (Fig. 6.2A) by digestion with *Xba*I and *Bam*HI. The segments were purified by gel electrophoresis and Geneclean, then assembled in a ligation reaction together with pRPV4M1.2 digested with *Xho*I and *Bam*HI. Recombinant clones derived from this ligation contained segments 2M, 3M and 4M in

Fig. 6.2. Construction of a full-length BYDV-RPV-Vic cDNA clone under transcriptional control of pCass sequences. Interpretation of the diagram is as for Fig. 6.1, except that small dark arrows indicate PCR primers. Viral cDNA sequences are represented with an arrow inside the box showing orientation. (A) Cloned BYDV-RPV-Vic 'M' segments in pBluescript. (B) Assembly of segment 1M and pCass transcriptional sequences. (C) Assembly of BYDV-RPV segments 2M, 3M and 4M. (D) Construction of the full-length clone from clones synthesised in (A), (B) and (C). Abbreviations for PCR primers: R5T refers to RPV5TERM; R2289M is RPV 2289M.


A
SacI
BstXI
SacII
EagI
NotI
XbaI
SpeI
BamHI
pRPVseg1M
XhoI
PstI
EcoRI
EcoRV
HindIII
ClaI
SalI
AccI
XhoI
DraII
ApaI
Asp718
pRPVseg2M
Asp718
600
XbaI
pRPVseg3M
BamHI
pRPVseg4M
B
pRPVseg1M
XhoI
2289
pC8
StuI
PCR
primers R5T and R2289M
Vent DNA polymerase
BamHI
P
pC:RPV1M
EcoRI
T
XhoI
2289
Asp718
SacI
PvuII
pUC18
BamHI +
EcoRI
BamHI + EcoRI
pC:RPV1.1M
(pUC18)

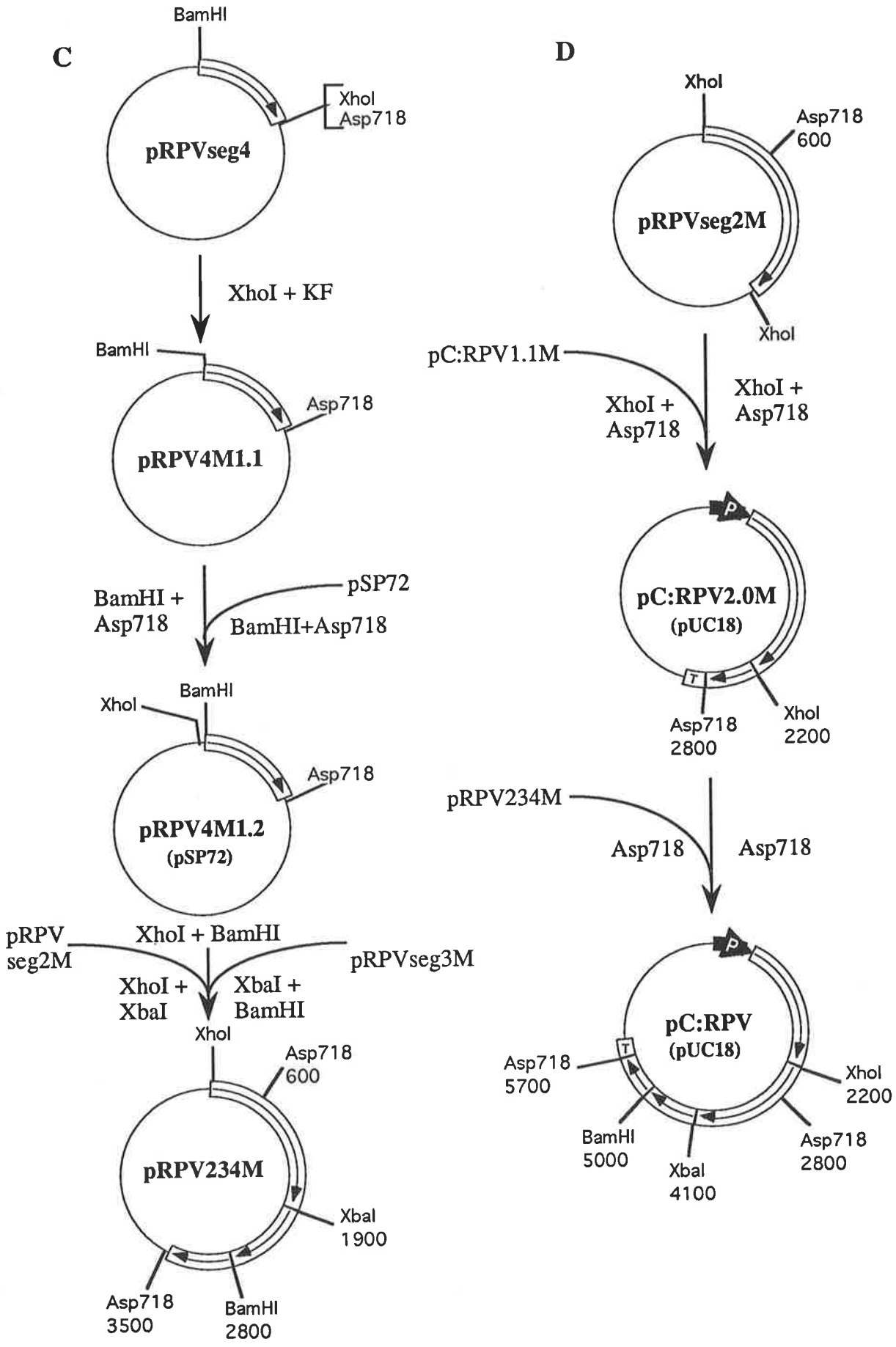

C
BamHI
XhoI
Asp718
pRPVseg4
XhoI + KF
BamHI
Asp718
pRPV4M1.1
BamHI +
Asp718
pSP72
BamHI+Asp718
XhoI
BamHI
Asp718
pRPV4M1.2
(pSP72)
pRPV
seg2M
XhoI + BamHI
pRPVseg3M
XhoI +
XbaI
XbaI +
BamHI
XhoI
Asp718
600
pRPV234M
XbaI
1900
Asp718
3500
BamHI
2800
D
XhoI
Asp718
600
pRPVseg2M
XhoI
pC:RPV1.1M
XhoI +
Asp718
XhoI +
Asp718
P
pC:RPV2.0M
(pUC18)
T
Asp718
2800
XhoI
2200
pRPV234M
Asp718
Asp718
P
pC:RPV
(pUC18)
T
Asp718
5700
XhoI
2200
BamHI
5000
XbaI
4100
Asp718
2800

Fig. 6.2 (cont.)

order and were verified by restriction digestion. The clone selected in this way was designated pRPV234M (Fig. 6.2C).

Assembly of the full-length clone. A fragment of ~600 nt was released from pRPVseg2M (Fig. 6.2A) by digestion with *Xho*I and *Asp*718, and cloned into pC:RPV1.1M digested with these same enzymes to create pC:RPV2.0M (Fig. 6.2D). The remaining BYDV-RPV-Vic cDNA sequences were obtained by digestion of pRPV234M with *Asp*718 to release a fragment of ~2.8 kb, which was cloned into pC:RPV2.0M also digested with *Asp*718. The final clone containing a full-length BYDV-RPV-Vic cDNA in the transcriptional sequences of pCass was designated pC:RPV (Fig. 6.2D), and was selected by restriction digestion with verification of recombinant junctions by dideoxy sequencing.

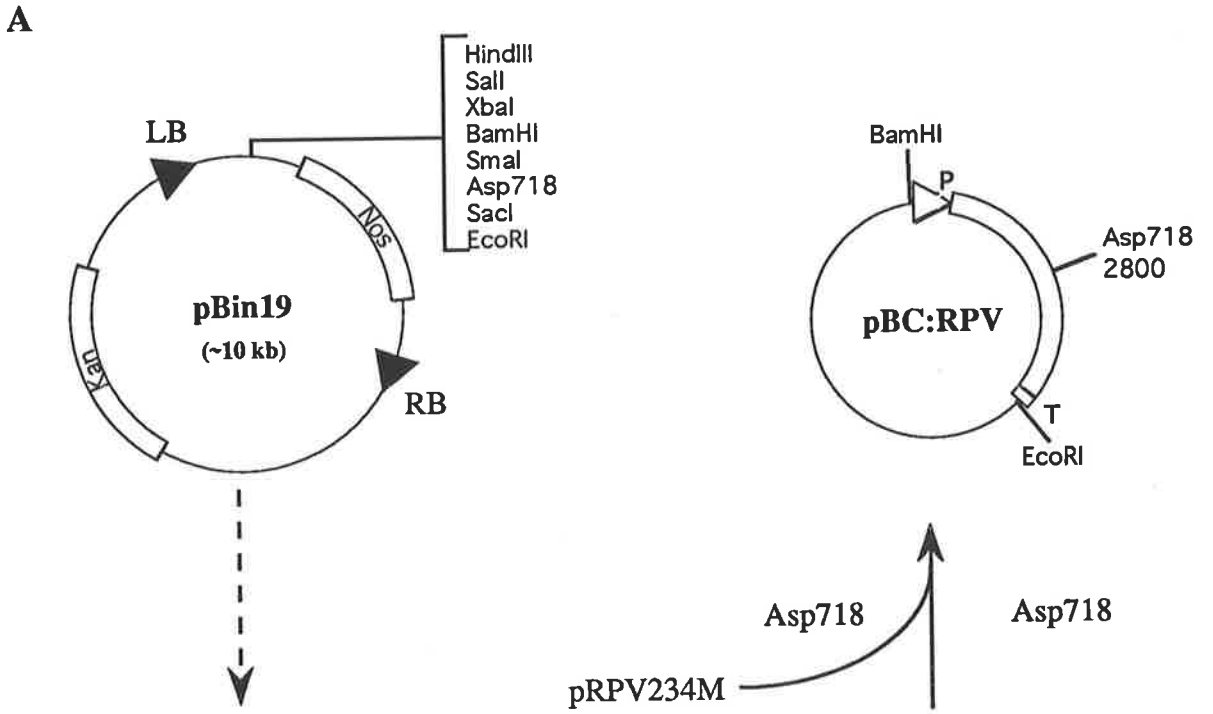
6.2.9 Restriction mapping of pC:RPV

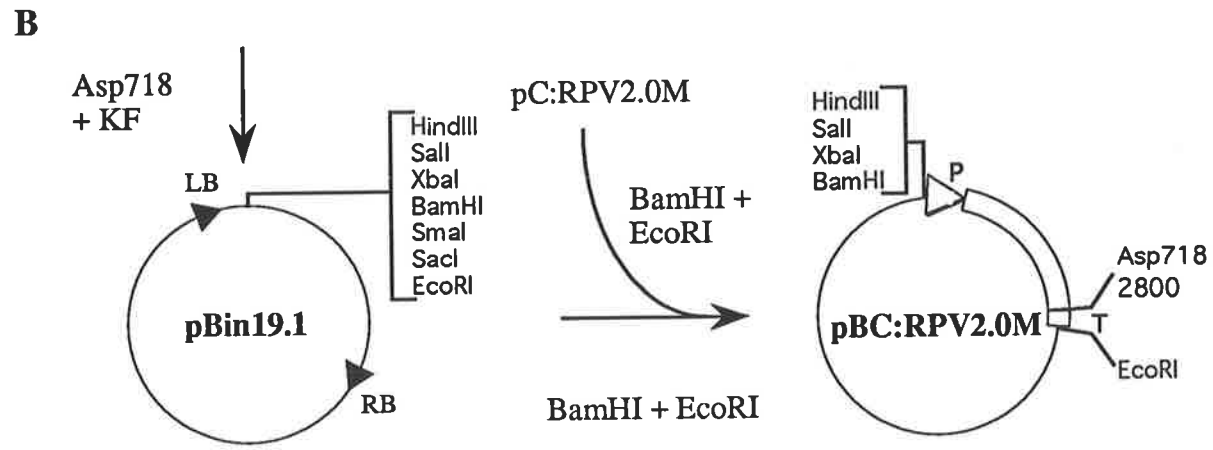
Purified pC:RPV (6.2.8) was treated singly or pairwise with the following restriction endonucleases; *Asp*718, *Bam*HI, *Cla*I, *Eco*RI, *Pst*I, *Sac*I, *Xba*I, *Xho*I. DNA fragments were resolved on large 1.0% agarose 1xTBE slab gels (15 x 20 cm) and visualised with ethidium bromide/UV. Sizing of restriction fragments and determination of the map was performed manually.

6.2.10 Cloning of a full-length BYDV-RPV-Vic cDNA into an *Agrobacterium*/ *E. coli* binary vector

The starting point of this work was the binary vector pBIN19 (Fig. 6.3A; Bevan, 1984), a wide-host range plasmid containing the tandem repeats of the *Agrobacterium* Ti plasmid T-DNA, allowing sequences cloned within the repeats to be transferred to the plant cell nucleus. pBIN19 contains the polylinker sequence from M13mp19 (Norrander *et al.*, 1983), which was modified here by restriction with *Asp*718 and endfilling followed by religation to create pBIN19.1 (Fig. 6.3B). The transcription cassette/viral cDNA fusion was released from plasmid pC:RPV2.0M (6.2.8) and inserted into pBIN19.1 also digested with these enzymes to create pBC:RPV2.0M. The

Fig. 6.3. Cloning of the full-length BYDV-RPV-Vic cDNA under control of pCass transcriptional elements into pBIN19. (A) Structure of pBIN19. Dark triangles represent sequence repeats necessary for transfer of the T-DNA to the plant nucleus (LB and RB stand for Left and Right Borders respectively). Boxes represent genes for kanamycin resistance (Kan) and Nopaline synthase (Nos) respectively. Unique restriction sites in the T-DNA polylinker (derived from M13mp19) are shown. (B) Construction of the full-length BYDV-RPV-Vic clone in pBIN19 (pBC:RPV). Open triangle (P) represents CaMV 35S promoter, T represents CaMV 35S terminator.


A
LB
HindIII
SalI
XbaI
BamHI
SmaI
Asp718
SacI
EcoRI
Nos
pBin19
(~10 kb)
Kan
RB
BamHI
P
Asp718
2800
pBC:RPV
T
EcoRI
Asp718
Asp718
pRPV234M


B
Asp718
+ KF
LB
HindIII
SalI
XbaI
BamHI
SmaI
SacI
EcoRI
pBin19.1
RB
pC:RPV2.0M
BamHI +
EcoRI
BamHI + EcoRI
HindIII
SalI
XbaI
BamHI
P
Asp718
2800
pBC:RPV2.0M
T
EcoRI

remaining sequences of the full-length BYDV-RPV-Vic cDNA were cloned as previously by insertion of the pRPV234M *Asp*718 fragment into *Asp*718 digested pBC:RPV2.0M to create pBC:RPV. The integrity of the final plasmid was checked by restriction digestion and partial sequence analysis.

6.2.11 Cloning of a full-length BYDV-PAV-Vic cDNA into a binary vector under the transcriptional control of pCass sequences

Cloning of BYDV-PAV-Vic 5' end sequences. A derivative of pCass was constructed in which the *Xho*I site was removed. This was achieved by cleavage with *Xho*I, followed by endfilling and ligation to create plasmid pC7 (Fig. 6.1). The 5' ~2 kb of the BYDV-PAV-Vic genome was amplified from a full-length cDNA clone (pBYDV19; Young *et al.*, 1991 (a kind gift of Drs P. Keese and W. Gerlach, CSIRO Division of Plant Industry, Canberra, Australia)) using primers PAV-1-H (5'-AGTGAAGATTGACCATCTCACAAAAGC-3'; homologous to nucleotides 1-27 of the BYDV-PAV-Vic genome) and PAV-1 (5'-TTAATGTCACCGGACATTCTGTGGCC-3'; complementary to nucleotides 2007-2032 of the BYDV-PAV-Vic genome) in a PCR reaction employing Vent DNA polymerase with each primer at a concentration of 2.9 μM. The thermal cycling for the reaction was [94°C/20 s; 60°C/5 s; 72°C/60 s]$_8$; 72°C/5 min; 25°C/5 min and was performed on a capillary DNA Thermal Sequencer under standard conditions except that the final concentration of $MgSO_4$ was 3 mM. The reaction product was purified by gel electrophoresis and Geneclean before cloning into the *Stu*I site of pC7 to create pC:PAV1 (Fig. 6.4A). The orientation and promoter/5' viral cDNA fusion was checked by restriction and sequence analysis respectively, before restriction of the clone with *Xho*I (cuts at nucleotide 535 of the BYDV-PAV-Vic genome) and *Asp*718 (cuts between the 3' end of the viral cDNA and the beginning of the 35S terminator sequence) to remove unwanted viral sequences. The restricted plasmid was purified from the released insert and treated with Klenow fragment and dNTPs to fill in ragged ends, then religated to restore the original *Xho*I site at nucleotide 535 of the BYDV-

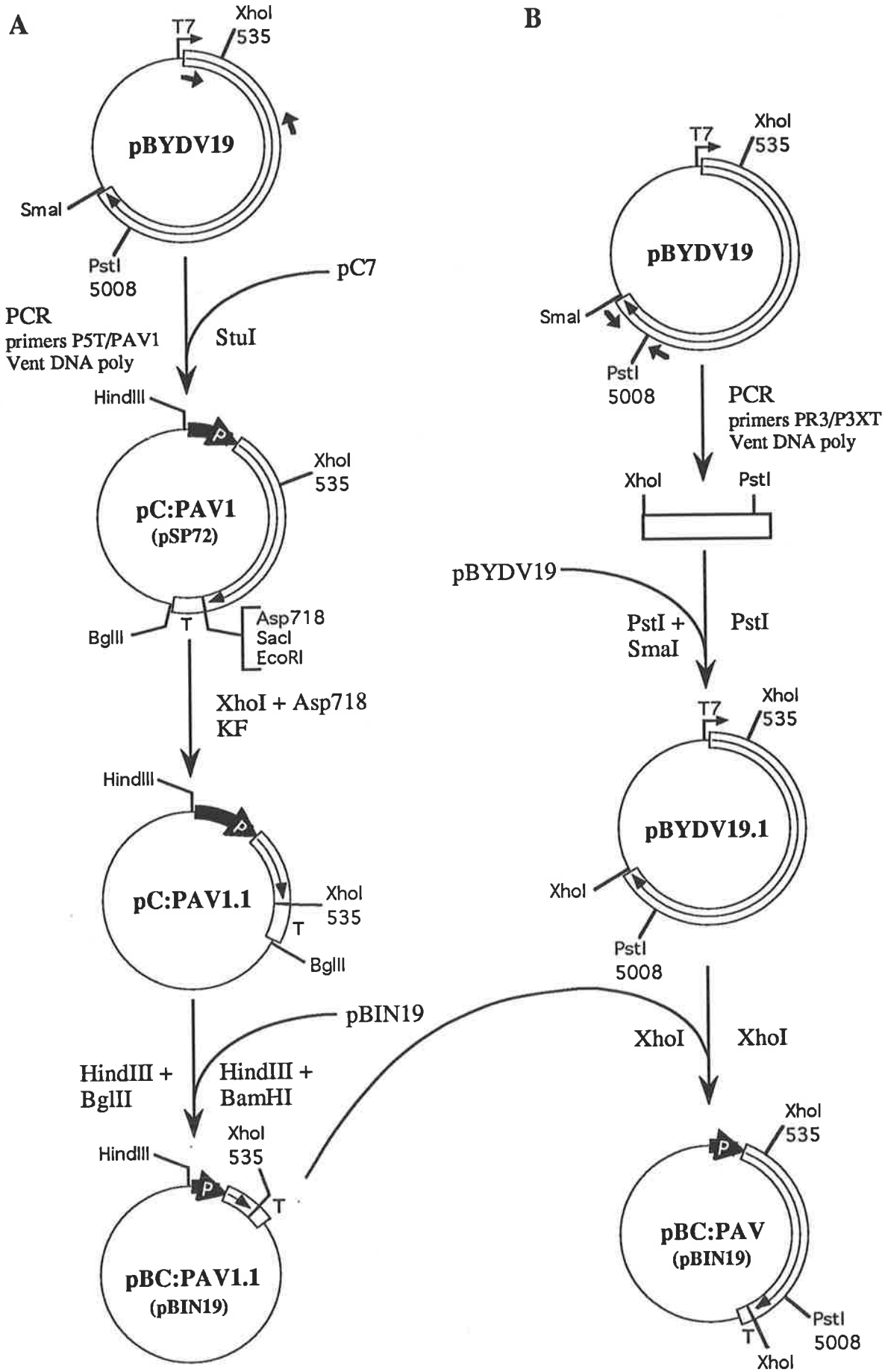
PAV cDNA. This clone was identified by restriction analysis and designated pC:PAV1.1. The insert of pC:PAV1.1 comprising the CaMV 35S transcriptional sequences and the 5' viral cDNA was excised by restriction with *Bgl*II and *Hind*III, then cloned into pBIN19 which had been cleaved with *Bam*HI and *Hind*III thus creating pBC:PAV1.1.

Cloning of the remaining BYDV-PAV sequences. pBYDV19 was modified to contain a *Xho*I site immediately downstream of the 3' terminal cDNA sequence (Fig. 6.4B). The 3' ~750 nt was amplified by PCR using primers PAVRACE3 (5'-ATGAATTCACGTTATCGCCGTTTGTAT-3'; homologous to residues 4933-4951 of the BYDV-PAV-Vic genome) and PAV-3X-TERM (5'-CTAGCTCGAGGGTTGCCGAACTGCTCTTTC-3'; complementary to residues 5657-5677 of the BYDV-PAV-Vic genome), each at 1.1 μM, employing Vent DNA polymerase as suggested by the manufacturer except that the final concentration of $MgSO_4$ was 4 mM. Fifty ng of pBYDV19 was used as template with a reaction profile of 94°C/1 min; [94°C/5 s; 60°C/5 s; 72°C/30 s]$_{20}$; 72°C/5 min; 25°C/5 min performed in a capillary DNA Thermal Sequencer. The reaction product was purified by gel electrophoresis and Geneclean, then restricted with *Pst*I (cuts at nucleotide 5008 of the BYDV-PAV-Vic genome) before cloning into pBYDV19 that had been restricted with *Pst*I and *Sma*I (cuts at nucleotide 5677 of the BYDV-PAV-Vic genome) to create pBYDV19.1. The 3' *Xho*I fragment (comprising nucleotides 535-5677) was excised from pBYDV19.1 by digestion with *Xho*I, and cloned into pBC:PAV1.1 that had also been restricted with this enzyme. Clones carrying the fragment in the correct orientation were identified by restriction analysis, with designation of the selected clone as pBC:PAV. The sequence at recombinant junctions of pBC:PAV was verified by dideoxy nucleotide sequencing.

6.2.12 Agroinfection procedures with clones pBC:RPV and pBC:PAV

Agrobacterium tumefaciens strain C58 was a gift of Dr I.B. Dry, CSIRO Division of Horticulture, Adelaide, South Australia. Full-length BYDV clones

Fig. 6.4. Cloning of the full-length BYDV-PAV-Vic cDNA (Young *et al.*, 1991) into pBIN19 under the control of pCass transcriptional elements. (A) Fusion of the 5' BYDV-PAV cDNA sequences to pCass transcriptional elements, and transfer to pBIN19. (B) Assembly of the remainder of BYDV-PAV cDNA sequences to complete the full-length clone in pBIN19 (pBC:PAV). Abbreviations for PCR primers: P5T refers to PAV5TERM; PAV1 is PAV-1; PR3 is PAVRACE3; P3XT is PAV-3X-TERM.


A
T7
XhoI
535
pBYDV19
SmaI
PstI
5008
PCR
primers P5T/PAV1
Vent DNA poly
pC7
StuI
HindIII
P
XhoI
535
pC:PAV1
(pSP72)
BglII
T
Asp718
SacI
EcoRI
XhoI + Asp718
KF
HindIII
P
pC:PAV1.1
XhoI
535
T
BglII
pBIN19
HindIII +
BglII
HindIII +
BamHI
XhoI
535
HindIII
P
T
pBC:PAV1.1
(pBIN19)
B
T7
XhoI
535
pBYDV19
SmaI
PstI
5008
PCR
primers PR3/P3XT
Vent DNA poly
XhoI
PstI
pBYDV19
PstI +
SmaI
PstI
T7
XhoI
535
pBYDV19.1
XhoI
PstI
5008
XhoI
XhoI
XhoI
535
P
pBC:PAV
(pBIN19)
PstI
5008
T
XhoI

pBC:RPV and pBC:PAV were transformed into this strain by electroporation exactly as described by Wen-jun and Forde (1989). Transformants were recovered by selective growth on LB agar plates containing 50 µg/ml kanamycin and 25 µg/ml rifampicin after incubation at 30°C for 2 days. Single colonies were selected and grown in 20 ml 2YT containing 50 µg/ml kanamycin and 25 µg/ml rifampicin for two days at 30°C with shaking, then subcultured 1:20 into 10 ml of new growth media containing antibiotics as previously, and incubated overnight at 30°C. Overnight cultures were centrifuged at 5,000 rpm for 10 min in an HB4 rotor at 4°C, after which the supernatant was discarded and the bacterial pellet drained. The bacteria were resuspended in 500 µl of 50 mM Tris-HCl pH 7.0 and kept on ice until used for agroinoculation.

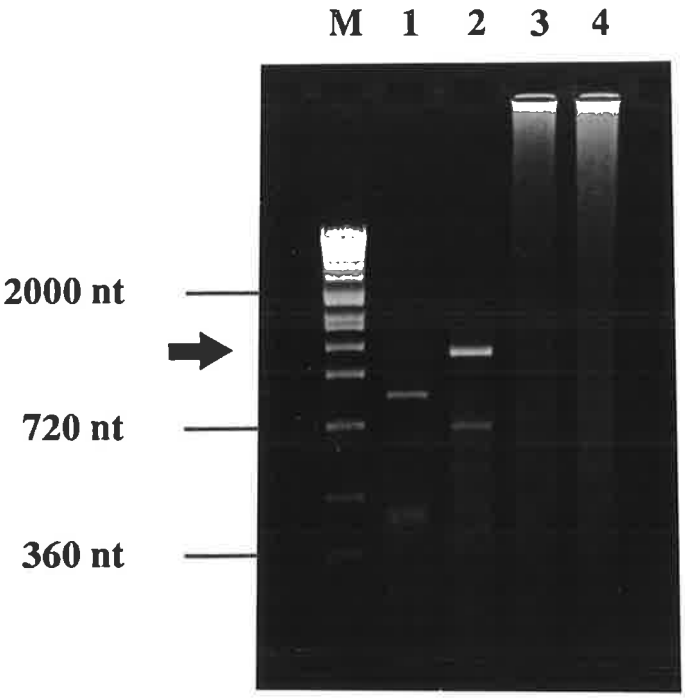
Thick suspensions of *Agrobacterium* as outlined above were inoculated onto week-old oat seedlings (*Avena sativa* cv. Stout) by injection into the area at the base of the shoot containing the vegetative meristem (Marks *et al.*, 1989). Two to five injections of five µl each were injected into each meristem using a 20 µl Hamilton syringe and a disposable 25 gauge needle. The needle was discarded after injection of the seedling, and the syringe washed with 70% ethanol and sterile water before performing the next inoculation. Inoculated plants were grown in insect-proof cages in a C1 containment glasshouse under natural lighting conditions and a constant temperature regime of 21°C. Infection was monitored by dot-blot detection of viral RNA in crude plant nucleic acid extracts using radioactive RNA probes (2.2.13).

6.3 Results

6.3.1 Isolation of BYDV-RPV-Vic dsRNA

The double-stranded viral RNA associated with infection by single-strand plant RNA viruses represents an intermediate form of the viral genome generated during replication of the positive-sense RNA. dsRNA is easily purified because of its affinity to CF-11 cellulose under high salt and ethanol conditions (*e.g.* Morris and Dodds, 1979; Smith *et al.*, 1991), so offers a convenient alternative to the purification of virus particles for the isolation of luteoviral RNA. Proof that both minus and plus strands of

Fig. 6.5. PCR amplification from positive and negative strands of BYDV-RPV-Vic dsRNA. First-strand cDNA synthesis was primed from minus-strand (lane 1) or positive strand (lane 2) BYDV-RPV-Vic RNA. Lanes 3 and 4 used the same first-strand cDNA oligonucleotide primer as for lanes 1 and 2 respectively, except that no dsRNA was added as template. Sizes of three molecular weight standards (lane M) are indicated. The PCR product amplified from the first-strand cDNA (most visible in lane 2) is ~1100 nt in length and migrates just ahead of the 1160 nt molecular weight standard (not indicated).


M 1 2 3 4
2000 nt
720 nt
360 nt

the genomic RNA were present in the dsRNA preparation described here was obtained by PCR amplification of cDNAs derived from each strand. The PCR product of 1100 nt was amplified from cDNAs primed from oligonucleotides hybridising to either the plus or the minus strand genomic RNA (Fig. 6.5). No PCR product was obtained from either first-strand cDNA primer when dsRNA was not added to the reverse transcription reaction. Less DNA was amplified when cDNA synthesis was primed from the minus-strand; this may be because less minus-strand RNA was present in the dsRNA preparation, or reflect a poorer efficiency of cDNA synthesis from the minus strand. This result shows that BYDV-PAV dsRNA was successfully purified from infected oat tissue, in sufficient quantity for amplification by PCR.

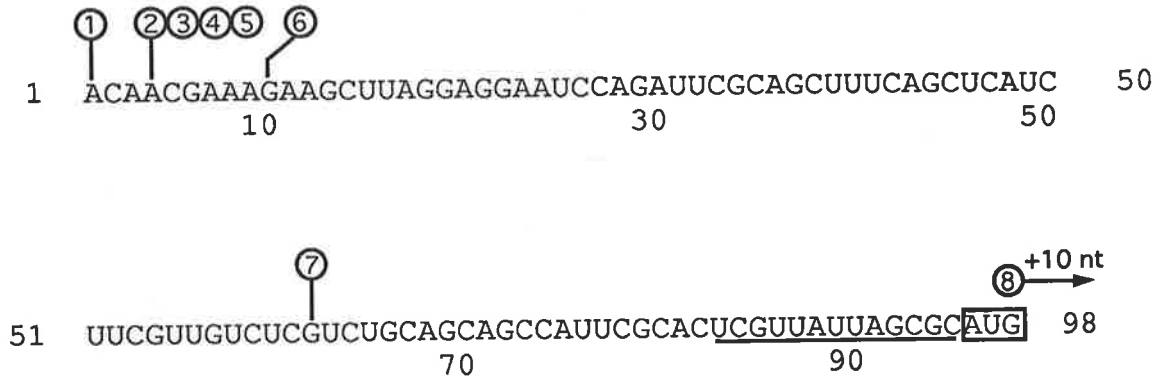
6.3.2 Determination of the 5' and 3' terminal sequences of the BYDV-RPV-Vic RNA genome

A RACE protocol was used to determine the sequences at the termini of the BYDV-RPV-Vic genomic RNA. This was carried out largely as described for SDV (Chapter Three), but using dsRNA (6.3.1) as an initial template for cDNA synthesis. This allowed direct priming of first-strand cDNA synthesis off genomic RNA strands of each polarity, obviating the need for treatment of the positive-strand genomic RNA with poly(A) polymerase as in Chapter Three. The oligonucleotides used for first-strand cDNA synthesis were also used in the RACE-PCR reaction. As no BYDV-RPV sequence was known when this work was underway, the sequences of the oligonucleotides were designed to conserved regions in the deduced amino acid sequence of ORFs of BWYV and a Chinese BYDV-RPV-like isolate, BYDV-GPV (Cheng *et al.*, 1994b). A single band was obtained after the PCR reaction for both genomic termini, and subsequent clones were identified by sequence analysis.

Eight clones corresponding to the 5' genomic sequences of BYDV-RPV-Vic were sequenced. Only one of the clones contained all the sequence that was deemed by comparison to other luteovirus sequences to constitute the full-length 5' leader sequence of BYDV-RPV. Four other clones initiated at +4 nucleotides, one at +10, and the

Fig. 6.6. RACE determination of BYDV-RPV-Vic 5' genomic RNA sequence. (A) Nucleotide sequence of the 5' genomic region upstream of the first (ORF 0) initiation codon. The codon corresponding to the AUG of ORF 0 in BYDV-RPV-NY is boxed. The only stretch of sequence with detectable similarity to BYDV-RPV-NY is underlined. The 5' extremity of each of eight 5' RACE clones is indicated. Clone 8 maps 10 nucleotides downstream of the ORF 0 AUG codon (arrow). (B) Alignment of the 5' terminal nucleotide sequences of SBMV, subgroup II luteoviruses including BYDV-RPV-Vic, and RCNMV. The sequences were aligned manually. Stars indicate consensus nucleotides in the sequence of subgroup II luteoviruses in this alignment. Conserved nucleotides are indicated in bold.

A



B

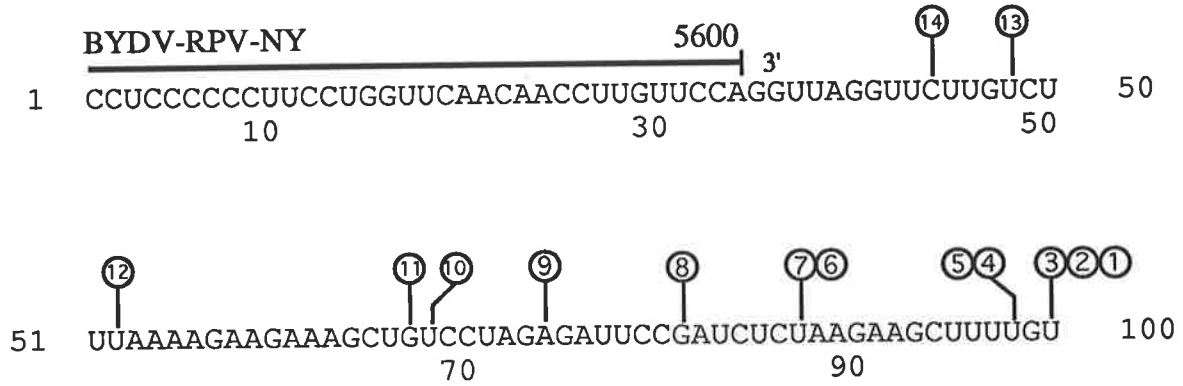
SBMV	C**ACAAAA**U**A**UA**A**GA**AG**GAAA
BWYV	**ACAAAA**GA**A**-**A**CC**AG**GAGG
PLRV-A	**ACAAAA**GA**A**UA**A**CC**AG**GAGA
CABYV	**ACAAAA**GA**A**-UA**A**CG**AG**CGGG
	****** * * **
RPV-VIC	**ACAA**-CG**A**AA**A**GA**AG**CUUA
RPV-NY	CGGCUCCA**A**AGCUU**A**AGCUC
RCNMV	G**ACAAA**CGUUUUACC**G**GUUU

remaining two clones represented more severely truncated products of the RACE reaction (Fig. 6.6A). Definition of the 5' genomic terminus in this way gives a 5' leader sequence of 101 nt before the AUG initiation codon of ORF 0, which compares to 114 nt for BYDV-RPV-NY. It is possible that the published 5' terminal sequence of BYDV-RPV-NY does not represent the true extent of that sequence (Vincent *et al.*, 1991). Alignment of the 5' terminal nucleotide sequence of BYDV-RPV-Vic with those of other subgroup II luteoviruses (Fig. 6.6B) reveals that only the first four nucleotides of the BYDV-RPV-Vic sequence correspond to the luteovirus consensus sequence ACAAAa defined by Miller *et al.* (1994). BYDV-RPV-Vic shares varying homology with SBMV and subgroup II luteoviruses over the first ~20 nt of the genomic sequence, but the sequence of BYDV-RPV-NY published for this region is completely dissimilar. The 5' terminal sequence of RCNMV RNA 1 also shows homology to the sequence of BYDV-RPV-Vic in particular, but also to other subgroup II luteoviruses (as pointed out by Miller *et al.*, 1994). This is of interest because RCNMV possesses a carmovirus-like RdRp ORF, as against the sobemovirus-like ORF of the subgroup II luteoviruses.

Variable length of clones was also a feature of the RACE determination of the BYDV-RPV-Vic 3' genomic terminus. Of 14 clones sequenced, only three contained the sequence deemed to represent full-length, with the others variously spaced inward from the genomic terminus over 56 nt (Fig. 6.7A). The length of the 3' UTR given by this 3' end determination is 168 nt, compared with 102 nt for BYDV-RPV-NY. The 3' terminal sequence of BYDV-RPV-NY closely matches that of BYDV-RPV-Vic to its last nucleotide, leaving an extra 66 nt at the 3' end of BYDV-RPV-Vic (Fig. 6.7A). This suggests that the published sequence of BYDV-RPV-NY (Vincent *et al.*, 1991) does not represent the complete extent of the BYDV-RPV genome. Strikingly, the reverse complement of the extreme 3' terminal sequence encodes the conserved 5'-ACAAAAG-3' that is found at the 5' genomic terminus of BWYV, PLRV and CABYV, which is closely related to that at the 5' end of the BYDV-RPV-Vic genome

Fig. 6.7. RACE determination of BYDV-RPV-Vic 3' genomic RNA sequence. (A) Nucleotide sequence of the 3' genomic region. The 3' extent of the BYDV-RPV-NY genome is shown (thick line; Vincent *et al.*, 1991). The 3' extremity of each of 14 3' RACE clones is indicated. (B) Alignment of the reverse complement of the BYDV-RPV-Vic 3' genomic sequence with the 5' genomic sequence of SBMV and subgroup II luteoviruses. The sequences were aligned manually. Conserved nucleotides are indicated in bold.

A



B

SBMV	C**ACAAAA**U**A**UA**AG**A**AG**GAAA
BWYV	**ACAAAAGAA**-**A**CC**AG**GAGG
PLRV-A	**ACAAAAGAA**UACC**AG**GAGA
CABYV	**ACAAAAGA**-UA**C**G**AG**CGGG
RPV-Vic	**ACAA**-C**GAAAA**G**AAG**CUUA
RPV-Vic 3'RC	**ACAAAAGAA**--GA**A**UCUCA

(Fig. 6.7B). This has not been previously reported for any subgroup II luteovirus although all contain the sequence 5'-GU-3' at their respective 3' genomic termini.

6.3.3 PCR-mediated construction of a full-length BYDV-RPV-Vic cDNA clone

The entire genome of BYDV-RPV-Vic was amplified in four segments from first-strand cDNA derived from dsRNA (Fig. 6.8). In most cases, success of the amplification depended on the use of high concentrations of the oligonucleotide primers in the PCR reaction. The region encompassing segments 3 and 4 was originally intended to be amplified as a single fragment, but the failure of this approach (data not shown) led to amplification of the segments as two pieces. The four segments were subjected to restriction analysis to design a strategy for assembly of the full-length cDNA clone using standard cloning techniques. The segments were restricted without prior cloning in order to avoid artefactual results due to PCR-generated mutations. The analysis was intended to identify restriction enzymes without recognition sites in adjacent fragments, so that these enzymes could be used to join the segments after appropriate mutagenesis to the segments. This approach negates the requirement for determination of the complete nucleotide sequence of BYDV-RPV-Vic before construction of a full-length cDNA is possible. Table 6.1 summarises the restriction data for each segment, and reveals the following; *Xho*I does not cut in either segment 1 or 2; *Xba*I does not cut in segments 2 or 3; and *Bam*HI does not cut in segments 3 or 4. Therefore, restriction sites for these three enzymes were engineered into the segment cDNAs by PCR to facilitate assembly of the full-length clone.

PCR primers carrying engineered restriction sites were designed as follows. Sequences in the coding regions of BYDV-RPV-NY suitable for mutagenesis were selected with regard to two criteria: The site must be close to the primer binding site of the initial cDNA segments (1-4), and must require minimal modification to incorporate the restriction site, without affecting the deduced amino acid sequence of the ORF. The sequence of the selected site was also determined in BYDV-RPV-Vic from clones

Fig. 6.8. Amplification of cDNAs covering the entire BYDV-RPV-Vic RNA genome. Coverage of the genome and size of the amplified fragments is indicated. Sizes were determined by comparison to the sequence of BYDV-RPV-NY. Segment 4 is greater than the indicated (deduced) size of 516 nt because of the extra sequences at the 3' end of the BYDV-RPV-Vic genome (see text).

BYDV-RPV-NY (5600 nt)

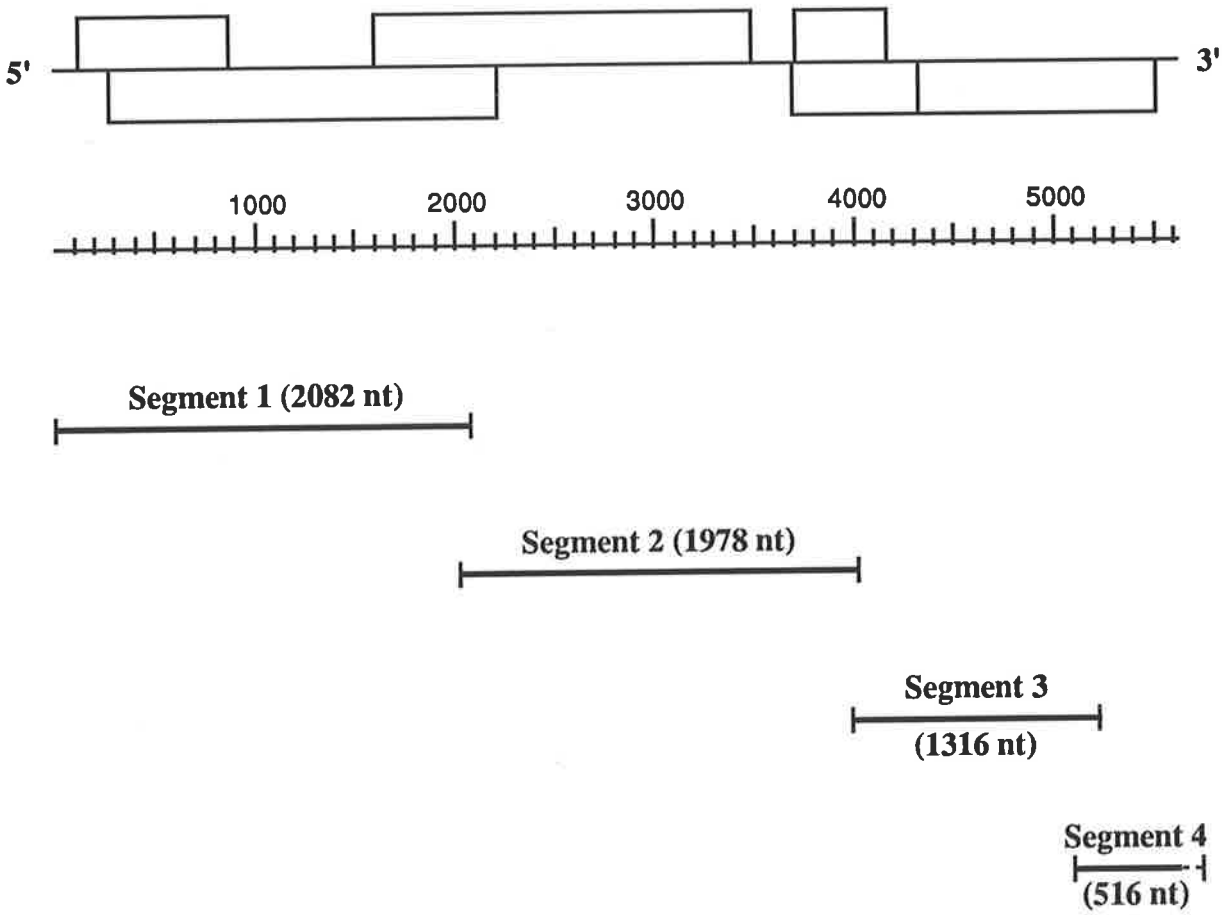


Table 6.1. Restriction sites in the four BYDV-RPV-Vic cDNA segments

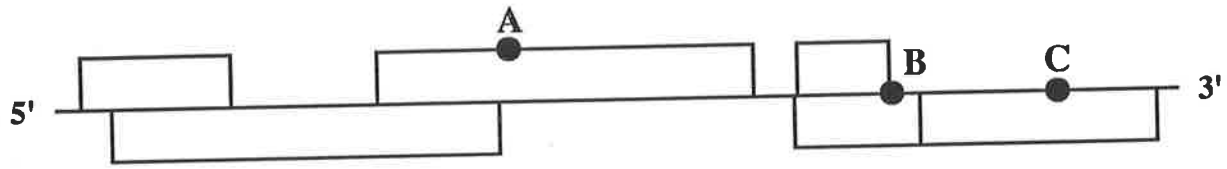
Segment	*Bam*HI	*Bcl*I	*Bgl*II	*Cla*I	*Eco*RI	*Pst*I	*Sal*I	*Sph*I	*Xba*I	*Xho*I
1	-[a]	ND[b]	ND	ND	-	+[c]	+	?[d]	ND	-
2	+	+	+	+	+	-	-	?	-	-
3	-	+	-	-	+	+	+	?	-	-
4	-	ND	ND	ND	+	ND	ND	ND	ND	ND

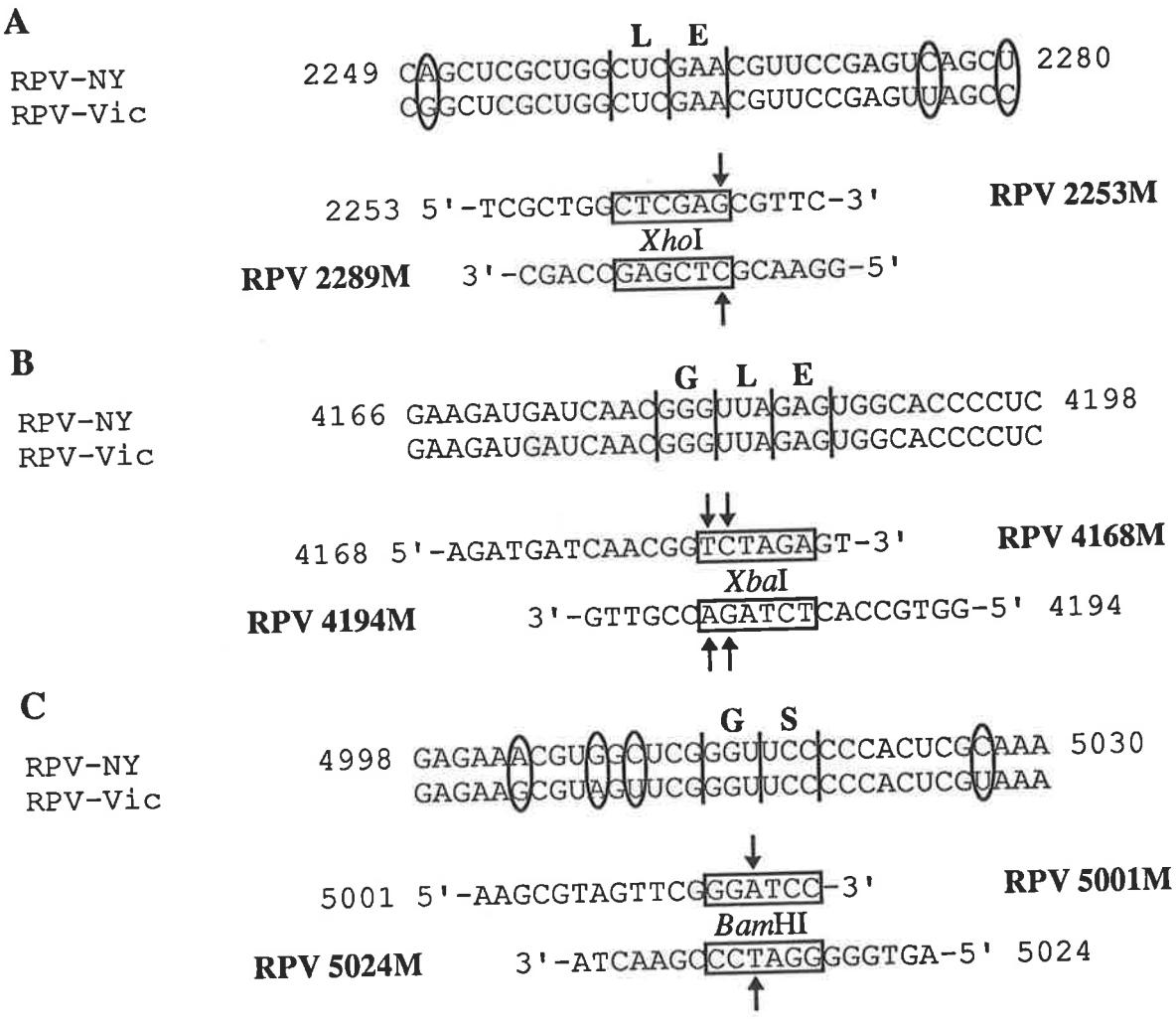
[a]Restriction site not detected; [b]not determined; [c]restriction site detected; [d]result unclear.

constructed in 6.2.5 (pRPVseg1-4) to ensure that it was conserved, so that mutagenesis would not change the identity of the encoded amino acid. The following modifications were made (Fig. 6.9); the sequence (2260) 5'-CUC GAA-3' encoding the amino acids **LE** in ORF 2 was changed to 5'-CUC GA*G*-3' at the junctions of segments 1M and 2M, creating a recognition site for *Xho*I; the sequence (4179) 5'-GGG UUA GAG-3' encoding the amino acids **GLE** in ORF 3 was changed to 5'-GG*U*-*C*UA-GAG-3', thus encoding an *Xba*I recognition site; and the sequence (5013) 5'-GGU UCC-3' encoding the amino acids **GS** in ORF 5 was modified to 5'-GG*A* UCC-3', creating a recognition site for *Bam*HI. The mutant segments were amplified and cloned before restriction with the relevant endonuclease, firstly to confirm the incorporation of the mutation, and also to preclude the existence of further sites in the clone. Both points were established for all four clones (data not shown).

Construction of the full-length cDNA clone of BYDV-RPV-Vic proceeded with fusion of the 5' end of segment 1M (representing the 5' end of the viral genome) to the 3' end of the modified CaMV 35S promoter of pCass (Fig. 6.2). This cloning step was mediated by PCR as described in Chapter Five, using a large amount of cloned segment

Fig. 6.9. Mutant primers for the synthesis of overlapping cDNAs of BYDV-RPV-Vic genomic RNA. Position of the primer sites in the genome is indicated (filled black circles). The nucleotide sequences of BYDV-RPV-NY and -Vic at the primer sites are aligned, with sequence variations circled. The deduced amino acid sequence of the ORF at the mutation site is represented. Mutations introduced to the primer sequences are indicated by arrows, and restriction sites are boxed. The name of each primer occurs at the 3' end of the primer sequence. (A) Primers at the overlap of segments 1M and 2M. A recognition site for *Xho*I has been created at the amino acid sequence **LE**. (B) Primers at the overlap of segments 2M and 3M. A recognition site for *Xba*I has been created at the amino acid sequence **GLE**. (C) Primers at the overlap of segments 3M and 4M. A recognition site for *Bam*HI has been created at the amino acid sequence **GS**.


A
B
C
5'
3'


A
L E
RPV-NY
RPV-Vic
2249 CAGCUCGCUGGCUCGAACGUUCCGAGUCAGCU 2280
CGGCUCGCUGGCUCGAACGUUCCGAGUUAGCC
2253 5'-TCGCTGGCTCGAGCGTTC-3'
RPV 2253M
XhoI
RPV 2289M
3'-CGACCGAGCTCGCAAGG-5'
B
G L E
RPV-NY
RPV-Vic
4166 GAAGAUGAUCAACGGGUUAGAGUGGCACCCCUC 4198
GAAGAUGAUCAACGGGUUAGAGUGGCACCCCUC
4168 5'-AGATGATCAACGGTCTAGAGT-3'
RPV 4168M
XbaI
RPV 4194M
3'-GTTGCCAGATCTCACCGTGG-5' 4194
C
G S
RPV-NY
RPV-Vic
4998 GAGAAACGUGGCUCGGGUUCCCCCACUCGCAAA 5030
GAGAAGCGUAGUUCGGGUUCCCCCACUCGUAAA
5001 5'-AAGCGTAGTTCGGGATCC-3'
RPV 5001M
BamHI
RPV 5024M
3'-ATCAAGCCTAGGGGGTGA-5' 5024

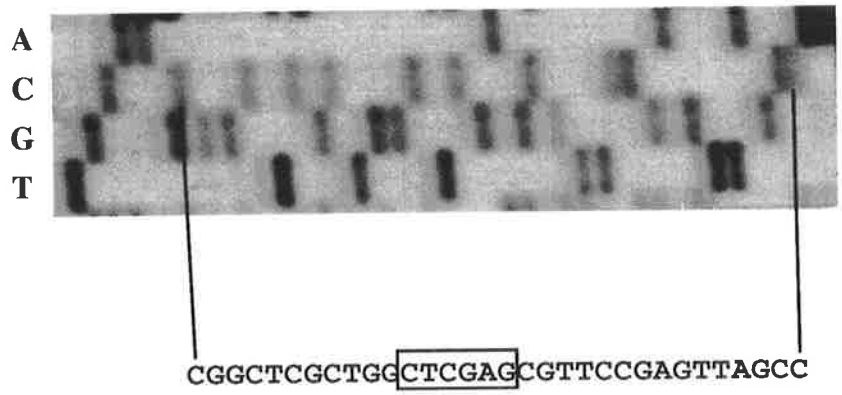
1M as template to restrict the number of cycles in PCR to 10. This reduces the probability of introducing mutations to the cDNA during PCR. The rest of the cDNA was assembled in a straight forward manner using the three introduced restriction sites, and an *Asp*718 site ~600 nt downstream of the *Xho*I site. The recombinant junctions at the 5' and 3' ends of full-length cDNA, as well as the *Xho*I, *Xba*I and *Bam*HI sites, were checked by sequencing and found to faithfully represent the sequences represented by the PCR primers (Fig. 6.10). The fidelity of the sequence at the *Asp*718 site was subsequently checked in the binary vector clone pBC:RPV (data not shown).

6.3.4 Restriction analysis of pC:RPV

The full-length BYDV-RPV cDNA clone pC:RPV was cleaved singly and pairwise with different restriction enzymes, after which a restriction map of the clone was constructed. This served to verify the correct assembly of the mutated cDNA segments, and allows a basis for comparison of the BYDV-RPV-Vic sequence with that of BYDV-RPV-NY (Fig. 6.11). The restriction map derived from this analysis shows that the introduced mutations occur in correct order and with accurate spacing. Thus the single *Xho*I site maps to 2200 nt (expected 2275), the *Xba*I site maps 1900 nt downstream (expected 1879) at 4100 nt, and the *Bam*HI site occurs 800 nt further along (expected 850 nt) at 4900 nt. This leaves approximately 800 nt at the 3' end (expected 700 nt), and gives a total map length of 5700 nt. Restriction sites for *Pst*I, *Cla*I and *Asp*718 were also placed on the map. Computer analysis of restriction sites for these enzymes in the BYDV-RPV-NY genome reveals few similarities with the genome of the Victorian isolate. The collection of seven restriction enzymes cuts the BYDV-RPV-Vic genome in only 15 places, which includes the four introduced sites. In contrast, the BYDV-RPV-NY genome is cut in 17 places by the same group of enzymes. Only seven out of the total number of sites appeared to be present in both genomes. However, it should be noted that restriction fragments of less than 50 nt were unlikely to be detectable in this analysis because of the relatively low resolution of agarose gel electrophoresis.

Fig. 6.10. Sequence integrity at the mutated overlap regions in the full-length clone pC:RPV. Dideoxynucleotide sequencing of pC:RPV at the regions of overlap (see Fig. 6.9) are shown. The lanes read A, C, G, T downwards as indicated. The sequence of pC:RPV is given above the sequence of native BYDV-RPV-Vic, with mutated restriction sites boxed. Position of the sequences in the BYDV-RPV-NY genome is indicated (brackets). (A) Overlap between segments 1M and 2M. (B) Overlap between segments 2M and 3M. (C) Overlap between segments 3M and 4M. Refer to Fig. 6.9 for details of mutagenesis.

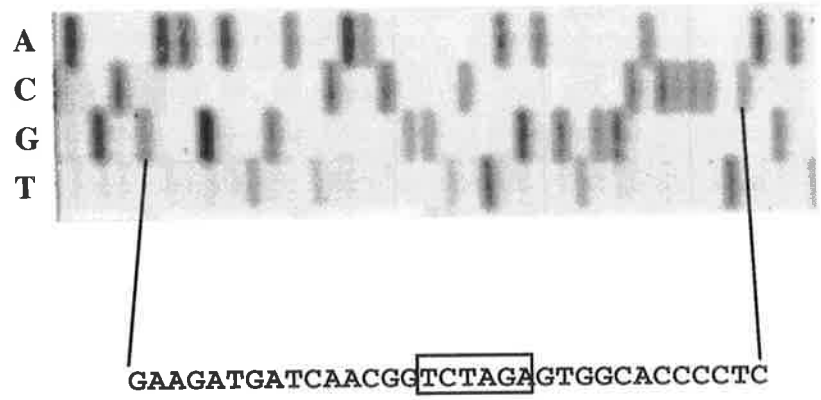
A



pC:RPV CGGCTCGCTGGCTCGAGCGTTCCGAGTTAGCC

RPV-Vic (2249) CGGCUCGCUGGCUCGAACGUUCCGAGUUAGCC (2280)

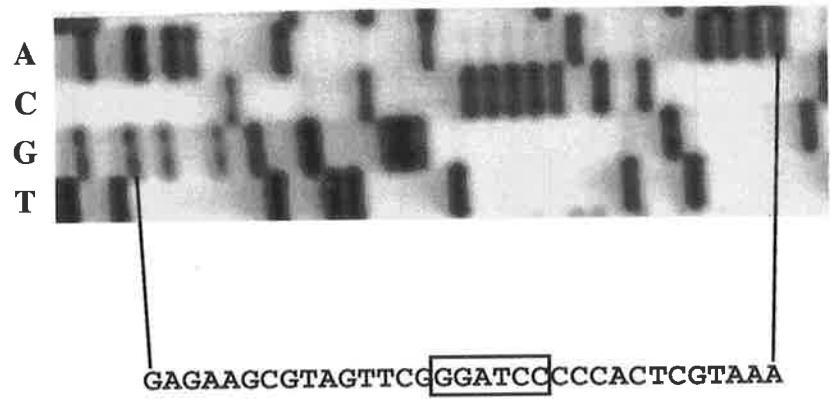
B



pC:RPV GAAGATGATCAACGGTCTAGAGTGGCACCCCTC

RPV-Vic (4166) GAAGAUGAUCAACGGGUUAGAGUGGCACCCCUC (4198)

C



pC:RPV GAGAAGCGTAGTTCGGGATCCCCCACTCGTAAA

RPV-Vic (4998) GAGAAGCGUAGUUCGGGUUCCCCCACUCGUAAA (5030)

6.3.5 Transfer of full-length BYDV cDNAs to the binary vector pBIN19

Full-length cDNAs of BYDV-RPV and BYDV-PAV were transferred to the binary vector pBIN19 (Bevan, 1994) using similar procedures (Figs 6.3 and 6.4). As for all cloning procedures using pCass or its derivates, the first step involved precise fusion of the 5' viral cDNA to the 35S promoter mediated by PCR and blunt-end cloning. The cloned 5' cDNAs were then modified so that the remaining sequences of the cDNA could be cloned in a single step. This is important because cloning into pBIN19 is technically difficult because of its large size (~10 kb), and the difficulty in purifying large amounts of the plasmid due to its wide host-range (= low copy number) origin of replication. The BYDV-RPV 5' cDNA clone (pC:RPV1.1M) was modified by addition of a *Xho*I-*Asp*718 fragment (pC:RPV2.0M), allowing cloning of the 3' sequences as an *Asp*718 fragment. The BYDV-PAV 5' cDNA (pC:PAV1) was modified by excision of the 3' ~1500 nt of the viral cDNA using *Xho*I and *Asp*718, followed by endfilling and religation to recreate the *Xho*I site (pC:PAV1.1M). The remaining 3' cDNA section could then be cloned as a *Xho*I fragment, after modification of the full-length cDNA clone (pBYDV19) to contain a recognition site for *Xho*I downstream of the 3' terminal nucleotide. The 5' cDNA fragments cloned into the pCass transcriptional sequences were cloned into pBIN19 before assembly of the full-length cDNA using *Asp*718 (BYDV-RPV) or *Xho*I (BYDV-PAV). The integrity of all recombinant junctions was checked both in pBC:RPV and pBC:PAV by dideoxy nucleotide sequencing (data not shown).

6.3.6 Initial agroinfection experiments with pBC:RPV and pBC:PAV

Full-length BYDV cDNA clones pBC:RPV and pBC:PAV were electroporated into *Agrobacterium tumefaciens* strain C58, and single colonies selected on solid media containing kanamycin (kanamycin resistance is carried on the parent plasmid pBIN19). Fresh cultures of transformed *Agrobacterium* were concentrated in 50 mM Tris-HCl pH 7.0 and inoculated to week-old oat seedlings by injection with a Hamilton syringe into the approximate area of the vegetative meristem of the seedling. Seedlings were

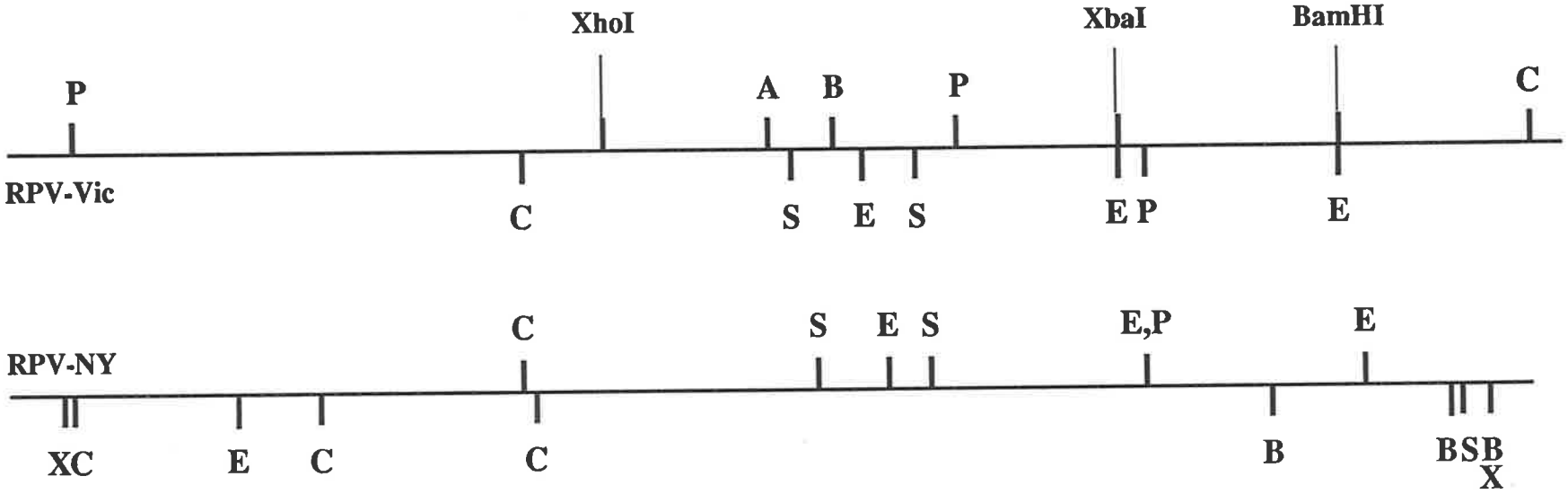
injected 2-5 times with 5 µl of the concentrated bacterial suspension, then planted into pots containing fresh sterile soil. The plants were grown for approximately one month after inoculation after which they were assayed for viral RNA content using a northern dot-blot procedure. Infection was not detected in any of the agroinoculated plants for either construct. Control plants infected by infestation with aphids viruliferous for each strain of BYDV gave positive signals in the dot-blot assay, while healthy controls were negative (data not shown). Lack of time prevented further attempts to find conditions for the successful agroinoculation of plants using pBC:RPV and pBC:PAV.

6.4 Discussion

The major part of this Chapter describes construction of a full-length cDNA clone corresponding to the genomic RNA of a Victorian isolate of BYDV-RPV of unknown sequence. The strategy involved affinity purification of dsRNA from BYDV-RPV-Vic infected plants, followed by determination of the nucleotide sequence at the genomic termini, which allowed a PCR-based approach for synthesis of cDNAs covering the genomic RNA. This compares with a more conventional approach (*cf.* Chapter Three) where cDNAs are primed randomly on genomic RNA purified from virus particles, allowing sequence determination after which full-length cDNA construction can proceed. The principal advantage of the procedure described here is the comparative rapidity in which the full-length clone can be synthesised, although its success is predicated on the existence of sequence data for a closely related viral isolate. The drawback of the method is the reliance on PCR amplification with its potential to introduce unwanted nucleotide changes into the cDNA sequence, although this was offset here to some extent by use of a high-fidelity DNA polymerase.

The use of dsRNA provides an attractive alternative to purification of virus particles for the isolation of viral RNA. The low titre of luteoviruses in host tissue results in a poor yield of virus particles during purification, typically 0.5-0.8 mg per kilogram of tissue (Hammond *et al.*, 1983). While the yield of dsRNA was not quantified here, it was adequate for PCR amplification of large cDNA segments (≈2 kb)

Fig. 6.11. Comparative restriction maps of the RNA genomes of BYDV-RPV-Vic and -NY. Scale indicates the position of each restriction site in the genome. Vertical lines indicate the position of restriction sites. Restriction sites likely to correspond in each genome are represented on the inside of the lines, others are indicated outside. Introduced restriction sites in the sequence of pC:RPV are indicated in full. Abbreviations: A is *Asp*718; B is *Bam*HI; C is *Cla*I; E is *Eco*RI; P is *Pst*I; S is *Sac*I; X is *Xba*I.


XhoI
XbaI
BamHI
P
A
B
P
C
RPV-Vic
C
S
E
S
E P
E
C
S
E
S
E,P
E
RPV-NY
XC
E
C
C
B
BSB
X
1000
2000
3000
4000
5000

in a minimal number of cycles compared to that possible with total RNA preparations (data not shown). dsRNA was isolated from small amounts of infected tissue (~30 g) in sufficient purity for specific amplification in PCR employing a single sequence-specific primer. A further advantage is that both positive and negative strands of the genomic RNA are available for cDNA synthesis, which can be used to circumvent difficulties with first-strand cDNA synthesis due to varying properties of the template RNA.

The genomic termini of BYDV-RPV-Vic were successfully amplified from dsRNA using a RACE-PCR protocol. Sequencing of clones derived from this work revealed that the published sequence of BYDV-RPV (NY isolate) is likely to be incomplete at both 5' and 3' ends. In particular, a further 68 nt of sequence at the genomic 3' end was found in the work presented here. This is not surprising because the authors of the BYDV-RPV-NY do not claim to have elucidated the entire sequence of BYDV-RPV (Vincent *et al.*, 1991); their strategy employing random primers for cDNA synthesis was unlikely to yield clones covering the entire 3' genomic terminus, because of the low probability of a random primer binding precisely to the end of the genomic RNA. Although the 5' leader sequence defined here for BYDV-RPV-Vic is shorter than that published for BYDV-RPV-NY, it is more likely to contain the full 5' sequence. This is because of the presence of a sequence at the extreme 5' end of the BYDV-RPV-Vic genome that is related to the conserved sequence motif that occurs at the first few nucleotides of the subgroup II luteoviruses (Keese *et al.*, 1990). This sequence is absent at the 5' end of the BYDV-RPV-NY sequence, strengthening the argument that the published 5' terminus does not reflect the true start of the genomic RNA.

Clones derived from the RACE reactions contained cDNA inserts of variable length. The reason for this is unclear, but is unlikely to reflect degradation of the RNA template because dsRNA is highly stable and generally not susceptible to attack by nucleases. While it is possible for short RACE products to be artefactually generated in the PCR, no such molecules were found in determination of the SDV genomic termini (Chapter Three), or in RACE cloning of the 3' genomic end of BYDV-PAV-Vic (data

not shown). It is possible that the variable ends found here for BYDV-RPV are a true feature of the dsRNA population, and may reflect either a replicatory or translational strategy of the virus, or alternatively a deficiency of the viral RdRp in the production of full-length copies of the viral genome.

The sequence motif 5'-ACAAAAG-3' is conserved at the 5' genomic terminus of the subgroup II luteoviruses BWYV, CABYV and PLRV, while related sequences are found at the same position in BYDV-RPV-Vic and SBMV. The presence of the reverse complement of this sequence at the extreme 3' terminus of the BYDV-RPV-Vic genome is unprecedented in the luteoviruses. While it seems unlikely that BYDV-RPV would possess a different strategy for replication from the other subgroup II luteoviruses, the length and positioning in the genome of this conserved sequence suggest that it does not occur purely by chance. If it is active in replication of BYDV-RPV genomic RNA then it must be questioned why a similar sequence is not present in the genomes of the remaining subgroup II luteoviruses. In any case, the sequence motif 5'-GU-3' is present in the final two nucleotides of all sequenced subgroup II luteoviruses other than BYDV-RPV-NY, for which it is argued above that the published sequence is short of full-length.

Extensive use of PCR was made in construction of the full-length BYDV-RPV clone. This was an intrinsic part of the strategy of using dsRNA as the starting template for cDNA construction. A feature of amplification was the PCR of large (≈2 kb) cDNA fragments, which was achieved firstly by the use of dsRNA, and secondly by the high concentration of oligonucleotide primers in the PCR. It is not clear why high primer concentration should favour the formation of long products in the PCR. Subcloning of cDNA sequences mediated by PCR was approached here also with high primer concentrations, and by the use of large amounts of starting template (as cloned circular cDNA) combined with few cycles of amplification, in order to minimise the potential for the introduction of mutations during PCR. Addition of large amounts of nucleic acid to the PCR in the form of primers and template appeared to titrate out Mg^{2+} ions, which was countered here by the addition of $MgSO_4$.

Engineering of restriction sites into the sequence of PCR primers provided a rapid method for construction of the full-length BYDV-RPV cDNA clone. The strategy was successful as measured by correct ordering of the segments in the genome, and the integrity of the sequence at the recombinant junctions. Mutagenesis to create restriction sites in the PCR primers was designed not to disrupt the deduced amino acid sequence of the ORF in which the primer binding site occurred. However, the possibility that the mutations might affect some other aspect of viral function, for example disruption of an unrecognised ORF, or mutation of an unknown regulatory element, can not be excluded. Restriction mapping of the full-length clone verified its correct assembly and provided points for comparison with the published BYDV-RPV-NY sequence. The limited similarity between the genomes of the different isolates in the presence, number and order of restriction sites for the seven enzymes assayed here is unsurprising. Only a limited difference in nucleotide sequence is necessary for creation or disruption of six-base recognition sites; typical variation in nucleotide sequences of luteovirus isolates is in the order of 5-10% (Keese *et al.*, 1990; Vincent *et al.*, 1990). The poor correlation between the restriction maps of the respective BYDV-RPV genomes demonstrates the limited usefulness of knowledge of the BYDV-RPV-NY sequence in construction of a full-length cDNA clone of BYDV-RPV-Vic, and supports the strategy of introducing new restriction sites to the genomic sequence that was used here.

Cloning of full-length BYDV sequences into pBIN19 was straightforward. Both BYDV-RPV and BYDV-PAV sequences were cloned in a two step procedure, the first involving transfer of the viral 5' cDNA with the 35S transcription sequences, and the second to insert the remainder of the viral cDNA. A minor modification to the BYDV-PAV cDNA clone was required to allow single-step cloning of the viral 3' cDNA in a single piece. Correct assembly of the viral sequences was verified in both clones by nucleotide sequencing.

The failure of the agroinfection procedure to establish BYDV infection from cloned cDNAs as described here should be regarded as a preliminary result. The most

important variable, the species and strain of *Agrobacterium* (Marks *et al.*, 1989), requires further investigation. *Agrobacterium rhizogenes* appears to give higher frequencies of agroinfection of wheat dwarf geminivirus in wheat seedlings (Marks *et al.*, *ibid.*), so would be a suitable vector for investigation. Further variables which might be investigated are the procedure for inoculation of the plant, and the identity of the promoter driving transcription of the viral cDNA. The strength of the CaMV 35S promoter is significantly less in monocot than in dicot tissues (Vasil, 1994), although suitable strong promoters from monocot-infecting DNA plant viruses characterised to date appear to exhibit tissue specificity so are unlikely to be appropriate for agroinfection. It is also possible that PCR errors have destroyed the infectivity of the viral cDNAs. This is more likely for the BYDV-RPV than the BYDV-PAV clone, which was assembled with minimal use of PCR and few cycles of amplification where PCR was employed. However, the significant usage of PCR in construction of CMV infectious clones (Chapter Five) failed to abolish infectivity. The success of Leiser *et al.* (1992) in establishing an agroinfection protocol for BWYV suggests that attention to detail will also bring success with the BYDV clones described here.

CHAPTER SEVEN

FINAL DISCUSSION

7.1 Iterative recombination in the evolution of luteovirus genomes

The principle of RNA recombination in the evolution of plant RNA virus genomes is well established. There are two types of RNA recombination (King, 1988); legitimate, where recombination occurs between homologous nucleotide sequences, and illegitimate, which can take place at any position between unrelated RNA sequences. Legitimate recombination can further divided into two types (Lai, 1992); homologous (or symmetrical), occurring at equivalent positions in the parental genomes, or aberrant homologous (assymetrical), occurring at non-equivalent positions in the parental genomes but at regions of local nucleotide homology. Homologous legitimate RNA recombination is thus analogous to homologous DNA (sexual) recombination, whereas aberrant homologous RNA recombination is more similar in effect to illegitimate recombination, in that formation of new genome structures results. Evidence exists for the occurrence of each of these mechanisms in the evolution of the luteoviruses.

The recombination event proposed by Miller *et al.* (1994) to have led to the divergence of the luteovirus genome subtypes is best categorised as aberrant homologous. This is because the exchange of replicase ORFs is postulated to have occurred between dissimilar genomes, albeit at homologous sequences present in subgenomic RNA promoters (reviewed in Chapter One). Illegitimate recombination between chloroplastic and viral genomic RNAs has been observed in the 5' untranslated region of a Scottish isolate of PLRV (Mayo and Jolly, 1991). The recombination event postulated in Chapter Three leading to formation of the SDV genome from subgroup I and II parents appears to fall into the homologous legitimate class. Because this event reproduced that leading to formation of the two subgroup genomes, it is termed here 'iterative'. Iterative RNA recombination between luteovirus genomes may have important biological significance.

The exchange of homologous sequences is essential to the Darwinian evolution of sexually reproducing species. Homologous recombination results in the reassortment of pre-existing variant sequences, thus increasing the range of genetic variation that is subject to selection. Two artificial systems demonstrate the importance

of recombination (coupled with a low level of random mutation) in the evolution of linear sequences. Firstly, evolution may be simulated by computer programs termed genetic algorithms (Holland, 1992; Forrest, 1993). In genetic algorithms, selection operates on strings of binary digits representing individual characteristics. The strings can be made to evolve over time by allowing 'mutation' in the binary code coupled with recombination between varying strings. Repeated cycles of evolution coupled with selection result in strings with optimal combinations of characters. In this way, complex structures arise with functional applications, for example strategies for playing games (Forrest, 1993). However, genetic algorithms relying on random mutation in the absence of recombination fail to evolve functional strings (Holland, 1992).

The second demonstration of the importance of recombination to sequence evolution involves *in vitro* DNA amplification. Conventional random mutagenesis of PCR products employs conditions favouring low fidelity replication known as error-prone PCR (Caldwell and Joyce, 1992). Incorporation of random *in vitro* homologous recombination to error-prone PCR substantially advances functional evolution over that obtained by error-prone PCR alone (Stemmer, 1994a, 1994b). For example, a bacterial antibiotic resistance gene mutated by recombinant error-prone PCR was 32,000-fold more effective than the original sequence, whereas that mutated by error-prone PCR in the absence of recombination resulted in only a 16-fold increase in effectiveness.

Although RNA recombination has a recognised role in plant virus genome evolution, it is a sporadic event predominately resulting in the transfer of novel genes (Koonin and Dolja, 1993). Continuing evolution of RNA genomes relies instead on the high mutation rate intrinsic to RdRps (Steinhauer *et al.*, 1992). Routine homologous recombination, as occurs in sexually reproducing organisms, is not known to occur. However, organisation of the luteovirus genome into functional modules as argued in Chapter Four increases the likelihood of productive (iterative) RNA recombination between the modules. Such iterative RNA recombination provides a mechanism similar to sexual recombination in that greater usage of existing nucleotide variation

can be made. The divergence of the various luteovirus strains provides a large pool of genetic diversity for exploitation through recombination.

The importance of iterative recombination in the evolution of the luteoviruses can be measured to some extent by documentation of its occurrence. To date, SDV represents the only known example resulting from recombination of existing luteovirus strains. However, two points are relevant: Firstly, further characterisation of luteovirus sequences may uncover genomes derived from independent iterative recombination events. Secondly, the structure of known luteovirus genomes may have resulted from iterative recombination. Such events occurring early in the evolution of the virus group may be difficult to trace, especially where members of the same luteovirus subgroup are involved, or where the sequence of parent genomes is not known. The success of luteoviruses as agriculturally important pathogens could possibly be explained in part by invocation of iterative recombination, allowing full exploitation of nucleotide variation between luteovirus strains.

7.2 Viral ORFs conditioning interaction with resistance genes

The aim of this thesis was to identify the ORF of BYDV-PAV responsible for interaction with the Yd2 resistance gene of barley. This aim was not achieved. However, other workers have exploited resistance-breaking isolates of plant RNA viruses to elucidate the viral ORF that specifies the virus-resistance gene interaction. This work has followed a general strategy. The nucleotide sequence of mutant virus isolates able to overcome plant resistance genes, generated either naturally or artificially, have been determined and compared to that of the parent virus which is susceptible to the resistance. Nucleotide differences between the RNA genomes of mutant and parent isolates are determined, and changes reintroduced to the parent virus by *in vitro* mutagenesis. This allows verification that the observed mutation is in fact responsible for the change in interaction with the resistance gene. Viral ORFs conditioning interaction with the Tm-1, Tm-2, and Tm-2^2 of tomato, the *N* and *N′* genes

Table 7.1. Genes for resistance to plant RNA viruses , and viral ORFs mediating the resistance interaction

R gene	Virus	Host	Active in protoplasts?	Viral ORF	Ref.
Tm-1	TMV	tomato	yes	replicase*	a,b,c
Tm-2	TMV	tomato	no	MP	d,e
$Tm\text{-}2^2$	TMV	tomato	no	MP	e,f
N	TMV	tobacco	no	replicase†	g
N'	TMV	tobacco	ND	CP	h
Nx	PVX	potato	?‡	CP	i,j
Rx	PVX	*Solanum* spp.§	yes	CP	i

*180 K and 130 K ORFs; †130 K ORF; ‡conflicting reports in literature; §*S. andigena* and *S. acaule.*

[a]Watanabe *et al.*, 1987; [b]Meshi *et al.*, 1988; [c]Yamafuji *et al.*, 1991; [d]Meshi *et al.*, 1989; [e]Calder and Palukaitis, 1992; [f]Weber *et al.*, 1993; [g]Padgett and Beachy, 1993; [h]Saito *et al.*, 1987; Knorr and Dawson, 1988; [i]Kavanagh *et al.*, 1992; [j]Santa Cruz and Baulcombe, 1993.

Abbreviations: TMV - tobacco mosaic tobamovirus; MP - movement protein (30 K); ND - not determined; CP - viral coat protein; PVX - potato virus X potexvirus.

of tobacco, as well as the *Nx* gene of potato and the *Rx* gene of some *Solanum* species have been determined in this way (Table 7.1).

The data reveal the following. The large proportion of resistance genes investigated in this manner are active against tobacco mosaic tobamovirus (either tobacco or tomato strains). Additionally, the resistance interaction is mediated by a variety of viral ORFs, including presumed replicase components, viral cell-cell movement protein and coat proteins. There does not appear to be any relationship between the viral ORF targeted by the resistance gene and the activity of the gene in plant protoplasts.

The mode of action of the respective virus resistance genes is not clear from these data. It is unknown if the product of the resistance gene interacts directly with the product of the viral ORF, or if the viral ORF product acts as an elicitor of some other resistance mechanism. Evidence exists to suggest that resistance requires recognition of the relevant viral protein by a host factor. Firstly, the net local charge of the replicase proteins is altered in a TMV strain that overcomes the Tm-1 resistance gene of tomato (Meshi *et al.*, 1988). The replicase proteins are not degraded in plant protoplasts homozygous for the resistance gene (Yamafuji *et al.*, 1991). Taken together, these results suggest that the viral replicase proteins interact electrostatically with a host factor prior to the induction of resistance, with resistance not necessarily involving destruction of the viral factor. Similarly, strains of TMV able to overcome the allelic Tm-2 and Tm-2^2 resistance genes of tomato show changes in net charge of their respective movement proteins (MP; Meshi *et al.*, 1989; Calder and Paulkaitis, 1992; Weber *et al.*, 1993). A mutant TMV strain sensitive to the *N* gene of tobacco (derived by random mutagenesis from a resistance-breaking strain) did not contain a net charge change in a putative replicase protein, however the mutation responsible for resistance sensitivity involved a proline-to-leucine substitution (Padgett and Beachy, 1993). This is likely to alter the structural conformation of the protein, which in turn could promote interaction of the protein with a host component.

Thus induction of the resistance phenotype appears to be mediated by a protein-protein recognition event involving a host and a viral factor. Whether the data for TMV and PVX resistances can be extrapolated to the interaction between BYDV-PAV and the Yd2 gene is not clear. Further investigation of the resistance response, including cloning of viral resistance genes (7.3), is necessary before generalisations or predictions can be made.

7.3 Plant genes specifying resistance to disease

Plant disease resistance genes control recognition of invading pathogens and subsequent activation of plant defences (Keen, 1992). Each resistance gene acts in a highly specific manner, recognising only particular strains of viral, bacterial, fungal or nematode pathogens. Flor (1947) formulated the gene-for gene hypothesis to describe specific resistance in plant-pathogen interactions, in which resistance in the plant is dependent on recognition of a specific avirulence factor in the pathogen by a specific resistance gene in the plant. Three such genes fulfilling the criteria of the gene-for-gene hypothesis conditioning resistance to bacterial and viral pathogens have been cloned. A summary of the genes and their putative biochemical properties is presented in Table 7.2.

Little is known of the cellular function(s) of resistance genes cloned to date, although predictions can be made from the presence of certain motifs in the deduced amino acid sequences. Thus serine/threonine protein kinase motifs in the sequence of *PTO* (Martin *et al.*, 1993), and ATP/GTP binding motifs in the sequences of *RPS2* and *N* (Bent *et al.*, 1994; Mindrinos *et al.*, 1994; Whitham *et al.*, 1994), suggest roles in signal transduction for each of these proteins. This is consistent with the role of each gene in the induction of the hypersensitive response (HR), the formation of necrotic lesions and antimicrobial products around the foci of infection. The relationship between the deduced amino acid sequences of each gene is surprising. *PTO* and *RPS2*, which both specify resistance to the bacterial pathogen Pseudomonas syringae, do not share significant homology (Martin *et al.*, 1993; Mindrinos *et al.*, 1994). Conversely,

Table 7.2. Properties of cloned plant disease resistance genes

R gene	plant	pathogen	M_r	Putative functional domains	Ref.
PTO	tomato	*Pseudomonas syringae* pv. *tomato*	35 K	- serine/threonine protein kinase	a
RPS2	*Arabidopsis thaliana*	*Pseudomonas syringae*	105 K	- leucine zipper - ATP/GTP binding - membrane spanning - LRR (receptor)	b,c
N	tobacco	TMV	131	- leucine zipper - ATP/GTP binding - LRR (receptor)	d

[a]Martin *et al.*, 1993; [b]Bent *et al.*, 1994; [c]Mindrinos *et al.*, 1994; [d]Whitham *et al.*, 1994.

RPS2, which confers resistance to a bacterial pathogen in Arabidopsis, and *N*, which confers resistance to a viral pathogen in tobacco, show significant amino acid sequence homology and conservation of putative amino acid motifs (Table 7.2; Bent *et al.*, 1994; Mindrinos *et al.*, 1994; Whitham *et al.*, 1994). The major (tentative) difference between the proteins encoded by these genes is that *RPS2* is possibly membrane bound, whereas *N* appears to be cytoplasmic. Of the three genes, only *RPS2* and *N* appear to encode receptor domains, which is surprising given the historical view that resistance genes are likely to encode receptors for pathogenic elicitors (Keen, 1992).

The significance of these results to the function of the Yd2 gene of barley is not clear. The fact that each of the resistance genes cloned thus far appears to play a role in signal transduction suggests that each is a member of similar pathways. However, Yd2 does not appear to induce HR unlike the resistance genes cloned to date. Therefore it is not certain that the resistance encoded by Yd2 will operate via the same general mechanism. Of the viral resistance genes discussed in section 7.2, only *N*, *N'* and *Nx* induce HR. Cloning of non-HR resistance genes is therefore necessary to establish whether more than one pathogen resistance mechanism has evolved in higher plants.

7.4 Future work

Full-length cDNA clones of BYDV-RPV and BYDV-PAV assembled under the transcriptional control of CaMV 35S sequences in the binary vector pBIN19 failed to establish infection when used here in agroinfection experiments. Therefore, conditions for the use of these clones in agroinfection must be established. Three variables likely to control successful agroinfection identified in Chapter Six are the strength of the CaMV 35S promoter in monocot cells, the correct delivery of the *Agrobacterium* to the vegetative meristem of the seedling, and the efficiency of different strains of *Agrobacterium* in delivering T-DNA to the plant nucleus. Difficulties concerning delivery of the T-DNA to the plant (*Agrobacterium* strain, correct injection of bacterial suspensions) could be circumvented by transfection of plant protoplasts with the binary clones, allowing investigation of the intrinsic infectivity of the constructs. Transfection

of protoplasts with the clones will establish whether the CaMV 35S promoter is strong enough to allow infection with the cloned BYDV cDNAs in monocot cells. *In vitro* transcripts derived from the full-length BYDV-PAV cDNA clone pBYDV19 (Chapter Six) would serve as a suitable positive control in such experiments.

If it is not possible to establish infection with the BYDV binary vector clones in protoplasts, then a suitable promoter for this purpose must be found. This may be difficult given the requirements for expression without tissue specificity, the ability to control the sequence identity at the first nucleotide of transcription, and strong transcription in monocot cells. Alternatively, if protoplast infection with the BYDV binary vector clones is possible, then strains of *A. tumefaciens* and *A. rhizogenes* must be screened for their effectiveness in establishing infection with the clones in intact plants. Successful agroinfection with the clones will allow the module-swapping experiments described in Chapter Four to proceed.

An alternative to the agroinfection/module-swapping approach described above is to use random mutagenesis to create a BYDV-PAV isolate not sensitive to the Yd2 gene. This could follow a similar protocol to that of Padgett and Beachy (1993) who used hydroxylamine to mutate a resistance-breaking TMV cDNA clone to become sensitive to resistance conferred by the *N* gene in tobacco. The large numbers of variant sequences required by this approach would be better suited to protoplast infection than agroinfection, although it would be necessary to screen the mutants in intact plants as Yd2 does not appear to operate in protoplasts (Larkin *et al.*, 1991). Success in isolation of a Yd2 insensitive mutant of BYDV-PAV and determination of its genomic nucleotide sequence would allow discovery of the mutation(s) responsible for the change in Yd2 sensitivity.

Additional work relevant to this thesis is to further characterise the resistance conferred by Yd2. This would involve measurement of rates of viral replication and spread in BYDV-PAV resistant and susceptible barley lines near-isogenic for the Yd2 gene. Quantitative molecular techniques including northern dot-blot and RNase protections would be suitable for this type of investigation. Furthermore, *in situ*

hybridisation using specific nucleic acid and antibody probes for viral products at the electron- or light microscope level might indicate the point in the viral life cycle at which Yd2 acts.

Finally, the specificity of the resistance mechanism induced by Yd2 could be investigated. Dual infections of BYDV-PAV and BYDV-RPV could be used to determine if BYDV-PAV is able to induce Yd2 action against BYDV-RPV. This would indicate if the antiviral mechanism induced by Yd2 is specific for BYDV-PAV, or if specificity acts only in recognition of the viral elicitor. The degree of specificity may indicate the complexity of the resistance pathway, *i.e.* if there is a functional separation between elicitation of resistance and the antiviral resistance activity.

LITERATURE CITED

Aapola, A.I.E., and Rochow, W.F. (1971). Relationships among three isolates of barley yellow dwarf virus. *Virology* **46**, 127-141.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403-410.

Ashby, J.W., and Johnstone, G.R. (1985). Legume luteovirus taxonomy and current research. *Aust. Plant Path.* **14**, 2-7.

Bahner, I., Lamb, J., Mayo, M.A., and Hay, R.T. (1990). Expression of the genome of potato leafroll virus: readthrough of the coat protein termination codon *in vivo*. *J. gen. Virol.* **71**, 2251-2256.

Baldi, G., Moletti, M., and Osler, R. (1990). Inheritance of resistance to giallume in rice. *In* Burnett, P.A. (ed.) *World Perspectives on Barley Yellow Dwarf*. pp. 429-433. CIMMYT, Mexico, D.F., Mexico.

Baldi, G., Moletti, M., and Callegarin, A. (1991). Selection for "giallume" resistance in rice *(Oryza sativa* L.). *J. Genet. Breed.* **45**, 107-110.

Baltenberger, D.E., Ohm, H.W., and Foster, J.E. (1987). Reactions of oat, barley, and wheat to infection with barley yellow dwarf virus isolates. *Crop Science* **27**, 195-198.

Banks, P.M., Waterhouse, P.M., and Larkin, P.J. (1992). Pathogenicity of three RPV isolates of barley yellow dwarf virus on barley, wheat and wheat alien addition lines. *Ann. appl. Biol.* **121**, 305-314.

Barker, H. (1987). Invasion of non-phloem tissue in *Nicotiana clevelandii* by potato leafroll luteovirus is enhanced in plants also infected with potato Y potyvirus. *J. gen. Virol.* **68**, 1223-1227.

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J., Staskawicz, B.J. (1994). *RPS2* of *Arabidopsis thaliana:* A leucine-rich repeat class of plant disease resistance genes. *Science* **265**, 1856-1860.

Bevan, M. (1984). Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* **12**, 8711-8721.

Boyer, J.-C., and Haenni, A.-L. (1994). Infectious transcripts and cDNA clones of RNA viruses. *Virology* **198**, 415-426.

Brault, V., and Miller, W.A. (1992). Translational frameshifting mediated by a viral sequence in plant cells. *Proc. Natl. Acad. Sci. USA* **89**, 2262-2266.

Brettell, R.I.S., Banks, P.M., Cauderon, Y., Chen, X., Cheng, Z.M., Larkin, P.J., and Waterhouse, P.M. (1988). A single wheatgrass chromosome reduces the concentration of barley yellow dwarf virus in wheat. *Ann. appl. Biol.* **113**, 599-603.

Burnett, P.A. (1984). Preface. *In* Burnett, P.A. (ed.) *Barley Yellow Dwarf, A Proceedings of the Workshop*. pp. 6-13. CIMMYT 1984.

Calder, V.L., and Palukaitis, P. (1992). Nucleotide sequence analysis of the movement genes of resistance breaking strains of tomato mosaic virus. *J. gen. Virol.* **73**, 165-168.

Caldwell, R.C., and Joyce, G.F. (1992). Randomisation of genes by PCR mutagenesis. *PCR Methods Applic.* **2**, 28-33.

Carrington, J.C., Hillman, B.I., and Morris, T.J. (1989). The genome structure of turnip crinkle virus. *Virology* **170**, 219-226.

Catherall, P.L., Jones, A.T., and Hayes, J.D. (1970). Inheritance and effectiveness of genes in barley that condition tolerance to barley yellow dwarf virus. *Ann. appl. Biol.* **65**, 153-161.

Chalhoub, B.A., Kelly, L., Robaglia, C., and Lapierre, H.D. (1994). Sequence variability in the genome-3'-terminal region for 10 geographically distinct PAV-like isolates of barley yellow dwarf virus: analysis of the ORF6 variation. *Arch. Virol.* (in press).

Cheng, S.-L., Domier, L.L., and D'Arcy, C.J. (1994a). Detection of the readthrough protein of barley yellow dwarf virus. *Virology* **202**, 1003-1006.

Cheng, Z.M., Keese, P., Waterhouse, P., He, X.Y., Zhou, G.G. (1994b). Nucleotide sequence of the barley yellow dwarf luteovirus GPV isolate coat protein gene. *Barley Yellow Dwarf Newsletter No. 5, June 1994*, p 31.

Chin, L.-S., Foster, J.L., and Falk, B.W. (1993). The beet western yellows virus ST9-associated RNA shares structural and nucleotide sequence homology with carmo-like viruses. *Virology* **192**, 473-482.

Citovsky, V., and Zambryski, P. (1991). How do plant virus nucleic acids move through intercellular connections? *BioEssays* **13**, 373-379.

Citovsky, V., McLean, G., Aupan, J.R., and Zambryski, P. (1993). Phosphorylation of tobacco mosaic virus cell-to-cell movement protein by a developmentally regulated plant cell wall-associated protein kinase. *Genes and Dev.* **7**, 904-910.

Comeau, A., and Jedlinski, H. (1990). Successful breeding for barley yellow dwarf resistance or tolerance: a systematic approach related to other agronomic characteristics. *In* Burnett, P.A. (ed.) *World Perspectives on Barley Yellow Dwarf*. pp. 441-451. CIMMYT, Mexico, D.F., Mexico.

Conti, M., D'Arcy, C.J., Jedlinski, H., and Burnett, P.A. (1990). The "yellow plague" of cereals, barley yellow dwarf virus. *In* Burnett, P.A. (ed.) *World Perspectives on Barley Yellow Dwarf*. pp. 1-6. CIMMYT, Mexico, D.F., Mexico.

D'Arcy, C.J. (1986). Current problems in the taxonomy of luteoviruses. *Micro. Sciences* **3**, 309-313.

D'Arcy, C.J., Torrance, L., and Martin, R.R. (1989). Discrimination among luteoviruses and their strains by monoclonal antibodies and identification of common epitopes. *Phytopathology* **79**, 869-873.

Damsteegt, V.D., and Bruehl, G.W. (1964). Inheritance of resistance in barley to barley yellow dwarf. *Phytopathology* **54**, 219-224.

Damsteegt, V.D., Hewings, A.D., and Sindermann, A.B. (1990). Soybean dwarf virus: experimental host range, soybean germ plasm reactions, and assessment of potential threat to U.S. soybean production. *Plant Dis.* **74**, 992-995.

Davies, C., and Symons, R.H. (1988). Further implications for the evolutionary relationships between tripartite plant viruses based on cucumber mosaic virus RNA 3. *Virology* **165**, 216-224.

Demler, S.A., and de Zoeten, G.A. (1991). The nucleotide sequence and luteovirus-like nature of RNA 1 of an aphid non-transmissible strain of pea enation mosaic virus. *J. gen. Virol.* **72**, 1819-1834.

Demler, S.A., Rucker, D.G., and de Zoeten, G.A. (1993). The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. *J. gen. Virol.* **74**, 1-14.

Demler, S.A., Borkhsenious, O.N., Rucker, D.G., and de Zoeten, G.A. (1994). Assessment of the autonomy of replicative and structural functions encoded by the luteo-phase of pea enation mosaic virus. *J. gen. Virol.* **75**, 997-1007.

Devereux, J., Haerbili, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**, 387-395.

Di, R., Dinesh-Kumar, S.P., and Miller, W.A. (1993). Translational frameshifting by barley yellow dwarf virus (PAV serotype) in *Escherichia coli* and in eukaryotic cell-free extracts. *Mol. Plant Microbe Interac.* **6**, 444-452.

Dinesh-Kumar, S.P., and Miller, W.A. (1993). Control of start codon choice on a plant viral RNA encoding overlapping genes. *Plant Cell* **5**, 679-692.

Dinesh-Kumar, S.P., Brault, V., and Miller, W.A. (1992). Precise mapping and *in vitro* translation of a trifunctional subgenomic RNA of barley yellow dwarf virus. *Virology* **187**, 711-722.

Ding, S.-W., Anderson, B.J., Hasse, H.R., and Symons, R.H. (1994). New overlapping gene encoded by the cucumber mosaic virus genome. *Virology* **198**, 593-601.

Domier, L.L., Shaw, J.G., and Rhoads, R.E. (1987). Potyviral proteins share amino acid sequence homology with picorna-, como- and caulimoviral proteins. *Virology* **158**, 20-27.

Domier, L.L., Lukasheva, L.I., and D'Arcy, C.J. (1994). Coat protein sequence of RMV-like strains of barley yellow dwarf virus separate them from other luteoviruses. *Intervirology* **37**, 2-5.

Domingo, E. (1992). Genetic variation and quasi-species. *Curr. Op. Genet. Dev.* **2**, 61-63.

Domingo, E., and Holland, J.J. (1988). High error rates, population equilibrium, and evolution of RNA replication systems. *In* Domingo, E., Holland, J.J., and Ahlquist, P. (eds) *RNA genetics, vol 3*. pp. 3-36. CRC, Boca Raton.

Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**, 735-744.

Dunsmuir, P., Bond, D., Lee, K., Gidoni, D., and Townsend, J. (1988). Stability of introduced genes and stability in expression. *In* Gelvin, S.B., Schilperoot, R.A., and Verma, D.P.S. (eds) *Plant Molecular Biology Manual*. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Esau, K. (1957). Phloem degeneration in graminae affected by the barley yellow-dwarf virus. *Amer. J. Bot.* **44**, 245-251.

Esau, K., and Hoefert, L.L. (1972). Development of infection with beet western yellows virus in the sugarbeet. *Virology* **48**, 724-738.

Eweida, M., Oxelfelt, P., and Tomenius, K. (1988). Concentration of virus and ultrastructural changes in oats at various stages of barley yellow dwarf virus infection. *Ann. appl. Biol.* **112**, 313-321.

Falk B.W., and Duffus, J.E. (1984). Identification of small single- and double-stranded RNAs associated with severe symptoms in beet western yellows virus-infected *Capsella bursa-pastoris*. *Phytopathology* **74**, 1224-1229.

Filichkin, S.A., Lister, R.M., McGrath, P.F., and Young, M.J. (1994). *In vivo* expression and mutational analysis of the barley yellow dwarf virus readthrough gene. *Virology* **205**, 290-299.

Flor, H.H. (1947). Inheritance of reaction to rust in flax. *J. Agr. Res.* **74**, 241-262.

Forrest, S. (1993). Genetic algorithms: principles of natural selection applied to computation. *Science* **261**, 872-878.

Francki, R.I.B., Randles, J.W., Chambers, T.C., and Wilson, S.B. (1966). Some properties of purified cucumber mosaic virus (Q strain). *Virology* **28**, 729-741.

Fraser, R.S.S. (1987). Genetics of plant resistance to viruses. *In* Evered, D., and Harnett, S. (eds) *Plant Resistance to Viruses*. pp. 6-22. Ciba Foundation Symposium 133. John Wiley and Sons, Chichester, UK.

Frohman, M.A. (1990). RACE: rapid amplification of cDNA ends. *In* Innes, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. (eds) *PCR Protocols*. pp. 28-38. Academic Press, San Diego.

Fujiwara, T., Giesman-Cookmeyer, D., Ding, B., Lommel, S.A., and Lucas, W.J. (1993). Cell-to-cell trafficking of macromolecules through plasmodesmata potentiated by the red clover necrotic mosaic virus movement protein. *Plant Cell* **5**, 1783-1794.

Garcia, A., van Duin, J., and Pleij, C.W.A. (1993). Differential response to frameshift signals in eukaryotic and prokaryotic translational systems. *Nucleic Acids Res.* **21**, 401-406.

Gibbs, A. (1987). Molecular evolution of viruses: 'trees', 'clocks' and 'modules'. *J. Cell Sci. Suppl.* **7**, 319-337.

Gildow, F.E. (1985). Transcellular transport of barley yellow dwarf virus into the hoemocoel of the aphid vector, *Rhopalosiphum padi*. *Phytopathology* **75**, 292-297.

Gildow, F.E. (1990). Barley yellow dwarf virus-aphid vector interactions associated with virus transmission and vector specificity. *In* Burnett, P.A. (ed.) *World Perspectives on Barley Yellow Dwarf*. pp. 111-122. CIMMYT, Mexico, D.F., Mexico.

Gildow, F.E. (1993). Evidence for receptor-mediated endocytosis regulating luteovirus acquisition by aphids. *Phytopathology* **83**, 270-277.

Gildow, F.E., and Rochow, W.F. (1980). Role of accessory salivary glands in aphid transmission of barley yellow dwarf virus. *Virology* **104**, 97-108.

Gildow, F.E., Ballinger, M.E., and Rochow, W.F. (1983). Identification of double-stranded RNAs associated with barley yellow dwarf virus infection of oats. *Phytopathology* **73**, 1570-1572.

Gill, C.C. (1967). Transmission of barley yellow dwarf virus isolates from Manitoba by five species of aphids. *Phytopathology* **57**, 713-718.

Gill, C.C., and Chong, J. (1975). Development of the infection in oat leaves inoculated with barley yellow dwarf virus. *Virology* **66**, 440-453.

Gill, C.C., and Chong, J. (1976). Differences in cellular ultrastructural alterations between variants of barley yellow dwarf virus. *Virology* **75**, 33-47.

Gill, C.C., and Chong, J. (1979a). Cytological alterations in cells infected with corn leaf aphid-specific isolates of barley yellow dwarf virus. *Phytopathology* **69**, 363-368.

Gill, C.C., and Chong, J. (1979b). Cytopathological evidence for the division of barley yellow dwarf virus into two subgroups. *Virology* **95**, 59-69.

Gill, C.C., and Chong, J. (1981). Vascular cell alterations and predisposed xylem infection in oats by inoculation with paired barley yellow dwarf viruses. *Virology* **114**, 405-413.

Gill, C.C., and Comeau, A. (1977). Synergism in cereals between corn leaf aphid-specific and aphid-nonspecific isolates of barley yellow dwarf virus. *Phytopathology* **67**, 1388-1392.

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

Gorbalenya, A.E., and Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure-function relationships. *Curr. Op. Struct. Biol.* **3**, 419-429.

Grimsley, N. (1990). Agroinfection. *Physiol. Plant.* **79**, 147-153.

Grimsley, N., and Bisaro, D. (1987). Agroinfection. *In* Hohn, T., and Schell, J. (eds) *Plant Gene Research: Plant DNA Infectious Agents*. pp. 87-107. Springer-Verlag, New York.

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

Gubler, U., and Hoffman, B.J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene* **25**, 263-239.

Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K., and Morris, T. (1985). Nucleotide sequence and genome organisation of carnation mottle virus RNA. *Nucleic Acids Res.* **13**, 6663-6677.

Guilley, H., Wipf-Scheibel, C., Richards, K., Lecoq, H., and Jonard, G. (1994). Nucleotide sequence of cucurbit aphid-borne yellows luteovirus. *Virology* **202**, 1012-1017.

Habili, N., and Symons, R.H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res.* **17**, 9543-9555.

Halbert, S.E., Connelly, B.J., Bishop, G.W., and Blackmer, J.L. (1992). Transmission of barley yellow dwarf virus by field collected aphids (Homoptera: Aphididae) and their relative importance in barley yellow dwarf epidemiology in southwestern Idaho. *Ann. appl. Biol.* **121**, 105-121.

Halstead, B.E., and Gill, C.C. (1971). Effect of inoculation of oats with paired combinations of barley yellow dwarf virus isolates. *Can. J. Bot.* **49**, 577-581.

Hammond, J., Lister, R.M., and Foster, J.E. (1983). Purification, identity and some properties of an isolate of barley yellow dwarf virus from Indiana. *J. gen. Virol.* **64**, 667-676.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**, 557.

Harris, K.F., and Bath, J.E. (1972). The fate of pea enation mosaic virus in its pea aphid vector, *Acyrthosiphon pisum* (Harris). *Virology* **50**, 778-790.

Harris, K.F., Bath, J.E., Thottappilly, G., and Hooper, G.R. (1975). Fate of pea enation mosaic virus in PEMV-injected pea aphids. *Virology* **65**, 148-162.

Helms, K., Waterhouse, P.M., and Carver, M. (1983). *Aulacorthum (Neomyzus) circumflexum*, a vector of subterranean clover red leaf virus. *Aust. Plant Path.* **12**, 66-67.

Henry, M., Francki, R.I.B., and Wallwork, H. (1992). Occurrence of barley yellow dwarf virus in cereals and grasses of the low-rainfall wheatbelt of South Australia. *Plant Pathology* **41**, 713-721.

Herrera, G., and Plumb, R. (1989). The response of winter and spring barleys, with and without the Yd2 gene, to 3 British isolates of BYDV. *Barley Yellow Dwarf Newsletter No. 2, April 1989,* pp. 38-40.

Higuchi, R., Krummel, B., and Saiki, R.K. (1988). A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res.* **16**, 7351-7367.

Holland, J.H. (1992). Genetic algorithms. *Sci. Am.* **267**, 44-50.

Holland, J.J., De La Torre, J.C., and Steinhauer, D.A. (1992). RNA virus populations as quasispecies. *Curr. Topics Micro. Immuno.* **176**, 1-20.

Irwin, M.E., and Thresh, J.M. (1990). Epidemiology of barley yellow dwarf: a study in ecological complexity. *Ann. Rev. Phytopathol.* **28**, 393-424.

Jedlinski, H., and Brown, C.M. (1965). Cross protection and mutual exclusion by three strains of barley yellow dwarf virus in *Avena sativa* L. *Virology* **26**, 613-621.

Jensen, S.G. (1969). Occurrence of virus particles in the phloem tissue of BYDV-infected barley. *Virology* **38**, 83-91.

Johnson, R.A., and Rochow, W.F. (1972). An isolate of barley yellow dwarf virus transmitted specifically by *Schizaphis graminum*. *Phytopathology* **62**, 921-925.

Johnstone, G.R., Sward, R.J., Farrell, J.A., Greber, R.S., Guy, P.L., McEwan, J.M., and Waterhouse, P.M. (1990). Epidemiology and control of barley yellow dwarf in Australia and New Zealand. *In* Burnett, P.A. (ed.) *World Perspectives on Barley Yellow Dwarf*. pp. 228-239. CIMMYT, Mexico, D.F., Mexico.

Jones, A.T., and Catherall, P.L. (1970a). The relationship between growth rate and the expression of tolerance to barley yellow dwarf virus in barley. *Ann. appl. Biol.* **65**, 137-145.

Jones, A.T., and Catherall, P.L. (1970b). The effect of different virus isolates on the expression of tolerance to barley yellow dwarf virus in barley. *Ann. appl. Biol.* **65**, 147-152.

Kamer, G., and Argos, P. (1984). Primary structural comparison of RNA-dependent RNA polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12**, 7269-7282.

Kashiwazaki, S., Minobe, Y., Omura, T., and Hibino, H. (1990). Nucleotide sequence of barley yellow mosaic virus RNA 1: a close evolutionary relationship with potyviruses. *J. gen. Virol.* **71**, 2781-2790.

Kashiwazaki, S., Minobe, Y., and Hibino, H. (1991). Nucleotide sequence of barley yellow mosaic virus RNA 2. *J. gen. Virol.* **72**, 995-999.

Kavanagh, T., Goulden, M., Santa Cruz, S., Chapman, S., Barker, I., and Baulcombe, D. (1992). Molecular analysis of a resistance-breaking strain of potato virus X. *Virology* **189**, 609-617.

Kawchuk, L.M., Martin, R.R., Rochon, D.M., and McPherson, J. (1989). Identification and characterisation of the potato leafroll virus putative coat protein gene. *J. gen. Virol.* **70**, 783-88.

Keen, N.T. (1992). The molecular biology of disease resistance. *Plant Mol. Biol.* **19**, 109-22.

Keese, P.K., and Gibbs, A. (1992). Origin of genes: "big bang" or continuous creation? *Proc. Natl. Acad. Sci. USA* **89**, 9489-9493.

Keese. P.K., Martin, R.R., Kawchuk, L.M., Waterhouse, P.M., and Gerlach. W.L. (1990). Nucleotide sequences of an Australian and a Canadian isolate of potato leafroll luteovirus and their relationships with two European isolates. *J. gen. Virol.* **71**, 719-724.

Kelly, L., Gerlach, W.L., and Waterhouse, P.M. (1994). Characterisation of the subgenomic RNAs of an Australian isolate of barley yellow dwarf luteovirus. *Virology* **202**, 565-573.

King, A.M.Q. (1988). Recombination in positive strand RNA viruses. *In* Domingo, E., Holland, J.J., and Ahlquist, P. (eds) *RNA genetics.* pp. 149-165, CRC Press, Inc., Boca Raton.

Knorr, D.A., and Dawson, W.O. (1988). A point mutation in the tobacco mosaic virus capsid protein gene induces hypersensitivity in *Nicotiana sylvestris*. *Proc. Natl. Acad. Sci. USA* **85**, 170-174.

Koonin, E.V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *J. gen. Virol.* **72**, 2197-2206.

Koonin, E.V., and Dolja, V.V. (1993). Evolution and taxonomy of positive-stranded RNA viruses: implications of comparative analysis of amino acid sequences. *CRC Crit. Rev. Biochem. Molec. Biol.* **28**, 375-430.

Kuhn, R.J., and Wimmer, E. (1987). The replication of picornaviruses. *In* Rowlands, D.J., Mayo, M.A., and Mahy, B.W.J. (eds) *The molecular biology of the positive strand RNA viruses*. Academic Press, Inc., London.

Kujawa, A.B., Drugeon, G., Hulanicka, D., and Haenni, A.-L. (1993). Structural requirements for efficient translational frameshifting in the synthesis of the putative viral RNA-dependent RNA polymerase of potato leafroll virus. *Nucleic Acids Res.* **21**, 2165-2171.

Lai, M.M.C. (1992). RNA recombination in animal and plant viruses. *Micro. Reviews* **56**, 61-79.

Larkin, P.J., Young, M.J., Gerlach, W.L. and Waterhouse, P.M. (1991). The Yd2 resistance to barley yellow dwarf virus is effective in barley plants but not in their leaf protoplasts. *Ann. appl. Biol.* **118**, 115-125.

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

Leiser, R.-M., Ziegler-Graff, V., Reutenauer, A., Herrbach, E., Lemaire, O., Guilley, H., Richards, K., and Jonard, G. (1992). Agroinfection as an alternative to insects for infecting plants with beet western yellow luteovirus. *Proc. Natl. Acad. Sci. USA* **89**, 9136-9140.

Lister, R.M., and Rochow, W.F. (1979). Detection of barley yellow dwarf virus by enzyme-linked immunosorbent assay. *Phytopathology* **69**, 649-654.

Lommel, S.A., Weston-Fina, M., Xiong, Z., and Lomonossoff, G.P. (1988). The nucleotide sequence and gene organisation of red clover necrotic mosaic virus RNA-2. *Nucleic Acids Res.* **16**, 8587-8602.

MacFarlane, S.A., Gilmer, D., and Davies, J.W. (1992). Efficient inoculation with CaMV 35 S promoter-driven DNA clones of the tobravirus PEBV. *Virology* **187**, 829-831.

Maes, M., and Messens, E. (1992). Phenol as grinding material in RNA preparations. *Nucleic Acids Res.* **20**, 4374.

Mang, K., Ghosh, A., and Kaesberg, P. (1982). A comparative study of the cowpea and bean strains of southern bean mosaic virus. *Virology* **116**, 264-274.

Marks, M.S., Kemp, J.M., Woolston, C.J., and Dale, P.J. (1989). Agroinfection of wheat: a comparison of *Agrobacterium* strains. *Plant Science* **63**, 247-256.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. (1993). Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432-1436.

Martin, R.R., Keese, P.K., Young, M.J., Waterhouse, P.M., and Gerlach, W.L. (1990). Evolution and molecular biology of luteoviruses. *Annu. Rev. Phytopathol.* **28**, 341-63.

Mathre, D.E. (1982). Compendium of barley diseases (ed.). APS Press, Minnesota.

Mayo, M.A., and Jolly, C.A. (1991). The 5' terminal sequence of potato leafroll virus RNA: evidence of recombination between virus and host RNA. *J. gen. Virol.* **72**, 2633-2638.

Mayo, M.A., Barker, H., Robinson, D.J., Tamada, T., and Harrison, B.D. (1982). Evidence that potato leafroll virus RNA is positive-stranded, is linked to a small protein and does not contain polyadenylate. *J. gen. Virol.* **59**, 163-167.

Mayo, M.A., Robinson, D.J., Jolly, C.A., and Hyman, L. (1989). Nucleotide sequence of potato leafroll luteovirus RNA. *J. gen. Virol.* **70**, 1037-1051.

McElroy, D., and Brettell, R.I.S. (1994). Foreign gene expression in transgenic cereals. *TIBTECH.* **12**, 62-68.

Meshi, T., Motoyoshi, F., Adachi, A., Watanabe, Y., Takamatsu, N., and Okada, Y. (1988). Two concomitant base substitutions in the putative replicase genes of tobacco mosaic virus confer the ability to overcome the effects of a tomato resistance gene, *Tm-1*. *EMBO J.* **7**, 1575-1581.

Meshi, T., Motoyoshi, F., Maeda, T., Yoshiwoka, S., Watanabe, H., and Okada, Y. (1989). Mutations in the tobacco mosaic virus 30-kD protein gene overcome *Tm-2* resistance in tomato. *Plant Cell* **1**, 515-522.

Miller, J.S., and Mayo, M.A. (1991). The location of the 5' end of the potato leafroll luteovirus subgenomic coat protein mRNA. *J. gen. Virol.* **72**, 2633-2638.

Miller, W.A. (1994). Luteoviruses. *In* Webster, R.G., and Granoff, A. (eds) *Encyclopedia of Virology*. pp. 792-798. Academic Press, London.

Miller, W.A., Waterhouse, P.M., and Gerlach. W.L. (1988a). Sequence and organisation of barley yellow dwarf virus genomic RNA. *Nucleic Acids Res.* **16**, 6097-6111.

Miller, W.A., Waterhouse, P.M., Kortt, A.A., and Gerlach. W.L. (1988b). Sequence and identification of the barley yellow dwarf virus coat protein gene. *Virology* **165**, 306-309.

Miller, W.A., Hercus, T., Waterhouse, P.M., and Gerlach. W.L. (1991). A satellite RNA of barley yellow dwarf virus contains a novel hammerhead structure in the self-cleavage domain. *Virology* **183**, 711-720.

Miller, W.A., Dinesh-Kumar, S.P., and Paul, C.P. (1994). Luteovirus gene expression. *Crit. Rev. Plant Sci.* (in press).

Mindrinos, M., Katagiri, F., Yu, G.-L., and Ausubel, F.M. (1994). The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**, 1089-1099.

Mori, M., Mise, K., Kobayashi, K., Okuno, T., and Furusawa, I. (1991). Infectivity of plasmids containing brome mosaic virus cDNA linked to the cauliflower mosaic virus 35S RNA promoter. *J. gen. Virol.* **72**, 243-236.

Morris, T.J., and Dodds, J.A. (1979). Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology* **69**, 854-858.

Murant, A.F. (1993). Complexes of transmission-dependent and helper viruses. *In* Matthews, R.E.F. (ed.) *Diagnosis of Plant Virus Diseases*. pp. 334-367. CRC Press, Boca Raton.

Murphy, J.F., D'Arcy, C.J., and Clark, J.M. (1989). Barley yellow dwarf virus RNA has a 5'-terminal genome-linked protein. *J. gen. Virol.* **70**, 2253-2256.

Needleman, S.B., and Wunsch, C.D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443-453.

Neeleman, L., van der Vossen, E.A.G., and Bol, J.F. (1993). Infection of tobacco with alfalfa mosaic virus cDNAs sheds light on the early function of the coat protein. *Virology* **196**, 883-887.

Norrander, J., Kempe, T., and Messing, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101-106.

Oswald, J.W., and Houston, B.R. (1951). A new disease of cereals, transmissible by aphids. *Plant Disease Reporter* **35**, 471-475.

Padgett, H.S., and Beachy, R.N. (1993). Analysis of a tobacco mosaic virus strain capable of overcoming *N* gene-mediated resistance. *Plant Cell* **5**, 577-586.

Paliwal, Y.C., and Comeau, A. (1987). Yellow dwarf of cereals. Ministry of Agriculture and Food Fact Sheet, Canada.

Passmore, B.K., Sanger, M., Chin, L.-S., Falk, B.W., and Bruening, G. (1993). Beet western yellows virus-associated RNA: an independently replicating RNA that stimulates virus accumulation. *Proc. Natl. Acad. Sci. USA* **90**, 10168-10172.

Plumb, R.T. (1990). The epidemiology of barley yellow dwarf in Europe. *In* Burnett, P.A. (ed.) *World Perspectives on Barley Yellow Dwarf.* pp. 215-227. CIMMYT, Mexico, D.F., Mexico.

Prill, B., Maiss, E., Katul, L., and Casper, R. (1990). Nucleotide sequence of the bean leafroll luteovirus coat protein gene. *Nucleic Acids Res.* **18**, 5544.

Prüfer, D., Tacke, E., Schmitz, J., Kull, B., Kaufmann, A., and Rohde, W. (1992). Ribosomal frameshifting in plants: a novel signal directs the -1 frameshift in the synthesis of the putative viral replicase of potato leafroll luteovirus. *EMBO J.* **11**, 1111-1117.

Rasmusson, D.C., and Schaller, C.W. (1959). The inheritance of resistance in barley to the yellow-dwarf virus. *Agron. J.* **51**, 661-664.

Reisman, D., and de Zoeten, G.A. (1982). A covalently linked protein at the 5'-ends of the genomic RNAs of pea enation mosaic virus. *J. gen. Virol.* **62**, 187-190.

Reutenauer, A., Ziegler-Graff, V., Lot, H., Scheidecker, D., Guilley, H., Richards, K., and Jonard, G. (1993). Identification of beet western yellows luteovirus genes implicated in viral replication and particle morphogenesis. *Virology* **195**, 692-699.

Revill, P.A. Davidson, A.D., and Wright, P.J. (1994). The nucleotide sequence and genome organisation of mushroom bacilliform virus: a single-stranded RNA virus of *Agaricus bisporus* (Lange) Imbach. *Virology* **202**, 904-911.

Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R., and Symons, R.H. (1984). Nucleotide sequence of cucumber-mosaic-virus RNA 2 reveals a translation product significantly homologous to corresponding proteins of other viruses. *Eur. J. Biochem.* **143**, 277-284.

Rezaian, M.A., Williams, R.H.V., and Symons, R.H. (1985). Nucleotide sequence of cucumber mosaic virus RNA 1. Presence of a sequence complementary to part of the viral satellite RNA and homologies with other viral RNAs. *Eur. J. Biochem.* **150**, 331-339.

Rochow, W.F. (1969). Biological properties of four isolates of barley yellow dwarf virus. *Phytopathology* **59**, 1580-1589.

Rochow, W.F. (1979). Variants of barley yellow dwarf virus in nature: detection and fluctuation during twenty years. *Phytopathology* **69**, 655-660.

Rochow, W.F., and Carmichael, L.E. (1979). Specificity among barley yellow dwarf viruses in enzyme immunosorbent assays. *Virology* **95**, 415-420.

Saito, T., Meshi, T., Takamatsu, N., and Okada, Y. (1987). Coat protein gene sequence of tobacco mosaic virus encodes a host response determinant. *Proc. Natl. Acad. Sci. USA* **84**, 6074-6077.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). Molecular Cloning: a laboratory manual. Second Edition. Cold Spring Harbor Laboratory Press, USA.

Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H., and Roe, B.A. (1980). Cloning in a single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**, 161-178.

Santa Cruz, S., and Baulcombe, D.C. (1993). Molecular analysis of potato virus X isolates in relation to the potato hypersensitivity gene *Nx*. *Mol. Plant Microbe Interac.* **6**, 707-714.

Schaller, C.W. (1984). The genetics of resistance to barley yellow dwarf virus in barley. *In* Burnett, P.A. (ed.) *Barley Yellow Dwarf, A Proceedings of the Workshop*. pp. 93-99. CIMMYT 1984.

Schaller, C.W., Rasmusson, D.C., and Qualset, C.O. (1963). Sources of resistance to the yellow-dwarf virus in barley. *Crop Science* **3**, 342-344.

Schaller, C.W., Qualset, C.O., and Rutger, J.N. (1964). Inheritance and linkage of the Yd2 gene conditioning resistance to the barley yellow dwarf virus disease in barley. *Crop Science* **4**, 544-548.

Smith, H.C. (1963). Interaction between isolates of barley yellow dwarf virus. *N.Z. J. agric. Res.* **6**, 343-353.

Smith, O.P., and Harris, K.F. (1990). Potato leaf roll virus 3' genome organisation: sequence of the coat protein gene and identification of a viral subgenomic RNA. *Phytopathology* **80**, 609-614.

Smith, O.P., Hunst, P.L., Hewings, A.D., Stone, A.L., Tolin, S.A., and Damsteegt, V.D. (1991). Identification of dsRNAs associated with soybean dwarf virus-infected soybean. *Phytopathology* **81**, 131-134.

Smith, O.P., Damsteegt, V.D., Harris, K.F., and Vonder Haar, R. (1993). Nucleotide sequence and *E. coli* expression of the coat protein gene of the yellowing strain of soybean dwarf luteovirus. *Arch. Virol.* **133**, 223-231.

Staden, R. (1980). A new computer method for the storage and manipulation of DNA gel reading data. *Nucleic Acids Res.* **8**, 3673-3694.

Steinhauer, D.A., Domingo, E., and Holland, J.J. (1992). Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. *Gene* **122**, 281-288.

Stemmer, W.P.C. (1994a). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**, 10747-10751.

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

Suneson, C.A. (1955). Breeding for resistance to barley yellow dwarf virus in barley. *Agron. J.* **47**, 283.

Symons, R.H. (1979). Extensive sequence homology at the 3' termini of the four RNAs of cucumber mosaic virus. *Nucleic Acids Res.* **7**, 825-837.

Tacke, E., Prüfer, D., Salamini, F., and Rohde, W. (1990). Characterisation of a potato leafroll luteovirus subgenomic RNA: differential expression by internal translation initiation and UAG suppression. *J. gen. Virol.* **71**, 2265-2272.

Tacke, E., Prüfer, D., Schmitz, J., and Rohde, W. (1991). The potato leafroll luteovirus 17K protein is a single-stranded nucleic acid-binding protein. *J. gen. Virol.* **72**, 2035-2038.

Tacke, E., Schmitz, J., Prüfer, D., and Rohde, W. (1993). Mutational analysis of the nucleic acid-binding 17 kDa phosphoprotein of potato leafroll luteovirus identifies an amphipathic α-helix as the domain for protein/protein interactions. *Virology* **197**, 274-282.

Töpfer, R., Matzeit, V., Gronenborn, B., Schell, J., and Steinbiss, H.-H. (1987). A set of plant expression vectors for transcriptional and translational fusions. *Nucleic Acids Res.* **15**, 5890.

Turpen, T.H., Turpen, A.M., Weinzettl, N., Kumagai, M.H., and Dawson, W.O. (1993). Transfection of whole plants from wounds inoculated with *Agrobacterium tumefaciens* containing cDNA of tobacco mosaic virus. *J. Virol. Meth.* **42**, 227-240.

Ueng, P.P., Vincent, J.R., Kawata, E.E., Lei, C.-H., Lister, R.M., and Larkins, B.A. (1992). Nucleotide sequence analysis of the genomes of the MAV-PS1 and P-PAV isolates of barley yellow dwarf virus. *J. gen. Virol.* **73**, 487-492.

van der Wilk, F., Huisman, M.J., Cornelissen, B.J.C., Huttinga, H., and Goldbach, R. (1989). Nucleotide sequence and organisation of potato leafroll virus genomic RNA. *FEBS Lett.* **245**, 51-56.

Vasil, I.K. (1994). Molecular improvement of cereals. *Plant Mol. Biol.* **25**, 925-937.

Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K., and Jonard, G. (1988). Nucleotide sequence of beet western yellows virus RNA. *Nucleic Acids Res.* **16**, 9917-9932.

Veidt, I., Bouzoubaa, S.E., Leiser, R.-M., Ziegler-Graff, V., Guilley, H., Richards, K., and Jonard, G. (1992). Synthesis of full-length transcripts of beet western yellows virus RNA: messenger properties and biological activity in protoplasts. *Virology* **186**, 192-200.

Verwoerd, T.C., Dekker, B.M.M., and Hoekema, A. (1989). A small-scale procedure for the rapid isolation of plant RNAs. *Nucleic Acids Res.* **17**, 2362.

Vincent, J.R., Lister, R.M., Ueng, P.P., Wen, F., Lei, C.H., Klein, R.E., and Larkins, B.A. (1990). Genomic organisation of barley yellow dwarf viruses. *Barley Yellow Dwarf Newsletter No. 3, June 1990,* pp. 34-37.

Vincent, J.R., Lister, R.M., and Larkins, B.A. (1991). Nucleotide sequence analysis and genomic organisation of the NY-RPV isolate of barley yellow dwarf virus. *J. gen. Virol.* **72**, 2347-2355.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. (1982). Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO. J.* **1**, 945-951.

Waterhouse, P.M., and Helms, K. (1984). Purification of particles of subterranean clover red leaf virus using an industrial-grade cellulase. *J. Virol. Methods* **8**, 321-329.

Waterhouse, P.M., Gerlach, W.L., and Miller, W.A. (1986). Serotype-specific and general luteovirus probes from cloned cDNA sequences of barley yellow dwarf virus. *J. gen. Virol.* **67**, 1273-1281.

Waterhouse, P.M., Gildow, F.E., and Johnstone, G.R. (1988). The luteovirus group. *Commonwealth Mycological Institute/Association of Applied Biologists Descriptions of Plant Viruses*. No. 339.

Weber, H., Haeckel, P., and Pfitzner, A.J.P. (1992). A cDNA clone of tomato mosaic virus is infectious in plants. *J. Virol.* **66**, 3909-3912.

Weber, H., Schultze, S., and Pfitzner, A.J.P. (1993). Two amino acid substitutions in the tomato mosaic virus 30-kilodalton movement protein confer the ability to overcome the *Tm-2*^{2} resistance gene in the tomato. *J. Virol.* **67**, 6432-6438.

Wen-jun, S., and Forde, B.G. (1989). Efficient transformation of *Agrobacterium* spp. by high voltage electroporation. *Nucleic Acids Res.* **17**, 8385.

Whitham, S., Dinesh-Kumar, S.P., Choi, D., Hehl, R., Corr, C., and Baker, B. (1994). The product of the tobacco mosaic virus resistance gene *N*: similarity to Toll and the interleukin-1 receptor. *Cell* **78**, 1101-1115.

Wiese, M.V. (1987). Compendium of wheat diseases (ed.). APS Press, Minnesota.

Wu, S., Rinehart, C.A., and Kaesberg, P. (1987). Sequence and organisation of southern bean mosaic virus genomic RNA. *Virology* **161**, 73-80.

Xin, Z.Y., Brettell, R.I.S., Cheng, Z.M., Waterhouse, P.M., Appels, R., Banks, P.M., Zhou, G.H., Chen, X., and Larkin, P.J. (1988). Characterisation of a potential source of barley yellow dwarf virus resistance for wheat. *Genome* **30**, 250-257.

Xiong, Z., and Lommel, S.A. (1989). The complete nucleotide sequence and genome organisation of red clover necrotic mosaic virus RNA 1. *Virology* **171**, 543-554.

Xiong, Z., Kim, K.H., Kendall, T.L., and Lommel, S.A. (1993). Synthesis of the putative red clover necrotic mosaic virus RNA polymerase by ribosomal frameshifting *in vitro*. *Virology* **193**. 213-221.

Yamafuji, R., Watanabe, Y., Meshi, T., and Okada, Y. (1991). Replication of TMV-L and Lta1 RNAs and their recombinants in TMV-resistant *Tm-1* tomato protoplasts. *Virology* **183**, 99-105.

Yamaya, J., Yoshioka, M., Meshi, T., Okada, Y., and Ohno, T. (1988). Expression of tobacco mosaic virus RNA in transgenic plants. *Mol. Gen. Genet.* **211**, 520-525.

Yanisch-Perron, C. Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103.

Young, M.J., Kelly, L., Larkin, P.J., Waterhouse, P.M., and Gerlach, W.L. (1991). Infectious *in vitro* transcripts from a cloned cDNA of barley yellow dwarf virus. *Virology* **180**, 372-379.

Zimmern, D. (1988). Evolution of RNA viruses. *In* Domingo, E., Holland, J.J., and Ahlquist, P. (eds) *RNA genetics, vol 3*. pp. 3-36. CRC, Boca Raton.

APPENDIX

Publications arising from this thesis:

Rathjen, J.P., Karageorgos, L.E., Habili, N., Waterhouse, P.M., and Symons, R.H. (1994). Soybean dwarf luteovirus contains the third variant genome type in the luteovirus group. *Virology* **198**, 671-679.

Ding, S.-W., Rathjen, J.P., Li, W.-X., Swanson, R., Healy, H., and Symons, R.H. (1995). Efficient infection from cDNA clones of cucumber mosaic cucumovirus RNAs in a new plasmid vector. *J. gen. Virol.* **76**, (in press).

Randles, J.W., and Rathjen, J.P. (1994). Genus luteovirus. *In* International Committee on Taxonomy of Viruses, VIth Report. (in press).