Characterisation of novel, fungal dsRNA viral elements associated with the Amasya cherry disease and *Phlebiopsis gigantea* ;

Genome-wide mapping of hypersensitive sites in *Drosophila melanogaster.*

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

by

Zisis Kozlakidis

December 2005

Faculty of Life Sciences
Division of Biology
Imperial College London
Imperial College Road
London, SW7 2AZ

Στην οικογένειά μου...

Abstract

The thesis is divided into three, independent projects, two of which concern mycoviruses.

Amasya cherry disease (ACD) is a disease of cherry trees (*Prunus avium*) first described in the Amasya region of Turkey in the 1950s. The causative agent of the disease is unknown. ACD is associated with twelve dsRNA species that are found in the tissue of infected trees. A similar pattern of dsRNA species has been observed in the tissue of the trees infected with the Cherry chlorotic rusty spot (CCRS) disease in Italy. In this project the five largest and the two smallest dsRNA species have been isolated and sequenced. Their origin, function and relation to the ACD disease are investigated. The ACD dsRNAs are also compared to the equivalent sized dsRNA species found in CCRS diseased tissue. A PCR-based diagnostic test for the detection of ACD in cherry tree leaf tissue based on the sequences of the associated dsRNAs has been developed and tested.

Mycoviruses are found in a large number of fungi. There are two large, asymptomatic dsRNA species isolated from the economically important fungus *Phlebiopsis gigantea* that may be useful as a tool for the genetic manipulation of the fungus. As a prelude to these investigations the two related dsRNAs are sequenced and characterised in terms of structure, function and origin.

The final project aims to establish a method for the creation of libraries containing fragments of *Drosophila melanogaster* DNA containing DNaseI hypersensitive sites (HS). The existing indirect end-labelling technique allows for the detection of HS only in a limited genomic area with limited accuracy of HS mapping and no sequence information. The creation of libraries containing clones with HS is a novel, high-throughput approach which can lead in the rapid, genome-wide mapping and detection of HS. This could assist in the study and characterisation of any given genomic area in terms of its three-dimensional DNA structure and in the actual, and the *in silico,* location of elements such as enhancers and promoters. *D. melanogaster* was chosen for the creation of the genomic HS library, its genome being sequenced and well characterised. The project presents the first map of the known HS in *D. melanogaster.*

Table of Contents

Abstract 3

Contents 4

Introduction 11

Chapter one: Introduction to mycoviruses

1.1 Viral diversity and general properties 14

1.1.1 Double-stranded RNA viruses 14

1.2 Double-stranded RNA mycoviruses 16

1.2.1 General characteristics, discovery and transmission 16

1.2.2 Taxonomy of dsRNA mycoviruses 18

1.2.2.1 *Chrysoviridae* 18

1.2.2.2 *Hypoviridae* 20

1.2.2.3 *Narnaviridae* 21

1.2.2.4 *Partitiviridae* 21

1.2.2.5 (*Myco-*) *Reoviridae* 23

1.2.2.6 *Totiviridae* 24

1.2.2.7 Unassigned members 26

1.3 Defective and satellite dsRNA elements 27

1.4 Evolution and phylogeny of mycoviruses 29

1.4.1 Forces shaping the viral dsRNA evolution 31

1.4.2 Virus-host co-evolution 32

1.4.3 Horizontal transmission of dsRNA mycoviruses 33

1.5 dsRNA elements in plant pathogenic fungi and their potential as biological control agents 34

1.6 Plant and fungal dsRNA viruses 35

1.6.1 *Endornaviridae* 36

1.7 Aims of the project 37

Chapter two: Materials and methods

2.1 Working conditions and materials 46

2.1.1 Safety considerations 46
2.1.2 Sterilisation and precautions of working with RNA 46
2.1.3 Centrifugation 46
2.1.4 Bacterial strains and cloning vectors 46
2.1.5 Bacterial media and antibiotics 47
2.1.6 Fungal media 47
2.1.7 Plant material 47
2.1.8 Cell line material 48
2.1.9 RNA material 48
2.2 Basic nucleic acid manipulation 48
2.2.1 Phenol/chloroform extraction 48
2.2.2 Nucleic acid precipitation in the presence of sodium acetate 48
2.2.3 Non-denaturing agarose gel electrophoresis 49
2.2.4 Recovery of DNA from agarose gels 49
2.2.5 Recovery of dsRNA from agarose gels 49
2.2.6 Small-scale isolation of plasmid DNA (mini-prep) 50
2.2.7 Restriction of DNA fragments 50
2.2.8 Cleaning of DNA fragments 51
2.3 Extraction of total RNA from plants 51
2.4 Extraction of viral dsRNA 51
2.4.1Extraction of viral dsRNA from fungal tissue (LiCl fractionation) 51
2.4.2 Extraction of viral dsRNA from plant tissue (CF-11extraction) 52
2.5 Extraction of viral dsRNA from 50% glycerol stock suspensions of viruses 53
2.6 *DNase*I digestion of isolated Schneider 2 cell nuclei 53
2.6.1 Carey and Smale (2000) 54
2.6.2 Aasland and Stewart (1999) 54
2.6.3 Jack *et al.* (1991) 55
2.6.4 Talbot *et al.* (1990) 55
2.7 Hybridisation of nucleic acids from non-denaturing agarose gels 56
2.7.1 Nucleic acid blotting 56
2.7.2 Radioactive oligolabelling of DNA 57

2.7.3 Radioactive nick-translation labelling of DNA 57
2.7.4 Primer [α^{32}]P labelling 58
2.7.5 Washing of the membranes and autoradiography 58
2.8 PCR amplification from total genomic DNA 58
2.9 Reverse transcription polymerase chain reaction (RT-PCR) from total RNA 59
2.9.1 Amplification of known sequences 59
2.9.2 Extension of known sequences 61
2.10 RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) 62
2.11 cDNA library construction 63
2.12 Cloning of PCR products 65
2.12.1 Preparation of competent JM109 cells 65
2.12.2 Preparationof competent BL21 cells 65
2.12.3 Preparation of DNA fragments for ligation 66
2.12.4 Transformation of competent cells 66
2.12.5 Preparation of the sequencing reactions 67
2.12.6 Nucleotide sequence analysis 67
2.13 SDS-polyacrylamide gel electrophoresis (SDS-PAGE) 67
2.14 Viral particle isolation from fungal tissue 68
2.14.1 Isolation *via* direct pelleting 69
2.14.2 PEG_{6000} precipitation 69
2.15 Staining of protein gels 69
2.15.1 Coomassie blue staining 69
2.15.2 Silver staining 70
2.16 Sterol glycosyltransferase activity assay 70
2.17 Abbreviations 71
2.18 The Genetic code 72
2.19 Amino acid codes 72
2.20 List of suppliers 73

Chapter three: Characterisation of five dsRNA elements associated with the Amasya cherry disease

3.1 Introduction 76
3.1.1 Amasya cherry disease (ACD), symptoms and spread of the disease 76
3.1.2 ACD associated dsRNA elements 77
3.2 Results and discussion 79
3.2.1 Sequencing the four largest dsRNA elements 79
3.2.1.1 Sequencing approach 79
3.2.1.2 Sequence information 80
3.2.1.2.1 ACD dsRNA1a 80
3.2.1.2.2 ACD dsRNA1b 83
3.2.1.2.3 ACD dsRNA2 86
3.2.1.2.4 ACD dsRNA3 93
3.2.1.2.5 ACD dsRNA4 100
3.2.1.3 The 5' and 3' UTR of the ACD dsRNA elements 101
3.2.1.4 Northern blots 106
3.2.1.5 Viral particle isolation 107
3.2.1.6 Phylogeny and viral origin 108
3.2.2 Sequencing the two smallest dsRNA elements 110
3.2.2.1 Sequence information 110
3.2.2.1.1 The 0.6 kbp dsRNA elements 111
3.2.2.1.2 The 1.1 kbp dsRNA element 116
3.3 Comparison of ACD-associated and CCRS-associated dsRNA elements 117
3.3.1 Cherry chlorotic rusty spot (CCRS), symptoms and spread of the disease 117
3.3.2 Relation between the ACD and CCRS dsRNA elements 118
3.3.3 Comparison between the largest dsRNA molecules 119
3.3.4 Comparison between the smallest dsRNA molecules 120
3.4 Development of an RT-PCR diagnostic test for ACD based on the dsRNAs associated with the disease 121
3.4.1 Design of an RT-PCR based approach 121
3.4.2 RT-PCR diagnostic tests results 124
3.4.3 Results of diagnostic RT-PCR tests with dsRNA extracts

from CCRS-diseased leaf tissue 128
3.5 Chapter summary 130
3.5.1 Aetiology of the disease 131

Chapter four: Characterisation of the two dsRNA elements from *Phlebiopsis gigantea*

4.1 Introduction 135
4.1.1 *Phlebiopsis gigantea* and its industrial use 135
4.1.2 The dsRNA elements 136
4.2 Results and discussion 137
4.2.1 Viral capsid detection 137
4.2.2 Sequence information 140
4.2.2.1 dsRNA species 1 141
4.2.2.2 dsRNA species 2 142
4.2.3 Genes and possible functions 144
4.2.3.1 Gene prediction 144
4.2.3.2 *In vitro* translation of purified dsRNA 148
4.2.4 Viral phylogeny 148
4.2.5 Chpater summary and future prospects 153

Chapter five: Reviewing the CYAV dsRNA element sequence, CYSDV RNA2 assembly and PEV1, UGT gene expression assay

5.0 Introduction 153
5.1 Reviewing the CYAV dsRNA element sequence 153
5.1.1 The CYAV dsRNA element 153
5.1.2 Results and discussion 154
5.1.2.1 Sequencing of the CYAV dsRNA derived cDNA library 154
5.1.2.2 CYAV sequence information and viral phylogeny 154
5.2 The CYSDV RNA2 assembly 158
5.2.1 The CYSDV ssRNA virus 158

5.2.2 The CYSDV RNA2 assembly 159
5.3 PEV1, UGT gene expression assay 164
5.3.1 The PEV1 dsRNA virus 164
5.3.2 PEV1, UGT gene cloning and expression 166
5.3.3 UGT activity assay 171
5.4 Chapter summary 172

Chapter six: Genome-wide mapping of hypersensitive sites in *Drosophila melanogaster*

Abstract 175
6.1 Introduction 176
6.1.1 General characteristics of the genomic DNA and hypersensitive sites (HS) 176
6.1.1.1 Morphological features of HS 176
6.1.1.2 Biology of HS 177
6.1.1.3 Functional sequences associated with HS 183
6.1.1.4 Mechanisms of establishment of HS 187
6.1.2 Current technique, its limitations and the novel approach for mapping HS 191
6.1.2.1 Current technique for HS mapping and its limitations in locating HS 191
6.1.2.2 Advantages of the novel approach 193
6.1.2.3 Outline of HS library generation 194
6.1.3 Choice of model system and available information re-examined 196
6.1.3.1 Choice of the model system 196
6.1.3.2 HS and *Drosophila melanogaster*, available information re-examined 197
6.1.4 Project aim 198
6.2 Results 199
6.2.1 Cell culture conditions for the Schneider 2 cell line 199
6.2.2 DNaseI digests of genomic DNA *in situ* and *in vivo* 200

6.2.3 Radioactive labelling of DNA fragments 203
6.2.4 Mapping of HS in *Drosophila melanogaster* 206
6.2.4.1 Arbitrarily primed, cold PCR 206
6.2.4.2 'Wet' mapping of HS 208
6.2.4.3 'Dry' mapping of HS 210
6.3 Discussion 213
6.3.1 Experimental overview 213
6.3.2 New methods, data and future prospects 215

Acknowledgments 218

References 219
A. References for chapters one-five 219
B. References for chapter six 236

List of Figures and Tables 252
A. Figures 252
B. Tables 256

Appendix 258
A. Pictures of ACD-infected leaf tissue 258
B. Manuscript 260

INTRODUCTION

The research detailed in the present thesis does not consist of a single project nor is it dedicated to a unique line of investigation. However, all of the projects/investigations have a central unifiying theme. This theme is the development of various experimental protocols and their application in specific aspects of molecular biology, namely in virology and eukaryote genomics. The common nature underlying living organisms, the genetic material, allows for similar approaches to be adopted for a wide spectrum of experiments. However, there is always a need to optimise these approaches in an organism/model-specific fashion.

A large number of the protocols used in the different investigations presented here are common to all projects, *e.g.* agarose gel electrophoresis, or are variants based on common biological processes, *e.g.* molecular cloning. Hence, the thesis contains a single chapter (Chapter 2) detailing the materials and methods used in all of the projects. Chapter 1 provides a general introduction to mycoviruses as a comprehensive background to the virology projects. Each of Chapters 3 to 6 is structured around the individual project in question.

Chapter 3 concerns the characterisation of seven novel dsRNA elements that are associated with the Amasya cherry disease, found in Turkey. The complete sequence information of these elements is presented and analysed in detail. These dsRNA elements are then compared to those associated with the Cherry chlorotic rusty spot disease, found in Italy, and a diagnostic approach is designed and applied for their detection in tissue samples. The aetiology of the disease is also discussed. In Chapter 4, two novel dsRNA elements, found in the fungus *Phlebiopsis gigantea* are described. Their sequence information is presented and analysed in detail. Their potential for use as biological control agents is also discussed. Chapter 5 contains three smaller investigations that are part of a wider series of experiments. However, they have been used as a way of developing and optimising experimental apporaches that were subsequently used in the main body of research.

The research in Chapter 6 is focused on eukaryote genomics, investigating a novel method of high-troughput mapping for DNase I hypersensitive sites on the *Drosophila melanogaster* genome. This technique is compared and contrasted to the older established one, the end-labelling detection of hypersensitive sites. The results of are discussed extensively, while other developments in the high-throughput mapping of hypersensitive sites in eukaryotes are also presented and analysed.

CHAPTER ONE

Introduction to mycoviruses

1 Introduction

1.1 Viral diversity and general properties

The increasing number of sequenced viral (and cellular) genomes has lead to a drive to utilize this information to rationalize the organization of viral life forms. These genome and gene comparisons have created a wealth of information that has made it possible to construct gene and genome phylogenies. However, it is also becoming apparent that although eukaryotic viruses exist in extraordinary diversity, the individual virus isolates can be divided into a relatively small number of viral lineages and families, including members that infect hosts belonging to all different forms of life.

Viruses are in basic terms genetic information surrounded, in most cases, by a protein coat; they may also contain external structures and a membrane. Viruses are obligate intracellular parasites, requiring a host in which to replicate. They are classified according to features such as morphology (size, shape, enveloped or non-enveloped), genome (RNA, DNA, segmented or non-segmented, single- or double-stranded), genome content (protein composition and function), antigenic properties, biological properties (host range) and physiochemical properties (molecular mass, pH, ionic stability).

1.1.1 Double-stranded RNA viruses

Double-stranded RNA (dsRNA) viruses infect all major groups of organisms, containing a diverse range of pathogens, several of which are of scientific, agricultural or medical importance. dsRNA viruses are also characterized according to a wide variety of genome size and structure, as well as their infectivity. For example, viruses in the family *Reoviridae* (including reoviruses and rotaviruses), which are important causes of death in farm crops and morbidity in animals and humans, have 10 – 12 dsRNA genomic segments. The genome of viruses in the family *Birnaviridae* (such as the infectious bursal disease virus of chickens or the pancreatic necrosis virus of fish) has two segments, while in the *Totiviridae* family (such as the L-A virus of yeast and other fungal and parasite viruses) the genome is unsegmented (Wickner 1993). There are currently eight distinct families of dsRNA viruses recognized by the International Committee for the Taxonomy

of Viruses (ICTV); of those the largest family by far is that of the Reoviridae, which has the largest number of characterized members.

The wide variety of organisms infected by dsRNA viruses necessitates equally diverse methods of viral transmission and infection. dsRNA (viral) transmission can be extracellular or intracellular. Extracellular transmission is either natural, *e.g.* infections predominantly in the respiratory and enterical regions as is the case of reoviruses, or artificial, *i.e.* by mechanical means, through seed, by vegetative propagation, by parasitic plants, invertebrates or fungi. Alternatively, dsRNA elements can also be transmitted in an intracellular fashion, *i.e.* by hyphal anastomosis (cell fusion) in vegetatively compatible filamentous fungi, the formation of spores or by cytoplasmic inheritance, in fungal and in some plant species cultivars (Hull 2001).

dsRNA viruses are characterized by necessary similarities in replication strategy and some similarities in virus structure and certain proteins (usually the RNA-dependent RNA polymerase and capsid proteins) due to the common nature of their genetic material. However, they still represent a very diverse group of viruses in a genomic perspective. The level of amino acid and nucleotide sequence similarity between different genera is low and often approaches that of random sequence comparisons (Mertens 2004). Thus, in many cases it becomes very difficult to assign a tentative function to viral proteins by sequence comparison alone. Even within a single genus there can be very high levels of sequence difference between species.

dsRNA elements can sometimes be correlated with a phenotype, *e.g.* the '447' cytoplasmic male sterility trait in *Vicia faba* is associated with the presence of a high molecular weight dsRNA (Pfeiffer 1998), or the more severe phenotype of La France disease in cultivated mushroom, *Agaricus bisporus*, which is associated with nine dsRNA elements (Romaine *et al.* 1993). However, a large proportion of the characterized dsRNA elements do not have a significant or noticeable phenotype associated with them. For example, although three quarters of the brown rot fungus *Monilinia fructicola* isolates contained one or more dsRNA elements, there was not a single case where their presence

could be associated with a difference in pathogenicity (Tsai 2004). A similar case was reported in a study on the human intracellular parasite *Leishmania braziliensis*, where it was shown that although the dsRNA Leishmania virus was found in approximately a quarter of patient samples examined, there was no association between the presence of the virus or the viral load and the type of lesion (Ogg *et al.* 2003).

This thesis concentrates mainly on dsRNA viruses that are found in fungi and plants and are sometimes associated with a phenotype. These include four dsRNA elements associated with the Amasya Cherry Disease and the Cherry chlorotic rusty spot disease (presented in detail in chapter 3), as well as the cucumber yellows-associated virus (CYAV), co-infecting plants of the cucurbit family with the unrelated pathogenic closterovirus, Beet pseudo-yellows virus (BPYV), (presented in detail in chapter 5). Furthermore, two novel, asymptomatic dsRNA elements are characterized from the economically important fungus *Phlebiopsis gigantea* (presented in detail in chapter 4). An attempt to produce full length infectious clones of the *Cucurbit yellow stunting disorder virus* (CYSDV) is also described (presented in detail in chapter 5). The cloning and functional assay of the cloned UGT gene, found in the PEV1 endornavirus is also described (presented in detail in chapter 5).

1.2 Double-stranded RNA mycoviruses

1.2.1 General characteristics, discovery and transmission

Fungal viruses or 'mycoviruses' have been discovered relatively recently; the first fungal viruses were isolated from cultivated mushrooms in 1962 (Hollings 1962), followed five years later by the discovery of viruses in *Penicillium funiculosum* (Lampson *et al.* 1967). Since then it has been shown that the genome of most mycoviruses consists of dsRNA, which can be encapsidated or non-encapsidated. The viral dsRNA genomes can be segmented or non-segmented, and in the case of multiple dsRNA infections each dsRNA species is separately encapsidated. dsRNA mycoviruses are transmitted intracellularly by hyphal dissemination (lateral transmission) or by the formation of sexual or asexual spores (serial transmission) and are usually associated with latent infections of their hosts.

The mycoviruses' natural host range is limited to individuals within the same or closely related vegetative compatibility groups. A large number of fungi have an 'apoptosis-like' mechanism to prevent cytoplamsic fusion and exchange between two incompatible strains (heterokaryon incompatibility). Thus the host range of mycoviruses remains restricted. Mixed infections with two or more unrelated viruses are common, but heterologous encapsidation has not been reported in mixed infections, there are apparently no structural interactions between co-infecting viruses (McCabe 1999). As different fungi have different ways of replication and growth, depending in each case on environmental conditions, and in certain cases can also accommodate cytoplasmic exchange, *e.g.* the black *Aspergilli* (van Diepeningen 1998), they are excellent hosts for the efficient transmission and spread of the viruses. Thus, mycoviruses are very successful as, despite lacking an extracellular phase to their life cycles, they are a very common natural occurrence (Ghabrial 1998).

Encapsidated dsRNA fungal viruses are typically isometric particles, 25 - 50 nm in diameter (up to 80 nm for the double-shelled *Reoviridae*), with a capsid consisting of 120 or 60 subunits arranged in a T = 2 or T = 1 lattice (*Totiviridae* and *Chrysoviridae* respectively) (Caston *et al.* 2003). The capsid is essential for the proper inclusion of the genomic dsRNA and the replicative machinery into the core of the viral particle. Most capsid proteins have a non-specific RNA binding activity with the RNA-binding domain facing towards the interior of the viral particle core. Multi-segmented genomes are packaged as a single segment per viral capsid, each segment forming a complex with the polymerase protein, while in each capsid there are one or two identical such complexes present. This observation, *i.e.* a single segment per viral capsid, indicates that there is no specific strong interaction (or preference) of a particular genomic segment with the polymerase. Structural differences of full and empty capsids show that RNA-protein interactions contribute towards increasing the stability of the virion structure (Caston *et al.* 2003)

Phenotypic consequences of mycoviruses can range from hypervirulence to hypovirulence and from symptomless to debilitating. However, in the case of multiple dsRNA viruses being present, it is difficult to attribute an observed phenotype to a particular dsRNA genome. Infections by mycoviruses are usually both persistent (benefiting the virus in the absence of infectivity) and latent (benefiting the host)- a phenomenon which might account for their relatively late discovery. The aspect of phenotypes arising from mycoviral infections and the interaction of the dsRNA elements with their hosts is discussed in greater detail in section 1.5.

1.2.2 Taxonomy of dsRNA mycoviruses

There are currently six families of dsRNA mycoviruses recognised by the ICTV, presented in brief below in alphabetical order (sections 1.2.2.1 to 1.2.2.6), as well as some dsRNA elements that are characterised but not yet assigned to a family (section 1.2.2.7). The genomes of the viruses that will be frequently used in comparisons to the dsRNA elements characterised in this thesis are also presented schematically (Figures 1.1 to 1.4). The viruses, their abbreviations, GenBank accession numbers and host taxonomic groups are presented in alphabetical order of their respective families in Table 1.1.

1.2.2.1 *Chrysoviridae*

This family consists of the single genus *Chrysovirus*, formerly classified in the *Partitiviridae* family. However, the genomic organisation and sequence comparison data indicate that the two families are only distantly related and can therefore be classified independently of one another. There are no known natural vectors for this family. The *Chrysoviridae* have a linear, dsRNA genome with four segments ranging in size from about 2.9 – 3.6 kbp. The virions are non-enveloped, icosahedral and 35 - 40 nm in diameter. The virion capsid consists of 60 protein subunits arranged in a genuine T = 1 lattice formation (Caston *et al.* 2003).

The genome is packaged as a single segment per viral capsid, thus in the host there are four populations of viral particles present. It is thought that all four dsRNA segments are essential for viral infection, but there is no mycoviral infectivity assay available yet to

test this hypothesis. Each dsRNA species contains a single open reading frame (ORF) encoding a single protein. However, the functions of only two out of the four encoded proteins have been comprehensively characterised- those of the RNA-dependent RNA polymerase (RdRp) and the coat protein. It has been suggested that the protein produced by the shortest RNA segment might be a virion associated capsid (Jiang and Ghabrial 2004).

Though the four Chrysovirus sequences published to date are either associated with a disease phenotype (Hv145SV (Soldevila *et al.* 2000), ACD– and CCRS- associated Chrysoviruses (Covelli *et al.* 2004)) or a fungal strain instability, there has been no evidence that the dsRNA viruses contribute to the phenotypic effects observed. The genomic organisation of the type member of the *Chrysovirus* genus, *Penicillium chrysogenum virus* (PcV), is shown below in Figure 1.1 (Caston *et al.* 2003).

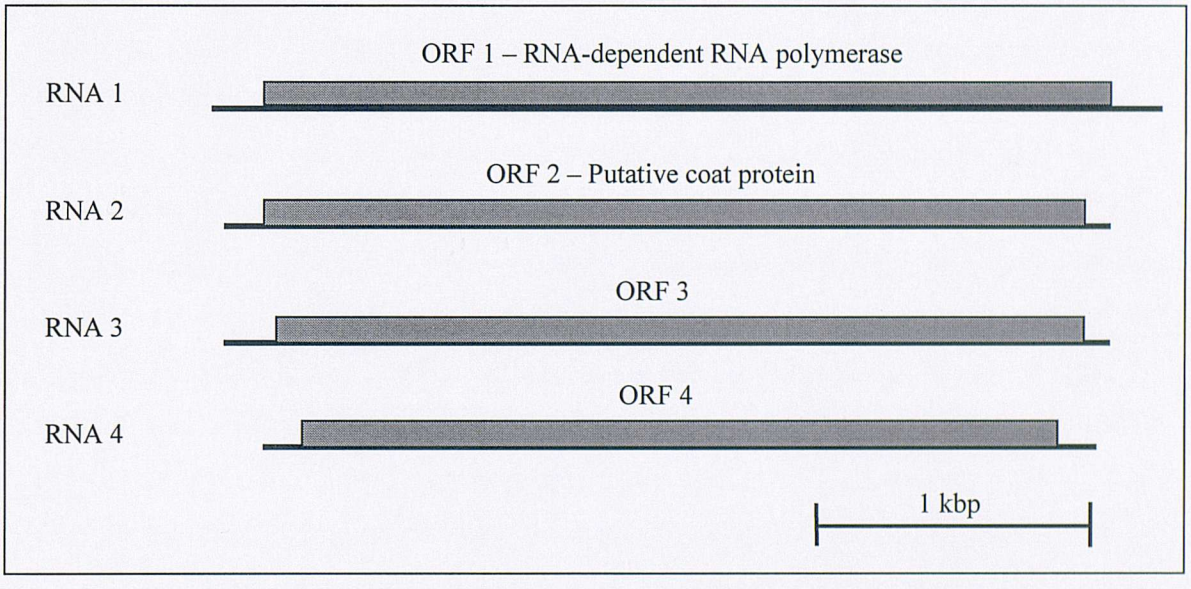


Figure 1.1: Genomic organisation of the *Penicillium chrysogenum* virus (strain ATCC 9480). The viral genome is quadripartite consisting of similar sized fragments. RNAs 1, 2, 3 and 4 are 3562, 3200, 2976 and 2902 bp long respectively. RNA 1 encodes an RNA-dependent RNA polymerase and RNA 2 a putative coat protein; the function of the translation products from RNAs 3 and 4 is unknown. The ORFs are represented by the shaded boxes, while the 5'- and 3'-untranslated regions are shown as a straight line.

1.2.2.2 *Hypoviridae*

This family consists of the single genus *Hypovirus*. The *Hypoviridae* do not encode a coat protein or form a discrete particle. Conversely, the hypovirus genome is predominantly associated with pleiomorphic membrane vesicles, ranging in diameter from 50 – 80 nm. *Cryphonectria hypovirus 1* (CHV1) was first described in 1977, when it was suggested that the dsRNA contained in a strain of the chestnut blight fungus, *Cryphonectria parasitica,* was responsible for the observed hypovirulent phenotype of the fungal strain (Nuss 1992). The known natural host range of *Hypoviridae* is currently limited to *C. parasitica*. Hypovirus infection is nonlytic, persistent and results in very consistent phenotypic changes in the host, *e.g.* reduced host pigmentation, reduced asexual reproduction (conidiation) and reduced host virulence (Nuss 1992; Chen and Nuss 1999).

Hypoviridae have a linear, non-segmented, dsRNA genome varying in size from 9 to 13 kbp. Based on their genomic organisation Hypoviruses can be divided into two groups. CHV1 and CHV2 form one group (Shapira *et al.* 1991; Hillman *et al.* 1994 respectively), with their genome residing in two ORFs, while CHV3 and CHV4 form a second group (Smart *et al.* 1999; Linder –Basso *et al.* 2005 respectively), with their genome residing in a single contiguous ORF. The major difference between the two groups in terms of their coding information is the replication of the 5'- proximal protease domain. The Hypoviruses with a single ORF contain one copy of the protease domain, while the Hypoviruses with two ORFs contain one protease domain per ORF. A UDP-glycosyltransferase domain has also been located in CHV4, downstream of the protease domain, however its activity and function remains unresolved (Linder-Basso *et al.* 2005).

Each ORF of the hypoviruses undergoes autoproteolytic processing upon translation. In the case of CHV1, the shorter, 5' proximal ORF (ORF A) is autocatalitically cleaved into two smaller products: p29 and p40, one being a protease and the other of unknown function. The longer, 3' proximal ORF (ORF B) contains N-terminally a short protease, p48, that is autocatalitically released, while the remaining polypeptide is assigned as the replication protein containing RNA-dependent RNA polymerase and helicase domains

deduced by amino acid sequence comparisons. It should be noted though, that the autoproteolytic processing cascades in the hypovirus-encoded polyproteins are not mapped in detail. Hypovirus infection results in alterations in the host signal transduction pathways and specifically the G-protein signalling pathway (Dawe *et al.* 2004). The genomic organisation of the type member of the genus, *Cryphonectria hypovirus 1* (CHV 1), is shown in Figure 1.2 (Nuss 1992).

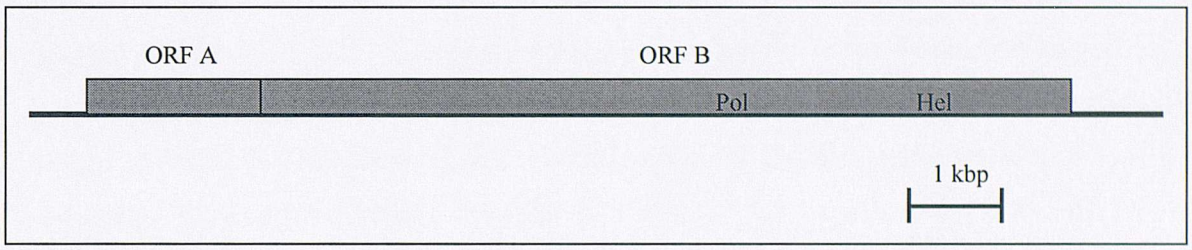


Figure 1.2: Genomic organisation of the *Cryphonectria parasitica* hypovirus 1 (isolate EP713). The viral genome is unsegmented, consisting of two ORFs. The ORFs are represented by the shaded boxes, while the 5'- and 3'- untranslated regions are shown as a straight line. ORF A is 1869 bp long and ORF 2 is 9498 bp long. The approximate positions of the RNA-dependent RNA polymerase and helicase motifs in ORF B are also shown.

1.2.2.3 *Narnaviridae*

The family consists of two genera: *Narnavirus* which infect fungi and *Mitovirus* which infect mitochondria. Narnaviruses are non-encapsidated and have a linear genome varying in size between 2.2 and 3.6 kbp. As both ssRNA and dsRNA forms can be isolated from fungi, they can be regarded as either ssRNA or dsRNA viruses. Their genome contains a single ORF, encoding only for the RdRp (RNA-dependent RNA polymerise). They are therefore the simplest of the dsRNA mycoviruses. The type member of the Narnavirus genus is the *Saccharomyces narnavirus 20S RNA* (SNV-20S) (Esteban and Fujimura 2003).

1.2.2.4 *Partitiviridae*

This family consists of three genera: *Partitivirus* (infecting fungi) *Alphacryptovirus* and *Betacryptovirus* (infecting plants). Members of the *Partitiviridae* have a linear, bi-partite dsRNA genome varying in size from 1.3 to 2.2 kbp The two segments are of similar size.

There are no known natural vectors for this family. The partitiviruses are some of the most studied mycoviruses as there are approximately twenty characterised members of the genus. The virions are non-enveloped, icosahedral and 30 – 35 nm in diameter. The virion capsid consists of 120 protein subunits arranged in a T = 2 lattice formation. Transmission through asexual spores of the host fungi is highly efficient; sexual spores are usually virus-free. However, it is not known how the mitotic or meiotic processes affect the viral viability (Ghabrial and Hillman 1999). The successful transfection of a dsRNA-free isolate of the white root rot fungus *Rosellinia necatrix* Prillieux with purified partitivirus particles has been recently demonstrated (Sasaki *et al.* 2005). It provides a useful tool in studying the processes affecting efficient viral transmission as well as the influence of the virus in the host fungus.

The genome is packaged as a single segment per viral capsid, thus in the host there are two populations of viral particles present. Each dsRNA species contains a single open reading frame (ORF), encoding for a single protein. The larger segment encodes the RdRp, while the shorter one encodes the coat protein (Ghabrial 1998).

It should be noted at this point that there are three partitiviruses described, from two different isolates of the fungus *Gremmeniella abietina* (Tuomivirta and Hantula 2003; Tuomivirta and Hantula 2005) and an isolate of the fungus *Discula destructiva* (Rong *et al.* 2002) with a tri-partite genome. The third dsRNA fragment is the smallest of the three, containing a single ORF. It is not known, though, whether the ORF is translated at all and what the function of its protein product is. The genomic organisation of the type member of the genus, *Atkinsonella hypoxylon virus* (AhV), is shown below in Figure 1.3 (Oh and Hillman 1995).

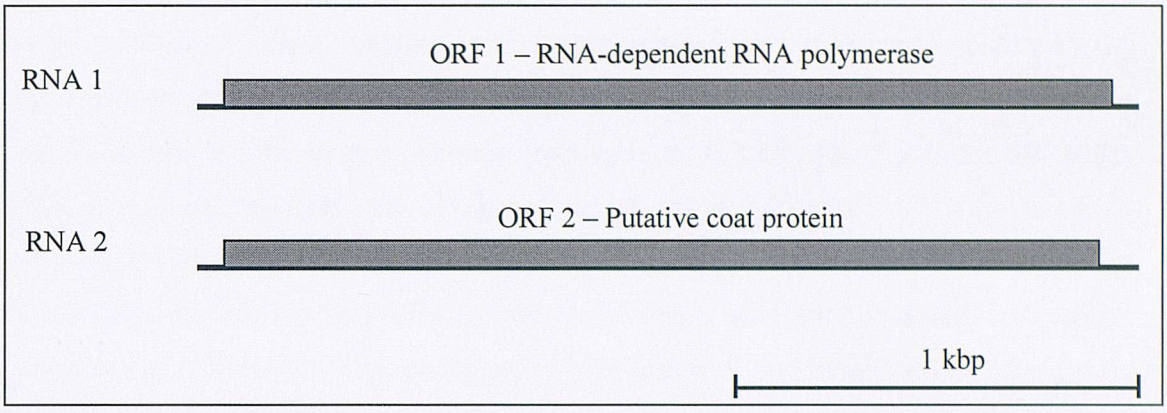


Figure 1.3: Genomic organisation of the *Atkinsonella hypoxylon partitivirus* (isolate H2). The viral genome is bipartite consisting of two similar sized fragments. RNA 1 is 2180 bp long and RNA 2 is 2135 bp long, encoding an RNA-dependent RNA polymerase and a putative coat protein respectively. The ORFs are represented by the shaded boxes, while the 5'- and 3'-untranslated regions are shown as a straight line.

1.2.2.5 *(Myco-) Reoviridae*

The seventh ICTV report lists nine genera in the large *Reoviridae* family. There have recently been three reoviruses infecting fungi reported, tentatively assigned as an additional, individual genus of Mycoreoviruses (Mertens *et al.* 2004). These are two viruses from two different, hypovirulent strains of *C. parasitica*, named Mycoreovirus 1 and 2 (MYRV1 and MYRV2) (Hillman *et al.* 2004) and one from a hypovirulent strain of *Rosellinia necatrix*, named MYRV3 (Wei *et al.* 2004). Although the virulence of the *C. parasitica* isolate was greatly attenuated by infection with MYRV1 and 2, pigmentation and conidiation of the host fungus were slightly enhanced. This phenotype is dramatically different from the *C. parasitica* hypovirulent strains infected by *Hypoviridae,* which suppress virulence, conidiation and pigmentation (Suzuki *et al.* 2004). The genome of MYRV 1 and 2 consists of 11 dsRNA segments, of MYRV3 consists of 12. MYRV1 and 2 have a double-shelled particle structure, and their genomic fragments range from *ca.* 0.7 to 4.2 kbp, characteristics typical of the *Reoviridae* family.

There is a considerable variation amongst family members of *Reoviridae* with protein functions that are unique within a genus. Thus, the functions of proteins encoded by a given dsRNA segment cannot be predicted solely on the basis of sequence comparison across genera. The incomplete nature of the sequencing information published to date has

not allowed for extensive sequence comparisons between all fragments of the MYRV1, 2 and 3 viruses. However, it has been shown that there is a single ORF in each of the dsRNA segments. The largest genomic fragment, *ca.* 4.2 kbp long, encodes the RdRp, while the fourth segment, *ca.* 2.3 kbp long, is myristoylated at the N-terminus. In addition, the mycoreoviruses characterised are closer in terms of genomic sequence conservation to Coltiviruses, a *Reoviridae* genus comprising vertebrate and tick viruses (Suzuki *et al.* 2004). Coltiviruses are among the less well understood *Reoviridae* genera in terms of their particle structure and protein function.

1.2.2.6 *Totiviridae*

This family consists of three genera: *Giardiavirus*- infecting protozoan parasites, *Leishmaniavirus*- infecting protozoan parasites and *Totivirus*- infecting fungi. There is considerable variation of sequence conservation, genomic structure and expression among members of the *Totiviridae* family, as well as among members of each individual genus. The *Totiviridae* have an undivided linear dsRNA genome varying in size from 3.2 to 6.2 kbp. There are no known natural vectors for this family. The virions are non-enveloped, icosahedral and *ca.* 40 nm in diameter. The virion capsid consists of 120 protein subunits arranged in a T = 1 lattice formation. Totiviruses lack an extracellular phase thus transmission is intracellular, during cell division, sporogenesis and cell fusion. All *Totiviridae* infections characterised to date have been asymptomatic. There are no known experimental host ranges for the fungal totiviruses, because of the lack of suitable infectivity assays (Ghabrial and Patterson 1999).

Information on the replication cycle of the *Totiviridae* has been derived from *in vitro* studies of virion-associated RNA polymerase activity and the isolation of particles representing various stages in the replication cycle (Huang *et al.* 1997). Though *Totiviridae* contain only the genes essential for their survival, *i.e.* the RdRp and the capsid protein, they use host proteins efficiently. It has been shown that the capsid protein in some cases, *e.g.* in the *Helminthosporium victoriae 190S virus* (Hv190SV), is subject to host-encoded post-translational phosphorylation and proteolytic cleavage. These two modifications are involved respectively in the interaction with the dsRNA genome and

the release of the (+) strand RNA transcripts from virions (Soldevila *et al.* 1998). However, the proposed life cycle of Hv190SV can not be representative of all the *Totiviridae* as the life cycle of Leishmaniaviruses was shown to be different. The Leishmaniaviruses capsid protein has an endoribonuclease activity (MacBeth and Patterson 1998) and their genome, expressed as a gag-pol fusion protein, is cleaved by a host-encoded cysteine protease (Carrion *et al.* 2003).

Another case of a totivirus-host interaction has been characterised in yeast, where a system of six chromosomal genes (termed superkiller or SKI) negatively control the copy number of the *Saccharomyces cerevisiae totivirus L-A* (ScV-L-A) and its satellite M-dsRNAs (Toh-E *et al.* 1978). The essential function of these genes is to block virus multiplication by affecting the initiation of translation, rather than the stability of the mRNA. About 30 chromosomal genes (termed maintenance of killer or MAK) are required for the stable replication of the satellite M-dsRNAs that encode a toxin, resulting in the killer phenotype of the strain. However, three of those are necessary for the maintenance of the stable replication of the helper virus ScV-L-A (Icho and Wickner 1989; Ghabrial and Patterson 1999).

The genome of the *Totiviridae* consists of two overlapping ORFs: ORF 1 encodes the capsid protein, while ORF 2 contains the RdRp. The overlap of these two ORFs can be quite variable. Viruses such as ScV-L-A (130 nt), LRV 1-1 and LRV 1-4 (71 nt each) and GLV (122 nt) have relatively long overlapping regions (Icho and Wickner, 1989; Stuart *et al.* 1992; Scheffter *et al.* 1994; Wang *et al.* 1993 respectively). Conversely, viruses such as SsRV1, SsRV2 (2 nt each), CmRV (4 nt), GaRV-L1 (1 nt) and Hv190SV (overlap directly) have short or directly overlapping regions (Preisig *et al.* 1998; Cheng *et al.* 2003; Tuomivirta and Hantula 2003; Huang and Ghabrial 1996 respectively). There are two different strategies suggested for the expression of the second ORF, however they are not clearly correlated yet with the size or sequence information of the overlap region. The first is the production of a fusion protein between the coat protein and polymerase regions (gag-pol) resulting by ribosomal frameshifting. This seems to be the case for the GLV, where the overlap region contained a slippery site and a pseudoknot structure was

found immediately downstream of it (Wang *et al.* 1993). The second suggested strategy is that of a coupled termination and (internal re-)initiation, as is the case of Hv190SV where the RdRp is expressed as a non-fused polypeptide independently of a coat protein (Soldevila *et al.* 2000).

It should be noted at this point that though the general genomic structure of *Totiviridae* is a single dsRNA fragment containing two ORFs, there are certain cases, *e.g.* the *Leishmania virus 1-1* (LRV1-1), where a small 5'- proximal ORF is present (Stuart *et al.* 1992). This, ORF 1 of LRV1-1, is 71 amino acids long with no significant similarity to characterised protein motifs. It could express a protein, though the high number of secondary structures present suggest a role other than protein coding, probably that of efficient ribosomal binding (Stuart *et al.* 1992).

The genomic organisation of the type member of the genus, *Saccharomyces cerevisiae virus L-A* (ScV-L-A), is shown below in Figure 1.4 (Icho and Wickner 1989).

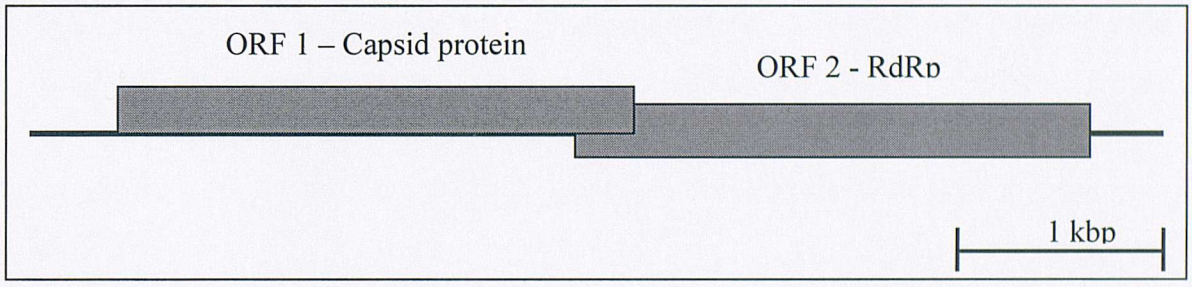


Figure 1.4: Genomic organisation of the *Saccharomyces cerevisiae L-A totivirus* (strain RE59). The viral genome is monopartite consisting of two ORFs. ORF 1 is 2180 bp long and ORF 2 is 2135 bp long, encoding respectively a putative coat protein and an RdRp. The ORFs are represented by the shaded boxes, while the 5'- and 3'- untranslated regions are shown as a straight line.

1.2.2.7 Unassigned members

There are a small number of dsRNA mycoviruses that have not yet been assigned to a specific genus, or that are not sufficiently distinguished from characterised genera so that they can be classified into a new family or genus. These are viruses that have some of

their key characteristics, but not all, known or are altogether poorly characterised. The present introduction provides two such examples below- included in the ICTV database- illustrating in parallel the fact that an exhaustive and universally applicable virus taxonomy is not yet complete.

The La France disease is a major limiting factor in the production of cultivated mushroom, *Agaricus bisporus*, with a viral aetiology suggested for the disease. There are nine dsRNA fragments associated with symptomatic cultures, collectively named as La France Isometric virus (LIV), ranging in size from 0.8 to 3.6 kbp. Some of these fragments are encapsidated in isometric particles with a 36 nm diameter. The main transmission of the disease agent is *via* basidiospores (Romaine *et al.* 1993).

The mushroom virus X disease (MVX) causes a combination of the following symptoms in commercially grown *A. bisporus*: premature veil opening, browning of the mushrooms, sporophore malformation and loss of crop yield. MVX was first recorded in 1996 it was associated with localised areas of pin suppression. It seriously affected the UK mushroom industry in 2000/01 and has now been reported in Ireland, Holland and a number of other mushroom growing countries. A total of 26 dsRNA bands have been identified on agarose gels as associated with MVX, suggesting that there is a mixture of different viruses and sub-genomic/satellite dsRNA elements associated with the disease (Grogan *et al.* 2003).

1.3 Defective and satellite dsRNA elements

A common feature of mycoviruses is the presence of defective and/or satellite dsRNA elements. The identification and sequencing of these smaller dsRNA elements provides clues for the minimal sequence required for the stable replication of a dsRNA element, though they are often thought to be unimportant as components of natural infection so their long term evolution has been less well studied. The information derived from studying defective and satellite dsRNA elements can be utilised in the design of dsRNA-based expression systems for the specific host systems in which these elements originated (Nuss and Koltin 1990). However, both satellite and defective dsRNA elements can

potentially interfere with the replication of the helper virus, probably by competing for replication factors (Buck 1998).

Satellite dsRNAs are sub-viral elements that depend for their multiplication on co-infection of a host cell with a helper virus. Satellite elements have substantially distinct nucleotide sequences from those of the genomes of either the helper virus or of their host. If a satellite element encodes its own capsid protein that will contain its genome, it is referred to as a satellite virus (Hull 2001). There are no satellite viruses found co-infecting fungi with dsRNA mycoviruses. Conversely, there has been a systematic study and characterisation of a variety of satellite dsRNA elements that are associated with different mycoviruses. The best studied of those are the M satellite dsRNA elements of yeast. These satellite elements are encapsidated in separate viral particles from those containing the yeast L-A helper virus. The satellite element genome encodes a secreted toxin, thus conferring an advantage to the yeast host. The M toxin is lethal for all non-M dsRNA containing yeast cells in the vicinity of the secreting host. There is a number of different M dsRNA elements, each one encoding a distinct toxin-immunity specificity-the immunity specificity factor being the protoxin itself. The production, processing, secretion and action of these toxins have been studied in depth. (Wickner 1996).

Similar toxin-coding dsRNA elements have also been described in *Ustilago maydis,* the toxins killing sensitive cells of the same and related species. There are three well-characterised viral subtypes, P1, P4 and P6, which differ in the specificity of their secreted killer toxins. Each subtype has three size classes of satellite dsRNA elements: H (heavy), M (medium) and L (light). The L segments are unique in being derived from one end of the larger M dsRNA elements (Park *et al.* 1996). Non-coding satellite dsRNA elements have been characterised in *Trichomonas vaginalis* virus 1 (TVV-1), *A. hypoxylon* partitivirus (AhV) and *C. parasitica* hypovirus 3 containing isolates (Tai *et al.* 1995; Oh and Hillman 1995; Hillman *et al.* 2000 respectively).

Defective dsRNA elements depend on the helper virus for their replication, but they are also derived from the helper virus, generally by internal deletions and/or recombination

of existing dsRNA information. Defective dsRNAs can be produced under specific conditions, such as the rapid, high-multiplicity of infection passage of virus through a highly susceptible host species (Hillman *et al.* 2000). The extent that these defective dsRNAs represent by-products of virus replication in culture as opposed to the normal ratio of these components found in mycoviral populations in nature is unknown. It can be speculated that because of their negative effect on the accumulation of the helper virus due to competition for replicating factors, defective dsRNA elements would not be favoured in natural populations unless they confer a significant advantage on the genomic RNA or the host (Celix *et al.* 1997). Defective dsRNA elements have been also characterised for the *C. parasitica* hypoviruses and the yeast L-A totivirus (Hillman *et al.* 2000; Wickner 1996 respectively).

1.4 Evolution and phylogeny of mycoviruses

Viral evolution: One or many lineages?

Fungal dsRNA viruses are thought to have originated at a very early stage in viral phylogeny. There are related viruses to be found in divergent clades of hosts. The *Totiviridae* family illustrates this argument characteristically well. The Leishmania viruses, though closely related, are found in both New World and Old World parasites, suggesting the presence of the ancestral virus in the cytoplasm of the parasite before the contemporary diversification and speciation took place (Scheffter *et al.* 1995). Totiviruses presence is a similar case, as high sequence similarities can be found between different species of fungi (as seen on Table 1.1). Thus, the sequence information available indicates that there are several independently evolving dsRNA lineages, characterised by the different genomic organisations.

It is unusual however that many independent evolutionary events to all end in a neutral state, *e.g.* most infections with *Totiviridae* and *Partitiviridae* are asymptomatic. This conclusion may be skewed by a number of factors (McCabe *et al.* 1999). Firstly, only a small number of these viruses have been studied to date. There are only five *C. parasitica* and no *Monilinia fructicola* dsRNA viruses characterised to date, though it has been

shown that the presence of dsRNA elements in nature can range between 28% (in N. American *C. parasitica* isolates (Linder-Basso *et al.* 2005)) and 73.5% of the total fungal population (in New Zealand isolates of *M. fructicola* (Tsai *et al.* 2004)).

Secondly, the ability of certain fungal species for inter- and intra-species horizontal transmission of their dsRNA elements- *e.g.* the Aspergilli (van Diepeningen *et al.* 1998), and the ability of the fungi to harbour multiple infections- *e.g.* the co-infection of *G. abietina* with three distinct dsRNA viruses (Tuomivirta and Hantula 2005), present severe difficulties in the construction of phylogenetic trees describing the evolution of dsRNA mycoviruses within a genus or a family. Further complexity and variation is added by the common presence of defective and satellite dsRNAs.

The comparative analysis of the amino acid sequences of proteins encoded by dsRNA viruses reveals surprisingly little similarity between viruses of different genera (Mertens 2004). The phylogenetic analysis of RdRps, which are necessarily among the most conserved proteins in dsRNA viruses, does not suggest a monophyletic origin for dsRNA viruses, thus mycoviruses could have originated at multiple events in evolution. It is therefore likely that there are different evolutionary trees with unclear links between them, rather than a single tree with unresolved, unclear subgroups (Ghabrial 1998).

It should be noted though that using sequence information alone for the construction of phylogenetic trees is feasible only for viruses that are closely related and/or are exchanging genetic information, so that sequence similarities can be detected and followed. For longer evolutionary distances, which is usually the case with fungal viruses (as they are considered representative of early viral phylogeny) the recognisable sequence similarity signals detected are insignificantly higher than those that can arise at random. In such cases, the detection of evolution needs to include other characteristics as well, such as overall protein structure and architecture (Bruenn 2003; Bamford 2003). For example, most fungal viruses retain their genome and polymerase enzymes within a stable closed (isometric) protein capsid shell. Many of the second (outer) capsid layers and proteins show specific adaptations to the host species and thus a greater degree of

diversity in organisation and structure than those of the inner core. Thus in sequence comparisons it is expected the inner capsid proteins to be significantly more conserved than the outer capsid proteins (Mertens 2004).

1.4.1 Forces shaping the viral dsRNA evolution

The current opinion that RNA viruses are very variable derives mostly from the analysis of viruses such as *Influenza A virus* and *Hepatitis C virus*. It is possible however, that these may be exceptions rather than the rule. The genomic sequence data for dsRNA viruses points to trends regarding the main forces shaping the viral dsRNA evolution. Firstly, RNA mycoviruses would be expected to evolve at a more rapid rate than their hosts' DNA genomes, as their RdRp proteins lack a proof-reading activity, resulting in higher mutation rates than DNA polymerases (Buck 1998). However, high mutation rates are not necessarily adaptive, as a large fraction of the mutations can be deleterious or lethal. The segmentation of the genome improves genetic diversity, as related strains can exchange genetic segments and also reduces the effect of lethal mutations to single genes or genomic fragments, rather than the whole virus genome. Extensive experiments analysing the effects of mutations and deletions have been conducted with the yeast ScV-L-A and ScV-L-BC totiviruses (Ribas *et al.* 1994 and Yao *et al.* 1995 respectively).

Secondly, genetic recombination does affect the mycoviral diversity and evolution, though it has not been studied in as much detail in dsRNA mycoviruses as in other RNA viruses. Genetic recombination is defined here as the formation of novel chimeric molecules from segments previously separated on the same molecule or present on different parental molecules (Coenen *et al.* 1997). Genetic recombination serves a dual purpose: in the long-term it generates novel molecules that might have a selective advantage over the parental ones, and in the short-term functions as a repair mechanism by rescuing functional sequences from parental mutated ones (Nagy and Simon 1997; Mikkelsen and Pedersen 2000). Most such recombination events are mediated by the viral RdRp, thus recombinants can be considered by-products of the replication machinery. As the presence of defective dsRNA elements in mycoviral populations is a common occurrence (Shapira *et al.* 1991; Hong *et al.* 1998; Hillman *et al.* 2000), it is

evident that genetic recombination does take place. However, the exact mechanisms by which it is facilitated, *e.g.* by the replicase-driven template switching model or the RNA breakage and ligation model, and the effects it has on the evolution of dsRNA populations so far remain unresolved

Thirdly, selection is not the only factor shaping the evolution of dsRNA viruses. A large number of mycoviruses are asymptomatic, *i.e.* their infection confers neither an advantage nor disadvantage to the fungal host. Thus host selection pressure would be much reduced allowing for other forces to shape the mycoviral evolution. Mycoviruses might have evolved towards benevolence in accordance with the theory that there is a trade-off between virus transmission and virulence. Mycoviruses are intracellular, thus their transmission rate is directly proportional to the fitness of their host; the smaller the reduction in fitness, the higher the viral transmission rate (Coenen *et al.*1997). In the case of hypovirulence, *e.g.* in the *C. parasitica* hypoviruses, the host plants of virus-infected fungi live significantly longer than the host plants of virus-free fungi, suggesting that a virus-induced reduction in virulence of the host might be compensated for by an expansion of both the host (tree) and the fungal life span.

Determining the mode of evolution of mycoviruses is complex and difficult as there are several modes of evolution suggested. The two main ones, virus-host co-evolution and horizontal transmission, are presented in brief in sections 1.4.2 and 1.4.3 below.

1.4.2 Virus-host co-evolution

Mycoviruses that have been associated with their specific host for long periods of evolutionary time would be expected to co-evolve with their host (Buck 1998). Such an example was shown in *G. graminis*, where hypovirulent isolates containing nine dsRNAs lost their sexual-reproductive capacity, while dsRNA-free isolates did not. These dsRNAs are poorly transmitted in sexual spores, thus the replication of dsRNA and/or its maintenance is integrated into the life cycle of *G. graminis*, by suppressing functions that are essential for host sexual reproduction. This indicates a long evolutionary relation and between the dsRNA and their fungal host (Nuss and Koltin 1990). The virus-host co-

evolution hypothesis is further supported by the RdRp sequence comparisons between Old- and New-world Leishmania infecting viruses. LRV1-1 and LRV1-4, both infecting New-world parasites, shared a 77% identity in their RdRp peptide sequence (Scheffter *et al.* 1994), while LRV1-4 and LRV2-1, the latter infecting an Old-world parasite, had a 43% identity in their RdRp peptide sequence (Scheffter *et al.* 1995). These data suggest the existence of a progenitor Leishmaniavirus predating the divergence of the Old- and New-world protozoan parasites and an extensive host-virus co-evolution, since the divergence between Old- and New-world parasites is reflected in the divergence of their respective dsRNA viruses.

1.4.3 Horizontal transmission of dsRNA mycoviruses

The transfer of genetic information, such as dsRNA viral genomes, between widely divergent taxonomic groups of organisms (for example between plants and bacteria) is referred to as horizontal transmission. The best studied example of this rare phenomenon is of the Ti plasmid transferred between *Agrobacterium tumefaciens* and plants. The hypothesis of inter-species horizontal transmissions at certain points in the past during mycoviral evolution is supported by the degree of relatedness between dsRNA viral groups with different host ranges. The Mycoreovirus infecting fungi are closely related to the Coltiviruses infecting vertebrates (Hillman *et al.* 2004). Partitiviruses infecting fungi are related to Cryptoviruses infecting plants, while Totiviruses infecting fungi are related to Leishmaniaviruses infecting protozoa (Buck 1998). Additionally, the presence of homologous dsRNA elements in vegetatively incompatible strains of fungi, such as *Fusarium poae* or *Helicobasidium mompa*, may also be the result of horizontal transfers of these elements, occurring under natural conditions within the same (or related) host species (Compel *et al.* 1998 and Suzaki *et al.* 2005 respectively).

The most effective way to distinguish between these two modes of evolution is by the comparative study of the phylogenies of the different viral lineages and of their hosts. The two phylogenetic trees can then be compared and if a high level of correlation is found between them, as in the case of Leishmaniaviruses and their hosts, this suggests that host-virus co-evolution has occurred. If the trees do not correlate, this might be due

to horizontal transmission having taken place. This method enables the detection of horizontal transfers, their approximate time in evolution and the direction in which they occurred (Buck 1998). However, this method might fail in cases when the origin of the viruses appears polyphyletic (Ghabrial 1998), and when horizontal transmission has taken place between hosts of related groups, where evolutionary distances are short.

1.5 dsRNA elements in plant pathogenic fungi and their potential as biological control agents

The use of fungicidal chemicals that, when applied to plants, protect them from infection or minimise invasion is an important method for the control of many fungal diseases. However, numerous fungal pests have evolved resistance to a wide range of fungicides. On the other hand, despite widespread release over many years, the efficacy of biological control agents in populations of target pests has remained relatively stable. For this reason, research into developing biological control methods has attracted increasing attention (Peever *et al.* 2000). In the case of plant pathogenic fungi, dsRNA elements may provide an ideal means (as adapted, host-specific vectors) for the genetic transformation or regulation of virulence of their host.

Although mycoviruses are widespread in nature and are stably replicated within the cytoplasm of their host, their phenotypic consequences can vary greatly. Most mycoviral infections are asymptomatic. However, in the case of *Nectria radicicola*, the presence of an unclassified 6 kbp dsRNA element is associated with high levels of sporulation, laccase activity and pigmentation, and significant increase in the virulence of the dsRNA infected host (Ahn and Lee 2001). Conversely, hypoviruses and the mycoreovirus1 (MYRV1) are the causal agents of hypovirulence in *C. parasitica* strains (Nuss 1992 and Hillman *et al.* 2004 respectively). The lack of infectivity assays for fungal viruses makes it difficult to study viral-induced phenotypes. Thus, to date very few viruses have been shown as the sole causal agents of a specific phenotype, such as the alteration in fungal pigmentation, which could be used as markers confirming a successful viral infection. In

the case of multiple dsRNA viruses being present it is even more difficult to attribute an observed phenotype to any particular dsRNA genome (McCabe *et al.* 1999).

Hypoviruses are an exception amongst mycoviruses as their infection is persistent and alters the host phenotype in a predictable and efficient manner. Thus they are currently the most studied in relation to the development of a biological control agent (Nuss 1992). Hypovirulent (hypovirus containing) strains of *C. parasitica* have been utilised successfully in Europe over the past three decades for the control of the chestnut blight fungus. The same approach in N. America however has not been as successful, probably due to the high degree of diversity in vegetative incompatibility groups within the natural population of the fungus, thus prohibiting the horizontal spread of the virus *via* anastomosis (Chen and Nuss 1999).

The recent ability to uncouple hypovirulence from associated traits, *e.g.* reduced levels of conidiation, established the feasibility of engineering infectious viral cDNA for the construction of hypovirulent fungal strains with specific phenotypic traits. The stable integration of the viral genome into the host genome can provide a way (*i.e.* the viral transmission to ascospore progeny) of circumventing the problem of viral dissemination, restricted by the variety of vegetative compatibility groups. However, it is also suggested that a successful and persistent hypovirulence-mediated biological control may require continual release of hypovirulent inocula into nature (Chen and Nuss 1999).

1.6 Plant and fungal dsRNA viruses

The number of dsRNA plant viruses characterised is small when compared to that of (+) ssRNA plant viruses. There are currently two families recognised by the ICTV: *Partitiviridae* and *Phytoreoviridae*, while the *Endornaviridae* family is also tentatively assigned to this group. However, this classification may be reviewed, as there is a recently characterised fungal endornavirus (Hacker *et al.* 2005). This thesis makes extensive references to the *Endornaviridae* (section 1.6.1) and are frequently used in comparisons with other presently characterised dsRNA elements. The viruses, their

abbreviations, GenBank accession numbers and host taxonomic groups are presented in alphabetical order in Table 1.2.

1.6.1 *Endornaviridae*

This viral family consists of the single genus *Endornavirus*. The *Endornaviridae* do not encode a coat protein or form a discrete particle. They are large (>10 kbp) cytoplasmic dsRNA elements that can replicate independently of their host and can be regarded as RNA plasmids. These dsRNA viruses, when found in plants, are in every tissue of the host, at every developmental stage and are maintained almost at a constant concentration (100 copies/cell) in most tissues, with the exception of pollen grains where the copy number increases 10 fold (Horiuchi *et al.* 2003). Unlike most plant viruses, the endornaviruses are efficiently transmitted intracellularly as well as to progeny via pollen and ova (Moriyama *et al.* 1999). There is no horizontal transmission incident recorded to date.

Endornaviridae infections are asymptomatic, with the exception of the *Vicia faba endornavirus* (VFV), which is associated with the cytoplasmic male sterility trait of the '447' line (Pfeiffer 1998). Their genome consists of a single large ORF (called the Long Protein, LP) containing motifs characteristic of an RNA-dependent RNA polymerase (RdRp) and an RNA helicase, deduced by amino acid sequence comparisons. However, very little is known about the post-translational modifications and enzymatic activities of the LP. The ORF could also contain protease domains. Due to the low degree of conservation, protease motifs are very difficult to identify in viral sequences. Thus, *in vitro* proteolysis studies are required to prove their existence (Pfeiffer 1998).

The (-)RNA strand of the endornaviruses is continuous, while the (+)RNA, coding strand is interrupted, containing a conserved nick. The conservation of this break on the genomic sequence among most endornaviruses and the fact that the LP gene continues in-frame through the break, suggests that it might be linked somehow to the overall dsRNA form. It is not known how this nick is generated or maintained, or if it is relevant to transcription (Gibbs *et al.* 2000).

The genomic organisation of the type member of the genus, *Vicia faba endornavirus* (VFV), is shown below in Figure 1.5.

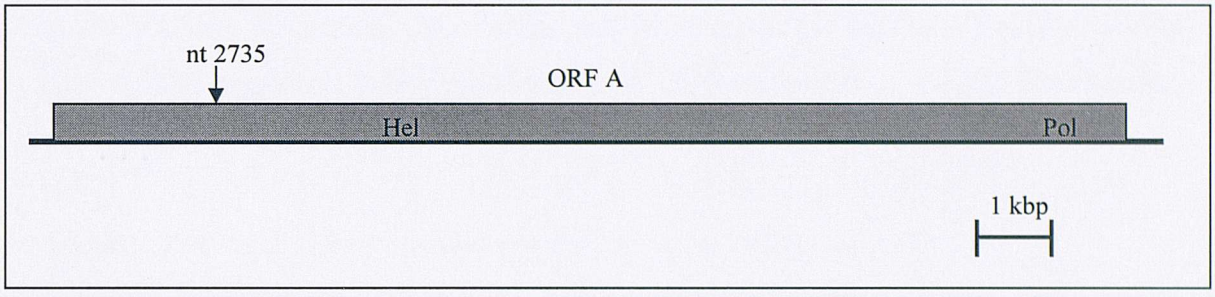


Figure 1.5: Genomic organisation of the *Vicia faba endornavirus*. The viral genome is monopartite consisting of one ORF. The ORF is 17475 bp long, containing single putative RNA helicase (Hel) and RdRp motifs. The ORF is represented by the shaded box, while the 5'- and 3'-untranslated regions are shown as a straight line. The site and the nucleotide position of the nick within the ORF on the coding strand is shown by a vertical arrow.

1.7 Aims of the project

The aim of this part of the thesis is the characterisation of novel dsRNA elements found in fungal hosts or related to known dsRNA mycoviruses. The aims of the projects are presented below in chapter order.

Chapter 3: Amasya cherry disease associated dsRNA elements.

Amasya cherry disease (ACD) is associated with twelve dsRNA elements, six of which have already been characterised (Coutts *et al.* 2004; Covelli *et al.* 2004). The aim was to characterise the remaining dsRNA elements, *via* their complete sequencing (section 3.2). The sequence information was used to determine their origin, function and relation to other known dsRNA elements. A second aim was to compare the sequence data of the ACD dsRNA elements to the equivalent dsRNA species found in the CCRS-diseased tissue (section 3.3). The sequence information was used to design a PCR-based diagnostic test for the detection of ACD and CCRS associated dsRNA elements in cherry tree leaf tissue (section 3.4).

Chapter 4: Phlebiopsis gigantea dsRNA elements

The economically important fungus *Phlebiopsis gigantea* (isolate TW2) contains two dsRNA species that may be useful as a tool for its genetic manipulation. As a prelude to these investigations the aim was to characterise these two dsRNA elements and get a more complete sequencing information, and an investigation for the presence of a viral capsid. The sequence information was used to determine their origin, function and relation to other known dsRNA elements.

Chapter 5: CYAV *dsRNA element and CYSDV RNA 2 cDNA assembly*

Cucurbit yellows-associated virus (CYAV) was found to co-infect cucurbits together with BPYV. The aim was to examine a cDNA library already available and review the sequence of the CYAV dsRNA element *via* partial genomic sequencing. The sequence information was used to determine the origin, function and relation of CYAV to other known dsRNA elements (section 5.1).

CYSDV is a ssRNA plant virus. The aim was to assemble a full-length cDNA of the RNA 2 genomic fragment from the dsRNA replicative intermediate form of the virus, developing a technology that can be applied in the future to dsRNA viruses. The strategy and efforts towards achieving this aim are presented and discussed in section 5.2.

Phytophthora endornavirus 1 (PEV1) is a newly characterised endornavirus (Hacker *et al.* 2005) that contains a UDP glucosyltransferase gene. The initial experiments towards cloning, expressing and assaying the activity of the gene *in vitro* are being described and discussed in section 5.3.

Table 1.1: Taxonomic presentation of dsRNA viruses of fungi. The table provides a list of the characterised dsRNA mycoviruses, classified according to the families and genera they belong to. There are viruses shown assigned to a family, but not to a specific genus within that family, as well as unassigned dsRNA viruses. In addition, dsRNA viruses (though not mycoviruses) that are frequently used in this thesis are included in the Table. They are all shown in alphabetical order together with their taxonomic group of their host. The GenBank accession numbers of their genomic sequence and the publication date are shown in the references list below the table.

Genus	Virus	Genomic fragments	Taxonomic group of the host	Reference number
Family: *Chrysoviridae*				
Chrysovirus	*Amasya cherry disease associated chrysovirus*	4	Putative fungus	1
	Cherry chlorotic rusty spot associated chrysovirus	4	Putative fungus	2
	Helminthsporium victoriae virus 145S (Hv145SV)	4	Filamentous fungus	3
	Penicillium chrysogenum virus (PcV)	4	Ascomycete	4
Family: *Hypoviridae*				
Hypovirus	*Cryphonectria hypovirus 1* (CHV 1)	1	Ascomycete	5
	Cryphonectria hypovirus 2 (CHV 2)	1	Ascomycete	6
	Cryphonectria hypovirus 3 (CHV-3)	1	Ascomycete	7
	Cryphonectria hypovirus 4 (CHV-4)	1	Ascomycete	8
Family: *Narnaviridae*				
Narnavirus	*Saccharomyces narnavirus 20S RNA* (SNV-20S)	1	Ascomycete	9
	Saccharomyces narnavirus 23S RNA (SNV-23S)	1	Ascomycete	10
Unassigned	Rhizoctonia virus M2 (RVM2)	1	Basidiomycete	11

Family: *Partitiviridae*				
Partitivirus	*Amasya cherry disease associated partitivirus*	2	Putative fungus	12
	Atkinsonella hypoxylon virus (AhV)	3	Ascomycete	13
	Ceratocystis polonica partitivirus (CpPV)	2	Ascomycete	14
	Cherry chlorotic rusty spot associated partitivirus	2	Putative fungus	15
	Discula destructiva virus 1	3	Ascomycete	16
	Discula destructiva virus 2	2	Ascomycete	17
	Discula destructiva virus 3	2	Ascomycete	18
	Fusarium poae virus 1 (FpV1)	2	Ascomycete	19
	Fusarium solani virus (FsV1)	2	Ascomycete	20
	Gremmeniella abietina RNA virus MS1	3	Ascomycete	21
	Gremmeniella abietina RNA virus MS2	3	Ascomycete	22
	Helicobasidium mompa partitivirus V1-1	2	Basidiomycete	23
	Helicobasidium mompa partitivirus V1-2	2	Basidiomycete	24
	Heterobasidion annosum partitivirus	2	Basidiomycete	25
	Penicillium stoloniferum virus S (PsV-S)	2	Ascomycete	26
	Rhizoctonia solani virus (RsV)	2	Basidiomycete	27
Unassigned	*Bryopsis cinicola* mitochondrial virus	2	Green alga	28
	Ophiostoma quercus partitivirus	2	Ascomycete	29
	Pleurotus ostreatus virus 1	2	Basidiomycete	30
	Radish partitivirus	2	Plant/ putative fungus	31

Family: *Reoviridae*				
Mycoreovirus	*Mycoreovirus 1* (MYRV1)	11	Ascomycete	32
	Mycoreovirus 2 (MYRV2)	11	Ascomycete	33
	Mycoreovirus 3 (MYRV3)	12	Ascomycete	34
Family: *Totiviridae*				
Giardiavirus	*Giardia lamblia virus* (GLV)	1	Protozoa	35
	Trichomonas vaginalis virus 1 (TVV-1)	1	Protozoa	36
	Trichomonas vaginalis virus 2 (TVV-2)	1	Protozoa	37
	Trichomonas vaginalis virus 3 (TVV-3)	1	Protozoa	38
Leishmania virus	*Leishmania RNA virus 1-1* (LRV1-1)	1	Protozoa	39
	Leishmania RNA virus 1-2 (LRV1-2)	1	Protozoa	40
	Leishmania virus 1-4 (LRV1-4)	1	Protozoa	41
	Leishmania RNA viruses 1-7 (LRV1-7), *1-8* (LRV1-8), *1-9* (LRV1-9), *1-10*(LRV1-10), *1-11*(LRV1-11), *1-13*(LRV1-13)	1	Protozoa	42
	Leishmania RNA virus 2-1 (LRV 2-1)	1	Protozoa	43
Totivirus	*Coniothyrium minitans mycovirus*	1	Ascomycete	44
	Gremmeniella abietina RNA virus L1 (GaRV-L1)	1	Ascomycete	45
	Gremmeniella abietina RNA virus L2 (GaRV-L2)	1	Ascomycete	46
	Helicobasidium mompa Tanaka virus (HmTV1-17)	1	Basidiomycete	47
	Helminthosporium victoriae virus 190S (HvV190S)	1	Filamentous fungus	48

	Saccharomyces cerevisiae virus L-A (ScV-L-A)	1	Ascomycete	49
	Saccharomyces cerevisiae virus L-BC (ScV-L-BC)	1	Ascomycete	50
	Ustilago maydis virus H1 (UmV-H1)	1	Basidiomycete	51
	Zygosaccharomyces bailii virus Z	1	Ascomycete	52
Unassigned	*Eimeria brunette RNA virus 1* (Eb-RV1)	1	Protozoa	53
	Magnaporthe grisea virus 1	1	Ascomycete	54
	Ophiostoma minus totivirus (isolate H2392)	1	Ascomycete	55
	Sphaeropsis sapinea RNA virus 1 (SsRV-1)	1	Ascomycete	56
	Sphaeropsis sapinea RNA virus 2 (SsRV-2)	1	Ascomycete	57
	Chalara elegans (Thielaviopsis basicola) RNA virus 1 (CeRV1)	1	Ascomycete	58
	Chalara elegans (Thielaviopsis basicola) RNA virus 2 (CeRV2)	1	Ascomycete	59
	Unnamed *Ribes* (blackcurrant) totivirus	1	Plant/ putative fungus	60
Unassigned dsRNA mycoviruses				
	Agaricus bisporus L1, L5 dsRNA	1	Basidiomycete	61
	Agaricus bisporus virus X	3	Basidiomycete	62
	Pyrus pyrifolia (Japanese pear) dsRNA	1	Eudicotyledonous plant	63
	Melampsora lini B3 dsRNA	1	Basidiomycete	64
	La France isometric virus (LIV)	9	Basidiomycete	65

References: **1.** Covelli *et al.* 2004, (GenBank accession number) AJ781163-6 **2.**Covelli *et al.* 2004, AJ781397 -AJ781400 **3.** Soldevila *et al.* 2000, AF297176-9 **4.** Caston *et al.* 2003, AF296439-AF296442 **5.** Chen and Nuss 1999, M57938 **6.** Hillman *et al.* 2004, L29010 **7.** Smart *et al.* 1999, AF188514-5 **8.** Linder-Basso *et al.* 2005, AY307099 **9.** Rodriguez-Cousino *et al.* 1998, AF039063 **10.** Rodriguez-Cousino *et al.*1998, U90136 **11.** Lakshman *et al.* 1998, U51331 **12.** Coutts *et al.* 2004, AJ781167-8 **13.** Oh and Hillman 1995, L39125 **14.** Marin *et al.*

unpublished, AY247204-5 **15.** Coutts *et al.* 2004, AJ781401-2 **16.** Rong *et al.* 2002, AF316992-5 **17.** Rong *et al.* 2002, AY033436-7 **18.** Rong *et al.* unpublished, partial sequence AF375469-AF375470 **19.** Compel *et al.* 1999, AF015924 and AF047013 **20.** Nogawa *et al.* 1996, D55668-9 **21.** Tuomivirta and Hantula 2003, AY089993-5 **22.** Tuomivirta and Hantula 2005, AY615211-3 **23.** Osaki *et al.* 2004, partial sequence AB110979 **24.** Osaki *et al.* 2004, partial sequence AB110980 **25.** Ihrmark *et al.* 2001, partial sequence AF473549 **26.** Kim *et al.* 2003, AY156521-2 **27.** Strauss *et al.* 2000, AF133290-1 **28.** Doherty *et al.* 2005, incomplete sequence AM111099 **29.**Koga *et al.* 2003, partial sequence AB070653 **30.** Lim *et al.* 2005, AY533036 and AY533038 **31.** Liu *et al.* unpublished, partial sequence AY748911 **32.** Suzuki *et al.* 2004, AB179636-AB179643 **33.** Suzuki *et al.* 2004, AY277888-AY277890 **34.** Osaki *et al.* 2002, AB073276-AB073283, AB098022-3, AB102674-5 **35.** Wang *et al.* 1993, L13218 **36.** Tai and Ip 1995, U08999 and U57898 **37.** Bessarab *et al.* 2000, AF127178 **38.** Bessarab and Tai unpublished, AF325840 **39.** Stuart *et al.* 1992, M92355 **40.** Zamora *et al.* 2000, partial sequence AF230881 **41.** Scheffter *et al.* 1994, U01899 **42.** Zamora *et al.* 2000, partial sequence for all, AF230882-6 and U01899 **43.** Scheffter *et al.* 1995, U32108 **44.** Cheng *et al.* 2003, AF521012 and AF527633 **45.** Tuomivirta and Hantula 2003, AF337175 **46.** Tuomivirta and Hantula 2005, AY615210 **47.** Nomura *et al.* 2003, AB085814-5 **48.** Huang and Ghabrial 1996, U41345 **49.** Diamond *et al.* 1989, M28353 **50.** Park *et al.* 1996, U01060 **51.** Bruenn 1993, U01059 **52.** Schmitt and Neuhausen 1994, AF224490 **53.** Fraga *et al.* unpublished, AF356189 **54.** Yokoi *et al.* unpublished, AB176964 **55.** Doherty *et al.* 2005, incomplete sequence AM111098. **56.** Preisig *et al.* 1998, AF038665 **57.** Preisig *et al.* 1998, AF039080 **58.** Park *et al.* 2005, AY561500 **59.** Park *et al.* 2005, AY556461 **60.** Cox *et al.* 2000 unpublished, partial sequence AJ133821 **61.** van der Lende *et al.* 1996, partial sequence X94362 **62.** Mikosch unpublished, incomplete sequence AJ421979-AJ421990 **63.** Osaki *et al.* 1998, incomplete sequence AB012616 **64.** Dickinson *et al.* 1993, incomplete sequence X64371 **65.** Harmsen *et al.* 1991 unpublished, incomplete sequence D10828-D10830 and U62640-4

Table 1.2 : List of the *Endornaviridae* viruses. The table provides a list of the four characterised endornaviruses to date. They are shown in alphabetical order with the GenBank accession number of their genomic sequence and the publication date shown below the table.

Genus	Virus	Genomic fragments	Taxonomic group of the host	Reference number
Family : *Endornaviridae*				
Endornavirus	*Oryza rufipogon endornavirus* (ORV)	1	Monocotyledonous plant	1
	Oryza sativa endornavirus (OSV)	1	Monocotyledonous plant	2
	Phaseolus vulgaris endornavirus (PVuV)	1	Eudicotyledonous plant	3
	Phytophthora endornavirus 1 (PEV1)	1	Oomycete	4

	Vicia faba endornavirus (VFV)	1	Eudicotyledonous plant	5
Unassigned	Melon endornavirus (Melend1 and Melend2)	1	Eudicotyledonous plant	6

References : **1.** Moriyama *et al.* 1999, AB014344 **2.** Fukuhara *et al.* 1993, D32136 **3.** Wakarchuk and Hamilton 1990, X16637 **4.** Hacker *et al.* 2005, AJ877914 **5.** Pfeiffer *et al.* 1998, AJ000929 **6.** Coutts 2005, partial sequence, AJ968730.

CHAPTER TWO

Characterisation of novel, fungal dsRNA viral elements associated with the Amasya cherry disease and *Phlebiopsis gigantea*

Genome-wide mapping of hypersensitive sites in *Drosophila melanogaster*

Materials and methods

2.1 Working conditions and materials

2.1.1 Safety considerations

Radioactive isotopes were handled behind a perspex screen wearing disposable gloves; areas were monitored with a hand-held Geiger-Müller counter. Personal exposure to radioactive material was monitored by a film-badge, which was developed and checked every three months. All radioactive waste was disposed off in accordance with Imperial College Safety Regulations.

A facemask was worn when weighing out hazardous powders or dispensing toxic or volatile liquids. Exposure to UV-radiation was minimised by wearing a full-face visor in conjunction with laboratory coat and gloves.

2.1.2 Sterilisation and precautions of working with RNA

All plastic, glassware and solutions were autoclaved at 15 psi (1.03 bar), 121°C for 20-30 min. If the solutions were heat-sensitive, they were sterilised by passage through a disposable 0.2μm syringe filter (Schleicher & Schuell). Solutions to be used with RNA were prepared with autoclaved H_2O.

2.1.3 Centrifugation

Volumes less than 2 ml were centrifuged in an Eppendorf 5415 C benchtop centrifuge. Centrifugation of larger volumes was performed in a Beckman J2-21 centrifuge or a Beckman Ultima L-100 XP ultracentrifuge.

2.1.4 Bacterial strains and cloning vectors

The following *Escherichia coli* strains were used (Promega and Stratagene respectively):

JM109	*end*A1, *rec*A1, *gyr*A96, *thi, hsd*R17 (r_k-, m_k+), *rel*A1, *sup*E44, Δ(*lac-pro*AB)
BL21	B F^-, dcm, ompT, hsdS(r_B^- m_B^-) gal

Bacterial strains were stored in the short term at 4°C, while storage for longer periods was performed in 15% glycerol aliquots at -70°C.

PCR products were cloned into either a pBluescript II SK(+) vector (Stratagene) or a pGEM-T Easy plasmid vector (Promega). PCR products for protein expression were inserted into a pET28b plsmid vector (Novagen).

2.1.5 Bacterial media and antibiotics

Bacteria were typically grown at 37°C with shaking at 225 rpm in liquid LB media or on solid plates containing 15 g/l agar (OXOID). If antibiotic selection was required, ampicillin or kanamycin (100 μg/ml, Sigma-Aldrich) was added when the temperature of the media was 50-60°C.

LB media:	Bacto-tryptone	5 g/l
	Yeast extract	10 g/l
	NaCl	5 g/l

2.1.6 Fungal media

Phlebiopsis gigantea fungal cultures were typically grown at 22°C on solid plates containing 15 g/l of agar (OXOID) and 15 g/l of malt extract agar (Merck). Stock plates were stored for longer periods of time at 4°C. Plates used for the harvesting of tissue, were overlayed prior to inoculation with cellophane disks (Cellophaneroll D60/2, Whatman) that had been deplasticised by boiling in distilled water for 10 min. The plates with the disks were then left in a sterile environment overnight to dry out, and inoculation took place the following day by placing a small hyphal inoculum in the centre of the plate. The fungal tissue was harvested when the mycelia covered 80% of the plates.

Phytophthora spp. fungal cultures were typically grown at 22°C in liquid cultures, prepared after incubating 125 g of frozen garden peas (BirdsEye) in 1 l of H_2O at 80 °C for 1 h. The resulting liquid was collected, strained, aliquoted into the flasks and autoclaved. The fungal tissue was harvested approximately two weeks after inoculation, when it covered 80% of the surface of the liquid.

2.1.7 Plant material

Chilled samples of ACD infected and healthy leaves were supplied by Serap Açikgöz (Turkey). The leaf tissue was then immediately powdered in liquid nitrogen in a pre-chilled pestle and mortar. The powdered tissue was kept at -80°C.

2.1.8 Cell line material

For the detection and study of hypersensitive sites, the Schneider 2 cell line of *Drosophila melanogaster* was used (Promochem). The cell cultures were grown in canted neck flasks at room temperature in Schneider's insect medium (Sigma) containing, according to the ATCC instructions, 5 - 10% of heat inactivated FBS (v/v).

2.1.9 RNA material

Samples from total ds RNA extracts in 100% ethanol of ACD and CCRS infected leaves were supplied by Serap Açikgöz and Ricardo Flores respectively. The samples were washed once with 70% ethanol, air dried and resuspended in the desired volume of nuclease-free H_2O.

2.2 Basic nucleic acid manipulation

2.2.1 Phenol/chloroform extraction

Phenol was melted at 68°C and then equilibrated five times with an equal volume of 1 M Tris-HCl (pH 8.0) containing 0.1% 2-hydroxy-quinoline, before overlaying with 0.1 M Tris-HCl (pH 8.0). Proteins were removed from solution by vortexing with an equal volume of phenol until an emulsion was formed, and microcentrifugated for 3 min to separate the two phases. The aqueous phase was removed and extracted with an equal volume of Sevag (24:1 chloroform: iso-amyl alcohol), in order to remove the residual phenol.

2.2.2 Nucleic acid precipitation in the presence of sodium acetate

Nucleic acid solutions were precipitated by the addition of one tenth of the volume 3 M sodium acetate (pH 5.5) and 2.5 volumes of chilled ethanol 100%. The solution was thoroughly mixed and allowed to precipitate overnight at -20°C. The precipitation was

centrifuged at 4°C for 20 min. The supernatant was discarded, the pellet was washed in 70% (w/v) ethanol and the mixture was again centrifuged at 4°C for 20 min. The supernatant was discarded and the pellet was left to dry to remove any traces of ethanol. Pellets were resuspended in H_2O.

2.2.3 Non-denaturing agarose gel electrophoresis

Non-denaturing agarose gels were prepared by melting molecular grade agarose (Bioline) to a final concentration of 0.8-2.0% (w/v) in 1x TAE buffer (10x TAE: 400 mM Tris-base, 11.4 ml/l glacial acetic acid, 10 mM EDTA). Ethidium bromide was added to a final concentration of 50 μg/100 ml when the temperature was approximately 55°C. The samples were mixed in 5x loading buffer (5x loading buffer: 50% glycerol, 0.2% bromophenol blue) before loading. The samples were electrophoresed at 30-80 V submerged in 1x TAE buffer. The progress of the electrophoresis was monitored by the migration of the bromophenol blue dye.

2.2.4 Recovery of DNA from agarose gels

DNA was routinely recovered from non-denaturing agarose gels using the QIAEX II kit (QIAGEN). After electrophoresis, the desired band was excised from the gel, its weight was roughly estimated and three volumes of DNA binding salt buffer (QX1) were added. The solution was incubated at 50°C for 5 min in order to melt the agarose and 12 μl of DNA binding matrix (QIAEX II) capable to bind 2.5 μg of DNA were added. After washing twice with the QX1 solution and twice with the PE buffer, the pellet was air-dried. The pellet was resuspended in 20 μl of nuclease free H_2O and incubated at room temperature for 5 min. After centrifugation for 30 sec the clear supernatant was collected. The pellet was resuspended for a second time in 20 μl of nuclease free H_2O and the DNA containing supernatant was extracted as before. The two aliquots were pooled together.

2.2.5 Recovery of dsRNA from agarose gels

For the recovery of dsRNA from non-denaturing agarose gels, the BIO-101 RNaid kit (Anachem) was used. The band of interest was excised from the gel with a minimal exposure to UV light. The weight was roughly estimated and three volumes of RNA

binding salt were added. The suspension was incubated at 50°C for 10 min until the agarose was completely melted. An amount of 10 µl of RNA matrix solution (able to bind 10 µg of dsRNA) was added and further vortexed in order to create a homogeneous solution. The suspension was left to stand at room temperature for 10 min with occasional mixing of the contents to assist adsorption. After washing twice with wash buffer the pellet was resuspended in 65 µl of nuclease free H_2O. The mixture was incubated at 50°C for 5 min and following centrifugation the supernatant was collected. The pellet was resuspended for a second time in 65 µl of nuclease free H_2O and the dsRNA containing supernatant was extracted as before. The two aliquots were pooled together and precipitated overnight in the presence of sodium acetate. The dsRNA was resuspended the following day in the desired volume of nuclease free H_2O.

2.2.6 Small-scale isolation of plasmid DNA (mini-prep)

For the routine small-scale isolation of plasmid DNA (mini-prep), the QIAprep Spin Miniprep kit was used (QIAGEN). Bacterial cultures (1-3 ml), grown from a single colony overnight in a 15 ml sterile tube were transferred to Eppendorf tubes and the cells were pelleted by centrifugation for 1 min. After aspiration of the supernatant, the bacterial cells were resuspended in 250 µl of chilled P1 buffer. The cells were lysed by the addition of 250 µl P2 buffer and inversion of the mixture. When the suspension was clear, 350 µl of neutralisation N3 buffer were added and mixed as above. After centrifugation for 10 min, the supernatant was collected and passed through a QIAprep spin column. The column was washed once with 750 µl of PE buffer and the DNA was eluted in 50-70 µl of nuclease free H_2O by centrifugation for 1 min.

2.2.7 Restriction of DNA fragments

Restriction of DNA fragments was carried out in the appropriate buffer, with 5 U of restriction enzyme used per 500 µg of DNA, and by incubation at the optimal temperature for 1-2 h. Restriction enzymes were heat inactivated at 65°C for 20 min or after phenol/chloroform extraction. Digestions were checked by DNA electrophoresis on 1% agarose gels prior to any further manipulations.

2.2.8 Cleaning of DNA fragments

For the routine cleaning of PCR reactions the QIAquick Purification kit was used (QIAGEN). Five volumes of PB binding buffer were added to the sample and mixed by inversion. The mixture was passed through a QIAquick spin column and the column was washed once with 750 µl of PE buffer. The DNA was eluted in 35-50 µl of nuclease free H_2O by centrifugation for 1 min.

2.3 Extraction of total RNA from plants

For the total RNA extraction from powdered plant tissue the RNeasy Plant Mini kit was used (QIAGEN). Approximately 100 mg of plant tissue were mixed with 450 µl of buffer RLT to disrupt the tissue, until the mixture had a homogeneous appearance. The mixture was passed through a QIAshredder spin column, removing cell debris and further homogenising the sample. 0.5 volume of chilled 100% ethanol was added to the lysate and was mixed immediately by pipetting. The sample was then passed through a RNeasy mini column. The column was washed once with 700 µl of buffer RW1 and twice with 500 µl of buffer RPE. The dsRNA was eluted in 30-60 µl of nuclease free H_2O by centrifugation for 1 min.

2.4 Extraction of viral dsRNA

2.4.1 Extraction of viral dsRNA from fungal tissue (LiCl fractionation)

This protocol was adapted from Hull and Covey (1983) who used this method for the extraction of RNA transcripts of *Caulilflower mosaic virus* from turnip leaves.

Fungal mycelia were harvested and were ground under liquid nitrogen. The resulting powder was emulsified with 1 ml per 1 g of tissue with extraction buffer (50 ml/l Tris-HCl buffered phenol, 60 g/l 4-amino salycilate, made up to the desired volume with autoclaved 50 mM Tris-HCl, pH 8.4). After vigorous shaking on ice for 1 h the mixture was centrifuged for 40 min at 10,000 rpm (15300 *g*). The aqueous supernatant containing the total nucleic acids was transferred into a sterile centrifuge tube (Beckman) and one phenol:chloroform extraction was followed by two chloroform extractions to

deproteinase the nucleic acids. The mixture was centrifuged each time at 10,000 rpm for 30 min.

Following the extraction series, the supernatant was transferred to a sterile centrifuge tube and LiCl fractionation was carried out according to Diaz-Ruiz and Kaper (1978). The solution of total nucleic acids was added to an equal volume of 4 M LiCl and incubated at 4°C for 16 h. The solution was then centrifuged at 10,000 rpm (15300 *g*) for 30 min to pellet the precipitated higher molecular weight ss RNA.

The supernatant was transferred into a fresh, sterile centrifuge tube and an equal volume of 8 M LiCl was added to the tube. The solution was then incubated at 4°C for 16 h. The dsRNA precipitate was collected by centrifugation at 10,000 rpm for 30 min. The resulting dsRNA pellet was washed once with cold 70% (w/v) ethanol and was left to dry. The dsRNA was then completely resuspended in nuclease free H_2O, using approximately 20 µl of H_2O used per g of starting tissue. Purification of the dsRNA solution was then carried out using DNase I digestion and dsRNA extraction from a non-denaturing agarose gel.

2.4.2 Extraction of viral dsRNA from plant tissue (CF-11 extraction)

The protocol used was adapted from Coffin and Coutts (1992) for the extraction of dsRNA from plant tissue. Approximately 100 g of plant tissue were ground to a fine powder under liquid nitrogen. A mixture of 150 ml GPS (GPS buffer 5x: 87.66 g NaCl, 37.54 g glycine, 35.49 g Na_2HPO_4, [pH 9.5]), 150 ml GPS-saturated phenol, 150 ml Sevag (24:1 chloroform: isoamyl alcohol), 15 ml 10% SDS and 1.5 ml β-mercaptoethanol was added directly to the powder and was left to stir for 1 h at 4 °C. The mixture was then strained through a double layer of muslin in order to remove solid plant material, and was then centrifuged for 10 min at 10000 rpm (15300 *g*) at 4 °C. The aqueous phase was recovered and ethanol (96%) was added to a final volume of 17%, followed by the addition of CF-11 fibrous cellulose powder (Whatman), the dsRNA binding to the cellulose in the presence of ethanol. The mixture was stirred for 1 h at

room temperature to allow for the dsRNA to bind and was then centrifuged for 10 min at 15300 *g*.

The resulting cellulose pellet was resuspended into 50 ml Falcon tubes using *ca.* 45 ml of STE buffer with 17% ethanol (STE buffer 10x: 29.22 g NaCl, 30.29 g Tris-HCl, 1.86 g EDTA, 500 ml H_2O, [pH 7.5]). This was shaken for 5 min and the cellulose was re-pelleted after centrifuging at 2000 *g* for 2 min. This washing step was repeated 4 times. The dsRNA was eluted in 4 steps using 20 ml of STE buffer (without ethanol) and centrifuged as before, with the final elution being left for 1 h to maximise the yield. The collected supernatants were centrifuged for 30 min at 23700 *g* to remove any residual cellulose. The resulting supernatant was precipitated overnight in ethanol in the presence of sodium acetate. The dsRNA was resuspended the following day in the desired volume of nuclease free H_2O and was run on an 1% agarose gel in order to assess the quality of the dsRNA preparation.

2.5 Extraction of viral dsRNA from 50% glycerol stock suspensions of viruses

An aliquot from the stock suspension containing *Aspergillus foetidus* viruses F and S (povided by Prof. Buck) was diluted with nuclease free H_2O to 4% glycerol and SDS was added to 0.1% of the final volume. The mixture was incubated at 60°C for 20 min to disrupt the viral particles. The dsRNA was then extracted twice with phenol, followed by two Sevag extractions; dsRNA was then precipitated overnight in the presence of sodium acetate. The dsRNA was resuspended the following day in the desired volume of nuclease free H_2O.

2.6 DNase I digestion of isolated Schneider 2 cell nuclei

The isolated Schneider 2 cell nuclei were DNase I digested according to the following four protocols. The first protocol was the most commonly used one.

2.6.1 This protocol was adapted from Carey and Smale (2000) who used this method for the indirect labelling of hypersensitive sites in *D. melanogaster*.

The cells were harvested at OD 0.6 at 600 nm and centrifuged for 10 min at 1000 rpm (167 *g*) at 4°C. The supernatant was discarded and the pellet was washed once with ice-cold 1x PBS buffer (PBS buffer 1x: 0.15 M NaCl, 2.5 mM KCl, 40 mM Na_2HPO_4, 20 mM KH_2PO_4) and the cells were repelleted for 10 min at 1000 rpm (167 *g*) at 4°C. The pellet was resuspended in 2.5 ml ice-cold NP-40 lysis buffer (NP-40 lysis buffer: 10 mM Tris-HCl [pH 7.5], 10 mM NaCl, 3 mM $MgCl_2$, 0.15 mM spermine, 0.5 mM spermidine and 0.5% (v/v) NP-40 added dropwise to the stirred solution, addition of 300 mM sucrose to the solution is optional) and incubated on ice for 5 min. The nuclei were pelleted for 10 min at 1000 rpm (167 *g*) at 4°C and resuspended in 1.4 ml Buffer A (Buffer A: 100 mM NaCl, 50 mM Tris-HCl [pH 8.0], 3 mM $MgCl_2$, 0.15 mM spermine and 0.5 mM spermidine). They were repelleted for 10 min at 1000 rpm (167 *g*) at 4°C and subsequently resuspended in 500 µl Buffer A containing 1mM $MgCl_2$. 100 µl of the sample were transferred into tubes containing different dilutions of DNase I (1 U/µl, Promega). The samples were incubated at 37°C or on ice, for longer periods of time.

To stop the reaction 2 µl of 0.5 M EDTA and 100 µl of Buffer A were added to each tube followed immediately by 3 µl proteinase K (25 mg/ml, Fermentas) and 10 µl of 20% SDS. The DNA from the nuclei samples was phenol:chloroform extracted and precipitated overnight in the presence of sodium acetate.

2.6.2 This protocol was adapted from Aasland and Stewart (1999) who used this method for the *in situ* genomic digest of permeabilised cells.

The appropriate amount of diluted DNase I was added to tubes containing 2 ml NP-40 buffer (NP-40 buffer: 15 mM Tris-HCl [pH 7.5], 60 mM KCl, 15 mM NaCl, 0.2 mM EDTA, 0.2mM EGTA, 0.15 spermine and 0.5 mM spermidine) and mixed by pippeting. The culture medium was removed from the flasks by aspiration and for each 6 cm^2 of tissue culture, 1 ml of the NP-40/DNase I solution was added. The flasks were then

incubated for the desired amount of time at room temperature. The reaction was quenched by the addition of 0.2 ml 6x lysis buffer (6x lysis buffer: 6% SDS, 300 mM Tris-HCl, 120 mM EDTA). The flasks were swirled gently to mix the reagents and were left for 5 min at room temperature. 12 µl of proteinase K (20 mg/ml) were subsequently added to the solution and the flasks were incubated overnight at 37°C. The DNA from the samples was phenol:chloroform extracted and dialysed overnight in TE buffer (1x TE buffer: 10 mM Tris-HCl, 0.1 mM EDTA) at 16°C. If necessary the DNA was precipitated overnight in the presence of sodium acetate.

2.6.3 This protocol was adapted from Jack *et al.* (1991) who used this method for the DNase I digests of *D. melanogaster* embryo nuclei.

The cell nuclei are resuspended as described in section 2.6.1. The nuclei are resuspended in digestion buffer (digestion buffer: 50 mM Tris-HCl [pH 7.5], 50 mM NaCl, 4 mM $MgCl_2$), in an amount that is equal to the starting volume of the cell culture, at 37°C in the presence of 20 U/ml of DNase I. They are left for up to 2 min and the digests are terminated by the addition of EDTA to 25 mM, SDS to 0.5% and proteinase K (20 mg/ml) to 1 mg/ml. The samples are subsequently incubated at 37°C overnight. The DNA from the nuclei is then phenol:chloroform extracted and precipitated overnight in the presence of sodium acetate.

2.6.4 This protocol was adapted from Talbot *et al.* (1990) who used this method for the DNase I digest of human cell cultures.

The cell nuclei are resuspended as described in section 2.6.1. The nuclei are resuspended in 0.5 ml of DNase I digestion buffer (DNase I digestion buffer: 15 mM Tris-HCl [pH 7.5], 60 mM KCl, 15 mM NaCl, 0.2 mM EDTA, 0.2 mM EGTA, 0.15 mM spermine and 0.5 mM spermidine). DNase I (10 ng/ml, Sigma) is added to concentrations of 1 – 8 µg/ml and the reaction is initiated by the addition of $MgCl_2$ to 10 mM and $CaCl_2$ to 1 mM. The reaction is allowed to proceed for 5 min on ice and is stopped by the addition of equal volume of a 50 mM EDTA and 1% SDS containing solution. The DNA from the

digested nuclei is then phenol:chloroform extracted and precipitated overnight in the presence of sodium acetate.

2.7 Hybridisation of nucleic acids from non-denaturing agarose gels

2.7.1 Nucleic acid blotting

Non-denaturing agarose gels were electrophoresed until the bromophenol blue dye reached 2/3 the length of the gel. The nucleic acids were nicked by incubating the gel in 0.25 N for 15 min or until the bromophenol blue dye turned yellow. The nucleic acids were then denatured by soaking the gel in 50 mM NaOH for 30 min and were twice neutralised in 2x TBE (1xTBE: 90 mM Tris-base, 90 mM boric acid, 2 mM EDTA) for 15 min each time. The gel was reversed and was placed on top of three sheets of 3 mm Whatman paper (Whatman), which served as a bridge with the blotting buffer (20x SSC: 3 M NaCl, 0.3 M Na_2CO_3 [pH 7]). On top of the gel were placed in order a Hybond$^+$ membrane (Amersham) pre-wetted briefly in 2x SSC, three pieces of 3 mm Whatman paper cut exactly on the same size of the gel and paper towels 5 cm thick. The blot was covered with SaranWrap^{TM} film (DuPont) in order to reduce evaporation and a weight of 500 g was placed on top. The nucleic acids were capillary transferred onto the membrane overnight. After dismantling the assembly, the membrane was briefly rinsed in 2x SSC and was left to dry. Nucleic acids were fixed on the membrane by exposure to UV radiation for 60 sec using a Stratalinker^{TM} 1800 (Stratagene) at 120 mj/cm^2. Pre-hybridisation of the membrane was carried out in a Hybaid hybridisation oven (Amersham) in 5 ml of pre-hybridisation solution at 65°C for at least 3 h. The three first pre-hybridisation solutions were used interchangeably, the fourth one was used according to the manufacturers instructions.

Pre-hybridisation solutions:

i) 1.25 ml 20x SSPE (3 M NaCl, 0.2 M NaH_2PO_4, 0.02 M EDTA [pH 7.4])
0.50 ml 50x Denhardt's solution (for 0.5 l: 5 g Ficoll Type 400, 5 g PVP, 5 g BSA)
0.25% SDS

100 µl sonicated single stranded salmon sperm (1 mg/ml) [NBL]
up to 5 ml H_2O

ii) PerfectHyb™ Plus hybridization buffer (Sigma)

iii) Church buffer: 0.5 M NaH_2PO_4 (ph 7.2)
7% SDS
10 mM EDTA

iv) Rapid-hyb™ hybridization buffer (Amersham)

2.7.2 Radioactive oligolabelling of DNA

DNA cloned fragments were [α^{32}]P oligolabelled using the method described by Feinberg and Vogelstein (1983). 5 µl of the plasmid (20 µg/µl final concentration) was heat denatured at 95°C for 3 min. This was then placed immediately on ice for 2 min. The denatured DNA was mixed to the oligolabelling mix and 1 µl (5 U/µl) of the Klenow fragment (large) of DNA polymerase I (Promega) was added . The mixture was incubated at room temperature for 60 min. The probe was heat denatured at 95°C for 3 min and an aliquot (typically 5 µl, 10% of the total volume of the oligolabelling mixture) was added in the hybridisation tube containing the pre-hybridisation solution and the membrane. The blot was left to hybridise at 65°C overnight. The remainder of the mixture was stored in an appropriate container at -20°C.

Oligolabelling mix:

Oligolabelling buffer (5x, Promega)	10 µl
dGTP, dATP, dTTP (500 µM, Promega)	2 µl each
Nuclease free BSA (10 µg/µl, Promega)	2 µl
[α^{32}]P dCTP (50 µCi/mmol, Amersham)	2 µl
H_2O	24 µl

2.7.3 Radioactive nick-translation labelling of DNA

Cloned DNA fragments were [α^{32}]P nick-translation labelled using either one of the two following methods.

i) According to the method described by Sambrook *et al.* (1990), 5 µl of the plasmid (5 µg/µl final concentration) were mixed, on ice, with 2.5 µl nick translation buffer (10 x: 0.5 M Tris-HCl [pH 7.2], 0.1 M $MgSO_4$, 1 mM DTT, 500 µg/ml BSA), 2 µl of each of dATP, dTTP and dGTP (500µM), 2 µl of [α^{32}]P dCTP (50 µCi/mmol) and 6 µl of H_2O. 2.5 µl of diluted DNase I (10 ng/ml, Sigma). Then 0.5 µl of *E.coli* DNA polymerase 1 (10 U/µl, Invitrogen) were added to the mixture and it was incubated at 16°C for 1 h. The reaction was stopped by the addition of 1 µl 0.5 M EDTA.

ii) Nick-translation Kit (dCTP) (Amersham).

2.7.4 Primer [α^{32}]P labelling

2.5 µl of a primer (8 µM) were mixed on ice with 2.5 µl [α^{32}]P ATP (50 µCi/mmol, Amersham), 1µl of 10x T4 PNK buffer (10 x: 0.5 M Tris-HCl [pH 7.6], 100 mM $MgCl_2$, 100 mM β-mercaptoethanol), 1 µl of T4 PNK (10 U/µl, Promega) and 3 µl of H_2O. The mixture was incubated on ice for 20 min, at 37°C for 2 min and then at 90°C for 2 min. The unincorporated nucleotides were removed by using a pre-equilibrated Sephadex G-25 spin micro-column (Amersham).

2.7.5 Washing of the membranes and autoradiography

After hybridisation, the membranes were washed twice in 2x SSPE, 0.1% SDS at 65°C for 20 min, once in Ix SSPE, 0.1% SDS at 65°C for 20 min and once in 0.1x SSPE, 0.1% SDS at 65°C for 20 min. The membranes were wrapped in SaranWrap™ and were exposed overnight or for longer periods of time to X-ray film (Kodak) at -20°C, using a cassette with an intensifying screen.

2.8 PCR amplification from total genomic DNA

The EasyStart™ pre-aliquoted master mix tubes (MBio) were used for the PCR amplification of fragments from genomic DNA. 100 ng of template DNA were mixed on ice in the tube with 25 pmol of primers, H_2O and 0.5 µl of Biotaq DNA polymerase (5 U/µl, Bioline). The reaction was run in a thermocycler (Hybaid) as follows:

i) For the amplification of DNA fragments

94°C for 5 min	1 cycle
94°C for 2 min/ 55°C for 1 min/ 72°C for 2 min	32 cycles
72°C for 10 min	1 cycle

ii) For the extension of a sequence using a single [α^{32}]P labelled primer

94°C for 5 min	1 cycle
94°C for 1 min/ 55°C for 1 min/ 72°C for 1 min	32 cycles
72°C for 10 min	1 cycle

In each case one tenth of the PCR reaction was electrophoresed in a 1% non-denaturing agarose gel in order to check the amplified DNA products.

2.9 Reverse transcription polymerase chain reaction (RT-PCR) from total RNA

2.9.1 Amplification of known sequences

A 7 µl aliquot of total dsRNA extract was denatured by mixing with 2 µl of CH_3HgOH (100 mM, SERVA) and 1 µl (100 µM) of a sequence specific primer, oligo 1. It was incubated at room temperature for 20 min, and it was then chilled on ice for 1 min.

For the first strand cDNA synthesis, the denatured dsRNA was added to the first strand cDNA synthesis reaction mixture, that was pre-heated at 50-55°C for 1 min. The mix was incubated at 50-55°C for 1 min and then 1 µl (200 U/µl) of Superscript™ III RNase H- Reverse Transcriptase (Invitrogen) was added. The reaction was incubated at 50-55°C for 1 h.

First strand cDNA synthesis reaction mix:

Reverse transcription buffer (5x, Invitrogen)	10 µl
DTT (100 mM, Invitrogen)	4 µl
RNasin™ Ribonuclease inhibitor (40 U/µl, Promega)	1.25 µl
dNTP mix (20 mM, Promega)	1 µl
H_2O	23.75 µl

The solution was then size-fractionated and concentrated using an Amicon YM-30 column (Millipore). The volume of the first strand cDNA solution was increased to 500 µl with H_2O and centrifuged in the column at 2000g for 10 min, resulting in the recovery of *ca.* 150 µl of solution. A PCR reaction was then set using a third of the volume of the first strand cDNA synthesis reaction.

PCR mixture:

First strand cDNA	50 µl
PCR reaction buffer (10x, Bioline)	10 µl
$MgCl_2$ (50 mM)	4 µl
Oligo 1 (100 µM)	1 µl
Oligo 2 (100 µM)	1 µl
dNTP mix (20 mM, Promega)	1.2 µl
H_2O	32.3 µl
BioTaq polymerase (5 U/µl, Bioline)	0.5 µl

The reaction was run in a PCR thermocycler (Hybaid) as follows:

96°C for 5 min	1 cycle
96°C for 30 sec/ 50°C for 30 sec/ 72°C for 2 min	32 cycles
72°C for 10 min	1 cycle

One tenth of the PCR reaction was electrophoresed in a 1% non-denaturing agarose gel in order to check the amplified DNA products.

2.9.2 Extension of known sequences

A 7 µl aliquot of total dsRNA extract was denatured by mixing with 2 µl of CH_3HgOH (100 mM, SERVA) with 1 µl of a sequence specific primer, oligo 1, which is designed at *ca.* 100 nucleotides distance from the end of the known sequence. It was incubated at room temperature for 20 min, and it was then chilled on ice for 1 min.

The first strand cDNA synthesis was performed as described previously (section 2.7.1). The solution was then size fractionated and concentrated using an Amicon YM-30 column. A PCR reaction was set up using a third of the volume of the first strand cDNA synthesis reaction.

PCR mixture:

First strand cDNA	50 µl
PCR reaction buffer (10x)	10 µl
$MgCl_2$ (50 mM)	4 µl
Oligo 1 (100 µM)	1 µl
dNTP mix (20 mM)	1.2 µl
H_2O	33.3 µl
BioTaq polymerase	0.5 µl

The reaction was run in a PCR thermocycler as follows:

96°C for 5 min	1 cycle
96°C for 30 sec/ 42°C for 45 sec/ 72°C for 2 min	3 cycles
96°C for 30 sec/ 55°C for 45 sec/ 72°C for 2 min	32 cycles
72°C for 10 min	1 cycle

One tenth of the PCR reaction was electrophoresed in a 1% non-denaturing agarose gel in order to check the amplified DNA products.

2.10 RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE)

This protocol was adapted from Liu and Gorovsky (1993) and Coutts and Livieratos (2003). It involves the ligation of an oligo (LIG-REV) to the 3' end of the dsRNA molecule, followed by RT-PCR amplification. A 7 µl aliquot of a total dsRNA extract was heat denatured at 90°C for 2 min. in the presence of 1µl of the LIG-REV oligo (5'-PO_4 kinated, 3' –OH blocked, 100 µM). The solution was then chilled on ice for 1 min and added to the ligation reaction mixture and 100 U of T4 RNA ligase were added. The reaction was incubated for 16 h at 17°C.

Ligation reaction mixture:

$MgCl_2$ (2 M)	1 µl
DMSO (Sigma)	10 µl
Rnasin™ Ribonuclease inhibitor	2.5 µl
HEPES (2 M, pH 7.5)	2.5 µl
DTT (100 mM, Invitrogen)	5 µl
ATP (100 mM)	1 µl
H_2O	65 µl
T4 RNA ligase (20 U/ml, NEB)	5 µl

The solution was size fractionated and concentrated using an Amicon YM-30 column as described previously (section 2.9.1). It was then precipitated overnight in the presence of sodium acetate. The dsRNA was resuspended the following day in 7 µl nuclease free H_2O. It was denatured at room temperature for 20 min in the presence of 2 µl CH_3HgOH (100 mM) and 1 µl (100 µM) of the LIG-FOR oligo, whereafter the first strand cDNA synthesis was performed as described previously. The solution was size fractionated and concentrated using and Amicon YM-30 column and a third of the volume was used for the PCR reaction.

PCR mixture:

First strand cDNA	50 µl

PCR reaction buffer (10x)	10 µl
$MgCl_2$ (50 mM)	4 µl
LIG-FOR (100 µM)	1 µl
Oligo 1 (100 µM)	1 µl
dNTP mix (20 mM)	1.2 µl
H_2O	32.3 µl
BioTaq polymerase	0.5 µl

The reaction was run in a PCR thermocycler as follows:

96°C for 5 min	1 cycle
96°C for 30 sec/ 50°C for 30 sec/ 72°C for 2 min	32 cycles
72°C for 10 min	1 cycle

One tenth of the PCR reaction was electrophoresed in a 1% non-denaturing agarose gel in order to check the amplified DNA products. In the case of multiple fragments resulting from the PCR reaction, the following, more stringent PCR program was used:

96°C for 5 min	1 cycle
96°C for 30 sec/ 52°C for 45 sec/ 72°C for 2 min	32 cycles
72°C for 10 min	1 cycle

2.11 cDNA library construction

For the cloning and sequencing of individual RNA components cDNA libraries were generated by random priming as described by Coffin and Coutts (1992).

A 7 µl aliquot of total dsRNA extract was denatured by mixing with 2 µl of CH_3HgOH (100 mM, SERVA) in the presence of 2 µl of random hexamer primers (0.5 µg/µl, Promega). It was incubated at room temperature for 20 min, and then chilled on ice for 1 min.

The denatured dsRNA was added to the first strand cDNA synthesis reaction mixture. The mix was incubated at 25°C for 10 min to allow the random hexamers to bind efficiently to the dsRNA before the first strand synthesis is initiated. It was then heated at 42°C for 1 min and then 1 µl (200 U) of Superscript™ II RNase H- Reverse Transcriptase (Invitrogen) was added. The reaction was incubated at 42°C for 1 h.

First strand cDNA synthesis reaction mix:

Reverse transcription buffer (5x)	10 µl
DTT (100 mM)	4 µl
RNasin™ Ribonuclease inhibitor	1.25 µl
dNTP mix (20 mM)	1 µl
H_2O	23.75 µl

The first strand synthesis reaction (50 µl) was added to the second strand reaction mixture, prepared on ice, and was incubated at 16°C for 2 h. Then 2 µl of T4 DNA polymerase (5 U/µl, Invitrogen) were added to the reaction, which was left to incubate for a further 5 min. The reaction was stopped by a phenol/Sevag extraction and was followed by overnight precipitation in the presence of sodium acetate.

Second strand cDNA synthesis reaction mix:

Second strand buffer (5x, Invitrogen)	30 µl
dNTP mix (10 mM)	3 µl
E.coli DNA ligase (10 U/µl, Invitrogen)	1 µl
E.coli DNA polymerase I (10 U/µl, Invitrogen)	4 µl
E.coli RNAse H (2 U/µl, Promega)	1 µl
H_2O	61 µl

The pellet was resuspended the following day in 6 µl of nuclease free water and phosphorylated. 1 µl of PNK buffer (10x, Promega) and 1 µl dATP (10 mM, Promega) were added to the resuspended pellet and the mix was heated at 70°C for 5 min. It was then placed on ice, 2 µl of T4 PNK (5-10 U/µl. Promega) were added to it and it was

incubated at 37°C for 30 min. The reaction was terminated by heating at 68°C for 10 min and was cleaned using the QIAquick purification kit as described earlier (section 2.2.8). The product was eluted in 30 µl of H_2O. An adenine residue was added as follows to the phosphorylated DNA product allowing for its ligation to the pGEM-T Easy vector. The 30 µl of phosphorylated dsDNA were added to the 'A' tailing reaction mixture shown below.

'A' tailing reaction mixture:

Taq Polymerase buffer (10x, Promega)	4 µl
dATP (10 mM)	2 µl
$MgCl_2$ (50 mM)	1 µl
H_2O	2.8 µl
Taq DNA Polymerase (5 U/µl, Promega)	0.2 µl

The solution was incubated at 72°C for 20 min and was cleaned using the QIAquick purification kit. The dsDNA fragments were eluted in 30 µl of H_2O and half of this was used for a ligation reaction into the pGEM-T Easy vector.

2.12 Cloning of PCR products

2.12.1 Preparation of competent JM109 cells

A single colony of competent cells was picked and grown overnight at 37°C in 5 ml of LB medium. 0.5 ml of this culture was then added into 50 ml of LB, and the culture was incubated at 37°C with vigorous shaking (250 rpm) until the OD was 0.4. The cells were placed on ice for 5 min, they were centrifuged for 10 min at 3,000 rpm at 4°C and were resuspended in 5 ml of Transformation Storage Solution [TSS: 50% (v/v) 2x LB, 10% (w/v) PEG_{8000}, 50 mM $MgCl_2$, 5% (v/v) DMSO, filter sterilised]. The cells were then immediately aliquoted in and frozen in liquid nitrogen. The aliquots were stored at -80°C.

2.12.2 Preparation of competent BL21 cells

A single colony of competent cells was picked and grown overnight at 37°C in 5 ml of LB medium. 0.5 ml of this culture was then added into 10 ml of LB, and the culture was

incubated at 37°C with vigorous shaking (250 rpm) until the OD_{600} was 0.4. 1 ml of this culture was then added to 9 ml of LB, and the culture was incubated at 37°C with vigorous shaking (250 rpm) until the OD_{600} was 0.3. The cells were then put for 10 min on ice. They were centrifuged for 10 min at 4°C at 5000 rpm. The pellet was resuspended in 5 ml of 0.1 M $CaCl_2$ and was left for 20 min on ice. The cells were centrifuged for 10 min at 4°C at 5000 rpm. The pellet was resuspended in 1ml of 0.1 M $CaCl_2$ containing 15% glycerol. The cells were then immediately aliquoted in and frozen in liquid nitrogen. The aliquots were stored at -80°C.

2.12.3 Preparation of DNA fragments for ligation

DNA fragments of interest were digested with the appropriate restriction enzyme and electrophoresed in a non-denaturing agarose gel to confirm the success of the digestion reaction. After phenol/chloroform extraction and ethanol precipitation, DNA fragments with compatible ends were ligated in 1x T4 ligation buffer (10x ligation buffer: 300 mM Tris-HCl [pH 7.8], 100 mM $MgCl_2$, 100 mM DTT and 10 mM ATP) and 1 µl (0.5 U/µl) of T4 DNA ligase enzyme (Promega). The ligation reaction was incubated at 4°C for 16 h.

Alternatively, PCR products were ligated into pGEM-T Easy vector (Promega). After gel recovery (section 2.2.4) approximately 250 ng PCR DNA product, ligations were set up in a total volume of 10 µl containing 1x T4 ligation buffer (as above), 50 ng of pGEM-T Easy plasmid and 2 U of T4 ligase. Ligation reactions were incubated at 4°C for 16 h.

2.12.4 Transformation of competent cells

The competent cells were left to thaw on ice and an aliquot of 50 µl was added to chilled Eppendorf tubes containing half of the ligation reaction (typically 10 µl). They were mixed by flicking the tube once and placed immediately on ice for 20 min. The tubes were placed at 42°C for 45 sec and were returned on ice for 2 min. 500 µl of liquid LB medium was added in each tube and the cells were left to grow at 37°C on an orbital shaker (250 rpm) for 1 h. The cells were then pelleted at 12,000 rpm for 10 sec and were

spread on a Petri plate containing LB agar and the appropriate antibiotic selection (typically 100 μg/ml ampicillin or kanamycin).

2.12.5 Preparation of the sequencing reactions

For each sequencing reaction 1 μg of the purified template DNA was mixed with 12.8 pmol of the appropriate primer and made up to a final volume of 12 μl with H_2O.

2.12.6 Nucleotide sequence analysis

Plasmids containing an insert of interest were sequenced using the Universal primer. When the insert was large, the clone was also sequenced using the Reverse primer. In the cases of large inserts, internal primers were also designed and used for sequencing.

Sequence data were analysed, aligned and compared for similarity using the University of Wisconsin Genetics Computer Group (GCG) package of sequence programs, the BLAST package of programs on the NCBI Web server and the European Molecular Biology Open Software Suite (EMBOSS) package of sequence programs. Phylogenetic trees were constructed using the CLUSTAL_X program (Thompson *et al.*, 1997) and drawn using the TreeView version 1.6.6 freeware (Page 1996).

2.13 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

Separation of proteins was performed under denaturing conditions using SDS-PAGE, as described by Laemmli (1970). The Mini-Protean electrophoresis system was used (BioRad) and slab gels were cast between two glass plates (12x 10x 0.2 cm) separated by 1.5 mm spacers. The gels consisted of a resolving and a stacking gel. The resolving gel was prepared by mixing the following:

Resolving gel:

	10% Acrylamide
H_2O	4.00 ml
1.5M Tris-HCl (pH 8.8)	2.50 ml

10% SDS	0.10 ml
30% acrylamide/0.8% bis-acrylamide	3.33 ml
10% APS	100 µl
TEMED	10 µl

After the addition of a few drops of isopropanol on the surface of the gel, the system was left for 50 min to polymerise. After polymerisation the isopropanol was removed, the surface of the gel was washed a few times with SDW, the stacking gel was poured using a Pasteur pipette and a comb was inserted.

Stacking gel:

	4% Acrylamide
H_2O	2.77 ml
0.5M Tris-HCl (pH 6.8)	1.26 ml
10% SDS	50 µl
30% acrylamide/0.8% bis-acrylamide	0.83 ml
10% APS	50 µl
TEMED	5 µl

Samples were mixed in 2x Laemmli buffer (100 mM Tris-HCl [pH 6.8], 2% β-mercaptoethanol, 4% SDS, 0,2% bromophenol blue, 20% glycerol) and boiled for 5 min. After polymerisation of the stacking gel, the comb was carefully removed and samples were loaded into the wells. SDS-PAGE gels were immersed in 1x SDS-PAGE running buffer (25 mM Tris-HCl, 250 mM glycine, 0.1% SDS, pH of the solution adjusted to pH 8.3) and electrophoresed for 1-1.5 h at 80-100 V until the bromophenol blue dye reached the bottom of the gel.

2.14 Viral particle isolation from fungal tissue

The viral particle isolation processes shared a common first step. The fungal tissue was harvested, weighed and the mycelia were disrupted by grinding up in a chilled pestle,

using 7-10 g of Silica carbide C6 (Carborundum Corp.). The homogenate was resuspended in cold PBK buffer (0.03 M Na_2HPO_4, 0.3 M KCl [pH 7.6]) in 20 ml/g of wet weight of the fungal tissue. The homogenate was then stirred at 4°C for 4 h and was subsequently cleaned with a centrifugation at 12,000 rpm (22100 *g*) for 45 min.

2.14.1 Isolation *via* direct pelleting

The collected supernatant was centrifuged at 26,000 rpm (78590 *g*) for 16 h. The pellet was resuspended in 80 - 100 ml of PBK buffer by stirring at 4°C for 4 h. The solution was then cleaned by centrifugation at 16,000 rpm (30965 *g*) for 45 min and was repelleted at 26,000 rpm (78590 *g*) for 16 h. The pellet was resuspended in 1 ml PBK buffer and was cleaned by low speed centrifugation at 16,000 rpm (30965 *g*) for 30 min. The resulting supernatant was stored at 4°C. The preparation was further purified by a 20-50% sucrose gradient centrifugation at 40,000 rpm (197568 *g*) for 4 h.

2.14.2 PEG_{6000} precipitation

Solid NaCl was added to the collected supernatant to give a final concentration of 1 M, under constant stirring at 4°C. Filter sterilised 30% (w/v) PEG_{6000} was added dropwise while stirring to give a final volume of 10%. The solution was then stirred overnight at 4°C. The PEG_{6000} precipitate was collected by centrifugation at 12,000 rpm (22100 *g*) for 45 min and was resuspended in 60–100 ml of PBK buffer. The supernatant was cleaned by centrifugation at 26,000 rpm (78590 *g*) for 16 h. The pellet was resuspended in 1 ml of PBK buffer and was cleaned by low speed centrifugation at 16,000 rpm (30965 *g*) for 30 min. The resulting supernatant was stored at 4°C. The preparation was further purified by a 20-50% sucrose gradient centrifugation at 40,000 rpm (197568 *g*) for 4 h.

2.15 Staining of protein gels

2.15.1 Coomassie blue staining

Coomassie staining provided a simple and sensitive method (*ca.* 15 μg) for detecting electrophoresed proteins. The gel was stained at room temperature in 0.1% Coomassie brilliant blue R250 in methanol:SDW:glacial acetic acid (5:5:1 by volume), for at least 15

min on an orbital shaker. De-staining was achieved by diffusion overnight against methanol:SDW:glacial acetic acid (5:5:1 by volume), or with several changes of de-stain solution over 2-3 h.

2.15.2 Silver staining

For the silver staining of SDS-PAGE gels the Silver Stain Plus kit was used (BioRad). All the steps took place at room temperature. The gel was fixed for 20 min in fixative enhancer solution (50% methanol, 10% acetic acid, 10% fixative enahncer concentrate and 30% H_2O). The fixative enhancer solution was decanted and the gel was rinsed twice for 10 min in H_2O with gentle agitation. The H_2O was decanted and the staining and development step took place by immersing the gel for 20 min in the staining solution containing the development accelerator solution (2.5 ml silver complex solution, 2.5 ml reduction moderator solution, 2.5 image development reagent, 25 ml development accelerator solution and 17.5 ml H_2O). The staining step could be extended until the desired band intensity was reached. The reaction was stopped by placing the gel in 5% acetic acid for 15 min. After stopping the reaction the gel was rinsed for 5 min in H_2O. It was then dried and photographed.

2.16 Sterol glycosyltransferase activity assay

The protocol was adapted from Warnecke *et al.* (1999). 80 µl of the purified, His-tagged protein (or a same volume of the aqueous fraction from the cell homogenate, following sonication and centrifugation of the cell culture), were added to 10 µl of a 4 mM solution of cholesterol in ethanol (Sigma) and 10 µl of UDP-[U-^{14}C] glucose, equivalent to 0.25 µCi/ 9.25 kBq (Amersham). The mixture was incubated for 2 h at 30°C and the reaction was terminated by the addition of 0.2 ml of 0.5% NaCl solution (w/v in H_2O) and 40 µl of chloroform:methanol (2:1). 2.5 µl of the extracted lipids were spotted on a silica gel plate (Greyhound) and were separated by thin layer chromatography with chloroform:methanol (85:15 v/v). The radioactivity on the silica plate was detected by an IP phosphoimager scanner (AGFA).

2.17 Abbreviations

APS	ammonium persulphate	mg	milligram
ATCC	American type culture collection	µg	microgram
BSA	bovine serum albumin	min	minute(s)
°C	degrees Celsius	mj/cm^2	millijoule per centimetre
ca.	approximately	ml	millilitre
µCi	micro Curie	µl	microlitre
Cm	centimeter	mmol	millimolar
CS	Calf serum	N	normality
dATP	deoxyadenosine triphosphate	ng	nanogram
dCTP	deoxycytosine triphosphate	nm	nanometers
dGTP	deoxyguanosine triphosphate	OD	optical density
dTTP	deoxythymidine triphosphate	PCR	polymerase chain reaction
DNase I	deoxy-ribonuclease	PEG	polyethylene glycol
DNA	deoxy-ribonucleic acid	pmol	picomolar
DMSO	dimethyl sulfoxide	PNK	polynucleotide kinase
dNTPs	equimolar mixture of dATP, dCTP, dGTP, dTTP	psi	pounds per square inch
ds	double stranded	RNA	ribonucleic acid
DTT	dithiothreitol	rpm	revolutions per minute
EDTA	ethylene-diamene-tetra-acetic acid (disodium salt)	ss	single stranded
EGTA	Ethylene-glycol-tetra-acetic-acid	SDS	sodium dodecyl sulphate
FBS	foetal bovine serum	SDW	sterile dionised water
g	gravity	sec	second(s)
h	hours	TEMED	N, N' tetramethylethylene alcohol
HEPES	N-(2-hydroxyethyl)piperazine-N'-2-ethanesulfonic acid	Tris	trishydroxymethylaminoethane
LB	Luria-Bertani	U	unit
M	molar	UV	ultra violet
mM	millimolar	V	volt(s)
mm	millimetre		
mm	micromolar		

2.18 The Genetic code

First position (5' end)	Second position U	Second position C	Second position A	Second position G	Third position (3' end)
U	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	U
	Phe (F)	Ser (S)	Tyr (Y)	Cys (C)	C
	Leu (L)	Ser (S)	Stop	Stop	A
	Leu (L)	Ser (S)	Stop	Trp (W)	G
C	Leu (L)	Pro (P)	His (H)	Arg (R)	U
	Leu (L)	Pro (P)	His (H)	Arg (R)	C
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	A
	Leu (L)	Pro (P)	Gln (Q)	Arg (R)	G
A	Ile (I)	Thr (T)	Asn (N)	Ser (S)	U
	Ile (I)	Thr (T)	Asn (N)	Ser (S)	C
	Ile (I)	Thr (T)	Lys (K)	Arg (R)	A
	Met (M)	Thr (T)	Lys (K)	Arg (R)	G
G	Val (V)	Ala (A)	Asp (D)	Gly (G)	U
	Val (V)	Ala (A)	Asp (D)	Gly (G)	C
	Val (V)	Ala (A)	Glu (E)	Gly (G)	A
	Val (V)	Ala (A)	Glu (E)	Gly (G)	G

2.19 Amino acid codes

Alanine	Ala	(A)
Arginine	Arg	(R)
Asparagine	Asn	(N)
Aspartic acid	Asp	(D)
Cysteine	Cys	(C)
Glutamine	Gln	(Q)
Glutamic acid	Glu	(E)
Glycine	Gly	(G)
Histidine	His	(H)
Isoleucine	Ile	(I)
Leucine	Leu	(L)
Lysine	Lys	(K)
Methionine	Met	(M)
Phenylalanine	Phe	(F)
Proline	Pro	(P)
Serine	Ser	(S)
Threonine	Thr	(T)
Tryptophan	Trp	(W)
Tyrosine	Tyr	(Y)
Valine	Val	(V)

2.20 List of suppliers

Amersham, Amersham Biosciences UK Ltd., Amersham Place, Little Chalfont, HP7 9NA., Buckinghamshire.

Anachem, Anachem Ltd., Anachem House, Charles Street, Luton, LU2 0EB, Bedfordshire.

Beckman, Beckman Coulter Ltd., Oakley Court, Kingsmead Business Park, London Road, High Wycombe, HP11 1JU, Buckinghamsphire

Bioline, Bioline Ltd., 16 The Edge Business Centre, Humber Road, London, NW2 6EW.

BioRad, Bio-Rad Laboratories Ltd., Bio-Rad House, Maylands Avenue, Hemel Hempstead, HP2 7TD, Hertfordshire

Carborundum Corp., Saint-Gobain Abrasives Ltd. Unit 25 Anson Business Park, Cheltenham Road East, Staverton, GL2 9QN, Gloucestershire

DuPont, Du Pont Ltd., Wedgwood Way, Stevenage, SG1 4QN, Hertfordshire.

Fermentas, Helena Biosciences Europe, Colima Avenue, Sunderland Enterprise Park, SR5 3XB, Sunderland.

Greyhound Chromatography and Allied Chemicals, Analysis House, 6 Kelvin Park, Birkenhead, CH41 1LT, Merseyside.

Hybaid, Thermo Electron Corp., 1 St. George's Court, Hanover Business Park, Altrincham, WA14 5TP.

Invitrogen, Invitrogen Ltd., 3 Fountain Drive, Inchinnan Business Park, Paisley, PA4 9RF.

Kodak, Eastman-Kodak, Ltd., Kodak House, Hemel Hempstead, HP11 IJZ, Hertfordshire.

MBio, Molecular BioProducts Inc., 9880 Mesa Rim Road, San Diego, CA 92121-2979, USA.

Merck, Merck Sharp & Dohme Ltd., Hertford Road, Hoddesdon, EN11 9BU, Hertfordshire.

Millipore, Millipore Ltd., Units 3&5, The Courtyards, Hatters Lane, Watford, WD18 8YH.

NEB, New England Biolabs Ltd., 73 Knowl Piece, Wilbury Way, Hitchin, SG4 0TY, Hertfordshire.

Novagen, EMD Biosciences, Inc., 10394 Pacific Center Court, San Diego, CA 92121, USA

OXOID, Oxoid Ltd., Wade Road, Basingstoke, RG24 8PW, Hampshire.

Promega, Delta House, Chilworth Science Park, SO16 7NS, Southampton.

Promochem, LGC Promochem, Queens Road, Teddington, TW11 0LY, Middlessex.

QIAGEN, QIAGEN House, Fleming Way, Crawley, RH10 9NQ, West Sussex.

Schleicher & Schuell, Schleicger & Schuell Ltd., Unit 11, Brunswick Park Industrial Estate, N11 1JL, London.

SERVA, AMS Biotechnology Ltd., 63B Milton Park, Abingdon, OX14 4RX.

Sigma-Aldrich, Sigma-Aldrich Company Ltd., The Old Brickyard, New Road, Gillingham, SP8 4XT, Dorset.

Stratagene, 11011 N. Torrey Pines Road, La Jolla, CA. 92037, USA.

Whatman, Whatman plc., 27 Great West Road, Brentford, TW8 9BW, Middlessex.

CHAPTER THREE

Characterisation of five dsRNA elements associated with the Amasya Cherry disease.

3.1 Introduction

3.1.1 Amasya cherry disease (ACD), symptoms and spread of the disease

The *Prunus* genus belongs to the *Rosaceae* family, of which over four hundred species are known, most occurring in the northern temperate zone (Allaby 1992). Sweet cherry (*Prunus avium L.*) is susceptible to a number of diseases caused by different pathogens. Amongst them are virus and virus-like diseases which occur in the *Prunus* species as well as on sweet cherry, *e.g. Cherry rasp leaf virus* (CRLV) and the *Prunus necrotic rusty spot virus* (PNRSV). Sweet cherry is also susceptible to infections caused by virus-like pathogens, *e.g.* mycoplasma-like organisms, while a number of the diseases are caused by unknown agents, some of which are characterised as graft-transmissible pathogens.

Amasya cherry disease (ACD) was first reported in the Amasya region of Turkey in the 1950s. Initial investigations concluded that Pfeffinger disease was the probable cause (Blodgett *et al.* 1970). Within the first ten years of the disease front spreading, at least 65,000 mature cherry trees were killed. Studies revealed that ACD was not due to the Pfeffinger disease but to a graft-transmissible pathogen, but neither fungi nor bacteria could be correlated with its presence. It was therefore thought that it spread through natural root grafting, or soil organisms acting as vectors, favoured by the close proximity of tree spacing. Space is limited in the Amasya region and, since no pruning is done, trees grow very tall and very close together. Other cherry diseases known to occur in the Amasya area include Coryneum blight caused by *Coryneum beijerinckii*, brown rot caused by the *Monilia spp.* and root rot caused by *Armillaria mellea*. There is also evidence that PNRSV and rugose mosaic are present in the area. Pests known to occur in the area include the cherry fruit fly (*Rhagoletes spp.*), cherry slug (*Eriocampoides lamacina*), shot-hole borer (*Scolytis spp.*), black cherry aphid (*Myzus cerasi*) and the mites *Tetranychus telarius* and *T. viennensis* (Blodgett *et al.* 1970).

ACD infected trees have brick-coloured spots on the leaves, following the appearance of small chlorotic lesions in the early stages of infection (shown in Appendix, section A). These spots are *ca.* two cm in diameter, become necrotic and eventually cover most of the leaf area. A variable but significant amount of diseased foliage remains on the

diseased trees throughout winter. The ACD infection causes poor blossoms and small, light pink coloured fruit. ACD affected trees show severe limitation of terminal growth and considerable die-back of weak fruiting wood. The death of the infected tree is inevitable and usually occurs within three to five years after the appearance of the first symptoms (Citir 1987). It is estimated that approximately two years are required for symptoms to appear following bud inoculation (Blodgett *et al.* 1970).

The local cherry tree cultivar varieties in the Amasya region, *e.g.* Koroglu, Tabani Yank and Kara Kinaz, exhibit much more severe symptoms to ACD than imported cultivar varieties, *e.g.* Lambert, which are more tolerant and show milder symptoms. However, even newly planted orchards containing imported cultivar varieties eventually become infected, resulting firstly in a severe reduction of crop yield and eventually to the death of the trees (Citir 1987).

Mechanical transmission of ACD in *Solanaceae* (*Petunia spp.*) produced no visible symptoms on the leaves. However, infected plants did show some symptoms when flowering, produced less seed and dying much earlier. In *Antirrhinum majus* the plants infected with ACD, showed chlorotic yellow spots on the leaves approximately one month after inoculation and never grew sufficiently to produce blossom. The inability to study further the ACD effects on the inoculated *Solanaceae* plants and to back-inoculate the infectious agent on sweet cherry plants meant that Koch's postulates could not be satisfied for this pathogenic agent (Citir 1987). Transmission tests on *Gomphrena globosa*, *Chenopodium album*, *C. amaranticolor*, *C. quinoa*, *Zinnia elegans*, *Cucumis sativus*, *C. melo*, *Phaseolus vulgaris*, *Vigna sinensis*, *Nicotiana glutinosa*, *N. langsdorffii* and *N. tabacum* did not produce any positive results (Citir 1987).

3.1.2 ACD associated dsRNA elements

Low and high speed centrifugation resulted in the partial purification of the infectious agent. Further purification allowed for a UV spectrum to be obtained, with the absorption being characteristic for the presence of both viral proteins and nucleic acids. Electron microscopy however, did not find any viral particles (discussed in section 3.2.1.5). PAGE

analysis of dsRNA-rich preparations from ACD infected cherry trees showed a complex profile of dsRNA elements that were absent from healthy-looking trees (Açikgöz *et al.* 1994). This profile is typical of ACD infections and is shown below in Figure 3.1 A). Different ACD-affected tissues, from different cultivar varieties, contained the same, constant number of dsRNA species (Covelli *et al.* 2004 and personal observation).

Six out of the eleven dsRNA species have been recently isolated and their genomes sequenced. They include, as shown in Figure 3.1 B), two dsRNAs constituting a new species in the genus *Partitivirus* (Coutts *et al.* 2004) and four dsRNAs constituting a new species in the genus *Chrysovirus* (Covelli *et al.* 2004). Though the species of the genus *Chrysovirus* described so far infect fungi, the possibility exists that other members could infect plants, in a similar manner to the cryptovirus (*Alphacryptovirus* and *Betacryptovirus*) members of the *Partitiviridae* family.

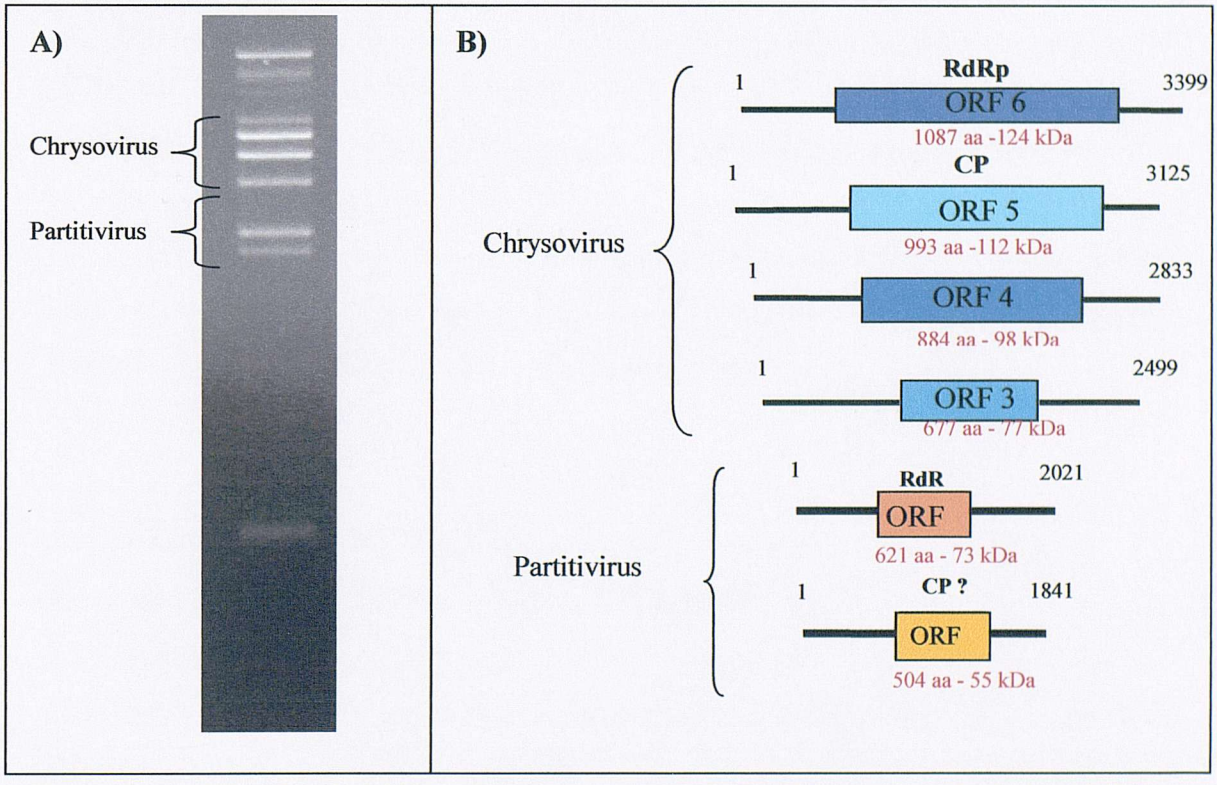


Figure 3.1: ACD related dsRNA elements. A) Analysis in 1% agarose gel of the dsRNAs associated with ACD, stained with ethidium bromide. The two characterised viruses are shown. B) Schematic representation of the two characterised viruses. The order of the dsRNA elements in Figure B follows the order of the dsRNA elements in Figure A. The length in bp of each

dsRNA element is shown. The ORFs are represented by the shaded boxes; the ORF size in amino acids is shown, as well as the molecular weight of the predicted protein product. The function of the predicted protein product is shown where known. The 5' and 3' UTRs are shown as a straight line. (Reproduced from Covelli *et al.* 2004)

In respect to the causal agent of ACD, both *Partitivirus* and *Chrysovirus* infections are asymptomatic, thus the dsRNA species described so far are unlikely to be solely responsible for the observed disease symptoms. Based mainly on circumstantial evidence, it is suggested that ACD has a fungal origin. Electron microscopic examinations of thin sections have revealed a fungus-like organism in ACD symptomatic leaves, and fluorescent microscopy observations have confirmed the presence of 'mycelium' in affected leaf areas, growing between the lower epidermis, palisade and spongy mesophyll cells, which was not found in vascular bundles or in asymptomatic samples (Alioto *et al.* 2003). However, efforts to isolate and grow the fungus-like organism observed on artificial media have failed (Covelli *et al.* 2004). As part of the ongoing investigation of ACD, the remaining five largest and two smallest dsRNA elements were sequenced to completion and are presented in this thesis. The proposed functions of their predicted encoding peptides and their viral origin are discussed. In terms of hypotheses tested in this chapter, there are two main ones: firstly, that the remaining five largest dsRNA elements represent dsRNA viruses, and secondly, that these viruses, being probably mycoviruses or of fungal origin, are not related to the disease.

3.2 Results and discussion

3.2.1 Sequencing the five largest dsRNA elements

3.2.1.1 Sequencing approach

The five largest dsRNA elements, named here in descending order of size as ACD dsRNAs 1a, 1b, 2, 3 and 4 (shown in Figure 3.2), were sequenced to completion with the exception of dsRNA4. The sequencing approach involved the construction of a random cDNA library from the gel extracted dsRNA template, and the RT-PCR extension of

overlapping clones or clones with similarity to other known viral sequences. The latter was used for the extension of clones containing characteristic RdRp protein motifs.

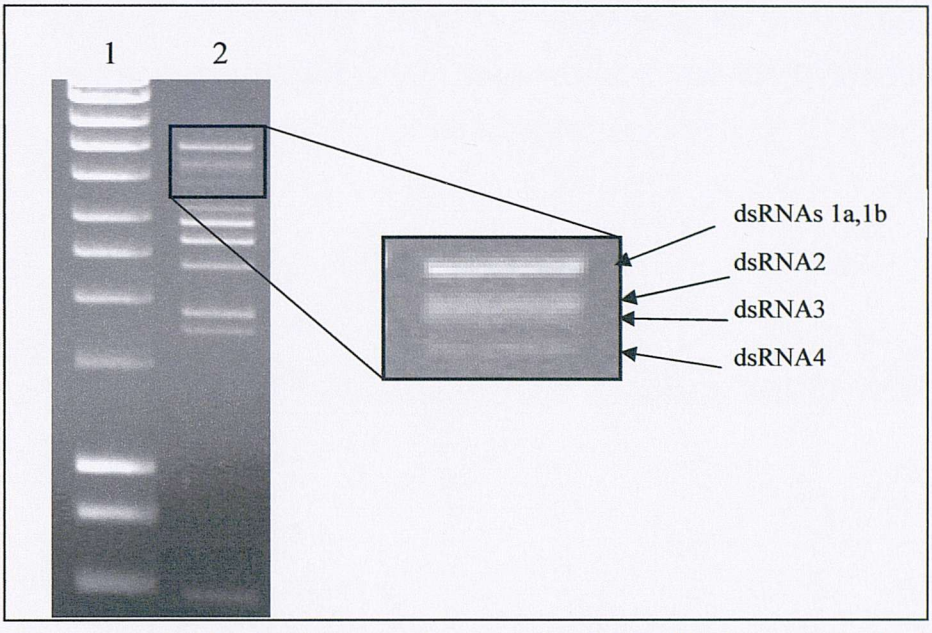


Figure 3.2: The five largest ACD-associated dsRNA elements. Analysis in 1% agarose gel of the ACD associated dsRNAs, stained with ethidium bromide (lane 2) together with the 1kb genomic size marker (lane1). The four largest dsRNA molecules of ACD are shown in the inset in order of ascending genomic size: a doublet of dsRNA1a and dsRNA1b, a doublet of dsRNA2 and dsRNA3, and dsRNA4. The dsRNAs have estimated sizes of approximately 5.1, 5.1, 4.4, 4.3 and 3.9 kbp respectively.

3.2.1.2 Sequence information

3.2.1.2.1 ACD dsRNA1a

The ACD dsRNA1a element is 5121 bp long, the sequence information is presented in full in the GenBank database, accession number: AM085136. The map in Figure 3.3 shows the number of clones generated, which were used to construct the contiguous sequence of the dsRNA1a element. The approximate position and the orientation of the primers used are also shown. Computer assisted translation of the cDNA sequence showed that the dsRNA1a element contained one single ORF, 1628 amino acids long with a predicted protein product of 185.59 kDa and pI = 6.03 (calculated using the ExPaSy Compute pI/Mw tool). The 3' UTR is 199 bp long, but the 5' UTR is presently 35 bp in length. There are two methionine residues at bp positions 195 and 558. The latter

is the first methionine in the optimal context for an AUG start codon according to Kozak's rules and is followed closely by a second methionine in a sub-optimal context at bp position 579 (Kozak 1995). This would make the 5' UTR 557 bp long, which is a comparatively long size when compared to the other dsRNA viruses present in the ACD-infected samples, *e.g.* the ACD-associated chrysovirus 5' UTR varies from 87 to 107 bp in length.

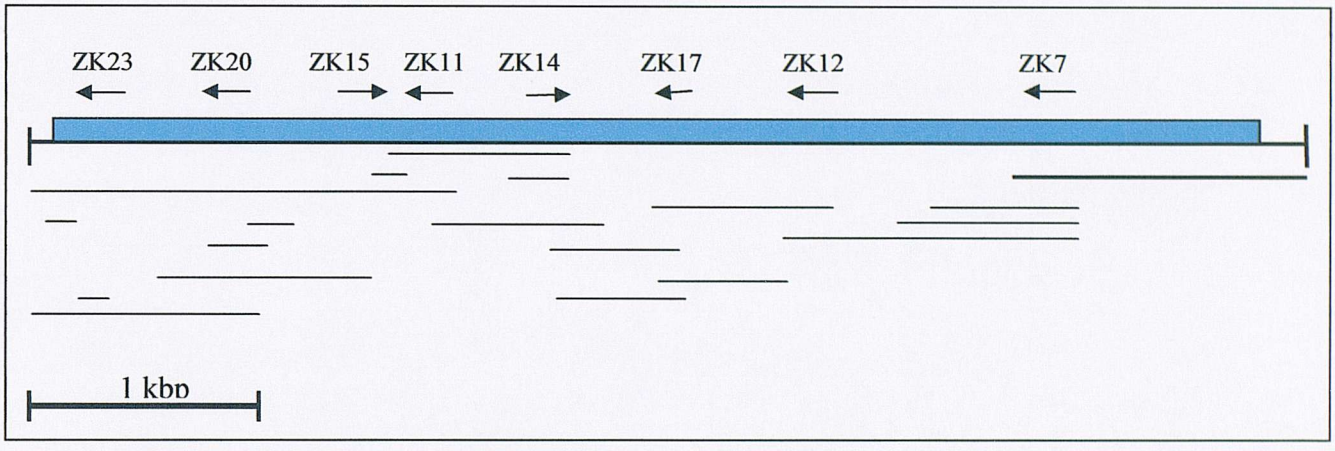


Figure 3.3: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of the ACD dsRNA1a species. The approximate position of the primers is shown. The shaded box represents the contiguous ORF, while the 5' and 3' UTRs are shown as straight lines. The horizontal lines below the ORF represent the number, size and overlap of the obtained clones.

Table 3.1: Sequence of the primers used in cDNA cloning of ACD dsRNA1a. The sequence of the primers designed for dsRNA1a is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
ZK7	GCTTCCCGGGTATGATCCTACG
ZK11	CTTAGCTCAAAGCCGTAACG
ZK12	GACATTGGTGGTCGTATCGC
ZK14	CAAGAAGACGATGCGGTACG
ZK15	GAAACCAGCCGAGCTGAGC
ZK17	CGCCGCGTACACCCTGTGC
ZK20	CAGCCGAGCTGAGCTGATCG
ZK23	GCTTGACGCGGTTGACACG

The predicted protein product of the dsRNA1a ORF did not show any similarities to proteins predicted from known viral dsRNA sequences, nor to any of the sequence information in the NCBI sequence database (using the BLASTp search tool). As no highly conserved domains or any other conserved areas are detectable in the predicted protein of the dsRNA1a ORF its functionality remains unknown. However, using the NCBI-BLAST2 pairwise protein alignment display tool, limited similarities were found to a few characterised proteins, though none of viral origin (shown in Figure 3.4). These results take into account the similarity found in a small part of the peptide sequence each time and as such they can only provide an idea for a possible role to the protein product and not a definite answer. It can however, be suggested that the dsRNA1a predicted protein product is involved in viral catalytic activity, based on the above mentioned NADH dehydrogenase and DNA helicase peptide sequence similarities.

Figure 3.4: Alignments results using the predicted peptide sequence of ACD dsRNA1a element ORF. The three highest scoring results are shown. They have been generated using the pairwise NCBI-BLAST2 protein alignment display. The identity and similarity scores between the two sequences are shown, as well as the UNIPROT accession number, the name and the organism of origin for each one of the aligned protein sequences.

1) UNIPROT: Q89NV7_BRAJA Q89NV7: NADH dehydrogenase (EC 1.6.99.3). Length = 440, Identities = 22/64 (34%), Positives = 30/64 (46%). (*Bradyrhizobium japonicum*).

```
dsRNA1:818 ILHAKTMGSEHVSELHDTESIVGAKPMDLAPANADVRSELVELASNLLKINGRRPSAQVK
           I+ A   G E  +ELH T   V A  +D      D+R  L+E A  +L     R S + +
Sbjct: 180 IIGAGATGVELAAELHRTTREVVAHGLDQVDPQKDIRITLIEAADRVLPALPERVSKETE

dsRNA1:878 KLSV 881
           KL V
Sbjct: 240 KLLV 243
```

2) UNIPROT: Q6UDM5_BRARE Q6UDM5: Starmaker. Length = 613, Identities = 21/84 (25%), Positives = 38/84 (45%). (Danio rerio)

```
dsRNA1:886 EGKNEQGFESLSESDNNAQEIDAESSGDVSMVGETSEQQVRDAIAHSEAESRLFDDTADV
           +  ++   +S S+SD+++   D E   D S      +    D+ ++SE+     +DT D
Sbjct: 450 DSDSDSDSDSDSDSDSDSNSRDKEEKKDKSSESRDEDSSDSDSKSNSESSETAEEDTNDD

dsRNA1:946 DDKLFEACSYVPGDASEVGASDVD 969
            D   E       D++ V A+D D
Sbjct: 510 KDSSVEKDKTDSSDSASVEANDSD 533
```

3) UNIPROT: Q6Y5A7_BRANA Q6Y5A7 Putative DNA helicase RecQsim. Length = 880, Identities = 24/123 (19%), Positives = 46/123 (36%). (*Brassica napus*(Rape))

```
dsRNA1:868 NGRRPSAQVKKLSVGNADEGKNEQGFESLSESDNNAQEIDAESSGDVSMVGETSEQQVRD
           +G+R +   K   +    EG+++ G+          + D E   + S+  + S
Sbjct: 376 SGKRKATGKKLAVISQGSEGQSDFGYHDAENIHETDNDDDDEEDPENSVAKQNSSNGKEM

dsRNA1:928 AIAHSEAESRLFDDTADVDDKLFEACSYVPGDASEVGASDVDITPKKSAIRPPMSSRRSS
           +  + E E+ +F    D D    E C+  P D  ++       TP +S       ++
Sbjct: 436 SEEYLEDETDIFQSVDDWDVACGEFCAMSPCDVLDIPDPSGKQTPDESGHNHSNKAKNLE

dsRNA1:988 TPT 990
            PT
Sbjct:  496 GPT 498
```

3.2.1.2.2 ACD dsRNA1b

The ACD dsRNA1b element is 5047 bp long, the sequence information is presented in full in the GenBank database, accession number: AM085137. The map in Figure 3.5 shows the number and position of the overlapping clones generated and used to construct the contiguous sequence of the dsRNA1b element. The approximate position and the orientation of the primers used are also shown. Computer assisted translation of the cDNA sequence showed that the ACD dsRNA1b element contained one single ORF, 1620 amino acids long with a protein product of predicted molecular weight 179.25 kDa and pI = 8.1. The 3' UTR is 149 bp long, while the 5' UTR is probably 35 bp long. There are a further two methionine residues at positions 312 and 408, but both are in a sub-optimal context. The first methionine residue at an optimal context according to Kozak's rules is found at position 423 (Kozak 1995).

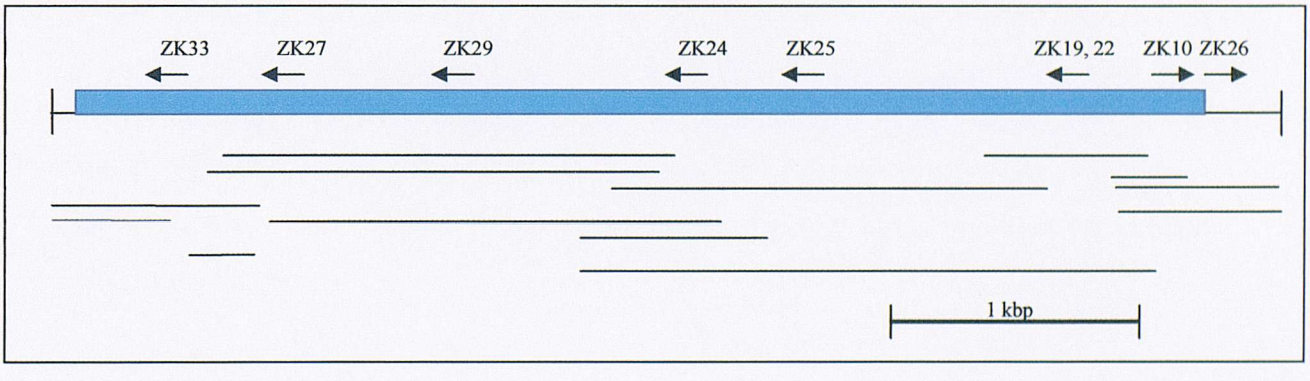


Figure 3.5: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of the ACD dsRNA1b species. The approximate position of the primers is shown. The shaded box represents the contiguous ORF, while the 5' and 3' UTR are shown as a straight line.

Table 3.2: Sequence of the primers used in the cDNA cloning of ACD dsRNA1b. The sequence of the primers designed for dsRNA1b is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
ZK10	GGATATAAAAGTGCGTCCG
ZK19	GTAGTGTACACTGGAGTCG
ZK22	GCCCCGTTGACTTTCGGCCG
ZK24	CTGGTGCTGAAGTTTTCGGG
ZK25	CATGGCGATGGCATCCTAGG
ZK26	CCGGCCGAAAGTCAACGGGC
ZK27	CAAGTTTACGACTGAGGCC
ZK29	CAACTCAGTGGCCAGGCC
ZK33	CGTTGTCTACCCACGCCATCAGG

The predicted protein product of the dsRNA1b ORF did not show any similarities to proteins predicted from known viral dsRNA sequences, nor to any of the sequence information in the NCBI sequence database (using the BLASTp search tool). As no highly conserved domains or any other conserved areas are detectable in the predicted protein of the dsRNA1b ORF its functionality remains unknown. However, using the NCBI-BLAST2 pairwise protein alignment display tool, limited similarities were found to a few characterised proteins, none of viral origin (shown in Figure 3.6). These results take into account the similarity found in a small part of the peptide sequence each time and as such they can only provide an idea for a possible role to the protein product and not a definite answer. Thus, it can be suggested that the dsRNA1b predicted protein product along with the dsRNA1 predicted protein product, might be involved in viral catalytic activity.

The two peptides predicted from the sequences of ACD dsRNA1a and ACD dsRNA1b when aligned against one other have a 38.9% sequence identity and 69.8% sequence similarity at a nucleotide level and 55.8% sequence similarity at a peptide level. The widest differences of the peptide sequence alignment are located between amino acid residues 1085–1280 and 1100-1280 for dsRNAs 1a and 1b respectively. Thus, it would appear that these two dsRNA elements encode at least one protein product, most probably with the same function and are evolutionarily very closely related, but are not identical, *i.e.* they are not variants of the same dsRNA molecule.

Figure 3.6: WU-BLAST2 alignments results using the peptide sequence of the ACD dsRNA1b element. The three highest scoring results are shown. They have been generated using the pairwise NCBI-BLAST2 protein alignment display. The identity and similarity scores between the two sequences are shown, as well as the UNIPROT accession number, the name and the organism of origin for each one of the aligned protein sequences.

1) UNIPROT: CO410_WHEAT P46524 Dehydrin COR410 (Cold-induced COR410 protein). Length = 262, Identities = 35/134 (26%), Positives = 60/134 (44%). (*Triticum aestivum*)

```
dsRNA4:  539 VKDVKMLGTEHIADIEYADEVLAEKQRRPSDRSE--DIREQMMLNANALVLKFSGRKGAH 596
             V++ ++   EH  D E  + + ++  R  S  S   D  E+ +++ N  V+K   +KG
Sbjct:    55 VEEPEVKKEEH-EDGEKKETLFSKLHRSSSSSSSSSDEEEEEVIDDNGEVIKRKKKKGLK 113

dsRNA4:  597 SK---KLSVDVETEQEECVDFDGGLTPSPVKTQGGESDSSAEVTSIREDIRINNAELVLE 653
              K   KL    +TE E          P+ V+T GG  D+   V  I  D++   A  V E
Sbjct:   114 EKLQGKLPGHKDTEGEHVTGLPAPAAPASVQTHGGHHDTDVVVEKIDGDVKTEAAPAVPE 173

dsRNA4:  654 PPRPEVSDEVADKL 667
               +    +++ +KL
Sbjct:   174 EEKKGFLEKIKEKL 187
```

2) UNIPROT: Q93ZQ2_ARATH Q93ZQ2 AT3g60110/T2O9_90. Length = 641, Identities = 47/206 (22%), Positives = 86/206 (41%). (*Arabidopsis thaliana*)

```
dsRNA4:  537 EDVKDVKMLGTEHIADIEYADEVLAEKQRRPSDRSEDIREQMMLNANALVL-KFSGRKGA 595
             +D KD K L  +H+ D++  ++ + EK    S      R+  +L  NA+V    S  +
Sbjct:   326 QDTKDYKRLIRQHL-DMKTIEKKM-EKGSYVSSSLSFYRDLKLLFTNAIVFFPTSSSESI 383


dsRNA4:  596 HSKKLSVDVETEQEECVDFDGGLTPSPVKTQGGESDSSAEVTSIREDIRINNAELVLEPP 655
              +++L   V  E ++     G L    +K++  ES  S + +S+   +         L+
Sbjct:   384 AAQELRTLVSNEMKKRT---GKLGHCVIKSEA-ESSVSRQKSSVLSLVPCKKKSSALKKT 439

dsRNA4:  656 RPEVSDEVADKLTQAKEFADSLTTTSSETRDKKSWDVPYMNAAEEKVYRTERVREQLGRI 715
              P  S    D+    +   + + TT++ T  + S       A   K  +T R +  + +
Sbjct:   440 SPSSSSRQKDEKKSQEVSEEKIVTTTATTSARSSRRTSKEIAVVAKDTKTGRAKNNIKK- 498
```

```
dsRNA4:  716 RRDLMEEQSSDDESFKKGNSVARPPT 741
             + D   E S DD+  K+ NS     T
Sbjct:   499 QTDTKTESSDDDDDEKEENSKTEKKT 524
```

3) UNIPROT:Q4SJM9_TETNG Q4SJM9 Chromosome 1 SCAF14573, whole genome shotgun sequence. (Fragment). Length = 247, Identities = 52/203 (25%), Positives = 99/203 (48%). (*Tetraodon nigroviridis*)

```
dsRNA4:  537 EDVKDVKMLGTEHIADIE-YADEV-LAEKQRRPSDRSEDIREQMMLNANALVLKFSGR-- 592
             E +K +K L       D+E Y  ++ L EKQR+  D  E + +Q  +  N +  +F  +
Sbjct:    41 ERMKGLKSLAKP--CDVEVYEKKIKLLEKQRK--DVLE-VNKQWDIQWNTMKSQFEQKIT 95

dsRNA4:  593 ----KGAHSKKLSVDVETEQEECV-DFDGGL--TPSPVKTQGGESDS-SAEVTSIREDIR 644
                 + A S+K  +++E E+E+   D+D  L    S ++   GE +  ++E T +++ IR
Sbjct:    96 DLRQRLAESQKTVLELEAEREQRQRDYDKKLLLAKSKIENVQGEKECLNSETTELKQKIR 155

dsRNA4:  645 INNAELVLEPPRPEVSDEVADKLTQAKEFADSLTTTSSETRDKKSWDVPYMNAAEE-KVY 703
                +L+    + E  ++   +L +A E A +L + SS  +     +  + +AA   K++
Sbjct:   156 YLQDQLLPLSKQREYQEKEIQRLNRALEEALNLHSPSSSQQPPGQGN--FADAANNVKIF 213

dsRNA4:  704 RTERVREQLGRIRRDLMEEQSSD 726
               E  R++  R  R+ M E+  D
Sbjct:   214 E-EDFRKE--RSDRERMNEEKED 233
```

3.2.1.2.3 ACD dsRNA2

The ACD dsRNA2 element is 4458 bp long, the sequence information is presented in full in the GenBank database, accession number: AM085134. The map in Figure 3.7 shows the number and position of the overlapping clones generated and used to construct the contiguous sequence of the dsRNA2 element. The approximate position and the orientation of the primers used are also shown. Computer assisted translation of the cDNA sequence showed that the dsRNA2 element contained one single ORF, 1364 amino acids long with a protein product of 151.83 kDa predicted molecular weight and pI = 8.05. The 3' UTR is 307 bp long, while the 5' UTR is probably 59 bp long. As mentioned previously, translation of the ORF is thought to initiate at the first AUG in the sequence at 59 bp which conforms to Kozak's rules since the second AUG at position 72 bp is in a sub-optimal context (Kozak 1995).

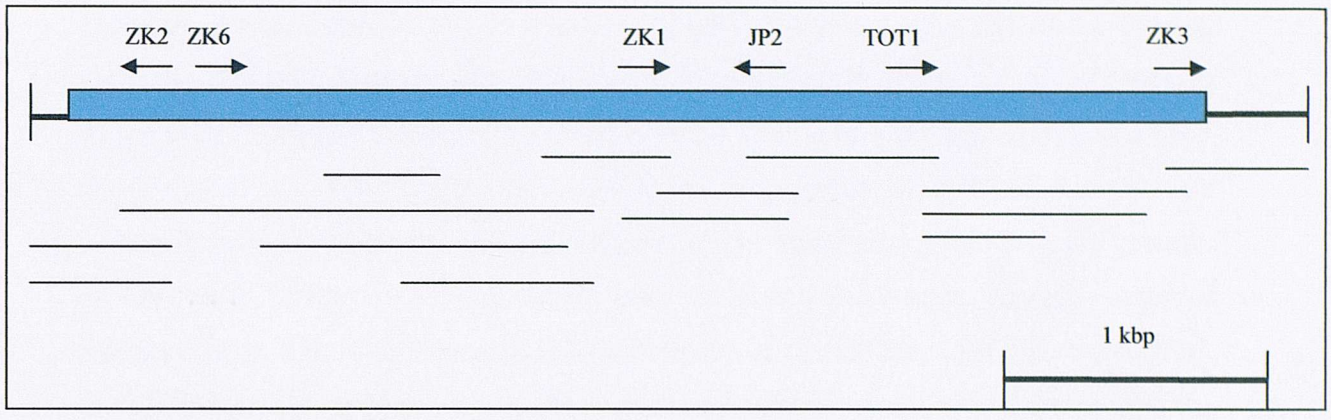


Figure 3.7: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of the ACD dsRNA2 species. The approximate position of the primers is shown. The shaded box represents the contiguous ORF, while the 5' and 3' UTRs are shown as straight lines. The horizontal lines below the ORF represent the number, size and overlap of the obtained clones.

Table 3.3: Sequence of the primers used in the cDNA cloning of ACD dsRNA2. The sequence of the primers designed for dsRNA2 is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
JP2	ACATACAGTTCAGGTGGCACTG
TOTI1	AAACATTGTCACTGCACTCT
ZK1	CCAGGTGCATCAAGGTTGG
ZK2	CTCATTCCCCTGAACACCCATCG
ZK3	GCAAGAGCGCTAACGTTCGC
ZK6	GCCCCTTGCTCATTCCC

The C-terminal dsRNA2 predicted protein product showed very significant similarity to the totiviral RdRp proteins, at the peptide sequence alignment level (Figure 3.8). Thus it can be safely proposed that part, or all, of the predicted protein product constitutes a novel, viral RdRp. As the highest RdRp similarity scores are to totiviruses and the dsRNA element is 4.4 kbp long, which is the average molecular size of viruses in the family *Totiviridae*, it was investigated whether dsRNA2 element represents a novel totiviral species.

Totivirus genomes consist of two overlapping ORFs. Although, the size of the overlap varies significantly, there are two distinct peptide sequences in different coding frames on the genome. Conversely, in ACD dsRNA2 element there is only a single contiguous ORF with no other significant ORFs to be found in any of the other two forward or the reverse coding frames. There are no internal transcription initiation sites (IRES) on the genome or any other detectable *cis*- elements, such as potential protease cleavage sites or 'slippery sites'. Those could suggest the possibility of two different peptides made from the one single ORF, initially expressed as a gag-pol fusion protein, as is the case with the yeast totiviruses. Therefore, it is suggested that dsRNA2 element, though it does not contain all of the *Totiviridae* characteristics, has significant similarities that make it distantly related to them and could represent a novel group within the particular viral family.

Nucleotide and protein sequence alignments of the part of the genome upstream of the conserved RdRp motifs (N-terminal peptide sequence), shown in Figure 3.9, did not reveal any conserved regions or any similarity to other known sequences. Thus, at present it is not known whether the ORF produces more than one protein product and if so, what the role of the additional product is.

Figure 3.8: CLUSTAL_ X alignment results of the peptide sequence for the ACD dsRNA2 ORF highlighted using GeneDoc tool. The entire peptide sequence is shown aligned against nine totiviral RdRp peptide sequences and the entire gag-pol fusion protein of ScV-L-A. The abbreviations of the viruses and the GenBank accession numbers of the peptide sequences are shown below the Figure. The conserved RdRp motifs (Bruenn 2003) are highlighted. Abbreviations of assigned viruses are italicised, of unassigned are not.

```
                        *       260         *       280         *       300         *       320
LRV 1-1 : ------------------------------------------------------------------------------:    -
LRV 2-1 : ------------------------------------------------------------------------------:    -
HmTV1-17: ------------------------------------------------------------------------------:    -
SsRV-1  : ------------------------------------------------------------------------------:    -
HvV190S : ------------------------------------------------------------------------------:    -
GaRV-L1 : ------------------------------------------------------------------------------:    -
SsRV-2  : ------------------------------------------------------------------------------:    -
ScV-L-A : AYRHALTVQDADEWIEGDRTDDQFRPPSSKVMLSALRKYVNHNRLYNQFYTAAQLLAQIMMKPVPNCAEGYAWLMHDALV:  320
ScV-L-BC: ------------------------------------------------------------------------------:    -
TVV-1   : ------------------------------------------------------------------------------:    -
dsRNA2  : ------------------------------------------MAHHMFMESLSEENEQGSPFIDSQALPDHKMAREAP:   36
```

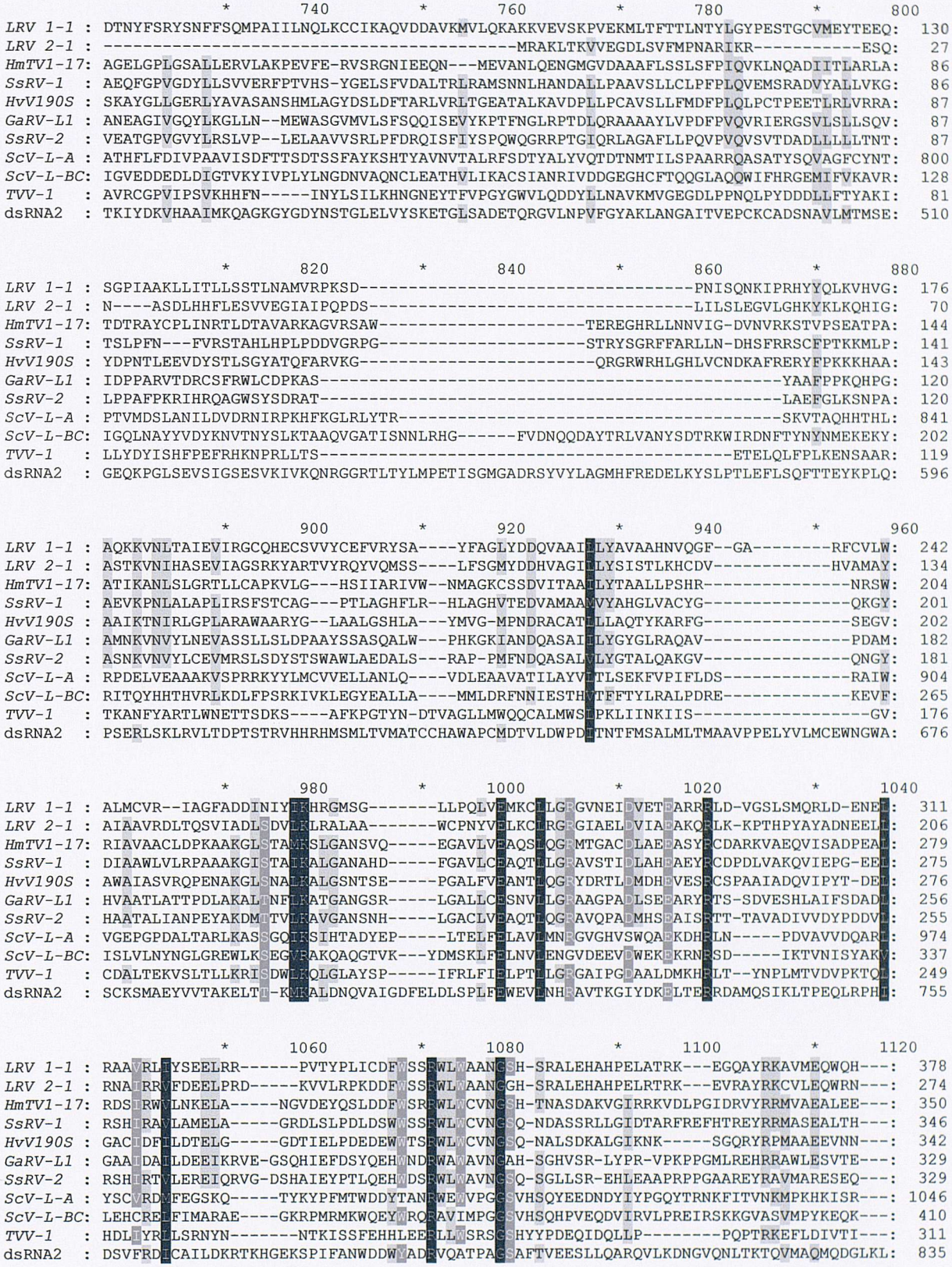
```
                  *       340         *       360         *       380         *       400
LRV 1-1 : ----------------------------------------------------------------------------:    -
LRV 2-1 : ----------------------------------------------------------------------------:    -
HmTV1-17: ----------------------------------------------------------------------------:    -
SsRV-1  : ----------------------------------------------------------------------------:    -
HvV190S : ----------------------------------------------------------------------------:    -
GaRV-L1 : ----------------------------------------------------------------------------:    -
SsRV-2  : ----------------------------------------------------------------------------:    -
ScV-L-A : NIPKFGSIRGRYPFLLSGDAALIQATALEDWSAIMAKPELVFTYAMQVSVALNTGLYLRRVKKTGFGTTIDDSYEDGAFL:  400
ScV-L-BC: ----------------------------------------------------------------------------:    -
TVV-1   : ----------------------------------------------------------------------------:    -
dsRNA2  : PAVVDSRLRGNPGAWMMYPDDQISELISYLANTQAQDMPTTSASAVSPHIIRALRTMGVQGNEQGAFSDPNYCVHDNQVR:  116

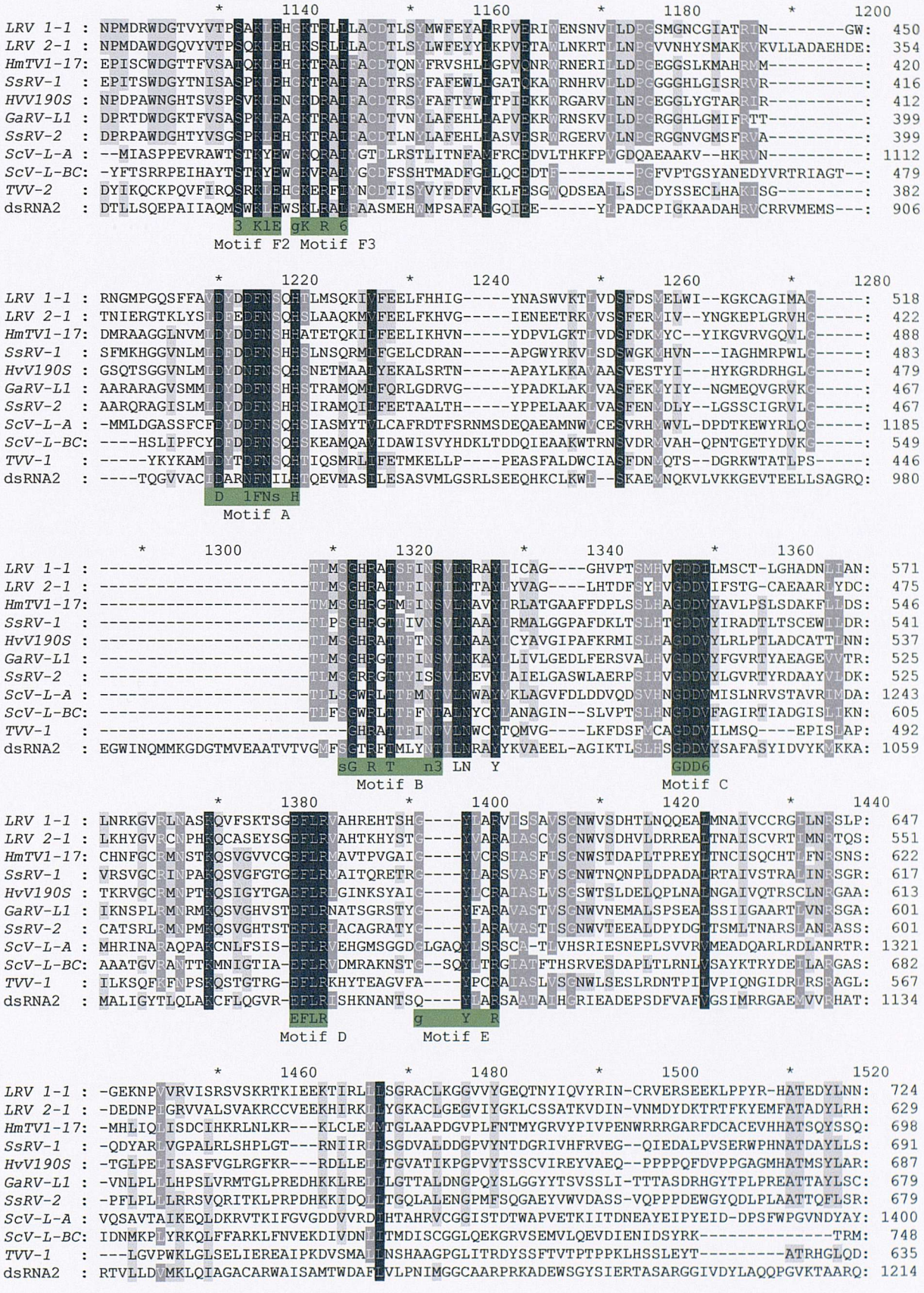
                  *       420         *       440         *       460         *       480
LRV 1-1 : ----------------------------------------------------------------------------:    -
LRV 2-1 : ----------------------------------------------------------------------------:    -
HmTV1-17: ----------------------------------------------------------------------------:    -
SsRV-1  : ----------------------------------------------------------------------------:    -
HvV190S : ----------------------------------------------------------------------------:    -
GaRV-L1 : ----------------------------------------------------------------------------:    -
SsRV-2  : ----------------------------------------------------------------------------:    -
ScV-L-A : QPETFVQAALACCTGQDAPLNGMSDVYVTYPDLLEFDAVTQVPITVIEPAGYNIVDDHLVVVGVPVACSPYMIFPVAAFD:  480
ScV-L-BC: ----------------------------------------------------------------------------:    -
TVV-1   : ----------------------------------------------------------------------------:    -
dsRNA2  : RIDRVKAAMNMRRWVGSAIVQRKKTSMGTYPVLSPIPNDDPMMMKFSPFPLQLELYFGGKRAMNADQAAQACQIVINQMT:  196

                  *       500         *       520         *       540         *       560
LRV 1-1 : ----------------------------------------------------------------------------:    -
LRV 2-1 : ----------------------------------------------------------------------------:    -
HmTV1-17: ----------------------------------------------------------------------------:    -
SsRV-1  : ----------------------------------------------------------------------------:    -
HvV190S : ----------------------------------------------------------------------------:    -
GaRV-L1 : ----------------------------------------------------------------------------:    -
SsRV-2  : ----------------------------------------------------------------------------:    -
ScV-L-A : TANPYCGNFVIKAANKYLRKGAVYDKLEAWKLAWALRVAGYDTHFKVYGDTHGLTKFYADNGDTWTHIPEFVTDGDVMEV:  560
ScV-L-BC: ----------------------------------------------------------------------------:    -
TVV-1   : ----------------------------------------------------------------------------:    -
dsRNA2  : EGMSSKEQALMHTVLLQGYQSRGSIARMEIKSRMGKTNPPIDSNMSISDKYSMHEPEYEAAYVRFCDANENTIRDYMARC:  276

                  *       580         *       600         *       620         *       640
LRV 1-1 : ----------------------------------------------------------------------------:    -
LRV 2-1 : ----------------------------------------------------------------------------:    -
HmTV1-17: ----------------------------------------------------------------------------:    -
SsRV-1  : ----------------------------------------------------------------------------:    -
HvV190S : ----------------------------------------------------------------------------:    -
GaRV-L1 : ----------------------------------------------------------------------------:    -
SsRV-2  : ----------------------------------------------------------------------------:    -
ScV-L-A : FVTAIERRARHFVELPRLNSPAFFRSVEVSTTIYDTHVQAGAHAVYHASRINLDYVKPVSTGIQVINAGELKNYWGSVRR:  640
ScV-L-BC: ----------------------------------------------------------------------------:    -
TVV-1   : ----------------------------------------------------------------------------:    -
dsRNA2  : IKVGVAAHGFEGTKTEEAAAMAKLARRKAGTYGEKRAVAPGDLGLVDSLVVKSQIDWWHCNVTMWVAAYKLSVTLAASCG:  356

                  *       660         *       680         *       700         *       720
LRV 1-1 : ----------------------------MQCPNQNHMLVNRAMVVAALDSFEDARRIISGVLDLSRLTNTSVKGQHTN:   50
LRV 2-1 : ----------------------------------------------------------------------------:    -
HmTV1-17: ------------------------------------------------------------------MKALRAAQDK:   10
SsRV-1  : --------------------------------------------------------------------MNKFIDR:    7
HvV190S : --------------------------------------------------------------------MSDPQER:    7
GaRV-L1 : ------------------------------------------------------------------MIESPVSVR:    9
SsRV-2  : ------------------------------------------------------------------MSNDRLHSR:    9
ScV-L-A : TQQGFRSGRSYDASCNAYRRTYSWRCPRRVDRTGGQCFSRVNVIEPSHGPRPTRYILQEPGTYPAWIRFRNRVQAVSRQK:  720
ScV-L-BC: ------------------------------FSCSSLRDTKFVGQQYAILTPSGTTTDIRSGRGTNQSYRRGRTSTGYR:   48
TVV-1   : ----------------------------------------------------------------------DPFLSK:    6
dsRNA2  : DMSFSDEEANEIIQYTFSRSTYRKLVASDAVMDSTRDIAASEVTQAASTPVRWARQVNPIVLVLDDAEYAITRNEANGVL:  436
```




Motif F2 Motif F3
Motif A
Motif B
Motif C
Motif D
Motif E

```
                    *      1540          *      1560          *      1580          *      1600
LRV 1-1  : HLADIEVMAVR-QYGSDIADIVAQASWKKSMSTEGAEDVSRLSLQRDKTLPCLHCITEKETSLLPVR-YGLFSSYPILMM:  802
LRV 2-1  : HTSYAESVMLR-QYGADIADIVAQSSWMKSMEASSRVSPSSMRVTLSRSDLGVFTMPVCDVERFPVR-KGVFEKYPLLMM:  707
HmTV1-17: HLPPVKVEALT-ETGTSAASAVQLASYKRAIDSMLSYEPRKVKFGKPRTMNALGLNLAKELFTHLLN-TGCLEAMPVLRL:  776
SsRV-1   : HVSEVERAALA-ATGTDAAALVLASSFSKGHSSLGSARRQVVRVAKRETTRAHGFADASDLLHRDRDKVGVLKAYPLVNL:  770
HvV190S  : HTTLVEAQARNRATGDQVPDALVILPEKPDLELKLDRTSRCRSLGANHPRVAVGFSMAHELTGRGVK-EGCLSGHPLLRL:  766
GaRV-L1  : AADPLEVNVLT-QAGVSVVSALEEASFRKSLPARYTGYET-LRLGPKSLIRSIGTASVVDLISASPP-RGVLERYPLLTL:  756
SsRV-2   : CAQPLEVKYLT-EAGIGLTAVVAEASYRKTFSSTFSRSDR-VILGPVRRRPVAGTASVERLLNTRPP-EGCLEKYPLLRL:  756
ScV-L-A  : KVWKNFGER---LEFNKIKDAVARGSRSTIALKRKARITSKKNEFANKSEWERTMYKAYKGLAVSYYANLSKFMSIPPMA: 1477
ScV-L-BC: IAKLIDKG--VGDYTAFLKTNFSEIADAITRETRVESVTKAYNVKKKTVVRAFRDLSAAYHERAVRHAWKGMSGLHIVNR:  826
TVV-1    : LCKHVPWKQLTANECNKLGQQIKKMSHRHCSQTKITYKCVYEVFKPSGLPTVLSEVSQSALSLVWWQAMLKEAMQDYSTK:  715
dsRNA2   : LVEALGIR----RYHKRVAQAVAAAIAPKGVLKNYGMIVRWMNKDDLEHLATVTGQLERIRQSREYILSKAAGLFNTLAI: 1290

                    *      1620          *      1640          *      1660          *      1680
LRV 1-1  : LKDRIPIKEALKLAVTIGYRPQPNSDLELDLWGESNNSCAIEGVLPYNEATSLAQKLPCGGVVIQVIHNVYV--------:  874
LRV 2-1  : IKERIPLREALDLAQGIGYIAPHG--CAEDLWGGTMRLCAIDGVLPYAMLVNWHAPVPCNGVRLVVDHNVYA--------:  777
HmTV1-17: IQDQIKRPLLRELVLLAGG-DGGAADLEVEAWGPQTRKCSFFGWIPYSDAAALQKKADHIKIMNHTPVRM----------:  845
SsRV-1   : VKNRLRKETISELLILAGH-TPG-ADPYRQAFGEHPESKNIIGSLSHSDASALGKATRSGNIYTLTPVFV----------:  838
HvV190S  : FERG------------------------------------------------------------------------:  770
GaRV-L1  : AKRRLPEYLVRWAVSIAGG-NPSAPDIGMEAWGEFKHGCMIATPMSYSDAATFGKRTVSTVLTCPLDAHV----------:  825
SsRV-2   : ARRRLPRSVLQRAIAEAGG-NPDAADLDLEAWGEHNHGCVVATPLSYADAATYGRRTTCGVLVCGLHFYV----------:  825
ScV-L-A  : NIEFG-----------------------------------QARYAMQAALDSSDPLRALQVIL--------------: 1505
ScV-L-BC: IRMGVS-------------------------------NLVMVVSKINPAKANVLAKSGDPTKWLAVLT-----------:  863
TVV-1    : KKD-----------------------------AHMYACNACTSSVSGDAFLRATSKMAGVLITSLISSSS---------:  756
dsRNA2   : NDHFWGDVTGVFRGIPSAWHARALTFALAKRGEEKALFDNRRFKLHTTVMGWIGVATHYSQTRHRLNLGQRPL-------: 1363
```

References: *Leishmania RNA virus 1-1* (***LRV 1-1)***, (GenBank accession number) M92355, Stuart *et al.* 1992; *Leishmania RNA virus 2-1* (***LRV 2-1)***, U32108, Scheffter *et al.* 1995; *Helicobasidium mompa Tanaka virus* (***HmTV1-17)***, AB085814, Nomura *et al.* 2003 ; *Sphaeropsis sapinea RNA virus 1* (***SsRV-1)***, AF038665, Preisig *et al.* 1998; *Helminthosporium victoriaea virus 190S* ***(HvV190S)***, U41345, Huang and Ghabrial; *Gremmeniela abietina RNA virus L1* (***GaRV-L1)***, L13218, Wang *et al.* 1993; *Sphaeropsis sapinea RNA virus 2* (***SsRV-2)***, AF039080, Preisig *et al.* 1998; *Saccharomyces cerevisiae virus L-A* (***ScV-L-A)***, M28353, Diamond *et al.* 1989; *Saccharomyces cerevisiae virus L-BC* (***ScV-L-BC)***, U01060, Park *et al.* 1996; *Trichomonas vaginalis virus 1* (***TVV-1)***, U08999, Tai and Ip 1995.

Figure 3.9: Alignments results using the N-terminal peptide sequence of the ACD dsRNA2 element ORF. The three highest scoring results are shown. They have been generated using the pairwise NCBI-BLAST2 protein alignment display. The identity and similarity scores between the two sequences are shown, as well as the UNIPROT accession number, the protein name and the organism of origin for each one of the aligned protein sequences.

1) UNIPROT: Q8IHY0_PLAF7 Q8IHY0, Protein phosphatase 2C. Identities = 24/87 (27%), Positives = 42/87 (47%), Gaps = 6/87 (6%). (*Plasmodium falciparum*)

```
dsRNA2:193 NQMTEGMSSKEQALMHTVLLQGYQSRGSIARMEIKSRMGKTNPPIDSNMSISDKY-----
           N+   GMSS  +AL+   + +    +G +   +IK+     +   DSN + ++
Sbjct: 352 NKQVNGMSSYNKALIEGAMGESIDEKGLLNLNDIKTNNILIDNNNDSNNNNNNNNNNNS

dsRNA2:248 SMHEPE-YEAAYVRFCDANENTIRDYM 273
           S ++P  +  AY   C+ NE TI DY+
Sbjct: 412 SSNDPSMFSTAYAHHCNNNEGTIGDYL 438
```

2) UNIPROT: TPOR_MOUSE Q08351 Thrombopoietin receptor precursor (TPO-R) (Myeloproliferative leukemia protein). Identities = 26/91 (28%), Positives = 48/91 (52%), Gaps = 10/91 (10%). (*Mus musculus*)

```
dsRNA2:4    HMFME--SLSEENEQGSPFIDSQALPDHKMAREAPPAVVDSRLRGNPGAWMMYPDDQISE
            H++++  SL++    Q    F+DS  LP        APP V+ +R     PG    ++ +     E
Sbjct: 108 HLWVKNVSLNQTLIQRVLFVDSVGLP-------APPRVIKARGGSQPGELQIHWEAPAPE

dsRNA2:62   LISYLANTQAQDMPTTSASAVSPHIIRALRT 92
            +  +L + + +  PT S++A +P +I+ L T
Sbjct: 161 ISDFLRH-ELRYGPTDSSNATAPSVIQLLST 190
```

3) UNIPROT: BGS4_SCHPO O74475 1,3-beta-glucan synthase component bgs4 (EC 2.4.1.34)(1,3-beta-D- glucan-UDP glucosyltransferase). Identities = 24/98 (24%), Positives = 44/98 (44%), Gaps = 4/98 (4%). (*Schizosaccharomyces pombe*).

```
dsRNA2:207 MHTVLLQGYQSRGSIARMEIKSRMGKTNPPIDSNMSISDKYSMHEPEYEAAYVRFCDANE
           +   L+ G+        R   K R+  + PI  +   SD  +M  P Y   Y++  DAN+
Sbjct:1110 LFAALIDGHSEIMENERRRPKYRIRLSGNPILGDGK-SDNQNMSLPFYRGEYIQLIDANQ

dsRNA2:267  NTIRDYMARCIKVGVAAHGFEGTKTEEAAAMAKLARRK 304
            +   +Y+  C+K+      FE  +T+     ++ AR +
Sbjct:1169  D---NYLEECLKIRSVLAEFEEMETDNVNPYSESARER 1203
```

3.2.1.2.4 ACD dsRNA3

The ACD dsRNA3 element is 4318 bp long, the sequence information is presented in full in the GenBank database, accession number: AM085135. The map in Figure 3.10 shows the number of clones generated, which were used to construct the contiguous sequence of the dsRNA element. The approximate position and the orientation of the primers used are also shown. Computer assisted translation of the cDNA sequence showed that the dsRNA 3 element potentially encodes one large ORF spanning most of the dsRNA sequence, 1294 amino acids in length with a protein product of 143.09 kDa predicted molecular weight and pI = 8.45. The 3' UTR is 345 bp long, while the 5' UTR is probably 73 bp long. Translation of the ORF is thought to initiate at the first AUG in the sequence at bp 74 which conforms to Kozak's rules since the second AUG at bp position 86 is in a sub-optimal context (Kozak 1995). The ACD dsRNA3 genome also contains a smaller ORF, the transcription start site of which is located at 897 bp with a stop codon at bp 1221. The putative encoded protein is 108 amino acid in length of 12.04 kDa predicted molecular weight and pI = 9.76.

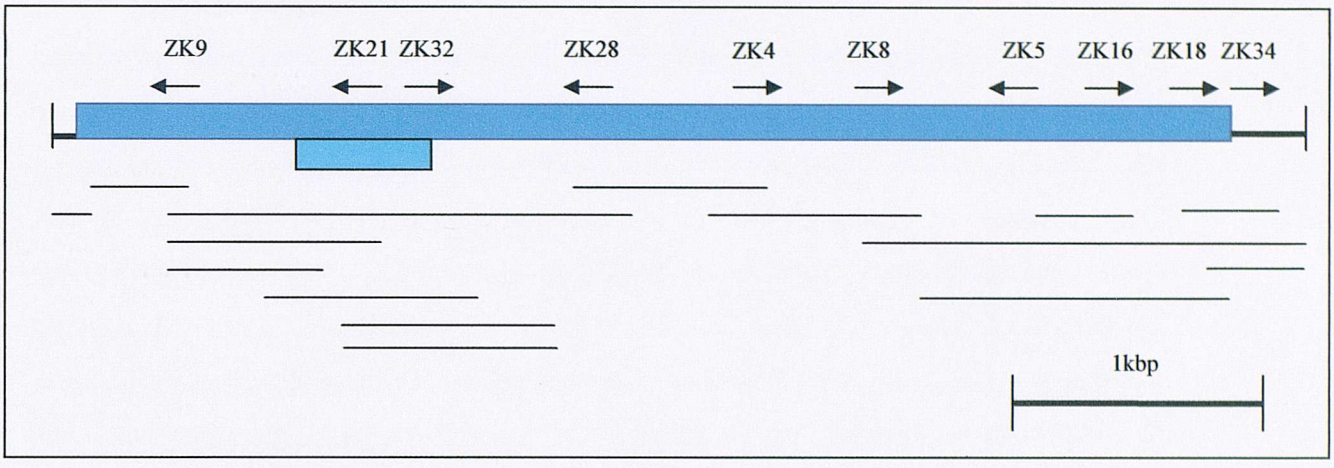


Figure 3.10: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of ACD dsRNA3 species. The approximate position of the primers is shown. The shaded boxes represent the two contiguous ORF, while the 5' and 3' UTR are shown as a straight line.

Table 3.4: Sequence of the primers used in the cDNA cloning of ACD dsRNA3. The sequence of the primers designed for dsRNA3 is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
ZK4	GGGCCAAGTGTGACTCTATGG
ZK5	GCGCTTGCGGTTCTCTCTATTG
ZK8	GCGTAAAGCATTCAGTCAGGG
ZK9	GTCATTATTGATGAATGGCGATC
ZK16	CTGCAGGCTCGGCTTTCACG
ZK18	GACGGGACTGTAGTTGAAGC
ZK21	CCACCAGCTCACTTGTGATGC
ZK28	GCGTCTGAATCCGAAGTGG
ZK32	GTTGCATCACAAGTGAGCTGG
ZK34	CAACTTGACTTACCAATCACG

The C-terminal region of the dsRNA3 predicted protein product showed very significant similarity to the totiviral RdRp proteins at the peptide sequence alignment level (Figure 3.11). Thus at least one part of the predicted protein product constitutes a novel, viral RdRp. It should be noted that all of the RdRp motifs are present in ACD dsRNAs 2 and

3, but not as stricty conserved as in other totiviruses. In general, though, these motifs seem to be more strictly conserved in some mycoviruses, such as partitiviruses, but not in others, *e.g.* totiviruses. As with ACD dsRNA2 the highest RdRp similarity scores are to totiviruses and the dsRNA element is 4.3 kbp long, which is within the average molecular size of *Totiviridae*, it was investigated whether dsRNA3 element represents a novel totiviral species. This hypothesis was dismissed, *i.e.* dsRNA3 element is not a novel totivirus, for the same reasons as for the ACD dsRNA2 element. Namely, ACD dsRNA3 is predicted to encode a single contiguous ORF, instead of two overlapping ones. There are no internal transcription initiation sites (IRES) on the genome or any other *cis*-elements, such as potential protease cleavage sites, that could suggest the possibility of two different peptides made from the one single ORF, initially expressed as a fusion protein. However, based on significant similarities to the totiviruses, dsRNA3 could represent a novel group of totivirus-like viruses.

Nucleotide and protein sequence alignments of the part of the genome upstream of the conserved RdRp motifs (N-terminal peptide sequence) did not reveal any conserved regions or any similarity to other known sequences. Thus, at present it is not known whether the large ORF produces more than one protein product and in the latter case, what the role of the additional product is. This result also indicates that the similarities shown in section 3.2.2.2 concerning the N-terminal part of ACD dsRNA2 might be artefactual.

The two peptide sequences of dsRNA2 and dsRNA3 when aligned against each other have a 66.8% sequence identity and 81.2% sequence similarity. Thus, it can be said that the two dsRNA elements encode at least one protein product with the same replication functions as afforded by an RdRp and are evolutionarily very closely related, but they are not identical, *i.e.* they are not variants of the same dsRNA molecule. Furthermore, their peptide sequence differences, though present mostly in the N-terminal part are not concentrated in a single specific area, but are spread evenly along the sequence. The smaller predicted ORF when aligned against other known proteins (WU-BLAST2 alignment) showed some limited similarity to tryptase protein precursors and some other

functionally uncharacterised plant proteins (Figure 3.12). Though there is the possibility of this product being expressed and potentially being catalytically active, this can only be confirmed by isolating and purifying the specific protein product from ACD-infected plant tissue.

Figure 3.11: CLUSTAL_ X alignment results of the peptide sequence for the ACD dsRNA3 element ORF highlighted using the GeneDoc software application. The entire peptide sequence is shown aligned against nine totiviral RdRp peptide sequences and the entire gag-pol fusion protein of ScV-L-A. The abbreviations and peptide sequence GenBank accession numbers are the same as in Figure 3.8. The conserved RdRp motifs (Bruenn 2003) are highlighted. Abbreviations of assigned viruses are italicised, of unassigned are not.

```
                    *       180         *       200         *       220         *       240
LRV 1-1  : ------------------------------------------------------------------------------:   -
LRV 2-1  : ------------------------------------------------------------------------------:   -
SsRV-1   : ------------------------------------------------------------------------------:   -
HvV190S  : ------------------------------------------------------------------------------:   -
HmTV1-17 : ------------------------------------------------------------------------------:   -
GaRV-L1  : ------------------------------------------------------------------------------:   -
SsRV-2   : ------------------------------------------------------------------------------:   -
TVV-1    : ------------------------------------------------------------------------------:   -
ScV-L-A  : VTIDDTAGPTQFAWPSDRSTDSYPDWAQFSESFPSIDVPYLDVRPLTVTEVNFVLMMMSKWHRRTNLAIDYEAPQLADKF:  240
ScV-L-BC : ------------------------------------------------------------------------------:   -
dsRNA3   : ------------------------------------------------------------MAHHMFMENLSVSNDQ:   16

                    *       260         *       280         *       300         *       320
LRV 1-1  : ------------------------------------------------------------------------------:   -
LRV 2-1  : ------------------------------------------------------------------------------:   -
SsRV-1   : ------------------------------------------------------------------------------:   -
HvV190S  : ------------------------------------------------------------------------------:   -
HmTV1-17 : ------------------------------------------------------------------------------:   -
GaRV-L1  : ------------------------------------------------------------------------------:   -
SsRV-2   : ------------------------------------------------------------------------------:   -
TVV-1    : ------------------------------------------------------------------------------:   -
ScV-L-A  : AYRHALTVQDADEWIEGDRTDDQFRPPSSKVMLSALRKYVNHNRLYNQFYTAAQLLAQIMMKPVPNCAEGYAWLMHDALV:  320
ScV-L-BC : ------------------------------------------------------------------------------:   -
dsRNA3   : GSPFINNDELPEHQRARRAPNSSTGASGRGNPGAWVMYPDDKLADFISFLANTEIDDMCSASEVGIAPELIKVLRSTGVR:   96

                    *       340         *       360         *       380         *       400
LRV 1-1  : ------------------------------------------------------------------------------:   -
LRV 2-1  : ------------------------------------------------------------------------------:   -
SsRV-1   : ------------------------------------------------------------------------------:   -
HvV190S  : ------------------------------------------------------------------------------:   -
HmTV1-17 : ------------------------------------------------------------------------------:   -
GaRV-L1  : ------------------------------------------------------------------------------:   -
SsRV-2   : ------------------------------------------------------------------------------:   -
TVV-1    : ------------------------------------------------------------------------------:   -
ScV-L-A  : NIPKFGSIRGRYPFLLSGDAALIQATALEDWSAIMAKPELVFTYAMQVSVALNTGLYLRRVKKTGFGTTIDDSYEDGAFL:  400
ScV-L-BC : ------------------------------------------------------------------------------:   -
dsRNA3   : TSPLDAFADPNYCVCDNELRFIDHREAALNLKKWVGSAIVQKKQYAMGVHPMTSPVPNEDPMLLSFAPQSLKTIAVLGGA:  176
```

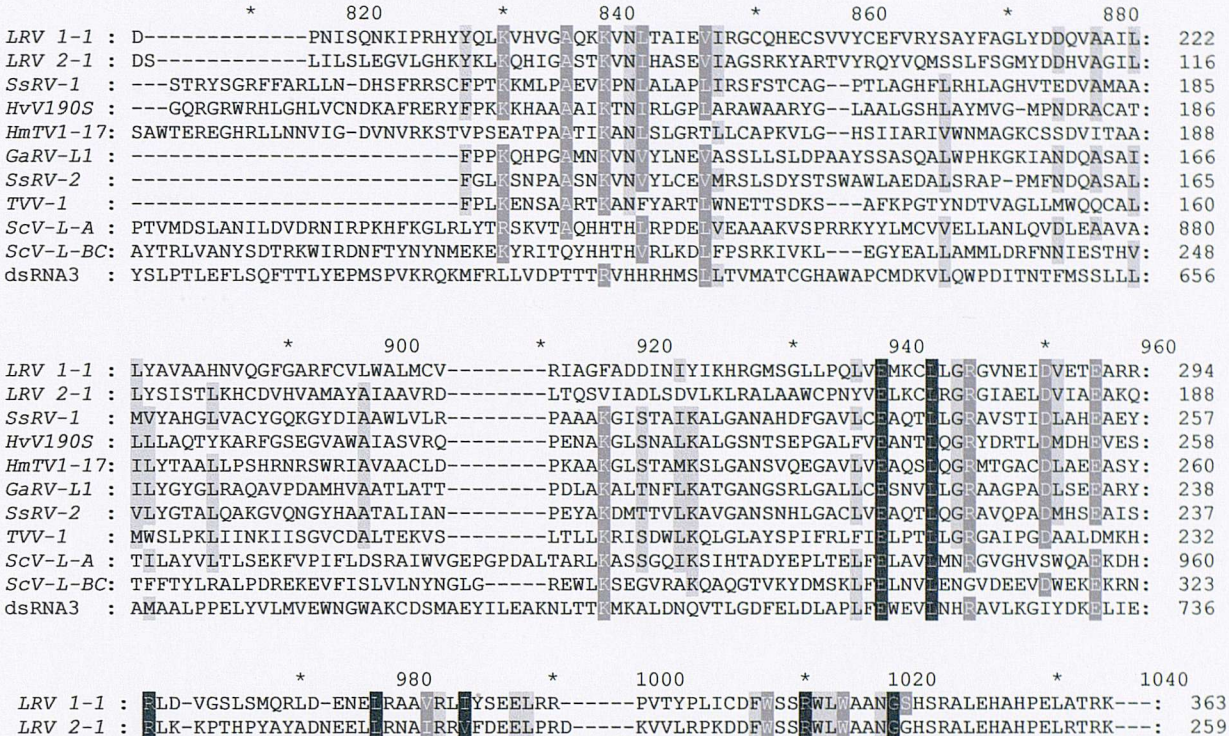
```
                    *       420         *       440         *       460         *       480
LRV 1-1 : ------------------------------------------------------------------------------:    -
LRV 2-1 : ------------------------------------------------------------------------------:    -
SsRV-1  : ------------------------------------------------------------------------------:    -
HvV190S : ------------------------------------------------------------------------------:    -
HmTV1-17: ------------------------------------------------------------------------------:    -
GaRV-L1 : ------------------------------------------------------------------------------:    -
SsRV-2  : ------------------------------------------------------------------------------:    -
TVV-1   : ------------------------------------------------------------------------------:    -
ScV-L-A : QPETFVQAALACCTGQDAPLNGMSDVYVTYPDLLEFDAVTQVPITVIEPAGYNIVDDHLVVVGVPVACSPYMIFPVAAFD:  480
ScV-L-BC: ------------------------------------------------------------------------------:    -
dsRNA3  : RAVTTDIAAHACQLLLNQQAEGRSERDQVILNAVLLHTMESRASIRPMKIQSRMGSNHAVVDPMNAISVKHSVTEDGYLR:  256

                    *       500         *       520         *       540         *       560
LRV 1-1 : ------------------------------------------------------------------------------:    -
LRV 2-1 : ------------------------------------------------------------------------------:    -
SsRV-1  : ------------------------------------------------------------------------------:    -
HvV190S : ------------------------------------------------------------------------------:    -
HmTV1-17: ------------------------------------------------------------------------------:    -
GaRV-L1 : ------------------------------------------------------------------------------:    -
SsRV-2  : ------------------------------------------------------------------------------:    -
TVV-1   : ------------------------------------------------------------------------------:    -
ScV-L-A : TANPYCGNFVIKAANKYLRKGAVYDKLEAWKLAWALRVAGYDTHFKVYGDTHGLTKFYADNGDTWTHIPEFVTDGDVMEV:  560
ScV-L-BC: ------------------------------------------------------------------------------:    -
dsRNA3  : EYSDFCEANRSTISTMMAKCMKAGIAHHGLDGLRTEEASASSKLASKRAGDVGQSKAMAAGDISRVDAMVVASQVSWWHP:  336

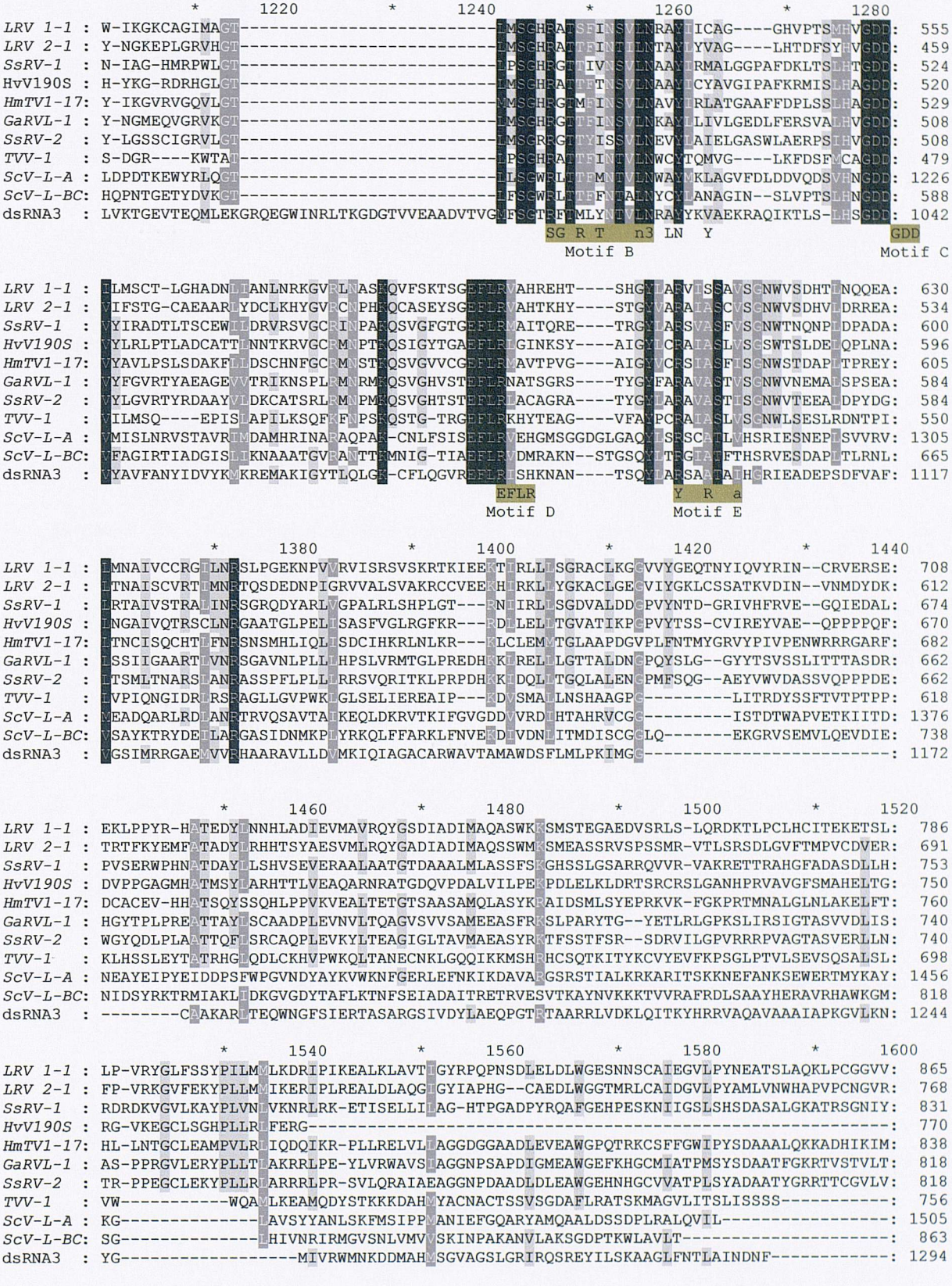
                    *       580         *       600         *       620         *       640
LRV 1-1 : ------------------------------------------------------------------MQCPNQ:    6
LRV 2-1 : ------------------------------------------------------------------------------:    -
SsRV-1  : ------------------------------------------------------------------------------:    -
HvV190S : ------------------------------------------------------------------------------:    -
HmTV1-17: ------------------------------------------------------------------------------:    -
GaRV-L1 : ------------------------------------------------------------------------------:    -
SsRV-2  : ------------------------------------------------------------------------------:    -
TVV-1   : ------------------------------------------------------------------------------:    -
ScV-L-A : FVTAIERRARHFVELPRLNSPAFFRSVEVSTTIYDTHVQAGAHAVYHASRINLDYVKPVSTGIQVINAGELKNYWGSVRR:  640
ScV-L-BC: ----------------------------------------------------------FSCSSLRDTKF:   11
dsRNA3  : GVTMWIAAYKMSVAMLSSAKALTLSDKEVQCLVEYTFSRTTYRKLVASNALMDSTRDIAASEVTQAASTPIRWERQVHPL:  416

                    *       660         *       680         *       700         *       720
LRV 1-1 : NHMLVNRAMVVAALDSFEDARRIISGVLDLSRLTNTSVKGQHTNDTNYFSRYSNFFSQMPA----------IILNQLKC:   75
LRV 2-1 : ------------------------------------------------------------------------------:    -
SsRV-1  : ------------------------------------MNKFIDRAEQFGPVGDYLLSVVER----------FPTVHS-Y:   31
HvV190S : ------------------------------------MSDPQERSKAYGLLGERLYAVASA----------NSHMLAGY:   32
HmTV1-17: ----------------------------------MKALRAAQDKAGELGPLGSALLERVLA----------KPEVFERV:   35
GaRV-L1 : ----------------------------------MIESPVSVRANEAGIVGQYLKGLLN------------MEWASGV:   32
SsRV-2  : ----------------------------------MSNDRLHSRVEATGPVGVYLRSLVP------------LELAAVV:   32
TVV-1   : --------------------------------DPFLSKAVRCGPVIPSVKHHFNINY---------LSILKHNG:   33
ScV-L-A : TQQGFRSGRSYDASCNAYRRTYSWRCPRRVDRTGGQCFSRVNVIEPSHGPRPTRYILQEPGTYPAWIRFRNRVQAVSRQK:  720
ScV-L-BC: VGQQYAILTPSGTTTDIRSGRGTNQSYRRGRTSTGYRIGVEDDEDLDIGTVKYIVPLYLNGDNVAQNCLEATHVLIKACS:   91
dsRNA3  : VLVLDDAEYSITRKTASVELTSVYNKVHAYMALGLGAMYGDIAKTGMQQPNSIGTGLLARSGRSQRASPVFARVQLTESA:  496

                    *       740         *       760         *       780         *       800
LRV 1-1 : CIKAQVDDAVKMVLQKAKKVEVSKPVEKMLTFTTLNTYLGYPESTGCVMEYTEEQSGPIAAKLLITLLSSTLNAMVRPKS:  155
LRV 2-1 : ----------------MRAKLTKVVEGDLSVFMPNARIKR----------ESQN----ASDLHHFLESVVEGIAIPQP:   48
SsRV-1  : GELSFVDALTRLRAMSNNLHANDALLPAAVSLLCLPFPLQVEMSRADVYALLVKGTSLPFNFVRSTAHLHPLPDDVGRPG:  111
HvV190S : DSLDFTARLVRLTGEATALKAVDPLLPCAVSLLFMDFPLQLPCTPEETLRLVRRAYDPNTLEEVDYSTLSGYATQFARVK:  112
HmTV1-17: SRGNIEE----QNMEVANLQENGMGVDAAAFLSSLSFPIQVKLNQADIITLARLATDTRAYCPLINRTLDTAVARKAGVR:  111
GaRV-L1 : MVLSFSQQISEVYKPTFNGLRPTDLQRAAAAYLVPDFPVQVRIERGSVLSLLSQVIDPPARVTDRCSFRWLCDPKASYAA:  112
SsRV-2  : SRLPFDRQISFIYSPQWQGRRPTGIQRLAGAFLLPQVPVQVSVTDADLLLLLTNTLPPAFPKRIHRQAGWSYSDRATLAE:  112
TVV-1   : NEYTFVPGYGWVLQDDYLLNAVKMVGEGDLPPNQLPYDDDLLFTYAKILLYDYISHFPEFRHKNPRLLTSETELQL----:  109
ScV-L-A : ATHFLFDIVPAAVISDFTTSDTSSFAYKSHTYAVNVTALRFSDTYALYVQTDTNMTILSPAARRQASATYSQVAGFCYNT:  800
ScV-L-BC: IANRIVDDGEGHCFTQQGLAQQWIFHRGEMIFVKAVRIGQLNAYYVDYKNVTNYSLKTAAQVGATISNNLRHGFVDNQQD:  171
dsRNA3  : VTVTACPTSDSDAVLMTINKGVEKAGLHNIVLGNEVVKVIKKPEYGMTLTYLIPSTISGKGADKSYVYLAGMHFREDSLS:  576
```



3 KlE gK R 6
Motif F2 Motif F3
D lFNs H
Motif A


LRV 1-1 : W-IKGKCAGIMAGT----------LMSGHRATSFINSVLNRAYIICAG----GHVPTSMHVGDD: 555
LRV 2-1 : Y-NGKEPLGRVHGT----------LMSGHRATTFINTILNTAYLYVAG----LHTDFSYHVGDD: 459
SsRV-1 : N-IAG-HMRPWLGT----------LPSGHRGTIVNSVLNAAYLRMALGGPAFDKLTSLHTGDD: 524
HvV190S : H-YKG-RDRHGLGT----------LMSGHRATTFTNSVLNAAYICYAVGIPAFKRMISLHAGDD: 520
HmTV1-17: Y-IKGVRVGQVLGT----------MMSGHRGTMFINSVLNAVYIRLATGAAFFDPLSSLHAGDD: 529
GaRVL-1 : Y-NGMEQVGRVKGT----------LMSGHRGTTFINSVLNKAYLLIVLGEDLFERSVALHVGDD: 508
SsRV-2 : Y-LGSSCIGRVLGT----------LMSGRRGTTYISSVLNEVYLAIELGASWLAERPSIHVGDD: 508
TVV-1 : S-DGR----KWTAT----------LPSGHRATTFINTVLNWCYTQMVG----LKFDSFMCAGDD: 479
ScV-L-A : LDPDTKEWYRLQGT----------LLSGWRLTTFMNTVLNWAYMKLAGVFDLDDVQDSVHNGDD: 1226
ScV-L-BC: HQPNTGETYDVKGT----------LFSGWRLTTFFNTALNYCYLANAGIN--SLVPTSLHNGDD: 588
dsRNA3 : LVKTGEVTEQMLEKGRQEGWINRLTKGDGTVVEAADVTVGMFSGTRFTMLYNTVLNRAYYKVAEKRAQIKTLS-LHSGDD: 1042
SG R T n3 LN Y GDD
Motif B Motif C
LRV 1-1 : LLMSCT-LGHADNLIANLNRKGVRLNASHQVFSKTSGEFLRVAHREHT----SHGTLAEVISSAVSCNWVSDHTLNQQEA: 630
LRV 2-1 : VIFSTG-CAEAARLYDCLKHYGVRCNPHRQCASEYSGEFLRVAHTKHY----STGYVAEAIASCVSCNWVSDHVLDRREA: 534
SsRV-1 : VYIRADTLTSCEWILDRVRSVGCRINPARQSVGFGTGEFLRMAITQRE----TRGYLAESVASFVSCNWTNQNPLDPADA: 600
HvV190S : VYLRLPTLADCATTLNNTKRVGCRMNPTRQSIGYTGAEFLRLGINKSY----AIGYLCEAIASLVSCSWTSLDELQPLNA: 596
HmTV1-17: VYAVLPSLSDAKFLLDSCHNFGCRMNSTRQSVGVVCGEFLRMAVTPVG----AIGYVCESIASFISCNWSTDAPLTPREY: 605
GaRVL-1 : VYFGVRTYAEAGEVVTRIKNSPLRMNRMRQSVGHVSTEFLRNATSGRS----TYGYFAEAVASTVSCNWVNEMALSPSEA: 584
SsRV-2 : VYLGVRTYRDAAYVLDKCATSRLRMNPMRQSVGHTSTEFLRLACAGRA----TYGYLAEAVASTISCNWVTEEALDPYDG: 584
TVV-1 : VILMSQ----EPISLAPILKSQFKFNPSRQSTG-TRGEFLRKHYTEAG----VFAYPCEAIASLVSCNWLSESLRDNTPI: 550
ScV-L-A : VMISLNRVSTAVRIMDAMHRINARAQPAR-CNLFSISEFLRVEHGMSGGDGLGAQYLSESCATLVHSRIESNEPLSVVRV: 1305
ScV-L-BC: VFAGIRTIADGISLIKNAAATGVRANTFMNIG-TIAEFLRVDMRAKN--STGSQYLTEGIAIFTHSRVESDAPLTLRNL: 665
dsRNA3 : VYAVFANYIDVYKMKREMAKIGYTLQLGK-CFLQGVREFLRISHKNAN----TSQYLAESAATATHCRIEADEPSDFVAF: 1117
EFLR Y R a
Motif D Motif E
LRV 1-1 : LMNAIVCCRGILNRSLPGEKNPVVRVISRSVSKRTKIEEKTIRLLLSGRACLKGCVVYGEQTNYIQVYRIN--CRVERSE: 708
LRV 2-1 : LTNAISCVRTIMNRTQSDEDNPIGRVVALSVAKRCCVEEKHIRKLLYGKACLGECVIYGKLCSSATKVDIN--VNMDYDK: 612
SsRV-1 : LRTAIVSTRALINRSGRQDYARLVGPALRLSHPLGT---RNIIRLLSGDVALDDGPVYNTD-GRIVHFRVE--GQIEDAL: 674
HvV190S : LNGAIVQTRSCLNRGAATGLPELISASFVGLRGFKR---RDLLELLTGVATIKPGPVYTSS-CVIREYVAE--QPPPPQF: 670
HmTV1-17: LTNCISQCHTLFNRSNSMHLIQLISDCIHKRLNLKR---KLCLEMVTGLAAPDGVPLFNTMYGRVYPIVPENWRRRGARF: 682
GaRVL-1 : LSSIIGAARTLVNRSGAVNLPLLLHPSLVRMTGLPREDHKKLRELLLGTTALDNGPQYSLG--GYYTSVSSLITTTASDR: 662
SsRV-2 : LTSMLTNARSLANRASSPFLPLLLRRSVQRITKLPRPDHKKIDQLLTGQLALENGPMFSQG--AEYVWVDASSVQPPPDE: 662
TVV-1 : LVPIQNGIDRLRSRAGLLGVPWKLGLSELIEREAIP---KDVSMALLNSHAAGPG---------LITRDYSSFTVTPTPP: 618
ScV-L-A : MEADQARLRDLANRTRVQSAVTALKEQLDKRVTKIFGVGDDVVRDLHTAHRVCGG---------ISTDTWAPVETKIITD: 1376
ScV-L-BC: VSAYKTRYDEILARGASIDNMKPLYRKQLFFARKLFNVEKDIVDNLITMDISCGGLQ-------EKGRVSEMVLQEVDIE: 738
dsRNA3 : VGSIMRRGAEMVVRHAARAVLLDVMKIQIAGACARWAVTAMAWDSFLMLPKIMGG----------: 1172
LRV 1-1 : EKLPPYR-HATEDYLNNHLADIEVMAVRQYGSDIADIMAQASWKKSMSTEGAEDVSRLS-LQRDKTLPCLHCITEKETSL: 786
LRV 2-1 : TRTFKYEMFATADYLRHHTSYAESVMLRQYGADIADIMAQSSWMKSMEASSRVSPSSMR-VTLSRSDLGVFTMPVCDVER: 691
SsRV-1 : PVSERWPHNATDAYLLSHVSEVERAALAATGTDAAALMLASSFSKGHSSLGSARRQVVR-VAKRETTRAHGFADASDLLH: 753
HvV190S : DVPPGAGMHATMSYLARHTTLVEAQARNRATGDQVPDALVILPEKPDLELKLDRTSRCRSLGANHPRVAVGFSMAHELTG: 750
HmTV1-17: DCACEV-HHATSQYSSQHLPPVKVEALTETGTSAASAMQLASYKRAIDSMLSYEPRKVK-FGKPRTMNALGLNLAKELFT: 760
GaRVL-1 : HGYTPLPREATTAYLSCAADPLEVNVLTQAGVSVVSAMEEASFRKSLPARYTG--YETLRLGPKSLIRSIGTASVVDLIS: 740
SsRV-2 : WGYQDLPLAATTQFLSRCAQPLEVKYLTEAGIGLTAVMAEASYRKTFSSTFSR--SDRVILGPVRRRPVAGTASVERLLN: 740
TVV-1 : KLHSSLEYTATRHGLQDLCKHVPWKQLTANECNKLGQQIKKMSHRHCSQTKITYKCVYEVFKPSGLPTVLSEVSQSALSL: 698
ScV-L-A : NEAYEIPYEIDDPSFWPGVNDYAYKVWKNFGERLEFNKIKDAVARGSRSTIALKRKARITSKKNEFANKSEWERTMYKAY: 1456
ScV-L-BC: NIDSYRKTRMIAKLIDKGVGDYTAFLKTNFSEIADAITRETRVESVTKAYNVKKKTVVRAFRDLSAAYHERAVRHAWKGM: 818
dsRNA3 : --------CAAKARLTEQWNGFSIERTASARGSIVDYLAEQPGTFTAARRLVDKLQITKYHRRVAQAVAAAIAPKGVLKN: 1244
LRV 1-1 : LP-VRYGLFSSYPILMVLKDRIPIKEALKLAVTIGYRPQPNSDLELDLWGESNNSCAIEGVLPYNEATSLAQKLPCGGVV: 865
LRV 2-1 : FP-VRKGVFEKYPLLMVIKERIPLREALDLAQGIGYIAPHG--CAEDLWGGTMRLCAIDGVLPYAMLVNWHAPVPCNGVR: 768
SsRV-1 : RDRDKVGVLKAYPLVNLVKNRLRK-ETISELLILAG-HTPGADPYRQAFGEHPESKNIIGSLSHSDASALGKATRSGNIY: 831
HvV190S : RG-VKEGCLSGHPLLRLFERG----------: 770
HmTV1-17: HL-LNTGCLEAMPVLRLIQDQIKR-PLLRELVLLAGGDGGAADLEVEAWGPQTRKCSFFGWIPYSDAAALQKKADHIKIM: 838
GaRVL-1 : AS-PPRGVLERYPLLTLAKRRLPE-YLVRWAVSLAGGNPSAPDIGMEAWGEFKHGCMIATPMSYSDAATFGKRTVSTVLT: 818
SsRV-2 : TR-PPEGCLEKYPLLRLARRRLPR-SVLQRAIAEAGGNPDAADLDLEAWGEHNHGCVVATPLSYADAATYGRRTTCGVLV: 818
TVV-1 : VW----------WQALKEAMQDYSTKKKDAHMYACNACTSSVSGDAFLRATSKMAGVLITSLISSSS----------: 756
ScV-L-A : KG----------IAVSYYANLSKFMSIPPMANIEFGQARYAMQAALDSSDPLRALQVIL----------: 1505
ScV-L-BC: SG----------LHIVNRIRMGVSNLVMVVSKINPAKANVLAKSGDPTKWLAVLT----------: 863
dsRNA3 : YG----------MIVRWMNKDDMAHMSGVAGSLGRIRQSREYILSKAAGLFNTLAINDNF----------: 1294

```
LRV 1-1 : IQVIHNVYV:  874
LRV 2-1 : LVVDHNVYA:  777
SsRV-1  : TLTPVFV--:  838
HvV190S : ---------:    -
HmTV1-17: NHTPVRM--:  845
GaRVL-1 : CPLDAHV--:  825
SsRV-2  : CGLHFYV--:  825
TVV-1   : ---------:    -
ScV-L-A : ---------:    -
ScV-L-BC: ---------:    -
dsRNA3  : ---------:    -
```

Reference: The designated viral names, GenBank accession numbers and references are the same as in Figure 3.8.

Figure 3.12: WU-BLAST2 alignment results of the short peptide sequence of the ACD dsRNA3 element. The complete peptide sequence is shown aligned pairwise to other peptide sequences. The three highest scoring alignments are shown. The identity and similarity scores between the two sequences are shown, as well as the UNIPROT accession number, the protein name and the organism of origin for each one of the aligned protein sequences.

1) UNIPROT: Q69KY6_ORYSA Q69KY6, Hypothetical protein OSJNBa0095O23.20. Length = 337, Identities = 23/66 (34%), Positives = 29/66 (43%) (*Oryza sativa*)

```
dsRNA3: 5 WYSAPRPGRAAYRGGICKFEASKQEGRRRWPKQGHGRWRHKSCGCN-GGCITSELVAPWR
          W +A  P  AA  GG  +  A+  E RRR       R+R K  G   GG +T E+   WR
Sbjct :62 WRTATLPRAAAANGGDRQRAAAHGERRRRNATTRRRRYRRKREGEERGGVLTEEVRPAWR

dsRNA3: 64 HDVDSG 69
                G
Sbjct: 122 ETTLEG 127
```

2) UNIPROT: Q41993_ARATH Q41993 LHCII, A/B BINDING PROTEIN TYPEIII (Fragment). Length = 77, Identities = 21/61 (34%), Positives = 24/61 (39%) (*Arabidopsis thaliana*)

```
dsRNA3: 39 HGRWRHKSCGCNGGCITSELVAPW-RHDVDSGIQ-DVRSHAVQRKGFDALGQGSTVPSGV
           HGRW     G  G CIT E++  W R D    +    R H    K F  L Q    P
Sbjct:  20 HGRWA--MLGAFG-CITPEVLQKWVRVDFKEPVWLQSRVHKSSPKRFGLLRQPKPSPCSE

dsRNA3: 97 H 97
           H
Sbjct:  77 H 77
```

3) UNIPROT, TRYT_SHEEP Q9XSM2 Tryptase 2 precursor (EC 3.4.21.59). Length = 273, Identities = 30/107 (28%), Positives = 41/107 (38%) (*Ovis aries*)

```
dsRNA3: 7 SAPRPGRAAYRGGICKFEASKQEGRRRWPKQGHGR-----WRHKSCGCNGGCITSELVAP
          +AP PG+A  R GI      K+    RWP Q   R     WRH+   C G  I  + V
Sbjct: 16 AAPAPGQALQRSGII---GGKEAPGSRWPWQVSLRVRDQYWRHQ---CGGSLIHPQWVLT

dsRNA3: 62 WRHDVDSGIQDVRSHAVQRKGFDALGQGSTVPSGVHIQQNHLQEVSS 108
             H +   +Q+     VQ +      Q   +P    I   H   V +
Sbjct:  70 AAHCIGPELQEPSDFRVQLREQHLYYQDRLLPISRVIPHPHYYMVEN 116
```

3.2.1.2.5 ACD dsRNA4

One of the products of the RLM-RACE reactions, performed using the gel extracted ACD dsRNA elements 1-4 as template, was not related to these totivirus-like dsRNAs neither to the characterised chrysovirus and partitivirus species. Sequence alignment (using the BLAST search tool) showed that it is very similar to the 5' end of *Ophiostoma novo-ulmi* mitovirus 6-Ld mRNA encoding for a putative RdRp (Hong *et al.* 1999). The two sequences share 97% sequence identity, as shown in Figure 3.13.

Table 3.5: Sequence of the primers used in the cDNA cloning of ACD dsRNA4. The sequence of the primers designed for dsRNA3 is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
ZK38	GTACATTGCTACATGAGAG
ZK39	CACGACAGTGACGAATGTC

Figure 3.13: BLAST sequence alignment results using the dsRNA4 fragment. The dsRNA4 fragment is designated as dsRNA4, while the *O.novo-ulmi mitovirus 6-Ld* is shown as 6-Ld.

1) GenBank accession number: AJ132756, *Ophiostoma novo-ulmi mitovirus* 6-Ld mRNA for putative RNA-dependent RNA polymerase, Length=2343, Score = 412 bits (208), Identities = 228/234 (97%)

```
dsRNA4      TGACGTTTAATTGATGGNAAAAAGACAATCAATAATATTCTTAGGTATTCGATATGTTTT  63
            ||||||||||||||||| ||||||||||||||||||||||||  |||| ||||||||||
6-Ld        TGACGTTTAATTGATGGAAAAAAGACAATCAATAATATTCTTAT-TATTTGATATGTTTT  175

dsRNA4      ATTTTGCATTAATTTTAATTGNTTAAGTTTCATTACTTGACATTCGTCACTGTCGTGAGT  123
            |||||  ||||||||||||||| ||||||||||||||||||||||||||||||||||||
6-Ld        ATTTTTCATTAATTTTAATTGTTTAAGTTTCATTACTTGACATTCGTCACTGTCGTGAGT  115

dsRNA4      TGCCGATTAAAAGGTAAAGAACTCCGTAATGTAGGACTATCTTTTTCTTGCGGTCATGTA  183
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
6-Ld        TGCCGATTAAAAGGTAAAGAACTCCGTAATGTAGGACTATCTTTTTCTTGCGGTCATGTA  55

dsRNA4      CATTCGTACATGAGAGTTTTCCAAAACTCAGCTATCCCTTGTAAAAGGGATAGT  237
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
6-Ld        CATTCGTACATGAGAGTTTTCCAAAACTCAGCTATCCCTTGTAAAAGGGATAGT  1
```

3.2.1.3 The 5' and 3' UTR of the ACD dsRNA elements

The 5' and 3' UTR sequences of the four ACD dsRNA elements were studied further for potential secondary structures. As these areas are likely to contain transcription

recognition, initiation and packaging signals, any secondary structures might be directly involved or play a role in the above viral functions. The mfold program was used (Zuker 2003) to produce a range of potential secondary structures. The potential structure with the lowest loop-free energy was considered in each case.

The 5' UTR regions of the dsRNA molecules all seem to adopt a stable hairpin loop structure, while the 3' UTR regions of the dsRNA molecules all seem to adopt a less stable hammerhead-like structure. On the nucleotide sequence level, the 5' UTR regions are more strongly conserved than the 3' UTR regions, though both sets of ends contain distinct areas with nucleotide sequence identity. It is not known whether these are transcription related signals or play a role in the secondary structure formation. The similarity in the sequence as shown by the CLUSTAL_X alignments is reflected in the similarity of the observed secondary structures.

Both ACD dsRNAs 2 and 3 and ACD dsRNAs 1a and 1b are each related to one another and secondary structure comparisons were made between each of the two pairs of sequences. Accordingly, Figures 3.14, 3.15 show the pairwise comparisons of predicted secondary structures for the 5' UTRs of ACD dsRNA species 2 and 3 and of the 5' UTRs of ACD dsRNAs 1a and 1b respectively. Figures 3.16, 3.17 show the pairwise comparisons of predicted secondary structures for the 3' UTRs of dsRNA species 2 and 3 and of the 3' UTRs of dsRNAs 1a and 1b respectively. Figures 3.18 and 3.19 show the sequence alignment of the 5' and 3' UTR ends of the four dsRNA species respectively.

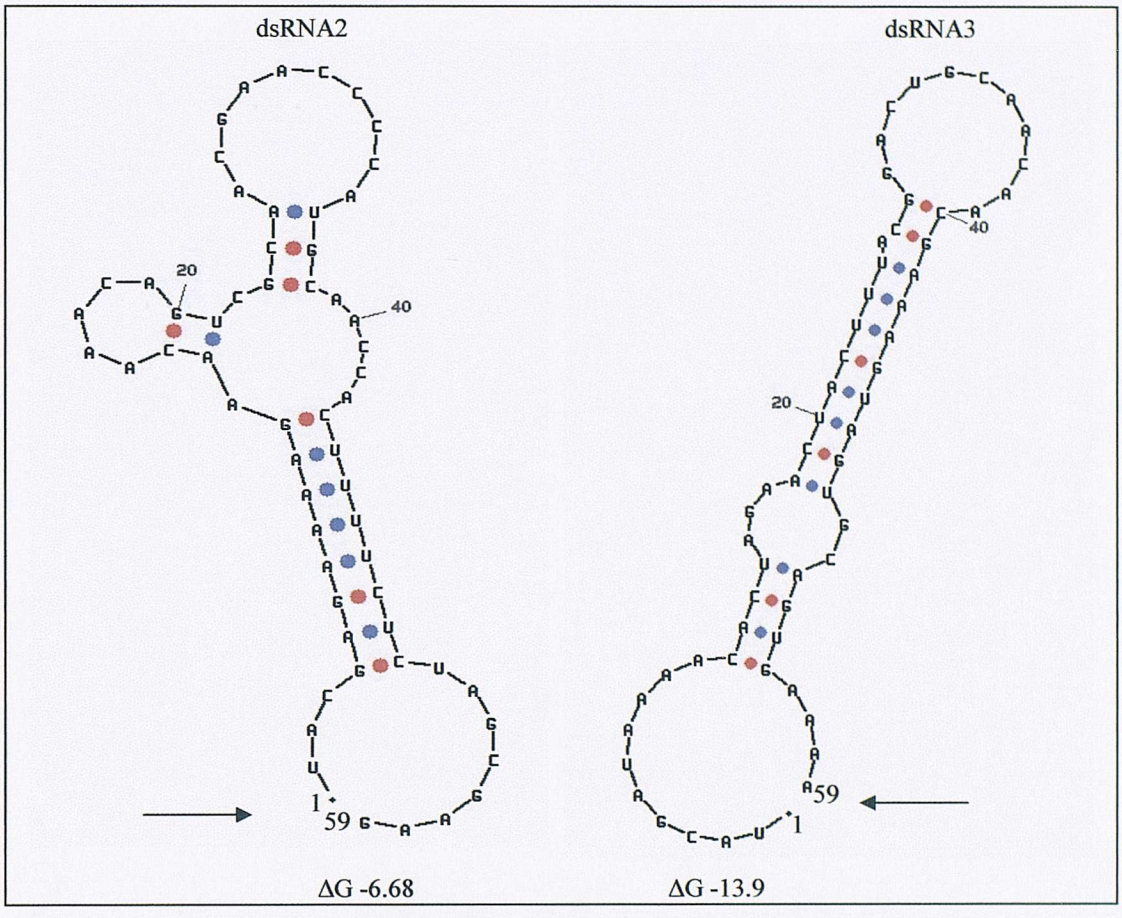


Figure 3.14 (previous page): Predicted folding of the 5' UTR regions for ACD dsRNAs 2 and 3. The potential secondary structures of the 5' termini of the positive strands of dsRNAs 2 and 3 are shown. ΔG values (kcal/mol) were calculated using the mfold program. In both cases a potential stem-loop structure is present. The horizontal arrow points to the start of the sequence.

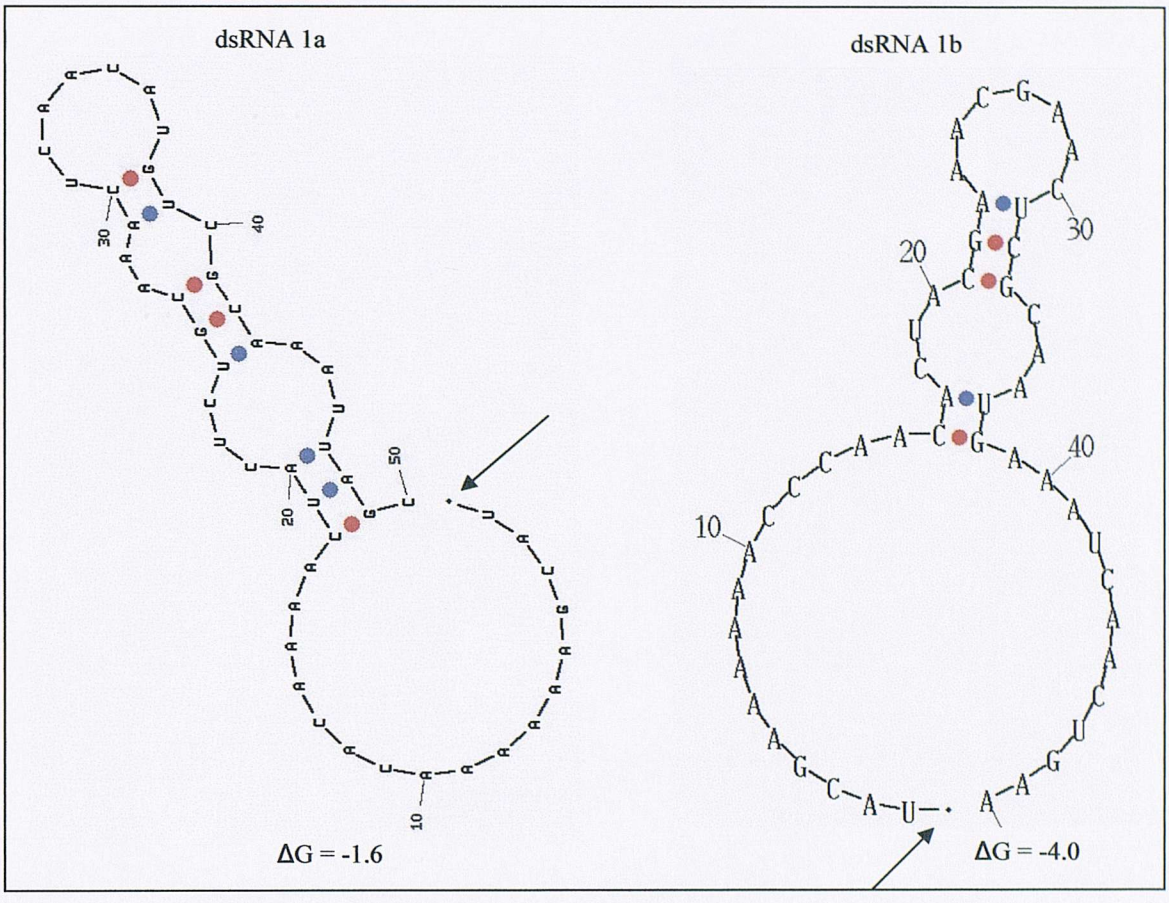


Figure 3.15: Predicted folding of the 5' UTR regions for ACD dsRNAs 1a and **1b.** The potential secondary structures of the 5' termini of the positive strands of dsRNAs 1a and 1b are shown. ΔG values (kcal/mol) were calculated using the mfold program. The horizontal arrow points at the start of the sequence.

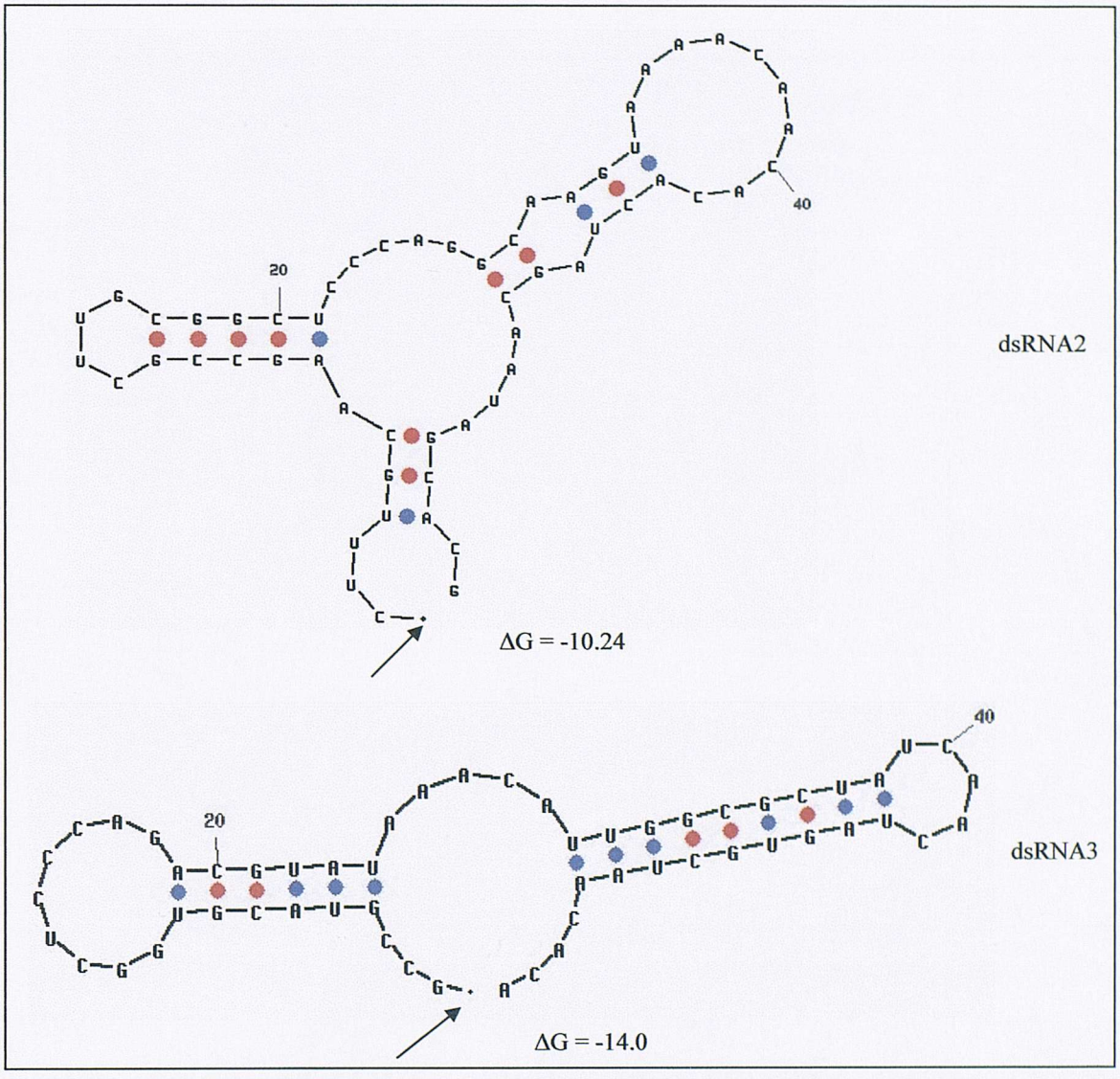


Figure 3.16: Predicted folding of the 3' UTR regions for ACD dsRNAs 2 and 3. The potential secondary structures of the 3' termini of the positive strands of dsRNAs 2 and 3 are shown. ΔG values (kcal/mol) were calculated using the mfold program. The horizontal arrow points at the end of the sequence.

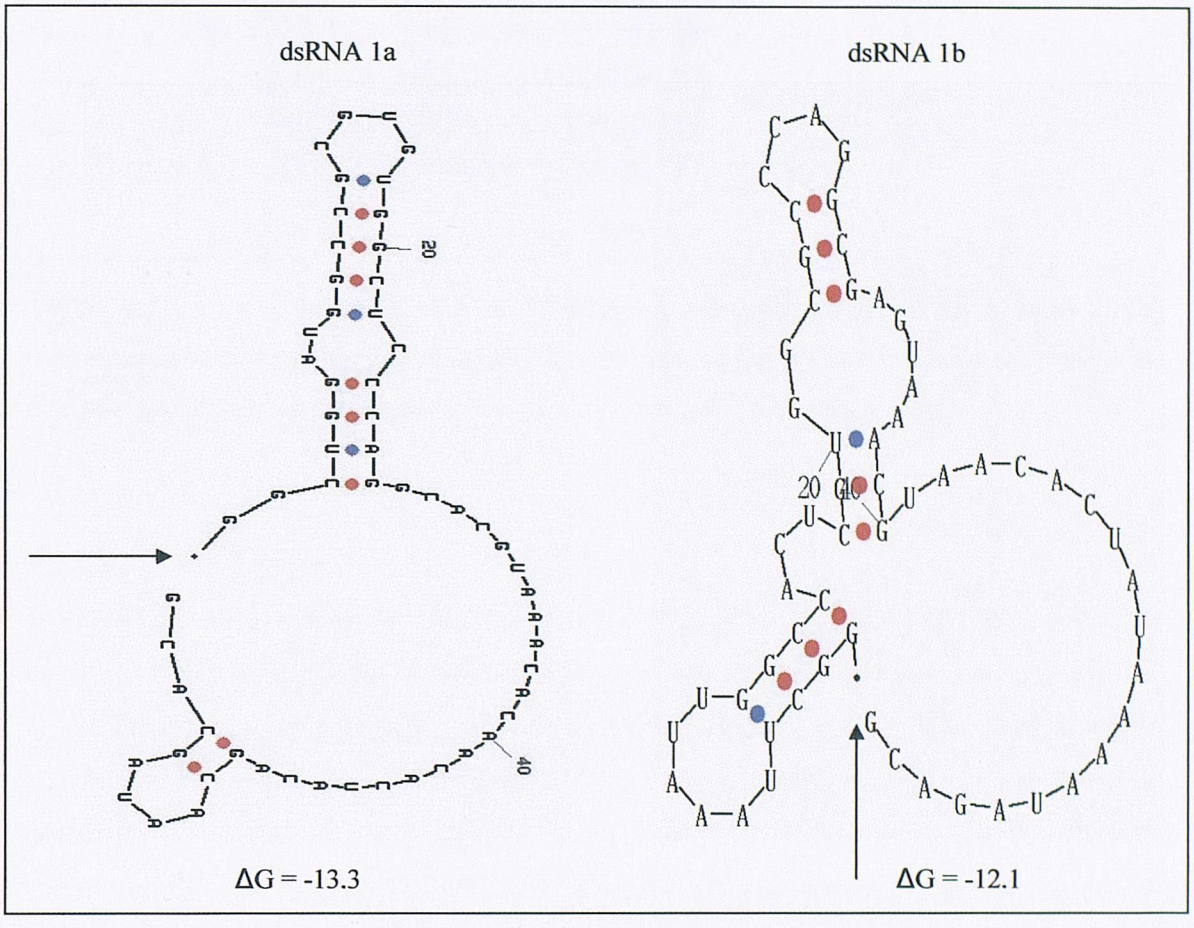


Figure 3.17: Predicted folding of the 3'UTR regions for ACD dsRNAs 1a and 1b. The potential secondary structures of the 5' termini of the positive strands of dsRNAs 1a and 1b are shown. ΔG values (kcal/mol) were calculated using the mfold program. The horizontal arrow points at the end of the sequence.

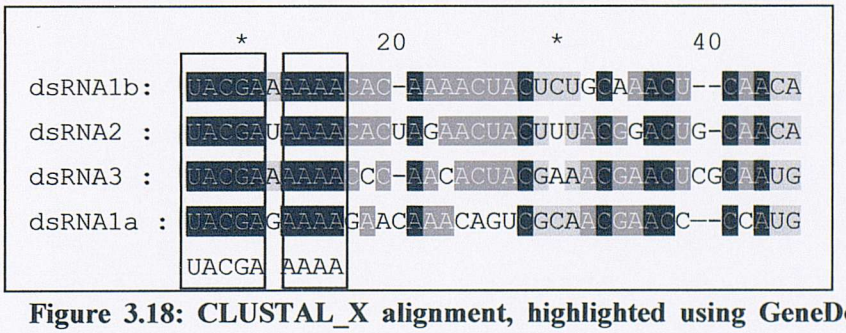


Figure 3.18: CLUSTAL_X alignment, highlighted using GeneDoc, of the 5' ends of the totivirus-like dsRNAs. The first 45 nucleotides of each sequence are shown, the conserved sequences are highlighted, the strictly conserved sequences are boxed.

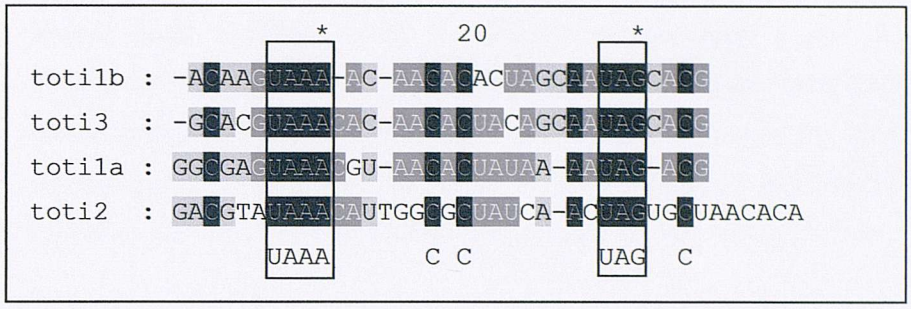


Figure 3.19: CLUSTAL_X alignment, highlighted using GeneDoc, of the 3' ends of the totivirus-like dsRNAs. The last approsimately 35 nucleotides of each sequence are shown, the conserved sequences are highlighted, the strictly conserved sequences are boxed.

3.2.1.4 Northern blots

A series of northern blots was performed using PCR generated fragments from each one of the five dsRNA elements. The fragments were ligated into a cloning vector (pGEM-T Easy) and the resulting plasmid was oligo-labelled with [α^{32}]P dCTP. The preparations of isolated and purified dsRNA were used in the northern blots, while the membranes, following hybridisation, were exposed to an X-Ray film for various lengths of time, depending on the quality and strength of radioactive signalling. The results of the northern blots are shown in Figure 3.20.

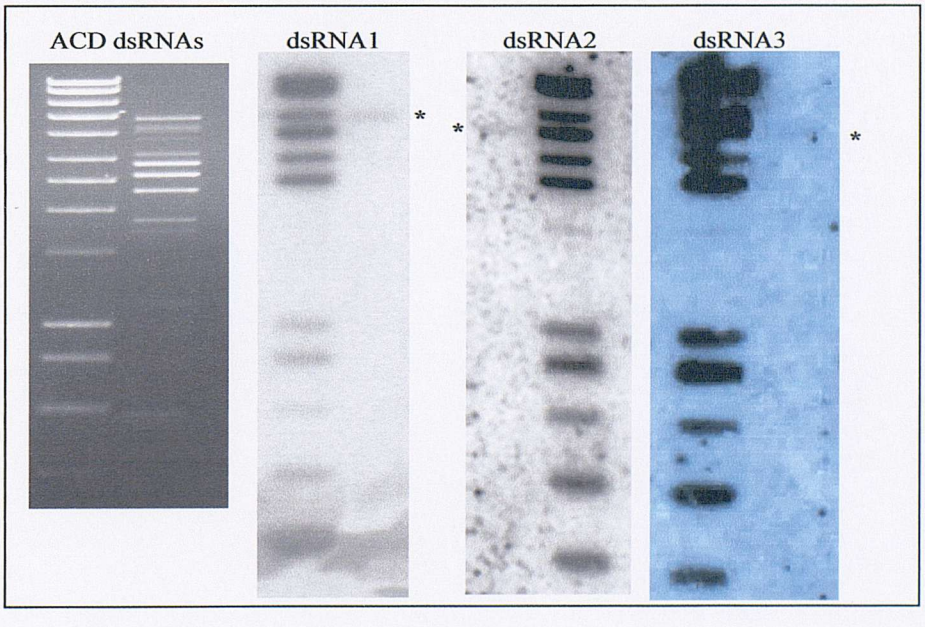


Figure 3.20: Northern blots for dsRNA species 1(a), 2 and 3. The dsRNA species complementary to the dsRNA sequences used for each blot are denoted with an asterisk. The 1 kb ladder is shown in each picture; not every DNA fragment in the ladder hybridised with the labelled plasmid, as not all of the bands derive from a single plasmid. Signal strength is directly related to the concentration of the dsRNA present in the sample each time.

The northern blots confirmed that the sequences obtained corresponded to the ACD-related dsRNA elements and that their initial classification according to the size was correct. The northern blots for ACD dsRNA1b resulted in a strong signal from the same band as ACD dsRNA1a. The blotting for dsRNA4 has been unsuccessful as the particular dsRNA molecule is usually present in very small amounts and therefore its detection is extremely difficult.

3.2.1.5 Viral particle isolation

The aim was to investigate whether any of the ACD associated viruses were encapsidated. Knowing the sizes of the expected viral capsid proteins for each viral species meant that the identification of the potential protein products would be made easier. Viral particles were isolated directly from infected leaf material, provided by Dr. Açikgöz, utilising the method described by Gispert *et al.* (1998) and using PEG_{6000}. The results are shown in Figure 3.21. There was only one clearly visible protein band, corresponding to the expected size of the chrysovirus capsid protein. There was also a much fainter band present, but not in all samples, probably due to much lower concentrations of the viral particle in the infected tissue, corresponding to the expected size of the partitivirus capsid protein. There were no other protein products visible, thus this result indicates that the two novel viral species encoded by dsRNAs 1a, 1b, 2 and 3 are non-encapsidated.

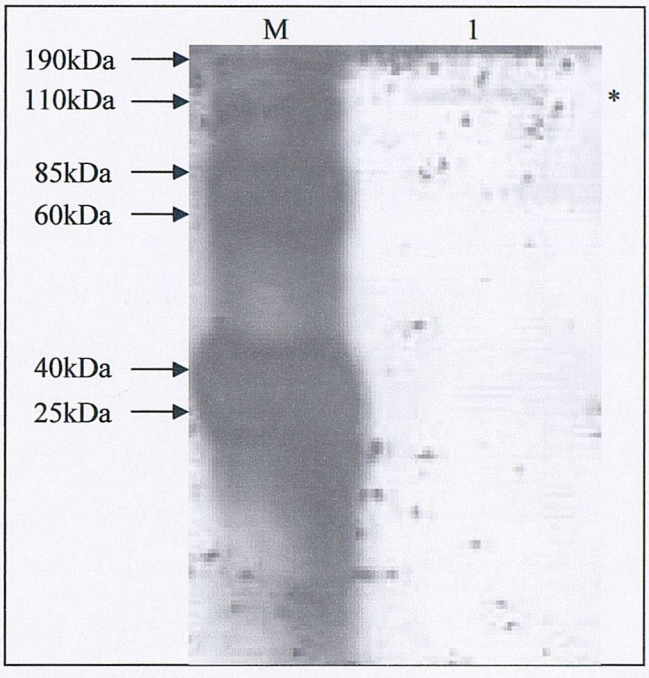


Figure 3.21: Viral particlel isolation from ACD-infected leaf material. The 10% SDS-PAGE gel contains the protein ladder with the sizes of the proteins shown and the sample on lane 1. There is one clearly visible band present, indicated with an asterisk and corresponding to the chrysovirus capsid protein (*ca.* 115 kDa)

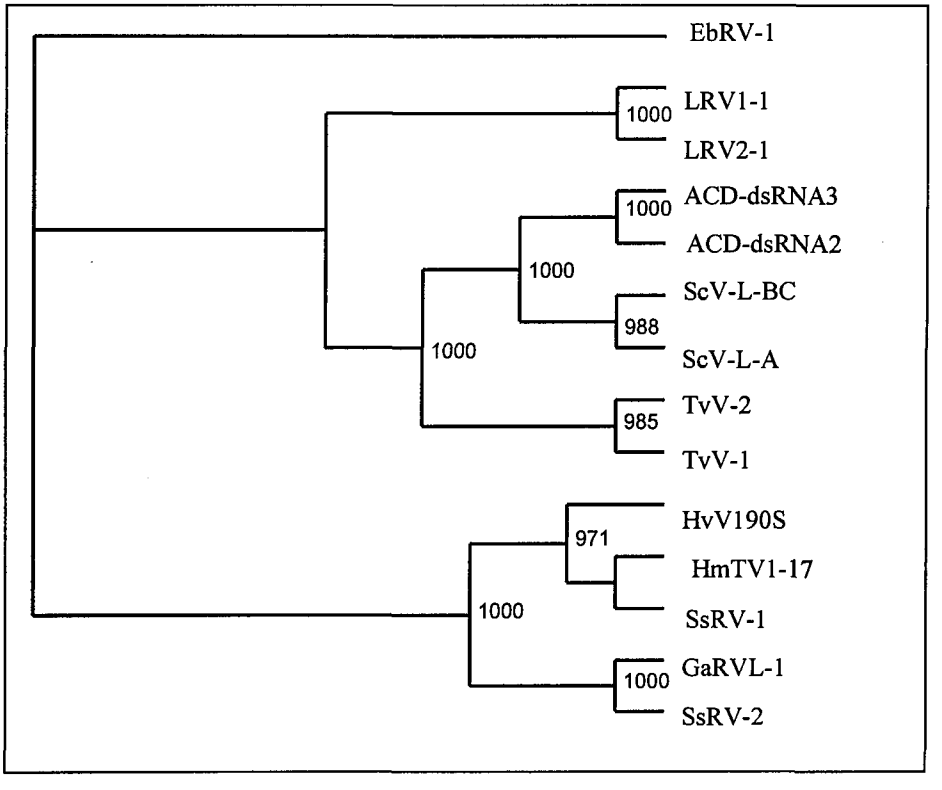
Previous attempts to isolate viral particles have failed, probably due to the relative concentration of the viral particles in infected leaves. The concentration of the obtained protein products suggests that viral particles exist in small quantities in the infected leaf tissue. The tissue itself is woody that makes processing of the material difficult and time-consuming. A further step would be to try and visualise these viral particles using EM.

3.2.1.6 Phylogeny and viral origin

The origin of ACD dsRNAs 1a and 1b is difficult to guess as the two dsRNA elements do not share any similarity with viral or any other sequences present in the GenBank database. It can be said, however that based on the nature of their genome, *i.e.* dsRNA, and their association with the other dsRNAs found in the ACD infected tissues, they are members of a novel mycoviral family, bearing some resemblance to the *Totiviridae*.

Thus, phylogeny could only be investigated for ACD dsRNAs 2 and 3, whose sequence had significant similarity to that of other dsRNA RdRp elements. An unrooted phylogenetic tree based on the neighbour-joining method with a 10,000 replicate bootstrap search was constructed using the same viral sequences as in the individual CLUSTAL_X sequence alignments (Figures 3.8 and 3.11). The result is shown in Figure 3.22. The RdRp regions of dsRNAs 2 and 3 were used for the construction of the tree, because of the absence of any homology in the N-terminal regions of these dsRNA elements. In the tree the totiviral RdRp sequences fall in two major clades. Within the clade containing them, dsRNAs 2 and 3 are grouped together with the yeast totiviral polymerases, but are distinct from the protozoan totiviral polymerases. However, the RdRp sequences of the ACD related dsRNA species form a distinct subgroup, indicating that there is a significant evolutionary distance separating them from their closest related viruses.

Figure 3.22: Unrooted phylogenetic tree based on the Neighbour-joining method with a 10,000 replicate bootstrap search. The tree was constructed using the RdRp peptide sequence of the viruses shown in Figure 3.8.



3.2.2 Sequencing the two smallest dsRNA elements

The dsRNA-rich preparations from ACD infected trees contained (except of the eleven larger dsRNA species) two smaller dsRNA elements, *ca.* 0.6 kbp and 1.1 kbp in size as shown in Figure 3.23. The 0.6 kbp dsRNA species was present in all ACD preparations; the 1.1 kbp species was present in most but not all, and often in much smaller quantities. Both of these dsRNA elements were gel purified and used as a template for RLM-RACE PCR amplification. The two small dsRNA elements are small enough for RLM-RACE to be attempted, without the initial use of any internal primers. However, a single primer internal to the 0.6 kbp dsRNA sequence A was also used (Table 3.6).

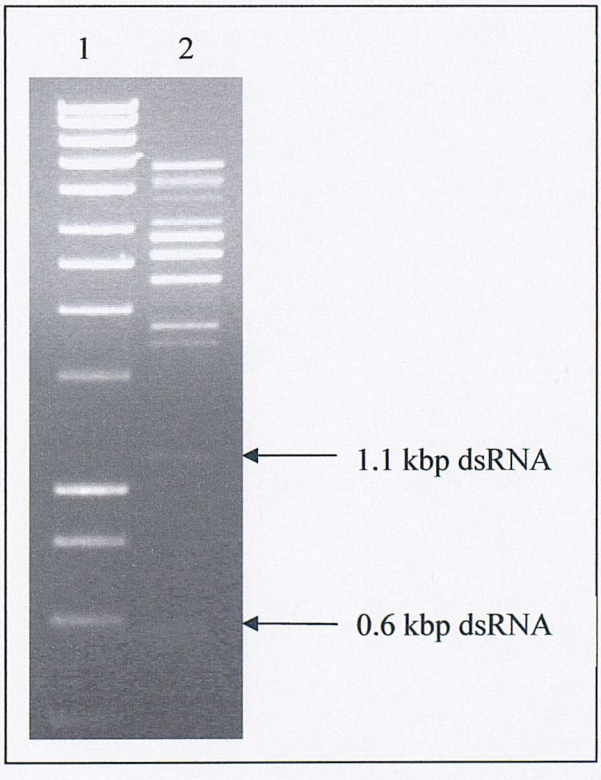


Figure 3.23: The two smallest ACD-related dsRNA elements. Analysis in a 1% agarose gel of the ACD associated dsRNAs, stained with ethidium bromide (lane 2) against the 1 kb genomic size marker (lane1). The 1.1 kbp and 0.6 kbp ACD dsRNA molecules are shown.

Table 3.6: Sequence of the primers used in the cDNA synthesis of the 0.6 kbp dsRNAs. The sequence of the primers is shown in the 5'- 3' orientation.

Primer name	Nucleotide sequence
Hexamer	ATGCGT
ZK31	CCCAGGAGGTCGTCGGGTCC

3.2.2.1 Sequence information

3.2.2.1.1 The 0.6 kbp dsRNA elements

The 0.6 kbp dsRNA band consisted of two dsRNA elements of very similar size, designated as A and B.

The 0.6 kbp dsRNA A element

The 0.6 kbp dsRNA A is 573 bp long and contains a single 78 amino acid ORF but it is not expressed as its sequence does not include a stop codon (GenBank accession number: AM085139). This element is probably a non-coding satellite of the ACD related dsRNA species. However, the absence of terminal region homology with the ACD related dsRNAs does not allow to specify which virus this element would be a satellite of. The sequence of the 0.6 kbp dsRNA A element did not show any significant similarities to known viral dsRNA sequences, nor to any of the sequence information in the NCBI sequence database (using the BLAST search tool), thus its origin remains unknown. The 3' terminus of this molecule contains three repeats, one of them inverted and another imperfect, inverted and homologous. These are shown in Figure 3.24, and indicate the formation of secondary structures.

```
1st repeat: 496:CGAAAGAAAAAAACAG: 511
                ||||||||||
2nd repeat: 548:CCAAAGAAAAAAAGGT: 533
                |||  |||||  |
3rd repeat: 557:GATTTCTTTTTATTTC: 572
```

Figure 3.24: Alignment of sequence segments from the 3' terminus of 0.6 kbp dsRNA A element. The positive strand of the dsRNA molecule is used on this alignment. The numbers correspond to the bp position on the dsRNA sequence.

The 0.6 kbp dsRNA B element

The 0.6 kbp dsRNA B is 606 bp long and it does not contain any coding sequence (GenBank accession number: AM085138). This element is a non-coding satellite dsRNA of the ACD chrysovirus species as there is significant homology between them at the 5'-termini (identical 7 terminal nucleotides) although there is no similarity between them at the 3'-termini. This element has the potential of producing unstable (because of the limited sequence homology) dimers, as the positive strand 5' terminus is complementary to the negative strand 5' terminus in the opposite orientation, as shown below in Figure 3.25. The sequence of the 0.6 kbp dsRNA B element (excluding the 5' most 7 bp) did not show any significant similarities to known viral dsRNA sequences, nor to any of the sequence information in the NCBI sequence database (using the BLAST search tool), thus as in the case of dsRNA element A, its origin remains unknown. It should be noted at this point that it is possible that both of the 0.6 kbp dsRNA elements are present in vivo in a ssRNA form, but we have only isolated and characterised the more prevalent (or stable) dsRNA form that represents a replicative intermediate.

```
            50        40        30        20        10
(-) strand GTGTGTTGGGTTGGGAGACACCAGAAAATAAAAATTATGATCACAATTTTCA
                                           :::::: :::::: :::::: :
(+) strand                                TGAAAATTGTGATCATAATTTTTATTTTCTGG
                                             10        20        30
```

Figure 3.25: The complementary 5' terminal positive and negative strands of the 0.6 kbp dsRNA element B.

Secondary structure of the 0.6 kbp dsRNA elements A and B

The potential secondary structures formed by both the 0.6 kbp dsRNA elements A and B were predicted using the mfold program (Zuker 2003). The 0.6 kbp dsRNA A element forms a stem loop structure on both its 3' and 5' ends, as shown in Figure 3.26. The molecule does not form any panhandle structures or any other major secondary structures along its length.

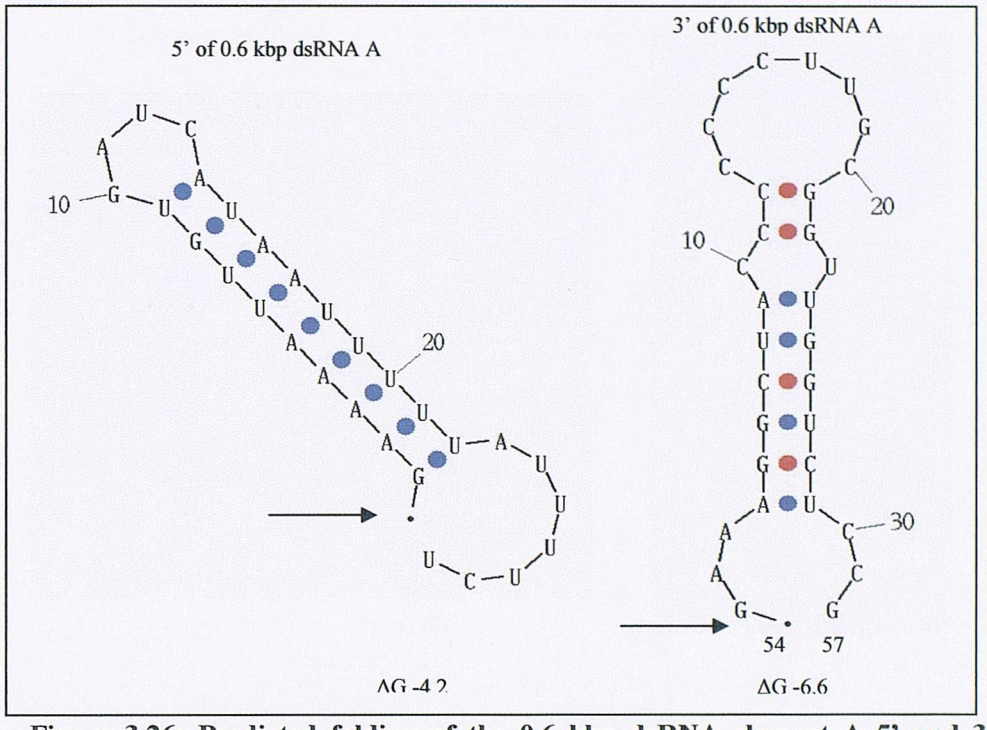


Figure 3.26: Predicted folding of the 0.6 kbp dsRNA element A 5' and 3' regions. The potential secondary structures of the 5' and 3' termini of the positive strand of 0.6 kbp dsRNA element A are shown. ΔG values (kcal/mol) were calculated using the mfold program. In both cases a potential stem-loop structure is present. The horizontal arrow points to the start of the sequence

The 0.6 kbp dsRNA B element forms a stem loop structure only at the 3' end. There is no major secondary structure predicted at the 5' end, though there is a potential panhandle secondary structure formed between the 3' and 5' ends of the molecule as shown in Figure 3.27. These results suggest a different secondary structure for the two satellite dsRNA elements. As the 3' and 5' secondary structures contain transcription recognition and initiation signals, it is probable that the two 0.6 kbp dsRNA molecules are recognised and transcribed by different RdRp molecules. The sequence similarity data presented above support this hypothesis.

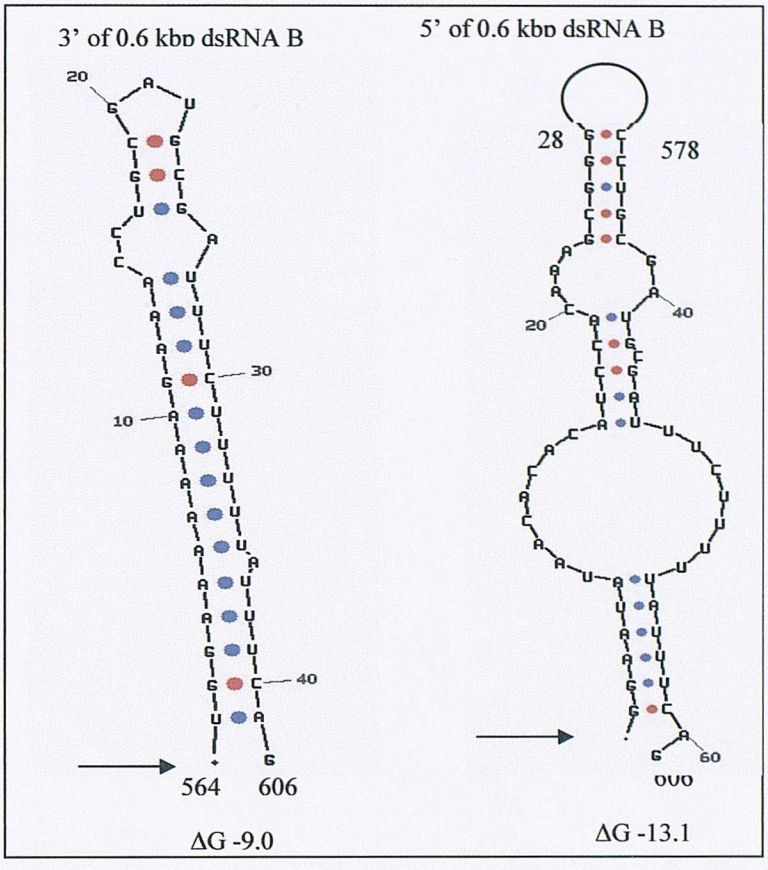


Figure 3.27: Predicted folding of the 0.6 kbp dsRNA element B 3' region as well as the aligned 5' and 3' regions. The potential stem loop structure at the 3' terminus of the positive strand of 0.6 kbp dsRNA element B is shown, as well as the potential panhandle structure formed by the aligned 5' and 3' termini. ΔG values (kcal/mol) were calculated using the mfold program. The horizontal arrow points to the start of the sequence.

Though the two different satellite elements might have slightly different structures at their termini, they appear with a secondary structure along the entire molecule. They adopt a potential 'hammerhead-like structure' (Figure 3.28), in which some of the conserved nucleotides that form the catalytic core of the ribozyme present changes. Helix II appears as the least stable in both strands of the dsRNA element B (Figure 3.29). The 'hammerhead-like structure' of the dsRNA element A appears less stable than the one formed by the dsRNA element B.

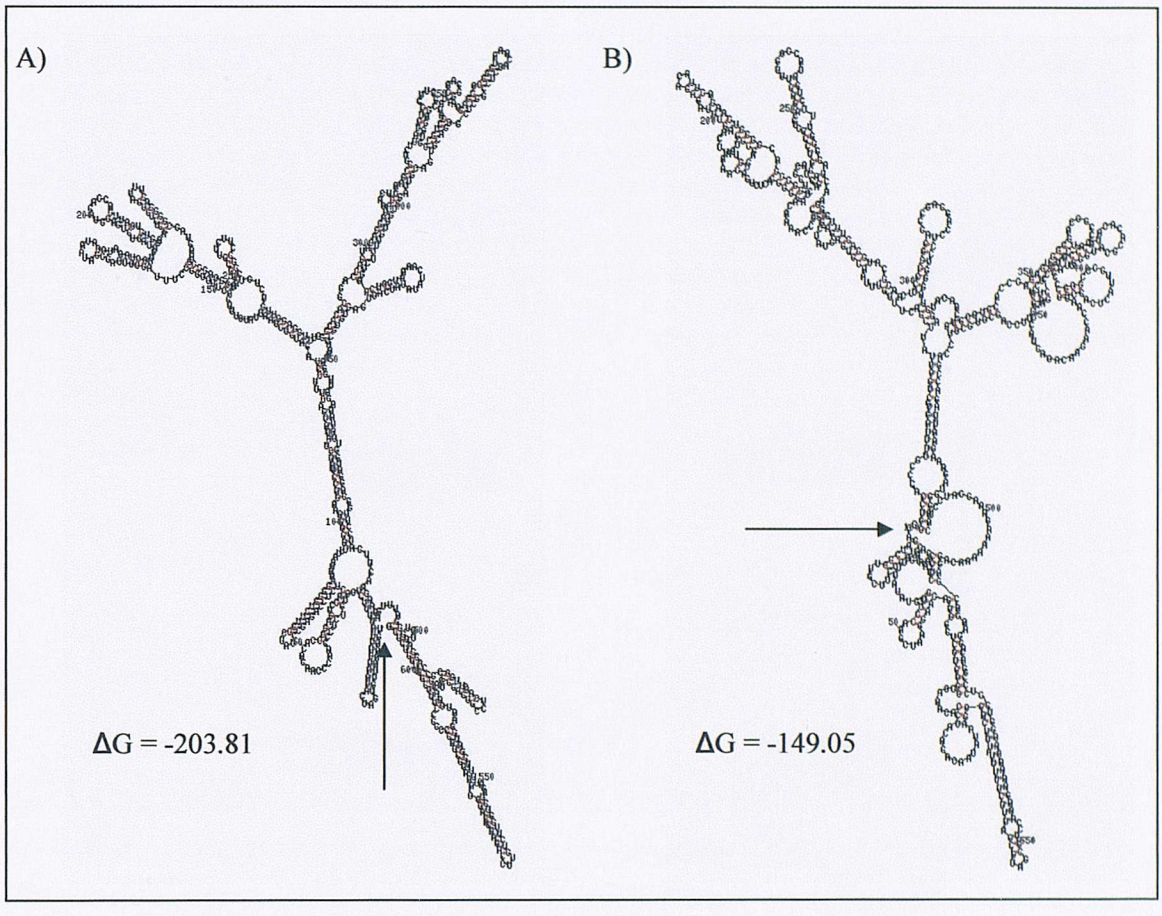


Figure 3.28: Predicted folding of the entire 0.6 kbp dsRNA elements A and B. The potential 'hammerhead-like' structures are shown for dsRNA element A [A)] and dsRNA element B [B)]. ΔG values (kcal/mol) were calculated using the mfold program. The arrows point to the start of the sequence.

UGAAAAUUGUGAUCAUAAUUUUUAUUUUCUGGUGUCUCCCAACCCAACACACGGGAGUCC
ACUCUCCAGAGCCUAUCGGUUCUGGGGAAAAAAUUUAGGGAAAGUCCUCUAUCUAUAUGA
CUUUCCUAGAUCUCUAAUUAUGUUUCACUUUUGGCCUUUACGGUUAGCCAUAAUAGGUAA
AUCGUUCUUACUCUGAGGUACACCACUUAGUGUGGAAGCUCUGUGAUUUCACGGGCGAUC
AAGCCGAGGGUAAGGCGUUCUUCCGCCUGUCUGUUGGAGGUUUCGGCACACGGACCAGUA
GACGAUAGACUCCAAGCGGUCUUAUCUCCUUCGCGAGGUGGCGUCCAUAUCAACGGAAGG
AGCGGGGGUAAACAGCUCCGGCUCGGGACGUGACGUUAGGUGUUGUUUUCUGGAACUGGU
AAAACUAUAUAUCAGUGAGUGUGCUUAAAAGGUUAGAUAUAUUGGUAGGACGGUUCCUAC
UUCAAGAAGAGUUUAGGGUGUGACCACCGGCGGUUCGACUCCGUCAGCCCUACAAGUAAG
GUAGCUUUUACAUCUGGUUCUGUUCCAGAACACAGGGAAAGGCUACCCCCCCUUGCGGUU
GGUCUCCG

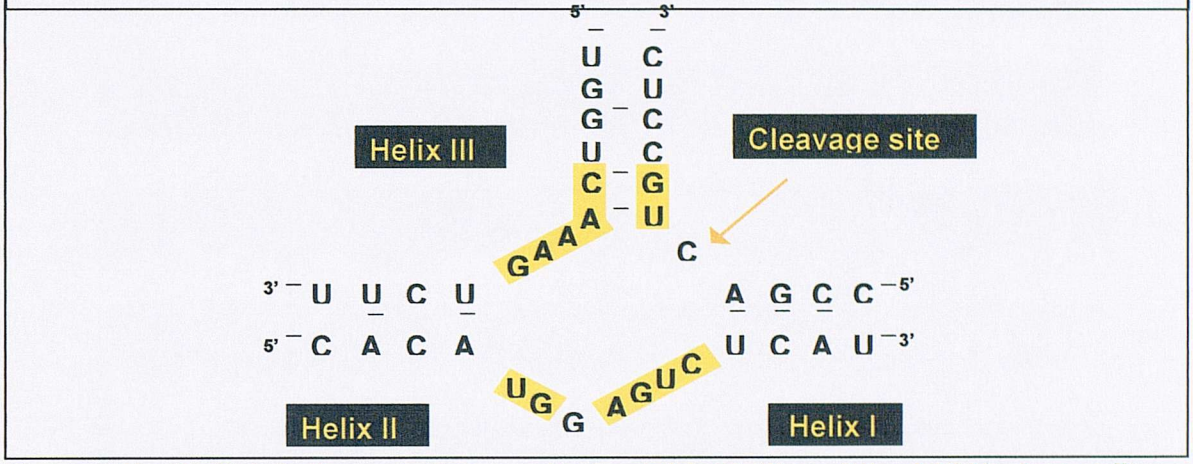
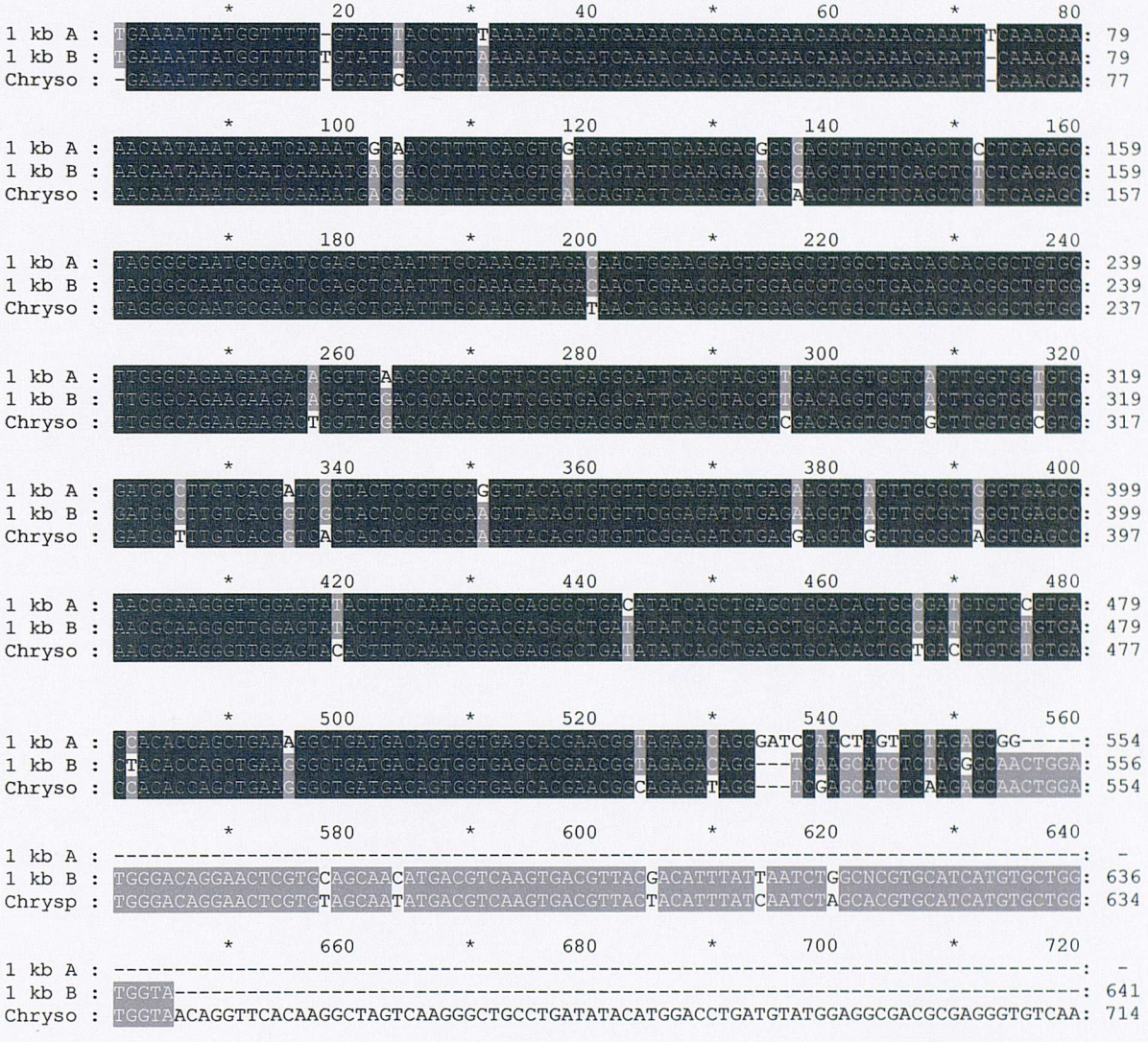


Figure 3.29: Predicted ribozyme cleavage site formed by folding of the dsRNA element B. The sequence of the dsRNA element is shown with the sides forming the ribozyme cleavage site being highlighted (adapted from Flores, personal communication to Coutts, 2005).

3.2.2.1.2 The 1.1 kbp dsRNA element

The 1.1 kbp dsRNA element proved very difficult to clone as when it was present, it was found in very small amounts. Gel extracted and purified dsRNA was used as a template for an RLM-RACE PCR amplification reaction, using the hexamer primer (Table 3.5) as a second internal primer. Two clones were generated successfully, shown in full in Figure 3.30. In both cases the sequence products generated were identical to the 5' portion of the ACD chrysovirus CP coding dsRNA element. The alignment of the cloned fragments is shown in Figure 3.30. Thus, while the 5' half of the 1.1 kbp dsRNA is sequenced, it became impossible to attempt the amplification and sequencing of the 3' half due to lack of dsRNA material. However, from the present data it is shown that the 1.1 kbp dsRNA is a defective dsRNA molecule originating from the chrysovirus CP coding dsRNA element.

Figure 3.30: CLUSTAL_X alignment, highlighted using GeneDoc, of the two 1 kbp dsRNA derived clones. The two clones, shown as 1 kbp (dsRNA) A and B, are compared against the ACD chryosivus CP coding dsRNA species, shown as Chryso.



3.3 Comparison of ACD-associated and CCRS-associated dsRNA elements

3.3.1 Cherry chlorotic rusty spot disease (CCRS), symptoms and spread of the disease

Cherry chlorotic rusty spot (CCRS) is a disease first reported in southern Italy in 1996. None of the known cherry viruses appear to be associated with it and initial attempts to

isolate and culture a bacterial or fungal pathogen were negative, thus the disease was presumed to be caused by an undescribed virus-like agent (Di Serio *et al.*, 1996). The symptoms observed in CCRS infected trees are very similar to those infected by ACD. Early leaf symptoms of CCRS disease are translucent spots of 1–2 cm diameter that develop into chlorotic spots with small rusty areas that finally cover their complete surface. Fruits from CCRS infected trees are generally smaller than average, and ripen later, while severely affected fruits exhibit deformations or shallow, irregular depressions- as shown in Figure 3.31 (Alioto *et al.* 2003).

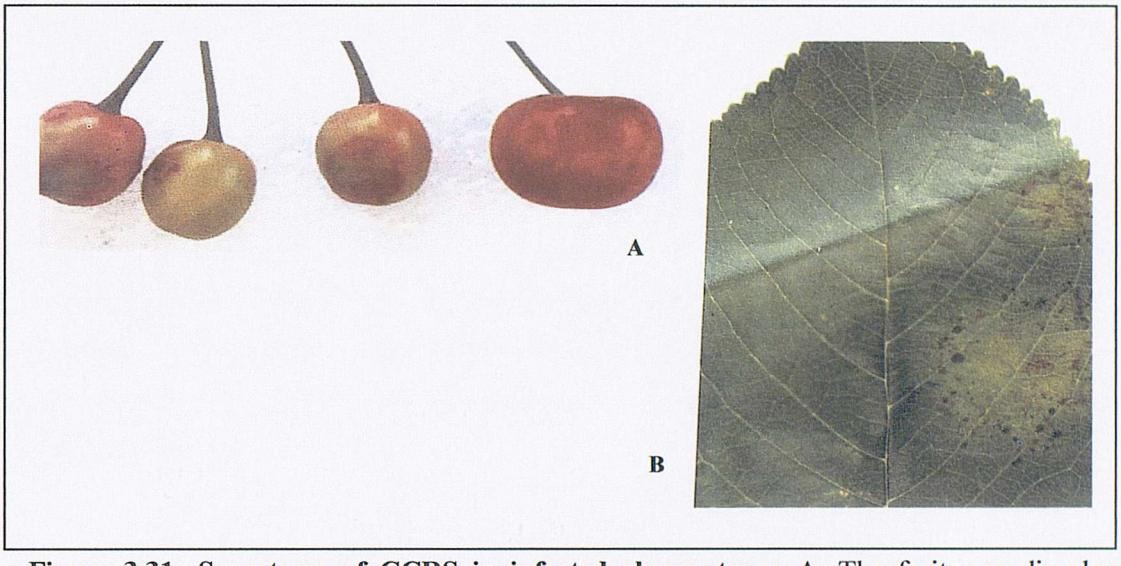


Figure 3.31: Symptoms of CCRS in infected cherry trees. A. The fruits are discoloured, elongated and smaller than average. B. Small red-brick coloured spots appear initially on infected leaf tissue. They eventually increase in size and cover the entire leaf area (Alioto *et al.* 2003).

PAGE analysis of dsRNA-rich preparations from cherry trees displaying the typical CCRS symptoms showed a complex profile of dsRNAs that were absent from healthy-looking trees. These dsRNAs ranged in size from approximately 1700 to 5500 bp (and two smaller dsRNAs of approximately 500 and 650 bp) and were identical in size and numbers to those observed for the ACD (Di Serio *et al.*, 1996). This similarity in the mobility of the dsRNA components suggested that the causative agent for the two diseases might be the same.

3.3.2 Relation between the ACD and CCRS dsRNA elements

To provide additional data supporting a close relationship between these two cherry disorders, a side-by-side examination of ACD- and CCRS-affected leaves of Turkish and Italian origin, respectively, was performed. The same six out of the ten dsRNA species were sequenced from ACD- and CCRS-infected leaves. Comparative analysis of the four chrysovirus ACD and CCRS dsRNAs showed a nucleotide identity greater than 93% and predicted amino acid sequences more than 95% identical. Similarly, the ACD and CCRS partitivirus predicted proteins showed a 94 - 98% identity. These results confirmed the very close relationship between the characterised ACD and CCRS viruses, which might even be considered as different isolates of the same virus (Coutts *et al.* 2004 and Covelli *et al.* 2004).

The same approach was used for the four largest and two smallest dsRNA elements found in both the ACD and CCRS infected tissues and is described in this thesis. The ACD sequences presented in this chapter (sections 3.2 and 3.3) were compared to the sequences obtained from some of the equivalent CCRS dsRNA elements, provided by Dr. Covelli and Prof. Flores. They are presented below according to the order of their molecular weight.

3.3.3 Comparison between the largest dsRNA molecules

The only available sequence of the CCRS equivalent from dsRNA 1a was produced by the diagnostic RT-PCR method (presented in detail in section 3.5.3). The two nucleotide sequences share 88.9 % identity, while the identity of the peptide sequences is 97.9 %.

The dsRNA2 species was sequenced to its entirety for both the ACD and CCRS diseases. The two nucleotide sequences have the exact same length and share a 98.16 % identity, while their peptide sequences are virtually identical, sharing a 99.4 % identity.

The dsRNA3 species was sequenced to its entirety only for the ACD, while two thirds of its length was sequenced for the CCRS. The two nucleotide sequences share a 96.73 % identity, while their peptide sequences are virtually identical sharing a 98.05 % identity.

There was no sequence information for a CCRS-associated equivalent of the ACD-associated dsRNA1b species since no such dsRNA has been described no comparisons can be made. However, based on the above data it is unlikely that the CCRS dsRNA1b sequence will be very different (if at all) to that of the ACD dsRNA1b species. The nucleotide and peptide sequence identity of the ACD and CCRS dsRNAs of higher molecular weight is higher than that of the partitivirus and chrysovirus species of the two diseases. This indicates a stronger evolutionary pressure for the conservation of the sequence than to the two already characterised viruses.

3.3.4 Comparison between the smallest dsRNA molecules

CCRS isolated RNA contains an abundance of a small circular RNA (termed cherry small circular RNA1 or csc RNA1), which was sequenced to completion and is 451 nt long. It has hammerhead ribozymes in both polarity strands, found in several viroid and viroid-like satellite RNA elements, but is probably a satellite of one of the dsRNAs. The same RNA was found in dsRNA-rich fractions of sour cherry trees tissue with similar symptoms to CCRS. The sour cherry trees were in the vicinity of CCRS infected sweet cherry trees (Di Serio *et al.* 1996). However, there was no evidence of this element present in either the ACD pooled RNA-rich fractions or the ACD-infected leaf tissue.

CCRS dsRNA preparations also contain three 0.6 kbp dsRNA molecules, termed A1, A2 and B. This terminology reflects sequence similarity, with A1 and A2 sharing 79.47% sequence identity while neither of which have any homology with element B.

ACD dsRNA samples contain only two of these three 0.6 kbp dsRNA molecules, namely A1 and B. The ACD 0.6 kbp dsRNA element A shares a 98.78 % sequence identity with the CCRS dsRNA A1. The ACD 0.6 kbp dsRNA element B shares a 96.865 % sequence identity with the CCRS dsRNA B. This data is consistent with the sequence identity ratios found when comparing the other ACD and CCRS dsRNA elements, further supporting the view of CCRS and ACD related dsRNAs represent different isolates of the same dsRNA elements/viruses. It should be noted that the CCRS 0.6 kbp small dsRNA elements displayed a much greater sequence variation than the ACD ones, where the

sequence disparity between clones obtained from different preparations was restricted to a very small number of single nucleotide changes. The sequence variation in the CCRS samples was such that allowed for the construction of a phylogenetic tree derived from multiple alignment of clones corresponding to the three groups of small dsRNA elements, termed A1, A2 and B (Figure 3.32).

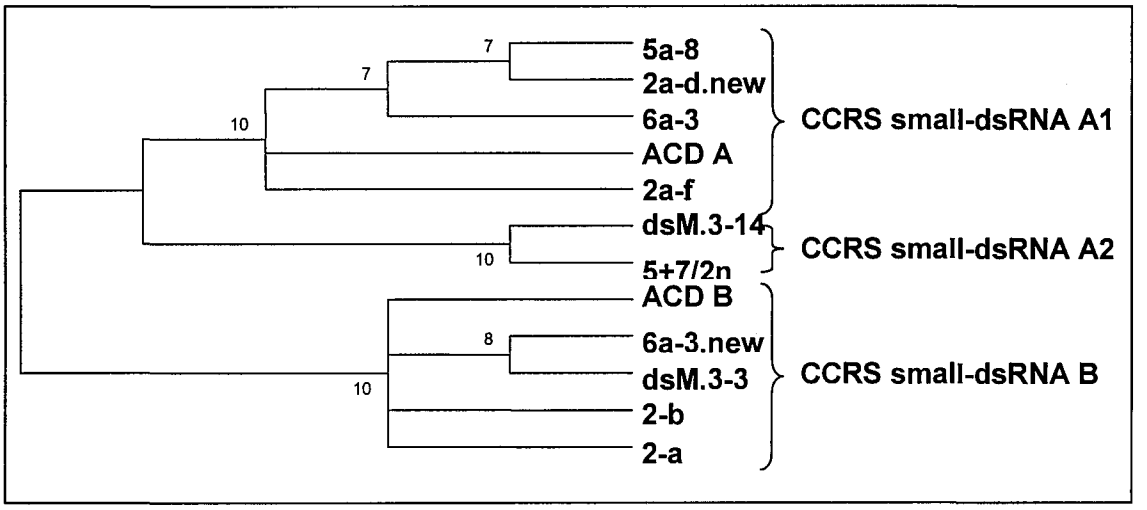


Figure 3.32: Phylogenetic tree constructed from the multiple sequence alignment of the sequenced ACD and CCRS 0.6 kbp dsRNA elements (alignment adapted from Flores, personal communication to Coutts, 2005). The ACD sequences are shown as ACD A and ACD B. The sequences are distinguished into three groups, with groups A1 and A2 being evolutionary closer to each other than to group B.

3.4 Development of an RT-PCR diagnostic test for ACD based on the dsRNAs associated with the disease.

3.4.1 Design of an RT-PCR-based approach

There is presently no economically viable treatment for ACD- or CCRS-affected trees in the field and elimination of spread is achieved by the eradication of infected stocks and replacement with trees sourced from certified, virus-free stock which are then at risk from infection with the disease from unidentified infected trees.

An understanding of the biology of the pathogenic agents involved in both diseases and their early identification will allow orchard management practices to be tailored towards minimising the risk of disease spread and the implementation of effective husbandry practices. As the first symptoms of ACD appear approximately two years after inoculation, a diagnostic method would be of great importance and economic benefit: the infected trees could be identified earlier, removed and replaced thus limiting the opportunity for the establishment of the diseases and any subsequent spread.

To this end a diagnostic RT-PCR-based approach has been developed which detects the presence of the ACD-related dsRNA elements. As these elements are always associated with ACD affected trees, their detection would also indicate the presence of the disease even in the absence of obvious symptoms. Three sets of primers were designed on unique areas of three of the dsRNA elements and were used to amplify fragments of known sizes *via* an RT-mediated step. One pair of primers was targeted to dsRNA1a- which is the most abundant dsRNA-associated species in ACD samples- and another to dsRNA1b– which is least abundant. A third primer pair targeted for amplification of the ACD chrysovirus dsRNA3, which is present in a consistently high concentration in infected tissue extracts, was also designed. The primer pairs were nominated ZK17 and ZK14 for dsRNA1a; ZK35 and ZK36 for dsRNA1b; ZK37 and ZK38 for Chrysovirus dsRNA3; producing 437 bp, 692 bp and 500 bp long fragments respectively (Table 3.7).

Table 3.7: Sequence of the primers used in the PCR-based detection of ACD-related dsRNA elements. The sequence of the primers designed for dsRNA1 and dsRNA4 are shown, in the 5'- 3' orientation.

Primer name	Primer sequence	PCR product size	dsRNA species
ZK14	CAAGAAGACGATGCGGTACG	437 bp	dsRNA1a
ZK17	CGCCGCGTACACCCTGTGC		
ZK35	CTTGTCATCATGGCTAGACG	692 bp	dsRNA1b
ZK36	GTCTGCAATTGTGCTCTGTGC		
ZK37	CCACCACGTGCAGAGCG	500 bp	Chrysovirus dsRNA3
ZK38	CAAGTTTAAGCGCAGTTGC		

Leaf tissue from ten different varieties of ACD-affected cherry trees and from two different varieties of healthy cherry trees was provided by Dr. Açikgöz from Turkey. The tissue was of differing quantity and quality, *i.e.* the symptoms of the disease were more severe in some samples than in others resulting in a large amount of necrotic leaf tissue in some instances. Healthy leaf tissue was sampled from two varieties: Lambert and Napolyon and ACD-affected leaf tissue samples were derived from the following varieties: Al Kinaz, El Napolyon, Er Tabani Yank, Kara Bodur, Kara Kirtik, Koroglu, Lambert, Napolyon, Tabani Yank and Yerli Kana. It has not been feasible to confirm whether the healthy leaf tissue samples originated from the same area of Turkey, *i.e.* the Amasya region, and whether the 'healthy' trees were visibly healthy, or from a certified virus-free stock. Although these details do not affect the quality of the data shown in this thesis, they do not allow for information, equally accurate to that provided with the virus-infected leaf tissue samples.

The RT-PCR amplification experiments were performed in such as a fashion so as to avoid any possible bias affecting the interpretation of the results. This was achieved by allocating numbers to the received samples the derivation of which were not known to the investigator. All PCR amplicons were cloned and sequenced and compared directly to the known sequence of ACD-associated dsRNAs.

3.4.2 RT-PCR diagnostic tests results

Primer combination ZK14-ZK17(for dsRNA element 1a).

The primer pair designed to detect dsRNA1a in ACD-affected tissue generated the PCR product of expected size with almost all of the diseased samples (Figure 3.33). The approach was used for different preparations of the same leaf tissue at different times and consistently gave the same results. It was thus considered a robust and reliable detection method for the presence of dsRNA1a in dsRNA-rich preparations from ACD-affected leaf tissue.

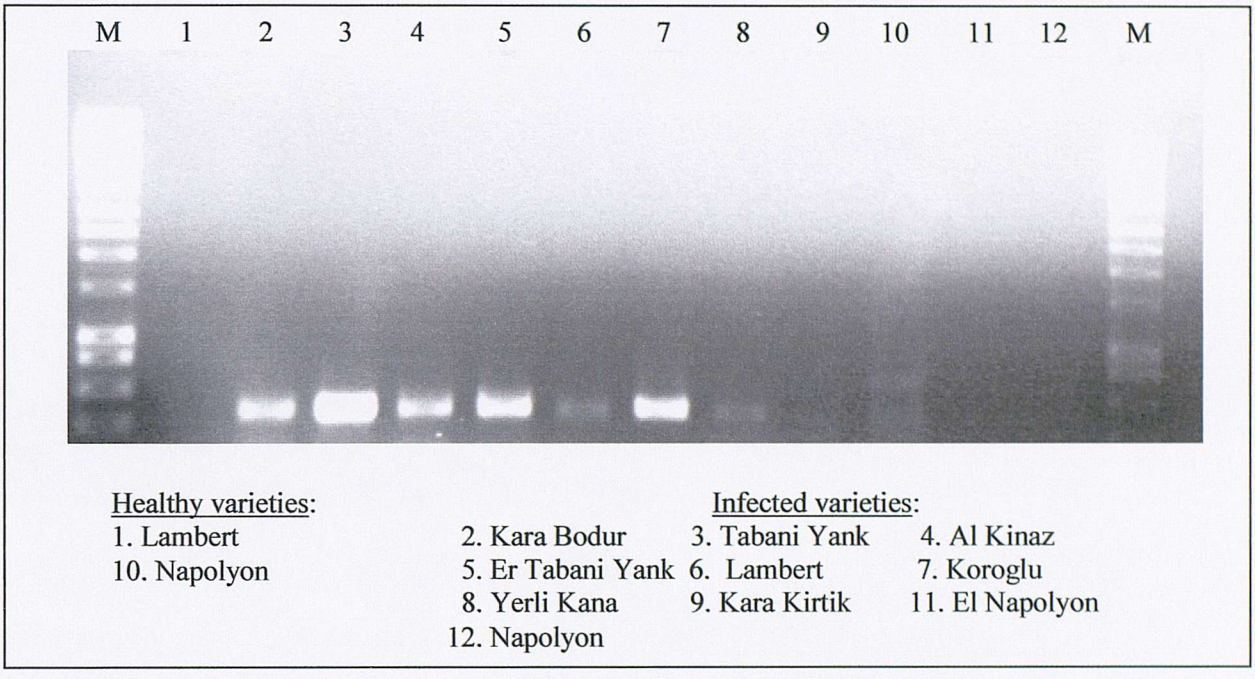


Figure 3.33: RT-PCR product of the ZK14-ZK17 set of primers using ACD-infected tissue. Lanes M contain the 1 kb Hyperladder I molecular size marker. For lanes 1 and 10 healthy leaf tissue was used. For lanes 2-9, 11 and 12 infected leaf tissue was used. Lane 1 did not give any products; lane 10 gave a product of different molecular size. Lanes 2-8 have a clear presence of the PCR product of predicted size. Lane 9 contains the same product as well, but in very small amounts. Lanes 11 and 12 did not give a PCR product.

PCR amplicons were not generated with leaf extracts of two ACD-affected trees of the varieties Napolyon and El Napolyon. However the leaves of both varieties were in small amounts and highly necrotic and it is thought that the absence of a PCR product was due to the quality of the sample rather than the detection ability of the system. In a small number of cases in the assay, one of the two healthy control samples generated a PCR amplicon of a different size to that expected lane (Figure 3.33: lane 10). Sequencing of the amplicon revealed it to be a fragment of cloned *Prunus* rRNA. Increasing the annealing temperature to 55°C from 50°C during the PCR reaction prevented the recurrent appearance of this product (Figure 3.34). The results of this approach were reproducible even when very small amounts of dsRNA were used as a template, < 20 ng of CF-11 extracted total dsRNA from leaf tissue, shown on lane 3 in Figure 3.34.

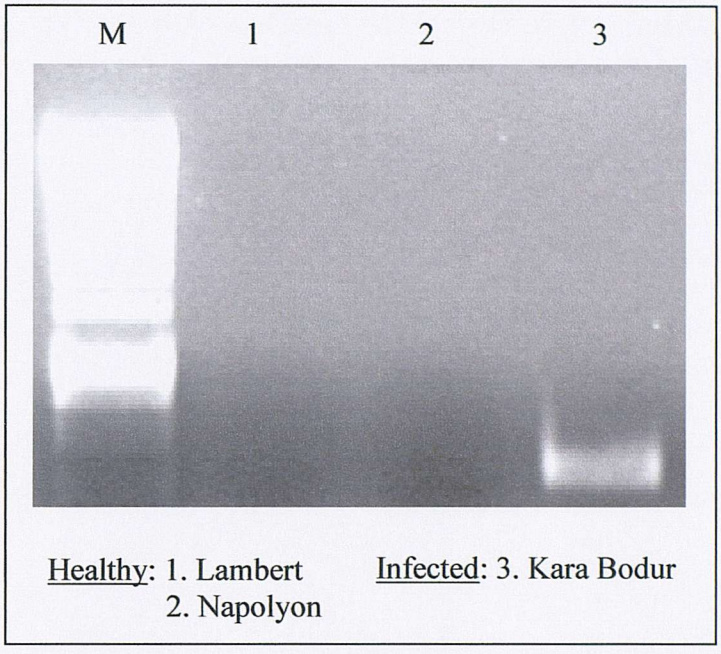


Figure 3.34: RT-PCR product of the ZK14-ZK17 set of primers, using 55°C as the annealing temperature and low amounts of dsRNA template. In lane M is the 1 kb Hyperladder I molecular size marker. Lanes 1 and 2 do not contain any PCR product. Lane 3 contains the PCR product of predicted molecular size. Small amounts (< 20 ng) of dsRNA were used for the RT-PCR reaction in lane 3.

Primer combination ZK35-ZK36 (for dsRNA element 1b)

The primer combination of ZK35-ZK36 had a limited success in producing fragments of the expected size, due to the low concentration of the dsRNA target band in the tissue. The experiment was repeated with concentrated, gel purified dsRNA resulting in a PCR product of expected size. Thus, a limiting factor of this approach is the relative abundance of the targeted dsRNA present in the infected leaf tissue extracts.

Primer combination ZK37-ZK38 (for chrysovirus dsRNA element 3)

The primer combination of ZK37-ZK38 gave initially a large number of PCR products. In order to minimise the number of PCR products, a more stringent PCR reaction was used. As in the case of primer pair ZK14-ZK17, the annealing temperature of the PCR reaction was increased to 55°C from 50°C. This resulted in a reduced number of PCR products (Figure 3.35), but not in a single product. The products of expected size, *i.e.* 500

bp, were isolated, cloned and sequenced and were correct. Thus this pair of primers can target the chosen dsRNA band, though the primer specificity might not be the highest possible.

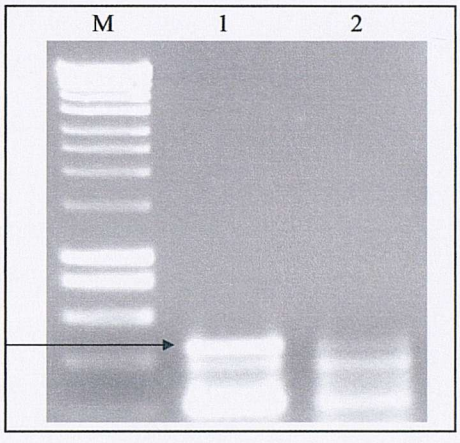


Figure 3.35: RT-PCR product of the ZK37-ZK38 set of primers, using 55°C as the annealing temperature and low amounts of dsRNA template. In lane M is the 1 kb Hyperladder I molecular size marker. Lanes 1 and 2 contain a ladder PCR products using an ACD (Lambert) and a CCRS dsRNA preparation. The product of expected molecular size is highlighted by the horizontal arrow. Small amounts (< 20 ng) of dsRNA were used for both RT-PCR reactions.

Alternative primer combinations

This RT-PCR method can potentially target any dsRNA species common to both ACD and CCRS diseases. The only limiting factors, as has been shown from the above results are the quality of the infected tissue, and indirectly of the dsRNA, as well as the concentration of the dsRNA species targeted. Pairs of unique primers have been designed for each one of the ACD-related dsRNA species. The sequence of these primers is shown in Table 3.8. The dsRNA species more likely to result in PCR products are those targeted to the partitivirus dsRNAs, and the dsRNA corresponding to ORF5 (the CP) of the chrysovirus (Figure 3.1), as they are to be found consistently in higher concentrations than the other dsRNA species.

Table 3.8: Sequence of the primers designed for the PCR-based detection of ACD-related dsRNA elements. The sequence of the primer pairs designed for the ACD partitivirus, ACD

chrysovirus and dsRNA2 and dsRNA3 elements are shown, in the 5'- 3' orientation. Odd numbers correspond to the 5' located primers, while even numbers correspond to the 3' located primers. The dsRNA molecules of the ACD-related partiti- and chrysovirus are numerically designated in order of ascending molecular size.

Primer	Sequence	PCR product size	dsRNA species
1	GGATCATCGCAGACACCATGC	501 bp	ACD partitivirus dsRNA1
2	GGTCGAAGCCCATCGCGG		
3	GACAGCCAATTCAAGGCCG	586 bp	ACD partitivirus dsRNA2
4	GTAAGGCTGCATGACGCGG		
5	GCAAGGCAGCGTGACAGCAGG	440 bp	ACD chrysovirus dsRNA1
6	CAACCACTCAGTCATCGC		
7	GACGATGTTGTGGCTTGGCG	637 bp	ACD chrysovirus dsRNA2
8	CTCAATTAGTGTCCTCCCG		
9	CTTATGGACATGTACGACC	619 bp	ACD chrysovirus dsRNA4
10	GGATCGTACGCCATGTGTTGC		
11	GGAAAGCGGGCACGTACG	475 bp	dsRNA2 element
12	CCTGCTGTTGTAGTCACCG		
13	GGTTGCATCACAAGTGAGC	625 bp	dsRNA3 element
14	CCTTCACTACCTCATTACC		

3.4.3 Results of diagnostic RT-PCR tests with dsRNA-extracts from CCRS-diseased leaf tissue

As the CCRS-related dsRNA elements are very closely related to the ACD-related ones in sequence (section 3.4), the PCR-based diagnostic test was attempted with dsRNA extracts from CCRS-diseased leaf tissue provided by Prof. Flores. Samples from two healthy cherry trees and six CCRS-diseased cherry trees, all of unknown varieties, were tested. The ZK14-ZK17 pair of primers produced the PCR product of expected size, which was sequenced and proved to be the first sequenced fragment from the CCRS equivalent of the ACD dsRNA1a species (Figure 3.36). As with the ACD-infected tissue there was a small amount of mispriming that could be attributed to the presence of contaminant mRNA within the total RNA extracted from the infected tissue, or even to

the low quality of the template (Figure 3.36 A). This mispriming was greatly improved when the annealing temperature was increased from 50°C to 53°C (Figure 3.36 B). The ZK37-ZK38 pair of primers produced the PCR product of expected size, though the specificity of the reaction would need to be improved further. It was not possible to investigate this further due to the limited amount of extracted plant material available. Thus, the RT-PCR method can be used to detect the presence of both the ACD- and CCRS-related dsRNA species and the respective diseases, with the only limiting factors for detection being the quality and concentration of the targeted dsRNA species itself.

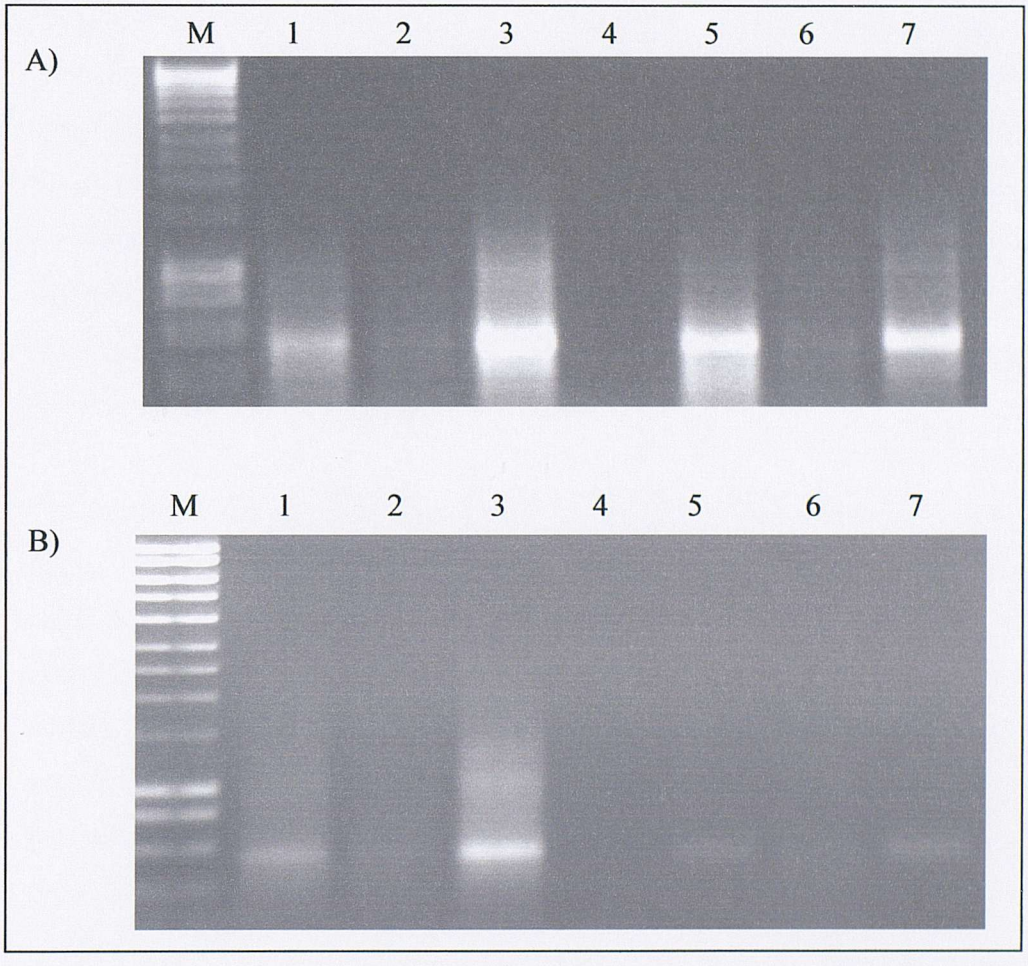


Figure 3.36: RT-PCR product of the ZK14-ZK17 set of primers using CCRS-infected tissue. Lanes M contain the 1 kb Hyperladder I molecular size marker. Lanes A4 and B4 contain the sample from healthy tree tissue. All other lanes contain samples were from CCRS-infected leaf tissue. An annealing temperature of 50°C was used for the samples in picture A). An annealing

temperature of 53°C was used for the samples in picture B). All other conditions of the diagnostic RT-PCR protocol were identical.

3.5 Chapter summary

The five largest ACD-associated dsRNA species have been cloned and sequenced, together with the two smallest dsRNA elements. The first working hypothesis, that the largest dsRNAs represent novel viral sequences has been confirmed. These novel dsRNA elements share some similarities in their RdRp regions to members of the *Totiviridae*. It is thus proposed that these five dsRNA elements represent two viruses of a novel mycoviral family related to the *Totiviridae* and a different isolate of a previously characterised mitovirus. However, it has not been possible to identify which two pairs of these dsRNA elements constitute the two new viruses. The phylogenetic and sequence similarity data presented in this chapter support this hypothesis. A possible method for the further study of these dsRNA species would be to use gel extracted and subsequently denatured dsRNA material in an *in vitro* translation system, such as the wheat germ system. This would allow for the identification of the proteins produced from the dsRNA species, especially those with novel sequence information.

The small dsRNA elements sequenced consist of two new dsRNA satellite species and one defective dsRNA element. The origin of the satellite dsRNA elements remains unknown, while the defective dsRNA is most probably derived from the ACD chrysovirus CP coding dsRNA. The small dsRNAs are most probably dsRNA satellite species. Satellite dsRNAs have been previously reported for partitiviruses and totiviruses, not for chrysoviruses. However, it seems that dsRNA element B is a satellite of the ACD-associated chrysovirus. It is not possible though to predict which chrysoviral dsRNA element(s) dsRNA A is a satellite of.

The previously noted high sequence identity between the ACD-associated and CCRS-associated dsRNA species was further supported by the comparisons of the newly cloned and sequenced ACD-associated dsRNA species to their equivalent CCRS-associated species. The high similarity is present in all of the four largest dsRNA species, as well as

in the smallest dsRNAs. This data supports the hypothesis that the two sets of dsRNA molecules constitute variant strains of the same viruses/elements (Coutts *et al.* 2004). A similar case of two differently named crop diseases attributed to different isolates of the same sets of viral elements has been observed before in Europe, *e.g.* the *Soil-borne rye mosaic virus* and the *European wheat mosaic virus* are two names characterising different isolates of a single furovirus (Koenig and Huth 2000).

Despite the extensive similarities there were also three differences found, all relating to the smallest dsRNA molecules. The chrysovirus defective dsRNA was not observed in the CCRS-derived dsRNA preparations. Likewise, the CCRS-associated 451 nt csc RNA was not observed in the ACD dsRNA preparations. Furthermore, the CCRS dsRNA fractions contained one additional 0.6 kbp dsRNA element that was not found in ACD dsRNA preparations. Though these differences are interesting, they are unlikely to contribute to the observed pathology of the infected trees.

The sequence data produced to date were used to design an RT-PCR diagnostic test for ACD based on the presence of mycoviral dsRNAs in diseased leaf-tissue. Whilst both ACD and CCRS most probably have a fungal aetiology (a hypothesis based on circumstantial observations) the only common, observed and detectable characteristic of the diseases is the pattern of dsRNAs extracted from infected tissue. Thus, the detection of the presence of these dsRNA molecules can directly diagnose the presence of the diseases.

3.5.1 Aetiology of the disease

The second working hypothesis was based on the previously published sequencing information regarding the ACD- and CCRS- associated chrysovirus and partitivirus species (Covelli *et al.* 2004 and Coutts *et al.* 2004 respectively). The hypothesis here is that the novel dsRNA elements described here would be mycoviral or mycovirus-related and not linked directly to the aetiology of the disease. The sequencing information of the dsRNA elements and the computer-assisted analyses of their sequence tend to support

this hypothesis. However, there are certain factors that need to be taken into account when considering the aetiology of the disease.

ACD was first reported to be of a viral origin, which was dismissed in favour of another transmissible agent, *i.e.* a fungus or bacterium. However, it has not been possible, since the first reports for the apearance of ACD in the 1950s, to grow the fungal-host in culture, with the saprophytic *Alternaria alternata* being sporadically detected (Alioto *et al.* 2003). The sequencing of the five largest dsRNA species followed an initial shotgun approach, where a large collection of cDNA clones was constructed. Although there were cherry tree sequences obtained, none of the clones generated belonged to a known fungal species (data not shown). The predicted protein products of the clones generated were not related to fungal proteins, *e.g.* chitin, or proteins that could denote a fungal presence. Furthermore, the evidence suggesting a fungal origin for the ACD- and CCRS-associated dsRNA elements and, indirectly, a fungal aetiology for the two diseases, is only circumstantial.

Firstly, electron microscopic examinations of thin sections have revealed a fungus-like organism in CCRS and ACD symptomatic leaves, and fluorescence microscopy observations have confirmed the presence in affected leaf areas of a 'mycelium' growing between the lower epidermis, palisade and spongy mesophyll cells that was not found in vascular bundles or in asymptomatic samples (Alioto *et al.* 2003). However, the relatively low amount of this 'mycelium' with respect to the entire leaf tissue cannot fully justify the consistency of the ACD detection reactions, unless the dsRNA elements are present in large quantities within their fungal host.

Secondly, treatment with systemic fungicides, though it diminished the adverse effects of ACD, it did not permit these effects to be fully reversed over a length of time (A. Citir, unpublished data). It should be noted though that the persistence of the disease despite the use of systemic fungicides could be attributed to other factors, such as recurrent infections of the diseased tree from the same fungus or a developed fungal resistance to the particular fungicide(s) used.

Collectively, the dsRNA sequencing data and the circumstantial evidence mentioned support the involvement of a fungus, probably as the causal agent, however this has not been firmly proven to date. Thus, the second working hypothesis is probably correct, but this cannot be supported with absolute certainty. A series of PCR reactions from ACD- or CCRS-infected leaf tissue for the detection of conserved motifs of typically fungal proteins might help in providing another line of evidence supporting the involvement of a fungal host.

Conclusively, whether the ACD- and CCRS-associated dsRNAs are avirulent for their fungal hosts, as with PcV (Ghabrial and Castón 2004), or whether they modulate fungal pathogenicity as previously reported for other mycoviruses (Ghabrial 1994, Nuss 1996, McCabe *et al.* 1999), thus contributing to the aetiology of the disease, remains an intriguing issue to be explored.

CHAPTER FOUR

Characterisation of the two dsRNA elements from *Phlebiopsis gigantea*

4.1 Introduction

4.1.1 *Phlebiopsis gigantea* and its industrial use

Phlebiopsis gigantea (previously called *Peniophora gigantean*) is a basidiomycete white rot fungus that can colonise freshly cut softwood. The fungus is a saprophyte, able to colonise the surface through invading hyphae which enter the tracheids and ray parenchyma cells. *P. gigantea* is used industrially to begin the process of paper pulping and as a biological control of the root and butt rot fungus *Heterobasidion annosum* (Behrendt *et al.* 2000). *H. annosum* is a common pathogen in temperate forests, causing large economic losses to many pine and spruce species in the north hemisphere (Stenlid and Waesterlund 1986). Once established, *H. annosum* is almost impossible to eradicate as the infected trees have fruiting bodies on them that release airborne basidiospores that spread the infection to new sites.

The problem is significant in managed coniferous forests, leading to the development of chemical and biological control methods (Korhonen and Stenlid 1998). Chemical treatments of felled softwood include the application of urea, boron compounds and ammonium sulphamate. However, this can be environmentally undesirable, thus biological control methods have also been actively developed. *P. gigantea* spores are applied to the face of freshly cut trunks and their successful fungal colonisation inhibits the colonisation of the exposed wood surface by *H. annosum* (Figure 4.1).

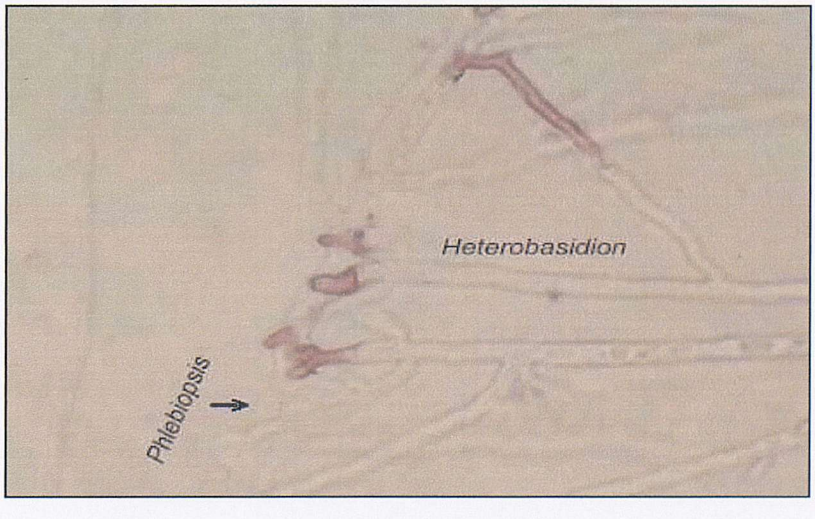


Figure 4.1: Hyphal interference of hyphae of *H. annosum* by hyphae of *P. gigantea* (Reproduced from Deacon 1998). Hyphae of *H. annosum* were antagonised when in contact with *P. gigantea* hyphae. The agar plate was subsequently flooded with a dilute solution of neutral red, which was taken up exclusively by the damaged hyphae of *H. annosum*.

P. gigantea is an attractive biological control method because it is not a biocidal agent, *i.e.* unlike many other commercially available agents it does not kill the target organism. It competes for resources that the pathogen would otherwise use, thus providing an extension to a naturally occurring process (Pratt *et al.* 1999). *P. gigantea* was registered in the UK as a bio-control agent in 1998 (as a suspension: PG Suspension; or in a dry form: Rotstop^{TM}) for use on pine trees, such as Scots pine (*Pinus sylvestris*). The fungus is not permitted to be used on Sitka spruce (*Picea sitchensis*), which is also an important conifer species in the UK. Rotstop^{TM} is currently commercially available in Britain, Finland, Norway, Sweden, Switzerland and the USA.

Though the *P. gigantea* treatments are effective in inhibiting new *H. annosum* infections, they have little effect in previously infected areas, where the established fungus spreads *via* root contact (Roy *et al.* 2003). Therefore, further work on *P. gigantea* is important to fully understand the fungus and explore any future prospects for more effective control of *H. annosum*. The *P. gigantea* TW2 isolate contains two large dsRNA elements that have not been shown to have a phenotypic effect on the host (Prof. Buck, personal communication 2005). Thus, their molecular characterisation can be used in the future development of means to genetically control and improve the efficacy of their host as a biological control agent.

4.1.2 The dsRNA elements

The *P. gigantea* TW2 isolate contains two separate dsRNA species approximately 10 – 13 kbp in length (Figure 4.2). Contigs of approximately 6.5 and 5.9 kbp had been assembled from these RNA species in previous work by Dr. Hacker and Prof. Buck. Based on their molecular weight and some limited sequence similarities, it was originally thought that these dsRNA elements may be closely related to the *Endornaviridae*; this

was the working hypothesis for this project. There is only one Endornavirus infecting a fungus-like organism (*Phytophthora spp.*) described in literature (Hacker *et al.* 2005). These viruses have not been extensively studied and thus are not well understood. The aim of the present project was to confirm the sequence obtained (where ambiguities were present), to continue and complete where possible the molecular characterisation of these two dsRNA elements.

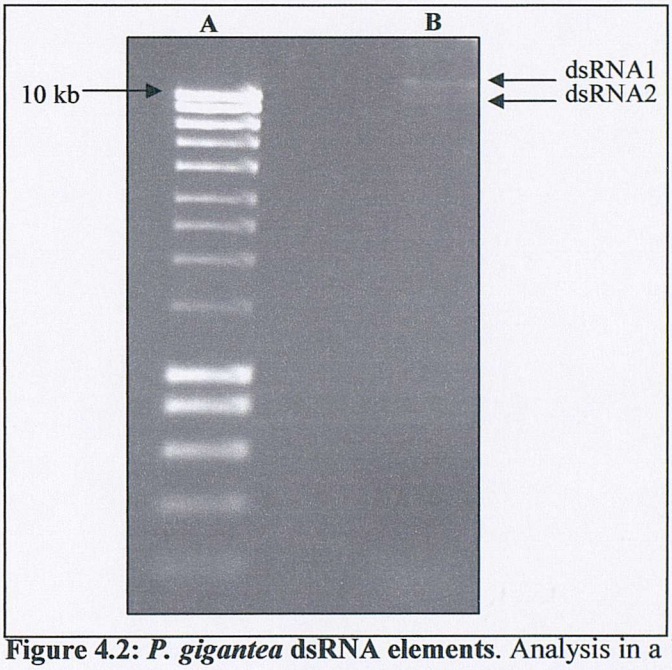


Figure 4.2: *P. gigantea* dsRNA elements. Analysis in a 1% agarose gel of the *P.gigantea* (TW2) dsRNA species, stained with ethidium bromide. Lane A contains the 1 kb Hyperladder I molecular size ladder, the highest molecular weight band of the marker is shown with a horizontal arrow.

4.2 Results and Discussion

4.2.1 Viral capsid detection

The *P. gigantea* dsRNA elements are similar in terms of size and some coding information to the *Endornaviridae*. The latter are unencapsidated, cytoplasmic dsRNA viruses. Thus, viral particle isolations were attempted in order to investigate whether the *P. gigantea* dsRNA elements are associated with any virus-like particles. Two different methods were used in parallel: direct pelleting and PEG_{6000} viral particle preparation. The first method provides crude but highly concentrated material, while the second one

results in a much cleaner but also less concentrated material. In both cases a 10 ml, 20-50% sucrose gradient separation final step was included. 0.5 ml fractions were collected from the sucrose gradient column; the dsRNA material was extracted from 50 µl of each one of these aliquots, concentrated and analysed on an agarose gel. A second 50 µl aliquot of each fraction was run on an SDS-PAGE gel. The presence of virus-like particles should be indicated by increased protein presence in the fractions that contain the dsRNA elements. A 100 µl aliquot of the dsRNA containing fractions was also sent to be photographed by electron microscopy (EM) by Dr. Booy Figures 4.3 A and 4.3 B show the dsRNA presence in specific sucrose gradient fractions. The Coomassie brilliant blue staining and silver staining of the SDS-PAGE gels electrophoresed with the dsRNA containing fractions are shown in Figures 4.4 A and 4.4 B respectively. Figures 4.5 A and 4.5 B are the EM obtained photographs of the dsRNA containing fractions.

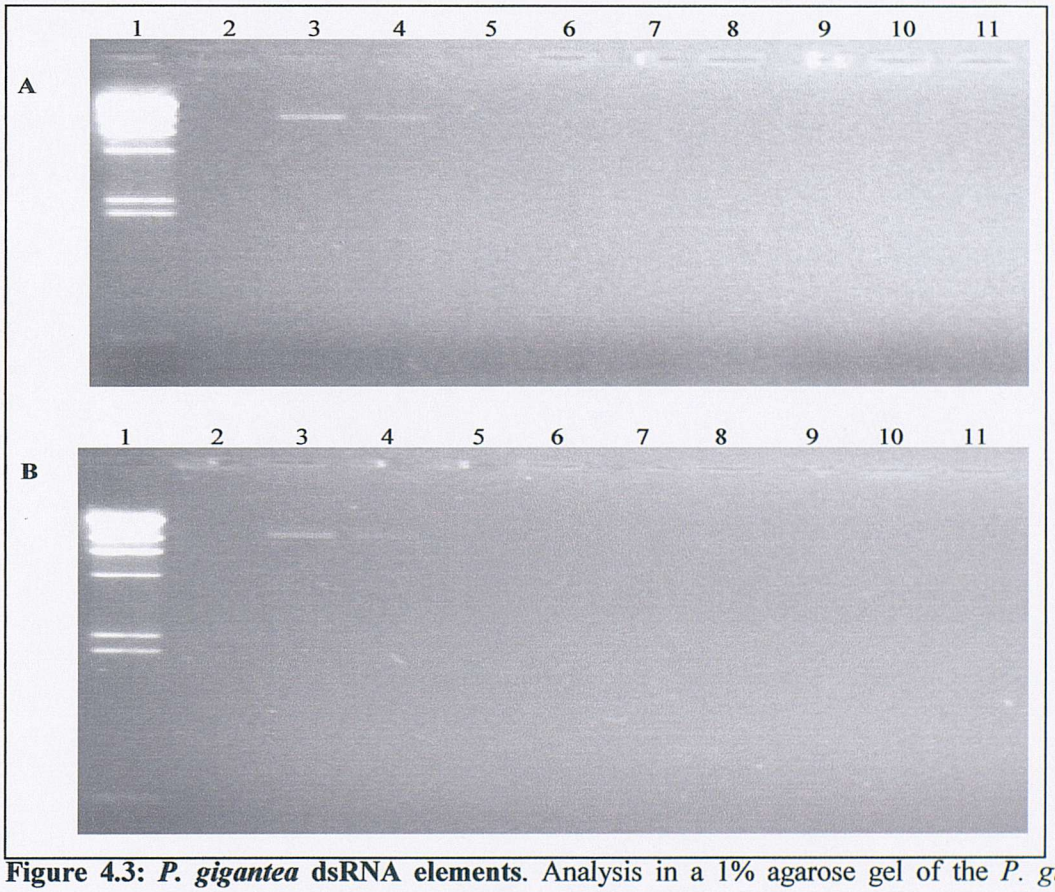


Figure 4.3: *P. gigantea* dsRNA elements. Analysis in a 1% agarose gel of the *P. gigantea* dsRNA species, stained with ethidium bromide. Lanes A1 and B1 contain the λ/*Hind*III

molecular size marker. Lanes A2 – 11 and B2 – 11 contain the sucrose gradient fractions after direct pelleting and PEG_{6000} viral particle isolation respectively.

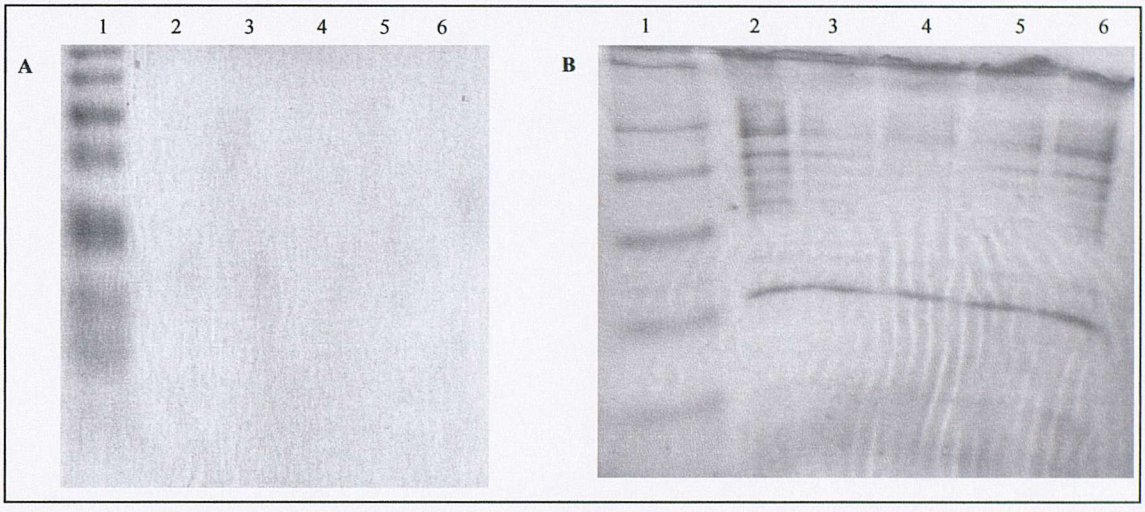


Figure 4.4: SDS-PAGE gels of fractions A2 – 6 (as shown in Figure 4.3). **A** Coomassie brilliant blue staining of the SDS-PAGE gel produced very diffuse and faint stained bands. **B** Silver staining of the SDS-PAGE gel produced better defined bands, but none associated specifically in high amounts with the dsRNA containing fractions.

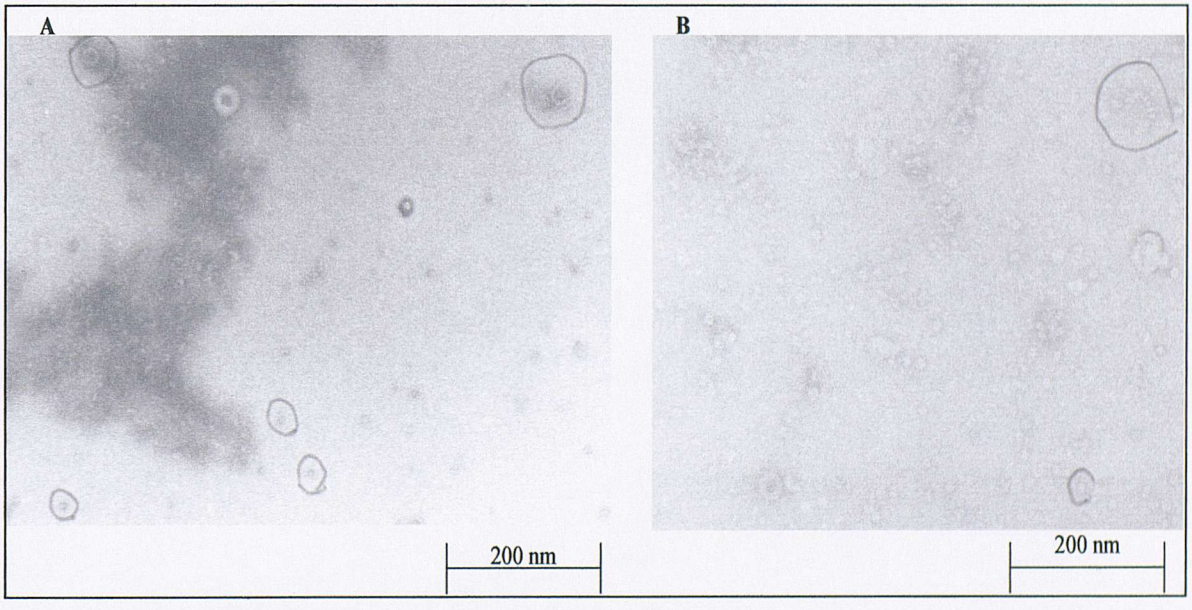


Figure 4.5: EM photographs of the dsRNA containing fractions (as shown in Figure 4.3). **A** EM photograph of fraction A3. **B** EM photograph of fraction B3. In both pictures interesting

features, such as possible symmetrical structures are circled. The PEG_{6000} preparation results are much clearer. There was no single specific structure (symmetrical or filamentous) found in abundance, thus the features circled are likely to be derived from proteins or lipids.

From the above results it can be seen that there is no protein specifically associated with the dsRNA containing fractions. The EM photogpraphs did not reveal any filamentous structures or any clear isometric/symmetrical structures. Thus, from this set of results it can be proposed that the *P. gigantea* dsRNA elements are not associated with any detectable virus-like particles. They are cytoplasmic dsRNA elements, most probably maintained in a similar manner to the *Endornaviridae*.

4.2.2 Sequence information

The two *P. gigantea* dsRNA species have been tentatively named in order of descending molecular size as dsRNAs 1 and 2 (Figure 4.2). Computer assisted translation of the initially obtained cDNA sequence information, indicated the presence of multiple ORFs. The C-terminal regions of the last ORFs showed significant similarity to the conserved mycoviral RdRp motifs (Bruenn 2003). The C-terminal regions were the only genomic regions where a sequence similarity could be detected between the two dsRNA elements.

The positions of the stop codons relative to the RdRp motifs are shown in Figure 4.6 for each dsRNA element. As neither internal stop codons nor multiple ORFs are usually present in mycoviruses, the areas containing the stop codons were sequenced multiple times. This was achieved by generating clones internal to the available contig sequence using different dsRNA preparations. The multiple sequence fragments were then aligned resulting in the final consensus contiguous cDNA sequence. The same approach was used in resolving the ambiguities present in the assembled genomic sequences.

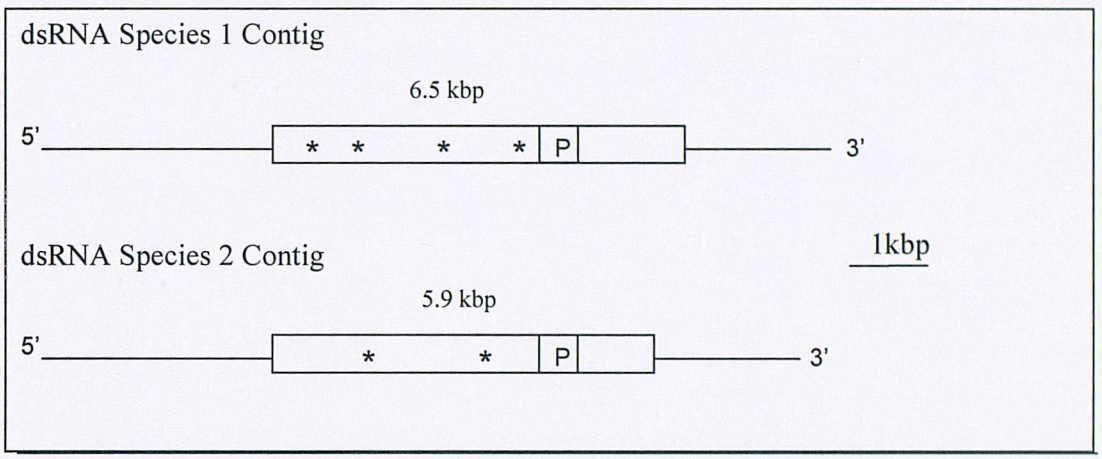


Figure 4.6: The two dsRNA species in the TW2 *P. gigantea* isolate (as provided by Dr. Hacker). The contig sequences are represented by rectangular boxes and are tentatively placed on the genome. The polymerase-like domains (P) are aligned between with the two species. * represents a stop codon found in the translated cDNA sequence.

4.2.2.1 dsRNA species 1

The available dsRNA species 1 sequence is currently 8175 bp long and presented in full in the GenBank database, accession number: AM111096. Computer assisted analysis of the available sequence has revealed the presence of three large ORFs, at least 4688 bp, 2796 bp and at least 627 bp long. Two of these ORFs are as yet undefined at their one end. The dsRNA sequence also contains a number of much smaller ORFs. The map in Figure 4.7 shows the numbers of clones generated as well as the size and location of the three overlapping ORFs. The sequences of the primers used are shown in Table 4.1.

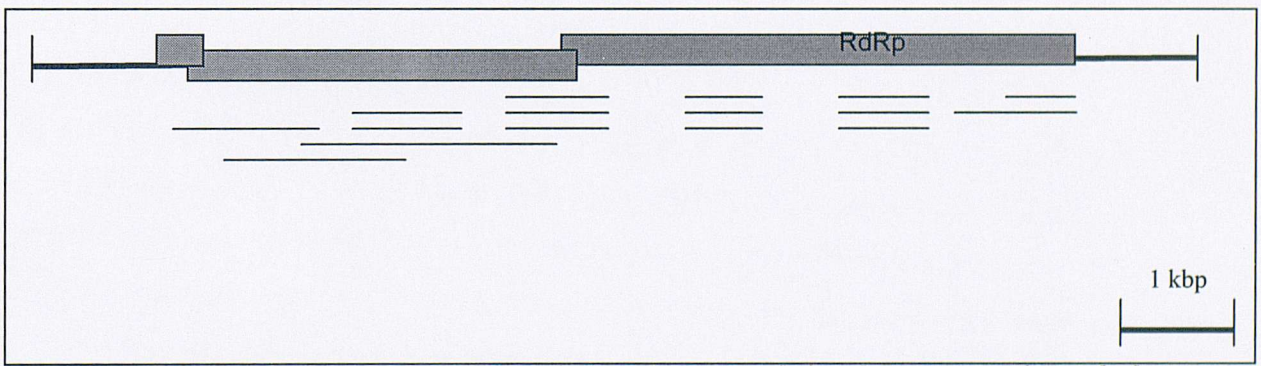


Figure 4.7: Schematic representation of the clones used in the construction of the contiguous sequence of the *P. gigantea* dsRNA1 species. The positions of the overlapping ORFs are shown. The ORFs are tentatively placed on the genome. The horizontal lines below the

ORF represent the number, size and overlap of the obtained clones. The approximate position of the RdRp motifs is shown. The undefined dsRNA sequence is shown as a straight line.

Table 4.1: Sequence of the primers used in the cDNA cloning of *P. gigantea* dsRNA1. The sequence of the primers designed for dsRNA1 is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
Phle1	CTGACGTGACCAACGCCTACG
Phle2	GGTGCAGATCGGCGCCGGC
Phle9	GGTTACGAAGGTGCGTCTGC
Phle10	GCTGTACGCGGCACCTATCC
Phle11	GGCAGTATTAGGTTTACTGC
Phle12	GCTTCTTCAACGGTGCTCTCGC
Phle13	GAACCAGAGGTCGACATACCTGG
Phle14	GAGATTGGCCGATCAATCC
Phle15b	CGACGAGGTTAGCCAAGGGG
Phle16	CGACACATTGCTAAGAACC
Phle19	CGGTTACGTAAAGCAACGG
Phle22	GAGTATAATTTGAACGACG
Phle23	CCAGAGACTGCATTCGCG
Phle24	GCTCGAAGGCGTCCAACAGC
Phle25	GACCACGGTTACGAATTACC
Phle30	CTTGATGTTTGTCTGCTACACG
Phle31	GCCAACCAGCGACAGGTGCCG

4.2.2.2 dsRNA species 2

The available dsRNA species 2 sequence is currently 6508 bp long and presented in full in the GenBank database, accession number: AM111097. Computer assisted analysis of the available sequence has revealed a dissimilar genomic organisation to dsRNA1, with the presence of two large ORFs, at least 3268 bp and 3247 bp long, both being undefined at their one end, and a number of smaller ones. The absence of a third ORF can be attributed to the amount of sequence information presently available for this dsRNA element. The map in Figure 4.8 shows the clones generated as well as the size and

location of the two overlapping ORFs. The sequences of the primers used are shown in Table 4.2.

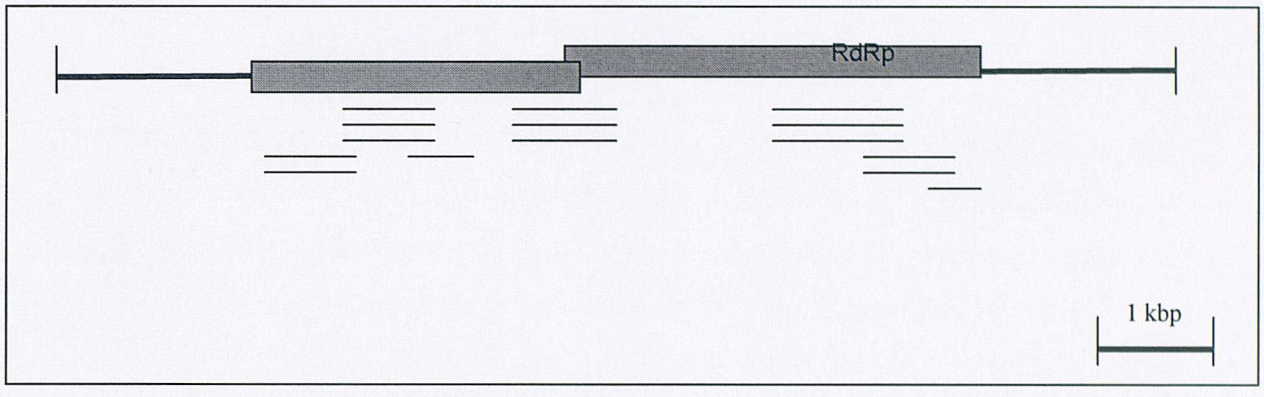


Figure 4.8: Schematic representation of the clones used in the construction of the contiguous sequence of the *P. gigantea* dsRNA2 species. The positions of the overlapping ORFs is shown, tentatively placed on the genomic map. The horizontal lines below the ORF represent the number, size and overlap of the obtained clones. The approximate position of the RdRp motifs is shown. The undefined dsRNA sequence is shown as a straight line.

Table 4.2: Sequence of the primers used in the cDNA cloning of *P. gigantea* dsRNA2. The sequence of the primers designed for dsRNA2 is shown, in the 5'- 3' orientation.

Primer name	Nucleotide sequence
Phle3	GATAAGTAGACCCGAGATAGG
Phle4	CAGTACAATCTGTTCCGATCG
Phle5	GCAAAACTCGTCCCGGAGTCC
Phle6	GTCCGTACTGCCCCTCTCG
Phle7	CACGTGCCCCACTTAACACG
Phle8	CATTCCCCCAATCATTAACACC
Phle15	CTGATTTCCTCAGATTACG
Phle16b	GATGTCGAAGGCTTTGC
Phle17	CCTGCTCCCAATATTCGC
Phle18	GCCAAGACACAGTCC
Phle20	GATTGCTAGCGATTGCACCC
Phle26	GGTAATGGAGCCGGGTTCGCC
Primer name	Nucleotide sequence
Phle27	CGACATCAGCGTTGGTGACTGC

Phle28	CGGCTTCAACCACGCCTTAGC
Phle29	CGACTAGAGCTGATCATGG
Phle30b	GGACCAAGCCAGTAAGG
Phle32	CCAGGCCAGTTAGGCAAGTG

The two dsRNA species have an overall 42.3% nucleotide sequence similarity, calculated using the Needleman-Wunsch global alignment tool. However, this similarity is not distributed evenly on the genome. It is strongest at the RdRp motifs and much less conserved in the remainder of the genome. The ORFs containing the RdRp motifs share a 22.5% peptide sequence identity and a 36.9% sequence similarity. The ORFs immediately upstream (ORF2 of dsRNA1 and ORF1 of dsRNA2) share a 21.1% peptide sequence identity and a 34.2% sequence similarity. These were calculated using the Waterman-Smith local alignment tool. These data suggest that although the two dsRNA elements are related they do not represent two different isolates of the same species, but two distinct dsRNA species belonging to a novel mycoviral genus.

4.2.3 Genes and possible functions

4.2.3.1 Gene prediction

The predicted protein product of the 3' most ORFs on both dsRNA elements showed very significant similarity to mycoviral RdRp proteins at the peptide sequence alignment level (Figure 4.9). Thus, it can be safely proposed that part, or all, of these predicted protein products constitute a novel viral RdRp. The RdRp had the highest alignment scores with the totiviral RdRp protein products, which is also reflected in the phylogenetic tree that was constructed based on the RdRp sequences (Figure 4.11).

The ORFs located 5' to the RdRp-containing ones were compared against the global nucleotide and peptide databases for their predicted protein products. The predicted protein products did not show any similarities to proteins predicted from known viral dsRNA sequences, nor to any of the sequence information in the NCBI sequence database (using the BLASTp search tool). Thus, the potential function of their protein products remains unknown. However, using the NCBI-BLAST2 pairwise protein

alignment display tool, limited similarities were found to a few characterised proteins. They are all derived from a phage (for dsRNA1) and from dsDNA viruses (for dsRNA2). These alignments are shown in Figure 4.10, and should be taken cautiously into consideration as they only refer to restricted areas of the entire ORF.

Family	Virus	1	2	3	4	5	6	7	8
Chrysoviridae									
	HvV145S	CLIGRR 73	WMTKGSLVS 63	GHKDRVL 52	DWANFNVQHS 52	GLYSGWRGTTWDNTVLNGCY 20	DQCGDDVD 34	EFFRV 12	RGLAT
Totiviridae									
	UmV-H1	ALYGRG 66	WLVSGSSAG 61	GKARAIY 55	DYPDFNSMHT 63	GLYSGDRDTTLINTLLNIAY 20	LCHGDDII 35	EYLRI 13	RCVAT
	HvV190S	TLQGRY 61	WCVNGSQNA 46	GKDRAIF 55	DYDNFNSQHS 45	TLMSGHRATTFTNSVLNAAY 15	LHAGDDVY 35	EFLRL 12	RAIAS
	ScV-L-A	VLMNRG 57	WVPGGSVHS 50	GKQRAIY 52	DYDDFNSQHS 52	TLLSGWRLTTFMNTVLNWAY 15	VHNGDDVM 34	EFLRV 16	RSCAT
	ScV-La	VLENGV 58	IMPGGSVHS 51	GKPRALY 51	DFDDFNSQHS 52	TLFSGWRLTTFFNTALNYCY 13	LHNGDDVF 34	EFLRV 14	RGIAT
	CmV	TLQGRG 60	WCVNGAHTS 51	GKTRAIF 55	DYDDFNSQHS 45	TLMSGHRGTTFFNSILNGVY 15	MHTGDDVY 35	EFLRV 12	RSIAA
	GaRV-L1	VLLGRA 65	WAVNGSQSG 49	GKTRAIF 55	DYDDFNSHHS 46	TLMSGHRGTTFINSVLNKAY 15	LHVGDDVY 35	EFLRN 12	RAVAS
	GaRV-L2	VLLGRA 65	WAVNGSQSG 49	GKTRAIF 55	DYDDFNSHHS 46	TLMSGHRGTTFINSVLNKAY 15	LHVGDDVY 35	EFLRN 12	RAVAS
	SsRV1	TLLGRA 61	WCVNGSQND 51	GKTRAIF 55	DFDDFNSHHS 45	TLPSGHRGTTIVNSVLNAAY 15	LHTGDDVY 35	EFLRM 12	RSVAS
	SsRV2	TLQGRA 65	WAVNGSQSG 50	GKTRAIF 55	DYDDFNSHHS 46	TLMSGRRGTTYISSVLNEVY 15	IHVGDDVY 35	EFLRL 12	RAVAS
	HmTV1-17	SLQGRM 62	WCVNGSHTN 51	GKTRAIF 55	DYDDFNSHHA 46	TMMSGHRGTMFINSVLNAVY 15	LHAGDDVY 35	EFLRM 12	RSIAS
Unassigned									
	ACD-dsRNA 2	VLNHRA 67	ATPAGSAFT 55	SKLRALF 52	DARNFNILHT 78	GMFSGTRFTMLYNTILNRAY 14	LHSGDDVY 34	EFLRI 12	RSAAT
	ACD-dsRNA 3	VLNHRA 67	TTPAGSAFT 55	SKLRALF 52	DAKNFNILHT 78	GMFSGTRFTMLYNTVLNRAY 14	LHSGDDVY 34	EFLRI 12	RSAAT
	AbV	LLIGRR 54	RKPGHSQVP 75	NASRSLW 53	DYANFNEQHS 53	GLLSGWRCTAYINNLINIAQ 20	DTGGDDGC 35	EFFRL 12	RMLAS
		g	p	s ly	n i	gm m st	c i	y	T
	Consensus	V* GRA	W VN**Q	G*T**IF	*YDD**SH*S	TLM**H*G*TFINSV** A*	LH ***VY	*F**	*SV*S
P.gigantea									
	dsRNA1	LLVGRA 69	VAPRGSIGT 50	GRLRQII 57	DYADFNYLHT 64	SLWSGWRTTTMINNTMNLVY 20	HVNGDDGD 35	EYLRI 12	RSCAS
	dsRNA2	FDPQRA 63	WAASGGAPG 70	GKRRAIW 62	DFSDFNINHQ 63	CLASGERATSFTNTILSRVY 19	SQQGDDVF 36	EFLRL 14	RAGLG

Figure 4.9: RNA-dependent RNA polymerase peptide sequence alignments. The peptide sequences from the motif areas of twelve sequenced viruses are aligned. The distances between consecutive motifs are shown in numbers of peptides. Viruses are categorized according to family. Asterisks indicate identical amino acids, while capital and lower case letters denote those found in at least five sequences or represent the alternative amino acids.

Sequences: *Helminthosporium victoriae virus 145S* (HvV145S) (GenBank accession number)NC005978; *Ustilago maydis virus H1* (UmV-H1) NC003823; *Helminthosporium victoriae 190S virus* (HvV190S), U41345; *Saccharomyces cerevisiae virus L-A* (ScV-L-A), M28353; *Saccharomyces cerevisiae virus L-a* (ScV-L-a), U01060; *Coniothyrium minitans virus* (CmV), AF527633; *Grammeniella abietina RNA virus L1* (GaRV-L1), L13218; *Grammeniella abietina RNA virus L2* (GaRV-L2), AF337175; *Sphaeropsis sapinea RNA virus 1* (SsRV1), AF038665; *Sphaeropsis sapinea RNA virus 2* (SsRV2), AF039080; *Helicobasidium mompa* Tanaka *virus* (HmTV1-17), U41345; Amasya cherry disease related dsRNA2 (ACD-dsRNA2), AM085134; Amasya cherry disease related dsRNA3 (ACD-dsRNA3), AM085135; Agaricus bisporus mycovirus (AbV), X94362

Figure 4.10: Alignments results using the predicted peptide sequence of *P. gigantea* dsRNA elements 1 and 2, ORFs 2 and 1 respectively. The three highest scoring results are shown. In the case of dsRNA1 there was only one high scoring result. They have been generated using the pairwise NCBI-BLAST2 protein alignment display. The identity and similarity scores between the two sequences are shown, as well as the UNIPROT accession number, the name and the organism of origin for each one of the aligned protein sequences.

A) *P. gigantea* dsRNA1

UNIPROT: Q6VT13_9VIRU Q6VT13 Hypothetical protein orf4C, *Vibrio parahaemolyticus phage VP16C*. Identities = 22/66 (33%), Positives = 32/66 (48%)

```
dsRNA1:  LAVQMRAGIDAVTASVDLNASTLRLRTGEAQTGVVLIDRVLEDTNEYVRFGKMGDYHSAY
         LA  + A +DA+TA +  N S   L TG+    V  +D  L+   E   +G +   + A
Sbjct:   LAAILNAYLDAITALLTANVSEYTLNTGQGSQRVTRLD--LQQLQE--NYGLLYRQYDAL

dsRNA1:  VSLAGN 416
         VS  GN
Sbjct:   VSRCGN 74
```

B) *P. gigantea* dsRNA2

1) UNIPROT: Q4A2Z7_9PHYC, *Emiliania huxleyi virus 86*, Putative membrane protein precursor, Identities = 32/128 (25%), Positives = 49/128 (38%)

```
dsRNA2:  PSKPAPPPSPEPQNTPPTHNTRSGSPTLSYESHNPRHVSEVPTMPATITTQTVDGAAKIN
         PS P PP  P P + PP   + S  P     S  P  +S  P  P +I+      +   +
Sbjct:   PSPPYPPSLPSPPSLPPPSASPSPPPPSLSPSPPPPSLSPSPP-PPSISPSPPPPSLSPS

dsRNA2:  APAYQYHGASPSAIRAEAQRLTELAAAAAESSNFALPPPSSIPPVEHSTTPPVIGPKSGPI
          P      + P    + +     L+ +    + +  PP S PP   S +PP   P S
Sbjct:   PPPPSVSPSPPPPSLSPSPPPPSLSPSPPPPSLSPSPPPSPPPPSTSPSPPPRPPPSASP

dsRNA2:  SRDSTSIT 880
         S    S++
Sbjct:   SPPPPSLS 329
```

2) UNIPROT: Q9DVW0_9BACU, *Plutella xylostella granulovirus*, Identities = 23/64 (35%), Positives = 25/64 (39%)

```
dsRNA2:   PAGSRLPGNPVAAASKAAWATPPPKPPPRAPLNTEGLRPNTPWLP--NPPTMPVTTQPPP
          P     P  P          TPPP PPP  P       P  P +P  +PPT P    PPP
Sbjct:    PTPPTSPATPPPTFPPTPPPTPPPTPPPTPPPTPPPTPPPPPIIPAPSPPTAPPQPPPPP

dsRNA2:   PAIP 977
          P  P
Sbjct:    PPPP 104
```

3) UNIPROT:Q8UZ11_PRVKA Q8UZ11, UL36 protein. *Pseudorabies virus* strain Kaplan, Length = 3084, Identities = 62/250 (24%), Positives = 90/250 (36%)

```
dsRNA2:   PHEPTRPSFLPKGRPSKPAP--PPSPEPQNTPPTHNTR--SGSPTLSYESHNPRHVSEVP
          P     P   P  +P +  P   P P+PQ  PP  + +  +G+   +     PR  +  P
Sbjct:    PAAKLAPPAPPPAKPVETTPLTKPQPQPQQGPPAAHKKPAAGTKAAAAPKQQPREPAPKP

dsRNA2:   -TMPATITTQT---VDGAAKINAPAYQYHGASPSAIRAEAQRLTELAAAAESSNFALPPP
           + P  I       ++    +  P Y     +P     EA +    AAA  +     PPP
Sbjct:    HSSPRKIQPSLKARIEPPPPVINPPYPATAPAPETAPPEAPQAQPPAAAKPTPQPQGPPP

dsRNA2:   SSIPPVEHSTTPPVIGPKSGPISRDSTSITYARIPRHNSVVRDGILDSNNDRLTKNVIVD
             PP   +  PP   P + P +  +T+   A  P+     R
Sbjct:    PPQPP--SAQAPPAQKPPAQPATAAATTAPKAT-PQTQPPTRA----QTQTAPPPPSAAT

dsRNA2:   GIVRYPAGSRLPGNPVAAASKAAWATPPPKPPPRAPLNTEGLRPNTPWLPNPPTMPVTTQ
             + P   + P +  AA  + A   PP  PPP A   T   RP  P  P PP+   T
Sbjct:    AAAQVPP--QPPSSQPAAKPRGAPPAPPAPPPPSA--QTTLPRPAAPPAPPPPSAQTTLP

dsRNA2:   ---PPPPAIP
             PPPP+ P
Sbjct:    RPAPPPPSAP
```

4.2.3.2 *In vitro* translation of purified dsRNA

As an attempt to verify the suggestion that each dsRNA element from *P. gigantea* translates at least one protein product an *in vitro* translation was carried out, using gel extracted, denatured dsRNA as template in a cell-free translation system (wheat germ extract). The translation reaction was monitored every 30 min over a time period of 3 h. There are multiple translation products present, which show that the dsRNA molecules have a protein coding capacity (data not shown). However, this experiment would need to be repeated using each of the two dsRNA species separately. That would be a better indication for the number and relative abundance of the *in vitro* produced viral proteins from each individual dsRNA element.

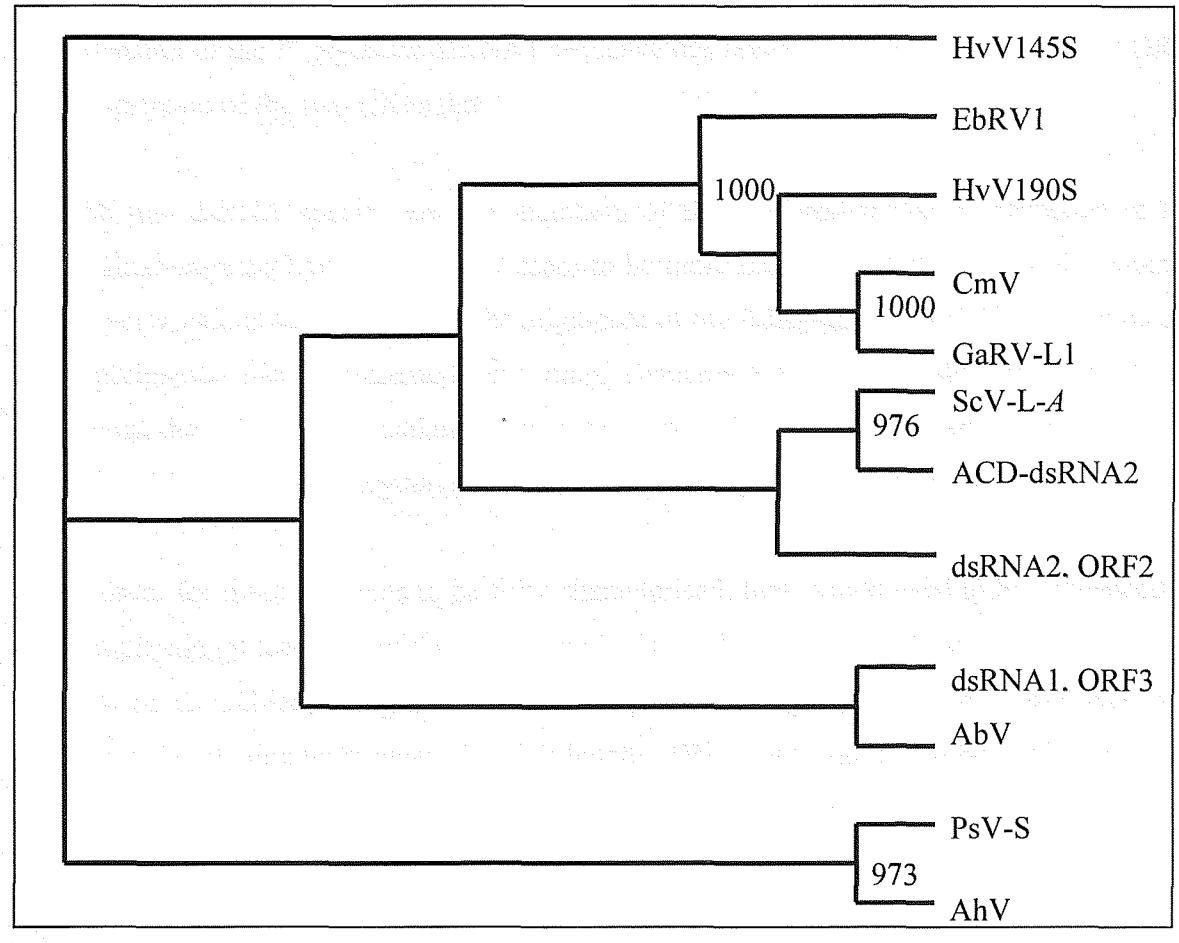
4.2.4 Viral phylogeny

The origin of the *P. gigantea* (isolate TW2) dsRNA elements is difficult to determine as they do not share any similarity with viral or any other sequences currently present in the GenBank database. It can be said that based on the nature (unencapsidated dsRNA) and size of their genome they bear some general resemblance to the *Endornaviridae*. However, the RdRp sequence alignments showed that they do not share any significant sequence homology with the *Endornaviridae*. Phylogeny could only be investigated for

the ORFs whose sequence had significant similarities to that of other dsRNA RdRp elements. An unrooted phylogenetic tree based on the neighbour-joining method with a 1000 replicative bootstrap search was constructed using viral sequences belonging to the *Totiviridae, Partitiviridae*, the unassigned mycoviral sequence of *A. bisporus* dsRNA and the RdRp encoding sequence of ACD-related dsRNA2. The HvV145*S* Chrysovirus was assigned as an outgroup. The results can be seen in Figure 4.11.

In the phylogenetic tree, the *Partitiviridae* (PsV-S and AhV) and *Totiviridae* (EbRV1, HvV190S, CmV, GaRV-L1, ScV-L-A) sequences cluster together. The ACD-related dsRNA encoded RdRp also clusters in the same group. The *P. gigantea* polymerase sequences show that they are not very closely related as they only share a 22.5% sequence identity. The dsRNA1 RdRp is closest related to the *A. bisporus* unclassified dsRNA mycovirus, while the dsRNA2 RdRp is closely related to the ACD-related totivirus-like polymerase sequence. Although the phylogram shows that these dsRNA elements are distinct from the known mycoviral genera, it cannot be considered as entirely representative of a true evolutionary relation. It does not take into account the remainder of the genomic sequence that has no similarity to the known mycoviral sequences, or the genomic organisation of the dsRNAs themselves, that is atypical. Thus, it is suggested that *P. gigantea* dsRNAs are only distantly related to the characterised mycoviral families and they have probably diverged a considerable time ago in evolutionary terms.

Figure 4.11: Unrooted phylogenetic tree based on the Neighbour-joining method with a 10,000 replicate bootstrap search. The tree was constructed using the RdRp peptide sequence of the viruses, their full names, abbreviations and GenBank accession numbers shown below the Figure.



References: *Helminthosporium victoriae* 145S virus (HvV145S), (GenBank accession number) NC005978; *Eimeria brunetti RNA virus 1* (Eb-RV1), AF356189; *Helminthosporium victoriae 190S virus* (HvV190S), U41345; *Coniothyrium minitans virus* (CmV), AF527633; *Grammeniella abietina RNA virus L1* (GaRV-L1), L13218; *Saccharomyces cerevisiae virus L-A* (ScV-L-A), M28353; *Agaricus bisporus* mycovirus (AbV), X94362; *Penicillium stoloniferum virus S* (PsV-S), AY156521; *Atkinsonella hypoxylon virus* (AhV), L39125.

4.2.5 Chapter summary and future prospects

The sequences of the two dsRNA elements isolated from the fungus *P. gigantea* (isolate TW2) have been reviewed and expanded. The previously available sequence contained multiple stop codons, unlike most other mycoviruses described to date, as well as areas where only a single coverage of the genome had been achieved. These latter areas were re-sequenced and their information was confirmed and consolidated. Additionally, the number and position of the stop codons on the genome was also elucidated. The

extension of the *P. gigantea* dsRNA1 sequence has revealed the presence of a third ORF, 5' upstream of the two ORFs that have been confirmed.

The two dsRNA species are not members of the *Endornaviridae* as proposed in the original working hypothesis. They seem to be more closely related to totiviral elements, however, this is solely based on the alignment of the RdRp-encoding ORFs. Based on the experimental results presented here, these elements are most probably unencapsidated, though the viral particle isolation experiments would need to be repeated in parallel with a dsRNA virus-free *P. gigantea* isolate, so that a complete set of data can be obtained.

In order for these elements to be fully characterised, they would need to be sequenced to completion (at least one of the two is required). A cDNA full-length copy of the virus can then be assembled, using RT-PCR with a proof-reading enzyme (*e.g.* Bio-X-Act) and molecular cloning techniques. The full-length cDNA copy could then be used as a vector or serve as the basis for the construction of vectors that can be introduced into a virus-free *P. gigantea* isolate.

Potential for utilising the dsRNA elements

The *P. gigantea* fungus is already used as a bio-control agent. However, its efficiency in the field is limited by the amount of antagonistic saprophytic fungi present at the site of application. The two dsRNA molecules could help understand the ways in which *P. gigantea* functions in order to outcompete other fungi. This can be achieved by using the two dsRNA elements as the basis for constructing vectors (or using the dsRNA elements themselves) for RNA-silencing studies. The dsRNA elements can also form the basis for constructing expression vectors for *P. gigantea.* Their presence is constant in the cytoplasm and they do not have an extracellular life-cycle, thus they are good candidates for creating stable transgenic *P. gigantea* lines that can then be tested for their effects utilising different isolates from the naturally existing *P. gigantea* and *H. annosum* population variation (Vainio and Hantula 2000 and Otrosina *et al.* 1993 respectively).

CHAPTER FIVE

Reviewing the CYAV dsRNA element sequence, CYSDV RNA2 assembly and PEV1, UGT gene expression assay

Introduction

The methods and technology already available in the laboratory, such as RLM-RACE and the construction of cDNA libraries, were optimised in three separate projects before being applied in the main body of research presented in chapters 3 and 4 of this thesis. These three separate projects are presented here, as individual entities. However, it should be noted that in the latter two cases they constitute part of a wider series of projects.

5.1 Reviewing the CYAV dsRNA element sequence

5.1.1 The CYAV dsRNA element

During the sequencing of the *Beet pseudo yellows virus* (BPYV) from infected lettuce plants (*Lactuca spp.*), a second, unrelated viral sequence was detected. It was named *Cucurbit yellows-associated virus* (CYAV) dsRNA element and was partially sequenced (Coffin 1992, PhD thesis). BPYV is a ssRNA closterovirus (Coutts and Coffin 1996), while CYAV showed extensive sequence similarities to the dsRNA totiviruses. Although it has not been possible to establish the origin of CYAV, the presence of multiple viruses co-infecting plant or fungal hosts has been previously documented (Tuomivirta and Hantula 2005; Park *et al.* 2005). It is likely that CYAV causes a latent, permanent infection to the plant, or to a fungus growing on it. It should be noted that the plants containing the dsRNA element were provided directly by the growers, therefore there is a multitude of ways in which CYAV could have infected the plants. Thus the possibility that CYAV might even be an insect associated virus can not be excluded.

The assembled, 1927 bp long, CYAV sequence was deposited in the GenBank database with Accession number X92203. However, it contained a small number of sequence ambiguities and a stop codon located proximally, 3' to the GDD conserved RdRp motif (motif C, Bruenn 2003), which is not seen in any other totivirus. In order to resolve these ambiguities an already existing library was analysed and the resulting sequences were assembled and compared to the original ones. The sequence information was then analysed and compared to other known totiviral sequences.

5.1.2 Results and Discussion

5.1.2.1 Sequencing of the CYAV dsRNA derived cDNA library

A cDNA library of clones was generated by random priming on the dsRNA-enriched plant tissue extract. The clones were available either in the form of a glycerol stock bacterial culture or as an isolated plasmid (miniprep). In the first case the bacterial cultures were re-plated before processing the clone; in the second, the plasmids were transformed into the JM109 competent cells before further processing. Of the 117 rescued clones, a total of 64 were sequenced, the remaining being duplicates. The relative size of the majority of the clones and their overlap on the existing CYAV sequence are shown in Figure 5.1

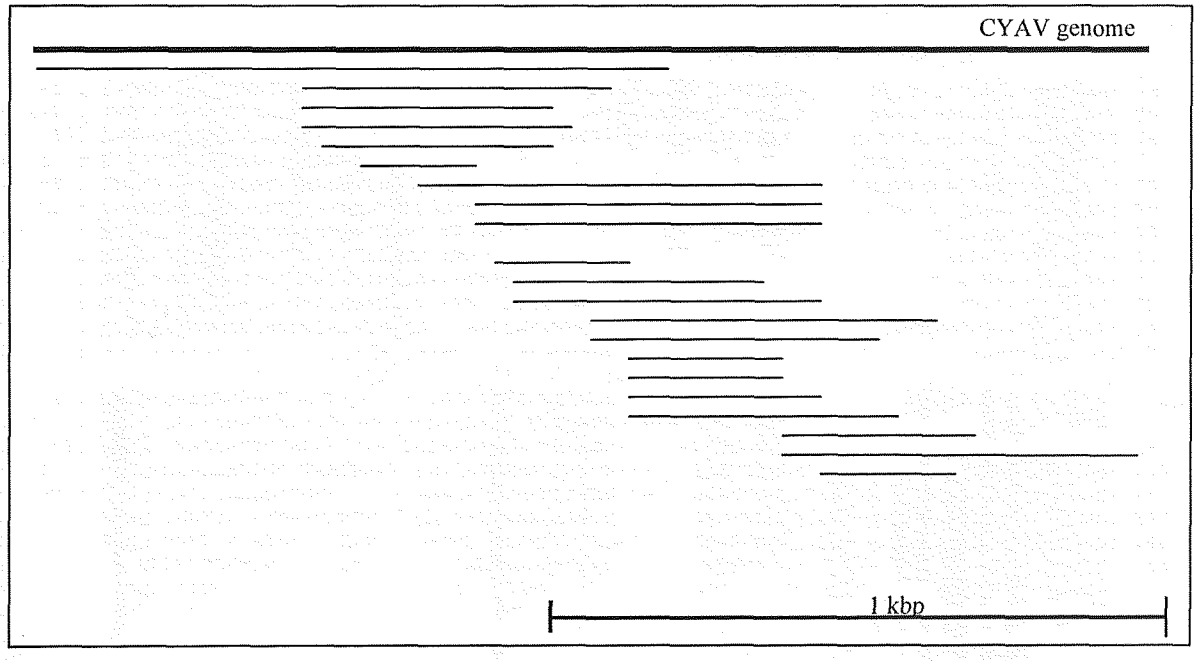


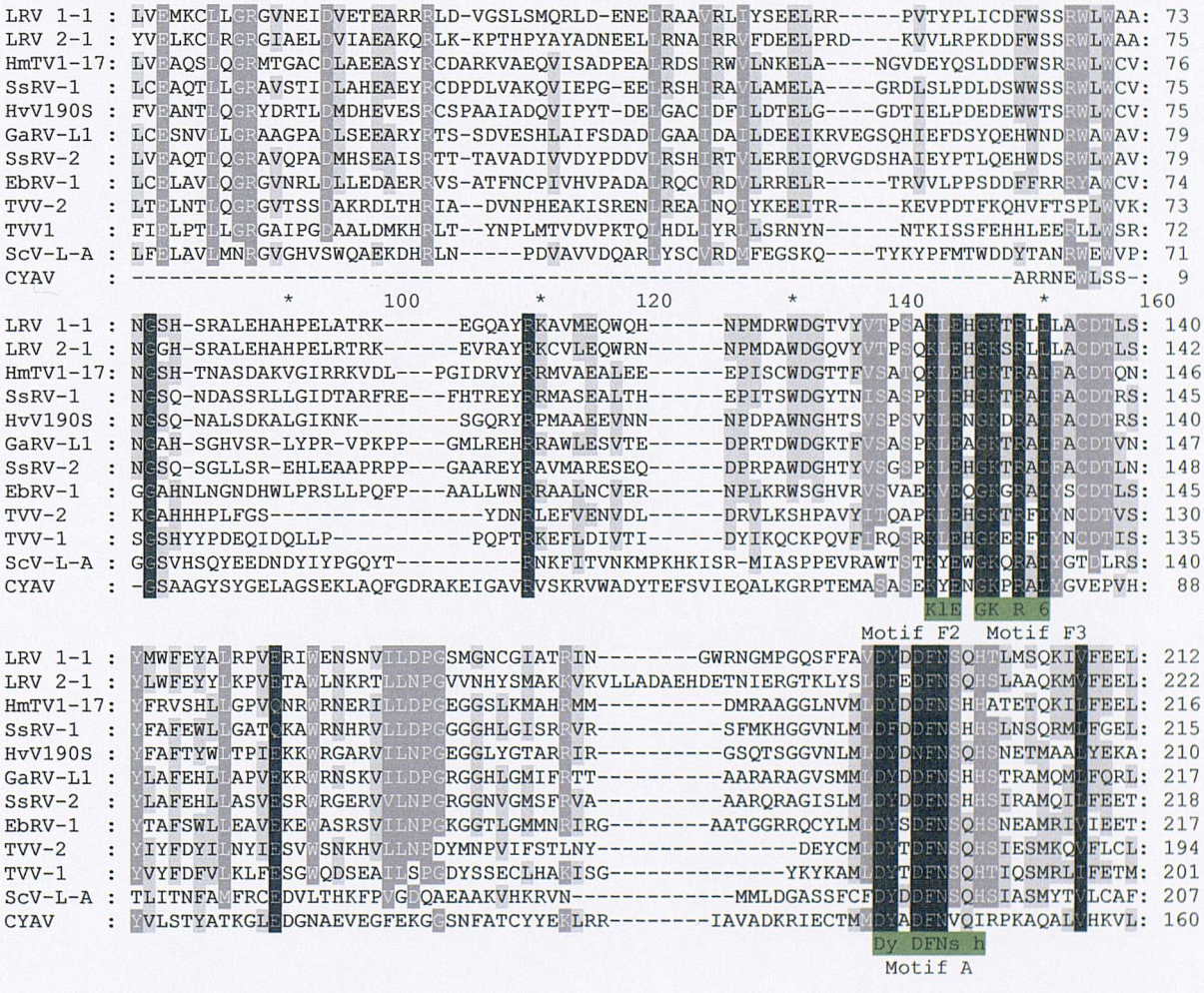
Figure 5.1: Schematic representation of the overlap of the clones used in the construction of the contiguous CYAV sequence. The approximate position of a third of the clones is shown by the horizontal lines representing the size and overlap of the obtained clones.

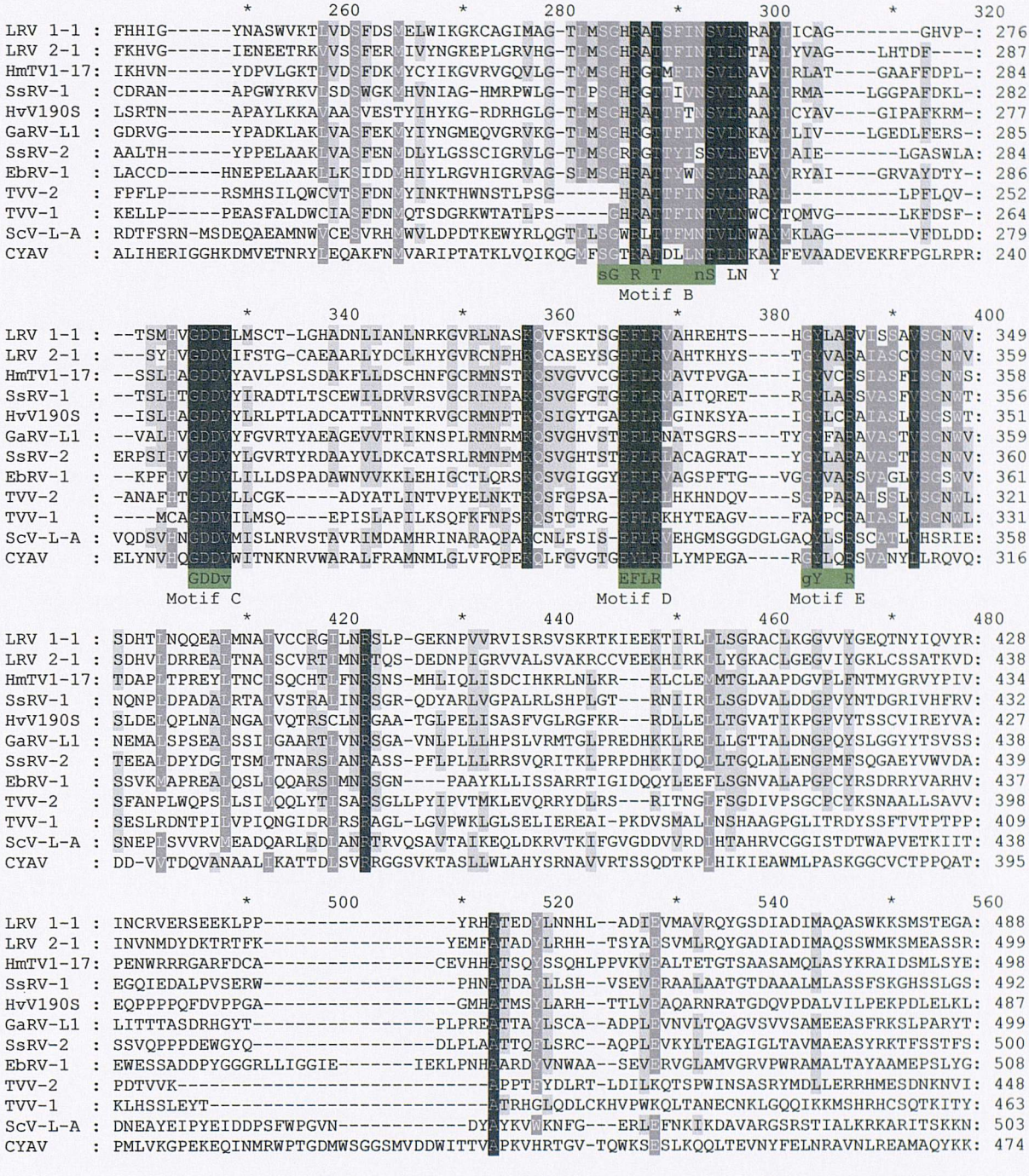
5.1.2.2 CYAV sequence information and viral phylogeny

Computer assisted analysis of the resulting contiguous sequence demonstrated the presence of a single contiguous ORF, not two as was the case in the original set of data. The small number of base ambiguities were also resolved. The revised 1903 bp long CYA*V* sequence was deposited in the GenBank database with the same Accession

number as the original sequence (X92203). The assembled CYAV sequence is approximately half of the average totiviral genome in length, thus the present sequence represents approximately half of the total viral genome. Sequence alignment with other known totiviral genomes confirmed the close relationship of the CYAV RdRp to totiviral RdRp genes (Figure 5.2). However, as there was no stop codon found in the assembled sequence either at the 5' or the 3' end of it, it is not possible at present to calculate accurately the size or to estimate the molecular weight of this protein product.

Figure 5.2: CLUSTAL_ X alignment results, highlighted using GeneDoc, of the peptide sequence of CYAV. The entire CYAV peptide sequence is shown aligned against the homologous, but partial, peptide sequence of ten totiviral RdRp peptide sequences and part of the gag-pol fusion protein of ScV-L-A. The GenBank accession numbers of the peptide sequences are shown below the Figure. The conserved RdRp motifs (Bruenn 2003) are highlighted.



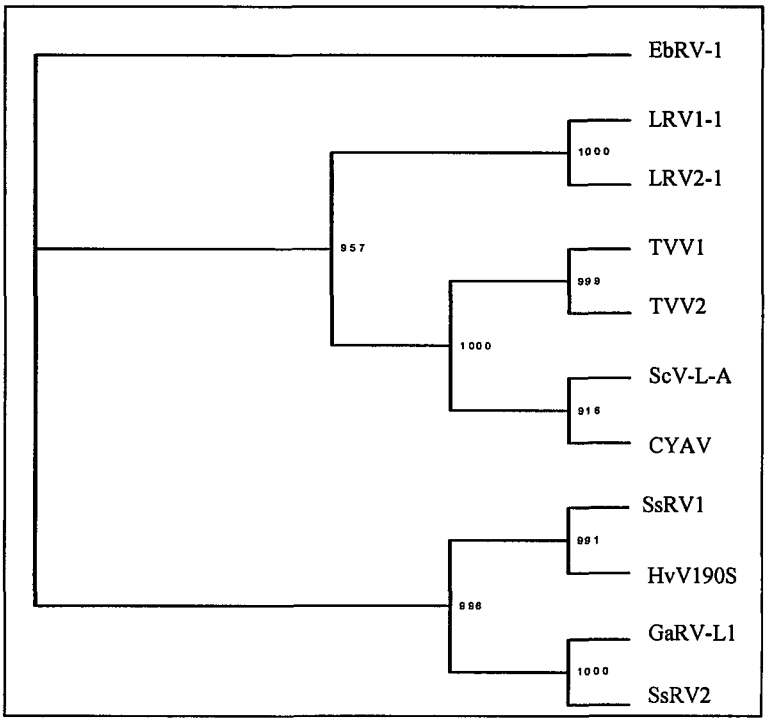

Motif B
Motif C
Motif D
Motif E

```
                    *       660         *       680         *       700         *       720
LRV 1-1 : SNNSCAIEGVLPYNEATSLAQKLPCGGVVIQVIHNVYV-----------------------------------------: 603
LRV 2-1 : TMRLCAIDGVLPYAMLVNWHAPVPCNGVRLVVDHNVYA-----------------------------------------: 612
HmTV1-17: RKCSFFGWIPYSDAAALQKKADHIKIMNHTPVRM---------------------------------------------: 608
SsRV-1  : HPESKNIIGSLSHSDASALGKATRSGNIYTLTPVFV-----------------------------------------: 604
HvV190S : -----------------------------------------------------------------------------:   -
GaRV-L1 : FKHGCMIATPMSYSDAATFGKRTVSTVLTCPLDAHV-----------------------------------------: 610
SsRV-1  : HNHGCVVATPLSYADAATYGRRTTCGVLVCGLHFYV-----------------------------------------: 611
EbRV-1  : WGGTNEGICVCGLLPYADAAGAASRVLGNAIHVSVPVAV-----------------------------------------: 622
TVV-2   : LRFATCRVGPATAIRLGSLWPANRINLIKPVYV-------------------------------------------------: 527
TVV-1   : MYACNACTSSVSGDAFLRATSKMAGVLITSLISSSS-----------------------------------------: 547
ScV-L-A : ---QARYAMQAALDSSDPLRALQVIL-----------------------------------------------------: 568
CYAV    : ASSVFKSAAATATALGTSTIDTLKFMAATTTSITAALEEAKTVTMERVRCGTQSDIKTLLDRNVGVLSNFGSLMPAALMG: 634
```

References: LRV 1-1, (GenBank accession number) M92355, Stuart *et al.* 1992; LRV 2-1, U32108, Scheffter *et al.* 1995; HmTV1-17, AB085814, Nomura *et al.* 2003 ; SsRV-1, AF038665, Preisig *et al.* 1998; HvV190S, U41345, Huang and Ghabrial; GaRV-L1, L13218, Wang *et al.* 1993; SsRV-2, AF039080, Preisig *et al.* 1998; EbRV-1, AF356189 Fraga *et al.* unpublished ; TVV-2, AF127178, Bessarab *et al.* 2000 ; TVV-1, U08999, Tai and Ip 1995; ScV-L-A, M28353, Diamond *et al* 1989.

An unrooted phylogenetic tree based on the neighbour-joining method with a 10000 replicate bootstrap search was constructed using the same peptide sequences as in the CLUSTAL_X sequence alignments (Figure 5.3). In the tree, totiviral sequences are divided into two main groups, those infecting yeast and protozoa and those infecting fungi. The EbRV-1 is the most distantly related RdRp protein sequence and is shown as an outgroup. The available RdRp CYAV peptide sequence is most similar to that of the yeast (ScV-L-A) and protozoan (TVV-1, TVV-2 and LRV1-1, LRV2-1) totivirus polymerases, rather than the fungal ones. This result suggests a higher probability of the virus having originated in a yeast (or a protozoan) host than in a fungal host. It should be noted at this point that the CYAV genomic sequence shares very limited sequence similarity with the ACD-related totivirus-like elements (53% and 50.7% similarity over the corresponding genomic areas of dsRNAs 2 and 3 respectively). However, this does not exclude the possibility that it might also be a totivirus-like element and not a strictly defined dsRNA totivirus.

Figure 5.3: Unrooted phylogenetic tree based on the Neighbour-joining method with a 10000 replicate bootstrap search. The tree was constructed using the RdRp peptide sequence of the viruses shown in Figure 5.2.



5.2 CYSDV RNA2 assembly

5.2.1 The CYSDV ssRNA virus

Cucurbit yellow stunting disorder virus (CYSD*V*) is a positive-stranded ssRNA virus with a bipartite genome. It belongs to the genus *Crinivirus* within the family *Closteroviridae*. These RNA viruses are transmitted by whiteflies in a semi-persistent manner (Karasev 2000). The sequence of each of the genomic components has been published: RNA1 is 9126 nt long (Coutts and Livieratos 2003; Genbank accession number AJ537493), while RNA 2 is 7976 nt long (Aguilar, *et al.* 2003: GenBank access. no. NC004810; Livieratos and Coutts 2002: GenBank access. no. AJ439690).

In an attempt to study the gene expression and dissect the molecular pathogenicity and infectivity of the virus, a full-length cDNA copy of the entire sequence for each genomic fragment would need to be constructed. The cDNA copy, incorporated in a selectable plasmid, would be translated into mRNA using a commercially available transcription kit and transformed into the desired plant protoplasts. This approach has been previously successful in the study of other plant viruses, *e.g.* the *Cassia yellow blotch virus* (CYB*V*),

genus *Bromovirus* (Iwahashi *et al.* 2005) and the *Citrus tristeza virus* (CTV), family *Closteroviridae* (Satyanaryana *et al.* 1999).

5.2.2 The CYSDV RNA2 assembly

Assembly approach

The dsRNA replicative form of the virus was extracted from infected leaf tissue and was used as a template for the RT-PCR amplification of specific genomic areas. These amplified areas were designed so that they contain an overlap and a restriction enzyme recognition site which is unique, if possible, for the entire genomic component. Thus, the digest of the fragments will allow complementary, 'sticky' ends to be created for overlapping fragments that can be then ligated and inserted into a vector of choice (pBluescript II SK (+) was used in this series of experiments). The CYSDV RNA1 genomic fragment had been previously assembled into a full-length cDNA clone (Eliasco 2005, PhD thesis). Thus the assembly of the remaining genomic component, RNA2, was attempted.

Assembly and expression

Five fragments were PCR amplified from the CYSDV RNA2 genomic component, including a restriction enzyme recognition site in each. These are shown in Figure 5.4. The fragments were amplified using a proofreading Taq polymerase enzyme (Bio-X-Act, Bioline), and then were sequenced, in order to avoid the generation and inclusion of point mutations in the assembled full-length RNA2 construct.

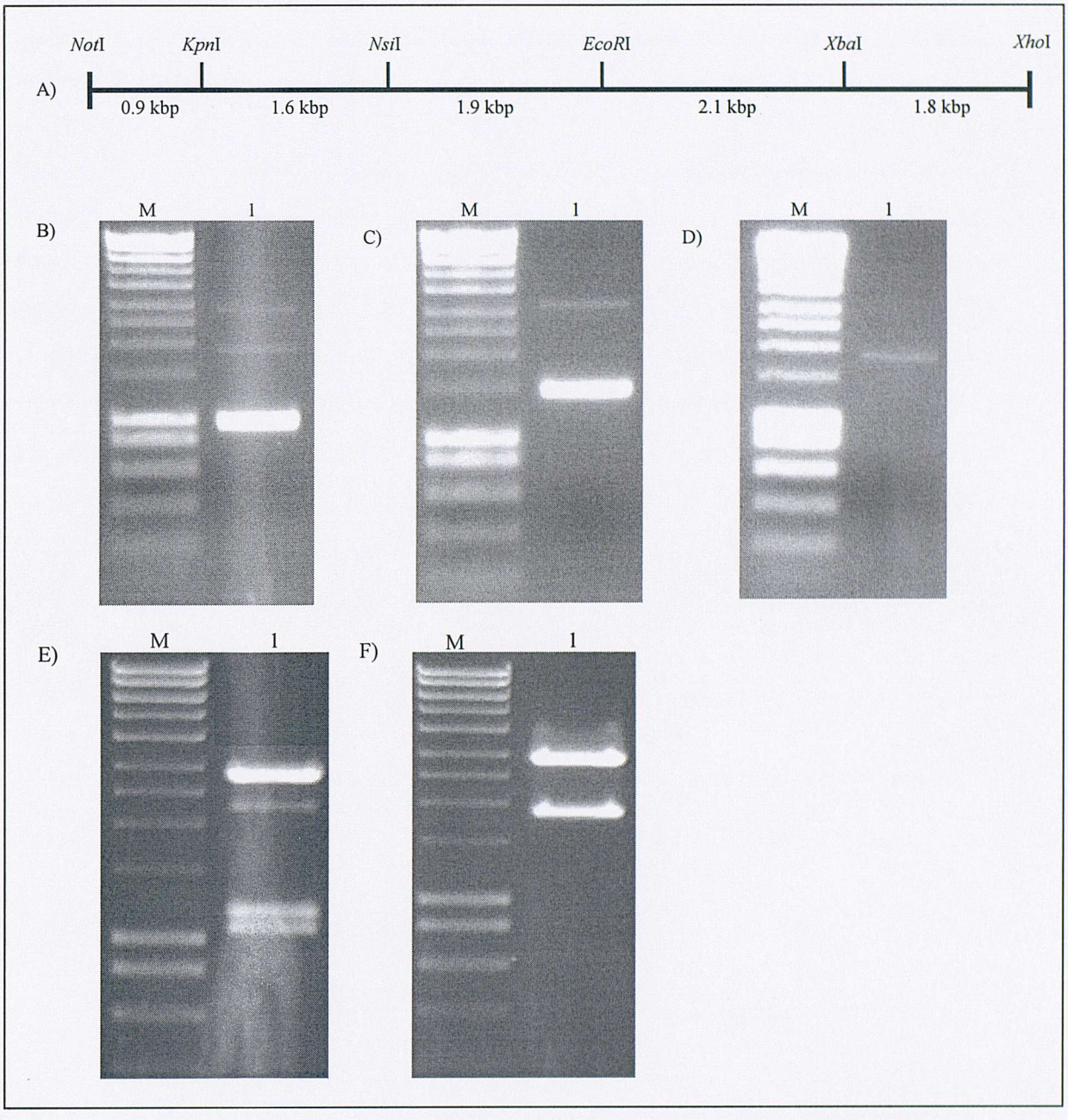


Figure 5.4: The RT-PCR amplified genomic fragments of CYSDV RNA2. A map of the CYSDV RNA2 amplified fragments is shown in A) containing the names and positions of the restriction enzymes used. Lanes M contain the 1 kb molecular marker ladder. B) In lane 1 is the 0.9 kbp *Not*I – *Kpn*I digested fragment. The fainter bands are dimeric and trimeric forms of the digested fragment. C) In lane 1 is the 1.6 kbp, *Kpn*I – *Nsi*I digested fragment. The fainter band is a dimeric form of the molecule. D) In lane 1 is the 1.9 kbp, *Nsi*I – *Eco*RI digested fragment. E) In lane 1 is the 2.1 kbp *Eco*RI – *Xba*I digested fragment. There are two bands present, of 1.0 and 1.1 kbp, due to the presence of an internal *Eco*RI site. The band of higher molecular weight is the

plasmid. F) In lane 1 is the 1.8 kbp, *Xba*I – *Xho*I digested fragment. The band of higher molecular weight is the plasmid.

The assembly of the RNA2 genomic fragment required two different approaches, as a key fragment combination resulting from the first approach (marked with an asterisk in Figure 5.5) proved toxic to the transformed cells. In this case different cell strains were initially used (JM109, BL21 and SURE™ Supercompetent cells [Stratagene]), in order to overcome the absence of transformed bacterial colonies, but the same cytotoxic effect was displayed in all cases. Thus, a second approach was followed, as shown in Figures 5.6 and 5.7. The fragment associated with the cell toxicity did not contain any of the ORF or individual genes present on RNA2 in their entirety. It is thought that a cryptic transcription initiation site, internal to these genes, might be responsible for the production of a truncated protein product that consequently results in the toxic phenotype.

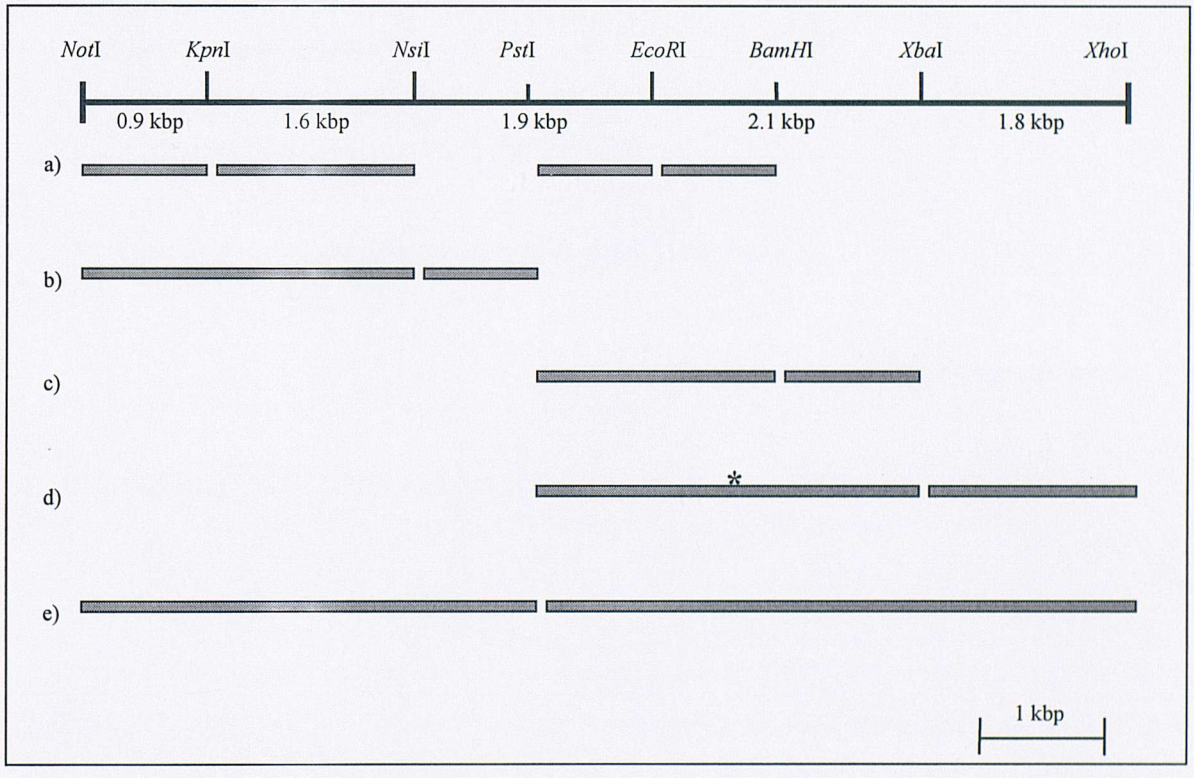


Figure 5.5: Schematic presentation of the first approach followed for the assembly of the full-length CYSDV RNA2 genomic segment. The thick horizontal line represents the RNA2genomic fragment. The vertical lines show the approximate location of the restriction sites

flanking each PCR generated fragment as well as the two restriction sites used that were internal to the PCR amplified fragments, represented as shaded boxes. The fragment that was associated with the cytotoxic effects is labelled with an asterisk. Each individual fragment and each ligation product were inserted into the vector.

The main difference between the two methods presented here is that in the second assembly approach the fragment resulting from step c), Figure 5.5, is cloned in a stepwise fashion to the partially obtained RNA2 cDNA sequence, *i.e.* steps b), c) and d), Figure 5.6. It is divided in three smaller units that are gradually incorporated to the reconstructed genome. This approach overcame the severe cytotoxic effects, though milder cytotoxic effects were still manifested in the form of slow colony growth and limited uptake of the transformed material from cells.

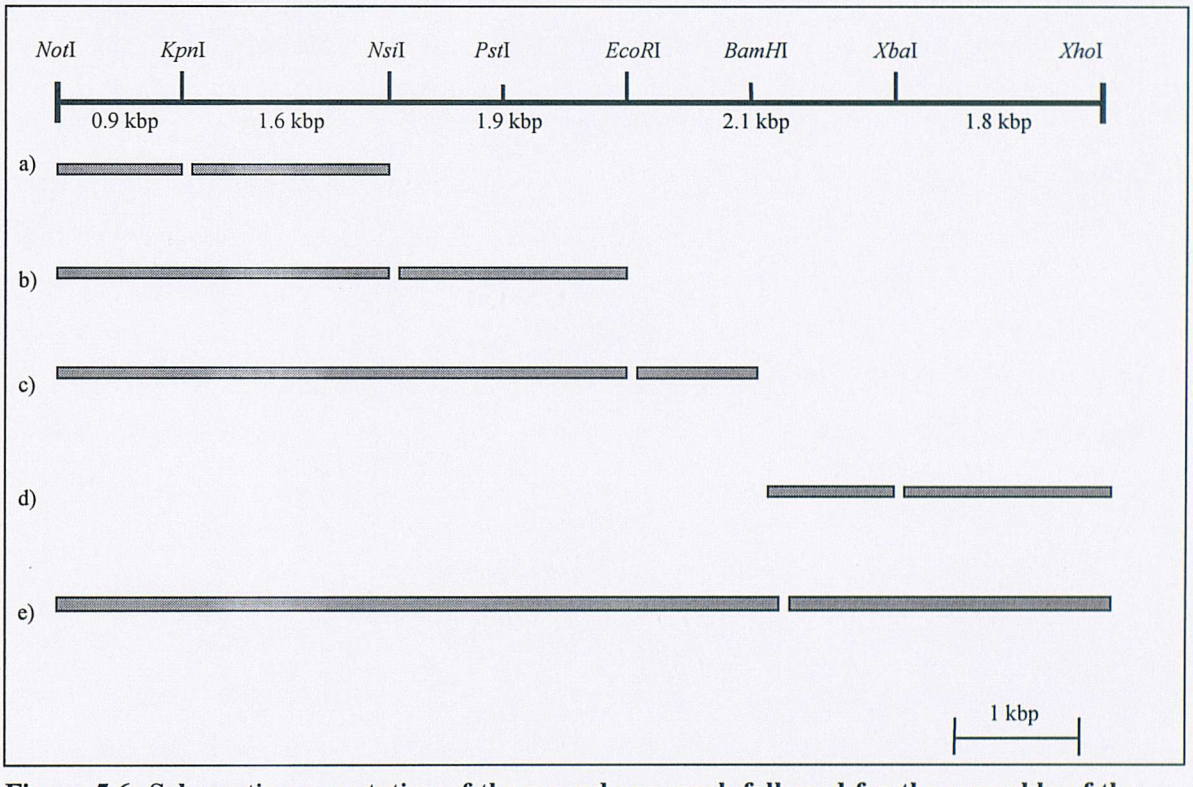


Figure 5.6: Schematic presentation of the second approach followed for the assembly of the full-length CYSDV RNA2 genomic segment. The thick horizontal line represents the RNA2 genomic fragment. The vertical lines show the approximate location of the restriction sites flanking each PCR generated fragment as well as the two restriction sites used that were internal

to the PCR amplified fragments, represented as shaded boxes. Each individual fragment and each ligation product were inserted into the vector.

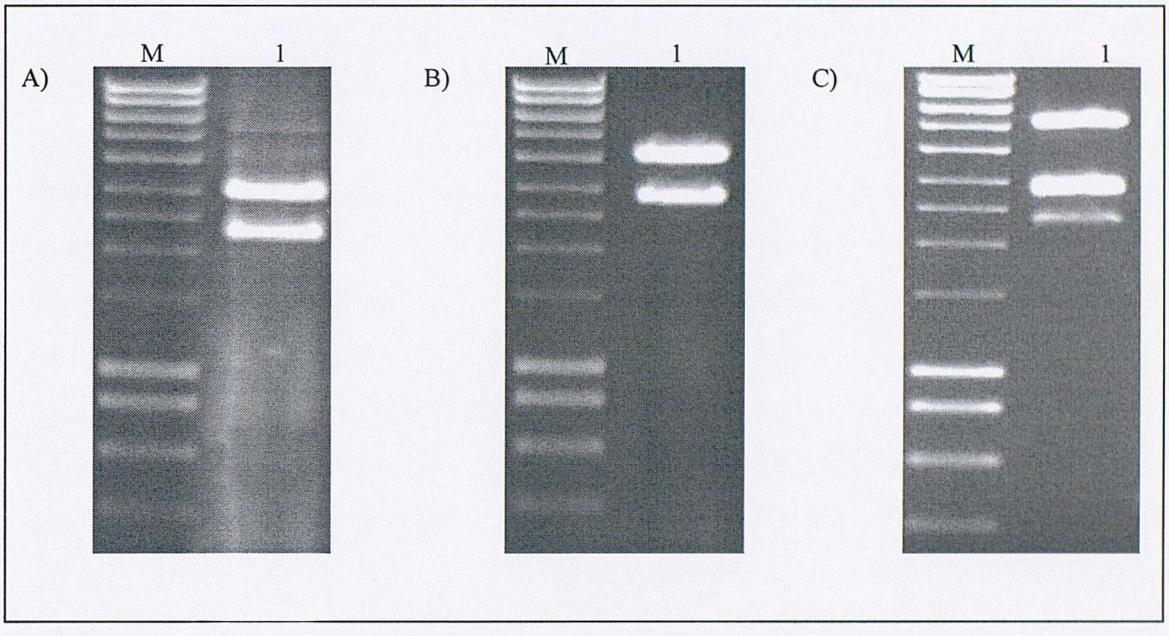


Figure 5.7: The cloned CYSDV RNA2 genomic fragments resulting from the second assembly approach. In lanes M is the 1 kb molecular marker ladder. All lanes 1 contain the plasmid with a weight of 3 kbp. All steps refer to the ones shown schematically in Figure 5.6. In A), lane 1 is the 2.5 kbp *Not*I – *Nsi*I fragment produced in step a). In B), lane 1 is the 4.4 kbp *Not*I – *Eco*RI fragment produced in step b). In C), lane 1 is the 5.5 kbp, NotI – BamHI fragment produced in step c). The lowest molecular weight band is the circular plasmid containing the fragment.

Future prospects

The effect of full length cDNA clone toxicity has been observed and studied to some extent in other viruses, *e.g.* CT*V* (Satyanaryana *et al.* 2003). An alternative method used in the example given above was the creation of a +1 frameshift mutation in the regions responsible for the cellular toxicity. The complete full-length cDNA copy of CYSDV RNA2 can be used to infect plant protoplasts with RNA1, which is responsible for the replication of both genomic components. The production of a stably transfected plant cell line can be used to study the pathogenicity and infectivity of the virus on a molecular level in detail.

5.3 PEV1, UGT gene expression assay

5.3.1 The PEV1 dsRNA virus

Phytophthora endornavirus 1 (PEV1) is a dsRNA endornavirus found in a *Phytophthora* isolate (P441) of an as yet uncharacterised *Phytophthora spp.* member, from Douglas fir. Unlike most other endornaviruses described, which infect plants, PEV1 infects *Phytophthora spp.*, a member of *Stamenopila*. The entire genomic sequence of PEV1 has been published (Hacker *et al.* 2005; GenBank accession number NC007069). The PEV*1* genome is 13883 bp long and contains the characteristic endornavirus motifs, *i.e.* helicase-like and RdRp-like. In addition to these two motifs, PEV1 contains a UDP glycosyltransferase gene (UGT gene), the first such gene to be found in a dsRNA virus, shown in Figure 5.8. A similar gene has since been described in another dsRNA virus, *Cryphonectria hypovirus 4* (Linder-Basso *et al.* 2005). Furthermore, when the sequences of the endornaviruses *Oryza sativa endornavirus* (OSV*)* and *Oryza rufipogon endornavirus* (ORV*)* were inspected, similar UGT-like motifs were found in a similar position in relation to the polymerase and helicase regions (Hacker *et al.*, 2005). It should be noted at this point that the type member of the *Endornaviridae Vicia faba endornavirus* (VFV) does not have a UGT gene. It is possible that the gene was lost during virus-host co-evolution as a result of the virus losing the necessity for the function and thus the presence in its genome.

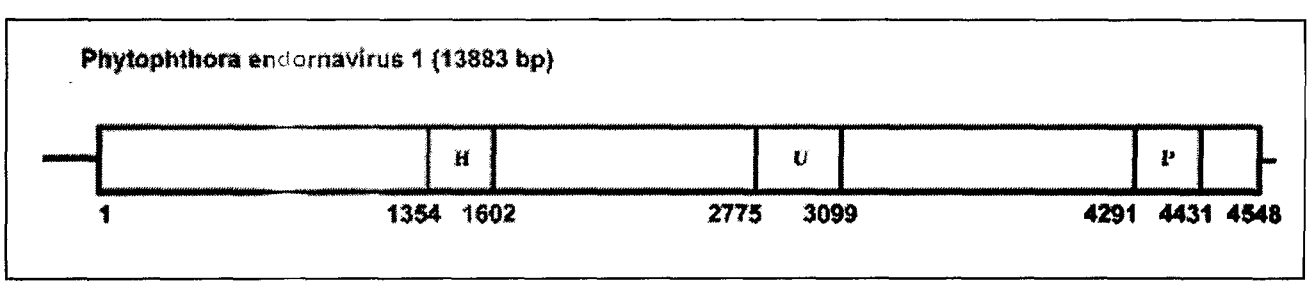


Figure 5.8: Schematic representation of the PEV1 genome. The domains highlighted include helicase-like (H), UDP glycosyltransferase (U) and RdRp-like (P) regions. The polyprotein shown as a rectangle has been started at the second methionine residue (adapted from Hacker *et al.*, 2005).

UGT genes

The function of the glycosyltransferase enzymes is to attach glucose molecules to a sterol backbone. This attachment of a glycosyl residue alters the physical properties of the acceptor lipid molecule. The knowledge base for sterol glycosides is rather limited compared with free sterols and sterol esters, where their synthesis, transport and functions have been extensively studied in animals (Duncan *et al.* 1998), plants (Hartmann 1998) and yeast (Leber *et al.* 1998). The isolation and characterisation of sterol glycosyltransferase genes has been achieved in some eukaryotic organisms. Thus, UGT genes have been found, successfully cloned and had their activity assayed in *S. cerevisiae*, *Pichia pastoris*, *Candida albicans* and *Dictiostelium discoideum* (Warnecke *et al.* 1999). However, there have not been any detailed studies regarding their potential expression and functions within known metabolic pathways in those organisms.

UGT genes and viruses

Glycosyltransferase genes have been found in several DNA virus families, *e.g.* in *Baculoviridae*. These are insect viruses with a circular, supercoiled dsDNA genome, which often contains an ecdysteroid UDP glycosyltransferase (UGT) gene. UGT genes are generally considered to have been acquired from their host at some time during virus-host co-evolution (Markine-Goriaynoff *et al.* 2004). The expression of the baculoviral UGT gene transfers glucose to an ecdysteroid insect moulting hormone, altering the host metabolism and allowing the virus to block moulting and pupation of virus-infected insect larvae.

The expression of UGT genes (α- and β- glycosyltransferases) in bacteriophages allows for their DNA to be glycosylated, thus rendering it resistant to host restriction endonucleases (Sommer *et al.* 2004). Temperate bacteriophages infecting pathogenic bacteria confer additional virulence factors to their host. For example, some bacteriophages modify the host-cell serotype through expression of virally encoded glycosyltransferases during lysogenic infection. This phenomenon may have resulted in the emergence of pandemic or the maintenance of some endemic bacterial diseases.

PEV1 UGT

Cellular sterol UGT genes found in two plant-pathogenic fungi, *Colletotrichum gleosporioides* and *Magnaporthe grisea*, were required for their pathogenicity (Kim *et al* 2002 and Sweigard *et al.* 1998 respectively). Hence the expression of the PEV1 UGT gene might modulate the pathogenicity of its *Phytophthora* host towards Douglas fir through gene silencing or other mechanisms. Homology dependent gene silencing of a pathogenicity gene in *P. infestans* severely reduced its virulence (Latinjhouwers *et al.* 2004). The PEV1 UGT gene could be involved in functions similar to the above.

However, before the PEV1 UGT gene can be studied further, it should be shown that it is active and being expressed by the virus. Thus, two genomic areas of PEV1 were RT-PCR amplified, one containing the entire UGT ORF and second, smaller one, avoiding regions of homology to other known UGT genes. The first fragment was named 1'2' (produced using primers 1' and 2', 1443 bp long) and accordingly the second one was named 3'4' (762 bp long). The entire gene (1'2') was used in an *in vitro* sterol glucosyltransferase time course assay in order to detect enzymatic activity, while the truncated version of the gene (3'4') will be used to raise antibodies, specific against the PEV1 UGT that can verify the results obtained from the assay.

5.3.2 PEV1, UGT gene cloning and expression

The PEV1 dsRNA virus was extracted from fungal tissue after a series of LiCl and phenol:chloroform extractions. This material, shown in Figure 5.9, was used as the template in the subsequent RT-PCR reactions.

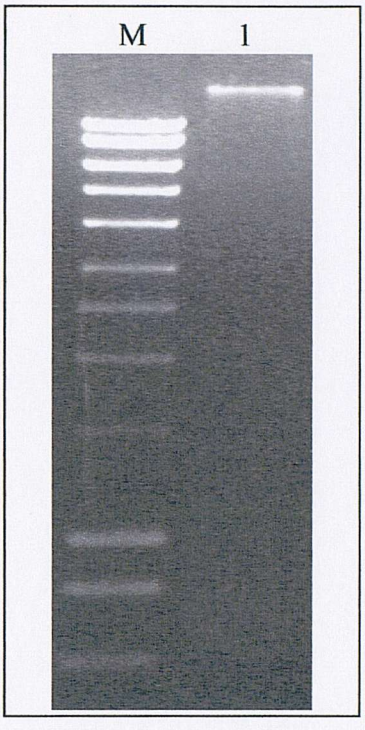


Figure 5.9: The extracted PEV1 dsRNA. In lane M is the 1 kb molecular marker ladder; in lane 1 is the extracted PEV1 dsRNA following DNaseI and S1 nuclease treatment. This material was used as the RT-PCR template.

Two areas of the PEV1 were RT-PCR amplified, as shown in Figure 5.10. The PCR amplification with the designed primers permitted the addition of restriction nuclease sites that were used in the direct cloning of the fragment into a pET28b+ expression vector. This allowed the addition of a 6x His tag that was used in the subsequent protein purification steps, via an ion exchange chromatography column. The ligated PCR products were transformed into BL21 (DE3) pLysS competent cells. The obtained 1'2' and 3'4' clones, shown in Figure 5.10 C), were sequenced in both the reverse and forward orientations in order to confirm the absence of polymerase induced point mutations and the insertion of the 6x His tag in the correct frame.

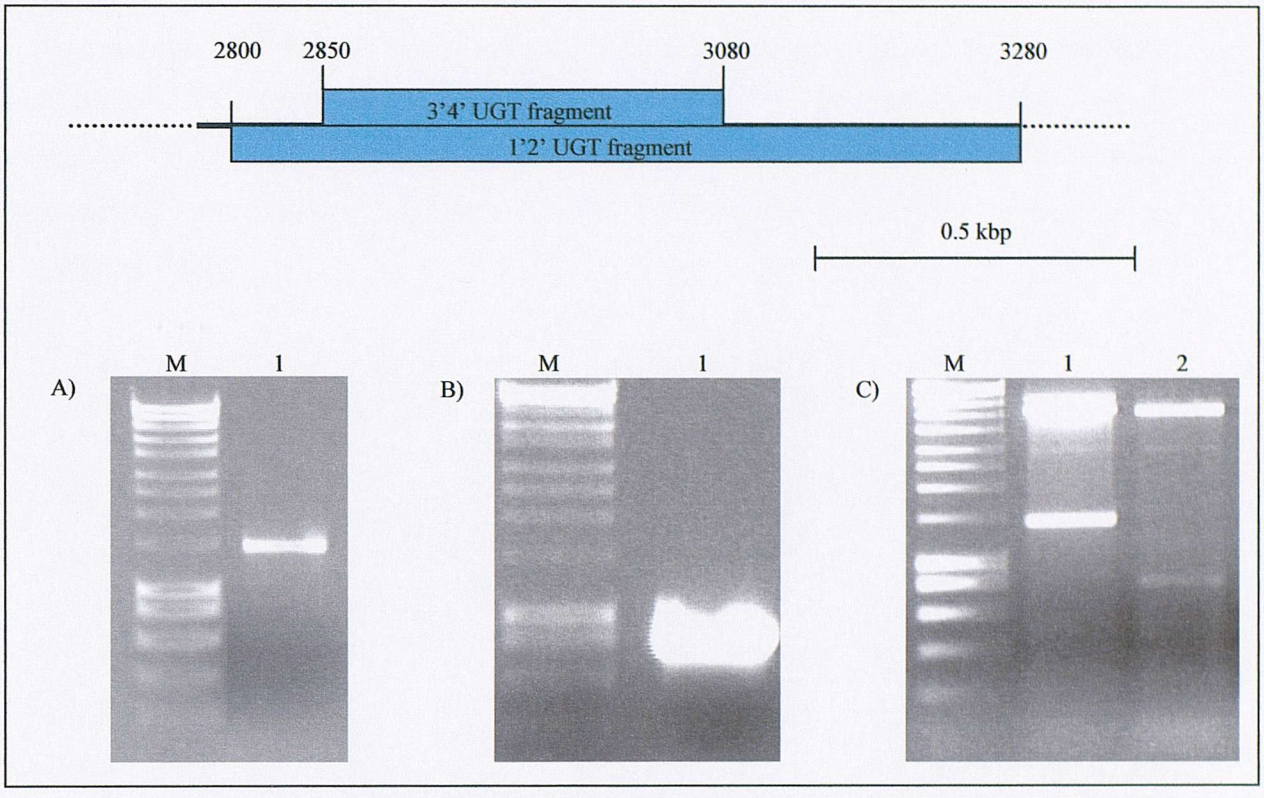


Figure 5.10: Schematic representation of the PEV1 UGT 1'2' and 3'4' amplified genomic regions; the resulting PCR products and clones. In the schematic representation of the PEV1 genomic area of interest, fragments 1'2' and 3'4' are shown as shaded boxes. The numbers indicate the amino acid number on the PEV1 polypeptide. Lanes M contain the molecular marker ladder. In A) lane 1 is the 1.4 kbp 1'2' PCR product. In B), lane 1 is the 0.8 kbp 3'4' PCR product. In C) lanes 1 and 2 are the *EcoR*I – *Nde*I digested clones of 1'2' and 3'4' respectively. The band of large molecular weight corresponds to the linearised plasmid, while the bands of lower molecular weight are of the same size as the PCR amplified products.

The constructs were under the direct control of an IPTG inducible promoter (T7 promoter). Cultures were grown from an agar plate containing freshly transformed bacteria and induced with a final concentration of 1 mM IPTG at mid-log phase. The protein expression was followed over a time course of 4 h. Each construct expression was repeated over a range of different temperatures (37°C, 30°C and 20°C) in order to establish the temperature that consistently gave the highest protein production. The samples were subsequently analysed in an SDS-PAGE gel (Figure 5.11). The construct containing the entire gene (1'2', with a predicted molecular weight of 53.31 kDa) expressed best at 30°C (Figure 5.11 A), while the protein product from the construct of

the truncated UGT protein (3'4', with a predicted molecular weight of 27.48 kDa) was insoluble at 37°C and 30°C. It was soluble however at 20°C, though expressing in small amounts (Figure 5.11 B). The decreased solubility of the truncated UGT protein could potentially occur due to the role of excluded protein regions in the correct folding of the protein product.

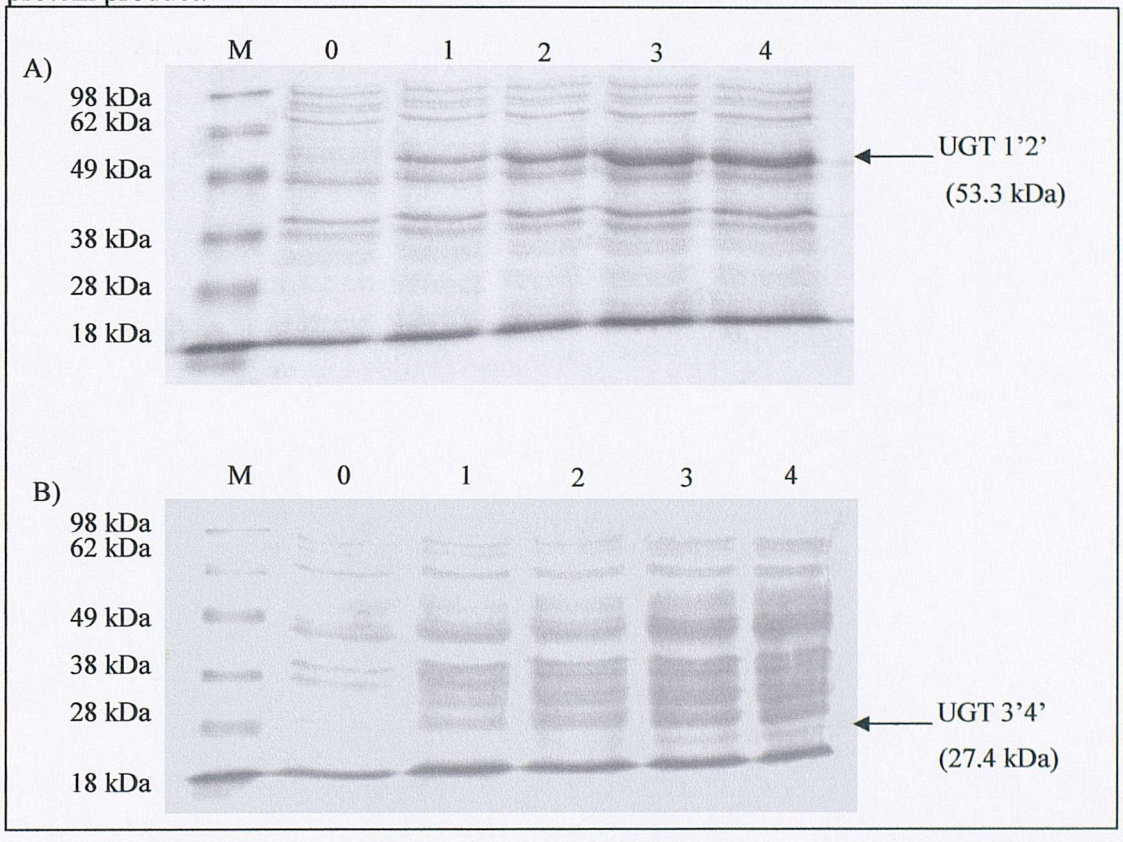


Figure 5.11: Time course protein expression from the 1'2' and 3'4' UGT constructs. The soluble fractions of the samples are run in a 4% SDS-PAGE gel. In lane M is the SeeBlue™ pre-stained protein marker, with the sizes of the marker ladder bands shown. The sample lane numbers correspond to the number of hours after IPTG induction. In lanes 0 are samples from the uninduced cultures. In A) UGT construct 1'2' was expressed at 30°C. In B) UGT construct 3'4' was expressed at 20°C.

Following the expression of the construct the 1'2' protein product was column purified using an ion exchange chromatography column (His Trap HP™) and an automated fraction collector (ÄKTA). The results from the purification step are shown in Figure 5.12 and 5.13. The final concentration of the protein product was adequate to attempt the UGT activity assay. It should be noted that there is an unknown protein product co-

purifying with the 1'2' UGT protein product after the first round of ion exchange column purification (shown in Figure 5.13). This product is consistently in lower amounts than the UGT protein and it did not seem to have an effect in the subsequent activity assay steps. It might be a protein produced by the *E. coli* vector that has a charge similar to the His-tagged 1'2' protein product, hence it co-purifies, or an amount of the actual 1'2' protein product that has been truncated by the action of proteases. Protease inhibitors have been included in the soluble fraction prior to the protein being loaded on the chromatography column.

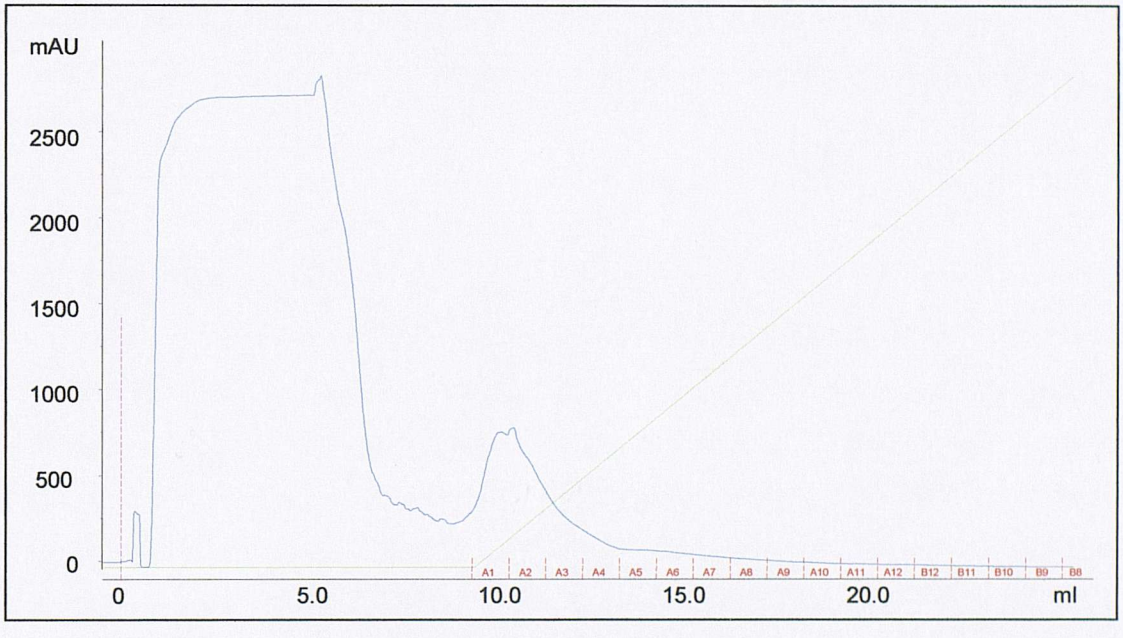


Figure 5.12: Ion exchange column purification of the His-tagged 1'2' *UGT* protein product. In the chart the green line shows the buffer gradient (increasing imidazole concentration and ionic strength). The blue line shows the presence of protein. The initial spike shows the unbound proteins washed off the column. The protein has been eluted in the first two fractions following the increase of ionic concentration in the buffer.

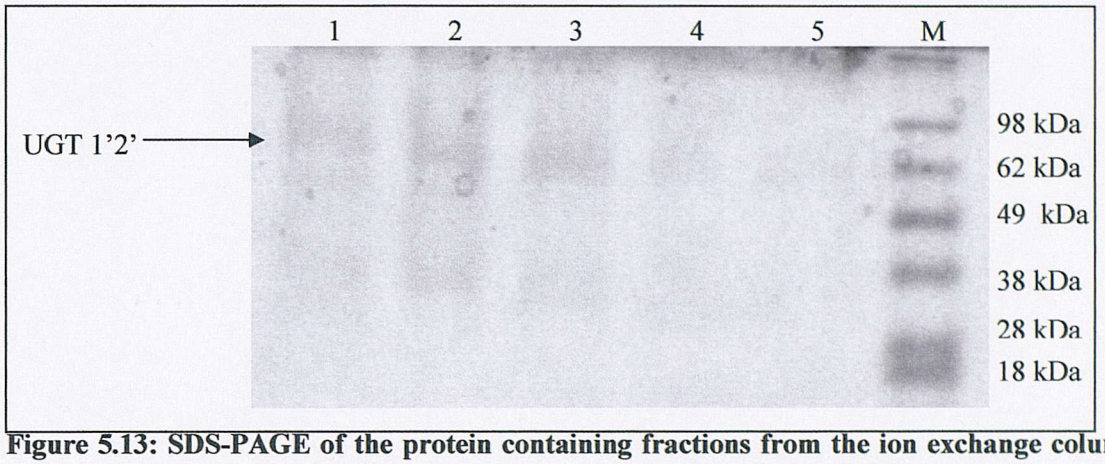


Figure 5.13: SDS-PAGE of the protein containing fractions from the ion exchange column purification. The fractions containing the protein samples were run on a 4% SDS-PAGE gel. In lane M is the SeeBlue™ pre-stained protein marker, with the sizes of the marker ladder bands shown.

5.3.3 UGT activity assay

In order to measure a possible sterol glycosyltransferase activity from the PEV1 UGT protein product, an *in vitro* activity enzyme assay was performed (adapted from Warnecke *et al.* 1999). Radioactively labelled UDP-[U-^{14}C]-glucose was used as the donor and cholesterol as the acceptor, while the column purified PEV1 UGT samples were used directly in the assay. Thin layer chromatography (TLC) was performed on the reaction samples and the resulting radioactivity in the TLC plate was measured with an IP phosphoimaging radio scanner (Figure 5.14). The assay shows the presence of UDP-glucose and radiolabelled cholesterol, thus demonstrating that the PEV1 UGT does have a glycosyltransferase enzymic activity. The insoluble fraction from an induced bacterial culture, the soluble fraction from a culture containing an uninduced plasmid and the soluble fraction of the *E. coli* host cell line without the plasmid were used as the negative controls in the TLC assays.

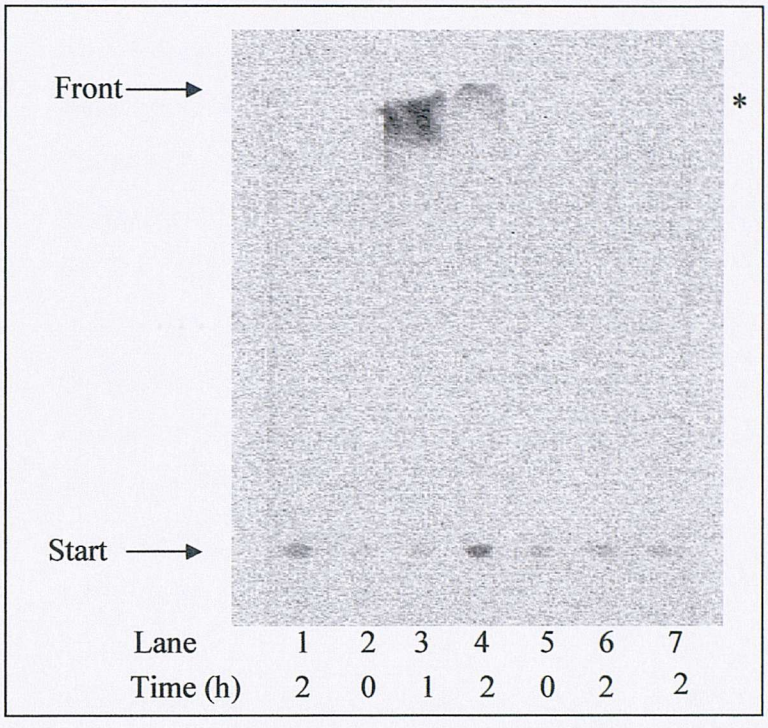


Figure 5.14: TLC enzyme activity assay for the PEV1 UGT protein product. Lane 1 contains the soluble fraction of an uninduced, plasmid containing culture. Lanes 2-5 contain the soluble fraction of induced, plasmid containing cultures. Lane 6 contains the soluble fraction of uninduced, culture, without the plasmid. Lane 7 contains the insoluble fraction of the induced, plasmid containing culture.

Future prospects

The actual substrate of the PEV1 UGT gene cannot be predicted due to the multitude of potential substrates within the *Phytophthora* host itself, as well as the lack of detailed studies on eukaryotic UGT genes. However, it is likely that the PEV1 UGT gene is expressed and active *in vivo*, at least at some point during the *Phytophthora* cell cycle. As the knowledge base for sterol glycosides is rather limited, it would be interesting to elucidate the action of the PEV1 UGT and compare it to that of related UGT genes. Such studies would need to use dsRNA-free *Phytophtora* cultures along with the PEV1 containing P441 isolate and compare, for example, the regulation and expression patterns of host UGT genes in the presence and absence of PEV1.

5.4 Chapter summary

The existing library of CYAV-derived clones was sequenced and analysed. The sequences obtained, though located on an area of the dsRNA mycovirus that was sequenced previously, allowed for a greater coverage of the genome and resolved the ambiguities present in the original set of data. The data were not sufficient to allow for the firm classification of CYAV within a viral family or genus. However, the sequence similarities can classify CYAV as a totivirus or totivirus-like dsRNA element. The experience gained from assembling the overlapping sequence fragments was directly applied in the sequence assemblies of the ACD- and *P. gigantea*- associated dsRNA elements (presented in chapters 3 and 4 respectively). The current sequence information is not complete but is added into an area of virology, that of dsRNA mycoviruses, where very few are characterised and the knowledge base is currently limited.

The efforts for the full-length assembly of the CYSDV RNA2, were impeded by the toxicity of the resuting sub-genomic fragments, hence different assembly approaches were necessary. The experience gained will be used directly in the design and application of a similar strategy for the creation of a full-length cDNA copy of one of the *P. gigantea* dsRNA elements. Additionally, the complete full-length cDNA copy of CYSDV RNA2, when achieved, can be used to infect plant protoplasts with RNA1, which is responsible for the replication of both genomic components. The production of a stably transfected plant cell line can then be used to study the pathogenicity and infectivity of the virus on a molecular level in detail.

The molecular cloning, expression and activity assay of the PEV1 UGT gene has shown that the sequence coded for by the virus has the potential to encode for a protein product with catalytic (transferase) activity. However, in order to study this protein product in detail, the assay itself needs to be optimised and substrates other than cholesterol, *e.g.* ergosterol, would need to be tested. Furthermore, it has to be shown that the UGT gene is active in the host *in vivo*. Once that activity is firmly established, it can be studied further, analysing the virus-host interactions, the timing of the gene expression and the effects thereof.

CHAPTER SIX

Genome-Wide Mapping of Hypersensitive Sites in *Drosophila melanogaster.*

Abstract

Chromatin fibres of folded, eukaryotic, genomic DNA are characterised by the existence of nuclease hypersensitive sites (HS). These are nucleosome-free chromatin segments adopting an open and accessible three-dimensional conformation. As a result they are more likely to interact with proteins, such as transcription factors, nucleases and nuclear scaffold proteins. HS frequently co-localise with transcriptional elements, enhancers and promoters to name two, thus their detection has been used as an indirect approach for the location of possible target areas where such elements can also be located.

Sites hypersensitive to digestion with DNase I were first characterised in *Drosophila melanogaster* over twenty five years ago (Wu *et al.,* 1979). However, the existing indirect end-labelling technique allows for the detection of HS only in a limited genomic area with shortcomings in the accuracy of HS mapping and no sequence information, preventing the quick creation of an integrated genomic and epigenomic map.

The creation of libraries enriched in clones containing HS is a novel, high-throughput approach, which can lead to the rapid, genome-wide mapping and detection of HS. This will assist in the study and characterisation of any given genomic area in terms of its three-dimensional DNA structure and location of elements such as enhancers and promoters, which play an active role in transcriptional regulation.

D. melanogaster was chosen for the creation of the genomic library of HS as it has a relatively small genome, but is still being a higher organism; its genome is fully sequenced and parts of it are very well characterised.

6.1 INTRODUCTION

6.1.1 General characteristics of the genomic DNA and of Hypersensitive Sites (HS).

6.1.1.1 Morphological features of HS.

The nuclear organisation of the DNA and of HS.

When compared to the relative size of the cells that contain them eukaryotic genomes are very large. As a result, the task of packaging the DNA within the nucleus is formidable. Moreover this packaging procedure must at least accommodate, if not contribute to, a system of regulated gene expression supporting the functioning and development of a multicellular organism, such as a change in morphology (development), differentiation of cell types (specialisation of tissue) and responses to external stimuli (dynamic metabolism and repair mechanisms). The folding of DNA into chromatin fibres limits the accessibility of many transcription factors to *cis*-regulatory elements and is thus functionally significant in the control of transcription, replication, repair and recombination.

A characteristic of the chromatin fibres of genomic DNA (folded) is the existence of nuclease hypersensitive sites (HS). These are segments of chromatin that are nucleosome-free regions and, as a result, adopt an open and accessible three-dimensional conformation. This three-dimensional structure of the HS makes them accessible to interactions with proteins such as transcription factors and the action of nucleases, thus HS are two orders of magnitude more sensitive to nuclease digestion than the bulk of chromatin, hence their name (Svetlova *et al.* 1998). Each HS usually spans a genomic area about 100 bp to 200 bp long, although there are instances where an HS may be much longer than this. For example, in the chicken ovalbumin gene, the size of the HS is estimated to be 1.3 kb long (Bellard *et al.* 1986). The digest of HS by nucleases occurs at the same rate as that of naked DNA, *i.e.* unbound to proteins like transcription factors or nucleosomes that might protect it from the action of endonucleases (Bonifer 1999). HS are a local, structural characteristic of the chromatin fibre and should not be confused with the long segments of chromatin fibre that are associated with potentially active or actively transcribing genes or gene clusters. The latter extend over many kb of genomic

sequence exhibiting a heightened general sensitivity to nucleases of perhaps an order of magnitude when compared to the bulk of chromatin (Stalder *et al.* 1980).

Biological reagents used for the detection of HS.

The method most commonly used for the HS mapping, and the one used in the current study, is the DNase I digestion of isolated nuclei. DNase I has only a very slight nucleotide preference with a minor effect (if any) on the large scale genomic digests. Additionally, its cuts do not differentiate between DNA sequences, in three dimensional or linear chromatin fibres, and thus provide a reliable way for the detection of HS (Wu *et al.* 1979). Although DNase I has been used extensively for the mapping of HS, other nucleases can also be used for the selective cleavage of HS. Mung bean and S1 nucleases, for example, have been used for the detection of HS in the vicinity of the *Drosophila melanogaster* hsp23, hsp26, hsp28 genes and the chicken alpha-globin gene (Siegfried *et al.*, 1986; Selleck *et al.* 1984; Targa *et al.* 1994 respectively and Drew 1984). In the case of restriction enzymes, the result will be influenced by the presence or absence of a recognition site as well as the DNA base bias on specific areas of the genome. The latter could be seen in the mammalian genome, which is heterogeneous with respect to base composition. In particular, in humans the GC content of large genomic DNA segments (>100 kb) ranges from 35% to 65%, although even these areas, termed isochores, are not strictly compositionally uniform (Nekrutenko and Li 2000). Nevertheless, within the cell nucleus, the recognition sequences within the HS will be cut immediately, while the sites that are not cut within the first few minutes of digestion will be resistant to the presence of restriction enzyme for long periods - even though these sites can be cut in naked DNA (Jack *et al.* 1991).

6.1.1.2 Biology of HS.

HS position and genomic distribution.

There are various hypotheses regarding HS and in particular their location and distribution on the genome. Approximately 1% of the genome of any organism can be present in the form of HS at any one time. This was estimated, though, using solely the

globin chromatin domain (Stalder *et al.* 1980, Gross and Garrard 1988). A more indicative estimate could be calculated perhaps by the parallel investigation of a number of chromatin domains that have since been characterised. HS are not restricted only to a single specific genomic area; areas of hypersensitivity have been found across conserved non-coding sequences of the mouse genome (Loots *et al.* 2000 and Hardison 2000). It has also been shown that HS are often found on or near CpG islands (Mucha, *et al.* 2000) and on chromosomal recombinational 'hotspots' (Svetlova *et al.* 1998). However, the nature of the information regarding HS is fragmented. This is because they are usually studied as part of research focusing on a particular gene, *e.g.* the human interleukin-5 gene (Valentine and Sewell 1998), on a genomic locus, *e.g.* the 44D Drosophila larval cuticle gene locus (Eissenberg *et al.* 1984), or on a locus and its developmental regulation, *e.g.* the human beta-globin locus control region (Philipsen *et al.* 1990, Kim and Murray 2001; other examples are shown in Figure 6.1). Thus the recorded distribution of characterised HS areas over the whole genome is usually limited to small genomic regions with no or very limited relation to each other.

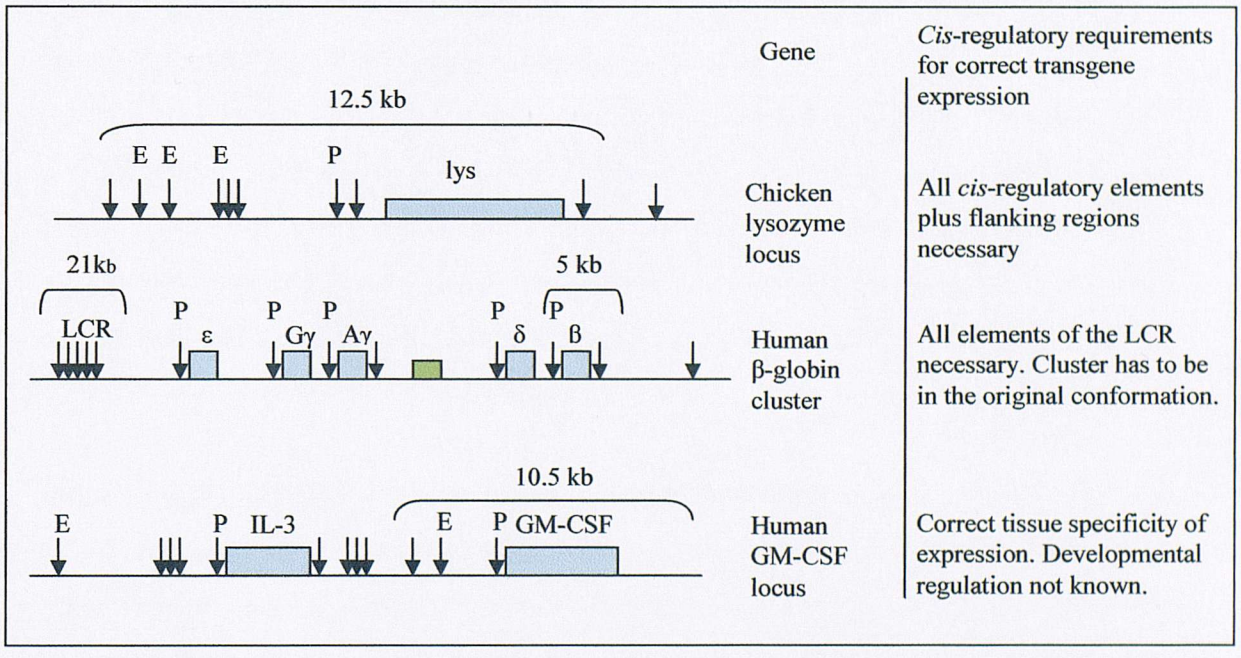


Figure 6.1: The organisation of coding regions with their *cis*-regulatory elements for three eukaryotic gene loci is schematically presented (adapted from Bonifer 2000). Each one shows that different *cis*-regulatory requirements are needed for the correct expression of the locus (also necessary for developmentally correct, position independent transgene expression in mice). The

individual elements are not drawn to scale. The shaded regions correspond to genes- lighter in the case of pseudogenes- with their name indicated above them. Most HS (indicated as vertical arrows) coincide with the positions of regulatory elements like locus control regions (LCR), enhancers (E) and promoters (P). Generating a library of HS could provide quick location (and sequence) information for proximal and distal sites to the genes of interest within the locus, with a high probability of them being involved in the regulation of gene expression.

Constitutive HS.

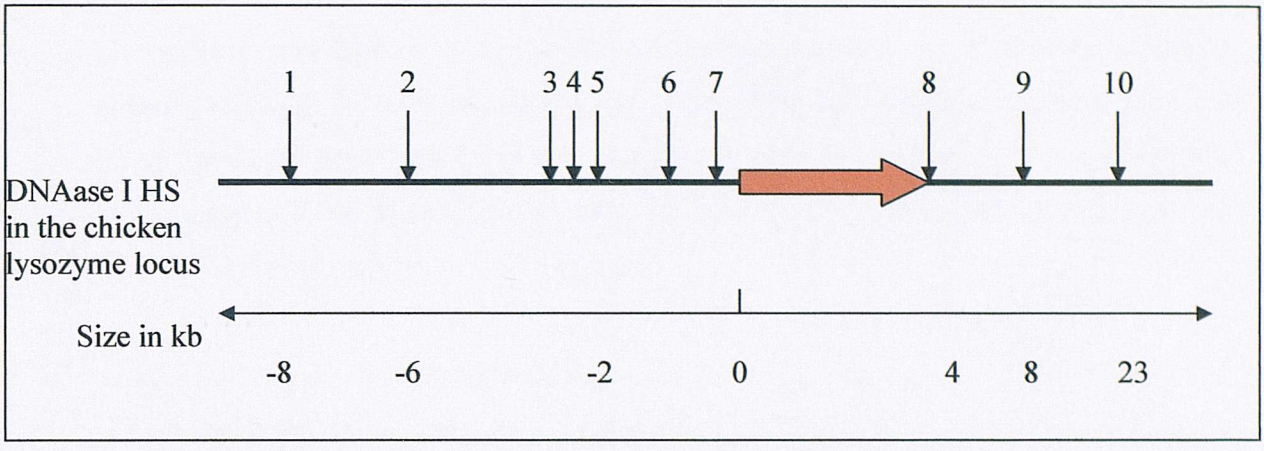
The presence of HS can be constitutive throughout development, as in the case of the HS in the beta-globin gene, where the continuous presence (associated with insulator elements) allowed for successful cross-species comparison between corresponding areas of the human and mouse genomes (Farrell *et al.* 2002). Two constitutive HS have been found in the case of the human 4F2 cell surface antigen gene (4F2HC). The gene itself belongs to the family of inducible cell surface molecules involved in T-cell lymphocyte activation and growth. One HS is in the 5' promoter region of the gene, while the other is in the first intron; the presence of these two HS is constitutive in both 4F2HC expressing and non-expressing cells (Gottesdiener *et al.* 1988). Constitutive HS are usually associated with the low, basal transcriptional levels of nearby genes, such as housekeeping genes. This is the case for the hamster dihydrofolate reductase gene (Azizkhan *et al.* 1986 and Pemov *et al.* 1995), as well as for the *D. melanogaster* histone genes (Samal *et al.* 1981 and Worcel 1981). Constitutive HS have also been located on the promoter regions of genes poised for transcriptional activation, although the genes are not actively transcribed, *e.g.* the heat-shock inducible genes of *D. melanogaster* (Dubrovsky *et al.* 1994 and Lowenhaupt *et al.* 1983).

Tissue-specific, inducible and developmentally regulated HS.

HS are a dynamic characteristic of eukaryotic genomes, reflecting changes in local and/or global gene regulation, and thus they can be expressed in a temporal fashion that can be tissue-specific, inducible or developmentally regulated. An example of tissue-specific HS is that of the mouse acetylcholine receptor delta-subunit gene, where all the HS of the locus, except for the two located 5' distally to the transcriptional start site of the gene, are muscle specific (Crowder and Merlie 1986). The bone-specific osteocalcin gene locus is

also characterised by the absence of hypersensitivity in non-expressing cells, providing evidence for a functional relationship between the presence of HS and the active osteocalcin gene transcription (Montecino *et al.* 1994). Moreover, the HS intensity for the particular gene is directly related to transcriptional intensity, and since the transcription of the gene is inducible and reversible, so is the presence and intensity of the HS. Another such example is the HS located at the 5' upstream sequence of the murine alpha1 collagen (Colla1) gene which is tissue-specific, *i.e.* only present in the collagen producing cells, but not in non-producing cells, indicating that it is associated with transcription of the gene and may function in its regulation (Salimi-Tari *et al.* 1997).

The HS are usually studied as part of the chromatin changes in the development of a tissue and are related to different stages of its differentiation. Thus, in the case of the chromatin structure of the IL-4 locus, HS change (appear or disappear) during CD4 to Th1 cell type differentiation (Lee *et al.* 2001). Another example of that can be seen in the changes in HS locations in the chicken lysozyme locus. The HS here are considered to be the direct result of chromatin reorganisation (termed maturation) and they are linked to the stepwise gene activation during myeloid differentiation, before transcription factors bond onto active elements of the DNA and the actual gene expression takes place. In this particular locus the HS are co-localising with the active *cis*-regulatory elements used each time (Kontaraki *et al.* 2000), summarised in Figure 6.2. A third such example is the case of the mouse M-lysozyme gene, where myeloid and differentiation specific HS were mapped (Mollers *et al.* 1992).



Tissue or cell type	HS number										Functional state of the gene	Transcription
	1,	2,	3,	4,	5,	6,	7,	8,	9,	10		
Mature oviduct	+	+		+	+		+	+	+	+	Active	Yes
Macrophage	+	+	+			+	+	+		+	Constitutive	Yes
Liver, kidney, brain, embryo	+			+				+	+	+	Inactive	No
Erythrocyte								+			Dormant	No

Figure 6.2: Biology of the chicken lysozyme locus in respect to its HS. The gene is indicated by the shaded arrow; the ten HS by the ten numbered vertical arrows. Some HS correspond to *cis*-regulatory elements, e.g. numbers 2, 3 and 7 designate the position of transcriptional enhancers (also shown in Figure 1.1) while number 4 is a transcriptional silencer. The locus is not drawn to scale. In the table, a plus sign means that the HS is present, while a blank means that the HS is absent. As can be seen by this diagram, HS in a particular locus can be constitutive (HS 8) or temporal and their presence/absence can be influenced by tissue specificity, as well as developmental stage. It can be assumed thus that changes in HS patterns reflect changes in the underlying mechanism for the regulation of gene expression (Gross and Garrard, 1988).

The changes in local chromatin structure are also reflected upon the induction of genes, as they become evident by the mapping of HS in the vicinity of such sequences. The inducible presence and formation of HS, flanking the human interleukin-5 (IL-5) gene (Valentine and Sewell 1998), is one such example. The IL-5 gene production is one of the necessary factors contributing to eosinophilia associated with allergic conditions and parasitic infection. The HS formation was induced by the same stimuli that induced IL-5 expression, suggesting that the hypersensitive region is involved in the transcription of the gene or is affected by it. Similarly, the mitochondrial manganese superoxide dismutase (Mn-SOD) promoter contains five HS that appear only when the gene is active.

The gene is induced by a variety of mediators, including polysaccharides, tumour necrosis factor-alpha and interleukin-1 (IL-1). The latter are in turn produced or activated by the increased concentration of damaging superoxide radicals in the cell environment, being generated by aerobic metabolism, usually as a consequence of inflammatory disease (Kuo *et al.* 1999).

Clusters of HS and the synergistic effects on loci.

HS can be found individually, as well as clustered in areas of hypersensitivity (as shown in Figure 1.1). For example, there are eight HS across the T-cell receptor alpha (TCR-alpha) locus control region (LCR). The TCR-alpha gene is limited to T-cells and its locus exists in different chromatin configurations, in both expressing and non-expressing cell types. The removal of certain HS, individually or in a combinatorial fashion from the LCR, revealed that discrete changes in chromatin structure have definite and differing effects (depending on the DNA sequence removed) both on transcription and gene expression. Thus, a theory of a synergistic function between the HS areas (and the elements they contain) and transcriptional factors, for influencing the chromatin structure and subsequently gene expression has been put forward (Ortiz *et al.* 1997). However, more recent data suggest that the relationship may not be as simple as had been originally thought. For example, this theory was investigated by *in vivo* studies of individual HS, conducted in the human beta globin gene LCR, where it is shown that HS exert a co-operative effect in generating a functional LCR with an open chromatin conformation (Leach *et al.* 2001 and Li *et al.* 1998). A study of the murine beta globin gene LCR, similar in structure and length to the human one, showed that although the HS exert a synergistic effect on the locus, their formation is autonomous, thus the deletion of individual HS does not prevent the formation of the others (Bender *et al.* 2000). The complications in assigning roles to HS are a result of the nature of hypersensitivity itself, as there is a broad range of factors responsible for HS creation, *e.g.* the interaction of DNA with transcription factors prior to the active transcription of a coding sequence, the association of *cis*- elements with matrix attachment factors. This variety (and the variety of interactions between the factors themselves) is reflected directly in their appearance,

maintenance, propagation, and on the effects they exert on chromatin structure and gene expression.

It should be added at this point, that a change in gene expression is not necessarily followed and/or preceded by the formation of HS. This was demonstrated in the characterisation of the chromatin domain containing the rat phosphoenolpyruvate carboxykinase gene in the kidney, where gene expression did not result in any alteration of the HS pattern (Cissell and Chalkley 1999). It should also be made clear that the gene transcription is not reliant on HS presence or formation. However, as HS commonly contain *cis*-regulatory elements, their temporally incorrect influences might not greatly affect gene transcription, but they could prove detrimental on the correct regulation of gene expression in development (Bonifer 2000).

6.1.1.3 Functional sequences associated with HS.

HS and transcriptional elements.

HS are a characteristic of the structural formation of the DNA. If the genome was folded in a random fashion they could potentially be found stochastically placed anywhere along the genome. However, in the majority of HS that have been studied in various organisms, their sequences are either directly involved in functions such as transcription, replication, chromosomal segregation and recombination or are indirectly involved in the regulation of the above. HS are thus often found in areas close to the 3' or 5' end of genes, within introns and regions that coincide, by and large, with the location of transcription factor binding sites (Leibovitch *et al.* 2002).

Promoters and enhancers.

A typical example of the commonly observed co-localisation of HS on a promoter sequence is that of the major chicken vitellogenin gene (VTGII), where two HS are located 5' upstream of the gene coding region, corresponding to the proximal promoter sequence. These HS are present in cells of the organism expressing the genes and are absent in non-expressing cells, with the exception of a small cellular sub-population

consisting of differentiated oviduct cells that have oestrogen receptors (Burch and Weintraub 1983). Similarly, in a study of the murine recombination activating gene 2 (i.e. one of the genes involved in the V(D)J recombination machinery during lymphocyte ontogeny) 25 kb of genomic sequence flanking the gene were analysed for areas of hypersensitivity. Three, lymphoid cell-specific HS were located at the 5' region: one co-localised with the proximal promoter of the transcriptional unit and the other two contained within the sequence of two enhancer elements (Wei *et al.* 2002). Enhancers are another *cis*-regulatory element often located at or within HS areas of the genome. The latter case can be seen *D. melanogaster*, where the enhancer region of the ecdysone regulated glue gene Sgs-3 coincides with a HS. The Sgs-3 gene exhibits modifications of its chromatin structure that can be directly related to the changes in gene expression of Sgs-3 during development (Dretzen *et al.* 1993). Another example of HS located on an enhancer comes from the study of the mouse Pro-a2(I) collagen gene, where within the distal enhancer of the gene, the minimal functional region of the *cis*-element coincided with two HS (De Val *et al.* 2002).

The most frequent localisation of HS in the close vicinity of the promoter region is experimentally important in assisting with the location of the particular regulatory sequences. From the genomic data available, promoter organisation seems to be species specific, with little or no homology between promoter sequences across species. Thus, the use of *in silico* sequence comparison tools for their identification might not be as effective as that of the experimental mapping of HS along the genome (Grandieu *et al.* 1997).

Other cis-regulatory elements: silencers, insulators (boundary elements) and locus control regions.

Areas of hypersensitivity do not coincide exclusively with promoter and/or enhancer elements, but also with a wide range of other *cis*-regulatory factors, including silencers, boundary elements and locus control regions, of which some indicative examples will be given below. The regions of DNA sequences that silence the expression of specific transcriptional units in the genome, usually at a specific developmental stage, are termed

silencers. For example, HS have been located in silencer elements in the case of the human immunoglobulin kappa gene, where the silencer is active in non-B cells (Pierce *et al.* 1991), as well as in the silencer and enhancer elements of the corresponding murine genomic locus (Liu *et al.* 2002). In the case of the human cholinergic gene locus, the HS corresponds to a neuronal restrictive silencer accounting for the repression of the cholinergic gene locus in non-neuronal cells (Tanaka *et al.* 1998). Silencers are used in the regulation of the homeotic genes of the *D. melanogaster* bithorax complex, ensuring their segmentally restricted pattern of expression. The iab-7 PRE silencer is located within 0.8 kb of genomic sequence in the particular locus, which also contains a HS (Hagstrom *et al.* 1997).

HS that were mapped in the same locus, *i.e.* the *D. melanogaster* bithorax complex, co-localised with boundary elements or insulators (another type of a *cis*-regulatory element). As euchromatin is organised into domains that act independently of one another, the extent of each domain is thought to be defined and maintained by special DNA sequences termed chromatin insulators or boundary elements. Boundary elements have an enhancer-blocking activity which is orientation- and enhancer-independent. This HS/boundary element co-localisations were observed in the chicken beta-globin domain, where the 5' and 3' boundaries of the locus function as insulators and are characterised by constitutive HS (Saitoh *et al.* 2000). They were also studied in the murine Ad4BP/SF-1 and GCNF loci, where the boundary separating the two loci found in close proximity on the genome co-localises with three strong HS. In this case the chromatin architecture contributes to the distinct expression patterns of the two loci (Ishihara and Morohashi 2005). One of the best studied areas containing boundaries is that of the Abdominal–B (Abd-B) gene in *D. melanogaster*, which is part of the homeotic regulatory mechanism responsible for specifying the identity of the fly parasegments 5 to 14 (PS5 – PS14) which form the posterior half of the organism. The expression patterns are generated and maintained by a complicated network of *cis*-regulatory elements, separated into domains and spanning approximately 300 kb. These domains are defined by boundary elements and the latter coincide with HS (Barges *et al.* 2000 and Mishra *et al.* 2001).

Regulatory elements can also be clustered, rather than individually found, in areas that are called locus control regions (LCR) and are commonly located at some distance from the main transcriptional units, on which they exert their effects. LCR can have stimulatory, inhibitory or more complex effects on genes, and these often occur in connection to certain developmental stages or stimuli. Such an LCR is located 40 – 60 kb away from the active genes of the human (and murine) beta-globin locus and contains five HS. All the proximal regulatory sequences are required for the correct initiation of transcription, while the distal ones are required for the temporally correct and maximal gene expression. The human alpha-globin cluster is organised in a similar manner, as far as the LCR and the HS it contains are concerned (Cao and Moi 2002 and Tolhuis *et al.* 2002). The same arrangement, *i.e.* HS within an LCR, has been observed for the murine LCR that is located between the TCRdelta/TCRalpha gene segments and the ubiquitously expressed Dad1 gene (Ortiz *et al.* 2001).

Matrix attachment regions (MARs)/ Scaffold attachment regions (SARs) and HS.

Matrix Attachment Regions (MARs) and Scaffold Attachment Regions (SARs) are intergenic regions, initially characterised for their ability to interact with the nuclear matrix/scaffold and to facilitate the long-range chromatin remodelling often required for enhancer activity and maximal level of gene transcription. There is also evidence of a structural role for MARs/SARs in dividing eukaryotic chromosomes into topologically looped domains, thus having a potential role of physically defining chromosomal domains (Laemmli *et al.* 1992). MARs have been found co-localising with regulatory elements and with HS, though the functional significance of this last observation remains to be determined. Such is the case of the human CD8 gene, where three strong MARs were closely linked to two tissue-specific DNase I HS clusters (Kieffer *et al.* 2002) and of the upstream region of the chicken α-globin gene domain, where the MAR co-localises with an erythroid specific HS (Valadez-Graham *et al.* 2004). From the available data, the co-localisation of MARs and HS is not as commonly observed as that of HS and other regulatory elements. Genomic areas are to be found that, though they are characterised by a strong hypersensitivity, are completely devoid of MARs. For example, this was observed with the human growth hormone (hGH-N), where the main three LCR functions

for the gene locus mapped on three strong HS, though the entire 87 kb region did not contain any MARs (Shewchuck *et al.* 2001).

6.1.1.4 Mechanisms of establishment of HS.

Theories for the establishment of HS.

The hypersensitivity of a genomic region is often determined by its underlying or flanking DNA sequence, as proteins will recognise and bind to their unique corresponding motifs on the DNA sequence. There are a number of suggestions regarding the mechanisms for the generation and establishment of HS on the genome. These theories can be, by and large, divided into two groups: those suggesting the active generation of HS and those following the passive generation of HS. The passive generation model proposes that HS are simply the result of chromatin remodelling and nucleosome repositioning, leaving regions of the genomic DNA naked and thus exposed to the action of nucleases. This view was adopted, for example, in the study of the hamster cardiac beta-myosin heavy chain (MyHC) gene, where during the tissue-specific transcriptional activation of the gene, the nucleosomes located on the proximal promoter are reorganised, resulting in a HS co-localising to the proximal promoter sequence (Huang and Liew 1998). Similarly, the passive generation of HS was considered in the study of DNA accessibility on damaged chromatin. It was found that a specific protein, Gadd45, interacts with hyperacetylated mononucleosomes, *i.e.* it recognises damaged chromatin, and interacts with the histone core to alter locally the accessibility of the DNA, creating HS (Carrier *et al.* 1999).

The active model for the generation of HS suggests that the DNA sequence itself, *e.g.* the *cis*-regulatory elements it might contain, acts with one or more proteins to reorganise the chromatin on a local level and therefore leads to the creation of HS on the genomic sequence. This seems to be the case in the human beta-globin locus control region. The particular genomic area contains an LCR comprised of multiple HS sites, 200 bp to 400 bp long each, located 5' to the transcriptional units. It was observed that the HS were formed even in the absence of chromatin assembly and in an independent fashion from

one another. In other words, the deletion of different, single HS did not affect the formation of others on the same LCR, even thought the deletion itself could have had an effect on the regulation of transciption of the gene domain. The experimental data indicated that the sensitivity to nucleases in the particular genomic locus is not simply a consequence of nucleosome reorganisation, but also depends on other factors as yet not characterised (Leach *et al.* 2001). This data agrees with the model proposed by Felsenfeld. According to this model, regulatory elements, like the ones contained in the human beta-globin LCR, are structurally prepared by the binding of proteins to the DNA sequence shortly after a cycle of cellular replication. These proteins, as well as the protein-protein and protein-DNA interactions, form and maintain the HS, conferring resistance locally to the presence of repressive chromatin, even when areas of the neighbouring genomic DNA sequence are experimentally deleted (Felsenfeld 1992 and Felsenfeld *et al.* 1996). Similarly, at the *D. melanogaster hsp70* gene, there are strong HS, 5' to the transcriptional unit. Small sequence deletions demonstrated that the presence of the HS is correlated to the presence of specific sequences, thus the HS are actively formed through the binding of proteins to these sequences (Costlow *et al.* 1985).

HS describe a structural characteristic of the genome; as such they can include a wide range of cases that seem to agree either with the one or with the other model of HS generation presented previously. The passive creation model (based on the effects of nucleosome phasing) can potentially explain the formation of HS in genomic areas that do not coincide with known *cis*-regulatory elements. In the same way, for HS that do contain characterised elements, the theory for the active creation of HS due to the interaction of their sequence with a range of transcription factors remains a plausible explanation. An additional factor that complicates the formulation of any further models on the subject is that despite the number of *cis*-regulatory elements that have been characterised, not all of them have been studied in relation to the HS information they might contain.

In certain cases, the establishment of HS was studied in respect to the proteins that might be responsible for it, but evidence seems to be contradictory. The GAGA-binding factor

is essential for the establishment of hypersensitivity at the promoter of the hsp26 gene in *D. melanogaster* (Leibovitch *et al.* 2002 and Farkas *et al.* 2000). However, the formation of HS and the binding of the GAGA factor on a *D. melanogaster* alcohol dehydrogenase minimal promoter constitute two separate and unrelated events *in vivo* (Pile and Cartwright 2000). These findings suggest that a different mechanism for the generation of HS might exist in each of the above cases, even though the same proteins or protein complexes are involved. In addition, differentiating between the two main models presented (passive or active) and other possible ones could be further limited by the existing experimental techniques for the mapping of HS. In summary, it appears that any of several mechanisms may operate to generate hypersensitive sites depending, in part, on the single or multiple binding of *trans*-acting factors.

Maintenance of HS.

The mechanisms for the maintenance and propagation of HS remain still largely unknown, though it has been suggested in specific cases that the maintenance of hypersensitivity is the direct result of protein-DNA interaction. For example, in the case of the rat tyrosine aminotransferase gene (TAT) the presence of glucocorticoids induces the transcriptional activation of the gene. The enhancer element co-localises with a HS and comprises of binding sites for the glucocorticoid receptor and other additional transcription factors, which are only recognised after hormone administration. The presence of the HS is inducible and directly related to the presence and binding of the glucocorticoid receptor (Reik *et al.* 1991). It has also been shown, in a study of the human epsilon-globin gene domain that the maintenance of HS and the regulation of their associated elements, which are promoters in the particular locus, can occur both by distinct and independent mechanisms (Bushel *et al.* 1990). However, experiments on the avian [lambda] light chain gene showed that DNase I HS maintenance in this genomic area is not related to protein-DNA interaction, but is probably related to higher order chromatin structures (Li *et al.* 1997), as was suggested some years earlier by Goodwin (Goodwin *et al.* 1985). Thus it appears that several mechanisms may operate individually or co-operatively in order to maintain HS. The mechanism utilised each time might

depend both on the binding of *trans*-factors as well as on higher order structural domains within the cell nucleus.

More than twenty five years have elapsed since the first HS were described (Wu *et al.* 1979), nevertheless very little is known about the mechanisms that maintain them and even less about the mechanisms that propagate them. It has been reported that for certain sites, for example in the alpha- and beta-globin domains of chicken fibroblast cells, once the HS are formed, they can be propagated for at least twenty generations, even in the absence of an associated function (Groudine and Weintraub 1982). However, there is to date no suggested reliable model explaining this mechanism of propagation. The examples of HS given above, in relation to the different temporal character of their appearance are summarised below in Table 6.1, with the exception of all the *D. melanogaster* HS, which are examined separately in section 1.3.4.3.

Table 6.1: Summary of the HS examples given. The HS are categorised alphabetically according to the organism in which they were characterised. The HS biology and the *cis*-regulatory elements that co-localise with them are also shown, demonstrating the variety of functional elements located on a HS. It should be noted that not all HS within a cluster of HS co-localise with *cis*-regulatory elements. In cases where the function of the *cis*-regulatory element remains to be determined, a questionmark is present.

Organism	Gene/genomic locus	HS biology	Elements co-localising with the HS(s)
Chicken	α-globin locus, δ gene	Cell-type specific	*Trans*-factor recognition
Chicken	β-globin locus	Constitutive	Boundary elements
Chicken	Ig-λ, light chain	Dev. stage specific	S/MAR?
Chicken	Lysozyme	Dev. stage specific	*cis*-acting elements
Chicken	Ovalbumin	Organ specific, inducible	Proximal promoter?
Chicken	Major vitellogenin II gene	Expression specific	Proximal promoter
Hamster	β-myosin heavy chain	Tissue-specific	Proximal promoter
Hamster	Dihydrofolate reductase	Constitutive	Proximal promoter
Human	Ad4BP/SF-1 locus	Constitutive?	MAR, boundary element

Organism	Gene/genomic locus	HS biology	Elements co-localising with the HS(s)
Human	CD8	Tissue-specific	MAR
Human	Cholinergic gene locus	Cell-type specific	Neuronal restrictive silencer
Human	α-globin locus	Cell-type specific	LCR
Human	β-globin locus	Cell-type specific	LCR
Human	hGH-N (growth hormone)	Organ-specific	LCR
Human	Ig-κ	Cell-type specific	Transcriptional silencer
Human	IL-4, IL-5	Dev. stage specific	Enhancer element
Human	Mitochondrial Mn-SOD	Inducible	Promoter
Human	Osteocalcin gene	Tissue-specific	Proximal promoter
Human	4F2 surface antigen	Constitutive	Promoter
Human	TCR-α locus	Cell-type specific	LCR
Mouse	Acetylcholine receptor δ-subunit gene	Tissue-specific	Regulatory locus?
Mouse	β-globin locus	Cell-type specific	LCR
Mouse	α1 collagen	Tissue-specific	Transcriptional regulator
Mouse	Ig-κ	Cell-type specific	Transcriptional silencer
Mouse	M-lysozyme	Tissue-specific	Enhancer?
Mouse	Pro-α2(I) collagen gene	Cell-type specific	Enhancer
Mouse	Recombination activating gene-2	Cell-type specific	Proximal promoter
Mouse	TCRα/TCRδ	Cell-type specific	LCR
Rat	Phosphoenolpyruvate carboxykinase locus	Constitutive, organ specific?	Proximal promoter, LCR
Rat	Tyrosine aminotransferase	inducible	*Trans*-factor elements

6.1.2 Current technique, its limitations and the novel approach for mapping HS.

6.1.2.1 Current technique for HS mapping and its limitations in locating HS.

The main aim of this investigation was to develop and test a method for the rapid mapping and detection of HS. The need for the creation of such a method lies in the

absence of any high-throughput HS mapping protocol, and the fact that the traditional method for HS mapping, the indirect end-labelling, provides only local information, *i.e.* near a target gene. This indirect end-labelling has also limited accuracy in locating the HS and does not provide any sequence information. The current technique of locating HS, described in Figure 6.3, is widely used and contains a number of steps. Firstly, the genomic DNA from isolated nuclei is digested by DNase I- or any other enzyme of choice- under limited amounts of time and enzyme, resulting in a series of genomic DNA samples with an increasing level of digestion. The DNA is then further digested to completion by another restriction enzyme, *e.g. EcoR* I or *BamH* I. A specific DNA fragment is labelled and used as a probe in the Southern blotting of the genomic DNA. The HS increases in intensity as the amount of DNase I used to cleave the genome increases. In the case of more than one HS, the Southern blotting produces a number of fragments with different sizes.

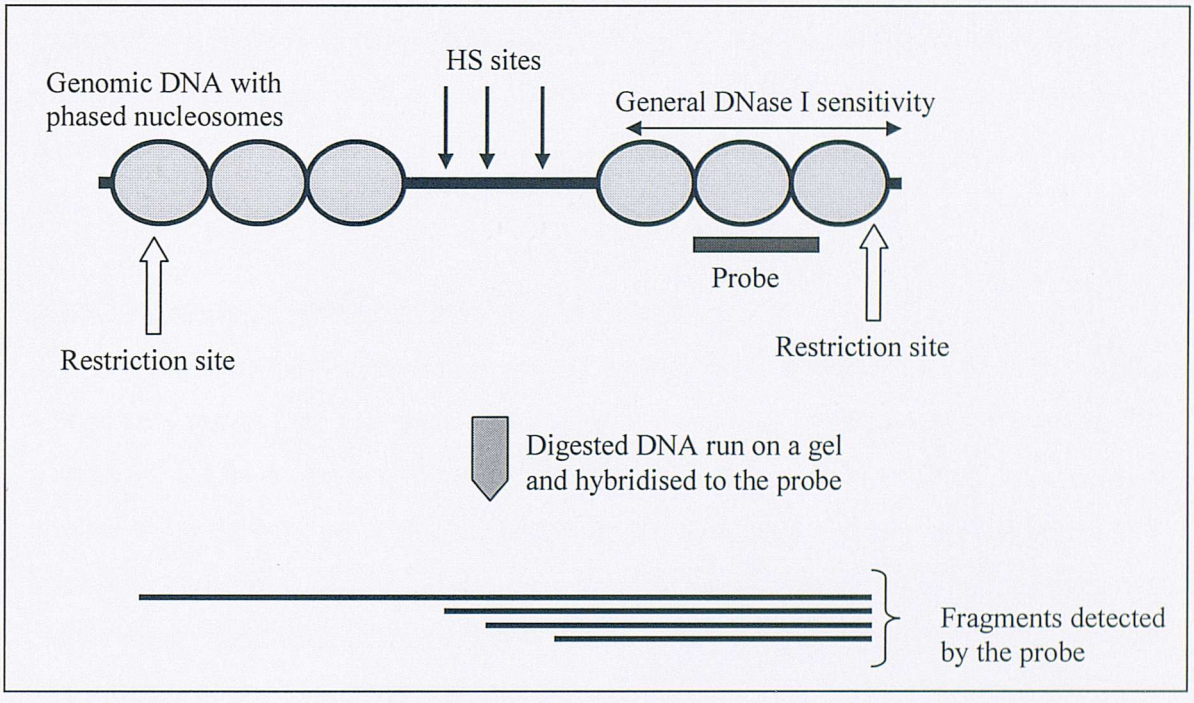


Figure 6.3: End-labelling mapping of HS. A nucleosomal array is disrupted by a region of exposed DNA, which will display hypersensitivity when incubated with limited amounts of a nuclease, *e.g.* DNase I. The restriction sites represent two recognition sites for the enzyme used to digest the genome to completion, *e.g. EcoR* I or *BamH* I. The common position of the probe used

in the Southern blotting step is indicated with the shaded box. The horizontal black lines represent the DNA fragments that would be detected by the probe. The shorter bands, those where the DNA is cleaved between the two restriction sites, will increase in intensity on the Southern blot as the amount of DNase I used to detect the HS increases.

The process described above is time consuming and technically cumbersome. The indirect end-labelling can only show that there exists a HS (or not) near the sequence of interest at an approximate distance from an enzyme restriction site (±50 bp calculated from the electrophoresis of the genomic fragments). It does not provide any information regarding the exact size, and therefore accurate positioning, of the HS or its nucleotide sequence (Gross and Garrard 1988). Furthermore, only one sequence can be studied at any one time and at a certain distance, usually *ca.* 1 kb, from the probe that is being used. In the case of clustered HS, the indirect end-labelling mapping sometimes requires the use of two probes, the second often used from the other end of the restriction fragment (Aasland and Stewart 1999). Additionally, as there is always a certain level of variability inherent in the experiments, it is essential to compare the cleavage patterns within one experiment, thus for each experiment at least three concentrations of enzyme should be used so that hypersensitivity is safely revealed.

6.1.2.2 Advantages of the novel approach

The indirect end-labelling approach can be significantly shortened in terms of time and resources required, by the creation of a library enriched in fragments containing HS (described below in section 6.1.2.3). This approach is practically robust, as it can be applied to all types of cell culture. The library contains information on thousands of HS- not solely on one restricted genomic area and once generated, the fragments with the HS can be easily identified and accurately located on the genome, as they can be sequenced. The creation of the library also ensures that a number of different genomic regions can be studied at once, without the need to repeat the protocol for each individual region. This allows a global analysis of HS patterns and their respective changes. Thus large-scale HS pattern comparisons can be investigated, for example, between different cell growth conditions and different cell types.

HS pattern alterations reflect the underlying causes for their formation, as well as the effects that such changes have in gene regulation. As such they could prove highly diagnostic. The sequence, and thus the accurate HS location on the genome, provides the opportunity to develop algorithms able to analyse the genome and predict new HS *in silico*. This information also allows the further characterisation of the cloned genomic fragments, regardless of whether they contain known *cis*- regulatory elements or other elements such as attachment regions, *e.g.* matrix attachment regions The sequences of HS cloned from the genomic libraries will allow for several questions to be answered, such as those relating to their usual positioning and genomic distribution and will be the first step towards a further, deeper understanding of the changes in HS patterns, the forces that cause them and their effects.

6.1.2.3 Outline of HS library generation

The generation of a library enriched for fragments containing HS has a number of steps, presented schematically in Figure 6.4. The genomic DNA present in cell nuclei is digested with a restriction enzyme under limited amounts of time and enzyme. The use of a frequent 4 bp cutter restriction enzyme, such as *Mbo* I or *Hae* III, ensures cleavage at a large proportion of the HS present in the genome and produces at the same time a 'sticky end' suitable for cloning. *Mbo* I has a four base recognition sequence (5' GATC 3'), thus it cuts on average once every 256 bp on the DNA sequence. As the average size or the HS core is approximately 200 bp, the *Mbo* I should statistically cut within 78% of any predicted number of HS. The use of DNase I is not be suitable for this approach, because the enzyme is non-specific in its DNA digestion and can cleave all four bases in heterogeneous DNA (although in practice it shows a slight sequence preference) leaving a 'blunt end'. However, this characteristic means that the enzyme can not be used for the generation of fragments that can subsequently be cloned into a plasmid.

The genomic DNA that has been cleaved by an enzyme, *e.g. Mbo* I, is ligated to a plasmid vector (pBluescript). The vector is cut with another enzyme that has a compatible overhang to the first one (in this case *BamH* I) and treated with alkaline phosphatase. The

DNA sequence adjacent to the *Mbo* I restriction site is joined to the plasmid DNA. However, this ligation step does not yet result in the formation of a functional recombinant DNA molecule, since the other end of the genomic fragment could be located kilobases away or might have been sheared during the genomic DNA extraction step and will thus be unsuitable for the specific joining to the other *BamH* I site in the linearised plasmid vector. Therefore, in order to generate genomic fragments suitable for cloning the ligation mixture is cut to completion with another restriction enzyme, for example *EcoR* I. This enzyme cuts at a defined site(s) within the plasmid vector, as well as at every restriction site in the genomic DNA fragments. This step creates the necessary conditions for the specific joining of the two complementary ends of the construct that are performed by a second ligation. The HS sequence is not affected as the DNA fragments adjacent to the *Mbo*I site are already ligated to the plasmid vector.

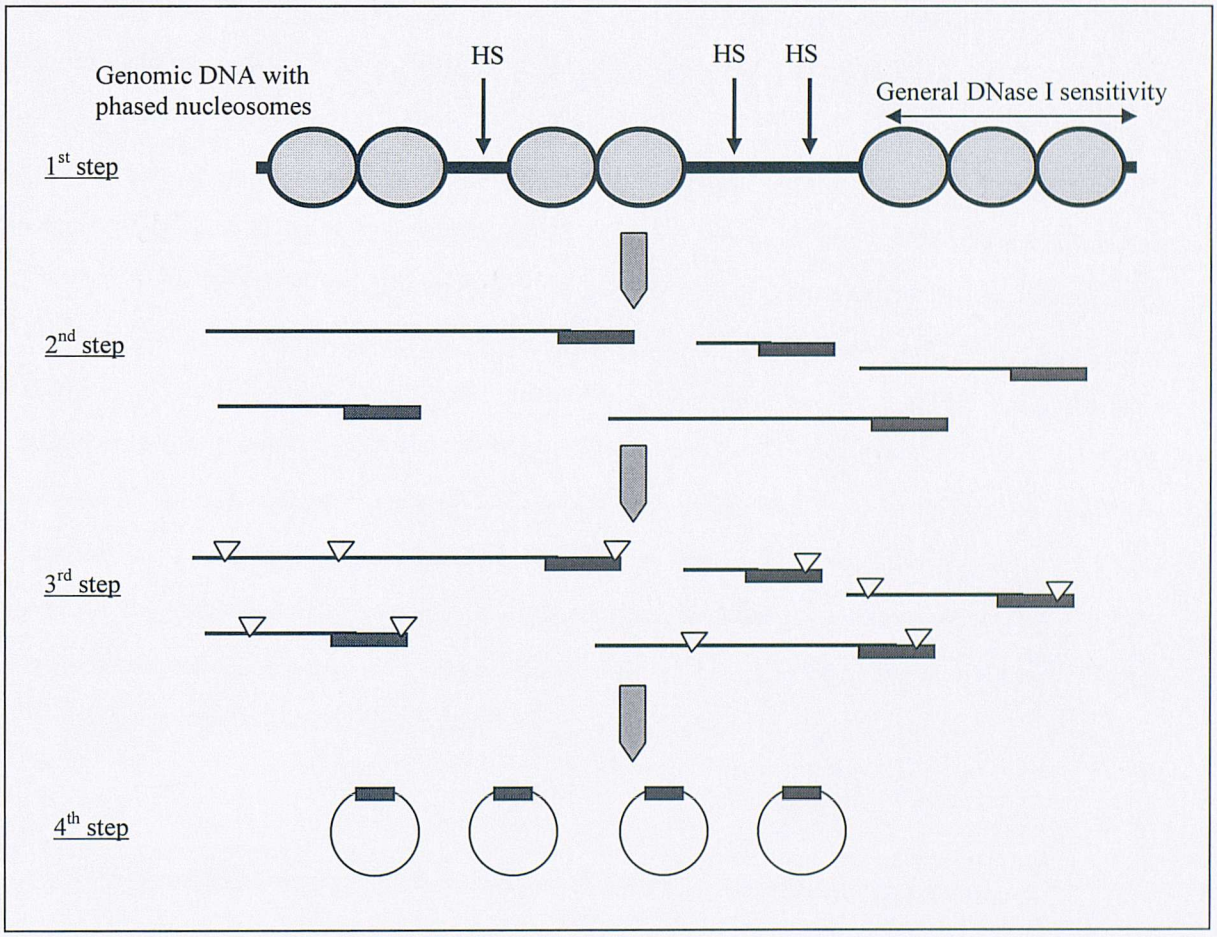


Figure 6.4: Schematic representation of the HS library generation. In the 1st step the genome is digested with limiting amounts of a nuclease, *e.g. Mbo* I, resulting in a wide range of genomic fragments. These fragments are ligated in the second step to a vector, *e.g.* pBluescript, shown as a shaded box. The recombinant DNA molecules are digested to completion in the 3rd step with a second enzyme, *e.g. EcoR* I, that cuts at a single specific site on the vector and at various sites on the genomic DNA fragment. In the 4th step a second ligation takes place resulting in a functional DNA molecule that can be transformed in the chosen competent cells, *e.g.* DH5-α.

As the orientation of the plasmid vector relative to the *Mbo* I fragment during the first ligation is at random, approximately 50% of the clones will be lost at the *EcoR* I digest in the third step (Figure 6.4). The creation of a number of libraries with the same method and enzyme should provide a large enough number of clones and therefore multiple coverage of the genome that can compensate for this loss. The ligation products can then be transformed into bacterial cells, *e.g. E.coli* DH5-α, resulting in the creation of a library containing genomic DNA fragments, derived from a variety of HS.

Since restriction sites on the genome are not evenly distributed, but can be found clustered or tens of bp or tens of kbp apart, there should be a wide size variation observed for the DNA inserts on the plasmid. The smaller sized genomic fragments are expected to be over-represented in the library, as there are technical difficulties involved in the ligation of larger sized fragments. Once the sequence that is adjacent to the *Mbo* I restriction site is determined, the position of each clone can be safely pinpointed by using a BLAST search against the appropriate genome databases (in the particular research against the *D. melanogaster* genome), as well as the precise genomic location of a HS around the *Mbo* I restriction site. In order to confirm that the clones in these libraries correspond indeed to HS, an attempt would have been made to locate a number of them by using the indirect end-labelling technique.

6.1.3 Choice of the model system and available information re-examined.

6.1.3.1 Choice of the model system.

The chosen model system is the fruit fly *D. melanogaster* and in particular the Schneider 2 cell line. *D. melanogaster* was chosen as its genome is relatively small, but still one of a higher organism; it has been completely sequenced and annotated, with candidate and identified genes. Additionally, published work on *D. melanogaster* is not extensive, but it is individually detailed. Thus, the method for the HS library creation can be readily tested for its effectiveness and the HS characterised can be immediately compared to the known ones. The spread of characterised HS on the *D. melanogaster* genome is sporadic as the HS studied were part of wider studies on specific gene loci, their expression and regulation. Nevertheless, a presence of HS has been detected in most regions of the genome, such as telomeric and centromeric regions on more than one chromosomes. In addition, most of the HS that are known are very well characterised. Thus, the model system is believed to be an appropriate one in testing the efficiency of library generating method. However, there are difficulties regarding the cloning and characterising of HS that this technique has to overcome successfully. To mention one example, there are genes and regulatory elements of the genome embedded in heterochromatin or found between a large number of repeats that are frequently difficult to locate and clone (Adams *et al.* 2000).

There are presently no studies comparing the presence/absence of a large number of HS between different *D. melanogaster* cell lines. Similarly, the change of HS pattern has not been studied between different developmental stages in the same tissues or between different tissues either. The HS sequences could be used in a bioinformatics approach, developing algorithms capable of predicting HS position *in silico*.

6.1.3.2 HS and *Drosophila melanogaster*, available information re-examined.

One of the first steps that needed to be accomplished was that of identifying all of the HS mapped on *D. melanogaster*. Since HS are studied individually as part of looking at the regulation mechanisms of a specific gene or gene locus, there is no cumulative record of the HS recorded in the particular model organism. Thus, information that could be useful regarding *cis*-regulatory elements or other genomic regions that they commonly co-localise with could have been overseen. In the same way, some initial conclusions

regarding overall genomic distribution and sequence similarity information could be drawn from the re-examination of the available HS data for *D. melanogaster*. The available information was recorded cohesively and this record is presented for the first time comprehensively in section 6.2.4.3, where HS areas are grouped according to the chromosome they are located on, followed by the visual aid of an annotated genomic map.

6.1.4 Project aim.

The lack of extensive data on HS prohibits the answering of a number of very basic questions. These questions are, to mention but a few, the typical number of HS found in any particular genome and the proportion of them being constitutive. Therefore, the amount of variation there is among organisms and between different developmental stages in respect to HS patterns remains an assumption. Even the usual location of HS in respect to coding sequences is speculation as the data that is available in the literature is highly fragmented. This research includes the first attempt to present comprehensively and cohesively all the information regarding HS in *D. melanogaster*, so that the already available information is easily accessible and readily comparable to the one resulting from the HS library generation.

The development of a high-throughput method that is much quicker than the current methods and robust in terms of its reproducibility is an essential first step for the better study of HS. Furthermore, the detection of HS in large numbers should provide enough evidence to analyse statistically their distribution on the genome and their positioning in respect to characterised genes and known regulatory elements. From the analysis of the genomic data on higher organisms it is becoming increasingly obvious that organisms have invested heavily in the regulation of gene expression; this is supported by the number of genes and the different mechanisms actively involved in regulation of gene expression in any developmental stage. Thus, the knowledge of HS positioning is of great importance, as it can provide direct access to sequences that contain regulatory elements or to the organisation of the chromatin at the locus of interest. Additionally, the location of HS can also provide information on the extent that this overall DNA structure changes

during development, cellular differentiation and the ways that these pattern alterations might have an effect on global gene expression.

6.2 RESULTS

6.2.1 Cell culture conditions for the Schneider 2 cell line.

A number of different growth conditions were investigated for the S2 cell culture. The medium suggested by ATCC (American Type Culture Collection) for the propagation of the line is the one containing heat inactivated 10% FBS. Four different media 5% or 10% FBS, 5% or 10% CS were used to grow the cells and the rate of growth of the cell culture for each of these media was investigated. Cell counts were taken for each culture at regular intervals; each culture was grown for a week before being split. As expected, FBS stimulated the growth of the S2 cells more than CS. In the case of the FBS containing media, the cultures grown with a 5% concentration of FBS had a faster rate of growth than those with the 10%. The 5% FBS concentration was used as optimum henceforth in the growth of the cells. These were grown as a monolayer in canted neck flasks. The results are summarised below in Figure 6.5.

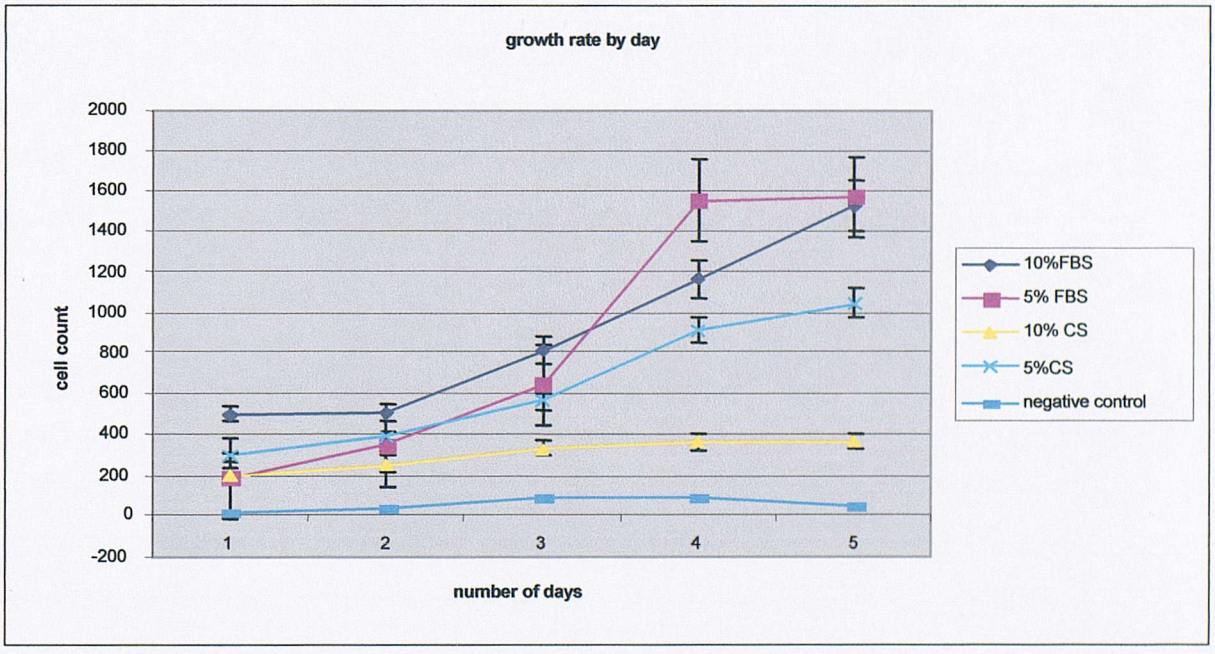


Figure 6.5: Chart summarising the cell culture condition results. The chart shows the rate of growth of the S2 cell cultures. Each of the media used is represented by a different line. The

negative control was the standard S2 medium, with no sera added to it. The 5%FBS and 10%FBS media have the best results in terms of rate of growth. The cell count was stopped after seven days, when the culture was split.

A suspension culture of the S2 cells, in the 5% FBS medium, was also set up in order to be used in the calculation of the optical density (OD). Optical readings at OD_{600nm} were taken from this cell culture, as optical readings are related to the cell density in the medium. This is a quicker method of determining cell density than cell counts. The results showed that a safe estimate of the concentration could be made by the OD_{600nm} measurement until the OD is 0.6. Also, there were small numbers of dead cells at any stage of culturing, until the OD_{600nm} of the culture solution reaches and surpasses 0.6, where the number of such cells becomes very large (verified using Trypan blue staining, 0.4% Trypan blue in PBS buffer). The inclusion of many dead cells in a culture means that there is a large portion of nuclei that have been broken and degraded or are breaking apart, thus their DNA conformation would be affected, thus affecting the results for the detection of HS.

Similar counts were also performed before and after centrifugation to pellet cells, prior to their DNase I digests. The aim was to investigate whether there is extensive cell damage in the first steps of the DNase I digest protocols. Thus involves the isolation of the cells and, subsequently, of their nuclei. It was shown that there is very little difference in cell mortality in the samples before and after each centrifugation.

6.2.2 DNase I digests of genomic DNA *in situ* and *in vivo.*

One practical difficulty facing this project was the absence of any standard DNase I digest protocol. As HS detection was only a part of previous studies focusing on particular genomic areas, each research followed a different protocol in achieving the analysis of the local chromatin structure. Therefore, various protocols resulting in the DNase I digestion of isolated nuclei were performed to allow for a direct comparison of the different techniques applied. This series of comparative experiments presented here,

summarised in table 6.2 on the next page, included some of the most common protocols cited in the literature, and their variations, and was necessary in order to find out the most efficient and reproducible method. The effects of different buffers on cell permeabilisation, nuclei resuspension and enzymic digest are considered. A comparison between the protocols involving nuclei isolation, an *in vivo* and an *in situ* digest is made.

Table 6.2: Summary of the different aspects of each of the DNase I digest protocols examined. The optimal condition sought was the one where there was the minimal amount of damage to the extracted and digested genomic DNA, within the shortest amount of time from the cell membrane permeabilisation step.

Step of protocol	Relevant protocols	Variations examined
Cell permeabilisation	Aasland and Stewart	10 mM Tris-HCl/ 15 mM Tris-HCl
	Carey and Smale	10 mM NaCl/ 15 mM NaCl
	Wu *et al.*	NP-40/ β-mercaptoethanol
Nuclei isolation	Aasland and Stewart	no sucrose/ 300 mM sucrose
	Carey and Smale	no sucrose/ 300 mM sucrose
DNase I digest	Aasland and Stewart	DNase I digest of isolated nuclei/ digest of cells *in situ* and *in vivo*

The first step of all the protocols involves the isolation of intact cell nuclei (approximately $1x10^7$ cells used per digest, yielding 60-100 μg of DNA per digest reaction) by the permeabilisation and lysis of the cells. The nuclei are subsequently digested by DNase I or any other restriction enzyme of choice. The differences between the various protocols were the different buffers used in the cell permeabilisation, and the nuclei isolation steps. The protocol from the first paper reporting the detection of HS (Wu *et al.* 1979) was the most extensive and laborious. However, a simplified version of it was attempted here, as parts of the protocol have been successfully replaced in later publications by more concise processes of equal results. The differences in the permeabilisation buffers relate to the varying amounts of sodium and potassium salt

concentration, NP-40 and β-mercaptoethanol, but these differences did not affect the subsequent digest of nuclei. The effectiveness of each buffer and its speed of permeabilising the cells were monitored in a haemocytometer by staining an aliquot of the cell culture with 0.4% Trypan blue solution in PBS buffer. Permeabilised cells stain blue, thus it was easy to observe whether the whole population of cells in the aliquot was stained during permeabilisation and how quickly. However, there was no significant difference observed in the time that it took for the whole cell culture to be permeabilised by the different buffers.

The difference between the nuclei isolation buffers used in the various protocols was the inclusion of sucrose or the lack thereof, used for the resuspension of the nuclei following cell lysis. It was thought that the PBS-sucrose containing buffer would protect the nuclei more than a plain PBS one, as it would allow for better osmolarity conditions in the solution of the nuclei during centrifugation. However, there was no overall difference observed in the condition of the digested DNA between the two solutions. The DNA was then further digested, to completion this time, by another set of enzymes, namely *Mbo* I, *Hind* III and *BamH* I (a range between 0.1 and 10 units of enzyme per ml of reaction), so that the comparison for the condition of the DNA is more efficient. Again, there was no visible difference in the condition of the differently treated genomic DNA.

The next comparison of protocols involved those where the DNase I digest occurs *in vivo* and *in situ*, in other words where the cell lysis solution containing the required amount of DNase I is directly added to the cell culture in the culture flask. These were compared to the protocols where the isolation of nuclei takes place first, followed by their digestion. It was shown that the digest *in situ* worked equally as effectively as the other protocols. Additionally, with the *in situ* digest there is less time elapsing between lysing the cells and digesting the genomic DNA. Time between nuclei isolation and genomic digest is of essence in this project, as we would like the HS identified to form a snapshot of the genomic DNA structure. The longer the isolated nuclei remain in suspension the more likely it is for some structures within the nucleus to collapse, thus a shorter protocol in terms of time is preferable.

The NP-40 buffer used in the protocols was tested to see whether it is an adequate digestion buffer for other nucleases, namely Mnase (Micrococcal nuclease) and *Mbo* I, and was effective in both cases. This was necessary as the generation of HS libraries requires that an enzyme other than DNase I is used. From the results it was clear that the genomic library of HS can be constructed by using this same NP-40 buffer. The genomic DNA was incubated overnight in each case with proteinase K, treated with *RNase* A and ethanol precipitated. Dialysis of the genomic DNA samples in 1 l of TE at 4°C overnight was also successful in replacing ethanol precipitation. In conclusion, the protocol used in this project was the one adapted by Carey and Smale (2000). This is because of its reproducible results, its simplicity and time efficiency, as well as on the fact that its buffer is a good digestion buffer for enzymes other than DNase I.

6.2.3 Radioactive labelling of DNA fragments.

The digestion protocol is to be followed by the actual detection of the HS fragments, achieved by Southern blotting and hybridisation. Therefore, once the probes to be used in the hybridisation were chosen, different types of labelling techniques were used in order to label each probe. The different techniques used in this project include primer labelling, PCR labelling, nick-translation and the commercially available Ready-To-Go radio-labelling kit, described in that order. The aim of this set of experiments was to obtain the probe with the highest specific activity. This is necessary as the HS that will be cleaved by the limited amounts of DNAase I form a small population of subgenomic sequences. Thus, in order to be able to successfully label these sequences among the total genomic DNA fragments, the probe needs to have the highest possible specific activity.

There were two genomic areas of the *D. melanogaste*r genome that were chosen for the testing of the different protocols: namely the (ribosomal protein large 32) *rpl32* and (heat shock protein 26) *hsp26* genomic regions. The relative positions of the various probes used, the gene coding areas and of the HS are shown below in Figure 6.6). The *rpl*32 gene was chosen as it is flanked by HS: more specifically, at its 3' end there are located

four strong, constitutive HS, while the promoter of the *hsp*26 gene contains a constitutive HS. Therefore, a successful digest of the HS will give fragments of known size, detected by Southern blotting as described in section 6.1.2.1.

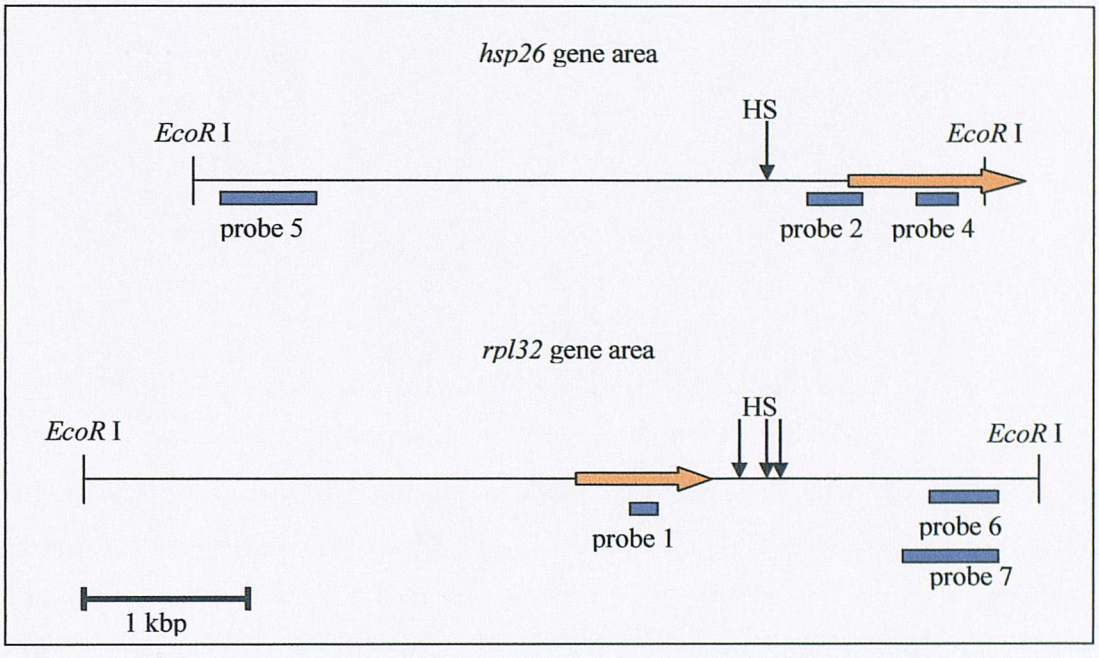


Figure 6.6: Genomic areas containing the *hsp26*and *rpl32* genes. The genomic areas have been drawn to scale. The gene is indicated by the shaded arrow; the HS by the vertical arrows. The shaded boxes show the position of the various probes that have been used. The *D. melanogaster* genome contains a single copy of both genes.

The first labelling attempt involved the polynucleotide kinase labelling of the DM1 and DM2 primers and the subsequent use of those primers in a PCR reaction, thus labelling the resulting DNA fragment. However, the specific activity of the labelled DNA was too low to detect the single copy *rpl*32 gene. The next attempt involved the addition of an [α^{32}P]-labelled dNTP in the PCR reaction mixture that produces the *rpl*32 fragment, in the particular case [α ^{32}P]-dGTP was used. The specific activity of the probe using this method can be substantial, but this would necessitate the use of large amounts of radioactive material. Thus amounts of non-radioactive dGTP (either a tenth or a fifth of the total amount of [α ^{32}P]-dGTP in the PCR reaction) were used alongside those of the radioactive one in the reaction mixture. The labelled DNA in this case detected and

hybridised to the *rpl*32 fragment. A result from the above series of comparative experiment is shown below in Figure 6.7.

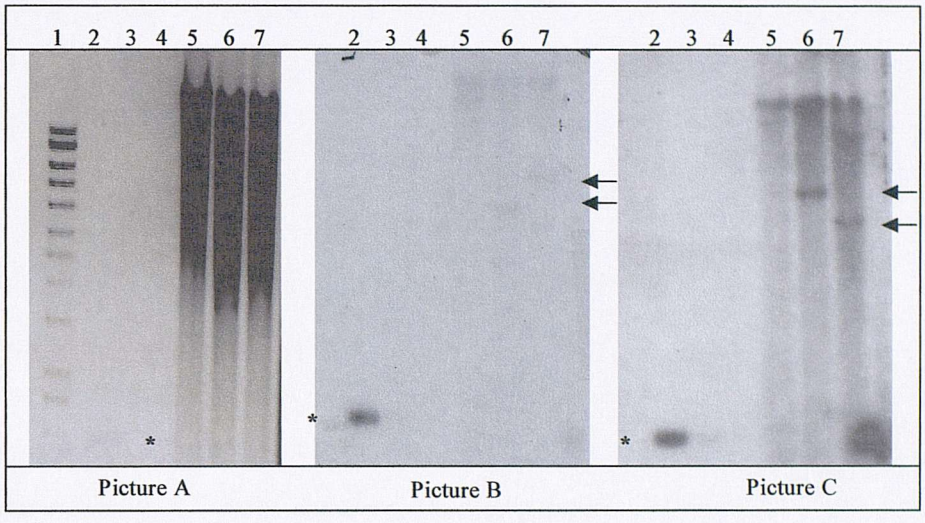


Figure 6.7: Comparing the strength of differently labelled radioactive probes. The lanes contain the 1 kb marker (lane 1), 100 ng, 10 ng and 1 ng of the PCR amplified *rpl32* genomic fragment (lanes 2, 3 and 4 respectively, denoted by an asterisk) and the total genomic DNA digests with *BamH* I, *EcoR* I and *Hind* III (lanes 5, 6 and 7 respectively). The gels were hybridised to the differently labelled *rpl32* probes. Picture A shows one of the agarose gels before Southern blotting., while pictures B and C are the autoradiographs after Southern blotting. The probe used in the case of Picture B was labelled by the inclusion of [α ^{32}P]-dGTP into the PCR reaction mixture. The probe used in the case of Picture C was labelled according to the nick-translation protocol. It is shown that the nick-translation protocol gives probes of higher specific activity, better able to detect single copy genes.

Nick translation of DNA is a method that produces a probe of high specific activity and equal distribution of labelled nucleotides along the DNA fragment. It is based on the ability of low concentrations of DNase I to induce single strand nicks, as opposed to double strand cleavage on the DNA fragments. The single strand nicks are repaired by the polymerase also included in the reaction mixture. The probes obtained with this method had higher specific activity than those of the previous approaches, and managed to hybridise to the PCR derived DNA fragment of *rpl*32, as well as to the copies of the gene in an undigested genomic DNA sample. However, the best results in terms of high

specific activity and percent incorporation of the probe (measured by TCA precipitation of the samples) were obtained by using the Radio-labelling kit, Ready-To-Go, with DNA labelling beads (using [α^{32}P]-dCTP). This kit utilizes the random primed DNA labelling approach, which is an alternative to the nick translation approach. As it also uses less time than the nick translation, the kit labelling was used to label the DNA probes for the rest of the research.

6.2.4 Mapping of HS in *Drosophila melanogaster.*

Southern blotting conditions.

The Church buffer and the Rapid–Hyb hybridisation buffer gave results with similar sensitivity and were used interchangeably. The Church buffer was used henceforth in the research as it is easy to produce in large amounts and gives as good a result in terms of sensitivity as its industrially available equivalent. The sensitivity of the blotting was further assisted by the subsequent high stringency wash, containing 0.5x SSC and 0.1% SDS.

6.2.4.1 Arbitrarily primed, cold PCR.

In parallel to the indirect end-labelling approach used to detect specific HS, involving Southern blotting, a PCR based approach was also investigated. This method examined the arbitrary primed PCR of genomic DNA. An arbitrary primer is used in a PCR reaction with low annealing temperature (40°C). This results in a specific band profile for a given genomic DNA sample. Depending on the levels of digestion of DNA, or on the digestion of the same sample by different enzymes, the band profile is altered. Therefore the change in bands corresponds to a change in the DNA structure.

The sensitivity of this method was investigated by subjecting the isolated, genomic DNA to a series of digests with the enzyme *Mbo* I for differing amounts of time (the amount of DNA and temperature were kept constant). This created a series of partial digests, in which the number of fragments increased as the amount of time increased. An aliquot from each of these digested DNA samples was then used as a template for an arbitrarily

primed PCR reaction. A series of PCR reactions were performed, using a single primer per PCR reaction. The digested DNA from gels with different band profiles can be transferred onto a nylon membrane and hybridised to one of the probes, for example the *rpl*32 probe, revealing whether the DNA fragments affected by the DNase I digestion in the band profiles corresponds to known HS. The sensitivity of this approach was not as high or reliable as expected as the band profile changed only at a high degree of genomic digestion, as shown below in Figure 6.8. The primer sequence itself could also have played a part in these results, but the optimisation of this method was not pursued any further.

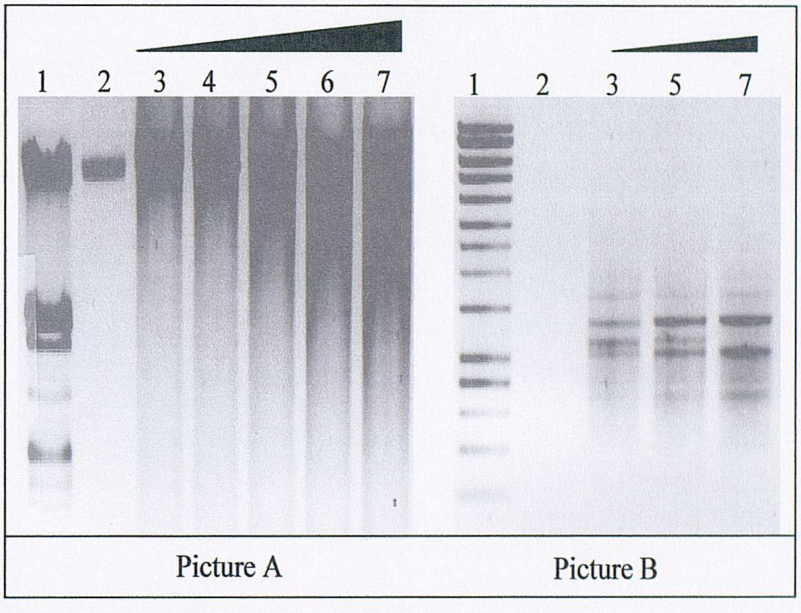


Figure 6.8: Different band profiles resulting from partially digested genomic DNA. In Picture A the lanes contain the λ/*Hind* III marker (lane 1), undigested genomic DNA (lane 2) and a series of partial genomic digests by increasing amounts of *Mbo* I (lanes 3 to 7). In Picture B the lanes contain the 1 kb marker (lane 1) and the cold PCR products using as template the partially digested genomic DNA from lanes 3, 5 and 7. It can be seen that although the partial genomic digest is successful, the sensitivity of the cold PCR is not very high, as the band profile only changes at the highest nuclease concentration.

HS in *Drosophila melanogaster* were mapped in two ways: 'wet' mapping, which involved the conventional, end-labelling based mapping, and the 'dry' mapping involving the collection and processing of all individually published HS data in *D. melanogaster*.

6.2.4.2 'Wet' mapping of HS.

The first step of the 'wet' mapping was to obtain a digestion profile of the *rpl*32 fragment when in a genome digested to completion; in this case enzymes *EcoR* I and *Hind* III were used. Such a profile will help avoid errors in interpreting the results when partial DNAase I digest takes place, differentiating it from a digestion to completion.

Subsequently, for the detection of HS and the confirmation of their being HS, the indirect end-labelling technique was used as described in section 6.1.2.1 and Figure 6.3. This involved firstly, the digestion of isolated nuclei under limited amounts of time and DNase I (a range between 0.1 ng and 1 μg of enzyme per ml of reaction, reaction stopped by the addition of 1 μl 0.5 M EDTA to the mixture). These genomic samples were then digested to completion with either *Hind* III (for indirect end-labelling of *rpl32*) or *EcoR* I (for *hsp26*) and were run on agarose gels. The gels were subsequently blotted and hybridised with their respective labelled probes. The results of this technique, as related to each probe are shown in Figure 6.9 for the *rpl32* gene and in Figure 6.10 for the *hsp26* genomic area.

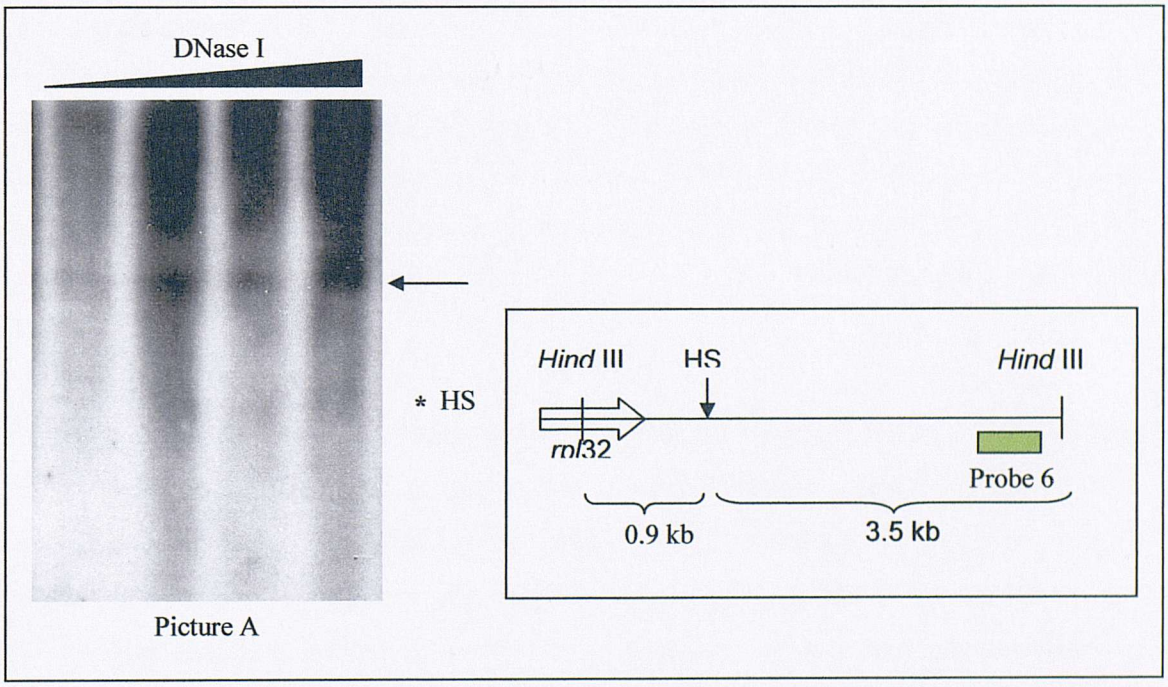


Figure 6.9: Mapping HS in the vicinity of the *rpl*32 gene. Picture A shows the autoradiograph following the Southern blot of a series of partially digested genomic DNA samples (by increasing

amounts of DNase I), subsequently digested to completion by *Hind* III. The horizontal arrow denotes the expected *Hind* III – *Hind* III fragment, while the asterisk shows the fragment resulting from the presence of a HS cleaved by DNase I. The inset box presents a map of the area. The gene coding region is represented as the clear, horizontal arrow, while the position of the HS is shown by the black, vertical arrow. The location of the probe used is also shown.

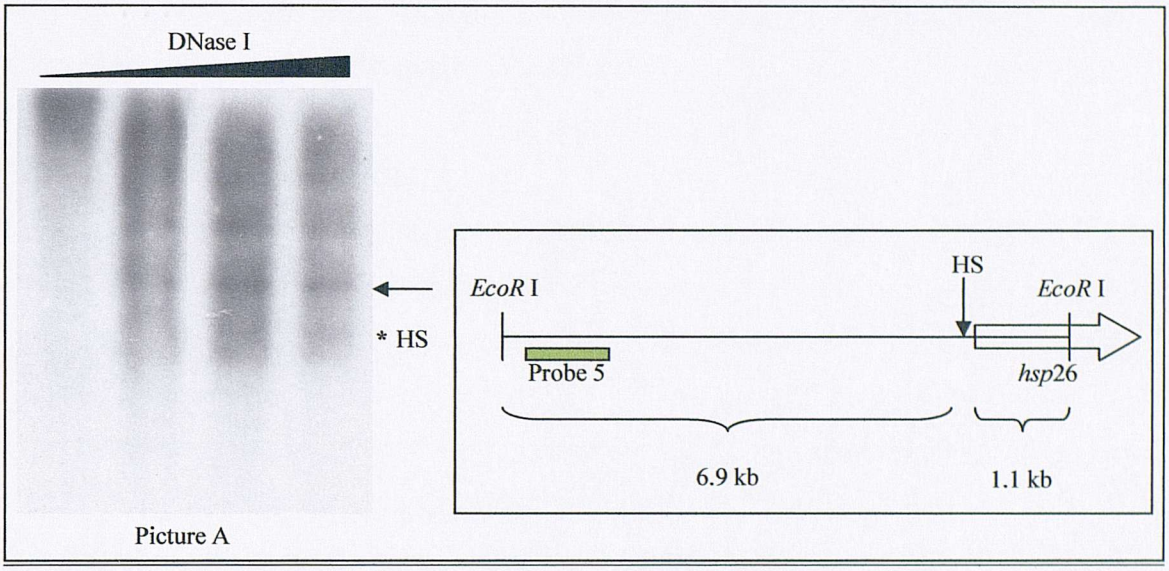


Figure 6.10: Mapping HS in the vicinity of the *hsp*26 gene. Picture A shows the autoradiograph following the Southern blot of a series of partially digested genomic DNA samples (by increasing amounts of DNase I), subsequently digested to completion by *EcoR* I. The horizontal arrow denotes the expected *EcoR* I – *EcoR* I fragment, while the asterisk shows the fragment resulting from the presence of a HS cleaved by DNase I. The inset box presents a map of the area. The gene coding region is represented as the clear, horizontal arrow, while the position of the HS is shown by the black, vertical arrow. The location of the probe used is also shown.

In order to answer further questions regarding the nature and distribution of the HS in *D. melanogaster*, a larger number of known and mapped HS need to become available. Then an estimate for the number of HS in the whole of the *D. melanogaster* genome could be made. Such an estimate can be calculated by using the generation of the HS library and the random sampling approach, based on the frequency of identical clones appearing upon randomly selecting and cloning part of the library. To give an example, the *D. melanogaster* genome can potentially contain 7030 HS. The latter estimate is based on

the prediction of 1% of the genome being in the form of HS (Gross and Garrard 1988) and the average size of a HS core being 200 bp. Thus, the expected probability for the appearance of identical clones is 0.014%. In other words, out of every 100 sequenced clones one would expect to find only one pair of identical sequences. Of course, this prediction is affected by the existence of weak and strong HS, since the latter would tend to be present more often. However, the random sampling method can still give a workable and representative rough estimate.

6.2.4.3 'Dry' mapping of HS.

The aim of the 'dry' mapping was to assemble all the existing information, code it and present it in a concise and comprehensive form, so that any new information regarding HS that might result from the generation of HS libraries could be easily added into this database. Furthermore, there is an amount of information to be extracted from the construction of such a genomic map of published HS. For example, the available data gave some clues for the genomic distribution of HS in *D. melanogaster*, while the DNA sequence of known HS that was unstudied was analysed for the first time. The list of HS in *D. melanogaster* is summarised below in Table 6.3.

Table 6.3: Table of all the published *D. melanogaster* HS to date, grouped according to the chromosome on which they are located. Chromosomes two and three are divided in their L and R arms The HS biology as well as the element(s) the HS co-localises with are also given when known. It should be noted that not all HS within a cluster of HS co-localise with *cis*-regulatory elements. In cases where the exact function of the element co-localising with the HS remains to be safely determined, a questionmark is present.

Chromosome	Gene/ gene locus	HS biology	Element(s) the HS co-localises with
X	actin 5C	Constitutive	Proximal promoter?
X	IMPDH ras	Constitutive?	Boundary element
X	notch	Constitutive	Insulator element
X	roX1	Male specific	MSL binding site
X	roX2	Male specific	MSL binding site

X	rudimentary gene	Constitutive	Transcription initiation site
X	Sgs4	Tissue specific	LCR?
2L	alcohol dehydrogenase	Dev. stage specific	Promoter
2R	eve	Dev. stage specific	Promoter, distal regulatory element
2R	Histone genes: H1, H2a, H2b, H3 and H4	Constitutive	Proximal promoter?
2R	Larval cuticle protein genes 1, 2, 3 and 4	Dev. stage specific	Proximal promoter?
3L	Heat shock proteins: *hsp*22, *hsp*23, *hsp*26, *hsp*27, *hsp*28, *hsp*82 and *hsp*83	Constitutive	Proximal promoter
3L	Sgs3	Tissue specific	Enhancer element
3R	*rpl*32	Constitutive	?
3R	abd-B	Constitutive	Boundary elements
Chromosome	**Gene/ gene locus**	**HS biology**	**Element(s) the HS co-localises with**
3R	Heat shock proteins: *hsp*70 Aa, Ab, Ba, Bb, and Bc	Constitutive	Promoter element
3R	tRNA gene cluster at 90BC: tRNAs for Val, Pro, Ala and Thr	Constitutive	Transcription initiation site?
Y	ribosomal gene repeat	Constitutive	Promoter elements?
There are currently no HS mapped for chromosome 4			

In Figure 6.11 the genomic location of each one of the genomic areas where HS have been mapped is shown. From the data available, it seems that there is not any specific area, such a centromeric or one telomeric region, where HS tend to accumulate on the *D. melanogaster* genome. Based on this latter observation, their distribution on each

chromosome and on various parts of the individual chromosomes is more consistent with a random distribution model. However, more data are necessary in order to confirm such a conclusion. From the attribution of function, it seems that a large part of the known HS coincide with promoter sequences. This observation could also be an experimental artefact, *i.e.* a result of the extensive study of the heat-shock sensitive gene cluster. The sequences coinciding with the HS contain elements that function individually, in a similar way to the elements such as insulators, enhancers, *etc.*, characterised in other various organisms. Thus, according to the existing data, the HS in *D. melanogaster* do not seem to have any additional role to play in terms of genomic organisation and transcriptional regulation; or a role that is very different to those already known from other organisms, as described in the introduction.

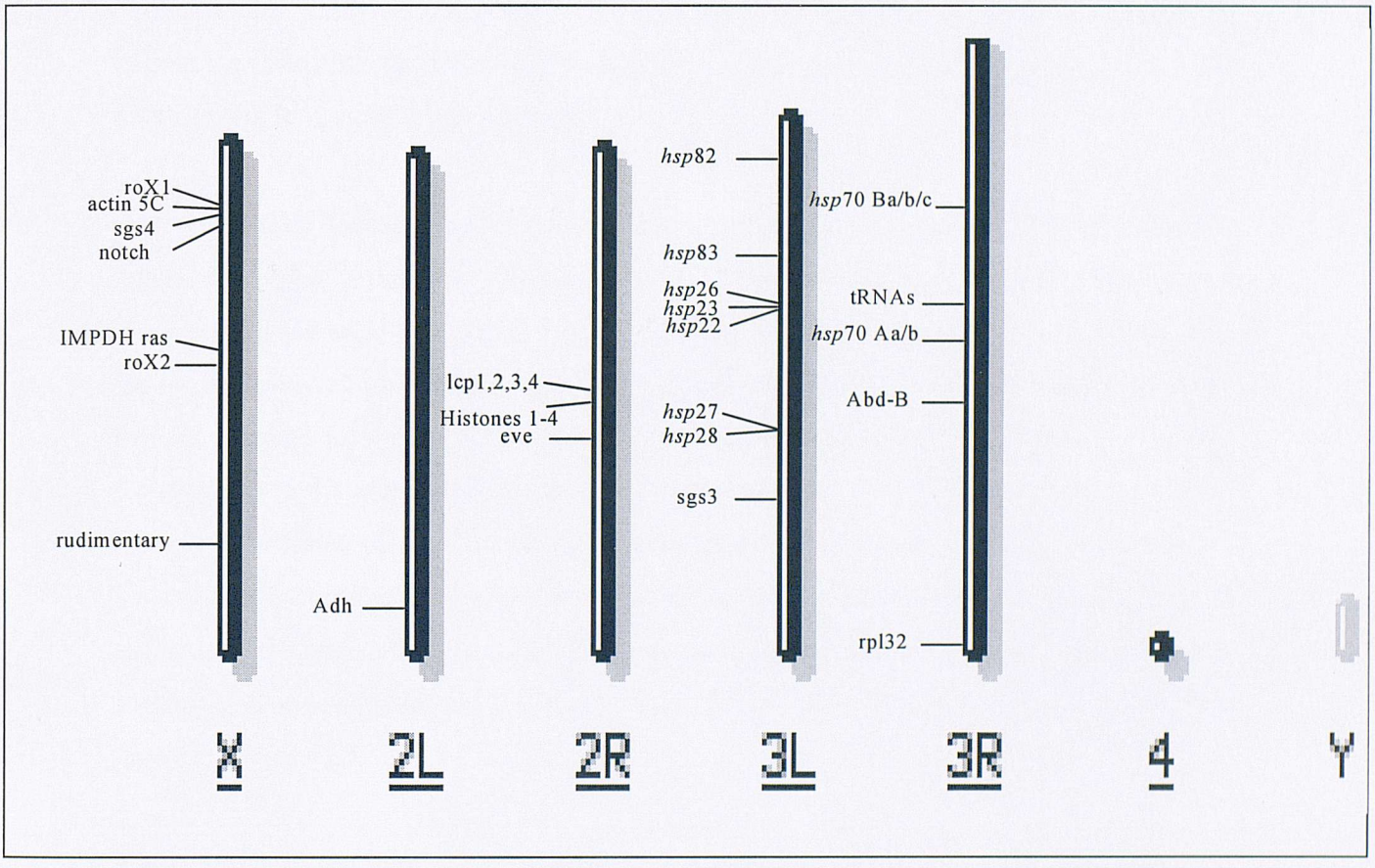


Figure 6.11: HS mapped on the *D. melanogaster* genome. The figure shows the location of the HS published, as presented in Table 6.3. The diagram is not drawn to scale, the location of the loci is to the best possible approximation, thus it is considered representative of the actual one.

6.3 DISCUSSION

6.3.1 Experimental overview.

As this project strived to introduce a new, practical approach in HS detection, its technical aspects had to be thoroughly investigated; accordingly it needed to be proved that the methods we use are the best suited ones for the purposes of each experimental step and that they are reproducible enough to be applied in a standard protocol for the generation HS libraries. The growth conditions of the S2 cell line were investigated, as this can be useful in the future when HS libraries from cell cultures grown under different conditions are generated. Furthermore, cell density within the culture can be quickly estimated by measuring the OD of a sample from the cell culture. Thus there is consistency between the different genomic DNA digests, as the same amounts of cells are exposed to the same amounts of nuclease.

The main issue facing this project from the beginning was the abundance of different protocols for performing the DNase I digests. That made it necessary to compare the most commonly used of them as well as some of their variations and decide which one should be followed or whether a new one should be considered. This was a laborious and time consuming task. The protocols were analysed and compared in a stepwise fashion, for example in the effects of the use of different buffers at the cell permeabilisation step, as well as in terms of their final result. Other parameters considered were the amount of time necessary for each of the protocol and the amount of DNA resuspensions; time is of essence, as the state of the genomic DNA can be affected. The protocol that gave the best results in a reproducible manner (Carey and Smale 2000) was applied throughout this project.

It was only right, though, after investing so much time on the protocol, that the rest of the techniques used in the project should also be thoroughly investigated. Different techniques for the radio-labelling of the DNA fragments were investigated, as the probe used in the HS detection needs to be of a very high specific activity. The nature of the

HS, being a small sub-population of unique DNA fragments necessitates such an aim in order for them to be detectable. After spending a considerable amount of time in comparing the experimental results, the best method for providing a radio-labelled probe with a very high specific activity was the commercially available one that is easy to use and readily available. This was a lesson in planning such experiments in the future.

The conditions for the Southern blotting were also examined in detail. There were four different hybridisation buffers that were compared, this time the commercially available one was included to the comparisons right from the beginning. In terms of sensitivity of the hybridisation conditions, the Church buffer and the bought one had a similar result. A novel approach in detecting HS was also examined, using the arbitrarily primed PCR. It was not as sensitive as originally anticipated, however since there are a couple of parameters that might be affecting the results, such as the primer sequence, it is thought worth to be revisited at a later date.

6.3.2 New methods, data and future prospects.

The detection of HS in the vicinity of the *rpl*32 and *hsp*26 genes using the indirect end-labelling method, followed the protocol optimisation steps. Those two genes were to be the positive control used to verify the success of the novel technique for generating libraries containing HS. Immediately after these sets of experiments the actual creation of the HS libraries was to follow, whereafter clones would have been isolated and sequenced. The research was terminated at this stage. It is interesting to note that a number of studies describing different approaches for the genome-wide detection and identification of HS have been recently published (they are described below).

Digital Analysis of Chromatin Structure (DACS) is a hybrid molecular-computational approach comprising two distinct phases: firstly the production of a genomic DNA tag library encompassing individual HS fragments and secondly the computational analysis of the genomic tag distributions to identify statistically significant clusters (Sabo *et al.* 2004a and Sabo *et al.* 2004b). In the particular study, just over 250,000 tags from human

K562 cells were sequenced in total, of which approximately 9,500 were statistically very highly probable to coincide with HS. Six of the predicted HS clusters were verified as indeed being HS clusters, through cross-referencing in the existing literature or by using the indirect end-labelling method. The *in silico* prediction model identified HS candidate sites based on repetition of the DNase I cut on the same place and by density, *i.e.* the amount of DNase I cuts 500 bp on either side of the tag, since many HS are clustered. Although this approach is the most efficient to date in predicting HS, it can not always differentiate successfully between HS and sensitive but not hypersensitive sites, non-specific DNase I cuts or random fragments. It is also more sensitive for HS clusters and less effective for single HS. Simulation runs of the *in silico* prediction model, using random sets of genomic tag libraries, showed that it can predict up to 90% of any existing HS at any one time, but above this threshold the background of false positives interferes with the results.

Human HS were also mapped from $CD4^+$ cells by vector assisted end cloning. The generated libraries were enriched in fragments containing DNase I HS using a protocol similar to the one presented in Figure 6.4 (Crawford *et al.* 2004). Approximately 5,500 clones were sequenced, 85% of them mapping to a single position on the genome and a third within 2 kb of an mRNA coding sequence, preferentially one that is expressed. Interestingly, half of the sequenced clones also mapped in areas that are conserved and can be found in the mouse genome. The authors suggested that the amount of HS present in this particular cell type is approximately 100,000. This means that six DNase I sites would correspond roughly to every expressed gene.

In a number of different studies, the active creation or disappearance of HS was studied indirectly by monitoring the degree of partial loss of histone H3 and H4 tetramers from the coding regions of heavily transcribed coding genes in yeast, *Saccharomyces cerevisiae* (Lee *et al.* 2004 and Bernstein *et al.* 2004) The promoters of genes repressed upon heat-shock (approximately 14% of all yeast genes) tended to have higher nucleosome occupancy after heat-shock by a factor of at least two. The data suggests an active and global mechanism for the creation of HS. In contrast, a study of the HS on the

left arm of chromosome III in *S. cerevisiae,* showed that most of the chromatin was very similarly organised between the different a- and α-cell mating types as well as during mating-type switching (Ercan and Simpson 2004). There was no indication of a global, generalised alteration in nuclease susceptibility during mating-type switching. There were also HS detected in promoters of untranscribed genes, suggesting that the establishment of transcriptional competency can proceed in the absence of extensive levels of transcription. It should be noted at this point that the genomic organisation of *S.cerevisiae* bears only broad similarities to the epigenetic organisation of higher eukaryotes, thus any predictions regarding HS location and function can not be generalised.

Two computational models for the prediction of HS were also recently published. In the first case, the frequent clustering of regulatory sites that often co-localise with HS, was used to create an *in silico* model able to choose HS candidate regions (Halfon *et al.* 2002). The model integrated a number of known *cis*-elements in *D. melanogaster*, looking for areas where at least five of those factors, including repetitions, coincide within 500 bp of the genomic sequence. Monte Carlo simulations and intra-species comparison for conserved non-coding genomic blocks with *Drosophila virilis* were used to filter the initial predictions. Although this method does not exclude false positive results, the filtering parameters ensure that they will occur in low frequency. In the second case, a similar approach, based on the clustering of enhancer elements only in *D. melanogaster*, was used to predict other possible locations where such clusters might be present, resulting also in the location of candidate HS (Rebeiz *et al.* 2002). Although this approach can be highly sensitive for predicting transcriptional regulatory loci and HS based on the frequency of the most commonly used transcription factor binding sites, it can not detect cis-regulatory elements that are unique and use only a single binding site for a transcription factor.

In conclusion, the feasibility of cloning and predicting DNase I HS has recently been demonstrated by using both vector-assisted end cloning and *in silico* subtractive pool enrichment respectively. However, the former method is not robust enough for the detection of all DNase I HS on a genome-wide scale because of the absence of a 1:1 ratio

between cutting and cloning events. Thus, for the efficient detection of a significant number of HS it requires the sequencing of a very large number of clones. The latter method however, critically requires an independent molecular validation for each candidate genomic area. Therefore a successful method for the genome-wide detection of HS would be one that can combine and integrate these two approaches in the most efficient manner.

Acknowledgments

I would like to thank Dr. Coutts for his guidance, support and good company throughout the years; for his encouragement not only for this work but also to attend and participate in various meetings around the world. Many thanks go to members of the Dr. Coutt's research group at Imperial College, namely Dr. Livieratos and Dr. Eliasco. I would also like to thank Dr. Lamb (Imperial College, London) and Dr. Werner (University College, London). Additionally, I would like to acknowledge the Division of Biology, Department of Biological Sciences, Imperial College, for its financial support.

References

A. References for chapters one - five.

Açikgöz, S., Doken, M., Citir, A., and Bostan, H. (1994). *EMBO-INRA Workshop 'Plants and viruses: partners in pathogenicity', Thierval-Grignon.*

Aguilar, J., Franco, M., Marco, C., Berdiales, B., Rodriguez-Cerezo, E., Truniger, V., and Aranda, M. (2003). Further variability within the genus Crinivirus, as revealed by determination of the complete RNA genome sequence of Cucurbit yellow stunting disorder virus. *J. Gen. Virol.* **84,** 2555 - 2564.

Ahn, I., and Lee, Y. (2001). A viral double-stranded RNA up regulates the fungal virulence of *Nectria radicicola. Mol. Plant Microbe Interact.* **14**(4), 496 - 507.

Alioto, D., Zacaria, F., Covelli, L., Di Serio, F., Ragozzino, A., and Milne, R. (2003). Light and electron microscope observations on chlorotic rusty spot, a disorder of cherry in Italy. *J. Plant Pathol.* **85**(3), 215 - 218.

Allaby, M. (1992). "The Concise Oxford Dictionary of Botany." Oxford University Press, Oxford, UK.

Annesi, T., Curcio, G., D'Amico, L., and Motta, E. (2005). Biological control of *Heterobasidion annosum* on *Pinus pinea* by *Phlebiopsis gigantea. For. Path.* **35**(2), 127 - 134.

Bamford, D. (2003). Do viruses form lineages across different domains of life? *Res. Microbiol.* **154,** 231 - 236.

Banks, G., Buck, K., Chain, E., Darbyshire, J., Himmelweit, F., and Ratti, G. (1970). Antiviral activity of double stranded RNA from a virus isolated from *Aspergillus foetidus. Nature* **227,** 505 - 507.

Behrendt, C., and Blanchette, R. (1997). Biological processing of pine logs for pulp and paper production with *Phlebiopsis gigantea. Appl. Environ. Microbiol.* **63**(5), 1995 - 2000.

Bessarab, I., Liu, H., Ip, C., and Tai, J. (2000). The complete cDNA sequence of a type II *Trichomonas vaginalis* virus. *Virology* **267,** 350 - 359.

Blodgett, E., Ismen, F., Gomec, B., Alay, K., Altinyay, N., and Wagnon, H. (1970). FAO Plant protection Bullettin Newsletter 18. Amasya cherry disease in Turkey.

Bruenn, J. (1993). A closely related group of RNA-dependent RNA polymerases from double-stranded RNA viruses. *Nucleic Acids Res.* **21**(24), 5667 - 5669.

Bruenn, J. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Res.* **31**(7), 1821 - 1829.

Buck, K. (1998). Molecular variability of viruses of fungi. *In* "Molecular variability of fungal pathogens." (P. Bridge, Ed.). CABI BIOSCIENCE, UK Centre, Egham.

Buck, K., and Brasier, C. (2001). Viruses of the Dutch elm disease fungi. *In* "DsRNA genetic elements: concepts and applications." (S. Tavantzis, Ed.), pp. 165 - 190. CRC Press, Florida.

Carrion, R. J., Ro, Y., and Patterson, J. (2003). Purification, identification, and biochemical characterization of a host-encoded cysteine protease that cleaves a leishmaniavirus gag-pol polyprotein. *J. Virol.* 77(19), 10448 - 10455.

Caston, J., Ghabrial, S., Jiang, D., Rivas, G., Alfonson, C., Roca, R., Luque, D., and Carrascosa, J. (2003). Three-dimensional structure of *Penicillium chrysogenum* virus: a double-stranded RNA virus with a genuine *T*=1 capsid. *J. Mol. Biol.* **331,** 417 - 431.

Celix, A., Rodriguez-Cerezo, E., and Garcia-Arenal, F. (1997). New satellite RNAs, but no DI RNAs, are found in natural populations of tomato bushy stunt tombusvirus. *Virology* **239,** 277 - 284.

Chen, B., and Nuss, D. (1999). Infectious cDNA clone of hypovirus CHV1-Euro7: a comparative virology approach to investigate virus-mediated hypovirulence of the chestnut blight fungus *Cryphonectria parasitica*. *J. Virol.* **73**(2), 985 - 992.

Cheng, J., Jiang, D., Fu, Y., Li, G., Peng, Y., and Ghabrial, S. (2003). Molecular characterisation of a dsRNA totivirus infecting the sclerotial parasite *Coniothyrium minitans*. *Virus Res.* **93,** 41 - 50.

Chu, Y., Lim, W., Yea, S., Cho, J., Lee, Y., and Kim, K. (2004). Complexity of dsRNA mycovirus isolated from *Fusariun graminearum*. *Virus Genes* **28**(1), 135 - 143.

Citir, A. (1987). Preliminary investigations about identity of the causal agent of Amasya cherry disease in Turkey. *J. Turk. Phytopathol.* **16**(1), 23 - 34.

Coenen, A., Kevei, F., and Hoekstra, R. (1997). Factors affecting the spread of double-stranded RNA viruses in *Aspergillus nidulans*. *Genet. Res. Camb.* **69,** 1 - 10.

Coffin, R., and Coutts, R. (1992). DsRNA cloning and diagnosis of beet pseudo-yellows virus by PCR and nucleic acid hybridization. *Intervirology* **33**(4), 197 - 203.

Coffin, R., and Coutts, R. (1995). Relationships among *Trialeuroides vaporarium*-transmitted yellowing viruses from Europe and North America. *J. Phytopathol.* **143,** 375 - 380.

Compel, P., Papp, I., Bibo, M., Fekete, C., and Hornok, L. (1999). Genetic interrelationships and genome organization of double-stranded RNA elements of *Fusarium poae*. *Virus Genes* **18**(1), 49 - 56.

Coutts, R. (2005). First report of an *Endornavirus* in the *Cucurbitaceae*. *Virus Genes* **31**(3), 361 - 362.

Coutts, R., and Coffin, R. (1996). Beet psuedo-yellows virus is an authentic closterovirus. *Virus Genes* **13**(2), 179 - 181.

Coutts, R., and Livieratos, I. (2003). A rapid method for sequnecing the 5' and 3'-termini of double-stranded RNA viral templates using RLM-RACE. *J. Phytopathol.* **151,** 525 - 527.

Coutts, R., and Livieratos, I. (2003). Nucleotide sequence and genome organization of Cucurbit yellow stunting disorder virus RNA1. *Arch. Virol.* **148**(10), 2055 - 2062.

Coutts, R., Covelli, L., Di Serio, F., Citir, A., Acikgoz, S., Hernandez, C., Ragozzino, A., and Flores, R. (2004). Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. II. Characterization of a new species in the genus *Partitivirus*. *J. Gen. Virol.* **85,** 3399 - 3403.

Coutts, R., and Livieratos, I. (2003). A rapid method for sequnecing the 5' and 3'-termini of double-stranded RNA viral templates using RLM-RACE. *J. Phytopathol.* **151,** 525 - 527.

Covelli, L., Coutts, R., Di Serio, F., Citir, A., Acikgoz, S., Hernandez, C., Ragozzino, A., and Flores, R. (2004). Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. I. Characterization of a new species in the genus *Chrysovirus*. *J. Gen. Virol.* **85,** 3389 - 3397.

Dawe, A., Segers, G., Allen, T., McMains, V., and Nuss, D. (2004). Microarray analysis of *Cryphonectria parasitica* Galpha- and Gbetagamma-signalling pathways reveals extensive modulation by hypovirus infection. *Microbiol.* **150,** 4033 - 4043.

Deacon, J. Course notes. (1998). Profiles of microorganisms - biological control: *Bacillus popilliae*. Edinburgh, UK: Biological Teaching Organisation, University of Edinburgh.

Diamond, M., Dowhanick, J., Nemeroff, M., Pietras, D., Tu, C., and Bruenn, J. (1989). Overlapping genes in a yeast double-stranded RNA virus. *J. Virol.* **63**(9), 3983 - 3990.

Diaz-Ruiz, J., and Kaper, J. (1978). Isolation of viral double-stranded RNAs using a LiCl fractionation procedure. *Prep. Biochem.* **8**(1), 1 - 17.

Di Serio, F., Daros, J., Ragozzino, A., and Flores, R. (1997). A 451-nucleotide circular RNA from cherry with hammerhead ribozymes in its strands of both polarities. *J. Virol.* **71**(9), 6603 - 6610.

Dickinson, M., Zhang, R., and Pryor, A. (1993). Nucleotide sequence relationships of double-stranded RNAs in flax rust, *Melampsora lini*. *Curr. Genet.* **24**(5), 428 - 432.

Esteban, R., and Fujimura, T. (2003). Launching the yeast 23S RNA narnavirus shows 5' and 3' cis-acting signals for replication. *Proc Natl Acad Sci U S A* **100**(5), 2568 - 2573.

Fukuhara, T., Moriyama, H., and Nitta, T. (1995). The unusual structure of a novel RNA replicon in rice. *J. Biol. Chem.* **270,** 18147 - 18149.

Garlapati, S., and Wang, C. (2004). Identification of a novel internal ribosome entry site in Giardiavirus that extends to both sides of the initiation codon. *J. Biol. Chem.* **279**(5), 3389 - 3397.

Ghabrial, S. (1998). Origin, adaptation and evolutionary pathways of fungal viruses. *Virus Genes* **16**(1), 119 - 131.

Ghabrial, S., and Hillman, B. (1999). Partitiviruses - Fungal (*Partitiviridae*). *In* "Encyclopedia of Virology", Vol. 2004. Academic Press.

Ghabrial, S., and Patterson, J. (1999). Totiviruses (*Totiviridae*). *In* "Encyclopedia of Virology", Vol. 2004. Academic Press.

Gibbs, M., Koga, R., Moriyama, H., Pfeiffer, P., and Fukuhara, T. (2000). Phylogenetic analysis of some large double-stranded RNA replicons from plants suggests they evolved from a defective single-stranded RNA virus. *J. Gen. Virol.* **81,** 227 - 233.

Gimenez-Pecci, M., Bogo, M., Santi, L., de Moraes, C., Correa, C., Vainstein, M., and Schrank, A. (2002). Characterization of mycoviruses and analyses of chitinase secretion in the biocontrol fungus *Metarhizium anisopliae*. *Curr. Microbiol.* **45,** 334 - 339.

Gispert, C., Perring, T., and Creamer, R. (1998). Purification and characterisation of peach mosaic virus. *Plant Disease* **82,** 905 - 908.

Grogan, H. M.; Adie, B. A. T.; Gaze, R. H.; Challen, M. P. and Mills, P. R. (2003). Double-stranded RNA Elements Associated with the MVX Disease of *Agaricus bisporus*. *Mycological Research*, **107** (2), 147 - 154.

Hacker, C., Brasier, C., and Buck, K. (2005). A double-stranded RNA from a Phytophthora species is related to the plant endornaviruses and contains a putative UDP glycosyltransferase gene. *J. Gen. Virol.* **86,** 1561 - 1570.

Hillman, B., Foglia, R., and Yuan, W. (2000). Satellite and defective RNAs of *Cryphonectria hypovirus* 3-Grnad Haven 2, a virus species in the family *Hypoviridae* with a single open reading frame. *Virology* **276,** 181 - 189.

Hillman, B., Supyani, S., Kondo, H., and Suzuki, N. (2004). A reovirus of the fungus *Cryphonectria parasitica* that is infectious as particles and related to the *Coltivirus* genus of animal pathogens. *J. Virol.* **78**(2), 892 - 898.

Hollings, M. (1962). Viruses associated with a die-back disease of cultivated mushroom. *Nature* **196,** 962 - 965.

Hong, Y., Cole, T., Brasier, C., and Buck, K. (1998). Evolutionary relationships among putative RNA-dependent RNA polymerases encoded by a mitochondrial virus-like RNA in the Dutch elm disease fungus, *Ophiostoma novo-ulmi,* by other viruses and virus-like RNAs and by the *Arabidopsis* mitochondrial genome. *Virology* **246,** 158 - 169.

Hong, Y., Cole, T., Brasier, C., and Buck, K. (1998). Novel structures of two virus-like RNA elements from a diseased isolate of the Dutch elm fungus, *Ophiostoma novo-ulmi. Virology* **242,** 80 - 89.

Hong, Y., Dover, S., Cole, T., Brasier, C., and Buck, K. (1999). Multiple mitochondrial viruses in an isolate of the Dutch elm disease fungus *Ophiostoma novo-ulmi. Virology* **258,** 118 - 127.

Horiuchi, H., Moriyama, H., and Fukuhara, T. (2003). Inheritance of *Oryza sativa endornavirus* in F1 and F2 hybrids between *japonica* and *indica* rice. *Genes Genet. Syst.* **78,** 229 - 234.

Huang, S., and Ghabrial, S. (1996). Organization and expression of the double-stranded RNA genome of *Helminthosporium victoriae 190S* virus, a totivirus infecting a plant pathogenic filamentous fungus. *Proc Natl Acad Sci U S A* **93,** 12541 - 12546.

Huang, S., Soldevila, A., Webb, B., and Ghabrial, S. (1997). Exrpession, assembly, and proteolytic processing of *Helminthosporium victoriae 190S* totivirus capsid protein in insect cells. *Virology* **234,** 130 - 137.

Hull, R., and Covey, S. (1983). Characterisation of cauliflower mosaic virus DNA forms isolated from infected turnip leaves. *Nucleic Acids Res.* **11**(6), 1881 - 1895.

Icho, T., and Wickner, R. (1989). The double-stranded RNA genome of yeast virus L-A encodes its own putative RNA polymerase by fusing two open reading frames. *J. Biol. Chem.* **264**(12), 6716 - 6723.

Iwahashi, F., Fujisaki, K., Kaido, M., Okuno, T., and Mise, K. (2005). Synthesis of infectious in vitro transcripts from Cassia yellow blotch bromovirus cDNA clones and a reassortment analysis with other bromoviruses in protoplasts. *Arch. Virol.* **150**(7), 1301 - 1314.

James, D., and Godkin, S. (1999). Electorn microscopic detection of novel, coiled viruslike particles associated with graft-inoculation of *Prunus* species. *Plant Disease* **83**(10), 949 - 953.

Jiang, D., and Ghabrial, S. (2004). Molecular characterization of *Penicillium chrysogenum virus*: reconsideration of the taxonomy of the genus *Chrysovirus*. *J. Gen. Virol.* **85,** 2111 - 2121.

Kang, J., Wu, J., Bruenn, J., and Park, C. (2001). The H1 double-stranded RNA genome of *Ustilago maydis virus-H1* encodes a polyprotein that contains structural motifs for capsid polypeptide, papain-like protease, and RNA-dependent RNA polymerase. *Virus Res.* **76,** 183 - 189.

Karasev, A. (2000). Genetic diversity and evolution of closteroviruses. *Annu. Rev. Phytopathol.* **38,** 293 - 324.

Khramtsov, N., Woods, K., Nesterenko, M., Dykstra, C., and Upton, S. (1997). Virus-like, double-stranded RNAs in the parasitic protozoan *Cryptosporidium parvum*. *Mol. Microbiol.* **26**(2), 289 - 300.

Kim, J., Kim, S., and Kim, K. (2003). Genome organization and expression of the *Penicillium stoloniferum* virus S. *Virus Genes* **27**(3), 249 - 256.

Koenig, R., and Huth, W. (2000). *Soil-borne rye mosaic virus* and *European wheat mosaic virus*: two names for a furovirus with variable genome properties which is widely distributed in several crops in Europe. *Arch. Virol.* **145,** 689 - 697.

Koga, R., Horiuchi, H., and Fukuhara, T. (2003). Double-stranded RNA replicons associated with chloroplasts of a green alga, *Bryopsis cinicola*. *Plant Mol. Biol.* **51,** 991 - 999.

Koonin, E., Choi, G., Nuss, D., Shapira, R., and Carrington, J. (1991). Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses. *Proc Natl Acad Sci U S A* **88,** 10647 - 10651.

Korhonen, K., and Stenlid, J. (1998). Biology of *Heterobasidion annosum*. *In "Heterobasidion annosum*: Biology, Ecology and Impact control." (S. Woodward, J. Stenlid, R. Karjalainen, and A. Hóttermann, Eds.), pp. 43 - 71. CAB International, Wallingford, UK.

Kozak, M. (1995). Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc Natl Acad Sci U S A* **92,** 2662 - 2666.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene* **234,** 187 - 208.

Lakshman, D., Jian, J., and Tavantzis, S. (1998). A double-stranded RNA element from a hypovirulent strain of *Rhizoctonia solani* occurs in DNA form and is genetically related to the pentafunctional AROM protein of the shikimate pathway.
Proc Natl Acad Sci U S A **95,** 6425 - 6429.

Lampson, G., Tytell, A., Field, A., Nemes, M., and Hilleman, M. (1967). Inducers of interferon and host resistance, I. Double-stranded RNA from extracts of *Penicillium funiculosum*. *Proc Natl Acad Sci U S A* **58,** 782 - 789.

Lim, W., Jeong, J., Jeong, R., Yoo, Y., Yie, S., and Kim, K. (2005). Complete nucleotide sequence and genome organisation of a dsRNA partitivirus infecting *Pleurotus ostreatus*. *Virus Res.* **108,** 111 - 119.

Linder-Basso, D., Dynek, J., and Hillman, B. (2005). Genome analysis of *Cryphonectria hypovirus 4*, the most common hypovirus species in North America. *Virology* **337**(1), 192 - 203.

Liu, X., and Gorovsky, M. (1993). Mapping the 5' and 3' ends of *Tetrahymena thermophila* mRNAs using RNA ligase mediated amplification of cDNA ends (RLM-RACE). *Nucleic Acids Res.* **21**(21), 4954 - 4960.

Macbeth, K., and Patterson, J. (1998). Overview of the *Leishmaniavirus* endoribonuclease and functions of other endoribonucleases affecting viral gene expression. *J. Exp. Zoology* **282,** 254 - 260.

Marienfeld, J., Unseld, M., Brandt, P., and Brennicke, A. (1997). Viral nucleic acid sequence transfer between fungi and plants. *Trends in Genet.* **13**(7), 260 - 261.

Markine-Goriaynoff, N., Gillet, L., van Etten, J., Korres, H., Verma, N., and Vanderplasschen, A. (2004). Glucosyltransferases encoded by viruses. *J. Gen. Virol.* **85,** 2741 - 2754.

Matthews, R., and Hull, R. (2001). "Matthew's Plant Virology." 4th ed. Academic Press, San Diego, CA.

Mayo, M. (2002). A summary of taxonomic changes recently approved by ICTV. *Arch. Virol.* **147,** 1655 - 1656.

McCabe, P., Pfeiffer, P., and Van Alfen, N. (1999). The influence of dsRNA viruses on the biology of plant pathogenic fungi. *Trends in Micr.* 7(9), 377 - 381.

Mel'nichuk, M., and Spiridonov, V. (2005). Molecular identity of double-stranded RNA-elements of viral nature isolated from the sugar beet. *Mikrobiol Z.* **67**(2), 55 - 62.

Mertens, P. (2004). The dsRNA viruses. *Virus Res.* **101,** 3 - 13.

Mikkelsen, J., and Pedersen, F. (2000). Genetic reassortment and patch repair by recombination in retroviruses. *J. Biomed. Sci.* **7**(2), 77 - 99.

Miyazaki, S., Iwabuchi, K., Pak, J., Fukuhara, T., and Nitta, T. (1996). Selective occurrence of endogenous double-stranded RNAs in insects. *Insect Biochem. Molec. Biol.* **26**(8), 995 - 961.

Moriyama, H., Horiuchi, H., Koga, R., and Fukuhara, T. (1999). Molecular characterization of two endogenous double-stranded RNAs and their inheritance by interspecific hybrids. *J. Biol. Chem.* **274**(11), 6882 - 6888.

Nagy, P., and Simon, A. (1997). New insights into the mechanisms of RNA recombination. *Virology* **235**, 1 - 9.

Nogawa, M., Kageyama, T., Nakatani, A., Taguchi, G., Shimosaka, M., and Okazaki, M. (1996). Cloning and characterization of mycovirus double-stranded RNA from the plant pathogenic fungus, *Fusarium solani* f. sp. *robiniae. Biosci. Biotech. Biochem.* **60**(5), 784 - 788.

Nomura, K., Osaki, H., Iwanami, T., Matsumoto, N., and Ohtsu, Y. (2003). Cloning and characterization of a totivirus double-stranded RNA from the plant pathogenic fungus, *Helicobasidium mompa* Tanaka. *Virus Genes* **26**(3), 219 - 226.

Nuss, D. and Koltin, Y. (1990). Significance of dsRNA genetic elements in plant pathogenic fungi. *Annu. Rev. Phytopathol.* **28**, 37 - 58.

Nuss, D. (1992). Biological control of chestnut blight: an example of virus-mediated attenuation of fungal pathogenesis. *Microbiol. Rev.* **56**(4), 561 - 576.

Nuss, D. (1996). Using hypoviruses to probe and pertrub signal transduction processes underlying fungal pathogenesis. *Plant Cell* **8**, 1845 - 1853.

Nuss, D. (1999). Hypoviruses (*Hypoviridae*). *In* "Encyclopedia of Virology", Vol. 2004. Academic Press.

Ogg, M., Carrion, R. J., Botelho, A., Mayrink, W., Correra-Oliveira, R., and Patterson, J. (2003). Short report: quantification of Leishmania virus RNA in clinical samples and its possible role in pathogenesis. *Am. J. Trop. Med. Hyg.* **69**(3), 309 - 313.

Oh, C., and Hillman, B. (1995). Genome organisation of a partitivirus from the filamentous ascomycete *Atkinsonella hypoxylon*. *J. Gen. Virol.* **76,** 1461 - 1470.

Osaki, H., Kudo, A., and Ohtsu, Y. (1998). Nucleotide sequence of seed- and pollen-transmitted double-stranded RNA, which encodes a putative RNA-dependent RNA polymerase, detected from Japanese pear. *Biosci. Biotech. Biochem.* **62**(11), 2101 - 2106.

Otrosina, W., Chase, T., Cobb, F. J., and Korhonen, K. (1993). Population structure of *Heterobasidion annosum* from North America and Europe. *Can. J. Bot.* **71,** 1064 - 1071.

Page, R.D.M. (1996). TREEVIEW: A n application to display phylogenetic trees on personal computers. *Computer appl. Biosc.* **12**, 357 – 358.

Park, C., Lopinski, J., Masuda, J., Tzeng, T., and Bruenn, J. (1996). A second double-stranded RNA virus from yeast. *Virology* **216,** 451 - 454.

Park, Y., Delano, J., and Punja, Z. (2005). Co-infection by two distinct totivirus-like double-stranded RNA elements in *Chalara elegans* (*Thielaviopsis basicola*). *Virus Res.* **109,** 71 - 85.

Payvar, F., and Schimke, R. (1979). Methylmercury hydroxide enhancement of translation and transcription of ovalbumin and conalbumin mRNA's. *J. Biol. Chem.* **254**(16), 7636 - 7642.

Peever, T., Liu, Y., Cortesi, P., and Milgroom, M. (2000). Variation in tolerance and virulence in the chestnut blight fungus-hypovirus interaction. *Appl. Environ. Microbiol.* **66**(11), 4863 - 4869.

Pena, M., Gago, S., and Flores, R. (2003). Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO* **22**(20), 5561 - 5570.

Pfeiffer, I., Kuscera, J., Varga, J., Parducz, A., and Ferenczy, L. (1996). Variability and inheritance of double-stranded RNA viruses in *Phaffia rhodozyma*. *Curr. Genet.* **30,** 294 - 297.

Pfeiffer, P. (1998). Nucleotide sequence, genetic organization and expression strategy of the double-stranded RNA associated with the '447' cytoplasmic male sterility trait in *Vicia faba*. *J. Gen. Virol.* **79,** 2349 - 2358.

Pratt, J., Gibbs, J., and Webber, J. (1999). Registration of *Phlebiopsis gigantea* as a forest biocontrol agent in the UK: recent experience. *Bioc. Sci. Technol.* **9**(1), 113 - 118.

Preisig, O., Moleleki, N., Smit, W., Wingfield, B., and MJ, W. (2000). A novel RNA mycovirus in a hypovirulent isolate of the plant pathogen *Diaporthe ambigua*. *J. Gen. Virol.* **81,** 3107 - 3114.

Preisig, O., Wingfield, B., and Wingfield, M. (1998). Coinfection of a fungal pathogen by two distinct double-stranded RNA viruses. *Virology* **252,** 399 - 406.

Reinisch, K. (2002). The dsRNA *Viridae* and their catalytic capsids. *Nature Struct. Biol.* **9**(10), 714 - 716.

Ribas, J., Fujimura, T., and Wickner, R. (1994). Essential RNA binding and packaging domains of the Gag-Pol fusion protein of the L-A double-stranded RNA virus of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **269**(45), 28420 - 28428.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European biology open software suite. *Trends in Genet.* **16**(6), 276 - 277.

Rodriguez-Cousino, N., Solorzano, A., Fujimura, T., and Esteban, R. (1998). Yeast positive-stranded virus-like RNA replicons. 20 S and 23 S RNA terminal nucleotide sequences and 3' end secondary structures resemble those of RNA coliphages. *J. Biol. Chem.* **273**(32), 20363 - 20371.

Romaine, C., Uhlirch, P., and Schlagnhaufer, B. (1993). Transmission of La France isometric virus during basidiosporogenesis in *Agaricus bisporus*. *Mycologia* **85**(2), 175 - 179.

Rong, R., Rao, S., Scott, S., Carner, G., and Tainter, F. (2002). Complete sequence of the genome of two dsRNA viruses from *Discula destructiva*. *Virus Res.* **90,** 217 - 224.

Rosewich, U., and Kistler, H. (2000). Role of horizontal gene transfer in the evolution of fungi. *Annu. Rev. Phytopathol.* **38,** 325 - 363.

Routhier, E., and Bruenn, J. (1998). Functions of conserved motifs in the RNA-dependent RNA polymerase of a yeast double-stranded RNA virus. *J. Virol.* **72**(5), 4427 - 4429.

Roy, G., Laflamme, G., Bussieres, G., and Dessureault, M. (2003). Filed tests on biological control of *Heterobasidion annosum* by *Phaeotheca dimorphospora* in comparison with *Phlebiopsis gigantea. For. Path.* **33,** 127 - 140.

Sasaki, A., Kanematsu, S., Onoue, M., Oyama, Y., and Yoshida, K. (2005). Infection of *Rosellinia necatrix* with purified viral particles of a member of *Partitiviridae* (RnPV1-W8). *Arch. Virol.*

Sasaki, A., Miyanishi, M., Ozaki, K., Onoue, M., and Yoshida, K. (2005). Molecular characterization of a partitivirus from the plant pathogenic ascomycete *Rosellinia necatrix. Arch. Virol.* **150**(6), 1069 - 1083.

Satyanarayana, T., Gowda, S., Boyko, V., Albiach-Marti, M., Mawassi, M., Navas-Castillo, J., Karasev, A., Dolja, Vainio, E., Hilf, M., Lewandowski, D., Morenos, P., Bar-Joseph, M., Garnsey, S., and Dawson, W. (1999). An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. *Proc Natl Acad Sci U S A* **96,** 7433 - 7438.

Satyanarayana, T., Gowda, S., Ayllon, M., and Dawson, W. (2003). Frameshift mutations in infectious cDNA clones of *Citrus tristeza virus*: a strategy to minimize the toxicity of viral sequences to *Escherichia coli. Virology* **313,** 481 - 491.

Scheffter, S., Ro, Y., Chung, I., and Patterson, J. (1995). The complete sequence of Leishmania RNA virus LRV2-1, a virus of an old world parasite strain. *Virology* **212,** 84 - 90.

Scheffter, S., Widmer, G., and Patterson, J. (1994). Complete sequence of Leishmania RNA virus 1-4 and identification of conserved sequences. *Virology* **199,** 479 - 483.

Schmitt, M., and Neuhausen, F. (1994). Killer toxin-secreting double-stranded RNA mycoviruses in the yeasts *Hanseniaspora uvarum* and *Zygosaccharomyces bailii*. *J. Virol.* **68**(3), 1765 - 1772.

Shapira, R., Ghoi, G., and Nuss, D. (1991). Virus-like genetic organisation and expression strategy for a double-stranded RNA genetic element associated with biological control of chestnut blight. *EMBO* **10**, 731 - 739.

Smart, C., Yuan, W., Foglia, R., Nuss, D., Fulbright, D., and Hillman, B. (1999). *Cryphonectria hypovirus* 3, a virus species in the family Hypoviridae with a single open reading frame. *Virology* **265**, 66 - 73.

Soldevila, A., and Ghabrial, S. (2000). Expression of te totivirus *Helminthosporium victoriae 190S* virus RNA-dependent RNA polymerase from its downstream open reading frame in dicistronic constructs. *J. Virol.* **74**(2), 997 - 1003.

Soldevila, A., Havens, W., and Ghabrial, S. (2000). A cellular protein with an RNA-binding activity co-purifies with viral dsRNA from mycovirus-infected Helminthosporium victoriae. *Virology* **272**(1), 183-90.

Soldevila, A., Huang, S., and Ghabrial, S. (1998). Assembly of the Hv190S totivirus capsid is indeendent of posttranslational modification of the capsid protein. *Virology* **251**, 327 - 333.

Sommer, N., Depping, R., Piotrowski, M., and Ruger, W. (2004). Bacteriophage T4 α-glucosyltransferase: a novel interaction with gp45 and aspects of the catalytic mechanism. *Bioch. Bioph. Res. Comm.* **323**, 809 - 815.

Stenlid, J., and Waesterlund, I. (1986). Estimating the frequency of stem rot in *Picea abies* using an increment borer. *Scand. J. Forest Res.* **1**, 303 - 308.

Stuart, K., Weeks, R., Guilbride, L., and Myler, P. (1992). Molecular organization of Leishmania RNA virus 1. *Proc Natl Acad Sci USA* **89**(18), 8596 - 8600.

Suzaki, K., Ikeda, K., Sasaki, A., Kanematsu, S., Matsumoto, N., and Yoshida, K. (2005). Horizontal transmission and host-virulence attenuation of totivirus in violet root rot fungus *Helicobasidium mompa*. *J. Gen. Plant Pathol.* **71,** 161 - 168.

Suzuki, N., Supyani, S., Maruyama, K., and Hillman, B. (2004). Complete genome sequence of *Mycoreovirus-1*/Cp9B21, a member of a novel genus within the family *Reoviridae*, isolated from the chestnut blight fungus *Cryphonectria parasitica*. *J. Gen. Virol.* **85,** 3437 - 3448.

Sweigard, J., Carroll, A., Farrall, L., Chumley, F., and Valent, B. (1998). Magnaporthe grisea pathogenicity genes obtained through insertional mutagenesis. *Mol. Plant Microbe Interact.* **11,** 404 - 412.

Tai, J., Chang, S., Ip, C., and Ong, S. (1995). Identification of a satellite double-stranded RNA in the parasitic protozoan *Trichomonas vaginalis* infected with *T.vaginalis* virus T1. *Virology* **208,** 189 - 196.

Thompson, J.D., Gibson, T., Plewniak, F., Jeanmougin, F., and Higgins, D. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**(24), 4867 - 4882.

Toh-E, A., Guerry, P., and Wickner, R. (1978). Chromosomal superkiller mutants of *Saccharomyces cerevisiae*. *J. Bacteriol.* **136**(3), 1002 - 1007.

Tsai, P., Pearson, M., and Beever, R. (2004). Mycoviruses in *Monilinia fructicola*. *Mycol. Res.* **108**(8), 907 - 912.

Tuomivirta, T., and Hantula, J. (2003). Two unrelated double-stranded RNA molecule patterns in *Gremmeniella abietina* type A code for putative viruses of the families *Totiviridae* and *Partitiviridae*. *Arch. Virol.* **148**(12), 2293 - 2305.

Tuomivirta, T., and Hantula, J. (2005). Three unrelated viruses occir in a single isolate of *Gremmeniella abietina* var. *abietina* type A. *Virus Res.* **110**(1-2), 31 - 39.

Tuomivirta, T., Uotila, A., and Hantula, J. (2002). Two independent double-stranded RNA patterns occur in the Finnish *Gremmeniella abietina* var. *abietina* type A. *For. Path.* **32,** 197 - 205.

Tzanetakis, I., Keller, K., and Martin, R. (2005). The use of reverse transcriptase for efficient first- and second-strand cDNA synthesis from single-a dn double-stranded RNA templates. *J. Virol. Meth.* **124,** 73 - 77.

Vainio, E., and Hantula, J. (2000). Genetic differentiation between European and North American populations of *Phlebiopsis gigantea. Mycologia* **92**(3), 436 - 446.

Valverde, R., Nameth, S., Abdallha, O., Al-musa, O., Desjardins, P., and Dodds, A. (1990). Indigenous double-stranded RNA from pepper (*Capsicum annuum*). *Plant Sci.* **67,** 195 - 201.

van der Lende, T., Duitman, E., Gunnewijk, M., Yu, L., and Wessels, J. (1996). Functional analysis of dsRNAs (L1, L3, L5 and M2) associated with isometric 34-nm virions of *Agaricus bisporus* (White button mushroom). *Virology* **217,** 88 - 96.

van Diepeningen, A., Debets, A., and Hoekstra, R. (1998). Intra- and interspecies virus transfer in Aspergilli via protoplast fusion. *Fungal Genet. Biol.* **25**(3), 171 - 180.

Vilches, S., and Castillo, A. (1997). A double-stranded RNA mycovirus in *Botrytis cinerea. FEMS Microbiol. Lett.* **155,** 125 -130.

Wakarchuk, W., and Hamilton, R. (1990). Partial nucleotide sequence from enigmatic dsRNAs in *Phaseolus vulgaris. Plant Mol. Biol.* **14**(4), 637 - 639.

Wang, A., and Wang, C. (1999). Giardiaviruses (*Totiviridae*). *In* "Encyclopedia of Virology", Vol. 2004. Academic Press.

Wang, A., Yang, H., Shen, K., and Wang, C. (1993). Giardiavirus double-stranded RNA genome encodes a capsid polypeptide and a gag-pol-like fusion protein by a translational frameshift. *Proc Natl Acad Sci U S A* **90,** 8595 - 8599.

Warnecke, D., Erdmann, R., Fahl, A., Hube, B., Mueller, F., Zank, T., Zaehringer, U., and Heinz, E. (1999). Cloning and functional expression of *UGT* genes encoding sterol glucosyltransferases from *Saccharomyces cerevisiae*, *Candida albicans*, *Pichia pastoris*, and *Dictiostelium discoideum*. *J. Biol. Chem.* **274**(19), 13048 - 13059.

Wei, C., Osaki, H., Iwanami, T., Matsumoto, N., and Ohtsu, Y. (2004). Complete nucleotide sequences of genome segments 1 and 3 of Rosellinia anti-rot virus in the family *Reoviridae*. *Arch. Virol.* **149**, 773 - 777.

Wickner, R. (1993). Double-stranded RNA virus replication and packaging. *J. Biol. Chem.* **268**(6), 3797 - 3800.

Wickner, R. (1996). Double-stranded RNA viruses of *Saccharomyces cerevisiae*. *Microbiol. Rev.* **60**(1), 250 - 265.

Yao, W., Muqtadir, K., and Bruenn, J. (1995). Packaging in a yeast double-stranded RNA virus. *J. Virol.* **69**(3), 1917 - 1919.

Zamora, M., Guilbride, L., Sacks, L., and Stuart, K. (2000). Phylogenetic analysis of the 5' subterminal region of isolates of Leishmania RNA virus-1. *An. Trop. Med. Parasitol.* **94**(2), 123 - 133.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**(13), 3406 - 3415.

B. References for chapter six.

Aasland, R.; Stewart, A. (1999). Analysis of DNaseI hypersensitive sites in chromatin by cleavage in permeabilized cells. *Methods Mol Biol.* **119**, 355 - 362.

Adams, M.; *et al.* (2000). The genome sequence of Drosophila melanogaster. *Science* **278**, 2185 - 2195.

Azizkhan, J.; Vaughn, J.; Christy, R. and Hamlin, J. (1986). Nucleotide sequence and nuclease hypersensitivity of the Chinese hamster dihydrofolate reductase gene promoter region. *Biochemistry* **25**, 6228 - 6236.

Barges, S.; Mihaly, J.; Galloni, M.; Hagstrom, K.; Muller, M.; Shanower, G.; Schedl, P.; Gyurkovics, H. and Karch, F. (2000). The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. *Developmnet* **127**, 779 - 790.

Bellard, M.; Dretzen, G.; Bellard, F.; Kaye, J.; Pratt-Kaye, S. and Chambon, P. (1986). Hormonally induced alterations of chromatin structure in the polyadenylation and transcription termination regions of the chicken ovalbumin gene. *EMBO* **5**, 567 - 574.

Bender, M.; Mehaffey, M.; Telling, A.; Hug, B.; Ley, T.; Groudine, M. and Fiering, S. (2000). Independent formation of DnaseI hypersensitive sites in the murine beta-globin locus control region. *Blood* **95**, 3600 - 3604.

Bonifer, C. (1999). Long-distance chromatin mechanisms controlling tissue-specific gene locus activation. *Gene* **238**, 277 - 289.

Bonifer, C. (2000). Developmental regulation of eukaryotic gene loci: which cis-regulatory information is required? *Trends Genet.* **16**, 310 - 315.

Burch, J. and Weintraub, H. (1983). Temporal order of chromatin structural changes associated with activation of the major chicken vitellogenin gene. *Cell* **33**, 65 - 76.

Bushel, P.; Rego, K.; Mendelsohn, L. and Allan, M. (1990). Correlation between patterns of DNase I-hypersensitive sites and upstream promoter activity of the human epsilon-globin gene at different stages of erythroid development. *Mol. Cell. Biol.* **10**, 1199 - 1208.

Cao, A. and Moi, P. (2002). Regulation of the globin genes. *Pediatr Res.* **51**, 415 - 421.

Carey, M., and Smale, S. (2000). "Transcriptional regulation in eukaryotes: concepts, strategies, and techniques." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Carrier, F.; Georgel, P.; Pourquier, P.; Blake, M.; Kontny, H.; Antinore, M.; Gariboldi, M.; Myers, T.; Weinstein, J.; Pommier, Y. and Fornace, A. J. (1999). Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol. Cell. Biol.* **19**, 1673 - 1685.

Cartwright, I. (1987). Developmental switch in chromatin structure associated with alternate promoter usage in the Drosophila melanogaster alcohol dehydrogenase gene. *EMBO* **6**, 3097 - 3101.

Cartwright, I. L. and Elgin, S. C. (1986). Nucleosomal instability and induction of new upstream protein-DNA association accompany activation of four small heat shock protein genes in *Drosophila melanogaster*. *Mol. Cell Biol.* **6**, 779 - 791.

Chen, S. and Corces, V. (2001). The gypsy insulator of Drosophila affects chromatin strucutre in a directional manner. *Genetics* **159**, 1649 - 1658.

Cissell, M. and Chalkey, R. (1999). Characterisation of a kidney specific pattern of chromatin structure in the rat phosphoenolpyruvate carboxykinase gene. *Biochim. Biophys. Acta* **1445**, 229 - 313.

Costlow, N. and Lis, J. (1984). High-resolution mapping of DNase I-hypersensitive sites of Drosophila heat shock genes in *Drosophila melanogaster* and *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **4**, 1853 - 1863.

Costlow, N.; Simon, J. and Lis, J. (1985). A hypersensitive site in hsp70 chromatin requires adjacent not internal DNA sequence. *Nature* **313**, 147 - 149.

Crawford, G.; Holt, I.; Mullikin, J.; Tai, D.; Blakesley, R.; Bouffard, G.; Young, A.; Masiello, C.; Green, E.; Wolfsberg, T.; Collins, F. and Center., N. I. O. H. I. S. (2004). Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl. Acad. Sci. USA* **101**, 992 - 997.

Crowder, C. and Merlie, J. (1986). DNase I hypersensitive sites surround the mouse acetylcholne receptor delta subunit-gene. *Proc. Natl. Acad. Sci. USA* **83**, 8405 - 8409.

Cuvier, O.; Hart, C.; Kas, E. and Laemmli, U. (2002). Identification of a multicopy chromatin boundary element at the borders of silenced chromatin domains. *Chromosoma* **110**, 519 - 531.

Cuvier, O.; Hart, C. and Laemmli, U. (1998). Identification of a class of chromatin boundary elements. *Mol. Cell. Biol.* **18**, 7478 - 7486.

De Lotto, R. and Schedl, P. (1984). A *Drosophila melanogaster* transfer RNA gene cluster at the cytogenetic locus 90BC. *J. Mol. Biol.* **179**, 587 - 605.

De Lotto, R. and Schedl, P. (1984). Internal promoter elements of transfer RNA genes are preferentially exposed in chromatin. *J. Mol. Biol.* **179**, 607 - 628.

De Val, S.; Pontikos, M.; Antoniv, T.; Wells, D.; Abraham, D.; Partridge, T. and Bou-Gharios, G. (2002). Identification of key regions within the mouse Pro-a2(I) collagen gene far-upstream enhancer. *J. Biol. Chem.* 277, 9286 - 9292.

Decoville, M.; Giacomello, E.; Leng, M. and Locker, D. (2001). DSP1, an HMG-like protein, is involved in the regulation of homeotic genes. *Genetics* **157**, 237 - 244.

Drew, H. (1984). Structural specificities of five commonly used DNA nucleases. *J. Mol. Biol.* **176**, 535 - 557.

Duanaway, M.; Hwang, J.; Xiong, M. and Yuen, H. (1997). The activity of scs and scs' insulator elements is not dependent on chromosomal context. *Mol. Cell. Biol.* **17**, 182 - 189.

Dubrovsky, E.; Dretzen, G. and Bellard, M. (1994). The Drosophila Broad-complex regulates developmental changes in transcription and chromatin structure of the 67B heat-shock gene cluster. *J. Mol. Biol.* **241**, 353 - 362.

Dubrovsky, E.; Dretzen, G. and Berger, E. (1996). The Broad-complex gene is a tissue specific modulator of the ecdysone response of the Drosophila hsp23 gene. *Mol. Cell. Biol.* **16**, 6542 - 6552.

Dubrovsky, E.; Dubrovskaya, V. and Berger, E. (2001). Selective binding of Drosophila BR-C isoforms to a distal regulatory element in the hsp23 promoter. *Insect Biochem. Mol. Biol.* **31**, 1231 - 1239.

Eissenberg, J.; Kimbrell, D.; Fristrom, J. and Elgin, S. (1984). Chromatin structure at the 44D larval cuticle gene locus in Drosophila: the effect of a transposable element insertion. *Nucleic Acids Res.* **12**, 9025 - 9038.

Ercan, S. and Simpson, R. (2004). Global chromatin structure of 45000 base pairs of chromosome III in a- and α-cell yeast and during mating-type switching. *Mol. Cell. Biol.* **24**, 10026 - 10035.

Farkas, G.; Leibovitch, B. and Elgin, S. (2000). Chromatin organisation and transcriptional control of gene expression in Drosophila. *Gene* **253**, 117 - 136.

Farrell, C.; West, A. and Felsenfeld, G. (2002). Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol. Cell. Biol.* **22**, 3820 - 3831.

Feinberg, A., and Vogestein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**(1), 6 - 13.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. *Nature* **355**, 219 - 224.

Felsenfeld, G.; Boyes, J.; Chung, J.; Clark, D. and Studitsky, V. (1996). Chromatin structure and gene expression. *Proc. Natl. Acad. Sci. USA* **93**, 9384 - 9388.

Feng, J.; Irving, J. and Villeponteau, B. (1991). A phosphatase inhibitor enhances the *DNase* I sensitivity of active chromatin. *Biochemistry* **30**, 4747 - 4752.

Galloni, M.; Gyurkovics, H.; Schedl, P. and Karch, F. (1993). The bluetail transposon: evidence for independent *cis*-regulatory domains and domain boundaries in the bithorax complex. *EMBO* **12**, 1087 - 1097.

Georgel, P.; Dretzen, G.; Jagla, K.; Bellard, F.; Dubrovsky, E.; Calco, V. and Bellard, M. (1993). GEBF-I activates the Drosophila Sgs-3 gene enhancer by altering a positioned nucleosomal core particle. *J. Mol. Biol.* **234**, 319 - 330.

Georgel, P.; Ramain, P.; Giangrande, A.; Dretzen, G.; Richards, G. and Bellard, M. (1991). Sgs-3 chromatin structure and trans-activators: developmental and ecdysone induction of a glue enhancer-binding factor, GEBF-I, in Drosophila larvae. *Mol. Cell. Biol.* **11**, 523 - 532.

Goetze, S.; Baer, A.; Winkelman, S.; Nehlsen, K.; Seibler, J.; Maass, K. and Bode, J. (2005). Performance of genomic bordering elements at predefined genomic loci. *Mol. Cell. Biol.* **25**, 2260 - 2272.

Goodwin, G.; Nicolas, R.; Cockerill, P.; Zavou, S. and Wright, C. (1985). The effect of salt extraction on the structure of transcriptionally active genes; evidence for a DNseI-sensitive structure which could be dependent on chromatin structure at levels higher than the 30 nm fibre. *Nucleic Acids Res.* **13**, 3561 - 3579.

Gottesdiener, K.; Karpinski, B.; Lindsten, T.; Strominger, J.; Jones, N.; Thompson, C. and Leiden, J. (1988). Isolation and structural characterization of the human 4F2 heavy-chain gene, an inducible gene involved in T-lymphocyte activation. *Mol. Cell. Biol.* **8**, 3809 - 3819.

Grandien, K.; Berkenstam, A. and Gustafsson, J. (1997). The estrogen receptor gene: promoter organization and expression. *Insect Biochem. Mol. Biol.* **29**, 1343 - 1369.

Gross, D. and Garrard, W. (1988). Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem.* **57**, 159 - 197.

Groudine, M. and Weintraub, H. (1982). Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: a possible mechanism for determination. *Cell* **30**, 131 - 139.

Hagstrom, K.; Muller, M. and Schedl, P. (1997). A Polycomb and GAGA dependent silencer adjoins the Fab-7 boundary in the Drosophila bithorax complex. *Genetics* **146**, 1365 - 1380.

Halfon, M.; Grad, Y.; Church, G. and Michelson, A. (2002). Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res.* **12**, 1019 - 1028.

Hardison, R. (2000). Conserved noncoding sequences are reliable guides to regulatory elements. *Trends Genet.* **16**, 369 - 372.

Huang, W. and Liew, C. (1998). Chromatin remodelling of the cardiac beta-myosin heavy chain gene. *Biochem. J* **330**, 871 - 876.

Ishihara, S. and Morohashi, K. (2005). A boundary for histone acetylation allows distinct expression patterns of the Adb4BP/SF-1 and GCNF loci in adrenal cortex cells. *Biochem Biophys Res Commun.* **329**, 554 - 562.

Ishii, H.; Sen, R. and Panzin, M. (2004). Combinatorial control of DNase I-hypersensitive site formation and erasure by immunoglobulin heavy chain enhancer-binding proteins. *J. Biol. Chem.* **279**, 7331 - 7338.

Jack, R.; Moritz, P. and Cremer, S. (1991). Restriction enzymes permit quantitative determination of defined chromatin structures within the chromosome. *Eur J Biochem.* **202**, 441 - 446.

Jupe, E.; Sinden, R. and Cartwright, I. (1995). Specialized chromatin structure domain boundary elements flanking a Drosophila heat shock gene locus are under torsional strain in vivo. *Biochemistry* **34**, 2628 - 2633.

Kageyama, Y.; Mengus, G.; Gilfillan, G.; Kennedy, H.; Stuckenholz, C.; Kelley, R.; Becker, P. and Kuroda, M. (2001). Association and spreading of the Drosophila dosage compensation complex from a discrete roX1 chromatin entry site. *EMBO* **20**, 2236 - 2245.

Karch, F.; Galloni, M.; Sipos, L.; Gausz, J.; Gyurkovics, H. and Schedl, P. (1994). Mcp and Fab-7: molecular analysis of putative boundaries of cis-regulatory domains in the bithorax complex of Drosophila melanogaster. *Nucleic Acids Res.* **22**, 3138 - 3146.

Keene, M.; Corces, V.; Lowenhaupt, K. and Elgin, S. (1981). DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription. *Proc. Natl. Acad. Sci. USA* **78**, 143 - 146.

Kelly, S. and Cartwright, I. (1989). Perturbation of chromatin architecture on ecdysterone induction of Drosophila melanogaster small heat shock protein genes. *Mol. Cell. Biol.* **9**, 332 - 335.

Kemper, B.; Jackson, P. and Felsenfeld, G. (1987). Protein-binding sites within the 5' DNase I-hypersensitive region of the chicken alpha D-globin gene. *Mol. Cell. Biol.* **7**, 2059 - 2069.

Kieffer, L.; Greally, J.; Landres, I.; Nag, S.; Nakajima, Y.; Kohwi-Shigematsu, T. and Kavathas, P. (2002). Identification of a candidate regulatory region in the human CD8 gene complex by colocalization of DNase I hypersensitive sites and matrix attachment regions which bind SATB1 and GATA-3. *J. Immunol.* **168**, 3915 - 3922.

Kim, A. and Murray, V. (2001). Chromatin structure at the 3'-boundary of the human beta-globin locus control region hypersensitive site-2. *Int J Biochem Cell Biol.* **33**, 1183 - 1192.

Kontaraki, J.; Chen, H.; Riggs, A. and Bonifer, C. (2000). Chromatin fine structure profiles for a developmentally regulated gene: reorganization of the lysozyme locus before trans-activator binding and gene expression. *Genes Dev.* **14**, 2106 - 2122.

Krebs, J. and Dunaway, M. (1998). The scs and scs' insulator elements impart a cis requirement on enhancer-promoter interactions. *Mol. Cell* **1**, 301 - 308.

Kuo, S.; Chesrown, S.; Mellott, J.; Rogers, R.; Hsu, J. and Nick, H. (1999). *In vivo* architecture of the manganese superoxide dismutase promoter. *J. Biol. Chem.* **274**, 3345 - 3354.

Laemmli, U. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227(5259), 680 - 685.

Laemmli, U.; Kas, E.; Poljak, L. and Adachi, Y. (1992). Scaffold-associated regions: cis-acting determinants of chromatin structural loops and functional domains. *Curr Opin Genet Dev.* **2**, 275 - 285.

Leach, K.; Nightingale, K.; Igarashi, K.; Levings, P.; Engel, J.; Becker, P. and Bungert, J. (2001). Reconstitution of human beta-globin locus control region hypersensitive sites in the absence of chromatin assembly. *Mol. Cell. Biol.* **21**, 2629 - 2640.

Lee, C.; Shibata, Y.; Rao, B.; Strahl, B. and Lieb, J. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics* **36**, 900-905.

Lee, G.; Fields, P. and Flavell, R. (2001). Regulation of IL-4 gene expression by distal regulatory elements and GATA-3 at the chromatin level. *Immunity* **14**, 447 - 459.

Leibovitch, B.; Lu, Q.; Benjamin, L.; Liu, Y.; Gilmour, D. and Elgin, S. (2002). GAGA factor and the TFIID complex collaborate in generating an open chromatin structure at the Drosophila melanogaster hsp26 promoter. *Mol. Cell. Biol.* **22**, 6148 - 6157.

Li, G.; Lim, K.; Engel, J. and Bungert, J. (1998). Individual LCR hypersensitive sites cooperate to generate an open chromatin domain spanning the human beta-globin locus. *Genes Cells* **3**, 415 - 429.

Li, Y.; Strahler, J. and Dodgson, J. (1997). Neither HMG-14a nor HMG-17 gene function is required for growth of chicken DT40 cells or maintenance of DNaseI-hypersensitive sites. *Nucleic Acids Res.* 25, 283 - 288.

Liang, G.; Gonzalgo, M.; Salem, C. and Jones, P. (2002). Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* **27**, 150 - 155.

Liu, Z.; George-Raizen, J.; Li, S.; Meyers, K.; Chang, M. and Garrard, W. (2002). Chromatin structural analyses of the mouse Igkappa gene locus reveal new hypersensitive sites specifying a transcriptional silencer and enhancer. *J. Biol. Chem.* **277**, 32640 - 32649.

Loots, G.; Locksley, R.; Blankespoor, C.; Wang, Z.; Miller, W.; Rubin, E. and Frazer, K. (2000). Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136 - 140.

Lowenhaupt, K.; Cartwright, I.; Keene, M.; Zimmerman, J. and Elgin, S. (1983). Chromatin structure in pre- and postblastula embryos of Drosophila. *Dev Biol.* **99**, 194 - 201.

Lu, Q.; Wallrath, L.; Allan, B.; Glaser, R.; Lis, J. and Elgin, S. (1992). Promoter sequence containing (CT)n.(GA)n repeats is critical for the formation of the DNase I hypersensitive sites in the Drosophila hsp26 gene. *J. Mol. Biol.* **225**, 985 - 998.

Lu, Q.; Wallrath, L.; Granok, H. and Elgin, S. (1993). (CT)n (GA)n repeats and heat shock elements have distinct roles in chromatin structure and transcriptional activation of the Drosophila hsp26 gene. *Mol. Cell. Biol.* **13**, 2802 - 2814.

McGinnis, W.; Shermoen, A. and Beckendorf, S. (1983). A transposable element inserted just 5' to a Drosophila glue protein gene alters gene expression and chromatin structure. *Cell* **34**, 75 - 84.

McNabb, S. and Beckendorf, S. (1986). Cis-acting sequences which regulate expression of the Sgs-4 glue protein gene of Drosophila. *EMBO* **5**, 2331 - 2340.

Mishra, R.; Mihaly, J.; Barges, S.; Spierer, A.; Karch, F.; Hagstrom, K.; Schweinsberg, S. and Schedl, P. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol. Cell. Biol.* **21**, 1311 - 1318.

Mollers, B.; Klages, S.; Wedel, A.; Cross, M.; Spooncer, E.; Dexter, T. and Renkawitz, R. (1992). The mouse M-lysozyme gene domain: identification of myeloid and differentiation specific DNaseI hypersensitive sites and of a 3'-cis acting regulatory element. *Nucleic Acids Res.* **20**, 1917 - 1924.

Montecino, M.; Pockwinse, S.; Lian, J.; Stein, G. and Stein, J. (1994). DNase I hypersensitive sites in promoter elements associated with basal and vitamin D dependent transcription of the bone-specific osteocalcin gene. *Biochemistry* **33**, 348 - 353.

Mucha, M.; Lisowska, K.; Goc, A. and Filipski, J. (2000). Nuclease-hypersensitive chromatin formed by a CpG island in human DNA cloned as an artificial chromosome in yeast. *J. Biol. Chem.* **275**, 1275 - 1278.

Muller, M.; Hagstrom, K.; Gyurkovics, H.; Pirrotta, V. and Schedl, P. (1999). The mcp element from the Drosophila melanogaster bithorax complex mediates long-distance regulatory interactions. *Genetics* **153**, 1333 - 1356.

Nekrutenko, A. and Li, W. (2000). Assessment of compositional heterogeneity within and between eukaryotic genomes. *Genome Res.* **10**, 1986 - 1995.

Nussinov, R.; Shapiro, B.; Lipkin, L. and Maizel, J. J. (1984). DNAase I hypersensitive sites may be correlated with genomic regions of large structural variation. *J. Mol. Biol.* **177**, 591 - 607.

Ortiz, B.; Cado, D.; Chen, V.; Diaz, P. and Winoto, A. (1997). Adjacent DNA elements dominantly restrict the ubiquitous activity of a novel chromatin-opening region to specific tissues. *EMBO* **16**, 5037 - 5045.

Ortiz, B.; Harrow, F.; Cado, D.; Santoso, B. and Winoto, A. (2001). Function and factor interactions of a locus control region element in the mouse T cell receptor-alpha/Dad1 gene locus. *J. Immunol.* **167**, 3836 - 3845.

Park, Y.; Mengus, G.; Bai, X.; Kageyama, Y.; Meller, V.; Becker, P. and Kuroda, M. (2003). Sequence-specific targeting of Drosophila roX genes by the MSL dosage compensation complex. *Mol. Cell* **11**, 977 - 986.

Parker, C. and Topol, J. (1984). A Drosophila RNA polymerase II transcription factor binds to the regulatory site of an hsp 70 gene. *Cell* **36**, 357 - 369.

Pelham, H. (1982). A regulatory upstream promoter element in the Drosophila hsp 70 heat-shock gene. *Cell* **30**, 517 - 528.

Pemov, A.; Bavykin, S. and Hamlin, J. (1995). Proximal and long-range alterations in chromatin structure surrounding the Chinese hamster dihydrofolate reductase promoter. *Biochemistry* **34**, 2381 - 2392.

Philipsen, S.; Talbot, D.; Fraser, P. and Grosveld, F. (1990). The beta-globin dominant control region: hypersensitive site 2. *EMBO* **9**, 2159 - 2167.

Pierce, J.; Gifford, A. and Baltimore, D. (1991). Silencing of the expression of the immunoglobulin kappa gene in non-B cells. *Mol. Cell. Biol.* **11**, 1431 - 1437.

Pile, L. and Cartwright, I. (2000). GAGA factor-dependent transcription and establishment of DNase hypersensitivity are independent and unrelated events *in vivo*. *J. Biol. Chem.* **275**, 1398 - 1404.

Ramain, P.; Bourouis, M.; Dretzen, G.; Richards, G.; Sobkowiak, A. and Bellard, M. (1986). Changes in the chromatin structure of Drosophila glue genes accompany developmental cessation of transcription in wild type and transformed strains. *Cell* **45**, 545 - 553.

Ramain, P.; Giangrande, A.; Richards, G. and Bellard, M. (1988). Analysis of a DNase I-hypersensitive site in transgenic Drosophila reveals a key regulatory element of Sgs3. *Proc. Natl. Acad. Sci. USA* **85**, 2718 - 2722.

Rebeiz, M.; Reeves, N. and Posakony, J. (2002). SCORE: A computational approach to the identification of cis-regulatory modules and target genes in whole-genome sequence data. *Proc. Natl. Acad. Sci. USA* **99**, 9888 - 9893.

Reik, A.; Schutz, G. and Stewart, A. (1991). Glucocorticoids are required for establishment and maintenance of an alteration in chromatin structure: induction leads to a reversible disruption of nucleosomes over an enhancer. *EMBO* **10**, 2569 - 2576.

Rowe, T.; Wang, J. and Liu, L. (1986). In vivo localization of DNA topoisomerase II cleavage sites on Drosophila heat shock chromatin. *Mol. Cell. Biol.* **6**, 985 - 992.

Sabo, P.; Hawrylycz, M.; Wallace, J.; Humbert, R.; Yu, M.; Shafer, A.; Kawamoto, J.; Hall, R.; Mack, J.; Dorschner, M.; McArthur, M. and Stamatoyannopoulos, J. (2004). Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc. Natl. Acad. Sci. USA* **101**, 16837 - 16842.

Sabo, P.; Humbert, R.; Hawrylycz, M.; Wallace, J.; Dorschner, M.; McArthur, M. and Stamatoyannopoulos, J. (2004). Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. *Proc. Natl. Acad. Sci. USA* **101**, 4537 - 4542.

Sackerson, C.; Fujioka, M. and Goto, T. (1999). The even-skipped locus is contained in a 16-kb chromatin domain. *Dev Biol.* **211**, 39 - 52.

Saitoh, N.; Bell, A.; Recillas-Targa, F.; West, A.; Simpson, M.; Pikaart, M. and Felsenfeld, G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO* **19**, 2315 - 2322.

Salimi-Tari, P.; Cheung, M.; Safar, C.; Tracy, J.; Tran, I.; Harbers, K. and Breindl, M. (1997). Molecular cloning and chromatin structure analysis of the murine alpha1(I) collagen gene domain. *Gene* **198**, 61 - 72.

Samal, B.; Worcel, A.; Louis, C. and Schedl, P. (1981). Chromatin structure of the histone genes of D. melanogaster. *Cell* **23**, 401 - 409.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). "Molecular cloning: a laboratory manual." Cold Spring Harbor Press, Cold Spring Harbor, NY.

Selleck, S.; Elgin, S. and Cartwright, I. (1984). Supercoil-dependent features of DNA structure at Drosophila locus 67B1. *J. Mol. Biol.* **178**, 17 -33.

Shermoen, A. and Beckendorf, S. (1982). A complex of interacting DNAase I-hypersensitive sites near the Drosophila glue protein gene, Sgs4. *Cell* **29**, 601 - 607.

Shewchuk, B.; Cooke, N. and Liebhaber, S. (2001). The human growth hormone locus control region mediates long-distance transcriptional activation independent of nuclear matrix attachment regions. *Nucleic Acids Res.* **29**, 3356 - 3361.

Siegfried, E.; Thomas, G.; Bond, U. and Elgin, S. (1986). Characterization of a supercoil-dependent S1 sensitive site 5' to the Drosophila melanogaster hsp 26 gene. *Nucleic Acids Res.* **14**, 9425 - 9444.

Stalder, J.; Larsen, A.; Engel, J.; Dolan, M.; Groudine, M. and Weintraub, H. (1980). Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* **20**, 451 - 460.

Sun, F.; Cuaycong, M. and Elgin, S. (2001). Long-range nucleosome ordering is associated with gene silencing in Drosophila melanogaster pericentric heterochromatin. *Mol. Cell. Biol.* **21**, 2867 - 2879.

Svetlova, E. ; Avril-Fournout, N. ; Ira, G. ; Deschavanne, P. and Filipski, J. (1998). DNase-hypersensitive sites in yeast artificial chromosomes containing human DNA. *Mol Gen Genet.* **257**, 292 - 298.

Talbot, D.; Philipsen, S.; Fraser, P. and Grosveld, F. (1990). Detailed analysis of the site 3 region of the human beta-globin dominant control region. *EMBO* **9**, 2169 - 2177.

Tanaka, H.; Zhao, Y.; Wu, D. and Hersh, L. (1998). The use of DNase I hypersensitivity site mapping to identify regulatory regions of the human cholinergic gene locus. *J. Neurochem.* **70**, 1799 - 1808.

Targa, F. ; Razin, S. ; de Moura Gallo, C. and Scherrer, K. (1994). Excision close to matrix attachment regions of the entire chicken alpha-globin gene domain by nuclease S1 and characterization of the framing structures. *Proc. Natl. Acad. Sci. USA* **91**, 4422 - 4426.

Tolhuis, B.; Palstra, R.; Splinter, E.; Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453 - 1465.

Udvardy, A.; Louis, C.; Han, S. and Schedl, P. (1984). Ribosomal RNA genes of Drosophila melanogaster have a novel chromatin structure. *J. Mol. Biol.* **175**, 113 - 130.

Udvardy, A.; Maine, E. and Schedl, P. (1985). The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains. *J. Mol. Biol.* **185**, 341 - 358.

Udvardy, A. and Schedl, P. (1984). Chromatin organization of the 87A7 heat shock locus of Drosophila melanogaster. *J. Mol. Biol.* **172**, 385 - 403.

Valadez-Graham, V.; Razin, S. and Recillas-Targa, F. (2004). CTCF-dependent enhancer blockers at the upstream region of the chicken á-globin gene domain. *Nucleic Acids Res.* **32**, 1354 - 1362.

Valentine, J. and Sewell, W. (1998). Characterisation of inducible DNase I hypersensitive sites flanking the human interleukin-5 gene. *Biochim. Biophys. Acta* **1442**, 218 - 219.

Vazquez, J. and Schedl, P. (1994). Sequences required for enhancer blocking activity of scs are located within two nuclease-hypersensitive regions. *EMBO* **13**, 5984 - 5993.

Vazquez, J. and Schedl, P. (2000). Deletion of an insulator element by the mutation facet-strawberry in Drosophila melanogaster. *Genetics* **155**, 1297 - 1311.

Villeponteau, B. (1989). Characterization of a topoisomerase-like activity at specific hypersensitive sites in the Drosophila histone gene cluster. *Biochem Biophys Res Commun.* **162**, 232 - 237.

Voelker, R.; Graves, J.; Gibson, W. and Eisenberg, M. (1990). Mobile element insertions causing mutations in the Drosophila suppressor of sable locus occur in DNase I hypersensitive subregions of 5'-transcribed nontranslated sequences. *Genetics* **126**, 1071 - 1082.

Wallrath, L.; Lu, Q.; Granok, H. and Elgin, S. (1994). Architectural variations of inducible eukaryotic promoters: preset and remodeling chromatin structures. *Bioessays* **16**, 165 - 170.

Weber, J.; Taxman, D.; Lu, Q. and Gilmour, D. (1997). Molecular architecture of the hsp70 promoter after deletion of the TATA box or the upstream regulation region. *Mol. Cell. Biol.* **17**, 3799 - 3808.

Wei, X.; Kishi, H.; Jin, Z.; Zhao, W.; Kondo, S.; Matsuda, T.; Saito, S. and Muraguchi, A. (2002). Characterization of chromatin structure and enhancer elements for murine recombination activating gene-2. *J. Immunol.* **169**, 873 - 881.

Wong, Y.; O'Connell, P.; Rosbash, M. and Elgin, S. (1981). DNase I hypersensitive sites of the chromatin for Drosophila melanogaster ribosomal protein 49 gene. *Nucleic Acids Res.* **9**, 6749 - 6762.

Worcel, A.; Gargiulo, G.; Jessee, B.; Udvardy, A.; Louis, C. and Schedl, P. (1983). Chromatin fine structure of the histone gene complex of Drosophila melanogaster. *Nucleic Acids Res.* **11**, 421 - 439.

Wu, C. (1980). The 5' ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**, 854 - 859.

Wu, C.; Bingham, P.; Livak, K.; Holmgren, R. and Elgin, S. (1979). The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell* **16**, 797 - 806.

Wu, C.; Wong, Y. and Elgin, S. (1979). The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell* **16**, 807 - 814.

Zerges, W.; Udvardy, A. and Schedl, P. (1992). Molecular characterization of the 5' end of the rudimentary gene in Drosophila and analysis of three P element insertions. *Nucleic Acids Res.* **20**, 4639 - 4647.

Zhao, K.; Hart, C. and Laemmli, U. (1995). Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell* **81**, 879 - 889.

List of figures and Tables

A. Figures:

Chapter one

Figure 1.1: Genomic organisation of the *Penicillium stoloniferum* virus. 19

Figure 1.2: Genomic organisation of the *Cryphonectria parasitica* hypovirus 1. 21

Figure 1.3: Genomic organisation of the *Atkinsonella hypoxylon partitivirus.* 22

Figure 1.4: Genomic organisation of the *Saccharomyces cerevisiae L-A totivirus.* 26

Figure 1.5: Genomic organisation of the *Vicia faba endornavirus.* 36

Chapter three

Figure 3.1: ACD related dsRNA elements. 78

Figure 3.2: The four largest ACD-associated dsRNA elements. 80

Figure 3.3: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of the ACD dsRNA1a species. 81

Figure 3.4: Alignments results using the predicted peptide sequence of ACD dsRNA1a element ORF. 82

Figure 3.5: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of the ACD dsRNA1b species. 83

Figure 3.6: WU-BLAST2 alignments results using the N-terminal peptide sequence of the ACD dsRNA2 element ORF. 85

Figure 3.7: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of ACD dsRNA2 species. 87

Figure 3.8: CLUSTAL_ X alignment results of the peptide sequence for the ACD dsRNA2 ORF highlighted using GeneDoc tool. 88

Figure 3.9: WU-BLAST2 alignment results of the short peptide sequence of the ACD dsRNA2 element. 92

Figure 3.10: Schematic representation of the primers and the overlap of the clones used in the construction of the contiguous sequence of the ACD dsRNA3 species. 94
Figure 3.11: CLUSTAL_ X alignment results of the peptide sequence for the ACD dsRNA3 ORF highlighted using GeneDoc tool. 96
Figure 3.12: WU-BLAST2 alignments results using the predicted peptide sequence of the ACD dsRNA3 element. 100
Figure 3.13: BLAST sequence alignment results using the dsRNA4 fragment. 102
Figure 3.14: Predicted folding of the 5' UTR regions for ACD dsRNAs 2 and 3. 103
Figure 3.15: Predicted folding of the 5' UTR regions for ACD dsRNAs 1a and 1b. 104
Figure 3.16: Predicted folding of the 3' UTR regions for ACD dsRNAs 2 and 3. 105
Figure 3.17: Predicted folding of the 3'UTR regions for ACD dsRNAs 1a and 1b. 106
Figure 3.18: CLUSTAL_X alignment, highlighted using GeneDoc, of the 5' ends of the totivirus-like dsRNAs. 106
Figure 3.19: CLUSTAL_X alignment, highlighted using GeneDoc, of the 3' ends of the totivirus-like dsRNAs. 107
Figure 3.20: Northern blots for dsRNA species 1(a), 2 and 3. 108
Figure 3.21: Viral particle isolation from ACD-infected leaf material. 109
Figure 3.22: Unrooted phylogenetic tree based on the Neighbour-joining method with a 10,000 replicate bootstrap search. 110
Figure 3.23: The two smallest ACD-related dsRNA elements. 111
Figure 3.24: Alignment of sequence segments from the 3' terminus of 0.6 kbp dsRNA A element. 112
Figure 3.25: The complementary 5' terminal positive and negative strands of the 0.6 kbp dsRNA element B. 113
Figure 3.26: Predicted folding of the 0.6 kbp dsRNA element A 5' and 3' regions. 114

Figure 3.27: Predicted folding of the 0.6 kbp dsRNA element B 3' region as well as the aligned 5' and 3' regions. 115
Figure 3.28: Predicted folding of the entire 0.6 kbp dsRNA elements A and B. 116
Figure 3.29: Predicted ribozyme cleavage site formed by folding of the dsRNA element B. 117
Figure 3.30: CLUSTAL_X alignment, highlighted using GeneDoc, of the two 1 kbp dsRNA derived clones. 118
Figure 3.31: Symptoms of CCRS in infected cherry trees. 119
Figure 3.32: Phylogenetic tree constructed from the multiple sequence alignment of the sequenced ACD and CCRS 0.6 kbp dsRNA elements 122
Figure 3.33: RT-PCR product of the ZK14-ZK17 set of primers using ACD-infected tissue. 125
Figure 3.34: RT-PCR product of the ZK14-ZK17 set of primers, using 55°C as the annealing temperature and low amounts of dsRNA template. 126
Figure 3.35: RT-PCR product of the ZK37-ZK38 set of primers, using 55°C as the annealing temperature and low amounts of dsRNA template. 127
Figure 3.36: RT-PCR product of the ZK14-ZK17 set of primers using CCRS-infected tissue. 131

Chapter four

Figure 4.1: Hyphal interference of hyphae of *H. annosum* by hyphae of *P. gigantea.* 135
Figure 4.2: *P. gigantea* dsRNA elements. 137
Figure 4.3: *P. gigantea* dsRNA elements. 138
Figure 4.4: SDS-PAGE gels of fractions A2 – 6. 139
Figure 4.5: EM photographs of the dsRNA containing fractions. 139
Figure 4.6: The two dsRNA species in the TW2 *P. gigantea* isolate. 141
Figure 4.7: Schematic representation of the clones used in the construction of the contiguous sequence of the *P. gigantea* dsRNA1 species. 141
Figure 4.8: Schematic representation of the clones used in the construction

of the contiguous sequence of the *P. gigantea* dsRNA2 species. 143
Figure 4.9: RNA-dependent RNA polymerase peptide sequence alignments. 146
Figure 4.10: Alignments results using the predicted peptide sequence of *P. gigantea* dsRNA elements 1 and 2, ORFs 2 and 1 respectively. 147
Figure 4.11: Unrooted phylogenetic tree based on the Neighbour-joining method with a 10,000 replicate bootstrap search. 150

Chapter five

Figure 5.1: Schematic representation of the overlap of the clones used in the construction of the contiguous CYAV sequence. 154
Figure 5.2: CLUSTAL_ X alignment results, highlighted using GeneDoc, of the peptide sequence of CYAV. 155
Figure 5.3: Unrooted phylogenetic tree based on the Neighbour-joining method with a 10000 replicate bootstrap search. 158
Figure 5.4: The RT-PCR amplified genomic fragments of CYSDV RNA2. 160
Figure 5.5: Schematic presentation of the first approach followed for the assembly of the full-length CYSDV RNA2 genomic segment. 161
Figure 5.6: Schematic presentation of the second approach followed for the assembly of the full-length CYSDV RNA2 genomic segment. 162
Figure 5.7: The cloned CYSDV RNA2 genomic fragments resulting from the second assembly approach. 163
Figure 5.8: Schematic representation of the PEV1 genome. 164
Figure 5.9: The extracted PEV1 dsRNA. 167
Figure 5.10: Schematic representation of the PEV1 UGT 1'2' and 3'4' amplified genomic regions; the resulting PCR products and clones. 168
Figure 5.11: Time course protein expression from the 1'2' and 3'4' UGT constructs. 169
Figure 5.12: Ion exchange column purification of the His-tagged 1'2' UGT protein product. 170
Figure 5.13: SDS-PAGE of the protein containing fractions from the ion exchange column purification. 171

Figure 5.14: TLC enzyme activity assay for the PEV1 UGT protein product. 172

Chapter six

Figure 6.1: The organisation of coding regions with their *cis*-regulatory elements for three eukaryotic gene loci. 178
Figure 6.2: Biology of the chicken lysozyme locus in respect to its HS. 191
Figure 6.3: End-labelling mapping of HS. 192
Figure 6.4: Schematic representation of the HS library generation. 195
Figure 6.5: Chart summarising the cell culture condition results. 199
Figure 6.6: Genomic areas containing the *hsp26*and *rpl32* genes. 204
Figure 6.7: Comparing the strength of differently labelled radioactive probes. 205
Figure 6.8: Different band profiles resulting from partially digested genomic DNA. 207
Figure 6.9: Mapping HS in the vicinity of the *rpl*32 gene. 209
Figure 6.10: Mapping HS in the vicinity of the *hsp*26 gene 209
Figure 6.11: HS mapped on the *D. melanogaster* genome. 213

B. Tables:

Chapter one

Table 1.1: Taxonomic presentation of dsRNA viruses of fungi. 39
Table 1.2: List of the *Endornaviridae* viruses. 43

Chapter three

Table 3.1: Sequence of the primers used in the cDNA cloning of ACD dsRNA1a. 81
Table 3.2: Sequence of the primers used in the cDNA cloning of ACD dsRNA1b. 84
Table 3.3: Sequence of the primers used in the cDNA cloning of ACD dsRNA2. 87
Table 3.4: Sequence of the primers used in the cDNA cloning of

ACD dsRNA3. 94

Table 3.5: Sequence of the primers used in the cDNA cloning of ACD dsRNA4 101

Table 3.6: Sequence of the primers used in the cDNA synthesis of the 0.6 kbp dsRNAs. 110

Table 3.7: Sequence of the primers used in the PCR-based detection of ACD-related dsRNA elements. 122

Table 3.8: Sequence of the primers designed for the PCR-based detection of ACD-related dsRNA elements. 126

Chapter four

Table 4.1: Sequence of the primers used in the cDNA cloning of *P. gigantea* dsRNA1. 142

Table 4.2: Sequence of the primers used in the cDNA cloning of *P. gigantea* dsRNA2. 143

Chapter six

Table 6.1: Summary of the HS examples given. 190

Table 6.2: Summary of the different aspects of each one of the DNase I digest protocols examined. 201

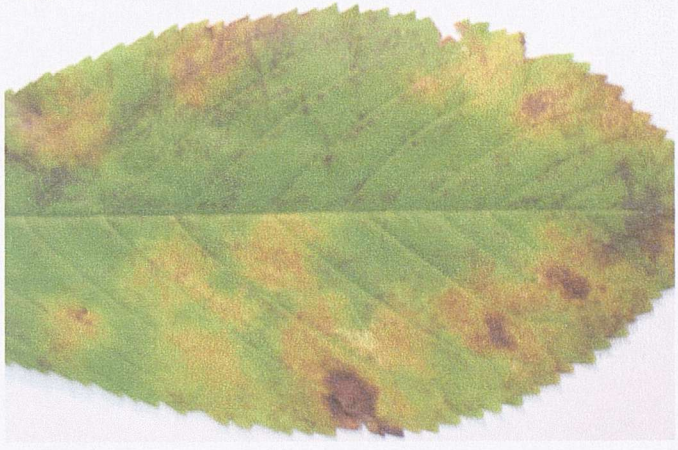
Table 6.3: Table of all the published *D. melanogaster* HS to date, grouped according to the chromosome on which they are located. 211

Appendix

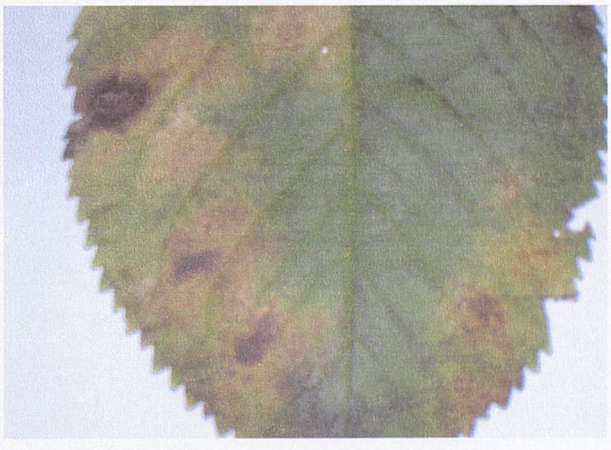
A. Pictures of ACD-infected leaf tissue.

The pictures show ACD-infected leaves at different sages of infection, collected from the Amasya region of Turkey (provided by Dr. Açikgöz). There is a wide range of symptoms demonsrated. Early symptoms (picture E) include small translucent spots that develop gradually into small areas of slight discolouration. These areas grow in size and number (pictures A and B) and the first symptoms of necrotic tissue appear. The spread of the ACD infection is not restricted in any specific area of the leaf (picture C is a photograph of the reverse side of the leaf shown in picture D). In the final stages of ACD infection (picture F), a substantial area of the leaf becomes necrotic, while the remainder loses its colour and eventually dries up.

A.



B



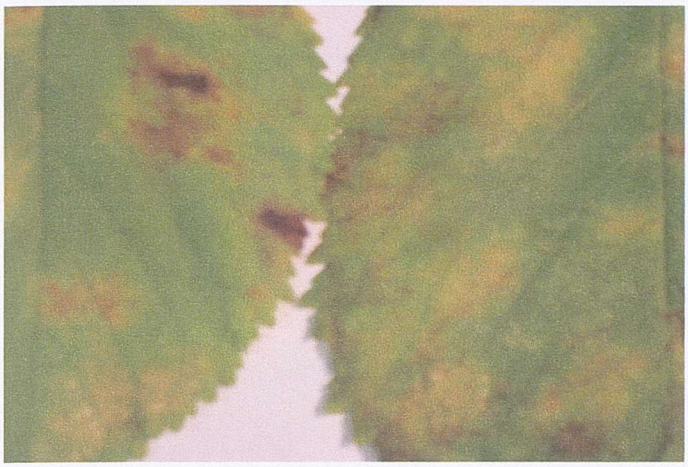
C.



D.



E.



F.



B. Manuscripts

(submitted to the J. of Phytopath.)

Molecular characterization of a totivirus from *Ophiostoma minus*

MICHELLE DOHERTY[1], KRISI SANGANEE[1], ZISIS KOZLAKIDIS[1], ROBERT H. A. COUTTS[1,*], CLIVE M. BRASIER[2] & KENNETH W. BUCK[1]

[1]Division of Biology, Faculty of Life Sciences, Sir Alexander Fleming Building, Imperial College London, London SW7 2AZ, UK

[2] Forest Research Agency, Alice Holt Lodge, Farnham, Surrey GU10 4LH, UK

[*]Author for correspondence: E-mail: r.coutts@imperial.ac.uk

A diseased isolate of *Ophiostoma minus* (formerly *Ceratocystis minor*) isolate H2392, which is a pathogenic ascomycete fungus and causes blue staining in the inner bark, phloem and xylem of pine trees [1] contains a double stranded (ds) RNA element 6.4 kbp in size. The partial nucleotide sequence (1314 nucleotides) of this dsRNA shows a single open reading frame (ORF) with the potential to encode a portion of an RNA-dependent RNA polymerase (RdRp). Five conserved amino acid motifs characteristic of RdRps [2] were identified in the sequenced region. The partial sequence of the RdRp encoded by the dsRNA is most closely related to a similar gene in a virus classified in the genus *Totivirus* of the family *Totiviridae* from *Sphaeropsis sapinea RNA virus 2* (SsRv2) with an overall sequence identity of 45%. In this report we describe the partial sequence of the RdRp gene of a new virus in *O. minus* which is the first described totivirus in the genus *Ophiostoma*.

O. minus was maintained on malt extract agar (Oxoid) at 22°C and total nucleic acids were isolated from fungal mycelium after 14 days growth on cellophane membranes overlaying the agar as described previously [4]. Single-stranded (ss) RNA and dsRNA were separated by precipitating ssRNA with 2 m LiCl and dsRNA with 5 M LiCl as described [5]. DsRNAs were separated by

polyacrylamide gel electrophoresis [4, 5] and extracted from the gel using an RNaid kit (Bio 101, Vista, CA, USA). Incubations with S1 nuclease and DNase 1 were as described [5] and the RNA was resistant to digestion with both enzymes (not shown), confirming that it is dsRNA. cDNA libraries were constructed using gel-purified dsRNA from *O. minus* isolate H2392 and the cDNA clones were sequenced as described before [5]. Sequence analysis was performed using the GCG version 8.1 programs [6]. Multiple sequence alignments, construction of phenograms, and bootstrapping analysis were accomplished with the CLUSTAL X programs [7]. The phylogenetic tree was displayed using the program TREEVIEW [8]. The EMBL database accession number for the assembled sequence is AM087551.

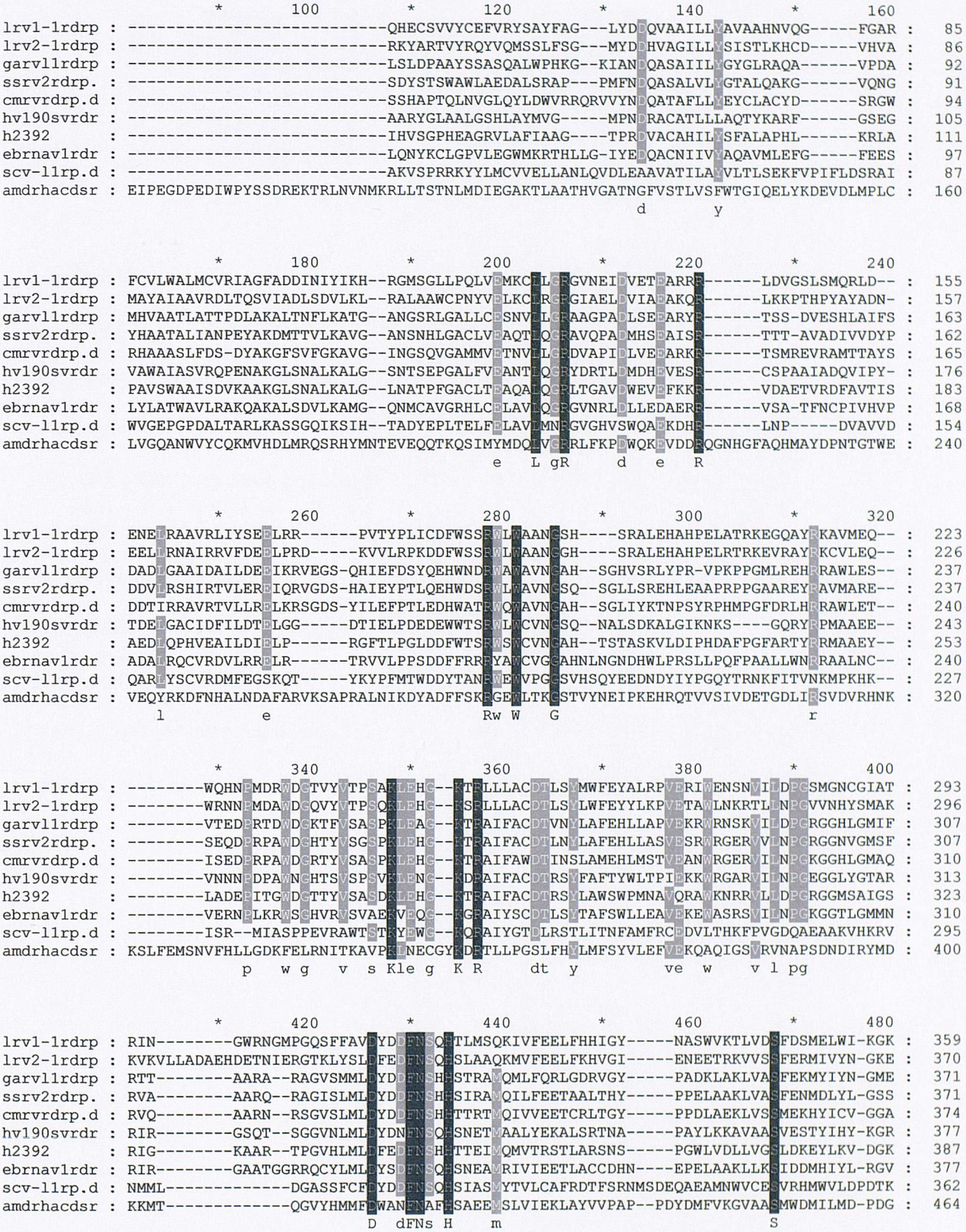
The sequence contained an ORF with the potential to encode a protein of 437 amino acids. Searches of the UniProt databases showed that the protein was related to RdRps of viruses in the genus *Totivirus* and further analyses revealed amino acid motifs typical of RdRps [2]. An alignment of these motifs with those in RdRps of other totiviruses is shown in Fig 1a. A phylogenetic tree, based on the conserved RdRp region, is shown in Fig 1b.

The family *Totiviridae* consists of viruses with isometric virions containing a single, linear, uncapped 4.6-7 kbp dsRNA genome [9]. The mRNA for coat protein and RdRp proteins is produced *via* a conserved mechanism [9]. In the phylogram shown in Fig 1b those totiviruses which cluster together share a similar translational strategy. For instance those viruses which utilise ribosomal frameshifting are grouped together e.g. *Saccharomyces cerevisae* virus L-A whilst those which utilise internal, initiation mechanisms e.g. *SsRV2* and *Gremmeniella abietina RNA virus L1* group together along with *O. minus* suggesting it may be a member of this grouping.

Acknowledgements. This work was supported by the UK Natural Environment Research Council as part of its Ecological Dynamics and Genetics (EDGE) program.

References

1. Masuya. H., Kaneko, S., and Yamaoka, Y., J. For Res *8*, 83-88, 2003.
2. Bruenn, J. A., Nucleic Acids Res *21*, 5667-5669, 1993.
3. Rogers, H. J., Buck, K. W., and Brasier, C. M. Plant Pathol *35*, 277-287, 1986.
4. Hong, Y., Dover, S. L., Cole, T. E., Brasier, C. M., and Buck, K. W., Virology, *242*, 80-89, 1998.
5. Hong, Y., Cole, T. E., Brasier, C. M., and Buck, K. W. Virology *246*, 158-169, 1998.
6. Devereux, J., Haeberli, P., and Smithies, O. Nucleic Acids Res *12*, 539-549, 1984.
7. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G., Nucleic Acids Res, *24*, 4876-4882, 1997.
8. Page, R. D. M. Comput Applic Biosci *12*, 357-358, 1996.
9. Ghabrial, S. A., Buck. K. W., Hillman, B. I. and Milne, R. G., in C. M. Fauquet, M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball (eds), *Virus Taxonomy, VIIIth Report of the ICTV*, Elsevier/Academic Press, London, 2004, 581-590.



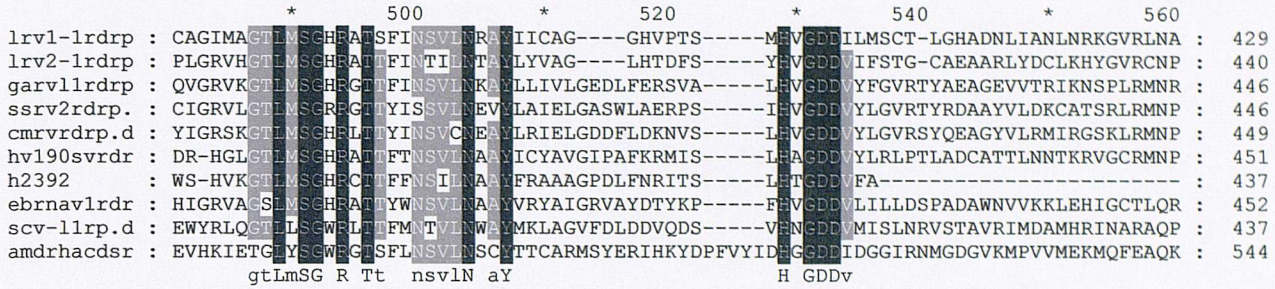


Fig. 1a. Multiple alignment of amino acid sequences of totivirus RdRp sequences. The alignment was carried out with the program CLUSTAL X and highlighted using the GeneDoc application. Sequence identities are denoted with upper case letters; sequence similarities and chemically similar residues are shown in lower case letters. as defined by the program using the BLOSUM series. Abbreviations (Database accession numbers in parentheses): OmV, *Ophiostoma minus totivirus* (AM087551); ScV-L-A, *Saccharomyces cerevisiae virus L-A* (M28353) ; LRV1-1, *Leishmania RNA virus 1-1* (M92355); LRV2-1, *Leishmania RNA virus 2-1* (U32108); CmV, *Coniothyrium minitans mycovirus* (AF527633); GaRV-L1, *Gremmeniella abietina RNA virus L1* (AF337175); SsRV2, *Sphaeropsis sapinea RNA virus 2* (AF039080); Hv190SV, *Helminthosporium victoriae virus 190S* (U41345); Eb-RV1, Eimeria brunette RNA virus 1, (AF356189).

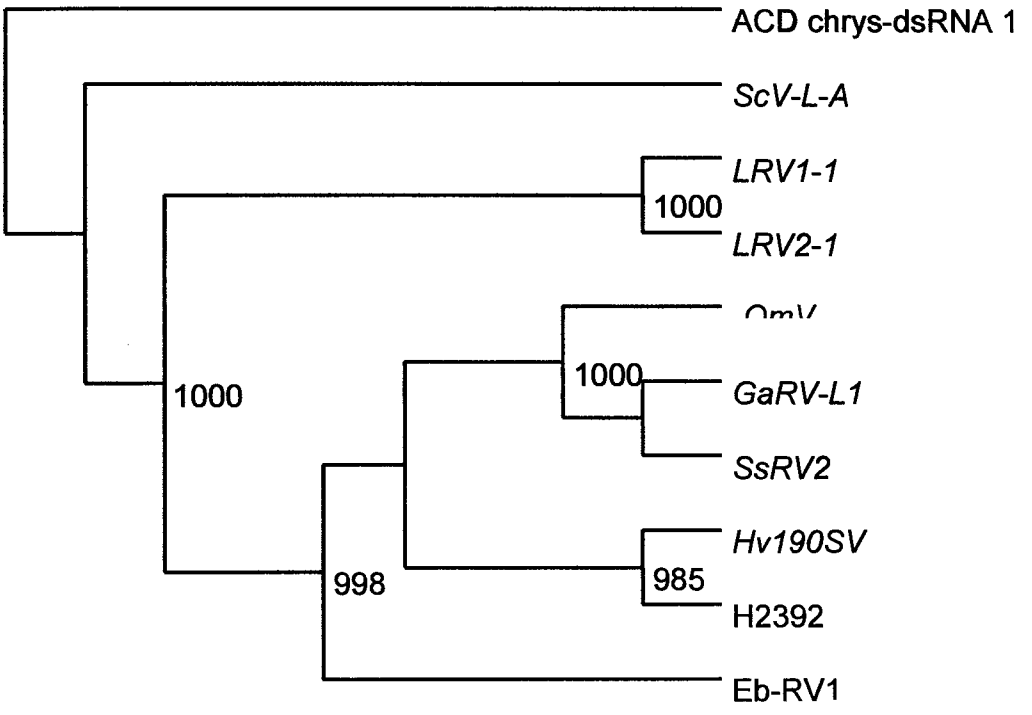


Fig. 1b. Neighbour-joining phylogenetic tree constructed using the RdRp sequences shown in Fig. 1a including the sequences between the motifs, using the program CLUSTAL X. Abbreviations are as in Fig. 1a. Bootstrap numbers out of 1000 replicates are given on the nodes. Only numbers greater than 900 are shown. Abbreviations are as shown in Fig 1a. The tree was rooted with the RdRp of Amasya cherry disease Chrysovirus dsRNA 1 (ACD chrys-dsRNA 1), a member of the genus *Chrysovirus* in the family *Chrysoviridae* [Accession no. AJ781166], which was included as an outgroup.