Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.
http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form and Deposit Licence.
Plant cell cake ‘infected’ with *Tobacco mosaic virus* (TMV).

Chocolate mud cake with chocolate ganache, fondant organelles and edible ink TMV ‘rigid rod’ virions.
AN INVESTIGATION OF MOLECULAR INDICATORS OF PLANT VIRUS INFECTION

SONIA TRACY LILLY

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Biological Sciences, The University of Auckland

2014
This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

**Chapter 5 - IDENTIFICATION AND VALIDATION OF REFERENCE GENES FOR VIRUS INFECTED ARABIDOPSIS**

Published as:

<table>
<thead>
<tr>
<th>Nature of contribution by PhD candidate</th>
<th>Development of methodologies, experimentation, interpretation of results and preparation of manuscript.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent of contribution by PhD candidate (%)</td>
<td>95%</td>
</tr>
</tbody>
</table>

**CO-AUTHORS**

<table>
<thead>
<tr>
<th>Name</th>
<th>Nature of Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revel Drummond</td>
<td>Advice on qPCR methodology and analysis</td>
</tr>
<tr>
<td>Mike Pearson</td>
<td>Provided samples, advice on plant viruses and comments on manuscript</td>
</tr>
<tr>
<td>Robin MacDiarmid</td>
<td>Advice on approach, data collection and analysis, and manuscript editing.</td>
</tr>
</tbody>
</table>

**Certification by Co-Authors**

The undersigned hereby certify that:
- the above statement correctly reflects the nature and extent of the PhD candidate’s contribution to this work, and the nature of the contribution of each of the co-authors; and
- in cases where the PhD candidate was the lead author of the work that the candidate wrote the text.

<table>
<thead>
<tr>
<th>Name</th>
<th>Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revel Drummond</td>
<td></td>
<td>17/02/2014</td>
</tr>
<tr>
<td>Mike Pearson</td>
<td></td>
<td>17/02/2014</td>
</tr>
<tr>
<td>Robin MacDiarmid</td>
<td></td>
<td>17/02/2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click here</td>
</tr>
</tbody>
</table>
ABSTRACT

Aims. This thesis sought to advance what is known about how best to monitor the health of plants in order to determine what could be used as a molecular marker to detect deviation from health due to plant virus infection. The specific objectives were to (i) confirm the ability of five dissimilar viruses to infect Arabidopsis thaliana, (ii) to develop a method for the accurate quantification and analysis of small RNAs (sRNAs) from low molecular weight RNA (LMW-RNA) in response to the five viruses, (iii) to develop a real-time quantitative polymerase chain reaction (qPCR) method for the quantification of gene transcripts of interest in response to the five dissimilar viruses, and (iv) extend qPCR assays to a further biotic stress and two abiotic stresses in Arabidopsis in order to determine specificity of assay to virus infection. Results. It was established by PCR and qPCR, sequencing, and Immunostrip® assay that each of the five viruses were successfully inoculated into Arabidopsis and were absent from mock-inoculated tissue. LMW-RNA components were accurately quantified but not all viruses could be detected at every time point. Analyses of the ratio of sRNA to rRNA as a proportion of averaged mock-inoculation predicted a 94% correlation with known virus presence. SGS3 showed a statistically significant change in transcript accumulation compared to mock-inoculation in response to all five viruses as assessed by qPCR. A decision tree predictive model was devised from the sRNA/rRNA ratio and SGS3 transcript accumulation, resulting in > 94% positive predictive value. Conclusions. It is concluded that calculating a ratio of sRNA to rRNA accumulation as a proportion of averaged mock-inoculation can predict known virus infection to a high degree of certainty, if this response proves specific to virus infection. The decision tree predictive model developed from the sRNA/rRNA ratio and SGS3 transcript accumulation increases the likelihood of predicting virus infection to > 94%. Given further investigation and analysis, our ability to detect generic plant virus infection is likely to benefit from this host plant based method.
ACKNOWLEDGMENTS

First and foremost I would like to acknowledge my family without whose loving support and understanding this thesis would never have been started, let alone finished. Thank you to our son Lewis for jumping on the bed at 6 am and telling me I needed to get onto the computer and finish my PhD. Thank you to my husband Barry, my mum Pam, brother Dean and our friends for reminding me I have the talent, tenacity and determination required to complete and submit this thesis.

I would like to acknowledge my two supervisors, Dr Robin MacDiarmid and Professor Mike Pearson, for their guidance, support and advice, comments on thesis drafts and collaboration on the published paper from this thesis. I cannot thank Robin enough for her ongoing support and encouragement and for enabling and promoting an accommodating environment for achieving completion of this thesis. Thank you also to my advisor Dr Lisa Ward for advice as to expectations of plant virus diagnostic assays for biosecurity.

Dr Revel Drummond deserves special mention for his assistance with the Beckman Coulter robot parameters and invaluable and meticulous advice with respect to qPCR methodology and analysis. Further, I would like to acknowledge Revel for his collaboration on the published manuscript from this thesis. Dr Jeremy McRae also deserves special mention for both assistance with the Beckman Coulter robot parameters and providing the custom-written Python V 2.6.3 script. Thank you also to Dr Nihal de Silva and Dr Barry Milne for statistical advice, direction and assistance.

Thank you to The New Zealand Institute of Plant and Food Research for providing conditions that enabled completion of laboratory work and writing of this thesis. Monica Dragulescu and G Wadasinghe (Wade) deserve special mention for their impeccable management of the glasshouse facilities, provision of a containment room and for providing equipment made to order. And a big thank you to the many colleagues Elaine Chan, Tracey Immanuel, Gardette Valmonte, Arnaud Blouin, Dan Cohen, Karmun Chooi, Sarah Jane Cowell, Mark Andersen, Linda Swift, Sakuntala Karunai retnam and Christina Hall, to name but a few, for the encouraging and supportive things you’ve all said and done.

Finally, this research was supported by Ministry of Science and Innovation (formerly, The New Zealand Foundation for Research, Science and Technology) grant C06X0710. I am grateful for the George Mason Scholarship in Biological Sciences, Pukehou Pouto Scholarship and the University of Auckland Partnership Appeal Award for providing stipend and funding to enable me to complete my final year of study.
TABLE OF CONTENTS

Abstract i
Acknowledgments ii
Table of Contents iii
List of Figures ix
List of Tables xiii
Abbreviations xv

CHAPTER ONE - GENERAL INTRODUCTION 1

1.1 Impact of plant viruses 1
1.2 Methods for the detection of plant viruses 5
1.3 Initial defence mechanisms of plants against viruses 6
1.4 Cellular stress responses 9
  1.4.1 The hypersensitive response (HR), defence genes, salicylic acid (SA) and systemic acquired resistance (SAR) 9
  1.4.2 Reactive oxygen species (ROS) 10
  1.4.3 Heat Shock Proteins (HSPs) 11
1.5 Developmental defect responses 13
  1.5.1 Phytohormones 13
  1.5.2 RNA interference (RNAi) 14
  1.5.3 Overview of the RNAi pathway in *Arabidopsis* 16
  1.5.4 Double-stranded RNA (dsRNA) as a trigger of RNAi 18
  1.5.5 The composition of the pool of small RNAs (sRNAs) and their biogenesis 21
  1.5.6 Function of *Arabidopsis* DCLs 22
    i. *DCL1* 23
    ii. *DCL2* 24
    iii. *DCL3* 25
    iv. *DCL4* 25
    v. Summary of *Arabidopsis* DCL functions, redundancy and hierarchy 26
  1.5.7 Guide strand selection and RISC assembly 27
  1.5.8 RISC assembly and the structure and function of Argonaute (AGO) proteins 28
    i. *AGO1* and *AGO2* 29
    ii. *AGO4* 30
  1.5.9 Amplification and local and systemic spread of RNAi 31
    i. *RDR1* and *RDR6* 32
    ii. *SGS3, SDE3* and *SDE5* 34
  1.5.10 Identity of the mobile silencing signal 36
1.6 Common suites of gene expression in virus infection 38
1.7 *Arabidopsis thaliana* as a model system for the study of plant-virus interactions 43
1.8 Five viruses selected for the present research 44
1.9 Aims 46
CHAPTER TWO – GENERAL MATERIALS AND METHODS

Preamble 48
2.1 Buffers, solutions and gels 48
2.2 Selection of maintenance and propagation host plants 49
2.3 Growth of Arabidopsis seedlings and maintenance host plants 50
2.4 Source of virus isolates 50
2.5 Mechanical inoculation 50
2.6 Experimental design
   2.6.1 Virus inoculations 52
   2.6.2 Abiotic stress treatments 53
   2.6.3 Further biotic stress inoculation 54
2.7 Nucleic acid manipulations
   2.7.1 RNA isolation, quantitation of concentration, determination of purity with
   Nanodrop and integrity with Bioanalyser 2100 RNA Nano LabChip and
   DNase I digest 55
   2.7.2 Reverse Transcription Polymerase Chain Reaction (RT-PCR) 56
2.8 Design of primers to amplify internal controls 57
2.9 Design of primers to amplify virus targets
   2.9.1 CaMV primers 60
   2.9.2 TMV primers 61
   2.9.3 TSWV primers 61
   2.9.4 TuMV primers 63
   2.9.5 TYMV primers 64
2.10 Confirmation of virus identity by gel electrophoresis and sequencing of amplicons 65
2.11 Agdia ImmunoStrip® for confirmation of TMV and TSWV identity 66
2.12 Isolation and quantification of sRNAs, selection reference genes for qPCR and qPCR
   method 66

CHAPTER THREE - CONFIRMING PRESENCE AND IDENTITY OF FIVE
VIRUSES IN ARABIDOPSIS

3.1 Introduction 67
3.2 Methods overview 67
3.3 Results
   3.3.1 CaMV 70
   3.3.2 TMV 72
   3.3.3 TSWV 74
   3.3.4 TuMV 76
   3.3.5 TYMV 78
3.4 Agdia ImmunoStrip® for confirmation of TMV and TSWV identity 80
3.5 qPCR analysis of viral transcript accumulation 81
3.6 Discussion 83
CHAPTER FOUR – EVALUATING METHODS FOR THE ISOLATION AND QUANTIFICATION OF SMALL RNAs (sRNAs)  

4.1 Introduction 86  
4.1.1 Isolation of sRNAs 87  
4.1.2 Detection and quantification of sRNAs 89  

4.2 Materials and Methods 89  
4.2.1. Isolation of sRNAs 89  
4.2.2 Development of a method for the detection and quantification of sRNAs 91  
4.2.3 Assessment of PAGE, Typhoon 9410 imaging and ImageQuant TL software (PAGE/Typhoon/ImageQuant method) for the quantification of sRNAs 92  
4.2.4 Polyacrylamide gel electrophoresis of LMW-RNA 92  
4.2.5 Gel scanning with Typhoon 9140 imager 93  
4.2.6 ImageQuant analysis 93  
4.2.7 Statistical analysis of data of the PAGE/Typhoon/ImageQuant method 94  
4.2.8 Assessment of the Agilent Bioanalyzer with Lab-on-Chip small RNA kit (Bioanalyzer method) for the quantification of accumulation of sRNA 94  
4.2.9 Small RNA Lab on Chip protocol 94  
   i) Gel preparation 95  
   ii) Preparation of gel-dye mix 95  
   iii) Loading the gel-dye mix 95  
   iv) Loading the conditioning solution and marker 96  
   v) Loading the Ladder and Samples 96  
4.2.10 Agilent Bioanalyzer 2100 expert B02.04 software – determining regions for each LMW-RNA component 96  
4.2.11 Statistical analyses of data from the Bioanalyzer method 97  

4.3 Results 98  
4.3.1 Nanodrop quantification and assessment of LMW-RNA 98  
4.3.2 Polyacrylamide gel electrophoresis of LMW-RNA 98  
4.3.3 Gel scanning with Typhoon 9140 imager and ImageQuant analysis 99  
4.3.4 Statistical analysis of PAGE/Typhoon/ImageQuant method data 101  
4.3.5 Agilent Bioanalyzer B02.04 software for determining region for each LMW-RNA component 104  
4.3.6 Statistical analyses of data from the Bioanalyzer method 105  
4.3.7 Analysis of the percentage LMW-RNA components of total LMW-RNA 109  
4.3.8 Tests of statistical significance 113  
4.3.9 Component LMW-RNA accumulation profile of *Tamarillo mosaic virus* (TamMV) 118  
4.3.10 Ratio of sRNA/rRNA accumulation as a tool predictive of virus infection 121  

4.4 Discussion 124  
4.4.1 Conclusions for PAGE/Typhoon/ImageQuant Method 124  
4.4.2 Conclusions for Agilent Bioanalyzer Method with sRNA chips 125  
4.4.5 Future direction and concluding remarks 130
CHAPTER FIVE – IDENTIFICATION AND VALIDATION OF REFERENCE GENES FOR VIRUS INFECTED ARABIDOPSIS

Preamble 132
5.1 Introduction 133
5.2 Materials and methods
   5.2.1 Virus isolates, plant material, and inoculation 136
   5.2.2 RNA isolation, DNase I digest, and cDNA synthesis 138
   5.2.3 Primer design 138
   5.2.4 Testing candidate reference gene primers with standard PCR 139
   5.2.5 qPCR conditions and analysis 141
   5.2.6 Primer efficiency and geNorm analysis of gene transcript accumulation stability 141
5.3 Results
   5.3.1 Initial selection of candidate reference genes 142
   5.3.2 Exclusion of genes affected by viral infection 145
   5.3.3 Establishing cDNA and primer pairs for qPCR 147
   5.3.4 Average transcript accumulation stability (M) as calculated by geNorm 151
   5.3.5 Quantity (Q) of transcript accumulation 154
   5.3.6 Choice of reference genes affects normalisation 157
   5.3.7 Amplification of Arabidopsis EF1α, F-BOX, SAND and PDF2 from alternative herbaceous indicator host plants 159
5.4 Discussion 162
5.5 Summary 169

CHAPTER SIX – TRANSCRIPT ACCUMULATION OF GENES OF INTEREST IN RESPONSE TO FIVE VIRUSES

6.1 Introduction 170
6.2 Materials and methods
   6.2.1 Sourcing genes commonly showing fold change in response to viruses in Arabidopsis 172
   6.2.2 Virus isolates, plant material, and inoculations 173
   6.2.3 Further biotic stress treatment - Pseudomonas syringae pv. tomato, plant material and inoculations 173
   6.2.4 Abiotic stress treatments - 100 mM /200 mM NaCl and drought 173
   6.2.5 RNA isolation, DNase I digest, and cDNA synthesis 174
   6.2.6 Primer design 174
   6.2.7 Testing target gene primers with standard PCR 174
   6.2.8 qPCR conditions and analyses 174
   6.2.9 Standardisation of variable biological replicates 175
6.3 Results

6.3.1 Selection of genes of interest for qPCR analysis of transcript accumulation  
6.3.2 Primer efficiency and transcript accumulation  
6.3.3 Transcript accumulation of genes commonly showing fold change as referenced in literature  
6.3.4 Transcript accumulation of components of RNAi and three additional genes of interest  
6.3.5 Log transformation, mean centering and auto-scaling of the quantity data (Q) of selected target genes  
6.3.6 Transcript accumulation of SGS3 in response to a further biotic stress and two abiotic stresses  
6.3.7 Combined sRNA/rRNA and SGS3 decision tree as a predictive model

6.4. Discussion

CHAPTER SEVEN – GENERAL DISCUSSION

7.1 Summary and main findings

1 – Confirm presence and identity of the five selected viruses in Arabidopsis  
2 – Evaluate methods for the accurate quantification and analysis of sRNAs in response to the five viruses  
3 – Develop a qPCR method for the quantification of genes showing fold change in response to five viruses  
4 – Extend assays to a further biotic and two abiotic stresses

7.2 Implications of the main findings

7.2.1 Bioanalyzer method  
7.2.2 Identification and validation of reference genes  
7.2.3 Changes in transcript accumulation of genes of interest  
7.2.4 Development of a sRNA/rRNA and SGS3 decision tree predictive tool  
7.2.5 The sRNA/rRNA and SGS3 decision tree predictive tool as a complement to next generation sequencing (NGS) technologies

7.3 Limitations of the present studies

7.3.1 Limitations of quantification of LMW-RNA components  
7.3.2 Limit to the number of genes investigated  
7.3.3 Limitations of microarray data as a source of genes of interest  
7.3.4 Limitations of qPCR data as a source of genes of interest and as a method for quantifying transcript accumulation  
7.3.5 Limitations of the combined sRNA/rRNA and SGS3 decision tree as a predictive model  
7.3.6 Limitations of transcriptomic analyses and the complexity of plant defence response pathways

7.4 Suggestions for further research

7.5 Concluding remarks

REFERENCES
APPENDICES

APPENDIX I  PRÉCIS OF EACH OF THE FIVE VIRUSES OF THIS STUDY

A1  *Cauliflower mosaic virus* (CaMV)
   A.1.1 Transmission, host range and particle morphology
   A.1.2 Symptomology
   A.1.3 Replication, transcription and translation
   A.1.4 VSRs – P6 and 8S RNA

A2  *Tobacco mosaic virus* (TMV)
   A.2.1 Transmission, host range and particle morphology
   A.2.2 Symptomology
   A.2.3 Replication, transcription and translation
   A.2.4 VSR – P126

A3  *Tomato spotted wilt virus* (TSWV)
   A.3.1 Transmission, host range and particle morphology
   A.3.2 Symptomology
   A.3.3 Replication, transcription and translation
   A.3.4 VSR – NSs

A4  *Turnip mosaic virus* (TuMV)
   A.4.1 Transmission, host range and particle morphology
   A.4.2 Symptomology
   A.4.3 Replication, transcription and translation
   A.4.4 VSR – HC-Pro

A5  *Turnip yellow mosaic virus* (TYMV)
   A.5.1 Transmission, host range and particle morphology
   A.5.2 Symptomology
   A.5.3 Replication, transcription and translation
   A.5.4 VSR – P69

APPENDIX II  COMPARISONS OF BIOANALYZER ELECTROPHORETIC OUTPUT OF LMW-RNA COMPONENTS

APPENDIX III  PROPORTION OF LMW-RNA COMPONENTS OF TOTAL LMW-RNA

APPENDIX IV  REGRESSIONS LOG DATA

APPENDIX V  PRINCIPAL COMPONENTS ANALYSIS DATA
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Host responses and altered gene expression associated with plant virus infections</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>Antiviral silencing in <em>Arabidopsis</em></td>
<td>17</td>
</tr>
<tr>
<td>1.3</td>
<td>Possible primary sources of RNA silencing as elicited by viruses</td>
<td>19</td>
</tr>
<tr>
<td>1.4</td>
<td>Summary model of the functions of DCLs in <em>Arabidopsis</em> RNA silencing and spread of silencing</td>
<td>24</td>
</tr>
<tr>
<td>1.5</td>
<td>Argonaute proteins (AGO) in <em>Arabidopsis</em> &amp; their association with given sRNAs</td>
<td>28</td>
</tr>
<tr>
<td>1.6</td>
<td>Schematic representation for dissecting TuMV-GFP infection foci into four zones for spatial analysis of host gene expression</td>
<td>39</td>
</tr>
<tr>
<td>1.7</td>
<td>Relationship of <em>Turnip mosaic virus</em> (TuMV) accumulation to the expression of major functional groups of genes</td>
<td>40</td>
</tr>
<tr>
<td>2.1</td>
<td>Annotation of primers designed to CaMV V5 and 35S regions</td>
<td>60</td>
</tr>
<tr>
<td>2.2</td>
<td>Annotation of primers designed to the TMV MP, CP and spanning both the MP and CP</td>
<td>62</td>
</tr>
<tr>
<td>2.3</td>
<td>Annotation of primers designed to the central region of the TSWV large (L) segment and the non-structural protein (NSm) of the medium (M) segment</td>
<td>63</td>
</tr>
<tr>
<td>2.4</td>
<td>Annotation of primer pairs designed to the CI and NIb regions of TuMV</td>
<td>64</td>
</tr>
<tr>
<td>2.5</td>
<td>Annotation of primers designed to the MP and CP regions of TYMV</td>
<td>64</td>
</tr>
<tr>
<td>3.1</td>
<td>Observations of symptom development on inoculated leaves and systemic tissues</td>
<td>69</td>
</tr>
<tr>
<td>3.2</td>
<td>CaMV in <em>Arabidopsis</em></td>
<td>71</td>
</tr>
<tr>
<td>3.3</td>
<td>TMV in <em>Arabidopsis</em></td>
<td>73</td>
</tr>
<tr>
<td>3.4</td>
<td>TSWV in <em>Arabidopsis</em></td>
<td>75</td>
</tr>
<tr>
<td>3.5</td>
<td>TuMV in <em>Arabidopsis</em></td>
<td>77</td>
</tr>
<tr>
<td>3.6</td>
<td>TYMV in <em>Arabidopsis</em></td>
<td>79</td>
</tr>
</tbody>
</table>
Figure 3.7. Agdia Immunostrip® detection of TMV and TSWV in *Arabidopsis* at 28 dpi

Figure 3.8. Normalised average quantity (Q) of three biological replicates of mock-inoculated and CaMV, TMV, TSWV, TuMV and TYMV inoculated samples rescaled to 1.0 and representing viral RNA transcript accumulation.

Figure 4.1. Image of an Agilent Technologies Small RNA Chip

Figure 4.2. Novex® TBE 8% (10 well) gel of LMW-RNA from two-fold serial dilution of *Arabidopsis* 21 days post inoculation with TuMV and stained with ethidium bromide

Figure 4.3. Novex® TBE 8% (10 well) gel of LMW-RNA from two-fold serial dilution of *Arabidopsis* 21 dpi with TuMV imaged with Typhoon 94100 and analysed with ImageQuant TL 7.0 automatic analysis

Figure 4.4. Scatterplot of ImageQuant data of four technical replicates of a two-fold ten point serial dilution of 21 dpi TuMV-inoculated *Arabidopsis* LMW-RNA

Figure 4.5. Scatterplots of band intensity reported from 10 replicates of 5.0, 2.5, 1.25 and 0.625 µg TuMV-inoculated LMW-RNA to determine within gel reproducibility of the PAGE/Typhoon/ImageQuant method

Figure 4.6. Example of Agilent Bioanalyzer 2100 electrophoretic gel output with LMW-RNA components labelled

Figure 4.7. Scatterplots of 10 point two-fold serial dilutions of mock- and TuMV-inoculated *Arabidopsis* LMW-RNA at 21 dpi

Figure 4.8 A-D. Proportion LMW-RNA components (sRNA and tRNA) of total LMW-RNA

Figure 4.8 E-H. Proportion LMW-RNA components (snoRNA and rRNA) of total LMW-RNA

Figure 4.9. Scatterplots of regression analysis of three replicates of the proportion of components of virus-inoculated LMW-RNA across dpi in comparison to mock-inoculation (n=21, TuMV n=15)

Figure 4.10. Principal component analysis plot

Figure 4.11 A-D. Proportion of LMW-RNA components of *Tamarillo mosaic virus* (TamMV)-inoculated Tamarillo LMW-RNA compared to LMW-RNA components of mock-inoculated Tamarillo LMW-RNA

Figure 4.12. Ratio of sRNA to rRNA of total LMW-RNA as a proportion of averaged mock-inoculation where mock-inoculation is rescaled to 1

Figure 4.13. Ratio of sRNA to rRNA of total TamMV-inoculated LMW-RNA as a proportion of averaged mock-inoculation where mock-inoculation is rescaled to 1
Figure 5.1. Flow diagram of method for cDNA synthesis, testing for integrity and absence of gDNA

Figure 5.2. Ranking of stability and transcript accumulation of candidate reference genes

Figure 5.3. Ranking of stability and transcript accumulation of candidate reference genes across time

Figure 5.4. Ranking of stability and transcript accumulation of candidate reference genes across treatments

Figures 5.5A. Relative accumulation stability (Q) of candidate reference genes/time course

Figures 5.5B. Relative accumulation stability (Q) of candidate reference genes/treatments

Figure 5.6. Comparison of raw and normalised accumulation of RDR1 and RDR6 gene transcripts

Figure 5.7. TBE Gel electrophoresis of four most stable reference genes (EF1α 137 bp, F-BOX 140 bp, SAND 127 bp and PDF2 104 bp) and ability to amplify from the Nicotiana species; N. benthamiana (Nb), N. clevelandii (Nc), N. occidentalis (No), N. sylvestris (Ns) and N. tabacum (Nt) compared to Arabidopsis (At)

Figure 5.8. TBE Gel electrophoresis of four most stable reference genes (EF1α 137 bp, F-BOX 140 bp, SAND 127 bp and PDF2 104 bp) and ability to amplify from Chenopodium amaranticolor (Ca), C. quinoa (Cq), P. vulgaris (Pv), G. globosa (Gg) and Cucumis sativum (Cs) compared to Arabidopsis (At)

Figure 6.1. Normalised average Q rescaled to 1 of T15B16.5, TAT3, EVERSHELD, TOUCH3, ZAT12 and PDX1.2 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock inoculation across time course (n=3)

Figure 6.2. Normalised average Q rescaled to 1 of AGO1 and AGO4 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock inoculation across time course (n=3)

Figure 6.3. Normalised average Q rescaled to 1 of DCL1, 2, 3 and 4 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across time course (n=3)

Figure 6.4. Normalised average Q rescaled to 1 of SGS3 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across time course (n=3)

Figure 6.5. Normalised average Q rescaled to 1 of ARF6, ARF8 and RBRI in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across time course (n=3)
Figure 6.6. Normalised average Q rescaled to 1 of three biological replicates of SGS3 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across the time course (2 to 42 dpi)

Figure 6.7. Log transformed, mean centered and auto-scaled SGS3 transcript accumulation (Q)

Figure 6.8. Normalised average Q of transcript accumulation of SGS3 in response to P. syringae pv. tomato, strain DC3000 compared to mock-inoculation of the same dpi.

Figure 6.9. Normalised average Q of transcript accumulation of SGS3 in response to NaCl and drought stress

Figure 7.1. Flowchart of methods from receipt of sample to decision tree as a predictive model

Figure 7.2. Decision tree as a predictive model

APPENDICES FIGURES

Figure AII.1. Schematic representation of the CaMV genome

Figure AII.2. Schematic representation of the TMV genome

Figure AII.3. Schematic representation of the three TSWV RNA segments of the tripartite genome

Figure AII.4. Schematic representation of the monopartite Potyvirus genome

Figure AII.5. Schematic representation of the Turnip yellow mosaic virus genome

Figure AIII.1. Comparisons of Agilent Bioanalyzer 2100 electrophoretic outputs of Mock, CaMV, TMV, TSWV, TuMV and TYMV-inoculated Arabidopsis LMW-RNA

Figure AIII.2. Proportion LMW-RNA components (sRNA and tRNA) of total LMW-RNA

Figure AIII.3. Proportion of LMW-RNA components (sRNA and tRNA) of mock-and Tamarillo mosaic virus (TamMV)-inoculated Tamarillo

Figure AIII.4. Proportion of LMW-RNA snoRNA and rRNA of mock-and Tamarillo mosaic virus (TamMV)-inoculated Tamarillo
## LIST OF TABLES

Table 1.1. Properties of the five viruses selected for the present research. Table summarises virus name and acronym, genus, genome sense and nucleic acid, virion morphology and virus-encoded proteins, including virus-encoded suppressors of RNAi (VSRs) known to interact with host plants  

Table 2.1. Maintenance and propagation plant hosts selected for the five viruses utilised in the present research  

Table 2.2. Experimental design – virus inoculation. Numbers of *Arabidopsis* plants per treatment (mock- or virus-inoculation) and tissue collection time point (2, 3, 7, 14, 21, 28, 35 and 42 dpi)  

Table 2.3. Experimental design - abiotic stresses. *Arabidopsis* plants with no treatment, salt (NaCl 100 mM and 200 mM) or drought and tissue collection time point (0, 15, 30 min, 1, 4, 24 h, or 7, 14, 21 d)  

Table 2.4. Experimental design - *Pseudomonas syringae* var. tomato DC3000 inoculation. Control, mock inoculated and *P. syringae* var. tomato inoculated *Arabidopsis* plants sampled over a time course of 1, 2, 6 and 10 days post inoculation (dpi)  

Table 2.5. Oligonucleotide primers used for cDNA amplification of viruses and internal controls  

Table 4.1. Amount loaded, amount reported (pg) and standard error of the mean of replicates of LMW-RNA. A. Three biological replicates of a 10 point two-fold serial dilution of mock-inoculated *Arabidopsis* LMW-RNA at 21 dpi. B. TuMV-inoculated *Arabidopsis* at 21 dpi. C. Four technical replicates of a five-point two-fold serial dilution of mock-inoculated *Arabidopsis* LMW-RNA at 21 dpi.  

Table 4.2. Summary table of the significance (from regression analysis) of the proportion of components (sRNA, tRNA, snoRNA and rRNA) of virus-inoculated LMW-RNA in comparison to components of mock-inoculated LMW-RNA across the time course (2-42 dpi).  

Table 4.3. Multivariate mixed model analysis output. Model information refers to dependent variable sRNA and subject effect replicate plants. Significance determined from Least Squares Means and Differences of Least squares Means with Dunnett-Hsu adjustment. Values of significance are shaded and highlighted in bold  

Table 5.1. Candidate reference genes and comparative genes (RDR1 and RDR6) selected for qPCR validation: specifications of primers and amplicon  

Table 5.2. geNorm ranking by (M) of candidate reference genes from qPCR validated microarray data from literature  

Table 5.3. Fold change of candidate reference genes following virus infection
Table 5.4. The nine candidate reference genes and two genes used for comparison (RDR1 and RDR6). AGI locus identifier number, descriptor and function of genes utilised in this study from information publically available from The Arabidopsis Information resource (TAIR)

Table 6.1. Genes commonly showing fold change as referenced in literature

Table 6.2. Components of RNAi pathway and miscellaneous genes selected for qPCR analysis of transcript accumulation

Table 6.3. Genes selected for qPCR analysis of transcript accumulation: specifications of primers and amplicons
ABBREVIATIONS

SI (Le Système International d’Unités) abbreviations are used for units, chemicals, elements and formulae. Other abbreviations used in the text are listed below:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenosine</td>
</tr>
<tr>
<td>abRNA</td>
<td>aberrant RNA</td>
</tr>
<tr>
<td>ACT</td>
<td>ACTIN</td>
</tr>
<tr>
<td>AGO</td>
<td>Argonaute</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>Arabidopsis thaliana</td>
</tr>
<tr>
<td>ARF</td>
<td>AUXIN RESPONSE FACTOR</td>
</tr>
<tr>
<td>ATF</td>
<td>aphid transmission factor</td>
</tr>
<tr>
<td>Avr</td>
<td>avirulence</td>
</tr>
<tr>
<td>AWCGS</td>
<td>Alan Wilson Centre Genome Serve</td>
</tr>
<tr>
<td>Bay-0</td>
<td>Bayreuth-0 ecotype (Arabidopsis)</td>
</tr>
<tr>
<td>BG3</td>
<td>B-1-3 GLUCANASE</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Alignment Search Tool</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>CAC</td>
<td>CLATHRIN ADAPTOR COMPLEX</td>
</tr>
<tr>
<td>CaLCuV</td>
<td>Cabbage leaf curl virus</td>
</tr>
<tr>
<td>CaMV</td>
<td>Cauliflower mosaic virus</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CI</td>
<td>cytoplasmic inclusions</td>
</tr>
<tr>
<td>CMV</td>
<td>Cucumber mosaic virus</td>
</tr>
<tr>
<td>Col-0</td>
<td>Columbia ecotype (Arabidopsis)</td>
</tr>
<tr>
<td>CP</td>
<td>coat protein</td>
</tr>
<tr>
<td>Cq</td>
<td>threshold cycle/ crossing threshold</td>
</tr>
<tr>
<td>CTAB</td>
<td>cetyltrimethylammonium bromide</td>
</tr>
<tr>
<td>CYTc-2</td>
<td>CYTOCHROME C2</td>
</tr>
<tr>
<td>DCL</td>
<td>DICER-LIKE</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>dNTPs</td>
<td>2’-deoxynucleotide 5’-triphosphates</td>
</tr>
<tr>
<td>dpi</td>
<td>days post inoculation</td>
</tr>
<tr>
<td>DRB</td>
<td>DOUBLE-STRANDED RNA BINDING PROTEIN</td>
</tr>
<tr>
<td>ds</td>
<td>double-stranded</td>
</tr>
<tr>
<td>dsRNA</td>
<td>double stranded RNA</td>
</tr>
<tr>
<td>DsMV</td>
<td>Dasheen mosaic virus</td>
</tr>
<tr>
<td>dsRBD</td>
<td>double-stranded RNA binding domain</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>EDS1</td>
<td>ENHANCED DISEASE SUSCEPTIBILITY 1</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylene diamine tetra-acetic acid</td>
</tr>
<tr>
<td>eEF1β</td>
<td>EUKARYOTIC ELONGATION FACTOR 1 BETA</td>
</tr>
<tr>
<td>EF1α</td>
<td>ELONGATION FACTOR 1 ALPHA</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>ER</td>
<td>endoplasmic reticulum</td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
</tr>
<tr>
<td>EXP5</td>
<td>EXPORTIN 5</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>F-Box</td>
<td>F-Box family protein</td>
</tr>
<tr>
<td>GAPDH</td>
<td>GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE</td>
</tr>
<tr>
<td>gDNA</td>
<td>genomic DNA</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GLRaV</td>
<td>Grapevine leafroll associated virus</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S-transferases</td>
</tr>
<tr>
<td>h</td>
<td>homolog</td>
</tr>
<tr>
<td>hai</td>
<td>hours after inoculation</td>
</tr>
<tr>
<td>HC</td>
<td>helper component</td>
</tr>
<tr>
<td>HC-Pro</td>
<td>HC protease</td>
</tr>
<tr>
<td>hc-siRNAs</td>
<td>heterochromatin siRNAs</td>
</tr>
<tr>
<td>HEL</td>
<td>HELICASE</td>
</tr>
<tr>
<td>HEN1</td>
<td>HUA-ENCHANCER 1</td>
</tr>
<tr>
<td>HMW-RNA</td>
<td>high-molecular-weight RNA</td>
</tr>
<tr>
<td>HR</td>
<td>hypersensitive response</td>
</tr>
<tr>
<td>HSP</td>
<td>heat shock protein</td>
</tr>
<tr>
<td>HST</td>
<td>HASTY</td>
</tr>
<tr>
<td>HYL1</td>
<td>HYPONASTIC LEAVES 1</td>
</tr>
<tr>
<td>ICTvDB</td>
<td>International Committee on Taxonomy of Viruses database</td>
</tr>
<tr>
<td>JAX1</td>
<td>JACALIN-TYPE LECTIN REQUIRED FOR POTEXVIRUS RESISTANCE1</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>LB</td>
<td>Luria Broth</td>
</tr>
<tr>
<td>Ler-2</td>
<td>Landsberg ecotype (<em>Arabidopsis</em>)</td>
</tr>
<tr>
<td>LMR</td>
<td>Lectin mediated resistance</td>
</tr>
<tr>
<td>LMV</td>
<td>Lettuce mosaic virus</td>
</tr>
<tr>
<td>LMW-RNA</td>
<td>low molecular weight RNA</td>
</tr>
<tr>
<td>M</td>
<td>gene stability value</td>
</tr>
<tr>
<td>MET</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>MID</td>
<td>Middle domain</td>
</tr>
<tr>
<td>MIQE</td>
<td>minimum standard for the provision of information for qPCR experiments</td>
</tr>
<tr>
<td>miRNA</td>
<td>micro-RNA</td>
</tr>
<tr>
<td>MP</td>
<td>movement protein</td>
</tr>
<tr>
<td>MPI</td>
<td>Ministry of Primary Industries</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>NAD5</td>
<td>Nicotinamide adenine dinucleotide dehydrogenase subunit five</td>
</tr>
<tr>
<td>nat-siRNAs</td>
<td>natural antisense transcript siRNAs</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>NGS</td>
<td>Next Generation Sequencing</td>
</tr>
<tr>
<td>NIA/b</td>
<td>nuclear inclusion protein a/b</td>
</tr>
<tr>
<td>NONII</td>
<td>non-conserved region II</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>OSRV</td>
<td>Oil seed rape virus</td>
</tr>
<tr>
<td>PAD4</td>
<td>PHYTOALEXIN DEFICIENT 4</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PAMP</td>
<td>pathogen-associated molecular pattern</td>
</tr>
<tr>
<td>PAZ</td>
<td>Piwi, Argonaute and Zwille</td>
</tr>
<tr>
<td>PCD</td>
<td>programmed cell death</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SOD</td>
<td>Superoxide dismutase</td>
</tr>
<tr>
<td>SPL</td>
<td>SQUAMOSA PROTEIN BINDING LIKE</td>
</tr>
<tr>
<td>sRNA</td>
<td>small RNA</td>
</tr>
<tr>
<td>sRSA</td>
<td>sRNA sequencing and assembly</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>SUC2</td>
<td>SUCROSE TRANSPORTER 2</td>
</tr>
<tr>
<td>SUMO</td>
<td>small ubiquitin-like modifier</td>
</tr>
<tr>
<td>TAE</td>
<td>tris-acetate-EDTA</td>
</tr>
<tr>
<td>TAIR</td>
<td>The Arabidopsis Information Resource</td>
</tr>
<tr>
<td>TamMV</td>
<td>Tamarillo mosaic virus</td>
</tr>
<tr>
<td>TAS</td>
<td>TRANSCRIPTIONAL SILENCING COMPLEX</td>
</tr>
<tr>
<td>tasiRNAs</td>
<td>trans-acting siRNAs</td>
</tr>
<tr>
<td>TATA</td>
<td>TATA BINDING PROTEIN</td>
</tr>
<tr>
<td>TAV</td>
<td>translational transactivator</td>
</tr>
<tr>
<td>TCV</td>
<td>Turnip crinkle virus</td>
</tr>
<tr>
<td>T-DNA</td>
<td>transfer DNA</td>
</tr>
<tr>
<td>TEV</td>
<td>Tobacco etch virus</td>
</tr>
<tr>
<td>TIP41</td>
<td>TIP41-like family protein</td>
</tr>
<tr>
<td>Tm</td>
<td>melting temperature</td>
</tr>
<tr>
<td>TMV</td>
<td>Tobacco mosaic virus</td>
</tr>
<tr>
<td>TMV-Cg</td>
<td>Tobacco mosaic virus crucifer strain</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>TRV</td>
<td>Tobacco rattle virus</td>
</tr>
<tr>
<td>TSWV</td>
<td>Tomato spotted wilt virus</td>
</tr>
<tr>
<td>TUB</td>
<td>β-TUBULIN</td>
</tr>
<tr>
<td>TuMV</td>
<td>Turnip mosaic virus</td>
</tr>
<tr>
<td>TuNI</td>
<td>TuMV necrosis-inducing factor</td>
</tr>
<tr>
<td>TVCV</td>
<td>Turnip vein clearing virus</td>
</tr>
<tr>
<td>TYMV</td>
<td>Turnip yellow mosaic virus</td>
</tr>
<tr>
<td>U</td>
<td>uridine</td>
</tr>
<tr>
<td>UBQ</td>
<td>UBIQUITIN</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>V</td>
<td>volt</td>
</tr>
<tr>
<td>vcRNA</td>
<td>viral complementary RNA</td>
</tr>
<tr>
<td>VPg</td>
<td>viral protein genome-linked</td>
</tr>
<tr>
<td>VRC</td>
<td>viral replication complex</td>
</tr>
<tr>
<td>VRGs</td>
<td>virus responsive genes</td>
</tr>
<tr>
<td>vRNA</td>
<td>viral RNA</td>
</tr>
<tr>
<td>vsiRNA</td>
<td>viral secondary RNA</td>
</tr>
<tr>
<td>VSR</td>
<td>virus-encoded suppressor of RNAi</td>
</tr>
<tr>
<td>WFT</td>
<td>Western flower thrip</td>
</tr>
<tr>
<td>XS</td>
<td>Rice gene X and SGS3 domain</td>
</tr>
<tr>
<td>YFP</td>
<td>Yellow fluorescent protein</td>
</tr>
<tr>
<td>YSL8</td>
<td>Yellow stripe like 8</td>
</tr>
</tbody>
</table>
CHAPTER ONE – GENERAL INTRODUCTION

1.1 Impact of plant viruses

New Zealand horticulture, agriculture and silviculture are dependent on exotic plant species, and require imported germplasm, for genetic material to initiate new industries or to introduce new desirable traits (e.g. disease resistance) into existing commercial cultivars. Plant viruses can be imported with plant material and while some viruses may be asymptomatic in their imported hosts; they could cause devastating losses in other plant hosts (Clover et al. 2007; Ministry of Primary Industries 2013).

Multiple agreements between countries aimed at reducing obstacles to international trade have resulted in an increased frequency of incursions of unwanted organisms and subsequent emergence of pests and diseases that are costly to control and/or eradicate (reviewed in MacDiarmid et al. 2013). The International Committee for the Taxonomy of Viruses database (ICTVdB) recognised a global total of around 2000 viral species in 2005 (Fauquet et al. 2005; reviewed in Rodoni 2009). However, metagenomic studies estimate \(10^{23}\) viral species are likely to be present globally and discovery of new viruses is occurring at an alarming rate (reviewed in Rodoni 2009). Of particular concern, in addition to unknown viruses, are emergent plant viruses, some appearing in response to climate change. The interactions between potential extreme weather phenomena and plant viruses and their effects cannot be predicted and the effects of unknown and emergent viruses cannot be anticipated (reviewed in Rodoni 2009).

The Biosecurity Strategy for New Zealand (Young 2003), states that “Biosecurity is the exclusion, eradication or effective management of risks posed by pests and diseases to the economy, environment and human health.” As a result, New Zealand Biosecurity seeks to protect plants, including native flora, of high economic, social, environmental and cultural value
(MacDiarmid et al. 2013). In addition, market access for New Zealand’s agricultural products can be affected by detection of quarantinable viruses. Therefore it is of importance that we are able to detect all plant viruses, previously known and unknown, in plant tissue at our borders. It would be beneficial for biosecurity and post-entry quarantine to have access to an assay, or tool, where the health status of plants can be monitored, any deviation from which might be suggestive of virus presence.

This leads to two questions central to this thesis:

i) How can we best monitor the health status of plants?

ii) What could be used as a molecular marker or indicator to detect deviation from health due to virus infection?

To begin to answer these questions it is necessary to consider the interaction between viruses and their plant hosts and the defence mechanisms that plants are known to deploy against virus infections. To realise their full replication cycle and maximise the likelihood of movement to a new host plant, plant viruses need to proliferate in newly infected cells, suppress host defences, and move into adjacent cells and through the vascular system in order to colonise the entire host plant. To achieve this, plant viruses commandeer plant proteins and other cellular resources for their own purposes. These redirected proteins and other cellular resources normally undertake routine, plant-specific processes (Citovsky et al. 2009). The obligate-intracellular and parasitic nature of plant viruses provides numerous opportunities for the proteins and nucleic acids of both the plant host and the virus to interact. The expression and activities of plant hormones, and plant and virus nucleic acids and proteins are in turn influenced by this interaction. These circumstances provide opportunity for plant pathologists to elucidate the
pathways involved in the deployment of defence and counter-defence mechanisms. From knowledge of these defence-related pathways we might identify molecular plant responses, or indicators, to virus presence from which novel tools for plant virus detection could be developed. However, plant-virus interactions are complicated by the heterogeneous mixture of plant cells responding to a virus as it replicates and spreads from the originally inoculated cells, further propagates and moves to adjacent and distant cells (Wise et al. 2007; Yang et al. 2007). Therefore an ideal indicator of plant virus infection needs to be present in all tissues across the time course of infection; else results could be greatly influenced by when and where test samples are taken and tissue type collected.

Most, if not all, plant species are susceptible to a number of viruses and a given virus may infect one or hundreds of plant species. A single plant can be infected by more than one virus concurrently, however, a plant is either a host for a particular virus or it is not. Species or cultivars of plants that do not provide one or more factors essential to the survival of a given virus, or that have mounted a resistance response, are unaffected by that virus and considered a non-host; and the plant-pathogen interaction considered incompatible. In contrast, successful viral infection and subsequent disease (compatible plant-pathogen interaction) will occur if environmental conditions are favourable, if the preformed defences of the plant are inadequate, and if the plant fails to detect the virus and activate an effective defence response (Agrios 2005).

In many cases, plant viruses cause a reduction in the growth, yield and quality of the infected plant and/or its fruit. Reduction in plant growth often results in reduced vigour that in turn increases plant sensitivity to abiotic stresses such as frost and drought, predisposing the infected plant to further pathogens (Hull 2002; Agrios 2005; Xu et al. 2008). Conversely, a reduction in size could reduce water requirements, for example, through reduced leaf area and
stomata, and as such improve plant survival in drought (Xu et al. 2008; Harb et al. 2010). In fact, despite the widely held belief that viruses are harmful to the host there is an increasing understanding that this is not always the case for all viruses. For example, Xu et al. (2008) found that four different RNA viruses extended the survival of eight plant hosts (beet, pepper, watermelon, cucumber, tomato, rice, Chenopodium and tobacco) under conditions of drought and cold stress. Improving tolerance to these abiotic stresses was evidenced by a correlation with increased concentrations of osmoprotectants and antioxidants in infected plants compared to mock-inoculated. Westwood et al. (2013) corroborated these results with Cucumber mosaic virus (CMV) infection in Arabidopsis thaliana (henceforth Arabidopsis). Studies of asymptomatic (cryptic) plant viruses, where there are no outward signs of disease, usually find no effect on the host. However, it is not always possible to find a plant of the same or similar genotype free of specific pathogens for comparison, and subtle effects of a virus are not likely detected. Asymptomatic viruses do have the potential to exert a profound effect on their plant hosts for example: as epigenetic elements providing novel genes and as sources for newly emerging viruses (Roossinck 2010; 2013).

Deep sequencing strategies have resulted in the discovery of numerous emergent and previously unrecognised viruses and the suggestion that viral infections are ubiquitous in plants and animals (Barzon et al. 2011; Roossinck 2011). Furthermore, deep sequencing has suggested that mixed viral infections may be the rule rather than the exception (reviewed in Syller 2012). Mixed virus infections may be synergistic, antagonistic or neither in their interaction. A synergistic infection usually facilitates replication of both viruses in a plant host. For example, the interaction between Potato virus Y (PVY) and Potato virus X (PVX) in tobacco plants results in enhanced disease symptoms and up to a 10-fold increase in titre of both viruses. In an
antagonistic infection only one virus benefits and the presence and activity of the second virus are reduced. Antagonistic infections occur when a previous infection with one virus prevents or interferes with subsequent infection by a second homologous virus (reviewed in Syller 2012).

Regardless, any reduction in size, yield and quality of produce as a result of virus infection frequently results in a reduction in market value and in an increase in costs of disease control (Hull 2002; Agrios 2005).

1.2 Methods for the detection of plant viruses

Methods for the detection and identification of plant viruses play a critical role in disease management, yet detection of asymptomatic and unknown viruses remains one of the most challenging problems in plant virology. Current methods for the detection of plant viruses usually require some prior knowledge of the virus or at least a strong suspicion as to what virus is causing symptoms. Enzyme-linked immunosorbent assay (ELISA) requires selection and preparation of appropriate antisera. Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR), hereon in termed PCR, requires that primers are designed to conserved genes of the suspected virus. Microarrays require that a probe for a suspected virus, and its close relatives, is present in the array (Boonham et al. 2007). Microscopy enables observation of virus particles and/or inclusion bodies indicative of different plant virus families however, has severe limitations and is less specific than ELISA and PCR (Hull 2002). The presence of double-stranded RNA (dsRNA) can be either indicative of virus or the replicative form of an ssRNA virus. Sequencing of dsRNA can be used to detect viral nucleic acids when a virus species cannot be predicted from symptoms or from a general survey using serology or PCR (Mielke and Muehlbach 2007), while identification of viral coat proteins is a useful method for the initial detection of both known and novel viruses (Blouin et al. 2010). All of these methods detect
properties of the plant virus itself. While advances have been made in sampling procedures and molecular characterisation of viral pathogens from plant tissues, and sequencing of dsRNA and next-generation sequencing (NGS) enable the detection of novel or unknown viruses, (Roossinck 2011), the biological impacts of asymptomatic viruses in novel hosts and environments remain unknown. It would be useful therefore to have a tool complementary to these methods that focuses on the defence response of plants themselves.

1.3 Initial defence mechanisms of plants against viruses

Plants are continually defending themselves against viruses and have developed effective mechanisms to prevent or limit damage due to viral infection. Plant innate immunity is extraordinarily complex and flexible in its capacity to recognise and counteract different pathogens, whether bacterial, fungal or viral (Pieterse and Dicke 2007). Each plant cell possesses both constitutive and inducible defences. This is in contrast to the vertebrate immune system where in addition to innate immunity, specialised cells dedicated to defence are rapidly mobilised to the site of infection (adaptive immunity), either eradicating the invading organism or limiting its spread (van Loon et al. 2006). Yet, the defence mechanisms of plants are thought to minimise infections as evidenced by wild plant populations where most plant species remain asymptomatic and healthy despite frequent infection with persistent viruses (Goldbach et al. 2003; Hammond-Kosack and Parker 2003; Nürnberger et al. 2004; van Loon et al. 2006; Pieterse et al. 2009; Roossinck 2010, 2011, and 2013). However, this suggests the association of viruses with wild plants could be more a case of tolerance, resulting in asymptomatic infections, rather than absolute resistance.
Because of the complexity of plant immunity, a comprehensive review of the current understanding of all defences that plants mount against viruses is beyond the scope of this thesis, however, key pertinent processes are highlighted. It is generally accepted that the alterations in plant physiology upon virus infection can be separated into four key divisions of defence (Figure 1.1) (Whitham et al. 2006). The first division incorporates the up-regulation of defence and resistance genes in response to a hypersensitive response (HR). The second division involves the induction of heat shock proteins (HSPs) in response to accumulating viral proteins. The third division involves the induction of hormone responsive genes in response to an imbalance of phytohormones (Peleg and Blumwald 2011). The fourth division encompasses the RNA interference (RNAi) pathway where small interfering RNAs (siRNAs), from a complex pool of small RNAs (sRNAs), guide sequence-specific cleavage of complementary target messenger RNA (mRNA). The functions of the microRNAs (miRNAs) of this RNAi pathway are thought to be disrupted by the presence of virus-encoded suppressors of RNAi (VSRs) resulting in defective plant development (Dunoyer et al. 2004). These four divisions provide a foundation from which to identify molecular indicators of virus infection, therefore they are the focus of this introduction and thesis.
Figure 1.1. Host responses and altered gene expression associated with plant virus infections.

Compatible interactions between susceptible hosts and virulent pathogens frequently lead to the expression of suites of defence-related (1) and heat shock proteins (2). The expression of these genes is controlled by signalling pathways in the plants associated with initiation of defence responses and by other cellular stress response pathways. Viral infections, through the action of phytohormone signalling or biosynthesis can lead to developmental defects (3) and the action of virus encoded silencing suppressors can also disrupt the functions of regulatory small RNAs, such as micro-RNAs (4). Adapted from Whitham et al. 2006.

The four divisions of defence have been grouped into two categories; cellular stress and developmental defects (Figure 1.1). Cellular stress responses include the induction of defence and pathogenesis-related (PR) genes in response to the HR and to increasing salicylic acid (SA), and by the induction of HSPs. Developmental defects are characterised by the induction of hormone responsive genes and alteration of miRNA function (Chellappan et al. 2005; Whitham et al. 2006; Groenenboom and Hogeweg 2008).

Each of the cellular stresses and developmental defects outlined above are considered in more detail below.
1.4 Cellular stress responses

1.4.1 The hypersensitive response (HR), defence genes, salicylic acid (SA) and systemic acquired resistance (SAR).

Early cellular stress responses are initiated by resistance genes immediately upon detection of a viral, fungal or bacterial pathogen or nematode bearing a cognate elicitor. The associated cellular and physiological features associated with this response have been well characterised (reviewed in Coll et al. 2011). Termed the HR, it is characterised by necrotic local lesions where cells within and surrounding an initial infection site undergo programmed cell death (PCD) as the result of a rapid burst of reactive oxygen species (ROS). This mechanism is thought to contribute to resistance by physically isolating an infection (reviewed in Soosaar et al. 2005). The HR response only occurs in incompatible host plant-pathogen interactions and is preceded by recognition of a pathogen-produced signal molecule (pathogen-associated molecular pattern (PAMP)) or avirulence (Avr) gene. In many cases this elicitor is the result of matching dominant gene products of the plant (resistance- or R-genes) and the elicitor of the pathogen (Bittel and Robatzek 2007; Sheen and He 2007; Tena et al. 2011).

R-genes confer resistance to pathogenic organisms, including viruses, bacteria, fungi, nematodes and insects that encode matching or recognised elicitors (Jones and Dangl 2006; Jones et al. 2008). Recognising and responding to pathogens in an immediate manner and conferring broad-spectrum resistance is similar to components of mammalian innate immunity. However, unlike the mammalian adaptive immune system, the HR does not confer long-lasting protective immunity to a host (Bittel and Robatzek 2007). Although several components of an R-gene-mediated signalling network have been identified, the mechanisms by which R-genes and their signalling components stimulate host plant defence remain poorly understood (He et al. 2007; Padmanabhan et al. 2009; Tena et al. 2011). R-gene resistance is typically associated with
an incompatible reaction and visual cues of this reaction are likely observable, meaning virus infection is likely to be contained or less likely to exist. Most \( R \)-genes are host/pathogen specific, suggesting \( R \)-gene resistance is not a defence mechanism used by all plants to all viruses. For these reasons, \( R \)-genes and their downstream signalling network have not been investigated in this thesis as potential indicators of plant virus infection.

Although plants do not produce circulating antibodies, and most biochemical defences require an elicitor from a given pathogen to be activated, plants have developed a generalised systemic resistance in response to infection or to treatment with specific exogenously applied natural or synthetic compounds (Agrios 2005). The phytohormone SA triggers signalling pathways that result in the production of long distance resistance signals remote from sites of initial infection to uninfected tissues. Long-lasting and broad-spectrum immunity is subsequently conferred (Kachroo et al. 2006; Park et al. 2007; Truman et al. 2007; Park et al. 2009; Pieterse et al. 2009). This phenomenon is termed systemic acquired resistance (SAR). SAR is analogous to the adaptive immunity of mammals in that a resistance response involving the recognition of ‘non-self’ occurs following an earlier exposure to a pathogen (Park et al. 2009). SAR functions in response to diverse pathogens and is not purely indicative of a response to virus infection. Therefore the SAR pathway has not been investigated in this thesis for potential indicators specific to plant virus infection.

1.4.2 Reactive oxygen species (ROS)

Plant-pathogen interactions are typically associated with an increased accumulation of ROS, which are small molecules or ions including free radicals and peroxides that are highly unstable. While produced as by-products of normal cellular functioning, and having important roles in cell signalling, excessive amounts of ROS can cause deleterious effects. During both
biotic and abiotic stress ROS levels have been demonstrated to increase dramatically above basal levels, resulting in oxidative stress and significant damage to cellular components and nucleic acids. A major contributor to oxidative damage is hydrogen peroxide (H$_2$O$_2$), which is converted from superoxide that leaks from mitochondria (Pnueli et al. 2003). Murphy et al. (1999) postulated that reduced electron flow through mitochondria in response to increased SA levels stimulates the generation of ROS and that ROS then act as signalling intermediates to stimulate defence responses. Cells are normally able to minimise the effects of ROS damage through the use of antioxidant enzymes such as superoxide dismutase (SOD) and glutathione S-transferases (GST) that convert the likes of H$_2$O$_2$ into oxygen and water. In fact, SOD and GST are induced in *Arabidopsis* and *Nicotiana benthamiana* in response to many different viruses (Whitham et al. 2006). However, ROS increase in response to a diverse range of pathogens and as such are not solely indicative of a response to virus infection. As such, specific ROS and the genes responding to them have not been investigated in this thesis as potential indicators of virus infection.

1.4.3 Heat Shock Proteins (HSPs)

All living organisms are continually challenged by endogenous and exogenous stresses that might disrupt cellular balance and lead to an instability of cellular protein machinery and incorrect protein folding. This destabilisation is corrected, in part, through the recruitment of a suite of chaperone proteins designed to either (i) refold or (ii) assist in targeting aberrant proteins for ubiquitin-proteasome degradation. One such class of chaperone proteins are the HSPs. HSPs are present in cells under normal conditions and function in normal growth and development, however, they are expressed at higher levels when cells are exposed to a temperature rise or other stress (Whitham et al. 2006; Yang et al. 2007; Verchot 2012). HSPs have been shown to
function in response to pathogens with HSP70 and HSP90 being essential components of the HR. Likewise, tombusvirus, tobamovirus and potyvirus genera require HSP70 and Bamboo mosaic potexvirus, carmovirus and dianthovirus genera require HSP90 to enable formation of membrane anchored replication complexes (Yang et al. 2007; Huang et al. 2012; Mine et al. 2012; Verchot 2012). In addition, some plant viruses encode chaperone proteins, or homologues thereof, in order to establish compatible infections. For example, members of the Closteroviridae utilise a self-encoded HSP70 homolog (HSP70h) as a virion component to assist the movement of virus particles through plasmodesmata (Aparicio et al. 2005).

Many plant RNA and DNA viruses have been shown to almost ubiquitously induce the expression of a number of HSPs in several diverse hosts, including pea (Pisum sativum), Nicotiana benthamiana and N. tabacum, squash (Cucurbita pepo) and Arabidopsis (Whitham et al. 2003; Senthil et al. 2005; Whitham et al. 2006; Verchot 2012). Whitham et al. (2003) performed microarray analyses of the interactions between five different positive single-stranded RNA (ssRNA) viruses (CMV; Oilseed rape virus, OSRV; Turnip vein clearing virus, TVCV; PVX; and Turnip mosaic virus, TuMV) inoculated to Arabidopsis leaves and provided further evidence that HSPs, more specifically HSP70, 83, 90, 17.4, 17.6A and 23.6, are rapidly induced in response to the two Tobamoviruses (OSRV and TVCV). The remaining viruses elicited HSPs at later time points of infection. It was concluded that a common mechanism controlling HSP expression in response to viral infection must exist however; this mechanism remains to be elucidated.

While there is much evidence to suggest that HSPs might be useful indicators of plant virus infection, HSPs accumulate in response to other biotic and to abiotic stresses (Pnueli et al. 2007).
As such, HSPs are not specifically indicative of virus presence and have not been investigated in this thesis as potential indicators of virus infection.

1.5 Developmental defect responses

1.5.1 Phytohormones

Phytohormones play important roles in regulating the complex networks of developmental processes and defence pathways. The SAR pathway that translates pathogen-induced early signalling events into the activation of defence responses depends on the coordinated action of many phytohormones, not just SA (Pieterse et al. 2009). The identification of several mutant plants affected in the biosynthesis, perception and signal transduction of a number of key phytohormones has enabled some progress to be made towards identifying phytohormones as key components of the plant response to pathogens (Bari and Jones 2009). Biotrophic pathogens are largely sensitive to defence responses regulated by SA whereas necrotrophic pathogens are sensitive to defence responses regulated by jasmonic acid and ethylene. To a lesser extent, auxin, abscisic acid, gibberellins, cytokinins, and brassinosteroids have also been implicated in plant defence (reviewed in Bari and Jones 2009 and Pieterse et al. 2009).

Pathogen induced hormone disruption could potentially create an imbalance in networks where every component is likely to be affected. Temporal and spatial studies quantifying hormone levels in detail across both the duration of an infection and across different cell and tissue types are lacking (Bari and Jones 2009; Grant and Jones 2009). Because of this complexity and the cross-talk, synergism and antagonism observed between phytohormone and defence pathways, (Mur et al. 2006; Koornneef and Pieterse 2008; Spoel and Dong 2008; Spoel et al. 2003; Atkinson and Urwin 2012).
(2009) and because phytohormone responses are not solely indicative of a response to plant virus infection, phytohormones have not been investigated in this thesis.

1.5.2 RNA interference (RNAi)

An important discovery in the area of plant defence against viruses is that of RNAi, also known as RNA silencing and post-transcriptional gene silencing (PTGS). RNAi describes conserved biological responses to dsRNA that occurs in all eukaryotes with the notable exception of the budding yeast, *Saccharomyces cerevisiae* (Baulcombe 2005; Großhans and Filipowicz 2008; Seo et al. 2013).

Briefly, RNAi is a sequence-specific mechanism of RNA degradation used to mediate resistance to endogenous and exogenous pathogenic nucleic acids resulting in both the negative regulation of gene expression and in defence against genetic parasites such as transposons and viruses (Bartel 2004; Baulcombe 2004; 2005; MacDiarmid 2005; Seo et al. 2013). The outcome of RNAi is either inhibition of gene expression at the transcriptional level, inhibition of gene expression at the translational level or in creating instability of mRNA. While there are several different processes that culminate at one of these three outcomes, in plants all processes share three key features: (i) the formation of dsRNA that may involve RNA-dependent-RNA-polymerase (RDR) activity, (ii) the cleavage of dsRNA by an RNase-III Dicer-like (DCL) enzyme into 21-24 nucleotide (nt) small RNAs (sRNAs) and (iii), the inhibitory action of a selected sRNA within an effector or RNA-induced silencing complex (RISC), containing an argonaute (AGO) RNA-binding protein (Brodersen and Voinnet 2006; Baulcombe 2007; Seo et al. 2013).

RNAi in plants is thought to act as an antiviral defence mechanism by surveying for aberrant RNAs including viral replication intermediates, dsRNA. Five lines of evidence support
the concept of RNAi as an antiviral mechanism in plants (reviewed in Hannon 2002). First, natural infection by plant viruses elicits strong PTGS (Lindbo et al. 1993). Second, viral replication can be suppressed by experimentally induced RNAi (Waterhouse et al. 1998; Waterhouse et al. 1999). Third, some components of the RNAi machinery for example, the RDRs, are up-regulated by virus infection (Schiebel et al. 1993). Fourth, mutations in *Arabidopsis* genes that encode some RNAi components, i.e. *RDR6* (*rdr6*), *SUPPRESSOR OF GENE SILENCING 3* (*sgs3*), and *SILENCING DEFECTIVE 3* (*sde3*), result in enhanced susceptibility to virus infection (Mourrain et al. 2000). Fifth, and perhaps the most convincing line of evidence, is that plant viruses encode a variety of inhibitors of RNAi, as counter-defence, known as virus-encoded suppressors of silencing (VSRs). Mutation or removal of these VSRs results in low titre virus replication (Voinnet et al. 1999; Goldbach et al. 2003; Kasschau et al. 2003; MacDiarmid 2005; Tomari and Zamore 2005; Blevins et al. 2006).

Because RNAi has been purported to serve as an antiviral mechanism in plants, the mechanism of RNAi has been investigated in this thesis as potential indicator of plant virus infection. The mechanism of the *Arabidopsis* RNAi pathway in response to RNA and DNA plant viruses is described in more detail below. An overview of the pathway is given followed by discussion of the function of key components based on the model of the RNAi pathway as it is currently understood (Figure 1.2).
1.5.3 Overview of the RNAi pathway in *Arabidopsis*

The RNAi pathway in *Arabidopsis* begins with the perception of dsRNA by a specific DCL and, in some cases, in association with a double-stranded RNA binding protein (DRB). As the trigger of RNAi, dsRNAs are derived from many different sources, including retrotransposons, endogenous RDR activity and viral replication. Long dsRNAs are ‘diced’ into primary sRNA species of 21-24 nt in length by a specific DCL (1, 2, 3 or 4). RNAi of RNA viruses in *Arabidopsis* is initiated by perception of viral dsRNA by *DCL4*. If *DCL4* is suppressed or inactive *DCL2* can replace it, generating 22-nt siRNAs. All four DCLs are involved in the production of 22–24 nt long siRNAs from DNA virus infections in *Arabidopsis*.

In the case of circular ssDNA geminiviruses, viral dsRNA formed by overlapping complementary viral RNA transcripts are recognised by *DCL3*, *DCL4* or *DCL2* (in order of preference). In dsDNA virus infections, such as *Cauliflower mosaic virus* (CaMV), RNAi is triggered by the 35S RNA leader sequence. Again, *DCL3* and *DCL4* are the most important dicers implicated in the production of viral secondary RNAs (vsiRNAs) derived from CaMV transcripts and *DCL2* activity is evident when *DCL4* is inactive. *DCL1* has a facilitating role, possibly rendering the 35S leader sequence more accessible to other DCLs. The *DCL3*-dependent 24 nt siRNAs are incorporated into *AGO4*. 
Figure 1.2. Antiviral silencing in Arabidopsis.

DCLs represented in association with known and unknown DRBs. DCL4 = primary DCL to detect RNA viruses, replaced by DCL2 if suppressed. AGO1 is the major antiviral slicer, but other AGOs are likely to be involved. All siRNAs are stabilised through HEN1methylation. Primary siRNAs (21 nt 1st) are amplified into secondary siRNAs (depicted as 21 nt 2nd) in the RDR6-dependent and DCL4-dependent silencing signal (movement) pathway. Movement may be enhanced through further rounds of amplification. Viral silencing suppressors are depicted in red. Adapted from (Ding and Voinnet 2007).

After a DCL catalyses the release of sRNA duplexes from their precursors, mature sRNAs are stabilised by methylation at their 3’ ends by HUA ENHANCER 1 (HEN1) in order to protect against 3’ uridylation and degradation. A single strand (termed guide strand) of a miRNA or siRNA duplex is selected and loaded by a DCL into an effector complex (RISC) containing
the major antiviral endonuclease slicer, \textit{AGO1}. Other members of the AGO family, such as \textit{AGO2} and \textit{AGO4}, may be also involved. The remaining (passenger) strand is degraded during the RISC assembly process. RISC uses the loaded sRNA strand as a sequence-specific guide to direct ‘slicing’ of the targeted sequence at the transcriptional or post-transcriptional level depending on what class of sRNA is loaded and which AGO protein is present.

Viral RNAs cleaved by RISC are then likely recognised as aberrant RNAs (abRNA) and converted to dsRNA by \textit{RDR6} action. This leads to the generation of secondary siRNAs (21 nt \textsuperscript{2nd}) that are thought to amplify RNAi and assist in cell-to-cell and systemic movement resulting in ‘immunisation’ of the whole plant ahead of an infecting virus. Secondary viral RNA (vsiRNA) biogenesis requires host \textit{RDR6} and its pathway cofactors \textit{SGS3, SDE3, SILENCING DEFECTIVE 5 (SDE5)} and \textit{SILENCING MOVEMENT DEFICIENT 1 through 3 (SMD1-3)}. While it is generally accepted that a mobile RNAi signal moves from cell-to-cell and systemically throughout the plant through plasmodesmata and vasculature, a demonstrated form of a mobile signal eludes researchers.

1.5.4 Double-stranded RNA (dsRNA) as a trigger of RNAi

As a distinct species of RNA, dsRNAs are derived from many different sources including viral replication (Figure 1.3) and/or endogenous RDR activity (Voinnet 2005a; Wassenegger and Krczal 2006).
For simplicity, full replication cycles are not shown. The ♦ symbol represents sources of dsRNA. a. Positive sense RNA viruses generate dsRNA due to partial or complete annealing of positive (+) and negative (-) viral RNA strands and folding of replicated, single-stranded genomic RNA forms secondary double-stranded structures. b. Geminiviruses (ssDNA) replicate bi-directionally and generate dsDNA intermediates. Transcription overlaps, converging at the 3' ends of sense/antisense transcripts, resulting in an overlap that forms dsRNA. c. The 35S RNA of CaMV has an extensive fold back structure at the 5' end called the translational leader that contributes significantly to dsRNA production. Adapted from Voinnet 2005.

Most plant viruses have genomes of positive, ssRNA that are replicated within the cytoplasm of the host. During replication of positive (+) single-stranded RNA (ssRNA) viruses, viral RNA (vRNA) is utilised as a template to generate complementary negative-stranded (-) genomic viral RNA (vcRNA), from which numerous copies of positive stranded RNA are reproduced. During processing, partial or complete annealing of positive vRNA and negative vcRNA strands constitutes the replicative form of RNA which provides one source of dsRNA. A second source is provided by the folding of replicated, single-stranded genomic RNA, which forms secondary double-stranded structures (Figure 1.3a). Viruses with genomes of negative ssRNA (not shown) follow a similar strategy, however genomic RNA must first be copied into a complementary, positive-stranded mRNA prior to translation of proteins (Voinnet 2005a).
Geminiviruses are ss-circular DNA (ssDNA) viruses and replicate bi-directionally in the nucleus through a rolling circle mechanism that generates dsDNA intermediates, templates for both transcription and replication. This bidirectional transcription overlaps, converging at the 3'-ends of sense/antisense transcripts, resulting in an overlap sufficient to form dsRNA to trigger RNAi (Figure 1.3b). The self-complementary fold-back structure of geminivirus transcripts could become a template for Dicer to cleave at specific locations and produce siRNAs directly (Chellappan et al. 2004; Vanitharani et al. 2005).

CaMV is a pararetrovirus with a dsDNA circular genome, where bidirectional transcription is not a feature of replication (Figure 1.3c). CaMV replicates by reverse transcription of an RNA intermediate by RNA polymerase II. This polymerase produces two viral RNA transcripts: 35S and 19S. The 35S RNA has an extensive fold back structure at the 5' end, called the translational leader. The translational leader facilitates ribosomal shunting required for expression of all open reading frames (ORFs) and contributes significantly to dsRNA production (Blevins et al. 2006; Moissiard and Voinnet 2006; Hohn and Rothnie 2013). In addition, terminal repeats at the ends of viral mRNAs form secondary double-stranded structures that could function as triggers for RNAi (Voinnet 2005a).

Regardless of the source of dsRNA ‘nature abhors a double-strand’ (Hutvágner and Zamore 2002) and dsRNAs are uncommon in healthy cells. Therefore dsRNAs have long been considered triggers of RNAi (Goldbach et al. 2003; Mello and Conte 2004; Zamore 2004; MacDiarmid 2005; Soosaar et al. 2005).
1.5.5 The composition of the pool of small RNAs (sRNAs) and their biogenesis

Once detected, DCLs cleave dsRNAs into smaller fragments, more specifically sRNAs of 21-24 nt in length. sRNA is a term that encompasses, to date, at least seven classes in plants: miRNAs, siRNAs, natural antisense transcript siRNAs (nat-siRNAs), repeat associated siRNAs (rasiRNAs), heterochromatin siRNAs (hc-siRNAs), trans-acting siRNAs (tasiRNAs) and vsiRNAs (Curtin et al. 2008; Meyers et al. 2008; Wang et al. 2010; Axtell 2013; Seo et al. 2013). sRNAs are not translated into proteins; instead, they negatively regulate a variety of biological processes essential to plant development, to stability of the genome and in responses to biotic and abiotic stress (Vaucheret 2006).

Plant virus infection can correlate with increased accumulation of sRNA species and in the accumulation and activity of enzymes involved in vsiRNA biogenesis. RNAi-mediated viral immunity in Arabidopsis requires host RDR1 or RDR6 for biogenesis of vsiRNAs following viral RNA replication-triggered biogenesis of primary siRNAs (Wang et al. 2010). As a result, and because vsiRNAs target an infecting virus; the siRNA fraction can increase in proportion to the miRNA fraction of the total accumulating sRNAs during virus infection (MacDiarmid 2005; Bazzini et al. 2007; Blevins et al. 2011).

Just as miRNAs and siRNAs have different functions, both have different modes of biogenesis. The precursors of siRNA are perfectly base-paired dsRNAs derived from virus replication, both the products of RDR activity, while miRNAs are first processed from specific genome-encoded precursors of long (c. 110 nt) ss-primary miRNAs (pri-miRNAs) then processed into 70 nt pre-miRNA. In Arabidopsis, pre-miRNA fold into hairpins containing imperfectly base-paired segments that are processed into a mature miRNA: miRNA passenger strand (miRNA: miRNA*) duplex in the nucleus by DCL1 in co-operation with the DRB,
HYponastic leaves 1 (HYL1), also known as DRB1, and henceforth termed DRB1 (Baulcombe 2003; 2004; Baulcombe and Molnár 2004; Baulcombe 2005; Vaucheret 2006; Eamens et al. 2008). Conversion to mature miRNAs in the nucleus is in contrast to animal miRNAs that require transportation from the nucleus to the cytoplasm by EXPORTIN 5 (EXP5) prior to maturation. The miRNA: miRNA* duplex of plants is exported to the cytoplasm by HASTY (HST), the plant orthologue of EXP5.

After DCL proteins catalyse the release of miRNA/miRNA* or siRNA duplexes from their precursors, mature miRNAs and siRNAs are methylated at their 3’ ends, a modification dependent on the RNA methyltransferase HEN1 that transfers a 2’-O-methylated base. Methylation is necessary to protect sRNA against 3’ end uridylation and subsequent degradation (Bartel 2004; Park et al. 2005; Mallory et al. 2008). Depending on the level of base pair complementation, miRNAs induce either mRNA cleavage or degradation (greater complementation), or repress translation (less complementation) without any appreciable destruction of mRNA (Zamore 2004; Zhang et al. 2006; Mallory and Bouche 2008).

1.5.6 Function of Arabidopsis DCLs

DCL proteins are key components in the miRNA and siRNA biogenesis pathways, processing long dsRNAs into mature sRNAs (Millar and Waterhouse 2005; Chapman and Carrington 2007). While considerable progress has been made in characterising Dicer (in animals) or DCL (plants) in model organisms such as Mus musculus (mouse), Drosophila melanogaster (fruit fly), Caenorhabditis elegans (nematode) and Arabidopsis, very little is known about plant species other than Arabidopsis (Yan et al. 2009). However, it has been postulated that all four DCLs present in Arabidopsis are likely to be present in all angiosperms and possibly all multicellular plants (Margis et al. 2006; Filiz and Koc 2013).
The number of dicer-like proteins varies among organisms. *Arabidopsis* encodes at least four DCLs; *DCL1, 2, 3 and 4* (Bartel 2004; Dunoyer et al. 2005; Blevins et al. 2006), while insects and fungi have genes that encode two Dicer-like proteins, and many animals including mammals have a single gene encoding one Dicer protein (Mukherjee et al. 2013).

A number of RNA viruses have been used to investigate DCL activity in *Arabidopsis*, including CMV, TuMV, TCV, TRV, and ORMV and the geminivirus; *Cabbage leaf curl virus* (CaLCuV), and caulimovirus; CaMV have been used to investigate DCL activity against DNA viruses. The functions of the four *Arabidopsis* DCLs are summarised in Figure 1.4 and discussed below.

i. *DCL1*

*DCL1* is involved in processing of both stem-loop pri-miRNAs and pre-miRNAs, in cooperation with *DRB1*, to mature miRNAs (Bartel 2004; Blevins et al. 2006; Liu et al. 2013). The resulting 21 nt miRNA/miRNA* duplex is loaded onto *AGO1* of the RISC (Baumberger and Baulcombe 2005). Blevins et al. (2006) have shown that *DCL1* generates 21 nt siRNAs from the CaMV leader region and therefore, *DCL1* might be essential to siRNA directed immunity against DNA viruses.

Moissiard and Voinnet (2006) reported that siRNA processing by DCL1 from CaMV is vastly suboptimal compared to processing by DCL3 (24 nt) and DCL4 (21 nt). A possible explanation for suboptimal viral siRNA processing is that the priority of DCL1 is to process and produce miRNAs in the nucleus of plant cells (Kurihara and Watanabe 2004). Therefore DCL1 can produce miRNAs and exogenous siRNAs (including vsiRNAs) whereas DCL2, 3 and 4 only produce siRNAs (Bouche et al. 2006).
Two adjoining plant cells are shown. The initially infected cell (left) contains high titres of virus. Arrows connecting dsRNA to DCL enzymes represent the relative contribution of each DCL to siRNA biogenesis. For a DNA virus, every DCL digests dsRNA into distinct size classes of siRNAs, with DCL3, DCL4 and DCL2 favoured (in that order). For RNA viruses, DCL4 prevails, with DCL2 and DCL3 able to compensate for any DCL4-deficiency. The 21 nt DCL4 product is potentially the signal required for silencing movement. Adapted from Blevins et al. 2006.

**ii. DCL2**

*DCL2* generates 22nt siRNAs from geminivirus and CaMV targets (Blevins et al. 2006) and 22nt siRNAs from diverse RNA virus transcripts (Bouche et al. 2006; Deleris et al. 2006; Aliyari and Ding 2009). *DCL2* is able to gain access to substrates normally processed by other DCLs especially when *DCL4* has been genetically removed or is inhibited by expression of a VSR such as *Turnip crinkle virus* (TCV) P38 (Blevins et al. 2006; Deleris et al. 2006). Further evidence to support *DCL2* ‘rescuing’ *DCL4* was found where *DCL2* processed 22 nt viral siRNAs from the 35S leader of CaMV when *DCL4* had been genetically inactivated. *DCL2* is
also thought to be antagonistic, partially suppressing the production of miRNAs and siRNAs by *DCL1* in certain circumstances (Bouché et al. 2006).

### iii. DCL3

*DCL3* processes endogenous dsRNA (mostly rasiRNAs from repetitive DNA loci) into 24 nt siRNAs, and is thought to mediate the establishment and maintenance of chromatin states by targeting both repetitive DNA and transposons (Hiraguri et al. 2005; Blevins et al. 2006). Deleris et al. (2006) suggested that *DCL3* is not associated with viral defence; however, Blevins et al. (2006) found that *DCL3* preferentially processes 24 nt geminivirus siRNAs and Moissiard and Voinnet (2006) found that *DCL3* processes CaMV 35S leader into 24nt siRNAs in absence of *DCL2* and *DCL4*.

### iv. DCL4

*DCL4* produces 21 nt primary siRNAs from RNA virus infections and produces tasiRNAs. Long range cell-to-cell spread or signalling of *Arabidopsis* RNAi is correlated with *DCL4* generated secondary 21 nt siRNAs in conjunction with *RDR6* amplification products and *HEN1*. This pathway has been shown to facilitate RNAi in uninfected cells ahead of an infection front and in new tissue growth (Dunoyer et al. 2005; Blevins et al. 2006 Bouché et al. 2006; Deleris et al. 2006). With respect to DNA viruses, *DCL4* efficiently processes CaMV 35S leader dsRNA into 21nt siRNAs (Moissiard and Voinnet 2006).
v. Summary of *Arabidopsis* DCL functions, redundancy and hierarchy

While the pathway for the production of specific sRNAs is generally considered to be associated with the particular DCL appropriate for the character of the sRNAs produced (Figure 1.4), considerable redundancy between DCLs has been observed. Functional redundancy masked findings from early studies using *dcl* mutants of *Arabidopsis* that were unable to convincingly link a given DCL to a given virus (Aliyari and Ding 2009).

Although virus derived dsRNA is accessible to all four DCLs under appropriate genetic circumstances, as aforementioned, *DCL2* and *DCL4* exhibit specific hierarchical antiviral activity. The primary dsRNA/viral RNA sensor is *DCL4* which produces 21 nt siRNAs. *DCL2* forms 22 nt siRNAs with antiviral activities that are most obvious when *DCL4* has been genetically removed or suppressed by VSRs (Blevins et al. 2006; Bouché et al. 2006; Deleris et al. 2006; Moissiard and Voinnet 2007), and in CaMV-infected *Arabidopsis* biogenesis of 35S leader derived siRNAs require the co-ordinated and hierarchical action of all four DCLs (Moissiard and Voinnet 2006).

In summary, viral siRNA biogenesis varies for different plant viruses and is determined by the DCL that cleaves the dsRNA (Blevins et al. 2006). DCLs not only process sRNAs, but also load sRNAs into the RISC (Qin et al. 2010). All four DCLs process replicating viral RNAs into siRNAs to some extent with a functional hierarchy of:

*DCL4>DCL2>DCL3>DCL1*

(Blevins et al. 2006; Bouché et al. 2006; Deleris et al. 2006; Curtain et al. 2008).

This multiplicity might help mediate plant responses to diverse virus infections (Blevins et al. 2006).
1.5.7 Guide strand selection and RISC assembly

Following cleavage by a DCL, a single guide strand of a miRNA or siRNA duplex is loaded by a DCL into a RISC containing an AGO, while the remaining passenger strand is degraded during RISC assembly. It remains unclear how the AGOs of plants remove the passenger strand during RISC maturation (Czech & Hannon 2011) however, specificity and selection of the guide strand has been investigated in several plant species.

While a duplex of sRNA is structurally symmetrical, asymmetry of the nucleotide sequence has been shown to generate a bias for preferred loading of a strand into a RISC complex. Chen et al. (2008) described an approach using HUMAN EMBRYONIC KIDNEY 293 cell lines where they were able to bias siRNA strand selection from duplex siRNA through phosphate sensing. Comparisons of 5′-O-methylated siRNA strands with 5′ phosphorylated strands showed that methylated siRNAs held reduced occupancy of the RISC complex compared to enhanced occupancy by phosphorylated strands. It was concluded that 5′-phosphorylation status of siRNA strands acts as an important determinant of strand selection (Chen et al. 2008).

While 5′-end phosphorylation and stability seem to contribute to strand selection, other results uncovered additional parameters influencing guide strand selection (Figure 1.5). Analyses of miRNA sequences that immunoprecipitate with AGO1 revealed that AGO1 preferentially selects 21 nt sRNAs with a 5′-terminal uridine (U), AGO2 shows a preference for 21 nt sRNAs with 5′ adenosine (A), AGO4 associates with 24 nt sRNAs with 5′ A and AGO5 preferentially selects sRNAs of 21, 22 and 24 nt with 5′ cytosine (C). It was not known what size sRNA or 5′ terminal preference AGO7 had (Kim et al. 2008) however, recent research found that AGO7 binds to miR390 by preferentially selecting miR390 with 5′ terminal adenosine (Endo et al.
It seems likely that a number of factors work in conjunction to sort sRNA strands for selection into RISC.

**Figure 1.5. Argonaute proteins (AGO) in Arabidopsis and their association with given sRNAs.**

The preferred 5′ terminal nucleotide for each AGO protein is indicated in red. Species and size of associated sRNAs are indicated underneath each AGO protein. AGO7 is thought to bind specifically to miR390 in a 5′ end-independent manner. However recent research shows AGO7 binds to miR390 with 5′ terminal adenosines (Endo et al. 2013). U, uridine; A, adenosines; C, cytosines; rasiRNA, repeat-associated short interfering RNA; tasiRNA, trans-acting siRNA. From Kim (2008).

### 1.5.8 RISC assembly and the structure and function of Argonaute (AGO) proteins

Once processed by a DCL from dsRNA substrates, a strand selected and modified, sRNAs are loaded into a Ribonuclease H (RNase-H)-like AGO protein to form the catalytic core of RISC. AGOs, the core components of RISC, are endonucleases - often referred to as ‘slicers’ - that play a central role in the cleavage of target RNA (Hock and Meister 2008). RISC uses the loaded sRNA as a sequence-specific guide to direct RNAi of the targeted sequence at the transcriptional or post-transcriptional level depending on what class of sRNA is loaded and which AGO protein family member is present (reviewed in Kim et al. 2011). The understanding
of the structure and function of AGOs remains far from complete. In fact, structural features have not been determined for any eukaryotic AGOs (Kim et al. 2011). However, the function of AGOs is clearly essential since genetic inactivation of a single AGO enhances virus susceptibility in many eukaryotic hosts (reviewed in Wang et al. 2011). Much of our current knowledge is based on ectopic expression studies of AGO purified from the bacteria Thermus thermophilis. Analyses have revealed that AGOs are large proteins comprised of a single variable N-terminal domain and conserved C-terminal domains: a Piwi, Argonaut and Zwille (PAZ) domain, a middle (MID) and a P-element induced wimpy testis domain (PIWI) (Vaucheret 2008; Wang et al. 2008). The N-terminal is thought to aid separation of the sRNA duplex after cleavage. The PAZ and MID domains recognise and anchor the 3’ and 5’ ends of the bound sRNA to its target mRNA. The third C-terminal domain specifies slicer activity.

In Arabidopsis, the AGO family comprises 10 members of which three (AGO1, 2 and 4) have been repeatedly shown to factor in protection against viruses. AGO1 and 2 factor in protection against RNA viruses while AGO4 factors in protection against DNA viruses (Qu et al. 2008; Takeda et al. 2008; Harvey et al. 2011; Wang et al. 2011; Schuck et al. 2013; Seo et al. 2013).

i. AGO1 and AGO2

AGO1 is the type member of the Arabidopsis AGO protein family, discovered in a forward genetics screen where plants exhibited developmental defects characterised by tubular shaped leaves that resembled tentacles of a squid of the genus Argonauta (reviewed in Kim et al. 2011). AGO1 directs sRNA-mediated gene expression regulation for all currently characterised Arabidopsis miRNAs (Vaucheret et al. 2004; Baumberger and Baulcombe 2005). AGO1 also
mediates the effector step for siRNA-directed RNAi where siRNAs may be derived from either an infecting virus or introduced transgenes (Kim et al. 2011). AGO1 is ubiquitous in plant tissues functioning in both the cytoplasm and the nucleus of plant cells but appears to function strictly in the cytoplasm when processing viral RNAs (Kim et al. 2011).

Studies have implicated the involvement of AGO2 in antiviral defence by showing that ago2 mutants are hyper-susceptible to TCV and CMV infection and that AGO2 expression can be induced with TCV and CMV inoculation of wild-type Arabidopsis (Harvey et al. 2011; Wang et al. 2011). The question was put forward; have plants evolved a back-up system against VSRs that target host AGO1? Harvey et al. (2011) did find that AGO2 provides a secondary antiviral mechanism that comes into play when the primary AGO1 layer is not active.

Wang et al. (2011) investigated the biogenesis of vsiRNAs with CMV mutants that were targeted for silencing by RDR6. Assessment of 25 single, double and triple mutants involving nine of the 10 Arabidopsis AGOs demonstrated a role for both AGO1 and AGO2 in defence against the mutant virus. Shuck et al. (2013) developed an in vitro system where RNAi could be reconstructed with a defined AGO protein (one of nine) and a siRNA of choice. It was found that the AGOs that showed the most slicer activity, AGO1 and AGO2, were also the most effective at inhibiting viral replication (Schuck et al. 2013). It appears then that AGO1 and AGO2 are the most effective at inhibiting viral replication.

ii. AGO4

Studies have indicated that some DNA plant viruses are targeted by an Arabidopsis AGO4-mediated gene silencing mechanism since Arabidopsis mutant ago4 plants show increased susceptibility to at least two geminiviruses, CaLCuV and Beet curly top virus. In Arabidopsis, AGO4 binds 24 nt sRNAs, and vsiRNAs detected in geminivirus-infected plants are
of 24 nt, therefore it is likely that \textit{AGO4} is involved in vsiRNA-directed inactivation of DNA viruses (Buchmann et al. 2009; Ding 2010, Raja et al. 2008). However, this finding remains to be demonstrated for CaMV.

\textbf{1.5.9 Amplification and local and systemic spread of RNAi}

In some organisms, including plants, RNAi is a non-cell-autonomous event meaning silencing initiated in one cell eventually leads to the silencing of the same sequence in a group of cells or whole organism (Kalantidis et al. 2008). There are two ways in which RNAi hinders viral replication in plants. The first is to limit the accumulation of viral RNA in infected cells by plasmodesmata–mediated short-range cell-to-cell silencing movement (reviewed in Nazim Uddin and Kim 2013). The second is to prime distant uninfected tissues for a rapid antiviral response by producing a phloem-mediated mobile and long distance systemic signal (Himber et al. 2003; Kalantidis et al. 2008; Chitwood and Timmermans 2010; Nazim Uddin and Kim 2013).

An important feature of the antiviral RNAi pathway in plants is referred to as amplification, whereby secondary siRNA substrates for DCLs are generated \textit{de novo} by RDRs (reviewed in Hunter et al. 2013). RDRs play a major role in synthesising the RNA strand complementary to a given ssRNA template, producing dsRNA. Secondary siRNAs arise by a process known as transitive silencing, where dsRNA is produced from a ssRNA template, outside of the region targeted by the initial dsRNA RNAi trigger, by RDR activity (Dunoyer et al. 2004; Mlotshwa et al. 2008).

\textit{Arabidopsis} encodes six RDRs designated \textit{RDR1–6}, of which three (\textit{RDR1, RDR2} and \textit{RDR6}), together with individual DCLs, control specific sRNA biogenesis pathways and have different molecular functions (Li and Ding 2006). In \textit{Arabidopsis} and other plants, \textit{RDR1} and
RDR6 contribute to antiviral RNAi, whilst RDR2 is involved in the control of chromatin structure and in establishment of transcriptional gene silencing (Hunter et al. 2013).

Wang et al. (2010) investigated the role of Arabidopsis RDRs in RNAi viral immunity by utilising a mutant of CMV that did not express the VSR, 2b, necessary for infection. The 2b-deficient mutant of CMV became pathogenic and accumulated to high levels in Arabidopsis when both RDR1 and RDR6 were not functional. This suggested an essential role for the amplification of viral siRNAs by either RDR1 or RDR6, or both. The findings of Wang et al. (2010) also demonstrate that in addition to components involved in dicing (DCL) and slicing (AGO), antiviral RNAi in plants requires RDR function to amplify vsiRNAs. In this regard RDRs of plants may have evolved for both a role in the biogenesis of endogenous siRNAs and in the amplification of vsiRNAs as a host adaptation to virus infection (Wang et al. 2010).

The purported roles of RDR1 and RDR6 in antiviral RNAi as reported in literature are discussed in more detail below.

i. RDR1 and RDR6

In Arabidopsis and N. tabacum, RDR1 is induced by exogenous SA treatment and by virus infection (reviewed in Xu et al. 2013). However, research to date has not managed to dissect the spatial and temporal antiviral mechanisms of RDR1 and no study has analysed the regulatory mechanism of RDR1 expression in plants. This is partly because little is known about the expression profile of RDR1 in different tissues. It is likely that RDR1 is under complex regulation by multiple endogenous signals (reviewed in Xu et al. 2013).

Although it is possible that RDR1 may contribute to SA-induced resistance to some viruses, RDR1 is unlikely to be essential to antiviral resistance. In some cases, RDR1 expression has been shown to enhance virus susceptibility (reviewed in Hunter et al. 2013). And while
evidence supports a general role of RDR1 in defence against viruses (Liu et al. 2009; He et al. 2010, reviewed in Xu et al. 2013), expression of RDR1 is also affected by various biotic and abiotic treatments. Hunter et al. (2013) found that wounding, specifically mechanical inoculation with virus, was sufficient to induce Arabidopsis RDR1 transcript. RDR1 induction in inoculated tissue is thought to be due to wounding, while induction in systemically infected tissue is thought to be due to localised induction of RNAi rather than a direct effect of the virus or its gene products (Hunter et al. 2013). Furthermore, leaves of Arabidopsis that have undergone mechanical inoculation exhibit strong resistance to the grey mould Botrytis cinerea, in a response independent of both tissue wounding and of known plant defence pathways (Benikhlef et al. 2013). In addition, RDR1 is regulated by a much broader range of phytohormones than previously thought. Hunter et al. (2013) showed that RDR1 gene expression was induced by not only SA but also by the phytohormones ethylene, abscisic acid and jasmonic acid, suggesting a role for RDR1 extending well beyond virus resistance.

While RDR1 has been implicated in viral siRNA biogenesis from several studies using Arabidopsis double and triple null mutants (Donaire et al. 2008; Qi et al. 2009; Wang et al. 2010; Blevins et al. 2011; Wang et al. 2011; Aregger et al. 2012), results are conflicting. Because of conflicting results and since evidence suggests a role for RDR1 in both biotic and abiotic stresses (Xu et al. 2013, Hunter et al. 2013), RDR1 has not been considered in the present research as a potential indicator of plant virus infection.

RDR6 is required for transitive silencing and enables plants to respond to the systemic silencing signal. Consistent with a role of host RDR6 in antiviral defence, loss of RDR6 activity increases susceptibility to virus infection (reviewed in Mlotshwa et al. 2008) and several VSRs function by blocking RDR6-dependent secondary siRNA biogenesis (Moissiard et al. 2007). Guo
et al. (2013) found that transient expression of the P6 protein of *Rice yellow stunt virus* in *N. benthamiana* leaves caused diminished accumulation of vsiRNAs and found that P6 binds both rice *RDR6 (OsRDR6)* and *Arabidopsis RDR6* in co-immunoprecipitation experiments. The results of Guo et al. (2013) suggest that P6 interferes with RNAi by altering the function of *RDR6* during vsiRNA synthesis thus restricting systemic RNAi. No effect of P6 on local RNAi was observed.

Host factors such as *SGS3, SDE3* and *SDE5*, in conjunction with *RDR6*, have been suggested to be crucial to the amplification of secondary mobile signals (Himber et al 2003). However, Guo et al. (2013) were unable to demonstrate any interaction between P6 and *SGS3* or *SDE3* in co-immunoprecipitation experiments. Discussion of *RDR6* interaction with *SGS3, SDE3* and *SDE5*, as reported by other researchers, follows.

**ii. *SGS3, SDE3* and *SDE5***

*SGS3* was initially isolated as a genetic loss-of-function mutant in *Arabidopsis* exhibiting increased susceptibility to CMV infection (Mourrain et al. 2000). It was later discovered that *SGS3* is the target of the *Tomato yellow leaf curl virus* VSR, V2, with a direct interaction blocking *SGS3* function and suppressing host RNAi (Glick et al. 2008).

Intensive sequence analysis found a novel XS domain (named after rice gene X and *SGS3*) to be conserved in plant *SGS3* proteins, rice gene X and other hypothetical proteins, although a function was still unknown (Zhang and Trudeau 2008). Utilising sequence analysis, fold recognition and structural modelling methods, Zhang and Trudeau (2008) demonstrated that the XS domain is an ssRNA-binding domain (RBD) that adopts a unique version of a RNA recognition motif (RRM) fold. The XS domain confers *SGS3* the ability to form a homodimer.
and bind to and stabilise newly generated dsRNA cleavage fragments with a 5’ overhang, derived from primary ss-siRNA precursors. When compared to a canonical RRM fold, the XS domain exhibits several unique structural characteristics that were concluded to define a distinct and specific RNA-binding mode adapted to the unique structural conformation of siRNA precursors (Zhang and Trudeau 2008).

Kumakura et al. (2009) observed co-localisation and interaction of SGS3 and RDR6 in specific cytoplasmic granules, which they named SGS3/RDR6-bodies. It was proposed that SGS3 not only stabilises RNAs, but also recruits them to RDR6 to initiate dsRNA synthesis in SGS3/RDR6-bodies. Yoshikawa et al. (2013) used an in-vitro assembled AGO1-RISC to demonstrate an interaction between SGS3 and RISC and established that SGS3 only binds within a RISC when a target sRNA has already bound to AGO1. The 3’ sRNA cleavage fragment was observed to be protected from degradation when within an AGO1-RISC-SGS3 complex. It was concluded that SGS3-dependent stabilisation of the 3’ fragment is crucial to siRNA production. The results of Zhang and Trudeau (2008), Kumakura et al. (2009) and Yoshikawa et al. (2013) suggest that SGS3 plays a pivotal role in RNAi however; the molecular function of SGS3 remains unclear.

Garcia et al. (2012) combined genetic, biochemistry and genomic studies to show that a putative RNA-helicase in Arabidopsis, SDE, assists RNAi amplification by RDR6. SDE3 possesses a series of WG/GW-rich motifs with conserved AGO-binding capacity and engages with AGO1 and AGO2 in RNAi amplification (reviewed in Karlowski et al. 2010). Strong silencing loss in the PVX-GFP- sde3 mutant system used indicated that RNA helicase function is central to the facilitating effects of SDE3 to RDR6 in RNAi. It was supposed that RDR6/SDE3 amplification machinery is likely to be required to enhance dsRNA production from limiting
amounts of transgenic or viral genomic RNA templates, thus attaining vsiRNA levels necessary to trigger optimal systemic silencing. SDE5 is thought to act with RDR6 in generating dsRNA from specific ssRNAs. SDE homologs found in Drosophila, C. elegans and humans contribute to silencing of viruses, transposons or recently duplicated genes. SDE5 in particular shares several sequence features with a human export factor responsible for RNA transport and as such is considered a putative plant homolog and transporter of ss-siRNAs to RDR6 that are then converted into dsRNA and vsiRNAs (Hernandez-Pinzon et al. 2007, reviewed in Jauvion et al. 2010). However, mechanisms of action, all protein partners and endogenous functions of SDE3 and SDE5 remain largely unknown (Jauvion et al. 2010).

1.5.10 Identity of the mobile silencing signal

A mobile silencing signal acts to immunise plants against systemic and secondary viral infections. Understanding of the silencing signal has stemmed from the study of Arabidopsis reporter systems in which inverted repeat hairpin derived RNAs expressed from the phloem companion cell specific promoter, SUCROSE TRANSPORTER 2 (SUC2), triggered non-cell-autonomous silencing of transgenes and endogenous transcripts (Himber et al. 2003; Dunoyer et al. 2005).

Studies with rdr6 mutant Arabidopsis reporter assays showed that silencing activity can only be detected at a short distance of approximately 10-15 cells outside the phloem. This short-range silencing has been found to be dependent on DCL4. Dependence on DCL4 led to the notion that 21 nt DCL derived siRNAs must be the signalling molecule, triggering production of dsRNA targets by RDR6. This dsRNA was then thought to be processed into secondary siRNAs by DCL4, which would then move cell to cell and further propagate systemic silencing through
amplification (Himber et al. 2003; Voinnet 2005b). Although reporter systems provide read outs of sRNA activity and demonstrate the presence of systemic RNAi, such assays have not been able to unequivocally demonstrate that 21 nt siRNAs constitute the silencing signal.

RNAi spreads over long distances from cell-to-cell and through the plant vasculature, presumably through plasmodesmata (Baulcombe et al. 2004). This spread probably immunises naïve cells surrounding infection (Deleris et al. 2006; Dunoyer et al. 2010). Using cell-specific rescue of DCL4 function and cell-specific inhibition of RNAi movement, Dunoyer et al. (2010), were able to establish that both exogenous and endogenous siRNAs (as opposed to their precursors) can act as mobile signals between plant cells.

It is believed that the main impact of RNAi in antiviral defence is at the level of systemic spread. However, the details of secondary siRNA production are not well understood and the identity of the silencing remains to be elucidated (Mlotshwa et al. 2008; Aregger et al. 2012). Furthermore, while it appears that sRNAs might function as non-cell autonomous signalling molecules, the understanding of the molecular mechanisms by which sRNAs are amplified and transported in plants remains in its infancy (reviewed in Uddin and Kim 2013).

Since RNAi is purported to serve as an antiviral mechanism in plants, the key components of Arabidopsis RNAi discussed above namely, sRNAs (specifically the accumulation of sRNAs), AGO1 and 4; DCL1-4; DRB1-5, HEN1, RDR2/SMD1, RDR6, SDE3 and 5; and SGS3 have been investigated in this thesis as potential indicators of plant virus infection.
1.6 Common suites of gene expression in virus infection

There has been considerable interest in identifying genes that play crucial roles in the signal transduction that controls activation of defence response pathways in plants (Glazebrook 2001). Genes that are either up- or down-regulated in response to the presence of plant viruses might prove valuable as potential indicators of virus infection.

As discussed in Chapter 1.1, analyses of gene expression in plant-virus interactions are complicated by the non-synchronous nature of a virus infection where plants are a heterogeneous mixture of cells responding to a virus as it replicates and spreads from the originally inoculated cells, propagates and moves to adjacent or distant cells (Wise et al. 2007; Yang et al. 2007). The progressive, time-dependent effects of virus infection on host gene expression have previously been investigated with in situ hybridisation of plant cotyledons. It was found that changes in gene expression occur at sites that contain the greatest amount of active viral replication and move as a wave with the replication front (Maule et al. 2002). While studies in cotyledons provide good resolution at a cellular level they are unable to provide a ‘global’ picture of host gene expression due to virus infection in true leaves and throughout a systemically infected plant (Yang et al. 2007).

Yang et al. (2007) explored the altered expression profiles of Arabidopsis genes in true leaves in relation to the spatial and temporal distribution of an infectious clone of TuMV expressing green fluorescent protein (GFP). Through GFP guided leaf dissection of four distinct zones of infection foci (Figure 1.6) at 120 hours after inoculation (hai), and DNA microarray profiling of mRNA transcript abundance, the authors were able to identify suites of genes with altered expression profiles during the time-course of TuMV infection.
Figure 1.6. Schematic representation for dissecting TuMV-GFP infection foci into four zones for spatial analysis of host gene expression.

Dissections taken at 120 hai using a 1.2-mm Harris punch. Zone 0 represents the epicentre, Zone 1 is adjacent to Zone 0 with a marginal area of intense GFP fluorescence, Zone 2 is at the infection periphery with weak GFP fluorescence and Zone 3 is distal with no direct contact to GFP fluorescent tissue. Adapted from Yang et al. 2007.

Genes that were identified as being up- or down-regulated by TuMV infection were classified according to Gene Ontology cellular component and the Munich Information Centre for Protein Sequences functional classification information. As expected, the genes identified spanned a variety of cellular locations and functions. Cell wall, intracellular localisation genes, genes encoding proteins localised to the chloroplast and involved in light harvesting and starch metabolism were overrepresented in the down-regulated suite of genes. Genes involved in cell rescue, defence and virulence - including HSPs and defence-related genes were overrepresented in the up-regulated suite of genes (Figure 1.7).
Figure 1.7. Relationship of *Turnip mosaic virus* (TuMV) accumulation to the expression of major functional groups of genes.

Gene functional classes are depicted in bold text in grey boxes and subcategories of genes within each functional class are depicted in plain text. The solid arrow represents virus accumulation and the dashed arrow represents the up- or down-regulation of the indicated gene functional groups. From Yang et al. 2007.

Results of quantitative real-time PCR (qPCR), hereon in termed qPCR, to evaluate expression profiles of up-regulated defence-related genes and HSPs showed that up-regulation was dependent on presence of virus and not in advance of a viral infection front. Therefore, TuMV does not induce systemic defence responses ahead of virus infection but rather in proportion to the amount of virus accumulation. This finding is in contrast to a study by Love et al. (2005) that showed CaMV induces expression of defence genes in systemic tissues ahead of the virus front.

Yang et al. (2007) identified, in *Arabidopsis* plants infected with TuMV, two functional classes of genes that had not previously been identified as being up-regulated in response to a
plant virus; genes involved in protein synthesis and protein fate. The protein synthesis genes that were identified, including *RECEPTOR LIKE PROTEIN 23 (RLP23)*, encoded ribosomal proteins while the protein fate genes encoded proteins assigned to targeting, sorting, translocation, assembly and degradation of proteins. It was postulated that these two suites of genes require presence of virus for their induction (Yang et al. 2007).

Whitham et al. (2003) found 114 genes in *Arabidopsis* that were induced in response to five RNA virus infections namely, CMV, ORMV, TVCV, PVX and TuMV. Babu et al. (2008) compared genes that were significantly up- or down-regulated in *Arabidopsis* in response to *Plum pox virus* (PPV) infection with the findings of studies of infections in *Arabidopsis* of the RNA viruses of Yang et al. (2007) and Whitham et al. (2003) and uncovered a common set of 416 genes responding to all six virus infections (i.e. CMV, ORMV, TVCV, PVX, TuMV and PPV).

Several other papers report suites of genes as differentially expressed in different plant species in response to diverse plant viruses. Agudelo-Romero et al. (2008) found 1727 genes displaying significant and consistent changes in expression levels in *Arabidopsis* infected with *Tobacco etch virus* (TEV). Microarray analysis of the *Arabidopsis* transcriptome in response to CaLCuV infection uncovered 5365 genes differentially expressed in infected rosette leaves at 12 dpi (Ascencio-Ibáñez et al. 2008). Alfenas-Zerbinii et al. (2009) inoculated tomato plants with *Pepper yellow mosaic virus* and constructed a subtractive library from inoculated leaves at 72 hai and identified several thousand genes as up or down-regulated including genes involved in plant defence, regulation of cell cycle, signal transduction, transcriptional regulation and stress responses.
More recently, Postnikova and Nemchinov (2012) performed a meta-analysis of data collected on compatible virus-host interactions in *Arabidopsis*. Viruses investigated were CalCuV, CaMV, CMV, *Lettuce mosaic virus* (LMV), PPV, TCV, TEV, TMV and TMV crucifer strain (TMV-Cg), *Tobacco rattle virus* (TRV), TuMV and ORMV. After subtraction of repeated genes, a list of 7639 genes responding to virus infection was obtained, representing 23% of the *Arabidopsis* genome. Each virus-host interaction was found to have different genes with altered expression. After subtraction of genes not common to all virus infections, the number of shared genes affected by all viruses was reduced to 198. Ultimately, the results of Postnikova and Nemchinov (2012) demonstrated that more than half of the 198 genes associated with all viral infections had previously been shown to respond to bacterial and fungal infections.

Similarly, Rodrigo et al. (2012) presented results of a meta-analysis of microarray data gathered from infections of *Arabidopsis* with seven RNA plant viruses (PPV, TEV, TuMV, TMV, TRV, TCV and TEV) and one DNA geminivirus (CaLCuV). Genes with altered expression levels were referred to as ‘virus-responsive genes’ (VRGs). Seven VRGs were found to be commonly up-regulated by six viruses however; each of these VRGs had previously been shown to respond to other biotic (including fungal and bacterial pathogens) and to abiotic stresses and as such not solely indicative of plant virus infection.

Despite the lack of encouraging results obtained in the references above, where most genes found to respond to numerous virus infections were also found to show generic stress responses, it was felt that pursuing genes with fold-change in response to virus infection was appropriate. There is likely to be hundreds, if not thousands of genes that have not been considered for their response to virus infection. For example, according to a review by Ouibrahim and Caranta (2013), the *Arabidopsis* genome presents a largely untapped source of
new genes and mechanisms involved in virus resistance. The method utilised to shortlist genes with fold-change in response to a number of virus infections for further investigation as part of the work of this thesis is described in Chapter 6.2.1.

### 1.7 Arabidopsis thaliana as a model system for the study of plant-virus interactions

Although *Arabidopsis*, a small flowering annual plant of the Cruciferae (formerly Brassicaceae) has little direct significance for agriculture, it has several traits that make it a useful model for understanding the genetic, cellular and molecular biology of flowering plants. First, being small in size and having simple growth requirements, *Arabidopsis* requires very little in the way of space, soil and maintenance and is convenient for the easy cultivation of large populations in laboratory conditions (Somerville and Koornneef 2002). Second, *Arabidopsis* has a very short life cycle: from seed germination, through formation of the rosette plant, bolting of the main stem, flowering and maturation of the first seeds in six to eight weeks (Meinke 1998; Leonelli 2007). Third, *Arabidopsis* has one of the smallest plant genomes making it amenable to detailed molecular analysis, mapping and sequencing, having been sequenced in 2000.

Annotation of the *Arabidopsis* genome is an ongoing process with The Arabidopsis Information Resource (TAIR) maintaining and updating assembly and annotation (Lamesch et al. 2012). Fourth, plant transformation in *Arabidopsis* using *Agrobacterium tumefaciens* to transfer DNA (T-DNA) to the plant genome is routine (Clough and Bent 1998; Leonelli 2007). Fifth, a collection of well characterised ecotypes have been collected from natural populations and are available for experimental analysis, with Columbia (*Col-0*) and Lansberg (*Ler-2*) ecotypes the accepted standards for genetic and molecular studies. Sixth, several thousand mutants defective in almost every aspect of plant growth and development have been identified and seed stocks maintained (Meinke et al. 1998). Seventh, because *Arabidopsis* is a widespread plant, and work
in the area of disease resistance has shown significant conservation across factors that mediate
defence between *Arabidopsis* and crop hosts (Carr 2007). *Arabidopsis* might harbour general
truths shared by many other plant species (Leonelli 2007). However, the extent to which
knowledge of cellular processes of *Arabidopsis* can be generalised to different plant species (and
to mammalian processes) is yet to be thoroughly tested (Bevan and Walsh 2004). *Arabidopsis*
was used as the experimental model in this thesis.

1.8 **Five viruses selected for the present research**

The above attributes that make *Arabidopsis* an excellent system for studies of molecular
genetics have contributed to its increasing usefulness as a model system for the study of plant-
virus interactions. *Arabidopsis* has been shown to serve as a compatible host for a diverse range
of viruses, including some of the most common and destructive viruses of widely grown crops
(Carr 2007; Ouibrahim and Caranta 2013).

The five viruses utilised in the present research; CaMV, TMV, TSWV, TuMV and
TYMV, were selected because they are known to readily infect *Arabidopsis*. The viruses selected
are also dissimilar in that they encompass both DNA and RNA viruses of different families and
genera with diverse replication strategies, gene expression, movement strategies and VSRs
(Table 1.1). This diversity lends itself to comprehensive studies of changes in *Arabidopsis* in
response to virus infection (Whitham et al. 2003). Précis of each of the five viruses with
reference to transmission, host range, particle morphology, symptomology, replication,
transcription, translation and VSRs are presented in Appendix I.
Table 1.1. Properties of the five viruses selected for the present research. Table summarises virus name and acronym, genus, genome sense and nucleic acid, virion morphology and virus-encoded proteins, including virus-encoded suppressors of RNAi (VSRs) known to interact with host plants. Information obtained from Brunt et al. 1996 onwards except where referenced in superscript.

<table>
<thead>
<tr>
<th>Virus name &amp; acronym</th>
<th>Genus</th>
<th>Genome sense &amp; nucleic acid</th>
<th>Virion morphology</th>
<th>Encoded proteins and RNAs known to interact with the host</th>
</tr>
</thead>
</table>
| **Cauliflower mosaic virus (CaMV)** | **Caulimovirus** | circular DNA | 50 nm isometric | MP (P1)\(^a\)  
ATF (P2\(^b\))  
DNA-binding (P3)\(^b\)  
TAV (P6)\(^a,c\) |
| **Tobacco mosaic virus (TMV)** | **Tobamovirus** | +ve ssRNA | 300 nm rigid rod | P126 –P183\(^d\)  
(MET, HEL & NONII)  
• P183 (replicase)  
• P50 (helicase)\(^d\)  
MP (P30)\(^d\)  
CP\(^d\) |
| **Tomato spotted wilt virus (TSVW)** | **Tospovirus** | ambisense ssRNA | 85 nm enveloped isometric | Cap-snatching\(^e\)  
RDR\(^e\)  
MP (NSm)\(^e\)  
NSs\(^e\) |
| **Turnip mosaic virus (TuMV)** | **Potyvirus** | +ve ssRNA | 720 nm flexuous rod | HC-Pro\(^f\)  
P3N-PIPO (P3)\(^g\)  
NII\(^b\) |
| **Turnip yellow mosaic virus (TYMV)** | **Tymovirus** | +ve ssRNA | 28 nm isometric | P140\(^i\)  
66\(^i\)  
tRNA-like structure\(^i\) |

\(^a\) Niehl et al. 2013  
\(^b\) Drucker et al. 2002  
\(^c\) Hohn 2013  
\(^d\) Liu and Nelson 2013  
\(^e\) Kormelink et al. 2011  
\(^f\) Lopez et al. 2008  
\(^g\) Chung et al. 2008  
\(^h\) Xiong and Wang 2013  
\(^i\) Jakubiec et al. 2004; 2007
1.9 Aims

As discussed in Chapter 1.1 New Zealand’s economy, and our crop and ornamental plant species are dependent on imported germplasm that needs to be free of plant viruses. While few of New Zealand’s native plants are cultivated as commercial crops, our native plant species maintain significant social, environmental and cultural value (MacDiarmid et al. 2013). As such New Zealand Biosecurity also seeks to protect native plants from virus infection. Current methods for the detection of plant viruses fail to address the need to be able to detect all plant viruses (known and unknown). It would be useful to have a tool based on a molecular indicator from plants that could be used to monitor their health status - any deviation from which might be suggestive of virus presence. Such a tool can be thought of as analogous to taking our own temperature rather than a pathogen specific test. This thesis documents the search for molecular indicators or ‘biomarkers’ of plant responses to virus infection with the long-term aim of developing tools for detecting virus-infected material.

This thesis addresses the development of assays to examine plant responses to virus presence by first checking that five dissimilar viruses (Cauliflower mosaic virus, CaMV; Tobacco mosaic virus, TMV; Tomato spotted wilt virus, TSWV; Turnip mosaic virus, TuMV; and Turnip yellow mosaic virus, TYMV) generate the same molecular responses in the model plant Arabidopsis - and later determining if the same responses occur in response to other biotic and abiotic stresses.
The specific objectives for this work were:

a) To confirm the ability of the five selected viruses (CaMV, TMV, TSWV, TuMV and TYMV) to infect *Arabidopsis* (Chapter 3).

b) To develop a method for the accurate quantification and analysis of sRNAs in response to five dissimilar viruses in *Arabidopsis* (Chapter 4).

c) To develop a qPCR method for the quantification of genes showing fold-change in response to five dissimilar viruses in *Arabidopsis* (Chapters 5 and 6).

d) To extend assays to a further biotic stress and two abiotic stresses *Arabidopsis* in order to determine if any potential indicators are solely indicative of virus infection (Chapter 6).
CHAPTER TWO – GENERAL MATERIALS & METHODS

Preamble

- All chemicals used were of analytical grade unless stated otherwise.
- Commercial supplies of reagents were prepared and used as advised by the manufacturers’ protocol.
- Arabidopsis thaliana ecotype Columbia (Col-0) plants are termed Arabidopsis and were the plant host selected for experimental studies undertaken.

2.1 Buffers, solutions and gels

Small RNA polyacrylamide gel electrophoresis

4x SDS PAGE Running buffer
- 12% (w/v) Tris
- 57.6% (w/v) glycine
- 0.1% (w/v) SDS
- This stock solution was diluted 4-fold with double-distilled water before use

2 x Gel loading Buffer
- 100 mM Tris, pH 6.8
- 4% (w/v) SDS
- 0.2% (w/v) Bromophenol blue
- 20% (v/v) glycerol
- 14 mM β-mercaptoethanol

Gel electrophoresis of RT-PCR products

50x TAE buffer
- 40 mM Tris, pH 8.0
- 20 mM acetic acid
- 50 mM EDTA
- This stock solution was diluted 50-fold with double-distilled water to make working solution

Agarose gel 1 and 1.5%
- 1x TAE buffer to 300 ml
- 0.30 g (1%) or 0.45 g (1.5%) agarose
- Dissolved and set as required

Ethidium bromide (EtBr) stain
- 200ml 1x TAE buffer
- 20 μl (0.5 mg/L) EtBr in a light protective bottle

Inoculation buffer
- 0.1 M potassium phosphate (K₂HPO₄) pH 7.4 at 4°C
- 0.1% (w/v) anhydrous sodium sulphite (Na₂SO₃) added to required volume immediate prior to use (e.g. 0.04 g Na₂SO₃ to 40 ml K₂HPO₄).
2.2 Selection of maintenance and propagation host plants

All susceptible herbaceous host plants chosen for mechanical inoculation of initial source viruses and as maintenance and propagation hosts of specific viruses were selected from the descriptions and lists from Plant Viruses Online VIDE Database: (Brunt et al. 1996 onwards) and are listed in Table 2.1.

Table 2.1. Maintenance and propagation plant hosts selected for the five viruses utilised in the present research.

<table>
<thead>
<tr>
<th>Virus name &amp; acronym</th>
<th>Maintenance &amp; propagation host</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauliflower mosaic virus (CaMV)</td>
<td>Brassica campestris subspecies (ssp). chinensis (pak choi)</td>
</tr>
<tr>
<td>Tobacco mosaic virus (TMV)</td>
<td>Nicotiana tabacum var. White Burley</td>
</tr>
<tr>
<td>Tomato spotted wilt virus (TSVW)</td>
<td>Lycopersicon esculentum (tomato)</td>
</tr>
<tr>
<td>Turnip mosaic virus (TuMV)</td>
<td>B. campestris ssp. chinensis</td>
</tr>
<tr>
<td>Turnip yellow mosaic virus (TYMV)</td>
<td>B. campestris ssp. chinensis</td>
</tr>
</tbody>
</table>

2.3 Growth of Arabidopsis seedlings and maintenance host plants

Arabidopsis seeds were sparsely scattered onto the surface of pots containing seed raising mix. Pots had base perforations for adequate drainage and were covered with translucent plastic sheeting to maintain required moisture and humidity until germinated seedlings were one week old. Plants were maintained in a shaded greenhouse set for 20 ± 2°C, where the ideal temperature range for growth of Arabidopsis is said to be 16-25°C (Weigel and Glazebrook 2002). A short day (SD) 8h photoperiod was implemented with 400W Philips SON-T Agro lights to provide photosynthetically active radiation of 120-150 μM m⁻² s⁻². Plants were grown under a SD photoperiod in order to suppress growth of reproductive tissue thus favouring growth of sufficient vegetative tissue mass required for all assays. With sub-irrigation preferable and to keep the soil moist, individually potted plants were placed in self-draining tubs containing capillary matting as described previously.
(Weigel and Glazebrook 2002; Eddy and Hahn 2006). An irrigation system with a timer was set to water the capillary matting for 12 min at three day intervals. This watering regime was optimised to provide appropriate moisture, growth and maintenance conditions for the Arabidopsis plants used in the present study. Maintenance host plants (Table 2.1) were grown under the same conditions.

2.4 Source of virus isolates

CaMV in freeze-dried Brassica campestris ssp. chinensis (pak choy) leaf material, TSWV in fresh chrysanthemum, and TuMV in freeze-dried turnip were kindly gifted by J. Fletcher and R. Lister (Plant & Food Research, Lincoln, New Zealand). TYMV in fresh B. campestris ssp. chinensis and TMV in frozen Nicotiana tabacum cv. White Burley were obtained from the School of Biological Sciences, University of Auckland, New Zealand. Upon receipt, the virus isolates were mechanically inoculated to their respective maintenance hosts as listed in Table 2.1 and to Arabidopsis, as described below.

2.5 Mechanical inoculation

Plants in the 3-5 leaf stage are said to be the most susceptible to effective virus inoculation as are plants that are inoculated in the evening (Walsh and Tomlinson 1985). Therefore all inoculations were undertaken in the early evening, on maintenance plant seedlings with 3-5 true leaves, and on SD Arabidopsis seedlings when rosettes measured approximately 4 cm in diameter (4-6 weeks of age).

A standard 0.1 M potassium phosphate (K$_2$HPO$_4$) inoculation buffer (pH 7.4), as described in Chapter 2.1, was prepared, autoclaved and stored at 4°C. Immediately prior to use, 0.1 % (w/v) anhydrous sodium sulphite (Na$_2$SO$_3$) was added to the buffer. This buffer was deemed suitable for mechanical inoculation of each of the five viruses onto Arabidopsis and maintenance host plants. Fresh plant tissue (100 mg) was harvested from plants showing
symptoms characteristic of each given virus, as described in Appendix I, (or the equivalent of freeze dried tissue with known infection for inoculation to maintenance hosts) and placed in a Bioreba mesh bag (Bioreba AG, Reinach, Switzerland) with 1 ml of inoculation buffer and carborundum (as an abrasive) and ground into an homogenate with a roller. Homogenates were drained from Bioreba bags into clean plastic weigh boats. Leaves to be inoculated were indicated by a small piercing with a pipette tip. Using a gloved hand, a small amount of homogenate was collected on a finger and gently brushed onto indicated leaves. Inoculation buffer and carborundum were used as a ‘mock-inoculation’ and plant tissue from mock-inoculated plants was used as a control. Shortly after inoculation leaves were misted with water to rinse off buffer and carborundum.

Post inoculation, all plants were put into individually allocated self-draining square tubs; i.e. one containing mock-inoculated plants only, one CaMV-inoculated, one TMV-inoculated, one TSWV-inoculated, one TuMV-inoculated and one containing TYMV-inoculated plants only. All efforts were made to keep trays of specific virus-inoculations separated to prevent co-infection. TMV-inoculated plants were kept behind a Perspex divider to prevent accidental mechanical inoculation. Mock-inoculated plants were kept separate from all virus-inoculated trays. Plants were observed on a daily basis for the development or absence of symptoms on inoculated leaves and in systemic tissues and the information recorded.
2.6 Experimental design

2.6.1 Virus inoculations

For each of the two main experiments 144 Arabidopsis plants were inoculated as previously described (Chapter 2.5) and the inoculated leaves from three plants harvested at 2, 3 and 7 dpi and systemic leaves from three plants harvested at 14, 21, 28, 35 and 42 dpi for each treatment. TuMV-inoculated plants had developed lethal necrosis by 35 dpi; therefore tissue from a total of 138 plants was collected (Table 2.2). Leaf tissue from each plant was individually packaged and labeled and immediately frozen in liquid nitrogen then transferred to −80°C for storage until RNA was isolated.

Table 2.2. Experimental design – virus inoculation. Numbers of Arabidopsis plants per treatment (mock- or virus-inoculation) and tissue collection time point (2, 3, 7, 14, 21, 28, 35 and 42 dpi).

<table>
<thead>
<tr>
<th>Treatment Time</th>
<th>Mock-inoculated</th>
<th>CaMV</th>
<th>TMV</th>
<th>TSWV</th>
<th>TuMV</th>
<th>TYMV</th>
<th># plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>3 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>7 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>14 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>21 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>28 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>35 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>42 dpi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>18</td>
<td>24</td>
<td>138 plants</td>
</tr>
</tbody>
</table>

For quantification of accumulation of sRNAs (Chapter 4), a time course of 2, 3, 7, 14, 21 and 28 dpi was used. For qPCR analysis of transcript accumulation of various genes (Chapters 5 and 6), a time course of 2, 3, 7, 14, 21, 28, 35 and 42 dpi was used, with the exception of TuMV treatment where plants had developed lethal necrosis by 28 dpi and no further tissue was available to harvest.
2.6.2 Abiotic stress treatments

For abiotic salt stress treatments, *Arabidopsis* plants were watered with either 100 mM NaCl or 200 mM NaCl at zero min. The leaves of three plants (total of 48 plants) were harvested after 15 min, 30 min, 1 h, 4 h, 24 h, 48 h, 7 d, and 14 dpi (Table 2.3). For drought stress treatments, watering ceased 7 d prior to leaf tissue collection at 7, 14 and 21 dpi (total of nine plants). Leaves from control plants (no treatment, i.e. normal watering regime) were collected across the entire time course (30 plants). Leaf tissue from a total of 87 plants was individually packaged, labeled and immediately frozen in liquid nitrogen then transferred to –80°C for storage until RNA was isolated. RNA isolation and cDNA synthesis was kindly performed by Gardette Valmonte (Plant and Food Research, Mt Albert, Auckland, New Zealand).

**Table 2.3. Experimental design - abiotic stresses.** *Arabidopsis* plants with no treatment, salt (NaCl 100 mM and 200 mM) or drought and tissue collection time point (0, 15, 30 min, 1, 4, 24 h, or 7, 14, 21 d).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control (no treatment)</th>
<th>100 mM NaCl</th>
<th>200 mM NaCl</th>
<th>Drought</th>
<th># plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>15 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>30 min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>1 h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>4 h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>24 h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>48 h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>7 d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>14 d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>21 d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>24</td>
<td>24</td>
<td>9</td>
<td>87 plants</td>
</tr>
</tbody>
</table>
2.6.3 Further biotic stress inoculation

In order to investigate the responses of genes of interest to a further biotic stress, 12 Arabidopsis plants were inoculated with *P. syringae* var. tomato strain DC3000, a pathogen of both tomato and Arabidopsis, in lysogeny Luria broth (LB). Arabidopsis plants that underwent no treatment (12 plants) and 12 mock-inoculated (LB only) plants were included as controls. Leaf tissue of three plants for each treatment (a total of 36 plants) was harvested at 1, 2, 6 and 10 dpi (Table 2.4). Plant inoculations, tissue collection, RNA isolation and cDNA synthesis were kindly undertaken by Gardette Valmonte (Plant and Food Research).

Table 2.4. Experimental design - *Pseudomonas syringae* var. tomato DC3000 inoculation. Control, mock inoculated and *P. syringae* var. tomato inoculated Arabidopsis plants sampled over a time course of 1, 2, 6 and 10 days post inoculation (dpi).

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Control (no treatment)</th>
<th>Mock-inoculated (LB)</th>
<th><em>Pseudomonas syringae</em> var. tomato DC3000</th>
<th># plants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 dpi</td>
<td><img src="image1" alt="Plant Image" /></td>
<td><img src="image2" alt="Plant Image" /></td>
<td><img src="image3" alt="Plant Image" /></td>
<td>9</td>
</tr>
<tr>
<td>2 dpi</td>
<td><img src="image4" alt="Plant Image" /></td>
<td><img src="image5" alt="Plant Image" /></td>
<td><img src="image6" alt="Plant Image" /></td>
<td>9</td>
</tr>
<tr>
<td>6 dpi</td>
<td><img src="image7" alt="Plant Image" /></td>
<td><img src="image8" alt="Plant Image" /></td>
<td><img src="image9" alt="Plant Image" /></td>
<td>9</td>
</tr>
<tr>
<td>10 dpi</td>
<td><img src="image10" alt="Plant Image" /></td>
<td><img src="image11" alt="Plant Image" /></td>
<td><img src="image12" alt="Plant Image" /></td>
<td>9</td>
</tr>
</tbody>
</table>

36 plants
2.7 Nucleic acid manipulations

2.7.1 RNA isolation, quantitation of concentration and determination of purity with Nanodrop and integrity with Bioanalyzer 2100 RNA Nano LabChip and DNAse I digest

Total RNA was isolated from no greater than 100 mg of frozen leaf tissue using a Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, U.S.A.) according to the manufacturer’s protocols. The concentration and purity and integrity of the RNA samples were determined using Nanodrop (ND-1000 Spectrophotometer; Nanodrop Technologies Inc., Wilmington, DE, U.S.A.). All samples had an absorbance ratio (absorbance at A260/280 nm) of between 1.8 and 2.2. Traditionally, RNA integrity has been measured by the absorbance ratio of 28S:18S ribosomal RNAs (rRNAs) where a value of 1.8 to 2.2 (2.1 = pure RNA) is considered acceptable for most downstream protocols (Udvardi et al. 2008). Following quantification, all RNA samples were normalised to 2.5 μg/μl. The integrity of RNA samples was further determined by micro-fluidics assessment with a Bioanalyser 2100 RNA Nano LabChip 6000 (Agilent Technologies, Santa Clara, CA, U.S.A.) where the software algorithm assigns a RIN determined from the entire electrophoretic trace of a total eukaryotic RNA sample, rather than determined solely from the 28S:18S rRNA ratio (Schroeder et al. 2006). All RNA samples were found to have RNA integrity (RIN) values of 6.0 or greater, where ten categories have been defined from 1 (totally degraded RNA) to 10 (fully intact RNA) (Schroeder et al. 2006). It has been recommended that only RNA samples with a RIN of ≥ 7.0 be used in downstream applications (Schroeder et al. 2006). However, total RNA from chloroplast containing tissue does not contain 28S rRNA but 25S rRNA and two additional rRNAs, 16S and 23S, thus adding two additional peaks that do not fit with the software algorithm designed to generate mammalian RIN values. As such, RIN values generated with earlier software do not apply to plant material, since this software was developed using mammalian tissue (De Keyser et al. 2013). It is recommended that
researchers consider both the RIN value and visually inspect the peaks of the electropherograms in order to determine RNA integrity of plants (Die and Román 2012). The peaks of the RNA of the present study were visually inspected and the peaks observed fitted with 25S, 23S and 16S RNAs thus determining integrity of plant RNA. A dedicated method to determine plant RIN values has since been introduced with the B.02.08 Service Release 2 (SR2) version of the Agilent 2100 Expert software in September 2012. Application notes were provided in 2013 (Babu and Gassmann 2013).

Any potential genomic DNA (gDNA) contamination was removed using DNase I (amplification grade; Invitrogen, San Diego, CA, U.S.A.) before 2.5 μg/μl of RNA was reverse transcribed into cDNA using SuperScript® VILO™ cDNA Synthesis Kit (Life Technologies-Invitrogen, San Diego, CA, U.S.A.) according to the manufacturers’ protocols.

2.7.2 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

In order to determine cDNA of plant origin and to confirm virus presence, a 25 μl reaction volume PCR was employed using 2.0 μl (2.5 ng/μl) cDNA, 12.5 μl of GoTaq® Green Master Mix (Promega Corp., Madison, WI, U.S.A.), 20 pmol each (sense and antisense) gene-specific primer and UltraPure™ DNase/RNase-free distilled water (Life Technologies – Invitrogen San Diego, CA, U.S.A.) to 25 μl (8.5 μl in cDNA template samples, 10.5 μl in cDNA deficient control reactions). Gene specific oligonucleotide primers were designed using NCBI Blast and Primer3 software (described in Chapter 2.8 and 2.9) and purchased from Sigma-Aldrich (Castle Hill, NSW, Australia). Primer pairs were reconstituted according to manufacturer’s specifications, and the sequences of each pair are presented in Table 2.5.

Amplification of cDNA employed the following conditions: initial denaturation for 5 min at 94°C followed by 38 cycles of denaturation 94°C for 30 s, annealing at 60°C for 30 s
and extension at 72°C for 30 s, followed by a final terminating extension for 5 min at 72°C. Negative control reactions were prepared, one omitting reverse transcriptase (RT-ve) to verify absence of DNA template and one omitting cDNA template (cDNA-ve) to verify absence of other, contaminating templates.

2.8 Design of primers to amplify internal controls

To determine whether the cDNA was of plant origin, and of good quality the NADH dehydrogenase subunit five from mitochondria of the higher plants (NAD5), with an expected amplicon of 181 bp was used as a positive control for cDNA from maintenance and propagation hosts. In later experiments Arabidopsis Elongation factor 1-alpha (At5g60390 – EF1α), with an expected amplicon of 137 bp, was used to determine that cDNA was of Arabidopsis origin and of good quality. The PCR product of each internal control primer pair (EF1α and NAD5) spanned an intron so that they would be amplified from all the known splice variants (as annotated on TAIR) of plant mRNA and not from plant gDNA. A larger amplicon could be expected if gDNA was present in samples.

To ensure maximum specificity and efficiency during PCR amplification under a standard set of reaction conditions, primers were designed using the Primer3 software, version 0.1 (Rozen and Skaletsky 2000). Parameters were optimised such that melting temperatures were of 60°C for amplicons greater than 200 bp and 65°C (< 150 bp), with a maximum difference of ± 3°C; had a length of 20 to 24 bases and GC content of 45 to 50%. Primers with amplicons of less than 150 bp were designed for qPCR. The descriptor, melting temperature, and amplicon size for each commonly targeted gene are presented in Table 2.5. Primer design for genes targeted specifically in Chapters 5 and 6 are detailed in each respective chapter.
2.9 Design of primers to amplify virus targets

Each virus required confirmation of infection and this was performed by PCR as described above (Chapter 2.7). Nucleic acid isolation for CaMV followed the same protocol for RNA isolation and cDNA generation outlined in Chapter 2.7. The rationale being that despite being a DNA virus, CaMV is a pararetrovirus with a dsDNA circular DNA genome and replicates through reverse transcription of a full length RNA intermediate. This 35S pre-genomic RNA is polycistronic and a complete transcript of the circular viral genome; therefore RNA isolation and cDNA generation are appropriate for amplification of the CaMV genome (Blevins et al. 2011). The designs of primer pairs to amplify each virus target are outlined in turn below and the descriptor, melting temperature, amplicon size and reference where applicable, for each virus primer pair are presented in Table 2.5.
Table 2.5. Oligonucleotide primers used for cDNA amplification of viruses and internal controls.

<table>
<thead>
<tr>
<th>Primer name and position</th>
<th>Targeted gene AGI and descriptor</th>
<th>5’→ 3’ oligonucleotide primer sequences and targeted nucleotides</th>
<th>Primer length (nt)</th>
<th>Tm (°C)</th>
<th>Amplicon size (bp) and reference where applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>*EF1α F(922-941)</td>
<td>At5g60390 – Elongation factor 1a. Elongation factor activity and involved in calmodulin binding.</td>
<td>CACCACCTGGAGGTTTTGAGG TGGAGTATTTGGGGGTGGT</td>
<td>20</td>
<td>64.5</td>
<td>137</td>
</tr>
<tr>
<td>*EF1α R(1158-1140)</td>
<td></td>
<td></td>
<td>19</td>
<td>64.0</td>
<td></td>
</tr>
<tr>
<td>NAD5 F(968-991)</td>
<td>AtMG00060 –Mitochondrial NADH dehydrogenase subunit 5. Electron transport, respiration, redox, photosynthesis.</td>
<td>GATGCTCTTCTGGGGCTTCTTGGT CTCCAGTCATCAATCCATGGCATTA</td>
<td>23</td>
<td>63.6</td>
<td>181</td>
</tr>
<tr>
<td>NAD5 R(1973-1949)</td>
<td></td>
<td></td>
<td>24</td>
<td>63.8</td>
<td></td>
</tr>
<tr>
<td>CaMV–V–F(4511-4530)</td>
<td>Cauliflower mosaic virus ORF V. Reverse transcriptase</td>
<td>CGAGAAGCGAAGGAAAGA ACACGAAATGCTTCGTCCAT</td>
<td>20</td>
<td>58.5</td>
<td></td>
</tr>
<tr>
<td>CaMV–V–R(4810-4971)</td>
<td></td>
<td></td>
<td>20</td>
<td>58.6</td>
<td>299</td>
</tr>
<tr>
<td>*CaMV – P35S– F(89-110)</td>
<td>Cauliflower mosaic virus 3SS RNA translational leader sequence. Constitutive promoter for transcription of genome.</td>
<td>CACGCTGAAATGCTTCGTCCAT</td>
<td>20</td>
<td>64.0</td>
<td></td>
</tr>
<tr>
<td>*CaMV – P35S – R(208-188)</td>
<td></td>
<td></td>
<td>20</td>
<td>63.8</td>
<td>120</td>
</tr>
<tr>
<td>TOBO (UNI2)-F(5479–5498)</td>
<td>Tobacco mosaic virus movement protein (MP)</td>
<td>GTYGGTGTAGTGGTCTTGGGA ATTTAAGTGAGGAAGAGGAAACCTAA</td>
<td>20</td>
<td>62.9</td>
<td></td>
</tr>
<tr>
<td>TOBO (UNI1)-R(6283-6260)</td>
<td>Tobacco mosaic virus coat protein (CP)</td>
<td>ATTTAAGTGAGGAAGAGGAAACCTAA</td>
<td>20</td>
<td>63.1</td>
<td>120</td>
</tr>
<tr>
<td>*TMV-F(5829-5848)</td>
<td>Tobacco mosaic virus coat protein</td>
<td>CAAGCTCGAACTGTCGTTCA</td>
<td>20</td>
<td>64.2</td>
<td></td>
</tr>
<tr>
<td>*TMV-R(5948-5929)</td>
<td></td>
<td></td>
<td>20</td>
<td>63.5</td>
<td>120</td>
</tr>
<tr>
<td>TSWV L2-F (4213-4233)</td>
<td>Tomato spotted wilt virus gene for RNA polymerase, complete cds</td>
<td>ATCGAGTCAAATGCTCGCA AATTGCTTTGCAACCAATTC</td>
<td>20</td>
<td>62.3</td>
<td>276</td>
</tr>
<tr>
<td>TSWV L1-R (4396-4376)</td>
<td></td>
<td></td>
<td>20</td>
<td>59.9</td>
<td></td>
</tr>
<tr>
<td>*TSWV MF(564–584)</td>
<td>Tomato spotted wilt virus medium RNA segment (M-RNA)</td>
<td>CCAACATGCATCTGGAAGAC AAACGTCATGTCATTTTG</td>
<td>21</td>
<td>64.0</td>
<td>137</td>
</tr>
<tr>
<td>*TSWV MR (678–700)</td>
<td></td>
<td></td>
<td>22</td>
<td>63.2</td>
<td>(Debreczeni et al. 2011)</td>
</tr>
<tr>
<td>TuMV F(5356-5376)</td>
<td>Turnip mosaic virus cytoplasmic inclusion protein (CI)</td>
<td>AGGTGTAGTCAGGTTTCGTGC</td>
<td>20</td>
<td>58.4</td>
<td>244</td>
</tr>
<tr>
<td>TuMV R(5599-5580)</td>
<td></td>
<td></td>
<td>20</td>
<td>58.4</td>
<td></td>
</tr>
<tr>
<td>*TuMV F(8463-8482)</td>
<td>Turnip mosaic virus Nlb protein nuclear inclusion protein B (Nlb). RNA replicase</td>
<td>CATGAATGGGATCGTCGAA TCAACCGCTGAAGC</td>
<td>20</td>
<td>63.8</td>
<td>121</td>
</tr>
<tr>
<td>*TuMV R(8583-8563)</td>
<td></td>
<td></td>
<td>20</td>
<td>63.8</td>
<td></td>
</tr>
<tr>
<td>TMYV-MP- F (247-288)</td>
<td>Turnip yellow mosaic virus movement protein (MP)</td>
<td>CACCATCCATCTCCTTTGC CGGTGATGGAGTAGGAGT</td>
<td>20</td>
<td>58.8</td>
<td>247</td>
</tr>
<tr>
<td>TMYV-MP- R(475-456)</td>
<td></td>
<td></td>
<td>20</td>
<td>58.5</td>
<td></td>
</tr>
<tr>
<td>*TMYV-CM-F(5978-5997)</td>
<td>Turnip yellow mosaic virus coat protein (CP)</td>
<td>GTGGCCAGATCTTCTGCAATT GGGGAGTCAAGGGTACTCAGA</td>
<td>20</td>
<td>64.1</td>
<td>119</td>
</tr>
<tr>
<td>*TMYV-CM-R(6096-6077)</td>
<td></td>
<td></td>
<td>20</td>
<td>64.1</td>
<td></td>
</tr>
</tbody>
</table>

* Primers designed to amplify shorter products (< 150 bp) for use in qPCR reaction
2.9.1 CaMV primers

One pair of CaMV primers was designed to target a 299 bp region of P5 (V) that encodes the viral reverse transcriptase, CaMV–V-F(4511-4530) and CaMV-V-R(4810-4971), while a second set was designed for qPCR reactions to target a region of 120 bp of the 35S RNA promoter, CaMV–P35S-F(89-110) and CaMV–P35S–R(208-188). The two primer pairs are annotated on their respective RNAs on the schematic diagram of the CaMV genome in Figure 2.1.

![Figure 2.1. Annotation of primers designed to CaMV V5 and 35S regions.](http://commons.wikimedia.org/wiki/File:CauliflowerMosaicRNA35S.png)
2.9.2 TMV primers

Universal tobamovirus primers, TOBO(UNI2)-F and (TOBO(UNI1)-R, as annotated in blue in Figure 2.2A, have previously been used to amplify a tobamovirus product of 804 bp spanning the subgenomic RNAs (sgRNAs) of the MP and CP (Letschert et al. 2002) and these primers were used to amplify TMV. TOBO(UNI2)-F is positioned at 5479 to 5498 of both the genomic and sgRNA MP, while TOBO(UNI1)-R is positioned at 6283 to 6260 (within the 3’UTR) on the genome and sgRNAs. A set of qPCR primers was designed to target the TMV CP; i.e. TMV-F(5829-5848) and TMV-R(5948-5929) of both the CP genomic and sgRNA. This primer pair is annotated in red on the schematic diagram of the TMV genome in Figure 2.2B.

2.9.3 TSWV primers

One pair of TSWV primers were designed to target a 236 bp of the central region of the large RNA segment (L-RNA) of the TSWV genome in its negative sense, TSWV L2 - F(4213-4233) and TSWV L1 -R(4396-4376) and one pair ( as described in Debreczeni et al. 2011) to target a 137 bp region of the non-structural (NSm) protein of the medium RNA segment (M-RNA) of the TSWV genome in its positive sense (TSWV MF(564–584) and TSWV MR(687–700)). The two primer pairs are annotated on the schematic diagram of the TSWV genome in Figure 2.3.
Figure 2.2. Annotation of primers designed to the TMV MP, CP and spanning both the MP and CP.

Schematic diagram of the TMV genome. A. Primer pair annotated in blue targeting a region spanning the MP and CP at nt position 5479-6283. B. Primer pair annotated in red targeting the CP at nt position 5829-5948. Annotated diagrams adapted from http://viralzone.expasy.org/viralzone/all_by_species/51.htm
Figure 2.3. Annotation of primers designed to the central region of the TSWV large (L) segment and the non-structural protein (NSm) of the medium (M) segment.

Schematic diagram of each of the three TSWV segments of the tripartite genome with primer pairs annotated on the L segment RNA and M segment RNA. One pair targets the region between nt positions 564-700 of the positive sense M segment (Debreczeni et al. 2011) while the other targets nt positions 4213-4396 of the negative sense L segment. Annotated diagram adapted from http://viralzone.expasy.org/viralzone/all_by_species/253.html

2.9.4 TuMV primers

One set of primers was designed to target the cytoplasmic inclusion protein (CI) of TuMV (TuMV-F(5356-5376) and TuMV-R(5600-5580)) and one set to target the nuclear inclusion protein B (NIb)/RNA replicase (TuMV F(8463-8482) and TuMV-R(8583-8563)) of the TuMV genomic RNA. The two primer pairs are annotated on the schematic diagram of the TuMV genome in Figure 2.4.
2.9.5 TYMV primers

One set of primers was designed to target the MP of TYMV (TYMV-MP-F(247-288) and TYMV-MP-R(475-456)) on the genomic RNA, and another pair to target CP (TYMV-CP-F(5978-5997) and TYMV-CP-R(6096-6077) on the CP genomic and sgRNAs. The two primer pairs are annotated on the schematic diagram of the TYMV genome in Figure 2.5.

Figure 2.5. Annotation of primers designed to the MP and CP regions of TYMV.

Schematic diagram of TYMV genome with primer pairs annotated to the MP region at nt position 247-475 and subgenomic RNA (sgRNA) CP region at nt position 5978-6096. Annotated diagram adapted from http://viralzone.expasy.org/viralzone/all_by_species/57.html
2.10 Confirmation of virus identity by gel electrophoresis and sequencing of amplicons

Agarose gel electrophoresis was performed using submerged horizontal gels. The presence of single products of appropriate size was confirmed by electrophoresis through 1.5% (w/v) UltraPure™ agarose (Life Technologies-Invitrogen) tris-acetate-EDTA (TAE)-buffered gels. Fragments were separated at 100 V for 50 min. Invitrogen TrackIt™ 1 Kb Plus DNA ladder was used for size determination of fragments. Agarose gels were stained for 30 min in 0.5 mg/L ethidium bromide solution in TAE. Bands were viewed under short wave ultraviolet light (250 – 310 nm) and digital images captured with a Gel Doc imager (Bio-Rad, Hercules, CA, U.S.A.). Bands were excised then purified using a GenElute Gel Extraction Kit (Sigma-Aldrich) as per the manufacturer’s protocol and amplicons sent to the Alan Wilson Centre Genome Service (AWCGS), Palmerston North, New Zealand for sequencing using an ABI3730 capillary sequencer, BigDye Terminator V3.1. Virus-specific amplification primers, as detailed in Table 2.5 and Figures 3.2C to 3.6C, were used to sequence PCR amplicons.

Sequence data in ABI file format (DNA chromatogram file/electropherogram trace) was accessed from the AWCGS website and viewed in the ABI file viewer, BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html, (Hall 1999). Sequences were proofread to ensure that any ambiguous sites were correctly called. Only quality sequences, determine by tall peaks with little to no overlap, were highlighted in BioEdit and copied and pasted to the enter query sequence dialogue box in the standard nucleotide basic local alignment search tool (BLAST) of the National Center for Biotechnology Information (NCBI) website http://blast.ncbi.nlm.nih.gov/ and optimised for highly similar sequences. The first sequence with the highest maximum percent identity and lowest E value (where E refers the chance of a random match) was selected as the most homologous to the query sequence and therefore the most likely corresponding virus and strain.
2.11 Agdia ImmunoStrip® for confirmation of TMV and TSWV identity

Agdia Immunostrips® monoclonal antibody strip systems were used to detect TMV and TSWV (Agdia Inc., Illinois USA) following the manufacturer’s instructions. Briefly, approximately 100 mg of sample taken from symptomatic *Arabidopsis* leaf tissues was placed between the mesh of a bag containing buffer (provided) and the bag rubbed with a pestle to homogenise the sample. A strip with the end marked “sample” was inserted into the extract for no more than two minutes. The appearance of an upper red/purple control line within two minutes confirms that the test has worked, if it does not appear the test is invalid. The appearance of a second line under the control line confirms that a sample is positive. Where no sample line appears but a control line does appear, a sample is negative. Positive test lines are red/purple and may differ in intensity depending on virus titre with high virus titre giving a band of darker intensity while low virus titre results in a band of lighter intensity.

2.12 Isolation and quantification of sRNAs, selection reference genes for qPCR and qPCR method

Specific materials and methods for the development of a method for the isolation and quantification of sRNAs (Chapter 4.3), for the selection of suitable reference genes for quantitative real-time PCR (qPCR) (Chapter 5.2) and for the method devised for qPCR of gene transcripts shown to have fold-change in response to virus infection and of RNAi components (Chapter 6.2) are documented in the respective chapters.
CHAPTER THREE – CONFIRMING PRESENCE AND IDENTITY OF FIVE VIRUSES IN ARABIDOPSIS

3.1 Introduction

In order to correlate data obtained from future experiments with infection and symptoms of each virus, it was important to verify the identity of each of the five viruses used in this study; CaMV, TMV, TSWV, TuMV and TYMV, and specific associated symptoms on infected Arabidopsis leaf tissue. Similarly, it was necessary to establish that mock-inoculated Arabidopsis leaf tissue demonstrated absence of each of the five viruses. This chapter gives an overview of the methodology used to achieve these aims before presenting and discussing results of symptom observations, gel electrophoresis images of PCR amplicons of virus genes, sequencing of PCR amplicon products, Agdia ImmunoStrip® detection and qPCR analysis of viral transcript accumulation.

3.2 Methods overview

Each of the five virus infections used in this thesis were established in Arabidopsis from infected maintenance plant hosts (Chapter 2.2 to 2.5). Virus identity and presence was confirmed by RNA isolation from leaves collected across a time course of 2, 3, 7, 14, 21, 28, 35 and 42 dpi, targeted amplification by PCR as described in Chapter 2.7 to 2.9 with primer pairs listed in Table 2.5 of Chapter 2.9, and sequencing of virus-derived amplicons and BLAST analysis against the NCBI database (Chapter 2.10). TMV and TSWV were also identified and their presence confirmed using an antibody based ImmunoStrip® method (Chapter 2.11). The five viruses were used in a large Arabidopsis infection experiment and the accumulation of each virus transcript was tracked across the time course as determined by real-time qPCR of cDNA as described in Chapter 5.2 with the specific virus primer pairs noted with asterisks in Table 2.5 of Chapter 2.9.
Data were normalised to the three reference genes validated in Chapter 5, *EF1α, F-BOX* and *SAND*, and primer sequences for these reference genes are provided in Table 5.1 of Chapter 5.2.3. In addition, qPCR assays were used to determine presence/ absence of co-infections with each of the five viruses.

**3.3 Results**

All virus inoculated plants were observed on a daily basis for the development of symptoms indicative of their respective virus in inoculated leaves and systemic tissues. Figure 3.1 illustrates observation of symptom absence (blue bars), characteristic symptoms in inoculated leaves (red) and in systemic tissue (green).

Symptoms were not observed on mock-inoculated *Arabidopsis* across the 2 to 42 dpi time course (0.1 on symptom scale). Systemic symptoms were observed on CaMV-inoculated *Arabidopsis* from 14 dpi (2.0 on symptom scale). Systemic symptoms were observed on TMV- and TSWV-inoculated *Arabidopsis* from 28 dpi (2.0 on symptom scale). Symptoms were observed on both TuMV-inoculated *Arabidopsis* leaves (1.0 on symptom scale) and newly developed, systemically infected leaves by 7 dpi. Systemic symptoms were observed on TYMV-inoculated *Arabidopsis* from 14 dpi (2.0 on symptom scale).

All inoculated plants, with the exception of some instances of TSWV, showed symptoms indicative of their respective inoculated virus (Appendix I, CaMV, A1.2; TMV, A2.2; TSWV, A3.2; TuMV, A4.2 and TYMV A5.2) by 28 dpi. Virus-inoculated *Arabidopsis* plants were photographed alongside mock-inoculated *Arabidopsis* plants at 28 dpi, in order to illustrate development of symptomology characteristic of each virus. In some cases, symptoms characteristic of specific virus were photographed earlier or later as required and provided as insets (Figures 3.2A to 3.6A).
Figure 3.1. Observations of symptom development on inoculated leaves and systemic tissues (n= 21; TuMV n=15).

Gel electrophoresis images of PCR amplicons of virus genes, a positive control for cDNA generated from plant RNA (using the NAD5 gene as a target), and two negative controls; one using an RT deficient reaction (except CaMV) and one cDNA deficient reaction are presented for each virus (Figures 3.2B to 3.6B). Amplicons had 96 - 98% nt identity to respective lodged NCBI database sequences as determined by BLAST analysis, thus confirming presence and identity of each of the five viruses. Figures 3.2C to 3.6C detail for each virus, the virus-specific amplification primers, amplicon size and sequence, percent nucleotide (% nt) identity, top BLAST accession number, description, virus strain and reference. The important features for the positive identification of each of the five viruses infecting Arabidopsis are detailed separately in sections 3.3.1 to 3.3.5 below.

3.3.1 CaMV

Figure 3.2A shows a photograph of symptoms indicative of CaMV, i.e. stunting and rugosity, while the inset shows a close up of chlorotic spotting and mottling of a rosette leaf, in virus-inoculated Arabidopsis at 28 dpi compared to mock inoculated. Figure 3.2B demonstrates absence of an amplification product for CaMV in the mock-inoculated plant (lane 2) and amplification of the appropriate size product for the CaMV reverse transcriptase (CaMV-V, 299 bp) in the CaMV-inoculated plant (lane 3). Both mock-inoculated and CaMV-inoculated samples contained intact Arabidopsis DNA/RNA as indicated by internal control bands of 181 bp (NAD5) in lanes 4 and 5 respectively. Absence of a CaMV amplification product in the negative (cDNA deficient) control reaction with CaMV-V specific primers (lane 6) demonstrates absence of CaMV-V cDNA. The identified CaMV isolate was most closely related to accession number V00140.1, virus strain CM1841 (Gardner et al. 1981) with a 98 % nt identity from a sequence length of 734 nt (Figure 3.2C).
Figure 3.2. CaMV in Arabidopsis.

A. Symptoms of CaMV. Mock-inoculation (Left) and CaMV virus-inoculation (Right). Stunting and leaf curling on inoculated plant compared to mock-inoculated plant. Photograph taken 28 dpi. Inset shows chlorotic mosaic characteristic of CaMV; photograph taken 14 dpi. B. Gel electrophoresis of RT-PCR amplicons to confirm CaMV in Arabidopsis (28 dpi). Lanes 1 and 7 (L), 1 Kb Plus DNA ladder. Lanes 2 and 3, CaMV amplicon from RNA extracted from mock-inoculated and CaMV-inoculated leaves, respectively. Lanes 4 and 5, NAD5 amplicon from RNA mock-inoculated and CaMV-inoculated leaves, respectively. Lane 6, CaMV-V amplicon from cDNA deficient reaction. C. Virus primers used to create sequenced amplicon (sequence provided in Chapter 2.9, Table 2.5); maximum percent nucleotide (% nt) identity; sequence length retrieved; accession number and description; virus strain and reference.
3.3.2 TMV

Inoculated *Arabidopsis* did not show the typical mosaic symptoms of TMV in virus-inoculated plants at 28 dpi compared to mock-inoculated *Arabidopsis* (Figure 3.3A). This is consistent with previous observations where the common TMV-U1 strain is often asymptomatic in the *Arabidopsis Col-0* ecotype, with virus reaching apical leaves after 21 dpi (Serrano et al. 2008). However, light mosaic symptoms suggestive of TMV were observed and photographed at 28 dpi (Figure 3.2A inset). Figure 3.3B demonstrates absence of an amplification product for TMV in the mock-inoculated plant (lane 2) and amplification of an appropriate size product spanning the TMV MP/CP (TMV-MP/CP, 804 bp) in the TMV-inoculated plant (lane 3). Both mock-inoculated and TMV-inoculated samples contained intact *Arabidopsis* DNA/RNA as indicated by internal control bands of 181 bp (*NAD5*) in lanes 4 and 5 respectively. Absence of an amplification product in the reverse transcriptase deficient control reaction (lane 6) reveals absence of DNA and other contaminating templates. Absence of a *NAD5* amplification product in the negative (cDNA deficient) control reaction with *NAD5* specific primers reveals absence of *Arabidopsis NAD5* (lane 7). Absence of a TMV amplification product in the negative (cDNA deficient) control reaction with TMV-MP/CP specific primers (lane 8) reveals absence of TMV-MP/CP cDNA. The identified TMV isolate was most closely related to accession number AJ429078.1, virus strain U1/vulgare (Letschert et al. 2002) with a 96% nt identity from a sequence length of 1028 nt (Figure 3.3C).
A. Symptoms of TMV. Mock inoculation (Left) Photograph taken 28 dpi. Inset shows light mosaic characteristic of late infection with TMV; photograph taken 28 dpi. B. Gel electrophoresis of PCR amplicons to confirm TMV in Arabidopsis (28 dpi). Lanes 1 and 9 (L), 1 Kb Plus DNA ladder. Lanes 2 and 3, TMV amplicon from RNA extracted from mock-inoculated and TMV-inoculated leaves, respectively. Lanes 4 and 5, NAD5 amplicon from RNA mock-inoculated and TMV-inoculated leaves, respectively. Lane 6, TMV-MP/CP amplicon from RT-deficient reaction. Lane 7, NAD5 amplicon from cDNA deficient reaction. Lane 8, TMV-MP/CP amplicon from cDNA deficient reaction. C. Virus primers used to create sequenced amplicon (sequence provided in Chapter 2.9, Table 2.5); maximum percent nucleotide (% nt) identity; sequence length retrieved; accession number and description; virus strain and reference.

**Figure 3.3. TMV in Arabidopsis.**

<table>
<thead>
<tr>
<th>Virus primers</th>
<th>% nt identity</th>
<th>Sequence length</th>
<th>Accession and description</th>
<th>Virus strain</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOBO (UNI2)-F(5479–5498)</td>
<td>96 %</td>
<td>1028 nt</td>
<td>AJ429078.1 TMV coat protein</td>
<td>U1/Vulgare</td>
<td>Letschert et al. 2002</td>
</tr>
</tbody>
</table>

1000 850 650 500 400 300 200 100

804 bp (TMV)

181 bp (NAD5)
3.3.3 TSWV

While leaf spotting symptoms indicative of TSWV were observed in TSWV-inoculated systemic Arabidopsis leaf tissue prior to 28 dpi, and subtle purpling of stems at 28 dpi, symptoms were mild and difficult to photograph. Figure 3.4A shows a photograph of symptoms indicative of earlier TSWV i.e. leaf spotting and leaf distortion (upper leaf), compared to mock-inoculated (lower leaf) Arabidopsis (German et al. 1995).

Mechanical transmission of TSWV is not very efficient in many plant species, including Arabidopsis (Brunt et al. 1996 onwards; Kikkert et al. 1999; Mandal et al. 2008). For this reason, nine Arabidopsis samples were compared with a TSWV-inoculated Lycopersicon esculentum (tomato) sample. Figure 3.4B demonstrates amplification of the appropriate size product for TSWV L-RNA in lanes 1-6 and 8 and 9 (TSWV-inoculated Arabidopsis) and lane 10 (TSWV-inoculated tomato). An appropriate size product for TSWV L-RNA failed to amplify in lane 7 (TSWV-inoculated Arabidopsis). Absence of an amplification product in the reverse transcriptase deficient control reaction (lane 11) reveals absence of DNA and other contaminating templates. Absence of an NAD5 amplification product in the negative (cDNA deficient) control reaction with NAD5 specific primers reveals absence of Arabidopsis NAD5 (lane 12). TSWV-inoculated samples contained intact Arabidopsis DNA/RNA as indicated by internal control band of 181 bp (NAD5) in lane 13. Absence of a TSWV amplification product in the negative (RT- deficient) control reaction with TSWV L-RNA specific primers (lane 14) reveals absence of gDNA. The identified TSWV isolate was most closely related to accession number D10066.1, virus strain BR-01 (de Haan et al. 1991) with 96 % nt identity from a sequence length of 382 nt (Figure 3.4C).
A. Symptoms of TSWV. Mock inoculation (below) TSWV-inoculation (above) (image from German et al. 1995). B. Gel electrophoresis of PCR amplicons to confirm TSWV in Arabidopsis (28 dpi). Lanes 1 and 15, 1 Kb Plus DNA ladder. Lanes 2-9 TSWV L-RNA amplicons from TSWV-Arabidopsis (276 bp). 10, TSWV L-RNA amplicon from TSWV-inoculated tomato. Lane 11, TSWV-L-RNA amplicon from mock-inoculated Arabidopsis. Lane 12, NAD5 amplicon from cDNA deficient reaction. Lane 13, NAD5 amplicon from RNA of TSWV-inoculated Arabidopsis. Lane 14, TSWV-L-RNA amplicon from RT-deficient reaction. C. Virus primers used to create sequenced amplicon (sequence provided in Chapter 2.9, Table 2.5); maximum percent nucleotide (% nt) identity; sequence length retrieved; accession number and description; virus strain and reference.

<table>
<thead>
<tr>
<th>Virus primers</th>
<th>% nt identity</th>
<th>Sequence length</th>
<th>Accession and description</th>
<th>Virus strain</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSWV L2F (4213-4233)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSWV L1R (4396-4376)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.4. TSWV in Arabidopsis.
3.3.4 TuMV

Figure 3.5A shows photographs of symptoms indicative of TuMV in *Arabidopsis* at 28 dpi, i.e. lethal necrosis in TuMV-inoculated *Arabidopsis* (right-hand plant (R)) compared to mock inoculated (left-hand plant (L)). It was expected that TuMV-inoculated plants would have suffered lethal necrosis by 21 - 28 dpi (Kim et al. 2010; Ouibrahim and Caranta 2013). The inset photograph taken at 14 dpi shows the light, dark green and cream mosaic characteristic of earlier TuMV infection. Figure 3.5B demonstrates absence of an amplification product for TuMV in the mock-inoculated plant (lane 2) and amplification of the appropriate size product for the TuMV cytoplasmic inclusion (CI, 244 bp) in the TuMV-inoculated plant (lane 3). Both mock-inoculated and TuMV-inoculated samples contained intact *Arabidopsis* DNA/RNA as indicated by internal control bands of 181 bp (*NAD5*) in lanes 4 and 5 respectively. Absence of an amplification product in the reverse transcriptase deficient control reaction (lane 6) reveals absence of DNA and other contaminating templates. Absence of a TuMV-CI amplification product in the negative (cDNA deficient) control reaction with TuMV-CI specific primers reveals absence of *Arabidopsis NAD5* (lane 7). The identified TuMV isolate was most closely related to accession number AB093612.1, virus strain NZ290 (Tomimura et al. 2003) with a 98% identity from a sequence length of 430 nt (Figure 3.5C).
Figure 3.5. TuMV in *Arabidopsis*.

Symptoms of TuMV. Mock inoculation (Left) and TuMV-inoculation (Right) Photograph taken 28 dpi. Note necrosis of TuMV-inoculated plant. Inset shows mosaic characteristic of early infection in a TuMV-inoculated leaf; photograph taken 14 dpi. B. Gel electrophoresis of PCR amplicons to confirm TuMV in *Arabidopsis* (28 dpi). Lanes 1 and 8 (L), 1 Kb Plus DNA ladder. Lanes 2 and 3, TuMV amplicon from RNA extracted from mock-inoculated and TuMV-inoculated leaves, respectively. Lanes 4 and 5, NAD5 amplicon from RNA mock-inoculated and TuMV-inoculated leaves, respectively. Lane 6, TuMV-CI amplicon from RT-deficient reaction. Lane 7, TuMV-CI amplicon from cDNA deficient reaction. C. Virus primers used to create sequenced amplicon (sequence provided in Chapter 2.9, Table 2.5); maximum percent nucleotide (% nt) identity; sequence length retrieved; accession number and description; virus strain and reference.
3.3.5 TYMV

Figure 3.6A shows photographs of symptoms indicative of TYMV in *Arabidopsis* at 28 dpi, i.e. chlorosis and stunting in TYMV-inoculated plants (L) compared to mock inoculated (R). Inset shows dark green islands characteristic of the TYMV mosaic, photograph taken 28 dpi. Figure 3.6B demonstrates absence of an amplification product for TYMV-MP in the mock-inoculated plant (lane 2) and amplification of the appropriate size product for the TYMV-MP (247 bp) in the TYMV-inoculated plant (lane 3). Both mock-inoculated and TYMV-inoculated samples contained intact *Arabidopsis* DNA/RNA as indicated by internal control bands of 181 bp (*NAD5*) in lanes 4 and 5 respectively. Absence of an amplification product in the reverse transcriptase deficient control reaction (lane 6) reveals absence of DNA and other contaminating templates. Absence of an *NAD5* amplification product in the negative (cDNA deficient) control reaction with *NAD5* specific primers reveals absence of *Arabidopsis NAD5* (lane 7). Absence of a TYMV-MP amplification product in the negative (cDNA deficient) control reaction with TYMV-MP specific primers (lane 8) reveals absence of TYMV-MP cDNA. The identified TYMV isolate was most closely related to accession number AF035403.1, virus strain Blue Lake isolate (TYMV-BL) (Skotnicki et al. 1992) with a 98% identity from a sequence length of 382 nt (Figure 3.6C).
A. Symptoms of TYMV. Mock inoculation (Left) and TYMV-inoculation (Right). Chlorosis and stunting in the TYMV-inoculated plant. Photograph taken 28 dpi. Inset shows mosaic of dark green islands characteristic of TYMV; 21 dpi.

B. Gel electrophoresis of PCR amplicons to confirm TuMV in Arabidopsis (28 dpi). Lanes 1 and 9 (L), 1 Kb Plus DNA ladder. Lanes 2 and 3, TYMV amplicon from RNA extracted from mock-inoculated and TYMV-inoculated leaves, respectively. Lanes 4 and 5, NAD5 amplicon from RNA mock inoculated and TYMV-inoculated leaves, respectively. Lane 6, TYMV amplicon from RT-deficient reaction. Lane 7, NAD5 amplicon from cDNA deficient reaction. Lane 8, TYMV amplicon from cDNA deficient reaction.

C. Virus primers used to create sequenced amplicon (sequence provided in Chapter 2.9, Table 2.5); maximum percent nucleotide (% nt) identity; sequence length retrieved; accession number and description; virus strain and reference.

**Figure 3.6. TYMV in Arabidopsis.**

A. Symptoms of TYMV. Mock inoculation (Left) and TYMV-inoculation (Right). Chlorosis and stunting in the TYMV-inoculated plant. Photograph taken 28 dpi. Inset shows mosaic of dark green islands characteristic of TYMV; 21 dpi. B. Gel electrophoresis of PCR amplicons to confirm TuMV in Arabidopsis (28 dpi). Lanes 1 and 9 (L), 1 Kb Plus DNA ladder. Lanes 2 and 3, TYMV amplicon from RNA extracted from mock-inoculated and TYMV-inoculated leaves, respectively. Lanes 4 and 5, NAD5 amplicon from RNA mock inoculated and TYMV-inoculated leaves, respectively. Lane 6, TYMV amplicon from RT-deficient reaction. Lane 7, NAD5 amplicon from cDNA deficient reaction. Lane 8, TYMV amplicon from cDNA deficient reaction. C. Virus primers used to create sequenced amplicon (sequence provided in Chapter 2.9, Table 2.5); maximum percent nucleotide (% nt) identity; sequence length retrieved; accession number and description; virus strain and reference.
3.4 Agdia ImmunoStrip® for confirmation of TMV and TSWV identity

In addition to virus symptom development, amplicon size and amplicon sequence, infection with TMV and TSWV were confirmed by Agdia Immunostrips® monoclonal antibody strip systems, as described in Chapter 2.11. Figure 3.7 illustrates Immunostrip® confirmation of the presence of TMV and TSWV in virus-inoculated Arabidopsis at 28 dpi. Lane 1 shows that mock-inoculated tissue is not infected with TMV as represented by the presence of a control line alone. Lane 2 shows that TMV is present in TMV-inoculated tissue as represented by the presence of both the control and test lines. Lane 3 shows that mock-inoculated tissue is not infected with TSWV as represented by the presence of a control line alone while Lane 4 shows TSWV is present in the TSWV-inoculated tissue as represented by the presence of both the control and test lines.

<table>
<thead>
<tr>
<th>Lane</th>
<th>TMV</th>
<th>TSWV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3.7. Agdia Immunostrip® detection of TMV and TSWV in Arabidopsis at 28 dpi.

From left to right, Lane 1, mock-inoculated test line absent (negative); 2, TMV-inoculated, test line present (positive); 3, mock-inoculated test line absent (negative) and; 4, TSWV-inoculated, test line present (positive). Control line present in all four lanes.
3.5 qPCR analysis of viral transcript accumulation

The transcript accumulation of each virus was tracked across the time course as determined by qPCR of Arabidopsis cDNA. The results of qPCR analysis of virus transcript accumulation (Figures 3.8A, B, C, D and E) show presence of each virus at variable accumulation across the time course compared to mock-inoculated plants. Data are normalised average quantity (Q) rescaled to 1.0 where 1.0 represents mock-inoculation at 2 dpi (Chapter 5.3.6). Data were normalised to the normalisation factor obtained from three reference genes; EF1a, F-BOX and SAND as calculated by the geNorm method (Chapter 5).

The transcript accumulation of CaMV 35S RNA increased from 2 dpi to 28 dpi, followed by a decrease in accumulation at 35 and 42 dpi. CaMV 35S RNA transcript accumulation was not detected from mock-inoculated Arabidopsis cDNA (Figure 3.8A). TMV CP transcript accumulation increased steadily to 14 dpi then plateaued for the rest of the time course. The maximum transcript accumulation of TMV CP is very low (Q = 0.6) compared to the accumulation of all other viruses investigated where the normalised rescaled Q values are in excess of 1200. TMV CP transcript accumulation was not detected from mock-inoculated Arabidopsis cDNA (Figure 3.8B). TSWV M-RNA transcript accumulation increased from 2 dpi, decreased then increased again at 21 and 28 dpi followed by a further decrease at 35 dpi. TSWV M-RNA transcript accumulation was not detected from mock-inoculated Arabidopsis cDNA (Figure 3.8C). TuMV N1b transcript accumulation increased steadily from 2 dpi and N1b transcript accumulation was not detected from mock-inoculated Arabidopsis cDNA (Figure 3.8D). TYMV CP transcript accumulation increased to 21 dpi then steadily declined to 42 dpi. TYMV CP transcript accumulation was not detected from mock-inoculated Arabidopsis cDNA (Figure 3.8E).
Figure 3.8. Normalised average quantity (Q) of three biological replicates of mock-inoculated and CaMV, TMV, TSWV, TuMV and TYMV inoculated samples rescaled to 1.0 and representing viral RNA transcript accumulation.

A. Cauliflower mosaic virus 35S RNA transcript accumulation between 2 and 42 dpi compared to mock-inoculated accumulation between 2 and 42 dpi. B. Tobacco mosaic virus CP transcript accumulation between 2 and 42 dpi compared to mock-inoculated accumulation between 2 and 42 dpi. C. Tomato spotted wilt virus M-RNA transcript accumulation between 2 and 42 dpi compared to mock-inoculated accumulation between 2 and 42 dpi. D. Turnip mosaic virus NIb transcript accumulation between 2 and 28 dpi compared to mock-inoculated accumulation between 2 and 42 dpi. E. Turnip yellow mosaic virus CP transcript accumulation between 2 and 42 dpi compared to mock-inoculated accumulation between 2 and 42 dpi.
The viral RNA transcript accumulation of each virus cannot be correlated with the observed symptoms or leaf collected (inoculated or systemic) as charted in Figure 3.1 (Chapter 3.3). CaMV leaf tissue was showing systemic symptoms from 14 dpi and while a steady increase in CaMV 35S RNA transcript accumulation was seen from 14 dpi, the accumulation declined from 35 dpi. Symptoms of CaMV were not visible on the inoculated leaves. TMV leaf tissue was showing systemic symptoms by 28 dpi however, TMV-CP transcript accumulation was increasing from 7dpi. Symptoms of TMV were not visible on the inoculated leaves. TSWV leaf tissue was showing systemic symptoms by 28 dpi; however TSWV L-RNA transcript accumulation was variable across the time course. Symptoms of TSWV were not visible on the inoculated leaves. TuMV leaf tissue was showing both symptoms in inoculated leaves and systemic leaf tissue by 7 dpi and TuMV N Ib transcript accumulation was at its highest in predominantly necrotic tissue at 28 dpi. TYMV leaf tissue was showing systemic symptoms from 14 dpi. Symptoms of TYMV were not visible on the inoculated leaves. The results of qPCR showed that there were no mixed infections with the five candidate viruses. However, it is possible that virus-inoculated tissue became co-infected with a further virus or viruses not tested for in this study.

3.6 Discussion

Taken together the results of symptom photography, PCR amplification and results of qPCR transcript accumulation show that each virus is present in either inoculated leaf and/or systemic Arabidopsis tissues and that none of the studied viruses was present in any mock-inoculated tissues. Sequencing of amplicon products determined that each virus shares between
96 - 98% nucleotide identity to respective virus species. Immunostrip® assay of TMV and TSWV indicate presence in virus-infected tissue and absence in mock-inoculated tissue.

Results of qPCR analysis of viral transcript accumulation demonstrate virus presence at each collection point (dpi) across the time-course of infection and suggest accumulation of each viral transcript is variable across the time course. This variability is likely to correlate with varying rates of local replication and systemic movement in both different viruses and different tissue collected at different time points. The variability of viral transcript accumulation might also correlate with the action of respective VSRs of RNAi. The action of VSRs and their possible impact on the findings of this thesis are discussed in Appendix I. In order to determine viral amount, or titre, it is necessary to calculate viral quantity using a standard curve generated using known genome equivalents or nucleic acid isolated from a titred virus control (reviewed in Atkinson et al. 2014). Viral titre was not quantified using a standard curve in the present study, therefore titre could not be correlated with observed symptoms or tissue collected at given time points. While all samples were tested by qPCR to determine absence of co-infection with each of the other five viruses of this study, and absence of co-infection confirmed, it is possible that virus-inoculated tissue became co-infected with a further virus or viruses not tested for in this study.

Mechanical transmission of TSWV is not very efficient in many plant species, including Arabidopsis (Brunt et al. 1996 onwards; Kikkert et al. 1999; Mandal et al. 2008) and this might explain variability of TSWV transcript accumulation in the present study. Poor efficiency of TSWV inoculation is likely to be due to the instability of virus particles once isolated from the plant (Brunt et al. 1996 onwards). Physically and chemically TSWV is one of the most unstable plant viruses. In sap, the thermal inactivation point (10 min) is 40-46°C, longevity in vitro at
room temperature is 2-5 h and the dilution end point between $2 \times 10^{-2}$ and $10^{-3}$. Infectivity rapidly falls at pH values below pH 5.0, and is maintained best at pH values near pH 7.0. Several factors such as the plant growth environment, plant growth stage, source of inoculum, antioxidants and abrasives have been found to influence the rate of mechanical transmission of TSWV. TSWV is stabilised in plant extracts by adding reducing agents such as Na$_2$SO$_3$ or sodium thioglycollate. All inoculations of the present study contained Na$_2$SO$_3$ as described in Chapter 2.5. Nonetheless, the results of PCR, sequencing, Agdia Immunostrip® and qPCR assays of TSWV-inoculated tissue confirm successful inoculation to and presence and identity in the tested samples of present study.

Once it had been established by way of PCR and sequencing (and Agdia Immunostrip® detection where applicable) that each of the five viruses to be utilised for the work of this thesis had been successfully inoculated into Arabidopsis - research was initiated on the quantification of accumulation of sRNAs in response to plant virus infection (Chapter 4), identification of reference genes that have stable transcript accumulation regardless of virus infection (Chapter 5), qPCR of transcript accumulation of genes reported in literature to respond to virus infection and transcript accumulation of selected RNAi components (Chapter 6).
CHAPTER FOUR – EVALUATING METHODS FOR THE ISOLATION AND QUANTIFICATION OF SMALL RNAs (sRNAs)

4.1 Introduction

During plant virus infection, replicative viral dsRNAs trigger host plant RNAi to cleave dsRNAs into virus-derived siRNAs that in turn target cognate viral siRNAs as part of the host’s defence against viruses (Ding 2010). Given the antiviral nature of these virus-derived siRNAs, it is not surprising that the amount of the sRNA population present within a virus infected plant has been noted to increase in some plant virus infections (Ding and Voinnet 2007; Mlotshwa et al. 2008; Bazzini et al. 2009; Ruiz-Ferrer and Voinnet 2009; Blevins et al. 2011; Hu et al. 2011; Schuck et al. 2013). For example, sRNAs have been shown to accumulate in plants infected with Citrus tristeza virus (Ruiz-Ruiz et al. 2011), Cabbage leaf curl virus (Aregger et al. 2012), CMV, PVX, TRV, TuMV, Cymbidium ringspot virus, Melon necrotic spot Carmovirus, Pepper mild mottle tobamovirus, Watermelon mosaic Potyvirus and Tomato yellow leaf curl Begomovirus amongst others (reviewed in Qu et al. 2010). To determine whether the accumulation of the total population of sRNAs (the composition of which is described in Chapter 1.5.5) could be developed as a generic tool for detecting virus presence or absence, sRNAs were quantified during a broad time-course of infection, from different plant virus infections and from different plant species.

Traditional methods for the detection and quantification of sRNAs, such as northern blotting and qPCR, target specific and known sRNA sequences. It would be useful to be able to quantify the total of all sRNAs present in a sample; known and unknown and endogenous and exogenous. The research aims of this chapter were to:
(i) isolate high quality low molecular weight RNA (LMW-RNA)

(ii) develop an assay for the accurate quantification of the total sRNA component (including siRNAs and miRNAs) of total isolated LMW-RNA.

(iii) determine if quantification of sRNA accumulation could be utilised as a tool for detecting virus presence or absence.

Traditional methods for the isolation and detection and quantification of sRNAs are discussed below, prior to the materials and methods for the work comprising this chapter.

4.1.1 Isolation of sRNAs

Traditionally, northern blot analysis has been used to isolate and detect sRNAs and involves RNA extraction, size separation by TBE polyacrylamide gel electrophoresis, transfer of sRNAs to a membrane and detection of specific sRNAs using labelled oligonucleotide or longer nucleic acid probe hybridisation (Cheng et al. 2010). A second, and preferred strategy to isolate sRNAs for downstream processing or separation and detection, is to chemically isolate low molecular weight (LMW-RNA), i.e. RNA of 4-150 nt. Typical LMW-RNA extraction protocols include stepwise precipitation with polyethylene glycol (PEG) to eliminate DNA and high-molecular-weight RNA (HMW-RNA). Since plant tissues are rich in polysaccharides, polyphenols and other secondary metabolites that can interfere with nucleic acid separation and result in poor yield and quality of LMW RNA, a cetyltrimethylammonium bromide (CTAB) based method combined with PEG and ethanol or isopropanol precipitation can be used to overcome this problem (Doyle and Doyle 1987).

The cationic properties of CTAB lyse cell walls and liberate cellular nucleic acids while the detergent properties of CTAB assist with cell wall solubilisation. CTAB also precipitates gDNA and removes polysaccharides from bacterial and plant preparations (reviewed in Tan and Yiap 2009). PEG is a non-denaturing water-soluble polymer that
reduces solubility and renders HMW-RNA insoluble. PEG precipitation to remove HMW-RNA is usually followed by sodium acetate and absolute ethanol (or isopropanol) precipitation of the aqueous supernatant to concentrate and recover LMW-RNA (reviewed in Tan and Yiap 2009). However, CTAB based extractions combined with chloroform/phenol and PEG precipitation are lengthy; taking at least two days. Lengthy extraction processes with many steps and reagents are liable to cause loss and/or degradation of LMW-RNA (Cheng et al. 2010; Fasold and Binder 2013, Wang et al. 2010).

A further choice is to use a commercial sRNA isolation kit based on CTAB and precipitation chemistry. Commercial extraction kits (e.g. mirVana™ and Purelink® (Ambion, Life Technologies, Carlsbad, CA, U.S.A)) and mirPremier™ (Sigma-Aldrich, St. Louis, MO, U.S.A.) are cost effective, rapid and efficient for purifying and enriching sRNAs (including miRNAs) from diverse biological sources, including tissues of many plant species. Ribonucleases and secondary metabolites are inactivated; HMW-RNA and gDNA remain insoluble and remaining sRNAs are captured on a silica binding column. However, the recovery of LMW-RNA does vary with different isolation protocols and different isolation kits. Masotti et al. (2009) used a lab-on-a-chip technology (Agilent Bioanalyzer) to characterise the LMW-RNA fractions obtained with different commercial miRNA extraction protocols, an acid phenol/guanidine isothiocyanate solution (TRIzol Reagent, Invitrogen), and MirVana and miRNEasy (Mini Kit, Qiagen) kits. The recovery of miRNAs was evaluated with specific qPCR and the Agilent Bioanalyzer. While total RNA recovery was similar for the three protocols, the composition of the LMW-RNA fractions was substantially different. TRIzol and MirVana allowed the highest recovery of LWM-RNA (20 to 34%) from total RNA while miRNEasy resulted in low (2.5 to 3%) recovery (Masotti et al. 2009).
4.1.2 Detection and quantification of sRNAs

The continuing discovery of sRNAs using NGS technologies has raised the need for methods that can reliably detect sRNAs and quantitate accumulation. Currently, northern blot hybridisation is the most widely used method for validating sRNAs that are identified by such technologies (Blevins 2010; Kim et al. 2010). Although northern blot is less sensitive than other analytical methods, it can quantitate the accumulation and size of specific sRNA sequences and their precursors. Several distinct northern blot protocols are currently used for sRNA detection, where methods primarily differ in the labelling and design of the oligonucleotide probes (the reverse complement of the sRNA sequence) to be used to detect RNA.

In addition to northern blot hybridization for sRNA detection, accumulation of specific known sRNA sequences can be quantitatively assessed using a qPCR assay of cDNA of total RNA extractions using a sRNA-specific stem-loop primer. This stem loop qPCR results in increased sensitivity over the traditional northern blot protocol and enables confirmation of sRNA sequencing results generated by NGS technologies (Smith and Eamens 2012).

4.2 Materials and Methods

4.2.1. Isolation of sRNAs

The mirPremier™ microRNA Isolation Kit (Sigma-Aldrich, MO, U.S.A.), miRNeasy Mini Kit (Qiagen, Venlo, Limburg, Netherlands) and Purelink® (Ambion, Life Technologies, Carlsbad, CA, U.S.A) commercial miRNA isolation kits were trialed. The mirPremier™ microRNA isolation kit provided the most consistent yields and best quality as determined by NanoDrop spectrophotometry (ND-1000 Spectrophotometer; Nanodrop Technologies Inc., Wilmington, DE, U.S.A.).
LMW-RNA was isolated from *Arabidopsis* leaf tissue samples either mock-inoculated or inoculated with either CaMV, TMV, TSWV, TuMV or TYMV and tissue collected at 2, 3, 7, 14, 21 and 28 dpi (three biological replicates of each and as previously described in Chapter Two, sections 2.2 to 2.5) with a mirPremier™ microRNA Isolation Kit (Sigma-Aldrich, MO, U.S.A.) according to manufacturer’s protocol for plant tissues as follows:

*Arabidopsis* tissue weighed to 100 mg was ground to a fine powder in liquid nitrogen using a mortar and pestle. Lysis mix for each sample was prepared from 650 µl microRNA Lysis Buffer, 300 µl Binding Solution and 50 µl molecular grade ethanol and 750 µl of this mix added to the powdered tissue of each sample in a 1.7 ml Eppendorf tube™ and vortexed thoroughly for 30 sec. Samples were incubated at 55 °C for 5 min to release LMW-RNA and to inactivate ribonucleases and any interfering secondary metabolites. Samples were transferred to spin columns within tubes provided and centrifuged at 16,000 g in a microcentrifuge for 5 min to remove cellular debris, gDNA, and HMW-RNA. The resultant supernatants were retained and pipetted into separate filtration columns within collection tubes and centrifuged at 16,000 g in a microcentrifuge for 1 min to remove residual debris. Each clarified flow-through lysate was retained and 1.1 volumes of molecular grade ethanol added to each and mixed by inversion in order to aid binding. LMW-RNA from each sample was captured on individual silica binding columns by pipetting 700 µl of the lysate/ethanol mix to each column and centrifuging at 16,000 g in a microcentrifuge for 30 sec. Flow through liquid was discarded. The binding step was repeated with any remaining binding mix and flow through discarded. Molecular grade ethanol (99.98%) to 700 µl was added directly to each silica binding column and centrifuged at 16,000 g in a microcentrifuge for 30 sec. The flow-through liquid was discarded and the column returned to the collection tube. 500 µl of binding solution was added to each column and centrifuged in a microcentrifuge at 16,000 g for 1 min. Each binding column was transferred to a fresh collection tube and 500 µl ethanol-
diluted ‘Wash Solution 2’ added. This was centrifuged at 16,000 g in a microcentrifuge for 30 sec. The flow through was discarded and each column returned to the collection tube and the ‘Wash Solution 2’ step repeated. The flow through liquid was again discarded and each column returned to each respective collection tube. Each column was dried by centrifugation at 16,000 g for 1 min then transferred to a clean 2 ml collection tube. Elution solution to 50 µl was added to the centre of each column filter. Each column was left to sit for 1 min prior to centrifugation at 16,000 g for 1 min to elute. The elution step was repeated by collecting each elutate in a pipette tip and reloading the solution directly onto the centre of each column filter followed by further centrifugation for 1 min at 16,000 g. Following elution, samples were quantified and assessed for quality using NanoDrop spectrophotometry as described in Chapter 2.7.1.

4.2.2 Development of a method for the detection and quantification of sRNAs

A thorough search of literature found no reference to methods for the accurate quantification of the accumulation of components of total LMW-RNA fraction, including sRNA (both siRNA and miRNA populations), that could be used to monitor their response to virus infection. It was therefore necessary to develop a robust method for quantifying the components of the total LMW-RNA population within the size range of 4 to 150 nt (equal to the range of the Agilent Bioanalyzer molecular weight ladder) from Arabidopsis samples across a time-course (2, 3, 7, 14, 21 and 28 dpi) and treatments (Mock, CaMV, TMV, TSWV, TuMV and TYMV). Two methods were assessed for feasibility as quantitative measures for sRNA and each methodology is outlined below.
4.2.3 Assessment of PAGE, Typhoon 9410 imaging and ImageQuant TL software (PAGE/Typhoon/ImageQuant method) for the quantification of sRNAs

PAGE of LMW-RNA samples isolated from Arabidopsis as described in Chapter 4.3.1 followed by gel image capture with a Typhoon 9410 bio-molecular imager (GE Healthcare Life Sciences, formerly Amersham Biosciences, Uppsala, Sweden) and band quantitation with ImageQuant TL 7.0 software (GE Healthcare Life Sciences, Uppsala, Sweden) was trialled to determine feasibility as a method of sRNA quantification. The method is termed the ‘PAGE/Typhoon/ImageQuant method’ from hereon in.

4.2.4 Polyacrylamide gel electrophoresis of LMW-RNA

Samples were prepared according to gel manufacturers’ recommendations as follows: 2 µl Novex® Hi-density TBE sample buffer (5x) (containing the tracking dyes Bromophenol Blue and Xylene Cyanol) was added to each 25 µl sample and samples denatured for 2 min at 70°C then placed immediately on ice. An XCell SureLock™ Mini Cell Electrophoresis system was assembled according to the manufacturer’s instructions (Invitrogen Carlsbad, CA, U.S.A.) and two precast Novex® TBE 8% (10 well) gels (Invitrogen Carlsbad, CA, U.S.A.) cassettes fitted. The upper buffer chamber was filled with 1X Novex® TBE running buffer (prepared by adding 200 ml 5X Novex® TBE running buffer to 800 ml double distilled deionised water) to completely cover the sample wells. A marker consisting of 5 µl of TrackIt™ 10 bp DNA Ladder (premixed with sample buffer and suitable for loading to agarose and polyacrylamide gels; Invitrogen Carlsbad, CA, U.S.A.) was added to the first well of each gel. 22 µl of each dilution point sample were loaded into individual wells following the 10 bp ladder in the first well. The lower chamber was filled with 600 ml 1X Novex® TBE running buffer and the apparatus run at 200V constant for 90 min.
The resultant electrophoretic gel was removed from the cassette and stained by immersion in TBE buffer with 0.5 µg/ml EtBr for 30 min. The stained gel was first imaged with the Bio-Rad Gel Doc and a digital gel image captured before scanning the gel with the Typhoon 9140 imager as per the Typhoon 9140 user manual.

4.2.5 Gel scanning with Typhoon 9140 imager

The PAGE sample was placed face down on the scanner’s glass platen with the top left corner cut as a reference point. Scanner control software was set to fluorescence acquisition mode. The emission filter 610 nm band-pass (610 BP 30) suitable for EtBr was selected in order to reject reflected and scattered excitation light and background fluorescence. The 610-nm band-pass filter transmits light between 595 nm and 625 nm and has a transmission peak centred at 610 nm. The area of the gel to be scanned was selected with the scanner software according to manufacturer’s instructions. The highest resolution available (10 µm pixel size) was selected. The ImageQuant TL 7.0 image analysis software was selected for automatic image analysis on completion of scanning. The gel was scanned and saved and an image of the sample appeared automatically in the image analysis software.

4.2.6 ImageQuant analysis

ImageQuant TL was launched and 1D gel analysis selected. The gel image was opened from file and automatic analysis selected. Quantity values were detected and the values exported to Excel for further analysis.
4.2.7 Statistical analysis of data of the PAGE/Typhoon/ImageQuant method

ImageQuant data was analysed by calculating mean and standard error of the mean in Microsoft Excel 2007/2010 in order to determine linearity, limits of detection and reproducibility. Data were plotted as a scatterplot with either a linear regression trend line fitted, or error bars representing standard error of the mean shown, as appropriate.

4.2.8 Assessment of the Agilent Bioanalyzer with Lab-on-Chip small RNA kit (Bioanalyzer method) for the quantification of accumulation of sRNA

The Agilent Bioanalyzer 2100 platform coupled with Agilent Pico or Nano RNA kits (Agilent Technologies, Santa Clara, CA, U.S.A.) and the use of an RNA integrity number (RIN) are being increasingly used in order to rapidly assess RNA quality for qPCR experiments (reviewed in Die and Román 2012). It was believed that the Agilent Bioanalyzer with small RNA lab-on-chip technology might be suitable for the quantification of sRNA accumulation for the present study and initial assessments to determine limits of detection, linearity and reproducibility were undertaken. The method is hereon in termed the ‘Bioanalyzer method’.

4.2.9 Small RNA Lab on Chip protocol

Small RNA chips were loaded following the manufacturer’s Small RNA Kit Quick Start Guide as below:

i) Gel preparation

The complete volume (approx. 650 μl) of the provided gel matrix was transferred to a spin filter column and centrifuged at 10,000 g for 15 min at room temperature. The filtered gel was stored at 4 °C to be used within 4 weeks.
ii) Preparation of gel-dye mix

Reagents were equilibrated to room temperature and the dye concentrate vortexed for 10 s and spun down before 2 μl of dye concentrate was pipetted into a 0.5 ml RNase- free microtube and 40 μl of the previously filtered gel added. The gel and dye were mixed by flicking the vial until homogeneous. The microtube was centrifuged at 13,000 g for 10 min at room temperature.

iii) Loading the gel-dye mix

A new sRNA chip was loaded on the chip priming station (provided) and 9 μl of gel-dye mix pipetted to the well labelled with a white G within a black circle (Figure 4.1). The plunger was positioned at 1 ml and the chip priming station closed. The plunger was pressed down until held by the clip. After 60 s the clip was released and the plunger moved back to the 0.3 ml mark, then slowly returned to the 1 ml position. The chip priming station was opened and 9 μl of gel-dye mix pipetted into the two wells labelled G (Figure 4.1).

Figure 4.1. Image of an Agilent Technologies Small RNA Chip.

Small RNA Chip showing wells for loading gel (G), conditioning solution (CS), ladder and 11 sample wells labelled as 1-3, 4-6, 7-9 and 10 and 11. Image from http://www.genomics.agilent.com/en/Bioanalyzer-DNA-RNA-Kits/Small-RNA-Analysis-Kits/.
iv) **Loading the conditioning solution and marker**

RNA conditioning solution (9 μl) was pipetted into the well labelled CS, then 5 μl RNA marker added to each of the 11 sample wells and the well labelled with a ladder symbol (Figure 4.1).

v) **Loading the Ladder and Samples**

Prepared ladder (1 μl) was pipetted to the well labelled with a ladder symbol and 1 μl of sample was pipetted to each respective well of the 11 sample wells (Figure 4.1). RNA marker (1 μl) was added to any unused sample wells. The chip was placed in an IKA vortexer (provided) and vortexed for 1 min at 2400 rpm, and then the chip was run in the Bioanalyzer instrument with the appropriate assay, sRNA, selected from the software menu.

4.2.10 **Agilent Bioanalyzer 2100 expert B02.04 software – determining regions for each LMW-RNA component**

A successful sRNA run results in one marker peak and two regions that are defined arbitrarily by the Bioanalyzer expert B02.04 software: the LMW-RNA region from 0 to 150 nt, and the miRNA (sRNA) region from 10 to 40 nt. These regions were modified by selecting the region table within the software and setting regions for analysis to extend the LMW-RNA region to 200 nt and to encompass the composition of LMW-RNA components i.e. 50 - 80 nt for transfer RNAs (tRNAs of 73 – 94 nt), 85 - 100 nt for small nucleolar RNAs (snoRNAs, e.g. U14 of 84 - 106 nt) and 110 - 180 nt for ribosomal RNAs (rRNAs; 5.8 S of 154 nt and 5 S of 120 nt) in addition to the sRNA region already set (10 – 40 nt). The LMW-RNA component regions were set to encompass average nt size reported for each LMW-RNA components across literature (Mackay et al. 1980; Leader et al. 1994; Wei et al. 2014).
4.2.11 Statistical analyses of data from the Bioanalyzer method

In order to determine linearity and limits of detection, scatter plots of the mean data of amount of LWM-RNA loaded and amount of LMW-RNA reported across three replicates of mock-inoculated and TuMV-inoculation 10-point two-fold serial dilutions were created and power trend-lines fitted and R² values reported. In order to determine reproducibility between sRNA chips, scatterplots of the data above were created and standard error of the mean reported. In order to determine reproducibility within sRNA chips the values of four technical replicates of five samples of mock-inoculated two-fold serial dilution (from 30 to 0.95 ng/µl) LMW-RNA were compared.

Regression analysis was performed to estimate the relationship between each LMW-RNA component (sRNA, tRNA, snoRNA and rRNA) and each virus-inoculated and mock-inoculated LMW-RNA from 120 observations. Mixed model repeated measures analyses of fixed (LMW-RNA components) and random (three biological replicates) effects was kindly performed by Dr Nihal de Silva using the statistical analysis software system SAS/STAT 9.22 to determine if there was a difference in LMW-RNA components profile between each of the five virus-inoculations and mock-inoculation across the 2 to 42 dpi time course.

To establish if the measured load on two components differed, principal component analysis was performed on the entire data set thus converting the complex and multidimensional data into a lower number of variables. If the load between the two calculated components differed it could be determined that the accumulation of LMW-RNA components of the five virus-inoculations can be distinguished from mock-inoculation.
4.3 Results

4.3.1 Nanodrop quantification and assessment of LMW-RNA

Following elution, all LMW-RNA samples were quantified and assessed for quality using NanoDrop spectrophotometry. All samples had an absorbance ratio (A260/280 nm) of between 1.8 and 2.2 where a ratio of ~2.0 is generally accepted as ‘pure’ or of good quality for RNA (NanoDrop Technologies 2007). A ratio appreciably lower may indicate the presence of protein, phenol or other contaminants that absorb strongly at or near 280 nm (NanoDrop Technologies 2007). Following quantification, all LMW-RNA samples were normalised to 30 ng/µl for the reasons discussed in the results section (4.4.6) of this chapter.

4.3.2 Polyacrylamide gel electrophoresis of LMW-RNA

In order to determine linearity, reproducibility and limits of detection of the PAGE/Typhoon/ImageQuant method, TuMV-inoculated Arabidopsis samples from 21 dpi collection were diluted in a 10-point two-fold serial dilution from a starting amount of 5.0 µg. TuMV samples were selected because TuMV gives strong symptomology in Arabidopsis at 21 dpi and it was supposed that strong symptomology might correlate with higher accumulation of sRNAs and as such, provide a positive yield. However, it could be argued that the antiviral mechanism of RNAi (and accumulation of sRNAs) would result in less viral symptoms. Figure 4.2 shows the resultant PAGE gel stained with EtBr prior to imaging with Bio-Rad Gel Doc 1000 imaging system (Bio-Rad Laboratories, Hercules, CA, U.S.A.).
Figure 4.2. Novex ® TBE 8% (10 well) gel of LMW-RNA from two-fold serial dilution of *Arabidopsis* 21 days post inoculation with TuMV and stained with ethidium bromide (EtBr).

Lane 1, 10 bp Ladder. Lanes 2-10 contain 5, 2.5, 1.25, 0.625, 0.313, 0.156, 0.078, 0.039 and 0.012 ug amounts of LMW RNA isolated from 21 dpi TuMV-inoculated *Arabidopsis*, respectively. The arrow indicates 20 bp.

In the Bio-Rad Gel Doc 1000 image of the two fold serial dilution LMW-RNA from 21 dpi TuMV-inoculated *Arabidopsis*, bands are visualised between 20 and 30 bp in lanes 2, 3 and 4. These bands are likely to represent sRNAs of between 21 to 24 nt. The bands visualised in lanes 2 to 9 at 100 bp (corresponding to the ladder in Lane 1) decrease two-fold in intensity with each serial dilution and are likely to represent rRNA. These results show that a two-fold serial dilution can be visualised by PAGE stained with EtBr.

**4.3.3 Gel scanning with Typhoon 9140 imager and ImageQuant analysis**

The previous gel (Figure 4.2) was transferred to the Typhoon imager and scanned and the image output with ImageQuant analysis and delineation of lanes and bands is presented in Figure 4.3.
Figure 4.3. Novex® TBE 8% (10 well) gel of LMW-RNA from two-fold serial dilution of *Arabidopsis* 21 dpi with TuMV imaged with Typhoon 94100 and analysed with ImageQuant TL 7.0 automatic analysis.

L, 10 bp, Lane 1 1Kb ImageQuant Ladder. Lanes 2-10 contain 5, 2.5, 1.25, 0.625, 0.313, 0.156, 0.078, 0.039 and 0.012 ug amounts of LMW RNA isolated from 21 dpi TuMV-inoculated *Arabidopsis*, respectively. The arrows indicates 20 and 100 bp.

In the Typhoon/ImageQuant image of the two fold serial dilution LMW-RNA from 21 dpi TuMV-inoculated *Arabidopsis*, 20 bp bands are visualised in lanes 2, 3 and 4. These bands are likely to represent sRNAs of between 20 to 30 nt. The results show that bands can be visualised more clearly with Typhoon/ImageQuant than with the Bio-Rad Gel Doc. Given that data was not reported across the entire dilution series for the sRNA fraction, data for determining linearity, reproducibility and limits of detection was obtained from the rRNA bands detected in lanes 2 – 9 corresponding to 100 bp of the 10 bp ladder (L).
4.3.4 Statistical analysis of PAGE/Typhoon/ImageQuant method data

In order to demonstrate the reliability of the quantification method, a linear relationship between LMW-RNA amount loaded (µg) and band density reported was required as was good reproducibility both between and within gels. Furthermore, it was necessary to establish the limits of detection of the PAGE/Typhoon/ImageQuant method. To determine linearity, reproducibility and limits of detection, data of each of four technical replicates of a ten point two-fold serial dilution of TuMV-inoculated Arabidopsis LMW-RNA (21 dpi) obtained from ImageQuant analysis software was used to generate a scatterplot. Linear trend-lines were fitted and $R^2$ values reported for each of the four replicates (Figure 4.4).

![Figure 4.4. Scatterplot of ImageQuant data of four technical replicates of a two-fold ten point serial dilution of 21 dpi TuMV-inoculated Arabidopsis LMW-RNA.](image)

Data of four replicates of a serial dilution; 5.0, 2.5, 1.25, 0.625, 0.313, 0.156, 0.078, 0.039, 0.12 and 0.001 pg LMW-RNA are represented by blue (Rep1), red (Rep2), green (Rep3) and purple (Rep4) data points respectively. Linear trend-lines are fitted for each respective replicate and regression values shown ($R^2$).
The $R^2$ values of the four replicates have acceptable degrees of linearity (0.957, 0.881, 0.968 and 0.974 respectively). However, the above results suggest poor reproducibility between gels where band intensity and $R^2$ values differ between all replicates at all dilution points. Data was obtained from the reported intensity of rRNA bands because data for all of the sRNA fractions was not acquired; this result alone suggests poor limits of detection. In addition, the results show that ImageQuant analysis was only able to report the highest eight amounts of rRNA (to 0.039 pg) therefore showing a narrow limit of detection.

In order to determine within gel reproducibility, scatterplots of the data of 10 replicates of each amount, 5.0, 2.5, 1.25 and 0.625 µg of TuMV-inoculated Arabidopsis LMW-RNA were plotted with error bars representing standard error of the mean. Figures 4.5A to 4.5D demonstrate variability at every concentration examined as evidenced, for the most part, by overlapping standard error of the mean.
Figure 4.5. Scatterplots of band intensity reported from 10 replicates of 5.0, 2.5, 1.25 and 0.625 μg TuMV-inoculated LMW-RNA to determine within gel reproducibility of the PAGE/Typhoon/ImageQuant method.

A. Band intensity reported for 10 technical replicates of 5.0 μg TuMV-inoculated LMW-RNA. B. Band intensity reported for 10 replicates of 2.5 μg TuMV-inoculated LMW-RNA. C. Band intensity reported for 10 replicates of 1.25 μg TuMV-inoculated LMW-RNA. D. Band intensity reported for 10 replicates of 0.625 μg TuMV-inoculated LMW-RNA. Error bars are standard error of the mean.
4.3.5 Agilent Bioanalyzer B02.04 software for determining region for each LMW-RNA component

Before initial assessments to determine good linearity, reproducibility and limits of detection of the Bioanalyzer method (Chapter 4.3.6), the accumulation of LMW-RNA components of mock- TuMV- and TYMV-inoculated LMW-RNA were analysed. The results of a chip run of three biological replicates of each mock-, TuMV- and TYMV-inoculated Arabidopsis LMW-RNA at 21 dpi shown in Figure 4.6 indicates that LMW-RNA can be quantified as distinct regions of banding. The band regions for each component of LMW-RNA (sRNAs, tRNAs, snoRNAs and rRNAs) were set as described in section 4.3.8 of this chapter.

![Figure 4.6. Example of Agilent Bioanalyser 2100 electrophoretic gel output with LMW-RNA components labelled.](image)

Three biological replicates (bioreps) of mock-, TuMV- and TYMV-inoculated Arabidopsis LMW-RNA at 21 dpi. L, Ladder corresponding to 4 to 150 nucleotides (nt), 1, mock-inoculated LMW-RNA biorep 1; 2, mock biorep 2; 3, mock biorep 3; 4, TuMV-inoculated LMW-RNA biorep 1; 5, TuMV biorep 2; 6, TuMV biorep 3; TYMV-inoculated biorep 1; 8, TYMV biorep 2; 9, TYMV biorep 3. Low molecular weight RNAs (LMW-RNAs), small RNAs (sRNAs; including miRNAs and siRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs) and ribosomal RNAs (rRNAs).
The electrophoretic gel output (Figure 4.6) suggests differences in the accumulation of all LMW-RNA components when comparing mock-inoculated LMW-RNA with TuMV-inoculated LMW-RNA data. More specifically, the region of dark banding between 10 – 40 nt (sRNAs) of TuMV-inoculated LMW-RNA (lanes 4-6) suggests a higher accumulation than the sRNAs of mock-inoculated LMW-RNA (no bands apparent between 10 - 40 nt in lanes 1-3). Conversely the rRNAs (c.110 nt) of TuMV (lanes 4-6) appear decreased in accumulation compared to mock (lanes 1-3). However, the gel output shows no difference in accumulation of sRNAs of TYMV-inoculated LMW-RNA while the rRNAs of TYMV-inoculated LMW-RNA (lanes 7, 8 and 9) appear decreased in accumulation compared to mock-inoculation (lanes 1-3). Combined data for chip runs of three biological replicates of mock, CaMV, TMV, TSWV, TuMV and TYMV-inoculated Arabidopsis LMW-RNA at 21 dpi are included for comparison in Figure AII.1 of Appendix II.

4.3.6 Statistical analyses of data from the Bioanalyzer method

In order to determine linearity, reproducibility and limits of detection, mean data of all three replicates (Rep 1, Rep 2 and Rep 3) of two-fold 10 point serial dilutions of mock-inoculated and TuMV-inoculated 21 dpi LMW-RNA were analysed. Figures 4.7.A and B present the resultant scatterplots with power trend-lines fitted and illustrate the limits of detection and linearity of the data of three biological replicates mock-inoculated and TuMV-inoculated LMW-RNA, respectively. The Bioanalyzer method was able to report concentrations for all 10 dilution points (of the LMW-RNA fraction) for both treatments with consistently high $R^2$ values ($R^2$ minimum 0.979 and maximum 0.996).
Figure 4.7. Scatterplots of 10 point two-fold serial dilutions of mock- and TuMV-inoculated *Arabidopsis* LMW-RNA at 21 dp to determine linearity and limits of detection of the Bioanalyzer method.

A. Scatterplot of serial dilution amount LMW-RNA (pg) loaded against amount LMW-RNA reported (pg) of three biological replicates (Rep 1, 2 and 3) of mock-inoculated *Arabidopsis* LMW-RNA at 21 dpi. 

B. Scatterplot of amount LMW-RNA (pg) loaded against concentration LMW-RNA reported (pg) of three biological replicates of TuMV-inoculated *Arabidopsis* at 21 dpi. Power trend-lines are fitted for each replicate and regression values shown ($R^2$).
In order to determine within chip reproducibility, the mean and standard error of the same biological replicate data of Figures 4.7A and B were calculated and presented in Table 4.1. The data outlines reproducibility between three sRNA chips and three biological replicates of mock-inoculated LMW-RNA and shows little variation between chips at every concentration as evidenced by minimal error (minimum error 0.1 from 234.4 pg, maximum 65.2 from 60,000 pg). Similarly, Table 4.1 outlines reproducibility between sRNA chips and biological replicates of TuMV-inoculated LMW-RNA and shows negligible variation as evidenced by minimal error (minimum error 1.5 from 117.2 pg, maximum 508.4 from 60,000 pg). In order to determine between chip reproducibility mean and standard error of the mean data of four technical replicates of a five point two-fold serial dilution of mock-inoculated 21 dpi LMW-RNA were calculated. The data (Table 4.1) show negligible variability within chips as evidenced by minimal error (minimum error 20.8 from 1563.0 pg, maximum 299.4 from 6250 pg).
Table 4.1. Amount loaded, amount reported (pg) and standard error of the mean of replicates of LMW-RNA. A. Three biological replicates of a 10 point two-fold serial dilution of mock-inoculated *Arabidopsis* LMW-RNA at 21 dpi. B. TuMV-inoculated *Arabidopsis* at 21 dpi. C. Four technical replicates of a five-point two-fold serial dilution of mock-inoculated *Arabidopsis* LMW-RNA at 21 dpi.

<table>
<thead>
<tr>
<th>Amount loaded (pg)</th>
<th>Mean of concentration reported (pg)</th>
<th>± Standard Error of the Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Mock-inoculated LMW-RNA within chip (n=3).</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60000</td>
<td>6197.9</td>
<td>65.2</td>
</tr>
<tr>
<td>30000</td>
<td>3611.3</td>
<td>38.5</td>
</tr>
<tr>
<td>15000</td>
<td>1463.9</td>
<td>7.3</td>
</tr>
<tr>
<td>7500</td>
<td>711.0</td>
<td>4.9</td>
</tr>
<tr>
<td>3750</td>
<td>356.8</td>
<td>5.7</td>
</tr>
<tr>
<td>1875</td>
<td>143.3</td>
<td>2.3</td>
</tr>
<tr>
<td>937.5</td>
<td>96.6</td>
<td>2.2</td>
</tr>
<tr>
<td>468.8</td>
<td>43.3</td>
<td>0.9</td>
</tr>
<tr>
<td>234.4</td>
<td>28.3</td>
<td>0.1</td>
</tr>
<tr>
<td>117.2</td>
<td>15</td>
<td>1.2</td>
</tr>
<tr>
<td><strong>B. TuMV-inoculated LMW-RNA within chip (n = 3).</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60000</td>
<td>12603.5</td>
<td>394.9</td>
</tr>
<tr>
<td>30000</td>
<td>7356.7</td>
<td>508.4</td>
</tr>
<tr>
<td>15000</td>
<td>3527.9</td>
<td>346.6</td>
</tr>
<tr>
<td>7500</td>
<td>1772.3</td>
<td>222.1</td>
</tr>
<tr>
<td>3750</td>
<td>670.4</td>
<td>35.3</td>
</tr>
<tr>
<td>1875</td>
<td>339.0</td>
<td>18.1</td>
</tr>
<tr>
<td>937.5</td>
<td>140.0</td>
<td>23.7</td>
</tr>
<tr>
<td>468.8</td>
<td>79.2</td>
<td>8.0</td>
</tr>
<tr>
<td>234.4</td>
<td>44.2</td>
<td>3.3</td>
</tr>
<tr>
<td>117.2</td>
<td>26.33</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>C. Mock-inoculated LMW-RNA between chip (n=4).</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25000</td>
<td>23846.9</td>
<td>105.4</td>
</tr>
<tr>
<td>12500</td>
<td>12529.7</td>
<td>145.8</td>
</tr>
<tr>
<td>6250</td>
<td>7345.9</td>
<td>299.4</td>
</tr>
<tr>
<td>3125</td>
<td>3222.2</td>
<td>174.1</td>
</tr>
<tr>
<td>1563</td>
<td>1121.2</td>
<td>20.8</td>
</tr>
</tbody>
</table>

Taken together the results shown in section 4.4.6 of this chapter suggest that the Bioanalyzer method provides data that are linear and reproducible between around 10000 to 100 pg (100 to 0.1 ng) and has high sensitivity as evidenced by good limits of detection. Given that the maximum yield of the LMW-RNA isolations of this study averaged between 50 and 100 ng as quantified by Nanodrop spectrophotometry as previously described (Chapter 4.3.1), 30 ng was
selected as the concentration to load for all future samples and all samples were normalised to 30 ng/µl as described in section 4.2.1.

4.3.7 Analysis of the proportion LMW-RNA components of total LMW-RNA

In order to quantify separate LMW-RNA components (sRNA, tRNA, snoRNA and rRNA) from *Arabidopsis* total LMW-RNA, samples of three biological replicates were loaded onto sRNA chips for analysis by the Bioanalyzer method. Each LMW-RNA component was analysed first as proportion of total LMW-RNA reported for mock-inoculation and all virus-inoculations at the same dpi and standard error of the mean calculated. Second, each LMW-RNA component was analysed as a proportion of total LMW-RNA where the average of all mock-inoculated data at each time point was rescaled to 1.0 (proportion of averaged mock-inoculation). Analysing LMW-RNA components as a proportion of averaged mock-inoculation enables inference of LMW-RNA component accumulation irrespective of dpi and has the practical advantage of not needing to know the history of the virus infection within a plant.

Since the results for proportion of mock- and virus-inoculated data demonstrated the same trend as the results for proportion of averaged mock-inoculation data, only the data of each LMW-RNA component as a proportion of averaged mock-inoculated data are shown. Data for proportions of mock- and virus-inoculated at the same time point dpi are presented in Appendix III. Figure 4.8A illustrates the proportion sRNA of total LMW-RNA compared to averaged mock-inoculation, where an increase is above the line marked at 1 and a decrease, below. CaMV sRNAs show an increase in proportion of total LMW-RNA compared to averaged mock-inoculation from 2-21 dpi, TMV at 2, 7, 14 and 42; TSWV between 14 and 42 dpi, TuMV at 2, 3, 14 and 21 dpi and TYMV at every time point, with the exception of 21 dpi.
Figure 4.8B illustrates the proportion tRNA of total LMW-RNA compared to averaged mock-inoculation. No change is seen in the proportion tRNA for CaMV, TMV, TSWV and TYMV while the proportion of TuMV tRNA decreases at 2 and 21 dpi, compared to averaged mock-inoculation.

Figure 4.8C illustrates the proportion snoRNA of total LMW-RNA compared to averaged mock-inoculation. CaMV snoRNAs show a decrease in proportion between 3 and 14 dpi, TMV between 2 and 28 dpi; TSWV at 3, 21 and 28 dpi, TuMV between 7 and 21 dpi and TYMV between 3 and 28 dpi.

Figure 4.8D illustrates the proportion of rRNA of total LMW-RNA compared to averaged mock-inoculation. CaMV rRNAs show a decrease in proportion at every time point except 21 dpi, TMV at every time point, TSWV at every time point except 14 dpi, TuMV between 3 and 21 dpi and TYMV at every time point except 21 dpi.

Taken together these data suggest a great deal of variability of all components of LMW-RNA at every time point for all virus inoculations. The data for TuMV shows the most consistent profile across the infection time course with the accumulation of sRNAs increasing and the accumulation of tRNA, snoRNA and rRNAs decreasing compared to averaged mock-inoculation, at most time points.
Figure 4.8. Proportion LMW-RNA components (sRNA and tRNA) of total LMW-RNA compared to averaged mock-inoculation rescaled to 1.0.

A. Proportion sRNA of CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA compared to averaged mock-inoculation. B. Proportion tRNA of CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA compared to averaged mock-inoculation. Black line at 1.0 corresponds to rescaled averaged mock-inoculation.
Figure 4.8. Proportion LMW-RNA components (snoRNA and rRNA) of total LMW-RNA compared to averaged mock-inoculation rescaled to 1.0.

C. Proportion snoRNA of CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA compared to averaged mock-inoculation. D. Proportion rRNA of CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA compared to averaged mock-inoculation. Black line at 1.0 corresponds to rescaled averaged mock-inoculation.
4.3.8 Tests of statistical significance

Regression analysis of the proportion of the individual components of total LMW-RNA across dpi and all five virus-inoculations compared to mock-inoculated LMW-RNA, was performed, and calculated p-values obtained from 120 observations, in order to determine significance of individual components (Figure 4.10; full data set Appendix IV). P-values were considered significant if less than the conventional level of significance, i.e. 0.05 (reviewed in Nuzzo 2014).

The results of regression analysis are presented in Figure 4.9 and summarised in Table 4.2. The results show a statistically significant increase in sRNA from TuMV-inoculated LMW-RNA (p<0.001) and a borderline significant increase in the proportion of sRNA from TYMV-inoculated LMW-RNA (p=0.010) compared to the proportion of sRNA from mock-inoculated LMW-RNA (Figure 4.10A). The results show a significant increase in the proportion of tRNA from TMV and a significant decrease in proportion of tRNA from TuMV-inoculated LMW-RNA (p<0.01) compared to the proportion of tRNA from mock-inoculated LMW-RNA (Figure 4.10B) CaMV-inoculated LMW-RNA has a borderline significance towards an increase in the proportion of tRNA (p=0.074), a significant decrease in the proportion of snoRNA from TMV-inoculated LMW-RNA (p=0.01) compared to the proportion of snoRNA from mock-inoculated LMW-RNA is seen (Figure 4.10C). A significant decrease in the proportion of rRNA of CaMV-(p<0.001), TMV-(p<0.001) and TYMV-inoculated (p<0.01) LMW-RNA compared to the proportion of rRNA from mock-inoculated LMW-RNA is seen (Figure 4.10D).
Table 4.2. Summary table of the significance (from regression analysis) of the proportion of components (sRNA, tRNA, snoRNA and rRNA) of virus-inoculated LMW-RNA in comparison to components of mock-inoculated LMW-RNA across the time course (2-42 dpi).

<table>
<thead>
<tr>
<th>LMW-RNA component</th>
<th>CaMV</th>
<th>TMV</th>
<th>TSWV</th>
<th>TuMV</th>
<th>TYMV</th>
</tr>
</thead>
<tbody>
<tr>
<td>sRNA</td>
<td></td>
<td></td>
<td></td>
<td>+ve</td>
<td>b +ve*</td>
</tr>
<tr>
<td>tRNA</td>
<td>b –ve*</td>
<td></td>
<td>+ve</td>
<td></td>
<td>-ve</td>
</tr>
<tr>
<td>snoRNA</td>
<td></td>
<td>-ve</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rRNA</td>
<td>-ve</td>
<td>-ve</td>
<td>+ve</td>
<td></td>
<td>-ve</td>
</tr>
</tbody>
</table>

+ve increase compared to mock-inoculated
-ve decrease compared to mock-inoculated
*borderline significance compared to mock inoculated
Blank cells = no significance
Figure 4.9. Scatterplots of regression analysis of three replicates of the proportion of components of total virus-inoculated LMW-RNA across the time course (2 - 42 dpi) in comparison to total mock-inoculation inoculated LMW-RNA (n=21, TuMV n=15).

A. Proportion sRNA of total LMW-RNA of CaMV, TMV, TSWV, TuMV and TYMV-inoculated LMW-RNA compared to mock-inoculated LMW-RNA across dpi. B. Proportion tRNA. C. Proportion snoRNA. D. Proportion rRNA. P values of <0.05 calculated from regression analysis are considered significant.
Table 4.3. Multivariate mixed model analysis output. Model information refers to dependent variable sRNA and subject effect replicate plants. Significance determined from Least Squares Means and Differences of Least squares Means with Dunnett-Hsu adjustment. Values of significance are shaded and highlighted in bold.

<table>
<thead>
<tr>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Dependent Variable</td>
</tr>
<tr>
<td>Covariance Structures</td>
</tr>
<tr>
<td>Subject Effect</td>
</tr>
<tr>
<td>Estimation Method</td>
</tr>
<tr>
<td>Residual Variance Method</td>
</tr>
<tr>
<td>Fixed Effects SE Method</td>
</tr>
<tr>
<td>Degrees of Freedom Method</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Least Squares Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Differences of Least Squares Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
<tr>
<td>Virus</td>
</tr>
</tbody>
</table>

To further investigate the data, principal component analysis was performed to determine if the measured load on two selected components, calculated from all possible combinations, differed and if so whether the five virus-inoculations could be distinguished from mock-inoculation. Principal component analysis was performed on the 18 variable data set: mock-inoculation; the five viruses, CaMV, TMV, TSWV, TuMV and TYMV; the time-course of sample
collection. 2, 3, 7, 14, 21, 28, 35, and 42 dpi and the LMW-RNA components, sRNA, tRNA, snoRNA and rRNA. In order to reduce the number of components Eigenvalues (\(\lambda\)) were calculated (full data set in Appendix V). There were two components out of four with Eigenvalues of greater greater than 1 \(\lambda\). Component one (2.17 \(\lambda\)) and component two (1.61 \(\lambda\)). High amounts of snoRNA (0.63) and rRNAs (0.64) were loaded on component one and low amounts of sRNA (- 0.77) and high amounts of tRNA (0.60) loaded on component two. Figure 4.12 shows the scatter plot generated from scores calculated for each component and demonstrates no clustering of separate colour markers (inoculations). There are however three scores in yellow (representing TuMV) that are separated from mock-inoculation (represented by white markers) towards component one (sRNA / tRNA). This is in keeping with the results of the regression and mixed model analyses undertaken where TuMV-inoculation resulted in a significant increase in sRNA (regression, \(p<0.001\); mixed model, \(P<0.0001\); Dunnett-Hsu adjustment \(P< 0.05\)) and significant decrease in tRNA (regression \(P<0.01\)) components compared to mock-inoculation. While three TuMV scores are separate from the mock-inoculated scores there is no defined separation between mock-inoculated and virus-infected Arabidopsis LMW-RNA components. These results establish that none of the LMW-RNA components or time points of collection (dpi) of five virus-inoculations are distinct from mock-inoculation.
Figure 4.10. Principal component analysis plot.

Scatterplot of scores on component one (high snoRNA, rRNA) and component two (low sRNA, high tRNA) for 120 samples of the 18 variable data set. Variables are mock-inoculation; the five viruses, CaMV, TMV, TSWV, TuMV and TYMV; the time-course of 2, 3, 7, 14, 21, 28, 35, and 42 dpi (to 28 dpi for TuMV) and LMW-RNA components, sRNA, tRNA, snoRNA and rRNA. Colours representing the scores of mock and five virus inoculations are presented in the plot legend.

4.3.9 Component LMW-RNA accumulation profile of *Tamarillo mosaic virus* (TamMV)

Since TuMV-inoculated LMW-RNA showed a significant increase in sRNA and decrease in tRNA components, a further potyvirus was investigated in order to determine if the accumulation pattern seen for TuMV might be indicative potyviruses in general. In order to quantify accumulation of separate LMW-RNA components from *Solanum betaceum* (tamarillo) in response to *Tamarillo mosaic virus* (TamMV), total LMW-RNA samples of three biological replicates across a time course of 2 to 28 dpi were loaded onto sRNA chips for analysis by the Bioanalyzer method. Each LMW-RNA component was analysed first as proportion of total
LMW-RNA reported for mock-inoculation and all virus-inoculations at the same dpi and standard error of the mean calculated. Second, each LMW-RNA component was analysed as a proportion of total LMW-RNA where the average of all mock-inoculated data at each time point was rescaled to 1.0 (proportion of averaged mock-inoculation). Data for proportions of mock- and virus-inoculated at the same time point dpi are presented in Appendix III.

Figure 4.11A illustrates the proportion sRNA of total LMW-RNA compared to averaged mock-inoculation, where an increase is above the line at 1.0 and a decrease, below. TamMV sRNAs show a decrease in proportion of total LMW-RNA compared to averaged mock-inoculation at 7 dpi and an increase from 14 to 28 dpi. Compared to averaged mock-inoculation, Figure 4.11B illustrates the proportion tRNA of total LMW-RNA compared, where TamMV tRNAs show a decrease in proportion of total LMW-RNA at every time point with the exception of 3 dpi.

Figure 4.11C illustrates the proportion sRNA of total LMW-RNA compared to averaged mock-inoculation, where TamMV snoRNAs show a decrease in proportion of total LMW-RNA at every time point. Compared to averaged mock-inoculation, Figure 4.11D shows a decrease in proportion rRNA accumulation of total LMW-RNA for TamMV at every time point dpi with the exception of 3 dpi.
Figure 4.11. Proportion of LMW-RNA components of *Tamarillo mosaic virus* (TamMV)-inoculated Tamarillo LMW-RNA compared to LMW-RNA components of averaged mock-inoculated Tamarillo LMW-RNA.

A. Proportion sRNA of TamMV-inoculated total LMW-RNA compared to averaged mock-inoculation. B. Proportion tRNA of TamMV-inoculated total LMW-RNA compared to averaged mock-inoculation. C. Proportion snoRNA of TamMV-inoculated total LMW-RNA compared to averaged mock-inoculation. D. Proportion rRNA of TamMV-inoculated total LMW-RNA compared to averaged mock-inoculation. Black line at 1 corresponds to rescaled averaged mock-inoculation.
4.3.10 Ratio of sRNA/rRNA accumulation as a tool predictive of virus infection

Because many sRNAs appear to increase in accumulation and many rRNAs appear to decrease in accumulation at most time points and in response to most virus-inoculations compared to averaged mock-inoculation, the ratio of sRNA to rRNA as a proportion of averaged mock-inoculation was calculated. The aim of this calculation was to determine whether there was a high correlation between a plant being infected and a ratio of sRNA accumulation / rRNA accumulation greater than averaged mock-inoculation rescaled to 1.0 (see equation below) that could be used as a predictive tool. Figure 4.12 shows that of 33 samples, 31 are greater than the averaged mock-inoculation, where mock inoculation equals 1.0. This shows that, in this instance, generating a ratio of sRNA to rRNA as a proportion of averaged mock-inoculation correlates with a 94% probability of detecting virus infection across the known virus inoculations and time course of the present study.

Equation

The probability that a sample is virus infected can be determined where:

\[
\frac{\frac{\text{virus-inoculated rRNA accumulation}}{\text{virus total LMW-RNA}}}{\frac{\text{mock-inoculated rRNA accumulation}}{\text{mock total LMW-RNA rescaled to 1.0}}} / \frac{\frac{\text{virus-inoculated sRNA accumulation}}{\text{virus total LMW-RNA}}}{\frac{\text{mock-inoculated sRNA accumulation}}{\text{mock total LMW-RNA rescaled to 1.0}}} > 1.0.
\]
Figure 4.12. Ratio of sRNA to rRNA of total LMW-RNA as a proportion of averaged mock-inoculation where mock-inoculation is rescaled to 1.0.

Ratio of sRNA to rRNA of CaMV, TMV, TSWV, TuMV and TYMV–inoculated total LMW-RNA compared to averaged mock-inoculation across the time course of 2 to 42 dpi. Black line at 1 corresponds to averaged mock-inoculation. The ratio sRNA to rRNA of TuMV-inoculated LMW-RNA compared to averaged mock at 14 dpi is 21.69 and the ratio of sRNA to rRNA of TuMV-inoculated LMW-RNA compared to averaged mock at 21 dpi is 78.78.

Similarly the same calculation was applied to the data for the ratio of sRNA to rRNA for TamMV-inoculated LMW-RNA. Figure 4.13 shows that of six samples, all six are greater than averaged mock-inoculation (1.0).
Figure 4.13. Ratio of sRNA to rRNA of total TamMV-inoculated LMW-RNA as a proportion of averaged mock-inoculation where mock-inoculation is rescaled to 1.0.

Ratio of sRNA to rRNA of TamMV–inoculated total LMW-RNA compared to averaged mock-inoculation across the time course of 2 to 28 dpi. Black line at 1 corresponds to averaged mock-inoculation.

The results of the calculation of the ratio of sRNA to rRNA of LMW-RNA as a proportion of averaged mock-inoculation suggests calculation of this ratio could, with further investigation, prove to be a useful tool for predicting the likelihood of virus presence from plant tissues.
4.4 Discussion

The experimental work of this chapter set out to evaluate methods to isolate high quality LMW-RNA and to accurately quantify sRNA accumulation. Trials of three commercial miRNA isolation kits determined that the mirPremier™ microRNA isolation kit provided the most consistent yields and best quality and was therefore utilised for all LMW-RNA isolations of this study. Two methods, (i) PAGE/Typhoon/ImageQuant method and (ii) the Bioanalyzer method, were assessed for feasibility as quantitative measures of sRNA accumulation from Arabidopsis samples across a time-course (2, 3, 7, 14, 21 and 28 dpi) and treatments (Mock, CaMV, TMV, TSWV, TuMV and TYMV). The conclusions for the two methods are discussed in the following sections; 4.4.1 and 4.4.2 respectively.

4.4.1 Conclusions for PAGE/Typhoon/ImageQuant Method

While bands could be visualised in the sRNA regions of the first three lanes of resultant Typhoon images, analysis of data revealed that the PAGE/Typhoon/ImageQuant method could not detect sRNA beyond the first three concentrated dilutions and subsequently had poor limits of detection, poor linearity and poor reproducibility both within and between gels. In addition, the processes were time consuming, used a substantial amount of reagents and samples (c. 25 µl) and the systems and software were not straightforward to use. As such, this method was deemed unsuitable and impractical for quantifying sRNA accumulation and was not pursued further.
4.4.2 Conclusions for the Bioanalyzer method

Analysis of preliminary data obtained from the Bioanalyser platform with sRNA chips demonstrated that the Bioanalyzer method had good linearity and good reproducibility both within and between chips and good limits of detection to 0.1 ng. In contrast to the PAGE/Typhoon/ImageQuant method, the microfluidics based Bioanalyzer method required 1 µl sample and provided a much faster, more sensitive and reproducible alternative. As such, the Bioanalyzer method was deemed suitable and practical for quantifying sRNA accumulation and was pursued further. In addition to sRNA quantification, the Bioanalyzer method allows the analysis of all components of LMW-RNA, as defined by the user. The accumulation of tRNA, snoRNA and rRNAs were therefore analysed in addition to sRNA.

The proportion of the individual LMW-RNA components of total LMW-RNA suggested a trend toward increased accumulation of sRNA compared to averaged mock-inoculation for most of the virus inoculations (especially from 14 dpi), decreased tRNA compared to mock-inoculation for TuMV-inoculation, decreased snoRNAs for all virus inoculations and decreased rRNA for most virus inoculations compared to mock-inoculation. While initial summary statistics showed changes in accumulation of some LMW-RNA components at some time points (dpi), the statistical significance from regression analysis of the proportion of sRNA across all time points of infection showed that only TuMV infection increases the sRNA proportion of LMW-RNA (TuMV; p < 0.001). This finding was supported by multivariate mixed model analysis where only TuMV infection increased the sRNA proportion of LMW-RNA compared to the sRNA proportion of mock-inoculated LMW-RNA. The statistical significance of the proportion of rRNAs across the time course of infection showed that most virus infections alter
the rRNA proportion (CaMV, TMV, TSWV and TYMV; \( p < 0.01 \)) of LMW-RNA. Principal component analysis could not differentiate all virus-inoculated LMW-RNA components from mock-inoculated LMW-RNA components at every time point (dpi).

The results of the present study showing that only TuMV infection increases the sRNA proportion of LMW-RNA is in contrast with studies that have shown that sRNAs accumulated in plants infected with a number of viruses (e.g. reviewed in Ruiz-Ferrer and Voinnet 2009 and reviewed in Qu et al. 2010). However, previous studies, based on northern blots, have shown that transgenic expression of VSRs from many different plant viruses often results in reduced accumulation of conserved miRNAs. The biological function of this down-regulated miRNA accumulation in viral infection and/or plant defence remains to be understood (reviewed in Du et al. 2011). It is possible that reduced accumulation of specific miRNAs in response to VSRs would result in a decrease in the overall sRNA component of LMW-RNA.

The results of the present study show that compared to mock-inoculation, the tRNA component of LMW-RNA decreases in response to TMV, TuMV and TamMV, and a borderline decrease in response to CaMV and virus infection. These findings are in partial agreement with other studies where tRNA synthesis has been shown to decrease in response to virus infection. For example, during viral synthesis of TMV in young inoculated *Nicotiana tabacum* leaves, tRNA synthesis decreased by more than 50% (reviewed in Nayudu 2008). Connections between tRNA biology and plant virology occur at many stages and in diverse ways. For example many viruses utilise tRNAs for viral function, either by encoding tRNAs in the viral genome to support viral protein synthesis or in the case of retroviral genomes, using host tRNAs as primers to prime reverse transcription and replication (reviewed in Dreher 2010). The genomic positive strand RNAs of plant viruses such as those of the *Tymoviridae, Virgaviridae* and *Bromoviridae* all have
a tRNA-like structure at their 3’ end. These tRNA-like structures mimic some tRNA functions, such as aminoacylation. There are three aminoacylation specificities, valine, histidine and tyrosine. For example, valine binds to the tRNA-like structure of the TYMV genome whilst tyrosine binds to the tRNA-like structure of the *Barley stripe mosaic virus* genome (Barends et al. 2003). The function of this ‘tRNA mimicry’ is to regulate viral transcription, translation initiation and termination, elongation and replication (reviewed in Nakamura and Ito 2011). Given host plant tRNAs have been redirected for viral purposes it is not surprising that a decrease in tRNA accumulation would be seen in response to virus infection.

The results of the present study show that compared to average mock-inoculation, the snoRNA component of LMW-RNA decreases in response to TMV. snoRNAs direct a series of nucleolytic cleavages and base modifications, such as site-specific methylation and pseudouridylation of rRNAs. Little is known about the role of snoRNAs in plant virus infection. However, an umbravirus-encoded ORF3 protein involved in long-distance RNA movement and protection of viral RNA from RNases, including from RNAi, has been shown to accumulate in host cell nucleoli. Functional analysis of mutants revealed the correlation between ORF3 protein nucleolar localisation and its ability to form cytoplasmic ribonucleoprotein particles with plant host snoRNAs and transport viral RNA long distances via the phloem (Kim et al. 2004). Again, given host plant snoRNAs have been redirected for viral purposes it would be expected that a decrease in snoRNA accumulation would be seen in response to virus infection, as was seen in the present study.

The results of the present study show that compared to mock-inoculation, the rRNA component of LMW-RNA decreases in response to CaMV, TMV, TSWV and TYMV. These findings are in accordance with previous studies where a decrease in chloroplast rRNAs have
been noted in response to virus infections (reviewed in Nayudu 2008). For example, TYMV infected *Brassica rapa* *cv. chinensis* shows a large decrease in chloroplast rRNAs in regions of chlorotic tissue compared to dark green islands. A similar change in chloroplast rRNA seen in senescent leaves suggests virus infection hastens senescence. In addition, studies of TMV infected *Nicotiana tabacum* leaves show inhibition of rRNA synthesis in young leaves shortly after inoculation while in older leaves rRNA synthesis was higher immediately after inoculation, decreased significantly during virus replication and increased again following virus accumulation (reviewed in Nayudu 2008).

Analyses of the ratio of sRNA to rRNA as a proportion of averaged mock-inoculation to determine correlation between virus infection and a ratio of sRNA accumulation / rRNA accumulation compared to averaged mock-inoculation of greater than 1.0 determined that this calculation could be a useful predictive tool. The ratio of sRNA to rRNA as a proportion of averaged mock-inoculation was calculated and a 94 % correlation with known virus infection obtained. However, until the ratio of sRNA to rRNA of averaged mock-inoculation is calculated for LMW-RNA subject to other biotic and abiotic stresses, it cannot be known if this ratio is specific to virus infection. This present result compares favourably with other biomarkers used for detecting and predicting outcomes of disease. For example, serum levels of cancer antigen 125 (CA-125) are often elevated in women with ovarian cancer. As such, CA-125 has found application as a tumour marker and is the only marker recommended for clinical use in the diagnosis and management of ovarian cancer (Suh et al. 2010).

However, CA-125 has limited specificity for ovarian cancer because elevated CA-125 levels can be found in individuals without ovarian cancer and may also be elevated in other cancers, including endometrial, fallopian tube, lung, breast, and gastrointestinal cancers. CA-125
may also be elevated in a number of relatively benign conditions, such as endometriosis and during menstruation and pregnancy and in the presence of any inflammatory condition of the abdominal area. Thus, CA-125 testing is not specific for ovarian cancer and often results in false positives. In addition, not every patient with ovarian cancer will have elevated levels of CA-125, where elevated is considered to above 35 U/mL (Nossov et al. 2008). For example, only 79% of all cases of advanced ovarian cancer are positive for elevated CA-125 and only around 50% of patients with early stage ovarian cancer have elevated CA-125 levels. Poor sensitivity can frequently lead to false negatives and these patients are unlikely to seek further treatment for their disease (Gupta and Lis 2010). However, despite low sensitivity and specificity a value above 35 U/mL is an indication that the patient should receive further diagnostic screening (ref (Gupta and Lis 2010; Patsner and Yim 2013). Further, serum values above 65 U/mL in post-menopausal women have shown to > 95% predicative of malignancy (reviewed in Patsner and Yim 2013). These findings indicate that elevated serum levels of CA-125 alone are not enough to predict malignancy to a high degree of probability and that taking two or more factors into consideration (e.g. serum CA-125 and menopausal status) increases probability of accurate assessment.

Similarly, the findings of the present study, where compared to averaged mock-inoculation the rRNA component of LMW-RNA decreased in response to most viruses, were not enough on their own to predict virus presence. Taking two LMW-RNA components into consideration and calculating the ratio of sRNA of rRNA of total LMW-RNA compared to average mock-inoculation resulted in a 94% correlation with known virus presence. Results such as these would provide an indication that the plant in question should receive further diagnostic screening.
While both potyviruses, TuMV and TamMV, investigated in the present study showed increased sRNA and decreased rRNA accumulation, the Bioanalyzer method might prove a useful tool for detection of viruses of the Potyviridae however, generic methods including microarray, antibodies and primers for the detection of Potyviridae are readily available (Wei et al. 2009).

4.4.5 Future directions and concluding remarks

Singh (2012) had the sRNA component of the same 7 and 21 dpi mock-inoculated and virus-inoculated LMW-RNA samples of this study sequenced. This sequencing was undertaken in order to determine which, if any, of accumulated sRNAs were of plant rather than virus origin and as such likely to be responding to virus infection and/or other stresses. Sequencing of samples identified one miRNA, miR156 that showed a significant negative fold change in all five virus-inoculated LMW-RNA compared to mock-inoculated LMW-RNA.

miR156 and miR172 target and regulate a group of plant-specific transcription factors called SQUAMOSA PROMOTER BINDING–LIKE (SPLs) that play important roles in plant phase transition, flower and fruit development and plant architecture. The miR156–SPL-miR172 interaction constitutes an evolutionarily conserved, endogenous cue for vegetative phase transition and flowering in many plant species (Chen et al. 2010; Huijser and Schmid 2011; Jung et al. 2014).

It would be interesting to quantify the accumulation of miR156 (as noted by Singh 2012) and miR172 in response to the five viruses and abiotic stresses of the present study, and to *P. syringae*, in order to determine if quantifying specific miRNAs might be a useful tool for detecting virus presence or absence. This could be achieved with stem-loop qPCR to assess
accumulation across an extensive time-course of very early (hai) infection until necrosis of the host plant (where applicable).

While it can be concluded that the Bioanalyzer method can accurately quantify separate LMW-RNA components, not all viruses could be detected at every time point. However, analyses of the ratio of sRNA to rRNA as a proportion of averaged mock-inoculation demonstrated that calculating the ratio of these two components correlates with known virus infection to a high degree of certainty. This calculation unquestionably warrants further investigation to determine if the response is specific to virus infection.
CHAPTER FIVE – IDENTIFICATION AND VALIDATION OF REFERENCE GENES FOR VIRUS INFECTED ARABIDOPSIS

This chapter comprises an extension of the peer reviewed article published as:


The structure of the published article is altered to ensure consistency between chapters of this thesis. Specifically, sections and subsections have been numbered (e.g. 5.1; 5.2.1) and figures and tables have been placed in the relevant parts of the text. In addition to the article, a preamble has been included to contextualise the research and the results section includes an additional experiment (5.3.7).

Preamble

In the process of developing a robust experimental design for qPCR analysis of target gene transcript accumulation (Chapter 6) the importance of selecting appropriate reference genes (formerly termed ‘housekeeping’ genes) that are stable, particularly in the presence of virus infection, was recognised. It is important to select appropriate reference genes not only to normalise for sample variation but also because transcript accumulation is influenced by abiotic and biotic stress. The selection of reference genes is of particular importance in a viral context because the host cell pathways used for reference genes can be modified by viruses (Watson et al. 2007). Using inappropriate reference genes can, and often does, result in misinterpretation and publication of misleading results. No detailed analysis of reference genes for qPCR of cDNA had been carried out in Arabidopsis plants subjected to diverse viral infections despite Arabidopsis becoming a more commonly used model to study plant–virus interactions. The
research that comprises this chapter addressed this issue and identified reference genes for qPCR
gene transcript accumulation experiments of cDNA from Arabidopsis leaf tissue either
uninfected or infected with one of five viruses, across an infection time course.

5.1 Introduction

Real-time quantitative qPCR using cDNA as template has enabled the rapid detection of
very low quantities of target transcripts and the ability to assess variations in accumulation.
However, with improved sensitivity and throughput come greater demands on experimental
design. In particular, the accurate quantification of transcript abundance obtained by qPCR
analysis is vitally dependent on the use of stably expressed reference genes to normalise for
sample-to-sample variation, variation in RNA integrity, reverse-transcriptase efficiency, and
variation in cDNA sample loading (Bustin et al. 2009).

A number of reference genes were identified for Northern analysis of Arabidopsis gene
expression: Actin2 and -8 (ACT2 and ACT8), β-actin, β-tubulin 2 and 6 (TUB2 and -6),
glyceraldehyde-3-phosphate dehydrogenase (GAPDH), elongation factor 1α (EF1α), and
polyubiquitin 10 (UBQ10). These genes, frequently referred to as “housekeeping” genes, are
required for cell survival and were assumed to undergo little or no variation in expression
(Thellin et al. 1999; Thellin et al. 2009). Unfortunately, the transcript levels of many
traditionally used housekeeping genes are not always stable, with transcript accumulation being
constant under some conditions and highly variable under others (Selvey et al. 2001; Glare et al.
2002; Lee et al. 2002; Czechowski et al. 2005). The ribosomal RNAs 18S and 28S (18S and
28S rRNA) have also been used as references (Xue et al. 2010). However, ribosomal subunit
transcription is affected by both endogenous and exogenous factors, including biotic stress.
Also these RNAs are not polyadenylated and, therefore, cannot be compared with cDNA that is derived from reverse transcribed RNA using oligo-dT primers (Huggett et al. 2005).

Normalisation of data against unstable references can result in incorrect conclusions, and ultimately, compromised studies of transcript accumulation (Czechowski et al. 2005; Huggett et al. 2005; Watson et al. 2007; Gutierrez et al. 2008a; Gutierrez et al. 2008b; Guenin et al. 2009; Bustin et al. 2010). The use of appropriate reference genes is of particular importance when using qPCR of cDNA to measure gene transcript abundance in the context of viral infection. Viruses typically commandeer and, thus, interfere with host cell pathways, the components of which are often encoded by traditional reference genes (Watson et al. 2007). Moreover, different viruses are likely to manipulate different transcription pathways, and the extent to which these pathways are perturbed is likely dependent on the timing of the virus infection and cell type infected. It is unlikely that any cellular gene maintains a constant steady state of transcript accumulation across all conditions and, as such, it is considered nearly impossible to identify a single reference gene for use in qPCR studies (Lee et al. 2002; Andersen et al. 2004; Faccioli et al. 2007; Klatte and Bauer 2009) let alone across a range of virus infections. Accordingly, careful consideration must be given when choosing genes as references for normalisation, and the reliability of a panel of potential genes should be determined for each novel set of experimental conditions. Ideally, at least the best three of these genes, from independent pathways, would be used for normalisation (Vandesompele et al. 2002; Guenin et al. 2009; Klatte and Bauer 2009). To date, five criteria have been outlined that define an appropriate reference gene transcript as one that is (i) stably accumulating across all analysed samples and unaffected by any experimental treatment; (ii) not associated with any pseudogene so as to avoid amplification of genomic DNA; (iii) able, on amplification, to reflect variations in
RNA quality, quantity, or cDNA synthesis; (iv) equivalent in accumulation to that of the target gene transcript; and v) accumulating to moderate amounts (i.e., a threshold cycle \( C_q \) of 15 to 30/35) (Wan et al. 2009).

Validating candidate reference genes presents a circular problem because the stability of a reference gene’s transcript abundance must be derived by comparison with other reference genes. Statistical algorithms for processing raw qPCR data have been developed to identify the best reference genes to use under given experimental conditions. Vandesompele et al. (2002) developed an algorithm, geNorm, which uses pairwise comparisons and geometric averaging across a series of reference genes and biological samples. Briefly, this algorithm is based on the principle that the expression ratio of two ideal reference genes should be identical in all samples regardless of changes in experimental conditions or cell or tissue type. Any variation suggests instability of either one or both candidate genes. The calculated gene stability value (\( M \)) is expressed as the average pair-wise variation of the log-transformed standard deviation of a given gene transcript accumulation ratio compared with the ratio of all other genes investigated. Genes with the lowest \( M \) value have the most stable expression, and stepwise exclusion of genes with the highest \( M \) value enables ranking of genes in order of increasing stability (Czechowski et al. 2005; Guénin et al. 2009; Gutierrez et al. 2008a and b; Vandesompele et al. 2002).

Validating reference genes has received little attention in plant studies despite evidence of the unreliability of conventionally used Arabidopsis housekeeping genes for normalisation. Recently, two groups demonstrated that several Arabidopsis genes are more stably expressed than the traditionally used housekeeping genes (Czechowski et al. 2005; Gutierrez et al. 2008a and b). Another study has highlighted a lack of rigor in plant science publications, where 96.8% of reference genes used in 188 studies were merely putatively stably expressed (Guénin et al.)
A few studies have been carried out to determine reliable reference genes in plant cells across various plant species, developmental stages, and root and shoot abiotic and biotic stresses, including inoculation with *Pseudomonas syringae* and *Phytophthora infestans* (Czechowski et al. 2005; Reid et al. 2006; Exposito-Rodriguez et al. 2008; Gutierrez et al. 2008a; Guenin et al. 2009; Wan et al. 2010). Additionally, reliable reference genes have been determined from mammalian and insect studies (Watson et al. 2007; Xue et al. 2010) and plant genes homologous to these might be considered for validation as reference genes. However, prior to this present study and to the best of our knowledge, no detailed analysis of reference genes for qPCR of cDNA had been carried out in *Arabidopsis* plants subjected to diverse viral infections despite *Arabidopsis* becoming a more commonly used model to study plant–virus interactions (Carr 2007). The aim of the present study was to identify reference genes for qPCR gene expression experiments of cDNA from *Arabidopsis* leaf tissue either uninfected or infected with one of five viruses across an infection time course.

5.2 Materials and methods

5.2.1 Virus isolates, plant material, and inoculation

CaMV in *Brassica campestris* subsp. *pekinesis* (Pak Choy) freeze-dried leaf material, TSWV in fresh chrysanthemum, and TuMV in freeze-dried turnip were kindly gifted by J. Fletcher and R. Lister (Plant & Food Research, Lincoln, New Zealand). TYMV in fresh *B. campestris* cv. *chinensis* and TMV in frozen *Nicotiana tabacum* cv. White Burley were obtained from the School of Biological Sciences, University of Auckland, New Zealand. Upon receipt, the virus isolates were mechanically inoculated to *Arabidopsis*, as described below, as a method for maintaining fresh stocks for inoculation of experimental plants. Virus identity was confirmed by
sequencing of reverse-transcriptase PCR products.

*Arabidopsis thaliana* plants (ecotype Columbia-0) were maintained in a shaded greenhouse set for 20 ± 2°C and a short-day, 8-h photoperiod with 400-W Philips SON-T Agro lights to provide photosynthetically active radiation of 120 to 150 µM m⁻² s⁻². At 4 weeks of age, plants were inoculated with each virus in 0.1 M potassium phosphate (K₂HPO₄) buffer (pH7.4) containing 0.1% (w/v) anhydrous sodium sulfite (Na₂SO₃) and carborundum (as an abrasive). Negative controls were mock-inoculated with buffer only. At 2, 3, 7, and 14 dpi, all of the leaf tissue from three predetermined plants per treatment was harvested and immediately frozen in liquid nitrogen then transferred to −80°C for storage until the RNA was isolated. Virus accumulation at these time points is reported in Chapter 3.5.

The Ministry for Primary Industries (MPI) Post Entry Quarantine (PEQ) laboratory (Morrin Rd, Auckland, NZ) provided fully expanded leaf tissue, of unknown age and immediately frozen in liquid nitrogen, for determining the efficacy of *Arabidopsis EF1α, F-BOX, SAND* and *PDF2* primers in the plant hosts *Chenopodium amaranticolor* (goosefoot, fat hen), *C. quinoa* (quinoa), *Nicotiana benthamiana* (37B), *N. clevelandii*, *N. tabacum*, *N. occidentalis*; *N. sylvestris*, *N. tabacum* White Burley (tobacco varieties), *P. vulgaris* (bean), *Gomphrena globosa* (globe amaranth) and *Cucumis sativum* (cucumber).
5.2.2 RNA isolation, DNase I digest, and cDNA synthesis

Total RNA was isolated from no greater than 100 mg of frozen leaf tissue using an RNeasy mini kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocols. The concentration and purity of the RNA samples was determined using Nanodrop (ND-1000 Spectrophotometer; Nanodrop Technologies Inc., Wilmington, DE, U.S.A.). All samples had an absorbance ratio (absorbance at 260/280 nm) of between 1.8 and 2.2. Following quantification, all RNA samples were normalised to 2.5 µg µl⁻¹. The integrity and intactness of RNA samples was determined by microfluidics assessment with a Bioanalyser 2100 Nano LabChip 6000 (Agilent Technologies, Santa Clara, CA, U.S.A.) and most were found to have RNA integrity numbers of 7.0 or greater. Any gDNA contamination was removed using DNase I (amplification grade; Invitrogen, San Diego, CA, U.S.A.) before 2.5 µg of RNA was reverse transcribed into cDNA using SuperScript III RNase H– reverse transcriptase (Invitrogen) according to the manufacturer’s protocols.

5.2.3 Primer design

To ensure maximum specificity and efficiency during qPCR amplification under a standard set of reaction conditions, primers were designed using the Primer3 software, version 0.1 (S. Rozen and H. J. Skaletsky). Parameters optimized were basic melting temperatures of 60°C, with a maximum difference of ±3°C; length of 20 to 24 bases; GC content of 45 to 50%; and product length of 100 to 150 bp. The PCR product of each primer pair spanned an intron and would be amplified from all the known splice variants (as annotated on TAIR). The AGI locus identifier number, descriptor, primer sequences, primer efficiencies, mean Cq, basic melting temperature, and amplicon size for each candidate reference are presented in Table 5.1.
5.2.4 Testing candidate reference gene primers with standard PCR

A 25-µl reaction volume PCR was employed using cDNA at 2.5 ng µl⁻¹, 12.5 µl of GoTaq Green Master Mix (Promega Corp., Madison, WI, U.S.A.), 2 pmol each gene-specific primer (Table 3), and UltraPure DNase/RNase-free distilled water (Invitrogen) to 25 µl. Amplification of cDNA employed the following conditions: initial 5 min of denaturation at 94°C; followed by 35 cycles of denaturation 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s; followed by a final extension for 5 min at 72°C. Negative control reactions, one omitting reverse transcriptase and one omitting cDNA template, were performed to confirm the absence of genomic DNA or other contaminants.

The presence of single products of appropriate size was confirmed by electrophoresis through a 1.0% agarose gel that was subsequently stained with ethidium bromide and viewed under ultraviolet light with a Gel Doc imager (Bio-Rad, Hercules, CA, U.S.A.).
<table>
<thead>
<tr>
<th>AGI†</th>
<th>Annotation†</th>
<th>5′→ 3′Primer sequences</th>
<th>Primer efficiency</th>
<th>Primer efficiency</th>
<th>Mean Cq (± SE)</th>
<th>Tm (°C)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g49240</td>
<td>ACTIN 8</td>
<td>F-(1078-1097) CCCAAAAGCCAACAGAGAGA R-(1313-1292) CATCACCAGAITCCACCACAAT</td>
<td>1.9 ± 0.04</td>
<td>1.987</td>
<td>24.3 ±0.26</td>
<td>60.40 59.90</td>
<td>140</td>
</tr>
<tr>
<td>At1g27450</td>
<td>APT1</td>
<td>F-(1777-1796) GATTGTTAAATGCTTGGCG R-(2012-1993) AGGATCAACACGATCCACAA</td>
<td>1.7 ± 0.02</td>
<td>1.929</td>
<td>28.3 ±0.32</td>
<td>60.3 59.2</td>
<td>143</td>
</tr>
<tr>
<td>At5g46630</td>
<td>CAC</td>
<td>F-(201-220) ATCGACCTACCGAGATGAC R-(514-493) CGATGTAACATTGCTTGGCG</td>
<td>1.7 ± 0.04</td>
<td>1.927</td>
<td>32.1 ±0.34</td>
<td>60.1 60.7</td>
<td>147</td>
</tr>
<tr>
<td>At5g60390</td>
<td>EF1α</td>
<td>F-(922-941) CACCACTGGAGGTTTTGAGG R-(1158-1140) TGGAGTATTTGGGGGTGGT</td>
<td>2.0 ± 0.00</td>
<td>2.145</td>
<td>20.3 ±0.14</td>
<td>60.5 60.0</td>
<td>137</td>
</tr>
<tr>
<td>At5g15710</td>
<td>F-BOX</td>
<td>F-(1463-1482) GGCTGAGAGGTTCGAGTGTT R-(1602-1583) GGCTGTTGCATGACTGAAGA</td>
<td>2.0 ± 0.01</td>
<td>1.858</td>
<td>29.0 ±0.32</td>
<td>59.5 60.0</td>
<td>140</td>
</tr>
<tr>
<td>At1g58050</td>
<td>HEL</td>
<td>F-(8121-8142) GGGAGAAGGAAGGAGAATTTTG R-(8316-8295) GACATAGTCCCGCGCATAATA</td>
<td>No amplification</td>
<td>No amplification</td>
<td>35+</td>
<td>60.4 60.0</td>
<td>112</td>
</tr>
<tr>
<td>At1g13320</td>
<td>PDF2</td>
<td>F-(3791-3811) TCATTCCGATAGTCGACCAAG R-(4020-4000) TTGATTTGCGAAATACCGAAC</td>
<td>1.8 ± 0.02</td>
<td>1.932</td>
<td>27.4 ±0.17</td>
<td>60.1 60.0</td>
<td>104</td>
</tr>
<tr>
<td>At1g59830</td>
<td>PP2A1</td>
<td>F-(1464-1483) CTGAGCTCGAGAAGCAGCATCA R-(1686-1678) TTCTAGGTGCTGGATCGAATTG</td>
<td>2.0 ± 0.00</td>
<td>2.226</td>
<td>25.6 ±0.22</td>
<td>60.2 60.2</td>
<td>134</td>
</tr>
<tr>
<td>At1g62930</td>
<td>PPR</td>
<td>F-(1846-1866) CATGATGGGAGATTGGAAAAA R-(1967-1947) AAAAGCGATTCACCTCACATA</td>
<td>No amplification</td>
<td>No amplification</td>
<td>2.537</td>
<td>35+</td>
<td>59.7 59.6</td>
</tr>
<tr>
<td>At2g28390</td>
<td>SAND</td>
<td>F-(3376-3396) GTTTGGTCTACACAGAGATTITGT R-(3611-3591) GCTCCTGGACAAGACATCA</td>
<td>2.0 ± 0.01</td>
<td>1.993</td>
<td>25.3±0.11</td>
<td>60.3 60.6</td>
<td>127</td>
</tr>
<tr>
<td>At1g55520</td>
<td>TATA-BP</td>
<td>F-(1922-1942) TTCCAGGAAGGTTCCACATCAA R-(2136-2116) CCAAGGAATATACACCCA</td>
<td>2.0 ± 0.01</td>
<td>1.997</td>
<td>26.4 ±0.17</td>
<td>59.7 60.8</td>
<td>127</td>
</tr>
<tr>
<td>At4g34270</td>
<td>TIP41</td>
<td>F-(1746-1767) TGAATACCTGAGAGAGAAGCA R-(2066-2045) TGCTTAACTCGAGAAGCTCCT</td>
<td>1.4 ± 0.3</td>
<td>No amplification</td>
<td>33.76 ±0.26</td>
<td>60.0 60.4</td>
<td>119</td>
</tr>
<tr>
<td>At1g14790†</td>
<td>RDR1†</td>
<td>F-(2750 - 2769) CGCACAAGACGTCTGAACCA R-(2962-2943) TCGCGATGATCCCTAAACT</td>
<td>2.0 ± 0.01</td>
<td>1.863</td>
<td>28.2 ±0.28</td>
<td>60.4 59.7</td>
<td>133</td>
</tr>
<tr>
<td>At3g49500†</td>
<td>RDR6†</td>
<td>F-(2139-2158) TGCTGACAAGGGAAAACACTG R-(3246-3227) TCTGCTCATGGCTCCATA</td>
<td>1.9 ± 0.01</td>
<td>1.912</td>
<td>27.2 ±0.48</td>
<td>59.9 60.5</td>
<td>145</td>
</tr>
</tbody>
</table>

a Arabidopsis Gene Initiative (AGI) locus identifier number.  
b Primer efficiency, mean ± standard error (SE).  
c Threshold cycle.  
d Melting temperature as reported by Primer3.  
e Candidate reference genes  
† potential genes of interest
5.2.5 qPCR conditions and analysis

qPCR reactions were performed in 384-well plates with a LightCycler 480 Real-Time PCR system using LightCycler 480 SYBR Green I Master (Roche Applied Science, Branchburg, NJ, U.S.A.). Reactions were in 10 µl volumes containing 2 µl of primer pair (5 µM each primer), 3 µl of cDNA (2.5 ng/µl), and 5 µl of SYBR Green I Master mix reagent. A Biomek 3000 Robot (Beckman Coulter, Fullerton, CA, U.S.A.) was used to aliquot all reagents, primers, and samples. Each qPCR reaction was completed in triplicate. Reactions were run using the manufacturer’s recommended cycling parameters, consisting of a 95°C pre-incubation for 5 min and 40 amplification cycles of denaturing at 95°C for 10 s, primer-dependent annealing at 65°C, and elongation at 72°C for 6 s (calculated by dividing largest amplicon length by 25). Florescence values were acquired at the end of the elongation phase. Amplification was followed by a single melting-curve analysis cycle of 95°C for 5 s and 65°C for 1 min with continuous acquisition of fluorescence data at five acquisitions per degree Centigrade at a ramp rate of 0.11°C s⁻¹ to 97°C. Finally, the samples were cooled at 40°C for 10 s.

5.2.6 Primer efficiency and geNorm analysis of gene transcript accumulation stability

Primer efficiency values were calculated with LinRegPCR (11.0) (Ruijter et al. 2009). The raw fluorescent data were baseline corrected and log transformed, and the window-of-linearity determined. Linear regression analysis was then used to fit a straight line through the data points within the window-of-linearity, and the PCR efficiency was calculated from the slope of the line for each individual sample. The mean PCR efficiency was calculated for each primer pair. LightCycler data output was prepared for LinRegPCR using a custom-written script (Python version 2.6.3; Jeremy McRae, Plant & Food Research, Auckland, New Zealand). The efficiency
of amplification from the primers was also calculated for all candidate genes using linear regression parameters from five-point standard curves generated via LightCycler 480 analysis.

To identify the transcript accumulation stability value \( M \), of each candidate reference gene, \( C_q \) values were converted to raw, non-normalised transcript abundance or \( Q \), using the comparative \( C_q (\Delta C_q) \) method and the calculated PCR efficiency, and rescaling the data such that the highest relative accumulation value is 1. Stability of gene transcript accumulation was analysed with geNorm v3.5 analysis software according to instructions in the operator’s manual.

In order to determine whether the choice of reference genes affects the normalisation of genes of interest, we normalised the RNA-dependent RNA polymerases, \( RDR1 \) and \( RDR6 \), to the three most stable candidate reference genes and the two least stable as determined by our geNorm analysis.

### 5.3 Results

#### 5.3.1 Initial selection of candidate reference genes

The candidate reference genes investigated were selected from two sources; first, the most stable gene transcripts published in studies of transcript accumulation in plants undergoing a range of abiotic and biotic stresses (Table 5.2), and second, those genes that had been investigated from any host-virus transcriptome microarray data or had previously been used as reference genes for qPCR of cDNA from virus-infected plants (Table 5.3). Data from microarray experiments were particularly useful because they compare a large number of gene transcripts within a single experiment, thus approaching the ideal of analysing gene expression data for all expressed genes of an organism from as many different organs and experimental conditions as possible (Czechowski et al. 2005). As such, the publicly available microarray datasets of
*Arabidopsis* (Affymetrix ATH1), *Glycine max* (Affymetrix GeneChip), and *Medicago truncatula* (Affymetrix mtr) were a rich source of information for previous authors to select potential qPCR reference genes (Czechowski et al. 2005; Kakar et al. 2008; Libault et al. 2008). Authors of studies investigating the species with no microarray data chose genes from those reported by Czechowski et al. (2005) to validate by qPCR and assess stability of transcript accumulation under their specific conditions (Nicot et al. 2005; Reid et al. 2006; Exposito-Rodriguez et al. 2008; Lee et al. 2008; Artico et al. 2010; Mallona et al. 2010). From this published information, a shortlist of 14 candidate reference genes was selected, including transcripts that were commonly reported as stable by geNorm analysis (Table 5.1).
Table 5.2. geNorm ranking by ($M$) of candidate reference genes from qPCR validated microarray data from literature.

<table>
<thead>
<tr>
<th>Plant species</th>
<th>Validation</th>
<th>Actin8</th>
<th>APT1</th>
<th>CAC</th>
<th>CYTc2</th>
<th>EF1α</th>
<th>F-Box</th>
<th>HEL</th>
<th>PDF2</th>
<th>PP2A1</th>
<th>PPR</th>
<th>SAND</th>
<th>TATA</th>
<th>TIP41</th>
<th>YLS8</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>Affymetrix ATH1, qPCR&lt;sup&gt;a&lt;/sup&gt;</td>
<td>20</td>
<td>5</td>
<td>21</td>
<td>9</td>
<td>17</td>
<td>6</td>
<td>8</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Lycopersicum esculentum</td>
<td>qPCR only&lt;sup&gt;c&lt;/sup&gt;</td>
<td>7</td>
<td>1=</td>
<td>10</td>
<td>4</td>
<td>3</td>
<td>1=</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td>1=</td>
<td></td>
<td></td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitis vinifera</td>
<td>qPCR only&lt;sup&gt;f&lt;/sup&gt;</td>
<td>8</td>
<td>1=</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>1=</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>1=</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gossypium hirsutum</td>
<td>qPCR only&lt;sup&gt;g&lt;/sup&gt;</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Petunia hybrida</td>
<td>qPCR only&lt;sup&gt;h&lt;/sup&gt;</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Glycine max</td>
<td>Affymetrix GeneChip, qPCR&lt;sup&gt;i&lt;/sup&gt;</td>
<td>13</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lolium perenne</td>
<td>SAGE tags and qPCR&lt;sup&gt;i&lt;/sup&gt;</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Solanum tuberosum</td>
<td>qPCR only&lt;sup&gt;k&lt;/sup&gt;</td>
<td>8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medicago truncatula</td>
<td>Affymetrix MTR &amp; qPCR&lt;sup&gt;l&lt;/sup&gt;</td>
<td>7</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> RDR1 and RDR6 are not considered candidate reference genes and are omitted from this table.

<sup>b</sup> Total number of genes validated by qPCR in the referenced publication.

<sup>c</sup> Microarray or qPCR validation. Array incorporated complementary DNA (cDNA) from development and shoot, root, hormone and nutrient stress samples; and qPCR assays incorporated cDNA from development and steady state leaf, shoot, root and floral tissues.

<sup>d</sup> Czechowski et al. (2008).

<sup>e</sup> Expósito-Rodriguez et al. (2008).

<sup>f</sup> Reid et al. (2006).

<sup>g</sup> Artico et al. (2010).

<sup>h</sup> Libault et al. (2008).

<sup>i</sup> Lee et al. (2008). SAGE = serial analysis of gene expression.

<sup>k</sup> Nicot et al. (2005).

<sup>l</sup> Kakar et al. (2008).
5.3.2 Exclusion of genes affected by viral infection

To determine whether each candidate reference gene had been reported to show transcript changes in the presence of virus infection, published virus transcriptome microarray data were reviewed for changes in their transcript accumulation (Whitham et al. 2003; Yang et al. 2007; Ascencio-Ibáñez et al. 2008; Babu et al. 2008; Alfenas-Zerbini et al. 2009; Baebler et al. 2009) (Table 5.3). Genes reported for human studies of virus infection, including human immunodeficiency virus and herpes viruses, were also considered where orthologous genes were present in Arabidopsis (Watson et al. 2007) (Table 5.3). Actin8 and EF1α are traditional housekeeping genes used as reference transcripts for qPCR analyses and were included as candidate reference genes despite evidence of increased transcript accumulation under Plum pox virus (PPV) infection compared with mock infection of Arabidopsis (threefold and fivefold, respectively) (Babu et al. 2008). Cytochrome C2 (CYTc-2) was excluded from this study because it showed a very large variation (threefold decrease to a fourfold increase) in response to seven different plant viruses as assessed by Arabidopsis Affymetrix microarrays (Ascencio-Ibáñez et al. 2008; Babu et al. 2008; Whitham et al. 2003) and in response to one virus in tomato (Alfenas-Zerbini et al. 2009). Yellow stripe like 8 (YSL8) was excluded, albeit due to a small flux (0.7-fold increase) in transcript accumulation in response to Cabbage leaf curl virus (CaLCuV) as assessed by microarray analysis of virus-infected Arabidopsis plants (Ascencio-Ibáñez et al. 2008). Omitting CYTc-2 and YLS8 resulted in a shortlist of 12 candidate reference genes.
Table 5.3. Fold change of candidate reference genes following virus infection

<table>
<thead>
<tr>
<th>Host species, virus</th>
<th>Actin8</th>
<th>APT1</th>
<th>CYTc-2</th>
<th>EF-1α</th>
<th>TATA</th>
<th>YLS8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arabidopsis thaliana</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Turnip vein-clearing virus</em></td>
<td></td>
<td></td>
<td></td>
<td>2.5 - 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Oilseed rape mosaic virus</em></td>
<td></td>
<td></td>
<td></td>
<td>2.5 - 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Potato virus X</em></td>
<td></td>
<td></td>
<td></td>
<td>2.5 - 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Cucumber mosaic virus</em></td>
<td></td>
<td></td>
<td></td>
<td>2.5 - 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turnip mosaic virus (TuMV)</td>
<td></td>
<td></td>
<td></td>
<td>2.5 - 4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TuMV</td>
<td>Ref</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plumpox virus</td>
<td>2.97</td>
<td>Ref</td>
<td>-3.3</td>
<td>4.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cabbage leaf curl virus</td>
<td>0.27</td>
<td></td>
<td></td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lycopersicon esculentum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PepYMV</td>
<td>Ref</td>
<td></td>
<td>2.00E-26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Solanum tuberosum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Potato virus Y(NTN)</em></td>
<td></td>
<td></td>
<td></td>
<td>Ref</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Homo sapiens</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human immunodeficiency virus</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Herpes simplex 1</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varicella zoster</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cytomegalovirus</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

= genes that were used as a reference gene in the given study. Candidate reference genes (*CAC, F-BOX, HEL, PDF2, PP2A1, PPR, SAND and TIP41*) were omitted from this table as no data were available.

\( ^{a} \) Ref

\( ^{b} \) Whitham et al. (2003).
\( ^{c} \) Yang et al. (2007).
\( ^{d} \) Babu et al. (2008).
\( ^{e} \) Ascencio-Ibáñez et al. (2008).
\( ^{f} \) Alfenas-Zerbini et al. (2009).
\( ^{g} \) Baebler et al. (2009).
\( ^{h} \) Watson et al. (2007).
5.3.3 Establishing cDNA and primer pairs for qPCR

Only high-quality RNA that met the following criteria was used in further experiments: i) free of gDNA contamination as assessed by PCR, ii) pure as assessed by light absorbance (ratio of 1.8 to 2.0 at 260 or 280 nm) (Nanodrop), and iii) intact as determined by distinct ribosomal banding visualised on a Nano LabChip 6000.

cDNA was synthesised from each RNA sample from each plant in triplicate and these were tested for integrity and absence of gDNA, then each of the cDNA samples of the triplicate were pooled in equimolar proportions (Figure 5.1). There were three biological replicates for each of the 28 treatments.

![Flow diagram of method for cDNA synthesis, testing for integrity and absence of gDNA.](image)

RNA was isolated from one plant then cDNA generated in triplicate. After found to be of high integrity and free of gDNA, cDNA triplicates were pooled in equimolar proportions and process repeated for each of the three biological replicates.
A 10-fold dilution series (from 250 ng to 25 pg per qPCR reaction) of each cDNA was prepared to determine the optimum cDNA concentration for subsequent qPCR analyses. Using cDNA at 250 and 25 ng/qPCR resulted in poor amplification of each PCR product, whereas 2.5 ng/qPCR resulted in consistent $C_q$ values in the range of 18 to 35. Therefore, 2.5 ng was selected as the optimal amount of cDNA per reaction.

Primer efficiency values for all candidate genes were calculated with LinRegPCR software and compared with efficiency values calculated using linear regression parameters from five-point standard curves generated via LightCycler 480 analysis (Table 5.1). Amplification efficiencies were similar from both methods but LinRegPCR offers the benefit of enabling primer efficiencies to be determined within each run; therefore, LinRegPCR primer efficiencies were used for this study. Amplification efficiencies between 1.7 (70% efficiency) and 2.0 (100% efficiency), which is considered acceptable (Pfaffl et al. 2004; Vandesompele et al. 2002), were achieved using 10 of the primer pairs under our PCR conditions.

The transcript accumulation, as determined by the number of cycles needed for amplification related fluorescence to reach a maximal point of a second derivative curve ($C_q$), was assessed for the candidate reference genes using cDNA template prepared from each of the treatments of *Arabidopsis* plants. Mean and standard error $C_q$ values for each gene amplified from each of the three biological replicates per leaf sample are given in Table 5.1. *EF1a* was the most abundant reference gene transcript, with amplification-related fluorescence reaching a maximum at cycle $20.3 \pm 0.14$. In contrast, TIP41-like family protein (*TIP41*) and Clathrin adaptor complex (*CAC*) were the least abundant transcripts, reaching maximum amplification fluorescence at cycles 32 to 34. Given that the qPCR protocol of this study used 40 cycles of amplification, $C_q$ values of 35 were considered late to amplify (late call) and as such, unreliable. Helicase (*HEL*), *YSL8*, and pentatricopeptide repeat-containing protein (*PPR*)
consistently resulted in $C_q$ values greater than 35 (Table 5.1) and were not included for further validation. *TIP41* was initially included for further validation because of its ranking as a stable gene transcript under developmental, shoot, root, hormone, and nutrient stress conditions (Table 5.1.) (Czechowski et al. 2005; Expósito-Rodríguez et al. 2008; Reid et al. 2006) but it had a low average primer efficiency (1.4 ± 0.3) and a relatively high $C_q$ (33.76 ± 0.26).

Furthermore, calculation of quantity (Q) values in preparation for geNorm analysis demonstrated that *TIP41* consistently gave very low Q values (0.002). Therefore, *TIP41* was removed from further analysis, resulting in a final list of nine candidate reference genes. The nine candidate reference genes and two genes used for comparison, RNA-dependent RNA polymerase 1 (*RDR1*) and *RDR6*, are listed by their *Arabidopsis* Gene Initiative (AGI) locus identifier number, gene descriptors, and function in Table 5.4.
Table 5.4. The nine candidate reference genes and two genes used for comparison (*RDR1* and *RDR6*). AGI locus identifier number, descriptor and function of genes utilised in this study from information publically available from The *Arabidopsis* Information resource (TAIR).

<table>
<thead>
<tr>
<th>AGI locus identifier</th>
<th>Descriptor</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g49240</td>
<td><em>ACTIN 8</em></td>
<td>Member of a subclass of actins composed of Actin 2 and Actin 8. Component of the cytoskeleton. Involved in cell motility and signalling.</td>
</tr>
<tr>
<td>At1g27450</td>
<td><em>APT1</em>: Adenine phosphoribosyl transferase 1</td>
<td>Involved in the one-step salvage of adenine to AMP</td>
</tr>
<tr>
<td>At5g46630</td>
<td><em>CAC</em>: Clathrin adaptor complex</td>
<td>Medium subunit family protein Functions in protein binding and intracellular protein transport.</td>
</tr>
<tr>
<td>At5g60390</td>
<td><em>EF1a</em>: Elongation factor 1-alpha</td>
<td>Involved in calmodulin binding. Translation elongation factor activity.</td>
</tr>
<tr>
<td>At5g15710</td>
<td><em>F-BOX</em>: F-box family protein</td>
<td>Molecular functions and biological processes unknown.</td>
</tr>
<tr>
<td>At4g04890</td>
<td><em>PDF2</em>: Protodermal factor 2</td>
<td>Encodes a 65-kD regulatory subunit of Serine/Threonine protein phosphatase 2A (PP2A). Also known as Protein phosphatase type 2A regulator subunit 3 (PP2AA3).</td>
</tr>
<tr>
<td>At1g13320</td>
<td><em>PP2A3</em>: Protein phosphatase 2A subunit 1</td>
<td>Encodes one of two 65 kDa isoforms of the catalytic subunit of protein phosphatase 2A.</td>
</tr>
<tr>
<td>At2g28390</td>
<td><em>SAND</em>: Sand family protein</td>
<td>Membrane protein with putative role in vesicle traffic and late endocytosis (<em>S. cerevisiae; C. elegans</em>).</td>
</tr>
<tr>
<td>At1g55520</td>
<td><em>TATA</em>: TATA-box binding protein</td>
<td>Required for basal transcription. Facilitates recruitment of TFIID to the promoter. Forms pre-initiation complex with RNA polymerase.</td>
</tr>
<tr>
<td>At1g14790</td>
<td><em>RDR1</em>: RNA-dependent RNA polymerase 1</td>
<td>Promotes turnover of viral RNAs in infected plants. Amplifies RNAi. Involved in the production of <em>Cucumber mosaic virus</em> siRNAs.</td>
</tr>
<tr>
<td>At3g49500</td>
<td><em>RDR6</em>: RNA-dependent RNA polymerase 6</td>
<td>Involved in trans-acting siRNA and other siRNA biogenesis. Required for post-transcriptional gene silencing and natural virus resistance.</td>
</tr>
</tbody>
</table>
5.3.4 Average transcript accumulation stability ($M$) as calculated by geNorm

Q values of the remaining nine candidate reference genes and $RDR1$ and $RDR6$, for all three biological replicates, were analysed using geNorm to generate a stability measure ($M$). Genes with the lowest stability of transcript accumulation, highest $M$, were excluded in a stepwise manner until the two most stably expressed genes remained. This process enabled ranking of genes in order of increasing transcript stability across the mock inoculation and virus infection time course (Fig. 5.2). Many of the candidate reference gene transcripts ($Actin8$, $APT1$, protein phosphatase 2A subunit 1 $PP2A1$, $CAC$, and TATA-box binding protein $TATA$) had low stability, with $M$ values of 4.30 to 2.54, above the default limit of 1.5 suggested by geNorm. $Actin8$ and $APT1$ were the first excluded as the least stable genes ($M$ values of 4.30 and 3.20, respectively). Protodermal factor 2 ($PDF2$), sand family protein ($SAND$), $EF1a$, and an Arabidopsis F-box family protein ($F-BOX$) were identified as the most stable genes, with $M$ values of 1.46, 1.32, 1.27, and 1.10, respectively. The $RDR1$ gene transcript was less stable ($M = 3.36$) than all candidate reference genes, with the exception of $Actin8$, while $RDR6$ was less stable than all other genes ($M = 4.82$).
Figure 5.2. Ranking of stability and transcript accumulation of candidate reference genes.

geNorm ranking of nine candidate reference genes based on quantitative polymerase chain reaction data. Four technical replicates of three biological replicates of Arabidopsis leaf tissue across treatment inoculations: mock inoculation, mock inoculated without Carborundum (no carb), Cauliflower mosaic virus (CaMV), Tomato spotted wilt virus (TSWV), Turnip mosaic virus (TuMV), Turnip yellow mosaic virus (TYMV), and Tobacco mosaic virus (TMV) across 2, 3, 7, and 14 days post inoculation (dpi) were included in the analysis. The black line at 1.5 represents the \( M<1.5 \), the default cut off value suggested by geNorm where \( M \) values above are considered unstable.

Average transcript stability and the range of \( M \) values of each gene across the time course (2, 3, 7, and 14 days post inoculation [dpi]) of inoculations are shown in Figure 5.3. Among the candidate reference genes, SAND has the most stable transcript accumulation across the time course, followed by PDF2, CAC, and F-BOX. In contrast, PP2A1 has the least stable transcript accumulation whereas EF1α, which is very stable across all treatments and time points, has only moderate stability across the time course of treatments.
Average transcript accumulation stability ($\Delta M$) across the time course of all treatments. The four candidate reference genes that were identified as most stable in Figure 5.1 are indicated by asterisks.

Average transcript stability, $M$, and the range of $M$ values of each gene across all treatments (mock, Cauliflower mosaic virus [CaMV], Tomato spotted wilt virus [TSWV], Turnip mosaic virus [TuMV], Turnip yellow mosaic virus [TYMV], and Tobacco mosaic virus [TMV] inoculation) are shown in Figure 5.4. PP2A1 had the most stable transcript accumulation in response to each treatment, followed by SAND and PDF2. Actin8 and APTI had the least stable transcript accumulation in response to each treatment ($\Delta M = 5.04$ and 3.10, respectively).
5.3.5 Quantity (Q) of transcript accumulation

The raw value of rate of transcript accumulation (Q) changes for each gene, depending on the stress the plant incurred, such that no gene transcript accumulated equally across all times and treatments. In order to determine how all the genes accumulate over the time course of all treatments; average Q values were plotted. Stable accumulation of a gene transcript across a time course would theoretically result in a straight and level line with minimal error. Average Q and the range of Q values and range of standard error for each gene are presented in Figure 5.5A. The results demonstrate that CAC, PDF2, TATA, and SAND have lines that are straight compared with Actin8, while Actin8 and APT1 are two of the least stably accumulating transcripts across time, with a range of ΔQ of 0.04 and 0.10, respectively, and both have a large range of error (0.04 and 0.03, respectively). Of the four most stable gene transcripts, as determined by M, PDF2 has the most stable accumulation (ΔQ = 0.02) and the smallest error (0.02) while EF1α is the least stable in transcript accumulation (ΔQ = 0.05) and has the largest range of error (0.04).
Accumulation stability of average quantity (Q)

A.  

- Actin8: $\Delta Q = 0.04$
- APT1: $\Delta Q = 0.10$ (0.03)
- PP2A1: $\Delta Q = 0.03$ (0.03)
- CAC: $\Delta Q = 0.02$ (0.02)
- TATA: $\Delta Q = 0.03$ (0.02)
- PDF2*: $\Delta Q = 0.02$ (0.02)
- SAND*: $\Delta Q = 0.03$ (0.02)
- EF1α*: $\Delta Q = 0.05$ (0.04)
- F-BOX*: $\Delta Q = 0.04$ (0.03)
- RDR1
- RDR8

Days post inoculation
Figures 5.5A & B. Relative accumulation stability of candidate reference genes.

Transcript accumulation of each candidate reference gene and RDR1 and RDR6, across A, the time course or B, all treatments. Error bars are standard error of mean quantities (Q) of three biological replicates (A, n = 12; B, n = 18). For each gene, the range of Q values is given at the top of the graph and the range of standard error is shown in brackets. The four most stable genes are identified by asterisks.
A range of stable accumulation of a gene transcript across treatments would theoretically result in a plot with similar data points and error for each treatment. Average Q and the range of Q values and range of standard error for each gene are presented in Figure 5.4B. The results demonstrate that RDR1 is the most stable in accumulation across all treatments, with the smallest range of standard error, while EF1α, APT1, and Actin8 have the most erratic accumulation.

5.3.6 Choice of reference genes affects normalisation

The accumulation of RDR1 and RDR6 transcripts were used as examples to show the effect of using different normalisation genes to estimate relative accumulation. The transcript quantities of RDR1 and RDR6 were determined by qPCR in the same samples used for the selection of reference genes. Average Q for RDR1 and RDR6, ranging from 0.03 to 1.0 and 0.005 to 1.0, respectively, across time and treatment prior to normalisation, are shown in Figure 5.6A and D. When transcript quantities of these two genes are normalised with the normalisation factor generated by geNorm for the three most stable genes (F- BOX, EF1α, and SAND), the range for RDR1 was between 0.16 and 1.0 and the standard error increased (Fig. 5.6B) compared with raw data while the range for RDR6 increased to between 0.01 and 1.0 (Fig. 5.6E) and the error remained similar to raw data (Fig. 5.6D). Interestingly, RDR6 reacted very differently to inoculation with CaMV than any other virus inoculation. When transcript quantities were normalised instead with the normalisation factor generated for the two least stable candidate reference genes (Actin8 and APT1), RDR1 transcript accumulation was overestimated further to between 0.2 and 1.0 and the reported error was greater in most cases (Fig. 5.6C). For RDR6, transcript quantity remains in a similar range (0.01 to 1.0) but the error was underestimated, most notably for all treatments at 14 dpi and TSWV at 2 dpi (Fig. 5.6F).
Figure 5.6. Comparison of raw and normalized accumulation of RDR1 and RDR6 gene transcripts.

Transcript accumulation across time course of all treatments (days post inoculation [dpi]) of A through C, RDR1 and D through F, RDR6 following A and D, no normalisation; B and E, normalisation to the three most stable references (F-BOX, EF1α, and SAND); or C and F, normalised to the two least stable references (Actin8 and APT1), using the normalisation factor generated by geNorm. Data were rescaled to 1.0, where 1.0 was the largest quantity (Q) value calculated for each gene transcript across all treatments.
5.3.7 Amplification of Arabidopsis EF1α, F-BOX, SAND and PDF2 from alternative herbaceous indicator host plants

The ability to amplify products of the correct band size was investigated for Nicotiana benthamiana (37B), N. clevelandii, N. tabacum, N. occidentalis; N. sylvestris, N. tabacum White Burley (tobacco varieties), Chenopodium amaranticolor (goosefoot, fat hen), C. quinoa (quinoa), P. vulgaris (bean), Gomphrena globosa (globe amaranth) and Cucumis sativum (cucumber) plant species. The efficacy of Arabidopsis reference gene primers; EF1α, F-BOX, SAND and PDF2, in alternative plant hosts were determined by RT-PCR and RT-PCR products were resolved on 20% TBE Gel Novex® TBE Gels (Invitrogen™, Life Technologies, Carlsbad, CA, USA) in order to obtain good separation and resolution of amplicon fragments into sharp bands.

Of the four Arabidopsis reference genes; EF1α, F-BOX, SAND and PDF2, only EF1α amplified products with the correct band size for all plant species investigated: i.e. N. benthamiana, N. clevelandii, N. occidentalis, N. sylvestris, C. amaranticolor, C. quinoa, P. vulgaris, G. globosa and C. sativum (Figures 5.7 and 5.8A and B; Lanes 2, 6 and 10).

Arabidopsis F-BOX amplified a product of correct size for all Nicotiana species investigated (Figure 5.7A; Lanes 3, 7 and 11 and 5.7; Lanes 3 and 7). While, Arabidopsis F-BOX and SAND amplified products of the correct size for all other plant species investigated, secondary non-specific products were also amplified (Figures 5.7. and 5.8A; Lanes 4, 5, 8, 9, 12 and 13 and Figures 5.7 and 5.8B; Lanes 4, 5, 8 and 9). PDF2 failed to amplify from N. benthamiana, N. occidentalis and C. sativum (Figures 5.7A, Lanes 5 and 13 and A, Lane 13).
Figure 5.7. TBE Gel electrophoresis of four most stable reference genes (EF1α 137 bp, F-BOX 140 bp, SAND 127 bp and PDF2 104 bp) and ability to amplify from the Nicotiana species; N. benthamiana (Nb), N. clevelandii (Nc), N. occidentalis (No), N. sylvestris (Ns) and N. tabacum (Nt) compared to Arabidopsis (At).

A. 1 and 15, 1Kb plus ladder. 2, 3, 4 and 5 Nb – EF1α, F-BOX, SAND and PDF2 respectively; 6, 7, 8 and 9 Nc - EF1α, F-BOX, SAND and PDF2; 10, 11, 12 and 13 No - EF1α, F-BOX, SAND and PDF2; 14; RT-deficient reaction (RT-ve; EF1α). B. 1 and 15, 1Kb plus ladder. 2, 3, 4 and 5 Ns - EF1α, F-BOX, SAND and PDF2; 6, Nt - EF1α, F-BOX, SAND, PDF2; 10, 11, 12 and 13 At – EF1α, F-BOX, SAND and PDF2. 14, At cDNA deficient reaction (cDNA –ve; EF1α).
Figure 5.8. TBE Gel electrophoresis of four most stable reference genes (EF1α 137 bp, F-BOX 140 bp, SAND 127 bp and PDF2 104 bp) and ability to amplify from Chenopodium amaranticolor (Ca), C. quinoa (Cq), P. vulgaris (Pv), G. globosa (Gg) and Cucumis sativum (Cs) compared to Arabidopsis (At).

A. Lanes 1 and 15, 1Kb plus ladder. 2, 3, 4 and 5, Ca – EF1α, F-BOX, SAND and PDF2 respectively. 6, 7, 8 and 9, Cq – EF1α, F-BOX, SAND and PDF2. 10, 11, 12 and 13, Pv - EF1α, F-BOX, SAND and PDF2. 14, blank. B. Lanes 1 and 15, 1Kb plus ladder. 2, 3, 4 and 5, Gg – EF1α, F-BOX, SAND and PDF2. 6, 7, 8 and 9 Cs - EF1α, F-BOX, SAND and PDF2. 10, 11, 12 and 13, At – EF1α, F-BOX, SAND and PDF2 (104 bp). 14, blank.
5.4 Discussion

For qPCR to be accurate, specific conditions and appropriate internal reference genes for normalisation must be determined. A reference gene should show minimal change, whereas a gene of interest may change greatly over the course of an experiment. Normalisation with multiple reference genes is becoming the minimum standard. Furthermore, a “minimum standard for the provision of information for qPCR experiments” (MIQE) has been proposed (Bustin 2010; Bustin et al. 2009). This minimum standard will encourage detailed auditing of the entire qPCR workflow from experimental design through technical details such as nucleic acid extraction, reverse transcription, qPCR target information, oligonucleotides, PCR cycling protocol, validation of appropriate reference genes, and data analysis. Bustin (2010) and Pfaffl (2010) propose that a current lack of rigorous standards of practice for qPCR has resulted in misinterpretation of data and publication of erroneous conclusions. By implementing the MIQE guidelines, qPCR can become a robust, accurate, and reliable method for reporting nucleic acid quantification. This is important if confidence in published results and conclusions is to be maintained.

Here, we evaluated the stability of accumulation of nine candidate reference genes, including two traditional housekeeping genes, in Arabidopsis subject to five diverse plant viral infections. DNA and RNA viruses with diverse replication, gene expression, and movement strategies were used. This diversity lends itself to a study of gene transcript accumulation in Arabidopsis in response to virus infection (Whitham et al. 2003).

Our findings show that Actin8, one of the most commonly used reference genes, showed the highest variability in transcript accumulation. These results caution against the use of Actin8 as a reference gene in virus-infected Arabidopsis. Similar results have been observed and reported from validation of reference genes in Arabidopsis and other plant species subject to various other stresses. For example, in Arabidopsis qPCR validated data,
Czechowski et al. 2005 ranked Actin8 20th for stability of 21 genes investigated. Nicot et al. (2005) ranked Actin8 as least stable of eight genes investigated in potato exposed to a biotic (late blight) and two abiotic (cold and salt) stresses. Similarly, Libault et al. (2008) ranked an actin ortholog of the Arabidopsis actin family as 13th most stable of 15 genes studied in soya bean subject to 130 different conditions. Mallona et al. (2010) ranked an Actin of petunia eighth for stability of nine genes investigated. Our results agree with these earlier findings where Actin8 shows poor transcript stability. Some studies have reported various actins with high transcript stability, suggesting that Actin8 might be considered a suitable reference gene given specific experimental conditions (Artico et al. 2010; Gutha et al. 2010; Wan et al. 2010). However, recent studies have revealed that the actins do not satisfy basic requirements for application as reference genes (Nicot et al. 2005; Gutierrez et al. 2008a; Thellin et al. 2009). This may be explained in part by the fact that actins, major components of cytoplasmic microfilaments in eukaryotic cells, participate in other cellular functions in addition to the provision of structural support and cellular shape (Sturzenbaum and Kille 2001).

PDF2, SAND, EF1α, and F-BOX act in distinct pathways and can be considered appropriate reference genes for normalisation given our experimental conditions for virus infection in Arabidopsis. These genes were stable across the experimental conditions and provided a range of transcript accumulation for comparison with genes of interest. The primer pairs used in this study to target F-BOX, EF1α, SAND, and PDF2 have high identity to genes of other species, especially Brassica and Nicotiana spp. and Vitis vinifera. Therefore, although not demonstrated in this study, they are likely candidates for use in other plant species.

Of our four most stably accumulating transcripts, PDF2 was the fourth, a finding similar to Czechowski et al. (2005) where PDF2 was ranked sixth of 21 transcripts of
Arabidopsis \((M < 0.75)\). PDF2 encodes a 65-kDa regulatory subunit of serine or threonine protein phosphatase and functions to maintain the identity of meristem L1 layer cells, and overexpression can delay flowering (Abe et al. 2003).

Our results show SAND to be the third most stable transcript, similar to the findings of Czechowski et al. (2005), where SAND ranked first of 21 Arabidopsis transcripts \((M < 0.75)\), and Remans et al. (2008), where SAND, in combination with F-BOX and YLS8, was identified as one of the best candidates for normalisation of data from Arabidopsis tissue exposed to cadmium (Cd) and copper (Cu) \((M < 0.4)\). Mallona et al. (2010) ranked SAND second most stable of nine transcripts in petunia \((M = 0.5)\). The encoded protein is thought to have a function in vesicle trafficking and late endocytosis as determined from Saccharomyces cerevisiae and Caenorhabditis elegans gene orthologs (Poteryaev and Spang 2005). Given that SAND plays a role in vesicle traffic, it may not be a suitable reference gene in studies of other pathogens where vesicle traffic may play an important role in defence.

EF1α is frequently used as a reference gene and was one of the four most stably accumulating transcripts in this study. Our finding that EF1α was stable in transcript accumulation is consistent with results from some other groups. For example, Nicot et al. (2005) ranked EF1α as the most stable of 8 genes \((M < 0.3)\) and Gutierrez et al. (2008a) ranked EF1α second most stable of 14 genes studied in Arabidopsis \((M < 0.5)\). Wan et al. (2009) ranked EF1α second most stable of 10 transcripts studied in cucumber \((M = 0.8)\).

However, Czechowski et al. (2005) reported that EF1α was the least stable of 21 transcripts of Arabidopsis \((M > 1.5)\) and Expósito-Rodríguez et al. (2008) ranked EF1α 10th of 11 transcripts of tomato. However, it should be noted that EF1α might not be a suitable reference gene in the context of all known plant virus infections. For example, EF1α has been shown to have a 4.92-fold change in response to PPV infection (Babu et al. 2008), is one of 226 suppressed genes in response to TuMV (Yang et al. 2007), and shows variability in
response to five viruses in tomato (Mascia et al. 2010). The $EF1\alpha$ protein is involved in calmodulin binding and has translation elongation factor activity (The Arabidopsis Information Resource [TAIR]).

$F$-BOX was identified as the most stable transcript in this study. This is consistent again with Remans et al. (2008), where $F$-BOX and SAND, along with $YLS8$, were identified as the best candidates for data normalisation ($M < 0.3$). Our results are also consistent with the findings of Libault et al. (2008), where $F$-BOX was ranked most stable of 15 transcripts tested from soya bean ($M < 0.3$). Our results are in contrast, however, with those of Artico et al. (2010), who ranked $F$-BOX seventh of nine ($M > 0.8$). The molecular functions and biological processes of this particular $F$-BOX protein in Arabidopsis are currently unknown (TAIR).

Recently, Mascia et al. (2010) examined the stability of eight candidate reference genes in tomato subjected to four different plant viruses at 9 dpi and one plant virus at 30 dpi. High variability was demonstrated for two widely used reference genes, 18S rRNA and $EF1\alpha$. However, three commonly used reference genes—Actin 3 ($ACT3$), GAPDH, and ubiquitin 3 ($UBQ3$)—showed high levels of stability throughout the infected tissues. Of the two novel candidate reference genes evaluated, uridylyl kinase (UK) did not respond to virus or viroid infection. Therefore, UK might be another appropriate candidate reference gene for trial as a transcript for normalisation across diverse plant–virus interactions.

$RDR1$ and $RDR6$ were included for comparison to reference genes. Arabidopsis requires host $RDR1$ and $RDR6$ to produce vsiRNAs. Wang et al. (2010) found that both $RDR1$ and $RDR6$ were involved in the production of Cucumber mosaic virus siRNAs and in the amplification of RNAi in infected plants. Given the putative roles of $RDR1$ and $RDR6$ in the course of plant viral infection, unstable $M$ values were expected in our study. Accordingly, our findings show that the stability of $RDR1$ falls between the two least stable
candidate reference genes, *Actin8* and *APT1*, whereas *RDR6* is the least stable of all genes considered.

When separating the data across the time course of treatments, our results for candidate reference gene stability, as measured by *M*, show that *SAND*, *PDF2*, and *F-BOX* are the best combination of three reference genes for normalisation. When separating out data across inoculation treatments, our results for candidate reference gene stability, as measured by *M*, show that *PP2A1*, *SAND*, and *PDF2* are the best combination for normalisation. A difference in ranking of appropriate sets of reference genes to use for normalisation of different datasets is not unexpected, given that no single gene or set of genes demonstrated constant transcript accumulation for all experimental conditions (Vandesompele et al. 2002; Czechowski et al. 2005; Remans et al. 2008).

Rather than investigating stability, by separating out measures of raw accumulation, or *Q*, across the time course of treatments, we demonstrated that *Actin8* and *APT1* were two of the least stable in transcript accumulation across time. *PDF2* had the most stable *Q* value across time. *PDF2* had one of the most stable *Q* values across the inoculations, while *Actin8* and *APT1* were again among the least stable in *Q* value across our treatments. *F-BOX* and *SAND* were moderately stable in transcript accumulation. These results serve to emphasise the unreliability of *Actin8* and *APT1* in our study.

Variation in the stability and transcript accumulation of reference genes between sample groups reduces the sensitivity of an assay to detect changes in transcript accumulation of target genes of interest and may give the appearance of false changes. To demonstrate the effect of using appropriate versus inappropriate reference genes to normalise data, we normalised our potential genes of interest, *RDR1* and *RDR6*, to the normalisation factor generated for the three most stable and to the two least stable candidate reference genes. Our results demonstrate a trend toward differences in reported transcript accumulation depending
on which set of genes was used to generate the normalisation factor. Interestingly, *RDR6* reacts very differently to inoculation with CaMV at every time point. This might be a function of CaMV being a double-stranded DNA virus and warrants further investigation. Our findings emphasise that, if a chosen reference gene has a large transcript fluctuation, normalisation will lead to erroneous transcript accumulation profiles of target genes of interest.

Our analyses have shown that different experimental conditions tested require different sets of reference genes. These results are consistent with other groups (Vandesompele et al. 2002; Czechowski et al. 2005; Thellin et al. 2009) and further emphasise the importance of validating reference genes for each experimental condition (for example, different tissues or organs) or for plant viruses that are phloem limited. Recently, a study investigated six putative stably expressed genes in qPCR of cDNA from phloem-limited *Grapevine leafroll associated virus* 3-infected grape leaves (Gutha et al. 2010). A combination of two genes, *Actin* and *NAD5*, was identified as suitable candidates for normalisation of data in both virus-free and virus-infected leaves. A further candidate reference gene might include sucrose/proton symporter gene 2 or phloem-specific reporter genes in transgenic plants (Ckurshumova et al. 2009). Changes in transcript accumulation within parenchyma cells adjacent to infected cells could be referenced to the gene transcripts identified in this study.

Nair (2013) aimed to identify appropriate reference genes for use in virus infection studies of *Colocasia esculenta* (taro). The four most stably accumulating genes of the present thesis chapter namely; *EF1a, F-BOX, SAND* and *PDF2* were assessed for stability in healthy *C. esculenta* and *N. benthamiana* using PCR. Various parameters including annealing temperature were tested using gradient PCR in order to optimise conditions for efficient amplification in *C. esculenta* and *N. benthamiana*. However, efficient amplification could not
be obtained using the stable primer pairs of the present chapter. Sequence analysis of homologous genes of *Arabidopsis EF1α, F-BOX, SAND* and *PDF2* against publicly available monocotyledonous (monocots) and dicotyledonous (dicots) sequences highlighted significant variation of target primer sequences, more specifically at the 3’ end of the reverse primers. It was believed that this variation accounted for the inefficient amplification of these genes in *C. esculenta* and *N. benthamiana* and new generic primers specific to monocot and dicot species for *EF1α* and *F-BOX* were designed accordingly. Due to limited availability of sequence information for *SAND* and *PDF2* in publicly available databases and due to significant variation between monocots and dicots, primers for *SAND* and *PDF2* were not designed. The new primers were tested on *C. esculenta, N. benthamiana* and *Arabidopsis* using PCR, conditions optimised and appropriate products amplified. *EF1α* and *F-BOX* were subsequently validated as reference genes in virus infected *C. esculenta* and *N. benthamiana* and statistical analysis of the data obtained showed that both genes had similar and stable accumulation from both healthy and *Dasheen mosaic virus* (DsMV) infected tissue (Nair 2013).

While *Arabidopsis EF1α* and *F-BOX* primers gave inconsistent results when tested on *C. esculenta* and *N. benthamiana* (Nair 2013), this finding is in contrast to results of the present study where *Arabidopsis EF1α* was found to amplify a product of appropriate size for *N. benthamiana* (37B), *N. clevelandii, N. tabacum, N. occidentalis; N. sylvestris, N. tabacum White Burley, C. amaranticolor, C. quinoa, P. sativum, G. globosa* and *C. sativum. Arabidopsis F-BOX* amplified a product of correct size for all *Nicotiana* species investigated. *Arabidopsis F-BOX, SAND* and *PDF2* amplified products of the correct size for all other plant species (with the exception of *PDF2* from *N. benthamiana, N. occidentalis* and *C. sativum*) however, secondary non-specific products were also amplified. These results suggest that *Arabidopsis EF1α* might be suitable as a universal reference gene primer and,
with further development *Arabidopsis* F-BOX, SAND and PDF2, or their homologs in other plant species, might prove useful as primers for commonly used indicator plant species. Further amplification studies of these primers with numerous monocot and dicot species are warranted.

### 5.5 Summary

Selection of appropriate reference genes is necessary when performing real-time qPCR. Accumulation of these reference gene transcripts should not vary in the investigated cells or tissues throughout experimental conditions. Reference gene transcript stability varies considerably, and this variability can and does lead to incorrect conclusions being drawn from data. It is recommended that at least three validated reference genes be used to calculate a normalisation factor (Guénin et al. 2009; Klatte and Bauer 2009; Vandesompele et al. 2002). Under our experimental conditions of virus-infected *Arabidopsis*, transcript accumulation of SAND, PDF2, EF1α, and F-BOX showed a high level of stability and, although it is imperative that reference genes be validated for every new set of experimental conditions, the genes investigated here provide a starting point for those wishing to evaluate gene transcript accumulation of virus-infected *Arabidopsis* by qPCR. Furthermore, these data will be beneficial to the selection and design of primer pairs targeting orthologous genes in other plant species and will aid in the selection of suitable genes for normalisation in studies of diverse plant–pathogen interactions. If the *Arabidopsis* primers are found to be orthologous to genes of other plant species, as they are in the present study, and if these genes exhibit stable transcript accumulation across various experimental conditions, it will prevent the need for designing new primers to suit each indicator plant species.
CHAPTER SIX – TRANSCRIPT ACCUMULATION OF GENES OF INTEREST IN RESPONSE TO FIVE VIRUSES

6.1 Introduction

Analyses of plant virus-host interactions, using microarray technologies that have made it possible to examine ‘gene expression’ (hereon in termed gene transcript accumulation) and the networks of plant defence responses triggered by viral infection, have focused mostly on the effects of viral infection on mRNA transcript abundance of susceptible plant hosts (Glazebrook 2001; Whitham et al. 2003; Whitham et al. 2006, Wise et al. 2007). Many large-scale Arabidopsis microarray experiments profiling genome-wide transcript accumulation across different developmental stages, cell types, and environmental conditions have resulted in vast amounts of data (reviewed in Reddy et al. 2013). Comparative analyses of publicly available microarray datasets, in addition to other databases of the Arabidopsis transcriptome during compatible interactions with plant viruses, have identified several thousand genes responding to virus infections and identified pathways commonly modified by virus infections (reviewed in Postnikova and Nemchinov 2012). Such comparative analyses have ascertained that genes associated with defence, metabolism and protein synthesis are typically up-regulated, while genes associated with translation, photosynthesis, development and storage proteins are typically down-regulated (reviewed in Yang et al. 2007).

Even though several summarising attempts have been made, a general picture of gene transcript accumulation changes common to susceptible virus-host interactions is lacking. As discussed in Chapter 1.6, Postnikova and Nemchinov (2012) assembled a comparative analysis of currently available microarray data on changes of gene transcript accumulation in Arabidopsis in response to infection with
various plant viruses of different families. After subtraction of genes not common to all virus infections investigated, the number of shared genes affected by all viruses was reduced to 198. However, more than half of the genes’ transcript accumulation changes associated with all viral infections also changed in response to other biotic and to abiotic stresses, a finding supported by comparative analyses of Rodrigo et al. (2012). Despite the lack of encouraging results where most genes responding to numerous virus infections also responded to generic stress responses, it was felt that pursuing novel genes with fold-change in response to virus infection was appropriate. It is purported that the *Arabidopsis* genome presents a largely untapped source of new genes and mechanisms involved in virus resistance (Ouibrahim and Caranta 2013) and therefore likely that numerous genes have not been considered for their response to virus infection.

To identify and investigate genes of interest as indicators of plant virus infection the aims of the work of the present chapter were to:

(i) Identify genes from comparisons of publicly available microarray data and corresponding literature with transcript fold-change in common with several plant viruses.

(ii) Assess the attributes of potential virus-responsive genes through Gene Ontology (http://www.geneontology.org/) and cross-reference to the Gene Expression Atlas of the European Bioinformatics Institute (http://www.ebi.ac.uk/) in order to eliminate genes responding to other biotic and to abiotic stresses.
(iii) Quantify shortlisted, target gene transcript accumulation by qPCR experiments of cDNA from *Arabidopsis* leaf tissue either mock-inoculated or inoculated with one of five viruses across the infection time course.

(iv) Quantify target gene transcript accumulation of genes showing fold-change in response to all five viruses of this study by qPCR experiments of cDNA from *Arabidopsis* leaf tissue subject to a further biotic stress: *Pseudomonas syringae* pv. tomato, strain DC3000.

(v) Quantify target gene transcript accumulation of genes showing fold-change in response to all five viruses of this study by qPCR experiments of cDNA from *Arabidopsis* leaf tissue subject to two abiotic stresses: NaCl (100 and 200 mM) and drought.

### 6.2 Materials and methods

#### 6.2.1 Sourcing genes commonly showing transcript fold change in response to viruses in *Arabidopsis*

Genes of interest were selected from references to microarray data reporting fold change, in common, in response to TVCV, ORMV, PVX, CMV, TuMV and PPV. Genes reported as induced in common in response to TVCV, ORMV, PVX, CMV and TuMV in inoculated *Arabidopsis* leaves by Whitham et al. (2003) were compared to spatial analysis findings of *Arabidopsis* in response to TuMV (Yang et al. 2007). AGI numbers of genes shown to respond to virus infection in both the studies of Whitham et al. (2003) and Yang et al. (2007) were searched for further references to response to virus infection in other literature. AGI numbers of selected genes were entered into the European Bioinformatics Institute Gene Expression Atlas.
to confirm virus-specific gene transcript accumulation patterns. Genes affected by other biotic stresses and abiotic stresses as reported by literature at the time, and from cross-referencing to the Gene Expression Atlas, were excluded from qPCR analysis of transcript accumulation. Components of RNAi, as described and listed in Chapter 1.5.10, and three further genes of interest, were also selected for investigation of transcript accumulation by qPCR in response to five viruses.

**6.2.2 Virus isolates, plant material, and inoculations**

Virus isolates and plant material were obtained, virus- and mock-inoculations performed and tissue harvested as previously described (Chapter 2.3 to 2.6.3).

**6.2.3 Further biotic stress treatment - Pseudomonas syringae pv. tomato, plant material and inoculations**

*Arabidopsis* plants (obtained as described in Chapter 2.3) were inoculated with either *P. syringae* var. tomato strain DC3000 or mock-inoculated and leaf tissue collected as described in Chapter 2.6.3 and Table 2.4. Plant inoculations, tissue collection, RNA isolation and cDNA generation were kindly undertaken by Gardette Valmonte (Plant and Food Research, Mt Albert, Auckland, New Zealand).

**6.2.4 Abiotic stress treatments - 100 mM / 200 mM NaCl and drought**

For abiotic salt stress and drought treatments and controls, *Arabidopsis* plants were watered and the leaf tissue harvested and stored as previously described (Chapter 2.6.2 and Table 2.3). RNA isolation and cDNA synthesis was kindly performed by Gardette Valmonte (Plant and Food Research, Mt Albert, Auckland, New Zealand).
6.2.5 RNA isolation, DNase I digest, and cDNA synthesis

RNA isolation, DNase I digest and cDNA synthesis, quantification and quality control performed as previously described (Chapter 2.7.1).

6.2.6 Primer design

To ensure maximum specificity and efficiency during qPCR amplification primers were selected under a standard set of reaction conditions as described in Chapter 5.2.3. The primer and amplicon specifications for the three reference genes used in this study, EF1α, F-BOX and SAND, are provided in Chapter 5.2.3, Table 5.1. All oligonucleotide primers were manufactured by Sigma-Aldrich, Castle Hill N.S.W., Australia, and reconstituted according to manufacturer’s instructions.

6.2.7 Testing target gene primers with standard PCR

Target gene primers were tested with 25-µl reaction volume PCRs as described in Chapter 5.2.4. Likewise, negative control reactions, one omitting reverse transcriptase and one omitting cDNA template, were performed to confirm the absence of gDNA and other contaminants. The presence of single products of appropriate size was confirmed by electrophoresis as described in Chapter 5.2.4.

6.2.8 qPCR conditions and analyses

Conditions for qPCR reactions were as described previously (Chapter 5.2.5). Primer efficiency values and transcript accumulation quantity of each candidate gene were calculated as previously described (Chapter 5.2.6). Stability of accumulation of the three reference genes (EF1α, F-BOX and SAND) was analysed with geNorm v3.5
analysis software according to instructions in the operator’s manual and normalisation factors generated for the data analysis of each plate.

Results data were normalised by first calculating an SD factor with the following formula for three reference genes in an Excel spreadsheet (where SD refers to the normalisation factor calculated from GeNorm for each sample):

$$\left(\frac{SD\_{\text{ref1}}}{3\times\text{mean}\_{\text{ref1}}}\right)^2 + \left(\frac{SD\_{\text{ref2}}}{3\times\text{mean}\_{\text{ref2}}}\right)^2 + \left(\frac{SD\_{\text{ref3}}}{3\times\text{mean}\_{\text{ref3}}}\right)^2.$$

Calculation of an SD factor was followed by calculating the normalised mean by dividing mean quantity (Q) by the respective normalisation factor. The normalised standard deviation was calculated with the following formula in the Excel spreadsheet:

$$\text{normalised mean} \times \left(\frac{\text{SD factor}}{\text{normalisation factor}}\right)^2 + \left(\frac{\text{SD}}{\text{mean}}\right)^2)^{0.5}$$

The normalised mean was rescaled to 1.0 by dividing each cell by the sample mean selected to represent 1.0 (i.e. mock 2 dpi) and the standard deviation rescaled in the same way. Finally, a rescaled standard error was generated by dividing the rescaled standard deviation by the square root of the number of replicate wells of samples in the qPCR (n=3). Bar charts of normalised rescaled mean Q were plotted and the rescaled standard error fitted.

6.2.9 Standardisation of variable biological replicates

The inherent variability of individual biological replicates poses problems and results in data that is often not meaningful or is overlooked. To date, standardisation of biological replicates has not been well studied by the qPCR community (reviewed in Ruijter et al. 2013). Assessment of statistical significance requires a standardised
data set because variation in data from multiple replicates might not result in statistically significant differences, even if a biological effect is discernable in each of the individual replicates (Willems et al. 2008). Based on a series of sequential corrections, including log transformation, mean centering, and auto-scaling, Willems et al. (2008) describe a robust standardisation method that can be applied to data sets with high variation between biological replicates, enabling appropriate statistical analysis and the drawing of relevant conclusions. The sequential corrections method (log-transformation, mean-centering and auto-scaling) of Willems et al. 2008 was used to standardise gene transcript accumulation data where high variation was observed between biological replicates.

6.3. Results

6.3.1 Selection of genes of interest for qPCR analysis of transcript accumulation

Initial information of fold-change of transcript abundance in response to virus infection was obtained from microarray data of *Arabidopsis* leaves either mock-inoculated or inoculated with CMV, ORMV, TVCV, PVX, or TuMV across a time course of 1, 2, 4, and 5 dpi (Whitham et al. 2003). Genes that play a role in cell rescue, defence, aging and death, metabolism, signal transduction and transcription were the most commonly reported. These commonly reported genes included *ENHANCED DISEASE SUSCEPTIBILITY 1 (EDS1), PHYTOALEXIN DEFICIENT 4 (PAD4)*, regulators of R gene-mediated plant defence, and salicylic acid pathway related genes such as *β-1-3 GLUCANASE (BG3)*, and *PATHOGENESIS-RELATED-5 (PR-5)*, that have been reported to respond to viruses in a number of studies (e.g. Agudelo-Romero et al. 2008; Ascencio-Ibáñez et al. 2008; Babu et al. 2008; Huang et al. 2005; Marathe et al. 2004; Rodrigo et al. 2012; Whitham et al. 2003; 2006; Wise et
al. 2007; Yang et al. 2007). Homologs of these *Arabidopsis* genes, have been reported to respond to a number of pathogens in a number of plant species including *Vitis vinifera* (Gao et al. 2010), *Glycine max*, *Zea mays* (Kong and Li 2011) and *Solanum lycopersicum* (Alfenas-Zerbini et al. 2009). However, since *EDS1*, *PAD4*, *BG3* and *PR-5* mediate broad-spectrum resistance to diverse pathogens these four genes listed by Whitham et al. (2003) as commonly responding to viruses, have not been investigated as possible indicators of plant virus infection in this study.

The remaining subset of genes reported by Whitham et al. (2003) as being induced in common by viral infection were compared to the spatial analysis findings of *Arabidopsis* in response to TuMV (Yang et al. 2007) and AGI numbers of genes that were shown to respond to virus infection in both the studies of Whitham et al. (2003) and Yang et al. (2007) were entered into the European Bioinformatics Institute Gene Expression Atlas and interrogated for virus-specific gene transcript accumulation patterns. Removal of those genes not responding to virus infection and responding to more than one further biotic or abiotic stress resulted in a shortlist of eight genes for further investigation in the present study. The AGI, annotation, function (as described on TAIR), dpi collected and fold-change reported, and associated references for each target gene shortlisted for investigation of transcript accumulation in response to the five viruses of this study by qPCR are presented in Table 6.1.

Of the genes listed in Table 6.1, Postnikova and Nemchinov (2012) identified one report of *T15B16.5*, nine reports of *RLP23*, four reports of *TAT3*, three reports of *EVERSHEDE*, two reports of *TOUCH3*, five reports of *ZAT12*, and three reports of *PDX1.2* responding to virus infection from a comparative analysis of microarray data of the *Arabidopsis* transcriptome during 11 plant virus interactions (Postnikova and
Each of these genes had (as of 2009) as reported in the Gene Expression Atlas, been shown to respond to only one or zero other biotic or abiotic stress in microarray experiments. It was therefore believed appropriate to investigate these genes further. Components of RNAi, listed in Chapter 1.5.10, were selected for investigation of transcript accumulation by qPCR in response to five viruses. The component genes investigated were AGO1 and 4, DCL1 to 4, DRB1 to 5, HEN1, RDR6, SDE3 and 5, SGS3 and SMD1/RDR2.

Three further genes of interest, AUXIN RESPONSE FACTOR 6 (ARF6), ARF8 and RETINOBLASTOMA-RELATED PROTEIN 1 (RBR1) were also investigated. The patterns of expression of both ARF6 and ARF8 are controlled by miRNA167 (Wu et al. 2006) which can be modulated by virus infection (Amin et al. 2011). Further, misregulation of ARF8 is thought to underlie the developmental abnormalities caused by three distinct VSRs (Jay et al. 2011); therefore both ARF6 and 8 were considered for further analysis in the present study. The Geminivirus AL1 replication protein binds to host RBR1, and relieves repression of specific transcription factors thus activating genes required for propagation of the DNA virus (reviewed in Sánchez-Durán et al. 2011; Zhao et al. 2012). RBR1 was included for further investigation in the present study, particularly to determine if changes in transcript accumulation would be seen in response to the DNA virus, CaMV. The AGI, annotation, description and function of each component of RNAi and the three additional genes of interest are provided in Table 6.2.
Table 6.1. Genes commonly showing fold change as referenced in literature*#. Genes selected for qPCR analysis of transcript accumulation. Changes in response to Turnip vein clearing virus (TVCV), Oilseed rape tobadovirus (ORMV), Potato virus X (PVX), Cucumber mosaic virus (CMV), Turnip mosaic virus (TuMV) and Plum pox virus (PPV) in Arabidopsis.

<table>
<thead>
<tr>
<th>AGI†</th>
<th>Annotation</th>
<th>Function</th>
<th>dpi (fold change)‡</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cell rescue, defence, death &amp; aging</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At4g01700</td>
<td>T15B16.5 – Putative chitinase</td>
<td>Putative chitinase class I chitinase - Cell wall macromolecule catabolic processes. SAR – mediated by NPR1</td>
<td>4(3.4-8.1)‡, (31.1)‡</td>
<td>a, b, d, e, f</td>
</tr>
<tr>
<td>At2g32680</td>
<td>RECEPTOR-LIKE PROTEIN 23 (RLP23)</td>
<td>Protein binding, kinase activity signal transduction, defence response.</td>
<td>2(23.0-177.0), 4(39.9-172.4), 5(13.5-80.3)‡ (-2.6)‡</td>
<td>a, c, d, e</td>
</tr>
<tr>
<td></td>
<td>Metabolism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At2g24850</td>
<td>TYROSINE AMINO-TRANSFERASE 3 (TAT3)</td>
<td>Biosynthetic process, response to jasmonic acid stimulus, response to wounding. Free radical scavenging and synthesis of vitamin E.</td>
<td>4(5.0-42.3), 5(4.2-19.8)a</td>
<td>a, d, f</td>
</tr>
<tr>
<td></td>
<td>Signal transduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At2g31880</td>
<td>EVERSHED</td>
<td>Encodes a leucine rich repeat transmembrane protein expressed in response to Pseudomonas syringae. May be required for silencing via lsiRNAs. Positive regulation of cell death and innate immunity</td>
<td>5(2.0-5.5)a</td>
<td>a, d</td>
</tr>
<tr>
<td>At2g41100</td>
<td>CALMODULIN LIKE 4, TOUCH3</td>
<td>Encodes a calmodulin-like protein, with six potential calcium binding domains. Response to dark, temperature and mechanical stimuli.</td>
<td>5(3.2-8.8)a</td>
<td>a, d</td>
</tr>
<tr>
<td></td>
<td>Transcription</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At5g59820</td>
<td>RESPONSIVE TO HIGH LIGHT 41 (RHL41), ZAT12</td>
<td>Response to chitin, cold, heat, light stimulus, oxidative stress and wounding</td>
<td>5(4.0-11.5)‡, 17(-3.3)‡</td>
<td>a, d, e, f</td>
</tr>
<tr>
<td></td>
<td>Unclassified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At2g32210</td>
<td>F22D22.4</td>
<td>Unknown function</td>
<td>4(2.3-6.0)a</td>
<td>a, b, d</td>
</tr>
<tr>
<td>At3g16050</td>
<td>PYRIDOXINE BIOSYNTHESIS 1.2 (PDX1.2), A37</td>
<td>Encodes a protein with pyridoxal phosphate synthase activity. Low abundance. Synthesis of vitamin B6 (pyridoxine).</td>
<td>5(3.0-4.4)a</td>
<td>a, d</td>
</tr>
</tbody>
</table>

† Arabidopsis Genome Initiative (AGI) number assigned.
‡ dpi, days post inoculation that each gene was induced or repressed for all viruses tested; (fold change) is the minimum and maximum fold change for corresponding dpi compared to mock inoculated.
* Whitham et al. 2003
# Marathe et al. 2004
*+ Huang et al. 2005
*+ Yang et al. 2007
*+ Babu et al. 2008 - Plum pox virus
*+ Sekine et al. 2008
Table 6.2. Components of RNAi pathway and miscellaneous genes selected for qPCR analysis of transcript accumulation. AGI number, annotation, description and function.

<table>
<thead>
<tr>
<th>AGI†</th>
<th>Annotation</th>
<th>Description and function (TAIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At1g48410</td>
<td>AGO1</td>
<td>Encodes an RNA Slicer that selectively recruits microRNAs and siRNAs. Targeted for degradation by silencing suppressor F-box-containing proteins.</td>
</tr>
<tr>
<td>At2g27040</td>
<td>AGO4</td>
<td>AGO4 is a member of a class of PAZ/PIWI domain containing proteins involved in siRNA mediated gene silencing. Response to bacterial pathogens.</td>
</tr>
<tr>
<td>At1g01040</td>
<td>DCL1</td>
<td>Encodes a Dicer homolog. Dicer is a RNA helicase involved in microRNA processing. DCL1 is able to produce miRNAs and siRNAs.</td>
</tr>
<tr>
<td>At3g03300</td>
<td>DCL2</td>
<td>Involved in the production of ta-siRNAs. Partially antagonizes production of miRNAs by DCL1. Substitutes for DCL4 to produce viral siRNA when DCL4 is missing or inhibited. Able to produce siRNAs but not miRNAs.</td>
</tr>
<tr>
<td>At3g43920</td>
<td>DCL3</td>
<td>Required for endogenous RDR2-dependent siRNA (but not miRNA) formation. Secondary siRNA production.</td>
</tr>
<tr>
<td>At5g20320</td>
<td>DCL4</td>
<td>Involved in production of 21-nt primary siRNAs from inverted-repeat constructs and endogenous sequences, as well as the RDR6-dependent 21-nt secondary siRNAs involved in long-range cell-to-cell signalling</td>
</tr>
<tr>
<td>At1g09700</td>
<td>DRB1/HYL1</td>
<td>Nuclear dsRNA binding protein. Involved in mRNA cleavage. Interacts with DCL1</td>
</tr>
<tr>
<td>At2G28380</td>
<td>DRB2</td>
<td>Function unknown. Cytoplasmic dsRNA-binding protein. A maternally expressed imprinted gene</td>
</tr>
<tr>
<td>At3G26932</td>
<td>DRB3</td>
<td>Function unknown. Intracellular dsRNA binding protein.</td>
</tr>
<tr>
<td>At3g62800</td>
<td>DRB4</td>
<td>Nuclear dsRNA-binding protein -interacts with DCL4. May regulate DCL4 function and affect miRNA biogenesis. DRB4 interacts with the P6 viral protein from CaMV and may be a target of viral silencing suppression.</td>
</tr>
<tr>
<td>At5g41070</td>
<td>DRB5</td>
<td>Function unknown. Intracellular dsRNA binding protein</td>
</tr>
<tr>
<td>At4g20910</td>
<td>HEN1</td>
<td>Methyltransferase. Methylates miRNAs and siRNAs on ribose of the last nucleotide. Likely protects and stabilises the 3' ends of the sRNAs from uridylation. dsRNA binding domain.</td>
</tr>
<tr>
<td>At3g49500</td>
<td>RDR6</td>
<td>Encodes RNA-dependent RNA polymerase. Involved in trans-acting siRNA and other siRNA biogenesis. Required for post-transcriptional gene silencing and natural virus resistance.</td>
</tr>
<tr>
<td>At1g05460</td>
<td>SDE3</td>
<td>Encodes a protein with similarity to RNA helicases. Mutants are defective in post-transcriptional gene silencing</td>
</tr>
<tr>
<td>At3G15390</td>
<td>SDE5</td>
<td>Encodes a novel protein that is similar to PRL1 interacting factor and is involved in virus induced silencing. Function unknown. Thought to be involved in production of ta-siRNAs involved in RNA interference</td>
</tr>
</tbody>
</table>

Continued over
<table>
<thead>
<tr>
<th>AGI$^1$</th>
<th>Annotation</th>
<th>Description and function (TAIR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At4g11130</td>
<td><em>SMD1/RDR2</em></td>
<td>Encodes RNA-dependent RNA polymerase required for endogenous siRNA (but not miRNA) formation</td>
</tr>
<tr>
<td>At1g30330</td>
<td>Auxin response factor 6-ARF6</td>
<td>Encodes a member of the auxin response factor family. Mediates auxin response via expression of auxin regulated genes. Acts redundantly with ARF8 to control stamen elongation and flower maturation. Expression of ARF6 is controlled by miR167.</td>
</tr>
<tr>
<td>At5g37020</td>
<td>Auxin response factor 8-ARF8</td>
<td>Encodes a member of the auxin response factor family. Mediates auxin response via expression of auxin regulated genes. Acts redundantly with ARF6 to control stamen elongation and flower maturation. Expression of ARF8 is controlled by miR167</td>
</tr>
<tr>
<td>At3G12280</td>
<td>Retinoblastoma-related protein 1 (<em>RBR1</em>)</td>
<td>Controls nuclear proliferation in female gametophyte. Required for correct differentiation of male gametophyte cells. Regulates stem cell maintenance in <em>Arabidopsis</em> roots. Involved in the determination of cell cycle arrest in G1 phase after sucrose starvation. Also involved in regulation of imprinted genes.</td>
</tr>
</tbody>
</table>
6.3.2 Primer efficiency and transcript accumulation

In order to determine primer efficiency, values were calculated with LinRegPCR for each primer pair across each of three biological replicates. The mean efficiency and standard error of the mean of each pair are presented in Table 6.3. Amplification efficiencies between 1.7 (70% efficiency) and 1.9 (90% efficiency), which is considered acceptable (Pfaf fl et al. 2004; Vandesompele et al. 2002), were achieved using 28 of the 30 primer pairs under the qPCR conditions of the present study. \textit{F22D22.4} and \textit{DRB2} failed to obtain efficiencies above 1.6 (60% efficiency) and were excluded from further analysis.

The mean and standard error C\textsubscript{q} values for each gene amplified from each of the three biological replicates per leaf sample are given in Table 6.3. \textit{TOUCH3} was the most abundant reference gene transcript, with amplification-related fluorescence reaching a maximum at cycle 20.97 ± 0.19. In contrast, \textit{DCL3} and \textit{SMD1} were the least abundant transcripts, reaching maximum amplification fluorescence at cycle 33. Given that the qPCR protocol of this study used 40 cycles of amplification, C\textsubscript{q} values of 35 were considered late to amplify (late call) and as such, unreliable. \textit{RLP23}, \textit{DRB3} and \textit{RDR6} consistently resulted in C\textsubscript{q} values greater than 35 (Table 6.3) and were not included in further analyses.
Table 6.3. Genes selected for qPCR analysis of transcript accumulation: specifications of primers and amplicons.

<table>
<thead>
<tr>
<th>AGI⁵</th>
<th>Annotation⁶</th>
<th>5’→3’ Primer sequences</th>
<th>Primer efficiency LinReg⁷ (mean ± SE)</th>
<th>Mean Cq⁸ (± SE)</th>
<th>Tm⁹ (°C)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
</table>
| At4g01700 | T15B16.5 | F(970-991) TTAGAATGGGATTTCAAGGAGG
R(1096-1075) AGGAGAAGGGTCTTGTGGTTC | 1.8 ± 0.01 | 29.23 ± 0.84 | 60.31 | 127 |
| At2g32680 | RLP23* | F(2430-2451) TTCCCTCTCAAAGGAAGTTC
R(2549-2522) CTCCGGTTCAACCTTCTTT | 1.7 ± 0.10 | 35+ | 60.00 | 120 |
| At2g24850 | TAT3 | F(2065-2086) TCTATGATGGCTAGCCATCTC
R(2179-2158) TAATGTGGACCTGTGGCATAGG | 1.8 ± 0.02 | 28.33 ± 0.59 | 61.00 | 115 |
| At2g31880 | EVERSHED | F(1832-1852) AGGTTGATGAAGTTCGTAGG
R(1954-1924) CACTGTGCTTTGATCTGGGACA | 1.8 ± 0.01 | 25.53 ± 0.27 | 60.10 | 123 |
| At2g41100 | TOUCH3 | F(1189-1210) AGAAGGCTTGAGACCGACGG
R(1297-1276) TCTTTCATGAAGTCGGAAATTG | 1.9 ± 0.02 | 20.97 ± 0.19 | 60.52 | 115 |
| At5g59820 | ZAT12 | F(674-695) CATCCGCAGATTTAGGAAACTT
R(814-794) ACGTCGGCAAAATAGGCTAATA | 1.8 ± 0.02 | 32.61 ± 1.14 | 60.47 | 142 |
| At3G16050 | PDX1.2 | F(1111-1132) CATGTGTTGCTGAGATGAG
R(1230-1211) TTGCAGCTTCGAACCTTTTC | 1.8 ± 0.01 | 27.92 ± 0.39 | 60.18 | 120 |
| At2g32210 | F22D22.4* | F(674-695) TCACTCTCTTTTCTTTTGGTG
R(796-744) CCACGTTGAGACCGTCTAAATA | < 1.6 | 35+ | 60.08 | 123 |
| At1g48410 | AGO1 | F(5781-5800) ACCACGTTCTTTGAGGAG
R(6137-6128) GAATCGAGCCCTAAAGCTG
AG | 1.9 ± 0.01 | 26.20 ± 0.39 | 60.9 | 150 |
| At2g27040 | AGO4 | F(5279-5297) CACCAGTGCCATTTCTGTTG
R(5480-5461) CACCATGGCTTGATGATGTC | 1.9 ± 0.01 | 26.41 ± 0.19 | 60.2 | 113 |
| At1g01040 | DCL1 | F(7652-7670) TGGCCAATGCCTTCTTACA
R(7895-7875) CGCAGCTGACATCTTACATC | 1.9 ± 0.04 | 30.13 ± 0.45 | 60.2 | 150 |
| At3g03300 | DCL2 | F(7664-7683) AGCGGTCTCACAATCTCCTTCA
R(7903-7884) AATATCCGGCGCATCACGAGA | 1.9 ± 0.04 | 28.08 ± 0.24 | 59.6 | 148 |
| At3g43920 | DCL3 | F(6975-6996) GATACTCGAAGAGAGTGAG
R(7112-7091) CTTCTACGCGCTATCTAATGGAT | 1.8 ± 0.03 | 33.66 ± 0.86 | 59.2 | 85 |
| At5g20320 | DCL4 | F(9391-9411) CGAATGTTGTAAGAGAAGAG
R(9590-9571) ATATGGGGCGACTCTTACAC | 1.9 ± 0.04 | 29.63 ± 0.39 | 59.7 | 94 |
| At1g09700 | DRB1 | F(2278-2300) GGGAGTGTGGAGAACAGAG
R(2417-2395) CAACAAGACTCCATCTCTGACTG | 1.9 ± 0.03 | 29.71 ± 0.79 | 60.7 | 140 |

Continued over
Table 6.3. continued

<table>
<thead>
<tr>
<th>AGIa</th>
<th>Annotationb</th>
<th>5'→ 3'Primer sequences</th>
<th>Primer efficiency LinRegb (mean ± SE)</th>
<th>Mean Cq*c (± SE)</th>
<th>Tm*d (°C)</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>At2G28380</td>
<td><strong>DRB2</strong>*</td>
<td>F(2118-2138) TGAGGTTGTTTTTCTGAAAATTG R(2230-2208) TTTCCAACCTCCCCAATATAA</td>
<td>&lt; 1.6</td>
<td>35+</td>
<td>60.11 59.96</td>
<td>113</td>
</tr>
<tr>
<td>At3G26932</td>
<td><strong>DRB3</strong>*</td>
<td>F(1633-1654) GGATCAACCCCATGATTAAGG R(1742-1721) ACAGAAACCCATGACCTGATT</td>
<td>1.8 ± 0.02</td>
<td>35+</td>
<td>59.55 59.73</td>
<td>110</td>
</tr>
<tr>
<td>At3g62800</td>
<td><strong>DRB4</strong></td>
<td>F(1750-1769) CTTGTTGACCCCAAAGCACA R(1869-1848) CCTTAGCATTACTGGCGG</td>
<td>1.8 ± 0.01</td>
<td>29.8 ± 0.68</td>
<td>59.3 60.6</td>
<td>120</td>
</tr>
<tr>
<td>At5g41070</td>
<td><strong>DRB5</strong></td>
<td>F(1427-1446) ACCAAACCCCACTTAAACC R(1539-1520) TCTCTCTGCTGCCCAAC</td>
<td>1.8 ± 0.00</td>
<td>29.5 ± 0.37</td>
<td>60.0 61.6</td>
<td>113</td>
</tr>
<tr>
<td>At4g20910</td>
<td><strong>HEN1</strong></td>
<td>F(3836-3855) AGTTTGACTCCAGGCTGCAT R(4035-4016) TTTTTCCCAAATCTCACCG</td>
<td>1.9 ± 0.01</td>
<td>32.1 ± 0.78</td>
<td>59.87 59.77</td>
<td>95</td>
</tr>
<tr>
<td>At3g49500</td>
<td><strong>RDR6</strong></td>
<td>F(2139-2158) TGCTGACCTCCAGGCTGACAT R(3246-3227) CTTCTGCCCCATGCTCCCAA</td>
<td>1.9 ± 0.01</td>
<td>35+</td>
<td>59.87 60.48</td>
<td>145</td>
</tr>
<tr>
<td>At1g05460</td>
<td><strong>SDE3</strong></td>
<td>F(3620-3641) AGGTGATTTTGTAGCCAAGGAA R(3733-3712) CCAAAATAGCCTTCTTTCAC</td>
<td>1.8 ± 0.01</td>
<td>32.7 ± 0.63</td>
<td>60.0 59.8</td>
<td>114</td>
</tr>
<tr>
<td>At3G15390</td>
<td><strong>SDE5</strong></td>
<td>F(3405-3426) AGGAAGATTCTGCAAGCAACG R(3632-3611) TCAATTCTTGGGGTCAATCTC</td>
<td>1.8 ± 0.05</td>
<td>32.6 ± 1.04</td>
<td>59.9 60.3</td>
<td>139</td>
</tr>
<tr>
<td>At5G23570</td>
<td><strong>SGS1</strong></td>
<td>F(2352-2373) GGAGAATTTCGAGATGTTGCAG R(2599-2578) CCACAACCTCCCTCTCATT</td>
<td>1.9 ± 0.03</td>
<td>28.3 ± 0.19</td>
<td>61.1 59.6</td>
<td>158</td>
</tr>
<tr>
<td>At4g11130</td>
<td><strong>SMD1/RDR2</strong></td>
<td>F(3760-3781) CATTCTTGAGTTTTCCCATGGAT R(3875-3854) GCCAATCTCAATGGATACAGTG</td>
<td>1.9 ± 0.01</td>
<td>33.2 ± 0.97</td>
<td>60.2 60.1</td>
<td>116</td>
</tr>
<tr>
<td>At1g30330</td>
<td><strong>ARF6</strong></td>
<td>F(4361-4382) ATTTGAGGATTTGACCCCAAACGA R(4560-4541) CCCATTGTGGACCTTCTTCGT</td>
<td>1.8 ± 0.01</td>
<td>30.4 ± 0.40</td>
<td>60.51 60.71</td>
<td>159</td>
</tr>
<tr>
<td>At5g37020</td>
<td><strong>ARF8</strong></td>
<td>F(3824-3843) TGGCGAGTTGATTTGCGTAC R(4030-4011) TGATTTCCATTTGGATGCAC</td>
<td>1.9 ± 0.02</td>
<td>28.1 ± 0.29</td>
<td>60.50 60.30</td>
<td>116</td>
</tr>
<tr>
<td>At3G12280</td>
<td><strong>RBR1</strong></td>
<td>F(5073-5094) GTGGTTGCTTGCAGATGG R(5189-5168) GATGGCAGATAGGCTCTTGGAA</td>
<td>1.8 ± 0.02</td>
<td>29.0 ± 0.31</td>
<td>59.86 60.45</td>
<td>116</td>
</tr>
</tbody>
</table>

a Arabidopsis Gene Initiative (AGI) locus identifier number.
b Primer efficiency, mean ± standard error (SE) n=3.
c Threshold cycle.
d Melting temperature °C as reported by Primer3.
e Annotation of target genes of interest.
6.3.3 Transcript accumulation of genes commonly showing fold change as referenced in literature

The results of qPCR analysis of transcript accumulation as normalised average quantity (Q) data, of three biological replicates rescaled to 1.0 (mock 2 dpi), for each target gene in Table 6.1 are presented in Figures 6.1 to 6.5. An arbitrary fold change of ≥ 2.0 is commonly deemed to be of biological significance (reviewed in Dalman et al. 2012) and is the limit determining biological significance in the present study.

Figure 6.1A presents data for the putative chitinase, T15B16.5 where compared to mock-inoculated data a two-fold increase in transcript accumulation for only CaMV-inoculated tissue at 14 dpi was found (denoted with an asterisk). Figure 6.1B presents normalised rescaled data for one biological replicate of TAT3 where TAT3 failed to amplify in TuMV and TYMV samples. A biologically significant change in transcript accumulation was not found TMV and TSWV-inoculated samples compared to mock-inoculated samples. Figure 6.1C presents data for EVERSHED and compared to mock-inoculated samples, no significant change in transcript accumulation was found for virus-inoculated samples. Figure 6.1D presents data for TOUCH3 and compared to mock-inoculated samples, while there appears to be an increase in transcript accumulation at every time point for TSWV, there is no significant difference as evidenced by overlapping error representing variation in biological replicates. Figure 6.1E presents data for ZAT12 and compared to mock-inoculated samples, there appears to be an increase in transcript accumulation from TMV 3, 14, 21, 28 and 35 and TSWV 14 dpi inoculated samples. However, the difference is not statistically significant, with the exception of TMV-inoculated 21 dpi samples (denoted with an asterisk), as evidenced by overlapping error representing variation in biological replicates.
Figure 6.1. Normalised average Q rescaled to 1 of T15B16.5, TAT3, EVERSHEED, TOUCH3, ZAT12 and PDX1.2 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock inoculation across time course (n=3).
Figure 6.1.F presents data for PDX1.2 and compared to mock-inoculated samples, only CaMV-inoculated 2 dpi samples show an increase in transcript accumulation of PDX1.2 that is biologically significant. Taken together, these data of Figure 6.1 suggest a great deal of variability of transcript accumulation of each target gene across all inoculations and time points and show no significant fold-change and as such, no common biological significance that would indicate virus infection.

6.3.4 Transcript accumulation of components of RNAi and three additional genes of interest

Figures 6.2 to 6.5 present data for the components of RNAi and for the three genes of interest, ARF6, ARF8 and RBR1. Figure 6.2 shows data for the two AGOs (AGO1 and AGO4) investigated in the present study. Figure 6.2A shows that compared to mock-inoculation data, accumulation of AGO1 increases in response to TMV and TSWV-inoculations at 35 and 42 dpi (fold change ≥ 2.0; denoted by asterisks) and there is no significant change of AGO1 in response to CaMV, TuMV and TYMV-inoculations. Figure 6.2B shows no significant change in accumulation of AGO4 in response to any of the five virus inoculations compared to mock-inoculation as evidenced by overlapping error representing variation in biological replicates.
Figure 6.2. Normalised average Q rescaled to 1 of $AGO1$ and $AGO4$ in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock inoculation across time course (n=3).

A. Transcript accumulation of three biological replicates of $AGO1$ from mock-inoculated and five virus inoculated samples. B. Transcript accumulation of three biological replicates of $AGO4$ from mock-inoculated and five virus inoculated samples. Asterisks denote statistical significance. Error bars are rescaled standard error of the normalised and rescaled mean.
Figure 6.3 presents data for the four DCLs (DCL1, 2, 3 and 4). Figure 6.3A shows that compared to mock-inoculation accumulation of DCL1 increases in response to TSWV-inoculation at 2 to 21 and 42 dpi however, this increase is not of biological significance. Compared to mock-inoculation, DCL1 does not change in accumulation in response to TSWV inoculation at other time points or to inoculation with CaMV, TMV, TuMV or TYMV. Figures 6.3B and C show that compared to mock-inoculation, DCL2 and DCL3 do not respond to inoculation with any of the viruses of the present study as evidenced by substantial error representing variation in biological replicates. Figure 6.3D shows that compared to mock-inoculation DCL4 increases slightly in response to CaMV at 2dpi, however this response of not of biological significance. Compared to mock-inoculation DCL4 does not change in accumulation in response to inoculation with CaMV at other time points or to inoculation with TMV, TSWV, TuMV or TYMV.
Figure 6.3. Normalised average Q rescaled to 1 of DCL1, 2, 3 and 4 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across time course (n=3).

A. Transcript accumulation of three biological replicates of DCL1 from mock-inoculated and the five virus inoculated cDNA samples. B. Transcript accumulation of DCL2. C. Transcript accumulation of DCL3. D. Transcript accumulation of DCL4. Error bars are rescaled standard error of the normalised and rescaled mean.
Of the components of RNAi with a role in dsRNA binding (DRB1, DRB4, DRB5) and methylation (HEN1), transcript accumulation did not change in response to any of the virus-inoculations at any time point compared to mock-inoculated (data not shown). Further, of the components of RNAi purported to have a role in amplification and signalling (SDE3, SDE5, SMD1/RDR2 and SGS3) SDE3, SDE5 and SMD1 did not show any significant change in accumulation in response to any of the five viruses of the present study at any time point (data not shown).

However, in comparison to mock-inoculation, SGS3 showed a trend towards decreasing transcript accumulation in response to inoculation with CaMV, TMV, TSWV, TuMV and TYMV at most time points from 14 dpi (Figure 6.4; statistical significance denoted by asterisks). This trend was not considered to be of biological significance since observed fold-changes were less than 2.0.

![Figure 6.4. Normalised average Q rescaled to 1 of SGS3 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across time course (n=3).](image)

Transcript accumulation of three biological replicates of SGS3 from mock-inoculated and CaMV, TMV, TSWV, TuMV, TYMV-inoculated cDNA samples. Error bars are rescaled standard error of the normalised and rescaled mean. Asterisks denote statistical significance.
Figure 6.5 presents normalised rescaled data for the three additional genes of interest namely, \( ARF6 \), \( ARF8 \) and \( RBR1 \). Compared to mock-inoculation, accumulation of \( ARF6 \) did not respond to CaMV, TMV, TSWV, TuMV or TYMV-inoculation at any time point. These findings are apparent from the similar accumulation levels observed compared to mock-inoculation in response to CaMV, TMV and TSWV-inoculation and by the variation in biological replicates evidenced by overlapping error in response to TuMV and TYMV-inoculation. Figure 6.5B shows that compared to mock-inoculation, \( ARF8 \) does not respond to CaMV, TMV, TSWV, TuMV or TYMV-inoculation at any time point.

Since \( RBR1 \) activity is modulated during geminivirus replication (Sánchez-Durán et al. 2011; Zhao et al. 2012), \( RBR1 \) was included for investigation in the present study, to determine if changes in transcript accumulation would be seen in response to CaMV. Figure 6.5 C shows that \( RBR1 \) does not respond to CaMV, TMV, TSWV, TuMV or TYMV-inoculation at any time point.

Taken together, these data (Chapter 6.3.4) show a great deal of variation in transcript accumulation of all genes of interest at every time point for all virus inoculations compared to mock-inoculation.
Figure 6.5. Normalised average Q rescaled to 1 of ARF6, ARF8 and RBR1 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across time course (n=3).

A. Transcript accumulation of three biological replicates of ARF6 from mock-inoculated and the five virus inoculated cDNA samples. B. Transcript accumulation of three biological replicates of ARF8 from mock-inoculated and the five virus inoculated cDNA samples. C. Transcript accumulation of three biological replicates of RBR1 from mock-inoculated and the five virus inoculated cDNA samples. Error bars are rescaled standard error of the normalised and rescaled mean.
6.3.5 Log transformation, mean centering and auto-scaling of the quantity data (Q) of selected target genes

In the present study, SGS3 transcript accumulation across three replicates (Figure 6.4) showed a trend towards decreasing transcript accumulation in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation at most time points from 14 dpi. While this trend was not of conventional biological significance (arbitrary fold change of ≥2.0), the same trend or biological effect was discernable in the separate biological replicates. Figure 6.6A, B and C present transcript accumulation data for SGS3 from separate biological replicates (BioRep1, BioRep2 and BioRep3, respectively) in response to mock-inoculation and inoculation with each of the five viruses, CaMV, TMV, TSWV, TuMV and TYMV. Compared to mock-inoculation, SGS3 shows a decrease in accumulation from 14 dpi for most virus-inoculations (with the possible exception of TYMV) in every replicate.

In order to standardise biological replicate data of the present study the sequential corrections method of log-transformation, mean-centering and auto-scaling were performed on the SGS3, AGO1, AGO4 and HEN1 data sets, from 7 dpi, as previously described (Willem et al. 2008). The data were normalised to a single reference gene, EF1α. While AGO1, AGO4 and HEN1 did not show discernable effects in separate biological replicates, data were log-transformed, mean-centred and auto-scaled for comparison. However, the error for these three genes was substantially increased by log transforming, mean centring and auto-scaling and the data were therefore not meaningful and not shown.

Figure 6.7 presents average log transformed, mean centred and auto-scaled Q data of the three biological replicates of SGS3 in virus-inoculated samples compared to mock-inoculated samples of the same time point (dpi). Error bars represent upper 95% confidence interval (+95% CI).
Figure 6.6. Normalised average Q rescaled to 1 of three biological replicates of SGS3 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation across the time course (2 to 42 dpi).

Figure 6.7. Log transformed, mean centred and auto-scaled SGS3 transcript accumulation (Q).

Transcript accumulation of three biological replicates of SGS3 from mock-inoculated and the five virus inoculated cDNA samples and normalised to the reference gene EF1a. Virus-inoculated data are compared to mock-inoculated data of the same dpi (7, 14, 21, 28, 35 and 42). Error bars are upper 95% confidence level (+95% CI).

The data in Figure 6.7 shows that compared to mock inoculation of the same time point, SGS3 transcript decreases between 14 and 28 dpi for all viruses (except TMV, 28 dpi). However, given that the observed fold-change of SGS3 is less than 2.0 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation of the same time point between 14 and 28 dpi, these data, while statistically significant, are not considered biologically significant.

6.3.6 Transcript accumulation of SGS3 in response to a further biotic stress and two abiotic stresses

In order to evaluate the transcript accumulation of SGS3 in response to a further biotic stress, the bacterium P. syringae pv. tomato strain DC3000, Arabidopsis plants were either mock-inoculated or inoculated with P. syringae. Accumulation of SGS3 at 1, 2, 6 and 10 dpi of Arabidopsis cDNA was assessed by qPCR. The data presented in Figure 6.8 show that compared to mock-inoculation SGS3 accumulation in response to P. syringae increases with statistical significance at 2 dpi, however at < 2.0-fold this finding is not biologically
significant. Compared to mock-inoculation SGS3 accumulation does not differ at 1, 6 and 10 dpi in P. syringae-inoculated samples.

![Graph](image)

**Figure 6.8.** Normalised average Q of transcript accumulation of SGS3 in response to P. syringae pv. tomato, strain DC3000 compared to mock-inoculation of the same dpi.

Transcript accumulation of three biological replicates of SGS3 from mock-inoculated and P. syringae pv. tomato strain DC3000 at 1, 2, 6 and 10 dpi. Error bars are standard error of the mean.

In order to assess the transcript accumulation of SGS3 from abiotic stresses namely, NaCl (100 mM and 200 mM concentrations) and drought stress, Arabidopsis plants were either i) watered following a normal regime (control), or ii) watered with 100 or 200 mM NaCl (salt stress), or iii) watering was withheld for 7 d prior to leaf tissue collection at 7, 14 and 21 dpi (drought stress). The data of three biological replicates presented in Figure 6.9 shows that compared to control treatment there is a statistically significant decrease in SGS3 transcript accumulation at every time point for 100 mM and 200 mM NaCl stress and at every time point for drought stress. The fold change compared to control treatment in every case is $> 2.0$, showing that these results are of biological significance.
Figure 6.9. Normalised average Q of transcript accumulation of SGS3 in response to NaCl and drought stress.

Transcript accumulation of three biological replicates of SGS3 from control (normal watering regime) and watering with either NaCl 100 mM or 200 mM across a time course of 1, 4, 24, 48 h and 7 and 14 days; and 7, 14 and 21 d of drought where watering was withheld from 7 d prior to 7 d collection. Error bars are standard error of the mean.

Taken together, these results suggest that SGS3 transcript accumulation does not significantly change in response to *P. syringae* pv. tomato but does change significantly, both biologically and statistically, demonstrating a decrease in accumulation, in response to the abiotic stresses evaluated in this chapter. Interestingly, SGS3 transcript accumulation decreases to a small degree (< 2.0 fold) in response to the five virus infections but to a large degree (> 2.0 fold) in response to both abiotic stresses investigated in the present study. It is possible then that a statistically significant small fold change might be indicative of virus infection whereas a large, biologically significant, fold change might be indicative of abiotic stress. However, this possibility needs to be evaluated at across further viruses, including cryptic viruses, and abiotic stresses.
6.3.7 Combined sRNA/rRNA and SGS3 decision tree as a predictive model

Since generating a ratio of sRNA to rRNA of total virus-inoculated LMW-RNA correlated with a 94 % probability of detecting known virus infection across the present inoculations and time course (Chapter 4.3.10), and SGS3 transcript decreased between 14 and 28 dpi for all viruses (except TMV at 28 dpi; Chapter 6; Figure 6.7), it is proposed that a decision tree/predictive model based on the two assays can be devised to predict likelihood of virus infection. Since this decision tree/predictive model draws the work of all chapters of this thesis together, the development and application of this model is presented and discussed in Chapter 7.2.3, 7.2.4, 7.3.6 and Chapter 7.5.

6.4. Discussion

The results of this chapter indicate that transcript accumulation of all the genes selected for evaluation is not significantly biologically affected by virus inoculation (< 2.0 fold) as determined by the arbitrary fold change significance of ≥ 2.0 (reviewed in Dalman et al. 2012).

Biologically and statistically significant fold-change was not found in the present study for the putative chitinase T15B16.5 (except ≥ 2.0-fold for CaMV at 14 dpi). This finding is in contrast to Whitham et al. (2003) who reported fold change range (3.4 to 8.1) at 4 dpi from microarray data confirmed with qPCR analysis. The present finding is also in contrast to Marathe et al. (2004) who found an overall (across a 0 to 18 h time course) increased fold change (two-way ANOVA p < 0.001) of T15B16.5 in response to CMV using whole genome microarray confirmed by qPCR and Babu et al. (2008) who found a 31.1 fold increase in PPV-infected Arabidopsis leaf tissue at 17 dpi (Babu et al. 2008).

Thilmoney et al. (2006) found a 3.03 to 6.06 fold-change of T15B16.5 in response to P. syringae pv. tomato strain DC3000 and a 2.64 increased fold-change in response to human
Escherichia coli O157:H7 in a genome wide microarray analysis of Arabidopsis. Recently, Hok et al. (2011) demonstrated a greater than two-fold up-regulation of T15B16.5 during both early and late stages of infection with Hyaloperonospora arabidopsidis (downy mildew). These latter findings indicate a role for T15B16.5 in response to general biotic stress and as such T15B16.5 is not solely indicative of plant virus infection.

Biologically and statistically significant fold-change was not found in the present study for TAT3. However, only one replicate was able to be analysed. The finding of the present study is in contrast to the findings of Whitham et al. (2003) where a fold change range of 5.0 to 42.3 at 4 dpi and 4.2 to 19.8 at 5 dpi in response to TVCV, ORMV, PVX, CMV and TuMV was found. Similarly, the findings are in contrast to Yang et al. (2007) who found a decreasing spatial expression pattern of TAT3 in response to TuMV (P<0.005). More recently, TAT3 has been shown to be a maker for wounding and jasmonic acid signalling, demonstrating a role for TAT3 in general stress response (Brosché and Kangasjärvi 2012).

Biologically and statistically significant fold-change was not found in the present study for the receptor like kinase, EVERSHED. This is in contrast to Whitham et al. (2003) who reported a range of fold change of 2.0 to 5.5 at 5 dpi in response to TVCV, ORMV, PVX, CMV and TuMV. Recently, EVERSHED has been shown to mediate responses to the fungal pathogens Sclerotinia sclerotiorum and to P. syringae in Arabidopsis (Zhang et al. 2013). In addition, the tomato orthologue of Arabidopsis EVERSHED, tomato SOBIR1/EVR and its close homolog Solanum lycopersicum (Sl)SOBIR1-like mediate resistance to the fungal pathogens Cladosporium fulvum and Verticillium dahliae, respectively (Liebrand et al. 2013) suggesting EVERSHED is not specific in response to virus infection.

Biologically and statistically significant fold-change was not found in the present study for the calmodulin-like TOUCH3 and this is in contrast to Whitham et al. (2003) who
reported, from microarray data confirmed by qPCR, a range of fold change between 3.2 and 8.8 in response to the five viruses of their study.

Biologically and statistically significant fold-change was not found in the present study for ZAT12 (except statistical significance TMV 21 dpi) and this is in contrast to Whitham et al. (2003) who reported, from microarray data confirmed by qPCR, range of fold change for ZAT12 between 4.0 and 11.5 at 5dpi in response to the five viruses of their study. The finding of the present study is also in contrast with Babu et al. 2008 who reported a negative fold change (-3.3) of ZAT12 at 17 dpi in response to PPV. However, Mehterov et al. (2012) employed platforms for multi-parallel qPCR analysis of 217 antioxidant and 180 ROS marker genes and subsequent qPCR analyses revealed increased expression of many ROS-responsive and abiotic stress signalling genes including transcription factors and co-regulators such as ZAT12 that showed a fold change range of 5.0 to 39.5. In addition, ZAT12 is described as a gene of the plant core environmental stress response (PCESR), a group of genes differentially expressed in almost any biotic or abiotic stress condition (Hahn et al. 2013). These recent findings suggest ZAT12 is not specific in its response to virus infection. Furthermore, Sewelam et al. (2013) have shown in qPCR results that ZAT12 has a 7.10 fold change response in the ethylene response factor 6 (erf6) mutant compared to wild-type Arabidopsis plants under H₂O₂ treatment. Consequently, ZAT12, which is regulated by a number of stresses, is not solely indicative of plant virus infection.

Biologically and statistically significant fold-change was not found in the present study for PDX1.2 (except biological and statistical significant response to CaMV at 2 dpi), and this contrasts with the findings of Whitham et al. (2003) who reported, from microarray data, a range of fold change between 3.0 and 4.4 at 5 dpi.

Compared to mock inoculation, the transcript accumulation of the four DCLs investigated in the present study did not show biologically or statistically significant changes
in response to the five viruses of this study. The lack of transcript accumulation change for 
*DCL1* is not unexpected given that the priority of *DCL1* is to process and produce miRNAs in 
the nucleus of plant cells. However, Azevedo et al. (2010) reported that the TCV VSR, P38, 
binds to and inhibits function of *AGO1* triggering up-regulation of *DCL1*. *AGO1* mediates the 
activity of cellular miRNAs, including that of miR162, which normally dampens *DCL1* 
accumulation. It is possible that other VSRs with different modes of action inhibited the 
accumulation of *DCL1* in the present study. The VSRs of the five viruses of the present study 
and their mode of action as they relate to components of RNAi are outlined in Appendix I.

No significant transcript accumulation change was found for *DCL2* in the present 
study. This is in contrast to studies that show *DCL2* is able to gain access to substrates 
processed by *DCL4* when *DCL4* function has been genetically inactivated or inhibited by a 
VSR (Blevins et al. 2006; Deleris et al. 2006). The lack of significant transcript accumulation 
change for *DCL3* in the present study is in contrast to Akbergenov et al., Blevins et al., and 
Moissiard and Voinnet (2006) who found that *DCL3* processes the dsRNA of DNA viruses. 
*DCL3* did not respond to CaMV of the present study. However, Deleris et al. (2006) 
suggested that *DCL3* is not associated with defence against virus, rather processing 
endogenous dsRNA. The lack of significant transcript accumulation change for *DCL4* in the 
present study is surprising given *DCL4* is the preferential DCL for processing replicating 
viral RNAs and generates secondary vsiRNAs in conjunction with *RDR6* amplification 
products (Blevins et al. 2006; Bouche et al. 2006; Deleris et al. 2006). However, all 
*Arabidopsis* DCLs respond to a number of biotic and abiotic stresses (reviewed in Liu et al. 
2009) and are therefore not solely indicative of plant-virus infection.

Compared to mock-inoculation the data of the present study showed a biologically 
and statistically significant increase in accumulation of *AGO1* in response to TMV and 
TSWV-inoculation at 35 and 42 dpi (fold change ≥ 2.0). There was no significant change of
AGO1 in response to CaMV, TuMV and TYMV-inoculations compared to mock-inoculation. These findings are in contrast to Várallyay et al. (2010); Várallyay and Havelda (2013); and Várallyay et al. (2014) who noted a significant down-regulation of endogenous AGO1 transcript accumulation in a transient expression assays where the VSR, P19, inhibited the translational capacity of AGO1 by modulating endogenous miR168. This finding is supported by earlier work where some VSRs have been noted to impair AGO1 activity leading to a reduction in AGO1 transcript accumulation, for example, Potyvirus HC-Pro (Xie et al. 2003), 2b of CMV (Zhang et al. 2006) and VSRs that mimic host-encoded glycine/tryptophan (GW) repeat-containing proteins normally required for RISC assembly and function, such as P38 of TCV (Azevedo et al. 2010).

AGO1 functions after miRNA maturation targeting mRNAs complementary to miRNA sequences and mediating gene expression regulation for all currently characterised Arabidopsis miRNAs in many aspects of plant biology, including development and response to stress (Vaucheret et al. 2004; Baumberger and Baulcombe 2005; Voinnet 2009). While siRNAs incorporate into AGO1 to target complementary viral RNA, the exact mechanisms involved and whether AGO1 slices viral RNA remain unknown. AGO1 also plays a role in response to P. syringae pv. tomato where infection of Arabidopsis induces accumulation of miR393. This miRNA loads into AGO1 to silence the auxin receptor TRANSPORT INHIBITOR RESPONSE 1; a negative regulator of basal generic defence (Pumplin and Voinnet 2013). Therefore, AGO1 function is not specific to virus infection.

The data of the present study showed no biologically or statistically significant change in accumulation of AGO4 in response to all five virus inoculations at any time point compared to mock-inoculation. This is not surprising given that AGO4 primarily interacts with 24-nt rasiRNAs that are not thought to be involved in antiviral defence (Mi et al. 2008) but rather initiate RNA-directed DNA methylation (reviewed in Pumplin and Voinnet 2013).
However, these results contrast with studies that indicate DNA plant viruses are targeted by *AGO4* (Buchmann et al. 2009; Ding 2010, Raja et al. 2008) and with separate studies that indicate *Arabidopsis* treatment with *flg22*, an inducer of plant biotic defence responses, causes a rapid and transient down-regulation of key RdDM pathway components, including *AGO4* (Peláez and Sanchez 2013; Pumplin and Voinnet 2013). These findings suggest *AGO4* responses are not specific to viral infection.

To avoid repetition of findings indicating no statistical or biological change in transcript accumulation in response to CaMV, TMV, TSWV, TuMV or TYMV-inoculation at any time point compared to mock-inoculation, the results of *DRB1, DRB4, DRB5, HEN1, SDE3, SDE5* and *SMD1* are not shown. The results of the present study showed no change for DRBs 1, 4 and 5 in response to the five viruses compared to mock-inoculation. The finding of the present study for *DRB1* is in accord with other studies where *DRB1*’s primary role is to direct guide strand selection from miRNA duplexes (Eamens et al. 2009) and where it was concluded from a study where no alterations to *DRB1* transcript accumulation were reported in response to *Beet curly top virus*-derived siRNAs, that *DRB1* is not required for synthesis or accumulation of siRNAs (Raja et al. 2014). The finding of the present study for *DRB4* is in contrast to other studies where it has been reported that *DRB4* interacts with *DCL4, AGO1, AGO7*, and *RDR6* to function in RNAi (Nakazawa et al. 2007; Qu et al. 2008; reviewed in Raja et al. 2014). Furthermore, *DRB4* function is suppressed by nuclear import of the CaMV VSR, P6 (Haas et al. 2008).

It is not known how the results of the present study with regard to *DRB5* compare with the findings other researchers. Very little is known about the function of *DRB5* and involvement in the parallel RNAi pathways of *Arabidopsis* remains to be determined. *DRB5* is considered dispensable for miRNA processing but has been shown to assist in silencing transcripts targeted by *DRB2*-associated miRNAs (Eamens et al. 2012).
The finding of the present study with respect to HEN1 is in contrast to previous studies where VSRs, such as the potyviral HcPro, were thought to cause alteration in siRNA and miRNA levels, as a consequence of perturbed HEN1 activity (reviewed in Jay et al. 2011). The results of the present study demonstrating no change for SDE3 and 5 in response to the five viruses compared to mock-inoculation is in contrast to Garcia et al. (2012) where combined genetic, biochemistry and genomic studies showed that RDR6/SDE3 amplification machinery, and possibly SDE5, is required to enhance secondary dsRNA production from limiting amounts of transgenic or viral genomic RNA templates. These machinery are required for increasing vsiRNA to levels necessary to trigger optimal systemic silencing and as such are targeted by VSRs, likely resulting in changes to transcript accumulation (Ding and Voinnet 2007; Hernandez-Pinzon et al. 2007, reviewed in Jauvion et al. 2010).

Secondary vsiRNA biogenesis also requires host SMD1 (RDR2) in addition to RDR6, SGS3, SDE3 and 5. SMD1 is involved the control of chromatin structure and is required for endogenous siRNA (but not miRNA) formation and in the establishment of transcriptional gene silencing (Hunter et al. 2013). No change in transcript accumulation was found in the present study for SMD1 and this finding is in agreement with Aregger et al. (2012) and Garcia-Ruiz et al. (2010) where no effect on transcript accumulation was associated with SMD1 (RDR2) or RDR6 in response to both geminivirus (Aregger et al. 2012) and TuMV (Garcia-Ruiz et al. 2010) and no detectable effect on the accumulation of TuMV-GFP-derived siRNAs was found (reviewed in Peláez and Sanchez 2013).

The results of the present chapter show that compared to mock-inoculation, SGS3 transcript decreases with statistical significance between 14 and 28 dpi for most viruses. However, given that the observed fold-change of SGS3 is less than 2.0 in response to CaMV, TMV, TSWV, TuMV and TYMV-inoculation compared to mock-inoculation between 14 and 28 dpi, these data are not of biological significance. These results are in partial accordance
with studies that have noted a decrease in transcript accumulation of \textit{SGS3} in the presence of specific VSRs. For example, V2 of \textit{Tomato yellow leaf curl virus} interacts directly with host \textit{SGS3} in planta, outcompeting \textit{SGS3} for dsRNA substrates and suppressing RNAi (Glick et al. 2008; Fukunaga and Doudna 2009); and P2 of \textit{Rice stripe virus} interacts with a rice homolog of \textit{SGS3} (Du et al. 2011).

Finally, of the three additional genes of interest, \textit{ARF6}, \textit{ARF8} and \textit{RBR1}, none showed any change in transcript accumulation in response to any virus inoculation at any time point compared to mock-inoculation. These findings are in contrast to previous literature proposing three distinct and unrelated VSRs; HcPro, P19 and P15, modulate \textit{ARF8} accumulation (Jay et al. 2011) and that geminivirus AL1 binds to host \textit{RBR1} resulting in down-regulation of \textit{RBR1} and subsequent propagation of virus (Sánchez-Durán et al. 2011; Zhao et al. 2012).

The findings of the present study might differ to the findings of previous publications as a result of limitations associated with microarray and qPCR data analyses. These limitations and how they might have affected the data of the studies of the present chapter are discussed in Chapter 7.3.1 to 7.3.5.

While the findings of the present study for \textit{SGS3} are not considered of biological significance as determined by the arbitrary cut off of fold-change $\geq 2.0$, there is a growing school of thought that in order to efficiently catalogue gene transcript accumulation changes, based on the chosen statistical significance or fold change cut-off; microarray analysis can provide more than one answer. Arbitrary fold change cut-offs of $\geq 2.0$ and statistical significance of $P<0.01$ for microarray and qPCR data have lead researchers to only consider genes that vary wildly amongst other genes and leads to questions as to which more important within the interpretation of data; statistical significance or biological fold-change cut-off (Dalman et al. 2012). Dalman et al. (2012) obtained a data set from a published paper.
and reanalysed the raw data using multiple different analytical approaches. From these analyses, it was shown that by changing the statistical significance level and fold change cut off; more than one interpretation of the data can be obtained. Dalman et al. (2012) concluded that future studies need to unravel which significance criteria are relevant, biological or statistical, in order to accurately explain biological phenomena.

As a further example, in a recent study the P value a common index for the strength of evidence was found to be 0.01. Sensitive to controversies over reproducibility the researcher decided to replicate the study with additional data and the P value came out as 0.59, nowhere near the conventional level of significance of 0.05 (reviewed in Nuzzo 2014). However, it was discovered the problem did not lie in the analyses but in the nature of the P value, which is neither as reliable nor as objective as assumed. This is worrying in light of reproducibility concerns. When Fisher introduced the P value in the 1920s, it was not intended to be a definitive test, rather as an informal way to judge whether evidence was worthy of review. It has been recommended that researchers recognise the limits of conventional statistics and instead bring into their analysis, elements of scientific judgement about the plausibility of a hypothesis and study limitations (reviewed in Nuzzo 2014).

For most experiments, there is an unstated requirement that the work be reproducible, from independent experiments, with a strong preference for reproducibility in at least three experiments. Although it is purported that reproducibility of an experiment means that it can be replicated, a distinction has been made between reproducibility and replication. Reproducibility refers to a phenomenon that recurs even when experimental conditions may vary, while replication describes the ability to obtain identical results when experiments are performed under identical conditions. When findings are dependent on replication of precise experimental conditions, the result may be less important than a phenomenon that can be reproduced by a variety of independent and non-identical approaches. If the results of an
experiment from two different groups are concordant, then the experiment can be considered to be successfully reproduced, despite an experiment not being replicated. To ensure that research findings are robust, it is desirable to demonstrate reproducibility in the face of variations in experimental conditions. Reproducibility remains central to science, since we are limited in our ability to achieve absolute predictability in the natural world (Casadevall and Fang 2010). Reproducibility has the potential to serve as a minimum standard for judging scientific claims when full independent replication of a study is not possible (Peng 2011). These suggestions have implications for the present study where findings of previous studies were not reproduced and where SGS3 transcript changes showed statistical but not biological significance. These implications are discussed in Chapter 7.2.3.

The results of this chapter and the discussions from recent publications (Postnikova and Nemchinov 2012; Rodrigo et al. 2012) of the genes commonly showing fold-change in response to virus infection show that these genes are not solely indicative of plant virus infection. The present results show that of the components of RNAi investigated in the present study, only SGS3 showed a statistically significant change in transcript accumulation compared to mock-inoculation in response to all five viruses however, this finding is not considered to be of biological significance. While, SGS3 transcript accumulation is not biologically or significantly affected by P. syringae var. tomato in the work of this chapter SGS3 does show statistically and biologically significant change in accumulation in response to the abiotic stresses evaluated. As such, SGS3 is likely a biological indicator of abiotic stress and therefore not specific to plant virus infection. However, it would be pertinent to repeat the SGS3 experiments with Arabidopsis, with other plant species, further viruses and in response to further abiotic stresses in order to determine whether the same response of the present study can be reproduced and as such best explain the biological trend.
CHAPTER SEVEN – GENERAL DISCUSSION

7.1 Summary and main findings

This thesis documents the investigation of potential molecular indicators or ‘biomarkers’ of plant responses to virus infection with the long-term aim of developing tools for detecting virus-infected material. There were four aims:

1. To confirm presence and identity of the five selected viruses (CaMV, TMV, TSWV, TuMV and TYMV) in Arabidopsis.
2. To develop a method for the accurate quantification and analysis of sRNAs in response to five dissimilar viruses in Arabidopsis.
3. To develop a qPCR method for the quantification of genes showing fold-change in response to five dissimilar viruses in Arabidopsis.
4. To extend assays to a further biotic stress and two abiotic stresses in Arabidopsis in order to determine if any potential indicators are specifically indicative of virus infection.

The main findings as they pertain to each of these aims are summarised as follows:

1 – Confirm presence and identity of the five selected viruses in Arabidopsis

- Results of symptom observation, symptom photography and PCR amplification of cDNA established that each virus, CaMV, TMV, TSWV, TuMV and TYMV, was present in both inoculated leaf and systemic Arabidopsis leaf tissue and that none of the studied viruses was present in any mock-inoculated Arabidopsis tissue.

- Sequencing of amplicons determined that each virus shared between 96-98% nucleotide identities to respective virus species.
• Immunostrip® assay of TMV and TSWV indicated presence in virus-infected tissue and absence in mock-inoculated tissue.

• Results of qPCR analysis of the accumulation of specific viral gene transcripts demonstrated virus presence at each collection point (dpi) across the time-course of infection. The accumulation of each viral transcript from sample leaves was variable across the time course.

2 – Evaluate methods for the accurate quantification and analysis of sRNAs in response to the five viruses

• The PAGE/Typhoon/ImageQuant method had poor limits of detection, acceptable linearity and poor reproducibility both within and between gels and was deemed unsuitable for quantifying sRNA accumulation.

• Preliminary data analysis demonstrated that the Bioanalyzer method had good limits of detection, good linearity and good reproducibility both within and between gels. The Bioanalyzer method could accurately quantify LMW-RNA components from virus-inoculated samples.

• Regression and mixed model statistical analyses determined that based on the change in accumulation of any single component of LMW RNA, not all viruses could be detected at all times points. Principal component analysis determined that some of the LMW-RNA component species of TuMV at some of the time points of collection (dpi) of five virus-inoculations were distinct from mock-inoculation.

• Calculating the ratio of sRNA to rRNA as a proportion of averaged mock-inoculated accumulation resulted in a 94% correlation with infection of five known viruses in Arabidopsis and 100% with one TamMV in tamarillo.
3 – Develop a qPCR method for the quantification of genes showing fold change in response to five viruses

- In the process of developing a robust experimental design for qPCR analysis of target gene transcript accumulation (Chapter 6) the importance of selecting appropriate reference genes (formerly ‘housekeeping’ genes) was recognised.

- Stability of qPCR gene transcript accumulation experiments of cDNA from Arabidopsis leaf tissue either uninfected or infected with one of five viruses across an infection time course were undertaken and stably accumulating reference genes identified. This research resulted in the publication:


- The Arabidopsis primers to amplify EF1a, F-BOX, and SAND were found to be orthologous to genes of other plant species, including, Nicotiana benthamiana (37B), N. clevelandii, N. tabacum, N. occidentalis; N. sylvestris, N. tabacum White Burley, Chenopodium amaranticolor, C. quinoa, P. vulgaris, Gomphrena globosa and Cucumis sativum thus preventing the need for designing new primers to suit these indicator plant species. The Arabidopsis PDF2 primers failed to amplify products from N. benthamiana, N. occidentalis and C. sativum but were found to be orthologous to genes of N. clevelandii, N. tabacum, N. sylvestris, N. tabacum, Chenopodium amaranticolor, C. quinoa, P. vulgaris, and Gomphrena globosa.
• Of the genes commonly showing fold-change in response to virus infection, and of the RNAi components investigated, only SGS3 showed change of statistical significance in qPCR transcript accumulation assays (i.e. 0.5 to 1.0-fold decreased transcript accumulation in response to the five viruses between 14 and 28 dpi) in comparison to mock inoculated.

4 – Extend assays to a further biotic and two abiotic stresses in order to determine if any potential indicators are solely indicative of virus infection

• SGS3 transcript accumulation was not significantly affected *P. syringae* pv. tomato but did change significantly, both statistically and biologically (> 2.0-fold) in response to NaCl (salt) and drought stress.

7.2 Implications of the main findings

7.2.1 Bioanalyzer method for the quantification of LMW-RNA components

The main findings of Chapter 4 showed that the Bioanalyzer method is suitable for accurately quantifying LMW-RNA components from mock- and virus-inoculated *Arabidopsis* samples and were extended successfully to tamarillo, a solaceous crop plant infected with the potyvirus TamMV. This method provides researchers with a rapid means for quantifying LMW-RNA components and given further assessment, is likely to be applicable to quantification of LMW-RNAs in a wider range of plant species. However, in the present study not all viruses could be detected at all times points. Calculating the ratio of sRNA to rRNA proportion of averaged mock-inoculation showed a 94% correlation with detection of five known virus infections. It would be worthwhile further investigating this calculation by extending the Bioanalyzer method to assess further plant species, such as herbaceous indicators utilised by biosecurity and important agricultural plant species, in the context of further plant viruses. A tool that provides a high probability of detecting plant
viruses might prove useful as a first assessment of imported plant tissue prior to implementation of other diagnostic methods to ascertain virus identity.

7.2.2 Identification and validation of reference genes for virus infected *Arabidopsis*

The data of Chapter 5, as published, furnish plant virologists with reference genes for normalisation of qPCR-derived gene transcript accumulation in virus-infected *Arabidopsis* and, as demonstrated, have been beneficial to the selection and design of primers targeting orthologous genes in other plant species. The *Arabidopsis* primers targeting EF1α, F-BOX, SAND and PDF2 were found to amplify products of the correct size from other plant species (Chapter 5.3.7) thus in some cases preventing the need for designing new primers to suit each indicator plant species. However, a caveat on this finding is that while the *Arabidopsis* primers targeting EF1α, F-BOX, SAND and PDF2 were found to amplify products of the correct size from other plant species in PCR assays, these results require validation by assessing primer efficiency, Cq and GeNorm stability values from qPCR assays.

It was of interest to review published manuscripts that had cited Lilly et al. (2011), in order to determine if EF1α, F-BOX, SAND and PDF2 were being trialled and validated as potential reference genes for qPCR assays in *Arabidopsis* and other plant species in response to virus infection. The following is a brief discussion of published manuscripts that have cited Lilly et al. (2011) to the date of submission of this thesis. Sestili et al. (2013) investigated EF1α of *Cucumis melo* (melon) as a potential reference gene in the context of infection with the fungus *Fusarium oxysporum* f.sp. melonis however, melon EF1α had the least stable transcript of 11 candidate reference genes evaluated. Mafra et al. (2012) evaluated several suitable reference genes with NormFinder from different citrus organs following treatment with different biotic stresses, including virus stress, and found that their data with regard to F-BOX and SAND corroborated with the data of Lilly et al. (2011). Bazzini et al. (2011)
selected tobacco *EF1α* as an internal control when validating a set of miRNAs from microarray data by qPCR. *EF1α* was the most consistent across the three algorithms used to determine gene stability (GeNorm, NormFinder and BestKeeper) and this finding was in agreement with Lilly et al. (2011). Lui et al. (2012) analysed gene stability of a number of genes in parallel with geNorm, NormFinder and BestKeeper and statistical analysis revealed *PP2A* and *F-BOX* to be the most stable. These results corroborate the findings of Lilly et al. (2011) where it was found that *F-BOX* to be one of the most stable. However, Lui et al. (2012) incorrectly stated that Lilly et al. (2011) reported *PP2A1* as one of the most stable transcripts. This interpretation is incorrect, *PP2A1* was one of the least stable reported in Lilly et al. (2011) with an *M* value of greater than 1.5 (Chapter 5.3.4, Figure 5.2). Other publications have cited Lilly et al. (2011) with regard to the importance of selecting appropriate reference genes but have not investigated the suggested reference genes in their validation studies (Moreno et al. 2011; Die and Román 2012; Podevin et al. 2012; Tucker and Yang 2012; Scholtz and Visser 2013; Tucker and Yang 2013; Wieczorek et al. 2013; Brulle et al. 2014). Citation by these publications indicates that publication of the work of Chapter 5 has provided, and continues to provide, a useful contribution to plant pathologists and the qPCR community.

### 7.2.3 Changes in transcript accumulation of genes of interest in response to five viruses and of *SGS3* in response to five viruses, *P. syringae*, NaCl (salt) and drought stress

Of the genes commonly showing fold-change in response to virus infection in microarray data, and of the RNAi components investigated in qPCR transcript accumulation assays of Chapter 6, *SGS3* showed decreased transcript accumulation of statistical significance between 14 and 28 dpi. This finding was not considered to be of biological significance as determined by the arbitrary fold change of ≥ 2.0. However, as discussed in Chapter 6.4 the arbitrary ≥ 2.0 fold change and statistical significance of *P*<0.01 cut-offs for
microarray and qPCR data have led to consideration of only genes that vary wildly amongst other genes. In order to accurately explain biological phenomena it is deemed necessary to resolve which significance criteria are relevant, biological or statistical. Further, and in order to ensure that research findings are robust; reproducibility in the face of variations in experimental conditions needs to be demonstrated (Casadevall and Fang 2010; Peng 2011; Dalman et al. 2012; Nuzzo 2014). Therefore, statistical, rather than biological significance may well be used to explain the biological phenomena of SGS3 of the present study on the proviso that findings for SGS3 prove to be reproducible across further experimental conditions. Suggestions for further research with regard to SGS3 are discussed in section 7.5.

SGS3 transcript accumulation is not significantly affected by infection with P. syringae pv. tomato but does change significantly in response to the NaCl (salt) and drought stresses of the present study. These findings show that a decrease in SGS3 transcript accumulation is not specific to virus infection. However, given further analysis a decrease of less than 2.0-fold might prove to be indicative of virus infection. Similarly, and given further analysis, a decrease in SGS3 transcript accumulation of greater than 2.0-fold might prove to be indicative of abiotic stresses.

7.2.4 Development of a sRNA/rRNA and SGS3 decision tree predictive tool

Since generating a ratio of sRNA to rRNA of total virus-inoculated LMW-RNA correlated with a 94 % probability of detecting known virus infection (Chapter 4.3.10), and SGS3 transcript decreased between 14 and 28 dpi for most viruses (Chapter 6; Figure 6.7), it is proposed that a decision tree based on the two assays be devised as a model to predict likelihood of virus infection. Since, it is not known whether generating a sRNA/rRNA ratio is specific to virus infection, it is essential that specificity to virus infection is first determined. This can be achieved through application of the Bioanalyzer method to further biotic (e.g. P.
syringae pv. tomato) and abiotic stresses (e.g. NaCl and drought) and subsequent calculation of a sRNA/rRNA ratio as a proportion of averaged mock-inoculation.

Figure 7.1 outlines the methods to be undertaken from receipt of sample to use of the decision tree as a predictive model. In plain text are additional activities required to validate the decision tree. In bold text are those activities required once the decision tree is validated. Total RNA and LMW-RNA is to be isolated from ‘unknown’ sample/s (as per manufacturer’s instructions of respective isolation kits) (7.1 – 1). The Bioanalyzer method is then applied to LMW-RNA samples as described in Chapter 4.2.9 and the proportion LMW-RNA components of total LMW-RNA calculated as described in Chapter 4.3.7. Following calculation, mock inoculation is averaged and rescaled to 1.0. A ratio of sRNA/rRNA accumulation as a proportion of average mock-inoculation is calculated using the equation presented in Chapter 4.3.10 (Figure 7.1 – 2). Concurrently, qPCR analysis of SGS3 transcript accumulation on cDNA synthesised from total RNA samples is to be undertaken as described in Chapter 6.2.8 (7.1 – 3). SGS3 transcript accumulation data is analysed and if substantial variation between biological replicates is noted, results are log transformed, mean centred and auto-scaled as described in Chapter 6.3.5 (7.1 – 4). Data can now be assessed using the decision tree/predictive model shown in Figure 7.2 (7.1 – 5).

Given that the results of the LMW-RNA study (Chapter 4.3.7) predict that known virus infected samples have a sRNA/rRNA ratio of greater than 1.0, and if given further validation using a broad range of virus infections and plant species the same result is achieved, it may not be necessary to assay known infected LMW-RNA samples or provide mock-inoculated samples. If given further qPCR analysis and validation, SGS3 transcript accumulation proves to consistently decrease to ~ 2.0-fold or less in response to virus infection and consistently to ~ 2.0-fold or greater in response to abiotic stresses and the selected reference genes remain stable, it may not be necessary to assay SGS3 transcript
accumulation from known abiotic stress, mock-inoculation and control cDNA samples from total RNA. Ultimately, once further analyses have been completed and confidence is gained in the tool, only the methods in bold text (Figure 7.1) will remain as requirements for application of the decision tree/predictive model (Figure 7.2).

Figure 7.1. Flowchart of methods from receipt of sample to decision tree as a predictive model.

Sample/s received and LMW-RNA of ‘unknown’ sample and total RNA is isolated from each unknown sample, known infected samples, known abiotic stress samples, mock-inoculations and controls (1). The Bioanalyzer tool is used to determine proportion of each LMW-RNA component of total LMW-RNA samples. Mock-inoculations are averaged and data rescaled to 1.0. A sRNA/rRNA ratio as a proportion of averaged mock-inoculation is calculated (2). qPCR analysis of SGS3 transcript accumulation in response to unknown sample/s, known infected, abiotic stress and mock inoculations and controls is undertaken (3) and data analysed and rescaled (4). Once stages 1 to 4 are complete the decision tree can be used as a predictive model (5). In plain text are additional assays required to validate the decision tree. In bold text are remaining required assays once the decision tree is validated.
Figure 7.2 shows the decision tree for use as a model for predicting likelihood of virus infection. The tree indicates that if a calculated sRNA/rRNA ratio of LMW-RNA as a proportion of averaged mock-inoculation is not greater than 1.0 then tested samples are unlikely to be virus-infected. On the other hand, if the calculated sRNA/rRNA ratio of LMW-RNA as a proportion of averaged mock-inoculation is greater than 1.0 then tested samples are more likely to be virus infected. Based on the conditions and the results of Chapter 4.3.10, the likelihood of virus infection was predicted to be 94% from known infections. The likelihood of virus infection can be further assessed by qPCR analysis of SGS3 transcript accumulation. If transcript accumulation does not decrease compared to mock-inoculation then virus infection is unlikely. On the other hand if SGS3 transcript accumulation does decrease compared to mock-inoculation then virus infection becomes increasingly likely. The likelihood of virus infection can then be further assessed by qPCR analysis of SGS3 transcript accumulation in response to abiotic stresses, such as salt and drought. If SGS3 transcript accumulation decrease in response to abiotic stresses is greater than 2.0-fold then the sample is likely to be responding to abiotic stress or stress in general. On the other hand, if SGS3 transcript accumulation decrease in response to abiotic stress is less than 2.0-fold then virus infection is increasingly likely and since the sample has already demonstrated a sRNA/rRNA of greater than 1.0, a positive predictive value of greater than 94% is obtained.
If the calculated sRNA/rRNA ratio of LMW-RNA as a proportion of averaged mock-inoculation is not greater than one then the sample is unlikely to be virus-infected. If the calculated sRNA/rRNA ratio of LMW-RNA as a proportion of averaged mock-inoculation is greater than one then the sample is more likely to be virus-infected. Likelihood can be further assessed by qPCR analysis of $SGS3$ transcript accumulation in response to virus infection. If $SGS3$ transcript accumulation does not decrease compared to mock-inoculation then virus infection is unlikely. If $SGS3$ transcript accumulation does decrease compared to mock-inoculation then virus infection is increasingly likely. Likelihood of virus infection can be further assessed by qPCR analysis of $SGS3$ transcript accumulation in response to the abiotic stresses, NaCl and drought. If $SGS3$ transcript accumulation does change in response to abiotic stresses then the sample is likely to be responding to stress in general. If $SGS3$ transcript accumulation does not change in response to abiotic stress then a positive predictive value of $> 94\%$ is obtained and virus infection is likely.

Figure 7.2. Decision tree as a predictive model.
Based on the outcome of this predictive tool, researchers or biosecurity can decide whether to destroy the plants from which ‘unknown’ samples were taken or investigate the samples further with, for example, high-resolution tandem mass spectrometry or NGS to determine presence or absence of viral coat proteins (Blouin et al. 2010). This decision tree predictive model deserves further investigation and the limitations and suggestions for further research and validation are discussed in sections 7.3.5 and 7.5. In summary:

- A decision tree for use as a predictive model for determining likelihood of virus infection was devised from both the calculation of a sRNA/rRNA ratio and transcript accumulation of SGS3 in response to biotic and abiotic stresses and a positive predictive value of greater than 94% obtained.

- Decisions can be made as to whether plants should be destroyed or further analyses undertaken as a result of the outcome of this predictive tool.

**7.2.4 The sRNA/rRNA and SGS3 decision tree predictive tool as a complement to next generation sequencing (NGS) technologies**

Next generation sequencing (NGS) technologies have led to a revolution in virus discovery and new potential for diagnostics and generic detection of viruses (Boonham 2014). In fact, the identification and characterisation of known and unknown viruses in infected plants are currently among the most successful applications of NGS technologies (Barba et al. 2014).

NGS platforms differ in the methods used for generating template and some, such as rolling circle amplification are limited to certain types of viruses (i.e., circular DNA viruses). Others platforms are more broadly applicable to any virus type, for example, virus genomes have been successfully sequenced from total mRNA (reviewed in Boonham 2014). However, a draw-back of sequencing total mRNA is that, especially with low-titre viruses, most of the
sequence is host RNA and therefore does not identify the virus. This problem has been addressed by using an uninfected plant for subtractive hybridisation; thereby enriching for non-plant RNA and limiting the amount of sequencing required. Despite such enhancements, total mRNA sequencing may not capture those viruses that lack the terminal polyA sequences used to enrich mRNA (reviewed in Boonham 2014). Another frequently used method has been dsRNA isolation followed by random cDNA synthesis (Roossinck et al. 2010; reviewed in Boonham 2014). Given that endogenous plant RNAs do not form extensive double stranded structures, but replicative intermediates of RNA viruses do, this approach strongly enriches for viral nucleic acids. Nevertheless, some viruses, such as DNA viruses, produce little or no dsRNA and again may be missed using this approach (Boonham 2014).

Plant viruses can also be detected indirectly where NGS of 21–24 nt siRNAs, known as sRNA sequencing and assembly (sRSA), enriches for viral sequences and offers good opportunities to identify viruses infecting plants, even at extremely low titres, in symptomless infections, and including previously unknown viruses (Barba et al. 2014; Boonham 2014). However, while sRSA is useful for the identification of new and known viruses, the short length of sRNA sequences poses challenges to the assembly of full genome sequences, especially when samples are infected with several closely related strains (Boonham 2014).

However, as it is currently, NGS has a number of limitations and consequently, it remains a requirement of MPI New Zealand that newer technologies like NGS complement existing methods of virus detection. NGS is not able to determine if the sequenced virus is alive and replicating and therefore pathogenic or if the sequenced virus is from surface contamination of plant samples prior to processing. Importantly, NGS is unable to provide information about the biological impacts of cryptic (persistent/asymptomatic) viruses on host plants. A concern for biosecurity is that while cryptic viruses might not have an effect in wild plant hosts, the impacts on new hosts in a different environment cannot be known.
This is where the decision tree/predictive tool of the present study (Figure 7.2) can complement NGS and existing methods for the detection of plant viruses even if accurate only 94% of the time. It is recognised that attaining 100% accuracy in a biosecurity diagnostic setting is not realistic since detection is affected by a number of factors including specificity and sensitivity of tests, condition of sample, pathogen concentration, quality of molecular extraction and importantly, knowledge of pathogens and tools available for detection (Lisa Ward, personal communication, 2014). The decision tree/predictive tool is based on host plant responses to virus infection rather than targeting aspects of viruses themselves and can therefore assist with a deciding whether a virus is pathogenic or not. If no plant response is determined by sRNA/ratio and SGS3 transcript accumulation then it is unlikely that any virus sequence detected by NGS is infectious and capable of replicating.

The decision tree predictive model might have a place in the range of tools available to diagnosticians and enable decisions to be made as to whether plants are likely to be healthy, should be destroyed or undergo further testing. However, as discussed (7.3.5 and 7.5) there is more research required to validate the decision tree/predictive model tool of the present study in order to determine host plant responses to viral, further biotic and abiotic stresses, more specifically, examining the response of different plant species to mild strains of viruses, to cryptic (asymptomatic/ persistent) viruses and latent viruses, to a number of bacterial and fungal pathogens and to numerous abiotic stresses.

7.3. Limitations of the present studies

7.3.1. Limitations of quantification of LMW-RNA components

A limitation of the experiments used to conduct the Bioanalyzer method is that LMW-RNA assays were not extended to further biotic or to abiotic stresses. Total RNA rather than LMW-RNA was isolated from P. syringae- inoculated Arabidopsis and plants subject to
abiotic stresses. Therefore, without further experiments, it is not currently known if responses observed, and the sRNA/rRNA ratio calculated from the present study, are specific to virus infection.

With regard to quantifying sRNAs, direct evidence for specific roles of endogenous sRNAs in plant antiviral immunity has been limited because of disturbance to biogenesis and function of host plant sRNAs by viruses. In several cases, VSRs alter the global accumulation of sRNAs through interaction with components of the sRNA biogenesis pathway that may affect sRNA production, function, or stabilisation (Kasschau et al., 2003; Chen et al., 2004; Takeda et al., 2005; Bortolamiol et al., 2007; Csorba et al., 2007; Vogler et al., 2007; Azevedo et al., 2010; Varallyay et al., 2010; Shivaprasad et al., 2012; Reviewed in Peláez and Sanchez 2013). The actions of VSRs specific to the five viruses utilised for the experimental work of this thesis are discussed in Appendix I.

7.3.2 Limit to the number of genes investigated

It was beyond the scope of this thesis to investigate all genes reported from comparative microarray and qPCR transcript accumulation data to play a role in plant virus defence. It is possible that one or more of the thousands of genes not investigated are specific in their response to plant virus infection. Instead, several genes representative of all four arms of plant defence (including defence genes, hormone responsive genes and components of RNAi) were selected. AGO2 was not considered in the present study, with preference given to investigating AGO1 and AGO4 since AGO1 is the primary slicer targeting RNA viruses and AGO4, DNA viruses. However, as discussed in Chapter 1.5.8, Harvey et al. (2011) have shown that AGO2 provides a secondary antiviral mechanism that comes into play when the primary AGO1 layer is not active. Furthermore, AGO2 has been implicated in antiviral defence through a studies showing that ago2 mutants are hyper-susceptible to TCV and CMV infection and AGO2 expression can be induced with TCV and CMV inoculation of wild-type
Arabidopsis (Harvey et al. 2011; Wang et al. 2011). In addition, RISC containing AGO2 have a clear preference and high affinity for viral siRNAs of 21 nt with a 5′-A (Harvey et al., 2011; Wang et al. 2011). Efficient RNA cleavage and inhibition of viral replication has also been observed in experiments where AGO2/RISC and a synthetic 21-nt vsiRNA have been investigated (Pumplin and Voinnet 2013). A role for AGO2 in virus defence has been demonstrated in studies on N. benthamiana plants, where NbAGO, an Arabidopsis AGO2 homologue was considered vital for antiviral defence against Tomato bushy stunt virus (TBSV) (Scholthof et al. 2011). And more recently, Bilichak et al. (2014) have shown down-regulation of AGO2 transcript in wild type Arabidopsis infected with TRV in qPCR analyses. An oversight of the present study was the exclusion of AGO2 from the qPCR assays of transcript accumulation of AGOs and it would be pertinent to investigate AGO2 in light of the findings of recent studies.

Additional genes of interest have recently come to light. Lectins are sugar-binding proteins that play essential roles in the innate immunity of animal cells. Since lectins recognise specific mono- or oligosaccharides, they are regarded as self-nonself discriminating molecules, suggesting an ability to recognise microorganisms, including pathogens (Yamaji et al. 2012; Sugawara et al. 2013). Yamaji et al. 2012 analysed the resistance of Arabidopsis and N. benthamiana to the potexviruses, PVX, Asparagus virus 3, White clover mosaic virus and Plantago asiatica mosaic virus (PlAMV). Results of map-based positional cloning revealed that the lectin gene, JACALIN-TYPE LECTIN REQUIRED FOR POTEXVIRUS RESISTANCE1 (JAX1), was responsible for the resistance observed, impairing accumulation of potexvirus RNA in inoculated leaves. JAX1-mediated resistance did not show the properties of conventional R-gene mediated resistance and was found to be independent of plant defence hormone signalling and RNAi. The resistance observed, termed lectin-mediated resistance (LMR), was suggested to comprise an uncharacterised branch of antiviral plant
innate immunity and has since been observed in a variety of plant–virus interactions (Yamaji et al. 2012; Sugawara et al. 2013).

A further lectin gene, *RESTRICTED TOBACCO ETCH VIRUS MOVEMENT 1* (*RTM1*), of *Arabidopsis* ecotypes *Col-0* and Bayreuth-0 (*Bay-0*), inhibits the systemic movement of potyviruses including TEV, PPV, and LMV (Cosson et al. 2012). In *RTM1* resistance, viral replication and cell-to-cell movement in inoculated leaves appear unaffected, SA is not involved and an HR and SAR are not triggered (reviewed in Cosson et al. 2012). In contrast to *JAX1*, where resistance is conferred at the very early stage of virus infection, *RTM1* confers virus resistance in the later stage of virus infection, impeding long distance movement of viruses (Cosson et al. 2012; Yamaji et al. 2012; Sugawara et al. 2013).

Since *JAX1* functions independently of plant defence hormone signalling and RNAi (Yamaji et al. 2012; Sugawara et al. 2013) and *RTM1* resistance does not involve SA or an HR and SAR is not triggered (Cosson et al. 2012), these two genes appear specific to restriction of potexvirus and potyvirus movement respectively. Future studies analysing the mechanism of LMR might uncover conserved defence mechanisms against plant viruses (Yamaji et al. 2012). As such, *JAX1* and *RTM1* are of interest for further investigation and elucidation of their roles and mechanisms in plant virus defence.

### 7.3.3 Limitations of microarray data as a source of genes of interest

Comparative analyses of microarray data have become an intense area of research, producing large quantities of gene expression data and potentially providing insight into defence against plant viruses. However, microarray data has lacked standard requirements for analysis and publication and as such might not be a reliable source for identifying molecular biomarkers. Different array platforms and experimental designs produce data in various formats and units of measure and are often normalised in different ways making comparisons
of data error prone (Brazma et al. 2001). Brazma et al. 2001 proposed the Minimum Information About a Microarray Experiment (MIAME) in order to describe the minimum information to ensure that microarray data can be easily interpreted and that results from analysis can be independently verified. This includes providing raw data, normalised data, sample annotation, experimental design, feature descriptors, and a detailed account of pre-processing and normalisation (Brazma et al. 2001; Brazma 2009). Accordingly, database submission (e.g. Gene Expression Omnibus or Array Express) and an concomitant accession number is now a requirement for publication of gene expression results and provision of MIAME are required by most journals as a prerequisite for review or acceptance (Leonelli et al. 2013; Witwer 2013).

Studies have shown that despite the wide adoption of MIAME principles by scientific journals, it has not always been possible obtain data or reproduce results, largely because data were not MIAME compliant (Brazma 2009; Witwer 2013). MIAME compliance was reviewed for 127 articles that included microarray-based miRNA profiling and were published from July 2011 through April 2012 from a number of journals across a number of disciplines. Overall, data submission was reported at publication for 40% of articles, and almost 75% of articles were not MIAME compliant. For several articles that were not MIAME compliant, data reanalysis revealed less than complete support for the published conclusions, leading to retraction of publication in at least one case (Witwer 2013).

In the present study, microarray data were used to inform new study design as it was from publicly available microarray data that some of the genes of interest for investigation in the present study were derived. It is possible that some arrays showed poor compliance to MIAME and in these cases, microarray data was likely an unreliable source of genes of interest.
It is typical in microarrays for only five or fewer replicates to be present in each group, and as such a single outlier observation could generate perceived accumulation differences of greater than two-fold. In addition, the high background level and limited dynamic range of detection typical of microarray leads to low sensitivity meaning accumulation of low abundance genes can be difficult to determine (Riedmaier and Pfaffl 2013). Therefore, many genes that might have been of interest to the respective array studies and to the present study might well have been missed.

A further limitation of microarray for the selection of genes of interest lies in the fact that comparison of microarray data from host-virus interactions are also complicated by the use of different plant hosts, platforms, and lack of studies encompassing an extensive time course. Most published transcriptome studies of plant-virus interactions have been conducted prior to or at the time of observable symptom development. The very early responses on the entire transcriptome as a result of plant-virus interactions are even less well understood (Golem and Culver 2003; Whitham et al. 2003; Senthil et al. 2005; Yang et al. 2007).

### 7.3.4 Limitations of qPCR data as a source of genes of interest and as a method for quantifying transcript accumulation

In contrast to microarray, qPCR has high sensitivity, specificity and wide linear dynamic range and as such is increasingly being used to detect and quantify microbial pathogens. However, pathogen abundance is often low and assay reproducibility is influenced by distribution statistics with sampling affecting the reliability of data (Johnson et al. 2013).

As discussed in Chapter 5.4 minimum standards for the provision of information for qPCR experiments (MIQE) have been proposed (Bustin 2010; Bustin et al. 2009). By implementing the MIQE guidelines, qPCR can become a robust, accurate, highly reproducible and reliable method for reporting nucleic acid quantification. However, despite MIQE, many reviews and meta-analyses on the use of qPCR tests for diagnostic assays have highlighted a lack of
homogeneity of methods used, lack of rigorous experimental controls and a prevalence of false-positive and false-negative results - all combining to hinder interpretation (reviewed in Johnson et al. 2013). In addition, it has been noted that it is difficult to have a qPCR reaction fail but alarmingly simple to produce poor quality data (Johnson et al. 2013). Therefore qPCR validation of microarray data and assay of transcript accumulation of individual genes of interest might not be a reliable method for identifying and quantifying molecular biomarkers.

MIQE and the 11 golden rules of qPCR (Udvardi et al. 2008) prompted the development of a rigorous method for qPCR analysis in the present studies, i.e. material was harvested from three biological replicates and stored appropriately, high-quality total RNA was isolated as determined by both Agilent Bioanalyzer RIN values and Nanodrop spectrophotometry, purified RNA was digested with DNase I to remove gDNA, RT reactions were performed with a robust reverse transcriptase to maximise cDNA yield, yield and quality of cDNA was assessed, gene-specific PCR primers were designed using a standard set of criteria, technical errors were reduced by using a robot where possible, reference genes were sought and validated resulting in publication and finally, and the relative transcript accumulation for each gene was calculated using qPCR efficiencies for the test gene and Cq values for both test and reference genes. While MIQE were fulfilled in the present studies it is not possible to account for the quality control of qPCR validation of microarrays of other researchers from which information with respect to some of the genes of interest for the present study was obtained.

7.3.5 Limitations of the combined sRNA/rRNA and SGS3 decision tree as a predictive model

A major limitation of the combined sRNA/rRNA and SGS3 predictive tool is that it is not currently known if calculating an sRNA/rRNA ratio as a proportion of average mock-inoculation is specific to virus infection. It remains to be determined if an sRNA/rRNA ratio
of greater than 1.0 is also predictive of further biotic and/or abiotic stresses and therefore indicative of a response to generic stress. It is also not known if a small decrease in SGS3 transcript accumulation (e.g. < 2.0-fold), such as that seen in response to virus infection in the present study is indicative of virus infection and a large (e.g. > 2.0-fold) change, such as that seen in response to abiotic stress in the present study, is indicative of abiotic stress.

A further limitation is that a decrease in SGS3 transcript accumulation compared to mock-inoculation was only seen between 14 and 28 dpi in the present study. It cannot be known for how long virus infection has been present in any plant material obtained at the border (if infected) therefore, any prospective tool for predicting virus infection needs to be derived from a molecular marker that is present across an extensive entire time course of infection. However, a way to circumvent this limitation, although time consuming and dependent on mechanical transmissibility of any ‘unknown’ viruses in the sample/s, would be to inoculate indicator plants with sap inoculum obtained from the sample/s in question and assay inoculated tissue between 14 and 28 dpi. At present, because SGS3 accumulation only changed compared to mock-inoculation between 14 and 28 dpi, it is not possible to average mock inoculation across the time-course to 1.0. If mock-inoculated SGS3 could be averaged then SGS3 transcript accumulation from ‘unknown’ samples could be rescaled to 1.0 and provision of mock-inoculated tissue for assay would not be necessary.

7.3.6 Limitations of transcriptomic analyses and the complexity of plant defence response pathways

The plant host defence pathways mounted against plant viruses and the network of interactions between plant host and plant virus are extraordinarily complex, making changes specific to plant virus infection difficult to distinguish. This complexity is exemplified by microarray and deep-sequencing comparative analyses of plant–virus interactions where
thousands of host gene transcripts differentially respond to plant virus infection (reviewed in Di Carli et al. 2012; Postnikova and Nemchinov 2012; Rodrigo et al. 2012; Fan et al. 2014).

Each plant virus encodes an average of between 4-10 proteins necessary to coordinate the complex biochemical and molecular interactions required for viral infection cycles. Although plant viruses usually encode only a few proteins, the expression profiles of the infected host usually show significant changes. As an example, the DNA genomes of geminiviruses encode 5–7 proteins that redirect host machineries and processes to establish a productive infection. These interactions reprogram plant cell cycle and transcriptional controls, inhibit cell death pathways, interfere with cell and hormone signalling and protein turnover, and suppress defence pathways (Hanley-Bowdoin et al. 2013). Gene transcription is a hugely dynamic process adapting rapidly to exogenous and endogenous changes of target cells or tissues (Riedmaier and Pfaffl 2013). Because of these constant changes it would be near impossible to find a single, let alone many, transcriptomic indicators of plant virus infection that are present across the time course of an entire time course of infection.

Until recently, functional genomics studies of plant–virus interactions have been limited to mainly transcriptomic analysis (Di Carli et al. 2012). Although changes at the transcript level can give good indications of plant defence responses indicative of specific signalling events, transcript level may not always be directly translatable into protein abundance or activity. Proteins may be regulated at the levels of transcription, translation, stability and activity; therefore transcriptomic, proteomic, and enzyme activity data all need to be considered in order to gain a complete view of plant responses to virus infection (Niehl et al. 2013, Wu et al. 2013). It is clear that studies based solely on transcriptional changes do not delineate the complexity of regulatory mechanisms taking place in host cells (Costa et al. 2013).
Only since 2010, due to rapid technical advances in both bioinformatics and proteomics tools (in particular, mass spectrometry), has an increasing number of papers focused on proteomic analysis of plant–virus interactions been observed (Di Carli et al. 2012). A review focused on comparative proteome-based studies of the pathogenesis of several viral genera, revealed widespread down-regulation of proteins associated with the photosynthetic apparatus, and up-regulation of proteins associated with energy metabolism/protein synthesis and turnover (reviewed in Di Carli et al. 2012). More recently, Wu et al. 2013 used 2-DE and matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS), to analyse the proteomic response of resistant and susceptible genotypes of maize to Sugarcane mosaic virus (SCMV) and found 96 protein spots with significant changes. Like the studies reviewed in Di Carli et al. (2102), a number of the proteins identified by Wu et al. (2013) were related to photosynthesis, energy and metabolism, and stress and defence response. Most proteins identified were located in chloroplasts and chloroplast membranes.

Recently, interaction analysis using both transcriptional regulatory network (TRN) and protein-protein interaction network (PPIN) have revealed that in Arabidopsis genes that are highly connected, central and organised in modules are preferentially perturbed. Termed ‘hub’ proteins, these are proteins with a large number of interactions within a network of the plant interactome (Elena and Rodrigo 2012; Rodrigo et al. 2012). Hub proteins that interact with SGS3, as annotated on the STRING database, include RDR6, DCL4 and DRB1. This supports the suggestion that it is unlikely that a single molecular indicator of plant virus infection will be found and instead a set of multiple indicators might be required to provide evidence of infection.

It has been suggested that future focus on the integration of data derived from different transcriptomic, proteomic and metabolomic approaches will contribute to the
understanding of the plant host–virus interactome, and may possibly reveal protein targets or biomarkers related to plant-virus interactions useful to the design of future diagnostic tools (Di Carli et al. 2013; Wu et al. 2013).

7.4 Suggestions for further research

The implications and limitations of the findings presented in this thesis suggest a number of directions for future research. First, the limitations of the sRNA/rRNA ratio calculation can be addressed by applying the Bioanalyzer method to LMW-RNA samples subject to further biotic stresses for example, *P. syringae* pv. tomato and *Hyaloperonospora arabidopsis* (powdery mildew) and to abiotic stresses such as NaCl and drought.

Second, analysis of LWM-RNA could be taken in a further direction where stem-loop qPCR might be used to quantify specific endogenous miRNAs, for example miR156 and miR172, in response to the five viruses, *P. syringae*, and abiotic stresses of the present study. This will determine if quantification of these miRNAs might be a useful tool for detecting virus presence or absence.

Third, it will be of interest to investigate the transcript accumulation of *AGO2* and the two LMR genes, *JAXI* and *RTMI*, in response the five virus infections, *P. syringae* pv. tomato, and to the abiotic stresses NaCl and drought in order to determine if these genes show a response specific to virus infection.

Fourth, a commercial antibody to *Arabidopsis SGS3* is available and as such *SGS3* might be investigated for changes in expression at the protein level. Differential In-Gel Electrophoresis (DIGE) technology might be considered. A multidisciplinary approach integrating data derived from transcriptome, proteome and metabolomic platforms could provide further insights toward the comprehension of viral pathogenic processes and towards the identification of host indicators specific to plant virus infection from complex pathways.
(Di Carli et al. 2012). It would be pertinent to investigate transcript accumulation and/or protein-protein interaction changes of *SGS3* in the context of changes to transcript and/or protein accumulation of its hub proteins for example, *RDR6, DCL4* and *DRB1*.

Fifth, in order to circumvent the limitations of the proposed decision tree predictive model, as discussed in the first suggestion for further research, specificity to virus infection needs to be determined by applying the Bioanalyzer method to further biotic and abiotic stresses and from further plant species, and sRNA/rRNA ratios calculated. It would also be necessary to inoculate indicator plants with sap inoculum from ‘unknown’ samples, provided the ‘unknowns’ are mechanically transmissible, and assay inoculated tissue between 14 and 28 dpi by qPCR.

Finally, further analyses of *SGS3* transcript accumulation might enable calculation of a mock-inoculated average where transcript accumulation data of future samples might be rescaled to 1.0, removing the need for mock-inoculated controls. Further qPCR assay of *SGS3* transcript accumulation in response to further virus infections, including mild strain and cryptic viruses, further abiotic stresses and in further plant species will determine if a small decrease in transcript accumulation (< 2.0 fold) is indicative of virus infection and a large decrease (> 2.0 fold) indicative of abiotic stress and will also determine reproducibility of *SGS3* data.
7.5 Concluding remarks

An ideal molecular biomarker or indicator of plant virus infection needs to be sensitive and specific and easily and accurately detectable from all plant tissues across the duration of an infection. It is unlikely that a single molecular indicator of plant virus infection from plant hosts will be found and instead, a set of multiple complementary indicators might be required (Riedmaier and Pfaffl 2013). With these factors in mind our ability to detect generic plant virus infection from altered physiology of host plants is likely to benefit from a multidisciplinary (e.g. transcriptomic, proteomic, metabolomic) or algorithmic approach. It is hoped that given further investigation and analysis, the decision tree predictive model proposed in this thesis will contribute to such an approach and find a place in the range of tools at the disposal of biosecurity diagnosticians.
REFERENCES


Filiz E, Koc I (2013) In silico analysis of dicer-like protein (DCLs) sequences from higher plant species. IUFS Journal of Biology 72: 10


Hernandez-Pinzon I, Yelina NE, Schwach F, Studholme DJ, Baulcombe D, Dalmay T (2007) SDE5, the putative homologue of a human mRNA export factor, is required for transgene silencing and accumulation of trans-acting endogenous siRNA. The Plant Journal 50: 140-148

245


Hwang J, Oh C-S, Kang B-C (2013) Translation elongation factor 1B (eEF1B) is an essential host factor for Tobacco mosaic virus infection in plants. Virology 439: 105-114


Lokesh B, Rashmi PR, Amruta BS, Srisathiyaranarayanan D, Murthy MRN, Savithri HS (2010) NSs encoded by Groundnut bud necrosis virus is a bifunctional enzyme. PloS ONE 5: e9757


Puustinen P, Mäkinen K (2004) Uridylylation of the Potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. Journal of Biological Chemistry 279: 38103-38110


Tucker ML, Yang R (2013) A gene encoding a peptide with similarity to the plant IDA signaling peptide (AtIDA) is expressed most abundantly in the root-knot nematode (Meloidogyne incognita) soon after root infection. Experimental Parasitology 134: 165-170


Xiong R, Wang A (2013) SCE1, the SUMO-conjugating enzyme in plants that interacts with Nlb, the RNA-dependent RNA polymerase of Turnip mosaic virus, is required for viral infection. J. Virol. 87: 4704-4715


APPENDIX I – PRÉCIS OF EACH OF THE FIVE VIRUSES OF THIS STUDY

Below are précis of each of the five viruses utilised in this study with particular reference to transmission, host range, particle morphology, symptomology, replication, transcription and translation and respective virus encoded suppressors of RNAi (VSRs).

A.1

*Cauliflower mosaic virus* (CaMV)

<table>
<thead>
<tr>
<th>Family:</th>
<th>Caulimoviridae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genus:</td>
<td>Caulimovirus</td>
</tr>
<tr>
<td>Species:</td>
<td><em>Cauliflower mosaic virus</em> (Type member)</td>
</tr>
<tr>
<td>Acronym:</td>
<td>CaMV</td>
</tr>
</tbody>
</table>

Viruses of the Caulimoviridae are plant pararetroviruses whose virions contain circular, dsDNA and replicate by reverse transcription of an RNA intermediate. Caulimoviruses are the only group of dsDNA viruses known to infect only plants (reviewed in Hapiak et al. (2008)).

A.1.1 Transmission, host range and particle morphology

Transmission of CaMV follows a ‘helper strategy’ in that an aphid transmission factor (ATF) interacts with both virus particle and a receptor in the aphid stylet. This forms a two-way molecular bridge enabling retention of infectious virus in the aphid mouthparts and subsequent egestion of infection to a new host plant (reviewed in Ziegler-Graff and Brault 2008). CaMV is readily transmitted by mechanical inoculation and by grafting, but not through seed or pollen (ICTV 2012).

CaMV particles are not enveloped, are spherical and 50 nm in diameter. The host range of CaMV is narrow, with the exception of *Nicotiana clevelandii* and *Datura stramonium*, only
members of the Cruciferae have been reported as hosts. Despite this narrow host range, CaMV is found worldwide due to the global distribution of a number of its aphid vectors (Brunt et al. 1996 onwards; ICTV 2012).

A.1.2 Symptomology

CaMV induces a variety of systemic symptoms in cruciferous crop plants particularly the various cultivars of *Brassica campestris* (e.g. mustards and Chinese cabbage) and *B. oleracea* (e.g. cabbage, cauliflower and broccoli). Symptoms range from chlorosis, streaking and mosaics to necrosis; with mosaic predominating. In *Arabidopsis* observed symptoms include vein clearing, chlorotic spotting, mosaic, stunting and diminished seed set (Melcher et al. 1986).

A.1.3 Replication, transcription and translation

Replication, transcription and translation of the genome of CaMV are complex. The α (minus) strand of DNA, used for transcription, has a discontinuity break of one to two nucleotides (nt). The complementary β (plus) strand has two breaks that have a short overlap of identical sequence to the end of a third segment (γ strand) (Hapiak et al. 2008; Hohn and Rothnie 2013; Shivaprasad et al. 2008). Post-infection, dsDNA moves to the cell nucleus where overlapping sequences are removed, and the single-stranded breaks in the viral DNA are repaired resulting in a circular supercoiled-minichromosome. The genome of CaMV (Figure A.1) encodes seven, possibly eight, proteins from seven ORFs. A host DNA-dependent RNA polymerase transcribes two RNAs: i) 19S that is subsequently translated into large quantities of P6 protein and ii) 35S that encodes and translates all other proteins of the virus (Hapiak et al. 2008; Hohn and Rothnie 2013).
The first protein, P7 is encoded and translated by the 35S RNA; however, the function of P7 remains unknown. The majority of gene VII can be deleted without obvious effects on viral infection and is undetectable in virus-infected plants. P7 undergoes cleavage by P5 (the reverse transcriptase) in vitro, suggesting P7 may be inherently unstable (Hapiak et al. 2008; Lutz et al. 2012; Shivaprasad et al. 2008).

ORF6 encodes a multifunctional protein, P6 that is essential for virus infection. In order to perform its tasks as viral RNA translator and determinant of host range, P6 interacts with both viral and host factors to form electron dense cytoplasmic inclusion bodies. These P6 inclusion bodies are thought to be complexes where virus replication, proteins synthesis and virus particle assembly and accumulation occur. Host plant microtubules are subverted to assist in the development and stability of these complexes and guide their trafficking along the endoplasmic reticulum and through plasmodesmata to neighbouring cells (Niehl et al. 2013). P6 also functions as a translational transactivator (TAV) interacting with host plant ribosome’s to reprogram them to reinitiate translation of 35S RNA and as a consequence, translation of all viral proteins (Hohn and Rothnie 2013). It is not surprising then, given the multitude of functions and interactions of P6, that P6 influences symptom development and severity including the formation of chlorotic symptoms (Haas et al. 2008; Hapiak et al. 2008; Kobayashi and Hohn 2003; Lutz et al. 2012).
Figure AI.1. Schematic representation of the CaMV genome.

mRNAs are indicated by the inner lines and the 3’ end indicated by arrowheads. MP, movement protein; ATF, aphid transmission factor; DB, DNA binding protein; CP, coat protein; RT, reverse transcriptase, TAV, translational transactivator / inclusion body protein; PR, promoter. Adapted from http://en.wikipedia.org/wiki/Cauliflower_mosaic_virus

Based on yeast two-hybrid and maltose binding protein pull down experiments P6 has been shown to interact with the HC/ATF (P2) and P3, a DNA- and dsRNA binding protein (DRB), to stabilise inclusion bodies. P2 bound via P3 to virions also constitutes the transmissible complex that forms in aphid mouthparts (Drucker et al. 2002; Leh et al. 2001; Leh et al. 1999). All three proteins interact with the MP encoded by ORF1 (P1) to establish systemic infection. P1 also facilitates cell-to-cell movement, with assistance from host plant microtubules, by modifying host cell plasmodesmata (Niehl et al. 2013).
A.1.4 VSRs – P6 and 8S RNA

Recently, it has been shown that in CaMV-infected plants, RNAi is counteracted by two types of suppressor: a protein-based suppressor under the control of the TAV protein and an RNA-based suppressor controlled by the viral 8S RNA (Blevins et al. 2011; Hohn 2013; Shivaprasad et al. 2008). TAV acts as a RNA-binding shuttle protein and is thought to transport CaMV 35S RNA from the nucleus to cytoplasm. It has been suggested that while in the nucleus, TAV interferes with the DCL4/DRB4/RDR6 siRNA-producing pathway. As a consequence, dicing of dsRNA is inhibited and viral dsRNAs accumulate (Hohn 2013; Vazquez and Hohn 2013). The 8S RNA is turned into a double-strand form likely by POL II and this gives rise to massive production of siRNAs that are thought to be decoy siRNAs, thus depleting plant AGO proteins (Blevins et al. 2011; Hohn 2013). It is not surprising then, given the multitude of functions and interactions of P6, that P6 influences symptom development and severity including the formation of chlorotic symptoms (Haas et al. 2008; Hapiak et al. 2008; Kobayashi and Hohn 2003; Lutz et al. 2012).
**A.2**

*Tobacco mosaic virus* (TMV)

**Family:** Virgaviridae – proposed 2007  
**Genus:** Tobamovirus  
**Species:** *Tobacco mosaic virus* (Type Member)  
**Acronym:** TMV

### A.2.1 Transmission, host range and particle morphology

TMV is transmitted by means that do not involve a vector including mechanical inoculation, grafting, contact between plants, and in most instances, by seed (Brunt et al. 1996 onwards). TMV is extremely persistent on clothing, glasshouse structures and agricultural equipment and therefore transmitted easily after handling contaminated tobacco products, implements, or infected plants. Soil-borne virus particles or fragments of infected tissue have also been shown to serve as a reservoir of infection by way of plant roots (Agrios 2005).

### A.2.2 Symptomology

TMV causes disease in a wide variety of plant species worldwide, many of which are commercially important. This virus causes serious losses by damaging leaves, flowers and fruit. TMV infection results in varying degrees of mosaic, mottling, chlorosis, curling, puckering, blistering, distortion, elongation and/or dwarfing of leaves, flowers and entire plants. Infection hinders chloroplast development, resulting in stunted plants with symptoms characteristic of the disease, a mosaic of light and dark green between the veins of young leaves followed quickly by development of the same mosaic across the rest of the leaf (Creager et al. 1999).

Tobacco cultivars including *N. tabacum* Samsun NN, Xanthi NN, and *N. glutinosa* and other plants including *Chenopodium amaranticolor* and *Phaseolus vulgaris* encode the tobacco
dominant TMV resistance gene, \( N \). The \( N \)-gene confers a HR and protects plants that encode the \( N \)-gene, and heterologous plants like tomato and \( N. \) \textit{benthamiana} that can be conferred the \( N \)-gene, against systemic spread of TMV (reviewed in Niemeyer et al. 2013). Symptoms in plants encoding, or conferred the \( N \)-gene, are restricted to necrotic lesions at the site/s of initial infection. Resistance proteins guard the plant against pathogens that contain an appropriate elicitor protein through a direct or indirect interaction with this elicitor protein. In the case of TMV the eliciting factor is the C-terminal helicase domain (P50) of the 126 kDa TMV replicase protein. The TMV \( N \)-gene is a member of the TIR-NB-LRR class of resistance genes and encodes two alternatively spliced transcripts, \( N(S) \) and \( N(L) \), both of which are required for resistance. Structural-functional analysis of the \( N \)-gene indicates that the toll interleukin 1 receptor (TIR), nucleotide binding (NB) and leucin-rich-repeat (LRR) domains are all indispensable for \( N \)-gene function (Marathe et al. 2002; Niemeyer et al. 2013)

Symptom development in \textit{Arabidopsis} is dependent on the ecotype infected. Dardick et al. (2000) identified 14 \textit{Arabidopsis} ecotypes displaying varying susceptibilities to TMV. Molecular analysis of all ecotypes indicates that susceptibility is controlled at the level of cell-to-cell virus movement. As an example, the Shahdara ecotype (Sha) supports rapid accumulation and movement of TMV and strong characteristic disease symptoms, while \textit{Col-0} supports intermediate accumulation of TMV with delayed movement and delayed systemic symptoms. These findings were verified by (Serrano et al. 2008) who compared the 14 \textit{Arabidopsis} ecotypes inoculated with the common TMV strain U1 (TMV-U1). The rate of TMV-U1 systemic movement was delayed in ecotype \textit{Col-0} when compared with all other 13 ecotypes. TMV-U1 reached apical leaves after 21 dpi in \textit{Col-0}, whereas TMV-U1 was detected at 9 dpi in the Uk-4 ecotype. Electron microscopy studies following virion movement in stems of \textit{Col-0} infected
plants showed the presence of curved virions, instead of the rigid rods typical of TMV-U1 (18 nm wide by 300 nm in length with a 4 nm inner channel). This was not observed in TMV-U1 infection of Uk-4 plants, where the observed virions are of the typical rigid rod morphology. The presence of defective assembled virions observed by electron microscopy of Col-0 infected plants correlates with delayed systemic movement of TMV-U1 in this ecotype. Incidentally, inclusion bodies within infected cells can be viewed within TMV infected cells in light microscopy and usually present as crystals or amorphous X-bodies containing virions (Brunt et al. 1996 onwards).

Since *Arabidopsis Col-0* and TMV-U1 are the host plant and virus respectively used in the research of this thesis, it was expected that symptoms indicative of TMV would not be observed until at least 21 dpi.

**A.2.3 Replication, transcription and translation**

Tobamoviruses, including TMV, are characterised by a positive sense ssRNA genome. TMV was the first ssRNA virus sequenced, revealing four closely packed ORFs encoding four multifunctional proteins (Figure A.2). All four proteins, the 126 and 183 kDa replication associated proteins, the MP and the CP (encoded through sgRNA of ORF4) interact with many host components to ensure a successful infection (Agrios 2005; Creager et al. 1999; Fauquet et al. 2005; Golemboski et al. 1990; Hull 2002; Liu and Nelson 2013; Zaitlin 1999). The amber stop codon (UAG) for the 126 kDa replicase protein is ‘leaky’ resulting in a second ORF (ORF2) and resultant read-through protein of 183 kDa. The 183 kDa protein also functions as a component of replicase and as such contains an RDR motif.
Like many other subgenomic plant virus genomes the 3’ terminus of the TMV genome has a transfer like RNA (tRNA-like) structure the function of which is to regulate plant virus replication, encapsulate RNA, increase RNA stability and enhance 3’-translation (Dreher 2009).

Figure AI.2. Schematic representation of the TMV genome.

Arrow denotes the amber read-through codon. Protein domains within the 126K protein are indicated (MT, methyltransferase; HEL, helicase). ● represents histidine t-RNA like structure, ○ represents 5’ cap M7G5’ppp5’Gp. Adapted from Zaitlin 1999, Creager et al. 1999; Fauquet et al. 2005.

The replication and assembly of many viruses occurs in an intracellular compartment termed a viral factory, viral inclusion body or a viral replication complex (VRC). Studies with TMV have shown that early TMV infection causes modifications of the filamentous and tubular endoplasmic reticulum (ER) of plants converting it into large irregular aggregates. The ER returns to filamentous at a later stage of infection and these changes are termed the ER transition (Chen et al. 2013). These large aggregates have been shown to contain virus particles; virus-
encoded 183 and 126 replicases, MP, viral RNA, ribosomes and the host elongation factor- EF1α (reviewed in Liu and Nelson 2013). It was postulated that these ER-derived structures are sites of TMV replication i.e. VRCs. How these VRCs are formed is not yet clear (Patarroyo et al. 2013), however, the *Arabidopsis* membrane proteins, tobamovirus multiplication protein 1 (*TOM1*) and ADP- ribosylation like factor (*ARL8*), have been shown to play an essential role in the formation of TMV VRCs and in TMV transmission (Nishikiori et al. 2011). P126 does not contain transmembrane domains but *Arabidopsis* *TOM1* contains transmembrane domains that interact with TMV P126, therefore TMV seems to hijack *TOM1* for anchoring viral replicases to the host plant ER (Chen et al. 2013). In addition, Chen et al. (2013) identified an *Arabidopsis* gene *PAP85* (vicillin-like seed storage protein) with up-regulated expression from 0.5- to 6-hr post TMV inoculation. TMV accumulation was reduced in *pap85*-RNAi mutant *Arabidopsis* and restored to wild-type expression levels when PAP85 was overexpressed in the *pap85*-RNAi plants. Furthermore, Chen et al. (2013) did not observe the ER transition on TMV-infected *PAP85*-knockdown protoplasts. However, *PAP85* only modified the morphology of the ER in the presence of TMV P126. Chen et al. (2013) purported therefore that co-expression of *PAP85* and P126 is required to induce the ER aggregates observed.

TMV-induced X bodies also contain the plant microtubule component β-tubulin and EF1α (Ding 1998). Furthermore plant translation elongation factors 1-alpha and 1-beta (*eEF1α* and *eEF1β*) interact with the 3’ UTR of viral RNA and with the methyltransferase (MET) domain of replication proteins (reviewed in Liu and Nelson 2013 and Patarroyo et al. 2013). *eEF1α* and *eEF1β* are required for TMV infection as evidenced by the reduction of accumulation of viral CP and spread of virus in *N. benthamiana* and capsicum plants where *eEF1α* and *eEF1β* were silenced (Hwang et al. 2013).
VRCs must travel from replication sites to neighbouring cells to start the replication cycle again if they are to initiate a systemic infection. It is generally accepted that plant viruses move from an infected cell to its neighbouring cell through PD (Hull 1989; reviewed in Patarroyo et al. 2013). However plant virus virions are too large to fit through unaltered PD. To facilitate spread of VRCs, most plant viruses encode at least one MP. In TMV the terminal codon of ORF2 overlaps a third ORF (ORF3) that encodes, via sgRNA (I₂ RNA), a 30 kDa movement protein (P30) important for cell-cell and long distance viral spread (Creager et al. 1999). P30 associates with plasmodesmata, dilating the channel and enlarging the size exclusion limit (SEL) thus enabling VRCs to move from cell to cell (Citovsky 1999).

A.2.4. VSR –P126

TMV ORF1 encodes a 126 kDa protein associated with virus replication and movement (P126). Biochemical assays of the TMV replicase have indicated that P126 binds 21 nt sRNA duplexes, interferes with HEN1 methylation of sRNAs resulting in reduced sRNA stability and prevents their incorporation into RISC (Ding et al. 2004; Molnar et al. 2005; Omarov and Scholthof 2012; Wang et al. 2012). From the action of P126 we might expect to see a decrease in sRNA accumulation, however, no significant change in the proportion of sRNAs was seen in response TMV in the present study. This finding might be explained by the observed low accumulation of TMV, compared to the other four viruses, of the present study as noted in Chapter 3.5 (Q = 0.6 compared Q = > 1200).
A.3

*Tomato spotted wilt virus* (TSWV)

Family: Bunyaviridae  
Genus: Tospovirus  
Species: *Tomato spotted wilt virus* (Type species)  
Acronym: TSWV

TSWV is a tospovirus within the arthropod-borne Bunyaviridae. Tospoviruses are unique among Bunyaviridae in that they infect plants where as all other Bunyaviridae infect animals.

A.3.1 Transmission, host range and particle morphology

TSWV has an extremely wide host range due to the global distribution of its vector, the western flower thrip (WFT), *Frankliniella occidentalis* (Brunt et al. 1996 onwards). TSWV particles are spherical and membrane-bound, 80-110 nm in diameter and covered with surface projections comprised of two glycoproteins, Gc from the C-terminal of a cleaved precursor protein (P127) and Gn from the N-terminal of P127 (Ribeiro et al. 2009; Snippe et al. 2007). Membrane-bound particles are an uncommon feature among plant viruses and are thought to reflect the requirement for uptake and replication of the virus in the body of its vector.

A.3.2 Symptomology

TSWV causes significant yield losses in a large number of economically important crops and ornamental plants (Mandal et al. 2008). Disease symptoms range from chlorosis, mottling, spotting, stunting and wilting to severe necrosis. In young infected plants symptoms consist of inward cupping of leaves and leaves develop a purple-bronze cast followed by dark spots. As the infection progresses additional symptoms develop including spotting of the leaves, wilting of the top portion of the plant and dark streaks on the main stem. Fruit may be deformed, often with
raised bumps on the surface, and show uneven ripening (Brunt et al. 1996 onwards). In Arabidopsis, common symptoms include the development of chlorotic spots, leaf curling, purpling of leaves and stems and an inhibition of plant growth (Cândido et al. 2005).

A.3.3 Replication, transcription and translation

TSWV virus particles harbour a tripartite genome of negative/ambisense polarity. The three ssRNA segments are termed small (S), medium (M) and large (L) RNA, according to their size of 2.9, 5.0 and 8.9 kb respectively (Figure A.3). All three ssRNA segments are tightly encapsidated with in a nucleocapsid protein and together with the viral RDR form the infectious ribonucleoprotein (RNP) (Kormelink et al. 2011; Ribeiro et al. 2008; Snippe et al. 2007).

Six functional proteins are derived from the genome. Expression of all TSWV viral genes occurs by the synthesis of sgRNAs and is initiated by a process termed ‘cap-snatching’. Cap-snatching is unique among plant viruses. During this process a nucleotide sequence between 10 and 20 nt is cleaved from the 5’ RNA capped leader of host mRNAs by endonuclease activity of the viral RDR (L RNA). The capped leader is subsequently used to prime transcription of the viral genome, ultimately leading to the synthesis of translatable viral mRNAs (Duijsings et al. 2001; van Knippenberg et al. 2005; Kormelink et al. 2011). From a sgRNA the ambisense S RNA segment encodes a non-structural protein (NSs) in the viral RNA (vRNA) sense (positive) that acts as a viral suppressor of silencing (Takeda et al. 2002), and the nucleocapsid protein in the vcRNA (negative) sense. The ambisense M RNA segment encodes, via sgRNA in the positive vRNA sense, a non-structural protein (NSm) implicated in cell-to-cell movement. In the negative vcRNA sense P127 is encoded and is the precursor protein for the two glycoproteins Gc
and Gn. The negative sense L RNA segment encodes the viral RDR (reviewed in Kormelink et al. 2011).

Figure A1.3. Schematic representation of the three TSWV RNA segments of the tripartite genome.

Small segment (S), encoding the nucleocapsid protein and non-structural small (NSs) viral suppressor of silencing (VSR); medium segment (M) encoding both glycoproteins Gn and Gc and the non-structural medium (NSm) movement protein (MP); and large segment (L) encoding the viral replicase. Adapted from http://viralzone.expasy.org/all_by_species/253.html

Development of reverse genetic screens for TSWV has been greatly impeded by the complexity of the unusual genome organization and replication/transcription strategies of TSWV. This, and the consequent absence of a TSWV infectious clone, has made it to date impossible to investigate the biological functions of the viral proteins within the context of its original three genomes (Zhang et al. 2012).
A.3.4 VSR – NSs

TSWV has a tripartite genome of negative/ambisense polarity where all three ssRNA segments, small (S-RNA), medium (M-RNA) and large (L-RNA) are tightly encapsidated with in a nucleocapsid protein and together with the viral RDR form the infectious ribonucleoprotein (Kormelink et al. 2011; Ribeiro et al. 2009; Snippe et al. 2007). The ambisense S RNA segment encodes, from a subgenomic RNA (sgRNA), a non-structural protein (NSs) in the viral RNA (vRNA) sense (positive) that acts as a VSR (Takeda et al. 2002). Using a GFP-silencing assay involving co-infiltration of *N. benthamiana* leaves with *A. tumefaciens*-GFP and NSs construct, Takeda et al. 2002 showed that expression of a NSs gene construct suppresses accumulation of GFP-siRNAs and NSs has an affinity for long dsRNAs and sRNAs. These data suggest that NSs functions as a VSR by interfering with by sequestering sRNAs thus preventing their loading to respective RISCs (reviewed in Kormelink et al. 2011 and de Ronde et al. 2013). In addition, Lokesh et al. (2010) found that the NSs protein encoded by Groundnut bud necrosis virus (GBNV) contains NTPase and phosphatase activity and speculated that NSs might remove the 5’ phosphate from dsRNA preventing dsRNA cleavage by DCLs into siRNAs and thier subsequent loading to RISCs. We would expect to see a decrease in the proportion of sRNA as a result of the function of NSs however, no significant change was seen in the present study.
A.4

*Turnip mosaic virus* (TuMV)

<table>
<thead>
<tr>
<th>Family:</th>
<th>Potyviridae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genus:</td>
<td>Potyvirus</td>
</tr>
<tr>
<td>Species:</td>
<td><em>Turnip mosaic virus</em></td>
</tr>
<tr>
<td>Acronym:</td>
<td>TuMV</td>
</tr>
</tbody>
</table>

TuMV is a member of the genus Potyvirus of which *Potato virus Y* (PVY) is the type species. The Potyviridae is the largest of all plant virus families recognised, containing at least 200 definitive and several tentative species.

A.4.1 Transmission, host range and particle morphology

TuMV is vectored by numerous species of aphid, especially *Myzus persicae* and *Brevicoryne brassicae*. Virus can be acquired and inoculated in seconds. As with CaMV, transmission of TuMV virus follows a helper strategy in that an HC interacts with both the virus particle and a receptor in the aphid stylet forming a molecular bridge enabling retention of virus in the vector mouthparts and subsequent egestion to a new host plant (reviewed in Ziegler-Graff and Brault 2008). TuMV is readily transmitted by mechanical inoculation but not seed transmitted (Brunt et al. 1996 onwards; ICTV 2012). Particles of TuMV are non-enveloped, filamentous and flexuous with a length of 720 nm, a diameter of 12-15 nm and helical symmetry. All potyviruses induce cytoplasmic inclusions (CI) appearing as pinwheel or scroll-shaped laminated aggregates in sections of infected plant cells. Amorphous inclusions containing aggregates of the virus-encoded protein helper component protease (HC-Pro) are also found in combination with degraded cell organelles and virus particles.
A.4.2 Symptomology

Potyviruses are among the most damaging of plant pathogens causing significant losses in a broad range of agricultural, pastoral, horticultural and ornamental hosts (Fauquet et al. 2005). TuMV was first reported in *Brassica campestris* species in the USA in 1921 (Brunt et al. 1996 onwards) and now has a worldwide distribution (ICTV 2012). In *Brassica nigra* and *B. campestris* ssp. *pekinensis* chlorotic local lesions, mosaic, mottling and puckering or rugosity is observed (Brunt et al. 1996 onwards). TuMV in *Arabidopsis* causes mosaic and necrosis. Mosaic often is accompanied by leaf distortion and stunting, while necrotic symptoms such as veinal necrosis or necrotic spot cause not only arrested plant development but also death of the leaf or lethal necrosis of the entire plant (Kim et al. 2010). TuMV-induced lethal necrosis as seen in the *Arabidopsis*–TuMV interaction has been shown to be determined by the plant host TuMV necrosis-inducing factor (*TuNI*). Necrosis typically starts from the base of the petiole and expands along veins in un-inoculated upper leaves. Because this necrosis is induced by a host factor and virus is localised in tissues in which cells eventually die, the lethal necrotic response is thought to be an active response of *Arabidopsis* to TuMV; namely, an HR rather than a passive damage from TuMV infection.

It is uncertain why a defence response that results in lethal necrosis in the TuMV-infected *Arabidopsis*, and conserved among many *Brassica* species, would benefit plant fitness. Three reasons have been put forward i) veinal necrosis is likely to prevent long-distance movement of the virus; ii) death of diseased plants could contribute to elimination of TuMV inocula from the field if the genotype with lethal necrosis is predominant in a population and iii) larvae of a vector are likely starve to death on plants with veinal necrosis therefore viral transmission could be
expected to decrease. Therefore, a defence system based on PCD against both virus and vector could be very effective in maintaining the fitness of a natural population (Kim et al. 2010).

**A.4.3 Replication, transcription and translation**

The genome of the Potyviridae is comprised of +ve ssRNA and is approximately 10 kb long with a viral protein genome-linked (VPg) 5' terminal to act as a primer during RNA synthesis, and a 3' poly-A tail. The genome consists of a single ORF encoding a large polyprotein (350 kDa) that is subsequently cleaved by three virus-encoded proteases: P1-pro, HC-pro and nuclear inclusion protein a (NIa) to generate 10 mature proteins (Hull 2002; ICTV 2012; Puustinen and Mäkinen 2004). The genome organisation of a typical member of the family indicating the 10 mature proteins and nine cleavage sites (arrowed) is shown in Figure A.4.

![Figure A.4. Schematic representation of the monopartite Potyvirus genome.](http://education.expasy.org/images/Potyvirus_genome.jpg)

Looping arrows represent cleavage sites of P1-pro and HC-pro, ▼ represents cleavage sites of NIa-pro. ■ represents PIPO at 3079-3258 nt position within P3. Adapted from [http://education.expasy.org/images/Potyvirus_genome.jpg](http://education.expasy.org/images/Potyvirus_genome.jpg).

P1-pro corresponds to the N-terminus of the Potyvirus genome and encodes a proteinase that enables its own cleavage from the polyprotein. Deletion and mutation analyses have shown
that P1 is not required for viral infectivity but is involved in enhancing amplification and movement of the virus, in defining host range and in suppression of RNAi. RNA binding activity has been reported for P1 with the same affinity for both dsRNA and ssRNA (Lopez et al. 2008; Urcuqui-Inchima et al. 2001).

HC-pro is a multifunctional protein originally described as an ATF. A modular distribution of functions has been proposed for the protein where the N-terminal region is essential for vector transmission, the central region is involved in suppression of silencing and the C-terminal region has a papain-like cysteine proteinase. HC-pro has also been implicated in genome amplification and long distance virus movement (Urcuqui-Inchima et al. 2001; Lopez et al. 2008).

The potyvirus P3 protein is cleaved from the polyprotein by HC-pro and NIa-pro, resulting in either the P3 protein or its precursor 6K-1. Products of the P3-6K1 region are thought to be involved in virus replication, accumulation, pathogenicity and symptomatology, resistance breaking and cell to cell movement. P3 is a likely determinant of potyvirus host range (Jenner et al. 2003; Lopez et al. 2008; Miyoshi et al. 2008). Chung et al. (2008) provided evidence of a new ORF embedded within P3 between 3079 and 3258 nt (A.4) and named it pipo (Pretty Interesting Potyviridae ORF). Pipo is expressed as a fusion product with the N-terminal of the P3 protein, P3N-PIPO through ribosomal frameshifting or transcriptional slippage at a highly conserved motif. P3N-PIPO has a central role as a viral MP and affects normal functions of RubisCO thus contributing to symptom development (Hillung et al. 2013; Lin et al. 2011; Patarroyo et al. 2013; Vijayapalani et al. 2012).

The CI is the largest potyviral gene product, is cleaved from the polyprotein by NIa-pro, and forms the distinctive pinwheel inclusions. CI exhibits RNA helicase activity and has been
proposed to function in RNA replication. A second, smaller peptide, 6K-2, follows the CI. This product has also been implicated in viral replication and may serve to retain a VRC in virus-induced membrane structures in the cytoplasm. Studies have revealed a role for 6K-2 in both viral movement and in symptom induction (Lopez et al. 2008).

Replication of potyviruses occurs in the cytoplasm of infected cells. Put simply, the virus penetrates a host cell, uncoats and releases viral genomic RNA into the cytoplasm. Viral RNA is translated to produce the polyprotein which is then processed by viral proteases.

The nuclear inclusion protein a (NIa) is subject to suboptimal cleavage resulting in the N-terminal VPg and C-terminal proteinase (NIa-pro) fragment. The VPg is also a pathogenicity determinant and facilitates long-distance movement and subsequent systemic infection (Lopez et al. 2008; Miyoshi et al. 2008). The nuclear inclusion protein b (NIb) is the potyviral RDR, responsible for viral replication and second component of nuclear inclusions and is cleaved by NIa-pro. Recruitment to the replication complex has been postulated to occur via interaction with NIa. SUMOylation is a transient, reversible post-translational protein modification that regulates cellular processes and is catalysed by small ubiquitin-like modifier (SUMO) enzymes. SUMOylation has been implicated in various pathogen-plant host interactions, however very little has been studied regarding its involvement in plant virus infections (reviewed in Xiong and Wang 2013). Xiong and Wang (2013) screened for positive interactions between all potyviral proteins and SCE1, the key SUMOylation enzyme in plants. It was found that the SUMO-conjugating enzyme 1, SCE1, of Arabidopsis (atSCE1) and N. benthamiana (NbSCE1) is an interaction partner of NIb and demonstrated that NIb could be SUMOylated. SCE1-silenced plants exhibited resistance to TuMV suggesting SCE1 is an essential host factor for TuMV infection.
The final cleavage product of NIa-pro is the CP, which in addition to encapsidation of virus particles, has been implicated in vector transmission, virus movement, and symptom development (Lopez-Moya et al. 2008).

A.4.4 VSR – HC-Pro

Virus species in the Potyviridae encode a VSR, HC-Pro that impedes methylation, binds dsRNAs, and inhibits the function of DCLs (Akbergenov et al. 2006; Chapman et al. 2004; Kasschau et al. 2003). It was expected that the action of TuMV HC-Pro would result in increased accumulation of sRNA in the present study. The regression results are in accordance with this assumption where the proportion sRNA of TuMV-inoculated LMW-RNA showed a significant increase compared to the proportion of sRNA from mock-inoculated LMW-RNA (p<0.001).
A.5

*Turnip yellow mosaic virus* (TYMV)

Family: Tymoviridae  
Genus: Tymovirus  
Species: *Turnip yellow mosaic virus*  
Acronym: TYMV

### A.5.1 Transmission, host range and particle morphology

Tymoviruses, including TYMV, have been recorded from tropical, subtropical and temperate regions of the world and only infect dicotyledonous plants. The virus first invades parenchymal tissues, replicates to high titres in infected leaves and eventually invades all main tissues of the host. The formation of double-membrane vesicles at the periphery of chloroplasts is also characteristic of the family (Martelli et al. 2002; Fauquet et al. 2005; Brunt et al. 1996 onwards). All tymoviruses are highly infectious TYMV and vectored by *Pyllotrettra* and *Psylliodes* (flea-beetles), *Phaedon* (mustard beetle) and *Pedilophorus* (pill beetle) species. In contrast to other tymoviruses where numerous plant species are infected, the host range of TYMV is almost entirely confined to the Cruciferae (Martelli et al. 2002; Fauquet et al 2005; Brunt et al. 1996 onwards).

Virus transmission is thought to take place by regurgitation into plant tissue from the foregut of feeding beetles (van der Want and Dijkstra 2006). Although vectored by beetles, TYMV is easily spread in the field by direct contact and by agricultural equipment and is readily transmissible by mechanical inoculation to the leaf and by grafting. However, transmission by seed and pollen is rare (Brunt et al. 1996 onwards; Martelli et al. 2002).
TYMV is a positive ssRNA virus encapsidated to form isometric particles that are non-enveloped and of 28-30 nm in diameter (Canady et al. 1996; Fauquet et al. 2005; Larson et al. 2005; Martelli et al. 2002).

**A.5.2 Symptomology**

The main macroscopic symptoms of TYMV are areas of bright yellow and green mosaic or mottling. The mosaic of TYMV is particularly striking and its development follows a well-defined pattern. Initial systemic symptoms appear as a yellowish vein clearing in indicator plants 7-10 days after inoculation in younger central leaves. In leaves longer than 1-2 cm at the time of systemic infection the full mosaic pattern does not appear with leaves only becoming yellowed (reviewed in Hull 2002). Leaves that are greater than 2 cm long at the time of infection develop numerous dark green islands of tissue. In contrast to all other colours of the mosaic tissue it has been established that the dark green areas (or islands) do not contain any virus or replicative viral RNA. Furthermore, chloroplasts appear normal under microscopic examination (reviewed in Moore and MacDiarmid 2006). The fifth true leaves develop a mosaic consisting of yellow-green and cream areas with numerous scattered dark green islands. This pattern has been shown to remain constant from the earliest stage at which it can be detected until leaf senescence.

Plants infected with TYMV do not recover as several properties are permanently altered by the virus infection. These properties include: a reduction in the rate of carbon fixation and diversion away from sugars into organic and amino acids and a reduction in chloroplast ribosomes and chlorophyll content (reviewed in Hull 2002).
A.5.3 Replication, transcription and translation

The ssRNA genome size of TYMV is 6.319 kb with two overlapping ORFs and a third ORF encoding the CP and replication polyprotein via sgRNA (Figure A.5). ORF1 encodes a 206 kDa protein of conserved sequence with motifs characteristic of methyltransferase (MT), helicase (HEL), papain-like protease (PLpro) and RDR. Toward the 3′-end of this large ORF is a highly conserved 16 nt sequence that functions as a subgenomic RNA (sgRNA) promoter known as the “tymobox”.

Cleavage of 206 K between residues 1259 and 1260 by PLpro leads to the synthesis of an N-terminal product that contains the MT, PRR, PLpro and HEL domains and a C-terminal product encoding an RDR domain (Figure A.5) (Jakubiec et al. 2007; Jakubiec et al. 2004). TYMV has a tRNA-like structure at the 3′ terminus of its genome (Figure A.5).

Viral RNA replication is dependent upon the assembly of a replication complex, comprised of both virus and host components that occurs in small vesicles induced at the periphery of chloroplasts (Jakubiec et al. 2004). These vesicles open into the cytoplasm where they have been shown to contain membrane-bound viral replicase (reviewed in Martelli et al. (2002). The 140K protein appears to be the coordinator of assembly of TYMV VRCs, which are associated with host membrane vesicles present at the chloroplast envelope. The 66K protein with RDR domain has a cytoplasmic distribution when expressed alone and depends of 140K for recruitment to sites of replication (reviewed in Jakubiec et al. 2004).
Figure A1.5. Schematic representation of the *Turnip yellow mosaic virus* genome.

The encoded 206K protein is proteolytically cleaved at a peptide bond denoted by the bidirectional arrow. Protein domains within the 140K protein are indicated (MT, methyltransferase; PRR, proline-rich region; PLpro, papain-like protease; HEL, helicase). • represents tRNA-like valine 3’OH terminal structure, ● represents 5’ cap 5’ M7G. Adapted from (Jakubiec et al. 2004 and Fauquet et al. 2005).

**A.5.4 VSR – P69**

The genome of TYMV has two overlapping ORFs and a third ORF encoding the CP and replication polyprotein via sgRNA. ORF2 initiates 7 nt upstream of ORF1 and encodes a 69 kDa proline-rich region (PPR) protein that influences symptom severity in infected plants and is recognised as a VSR (Cho 2006). P69 suppresses host RNAi by acting upstream of dsRNA, preventing the production of secondary vsiRNAs by *Arabidopsis RDR6* (Chen et al. 2004; Diaz-Pendon et al. 2007; Jakubiec et al. 2007). Interestingly, while P69 suppresses the host RNAi
antiviral defence it promotes RNAi of host genes mediated by miRNAs. It was expected therefore that the action of P69 would result in a decrease in siRNA accumulation and an increase in miRNA accumulation. This is in marginal agreement with one of the findings of the present study where the accumulation of the total sRNA region (both siRNAs and miRNAs) of TYMV-inoculated LMW-RNA was investigated and compared to the proportion of sRNA of mock-inoculated LMW-RNA and borderline significance found (p=0.010).
APPENDIX II - COMPARISONS OF BIOANALYZER ELECTROPHORETIC OUTPUT OF LMW-RNA COMPONENTS

Figure AII.1. Comparisons of Agilent Bioanalyzer 2100 electrophoretic outputs of Mock, CaMV, TMV, TSWV, TuMV and TYMV-inoculated Arabidopsis LMW-RNA.

Three biological replicates of mock, CaMV, TMV, TSWV, TuMV and TYMV-inoculated Arabidopsis LMW-RNA at 21 dpi. L, Ladder corresponding to 4 to 150 nucleotides (nt), 1, 2 and 3 mock-inoculated LMW-RNA biological replicates 1, 2 and 3 respectively; 4, 5, 6, CaMV-inoculated LMW-RNA biological replicates 1, 2 and 3 respectively; 7, 8, 9, TMV-inoculated LMW-RNA biological replicates 1, 2 and 3 respectively; 10, 11, 12, TSWV-inoculated LMW-RNA biological replicates 1, 2 and 3 respectively; 13, 14, 15, TuMV-inoculated LMW-RNA biological replicates 1, 2 and 3 respectively; 16, 17, 18, TYMV-inoculated LMW-RNA biological replicates 1, 2 and 3 respectively. Low molecular weight RNAs (LMW-RNAs), small RNAs (sRNAs; including miRNAs & siRNAs), transfer RNAs (tRNAs), small nucleolar RNAs (snoRNAs) and ribosomal RNAs (rRNAs).
The electrophoretic gel output image of Figure AII.1 suggests differences in the accumulation of all LMW-RNA components when comparing mock-inoculated *Arabidopsis* LMW-RNA with TuMV-inoculated *Arabidopsis* LMW-RNA data. More specifically, the region of dark banding between 10 – 40 nt (sRNAs) of TuMV-inoculated LMW-RNA (lanes 13-15) suggests a higher accumulation than the sRNAs of mock-inoculated LMW-RNA (no bands apparent between 10-40 nt in lanes 1-3). Conversely the rRNAs (c.110 nt) of TuMV (lanes 13-15) appear decreased in accumulation compared to mock-inoculated LMW-RNA (lanes 1-3).

The image shows a slight increase in signal intensity in the 10-40 nt region of TMV and TYMV-inoculated LMW-RNA (lanes 7-9 and 16-18 respectively) compared to mock-inoculated LMW-RNA. In contrast the image shows no to very little difference in the accumulation of the sRNA and tRNA components of CaMV (lanes 4-6) and TSWV (lanes 10-12) LMW-RNA compared to mock-inoculated LMW-RNA (lanes 1-3). Conversely, most rRNAs of CaMV, TMV, TSWV and TYMV-inoculated LMW-RNA appear decreased in accumulation compared to mock-inoculated LMW-RNA in *Arabidopsis*. 
APPENDIX III - PROPORTION OF LMW-RNA COMPONENTS OF TOTAL LMW-RNA

The data in this appendix show results of the proportion of LMW-RNA components of total LMW-RNA as discussed in Chapter 4.3.7 and 4.3.9. Figure AIII.1A illustrates an increase in proportion sRNA accumulation of total LMW-RNA compared to mock-inoculation for CaMV between 7 and 21 dpi, TMV between 14 and 42 dpi, TuMV at 3, 14 and 21 dpi and TYMV at every time point, with the exception of 14 dpi. Throughout the infection there is no discernable change in TSWV sRNA accumulation as a proportion of total LMW-RNA compared to mock-inoculation. Compared with mock inoculation, Figure AIII.1B illustrates an increase in percent tRNA accumulation of total LMW-RNA for CaMV at 3 and 7 dpi and a decrease at 28 dpi, TMV an increase at 2 and 3 dpi, TSWV at 3 dpi, TuMV decreasing tRNA accumulation at 2 and 21 dpi, and for TYMV a decrease in proportion tRNA at every time point. Compared to mock-inoculation, Figure AIII.2A illustrates a decrease in the proportion snoRNA accumulation of total LMW-RNA for CaMV between 3 and 21 dpi, TMV between 2 and 28 dpi with the exception of 7 dpi, TSWV at 3, 21 and 28 dpi, TuMV from 7 to 21 dpi and TYMV at 14 and 28 dpi. Compared to mock inoculation, Figure AIII.2B illustrates a decrease in the proportion rRNA accumulation of total LMW-RNA for CaMV between 2 and 14 dpi, TMV between 2 and 14 dpi and at 28 dpi, TSWV at every time point (except 14 dpi), TuMV between 3 and 21 dpi and TYMV at every time point except 21 dpi.
Figure AIII.1 Proportion LMW-RNA components (sRNA and tRNA) of total LMW-RNA.

A. Proportion sRNA of mock-inoculated, CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. B. Proportion tRNA of mock-inoculated, CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. Error bars are standard error of the mean.
Figure AIII.2 Proportion LMW-RNA components (snoRNA and rRNA) of total LMW-RNA.

A. Proportion snoRNA of mock-inoculated, CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. B. Proportion rRNA of mock-inoculated, CaMV, TMV, TSWV, TuMV and TYMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. Error bars are standard error of the mean.
Figure AIII.3A illustrates an increase in proportion sRNA accumulation of total LMW-RNA compared to mock-inoculation for TamMV at 3, 14, 21 and 28 dpi and a decrease at 7 dpi. Compared to mock-inoculation, Figure AIII.3B illustrates a decrease in tRNAs at every time point dpi, with the exception of 3 dpi. Compared to mock inoculation, Figure AIII.4A illustrates a decrease in proportion snoRNA accumulation of total LMW-RNA for TamMV at every time point dpi. Compared to mock inoculation, Figure AIII.4B illustrates a decrease in proportion rRNA accumulation of total LMW-RNA for TamMV at every time point except 3 dpi.
Figure AIII.3 Proportion of LMW-RNA components (sRNA and tRNA) of mock- and *Tamarillo mosaic virus* (TamMV)-inoculated Tamarillo.

A. Proportion sRNA of mock- and TamMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. B. Proportion tRNA of mock- and TamMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. Error bars are standard error of the mean.
Figure AIII.4 Proportion of LMW-RNA snoRNA and rRNA of mock-and _Tamarillo mosaic virus_ (TamMV)-inoculated Tamarillo.

A. Proportion snoRNA of mock- and TamMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. B. Proportion rRNA of mock- and TamMV-inoculated total LMW-RNA at 2, 3, 7, 14, 21, 28 and 42 dpi. Error bars are standard error of the mean.
Appendix IV - Regressions Log data

----------------------------------------
log: c:/regressions.log
log type: text
opened on: 21 Mar 2011, 21:19:18

. regress srna camv tmv tswv tumv tymv if del==0

Source |       SS       df       MS              Number of obs =
120
-------------+-----------------------------------------------------------------
         9.98 Model |  .395486156     5  .079097231           Prob > F      =
0.0000
         0.0000 Residual |  .903399518   114  .007924557           R-squared     =
0.3045
          0.3045 Total |  1.29888567   119  .010915006           Root MSE      =
  .08902

 srna |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
-------------+---------------------------------------------------------------
    .0908271 camv |    .036405   .0274722     1.33   0.188   -.0180172
    .0709958 tmv |    .016573    .027472     0.60   0.548   -.0378484
    .0591144 tswv |    .004692    .027472     0.17   0.865   -.0497299
    .2424659 tumv |   .1828494   .0300942     6.08   0.000     .123233
    .1264328 tymv |    .072011    .027472     2.62   0.010     .0175885
    .0979576 _cons |    .059475    .019426     3.06   0.003     .0209931

. regress trna camv tmv tswv tumv tymv if del==0

Source |       SS       df       MS              Number of obs =
120
-------------+-----------------------------------------------------------------
         6.81 Model |  .397561876     5  .079512375           Prob > F      =
0.0000

Residual | 1.33014807  114  .011667966           R-squared = 0.2301
----------+-----------------------------------------------------------------------------------
          Total | 1.72770995  119  .014518571           Root MSE = .10802

-----
          trna |        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
          -----+-----------------------------------------------------------------------------------
          camv |  .0600938   .0333352     1.80   0.074   -.005943    .1261305
          tmv |  .0956834   .0333352     2.87   0.005     .0296466   .1617201
          tswv |  .0553268   .0333352     1.66   0.100   -.0107099   .1213636
          tumv | -.1000213   .0365169    -2.74   0.007   -.1723609   -.0276817
          tytmv |  .0188713   .0333352     0.57   0.572    -.0471655   .084908
          _cons |  .6676253   .0235715    28.32   0.000     .6209303   .7143203

-----
. regress snorna  camv tmv tswv tumv if del==0

Source |       SS       df       MS
        Number of obs = 120
----------+-----------------------------------------------------------------------------------
        F(  5,   114) = 2.50
        Prob > F = 0.0345
        R-squared = 0.0988
        Adj R-squared = 0.0593
        Root MSE = 0.04113

-----
          snorna |        Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
          -----+-----------------------------------------------------------------------------------
          camv | -.0176026   .0126929    -1.39   0.168   -.0427471   -.0075419
          tmv | -.0431255   .0126929    -3.40   0.001   -.06827    -.017981
. regress rrna camv tmv tswv tumv tymv if del==0

Source | SS    df  MS
-------|-------+------+
Model  | .036673545  5  .007334709
Residual | .119723378  114  .001050205
Total  | .156396923  119  .00131426

F(  5,   114) = 6.98  Prob > F = 0.0000
R-squared = 0.2345
Adj R-squared = 0.2009
Root MSE = .03241

rrna | Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------|-------+--------+-----+--------+-----------------------+---------+
      |       |       |     |        |                      |        |
camv | -.0450765  .010001  -4.51  0.000  -.0648884  -.0252646
      |       |       |     |        |                      |        |
tmv  |  -.05222   .010001  -5.22  0.000  -.0720319  -.0324082
      |       |       |     |        |                      |        |
tswv |  -.0269629  .010001  -2.70  0.008  -.0467748  -.0071511
      |       |       |     |        |                      |        |
tumv |  -.0170325  .0109555  -1.55  0.123  -.0387353  -.0046703
      |       |       |     |        |                      |        |
tymv |  -.0301954  .010001  -3.02  0.003  -.0500073  -.0103836
      |       |       |     |        |                      |        |
  _cons |   .0807608   .0070718  11.42  0.000  .0667517  .0947699

. gen newscore = srna - snorna - rrna

. sum new

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
</table>
nscore | 126 | -.023872 | .1429268  | -.286107 | .7377186 |
. regress newscore camv tmv tswv tumv tymv if del==0

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 120</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F( 5, 114) = 5.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>.516030791</td>
<td>5</td>
<td>.103206158</td>
<td>Prob &gt; F = 0.0001</td>
</tr>
<tr>
<td>Residual</td>
<td>2.03388696</td>
<td>114</td>
<td>.017841114</td>
<td>R-squared = 0.2024</td>
</tr>
<tr>
<td>Total</td>
<td>2.54991775</td>
<td>119</td>
<td>.02142788</td>
<td>Adj R-squared = 0.1674</td>
</tr>
<tr>
<td></td>
<td>Root MSE = .13357</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| newscore | Coef.   | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|----------|---------|-----------|-----|-----|----------------------|
| camv     | .0990841 | .0412208  | 2.40| 0.018| .017426 .1807421    |
| tmv      | .1119192 | .0412208  | 2.72| 0.008| .0302611 .1935773   |
| tswv     | .0507307 | .0412208  | 1.23| 0.221| -.0309273 .1323888  |
| tumv     | .2247388 | .0451551  | 4.98| 0.000| .1352869 .3141907   |
| tymv     | .1307824 | .0412208  | 3.17| 0.002| .0491243 .2124405   |
| cons     | -.1218483| .0291475  | -4.18| 0.000| -.1795893 -.0641073 |

300
**PRINCIPAL COMPONENTS ANALYSIS**

```
pca srna trna snorna rrna if del==0
```

Principal components/correlation

<table>
<thead>
<tr>
<th>Component</th>
<th>Eigenvalue</th>
<th>Difference</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp1</td>
<td>2.16573</td>
<td>.55585</td>
<td>0.5414</td>
</tr>
<tr>
<td>0.5414</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp2</td>
<td>1.60988</td>
<td>1.44807</td>
<td>0.4025</td>
</tr>
<tr>
<td>0.9439</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp3</td>
<td>.161812</td>
<td>.0992339</td>
<td>0.0405</td>
</tr>
<tr>
<td>0.9844</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comp4</td>
<td>.0625777</td>
<td>.</td>
<td>0.0156</td>
</tr>
<tr>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Principal components (eigenvectors)

```
Variable | Comp1 | Comp2 | Comp3 | Comp4 | Unexplained
---------|-------|-------|-------|-------|---------------
srna     | -0.0832 | -0.7723 | 0.0216 | 0.6294 | 0
trna     | -0.4278 | 0.5972 | 0.0835 | 0.6733 | 0
snorna   | 0.6327 | 0.1794 | -0.6787 | 0.3270 | 0
rrna     | 0.6401 | 0.1215 | 0.7294 | 0.2086 | 0
```

```
predict f1 f2 if del==0
f1 already defined
r(110);
.drop f1 f2
.predict f1 f2 if del==0
(score assumed)
(2 components skipped)
```
Scoring coefficients
sum of squares(column-loading) = 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comp1</th>
<th>Comp2</th>
<th>Comp3</th>
<th>Comp4</th>
</tr>
</thead>
<tbody>
<tr>
<td>srna</td>
<td>-0.0832</td>
<td>-0.7723</td>
<td>0.0216</td>
<td>0.6294</td>
</tr>
<tr>
<td>trna</td>
<td>-0.4278</td>
<td>0.5972</td>
<td>0.0835</td>
<td>0.6733</td>
</tr>
<tr>
<td>snorna</td>
<td>0.6327</td>
<td>0.1794</td>
<td>-0.6787</td>
<td>0.3270</td>
</tr>
<tr>
<td>rrna</td>
<td>0.6401</td>
<td>0.1215</td>
<td>0.7294</td>
<td>0.2086</td>
</tr>
</tbody>
</table>

. regress f1 camv tmv tswv tumv tymv if del==0

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>45.0810851</td>
<td>5</td>
<td>9.01621703</td>
<td>F(5, 114) = 4.83</td>
</tr>
<tr>
<td>Residual</td>
<td>212.640799</td>
<td>114</td>
<td>1.86527017</td>
<td>Prob &gt; F = 0.0005</td>
</tr>
<tr>
<td>Total</td>
<td>257.721884</td>
<td>119</td>
<td>2.16573012</td>
<td>R-squared = 0.1749</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.1387</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 1.3657</td>
</tr>
</tbody>
</table>

| f1          | Coef.     | Std. Err. | t     | P>|t|   | [95% Conf. Interval] |
|-------------|-----------|-----------|-------|-------|---------------------|
| camv        | -1.300897 | .4214793  | -3.09 | 0.003 | -2.135845 -        |
| tmv         | -1.918401 | .4214793  | -4.55 | 0.000 | -2.753348 -        |
| tswv        | -.9608638 | .4214793  | -2.28 | 0.024 | -1.795811 -        |
| tumv        | -.4620798 | .4617074  | -1.00 | 0.319 | -1.376719 -        |
| tymv        | -1.083859 | .4214793  | -2.57 | 0.011 | -1.918806 -        |
| cons        | .9789636  | .2980309  | 3.28  | 0.001 | .3885668            |
|             |           |           |       |       |                     |

. regress f2 camv tmv tswv tumv tymv if del==0
### Summary Statistics

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 120</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F( 5, 114) = 8.58</td>
</tr>
<tr>
<td>Model</td>
<td>52.3809002</td>
<td>5</td>
<td>10.47618</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>139.194884</td>
<td>114</td>
<td>1.22100775</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R-squared = 0.2734</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.2416</td>
</tr>
<tr>
<td>Total</td>
<td>191.575784</td>
<td>119</td>
<td>1.60988054</td>
<td>Root MSE = 1.105</td>
</tr>
</tbody>
</table>

|        | Coef.     | Std. Err. | t    | P>|t| | [95% Conf. Interval] |
|--------|-----------|-----------|------|------|----------------------|
|        |           |           |      |      |                      |
| camv  | -0.1967418 | 0.341008  | -0.58 | 0.565 | -0.8722761           |
| tmv   | -0.005645  | 0.341008  | -0.02 | 0.987 | -0.6811792           |
| tswv  | 0.0685106  | 0.341008  | 0.20  | 0.841 | -0.6070236           |
| tumv  | -2.009572  | 0.373555  | -5.38 | 0.000 | -2.749583 -1.269561  |
| tymv  | -0.6608108 | 0.341008  | -1.94 | 0.055 | -1.336345 -0.0147234 |
| _cons | 0.3902667  | 0.2411291 | 1.62  | 0.108 | -0.0874081           |

### Component Scores

#### Component 1

<table>
<thead>
<tr>
<th>virus</th>
<th>Summary of Scores for component 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>0.97896364</td>
</tr>
<tr>
<td>2</td>
<td>-0.32193369</td>
</tr>
<tr>
<td>3</td>
<td>-0.93943688</td>
</tr>
<tr>
<td>4</td>
<td>0.01809979</td>
</tr>
<tr>
<td>5</td>
<td>0.51688388</td>
</tr>
<tr>
<td>6</td>
<td>-0.10489552</td>
</tr>
<tr>
<td>Total</td>
<td>8.380e-09</td>
</tr>
</tbody>
</table>

#### Component 2

<table>
<thead>
<tr>
<th>virus</th>
<th>Summary of Scores for component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>1</td>
<td>0.3902667</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
</tr>
<tr>
<td>2</td>
<td>.1935249</td>
</tr>
<tr>
<td>3</td>
<td>.38462172</td>
</tr>
<tr>
<td>4</td>
<td>.45877731</td>
</tr>
<tr>
<td>5</td>
<td>-1.6193053</td>
</tr>
<tr>
<td>6</td>
<td>-.27054404</td>
</tr>
</tbody>
</table>

-----------------------------------

| Total | 5.724e-10 | 1.2688107 | 120 | 302 |