

**Identification of reference genes and
quantification of gene expression changes in
Nicotiana glutinosa plants infected by
subgroups I & II of
lettuce necrotic yellows virus**

Anthony Hull

A thesis submitted to
Auckland University of Technology
In partial fulfilment of the requirements for the degree of
Master of Science (MSc)

2019

School of Science

Supervisor: Dr. Colleen Higgins

Abstract

Lettuce necrotic yellows virus (LNYV) is a plant virus that has been reported to cause widespread crop losses in lettuce in Australia and New Zealand for the last 60 years. Phylogenetic analysis has determined two subgroups of the virus exist within the population, identified as subgroup I and II. It appears subgroup II has emerged more recently than subgroup I and currently has a wider geographical distribution, with subgroup I appearing to now be extinct in Australia.

Limited research has been undertaken into understanding the molecular mechanisms by which the virus operates upon establishing infection within a host plant. It is not known whether the two different subgroups influence different molecular pathways which may explain the current distribution of the virus in the environment. This study was designed to determine the expression of four target genes in the host plant *Nicotiana glutinosa* after inoculation by subgroup I and subgroup II of LNYV.

A reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) assay was utilised to determine the relative gene expression of the target genes *CPK3*, *SGS3*, *WRKY26* and *WRKY70* in response to LNYV infection. For a RT-qPCR experiment, a set of validated reference genes are necessary to act as internal controls and need to be specifically selected for studies in a particular host. Currently, no validated reference genes have been reported in the literature for *N. glutinosa*, so candidates were selected to determine their suitability for this purpose; *Actin*, *EF1 α* , *F-BOX*, *L23*, *Ntubc2*, *PDF2*, *PP2A*, *SAND* and *Ubiquitin*. No full genome has been published for *N. glutinosa*, so molecular data from related species had to be obtained to infer the structure of these genes in order to design primers to amplify the genes in a qPCR experiment. Primers could not be designed for *F-BOX*, *L23* and *Ubiquitin*, and non-specific products were amplified during amplification of *WRKY26*, *EF1 α* and *PDF2*. The remaining candidate reference and target gene primers specifically amplified a single product and were considered suitable for testing in the gene expression study.

To obtain sufficient LNYV infected biological replicates for the qPCR experiment, *N. glutinosa* plants were grown from seed and inoculated with LNYV subgroup I or II. Infection rates varied between 0% and 15% for subgroup I and was 26.6% for subgroup II after 28 days of growth. After failing to grow enough replicates to study the virus across six time points, the

experimental design was amended to determine target gene expression after 28 days in both subgroups.

Normalised, outlier removed qPCR data was processed using GeNorm and Normfinder algorithms and the values obtained suggested of the remaining candidate reference genes tested, *SAND* and *Ntubc2* were suitable to be used in subsequent experiments based on their stable expression, though additional reference genes are required, and further biological replicates may be necessary to confirm this.

Using these reference genes and comparing the data of the genes of interest, it was determined that *CPK3*, *SGS3* and *WRKY70* were upregulated between uninfected and subgroup I infected conditions, and *CPK3* and *SGS3* were downregulated between uninfected and subgroup II infected conditions, whilst *WRKY70* was upregulated. Differences were identified between the subgroups, with all three genes being more highly expressed in subgroup I compared to subgroup II with an approximate 7-fold difference in *WRKY70* expression, suggesting that subgroup I isolates may induce transcription and signalling pathways in hosts to a higher degree than subgroup II during infection. Though this was a small scale study it indicates that the biological impact of the different subgroups of the LNYV subgroups may differ which may have influenced the current geographical distribution of the virus. Further research could focus on identifying additional reference genes for qPCR-based studies utilising *N. glutinosa*, or research additional target genes to try and identify additional molecular pathways the subgroups impact.

Keywords: Lettuce necrotic yellows virus, subgroup, gene expression, qPCR

Table of Contents

Abstract.....	3
List of Figures.....	12
List of Tables	23
Attestation of Authorship.....	27
Acknowledgements	28
List of Abbreviations	30
Chapter 1	34
1.1 Introduction.....	35
1.2 Family Rhabdoviridae.....	37
1.3 Genus <i>Cytorhabdovirus</i>	38
1.4 Lettuce necrotic yellows virus	39
1.4.1 LNYV host range	41
1.4.2 LNYV host symptoms	41
1.4.3 LNYV transmission	43
1.4.4 LNYV management	43
1.4.5 LNYV morphology	44
1.4.6 LNYV genome, genes and proteins	45
1.4.7 LNYV detection.....	47
1.4.8 LNYV localization in plant host cells.....	49
1.4.9 LNYV replication in plant hosts	49
1.4.10 LNYV subgroups I and II	51
1.5 Molecular responses in plant hosts to viral infection.....	52
1.5.1 Plant response pathways involving cellular stress	53
1.5.2 Plant response pathway causing developmental defects.....	55
1.6 <i>Nicotiana glutinosa</i>	56
1.7 Molecular methods for analysing gene expression	58
1.7.1 Low to mid-throughput detection techniques	59
1.7.1.1 Northern Blotting	59

1.7.1.2	Western Blotting	60
1.7.1.3	Fluorescent <i>in situ</i> hybridisation (FISH).....	61
1.7.1.4	RNase protection assay (RPA).....	62
1.7.1.5	Polymerase chain reaction (PCR)	62
1.7.1.6	Differential display of mRNA by PCR (DD-PCR).....	64
1.7.1.7	Quantitative real-time reverse transcription polymerase chain reaction (RT-qPCR) 64	
1.7.2	High-throughput techniques.....	67
1.7.2.1	Microarrays	67
1.7.2.2	Serial analysis of gene expression (SAGE)	68
1.7.2.3	High-throughput sequencing (HTS) / next generation sequencing (NGS) ..	69
1.7.2.4	RNA Sequencing (RNA-Seq).....	71
1.8	Summary of literature to determine the direction of the project.....	72
1.9	Aims of the study	73
Chapter 2	75
2.1	Introduction.....	76
2.1.1	Aims.....	78
2.2	Materials and methods	78
2.2.1	Identification of candidate genes of interest from existing literature	78
2.2.2	Identification of candidate reference genes from existing literature	78
2.2.3	Gene expression location for GOI and reference gene candidates in <i>Nicotiana</i> species 79	
2.2.4	Gene ontologies	79
2.2.5	Obtaining mRNA and gene sequences for GOI and reference genes from related plant species.....	79
2.2.6	Multiple sequence alignment of mRNA and gene sequences for GOI and candidate reference genes	80
2.2.7	Comparison of published primers with multiple sequence alignments	80
2.2.8	Primer design for GOI and candidate reference genes	81
2.2.9	Primer testing on LNYV infected and uninfected <i>N. glutinosa</i> leaf material..	82
2.2.9.1	Obtaining uninfected and LNYV infected <i>N. glutinosa</i> leaf material	82
2.2.9.1.1	<i>N. glutinosa</i> plant growth.....	82

2.2.9.1.2	Inoculations and sampling of <i>N. glutinosa</i> with LNYV	82
2.2.9.2	Extraction of total RNA from <i>N. glutinosa</i>	83
2.2.9.3	Gel electrophoresis.....	84
2.2.9.4	End point RT-PCR confirmation of primer amplification using Superscript ® III RT-PCR kit	84
2.2.9.5	Confirmation of correct primer amplification using RT-qPCR.....	85
2.3	Results.....	86
2.3.1	Identification of candidate reference genes	86
2.3.2	Identification of candidate GOIs.....	87
2.3.3	Projected gene expression patterns for GOIs and candidate reference genes in <i>Nicotiana</i> species	92
2.3.4	Gene ontologies for the GOIs and candidate reference genes	94
2.3.5	Obtaining mRNA and gene sequences for all candidate genes from related plant species	97
2.3.6	Multiple sequence alignment of mRNA and gene sequences for candidate reference genes.....	111
2.3.6.1	<i>Actin</i> multiple sequence alignment	112
2.3.6.2	<i>EF1α</i> multiple sequence alignment	113
2.3.6.3	<i>F-BOX</i> multiple sequence alignment	114
2.3.6.4	<i>L23</i> multiple sequence alignment	115
2.3.6.5	<i>Ntubc2</i> multiple sequence alignment	116
2.3.6.6	<i>PDF2</i> multiple sequence alignment.....	117
2.3.6.7	<i>PP2A</i> multiple sequence alignment	118
2.3.6.8	<i>SAND</i> multiple sequence alignment.....	119
2.3.6.9	<i>Ubiquitin</i> multiple sequence alignment	120
2.3.7	Multiple sequence alignment of mRNA and gene sequences for GOIs	121
2.3.7.1	<i>CPK3</i> multiple sequence alignment.....	122
2.3.7.2	<i>SGS3</i> multiple sequence alignment.....	123
2.3.7.3	<i>WRKY26</i> multiple sequence alignment.....	124
2.3.7.4	<i>WRKY70</i> multiple sequence alignment.....	125
2.3.8	Testing of existing primers on multiple sequence alignments.....	126
2.3.9	Primer design for candidate genes	126
2.3.10	Primer testing on LNYV infected and uninfected <i>N. glutinosa</i> leaf material	128

2.3.10.1	RNA concentration and quality assessment.....	128
2.3.10.2	Candidate reference gene primer RT-PCR confirmation at 50°C	128
2.3.10.2.1	<i>Actin</i> primers.....	129
2.3.10.2.2	<i>EF1α</i> primers	130
2.3.10.2.3	<i>Ntubc2</i> primers.....	131
2.3.10.2.4	<i>PDF2</i> primers.....	132
2.3.10.2.5	<i>PP2A</i> primers	133
2.3.10.2.6	<i>SAND</i> primers	134
2.3.10.3	GOIs primer RT-PCR confirmation at 50°C	135
2.3.10.3.1	<i>CPK3</i> primers.....	135
2.3.10.3.2	<i>SGS3</i> primers	136
2.3.10.3.3	<i>WRKY26</i> primers.....	137
2.3.10.3.4	<i>WRKY70</i> primers.....	138
2.3.11	Candidate reference gene primer RT-PCR confirmation at 60°C	139
2.3.12	GOI primer RT-PCR confirmation at 60°C.....	140
2.3.13	Confirmation of candidate reference gene primer amplification using RT-qPCR	141
2.3.14	Optimisation of qPCR experiments using alternative kits.....	143
2.3.15	Optimisation of qPCR experiment preparation using an autopipette	144
2.4	Discussion	146
Chapter 3		151
3.1	Introduction.....	152
3.1.1	Quantification of gene transcripts and impact on experimental design.....	152
3.1.2	Aims.....	154
3.2	Materials and methods	155
3.2.1	Leaf material used to inoculate <i>N. glutinosa</i> with LNYV from infected <i>L. sativa</i>	155
3.2.2	LNYV infection rate study.....	155
3.2.2.1	Inoculation and sampling of <i>N. glutinosa</i> with LNYV.....	155
3.2.2.1.1	Subgroup I.....	155
3.2.2.1.2	Subgroup II	156
3.2.2.2	Confirmation of LNYV infection	156

3.2.3	Two-Step RT-PCR for amplifying LNYV	156
3.2.3.1	cDNA synthesis with qScript™ Flex cDNA Synthesis Kit	156
3.2.3.2	RT-PCR amplification of cDNA with Promega GoTaq® Green Master Mix 157	
3.2.4	Alternative RNA extraction methods to reduce commercial kit use	157
3.2.4.1	CTAB extraction	157
3.2.4.2	CTAB extraction with commercial lysis buffer	158
3.2.4.3	Nucleic acid extraction column reuse	158
3.2.5	qPCR confirmation of candidate reference gene and GOI amplification using qScript™ One-Step SYBR® Green qRT-PCR kit	159
3.2.6	Analysis of RT-qPCR generated data	159
3.3	Results	161
3.3.1	Inoculation conditions	161
3.3.1.1	Establishing conditions for inoculating <i>N. glutinosa</i> with LNYV Subgroup I 161	
3.3.1.1.1	Inoculation with LNYV-Hv28	161
3.3.1.1.2	Inoculation with LNYV-Hv29	164
3.3.1.1.3	Inoculation with LNYV_Hv28-004a	167
3.3.1.1.4	Inoculation with LNYV-Hv14	168
3.3.1.1.5	Summary of subgroup I inoculations	170
3.3.1.2	Establishing conditions for inoculating <i>N. glutinosa</i> with LNYV Subgroup II 171	
3.3.1.2.1	Inoculation with LNYV-Hv19	171
3.3.1.2.2	Summary of subgroup II inoculations	175
3.3.1.3	Infection rates	176
3.3.1.3.1	Subgroup I infection rate study	176
3.3.1.3.2	Subgroup II infection rate study	179
3.3.1.3.3	Change to qPCR experimental design	181
3.3.2	Alternative RNA extraction methods to reduce commercial kit use	182
3.3.2.1	CTAB extraction	182
3.3.2.2	CTAB extraction with commercial lysis buffer	184
3.3.2.3	Nucleic acid extraction column reuse	185
3.3.3	Infection studies for preparing material for use in RT-qPCR studies	188

3.3.3.1	Obtaining RNA from uninfected and subgroup II infected <i>N. glutinosa</i> for qPCR experiment	188
3.3.3.2	Confirmation of infection of Hv19 inoculated <i>N. glutinosa</i> by one step RT-PCR	190
	190
3.3.3.3	Changes to experimental design	191
3.3.4	Analysis of <i>N. glutinosa</i> gene expression following infection by LNYV subgroups I & II	192
3.3.4.1	Experimental design.....	192
3.3.4.2	High resolution melt curve analysis of product amplification.....	193
3.3.4.2.1	HRM analysis of candidate reference genes	194
3.3.4.2.2	HRM analysis of GOIs.....	197
3.3.4.3	PCR amplification efficiencies of genes across all plates	199
3.3.4.4	Outliers.....	200
3.3.4.4.1	Standards.....	200
3.3.4.4.2	Candidate reference gene normalised data	201
3.3.4.4.3	GOI normalised data	204
3.3.4.5	Analysis of candidate reference genes	206
3.3.4.5.1	GeNorm.....	206
3.3.4.5.2	BestKeeper	207
3.3.4.6	Analysis of GOIs.....	208
3.4	Discussion	210
3.4.1	Inoculation of <i>N. glutinosa</i> using infected LNYV samples.....	210
3.4.2	Sampling of LNYV inoculated LNYV samples	211
3.4.3	RNA extraction methods.....	212
3.4.4	RNA integrity analysis using agarose gel electrophoresis.....	213
3.4.5	Infection rate studies	213
3.4.6	Amplification of reference genes and GOI using specifically designed primers in a gene expression assay	214
3.4.6.1	Functional analysis of gene expression changes in reference genes and GOI	214
	214	
3.4.6.1.1	Reference genes	214
3.4.6.1.2	GOI	215

3.5	Summary	217
Chapter 4	219
4.1	Final Discussion.....	220
References	227

List of Figures

Figure 1.1: Representation of the general genomic structure of rhabdoviruses	37
Figure 1.2: Diagram showing distribution of LNYV obtained field samples in Australia and New Zealand from papers sourced for this project. The location, year, and number of samples found at each site are displayed. Red samples denote samples belonging to LNYV subgroup I, and blue samples denote samples belonging to subgroup II (map adapted from information from Dietzgen 2007, Higgins et al, 2016, Ajithkumar, 2018 and Fletcher 2018).	39
Figure 1.3: Images showing infection of LNYV in field and experimental samples. (A) shows infected samples at different stages in the field next to healthy specimens, (B) the variance of infected and uninfected samples in the field, (C) close up of a mildly infected isolate and (D) close up of severely infected isolate (Dietzgen et al. 2007)..	42
Figure 1.4: Images showing exterior and interior representations of the LNYV virion. (adapted from information and images from Dietzgen 2007 and viralzone.expasy.org)	44
Figure 1.5: Representation of genomic structure of LNYV and approximate sizes of each gene within the genome.....	45
Figure 1.6: Representation of genomic structure of LNYV and order from which transcription of the genome occurs. The transcript of the N gene mRNA first makes it the most abundant cell transcript and creates a transcript abundance gradient in the order N>P>4b>M>G>L (modified from Dietzgen et al., 2017).	50
Figure 1.7: Overview of general pathways activated by environmental stressors within a plant host in the field (Atkinson and Urwin 2012). Experimentally, where conditions are controlled and fewer external factors are present at once, specific pathways may be identified as being attributable to a particular stressor (Martinelli et al. 2014).....	52
Figure 1.8: Schematic diagram showing general plant host pathways in response to a viral infection (Whitham et al, 2006)	53
Figure 1.9: Diagram showing the mechanisms of plant host responses to viral infection. Pathway (a) represents antiviral RNA silencing and RSSs, whereby viral dsRNAs or ds hairpin RNAs are processed to viral siRNAs (b) Viral resistance pathway that is triggered by the recognition of avirulence factors that subsequently activate defence	

mechanisms and inhibit downstream virus replication and movement. (c) Local resistance response to prevent aphid colonisation of a plant host (Boualem, 2016) ...53

Figure 1.10: Diagram of the northern blotting process (By Ilewieszoośmiornicach - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=48134046>59

Figure 1.11: Schematic diagram of the western blotting process. Reproduced from (Flint et al. 2015)60

Figure 1.12: Schematic representation of FISH (Image sourced from <http://www.abnova.com>)61

Figure 1.13: Schematic diagram of RPA (Qu and Boutjdir 2007b)62

Figure 1.14: Schematic diagram of the PCR process (Image sourced at <https://www.bosterbio.com/protocol-and-troubleshooting/molecular-biology-principle-pcr>).....63

Figure 1.15: Representation of the cDNA synthesis process. Image sourced at <https://www.thermofisher.com/uk/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcription-applications.html>63

Figure 1.16 Diagram of the DDPCR process (Miura and Scharf 2010).....64

Figure 1.17: Overview of the qPCR process (Image sourced at www.thermofisher.com/)65

Figure 1.18: Schematic diagram representing the microarray process (<https://bitesizebio.com/7206/introduction-to-dna-microarrays/>).....68

Figure 1.19: Diagram outlining the SAGE process (Harkin and Hall 2000).....69

Figure 1.20: Diagrammatic summary of the different NGS processes (<https://genotipia.com/wp-content/uploads/2019/03/NGS.pdf>).....70

Figure 1.21: Diagrammatic representation of RNA-Seq (Haas and Zody 2010)71

Figure 2.1: Expression maps for a) *CPK3* and b) *WRKY26* genes. The red areas indicate higher areas of likely expression for the gene under normal growth conditions.92

Figure 2.2: Expression maps for the candidate reference genes in *N. benthamiana*. The red areas indicate higher areas of expression for the gene a) *Actin* b) *EF1 α* c) *F-BOX* d) *L23* e) *Ntubc2* f) *PDF2* g) *PP2A* h) *SAND* i) *Ubiquitin*93

Figure 2.3: Schematic representation of the *Actin* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers. 112

Figure 2.4: Schematic representation of the *EF1 α* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers. 113

Figure 2.5: Schematic representation of the *F-BOX* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. 114

Figure 2.6: Schematic representation of the *L23* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. 115

Figure 2.7: Schematic representation of the *Ntubc2* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low

homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.116

Figure 2.8: Schematic representation of the *PDF2* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.117

Figure 2.9: Schematic representation of the *PP2A* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.118

Figure 2.10: Schematic representation of the *SAND* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.119

Figure 2.11: Schematic representation of the *Ubiquitin* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and

low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.120

Figure 2.12: Schematic representation of the *CPK3* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.122

Figure 2.13: Schematic representation of the *SGS3* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.123

Figure 2.14: Schematic representation of the *WRKY26* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.124

Figure 2.15: Schematic representation of the *WRKY70* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and

low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.125

Figure 2.16: 1.5% agarose TBE gel confirming amplification of the Actin mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.129

Figure 2.17: 1.5% agarose TBE gel confirming amplification of the *EF1α* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.130

Figure 2.18: 1.5% agarose TBE gel confirming amplification of the *Ntubc2* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.131

Figure 2.19: 1.5% agarose TBE gel confirming amplification of the *PDF2* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.132

Figure 2.20: 1.5% agarose TBE gel confirming amplification of the *PP2A* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.133

Figure 2.21: 1.5% agarose TBE gel confirming amplification of the *SAND* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.134

Figure 2.22: : 1.5% agarose TBE gel confirming amplification of the *CPK3* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.135

Figure 2.23: 1.5% agarose TBE gel confirming amplification of the *SGS3* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.136

Figure 2.24: 1.5% agarose TBE gel confirming amplification of the *WRKY26* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.....137

Figure 2.25: 1.5% agarose TBE gel confirming amplification of the *WRKY70* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature. Positive control for this experiment was the LNYV + sample.....138

Figure 2.26: a) 1.5% agarose TBE gel confirming amplification of the candidate reference genes in an uninfected *N. glutinosa* sample run at 60°C. b) 1.5% agarose TBE gel confirming no amplification in the NTCs run using the same primers run at 60°C. .139

Figure 2.27: a) 1.5% agarose TBE gel confirming amplification of the candidate target genes in an uninfected *N. glutinosa* sample run at 60°C. b) 1.5% agarose TBE gel confirming no amplification in the NTCs run using the same primers run at 60°C.....140

Figure 2.28: Normalised amplification curve testing six Quanta qScript kits (shown in red) in a RT-qPCR reaction against the combination kit previously used (shown in blue). .144

Figure 2.29: Normalised amplification curves for *EF1α* (red and green) and *CPK3* (blue and orange) in a RT-qPCR conducted with an autopipette. Red and green lines indicate uninfected and LNYV infected *N. glutinosa* samples with *EF1α* primer pair, respectively, and blue and orange lines indicate uninfected and LNYV infected *N. glutinosa* samples with *CPK3* primer pair, respectively.....145

Figure 2.30: A summary of the processes the candidate references genes and GOIs were subjected to throughout this part of the study. Arrows indicate where each gene successfully passed onto the next stage.150

Figure 3.1 a) LNYV Hv28 inoculated *N. glutinosa* plant at 28 dpi. Red arrow points to the area that was sampled for testing b) Mock inoculated plant for comparison.162

Figure 3.2 1% agarose gel electrophoresis of total RNA from LNYV Hv28 subgroup I inoculated *N. glutinosa* samples. (M) 100bp DNA Solis Biodyne ladder.162

Figure 3.3 Detection of LNYV-Hv28 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. a) Individual plants sampled at 28

dpi. b) LNYV_Hv19 positive sample and no template control analysed at the same time on a separate gel. (M) 100bp DNA Solis Biodyne ladder.	162
Figure 3.4: 1% agarose gel electrophoresis of total RNA from LNYV Hv19 subgroup II inoculated <i>N. glutinosa</i> samples. M: 100bp DNA Solis Biodyne ladder.....	163
Figure 3.5: Detection of LNYV-Hv28 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Hv28 subgroup I samples and a mock inoculated sample M: 100bp DNA Solis Biodyne ladder, NTC: No template control. LNYV-Hv29 used as positive control.	163
Figure 3.6: Images of asymptomatic Hv28 inoculated <i>N. glutinosa</i> plants that were tested, none of which were infected with LNYV. Red arrow points to area that was sampled at 28 dpi a) Sample 066 b) Sample 071 c) Sample 078 d) Sample 082 e) Sample 085 f) Mock inoculated sample.....	165
Figure 3.7: a) Detection of LNYV-Hv29 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, LNYV-Hv29 used as positive control. b) NTC: No template control sample run in the same RT-PCR reaction but on a separate gel.....	166
Figure 3.8: Detection of LNYV-Hv29 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. LNYV-Hv29 used as positive control.	166
Figure 3.9: a) Detection of LNYV-Hv28_4a by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. b) LNYV-Hv29 used as positive control	167
Figure 3.10: Images of Hv14 inoculated <i>N. glutinosa</i> plants that were tested based on differing symptoms to a mock inoculated plant, with a), b) and d) later confirmed with RT-PCR to be infected with LNYV. Red arrow points to area that was sampled for testing a) Sample 011 b) Sample 019 c) Sample 028 d) Sample 029 e) Mock inoculated plant.	168
Figure 3.11: a) Detection of LNYV-Hv14 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, LNYV-Hv29 used as positive control b) NTC: No template control run in same reaction.....	169

Figure 3.12 a) LNYV Hv19 inoculated *N. glutinosa* plant at 28 dpi. b) Mock inoculated *N. glutinosa* at 28 dpi. Red arrow points to the area that was sampled for testing.....171

Figure 3.13: 1% Agarose gel electrophoresis of total RNA from LNYV Hv19 subgroup II inoculated *N. glutinosa* samples. 100bp DNA Solis Biodyne ladder171

Figure 3.14: Detection of LNYV-Hv19 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. LNYV-H129 used as positive control172

Figure 3.15: Images of Hv19 inoculated *N. glutinosa* plants that were tested based on differing symptoms to a mock inoculated plant, and later confirmed with RT-PCR to be infected with LNYV. Red arrow points to area that was sampled for testing. A) Sample 36 B) Sample 43 C) Sample 54 D) Sample 57 E) Sample 59 F) Sample 60 G) Mock inoculated plant.....173

Figure 3.16 : 1% Agarose gel electrophoresis of total RNA from LNYV Hv19 subgroup II inoculated *N. glutinosa* samples. 100bp DNA Solis Biodyne ladder174

Figure 3.17: a) Detection of LNYV-Hv19 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. a) LNYV-Hv19 used as positive control run in the same RT-PCR on a separate area of the gel.....174

Figure 3.18: Detection of LNYV-Hv14 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Samples in gels A) and C) were tested in the same RT-PCR experiment, as were samples in gels B) and D). M: 100bp DNA Solis Biodyne ladder, NTC: No template control177

Figure 3.19: Images of LNYV Hv14 infected *N. glutinosa* samples confirmed by one step RT-PCR. A) Sample 011, B) Sample 029, C) Sample 019, D) Sample 023. Red arrows point to the areas that were sampled.178

Figure 3.20: Detection of LNYV-Hv19 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Samples in gels A) and C) were run in the same RT-PCR reaction, as were samples in gels B) and D). M: 100bp DNA Solis Biodyne ladder, NTC: No template control.....180

Figure 3.21: Images of LNYV Hv19 infected *N. glutinosa* samples confirmed by one step RT-PCR. Red arrows point to the areas that were sampled.181

Figure 3.22: 1% agarose gel electrophoresis of total RNA from LNYV infected <i>N. glutinosa</i> samples Hv19_034 and Hv19_058 using CTAB extraction. M: 100bp DNA Solis Biodyne ladder.	183
Figure 3.23: Detection of LNYV-Hv19_058b by RT-PCR using EF1 α and CPK3 primers with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control.	183
Figure 3.24: 1% agarose gel electrophoresis of total RNA from LNYV infected <i>N. glutinosa</i> samples Hv19_036, Hv19_056 and Hv19_057 using CTAB extraction with commercial lysis buffer. M: 100bp DNA Solis Biodyne ladder.....	185
Figure 3.25: 1% agarose gel electrophoresis of flow through from cleaned RNA extraction columns run against a previously extracted RNA sample. M: 100bp DNA Solis Biodyne ladder, FC: filtration column, BC: binding column.	186
Figure 3.26: Detection of LNYV-Hv14, Hv18 and Hv33 residual material by RT-PCR in the cleaned binding and filtration nucleic acid columns run with LNYV440F and 118R primers at an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control, FC: filtration column, BC: binding column.....	187
Figure 3.27: Detection of LNYV-Hv19_057 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Samples in gels a) and b) were run in the same RT-PCR reaction, as were samples in gels c) and d), and also e) and f). Gel g) has select samples run with 500 ng template RNA instead of 300 ng M: 100bp DNA Solis Biodyne ladder, NTC: No template control.	190
Figure 3.28: a) <i>Actin</i> , b) <i>EF1α</i> , c) <i>Ntubc2</i> , d) <i>PP2A</i> e) <i>SAND</i> melt curve analysis. Vertical blue lines indicate the presence of an amplicon at a particular temperature.	196
Figure 3.29: Analysis of RT-PCR by agarose gel electrophoresis of EF1 α primer with Hv19 and Hv14 infected samples to confirm amplification of single products. M: 100bp DNA Solis Biodyne ladder, NTC: No template control.....	197
Figure 3.30: a) <i>CPK3</i> b) <i>SGS3</i> c) <i>WRKY70</i> melt curve analysis. Vertical blue lines indicate the presence of an amplicon at a particular temperature.	198
Figure 3.31: Boxplot showing the distribution of Cq values of the standards, a sample identified as Mock_037, run across all plates in the qPCR experiments	200

Figure 3.32: Boxplots showing the distribution of Cq values of a) *Actin*, b) *Ntubc2*, c) *PP2A*, d) *SAND* under different experimental conditions across all plates in the qPCR experiments. Outliers are visible as filled in dots, and were removed from the subsequent analysis.....201

Figure 3.33: Boxplot of averaged Cq values for all candidate reference genes across all experimental samples. Outliers are visible as filled in dots.....203

Figure 3.34: Boxplots showing the distribution of Cq values of a) *CPK3* b) *SGS3* c) *WRKY70* under different experimental conditions across all plates in the qPCR experiments.204

Figure 3.35: GeNorm generated graph showing stability of candidate reference genes from least stable on the left of the graph to most stable on the right.....206

Figure 3.36: GeNorm generated graph of pairwise variation for the candidate reference genes analysed in this research.207

Figure 3.37: Flowchart of processes undertaken in this chapter, the samples grown and utilised and the outcomes of the gene expression studies. “X” denotes studies where no LNYV positive infections were obtained.....218

List of Tables

Table 1.1: Table summarising the features of the aforementioned techniques available to study gene expression (Moustafa and Cross 2016)	73
Table 2.1: Summary of candidate genes identified for this research, the papers they were identified from, their accession numbers, the identified functions of the encoded protein and the plant species that the genes have been studied in so far in response to viral infection.	87
Table 2.2: Summary of gene expression studies showing mRNAs found to increase or decrease in accumulation in response to virus infection. The fold change in accumulation of each mRNA, as measured by either microarray or RT-qPCR, is indicated. Viruses used in this study were <i>Impatiens necrotic spot virus</i> (INSV), <i>Sonchus yellow net virus</i> (SYNV), <i>Turnip vein clearing virus</i> (TVCV), <i>Oilseed rape mosaic virus</i> (ORMV), <i>Potato virus X</i> (PVX), <i>Cucumber mosaic virus</i> (CMV), <i>Turnip mosaic virus</i> (TMV), <i>Tomato ringspot virus</i> (ToRSV), <i>Potato virus Y</i> (PVY), <i>Tomato yellow leaf curl virus</i> (TYLCV), <i>Tobacco mosaic virus</i> (TMV), <i>Tomato spotted wilt virus</i> (TSWV), <i>Cauliflower mosaic virus</i> (CaMV), <i>Turnip mosaic virus</i> (TuMV), <i>Turnip yellow mosaic virus</i> (TYMV). Bold entries are the genes that were selected for analysis in this research.	89
Table 2.3: Summary of cellular location, function and processes of the candidate target genes from gene ontology feature on the QUT database	96
Table 2.4: Summary of cellular location, function and processes of the candidate reference genes from gene ontology feature on the QUT database	96
Table 2.5: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>Actin</i> gene.	98
Table 2.6: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>EF1α</i> gene.	99
Table 2.7: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>F-BOX</i> gene.	100
Table 2.8: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>L23</i> gene.....	101

Table 2.9: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>Ntubc2</i> gene.	102
Table 2.10: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>PDF2</i> gene.	103
Table 2.11: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>PP2A</i> gene.....	104
Table 2.12: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>SAND</i> gene.....	105
Table 2.13: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>Ubiquitin</i> gene..	106
Table 2.14: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>CPK3</i> gene.	107
Table 2.15: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>SGS3</i> gene.....	108
Table 2.16: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>WRKY26</i> gene. .	109
Table 2.17: Summary of gene and mRNA sequences from related <i>Nicotiana</i> species and the original <i>A. thaliana</i> sequence published from the literature for the <i>WRKY70</i> gene. .	110
Table 2.18: Summary of the multiple sequence alignments and information obtained for the candidate reference genes.	111
Table 2.19: Summary of the multiple sequence alignments and information obtained for the GOIs in this research.....	121
Table 2.20: Summary of characteristics of primers designed to amplify candidate genes in the qPCR study.	127
Table 2.21: C _q values for the candidate genes from a one-step SYBR Green RT-qPCR experiment to obtain initial amplification information about the primers. (‘-‘ means amplification was not achieved in the reaction)	142
Table 2.22: C _q values for the candidate genes <i>EF1α</i> and <i>CPK3</i> . Each sample was tested in triplicate.	145

Table 3.1: Names of LNYV isolates and their subgroups as published to date.....	155
Table 3.2: Summary of subgroup 1 inoculation experiments. ‘+’ denotes LNYV-like symptoms were observed during visual inspection or confirmed as being present in an RT-PCR. ‘-’ denotes no symptoms were observed or the RT-PCR did not confirm the presence of the virus in the sample.	170
Table 3.3: Summary of LNYV subgroup II inoculation experiments. ‘+’ denotes LNYV-like symptoms were observed during visual inspection or confirmed as being present in an RT-PCR. ‘-’ denotes no symptoms were observed or the RT-PCR did not confirm the presence of the virus in the sample.	175
Table 3.4: Summary of total RNA purity and concentrations of samples used in the CTAB extraction tests	182
Table 3.5: Summary of total RNA purity and concentrations of samples used in the CTAB extraction tests	184
Table 3.6: Summary of inoculum, sample day, plant ID and whether the sample tested positive for Hv19 by RT-PCR. ‘-ve’ denotes sample tested negative for infection.....	188
Table 3.7: Summary of Hv14 and Hv19 and uninfected leaf material used in RT-qPCR experiments	192
Table 3.8: RT-qPCR plate design for the LNYV gene expression study. NTC refers to the samples with no template for the primers. ‘-’ represents empty wells on the plates.	193
Table 3.9: Table showing the PCR amplification efficiency of all of the genes tested by RT-qPCR in this research.....	199
Table 3.10: BestKeeper generated values for the four candidate reference genes, showing the gene, number of samples, coefficient of correlation and p-value for each.	208
Table 3.11: Average Cq, ΔCq and $\Delta\Delta Cq$ values between reference genes and GOI under different experimental conditions.	209
Table 3.12: Averaged $\Delta\Delta Cq$ values of GOI under different experimental conditions against <i>SAND</i> and <i>Ntubc2</i> reference genes in uninfected and LNYV subgroup I and II infected <i>N. glutinosa</i>	209
Table 3.13: $2^{-\Delta\Delta Cq}$ values for each of the three GOIs against <i>SAND</i> and <i>Ntubc2</i> reference genes in uninfected and LNYV subgroup I and II infected <i>N. glutinosa</i>	209

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Anthony Hull

Acknowledgements

Though this project has one name on the front page, many people are responsible for the planning, undertaking, and support in completing it, all of whom to which I am extremely grateful.

To Dr. Colleen Higgins – thank you so much for giving me the opportunity to work on this project with you. It's been one of the most rewarding collaborations I've had the chance to be a part of. Professionally you always had the right advice at the right time and personally the right story at the right time. Your endless patience, expertise and guidance was appreciated every step of the way and thank you for keeping me on track as I frequently wandered off of it.

To Dr. Gardette Valmonte-Cortes, I really appreciate all of your support and suggestions throughout this research and thesis. You helped me out more times than I can count and always so clearly and thoroughly even when you had lots to do yourself.

To Shweta Shinde and Priya Ajithkumar, I can't thank you enough for teaching me so much in the lab and giving up so much of your time to help me out along the way. Not only are you two of the hardest working people I've met, you're also great people and you made many long days fly by. I wish you both the very best for your futures.

I would sincerely like to thank everyone in our research group; Liz Buckley, Lee Rabbidge, Dr. Brent Seale, Toni Darling, Kendall Mormon and Jules Tauevihi-Davis for being the best

support network you could ask for. You were great both in and out of the weekly meetings, you all made me laugh and I think the scientific world is richer with you being part of it.

I would also like to thank the technicians and staff at AUT who also always went out of their way to help me out throughout this research, particularly Saeedeh Sadooghy-Saraby, Sonya Popoff and Timothy Lawrence.

Lastly, I would like to thank my Mum and Dad, Kelly and Jack Wainwright, my partner in crime Alicia Grimes, as well as the ever supportive Martin Peaches - I know you all had no real idea of what I was doing all of this time but thanks for listening every time I talked about it and for encouraging me to keep going.

List of Abbreviations

aa	amino acid
bp	base pair
°C	degrees celsius
°C/s	degrees celsius per second
g	Gram
mg	milligram
ml	millilitre
mM	millimolar
M	mole
ng	nanogram
nt	nucleotide
µg	microgram
µl	microlitre
µM	micromolar
2-ME	2-mercaptoethanol
AGO	argonaute
ArMV	<i>Arabidopsis mosaic virus</i>
AU	Australia
AUT	Auckland University of Technology
Avr	avirulence
BEAST	Bayesian Evolutionary Analysis Sampling Trees
BLAST	Basic Local Alignment Search Tool
BWYV	<i>Beet western yellows virus</i>
BNYV	<i>Broccoli necrotic yellows virus</i>
CaMV	<i>Cauliflower mosaic virus</i>
cDNA	complementary DNA
cRNA	complementary RNA
CMV	<i>Cucumber mosaic virus</i>
CPK3	coat protein kinase 3
CTAB	Cetyl trimethylammonium bromide

CTV	<i>Citrus tristeza virus</i>
DAS-ELISA	double antibody sandwich enzyme linked immunosorbent assay
dpi	days post infection
DCL	dicer-like
DD-PCR	differential display of mRNA by PCR
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double stranded DNA
dsRNA	double stranded RNA
EDTA	ethylenediaminetetraacetic acid
<i>EF1α</i>	elongation factor 1 α
ELISA	enzyme linked immunosorbent assay
EM	electron microscopy
ER	endoplasmic reticulum
FISH	fluorescent in situ hybridization
G	glycoprotein
Gb	gigabases
GOI	gene of interest
HRM	high resolution melting
HR	hypersensitive response
HSP	heat shock protein
HTS	high throughput sequencing
IDT	Integrated DNA technologies
INSV	<i>Impatiens necrotic spot virus</i>
JA	jasmonic acid
L	polymerase
LBVV	<i>Lettuce big-vein virus</i>
LIVY	<i>Lettuce infectious yellows virus</i>
LMV	<i>Lettuce mosaic virus</i>
LNYV	<i>Lettuce necrotic yellows virus</i>
LYMoV	<i>Lettuce yellow mottle virus</i>
M	matrix protein
MAPK	mitogen-activated protein kinase

MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
MIMV	<i>Maize Iranian mosaic virus</i>
MPI	Ministry of Primary Industries
mRNA	messenger RNA
MUSCLE	multiple sequence comparison by log-expectation
MYSV	<i>Maize yellow striate virus</i>
N	nucleocapsid protein
NCBI	National Centre for Biotechnology Information
NGS	next generation sequencing
NTC	no template control
NZ	New Zealand
OD	optical density
ORMV	<i>Oilseed rape mosaic virus</i>
P	phosphoprotein
PCR	polymerase chain reaction
<i>PDF2</i>	protodermal factor 2
polyA	polyadenylated
<i>PP2A</i>	protein phosphatase 2A
PTGS	post transcriptional gene silencing
PVP	polyvinylpyrrolidone
PVX	<i>Potato virus X</i>
PVY	<i>Potato virus Y</i>
qPCR	quantitative polymerase chain reaction
qRT-PCR	quantitative real-time reverse transcription polymerase chain reaction
QUT	Queensland University of Technology
ROS	reactive oxygen species
R	resistance
rcf	relative centrifugal force
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	RNA interference / RNA silencing
RPA	RNase protection assay

RRSV	<i>Rice ragged stunt virus</i>
RSMV	<i>Rice stripe mosaic virus</i>
RSS	RNA silencing suppressor
RT-PCR	reverse transcriptase polymerase chain reaction
SA	salicylic acid
SAR	systemic acquired resistance
SAGE	serial analysis of gene expression
SCV	<i>Strawberry crinkle virus</i>
SDS-PAGE	sodium dodecyl sulfate-polyacridlamide gel
SGS3	suppressor of gene silencing 3
siRNA	small interfering RNA
SRBSDV	<i>Southern rice black-streaked dwarf virus</i>
ssDNA	single stranded DNA
ssRNA	single stranded RNA
SYNV	<i>Sonchus yellow net virus</i>
T _m	melting temperature
TBE	Tris/Borate/EDTA
TMV	<i>Tobacco mosaic virus</i>
TNV	<i>Tobacco necrosis virus</i>
ToRSV	<i>Tomato ringspot virus</i>
TSWV	<i>Tomato spotted wilt virus</i>
TuMV	<i>Turnip mosaic virus</i>
TVCV	<i>Turnip vein clearing virus</i>
TYLCV	<i>Tomato yellow leaf curl virus</i>
TYMV	<i>Turnip yellow mosaic virus</i>
UTR	untranslated region
UV	ultraviolet
vsRNA	virus derived small RNAs
w.p.i	weeks post inoculation
WYSV	<i>Wheat yellow striate virus</i>

Chapter 1

General Introduction +

Literature Review

1.1 Introduction

Viruses are the most abundant biological entities discovered so far on Earth (Koonin 2010). Nanoscopic in size, they form synergistic or antagonistic relationships intracellularly with living hosts, a requirement for their survival. Virus particles sequester regular cellular processes within organisms to construct new virions, often causing a negative impact on host health (Bao and Roossinck 2013). First identified in a plant host over 200 years ago, their capability of infecting most living organisms causes them to have significant biological and economic impact on humans and agriculture every year (Jones 2014; Martinelli et al. 2014; Sanfacon 2017).

As of 2018, 4852 species of virus have been identified globally, with approximately 1000 recognised as plant viruses, contributing to approximately 47% of all identified plant diseases (Boualem et al. 2016; ICTV 2018). Next generation sequencing (NGS) technology has suggested over 500,000 viruses have yet to be discovered, and with large gaps in knowledge still existing for many plant viruses and their biological operation, it is an area that demands continuing research due to their ongoing impact (Roossinck 2017). Plant viruses belonging to just one genus can result in billions of dollars of damage to crops alone (Kaur et al. 2016). Whilst not all viral infections in agricultural crops are as severe as one other, the rapid nature of virus replication, combined with fluctuating environmental conditions due to human induced climate change, means there is a possibility of a pathogen evolving and developing mechanisms that allow for increased infectivity in the future (Garrett et al. 2016; Jones 2012). Therefore, early recognition and understanding of how a virus operates can confer future benefits for disease management (Jones 2006; Varma 1993).

Lettuce (*Lactuca sativa* L.) is one of the top ten vegetables grown and consumed in New Zealand with 8,600 tonnes produced annually (Horticulture New Zealand 2017). The short shelf life of the crop means it is not a major export, with less than 1% of total New Zealand production being exported to Fiji, Samoa and Hong Kong, but is a staple food source in the New Zealand diet. Recently, lettuce growers in New Zealand have reported seasonal crop collapses with losses as high as 50%, and this has been attributed to viral infection (Fletcher et al. 2017). Viruses infecting lettuce in New Zealand have been studied for decades and those identified in field samples include *Arabidopsis mosaic virus* (ArMV) *Beet western yellows virus*

(BWYV), *Cucumber mosaic virus* (CMV) *Lettuce big-vein virus* (LBVV), *Lettuce mosaic virus* (LMV), *Tobacco necrosis virus* (TNV), *Tomato spotted wilt virus* (TSWV) and *Turnip mosaic virus* (TuMV) (El-Wahab 2012; Fletcher et al. 2005). With the use of existing molecular techniques and developments of new technologies, novel information continues to be gathered in relation to these viruses, though many factors relating to how their biological operation remain unknown and are the sources of ongoing research.

Another virus identified in lettuce in both New Zealand and Australia is *Lettuce necrotic yellows virus* (LNYV), the type species of the genus *Cytorhabdovirus*, part of the *Rhabdoviridae* family of viruses (Dietzgen et al. 2006). Currently, there is comparatively little understanding of the molecular impact LNYV has in the hosts it is able to infect in comparison to other more widely studied plant viruses. Reports from lettuce growers in New Zealand suggest up to 50% of crops can be lost upon infection of lettuce by LNYV; however, a specific economic cost of the impact of the virus has not been published, nor have these losses been solely attributed to the virus (Fletcher et al. 2017; Fletcher et al. 2018). However, it is an opportune time to study LNYV so further information may be elucidated to understand and potentially develop methods to control the pathogen before it becomes more prevalent in Australasia, or develop the ability to spread to other parts of the world.

Sections 1.2 to 1.4.10 will provide an overview of the current understanding of the taxonomy and biological understanding of LNYV to provide background information of the virus, before the more specific aims of this research are discussed.

1.2 Family Rhabdoviridae

Based on obtained data and current methods of analysis, rhabdoviruses are classified into eighteen genera that infect vertebrates, invertebrates and plant hosts (Walker et al. 2018). There is thought to be over 185 different viruses belonging to the *Rhabdoviridae* family with approximately 100 species having been identified as causing systemic infection in plant hosts (Dietzgen et al. 2017; Kormelink et al. 2011). Rhabdoviruses are negative sense, single stranded ribonucleic acid (-ssRNA) viruses with monopartite or bipartite genomes approximately 11-16 kilobases (kb) long and form a bacilliform shape (Dietzgen et al. 2017). They have a common genome structure consisting of five genes that code for five polypeptides; a nucleocapsid protein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G) and an RNA-dependent RNA polymerase protein (L) (Figure 1.1). These genes are flanked by a regulatory 3' leader and 5' trailer, with some viruses having supplementary accessory genes, many of which have an unknown function (Walker et al. 2011). For example, LNYV has an additional gene which codes for a protein identified as the 4b protein. The LNYV genome is discussed in more detail in Section 1.4.6. Rhabdoviral genomes lack 5' caps and 3' polyadenylated (polyA) tails, so are unable to function as messenger RNA (mRNA) templates. They require mRNAs within the host to translate their genome during replication. Though many full genome sequences have been published for rhabdoviruses, the specific mechanisms by which they operate at the molecular level in plant hosts broadly remain to be elucidated (Dietzgen et al. 2017).



Figure 1.1: Representation of the general genomic structure of rhabdoviruses

Phylogenetic analysis of the L gene, the most conserved of the rhabdovirus genes, has indicated, of the eighteen rhabdoviral genera, four infect plants and are transmitted via arthropods, mites, soil or fungi (Walker et al. 2018). It was hypothesised that rhabdoviruses group in a manner consistent with their insect vectors rather than plant species (Brault et al. 2010), although more recent reports don't support this (Mann and Dietzgen 2014). *Cytorhabdovirus*, *Dichorhavirus*, *Nucleorhabdovirus* and *Varicosavirus*, have all been shown

to infect a variety of plant hosts, and have been identified as having unique antigenic properties, genomic organization, replication sites, host ranges, and mechanisms of transmission (Ammar et al. 2009). This range of variance means that information gathered from previous studies into related rhabdoviral genera may not necessarily be applicable to another rhabdovirus genus, and so specific studies are necessary to understand individual viral species further.

Dicorhavirus and varicosaviruses are bipartite plant rhabdoviruses, whereas cytorhabdoviruses and nucleorhabdoviruses are monopartite plant rhabdoviruses (Walker et al. 2018). LNYV replicates in the cytoplasm of infected plant cells, so is classified as a cytorhabdovirus, compared to those that replicate in the nucleus, which are classed as nucleorhabdoviruses (Dietzgen et al. 2007) (Martin et al. 2012).

1.3 Genus *Cytorhabdovirus*

There are 11 virus species within the *Cytorhabdovirus* genera, of which LNYV is the type species (Walker et al. 2018). Demarcation into individual species is based on having a minimum nucleotide divergence of 50% in cognate genes, being able to infect different environments due to their host and vector range, and by being able to be identified individually by serological or nucleic acid hybridization tests (Walker et al. 2018). Different cytorhabdoviruses have individual host ranges and are predominantly transmitted by leafhoppers, aphids and planthoppers, which they can also replicate in (Yang et al. 2017). This project specifically studied the impact LNYV has on plant hosts.

1.4 Lettuce necrotic yellows virus

LNYV was first identified in infected lettuce in Australia in 1954 and was formally classified nearly a decade later (Stubbs and Grogan 1963). Subsequently the virus was discovered in lettuce and sow thistle isolates in the North Island of New Zealand (Fry et al. 1973). Reports of the virus in Spain and Italy also mention crops being infected with LNYV; however, no further experimental work was found relating to this, so the virus is currently thought to be confined to Australia and New Zealand (Ragozzino et al. 1989; Rubio-Huertos and Garcia-Hidalgo 1982). The geographical origin of LNYV remains unclear, but with the current known distribution of the virus to date it is likely to be antipodean (Figure 1.2).



Figure 1.2: Diagram showing distribution of LNYV obtained field samples in Australia and New Zealand from papers sourced for this project. The location, year, and number of samples found at each site are displayed. Red samples denote samples belonging to LNYV subgroup I, and blue samples denote samples belonging to subgroup II (map adapted from information from Dietzgen 2007, Higgins et al, 2016, Ajithkumar, 2018 and Fletcher 2018).

Lettuce necrotic yellows virus is the type species of the genus *Cytorhabdovirus*, part of the *Rhabdoviridae* family (Dietzgen et al. 2006). Host to host transmission for plant rhabdoviruses requires specific insect vectors in order to support the replication of the virus whilst it is in transit. LNYV is principally transmitted in a circular, propagative manner by the blackcurrant sow thistle aphid *Hyperomyzus lactucae* but has also been identified as being spread to a limited degree in *Hyperomyzus carduellinus* (Theob.) and the currant aphid *Nasonovia ribisnigri* (Dietzgen et al. 2007; Fletcher et al. 2017; Randles and Carver 1971). LNYV can be transmitted transovarially between the parent and offspring, with approximately a 20% transmission rate in offspring for up to two generations (Dietzgen et al. 2007). In early reports LNYV had been reported to be the source of extensive crop losses (Stubbs and Grogan 1963; Wetzel et al. 1994), though later reports have identified the virus often being a part of mixed infections with other lettuce infecting viruses. This suggests that it may not exclusively have caused these losses, but was a joint contributor (Fletcher et al. 2005). No studies have yet experimentally determined LNYV as being the sole contributor to field crop losses, with early field reports only hypothesising it was responsible. However, this is not to say this will not be the case in the future if the virus mutates and becomes more virulent, so there is benefit to studying characteristics of a virus that is not yet exclusively detrimental to host health.

Whilst LNYV was first detected over fifty years ago, papers relating to how the virus operates in the field are relatively limited compared to other viruses, possibly due to the lack of samples that have been obtained so far. As of 2018, there are approximately 150 published papers referencing LNYV, though the majority of these are not novel research, instead referencing the virus within a general introduction or discussion for other related experimental work. The majority of experimental work that directly studied LNYV used experimental plant hosts or aphid hosts inoculated with the virus to determine its features under controlled conditions, as field samples are difficult to come by due to diagnostic methods not yet existing to detect the virus *in situ*, and the difficulties associated with mechanically inoculating lettuce with LNYV (Dietzgen et al. 2007). Despite this, LNYV is the most widely studied cytorhabdovirus, and the papers that have focussed on the virus have provided important information that help understand the structure, operation and phylogenetic history of the virus (Jackson et al. 2005).

1.4.1 LNYV host range

LNYV has been identified in field samples and experimentally in single and mixed infection studies as being able to infect monocotyledonous and dicotyledonous plants (Dietzgen et al. 2007). It has been reported that LNYV can infect lettuce species (*L. sativa* and prickly lettuce (*L. serriola*)), sow thistle species (*Sonchus oleraceus* and *S. hydrophilus*) including New Zealand native puha (*S. kirkii*), lupin species (*Lupinus albus* and *L. angustifolius*), garlic (*Allium sativum*), safflower (*Carthamus tinctorius*), chickpea (*Cicer arietinum*), peanut (*Arachis hypogaea*), common marigold (*Calendula officianlis*), jimsonweed (*Datura stramonium*), petunia (*Petunia hybrida*), spinach (*Spinacia oleracera*), globe amaranth (*Gomphrena globose*), tobacco species (*N. benthamiana* and *N. glutinosa*), tomato (*Lycopersicon esculentum*), and false sow thistle (*Reichardia tingitiana*) (Dietzgen et al. 2007; Dietzgen et al. 1989; Fry et al. 1973; Higgins et al. 2016; Martin 1983; Sward 1990).

Lettuce has previously been reported as being the most economically important host of LNYV, though the papers that report this cite no source for this (Dietzgen et al. 2007). Experiments and observations have ascertained that *S. oleraceus* is the main reservoir host of LNYV and determined that its removal from adjacent lettuce crops decreased the incidence of the virus in lettuce (Coutts et al. 2004). The aphids *H. lactucae*, *H. carduellinus* and *N. ribisnigri* do not colonize on lettuce but have been found to probe lettuce species near to *Sonchus* plants and subsequently spread the virus.

Currently, there are very few reports about molecular responses of these plant hosts in response to LNYV infection, thus, it is an area of interest for further study.

1.4.2 LNYV host symptoms

LNYV infection can result in either symptomatic or asymptomatic responses in the aforementioned plant hosts. LNYV infects sow thistle and garlic asymptotically (Sward 1990). In safflower, LNYV infection generates mosaic like symptoms (Irwin and Jackson 1977), while in lupin species, interveinal chlorosis and stunting can be recognised. In chickpea, systemic leaves present bleached and necrotic symptoms (Dietzgen et al. 2007). LNYV causes severe disease in lettuce and causes healthy, shiny green leaves to have stunted, flattened

growth with a loss of green and an increase of pale yellow, often accompanied with necrosis followed by death (Figure 1.3) (Dietzgen et al. 2007; Stubbs and Grogan 1963).

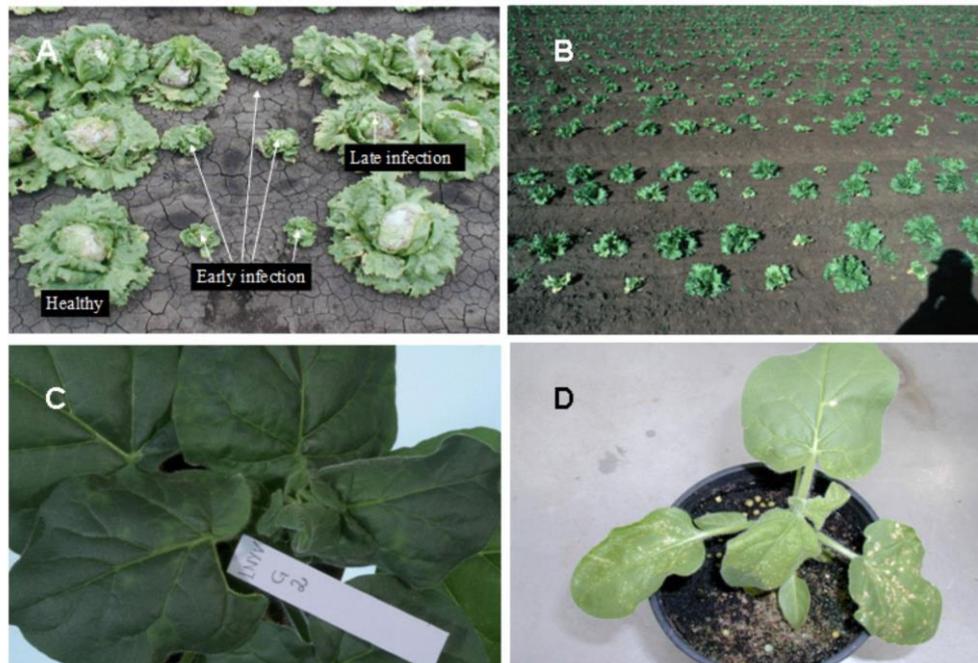


Figure 1.3: Images showing infection of LNYV in field and experimental samples. (A) shows infected samples at different stages in the field next to healthy specimens, (B) the variance of infected and uninfected samples in the field, (C) close up of a mildly infected isolate and (D) close up of severely infected isolate (Dietzgen et al. 2007)

Under experimental conditions utilising *N. glutinosa*, symptoms have been reported to be most prevalent on systemic leaves from as early as 7 days in symptomatic plant hosts. Symptoms include stunted growth, discolouration, downward curled leaves, distortion, mottling, crinkling, and systemic necrosis (Dietzgen et al. 2007; Fry et al. 1973; Randles and Coleman 1970). Severe isolates have been shown to cause necrotic lesions on inoculated leaves (Dietzgen et al. 2007). Experimentally under continuous light conditions, on leaves larger than 3-5cm, symptoms were apparent after 7 days, though other protocols with different light cycles have reported symptoms at 6-8 days. No information relates to field infections as it would be extremely difficult to determine this in the natural environment not knowing the day infection is being established (Dietzgen et al. 1989; Randles and Coleman 1970).

1.4.3 LNYV transmission

LNYV has been shown to be transmissible by mechanical inoculation, aphid transmission and sap inoculation (Stubbs and Grogan 1963). Whilst possible to inoculate many of the hosts mentioned earlier including, *S. oleraceus*, *L. esculentum*, *D. stramonium* and *G. globose*, it is easier to establish infection in some species over others using different techniques (Fry et al. 1973). For example, it is difficult to inoculate the virus into lettuce and sow thistle using sap inoculation, whilst easier to mechanically inoculate into *N. glutinosa*. For this reason, *N. glutinosa* is considered an indicator species for the virus and has been widely used in many experiments studying LNYV (Dietzgen et al. 2007).

1.4.4 LNYV management

Controlling the spread of a viral infection in a field environment can be difficult due to the complexity of insect vector behaviour coupled with the many biotic and abiotic factors present influencing the characteristics of both virus and vector at any given time (Atkinson and Urwin 2012). Under field conditions, reservoirs of LNYV have been identified in *S. oleraceus*, with the aphid *H. lactucae* feeding on infected material and transmitting it to uninfected lettuces as they probe for food sources. *S. oleraceus* grows very close to lettuce, and the eradication of the surrounding sow thistle or keeping it a distance from lettuce has been shown to cause a steep decline in the prevalence of LNYV in lettuce to as low as 4 to 5% (Coutts et al. 2004; Fry et al. 1973). No other techniques are available to halt the spread of the virus, and no studies were found that investigated the treatment of LNYV infected lettuces to stop additional transmission. No LNYV resistant plants have yet been found, though one report misreported a resistant response in *L. saligna*, but the paper it was referencing reported a resistance response to Lettuce infected yellows virus (LIYV), and not LNYV (Lebeda et al. 2013). Genetic modification of lettuce for resistance to LNYV has been attempted, but no definitive resistant lines were obtained and no subsequent research following this up was identified in this literature review (Campbell 2003).

Manipulation or controlling the population of the *H. lactucae* aphid around lettuce crops may be an area of future research, but currently little is known about the molecular mechanisms by which the aphid transmits the virus, as well its role in other biological processes in the field.

As AUT is not currently set up for insect studies, the focus of this project has been on plant host responses and not of the vectors to the virus.

1.4.5 LNYV morphology

In keeping with other members of the Rhabdoviridae family, the LNYV virion is bacilliform, with a monopartite, negative sense, single stranded RNA genome approximately 13 kb in length. The genome is contained in an infectious nucleocapsid core, replicating in the cytoplasm of infected cells (Dietzgen et al. 2007). Electron microscopy (EM) studies suggested individual LNYV virions were approximately 66 nm in diameter and 227 nm in length in *N. glutinosa*, and 56 nm in diameter and 380 nm in length in lettuce, with the difference in size being attributable to the density in which the virus was found in the sections taken, or the manner in which the slides were prepared (Chambers et al. 1965). As shown in

Figure 1.4, the particle consists of several proteins, each of which are discussed in more detail in the following section.

This content has been removed by the author due to copyright issues.

Figure 1.4: Image showing exterior and interior representations of the *Sonchus yellow net virus* (SYNV) virion, which as a cytorhabdovirus would be similar in structure to the LNYV virion. (Jackson et al., 2005)

LNYV genome, genes and proteins

The RNA genome of LNYV (accession no. AJ867584) has 12,807 nucleotides that encode six genes, flanked by two untranslated (UTR) regions consisting of a 3' leader and 5' trailer 84 nucleotides (nt) and 187 nt long respectively (Dietzgen et al. 2007). Each gene is separated by a polyadenylation signal (AUUCUUUU) and a conserved short intergenic region (GNU(C/U)(N)_nACU), both of which have roles in the regulation of transcription and translation. The genes follow the standard structure for rhabdoviruses and encode the five functionally conserved proteins, as well as an additional auxiliary protein, 4b (Dietzgen et al. 2006), referred to as P3 in some later literature (Mann et al. 2016a). The LNYV genome is structured as 3' leader-N-P-4b-M-G-L-5' trailer (Figure 1.5).

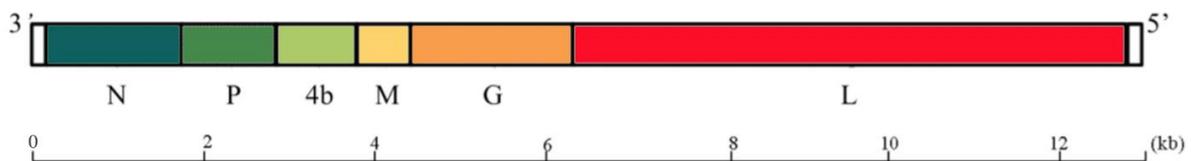


Figure 1.5: Representation of genomic structure of LNYV and approximate sizes of each gene within the genome.

Rhabdovirus genes are transcribed from the 3' -end in decreasing amounts ($N > P > 4b > M > G > L$) (Hull, 2014; Wagner and Jackson, 1997). As gene products at the 3' -end are those that are most highly transcribed, not only does it denote that they are of some importance, but also, they are the most highly abundant in the cell. This is confirmed in LNYV, with the N gene having the most abundant amount of gene product present within infected cells (Dietzgen et al. 1989).

The proteins encoded by the LNYV genes have not all yet had their biological functions explored experimentally yet. Their similarity to other proteins in the rhabdovirus family suggests their properties would be similar, but as the most similar protein shares only 40% nucleotide similarities with other sequenced plant rhabdoviruses, further experimentation would be necessary to elucidate precisely what roles the proteins play in the LNYV life cycle. The information so far understood about the genes and proteins are outlined in the following paragraphs.

The nucleocapsid (N) gene is 1530 nt in length and was the first LNYV gene to be fully sequenced (Wetzel et al. 1994). When expressed, it results in the formation of the nucleoprotein. The nucleoprotein has a molecular weight of approximate 57 kilodaltons (kDa). It has little direct sequence homology when compared to the nucleoprotein of other rhabdoviruses. Phylogenetic analysis of the N genes from eight LNYV infected plant isolates sampled in Australia between 1985 and 2000 suggested that two distinct subgroups, subgroups I and II, of the virus exist based on differences at the nucleotide level (Callaghan and Dietzgen 2005). Infected isolates containing LNYV belonging to subgroup I and subgroup II have also been identified in New Zealand (Higgins et al. 2016); however, no new isolates infected with LNYV belonging to subgroup I have been found in Australia since 1993 and is hypothesised to be extinct there (Dietzgen et al. 2007) (Higgins et al. 2016).

The phosphoprotein (P) gene codes for a putative phosphoprotein and is 1085 nt in length. The P protein has a molecular weight of approximate 38 kDa. Studies have identified it as having a role in nucleocapsid protein-RNA interactions with the L protein. It appears to have a regulatory role in RNA silencing mechanisms in plant hosts in response to viral infection, attaching to argonaute (AGO), RNA dependent RNA polymerase (RDRP) and silencer of gene suppression (SGS) proteins in the hosts (Dietzgen et al. 2007; Mann and Dietzgen 2017; Mann et al. 2016b; Mann et al. 2015)

The 4b gene encodes a movement protein, referred to in some literature as “P3”, responsible for movement of the virus, and is highly similar to protein structures of other plant movement proteins (Mann et al. 2016a). It is 1046 nt in length and is specific to LNYV. Visualisation of the protein has shown it to interact with the M protein (discussed next) in the nucleus of *N. benthamiana*, though the reasons why remain unclear as LNYV replicates in the cytoplasm.

The matrix (M) gene codes for the M protein and is 631 nt in length with a molecular weight of 19 kDa. The LNYV M protein has not been widely studied but, in related rhabdoviruses, it aids in condensing the nucleocapsid as well as associating with lipid bilayers and G protein, suggesting it has a role linking the nucleocapsid and glycoproteins in the viral envelope (Assenberg et al. 2010). It has also been suggested that it is involved in initiating both the change from transcription to replication of rhabdoviral genomes and structural changes in host cells (Dietzgen et al. 2017).

The glycoprotein (G) gene codes for the G protein and is 1836 nt in length. G proteins extend from the rhabdovirus envelope to form a network of glycoprotein spikes (Dietzgen et al. 2007; Dietzgen and Francki 1988). The G protein contains a signal peptide that targets the virion to the endoplasmic reticulum for cleavage and post-translational modification, with both processes thought to have a role in determining viral infectivity (Dietzgen et al. 2007).

The L gene is the RNA dependent RNA polymerase gene and codes for the L protein which is 6332 nt in length and has a molecular weight of 241 kDa. The L proteins have been found to perform RNA synthesis, mRNA capping and have polyadenylation and enzymatic RNA polymerase activity (Assenberg et al. 2010).

1.4.7 LNYV detection

Detection methods for LNYV include visual inspection for symptoms in some plant species such as lettuce, *N. glutinosa*, safflower, lupin species and chickpea but as the virus presents itself asymptotically in other species, such as *S. oleraceus* or garlic, it is not a robust method for all species. Also, known LNYV symptoms share similar characteristics to those caused by tomato spotted wilt virus (TSWV) or lettuce mosaic virus (LMV), and as there is no visual means currently by which to discriminate between these viruses, visual inspection is not necessarily a reliable method, and is also open to interpretation by an individual (Dietzgen et al. 2007).

EM was used in early LNYV studies in an attempt to identify characteristics of the virus particle such as the approximate location, morphology, size and structure of the virus. However, a visual diagnosis such as this is also open to interpretation as it appears similar to other viruses such as *Broccoli necrotic yellows virus* (Chambers et al. 1965; Harrison and Crowley 1965; Toriyama and Peters 1981). The technique was sensitive enough to differentiate between LNYV, TSMV and LMV (Chu and Francki 1982).

Other early researchers tried to apply existing serological diagnostic tests such as utilising gel double-diffusion in agar gel or immunodiffusion to detect LNYV, though the authors noted that either of the results obtained were not reliable enough to use as a routine diagnosis for LNYV infection (Harrison and Crowley 1965; McLean et al. 1971).

A later study developed a technique to detect LNYV using enzyme linked immunosorbent assay (ELISA), which could discriminate the virus within *N. glutinosa* within a laboratory based setting, *S. oleraceus* from field samples and the aphid *H. lactucae* in an experimental glasshouse (Chu and Francki 1982). An adaptation of this method, a double antibody sandwich ELISA (DAS-ELISA) was developed and has been used to diagnose suspected LNYV infected lettuce field samples from Canterbury, New Zealand (Fletcher et al. 2018) Whilst ELISA is a sensitive, fast and robust technique, it has limitations including the expense of generating the antibodies, the need for a large amount of sample and its inability to discriminate between strains of the same virus (Jeong et al. 2014). As a result, it cannot distinguish between subgroups I and II of LNYV.

In the last two decades, the polymerase chain reaction (PCR) has been an important technique for the detection of LNYV, amongst the multitude of other applications it has (Ajithkumar 2018; Dietzgen et al. 2007; Higgins et al. 2016). Generally, the principle of PCR involves denaturation of a dsDNA template through an increase in temperature, annealing of specifically designed primers to a target sequence, followed by the synthesis of complementary strands from free deoxyribonucleotide triphosphates (dNTPs) using the enzyme *Taq* DNA polymerase, as the primers extend in opposing directions within a buffer mixture (Jeong et al. 2014). This cycle of denaturation, annealing and extension is repeated a number of times, anywhere from 20 to above 40, to generate rapid amplification of a DNA target, with the ability to manipulate the components according to the specific experiment. As LNYV requires an intermediate molecule during transcription due to its negative sense RNA genome, a complementary DNA (cDNA) intermediate is required before amplification by PCR. Reverse transcriptase (RT) generates this cDNA before amplification (Martinelli et al. 2014). The whole process can be conducted in one reaction, a one-step RT-PCR, or in a two-step reaction where the cDNA is synthesised first in one reaction, followed by a separate amplification reaction (Wacker and Godard 2005). Whilst one-step reactions reduce the chance of contamination and are quicker to conduct, two-step reactions allow for the cDNA to be utilised for a more diverse range of downstream applications, so the use of a particular technique will depend on the experimental design. The final end product of a reaction is loaded into an agarose gel and using gel electrophoresis, visualized and compared against a DNA ladder to determine if a particular sized product has been obtained.

Protocols have been developed to detect LNYV utilising RT-PCR, initially through the amplification of the entirety of the N gene using a pair of primers identified as BCNG1 and BCNG2 (Callaghan and Dietzgen 2005). Subsequently, a pair of primers (LNYV440F and 1185R) were designed to amplify a portion of the N gene, as shorter fragments are easier to amplify than longer ones (Higgins et al. 2016). More recently primers have been designed that can amplify products specific to subgroups I and II of the virus, which allows individual samples to have their subgroups identified (Ajithkumar 2018). This means virus belonging to the different subgroups is more easily identifiable and obtainable, which allows subgroup specific work to be conducted from isolates obtained.

1.4.8 LNYV localization in plant host cells

Initial studies into the localization of LNYV noted that viral particles were most commonly found in mesophyll cells, epidermal cells, and in high concentrations in small multicellular glandular hairs of plant hosts (Chambers et al. 1965). Later studies utilising EM detected the virus in salivary glands, brain, muscle fat body, mycetomes, ovaries and oesophagus of the insect vector *Hypermyzous lactucae* (Dietzgen et al. 2007).

At the cellular level, studies involving *N. glutinosa* and lettuce demonstrated that LNYV particles detected were restricted to the cytoplasm, though did cluster close to the nucleus (Chambers et al. 1965). Later studies using fluorescent based molecular techniques confirmed the localisation of the individual proteins during the infection process, as well as the interactions they have with one another in a time dependent manner in *N. benthamiana* (Martin et al. 2012). With greater understanding of the pathways that are activated or suppressed in host plants, it may be possible to eventually cross reference the position of the different proteins and attribute their presence as the causative factor of these events.

1.4.9 LNYV replication in plant hosts

Experimentally there have been few studies that have looked at the specific mechanisms by which LNYV replicates in both plant and insect hosts. However, from other studies into other rhabdoviruses including vesiculoviruses and lyssaviruses, many conserved and similar pathways have been determined, so it is thought that replication mechanisms are broadly the

same across the family (Dietzgen et al. 2017; Walker et al. 2018). Being a cytorhabdovirus, LNYV replicates in the cytoplasm of cells in association with inclusion bodies or networks of proteins, known as viroplasm (Dietzgen et al. 2007). The model of cytoplasmic replication for rhabdoviruses is broadly split into five stages; cell entry, uncoating of the coat protein, transcription and translation, replication of the genome followed by encapsidation, and finally assembly and release of new virion particles.

Virus entry into the cell occurs via penetration by an aphid vector or mechanical inoculation into uninfected plant material. Upon entry, the G and M protein dissociate from the outer region of the virion particle away from the nucleocapsid, which allows for transcription to occur through interaction with the endoplasmic reticulum (ER) membrane. The L protein initiates transcription to synthesise the intermediate positive sense mRNA transcripts for each open reading frame (ORF), generating transcripts for each gene in the order mentioned in Section 1.4.6 and generating a transcript gradient as represented in Figure 1.6.

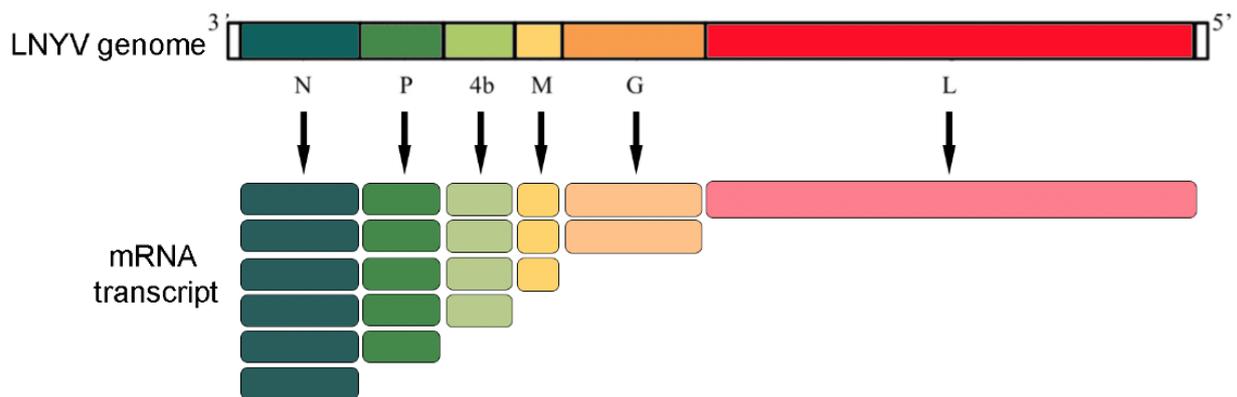


Figure 1.6: Representation of genomic structure of LNYV and order from which transcription of the genome occurs. The transcript of the N gene mRNA first makes it the most abundant cell transcript and creates a transcript abundance gradient in the order N>P>4b>M>G>L (modified from Dietzgen et al., 2017).

Translation of the mRNA transcripts occurs within viroplasm, which form as stable thread-like structures at multiple sites, likely near the endoplasmic reticulum (ER), in the cytoplasm of infected cells (Martin et al, 2012; Dietzgen et al, 2017; Mann and Dietzgen, 2014). The synthesized viral proteins and mature virion particles are subsequently transported to different intracellular locations after budding from the ER membranes (Martin et al, 2012).

1.4.10 LNYV subgroups I and II

RNA viruses replicate rapidly, have large population sizes and are prone to high rates of error (Elena et al. 2011). As a result, an organism infected with a virus is likely to have a multitude of virus particles with varying nucleotide sequences, forming a cloud population or quasi species (Sanfacon 2017). These intra-variant populations can give rise to competitive conditions and so a selective advantage is therefore potentially able to develop which may lead to differences arising in the same virus species (Garbutt et al. 2011; Syller 2012; Syller and Grupa 2016).

Callaghan and Dietzgen (2005) initially analysed the N gene of LNYV utilising isolates from Australia to determine the phylogenetic history of the virus. Their findings suggested that LNYV is made up of two distinct subgroups. Using further isolates collected from Australia and additional samples from NZ, phylogenetic and BEAST analysis confirmed separation of the isolates into the two subgroups based on variability in the N gene (Higgins et al. 2016). The last common ancestor for all samples was placed approximately 500 years ago, with subgroup I emerging approximately 150 years ago and subgroup II 75 years ago. With the suspected extinction of subgroup I from Australia, with no new isolates being found since 1993, and subgroup II's rapid dispersal, it has been hypothesised that subgroup II has developed a transmission or replication based advantage, within either the aphid or plant host (Higgins et al. 2016). Further sampling both countries and outside of it may provide further insight into the phylogenetic history of the different subgroups. Recently, further samples belonging to subgroup I and subgroup II have been collected from Canterbury in New Zealand, which are the first LNYV-positive samples obtained from the South Island (Ajithkumar 2018).

Viruses belonging to both subgroups have been individually identified in lettuce using molecular techniques (Ajithkumar 2018; Callaghan and Dietzgen 2005; Higgins et al. 2016). Experimental work mentioned in earlier sections involving the impact of LNYV on plant hosts did not identify the subgroup the inoculum belonged to. Whilst the findings of these studies have determined features of the virus as a whole, as yet, no individual discrimination of the impacts the individual subgroups have on plant hosts have been identified. By determining experimentally what differences there are in plant host responses to the different subgroups, it may provide further insight into the mechanisms by which the two subgroups of the virus possibly operate which has led to the more rapid radiation of subgroup II.

Within field infections, it is more common for there to be a population of viruses within a host rather than a single infection by a single virus (Ben-Ami et al. 2008; Ben-Ami and Routtu 2013; Hily et al. 2016; Salvaudon et al. 2013) (Tollenaere et al. 2016). So far, no mixed infections containing LNYV subgroup I and subgroup II have been identified in the literature, however a contributing factor of this could be the lack of samples obtained so far for the virus and the relatively new diagnostic techniques having not been applied yet. To determine the impacts of a particular subgroup on a plant host, it will be necessary to inoculate uninfected plant material with virus from each subgroup individually and examine them under controlled conditions.

1.5 Molecular responses in plant hosts to viral infection

To carry out such a plant host study, it is important to understand how viruses tend to impact other plant hosts, consider the most appropriate experimental design for an LNYV based study and review current experimental methods that can be used to detect and analyses molecular changes in plant hosts in response to viral infection.

A viral infection is an active biological process on the part of both virus and organism it infects, and this is no different with plant viruses in plant hosts. The establishment of an infection is not a straightforward process, with a series of pathways and responses being activated and coordinated in both virus and host to either establish or prevent infection. Many of these pathways and responses are impacted by abiotic and biotic factors, including whether the pathogen is entering a host that already has established viral populations within (Elena et al. 2014). The complex network of interactions between different viral and plant species leads to a broad range of symptomatic and asymptomatic responses (Figure 1.7) being seen in established infections in nature, up to and including host cell death (Alexander and Cilia 2016).

This content has been removed by the author due to copyright issues.

Figure 1.7: Overview of general pathways activated by environmental stressors within a plant host in the field (Atkinson and Urwin 2012). Experimentally, where conditions are controlled and fewer external factors are present at once, specific pathways may be identified as being attributable to a particular stressor (Martinelli et al. 2014).

The number of possible outcomes to review is beyond the scope of this study, however, a general set of responses have been identified in the literature that occur in many plant hosts in response to viral infection (Figure 1.8). The specific pathways that elicit these responses are likely to be unique for a particular virus-host interaction, but identification of general pathways allow for more targeted research to be conducted. Split into general cellular stress responses and responses that cause developmental defects, these general responses can help form the framework of what particular areas to focus on in this study.

This content has been removed by the author due to copyright issues.

Figure 1.8: Schematic diagram showing general plant host pathways in response to a viral infection (Whitham et al, 2006)

1.5.1 Plant response pathways involving cellular stress

At the area and surrounding areas where a virus is introduced into a host, a hypersensitive response (HR) has been identified where apoptotic pathways are activated via the upregulation of defence and resistance (R) genes. This is known as the “incompatible reaction” between pathogen and host. R genes within the plant host pathogen factors encoded by avirulence (Avr) genes, triggering mitogen-activated protein kinase (MAPK) signalling cascades (Figure 1.9b). This causes a downstream increase in levels of reactive oxygen species (ROS), defence hormones, salicylic acid (SA) and jasmonic acid (JA), triggering apoptotic signalling cascade

This content has been removed by the author due to copyright issues.

Figure 1.9: Diagram showing the mechanisms of plant host responses to viral infection. Pathway (a) represents antiviral RNA silencing and RSSs, whereby viral dsRNAs or ds hairpin RNAs are processed to viral siRNAs (b) Viral resistance pathway that is triggered by the recognition of avirulence factors that subsequently activate defence mechanisms and inhibit downstream virus replication and movement. (c) Local resistance response to prevent aphid colonisation of a plant host (Boualem, 2016)

pathways, resulting in necrotic lesions around the site of initial infection (Boualem et al. 2016; Caarls et al. 2015; Hernández et al. 2016). This response is a form of resistance that occurs between particular pathogens and particular viruses, but not between all plant-virus combinations. It does not provide immunity to a plant host, but is a rapid response to prevent the replication of a virus particle upon initially entering a host (Syller and Grupa 2016). There has been little research of the resistance responses to LNYV and cytorhabdoviruses in plant hosts, particularly of R genes in particular. In general, the mechanisms by which the hypersensitive response occurs is poorly understood (Senthil et al. 2005; Wang et al. 2015) . Therefore, this area does not seem a suitable area to explore in this project given the lack of existing experimental work to design subsequent research on, though certainly may be an interesting area of future research.

The increase in both ROS and SA as part of the HR does, however, generate a systemic acquired resistance (SAR) response that causes future pathogenic attacks to be less effective (Hernández et al. 2016). SAR is similar to the innate immune system of animals, though is not antigenic in its basis (Muthamilarasan and Prasad 2013). The relative expression of the PR-1 gene has been identified as a useful and stable molecular marker for activation of this pathway and has been used in studies of different virus-host interactions (Chavez-Calvillo et al. 2016). However, the full pathway is still not fully understood and is not a response unique to a viral infection, being though to be activated by several non-pathogenic factors as well (Muthamilarasan and Prasad 2013). No studies were identified in the literature describing the SAR response in plants infected by rhabdoviruses, suggesting another area for future research.

An increase in ROS in plant hosts is also a hallmark of plant-virus interactions, usually being found at a baseline level during regular cellular processes as they are components of several signalling pathways and constantly produced by aerobic processes in organelles (Jiang et al. 2017). However, elevated ROS levels are not exclusively a result of a virus infection, having been associated with other biotic and abiotic stressors such as excess light, UV radiation, extreme heat, lack of water, hypersalinity and mechanical stress (Khraiweh et al. 2012). Therefore, they are not suitable candidates for examining specific plant host responses to viral infection.

Endogenous molecules with a regular biological functioning will also alter in amount due to biotic or abiotic responses (Cramer et al. 2011; Khraiweh et al. 2012) Though present

throughout cells under regular conditions, but in lower amounts, chaperone heat shock proteins (HSPs) are induced to a much higher degree in response to environmental, physical and chemical stresses, including viruses (Liu et al. 2012; Senthil et al. 2005). However, specific HSPs will be induced to particular stressors in particular species, so without conducting a microarray or transcriptomic study on a particular host infected with a particular pathogen, it may not be easy to determine which proteins to study for research in new host species. Also, whilst studies have noted that the upregulation of HSPs could be viable indicators for general plant health, they may not necessarily be solely attributable to a virus infection.

1.5.2 Plant response pathway causing developmental defects

RNA silencing or RNAi interference (RNAi) or post transcriptional gene silencing (PTGS) is a conserved mechanism in eukaryotes activated by the recognition of double stranded (ds) RNA molecules. RNAi is a process whereby protein coding genes in a host are silenced on a short-term basis, with small interfering RNA (siRNA) molecules approximately 20-25 nt in length specifically targeting existing mRNA transcripts from the genes and degrading them before translation, rendering them useless to invading pathogens (Bivalkar-Mehla et al. 2011; Incarbone and Dunoyer 2013).

To counteract the RNA silencing mechanisms that plants have developed, plant viruses including rhabdoviruses have evolved RNA silencing suppressors (RSSs), proteins that prevent the silencing mechanisms in the plant host (Figure 1.9a). The phosphoprotein (P) protein of LNYV has been identified as an RSS that does not impact siRNA accumulation, instead interacting with multiple plant RNA silencing machinery proteins to inhibit RNA-induced silencing complex (RISC) cleavage, which can ultimately silence a whole viral genome (Mann et al. 2016b; Mann et al. 2015; Boualem et al. 2016)). Specifically, the LNYV RSS's interact with proteins AGO 1, 2 and 4, RDR6 and SGS3. Therefore, one of these genes could be a focus of a plant host response study to determine gene expression activity under non-infected and infected conditions.

The most widely studied responses to a viral infection are the interference of the accumulation or function of host proteins, nucleic acids or aforementioned defence mechanisms (Havelda et al. 2008). Many early studies focused on the global profiling of mRNA transcripts in uninfected

and virus infected plant hosts to determine what particular genes are downregulated as a response to viral infection (Whitham et al. 2003; Whitham et al. 2006). With development in technologies such as microarrays, it is more widely understood that a spectrum of host-specific genes in plants are either up or downregulated in response to viral infection. These generally appear specific to a particular virus-host interaction, though common sets of genes have been found to be impacted in related species (Lilly et al. 2011; Liu et al. 2012; Senthil et al. 2005; Whitham et al. 2003). This information can be utilised to possibly select candidate genes to look at in a limited study in similar plant species, such as *N. glutinosa* in response to LNYV infection, which is likely to have similar responses to previously studied plants such as the related *N. benthamiana* or *Arabidopsis thaliana*, both widely studied model species.

1.6 *Nicotiana glutinosa*

Tobacco plants (*Nicotiana* spp.) are widely used model plant organisms that have been used in multiple studies of biological processes and plant diseases. *N. benthamiana* is the most common plant host utilised for plant virology studies with its ability to be infected by many viruses as a result of a mutation in an RNA-dependent RNA polymerase gene (Goralski et al. 2016; Huang et al. 2012). However, limited studies exist related to the infection of *N. benthamiana* by LNYV (Mann et al. 2016b; Martin et al. 2012). Over time, the majority of LNYV studies have utilised *N. glutinosa*, as the plant host for their studies. *N. glutinosa* has been identified as a good differential host and indicator species for LNYV, as it is easily infected by the virus and has the ability to be used as inoculum for infecting other hosts (Dietzgen et al. 2007). *N. glutinosa* has been utilised as a host to study symptomology, to propagate and maintain LNYV, with it, and naturally infected lettuce, being a robust source from which to purify the virus (Dietzgen et al. 1989). It is easy to infect and is susceptible to a wide range of viruses. It was utilised when the structure of the LNYV virus particle was determined (Harrison and Crowley 1965), and more recently for cloning and sequencing the viral genome (Dietzgen et al. 2006; Dietzgen et al. 1989). As there is more material related to the use of this host plant for studying LNYV, it seems suitable to continue to utilise *N. glutinosa* for further work, though it would be interesting in the future to determine whether the findings from these experiments are applicable to other host species.

With plant-virus interaction studies, findings from experimental work using model species cannot necessarily be applied to the host that a virus may infect in the field, or other model species. Plant-virus interactions have been found to be highly specific between a particular virus in a particular host (Bose et al. 2016; Hily et al. 2016; Syller 2012; Syller and Grupa 2016) . As a result, conclusions and findings obtained from model systems under controlled conditions in experimental research need to be confirmed in the complex systems before fully drawing conclusions (Sanfacon 2017)

There is currently no fully published genome for *N. glutinosa*, so it is unknown how many genes the species has, nor the size of the genome. Other *Nicotiana* species have had their genomes published; *N. tabacum* has a genome approximately 4.5 gigabases (Gb) in size, the *N. benthamiana* genome is approximately 5.1Gb, *N. sylvestris* has a genome of approximately 4.4Gb, and the *N. tomentosiformis* genome is 4.1Gb (Edwards et al. 2017). Within these variable sized genomes, the number of genes found varies between 36,509 genes in *N. tomentosiformis* and 69,500 in *N. tabacum*. It is unknown if *N. glutinosa* is similar to any of these related species in terms of genome size or number of genes without fully sequencing the genome. As of 2018, only 400 published mRNA and gene sequences from *N. glutinosa* had been published on the NCBI database, so comparatively little information exists for genes and mRNA transcripts from this species compared to other related species.

As field samples are difficult to come by without knowing likely sources of the virus and as existing samples remain scarce, experimentally infecting a host plant under controlled conditions is a good way to conduct more studies. The use of *N. glutinosa* as a model to study plant host responses to LNYV is valid since it has been shown to be a suitable model host. There is no fully published genome for *N. glutinosa*, and so to conduct plant host response studies using molecular techniques, comparison of mRNA sequences from related species must be used in order to design primers to amplify regions of interest.

1.7 Molecular methods for analysing gene expression

For plant host-virus interactions, many diverse responses within plants have been identified over time, ranging from morphological and developmental through to changes at the molecular level (Section 1.5). Technologies have been developed to studies these different responses, and the selection of an appropriate method is important in order to obtain and process relevant data to the question being studied. The study of changes in gene expression, the variation in levels of functional gene product, can provide extensive information related to specific pathways that may be activated or suppressed in a host by the presence of a particular pathogen in a viral infection, though differential changes in gene expression does not necessarily mean there will be an observable difference in related protein accumulation. However, the information that can be obtained from the expression changes in a host plant can elucidate functional information about the processes by which a virus replicates and transmits itself from host to host. These findings could be utilised to development specific management tools to help control or prevent the spread of the virus in the future if the pathways the virus targets are known, as well as more broadly contribute to the way other related viruses may work as well.

There are a number of procedures available for the detection and quantification of a specific mRNA or protein to determine its expression level in a sample. Each of these methods has both advantages and limitations over the others. The use of particular techniques are more appropriate for certain studies than others, but equally limitations in these same techniques may inhibit the applicability of its use and so an existing alternative must be utilised. Also, though newer techniques may provide more information than older established techniques, the downstream analysis of the information may require entirely different skills to analyse and thus may not be suitable for the question being asked in a study.

In the following sections, a brief outline of available methods to quantify gene expression in biological samples is given, along with a brief summary of existing research that has been undertaken in the field of plant virology and the strengths and weaknesses of the techniques that were reported. The techniques can be categorised into low to mid-throughput, those that process fewer samples, and high-throughput, those that can process samples in the thousands or greater, though there are no published guidelines to this discrimination.

1.7.1 Low to mid-throughput detection techniques

1.7.1.1 Northern Blotting

Northern blotting was developed in the 1970s as a progression from the technique known as Southern blotting. Northern blotting utilises the same principle only using RNA instead of DNA. The technique is able to detect specific RNAs within a mixture of total RNA, providing information about the tissue, the developmental stage of the gene when it is expressed, and the size and number of RNAs (Moustafa and Cross 2016).

RNA samples are separated by loading into a denaturing agarose gel and separating on the basis of their size, with smaller RNA units travelling further through the gel relative to larger ones in a fixed time period (Figure 1.10). The RNA is transferred onto a membrane made of nylon or nitrocellulose, whereby it hybridizes to specifically labelled probes consisting of the reverse complement of a target RNA sequence. After incubation at a specific temperature the membrane is washed and exposed to film or fluorescence is detected, allowing the RNA-probe hybrids to be detected or quantified, as the intensity of the probe signal bound to target RNA is proportional to the starting RNA.

This content has been removed by the author due to copyright issues.

Figure 1.10: Diagram of the northern blotting process (By Ilewieszoośmiornicach - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=48134046>)

Several studies were identified in the literature that utilised northern blotting to determine plant host responses during plant rhabdoviral infections (Havelda et al. 2008; Liu et al. 2018; Mann et al. 2015; Tsai et al. 2005), and within *Nicotiana* species (Jovel et al. 2007; Ma et al. 2014). For example, northern blotting was utilised to detect changes in AGO mRNAs in *N. benthamiana* during infection by tomato ringspot virus (ToRSV) at two different temperatures (Paudel et al. 2018).

Amongst other advantages, northern blotting is cost effective, can identify both the size and amount of RNA in a sample, and can detect alternative splicing of mRNAs. However, even

slight degradation of the RNA due to improper storage or RNAses in the laboratory can lead to compromises in the quality of the data obtained. The technique also utilises chemicals that need to be handled with care as they are hazardous to health, a factor that should be taken into consideration (Moustafa and Cross 2016). Also, the procedure, in comparison to more recently developed techniques, requires approximately 100,000 copies of an RNA sequence, making it one of the less sensitive methods available for molecular analysis currently available, and is only able to detect one RNA at a time so is a time-consuming method.

1.7.1.2 Western Blotting

Proteins can be detected within complex samples utilising antibody based probes, and data related to quantities, molecular weights and post-translational modifications of the targets can be yielded (Jensen 2012). Western blotting is again a development from Southern and northern blotting, having been modified to allow proteins to be separated after being loaded into a denaturing polyacrylamide gel and using electrophoresis, followed by transferral onto a nitrocellulose membrane (Jensen 2012) (Figure 1.11). Once attached to the membrane the individual proteins are detected through the use of specific antibodies that will target them.

This content has been removed by the author due to copyright issues.

Figure 1.11: Schematic diagram of the western blotting process. Reproduced from (Flint et al. 2015)

Western blotting is a highly sensitive technique and can detect very small amounts of proteins within a sample (approximately 0.1 nanograms); however, if samples are limited, this can be a restrictive factor (Anderson et al. 2011). The process is time consuming, technically demanding, and expensive compared to other techniques, as generally only one protein can be identified at a time, and can also only be carried out if the primary antibody against the protein of interest is available (Ghosh et al. 2014). Antibody specificity is important, and another limitation is that some antibodies have been shown to interact with non-target proteins (Mahmood and Yang 2012).

Western blotting has been used in plant virus gene expression studies to determine both viral accumulation and host protein accumulation in infected plant leaf material (Mann et al. 2015;

Wu et al. 2018). Likely due to the limitations already mentioned, the majority of the literature found only reported findings related to one specific protein per study.

1.7.1.3 Fluorescent *in situ* hybridisation (FISH)

Fluorescent *in situ* hybridisation (FISH) allows visualisation of nucleic acid content within cells through the use of a specially designed fluorophore labelled DNA or RNA probe, that hybridises to a region of interest on an interphase or metaphase chromosome preparation and after incubation, can be visualised through a microscope (Kliot et al. 2014) (Figure 1.12).

Whilst other quantification methods such as northern blots, RT-PCR and microarrays provide information on average mRNAs in an a heterogenous population, FISH can specifically count and localize mRNAs in individual cells, so is highly specific. This specificity allows researchers to determine where they operate intracellularly and perhaps elucidate information about what molecular pathways they may influence (Shargil et al. 2015). For example, gene expression in virus infected grapevines was analysed using FISH and upregulation of genes known as invertase and pyrophosphorylase was observed at a site specific location in the tissues examined (Vorwerk et al. 2008). A quantitative version of FISH, Q-FISH, has been developed that can quantify the intensity of the fluorescence, but no studies utilising it have been used in conjunction with plant virus studies as of 2019.

Limitations to the method are that it is very time consuming as only one RNA can be detected at a time and hybridisation times of the probes are long. Further, inaccurate preparation of source material can have an impact on the downstream outcome of the assay (Gao et al. 2012; Huber et al. 2018).

This content has been removed by the author due to copyright issues.

Figure 1.12: Schematic representation of FISH (Image sourced from <http://www.abnova.com>)

1.7.1.4 RNase protection assay (RPA)

RPA is a sensitive molecular method to identify RNA molecules in a heterogeneous mixture. The technique relies on a specifically designed probe for a gene of interest annealing to the target mRNA sequence within extracted total RNA. Ribonucleases digest all single stranded non-hybridised products, proteinase K digests RNases, and the remaining hybridised products contain the sequence of interest and are obtained using phenol / chloroform extraction, followed by visualisation on an acrylamide gel (Figure 1.13) (Qu and Boutjdir 2007a).

RPA is a more sensitive technique compared to some other techniques and slightly degraded RNA samples can still be run using the technique and give valid results. Multiple RNAs can be detected in one run with appropriate probe design and alternate mRNA splices can also be detected. However, the size of the transcript is not obtained utilising the technique and it is a time-consuming assay (Ahne et al. 1998) (Qu and Boutjdir 2007a).

This content has been removed by the author due to copyright issues.

Figure 1.13: Schematic diagram of RPA (Qu and Boutjdir 2007b)

1.7.1.5 Polymerase chain reaction (PCR)

PCR is a widely used application in molecular biology. As previously outlined in Section 1.4.7, the assay involves DNA samples being rapidly and repetitively amplified via a set number of cycles consisting of denaturation, annealing and extension stages, each at different temperatures. Specifically designed primers anneal to a target sequence after denaturation and amplify complementary cDNA strands through the synthesis of new template strands via *Taq* DNA polymerase and free deoxynucleotide triphosphates (dNTPs) (Figure 1.14). The conditions under which this can occur are highly variable depending on the size and structure of the region of interest and so are very customisable and frequently optimised during the course of research. A PCR experiment generally has three recognised phases; a lag phase, an exponential extension stage, followed by a plateau phase. Once all the cycles have been completed, the products can be visualised on an agarose gel using electrophoresis. Multiple

primers can be included in one reaction to simultaneously amplify several targets, known as multiplex PCR.

This content has been removed by the author due to copyright issues.

Figure 1.14: Schematic diagram of the PCR process (Image sourced at <https://www.bosterbio.com/protocol-and-troubleshooting/molecular-biology-principle-pcr>)

RNA sequences cannot be directly amplified by conventional PCR, as a DNA template is required; however, the addition of a reverse transcriptase (RT) step has been developed. RT-PCR involves a complementary DNA (cDNA) being synthesised using the RT enzyme and a primer, either oligo(dT) or random primers, designed to amplify a specific sequence (Figure 1.15). Once the cDNA sequence has been generated the PCR step can follow. A one-step RT-PCR reaction carries out both the cDNA synthesis and PCR reactions in one closed tube in the thermocycler, whereas a two-step RT-PCR involves the synthesis of the cDNA in one reaction, with the PCR occurring in a separate reaction. Whilst a one-step RT-PCR reaction reduces the potential for contamination to occur, a two-step RT-PCR reaction generates cDNA which means less of the original sample needs to be utilised if multiple experiments need to be run, useful if samples are in short supply.

PCR is highly specific but prior knowledge of the sequence to be amplified is necessary in order to design the primers. Standard PCR and end point RT-PCR are not quantification assays; while they have been used extensively in gene expression assays, they cannot be utilised for quantitative studies with any confidence.

This content has been removed by the author due to copyright issues.

Figure 1.15: Representation of the cDNA synthesis process. Image sourced at <https://www.thermofisher.com/uk/en/home/life-science/cloning/cloning-learning-center/invitrogen-school-of-molecular-biology/rt-education/reverse-transcription-applications.html>

1.7.1.6 Differential display of mRNA by PCR (DD-PCR)

Total RNA can be extracted from samples under different experimental conditions, converted into cDNAs using anchored oligo dT primers. Some of the cDNAs are amplified using a general set of conditions with the primers, giving a subset of the mRNAs in the samples, that can be visualized on a gel (Figure 1.16). Limited studies were identified in the literature that utilised this technique in plant virus research (Benito et al. 1996). The assay is rapid and only requires small amounts of RNA to compare samples from different conditions and can discriminate the expression of different genes from multiple populations (Moustafa and Cross 2016). However, the assay preferentially amplifies mRNAs that have an initially large copy number, and due to the use of degenerate, non-specific primers, as well as a general set of amplification conditions, may make it susceptible to false positive results (Moustafa and Cross 2016).

This content has been removed by the author due to copyright issues.

Figure 1.16 Diagram of the DDPCR process (Miura and Scharf 2010)

1.7.1.7 Quantitative real-time reverse transcription polymerase chain reaction (RT-qPCR)

Quantitative real-time reverse transcription polymerase chain reaction (RT-qPCR) is a highly sensitive, commonly used molecular technique for measuring gene expression and has been used to quantify mRNA transcript levels present within samples in several plant virus studies including rhabdoviruses (Hashimoto et al. 2017; Liu et al. 2013; Wang et al. 2018). The assay is built upon the use of conventional PCR principles, but qPCR utilises the fluorescence of a sample at each cycle of amplification as a means of proportionally analysing the amount of material present in a sample at a given time. This is based on the intensity of the fluorescence, and so can be analysed during the run of the experiment. There are several ways to generate fluorescence. Two common ways are through the use of double stranded DNA intercalating dyes, such as SYBR green, that fluoresce upon binding to dsDNA, or through the use of specifically designed TaqMan probes that fluoresce upon binding to a specific target sequence or via hydrolysis through the extension phase (Figure 1.17). The method is highly reproducible

and sensitive and can also be combined into a multiplex reaction to reduce the possibility of contamination between the preparation of sample.

A well-designed RT-qPCR experiment can determine the differences in expression levels across different conditions, if any exist. The robust design of primers and the addition of a melt curve analysis at the end of the experiment to accurately determine the melting temperature (T_m) and number of the products amplified can reduce the possibility of obtaining a false positive result using SYBR green dye, which non-selectively binds to dsDNA. The relative ease of conducting the experiments has led to a significant number of studies and research methods utilising the technology; however, not all of the data presented from these studies has been necessarily reproducible due to poor experimental design or invalid and inappropriate controls being applied (Pabinger et al. 2014). As a result, a set of guidelines were established in 2009 that outlined the minimum requirements the experimental design and subsequent report needed to achieve in order to be considered a valid qPCR experiment, known as the MIQE guidelines (Bustin et al. 2009).

This content has been removed by the author due to copyright issues.

Figure 1.17: Overview of the qPCR process (Image sourced at www.thermofisher.com/)

One of the key MIQE guidelines is that RT-qPCR studies require a validated set of reference genes, genes which remain stably expressed across different experimental conditions. These are to be analysed during the course of the experiments alongside the genes being studied, for use as a point of comparison and to act as internal controls. These sets of reference genes have been shown in the literature to be specific to a plant host, and sometimes specific to a particular virus-host combination. RT-qPCR studies have been conducted in some *Nicotiana* species, such as *N. benthamiana*, and on other plant virus model plants such as *A. thaliana*, and sets of reference genes have been identified for these for researchers to use (Kozera and Rapacz 2013; Lilly et al. 2011; Liu et al. 2012; Schmidt and Delaney 2010). However, there is currently no set of reference genes for *N. glutinosa*. As it has potential to also be a model host, there is an opportunity to identify a set of specific reference genes during the course of this study for use in future experiments.

qPCR currently requires the use of a thermocycler and reagents obtained as kits from commercial suppliers and so may be costly for smaller laboratories compared to other methods.

1.7.2 High-throughput techniques

1.7.2.1 Microarrays

The development of microarray technology has allowed for the expression of large numbers of genes to be studied simultaneously, in a cost effective manner (Groen 2001). Microarrays have hundreds of thousands of spots of oligonucleotides, also known as probes, of varying lengths and combinations attached to them. These are complementary to known RNA sequences (Jaksik et al. 2015). To measure the relative amount of a specific RNA in a test sample, compared to a control, RNA must be extracted from each sample and each labelled with a different fluorescent label. For example, the sample of interest may be labelled with a red fluorescent dye, while the control with a green fluorescent dye. Each nucleic acid sample is hybridised with the microarray, so that individual sequences within the sample will hybridise to complementary oligonucleotides. Hybridisation can then be visualised by detection of the fluorescent signals for each spot on the array. The amount of each label can be quantified so that the relative amount of their sequence in a test sample can be measured against that in a control sample. Using the dye example above, if the expression of a gene is higher in the experimental sample than that in a control sample, the spot on the microarray will appear red, if it is lower, it appears green and if it is equally expressed it appears yellow (Figure 1.18).

Microarrays have a high degree of sensitivity; however, they have been reported to have decreased specificity meaning there may be inaccurate discrimination between samples with similar sequences (Dacheux et al. 2010). Experiments can yield information on the expression of thousands of genes, though only whether they were up or down regulated. No information is provided directly about their function or what wider implications these changes may have (Senthil et al. 2005). Determining the role specific RNAs of interest is a matter of interpretation, which needs to be undertaken by molecular specialists. Microarrays are more expensive than some other techniques such as RNA-Seq (Section 1.7.2.4); however, some of the steps in those alternative procedures, such as length hybridisation times, are used in microarrays, so the same limitations may be present in the alternatives (Jaksik et al. 2015)

There are limited studies related to the use of microarrays in *N. glutinosa*, with these studies focusing on wound induced gene responses, and there is no full genome study of *N. glutinosa*;

however, gene expression microarrays exist for *N. benthamiana* (Goralski et al. 2016) and *N. tabacum* (Edwards et al. 2010), as well as Solanaceae species such as tomato (Moore et al. 2005), and INSV and SNYV infected potato (Senthil et al. 2005). This latter study determined that genetically distinct enveloped viruses elicit unique change in plant host expression (Senthil et al. 2005). This information can be used to determine how particular genes in species related to *N. glutinosa* may be under or over expressed in response to viral infection, and so may provide a good starting point to identify candidate genes to study in *N. glutinosa* during an LNYV infection, or else candidate reference genes that are stably expressed across different experimental conditions.

This content has been removed by the author due to copyright issues.

Figure 1.18: Schematic diagram representing the microarray process (<https://bitesizebio.com/7206/introduction-to-dna-microarrays/>)

1.7.2.2 Serial analysis of gene expression (SAGE)

SAGE can simultaneously identify the transcript levels of thousands of genes in rapid succession, and has been utilised in multiple plant virus studies to analyse transcriptome changes between uninfected and infected samples (Fregene et al. 2004; Lu et al. 2012; Senthil et al. 2005). mRNA fragments are extracted from a sample to synthesise double stranded complementary DNA strands that are cleaved with anchoring enzymes. The upstream components are ligated to two adapters in separate reactions (Moustafa and Cross 2016). The separate reactions are combined and subsequently cleaved with a tagging enzyme to generate approximately 9-14 bp fragment tags that join via ligation to form di-tags. These are then amplified using PCR (Yamamoto et al. 2001). The PCR products can be analysed using Sanger sequencing to determine the gene expression profiles of various samples without prior knowledge as to what the genes may be, so is a highly informative tool (Tarasov et al. 2007).

SAGE requires a large amount of RNA sample. Thus, if samples are scarce it is not a viable technique to use. Further, the assay is expensive and time consuming. Revised techniques and

updated methodologies are continuing to be developed that rely on increasing the length of the fragment tags to make them more specific, as well as versions of the assay that require less starting material and fewer amplification cycles (Moustafa and Cross 2016).

This content has been removed by the author due to copyright issues.

Figure 1.19: Diagram outlining the SAGE process (Harkin and Hall 2000)

1.7.2.3 High-throughput sequencing (HTS) / next generation sequencing (NGS)

Next generation sequencing is a broad term for a set of techniques developed within the last two decades that have widespread applications in genomic, transcriptomic and epigenomic studies in many different organisms (Moustafa and Cross 2016). As “next generation” suggests future developments to come, despite the techniques already being in widespread use, the term “high-throughput sequencing” is beginning to be adopted.

HTS has the ability to sequence millions of sequences in parallel depending on the experimental design. The rapid nature of the technologies now allows for rapid sequencing of entire genomes in hours. Since the initial development of the technology, different techniques utilising different nucleic acid preparations have been established and continue to be optimised by different research groups (Figure 1.20). The use of total RNA, ribosomal RNA, double stranded RNA (dsRNA), and polyadenylated RNA (poly(A) RNA) have been identified as the most common in the literature (Pecman et al. 2017) and different methods of preparation of these nucleic acid samples have been the subject of several papers and reviews (Kesanakurti et al. 2016). Therefore, whilst there is an ability to obtain a significant amount of information utilising NGS techniques, particular design remains necessary depending on the type of information that is to be yielded. A current limitation for most NGS applications is the cost, but as with any new application this is becoming less prohibitive over time as new companies compete and developments occur. The individual NGS techniques have widespread applications but each will be looked at briefly in terms of their applications to gene expression studies in host plants for the purpose of this research.

Second-generation HTS platforms include 454 Sequencing (Roche) vHiSeq, MiSeq and NextSeq (Illumina), or Ion Torrent (Thermo Fisher Scientific) (Figure 1.20a). Massively parallel sequencing is used to obtain high throughput and has high base-calling accuracy. Disadvantages are that the sequencing reads are short and split contigs in repetitive regions during sequence assembly can be a problem (Ronholm et al. 2016). Third-generation sequencers, including PacBio (Pacific Biosciences) and MinION, PromethION, and SmidgION (Oxford Nanopore Technologies) (Figure 1.20b), can sequence single-molecule templates, achieving longer read length at a high throughput (Ronholm et al. 2016). Third-generation sequencers have high error rates relative to the other technologies, though their speed compared to second generation sequencers is much higher, so the use of a particular method is dependent on the situation the application is being utilised in.

This content has been removed by the author due to copyright issues.

Figure 1.20: Diagrammatic summary of the different NGS processes (<https://genotipia.com/wp-content/uploads/2019/03/NGS.pdf>)

1.7.2.4 RNA Sequencing (RNA-Seq)

RNA-Seq, or massively parallel cDNA sequencing, is a HTS technique that provides precise measurements of different types of RNA molecules including RNA, non-coding RNAs, microRNAs, and small RNAs (sRNAs), as well as their isoforms. It can also help elucidate information about changing expressions levels under varying conditions (Wang et al. 2009a). Generally, total RNA from a sample is reverse transcribed into cDNA, which is then fragmented to create a cDNA fragment library (Figure 1.21). Adapters are added to the cDNA ends. Because the adapters are common to every cDNA, HTS sequencing can be done and millions of reads can be obtained (Roberts et al. 2011).

This content has been removed by the author due to copyright issues.

Figure 1.21: Diagrammatic representation of RNA-Seq (Haas and Zody 2010)

Advantages of RNA Seq are that no prior knowledge of expressed sequences is required, and it has high sensitivity for genes expressed at high or low levels, as well as a more dynamic range of expression levels over which transcripts can be detected (Wang et al. 2009b). For examples, one plant rhabdovirus study discriminated twelve single nucleotide polymorphisms (SNPs) across a population of *Maize Iranian mosaic virus* (MIMV) infected maize samples (Ghorbani et al, 2018), information that can indicate variations of a virus that could be of importance in determining the evolution of the virus. Other studies have identified new plant rhabdoviruses and used RNA-Seq to elucidate the genome of these virus, such as “maize associated cytorhabdovirus” (Willie and Stewart, 2017), *Rice stripe mosaic virus* (Yang et al, 2017), and *Wheat yellow striate virus* (Liu et al, 2018). Disadvantages of utilising the technique are the large amounts of data generated that require large computational power and storage (Ozsolak and Milos 2011). The process is also costly and time intensive due to the assay design. The RNA fragmentation step that is involved in the mRNA library preparation has been noted for introducing 3' end bias as the fragments may not accurately reflect the transcripts they represent (Roberts et al. 2011).

1.8 Summary of literature to determine the direction of the project

There is currently a lack of information relating to plant host responses to cytorhabdoviruses. LNYV has been the focus of many studies over the last 53 years, though there has been little focus on the impact it has on plant hosts at the molecular level. With the identification of subgroups I and II of the virus, there is an opportunity to examine changes in gene expression in response to different viral subgroups, as currently few studies exist related to this area also. The hypothesis of Higgins et al, 2016, that subgroup II may have a transmission or replication based advantage over subgroup I, can be explored to a limited degree by analysing the gene expression data and determining differences between the subgroups.

In the published literature, little information was found regarding *N. glutinosa* as a model plant for virus studies. However, several LNYV based studies have utilised *N. glutinosa* for experimental work as it is easy to inoculate the virus into and several molecular techniques exist that outline methods to work with the model species and virus. In other studies looking at molecular processes within plant hosts, there is as yet no set of reference genes identified for *N. glutinosa* that have been validated as being candidates as internal controls in RT-qPCR experiments. This is important as every plant host will have a different set of genes that do not vary across different experimental conditions in their expression levels.

Many different molecular assays are available to study gene expression in plant hosts, each with advantages and disadvantages (Table 1.1). With the equipment and timescale available for the project, a limited in scope study utilising RT-qPCR was conducted to give an initial impression of gene expression in LNYV infected *N. glutinosa*. The MIQE guidelines formed the basis for the general design of the research presented here and every effort was made to adhere to the minimum standard established in those guidelines.

Table 1.1: Table summarising the features of the aforementioned techniques available to study gene expression (Moustafa and Cross 2016)

	Northern Blot	RPA	DD-PCR	SAGE	DNA Arrays	qPCR	NGS
No. of genes	low	low	medium	high	high	medium	high
Specificity	high	high	high	medium	medium	high	high
Targeted	yes	yes	no	no	yes	yes	No
Scalability	medium	medium	medium	medium	high	high	high
Difficulty	low	high	high	high	high	medium	high
Cost	low	low	Low	medium	high	medium	high

1.9 Aims of the study

This study had two main aims that are addressed in the following two chapters.

1. Identification of suitable target and reference genes in *N. glutinosa*, and design and testing of primers for these genes for gene expression studies using RT-qPCR (Chapter 2).

To date, there is no set of validated reference genes in *N. glutinosa* to run in RT-qPCR experiments, and no experiments have studied gene expression in plant hosts during LNYV infection. As the MIQE guidelines require a set of internal controls for a valid RT-qPCR experiment, these reference genes must be determined based on previous research into similar plant species. Also, as no published genome for *N. glutinosa* exists, databases containing gene information related to the identified genes must be searched for information to construct multiple sequence alignments to determine the likely nucleotide sequence for these genes in order to design primers. These primers must be tested and optimised before being used in a RT-qPCR experiment. If successful, these primers can be utilised by future researchers using *N. glutinosa* for RT-qPCR-based expression studies.

Objectives:

- Use the literature to identify candidate reference genes that maintain stable expression levels in similar plant species during viral infection, based on existing studies.

- Use the literature to identify candidate genes of interest that significantly up or down regulated in related plant species during viral infection, based on existing studies.
 - Obtain and align molecular information from related plant species for these candidate genes and genes of interest (GOI) to enable primer design to amplify genes in a RT-qPCR study.
 - Test primers on *N. glutinosa* using end point RT-PCR to determine appropriate amplification conditions.
2. Analyse the difference in expression of a select group of target genes in response to infection by LNYV subgroups I and II with *N. glutinosa* inoculated plant material (Chapter 3).

The second aim was to conduct a gene expression study using RT-qPCR on a limited set of target genes in *N. glutinosa* plants inoculated with LNYV virus belonging to subgroup I or subgroup II. The difference in gene expression levels, if any, between uninfected and plants infected with LNYV subgroups I or II were quantified and analysed using statistical tests to determine if any RNAs show significant differences in accumulation.

Objectives:

- Grow *N. glutinosa* from seed and inoculate with LNYV subgroups I and II obtained from lettuce samples collected in the field.
- Determine the infection rate of LNYV subgroups I and II in *N. glutinosa*.
- Using the primers developed above, compare the relative expression of GOI in *N. glutinosa* infected with LNYV subgroup I vs. subgroup II infected plants relative to mock infected plants.

Chapter 2

Identification and primer design of reference and target genes for *Nicotiana glutinosa* gene expression studies

2.1 Introduction

Gene expression studies focusing on the impact of plant viruses on living hosts are numerous within the published literature, with many assays available capable of detecting changes in the levels of a gene transcript in response to abiotic and biotic factors. One of these assays, RT-qPCR, has been widely used in plant virus studies to measure host responses to different stimuli (Baek et al. 2017; Die and Roman 2012; Noris and Miozzi 2015; Valmonte 2016a; Valmonte 2016b).

As of 2018, there had been no gene expression studies examining the impact of LNYV on a plant host, and, further to this, no gene expression studies to examine differences that occur between infection by subgroup I and subgroup II isolates of the virus. Changes in gene expression as a result of a viral infection have been shown to be specific to a particular plant-virus combination, with different sets of genes with different functions being up or down regulated during a particular infection (Hull 2013; Kamitani et al. 2016). It is not known whether this would apply to different subgroups of the LNYV virus, so a limited study was conducted of a select set of genes to initially determine if any significant differences could be discovered. *N. glutinosa* was the plant host infected in these studies as previous studies have utilised it as the indicator species for the virus, meaning it is easily grown and infectible for studying LNYV. Further, no studies have reported responses of *N. glutinosa* to infection. The candidate genes that were studied for the gene expression study were selected from existing published microarray and RT-qPCR data in related plant host species, where significant up or down regulation of genes in response to viral infection had previously been reported.

The designing and implementation of a RT-qPCR experiment involves many different processes, each of which needs to be considered and conducted carefully in order to obtain valid and robust data (Broeders et al. 2014). Whilst many studies have published data utilising RT-qPCR, it has been identified that inaccurate conclusions have been drawn in some studies due to the experimental design and lack of appropriate controls (Bustin et al. 2009; Kozera and Rapacz 2013; Taylor et al. 2009). A set of guidelines, the MIQE guidelines, were published to give researchers a checklist of the necessary components of qPCR experiments that need to be considered in order to obtain valid data and draw correct conclusions. These need to be clearly articulated when data are published. However, it remains difficult to ensure that researchers

correctly adhere to the guidelines, with it being reported that journals with higher impact factors tend to provide less of the required information (Bustin et al. 2013). Those that do have higher transparency in experimental reporting allow for the conclusions drawn to be more trusted by subsequent researchers. It is the intention of this study to adhere to as many of the MIQE guidelines as possible, and so the experimental design has been developed with them in mind.

The MIQE guidelines state that studies utilising RT-qPCR require several reference genes to be used within the experiment to act as internal controls. Ideal reference genes to use for RT-qPCR studies are genes that have no change in transcript accumulation under different conditions, such as between a host being infected or uninfected, not be associated with pseudogenes to avoid genomic DNA amplification, reflect variations in RNA quality and quantity, and accumulate past a certain threshold but not over accumulate. The MIQE guidelines suggest between two and five reference genes are necessary for a RT-qPCR experiment. Within the literature initially reviewed for this project, the number of reference genes that were utilised by different researchers varied. Whilst several papers selected a group of reference genes based on those used in earlier papers, it has been noted by other researchers that, due to molecular variances in conserved genes between different plant species, how a gene is expressed in one host may not be the same in another. Therefore, if a gene is stably expressed in one species it may not remain stable in another. In addition, primers designed to amplify a target transcript in one species may not work in other species and would likely perform optimally if specifically designed for use in a specific species.

As of 2019, no full genome had been published for *N. glutinosa*. Further, there is not a standard set of reference genes to use as internal controls in RT-qPCR studies of *N. glutinosa*. Thus, the candidate genes identified for the gene expression study and the candidate reference genes may not have published sequence information from which to design specific primers for their amplification. As a result, primer sequences for these genes must be inferred for *N. glutinosa* based on published mRNA and gene sequences for these genes in other related plant species. Once designed, primers for both reference and target genes must be tested and validated for their suitability in a RT-qPCR experiment. Identification of candidate genes and primer design occurred concurrently in this study, as neither had been previously studied with this particular virus and plant host. This chapter will report on the design and testing of primers for both the

candidate reference and target genes as the process was the same for each set, with minor variances between the methodologies which will be reported when appropriate.

2.1.1 Aims

- Using published studies, identify candidate reference genes that maintain stable expression levels in similar plant species during viral infection.
- Using published studies, identify candidate target genes that significantly up or down regulate in related plant species during viral infection.
- As no full genome for *N. glutinosa* has been published, obtain and align molecular information from related plant species for these candidate genes to design primers for amplification of those genes in a qPCR study.
- Test primers on *N. glutinosa* using end point RT-PCR to confirm amplification of correct product sizes.

2.2 Materials and methods

2.2.1 Identification of candidate genes of interest from existing literature

No gene expression studies have been conducted studying LNYV in *N. glutinosa*, nor in other plant species, so it is unknown what types of genes may be up or down regulated in response to this particular virus in this plant host. As a result, an examination was carried out of the literature, describing gene expression responses of *Solanaceae* plant hosts to virus infection. This examination was carried out to identify genes that showed significant expression changes between uninfected and infected states when using qPCR or microarray assays were used. Four genes were selected as candidate target genes of interest (GOI) to have their expression analysed between uninfected and LNYV subgroup I or II infected samples.

2.2.2 Identification of candidate reference genes from existing literature

Currently, there is not a standard set of reference genes to use in RT-qPCR studies of *N. glutinosa*. A literature review was also conducted to identify stable reference gene candidates

from *Solanaceae* species; genes that had been identified and validated across multiple species were studied further.

2.2.3 Gene expression location for GOI and reference gene candidates in *Nicotiana* species

To confirm the candidate genes that would likely be expressed at sufficient levels in leaf material taken from *N. glutinosa*, gene expression maps were obtained from the Queensland University of Technology (QUT) Gene Atlas version 6 feature on the database (<http://sefapps02.qut.edu.au/atlas/tREX6.php>). The accession numbers for each gene from NCBI were taken from the existing literature and used to obtain their related sequences. These were used as query terms in a BLASTn search of the National Centre for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov). The top *Nicotiana* hit was subsequently entered into the QUT *N. benthamiana* transcriptome database (version 6) (www.benthgenome.qut.edu.au) to determine the most likely sequences in *N. benthamiana*. Potential expression maps for *N. benthamiana* were obtained for each GOI using the Gene Atlas version 6.

2.2.4 Gene ontologies

For both GOI and reference gene candidates, the functions of each gene product were determined using a gene ontology database (<http://amigo.geneontology.org/>), for the purposes of determining what functional impact an LNYV infection may have on a plant host if significant over or under expression in a particular gene were to be observed in the study.

2.2.5 Obtaining mRNA and gene sequences for GOI and reference genes from related plant species

No full genome currently exists for *N. glutinosa*, though sequence data for some genes have been published. However, the sequences for the candidate GOI in this study were not available for *N. glutinosa* in any of the existing databases reviewed. Thus, it was necessary to infer what the sequence of the candidate reference genes might be from available sequences of related *Nicotiana* species. Five complete genomes were available for related *Nicotiana* species; *N. attenuata*, *N. benthamiana*, *N. sylvestris*, *N. tabacum* and *N. tomentosiformis*. These were

available from the NCBI, the QUT *Nicotiana benthamiana* genome website (www.benthgenome.qut.edu.au) and the Sol Genomics website (www.solgenomics.net).

The accession number for each GOI in a related host species was obtained from the existing literature and used to identify the gene and/or mRNA sequences in the NCBI. These sequences were used as query terms in a BLASTx search to identify the likely protein sequence. The protein sequence was then used as a query term in a BLASTp search of the NCBI website to determine the closest species in which the protein was found. The top *Solanum* or *Nicotiana* hit was identified and the mRNA sequence from this entry was then used as a query term against the existing *Nicotiana* genomes on the NCBI, QUT or Sol Genomics websites. This was done to obtain as many mRNA and gene sequences as possible from the different *Nicotiana* species for the particular candidate gene. This included variants of the mRNA sequences that existed in the databases, which arise due to alternative splicing and other modifications to the mRNA.

The same methodology was applied to obtain mRNA and gene sequences for the candidate reference genes from *Nicotiana* species.

2.2.6 Multiple sequence alignment of mRNA and gene sequences for GOI and candidate reference genes

To construct the likely intron/exon structure of each gene in *N. glutinosa* in order to be able to select an appropriate location to design primers around an intron/exon junction, Geneious (Biomatters NZ, version 6.0.6) was used. A multiple sequence alignment using the MUSCLE parameters within Geneious using at most eight iterations was conducted. The intron/exon structures within gene and mRNA sequences from each plant species obtained as described in Section 2.2.5 were aligned.

2.2.7 Comparison of published primers with multiple sequence alignments

Previous gene expression studies using related *Nicotiana* species described primers to amplify the candidate genes using RT-PCR or qPCR assays. Using the Primers tool feature in the Geneious software, it was possible to determine whether these primers would potentially bind

to all *Nicotiana* sequences in the multiple sequence alignment and therefore potentially be usable primers for studies in *N. glutinosa*.

2.2.8 Primer design for GOI and candidate reference genes

RT-qPCR primers for all candidate genes were designed towards the 3' end of the multiple sequence alignments, with a primer pair spanning an intron, whereby one primer was based either side of an intron. The primer pairs were designed to be each approximately 17 to 30 nucleotides (nt) long where possible, amplifying an amplicon in the region of 75 base pairs (bp) to 200 bp (Bustin and Huggett 2017; Rodríguez et al. 2015). The annealing temperature of the primers was designed so the primers would anneal at 60°C, so had melting temperatures (T_m) between 57°C and 63°C, with a GC content between 40% and 60%, no more than four dinucleotide repeats and no more than four consecutive runs of the same nucleotide. Primers were designed to have low GC content at their 3' ends, and efforts were made to start the primer with a C or G nucleotide to generate a stronger bond upon annealing.

The primer pairs were subjected to additional analysis in an effort to confirm their potential suitability for RT-PCR and RT-qPCR assays. Their characteristics and the products they were likely to generate were analysed using both Geneious and the Oligoanalyzer 3.1 software on the Integrated DNA technologies (IDT) website (<https://www.idtdna.com/>). Each gave individual results and the data was analysed to give likely mean values to judge. Characteristics such as the likelihood of homo and heterodimers forming in a reaction, the temperature at which hairpin loops may form and whether or not there was a BLAST match with *Homo sapiens* were all undertaken. The primers were modified as necessary when these criteria were not appropriate.

If any one of these features could not be attained the location, length or content of the primer pair was modified until all of these characteristics were achieved. The primers were ordered from a commercial supplier.

2.2.9 Primer testing on LNYV infected and uninfected *N. glutinosa* leaf material

To confirm all primers amplified the correct product size, the primer pairs were tested in a one-step RT-PCR reaction with total RNA extracted from uninfected or LNYV infected *N. glutinosa* leaf samples, as described below in Sections 2.2.9.1 to 2.2.9.5.

2.2.9.1 Obtaining uninfected and LNYV infected *N. glutinosa* leaf material

2.2.9.1.1 *N. glutinosa* plant growth

N. glutinosa seeds were kindly provided by John Fletcher (The Institute of Plant and Food Research, New Zealand). Seeds were sown in soil (Kings Plant Barn Potting Mix) supplemented with growth supplement (One scoop of Scott's Osmocote per 40 l bag of soil). The seedlings were housed in a plant growth chamber (Conviron CMP5090) at 20 °C, 80% humidity and light at 4 μ MO for three weeks on a 16/8 hr light/dark cycle. Plants were watered every three days. After four weeks, seedlings were transplanted into black plant bags with a fresh soil supply. Plants were grown for a further three days until 5-6 leaves had grown. Larger plants with healthy looking leaves were chosen for mechanical inoculation

2.2.9.1.2 Inoculations and sampling of *N. glutinosa* with LNYV

Of the 35 plants grown for inoculation, 30 plants were inoculated with LNYV subgroup II (Hv19) infected leaf material and five plants were mock inoculated with buffer. Infected leaf material was ground in 2 ml 0.1M phosphate buffer pH 7 containing 0.01% sodium sulphite, with a few pinches of 600 mesh carborandum. The mixture was then gently rubbed onto three leaves and left for six hours, before spraying off the carborandum with water.

Visual inspection of plants was undertaken up 28 days. Systemic leaves were analysed for symptoms by identifying signature features of LNYV infection such as reduced leaf size, yellowing of leaves, necrosis of leaves and mosaic like symptoms as outlined by Dietzgen et al (2007). Plants suspected of being infected had three systemic leaves removed into a ziplock bag and immediately placed into storage at -80 °C.

2.2.9.2 Extraction of total RNA from *N. glutinosa*

Total RNA was extracted from 100 mg LNYV positive and mock inoculated *N. glutinosa* leaves using the Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St Louis, MO, USA) following the manufacturer's instructions with minor variations. The 28 days post inoculation (dpi) systemic leaf material was removed from the -80°C freezer, immediately submerged in liquid nitrogen and ground with a sterile pestle and mortar to a fine powder. A solution containing 500 µl of lysis buffer and 5 µl of 2-mercaptoethanol was pipetted onto the powder. The mixture was pipetted into a 1.5 ml tube and vortexed for 30 seconds. The mixture was centrifuged at 15,000 g (Eppendorf 5417 R Centrifuge, Eppendorf) for 3 minutes to pellet the cellular debris. The supernatant was transferred into a filtration column and centrifuged at 15,000 g for 1 minute to remove the residual debris. The filtration column was removed and 500 µl binding solution was added to the clarified flow through lysate and mixed with a pipette. The mixture (700 µl) was pipetted into a binding column in a 2 ml tube and centrifuged at 15,000 g for 1 minute to allow RNA to bind to the column. The binding column was removed, the flow through discarded and residual liquid was removed by tapping the collection tube upside down on clean absorbent paper. The binding column was returned to the tube and the remainder of the clarified flow was subjected to the same process. Wash solution I (300 µl) was added to the binding column and centrifuged at 15,000 g for 1 minute. Genomic DNA was removed by pipetting a mixture of 10 µl of DNase I and 70 µl of on-column DNase digestion buffer onto the centre of the binding column. The column was incubated at room temperature for 15 minutes, and another 500 µl of wash solution I was pipetted into the binding column before centrifuging at 15,000 g for 1 minute. The flow through was discarded, the binding column removed, and the residual liquid in the tube dried by turning the tube upside down and tapping on absorbent paper. Wash solution II (500 µl) containing ethanol was pipetted into the binding column and the tube was centrifuged at 15,000 g for 30 seconds. The flow through was discarded, the binding column removed, and the residual liquid in the tube dried by turning the tube upside down and tapping on absorbent paper. The wash solution II step was repeated, and the binding column dried by centrifuging the tube at 15,000 g for 1 minute. The binding column was removed and placed into a new 2 ml tube. Elution solution (25 µl) was pipetted into the column, left for 1 minute at room temperature and centrifuged at 15,000 g for 1 minute. This step was repeated to obtain a total of 50 µl of elution solution.

A subsample of the eluted RNA was immediately subjected to spectrophotometric analysis to determine RNA concentration and quality (GE Biosciences, Auckland, NZ). The remainder of the sample was stored at -80 °C for future usage.

2.2.9.3 Gel electrophoresis

The integrity of the extracted RNA was determined by electrophoresis in 1% agarose/Tris/Borate/EDTA (TBE) gels containing 0.1mg/ml of ethidium bromide (Mini-Sub[®] Cell GT Cell, BioRad, Auckland, NZ). Each RNA sample (3 µl) was mixed with 3 µl of loading dye, with 3 µl of 100 bp DNA ladder (Dnature, Gisborne, NZ) as a size marker. Gels were run for 50 minutes at 75 volts. The agarose gel was visualised by exposing it to UV light in an Alpha Imager chamber and photographed using an Alpha Innotech camera and viewed on a computer using Alpha Imager software (Version 5.0.1).

2.2.9.4 End point RT-PCR confirmation of primer amplification using Superscript[®] III RT-PCR kit

To confirm amplification of the correct product size, one-step RT-PCR was performed on infected and uninfected RNA samples using Superscript[®] III RT-PCR System with Platinum[®] *Taq* (Invitrogen). A total volume of 12.5 µl was used instead of the manufacturer recommendation of a 50 µl. RT-PCRs were set up as follows: 300 ng of RNA was combined with 6.25 µl of 2X reaction mix, 0.5 µl SuperScript[™] III RT/Platinum[™] *Taq* mix, 0.25 µl of 10 µM forward primer, 0.25 µl of 10 µM reverse primer and autoclaved distilled water to make a final volume of 12.5 µl. A positive control to a volume of 12.5 µl was included using the LNYV440F and LNYV1185R primers (Higgins et al. 2016), with a product size of approximately 750 bp anticipated for positive amplification. The primer sequences are 5'-TGACACAGATTCAGAACAACCTC-3' for LNYV-440F and 5'-CGGACAATCCATCTCCACTA-3' for LNYV-1185R. A no template control (NTC) was also included to a volume of 12.5 µl, with water replacing the RNA. The PCR tube was briefly centrifuged and placed into a thermocycler (Techne TC-512). The RT-PCR conditions were 1 x 30-minute cycle at 50°C for cDNA synthesis, 35 cycles consisting of 1 x 30 second cycle at 94°C for denaturation, 1 x 30 second cycle at 50°C or 60°C for annealing and 1 x 1 minute cycle at 68°C for extension, and a 1 x 5 minute cycle at 68°C for a final extension. The PCR

product was stored in a -20°C freezer until used. RT-PCR products were visualised by electrophoresis by using 1X TBE gels. The whole 12.5 µl PCR was loaded with 3 µl of loading dye.

2.2.9.5 Confirmation of correct primer amplification using RT-qPCR

After confirmation of correct product size amplification of each primer set utilising end point RT-PCR, the primers were used to amplify their targets in LNYV infected and uninfected *N. glutinosa* RNA in a one-step RT-qPCR. RT-qPCR reactions were performed using a LightCycler 96 Real-Time PCR machine (Roche Applied Science) using SYBR Green dye. Reactions were set up with 300 ng of target RNA, 6.25 µl SYBR Green (Thermofisher), 0.5 µl of Superscript III (Quanta Biosciences), 0.25 µl 10 µM forward primer and 10 µM 0.25 µl reverse primer, with autoclaved distilled water to make a total volume of 12.5 µl. Three technical replicates were loaded for each primer pair tested. The RT-qPCR conditions were 1 x 30-minute cycle at 50°C for cDNA synthesis, 1 x 2-minute cycle at 94°C for pre-incubation, 35 amplification cycles consisting of 1 x 30 second cycle at 94°C for denaturation, 1 x 30 second cycle at 60°C for annealing and 1 x 1-minute cycle at 68°C for extension, and a 1 x 5 minute cycle at 68°C for a final extension. Fluorescence values were obtained at the end of each cycle. A high resolution melt curve analysis was conducted after the amplification cycles, consisting of 1 x 5 second cycle at 94°C, 1 x 1 minute cycle at 65°C followed by a ramp to 97°C at a rate of 0.11°C/s, with 5 fluorescence values being obtained for each degree celcius point. The data from the LightCycler 96 thermocycler was transferred to a LightCycler 96 software for subsequent analysis of the amplification products.

2.3 Results

2.3.1 Identification of candidate reference genes

For gene expression studies involving RT-qPCR, several reference genes are necessary to act as internal controls, and these should be specific to the plant species being tested. No such reference genes have been identified in *N. glutinosa* so were developed for this study. Identification of candidate reference genes was conducted by undertaking a review of previous studies that analysed gene expression responses in the model plant *A. thaliana* and *Solanaceae* plants in response to virus infection. *Solanaceae* plants were the main focus since this is the family to which *N. glutinosa* belongs. The number of studies was limited; however a selection of genes that had been tested in different reports as reference genes and exhibited stable expression were identified (Table 2.1). *Actin*, *EF1 α* , *Ntubc2*, *PP2A*, *PDF2*, *SAND*, *L23*, *F-BOX* and *Ubiquitin* have been shown to have stable expression in several plant species (Abrahamian et al. 2013; Baek et al. 2017; Bubici et al. 2014; Chen et al. 2017; Chen et al. 2012; D'Ippolito et al. 2016; Liu et al. 2012; Maneechoat et al. 2015; Perez-Canamas et al. 2017; Schmidt and Delaney 2010; Wieczorek and Obrepalska-Stepłowska 2013). In contrast, some studies showed some of these genes to be not suitable as reference genes in a particular plant species (Abrahamian et al. 2013; Baek et al. 2017; Bubici et al. 2014; Chen et al. 2017; Chen et al. 2012; D'Ippolito et al. 2016; Liu et al. 2012; Maneechoat et al. 2015; Perez-Canamas et al. 2017; Schmidt and Delaney 2010; Wieczorek and Obrepalska-Stepłowska 2013). As no literature was found having tested these genes in *N. glutinosa* they were all included as potential candidates to determine their suitability as reference genes in this species.

Table 2.1: Summary of candidate genes identified for this research, the papers they were identified from, their accession numbers, the identified functions of the encoded protein and the plant species that the genes have been studied in so far in response to viral infection.

Paper	Gene	Accession	Protein Function	Tested In
Lilly et al, 2011	<i>PP2A</i>	At1g59830	Encodes an isoform of protein phosphatase 2A (PP2A)	<i>A. thaliana</i>
	<i>F-BOX</i>	At5g15710	Function unknown. Possible involvement in protein degradation via the proteasome.	<i>A. thaliana</i>
	<i>PDF2</i>	At1g13320	Encodes regulatory subunit of serine/threonine PP2A.	<i>A. thaliana</i>
	<i>SAND</i>	At2g28390	Membrane protein role in vesicle traffic and endocytosis.	<i>A. thaliana</i>
	<i>EF1α</i>	At5g60390	Involved in calmodulin binding and has translation elongation factor activity.	<i>A. thaliana</i>
	<i>Actin</i>	At1g49240	Cytoskeleton component and responsible for cell motility and signalling processes.	<i>A. thaliana</i>
Schmidt et al, 2010	<i>L23</i>	L18908	Mitochondrial ribosomal protein	<i>N. tabacum</i>
	<i>Ntubc2</i>	AB026056	Ubiquitin conjugating enzyme	<i>N. tabacum</i>
	<i>EF1α</i>	AF120093	See above	<i>N. tabacum</i>
	<i>PP2A</i>	X97913	See above	<i>N. tabacum</i>
	<i>Actin</i>	X69885	See above	<i>N. tabacum</i>
Liu et al, 2012	<i>PP2A</i>	At1g13320	See above	<i>N. benthamiana</i>
	<i>F-BOX</i>	At5g15710	See above	<i>N. benthamiana</i>
	<i>L23</i>	At2g39460	See above	<i>N. benthamiana</i>
	<i>SAND</i>	At2g28390	See above	<i>N. benthamiana</i>
	<i>EF1α</i>	At5g60390	See above	<i>N. benthamiana</i>
	<i>Actin</i>	At2g37620	See above	<i>N. benthamiana</i>

2.3.2 Identification of candidate GOIs

Identification of candidate GOIs was conducted by undertaking a literature review of previous gene expression plant virus studies utilising *Solanaceae* plants, and the model species *A. thaliana*. Whilst many studies found reported particular genes being up or down regulated in response to viral infection, the published papers sometimes omitted information such as accession numbers of genes tested, the fold change by which the genes were up or down regulated in the study or the days post infection that the studies were conducted. As this information was important for this study, those papers were omitted from the summary of the genes found below (Table 2.2). Where reported, the changes in expression were noted and genes with a significant expression change across multiple papers were considered for this research. Having reviewed the potential genes that may be up or down regulated to a significant degree and detected with a qPCR assay, it was decided that the genes *SGS3*, *CPK3*, *WRKY33* and *WRKY4*, (highlighted bold in Table 2.2), would be studied in this research, as each had previously been found to have a specific impact on defence responses in plant hosts in response to biotic and abiotic factors, including viral infection (Ando et al. 2013; Arimura and Sawasaki 2010; Brown et al. 2003; Li et al. 2006; Li et al. 2004; Valmonte 2016). The accession numbers

of the gene sequences were taken and a BLASTn search was conducted to identify the closest sequence published in *Nicotiana* species. Whilst *SGS3* and *CPK3* were named similarly in the top *Nicotiana* hit, *WRKY33* and *WRKY4* were identified in the NCBI database as *WRKY26* and *WRKY70*, respectively. For these genes, these are the names that will be referenced throughout the remainder of this thesis.

Table 2.2: Summary of gene expression studies showing mRNAs found to increase or decrease in accumulation in response to virus infection. The fold change in accumulation of each mRNA, as measured by either microarray or RT-qPCR, is indicated. Viruses used in this study were *Impatiens necrotic spot virus* (INSV), *Sonchus yellow net virus* (SYNV), *Turnip vein clearing virus* (TVCV), *Oilseed rape mosaic virus* (ORMV), *Potato virus X* (PVX), *Cucumber mosaic virus* (CMV), *Turnip mosaic virus* (TMV), *Tomato ringspot virus* (ToRSV), *Potato virus Y* (PVY), *Tomato yellow leaf curl virus* (TYLCV), *Tobacco mosaic virus* (TMV), *Tomato spotted wilt virus* (TSWV), *Cauliflower mosaic virus* (CaMV), *Turnip mosaic virus* (TuMV), *Turnip yellow mosaic virus* (TYMV). **Bold entries** are the genes that were selected for analysis in this research.

Study	Accession number	Gene ID	Protein Function	Virus	Virus source	Plant host	Microarray fold change up/down	RT-qPCR fold change	DPI
Senthil et al., 2015	NP_001234404.1	CHI3 chitinase	Cell defence	SYNV	<i>S. lycopersicum</i>	<i>N. benthamiana</i>	6.5 ↑	-	14
				INSV			33.2 ↑	-	5
	NP_567347.1	cytosolic invertase 2	Metabolism	SYNV	<i>A. thaliana</i>	<i>N. benthamiana</i>	10.8 ↑	-	14
				INSV			27.4 ↑	-	5
	NP_192902.1	OSM34 osmotin 34	Cell defence	SYNV	<i>A. thaliana</i>	<i>N. benthamiana</i>	4.3 ↑	-	14
				INSV			19.8 ↑	-	5
	NP_001234805.1	beta-1,3-glucanase	Cell defence	SYNV	<i>S. lycopersicum</i>	<i>N. benthamiana</i>	5.7 ↑	-	14
				INSV			19.4 ↑	-	5
	NP_178199.1	WRKY40 WRKY DNA-binding protein 40	Transcription	SYNV	<i>A. thaliana</i>	<i>N. benthamiana</i>	7.4 ↑	-	14
				INSV			15.6 ↑	-	5
	NP_180747.1	Leucine-rich repeat protein kinase family protein	Cell signalling, cell death, innate immunity	SYNV	<i>A. thaliana</i>	<i>N. benthamiana</i>	6.6 ↑	-	14
				INSV			12.2 ↑	-	5
	NP_199107.5	alpha/beta-Hydrolases superfamily protein	Metabolism	SYNV	<i>A. thaliana</i>	<i>N. benthamiana</i>	8.5 ↑	-	14
				INSV			11.6 ↑	-	5
NP_001233989.1	H2B-1 histone H2B	Cell biogenesis	SYNV	<i>S. lycopersicum</i>	<i>N. benthamiana</i>	6.6 ↓	-	14	
			INSV			12.2 ↑	-	5	

Whitham et al., 2003	AT2G45570	putative cytochrome p450	Metabolism	TVCV, ORMV, PVX, CMV, TuMV	<i>A. thaliana</i>	<i>A. thaliana</i>	8-41.5	-	5
Dardick, 2007	STMEF49	WRKY4	Transcription	PPV	<i>A. thaliana</i>	<i>N. benthamiana</i>	1.75 ↑	6.61 ↑	14
				ToRSV			1.37 ↑	49.9 ↑	14
Baebler et al., 2009	STMET20	WRKY4	Transcription	PYV (NTN)		<i>S. tubersum</i>	-	↑	0.5
Chen et al., 2013	Solyc03g116890.2	ja-induced WRKY protein	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-7.89 ↓	
	Solyc04g051690.2	probable WRKY transcription factor 51-like	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-7.11 ↓	
	Solyc09g014990.2	probable WRKY transcription factor 33-like	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-6.52 ↓	
	Solyc04g072070.2	WRKY DNA-binding protein 51	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-6.50 ↓	
	Solyc08g082110.2	probable WRKY transcription factor 53-like	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-6.34 ↓	
	Solyc08g008280.2	probable WRKY transcription factor 53-like	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-5.98 ↓	
	Solyc08g062490.2	WRKY DNA-binding protein	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-5.90 ↓	
	Solyc10g011910.2	WRKY transcription factor 22-like	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-5.80 ↓	
	Solyc09g015770.2	WRKY transcription factor	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.77 ↓	
	Solyc02g080890.2	WRKY transcription factor 6-like	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.44 ↓	
	Solyc06g066370.2	double WRKY type transfactor	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.39 ↓	
	Solyc01g095630.2	WRKY transcription factor	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.21 ↓	
	Solyc08g006320.2	WRKY transcription factor 11	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.21 ↓	

	Solyc03g095770.2	probable WRKY transcription factor 70-like	transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.14 ↓	
	Solyc06g068460.2	WRKY transcription factor 1	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-4.13	
	Solyc10g009550.2	WRKY transcription factor	Transcription	TYLCV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	-3.84	
Sun Un Huh et al., 2012	AY071920	<i>Capsicum annuum</i> WRKY transcription factor <i>d</i> (<i>CaWRKYd</i>)	Transcription	TMV	<i>Capsicum annuum</i>	<i>Capsicum annuum</i>	-	41.99	
Huang et al., 2016		SIWRKY31	Transcription	TSWV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	significantly ↑	
		SIWRKY33	Transcription	TSWV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	significantly ↑	
		SIWRKY39	Transcription	TSWV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	significantly ↑	
		SIWRKY41	Transcription	TSWV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	significantly ↑	
		SIWRKY08	Transcription	TSWV	<i>S. lycopersicum</i>	<i>S. lycopersicum</i>	-	significantly ↑	
Valmonte, 2016		CPK3	Signal transduction	CaMV, TMV, TSMV, TuMV, TYMV	<i>A. thaliana</i>	<i>A. thaliana</i>	-	↑ 1.5 to 4 fold	
Lilly, 2014		SGS3	Silencer of gene suppressor	CaMV, TMV, TSMV, TuMV, TYMV	<i>A. thaliana</i>	<i>A. thaliana</i>	-	↓ 1-2 fold	

2.3.3 Projected gene expression patterns for GOIs and candidate reference genes in *Nicotiana* species

Available expression maps for the GOIs (Figure 2.1) and reference genes (Figure 2.2) were obtained using the Gene Atlas version 6 feature on the QUT database (<http://sefapps02.qut.edu.au/atlas/tREX6.php>). Maps for the candidate genes *CPK3* and *WRKY26* were available but not for *SGS3* and *WRKY70*.

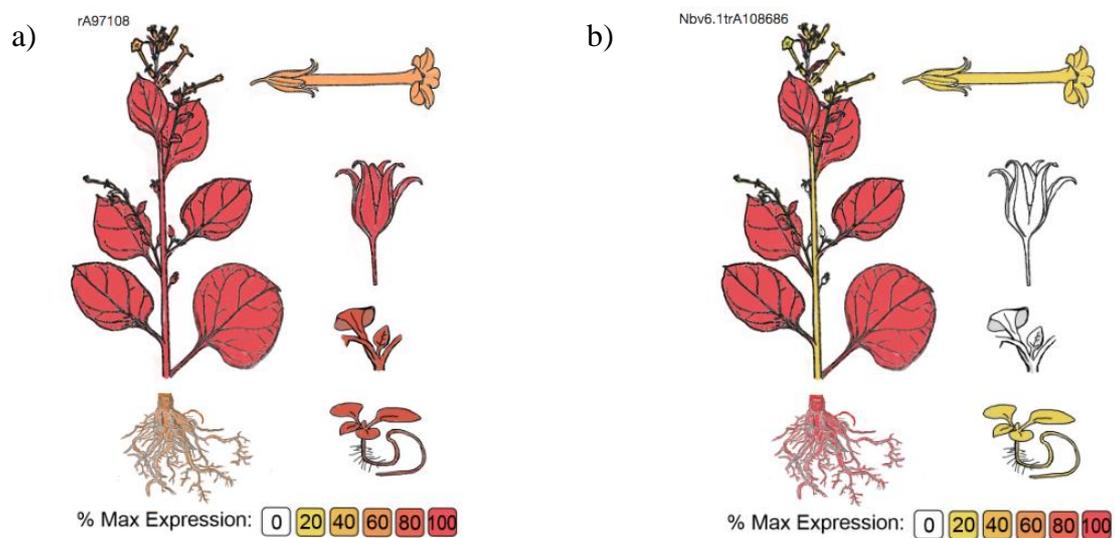


Figure 2.1: Expression maps for a) *CPK3* and b) *WRKY26* genes. The red areas indicate higher areas of likely expression for the gene under normal growth conditions.

These data suggested high levels of leaf expression for the GOIs *CPK3* and *WRKY26*, and the candidate reference genes *SAND* and *Ubiquitin*. The candidate reference genes for *Actin*, *EF1 α* , *F-BOX*, *L23*, *Ntubc2*, *PDF2* and *PP2A* would be expected to be lower. Regardless, the different gene maps obtained for both the candidate reference genes and candidate target genes indicated that all genes would be expressed in plant leaf material, which was important for this research as this was the tissue of interest for analysing mRNA responses to LNYV infection. As all candidate genes were identifiable in leaves in *N. benthamiana*, it was assumed this would likely be the case also in *N. glutinosa*. However, it should be noted that since there appear to be expression differences between the different genes according to the gene maps, this may not necessarily be indicative of the expression fold change that may be seen between uninfected

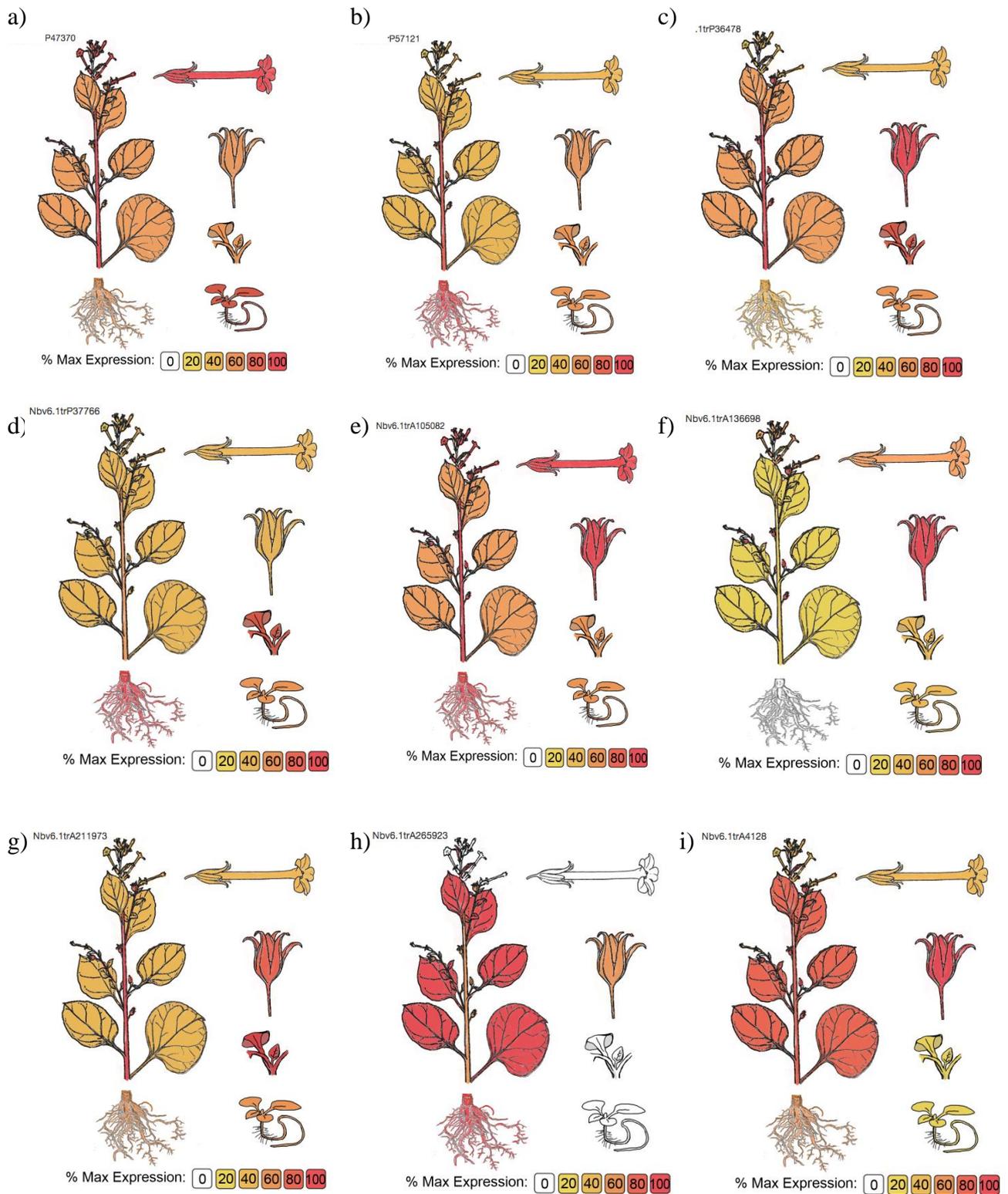


Figure 2.2: Expression maps for the candidate reference genes in *N. benthamiana*. The red areas indicate higher areas of expression for the gene a) *Actin* b) *EF1 α* c) *F-BOX* d) *L23* e) *Ntubc2* f) *PDF2* g) *PP2A* h) *SAND* i) *Ubiquitin*

and LNYV infected material; every gene will have a different amount of expression, and the relative change between the uninfected and infected conditions is the important factor to consider.

2.3.4 Gene ontologies for the GOIs and candidate reference genes

Whilst existing papers reported information related to the function of the candidate genes, this was limited (Schmidt et al, 2010; Lilly et al, 2011; Liu et al, 2012) . Gene ontologies provide information about the cellular location, molecular function and biological processes for gene products. By finding gene information in *Nicotiana* species using a gene ontology database, further information was obtained related to the likely function of the genes in *N. glutinosa*. By understanding their function, when the expression levels of these genes are obtained, information about the cellular processes LNYV may impact in a host can be further understood. Similarly, differences in expression of a target gene between the subgroups may help identify what affected processes differ between isolates and provide further information about the individual workings of the two subgroups. The gene ontologies from the QUT database for the candidate GOIs and candidate reference genes are shown in Table 2.3 and Table 2.4, respectively.

For the candidate GOIs, it was found that *CPK3* is involved in posttranslational modification of proteins, signalling processes and metabolic processes and *SGS3* is involved in gene expression regulation and viral defence responses, both within the cytoplasm or interacting with or within organelles outside the nucleus. These features mean there is a possibility, particularly with LNYV being a cytorhabdovirus that has been found to operate outside the nucleus, that the expression levels of *CPK3* and *SGS3* may be impacted by the establishment of an LNYV infection, as the virus hijacks the host cell mechanisms to replicate more virion particles. *WRKY26* and *WRKY70* are involved in the regulation of transcription in the nucleus so may influence host gene expression in response to infection as they have been found to be involved with SA and JA pathways. Therefore, they are suitable targets to study in this research.

For the candidate reference genes, it was found that *Actin* had a role in ATP binding, *EF1 α* in regulating translational elongation, *F-BOX* in metabolic processes, *L23* in the formation of ribosomes, *Ntubc2* in metabolic processes, *PDF2* in transcriptional regulation, *PP2A* in cell

generation and phosphorylation regulation, *SAND* in signalling and *Ubiquitin* in protein modifications. It is possible that all of these candidate reference genes may be impacted by the presence of an LNYV infection, as all of these processes generally could be sequestered to make new virus particles. However, as other studies have used them as reference genes in plant virus studies, this may not necessarily be the case with LNYV either. It was decided to test them to determine their suitability as reference genes, with the aim to link back any findings to this functional information during the analysis of the data.

Table 2.3: Summary of cellular location, function and processes of the candidate target genes from gene ontology feature on the QUT database

Gene	Cellular location	Function	Processes
<i>CPK3</i>	Vacuole, cytosol, plasma membrane	Protein serine/threonine kinase activity, calcium ion binding activity, ATP binding	Response to salt stress, abscisic acid-activated signalling pathway, regulation of stomatal movement, regulation of anion channel activity, regulation of protein localisation, response to cadmium ion, protein autophosphorylation, serine family amino acid metabolic process
<i>SGS3</i>	Perinuclear region of cytoplasm	-	Posttranscriptional gene silencing by RNA, regulation of defence response to virus, production of small RNA involved in gene silencing by RNA
<i>WRKY26</i>	Transcription factor complex	Transcription factor, sequence-specific DNA binding	Regulation of transcription, DNA-templated
<i>WRKY70</i>	Transcription factor complex	Transcription factor, sequence-specific DNA binding	Regulation of transcription, DNA-templated

Table 2.4: Summary of cellular location, function and processes of the candidate reference genes from gene ontology feature on the QUT database

Gene	Cellular location	Function	Processes
<i>Actin</i>	-	ATP binding	-
<i>EF1α</i>	Ribosome	Translation elongation factor activity, GTPase activity, GTP binding	Regulation on translational elongation
<i>F-BOX</i>	-	Ligase activity	Metabolic process
<i>L23</i>	Ribosome	Nucleotide binding, structural constituent of ribosome, rRNA binding	Translation, ribosome biogenesis
<i>Ntubc2</i>	-	Ligase activity	Metabolic process
<i>PDF2</i>	Nucleus	Lipid binding, sequence specific DNA binding	Regulation of transcription, DNA-templated
<i>PP2A</i>	Cytosol, plasma membrane	Binding, protein phosphatase type 2A regulator activity	Cell morphogenesis, cell growth, regulation of protein phosphatase type 2A activity, regulation of phosphorylation, Golgi vesicle transport
<i>SAND</i>	-	-	Calcium ion transport, cellular zinc ion homeostasis
<i>Ubiquitin</i>	Nucleus, cytoplasm	Phosphomannomutase activity, protein binding	Fructose metabolic process, mannose biosynthetic process, protein ubiquitination involved in ubiquitin-dependent protein catabolic process

2.3.5 Obtaining mRNA and gene sequences for all candidate genes from related plant species

As no full genome for *N. glutinosa* has been published, in order to design primers to amplify the reference genes and GOIs, gene sequences were retrieved from the genomes of the related species *N. attenuata*, *N. benthamiana*, *N. sylvestris*, *N. tabacum*, and *N. tomentosiformis*. Multiple alignments of these sequences allowed conserved regions between all species to be identified. It was considered that if these sequences were conserved between these five species, they would likely also be conserved in *N. glutinosa*. Table 2.5 to

Table 2.13 summarises (organised alphabetically by candidate reference gene name) the mRNA or gene sequences used to infer these sequences in *N. glutinosa*. Table 2.14 to Table 2.17 are summaries (organised alphabetically by candidate reference gene name) of the mRNA or gene sequences used to predict these sequences in *N. glutinosa*. For all genes, except for *WRKY26* and *WRKY70*, sequences were found in the NCBI, QUT and SOL databases. For *WRKY26* and *WRKY70* sequences were only found in the NCBI database.

Table 2.5: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *Actin* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Arabidopsis thaliana chromosome 2 sequence	NC_003071.7	<i>A. thaliana</i>	2,621
	Arabidopsis thaliana actin 1 (ACT1), mRNA	NM_179953.3	<i>A. thaliana</i>	1,851
	PREDICTED: <i>Nicotiana attenuata</i> actin (LOC109242640), transcript variant X1, mRNA	XM_019409473.1	<i>N. attenuata</i>	1,602
	PREDICTED: <i>Nicotiana attenuata</i> actin (LOC109242640), transcript variant X2, mRNA	XM_019409474.1	<i>N. attenuata</i>	1,620
	<i>Nicotiana attenuata</i> strain UT unplaced genomic scaffold, NIATTr2, whole genome shotgun sequence	NW_017670720.1	<i>N. attenuata</i>	3,281
	PREDICTED: <i>Nicotiana tabacum</i> actin-like (LOC107809070), transcript variant X1, mRNA	XM_016633661.1	<i>N. tabacum</i>	1,577
	PREDICTED: <i>Nicotiana tabacum</i> actin-like (LOC107809070), transcript variant X2, mRNA	XM_016633663.1	<i>N. tabacum</i>	1,570
	<i>Nicotiana tabacum</i> cultivar TN90 unplaced genomic scaffold, Ntab-TN90	NW_015947443.1	<i>N. tabacum</i>	3,083
	<i>Nicotiana sylvestris</i> unplaced genomic scaffold, Nsyl Nsyl_scaffold138051, whole genome shotgun sequence	NW_009381083.1	<i>N. sylvestris</i>	3,320
	PREDICTED: <i>Nicotiana sylvestris</i> actin (LOC104229840), transcript variant X1, mRNA	XM_009782557.1	<i>N. sylvestris</i>	1,607
	PREDICTED: <i>Nicotiana sylvestris</i> actin (LOC104229840), transcript variant X2, mRNA	XM_009782558.1	<i>N. sylvestris</i>	1,454
	PREDICTED: <i>Nicotiana tomentosiformis</i> actin-like (LOC104113373), mRNA	XM_009623510.2	<i>N. tomentosiformis</i>	1,586
	<i>Nicotiana tomentosiformis</i> unplaced genomic scaffold, Ntom_v01 Ntom_scaffold26243, whole genome shotgun sequence	NW_008906090.1	<i>N. tomentosiformis</i>	3,092
QUT	<i>N. benthamiana</i> Actin 3 gene	Nbv6.1trP47370 actin 3	<i>N. benthamiana</i>	3,019
SOL	<i>N. attenuata</i> actin	NIATv7_g33830.t1	<i>N. attenuata</i>	1,511
	<i>N. benthamiana</i> Actin	Niben101Scf00096g04015.1	<i>N. benthamiana</i>	3,560
	<i>N. benthamiana</i> Actin 3	Niben101Scf04103g05008.1	<i>N. benthamiana</i>	1,134
	<i>N. tabacum</i> Actin like	Nitab4.5_0003600g0030.1	<i>N. tabacum</i>	3,054

Table 2.6: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *EF1 α* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Arabidopsis thaliana chromosome 5 sequence	NC_003076.8	<i>A. thaliana</i>	2,425
	Arabidopsis thaliana GTP binding Elongation factor Tu family protein mRNA	NM_125432.4	<i>A. thaliana</i>	1,826
	Nicotiana attenuata strain UT chromosome 11, NIATTr2, whole genome shotgun sequence	NC_031999.1	<i>N. attenuata</i>	2,782
	PREDICTED: Nicotiana attenuata elongation factor 1-alpha (LOC109233991), mRNA	XM_019399807.1	<i>N. attenuata</i>	1,779
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold1536, whole genome shotgun sequence	NW_009398358.1	<i>N. sylvestris</i>	2,803
	PREDICTED: Nicotiana sylvestris elongation factor 1-alpha (LOC104231890), mRNA	XM_009784954.1	<i>N. sylvestris</i>	1,776
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold57971, whole genome shotgun sequence	NW_015920008.1	<i>N. tabacum</i>	2,680
	PREDICTED: Nicotiana tabacum elongation factor 1-alpha-like (LOC107791623), mRNA	XM_016613715	<i>N. tabacum</i>	1,837
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold12145, whole genome shotgun sequence	NW_008852325.1	<i>N. tomentosiformis</i>	2,682
	PREDICTED: Nicotiana tomentosiformis elongation factor 1-alpha (LOC104089999), mRNA	XM_009595030.2	<i>N. tomentosiformis</i>	1,839
QUT	Nbv6.1trP57121 elongation factor 1-alpha 1	Nbv6.1trP57121	<i>N. benthamiana</i>	2,058
SOL	N.attenuata v2 annot v5 mRNA EF1a	NIATv7_g03737.t1	<i>N. attenuata</i>	1,835
	N.benthamiana EF1a	Niben101Scf04639g06007.1	<i>N. benthamiana</i>	3,455
	N.benthamiana EF1a	Niben101Scf07423g04011.1	<i>N. benthamiana</i>	3,246
	Translation elongation factor EFTu/EF1A, domain 2	Nitab4.5_0002129g0030.1	<i>N. tabacum</i>	1,808

Table 2.7: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *F-BOX* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Arabidopsis thaliana chromosome 5 sequence	NC_003076.8	<i>A. thaliana</i>	2,520
	Arabidopsis thaliana Galactose oxidase/kelch repeat superfamily protein mRNA	NM_121575.5	<i>A. thaliana</i>	2,040
	Nicotiana attenuata strain UT unplaced genomic scaffold, NIATTr2, whole genome shotgun sequence	NW_017670689.1	<i>N. attenuata</i>	2,385
	PREDICTED: Nicotiana attenuata F-box/kelch-repeat protein At5g15710 (LOC109242322), mRNA	XM_019409192.1	<i>N. attenuata</i>	2,385
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold19732, whole genome shotgun sequence	NW_009446936.1	<i>N. sylvestris</i>	2,380
	PREDICTED: Nicotiana sylvestris F-box/kelch-repeat protein At5g15710 (LOC104236713), mRNA	XM_009790698.1	<i>N. sylvestris</i>	2,294
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold34710, whole genome shotgun sequence	NW_015893957.1	<i>N. tabacum</i>	2,427
	PREDICTED: Nicotiana tabacum F-box/kelch-repeat protein At5g15710-like (LOC107772484), mRNA	XM_016591988.1	<i>N. tabacum</i>	2,427
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold3604, whole genome shotgun sequence	NW_008916975.1	<i>N. tomentosiformis</i>	2,398
	PREDICTED: Nicotiana tomentosiformis F-box/kelch-repeat protein At5g15710 (LOC104120019), mRNA	XM_009631666.2	<i>N. tomentosiformis</i>	2,398
QUT	N.benthamiana FBox f-box kelch-repeat protein at5g15710	Nbv6.1trP36478	<i>N. benthamiana</i>	2,111
SOL	F-box domain, Galactose oxidase, beta-propeller" Nitab4.5_0002978	Nitab4.5_0002978g0090.1	<i>N. attenuata</i>	1,317
	N.attenuata FBox	NIATv7_g29342.t1 F-box/kelch-repeat protein At5g15710	<i>N. attenuata</i>	2,309
	F-box family protein	Niben101Scf02738g07013.1	<i>N. benthamiana</i>	4,690

Table 2.8: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *L23* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Arabidopsis thaliana chromosome 2 sequence	NC_003071.7	<i>A. thaliana</i>	1,993
	Arabidopsis thaliana ribosomal protein L23AA (RPL23AA), mRNA	NM_001161093.2	<i>A. thaliana</i>	1,495
	Nicotiana attenuata strain UT unplaced genomic scaffold, NIATTr2, whole genome shotgun sequence	NW_017670277.1	<i>N. attenuata</i>	581
	PREDICTED: Nicotiana attenuata 60S ribosomal protein L23a-like (LOC109237129), mRNA	XM_019403392.1	<i>N. attenuata</i>	581
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold91299, whole genome shotgun sequence	NW_009583052.1	<i>N. sylvestris</i>	2,207
	PREDICTED: Nicotiana sylvestris 60S ribosomal protein L23a (LOC104227248), mRNA	XM_009779462.1	<i>N. sylvestris</i>	729
	PREDICTED: Nicotiana tabacum 60S ribosomal protein L23a-like (LOC107805175), mRNA	XM_016629168.1	<i>N. tabacum</i>	804
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold81435, whole genome shotgun sequence	NW_015940395.1	<i>N. tabacum</i>	2,282
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold31963, whole genome shotgun sequence	NW_008912445.1	<i>N. tomentosiformis</i>	762
	PREDICTED: Nicotiana tomentosiformis 60S ribosomal protein L23a-like (LOC104117422), mRNA	XM_018778021.1	<i>N. tomentosiformis</i>	762
QUT	60s ribosomal protein l23a	Nbv6.1trP37765	<i>N. benthamiana</i>	1,061
	N.benthamiana 60S ribosomal protein L23a	Nbv6.1trP37766	<i>N. benthamiana</i>	1,032
SOL	N.attenuata v2 annot v5 mRNA L23	NIATv7_g36492.t1	<i>N. attenuata</i>	438
	N.benthamiana 60S ribosomal protein L23a	Niben101Scf01444g02009.1	<i>N. benthamiana</i>	2,525
	N.benthamiana 60S ribosomal protein L23a	Niben101Scf01942g00001.1	<i>N. benthamiana</i>	333
	N.tabacum Ribosomal protein L25/L23	Nitab4.5_0004851g0040.1	<i>N. tabacum</i>	2,277

Table 2.9: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *Ntubc2* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Nicotiana attenuata strain UT unplaced genomic scaffold, NIATTr2	NW_017670988.1	<i>N. attenuata</i>	7,456
	PREDICTED: Nicotiana attenuata ubiquitin-conjugating enzyme E2 2-like (LOC109245034), transcript variant X1, mRNA	XM_019412208.1	<i>N. attenuata</i>	846
	PREDICTED: Nicotiana attenuata ubiquitin-conjugating enzyme E2 2-like (LOC109245034), transcript variant X2, mRNA	XM_019412209.1	<i>N. attenuata</i>	924
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold8714	NW_009578431.1	<i>N. sylvestris</i>	7,983
	PREDICTED: Nicotiana sylvestris ubiquitin-conjugating enzyme E2 2-like (LOC104226439), transcript variant X1, mRNA	XM_009778450.1	<i>N. sylvestris</i>	915
	PREDICTED: Nicotiana sylvestris ubiquitin-conjugating enzyme E2 2-like (LOC104226439), transcript variant X2, mRNA	XM_009778451.1	<i>N. sylvestris</i>	887
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold34512, whole genome shotgun sequence	NW_015892308.1	<i>N. tabacum</i>	13,882
	PREDICTED: Nicotiana tabacum ubiquitin-conjugating enzyme E2 2-like (LOC107772211), transcript variant X1, mRNA	XM_016591702.1	<i>N. tabacum</i>	879
	PREDICTED: Nicotiana tabacum ubiquitin-conjugating enzyme E2 2-like (LOC107772211), transcript variant X2, mRNA	XM_016591703.1	<i>N. tabacum</i>	886
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold45319, whole genome shotgun sequence	NW_008927285.1	<i>N. tomentosiformis</i>	10,118
	PREDICTED: Nicotiana tomentosiformis ubiquitin-conjugating enzyme E2 2-like (LOC104087162), transcript variant X1, mRNA	XM_009591564.2	<i>N. tomentosiformis</i>	915
	PREDICTED: Nicotiana tomentosiformis ubiquitin-conjugating enzyme E2 2-like (LOC104087162), transcript variant X2, mRNA	XM_009591565.2	<i>N. tomentosiformis</i>	912
	PREDICTED: Nicotiana tomentosiformis ubiquitin-conjugating enzyme E2 2-like (LOC104087162), transcript variant X3, mRNA	XM_009591566.2	<i>N. tomentosiformis</i>	905
QUT	N.benthamiana ubiquitin-conjugating enzyme e2 2	Nbv6.1trA105082	<i>N. benthamiana</i>	650
SOL	N.attenuata v2 annot v5 mRNA Ubiquitin	NIATv7_g65609.t1	<i>N. attenuata</i>	816
	N.benthamiana ubiquitin-conjugating enzyme 3	Niben101Scf00339g07001.1	<i>N. benthamiana</i>	6,203
	N.benthamiana ubiquitin-conjugating enzyme 3	Niben101Scf02253g03005.1	<i>N. benthamiana</i>	8,296
	N.tabacum Ubiquitin-conjugating enzyme, E2	Nitab4.5_0008519g0010.1	<i>N. tabacum</i>	8,032

Table 2.10: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *PDF2* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Arabidopsis thaliana chromosome 4 sequence	NC_003075.7	<i>A. thaliana</i>	6,058
	Arabidopsis thaliana protodermal factor 2 (PDF2), mRNA	NM_001340517.1	<i>A. thaliana</i>	3,178
	Nicotiana attenuata strain UT chromosome 7, NIATTr2, whole genome shotgun sequence	NC_031995.1	<i>N. attenuata</i>	7,606
	PREDICTED: Nicotiana attenuata homeobox-leucine zipper protein MERISTEM L1-like (LOC109228868), mRNA	XM_019394107.1	<i>N. attenuata</i>	3,174
	PREDICTED: Nicotiana sylvestris homeobox-leucine zipper protein MERISTEM L1-like (LOC104245814), transcript variant X4, mRNA	XM_009801497.1	<i>N. sylvestris</i>	2,858
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold30866, whole genome shotgun sequence	NW_009515905.1	<i>N. sylvestris</i>	7,143
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold5509, whole genome shotgun sequence	NW_015917398.1	<i>N. tabacum</i>	7,076
	PREDICTED: Nicotiana tabacum homeobox-leucine zipper protein MERISTEM L1-like (LOC107789652), transcript variant X1, mRNA	XM_016611511.1	<i>N. tabacum</i>	3,120
	PREDICTED: Nicotiana tomentosiformis homeobox-leucine zipper protein MERISTEM L1-like (LOC104093789), transcript variant X3, mRNA	XM_018769975.1	<i>N. tomentosiformis</i>	2,703
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold674, whole genome shotgun sequence	NW_008951818.1	<i>N. tomentosiformis</i>	7,026
QUT	N. benthamiana homeobox-leucine zipper protein meristem l1-like	Nbv6.1trA136698	<i>N. benthamiana</i>	3,084
	N.benthamiana homeobox-leucine zipper protein meristem l1-like	Nbv6.1trA271601	<i>N. benthamiana</i>	2,287
SOL	N.attenuata v2 annot v5 mRNA PDF2	NIATv7_g28564.t1	<i>N. attenuata</i>	1,600
	Nicotiana attenuata Homeobox-leucine zipper protein MERISTEM L1	NIATv7_g28564.t1	<i>N. attenuata</i>	2,252
	N.benthamiana Homeobox-leucine zipper family protein / lipid-binding START domain-containing protein	Niben101Scf00703g00003.1	<i>N. benthamiana</i>	9,118
	N.tabacum Homeobox domain, START domain, Homeodomain-like, START-like domain	Nitab4.5_0000091g0520.1	<i>N. tabacum</i>	4,640

Table 2.11: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *PP2A* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Arabidopsis thaliana chromosome 1 sequence	NC_003070.9	<i>A. thaliana</i>	4,422
	Arabidopsis thaliana protein phosphatase 2A subunit A3 (PP2AA3), mRNA	NM_101203.5	<i>A. thaliana</i>	2,217
	Nicotiana attenuata strain UT chromosome 12, NIATTr2, whole genome shotgun sequence	NC_032000.1	<i>N. attenuata</i>	7,422
	PREDICTED: Nicotiana attenuata serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform (LOC109235247), mRNA	XM_019401282.1	<i>N. attenuata</i>	2,261
	PREDICTED: Nicotiana sylvestris serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform (LOC104217882), mRNA	XM_009768212.1	<i>N. sylvestris</i>	2,212
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold55170, whole genome shotgun sequence	NW_009542910.1	<i>N. sylvestris</i>	7,362
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold93067, whole genome shotgun sequence	NW_015950002.1	<i>N. tabacum</i>	7,962
	PREDICTED: Nicotiana tabacum serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform (LOC107810392), mRNA	XM_016635168.1	<i>N. tabacum</i>	2,749
	PREDICTED: Nicotiana tomentosiformis serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A beta isoform (LOC104100878), transcript variant X1, mRNA	XM_009608235.2	<i>N. tomentosiformis</i>	2,278
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold13711, whole genome shotgun sequence	NW_008869725.1	<i>N. tomentosiformis</i>	8,603
QUT	N.benthamiana serine threonine-protein phosphatase 2a 65 kda regulatory subunit a beta isoform	Nbv6.1trA58277	<i>N. benthamiana</i>	2,727
	Nicotiana benthamiana serine threonine-protein phosphatase 2a 65 kda regulatory subunit a beta isoform	Nbv6.1trA211973	<i>N. benthamiana</i>	2,578
SOL	N.attenuata v2 annot v5 mRNA PP2A	NIATv7_g01358.t1	<i>N. attenuata</i>	2,156
	Nicotiana benthamiana Serine/threonine-protein phosphatase 4 regulatory subunit 1	Niben101Scf09716g01002.1	<i>N. benthamiana</i>	9,773
	Nicotiana tabacum HEAT, type 2, Armadillo-type fold, Armadillo-like helica	Nitab4.5_0004895g0070.1	<i>N. tabacum</i>	8,169

Table 2.12: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *SAND* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	PREDICTED: <i>Nicotiana sylvestris</i> protein SAND-like (LOC104244651), transcript variant X1, mRNA	XM_009800118.1	<i>N. sylvestris</i>	2,704
	PREDICTED: <i>Nicotiana sylvestris</i> protein SAND-like (LOC104244651), transcript variant X2, mRNA	XM_009800119.1	<i>N. sylvestris</i>	2,704
	PREDICTED: <i>Nicotiana sylvestris</i> protein SAND-like (LOC104244651), transcript variant X3, mRNA	XM_009800120.1	<i>N. sylvestris</i>	2,369
	<i>Nicotiana sylvestris</i> unplaced genomic scaffold, Nsyl Nsyl_scaffold29225, whole genome shotgun sequence	NW_009514082.1	<i>N. sylvestris</i>	10,639
	PREDICTED: <i>Nicotiana tabacum</i> protein SAND-like (LOC107818978), mRNA	XM_016645059.1	<i>N. tabacum</i>	2,368
	<i>Nicotiana tabacum</i> cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold14666, whole genome shotgun sequence	NW_015820785.1	<i>N. tabacum</i>	9,225
	PREDICTED: <i>Nicotiana tomentosiformis</i> protein SAND-like (LOC104108944), transcript variant X1, mRNA	XM_009618103.2	<i>N. tomentosiformis</i>	2,403
	PREDICTED: <i>Nicotiana tomentosiformis</i> protein SAND-like (LOC104108944), transcript variant X3, mRNA	XM_009618104.2	<i>N. tomentosiformis</i>	2,365
	PREDICTED: <i>Nicotiana tomentosiformis</i> protein SAND-like (LOC104108944), transcript variant X2, mRNA	XM_009618105.2	<i>N. tomentosiformis</i>	2,288
	<i>Nicotiana tomentosiformis</i> unplaced genomic scaffold, Ntom_v01 Ntom_scaffold21083, whole genome shotgun sequence	NW_008900357.1	<i>N. tomentosiformis</i>	7,986
QUT	<i>Nicotiana benthamiana</i> protein sand	Nbv6.1trA265923	<i>N. benthamiana</i>	3,077
SOL	<i>Nicotiana benthamiana</i> protein sand-like isoform x2	Nbv6.1trA77284	<i>N. benthamiana</i>	2,489
	<i>N.attenuata</i> v2 annot v5 mRNA SAND	NIATv7_g20796.t1	<i>N. attenuata</i>	1,300
	<i>Nicotiana benthamiana</i> Vacuolar fusion protein	Niben101Scf00063g06034.1	<i>N. benthamiana</i>	10,615
	<i>Nicotiana benthamiana</i> Vacuolar fusion protein	Niben101Scf00519g00015.1	<i>N. benthamiana</i>	8,449
	<i>Nicotiana tabacum</i> Vacuolar fusion protein MON1	Nitab4.5_0005900g0050.1	<i>N. tabacum</i>	11,001

Table 2.13: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *Ubiquitin* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	PREDICTED: <i>Nicotiana attenuata</i> polyubiquitin 4 (LOC109210540), mRNA	XM_019373968.1	<i>N. attenuata</i>	1,455
	<i>Nicotiana attenuata</i> strain UT unplaced genomic scaffold, NIATTr2, whole genome shotgun sequence	NW_017671849.1	<i>N. attenuata</i>	2,472
	PREDICTED: <i>Nicotiana tabacum</i> polyubiquitin 4-like (LOC107816905), mRNA	XM_016642653.1	<i>N. tabacum</i>	1,459
	<i>Nicotiana tabacum</i> cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold138007, whole genome shotgun sequence	NW_015815015.1	<i>N. tabacum</i>	2,513
	<i>Nicotiana tomentosiformis</i> unplaced genomic scaffold, Ntom_v01 Ntom_scaffold48370, whole genome shotgun sequence	NW_008930675.1	<i>N. tomentosiformis</i>	2,582
	PREDICTED: <i>Nicotiana tomentosiformis</i> polyubiquitin 4 (LOC104088355), transcript variant X1, mRNA	XM_009593017.2	<i>N. tomentosiformis</i>	1,465
	PREDICTED: <i>Nicotiana tomentosiformis</i> polyubiquitin 4 (LOC104088355), transcript variant X2, mRNA	XM_018768157.1	<i>N. tomentosiformis</i>	1,237
QUT	<i>Nicotiana benthamiana</i> polyubiquitin 4-like	Nbv6.1trA4128	<i>N. benthamiana</i>	1,722
SOL	<i>N.attenuata</i> v2 annot v5 mRNA ubiquitin	NIATv7_g25670.t1	<i>N. attenuata</i>	1,388

Table 2.14: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *CPK3* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	AtCPK3 nucleotide	AT4G23650	<i>A. thaliana</i>	1,590
	Nicotiana attenuata strain UT chromosome 3, NIATTr2, whole genome shotgun sequence	NC_031991.1	<i>N. attenuata</i>	6,121
	PREDICTED: Nicotiana attenuata calcium-dependent protein kinase 1-like (LOC109223654), mRNA	XM_019388046.1	<i>N. attenuata</i>	1,939
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold17365, whole genome shotgun sequence	NW_009420636.1	<i>N. sylvestris</i>	5,658
	PREDICTED: Nicotiana sylvestris calcium-dependent protein kinase 3 (LOC104234199), mRNA	XM_009787733.1	<i>N. sylvestris</i>	1,903
	Nicotiana tabacum calcium-dependent protein kinase 3-like (LOC107831716), mRNA	NM_001326266.1	<i>N. tabacum</i>	1,885
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold22001, whole genome shotgun sequence	NW_015861485.1	<i>N. tabacum</i>	5,629
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold24639, whole genome shotgun sequence	NW_008904307.1	<i>N. tomentosiformis</i>	6,126
	PREDICTED: Nicotiana tomentosiformis calcium-dependent protein kinase 1 (LOC104112064), mRNA	XM_009621894.2	<i>N. tomentosiformis</i>	1,954
QUT	N.benth calcium-dependent protein kinase 3	Nbv6.1trA97108	<i>N. benthamiana</i>	1,912
SOL	N.benth calcium-dependent protein kinase 6	Niben101Scf01027g01014.1	<i>N. benthamiana</i>	6,706

Table 2.15: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *SGS3* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Nicotiana sylvestris unplaced genomic scaffold, Nsyl Nsyl_scaffold2858, whole genome shotgun sequence	NW_009513364.1	<i>A. thaliana</i>	2,964
	Arabidopsis thaliana XS domain-containing protein / XS zinc finger domain-containing protein-like protein (<i>SGS3</i>), mRNA	NM_001343816.1	<i>A. thaliana</i>	3,240
	Nicotiana attenuata strain UT chromosome 6, NIATTr2, whole genome shotgun sequence	NC_031994.1	<i>N. attenuata</i>	5,090
	PREDICTED: Nicotiana tomentosiformis protein SUPPRESSOR OF GENE SILENCING 3 (LOC104105584), mRNA	XM_009613932.2	<i>N. attenuata</i>	2,327
	Arabidopsis thaliana chromosome 5 sequence	NC_003076.8	<i>N. benthamiana</i>	1,908
	Nicotiana tabacum protein SUPPRESSOR OF GENE SILENCING 3-like (LOC107800768), mRNA	NM_001325691.1	<i>N. sylvestris</i>	5,237
	PREDICTED: Nicotiana sylvestris protein SUPPRESSOR OF GENE SILENCING 3 (LOC104244181), mRNA	XM_009799542.1	<i>N. sylvestris</i>	2,217
	Nicotiana benthamiana suppressor of gene silencing 3 (<i>SGS3</i>) mRNA, complete cds	KJ190939.1	<i>N. tabacum</i>	1,908
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold17794, whole genome shotgun sequence	NW_008896701.1	<i>N. tabacum</i>	5,279
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold7288, whole genome shotgun sequence	NW_015933080.1	<i>N. tomentosiformis</i>	4,483
	PREDICTED: Nicotiana attenuata protein SUPPRESSOR OF GENE SILENCING 3 (LOC109226845), mRNA	XM_019391730.1	<i>N. tomentosiformis</i>	1,918
QUT	Protein suppressor of gene silencing 3	Nbv6.1trA43726 protein suppressor of gene silencing 3	<i>N. benthamiana</i>	2,349
SOL	Niben101Scf05468	Niben101Scf05468	<i>N. benthamiana</i>	2,100
	Protein SUPPRESSOR OF GENE SILENCING 3	Niben101Scf05468g10025.1	<i>N. benthamiana</i>	5,314

Table 2.16: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *WRKY26* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
NCBI	Nicotiana attenuata strain UT chromosome 7, NIATTr2, whole genome shotgun sequence	NC_031995.1	<i>N. attenuata</i>	3,057
	WRKY transcription factor 1	Niben101Scf01297:502449..506682	<i>N. attenuata</i>	2,084
	Nbv6.1trA108686 probable wrky transcription factor 26 isoform x2	Nbv6.1trA108686	<i>N. benthamiana</i>	1,983
	Nicotiana sylvestris unplaced genomic scaffold, Nsyl1 Nsyl_scaffold28606, whole genome shotgun sequence	NW_009513394.1	<i>N. benthamiana</i>	748
	Nicotiana tabacum cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold32038, whole genome shotgun sequence	NW_015889972.1	<i>N. benthamiana</i>	4,234
	PREDICTED: Nicotiana tomentosiformis probable WRKY transcription factor 26 (LOC104108277), transcript variant X2, mRNA	XM_009617270.2	<i>N. sylvestris</i>	2,046
	PREDICTED: Nicotiana attenuata probable WRKY transcription factor 26 (LOC109229576), mRNA	XM_019395064.1	<i>N. sylvestris</i>	3,057
	Nicotiana tabacum probable WRKY transcription factor 26 (LOC107769908), mRNA	NM_001325041.1	<i>N. tabacum</i>	1,988
	PREDICTED: Nicotiana sylvestris probable WRKY transcription factor 26 (LOC104244191), transcript variant X2, mRNA	XM_009799562.1	<i>N. tabacum</i>	3,083
	WRKY transcription factor 1	Niben101Scf01297	<i>N. tabacum</i>	2,075
	PREDICTED: Nicotiana tabacum probable WRKY transcription factor 26 (LOC107769908), transcript variant X1, mRNA	XM_016589165.1	<i>N. tomentosiformis</i>	2,059
	PREDICTED: Nicotiana tomentosiformis probable WRKY transcription factor 26 (LOC104108277), transcript variant X1, mRNA	XM_009617269.2	<i>N. tomentosiformis</i>	2,055
	Nicotiana tomentosiformis unplaced genomic scaffold, Ntom_v01 Ntom_scaffold2048, whole genome shotgun sequence	NW_008899686.1	<i>N. tomentosiformis</i>	2,889

Table 2.17: Summary of gene and mRNA sequences from related *Nicotiana* species and the original *A. thaliana* sequence published from the literature for the *WRKY70* gene.

Database	Sequence name / identifier	Accession number	Organism	Sequence length (nt)
SOL	WRKY transcription factor 55	Niben101Scf06603g03002.1	<i>N. attenuata</i>	1,288
NCBI	PREDICTED: <i>Nicotiana sylvestris</i> probable WRKY transcription factor 70 (LOC104240678), mRNA	XM_009795552.1	<i>N. attenuata</i>	3,936
	Niben101Scf06603	Niben101Scf06603	<i>N. benthamiana</i>	984
	<i>Nicotiana attenuata</i> strain UT unplaced genomic scaffold, NIATTr2, whole genome shotgun sequence	NW_017671680.1	<i>N. benthamiana</i>	747
	<i>Nicotiana attenuata</i> strain UT unplaced genomic scaffold, NIATTr2, whole genome shotgun sequence	NW_009497837.1	<i>N. benthamiana</i>	2,041
	PREDICTED: <i>Nicotiana tomentosiformis</i> probable WRKY transcription factor 70 (LOC104093636), mRNA	XM_009599408.2	<i>N. sylvestris</i>	1,152
	<i>Nicotiana tomentosiformis</i> unplaced genomic scaffold, Ntom_v01 Ntom_scaffold6676, whole genome shotgun sequence	NW_008951107.1	<i>N. sylvestris</i>	2,831
	probable wrky transcription factor 70	Nbv6.1trA140049	<i>N. tabacum</i>	1,146
	PREDICTED: <i>Nicotiana attenuata</i> probable WRKY transcription factor 70 (LOC109209659), mRNA	XM_019372971.1	<i>N. tabacum</i>	2,825
	PREDICTED: <i>Nicotiana tabacum</i> probable WRKY transcription factor 70 (LOC107820490), mRNA	XM_016646779.1	<i>N. tomentosiformis</i>	1,094
	<i>Nicotiana tabacum</i> cultivar TN90 unplaced genomic scaffold, Ntab-TN90 Ntab-TN90_scaffold153748, whole genome shotgun sequence	NW_015825287.1	<i>N. tomentosiformis</i>	3,464

2.3.6 Multiple sequence alignment of mRNA and gene sequences for candidate reference genes

Multiple sequence alignments of the gene and mRNA sequences for the candidate reference genes were carried out. This was done to ascertain the likely structure of the gene in *N. glutinosa* (Figure 2.3 to Figure 2.11 organised alphabetically by candidate reference gene name) in order to facilitate designing the primers around an intron, preferably towards the 3' end of the gene. Exons were identified by regions with a high consensus identity, marked as green on the top bar of the alignments, and introns identified by highly variable stretches of low consensus, marked as red on the top bar of the alignments. The longer the stretch of green, the longer the conserved sequence.

Whilst these are not necessarily the actual gene structures that may be found in *N. glutinosa*, they provide the alignment and indication of where introns may be located, allowing primers to be designed to the conserved sequences around them. By selecting regions conserved across all gene and mRNA sequences either side of an intron, it is likely a primer will anneal to the *N. glutinosa* mRNA sequence. It is important to note that the multiple sequence alignments had limited data, and are not definitive alignments, and that further sequences as published in the future would provide further support for these alignments. Alternatively, gene structures could be confirmed by sequencing genes or full genome of *N. glutinosa*.

As no information has been published related to these genes specifically for *N. glutinosa*, a brief summary of the structure of each alignment is provided. Table 2.18 summaries the components of the multiple sequence alignments, and number of introns and exons identified.

Table 2.18: Summary of the multiple sequence alignments and information obtained for the candidate reference genes.

Gene	Number of gene sequences included	Number of mRNA sequences included	Length of multiple sequence alignment	Number of exons/introns identified
<i>Actin</i>	7	9	3638	5/4
<i>EF1α</i>	6	7	3536	3/2
<i>F-BOX</i>	6	6	2516	1/0
<i>L23</i>	4	7	2598	4/3
<i>Ntubc2</i>	7	12	16005	5/4
<i>PDF2</i>	6	8	10327	12/11
<i>PP2A</i>	6	7	11374	13/12
<i>SAND</i>	6	11	14550	14/13
<i>Ubiquitin</i>	3	6	2914	2/1

2.3.6.1 *Actin* multiple sequence alignment

Four introns and five exons were identified from the multiple sequence alignment of the sixteen *Actin* sequences, consisting of seven gene sequences and nine mRNA sequences (Figure 2.3). Consensus identity was high for the exons. Areas around the intron/exon junctions seemed to have potential sites for primer design. Conserved sequences either side of intron 3 appeared suitable for primer design.

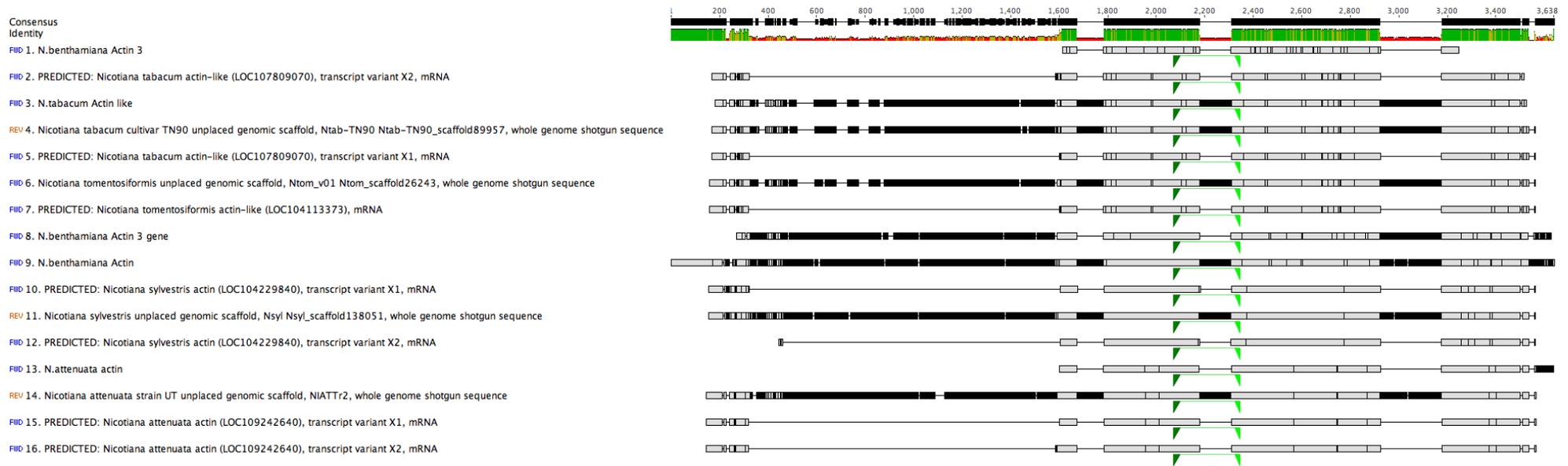


Figure 2.3: Schematic representation of the *Actin* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.6.2 *EF1α* multiple sequence alignment

Two introns and three exons were identified from the multiple sequence alignment of the thirteen *EF1α* sequences, consisting of six gene sequences and seven mRNA sequences (Figure 2.4). Consensus identity was high for the exons, becoming more variable towards the 3' end of the sequences. Conserved sequences either side of intron 2 appeared suitable for primer design.

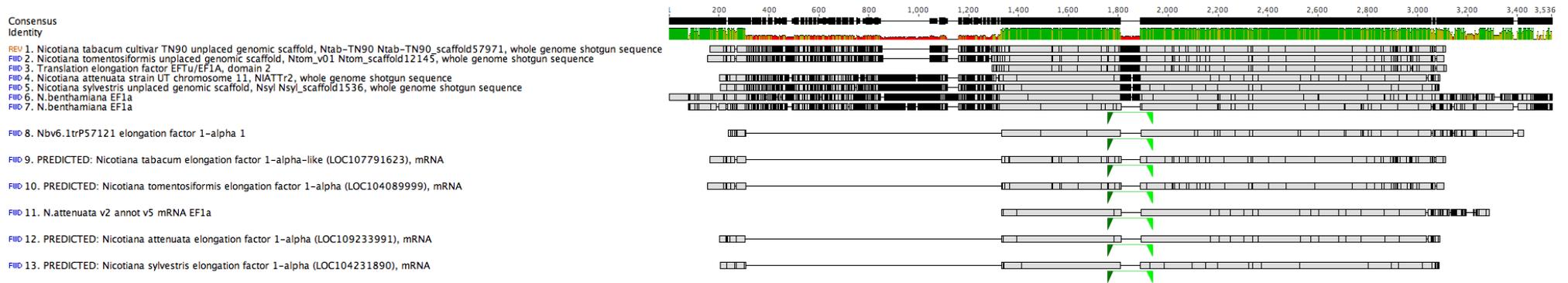


Figure 2.4: Schematic representation of the *EF1α* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.6.3 *F-BOX* multiple sequence alignment

No introns and only one exon were identified from the multiple sequence alignment of the twelve *F-BOX* sequences, consisting of seven gene sequences and six mRNA sequences (Figure 2.5). Consensus identity was generally high for the exon. However, as there was no intron-exon junction around which to design a primer to reduce the possibility of DNA amplification, it was decided not to use this gene in the gene expression study.

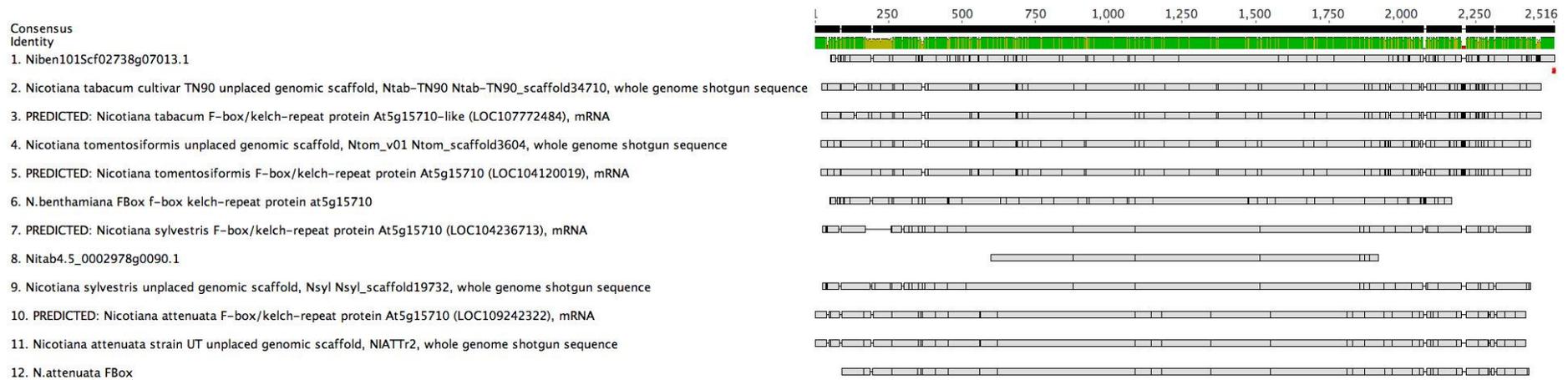


Figure 2.5: Schematic representation of the *F-BOX* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences.

2.3.6.4 *L23* multiple sequence alignment

Three introns and four exons were identified from the multiple sequence alignment of the thirteen *L23* sequences, consisting of four gene sequences and seven mRNA sequences (Figure 2.6). Consensus identity was variable for the exons, becoming more variable towards the 3' end of the sequence. Areas around the intron/exon junctions had limited potential sites for primer design due to short exon length and variable regions towards the edges of the exons. However, no suitable primer pair could be designed so the gene was removed from the research.

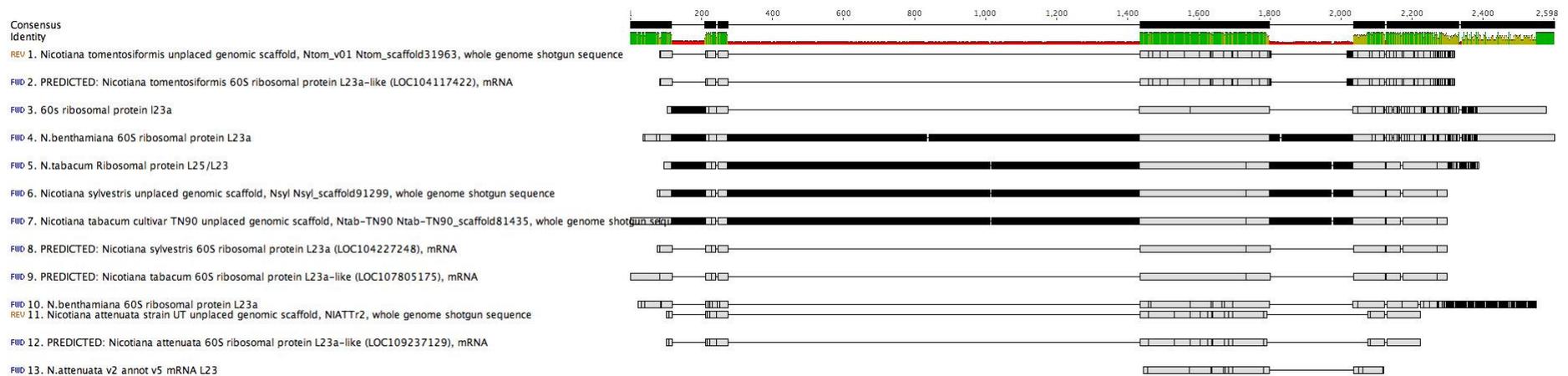


Figure 2.6: Schematic representation of the *L23* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences.

2.3.6.5 *Ntubc2* multiple sequence alignment

Four introns and five exons were identified from the multiple sequence alignment of the nineteen *Ntubc2* sequences, consisting of seven gene sequences and twelve mRNA sequences (Figure 2.7). Consensus identity was medium for the exons, most of which were very short compared to other genes analysed in this study. Areas around the intron/exon junctions had limited sites for primer design, due to the shortness of the exons, however a primer pair was designed around intron 4.

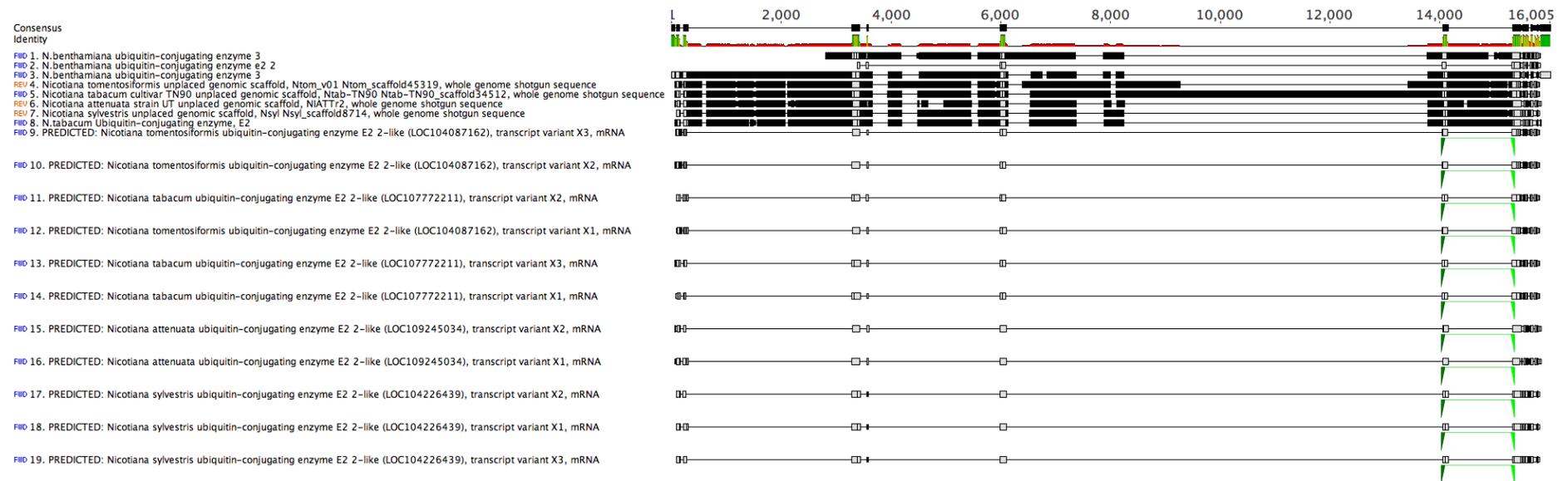


Figure 2.7: Schematic representation of the *Ntubc2* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.6.6 *PDF2* multiple sequence alignment

Eleven introns and twelve exons were identified from the multiple sequence alignment of the fourteen *PDF2* sequences, consisting of six gene sequences and eight mRNA sequences (Figure 2.8). Consensus identity was variable within some of the exons. Areas around the intron/exon junctions seemed to have potential sites for primer, with several short introns between exons being present in the alignments. A primer pair was designed around intron 7, due to the high sequence consensus available either side of this intron compared to others.

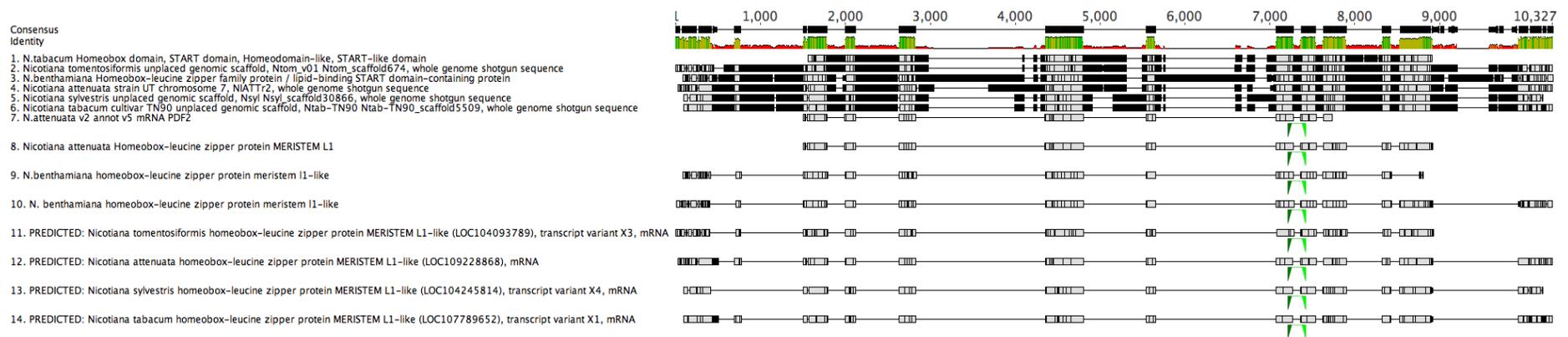


Figure 2.8: Schematic representation of the *PDF2* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.6.7 *PP2A* multiple sequence alignment

Twelve introns and eleven exons were identified from the multiple sequence alignment of the thirteen *PP2A* sequences, consisting of six gene sequences and seven mRNA sequences (Figure 2.9). Consensus identity was variable within some of the exons, most of which were relatively short compared to other exons on other multiple sequence alignments in this study. Areas around the intron/exon junctions seemed to have potential sites for primer design. A primer pair was designed around intron 8, due to the high sequence consensus available either side of this intron compared to others.

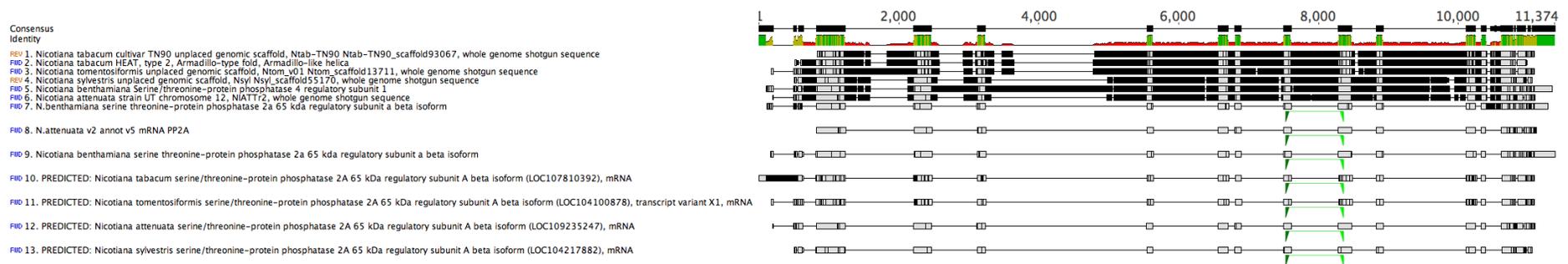


Figure 2.9: Schematic representation of the *PP2A* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.6.8 SAND multiple sequence alignment

Thirteen introns and twelve exons were identified from the multiple sequence alignment of the seventeen *SAND* sequences, consisting of six gene sequences and eleven mRNA sequences (Figure 2.10). Consensus identity was variable within some of the exons, most of which were relatively short compared to other exons in other multiple sequence alignments in this study. Though many intron/exon junctions were present, limited areas were identified as candidate regions for primer design. A primer pair was designed, around position 970 on the alignment due to the high sequence consensus available either side of this intron compared to others.

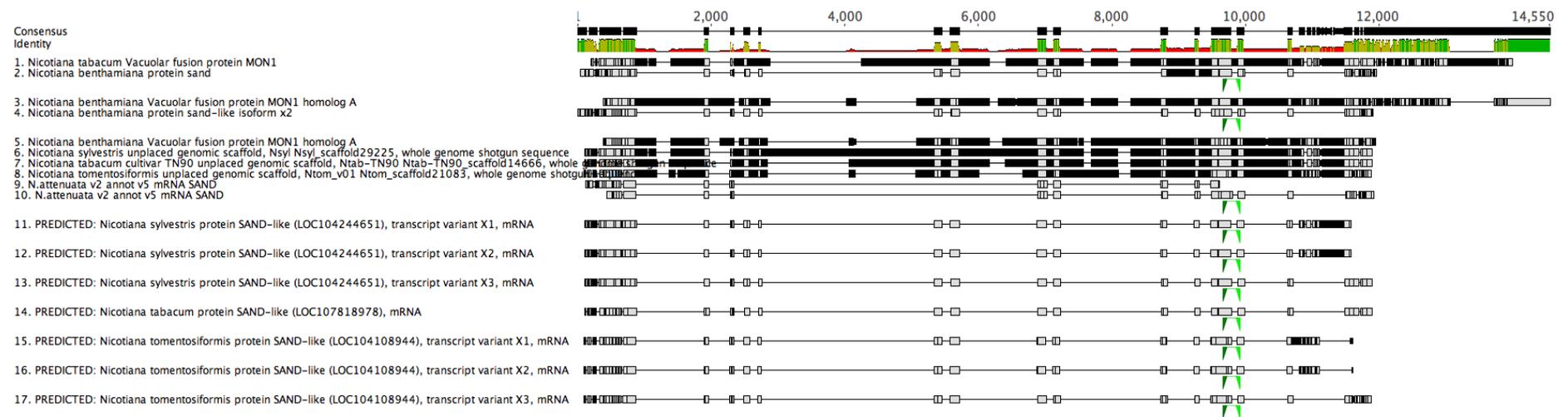


Figure 2.10: Schematic representation of the *SAND* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.6.9 Ubiquitin multiple sequence alignment

One intron and two exons were identified from the multiple sequence alignment of the nine *Ubiquitin* sequences, consisting of three gene sequences and six mRNA sequences (Figure 2.11). Consensus identity was high for the exons, becoming more variable towards the 3' end of the sequences. Areas around the intron/exon junctions seemed to have potential sites for primers to anneal to, with several nucleotide regions conserved across the mRNA and gene transcripts from the different species. A primer pair was designed around intron 1, due to the high sequence consensus available either side of this intron compared to others.

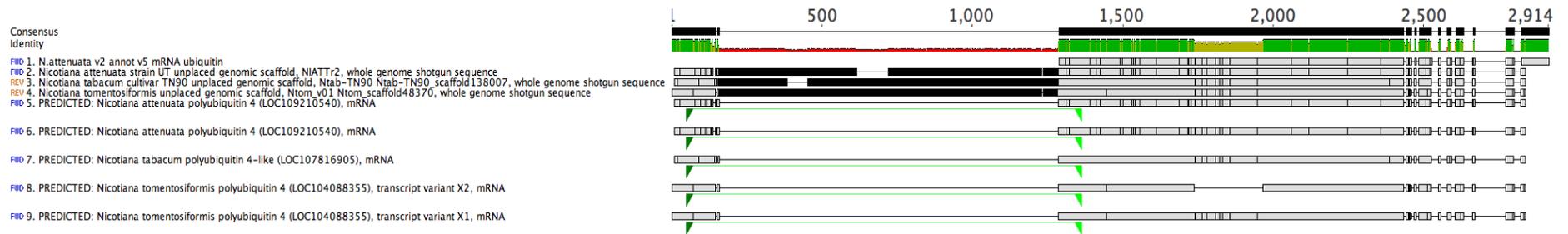


Figure 2.11: Schematic representation of the *Ubiquitin* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.7 Multiple sequence alignment of mRNA and gene sequences for GOIs

Multiple sequence alignments were also conducted for the GOIs in the same manner as for the candidate reference genes described in Section 2.3.6, presented below in Figure 2.12 to Figure 2.15, organised alphabetically by GOI gene name. Table 2.19 summarises the components of the multiple sequence alignments, and number of introns and exons identified.

Table 2.19: Summary of the multiple sequence alignments and information obtained for the GOIs in this research

Gene	Number of gene sequences	Number of mRNA sequences	Length of multiple sequence alignment	Number of exons/introns identified
<i>CPK3</i>	5	5	7655	8/7
<i>SGS3</i>	6	6	5835	6/5
<i>WRKY26</i>	5	7	4932	5/4
<i>WRKY70</i>	4	4	4250	3/2

2.3.7.1 *CPK3* multiple sequence alignment

Seven introns and eight exons were identified from the multiple sequence alignment of the ten *CPK3* sequences, consisting of five gene sequences and five mRNA sequences (Figure 2.12). Consensus identity was high for the exons, becoming more variable towards the 3' end of the sequences. Areas around the intron/exon junctions seemed to have potential sites for primers to anneal to. Sequences around intron 2 appeared to be suitable for primer design due to the high sequence consensus available either side of this intron compared to others.

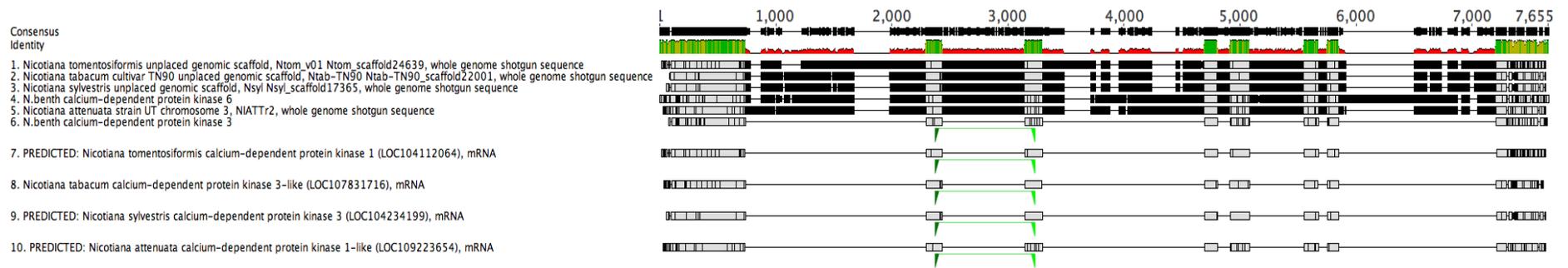


Figure 2.12: Schematic representation of the *CPK3* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.7.2 *SGS3* multiple sequence alignment

Five introns and six exons were identified from the multiple sequence alignment of the twelve *SGS3* sequences, consisting of six gene sequences and six mRNA sequences (Figure 2.13). Consensus identity was variable for some of the exons, with the sequence alignments being more variable at both the 5' and 3' ends of the alignment. Areas around the intron/exon junctions seemed to have potential as primer binding sites. Sequences around intron 2 appeared to be suitable for primer design due to the high sequence consensus available either side of this intron compared to others.

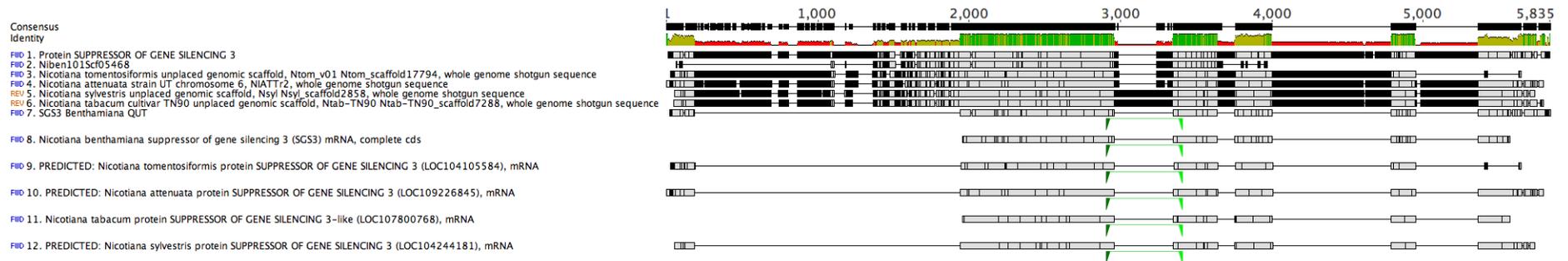


Figure 2.13: Schematic representation of the *SGS3* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.7.3 *WRKY26* multiple sequence alignment

Four introns and five exons were identified from the multiple sequence alignment of the twelve *WRKY26* sequences, consisting of five gene sequences and seven mRNA sequences (Figure 2.14). Consensus identity was high for the exons. Areas around the intron/exon junctions appeared to be sufficiently conserved for primer design, and sequences around intron 2 were selected for this.

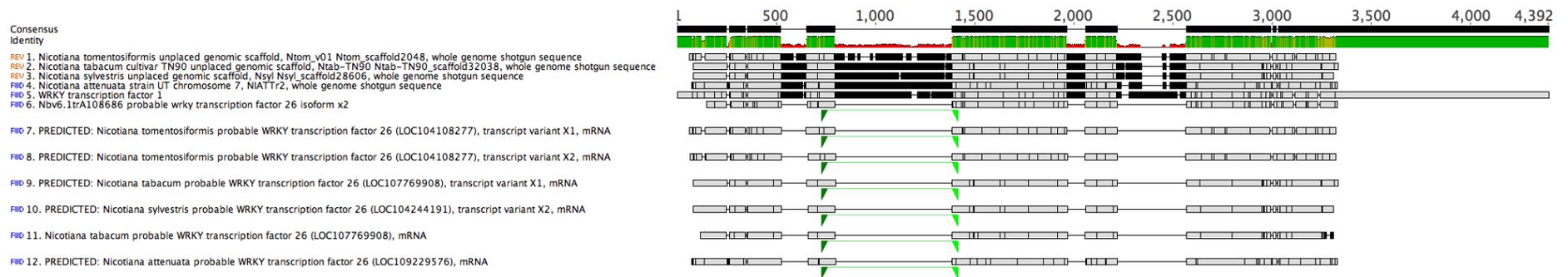


Figure 2.14: Schematic representation of the *WRKY26* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.7.4 *WRKY70* multiple sequence alignment

Two introns and three exons were identified from the multiple sequence alignment of the eight *WRKY70* sequences, consisting of four gene sequences and four mRNA sequences (Figure 2.15). Consensus identity was high for the exons, and it was noted that the alignment had one of the longest introns at nearly 3000nts, so primer design was not conducted around this intron as it may be difficult to amplify. Areas around the other intron/exon junctions seemed to have potential sites for primers to anneal too, with several nucleotide regions conserved across the mRNA and gene transcripts from the different species. Sequences around intron 2 appeared to be suitable for primer design due to the high sequence consensus available either side of this intron compared to others.

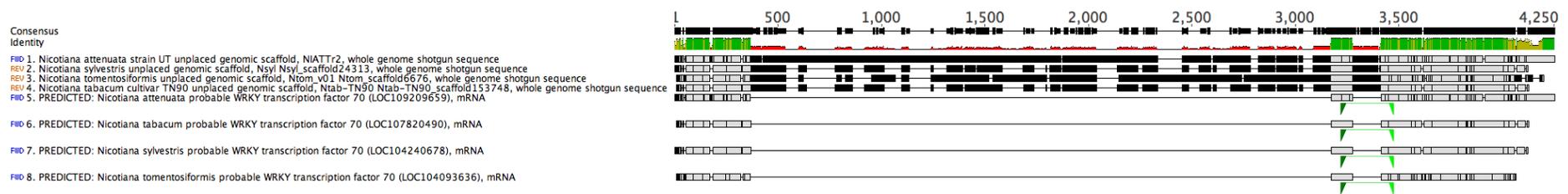


Figure 2.15: Schematic representation of the *WRKY70* multiple sequence alignment consisting of gene and mRNA sequences from related *Nicotiana* species. Regions with green above them indicate areas of high conservation (indicated as “consensus identity”) between the different species, ranging down to red for regions of high variability and low homology. Vertical black lines indicate individual nucleotide differences compared to other sequences. The joined dark green / light green bar on the individual sequences indicates the position of the designed primer pair - arrow heads indicate the forward and reverse primers.

2.3.8 Testing of existing primers on multiple sequence alignments

From previous research, primers to amplify the different genes had been designed and reported in the literature (Lilly et al. 2011; Liu et al. 2012; Senthil et al. 2005). These primer sequences were obtained and using the Primer function in Geneious, were annotated onto the multiple sequence alignments to determine if the primers would likely anneal around an exon and amplify a product size of the correct size for use in RT-qPCR (not shown). None of the primers bound to the multiple sequence alignments in a region that was conserved across all sequences; therefore, the primer specificity in *N. glutinosa* may not have been as high as desired if they were used experimentally. Therefore, the decision was made to redesign all primers for this experiment around intron/exon junctions identified in the multiple sequence alignments, in regions that were conserved across all mRNA and gene sequences to give the highest probability of successful amplification in *N. glutinosa*.

2.3.9 Primer design for candidate genes

Designing a pair of primers for each gene underwent many iterations, all of which are not individually reported here due to the number of variables involved. An intron/exon junction was selected, and the criteria mentioned in Section 2.2.8 were applied to design primer pairs that fulfilled the suggested characteristics primers should have for a qPCR experiment. For several of the primer pairs the majority of the characteristics were achieved, a summary of which are presented in Table 2.20. However, for the reference genes, *L23* and *Ubiquitin*, no suitable location for a primer pair was identified, meaning that these genes had to be removed from the research. This is not to say the genes cannot be amplified in the plant species, but it was not possible with the current multiple sequence alignment from the limited gene and mRNA sequences to design primers to do so.

Table 2.20: Summary of characteristics of primers designed to amplify candidate genes in the qPCR study.

	Gene name	Primer orientation	Sequence	Start position	Length(nt)	GC Content (%)	T _m (° C)	Min folding free energy at 25° C (kcal/mole)	Strongest folding temperature (° C)	Homo sapiens BLAST match?	Expected product size (bp)
Candidate reference genes	<i>Actin</i>	F	GCCAATCGAGAAAAGATGACTCAGATCATG	2070	28	42.9	58.5	-0.67	31.4	No	146
		R	GTGTGGCTGACACCATCACCAGAG	2347	24	58.3	61.7	-1.66	42.9	No	
	<i>EF1α</i>	F	GCAYTGCTTGCTTTACCCTTGG	1757	23	54.3	60.5	-0.96	36.7	No	104
		R	ACGATTCATCGTACCTAGCCTTGG	1939	25	48.0	58.5	-2.75	54.3	No	
	<i>Ntubc2</i>	F	GCTGATGGAAGTATTTGCTTGGACATC	14050	27	44.4	58.1	0.54	30.0	No	116
		R	GGCGAGTTAGGATTTGGATCACAGAGC	15350	27	51.9	60.5	2.09	44.3	No	
	<i>PDF2</i>	F	CCTGCTGGACTTTGGCATTATGTATCG	7470	29	44.8	59.7	-1.49	36.7	No	113
		R	AGGTCTGTAGATATTATGGACAGCTCTATC	7649	29	37.9	56.3	-0.94	31.6	No	
	<i>PP2A</i>	F	CGAGTTTCTGATGTGCGCCTGAAC	7549	25	56.0	62.1	0.77	5.2	No	149
		R	GTCCTCTGCTAGCTCAACAATAGCTGG	8349	27	51.9	60.5	-2.09	44.3	No	
<i>SAND</i>	F	CCTGCTGGACTTTGGCATTATGTATCG	9253	28	42.9	58.1	-1.49	36.7	No	142	
	R	CATGCATGGATGCATAAAGCTTCTGGTAAG	9916	30	43.3	59.7	-4.05	46.5	No		
Candidate GOIs	<i>CPK3</i>	F	GGAGTGCTGGAGTTATCCTGTACATTC	3208	27	48.1	58.6	-2.71	41.6	No	150
		R	GATCTTTAGCACTACTCGATACTGAAGGC	4716	29	44.8	58.2	0.25	20.3	No	
	<i>SGS3</i>	F	GCCACCAATGGTGATTATCATGAACAC	2905	27	44.4	58.3	-2.47	50.4	No	118
		R	AGCCTTGACAGCAGCATAAGAGC	3408	23	52.2	59.7	-0.32	29.5	No	
	<i>WRKY26</i>	F	TCTTTCCAAAGTAGGGCTGCTACTTCATC	728	29	44.8	59.9	-3.41	51.4	No	96
		R	GCATGTTGTTGRCTTGTCAATTAAYTCTTCC	1415	30	40.0	58.5	-0.91	8.4	No	
	<i>WRKY70</i>	F	GGYCATGCTTGGAGAAAATATGGAC	3222	25	46.0	57.2	-0.45	15.1	No	117
		R	CTGCACCTGTTTGGTTGCTTGAC	3474	23	52.2	52.2	-1.60	34.8	No	

2.3.10 Primer testing on LNYV infected and uninfected *N. glutinosa* leaf material

2.3.10.1 RNA concentration and quality assessment

Total RNA was extracted from confirmed LNYV infected and uninfected *N. glutinosa* samples to confirm the primers correctly amplify a product of the expected size with no non-specific products. The uninfected sample was a LNYV mock inoculated *N. glutinosa* plant sampled 28 days post inoculation and the infected sample was grown at the same time and also sampled after 28 days. RNA quality was confirmed using agarose gel electrophoresis, with visible 28S and 18S ribosomal RNA being observable for both samples (data not shown). These were deemed acceptable samples for use in the subsequent end point RT-PCR experiment.

2.3.10.2 Candidate reference gene primer RT-PCR confirmation at 50°C

All primers were initially tested in an end point RT-PCR using the LNYV positive and negative samples with an annealing temperature of 50°C. As there was no positive control to test the primers it was decided to use the primer pair identified as LNYV_440F and LNYV_1185R, designed by Higgins et al., 2016, on the positive sample to confirm correct amplification. Though this was an incorrect experimental design due to the newly designed primers being designed with an annealing temperature of 60°C, the experiments will be reported here as they provided initial information about the workings of the primers experimentally.

Since the optimal annealing temperature for LNYV-440F and LNYV-1185R from previous studies is 50°C, it was decided to also test the newly designed primers at 50°C so that all reactions could be run in the PCR machine simultaneously. The LNYV primers gave the expected product size of approximately 800 bp in LNYV infected leaves but not in the mock inoculated or NTC samples (Figure 2.16 to Figure 2.25).

2.3.10.2.1 *Actin* primers

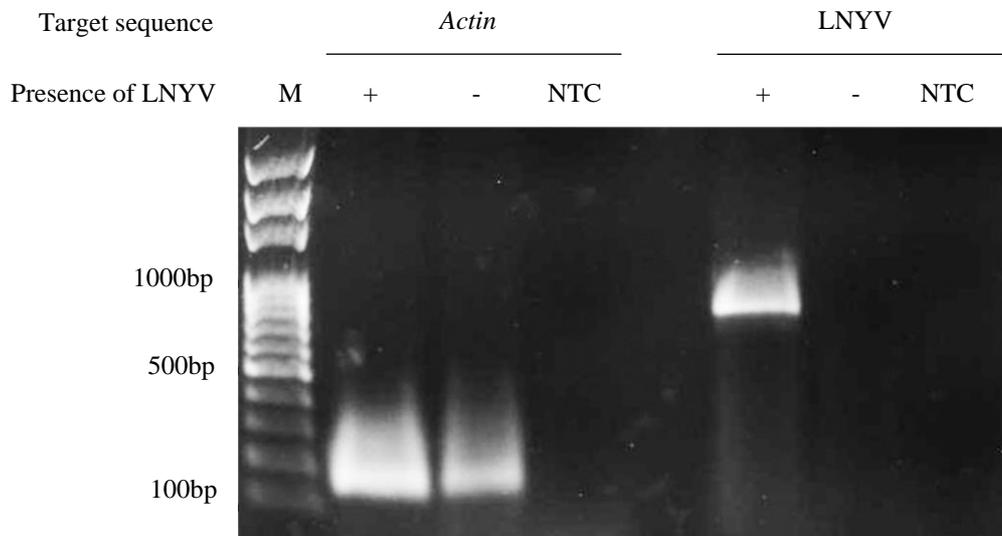


Figure 2.16: 1.5% agarose TBE gel confirming amplification of the *Actin* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Products of approximately 146 bp were amplified in both LNYV positive and negative samples, as expected (Figure 2.16). This is the expected size for amplification of mRNA rather than DNA, indicating no DNA contamination. The smeared bands may have been due to the discrepancy between the designed T_m of the primers and the annealing temperature used in the reaction leading to some non-specific binding of the primers to the template, or, it is efficient amplification and the gel was overloaded. Whilst amplification occurred, confirming that the primers do amplify a product of the expected size, it was necessary to test the *Actin* primers at 60°C, that is closer to the predicted T_m to confirm they would work optimally at the temperature required for the RT-qPCR reaction.

2.3.10.2.2 *EF1α* primers

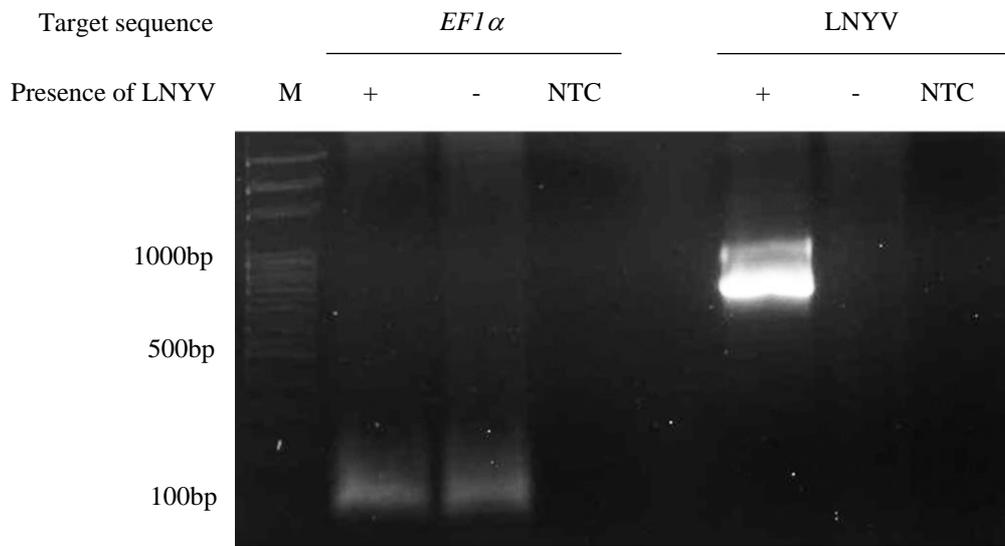


Figure 2.17: 1.5% agarose TBE gel confirming amplification of the *EF1α* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Figure 2.17 shows the expected product size of approximately 104 bp for the *EF1α* mRNA in both LNYV positive and negative leaf samples. Again, no amplification of genomic DNA is apparent since no higher molecular weight products were present. Testing the primers at 60°C is the next step to ensure amplification can occur under the conditions used for RT-qPCR.

2.3.10.2.3 *Ntubc2* primers

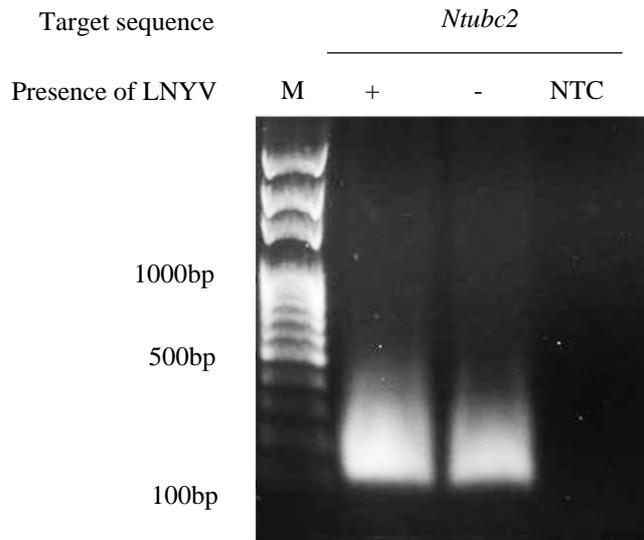


Figure 2.18: 1.5% agarose TBE gel confirming amplification of the *Ntubc2* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Products of approximately 116 bp were amplified in both LNYV positive and negative samples, as expected (Figure 2.18). The smeared bands may have been due to the same reasons as those outlined in Section 2.3.10.2.1. Testing with the LNYV specific primers was carried out at the same time; the results for this were run on a separate gel and can be seen in Figure 2.20. Again, no amplification of genomic DNA is apparent and it is necessary to test the primers at 60°C.

2.3.10.2.4 *PDF2* primers

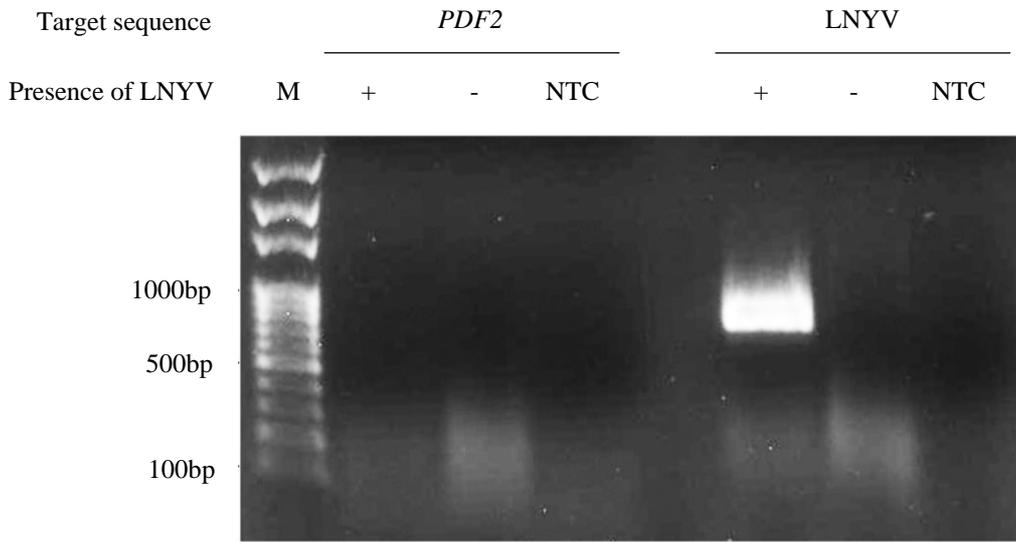


Figure 2.19: 1.5% agarose TBE gel confirming amplification of the *PDF2* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

The expected fragment size for the *PDF2* mRNA was 113 bp. Amplification with these primers resulted in a smear rather than a discrete band (Figure 2.19). As there is also primer dimer for LNYV amplification it is possible the primers were not the problem, rather the reaction conditions may need optimising. Though the amplification was non-specific in this reaction, it was decided to subsequently test the *PDF2* primers at 60°C to determine if they would work at the temperature required for the RT-qPCR reaction.

2.3.10.2.5 *PP2A* primers

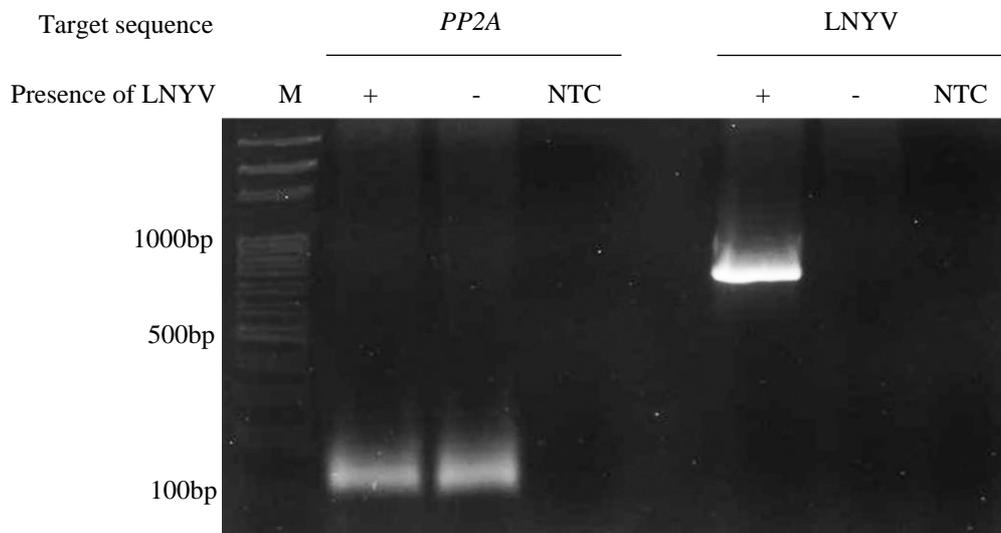


Figure 2.20: 1.5% agarose TBE gel confirming amplification of the *PP2A* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Products of approximately 149 bp were amplified in both LNYV positive and negative samples, as expected (Figure 2.20). The smeared bands suggest amplification was possibly less efficient, resulting in less product being formed. An additional product was identified by the presence of a faint band in the LNYV positive sample with *PP2A* primers at approximately 800 bp, a band not present in the negative sample, and was likely to be contamination with the LNYV specific primers. Despite this possible contamination, as the NTC's were negative, it was decided to test the primers at 60°C to determine their suitability for RT-qPCR reaction.

2.3.10.2.6 SAND primers

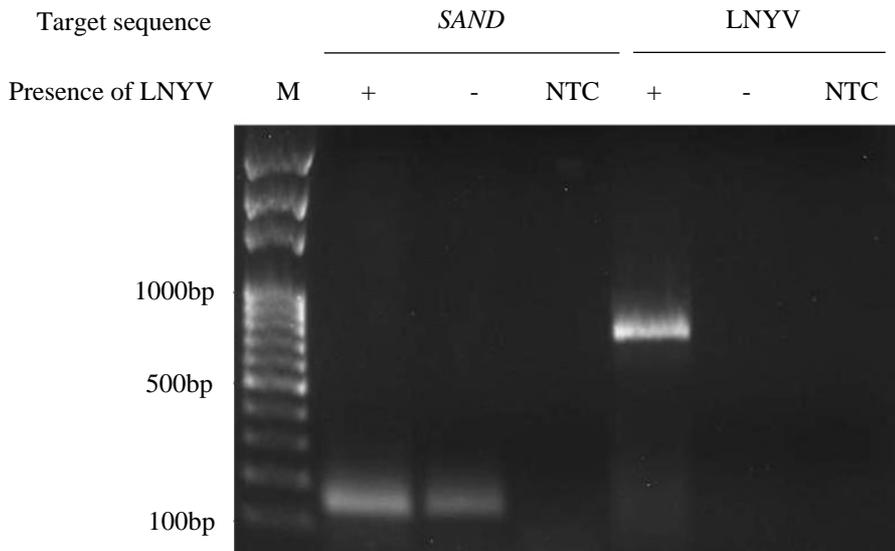


Figure 2.21: 1.5% agarose TBE gel confirming amplification of the *SAND* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Figure 2.21 shows products of approximately 142 bp were amplified in both LNYV positive and negative samples, as expected. Whilst 50°C is not the annealing temperature the primers were designed to be optimally used at, it demonstrates that they may be able to work over a range of temperatures, which can be explored in future studies. It was decided to subsequently test the *SAND* primers at 60°C to determine if they will work at the temperature required for the qPCR reaction.

2.3.10.3 GOIs primer RT-PCR confirmation at 50°C

All primers for the GOIs were also initially tested in an RT-PCR using the LNYV positive and negative samples with an annealing temperature of 50°C. The LNYV_440F and LNYV_1185R primers (Higgins et al., 2016) were again used as the positive control.

2.3.10.3.1 CPK3 primers

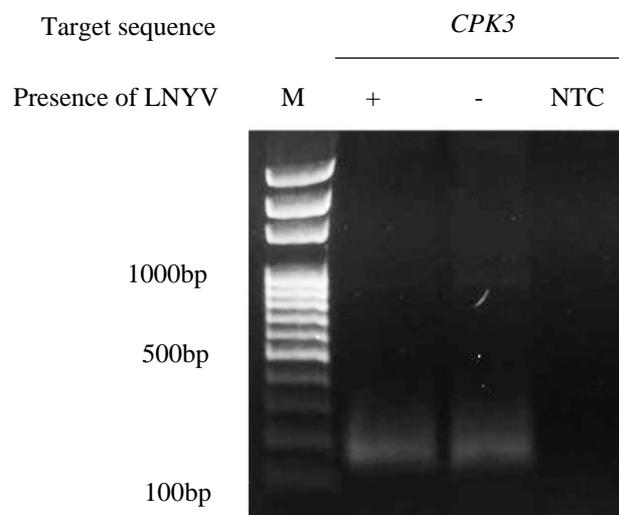


Figure 2.22: : 1.5% agarose TBE gel confirming amplification of the *CPK3* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Figure 2.22 shows the expected product size of approximately 150 bp for the *CPK3* mRNA in both LNYV positive and negative leaf samples. Testing with LNYV specific primers was carried out at the same time; the results were run on a separate gel and can be seen in Figure 2.23. Again, no amplification of genomic DNA is apparent since no higher molecular weight products were present. Testing the primers at 60°C is the next step to ensure amplification can occur under the conditions used for RT-qPCR.

2.3.10.3.2 SGS3 primers

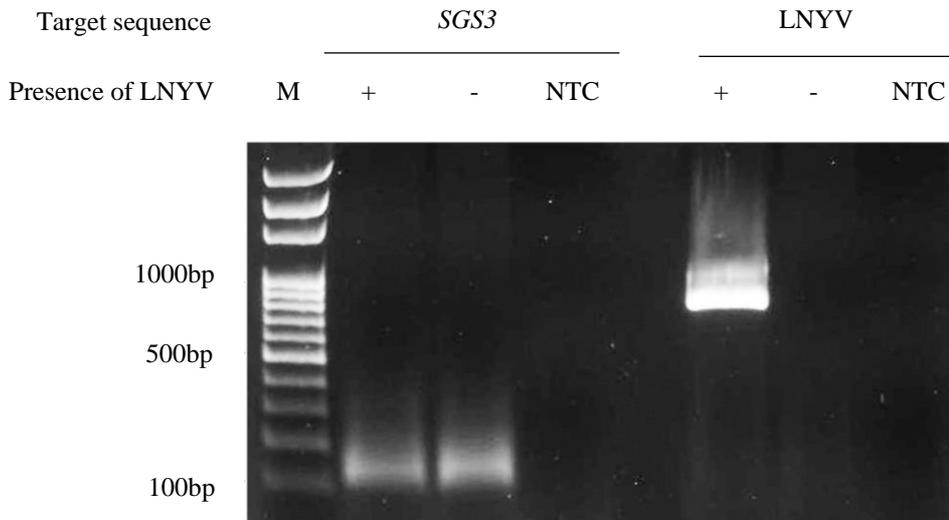


Figure 2.23: 1.5% agarose TBE gel confirming amplification of the *SGS3* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

mRNA products of approximately 118 bp were amplified in both LNYV positive and negative samples, as expected (Figure 2.23). No amplification of genomic DNA is apparent, but the smeared bands suggest possibly amplification was less efficient, resulting in less product being formed. However, it is necessary to test the primers at 60°C to determine their suitability for RT-qPCR reaction.

2.3.10.3.3 *WRKY26* primers

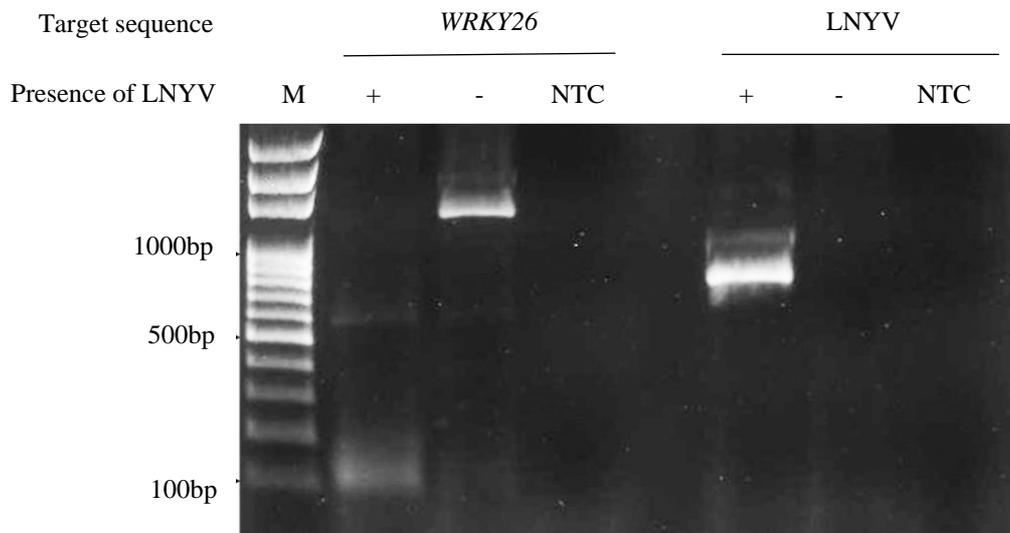


Figure 2.24: 1.5% agarose TBE gel confirming amplification of the *WRKY26* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

There was amplification of products in both the infected and uninfected samples using the *WRKY26* primers (Figure 2.24). However, multiple products were obtained in both samples instead of the expected one product of approximately 96 bp. The LNYV infected sample has a second product of approximately 600 bp in addition to a product of about 120 bp. The LNYV uninfected sample has a product of approximately 1500 bp and additionally a faint product at approximately 600 bp. This indicates possible non-specific binding of the primers to sequences unknown, indicating that 50°C may not be an appropriate annealing temperature, or the primers are not suitable to amplify this sequence. It was decided to subsequently test the *WRKY26* primers at 60°C to determine if more specific amplification would occur.

2.3.10.3.4 *WRKY70* primers

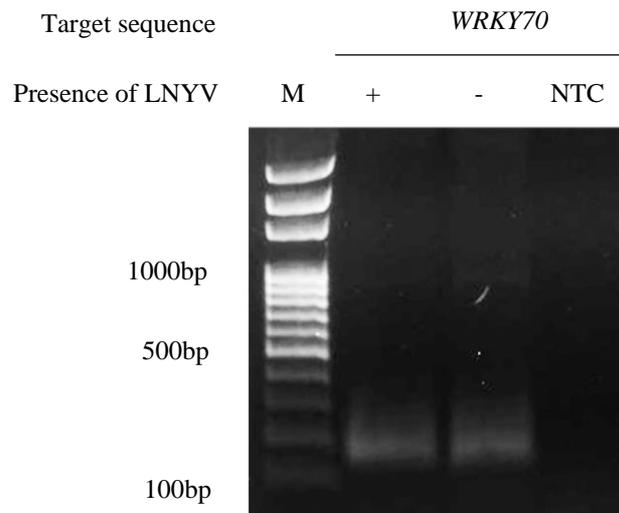


Figure 2.25: 1.5% agarose TBE gel confirming amplification of the *WRKY70* mRNA from LNYV infected (+) and uninfected (-) *N. glutinosa* sample. RT-PCR was carried out using 50°C as the annealing temperature.

Positive control for this experiment was the LNYV + sample.

Figure 2.25 shows the expected product size of approximately 117 bp for the *WRKY70* mRNA in both LNYV positive and negative leaf samples. Again, no amplification of genomic DNA is apparent since no higher molecular weight products are present. Testing with LNYV specific primers was carried out at the same time; the results were run on a separate gel and can be seen in Figure 2.24. Testing the primers at 60°C is the next step to ensure amplification can occur under the conditions used for RT-qPCR.

2.3.11 Candidate reference gene primer RT-PCR confirmation at 60°C

RT-PCR was repeated with uninfected *N. glutinosa* leaf samples, using the higher annealing temperature of 60°C. This temperature was tested as it is advised to carry out RT-qPCR experiments at higher annealing temperatures to avoid non-specific amplification (Ruiz-Villalba et al. 2017). Amplification of the expected product sizes was observed for the *EF1 α* , *SAND*, *PP2A*, *Ntubc2* and *Actin* genes (Figure 2.26a). There was no discrete band for *PDF2* at this temperature, which suggested that this primer pair is not capable of amplifying the correct product. This was in keeping with the observations when the annealing temperature was 50°C (Section 2.3.10.2.4). It was decided that, as these primers were unable to amplify the mRNA as intended, this sequence would not be included as a candidate reference gene. A single gene product was observed for the other genes. There was no contamination in any of the NTC reactions for this experiment (Figure 2.26b), thus, the other primers were considered suitable for use in RT-qPCR.

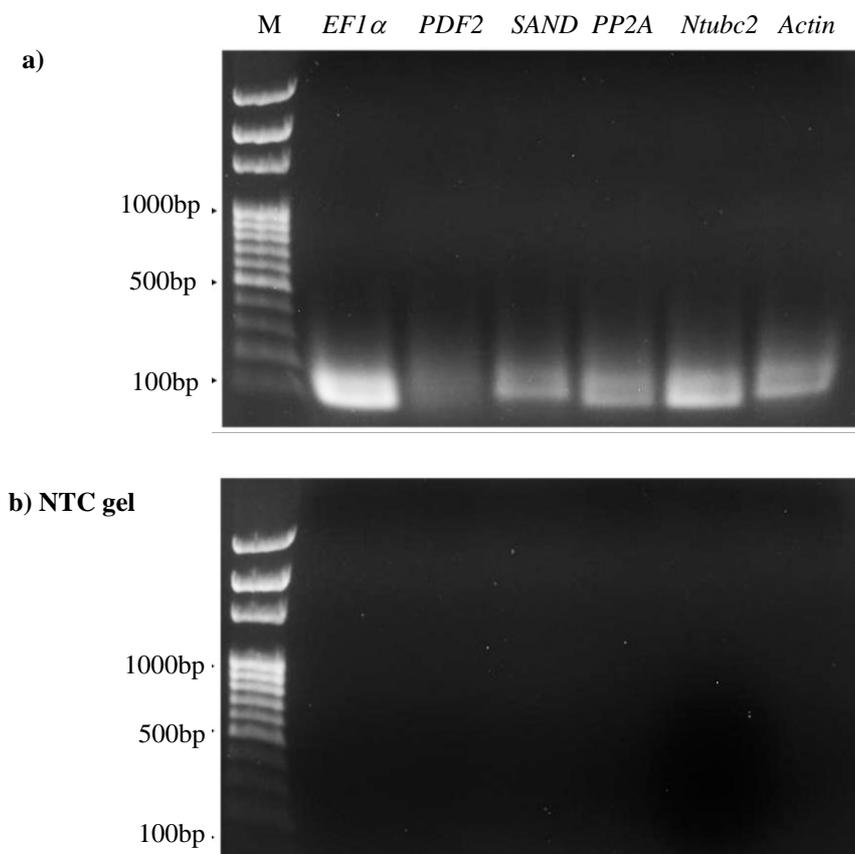


Figure 2.26: a) 1.5% agarose TBE gel confirming amplification of the candidate reference genes in an uninfected *N. glutinosa* sample run at 60°C. b) 1.5% agarose TBE gel confirming no amplification in the NTCs run using the same primers run at 60°C.

2.3.12 GOI primer RT-PCR confirmation at 60°C

Amplification of the expected size products for the *CPK3*, *SGS3* and *WRKY70* genes was observed when 60°C was used as the annealing temperature with uninfected samples (Figure 2.27a). There was no discrete band for *WRKY26* at this temperature, suggesting that the primer pair is not capable of amplifying the correct product at both 50°C and 60°C. It was decided that, as the primers were unable to amplify the *WRKY26* sequence as intended, it would be removed from this analysis. There was no contamination in any of the NTC reactions in this experiment, so the other primers were considered suitable for use in RT-qPCR (Figure 2.27b). For the remaining GOIs, *CPK3*, *SGS3* and *WRKY70*, a single product was observed, therefore, the primers for these sequences were considered for use in RT-qPCR.

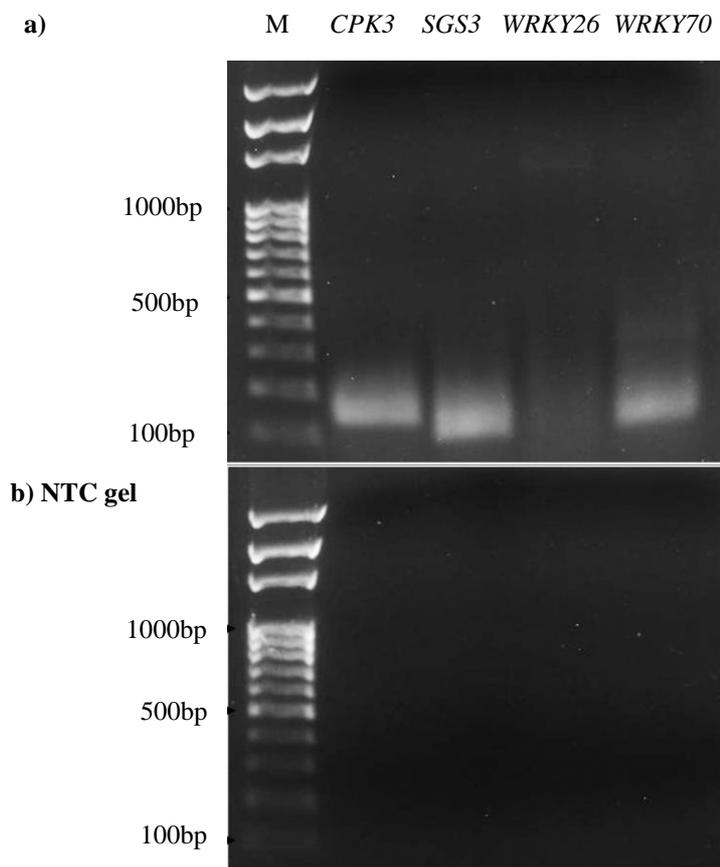


Figure 2.27: a) 1.5% agarose TBE gel confirming amplification of the candidate target genes in an uninfected *N. glutinosa* sample run at 60°C. b) 1.5% agarose TBE gel confirming no amplification in the NTCs run using the same primers run at 60°C.

2.3.13 Confirmation of candidate reference gene primer amplification using RT-qPCR

Preliminary tests were conducted on all primer pairs for each candidate gene utilising RT-qPCR to determine whether or not the primers amplified a product, and at what Cq value. They were tested on both LNYV infected and/or uninfected *N. glutinosa* samples from previous experiments, using three biological replicates and two technical replicates. A high-resolution melting (HRM) step was added at the end of the cycling to determine if the primers amplified one product as intended. The Cq values for each sample, that is the qPCR cycle at which the amplified samples fluoresce above a certain threshold, can be analysed to initially determine if the primers approximately remained stable and target genes had a change in expression values. Previous research generally accepts Cq values between 13 and 35 as being valid (Svec et al. 2015). The Cq values and their means for each gene are outlined in Table 2.21.

The initial data obtained from the qPCR experiment for the five remaining candidate reference genes indicated that, where tested, there may be an increase in expression of each of the reference genes between uninfected and LNYV infected samples. If so, none of them would be suitable. However, Cq values from replicate samples should be closer given that the same starting template was used. Examining the individual Cq values for these genes indicates a wide variation between replicates, and in some cases, such as the Actin gene in, no amplification was observed the infected sample.

It is likely this was a combination of pipetting error, improper preparation of the master mix, or due to evaporation of some of the samples due to the time it took to pipette all of the samples and reagents for this experiment. Additionally, the kit used for the qPCR study was a combination of reagents from two commercial kits, the SYBR Green from a ThermoFisher kit and the Superscript III from a Quanta kit. Whilst amplification did occur it may not have occurred as efficiently across all samples due to utilising an untested combination of kits.

As with the candidate reference genes the data obtained from the initial RT-qPCR experiment for the three candidate target genes indicated that between an LNYV infected and uninfected *N. glutinosa* sample, there may be differences in genes expression for *CPK3*, *SGS3* and *WRKY70* suggesting an increase in transcript levels in *CPK3* and *SGS3* and a decrease in transcript levels of *WRKY70* (Table 2.21).

Table 2.21: Cq values for the candidate genes from a one-step SYBR Green RT-qPCR experiment to obtain initial amplification information about the primers. ('-' means amplification was not achieved in the reaction)

	Target	Condition	Cq	Cq mean
Candidate reference genes	<i>Actin</i>	LNYV Subgroup II Infected	-	18.82
		LNYV Subgroup II Infected	-	
		LNYV Subgroup II Infected	-	
		Mock	20.19	
		Mock	17.45	
		Mock	-	
	<i>EF1α</i>	LNYV Subgroup II Infected	-	24.38
		LNYV Subgroup II Infected	27.38	
		LNYV Subgroup II Infected	21.38	
		Mock	14.93	14.72
		Mock	15.76	
		Mock	13.46	
	<i>Ntubc2</i>	LNYV Subgroup II Infected	22.96	21.75
		LNYV Subgroup II Infected	15.72	
		LNYV Subgroup II Infected	26.58	
		Mock	17.74	16.75
		Mock	16.30	
		Mock	16.20	
	<i>PP2A</i>	LNYV Subgroup II Infected	-	21.05
		LNYV Subgroup II Infected	-	
		LNYV Subgroup II Infected	21.05	
		Mock	16.83	16.57
		Mock	16.50	
		Mock	16.39	
<i>SAND</i>	LNYV Subgroup II Infected	28.09	26.02	
	LNYV Subgroup II Infected	23.19		
	LNYV Subgroup II Infected	26.77		
	Mock	21.86	21.88	
	Mock	22.00		
	Mock	21.77		
GOIs	<i>CPK3</i>	LNYV Subgroup II Infected	21.63	22.79
		LNYV Subgroup II Infected	-	
		LNYV Subgroup II Infected	29.15	
		Mock	21.42	20.94
		Mock	20.75	
		Mock	20.66	
	<i>SGS3</i>	LNYV Subgroup II Infected	21.47	25.00
		LNYV Subgroup II Infected	-	
		LNYV Subgroup II Infected	28.53	
		Mock	23.89	21.63
		Mock	20.81	
		Mock	20.19	
	<i>WRKY70</i>	LNYV Subgroup II Infected	19.76	20.28
		LNYV Subgroup II Infected	20.00	

		LNYV Subgroup II Infected	21.09	
		Mock	22.50	22.36
		Mock	22.22	
		Mock	-	

The HRM step also suggested that all of the candidate genes amplified only one product, except for *EF1 α* , which appeared to have a large and a small product (Sections 3.3.4.2.1 and 3.3.4.2.2 for further analysis and figures). However, it was determined that before any further experiments be conducted with more biological replicates to confirm these findings, it was necessary to find the optimal kits available to use and also reduce the pipetting error before any further studies were conducted.

2.3.14 Optimisation of qPCR experiments using alternative kits

A combination of kits was used for the initial RT-qPCR experiment and whilst amplification was achieved, it may be preferable to use kits from which all reagents come. Before several Quanta qScript kits (Quanta Biosciences, USA) that had been used for previous research were tested to determine if they were still viable to use, compared with the combination kit described in Section 2.2.9.5. Six kits were tested using uninfected *N. glutinosa* material and the *CPK3* primer pair, using the method and cycling conditions described in Section 2.2.9.5. The amplification curves are shown in Figure 2.28, whereby earlier Cq values were obtained for all samples. It appeared that nearly double the fluorescence was obtained in all of the samples utilising the different Quanta kits compared to the combination kit used previously. It was decided to proceed using the Quanta kits for the remainder of the experiments.

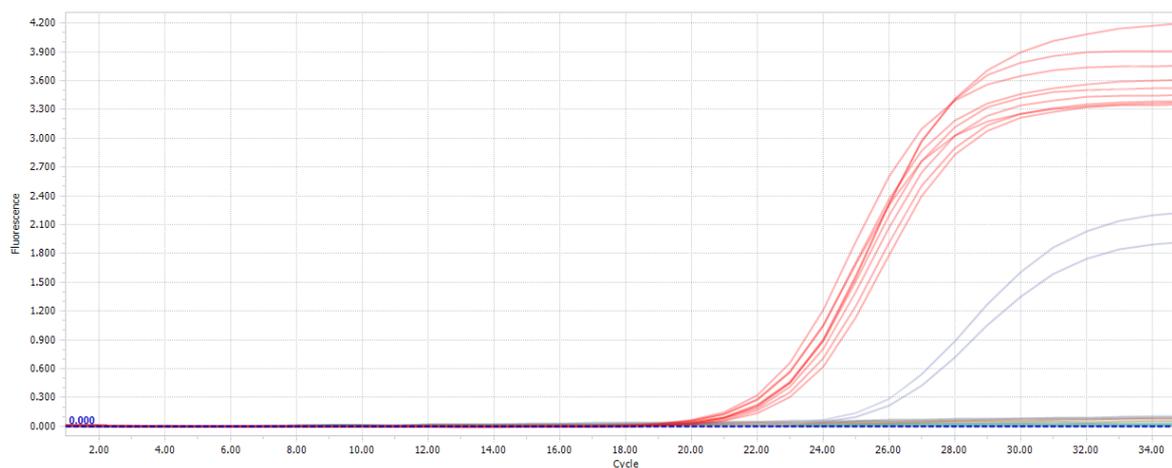


Figure 2.28: Normalised amplification curve testing six Quanta qScript kits (shown in red) in a RT-qPCR reaction against the combination kit previously used (shown in blue).

2.3.15 Optimisation of qPCR experiment preparation using an autopipette

An autopipette was used to dispense the samples and reagents in the next experiment, analysing only one reference and one target gene, *EF1 α* and *CPK3*, with the Quanta qScript kits as described in the previous section. Cq values were again analysed, as well as the normalised amplification curves and the HRM graphs. HRM analysis indicated only one product was amplified for *CPK3*, and a large and small product was amplified for *EF1 α* (data not shown), in keeping with the previous HRM analysis. Cq values (Table 2.22) indicated consistent amplification over a narrower range for the technical replicates for each condition. With the introduction of autopipetting, a large drop in the mean Cq values when testing the *EF1 α* primer on the same LNYV subgroup II infected samples was noted before (Table 2.21) and after (Table 2.22) optimisation, likely due to pipetting error in the earlier study causing inaccurate data generation. Further, *EF1 α* mRNA accumulation was potentially stable and with a possible upregulation of *CPK3* in an infection compared to a mock plant. This can be seen more clearly in the normalised amplification curves seen in Figure 2.29, where the profiles for the *EF1 α* replicates overlap (the red and green lines) while the *CPK3* curves (blue and orange lines) are separated between infected and uninfected.

As all of the primers tested had previously amplified a product, albeit with a range of Cq values between replicates, it was determined that given the optimisation of the kits and the use of the autopipette, the next step was to assess the usefulness of each candidate reference gene and to

assess the mRNA accumulation of *CPK3*, *SGS3* and *WRKY70* in LNYV infected *N. glutinosa* leaves relative to uninfected leaves.

Table 2.22: Cq values for the candidate genes *EF1 α* and *CPK3*. Each sample was tested in triplicate.

Target	Condition	Cq	Mean
<i>EF1α</i>	LNYV Subgroup II Infected	15.38	15.41
	LNYV Subgroup II Infected	15.43	
	LNYV Subgroup II Infected	15.42	
	Mock	15.13	15.11
	Mock	14.95	
	Mock	15.26	
<i>CPK3</i>	LNYV Subgroup II Infected	23.11	22.9
	LNYV Subgroup II Infected	22.90	
	LNYV Subgroup II Infected	22.77	
	Mock	21.76	21.6
	Mock	21.47	
	Mock	21.52	

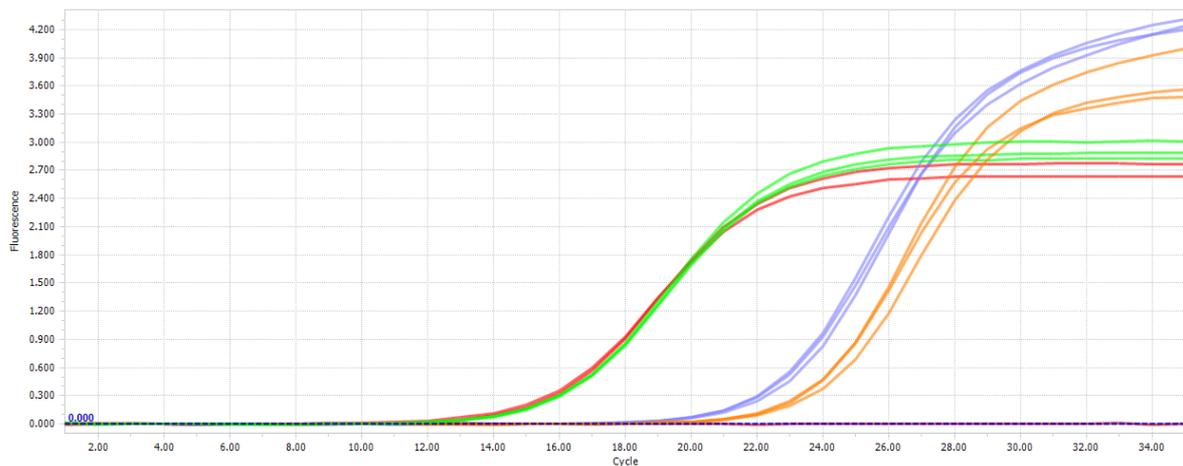


Figure 2.29: Normalised amplification curves for *EF1 α* (red and green) and *CPK3* (blue and orange) in a RT-qPCR conducted with an autopipette. Red and green lines indicate uninfected and LNYV infected *N. glutinosa* samples with *EF1 α* primer pair, respectively, and blue and orange lines indicate uninfected and LNYV infected *N. glutinosa* samples with *CPK3* primer pair, respectively.

2.4 Discussion

There have been no gene expression studies looking at host mRNA responses to infection by LNYV or its two subgroups to date. Currently, no reference genes have been identified in the literature for *N. glutinosa* based RT-qPCR studies. The focus of this chapter was to identify and amplify candidate reference and target genes to fulfil both of these criteria before conducting more in-depth RT-qPCR experiments to analyse the impact of LNYV infection and expression in more detail.

qPCR studies require a set of validated reference genes to be run alongside experimental target genes being studied, to act as internal controls as outlined by the MIQE guidelines (Bustin et al. 2013; Bustin et al. 2009). Previous research has suggested a set of reference genes has to be designed specifically for the species being studied in order to obtain reliable results. Further, the reference genes need to be useful for the biological states being examined. Currently, no valid set of reference genes exist for studies involving *N. glutinosa* nor for virus infected *N. glutinosa*, therefore these were designed as part of this study for determining the gene expression of four target genes in the host plant in response to LNYV infection. No full genome for *N. glutinosa* has been published, thus, for both the candidate reference genes and GOIs, mRNA and gene sequence data from related *Nicotiana* species were obtained to build multiple sequence alignments that would allow primers to be designed to specifically amplify the genes in an RT-PCR or RT-qPCR reaction.

Candidate genes were identified and selected based on previous plant virus studies that had been undertaken in similar host species. Though there were a large number of possible genes to choose from, it was decided to focus on previously studied genes first in order to utilise existing methods, protocols and information as a starting point. In addition, only a small subset of the genes identified in Table 2.2 were selected due to time and budget constraints. Many other potential candidate genes, particularly those identified in the Senthil et al (2015) study that induced a notable expression change in response to SYNIV, could be suitable for studying in future research. The candidate GOIs all had functional roles that may be impacted by viral infection, so were deemed suitable. It was noted that the candidate reference genes may also be impacted by the presence of a virus, but for the purposes of this research it was decided to test several of them with the notion that some, if not all, may be suitable for analysis.

Multiple sequence alignments

Of the 70+ species of *Nicotiana*, only seven have had their genomes fully published; *N. tomentosiformis*, *N. otophora*, *N. tabacum*, *N. obtusifolia*, *N. sylvestris*, *N. attenuata* and *N. glauca*. Partial gene sequences and mRNA sequences exist for many of the other *Nicotiana* species. For each of the candidate reference genes and GOIs, mRNA and gene sequences were obtained from three different databases, NCBI, QUT and SOL genomics. These sequences principally belonged to *N. attenuata*, *N. benthamiana*, *N. sylvestris*, *N. tabacum*, and *N. tomentosiformis*. The limited number of sequences available for the genes does not mean the genes are not present in other *Nicotiana* species but are yet to be identified through sequencing. Despite relatively few sequences existing to build multiple sequence alignments, one was constructed for each of the candidate reference genes as described in Sections 2.3.6 and 2.3.7. No previous information was found in the literature related to the intron/exon structure of each of the genes in *N. glutinosa*, so information related to this was inferred by constructing the alignments. It appeared that *F-BOX* had no intron, information which had not been reported in the literature reviewed for this research. Consensus identity levels were 100% for the majority of the exons identified so there was confidence that the alignments had been constructed correctly and confirmed by comparing to sequences in *A. thaliana*. The multiple sequence alignments constructed allowed primer design to be undertaken on all sequences except *F-BOX*, as it was necessary to design primers around an intron in order to prevent the amplification of genomic DNA. No comparison was made on the sizes of PCR products from *N. glutinosa* genomic DNA and mRNA, thus the sizes of the introns in each gene remains unknown. However, the expected product sizes for most mRNA based products were obtained with no apparent genomic DNA contamination. Fully sequencing *N. glutinosa* would confirm if the multiple sequence alignments for the genes studied here are accurate, but for now, are the most reliable source of information available to undertake this research.

Primer design

Primers for the candidate reference genes *Actin*, *EF1 α* , *Ntubc2*, *PP2A* and SAND and the GOIs *CPK3*, *SGS3*, *WRKY26* and *WRKY70* were designed and were amplified using RT-PCR and RT-qPCR. Expression maps indicated all genes should be detectable in leaves, which was confirmed by RT-PCR and RT-qPCR experiments. PCR products of the expected size were

achieved for the candidate genes, except *WRKY26*, but sequence identification was not confirmed and therefore product should be sequenced. Since they are all small, this would likely require cloning and sequencing. The expression profiles of these genes will be examined in the next chapter to determine if they remain stable between uninfected and LNYV infected *N. glutinosa*.

Primer design is one of the most important elements to successfully run a PCR assay (Bustin and Huggett 2017). The properties of each designed primer influence the specificity and sensitivity of a PCR reaction, and improper primer design can result in non-specific binding, false positives or intended targets failing to amplify. In order to ensure the primers work, several criteria have been previously reported in the literature that are necessary to follow in order to correctly design the primers. No reference gene primers had been previously reported to be designed for qPCR experiments utilising *N. glutinosa*, so it was necessary to design the primers from scratch, utilising these criteria as a basis. All of the primers designed were between 23 and 30 nucleotides in length, which are acknowledged as being acceptable in existing literature (Bustin and Huggett 2017). Whilst longer sequences increase the specificity of the binding of the primer to a target sequence, this may decrease the efficiency due to the finite numbers of free nucleotides within a reaction. Where possible, degenerate nucleotides were avoided, with only the forward *EF1 α* primer, the reverse *WRKY26* primer and the forward *WRKY70* primer containing them in order to maintain high specificity. Primers were designed around an intron to prevent genomic DNA from being amplified. A BLASTn search for all of the primer pairs suggested that there would be a low chance of the primers binding and amplifying to DNA sequences from humans, meaning that the possibility of false positives occurring whilst utilising the primers was minimised whilst also conducting the experiments in clean laboratory spaces with careful preparation. Whilst primers could not be designed for *L23* and *Ubiquitin*, alternative or additional sequence data from other *Nicotiana* species may help elucidate the structures of these genes to assist design of primers for future studies if further reference genes are required.

Amplification of candidate GOI and reference genes

From the initial RT-qPCR experiments conducted utilising the primers and uninfected and LNYV subgroup II infected *N. glutinosa* leaf material, an increase in the expression of each of the reference genes, as well as the GOIs *CPK3* and *SGS3*, was observed in LNYV infected

samples, and a decrease was seen in *WRKY70* compared to uninfected samples. Though this would suggest that none of the candidate reference genes identified are suitable to act as internal controls, the biological and technical limitations of the initial experiment, outlined in the following paragraph, mean further analysis is required before these conclusions can be considered to be robust.

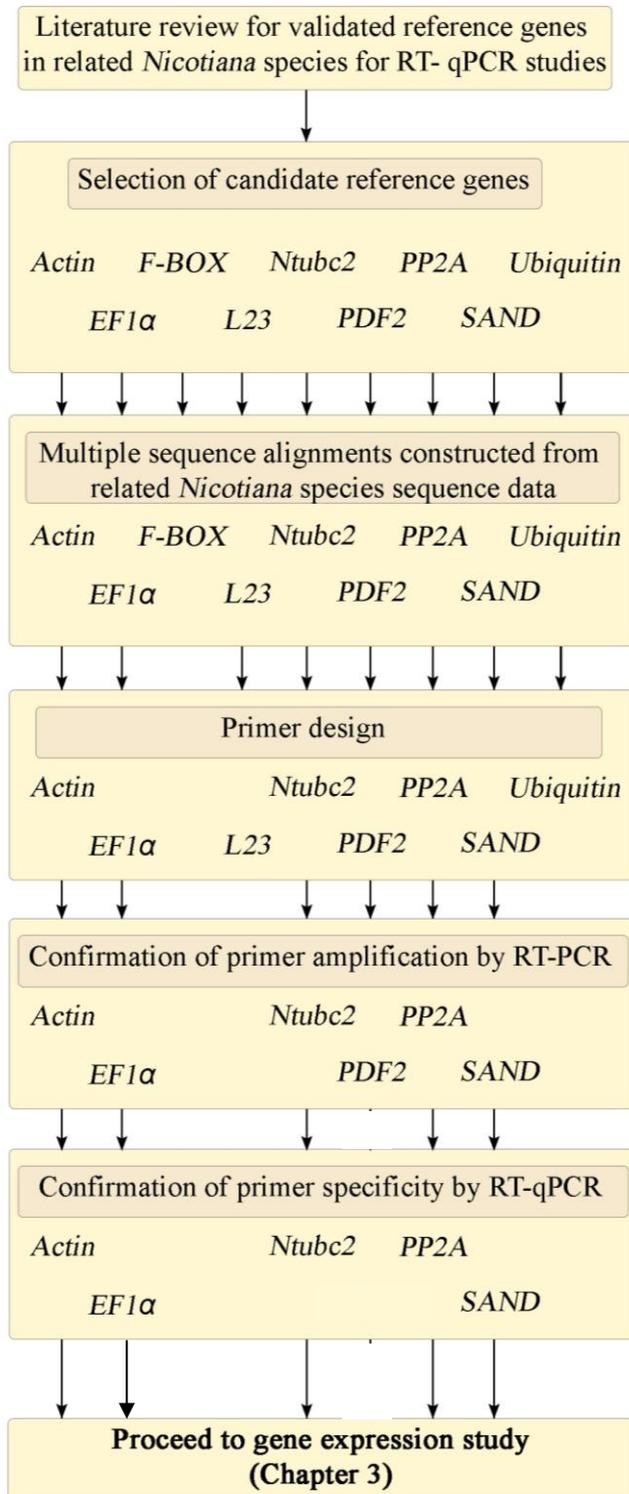
A summary flowchart of the processes the genes underwent is present in Figure 2.30.

RT-qPCR conditions

RT-qPCR is a sensitive molecular technique that requires optimisation of both biological and technical components in order to achieve the most robust results. For the initial RT-qPCR experiments run to test the primers on existing LNYV positive samples and *N. glutinosa* leaf material, several areas were identified as being problematic and efforts were made to optimise them, principally the use of fresh commercial kits for testing instead of mixing reagents and the use of an autopipette to minimise pipetting error in an effort not to produce unreliable data due to the volume of samples that need pipetting in these experiments.

Whilst suitability of the reference genes would usually be judged in advance of being used in a gene expression study, it was decided for the purposes of time, reagents and available facilities, to run the reference gene studies and target gene expression studies in parallel on LNYV infected and uninfected *N. glutinosa* grown under controlled conditions. This is described in Chapter 3.

Candidate Reference Genes



Genes of interest

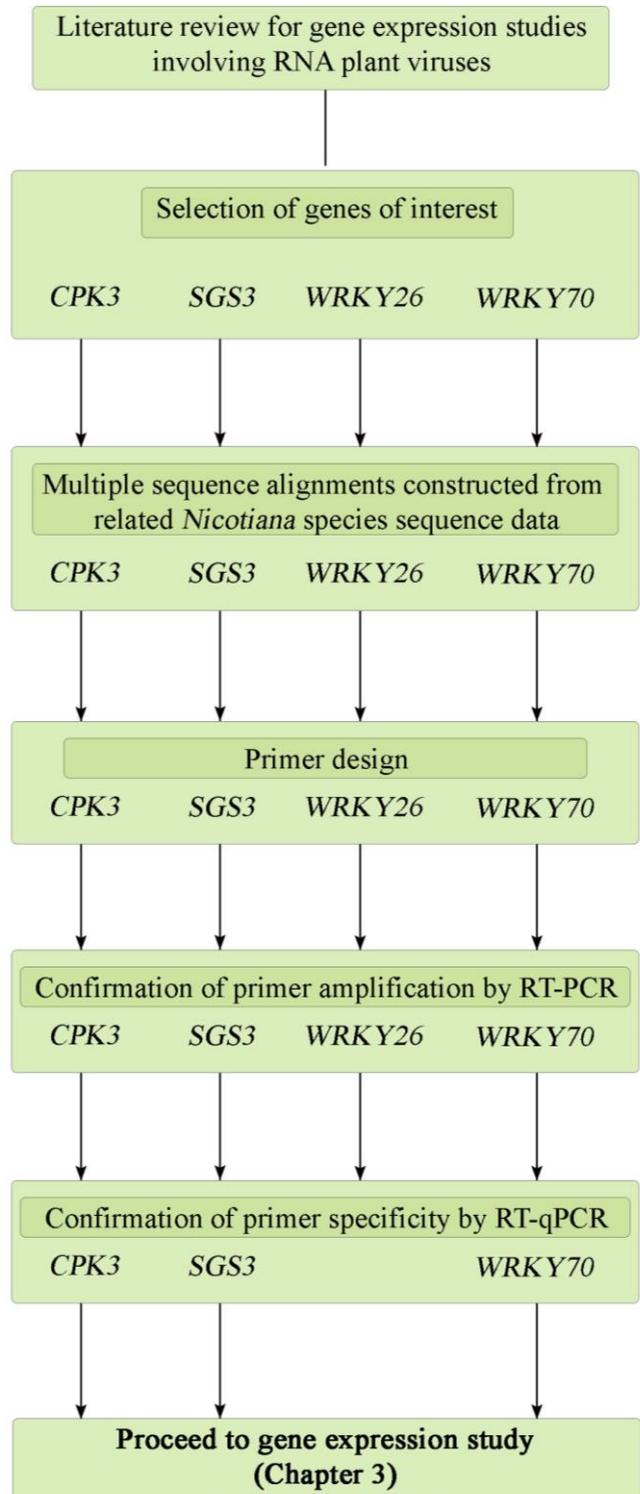


Figure 2.30: A summary of the processes the candidate references genes and GOIs were subjected to throughout this part of the study. Arrows indicate where each gene successfully passed onto the next stage.

Chapter 3

Analysis of *Nicotiana glutinosa* gene expression in response to LNYV subgroup I and II infection

3.1 Introduction

To date there have been no reported studies analysing host gene expression changes induced by LNYV infection. Further, no studies have assessed host responses to the different LNYV subgroups. For experimental studies into plant viruses it is important to have an easily infectible plant host in order to grow sufficient quantities of leaf material on which to inoculate. *N. glutinosa* has been used in a variety of LNYV studies over time and is a good experimental host species for this virus. This study assessed the mRNA changes in systemically infected *N. glutinosa* for the genes of interest described in Chapter 2, in response to infection by subgroups I and II of LNYV. Lettuce samples obtained from isolates collected in the field in New Zealand that were identified as infected with LNYV (Higgins et al, 2016) were used as inoculum for subsequent infection of *N. glutinosa*. Subgroup I and subgroup II isolates were used to establish infections in plants separately in order to compare the host's gene expression responses to each subgroup. The plants were sampled at various time points post inoculation to determine the expression of the GOIs over time to assess if any changes occurred in the expression of these genes during the establishment of infection. The primers described in Chapter 2 were used to amplify three target genes, *SGS3*, *CPK3* and *WRKY70*, and their expression levels relative to five reference genes; *Actin*, *EF1 α* , *Ntubc2*, *PP2A* and *SAND*.

3.1.1 Quantification of gene transcripts and impact on experimental design

According to the MIQE guidelines, appropriate quantification of gene transcripts is one of the key elements of the qPCR experimental design that needs to be addressed. The expression of genes can be measured either using absolute or relative quantification. Absolute quantification requires the generation of a standard curve from known concentrations of the amplified GOI and relates fluorescence values generated during a qPCR run to determine input copy number (Livak and Schmittgen 2001). Relative quantification measures changes in gene transcript expression between different experimental groups relative to stably expressed and previously validated reference genes (Pfaffl 2001; Pfaffl et al. 2004). Absolute quantification in this research was not possible as the amplification profiles of the candidate reference genes and GOIs were not known; to generate a standard curve all samples must share similar amplification properties (Svec et al. 2015).

Relative quantification requires the use of stably expressed reference genes as a point of comparison. The expression of the genes must not change between treatments being studied. In this case, reference gene expression must not be affected by infection by LNYV. Reference genes must be validated; this is generally done prior to use, however, in this study, due to the lack of time, this was done concurrently with analysis of the GOIs.

Selection of appropriate reference genes can be done utilising established software such as geNorm (Vandesompele, 2002), BestKeeper (Pfaffl *et al.*, 2004) or NormFinder (Andersen, 2004), or preferably a combination of these. The underlying calculations and assumptions of these generate a ranking of the stability of the candidate genes from most to least stable and provide confidence values in these rankings, as well as the number of genes that can be utilised from a given study for subsequent experiments. The relative expression of the GOI can then be determined against the most stable reference genes using software such as REST (Relative Expression Software Tool) (Pfaffl *et al.*, 2002) and qBase (Vandesompele *et al.*, 2007), or by manual calculations.

PCR amplification efficiency must also be considered. Early quantification models assumed 100% amplification efficiency at each cycle, meaning that each PCR cycle doubled the amount of transcript compared to the previous cycle. However, the presence of RT or PCR inhibitors in samples or nucleic acid degradation may impact the efficiency and so it is necessary to determine the amplification efficiency of a reaction after it has been run, made possible through linear regression analysis to determine how close to 100% each reaction was. A value of 2.0 in linear regression analyses indicates a 100% efficiency, with the value dropping with reduced efficiency. These values can provide a more accurate picture of the true expression levels of the gene being analysed.

Criteria established within the MIQE guidelines for designing robust RT-qPCR experiments indicate that sufficient biological and technical replicates are necessary in order to minimise variation and increase the statistical significance of the experiment (Bustin *et al.* 2009). In order to have sufficient biological replication in the RT-qPCR analysis, it was important to ensure enough plants could be generated at each time point that were infected with LNYV. To this end, the infection rates of each LNYV subgroup was assessed. This information helped ensure enough plants were inoculated so that enough infected plants could be obtained for analysis.

The areas of analysis briefly mentioned in this section, though mostly applied at the end of the study, influenced the experimental design of this project. Every effort was made to obtain quality nucleic acid material in a standardised manner and keep experimental variation to a minimum in order to obtain a valid dataset from which to analyse for the gene expression study. This chapter reviews the growth of LNYV infected and uninfected leaf material through to the end of the data analysis.

3.1.2 Aims

- Grow uninfected *N. glutinosa* from seed and inoculate with LNYV subgroups I and II obtained from infected field samples stored at -80°C.
- Determine the infection rate of LNYV subgroups I and II in *N. glutinosa*.
- Conduct RT-qPCR analysis of the relative expression levels of the target genes against the reference genes to determine the responses of the plant host *N. glutinosa* during LNYV infection, if any.
- Determine which of the candidate reference genes are most suitable and stable for future qPCR studies utilising *N. glutinosa*.

3.2 Materials and methods

3.2.1 Leaf material used to inoculate *N. glutinosa* with LNYV from infected *L. sativa*

In 2011, leaves from symptomatic lettuce plants were sampled in Harrisville, Auckland, New Zealand and stored at -80°C. Subsequently, they were identified as being infected with LNYV subgroup I or II by Higgins et al. (2016). A selection of these were used as inoculum in this study. Table 3.1 shows the sample identifications as published in Higgins et al (2016) and their laboratory identities as reported in Ajithkumar (2018). These samples are referred to in this thesis by the laboratory identifications as used by Ajithkumar (2018).

Table 3.1: Names of LNYV isolates and their subgroups as published to date.

Identification in Higgins et al, 2016	Identification in Ajithkumar, 2018	Subgroup
NZ1	Hv19	2
NZ2	Hv27	1
NZ3	Hv28	1
NZ4	Hv29	1
NZ5	Hv30	1
NZ6	Hv33	1

RNA was isolated from these samples (as per the method described in Section 2.2.9.2) and LNYV subgroup I or subgroup II infection confirmed by RT-PCR using the primers described by Ajithkumar (2018) and the method described in Section 2.2.9.4. RNA quality was assessed by spectrophotometry and agarose gel electrophoresis (see Section 2.2.9.3). Samples with the least degraded RNA were used as inoculum.

3.2.2 LNYV infection rate study

3.2.2.1 Inoculation and sampling of *N. glutinosa* with LNYV

3.2.2.1.1 Subgroup I

Plants were initially inoculated with the LNYV subgroup I isolate Hv28 or mock inoculated as described in Section 2.2.9.1.2, using 0.1% sodium sulphite. Systemic leaves were sampled as

described in Section 2.2.9.1.2 at 28 dpi. Subsequent inoculations were carried out with Hv29, Hv28, and Hv14 using 0.01% sodium sulphite to help improve the infection rate and systemic leaves sampled at 28 dpi. Plants were grown as described in Section 2.2.9.1.1.

3.2.2.1.2 Subgroup II

Plants were inoculated with the LNYV subgroup II isolate Hv19 or mock inoculated as described in Section 2.2.9.1.2, using 0.01% sodium sulphite. Systemic leaves were sampled as described in Section 2.2.9.1.1 at 28 dpi.

3.2.2.2 Confirmation of LNYV infection

LNYV infection of *N. glutinosa* was confirmed by RT-PCR. Total RNA was extracted from 100 mg frozen leaves utilising the method outlined in Section 2.2.9.2. RNA quality was assessed by spectrophotometry and agarose gel electrophoresis as described in Section 2.2.9.3.

3.2.3 Two-Step RT-PCR for amplifying LNYV

To conserve the use of RNA extracted from *N. glutinosa*, it was decided to use a two-step RT-PCR method to confirm LNYV infection.

3.2.3.1 cDNA synthesis with qScript™ Flex cDNA Synthesis Kit

cDNA was synthesised from total RNA using the qScript Flex cDNA synthesis kit (Quanta Biosciences). Approximately 300 ng RNA was combined with 2 µl 10 µM LNYV440F primer and 2 µl GSP enhancer. Nuclease free water was added to make a final volume of 15 µl. The tubes were briefly vortexed and centrifuged for 10 seconds, followed by incubation for 5 minutes at 65°C followed by incubation at 42°C. Following this, 4 µl of qScript Flex Reaction Mix (5X) and 1 µl qScript Reverse Transcriptase was added to give a final volume of 20 µl. The tubes were again briefly vortexed, centrifuged for 10 seconds and placed into a thermocycler (Techne TC-512) for 60 minutes at 42°C, followed by 5 minutes at 85°C, with a final hold at 4°C.

3.2.3.2 PCR amplification of cDNA with Promega GoTaq® Green Master Mix

cDNA template (2 µl) was combined with 6.25 µl 2X GoTaq® Green Master Mix, 0.25 µl of 10 µM LNYV_440F primer, 0.25 µl of 10 µM LNYV_1185R and 3.75 µl nuclease free water was added to a total volume of 12.5 µl. The tubes were briefly centrifuged and placed into a thermocycler (Techne TC-512). The PCR cycling conditions were 2 minutes at 94°C, 30 cycles of 30 seconds at 94°C, 30 seconds at 50°C and 1 minute at 68°C, with a final 5 minutes at 68°C. Reactions were held at 15°C. Conditions were later optimised to 35 cycles of 30 seconds at 94°C, 30 seconds at 50°C and 1 minute at 72°C with a final 5 minutes at 72°C.

3.2.4 Alternative RNA extraction methods to reduce commercial kit use

3.2.4.1 CTAB extraction

Prior to extraction, extraction buffer consisting of 2% cetyl trimethylammonium bromide; (CTAB), 2% soluble polyvinylpyrrolidone (PVP) K-40, 25 mM ethylenediaminetetraacetic acid (EDTA), 100 mM Tris-HCl (pH 8.0) and 2 M NaCl was heated to 65°C before adding β-mercaptoethanol (ME) to 3% (v/v).

N. glutinosa leaf material (250 mg) was ground to a powder and then in liquid nitrogen transferred to a 2 ml Eppendorf tube. Extraction buffer (1.2 ml) was added, and the sample vortexed vigorously for 1 minute. The sample was placed at 65°C for 30 minutes, vortexing every 5 minutes. The sample was centrifuged at 16,000 g at 18°C for 10 minutes and the supernatant was transferred to a new 2 ml Eppendorf tube. The supernatant was extracted twice with an equal volume of 24:1 chloroform-isoamyl alcohol by vortexing for 30 seconds, followed by centrifugation at 16,000 g for 15 minutes at 4°C. The aqueous phase was carefully transferred without disruption of the white interphase to a new 2 ml Eppendorf tube. The aqueous phase was carefully transferred without disruption of the white interphase to a new 2 ml Eppendorf tube and a third volume of 8 M LiCl was added. After mixing gently the sample was incubated at 4°C overnight. The sample was then centrifuged at 16,000 g for 60 minutes at 4°C. The supernatant was removed and the pellet was washed with 800 µl 70% ethanol. Following centrifugation for 5 minutes at 4°C, the ethanol was removed and the sample left to

air dry for 10 minutes. Once the ethanol had evaporated, the pellet was resuspended in 50 μ l of ddH₂O and stored at -80°C for future experimental use.

The quality of the RNA extracted using this method was assessed using a spectrophotometer and a 1% agarose gel as per the methodology described in Sections 2.2.9.2 and 2.2.9.3.

3.2.4.2 CTAB extraction with commercial lysis buffer

Due to poor quality RNA being extracted using the initial method as outlined in Section 3.2.4.1, the protocol was modified to use the lysis buffer from a Spectrum Plant Total RNA extraction kit (Sigma Aldrich) used earlier in this research, with minor modifications to the above protocol.

N. glutinosa leaf material (100 mg) was ground to a powder in liquid nitrogen and mixed with 800 μ l of the kit lysis buffer (combined with 8 μ l of 2-ME). The liquid was pipetted into a 1.5 ml Eppendorf tube and vortexed for 30 seconds, incubated at 56°C for 5 minutes, and centrifuged at 16,000 *g* at 18°C for 10 minutes. The remainder of the method was the same as that outlined in Section 3.2.4.1 from the addition of the chloroform-isoamyl alcohol step, onwards.

The quality of the RNA extracted using this method was assessed using a spectrophotometer and a 1% agarose gel as per the methodology in Sections 2.2.9.2 and 2.2.9.3.

3.2.4.3 Nucleic acid extraction column reuse

To allow for cleaning and reuse of the Spectrum Plant Total RNA extraction columns, it was necessary to first clean the Spectrum Plant Total RNA 1.5 ml tubes to remove residual nucleic acid material from previous extraction processes. Tubes from previous experiments were cleaned by soaking them twice for 5 minutes in 60°C water with bleach added to 10%, followed by soaking them twice in 99% alcohol for 5 minutes, followed by rinsing them twice with millipore water and autoclaving. They were then left to dry in a sterile environment.

The used nucleic acid extraction columns were inserted into a 1.5 ml tube, and 500 µl of the prewarmed buffer solution containing 0.2 M NaOH and 0.1 % Triton X-100TM (v/v) was heated to 75°C and added and incubated for 5 minutes at 75°C, followed by centrifugation for 1 minute at 10,000 g. The flowthrough was removed, and this step was repeated, but the columns were incubated for 10 minutes. The flowthrough was removed and 500 µl of 50 mM sodium acetate (pH4) was pipetted into the columns, which were centrifuged for 1 minute at 10,000 g. The sodium acetate was removed and 700 µl of RNase-free H₂O was pipetted into the columns and centrifuged for 1 minute. The H₂O was removed from the column and tubes and analysed using a spectrophotometer and 1% agarose gel to determine if any nucleic acid was carried over as per the methodology in Sections 2.2.9.2 and 2.2.9.3. The columns were left to air dry in a clean area with constant airflow and then placed into a clean bag for use in future extractions.

3.2.5 qPCR confirmation of candidate reference gene and GOI amplification using qScriptTM One-Step SYBR® Green qRT-PCR kit

Quantification of mRNA accumulation was carried out by one-step RT-qPCR using SuperScriptTM III PlatinumTM SYBRTM Green One-Step qPCR kit (Invitrogen) using a total volume of 12.5 µl instead of the manufacturer's recommendation of 50 µl. Each biological replicate was tested in triplicate to account for pipetting errors and triplicates of a standard were included on each plate to allow for plate to plate comparison. Total RNA (300 ng) was pipetted into wells of a 96 well plate, along with 6.25 µl of One-Step SYBR Green Master Mix (2X), 0.5 µl qScript[®] One-Step Reverse Transcriptase, 0.25 µl of 10 µM forward primer, 0.25 µl of 10 µM reverse primer and autoclaved distilled water to a final volume of 12.5 µl. The qPCR cycling conditions were the same as those outlined in Section 2.2.9.5, including the final HRM step at the end of the cycle.

3.2.6 Analysis of RT-qPCR generated data

The cycle at which each sample reached the set threshold was obtained (the C_q value). The C_q values for all of the standards were collated and analysed to identify and remove outlier values. The values of the three standards run on each plate were averaged and divided by the mean value of all of the standards from across all plates to obtain a normalisation value per plate. These individual plate normalisation values were then multiplied against the wells on each

plate to obtain a normalised RT-qPCR dataset. For each gene, the normalised Cq values for each experimental condition were averaged and outlier values were identified and removed from the dataset.

For each candidate reference gene, the normalised, outlier removed data was collated from across all of the plates. The grouped data for all of the reference genes was input into BestKeeper© software (version 1). The output gave an order of stability for the reference genes and a confidence level for each was obtained.

For each gene, the collated grouped data was also subjected to $2^{-\Delta\Delta Cq}$ analysis; the logarithm of the ratio of concentrations. For each gene, the values were ranked from lowest to highest, and the data was analysed using geNorm algorithm (Vandesompele et al. 2009). This gave an M value for each gene to determine the stability of each gene based on pairwise variation.

Genes identified as being the most stable were then utilised to determine gene expression changes in the target genes. The average Cq values of the target genes under each experimental condition were calculated, and the difference between those values and the average Cq values of each reference gene was calculated (ΔCq). The difference between each of these values (subgroup I vs. mock, subgroup II vs. mock, subgroup I vs. subgroup II) was then calculated ($\Delta\Delta Cq$). For each target gene, these values were averaged and inserted into the equation $2^{-\Delta\Delta Cq}$ to obtain an relative expression levels for each experimental condition.

3.3 Results

3.3.1 Inoculation conditions

The first step for the gene expression study was to identify suitable LNYV subgroup I and II isolates extracted from lettuce samples that were able to establish infection in the model plant *N. glutinosa*, and determine the optimal inoculation conditions to grow sufficient amounts of infected leaf materials for subsequent experiments.

3.3.1.1 Establishing conditions for inoculating *N. glutinosa* with LNYV Subgroup I

3.3.1.1.1 Inoculation with LNYV-Hv28

Lettuce leaves infected with the subgroup I Hv28 isolate of LNYV were used to inoculate 30 *N. glutinosa* plants. Visual inspection was carried out until 21 dpi to identify symptom development. Only one plant, identified as Hv28_004 (Figure 3.1a), displayed differences from mock inoculated plants (Figure 3.1b). Systemic leaves from this plant were sampled for molecular analysis, together with five apparently asymptomatic plants for comparison.

RNA was extracted from the leaves of the sampled plants, followed by RT-PCR for confirmation of LNYV infection. Figure 3.2 shows a typical RNA agarose gel indicating the extracted RNA was of good quality. The 28S and 18S rRNA bands are obvious with little evidence of degradation. RT-PCR analysis showed no amplification of the expected 746 bp product in any of the samples (Figure 3.3), indicating LNVY infection had not occurred. This suggested that the differences in leaf appearance in plant 004 were either not disease symptoms, or the virus was unevenly distributed in the leaf and not present in the leaf area analysed.

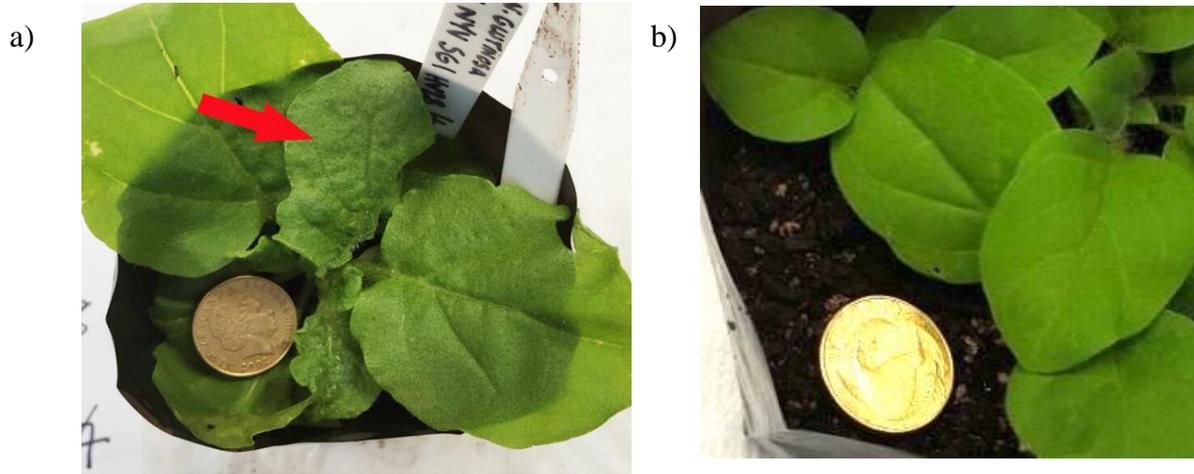


Figure 3.1 a) LNYV Hv28 inoculated *N. glutinosa* plant at 28 dpi. Red arrow points to the area that was sampled for testing b) Mock inoculated plant for comparison.

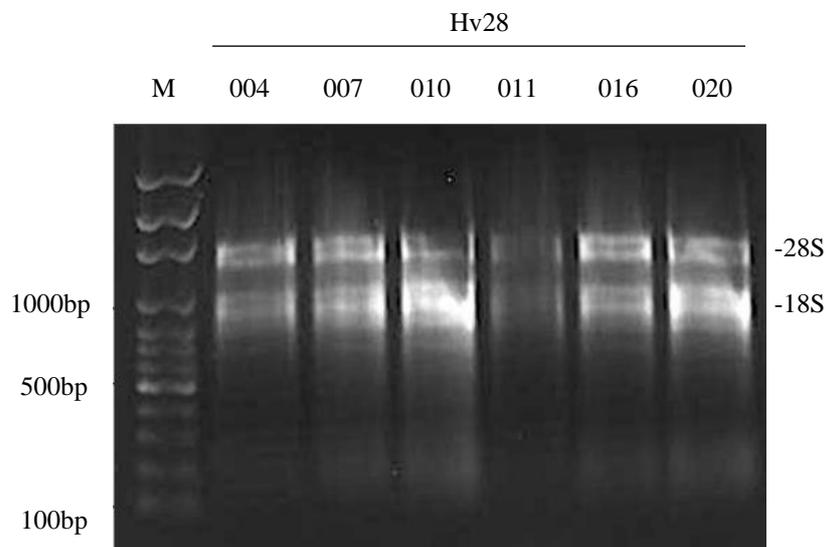


Figure 3.2 1% agarose gel electrophoresis of total RNA from LNYV Hv28 subgroup I inoculated *N. glutinosa* samples. (M) 100bp DNA Solis Biodyne ladder.

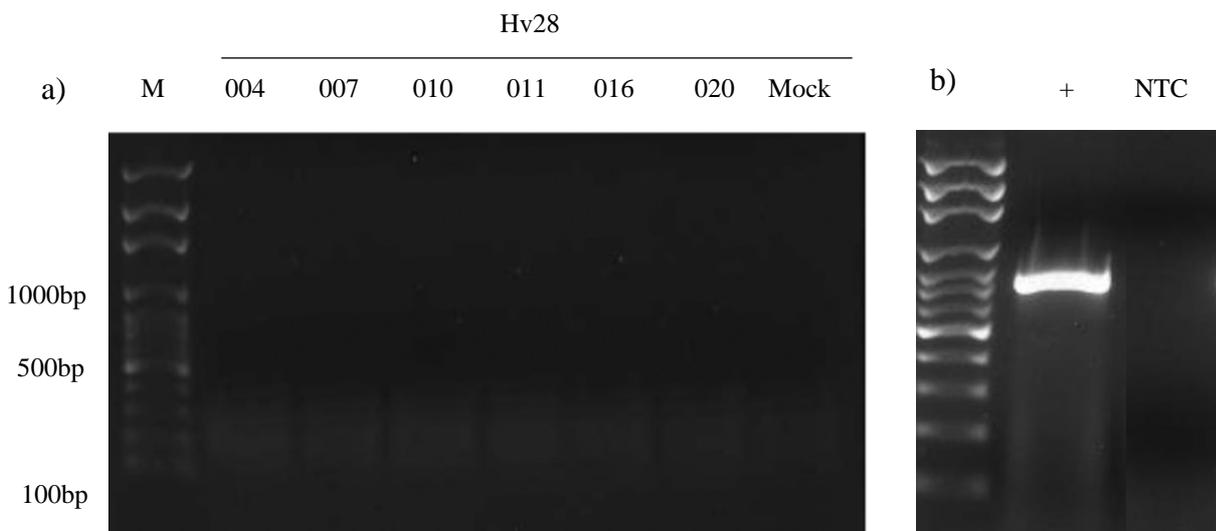


Figure 3.3 Detection of LNYV-Hv28 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. a) Individual plants sampled at 28 dpi. b) LNYV_Hv19 positive sample and no template control analysed at the same time on a separate gel. (M) 100bp DNA Solis Biodyne ladder.

To eliminate the possibility that LNYV was unevenly distributed in the *N. glutinosa* leaf tissue, other leaf regions were tested for the presence of LNYV. Figure 3.4 shows the extracted RNA was of good quality and Figure 3.5 shows amplification of the expected sized product in samples Hv28_004a, but not in Hv28_004b. This indicates that LNYV is unevenly distributed in the leaves so that care must be taken when sampling leaf material when carrying out RNA extractions.

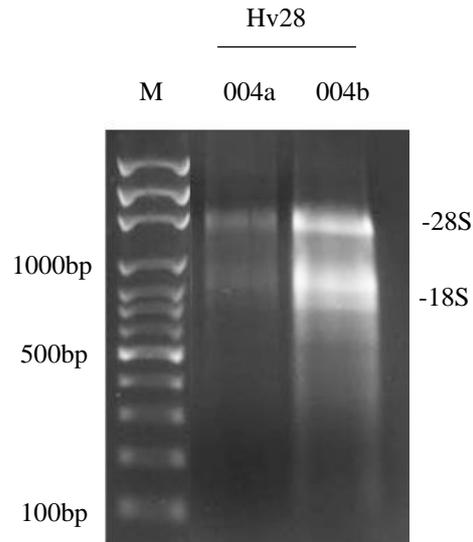


Figure 3.4: 1% agarose gel electrophoresis of total RNA from LNYV Hv28 subgroup I inoculated *N. glutinosa* samples. M: 100bp DNA Solis Biodyne ladder.

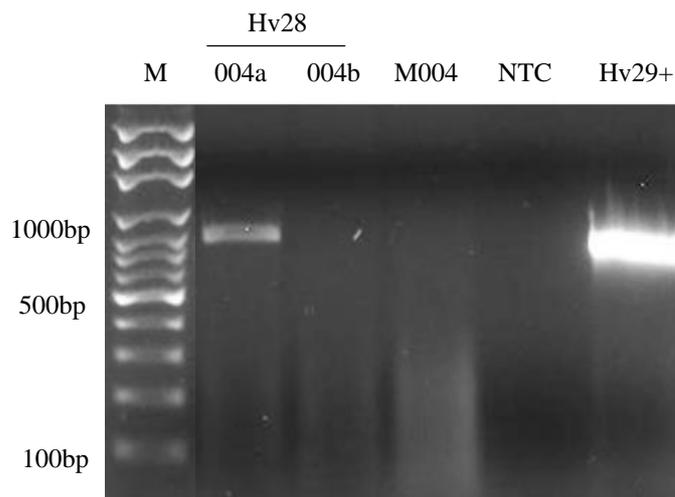


Figure 3.5: Detection of LNYV-Hv28 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Hv28 subgroup I samples and a mock inoculated sample M: 100bp DNA Solis Biodyne ladder, NTC: No template control. LNYV-Hv29 used as positive control.

Further, based on symptoms, only one plant from 30 inoculated showed evidence of infection. This plant was tested for LNYV together with five others; the remaining 24 plants were not tested. Therefore, it can't be assumed that other plants were not infected, but the lack of symptoms suggests they weren't. This indicates a low infection rate of around 3%, under the conditions used.

Upon reviewing the inoculation process it was determined that the sodium sulphite concentration in the inoculation buffer was too high at 0.1%. This may have interfered with the infection process, therefore, this was reduced to 0.01% for subsequent inoculations. Also, the lack of symptoms at 21 dpi may suggest a longer time is needed for symptoms to appear when transferring LNYV from lettuce to *N. glutinosa*. Thus, this time period was extended to 28 dpi for future inoculations.

3.3.1.1.2 Inoculation with LNYV-Hv29

Due to limited amounts of available Hv28 leaf material to inoculate further *N. glutinosa* plants, the inoculum for the next batch of subgroup I inoculum was changed to Hv29. After amending the inoculation buffer and increasing the period for infection to establish to 28 dpi, visual inspection of the plants was undertaken. No visual symptoms were identified in any of the 30 plants inoculated. The five smallest plants in comparison to the mock inoculated samples were sampled instead (Figure 3.6).

A one step RT-PCR was conducted on the Hv29 inoculated samples using the LNYV440F and 1185R primers. No amplification of the expected 746 bp product was observed for any of the samples (Figure 3.7). The positive and negative samples gave the expected results. This indicates that infection had not established in these plants.

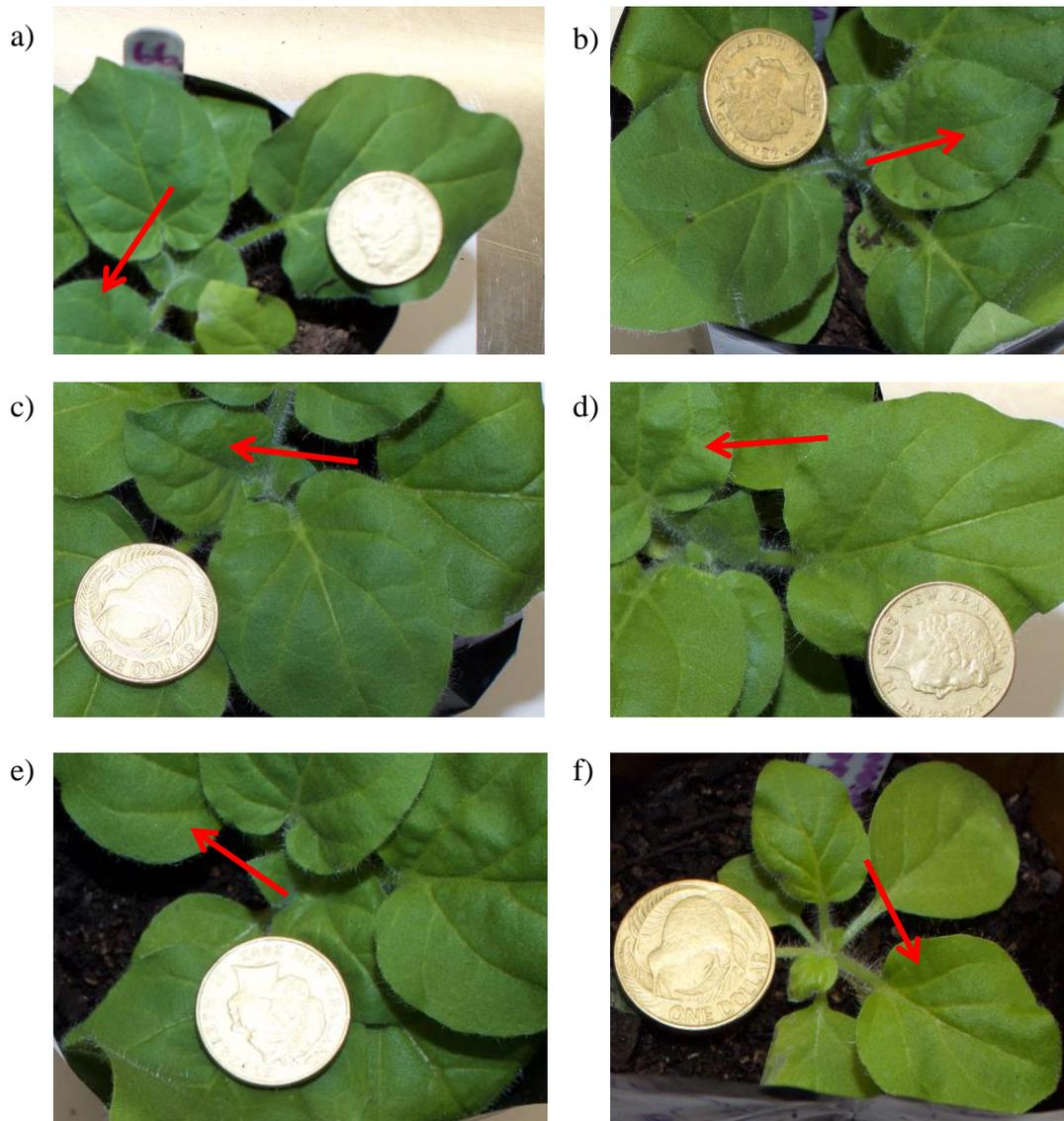


Figure 3.6: Images of asymptomatic Hv28 inoculated *N. glutinosa* plants that were tested, none of which were infected with LNYV. Red arrow points to area that was sampled at 28 dpi a) Sample 066 b) Sample 071 c) Sample 078 d) Sample 082 e) Sample 085 f) Mock inoculated sample.

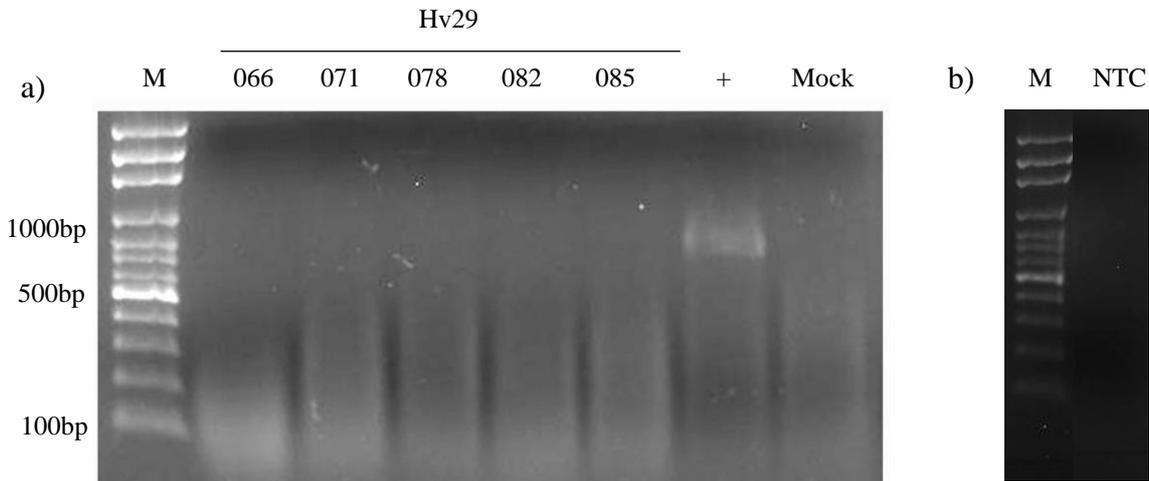


Figure 3.7: a) Detection of LNYV-Hv29 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, LNYV-Hv29, an original *L. sativa* LNYV infected sample was used as a positive control. b) NTC: No template control sample run in the same RT-PCR reaction but on a separate gel.

The *N. glutinosa* plants used for this inoculation had been initiated from seedlings grown outside of the conviron plant chamber, due to space limitations, and transferred into the chamber into individual plant bags before inoculation. The plants overall were smaller than previous plants grown; this may have impacted an establishment of LNYV infection. For the next set of inoculations, the inoculum was kept the same, but the seedlings were grown in the plant growth chamber as per previous experiments to prevent an environmental change at an important stage of growth. After 28 days, visual inspection of the Hv29 inoculated plants was undertaken but again, no visual symptoms were identified in any of the 30 plants inoculated. The five smallest plants in comparison to the mock inoculated samples were again sampled instead. RT-PCR analysis showed no evidence of LNYV-Hv29 infection in these plants (Figure 3.8).

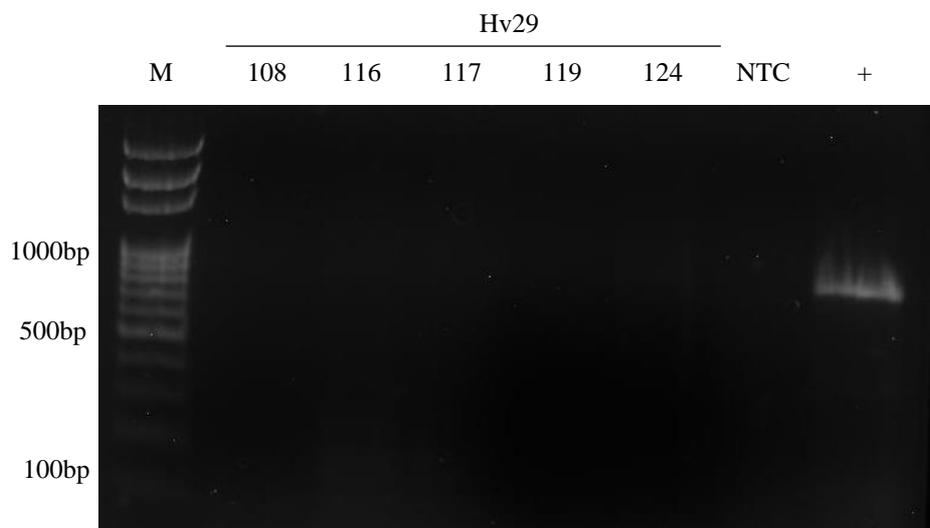


Figure 3.8: Detection of LNYV-Hv29 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. LNYV-Hv29 an original *L. sativa* LNYV infected sample was used as a positive control.

3.3.1.1.3 Inoculation with LNYV_Hv28-004a

As no LNYV positive samples were identified from either experiment using Hv29 inoculum, it was decided that another 30 plants should be inoculated with the positive sample from the initial round of subgroup I inoculations from this research, identified in Figure 3.5. It is possible that transfer of LNYV from a *N. glutinosa* infected plant may be more likely than from an infected lettuce plant. After 28 days, visual inspection of the Hv28_004a inoculated plants was undertaken but no visual symptoms were identified in any of the 30 plants inoculated. The five smallest plants in comparison to the mock inoculated samples were sampled instead and total RNA was extracted.

A one step RT-PCR was conducted on the Hv28-4a inoculated samples using the LNYV440F and 1185R primers. A product size of 746bp was expected when running the PCR product on a 1.5% agarose gel (Figure 3.9). No LNYV infection was detected in any of the five samples run, indicating that infection had not been established in this batch of plant growth.

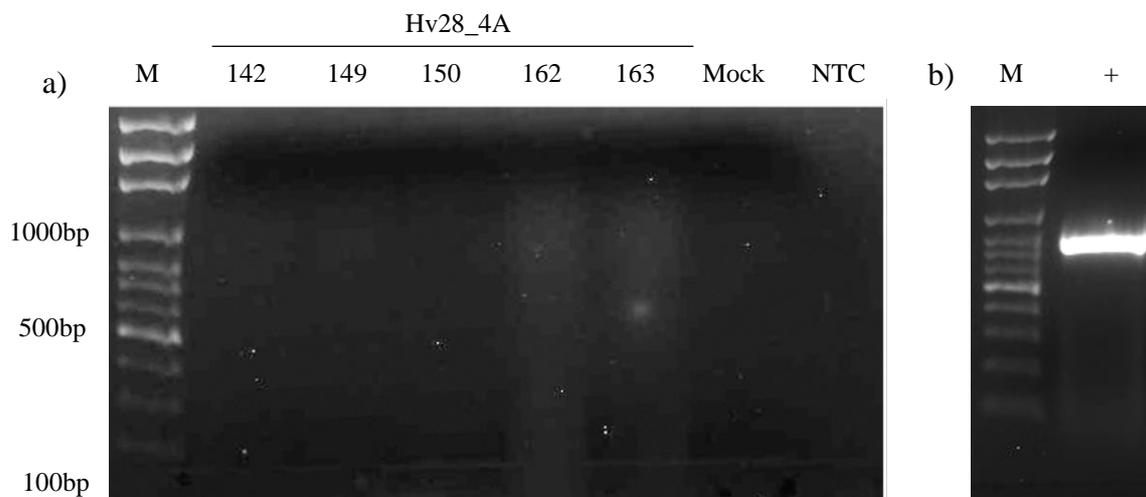


Figure 3.9: a) Detection of LNYV-Hv28_4a by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. b) LNYV-Hv29 used as positive control

As no positive samples had been yielded from Hv28_4a inoculum, it was decided to inoculate another 30 plants with another positive sample from Higgins et al., 2016. After 28 days, visual inspection of Hv14 plants were undertaken and five plants were sampled based on symptoms.

3.3.1.1.4 Inoculation with LNYV-Hv14

A one step RT-PCR was conducted on the Hv14 inoculated samples using the LNYV440F and 1185R primers. A product size of 746bp was expected when running the PCR product on a 1.5% agarose gel (Figure 3.11). LNYV infection was detected in three of the five samples run, 011, 019 and 029, indicating that infection had been established in this round of inoculation. Due to the low infection rate from previous subgroup I inoculated batches, the remainder of the plant batch was tested.

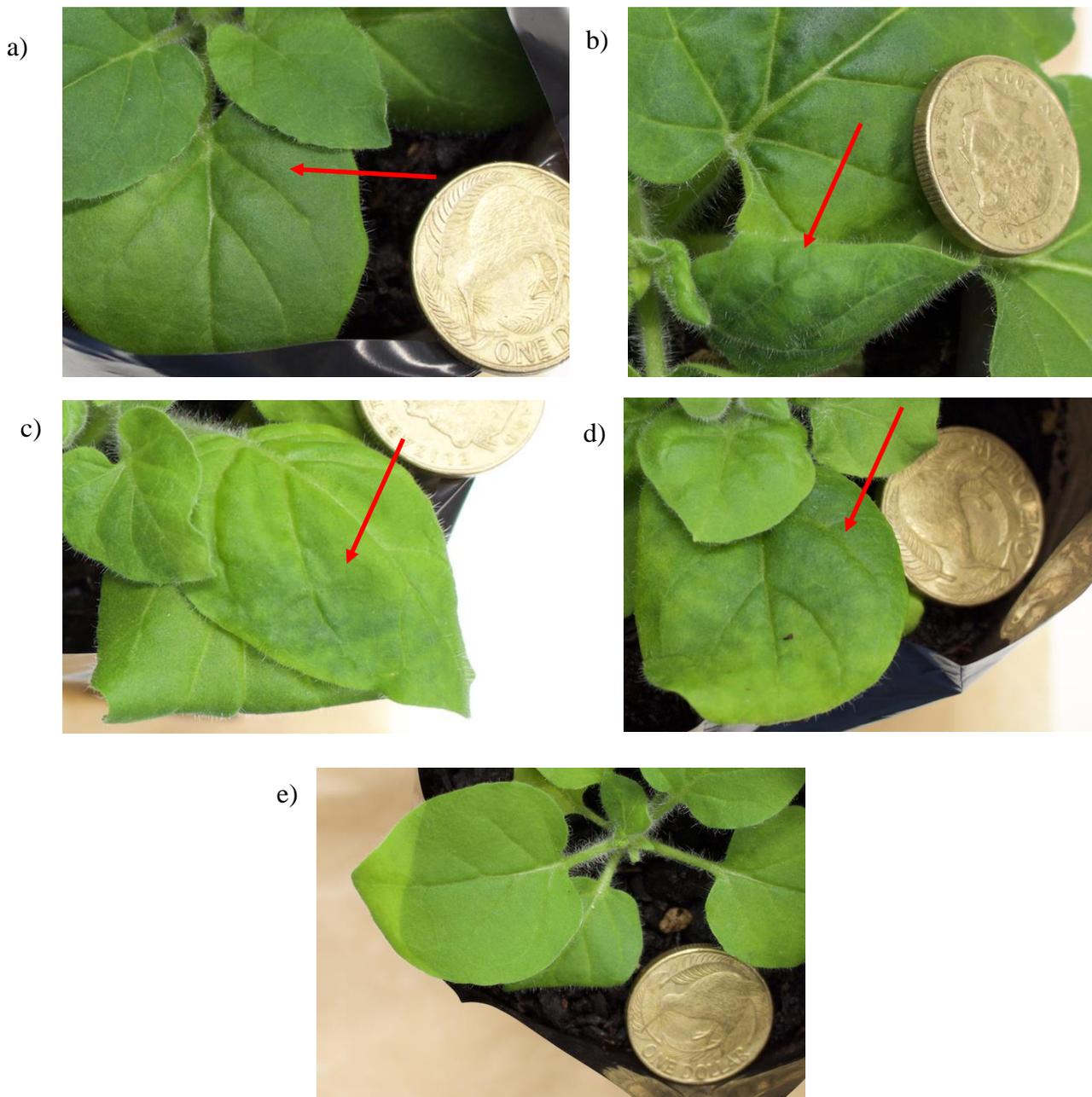


Figure 3.10: Images of Hv14 inoculated *N. glutinosa* plants that were tested based on differing symptoms to a mock inoculated plant, with a), b) and d) later confirmed with RT-PCR to be infected with LNYV. Red arrow points to area that was sampled for testing a) Sample 011 b) Sample 019 c) Sample 028 d) Sample 029 e) Mock inoculated plant.

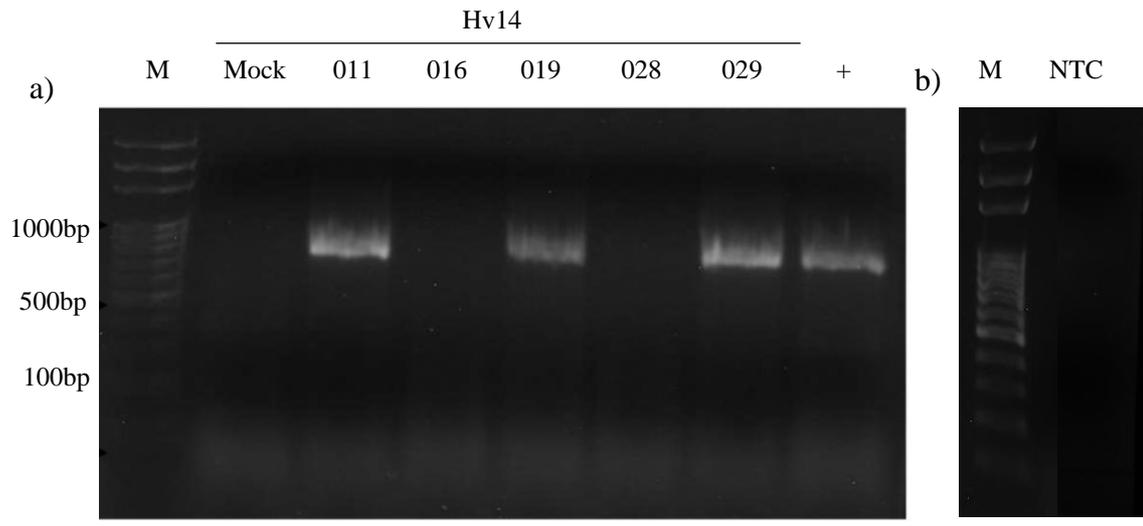


Figure 3.11: a) Detection of LNYV-Hv14 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, LNYV-Hv29 used as positive control b) NTC: No template control run in same reaction.

3.3.1.1.5 Summary of subgroup I inoculations

Table 3.2 shows a summary of all inoculation experiments with LNYV subgroup I inoculum to this point.

Table 3.2: Summary of subgroup 1 inoculation experiments. '+' denotes LNYV-like symptoms were observed during visual inspection or confirmed as being present in an RT-PCR. '-' denotes no symptoms were observed or the RT-PCR did not confirm the presence of the virus in the sample.

Inoculum	Plant #	LNYV symptoms	RT-PCR	Conclusion
Hv28	004	+	+	1 of 30 plants exhibited LNYV like symptoms, 1 of 6 tested positive for LNYV by RT-PCR
	007	-	-	
	010	-	-	
	011	-	-	
	016	-	-	
	020	-	-	0 of 30 plants exhibited LNYV like symptoms, 0 of 5 tested positive for LNYV by RT-PCR
Hv29	066	-	-	
	071	-	-	
	078	-	-	
	082	-	-	
	085	-	-	0 of 30 plants exhibited LNYV like symptoms, 0 of 5 tested positive for LNYV by RT-PCR
Hv28-4a	108	-	-	
	116	-	-	
	117	-	-	
	119	-	-	
	124	-	-	0 of 30 plants exhibited LNYV like symptoms, 0 of 5 tested positive for LNYV by RT-PCR
Hv28-4a	142	-	-	
	149	-	-	
	150	-	-	
	162	-	-	
	163	-	-	3 of 30 plants exhibited LNYV like symptoms, 3 of 5 tested positive for LNYV by RT-PCR
Hv14	011	+	+	
	016	-	-	
	019	+	+	
	028	-	-	
	029	+	+	

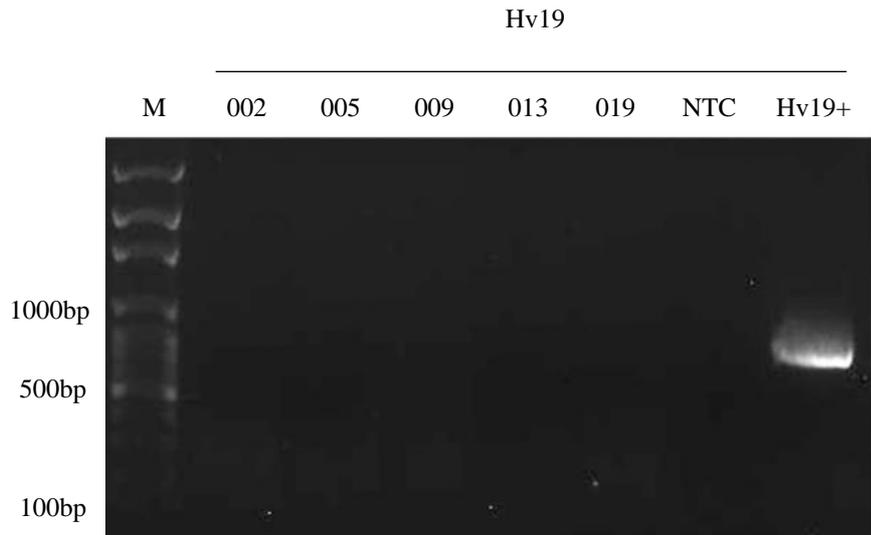


Figure 3.14: Detection of LNYV-Hv19 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. LNYV-H129 used as positive control

As with the initial subgroup I Hv28 inoculation procedure, after amending the inoculation buffer and increasing the inoculation time to 28 days before sampling, another 30 *N. glutinosa* Hv19 inoculated plants were visually assessed for LNYV like symptoms against five mock inoculated samples. This time six samples were selected based on symptoms, which, whilst not exhibiting the yellowing symptoms noted in previous papers, shared other characteristics compared to the mock plants such as reduced leaf size and mottling with light and dark patches (Figure 3.15). Two other samples, identified as 034 and 058 also exhibited possible LNYV symptoms but initially to save on using kit reagents these were not immediately tested.

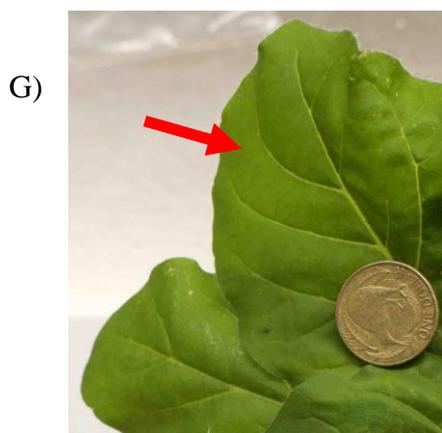


Figure 3.15: Images of Hv19 inoculated *N. glutinosa* plants that were tested based on differing symptoms to a mock inoculated plant, and later confirmed with RT-PCR to be infected with LNYV. Red arrow points to area that was sampled for testing. A) Sample 36 B) Sample 43 C) Sample 54 D) Sample 57 E) Sample 59 F) Sample 60 G) Mock inoculated plant.

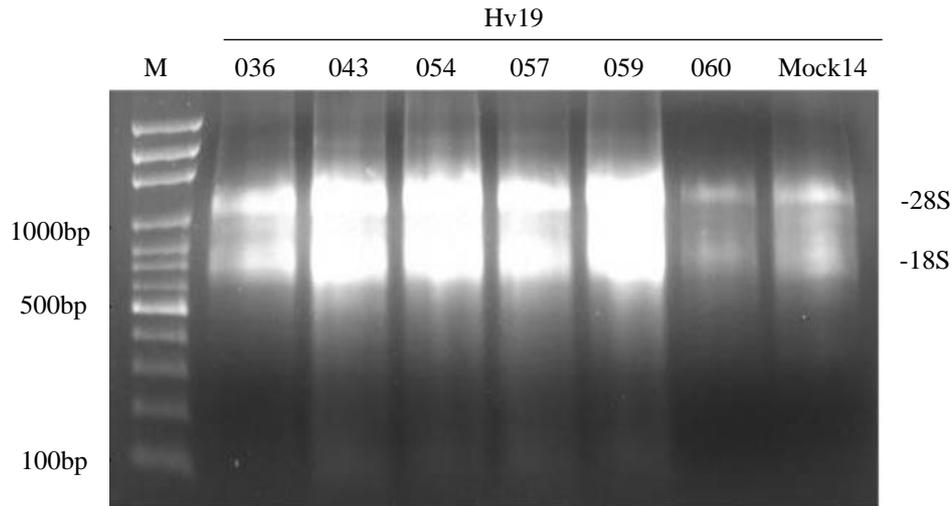


Figure 3.16 : 1% Agarose gel electrophoresis of total RNA from LNYV Hv19 subgroup II inoculated *N. glutinosa* samples. 100bp DNA Solis Biodyne ladder

Figure 3.16 shows the RNA for these plant samples was intact. A one step RT-PCR was conducted on the samples using the LNYV440F and 1185R primers. A product size of 746bp was expected when running the PCR product on a 1.5% agarose gel. LNYV infection was detected in all six samples run with product being visualised at approximately 750bp, confirming successful establishment of the virus in *N. glutinosa* (Figure 3.17). All could be used for future inoculations.

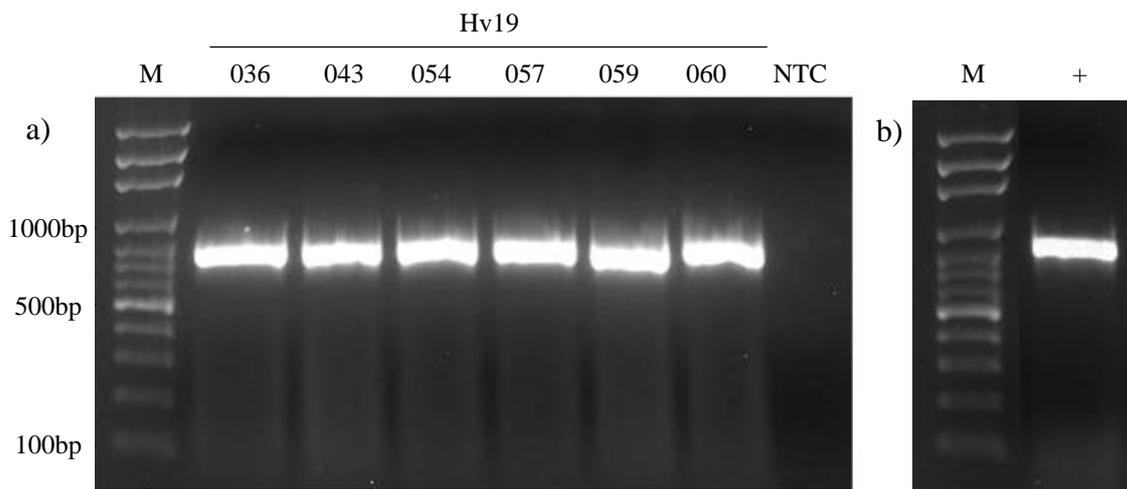


Figure 3.17: a) Detection of LNYV-Hv19 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control. a) LNYV-Hv19 used as positive control run in the same RT-PCR on a separate area of the gel.

3.3.1.2.2 Summary of subgroup II inoculations

Table 3.3 shows a summary of all inoculation experiments with LNYV subgroup II inoculum to this point.

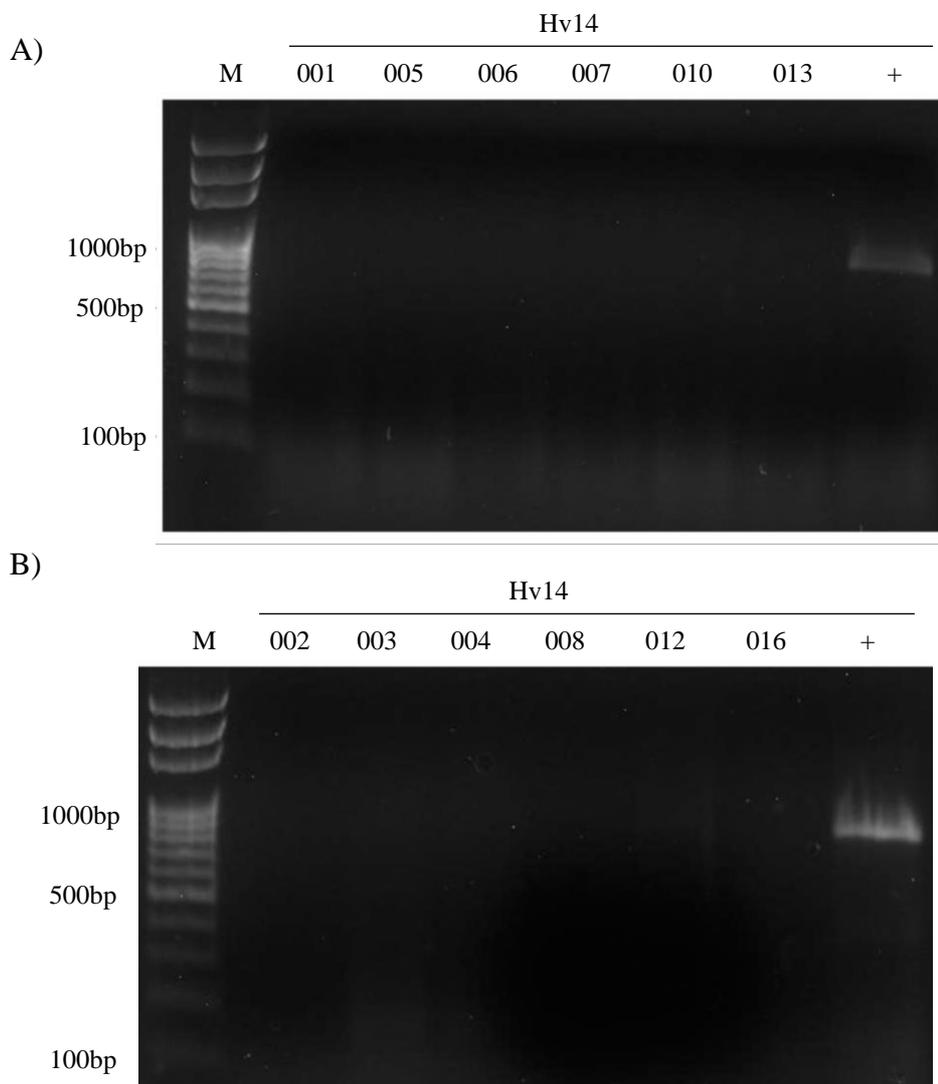
Table 3.3: Summary of LNYV subgroup II inoculation experiments. ‘+’ denotes LNYV-like symptoms were observed during visual inspection or confirmed as being present in an RT-PCR. ‘-’ denotes no symptoms were observed or the RT-PCR did not confirm the presence of the virus in the sample.

Inoculum	Plant #	LNYV symptoms	RT-PCR	Conclusion
Hv19	002	+	+	1 of 30 plants exhibited LNYV like symptoms, 1 of 6 tested positive for LNYV by RT-PCR
	005	-	-	
	009	-	-	
	013	-	-	
	019	-	-	
Hv19	034	+	Not tested	8 of 30 plants exhibited LNYV like symptoms, 6 of 6 tested positive for LNYV by RT-PCR
	036	+	+	
	043	+	+	
	054	+	+	
	057	+	+	
	058	+	Not tested	
	059	+	+	
	060	+	+	

3.3.1.3 Infection rates

3.3.1.3.1 Subgroup I infection rate study

The remainder of the 30 Hv14 inoculated *N. glutinosa* plants described in Section 3.3.1.1.4 were subsequently tested for infection to determine the approximate infection rate of subgroup I. A one step RT-PCR was conducted on the samples using the LNYV440F and 1185R primers. A product size of 746 bp was expected; LNYV infection was detected in one further sample, 023 (Figure 3.18d).



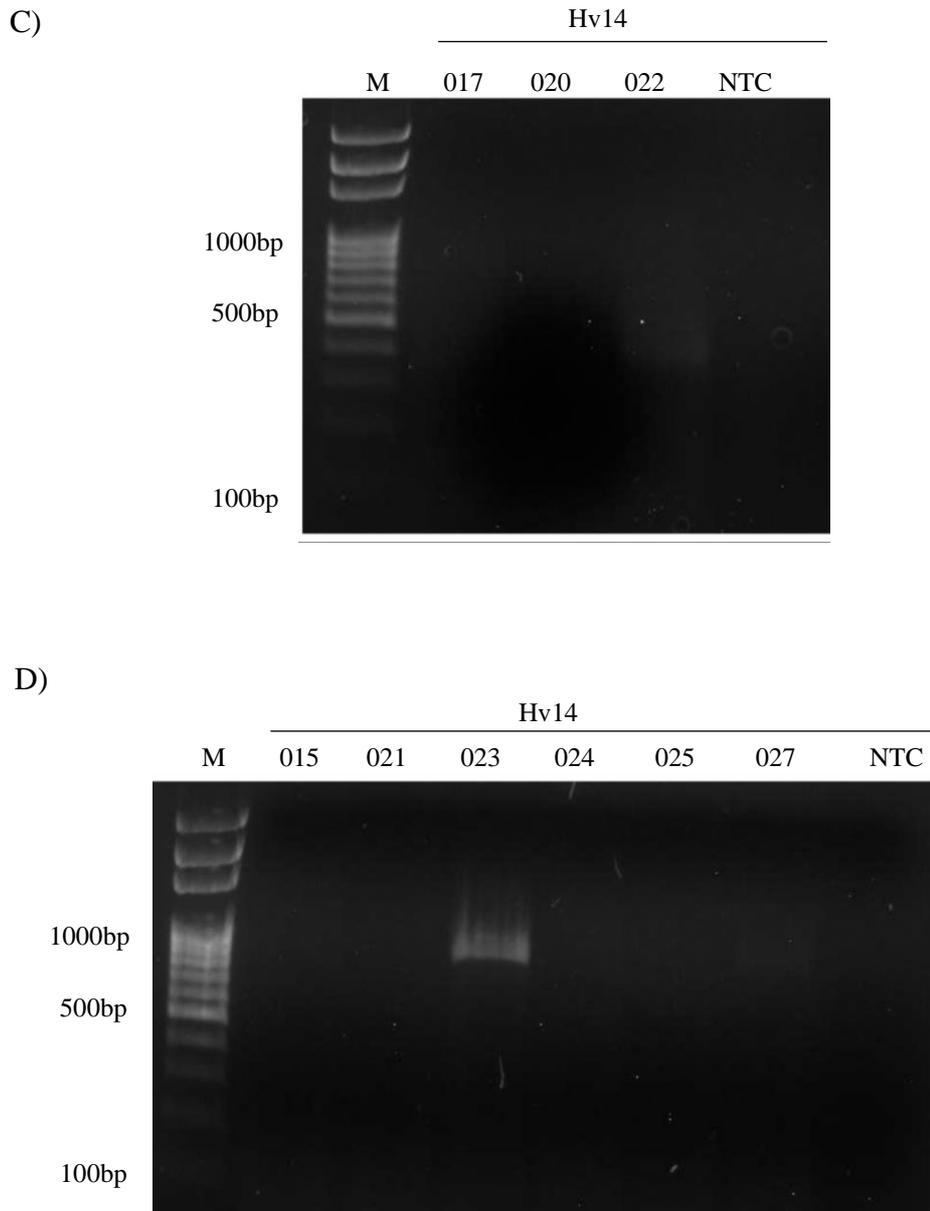


Figure 3.18: Detection of LNYV-Hv14 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Samples in gels A) and C) were tested in the same RT-PCR experiment, as were samples in gels B) and D). M: 100bp DNA Solis Biotyne ladder, NTC: No template control

With four out of thirty plants (Figure 3.19) being confirmed as infected with a subgroup I isolate of LNYV, an initial infection rate for subgroup I was determined to be approximately 13.3%. However, this is from just one of four inoculation experiments carried out using subgroup I isolates, not including the experiment conducted before inoculation conditions were optimised, where 0% infection rates were observed.

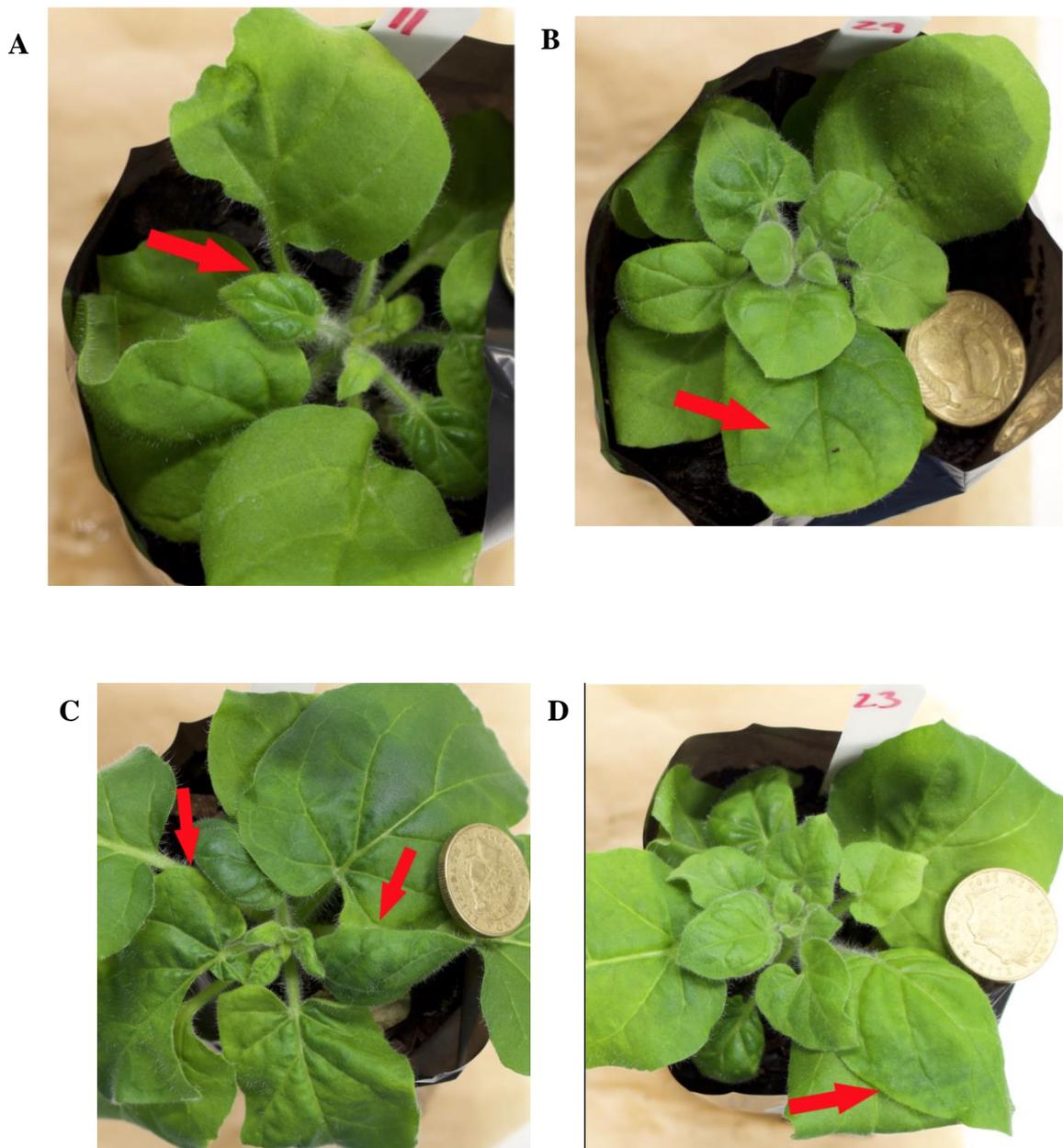
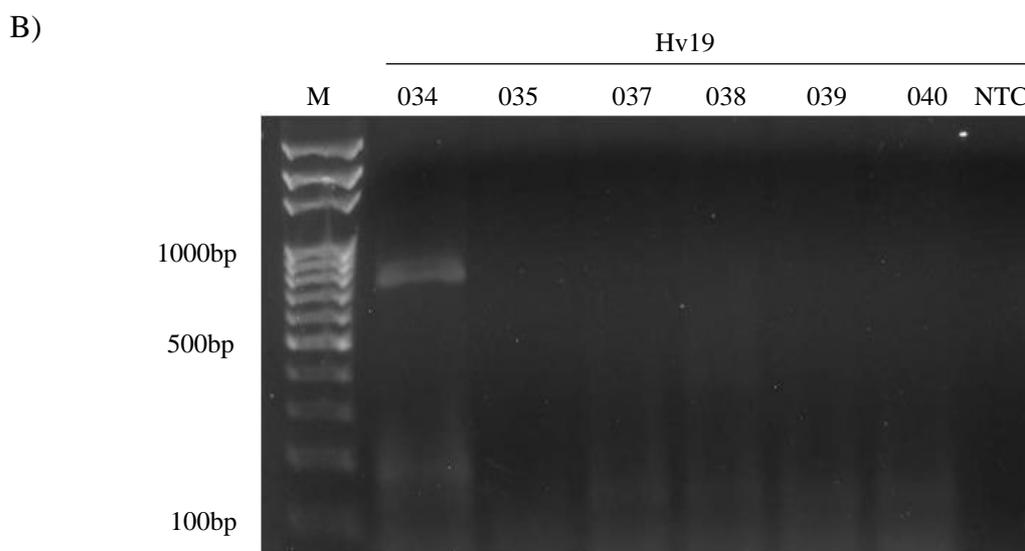
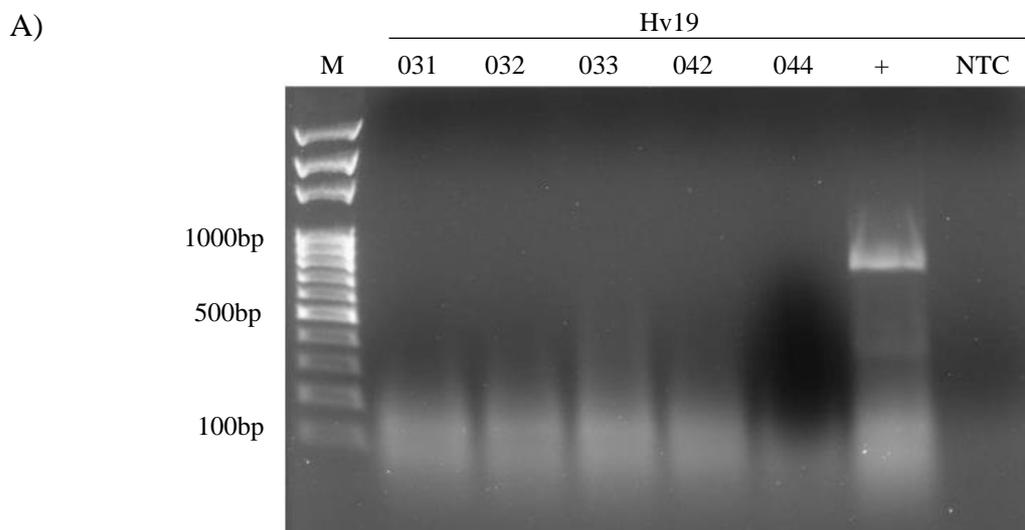


Figure 3.19: Images of LNYV Hv14 infected *N. glutinosa* samples confirmed by one step RT-PCR. A) Sample 011, B) Sample 029, C) Sample 019, D) Sample 023. Red arrows point to the areas that were sampled.

3.3.1.3.2 Subgroup II infection rate study

The remainder of the 30 Hv19 inoculated *N. glutinosa* plants from section were subsequently tested for infection to determine the approximate infection rate of subgroup II. A one step RT-PCR was conducted on the samples using the LNYV440F and 1185R primers. A product size of 746bp was expected when running the PCR product on a 1.5% agarose gel. LNYV infection was detected in two samples, 034 and 058, (Figure 3.20B, lane 1 and Figure 3.20D, lane 8 respectively) being visualised at approximately 750bp, confirming further successful establishment of the virus in *N. glutinosa*. The areas marked with red arrows in Figure 3.21 indicate the areas sampled and signs of mild symptoms are present at these locations.



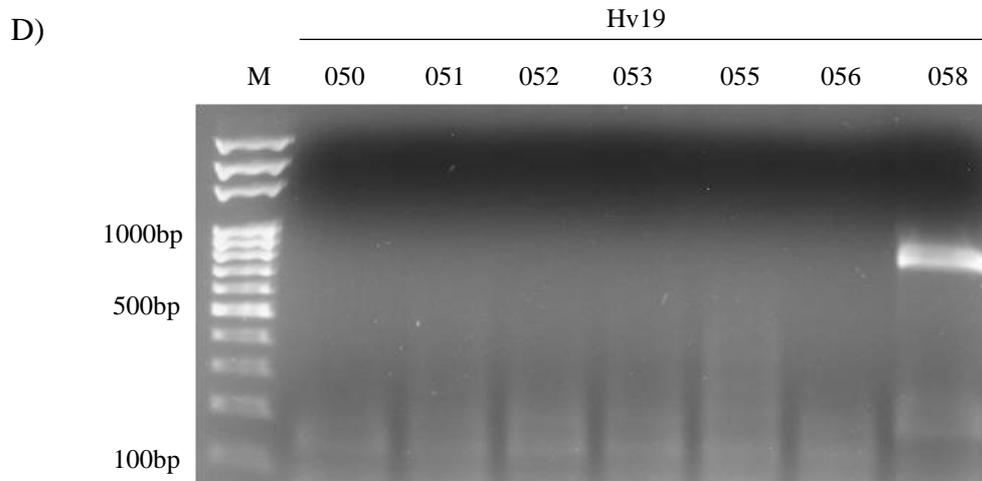
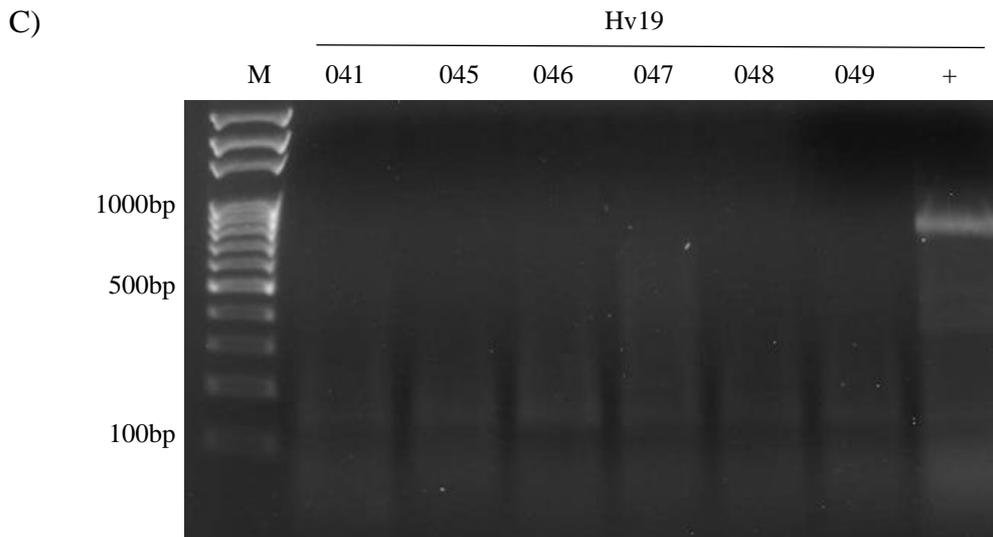


Figure 3.20: Detection of LNYV-Hv19 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Samples in gels A) and C) were run in the same RT-PCR reaction, as were samples in gels B) and D). M: 100bp DNA Solis Biodyne ladder, NTC: No template control.

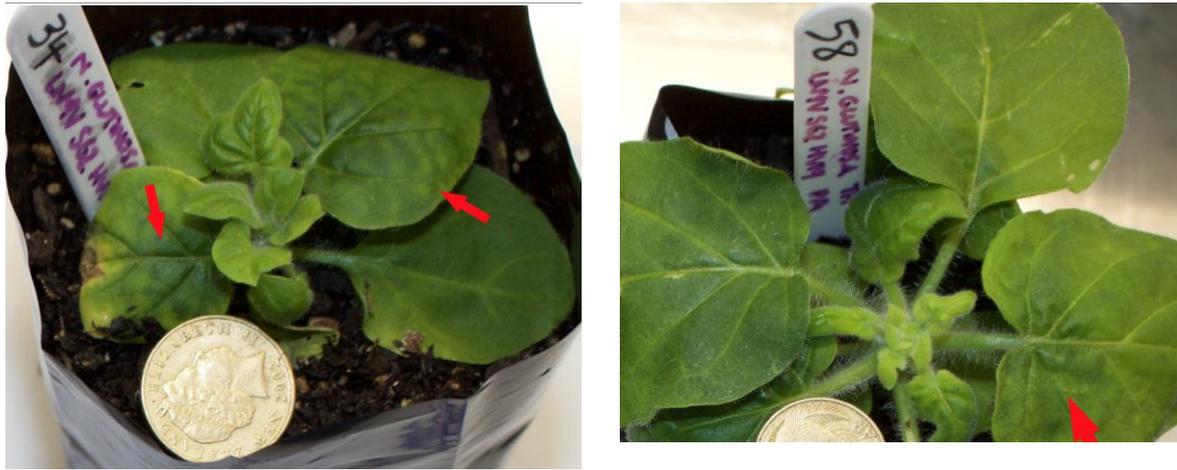


Figure 3.21: Images of LNYV Hv19 infected *N. glutinosa* samples confirmed by one step RT-PCR. Red arrows point to the areas that were sampled.

With eight out of thirty plants being confirmed as being infected with a subgroup II LNYV inoculum, an initial infection rate for subgroup II has been determined to be approximately 26.6%. This however is a very small-scale study under controlled conditions using only one inoculum sample. Further studies repeating the experiment are necessary to confirm the accuracy of the value; however, for the purposes of this research an approximation is all that is required in order to subsequently grow and inoculate enough plant material to obtain enough virally infected samples for the gene expression study.

3.3.1.3.3 Change to qPCR experimental design

The low infection rate obtained in these experiments, particularly for subgroup I meant that the amount of plant material that would need to be grown to obtain enough infected material for a RT-qPCR experiment would be difficult with the equipment available for the research. The plant growth chamber has room for approximately 120 plants, and it was estimated that at least 150 plants would have to be grown to try and obtain enough infected material. Also, this was using the infection rate value of 13.3%, but the lack of success overall with inoculating subgroup I meant the decision was made to not grow subgroup I isolates within the RT-qPCR study, and focus on differences in gene expression values in the host plant between uninfected plant samples and those inoculated with subgroup II samples.

3.3.2 Alternative RNA extraction methods to reduce commercial kit use

During the course of this research efforts were made to find alternatives to using commercially available kits in order to keep costs down or to reduce the amount of equipment used only once such as tubes and nucleic extraction columns. It had to be confirmed that robust results could be obtained using these alternative methods in order for them to be considered as viable alternatives to the existing protocols previously used.

3.3.2.1 CTAB extraction

A method for extracting total RNA utilising CTAB (White et al. 2008) was tested on fresh LNYV-infected leaf material, that is two samples from Hv19_034 and Hv19_058 described in Section 3.3.1.3.2. The RNA concentrations and absorbance ratios were determined by spectrophotometry (Table 3.4). Compared to the Spectrum Total RNA extraction kit, the RNA concentrations were highly variable. Samples 034a and 034b had A_{260}/A_{280} ratios around 2.0 indicating good purity RNA. However, the A_{260}/A_{230} values were all less than 2.0, indicating possible carbohydrate carryover meaning that cell lysis may not have occurred effectively. Gel electrophoresis showed the RNA quality to be variable (Figure 3.22).

Table 3.4: Summary of total RNA purity and concentrations of samples used in the CTAB extraction tests

Sample	Concentration (ng/μl)		A_{260} / A_{280} Ratio		A_{260} / A_{230} Ratio	
	Spectrum kit	CTAB	Spectrum kit	CTAB	Spectrum kit	CTAB
034a	253.6	25.2	2.142	2.107	2.331	0.649
034b		54.0		2.177		0.767
058a	185.2	809.2	2.134	2.075	2.134	1.999
058b		520.4		2.176		1.42

The sample on which the CTAB extraction appeared to work best from the spectrophotometry and agarose gel analyses, 58b, was amplified by RT-PCR with the *EF1α* primers and the *CPK3* primers alongside RNA extracted from another *N. glutinosa* plant using the Spectrum RNA extraction kit.

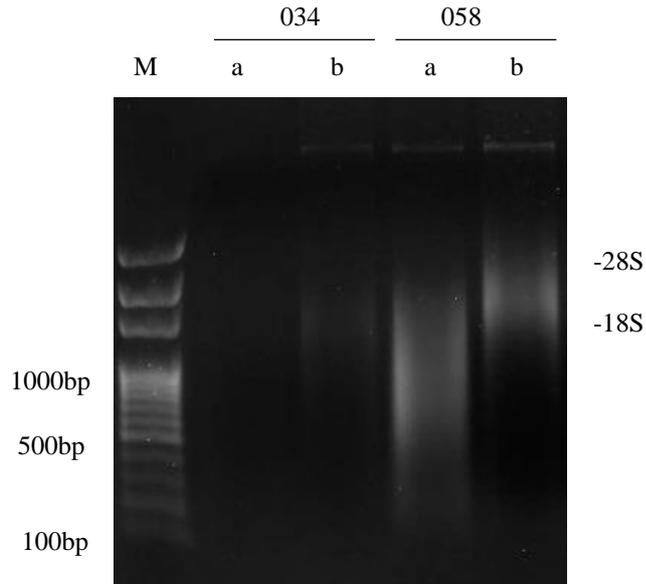


Figure 3.22: 1% agarose gel electrophoresis of total RNA from LNYV infected *N. glutinosa* samples Hv19_034 and Hv19_058 using CTAB extraction. M: 100bp DNA Solis Biodyne ladder.

Visualisation on a 1.5% agarose gel indicated that *EF1 α* was correctly amplified (Figure 3.23); however, there were two products amplified in the *CPK3* lane, indicating non-specific binding. This suggested RT-PCR may give unpredictable results, possibly due to poor lysis.

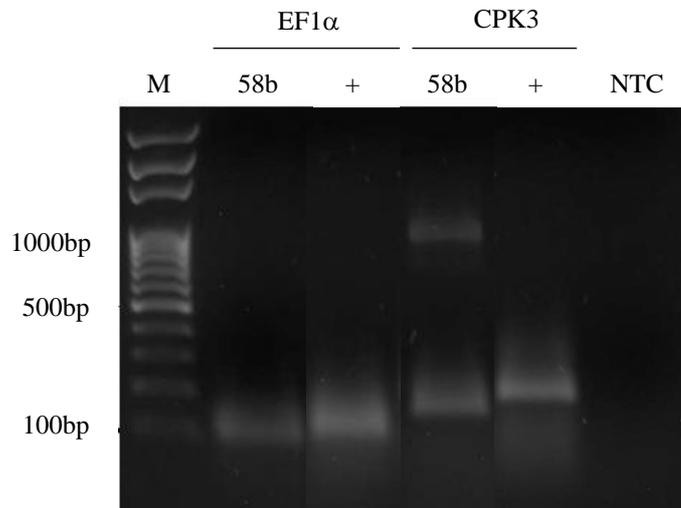


Figure 3.23: Detection of LNYV-Hv19_058b by RT-PCR using *EF1 α* and *CPK3* primers with an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control.

3.3.2.2 CTAB extraction with commercial lysis buffer

It was decided to modify the CTAB method using the buffer from the Spectrum Total RNA extraction kit to ensure effective lysis. Leaf material that had previously had RNA extracted from it was used for extraction with CTAB.

Table 3.5: Summary of total RNA purity and concentrations of samples used in the CTAB extraction tests

Sample	Concentration (ng/ μ l)		A ₂₆₀ / A ₂₈₀ Ratio		A ₂₆₀ / A ₂₃₀ Ratio	
	Spectrum kit	CTAB	Spectrum kit	CTAB	Spectrum kit	CTAB
Hv19_036	241.6	12.9	2.149	1.769	2.271	0.285
Hv19_056	625.6	14.4	2.151	1.967	2.293	0.571
Hv19_057	380.4	103.2	2.156	1.654	2.286	0.683

The concentrations, A₂₆₀ / A₂₈₀ and A₂₆₀ / A₂₃₀ values for all samples using the CTAB buffer were lower than the thresholds used throughout the rest of the project, except for the concentration of Hv19_057 and the A₂₆₀/A₂₈₀ of Hv19_056, indicating the purity of nucleic acid obtained from these samples are not a high enough quality necessary for downstream application. This was confirmed by running the samples on a 1% agarose gel (Figure 3.24), whereby no intact 28S/18S ribosomal subunits were identified in the lanes on the gel for any of the three samples. Due to the low integrity of the RNA determined from the spectrophotometer and the agarose gel, it was decided that it was unnecessary to carry out an RT-PCR. As the extraction method did not yield high quality RNA, this method was not an appropriate method to continue to use in this research. However, the method could continue to be explored and optimised in a future study, as it may work in other plant species.

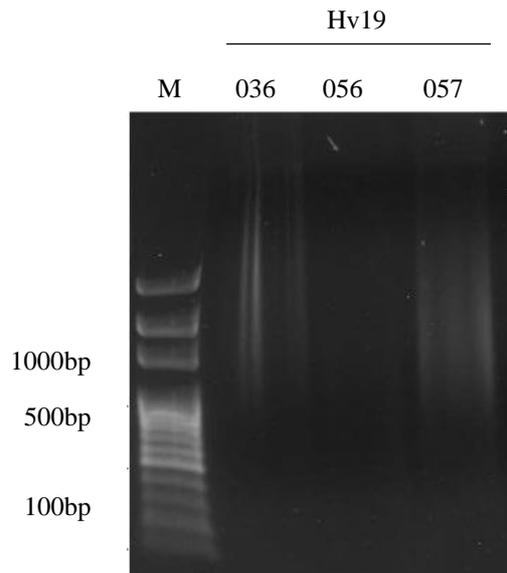


Figure 3.24: 1% agarose gel electrophoresis of total RNA from LNYV infected *N. glutinosa* samples Hv19_036, Hv19_056 and Hv19_057 using CTAB extraction with commercial lysis buffer. M: 100bp DNA Solis Biodyne ladder.

3.3.2.3 Nucleic acid extraction column reuse

Based on the infection rate studies (Sections 3.3.1.3.1 and 3.3.1.3.2), it was predicted that a large number of plants would have to be processed to identify LNYV plants that could be used in future RT-qPCR assays. Further, RNA extraction studies indicated that the Spectrum kit would be the most suitable to use for RNA extractions. Given the cost of commercial kits, and the fact that the columns always wear out first, it was decided to test if the columns could be reused. This analysis was based on a published method that reported a methodology to clean and reuse binding and filtration columns up to 15 times without degradation in the quality of the columns (Nicosia et al. 2010).

The filtration and binding columns from a commercial Spectrum Total RNA Extraction kit were saved after being used to extract RNA from LNYV-Hv14, -Hv18 and -Hv33 infected *N. glutinosa* samples. The RNA column cleaning protocol was applied and the final wash step flow through was kept and run on a 1% agarose gel to determine if any residual RNA was remaining on the columns (Figure 3.25). No residual RNA was detected in any of the lanes in the gel. Using a Nanovue spectrophotometer, all of the samples had an A_{260}/A_{280} ratio of less than 1.8 and RNA concentrations less than 10 $\mu\text{g}/\text{ml}$. Whilst a positive value for the

concentration would indicate possible residual RNA, experience within the AUT laboratory suggests that the Nanovue utilised has been thought to slightly overestimate the presence of RNA (Colleen Higgins, personal communication).

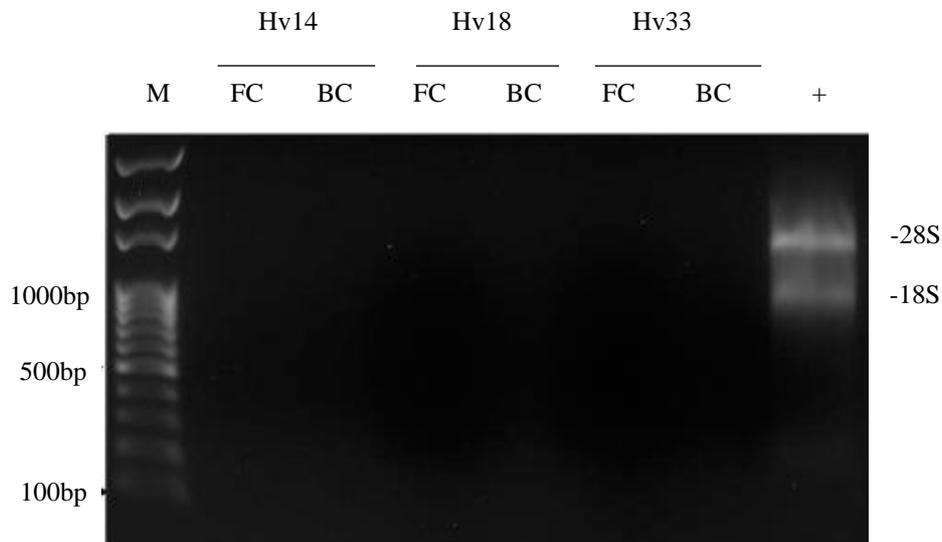


Figure 3.25: 1% agarose gel electrophoresis of flow through from cleaned RNA extraction columns run against a previously extracted RNA sample. M: 100bp DNA Solis Biodyne ladder, FC: filtration column, BC: binding column.

The flow through from the samples were run in an RT-PCR to determine if there was any amplification of LNYV that was previously present on the columns. No amplification was detected in any of the samples when run on a 1.5% agarose gel, indicating the columns were free of positively infected LNYV after the cleaning step (Figure 3.26).

The columns were subsequently used to extract total RNA from other *N. glutinosa* samples that had been inoculated with LNYV. The concentrations of the samples and the A_{260}/A_{280} ratios determined all indicated good quality RNA was obtained, though, in general, the values were lower than those previously seen upon first using the commercial columns. However, this could be due to techniques used in the RNA extraction method, and not necessarily the washing and reuse of the columns.

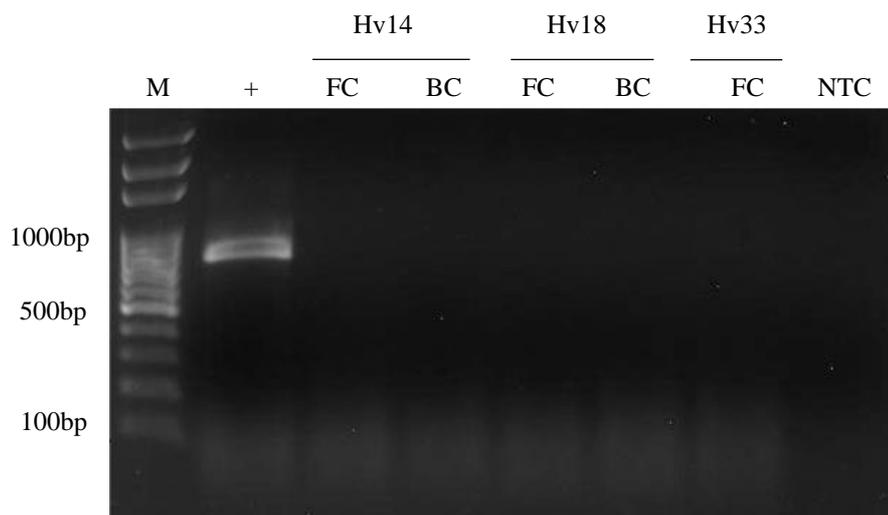


Figure 3.26: Detection of LNYV-Hv14, Hv18 and Hv33 residual material by RT-PCR in the cleaned binding and filtration nucleic acid columns run with LNYV440F and 118R primers at an annealing temperature of 50°C. M: 100bp DNA Solis Biodyne ladder, NTC: No template control, FC: filtration column, BC: binding column

A second reuse of the columns using the wash protocol was used but it was noted that afterwards some of the filters within the column appeared to be warped or have holes in them, rendering them unusable. It was decided that the columns should only be reused once going forward and that future studies could look at optimising the procedure for more uses of the columns, as it has been reported that up to 15 uses are possible (Nicosia et al. 2010), but this was not confirmed in this study. It was likely that incubating the columns at 75°C led to the observed warping, so alternatives to this may help optimise the procedure. However, reusing the columns once allows for leftover reagents from the commercial kits to be used, reducing wastage.

3.3.3 Infection studies for preparing material for use in RT-qPCR studies

This section reviews the assessment of the RNA extracted from this material, confirmation of virus infection by RT-PCR, and the RT-qPCR experiments to quantitatively determine the expression levels of both the candidate reference genes and the GOI.

3.3.3.1 Obtaining RNA from uninfected and subgroup II infected *N. glutinosa* for qPCR experiment

Ninety-five *N. glutinosa* plants were inoculated with LNYV-Hv19_057 while 30 were mock inoculated. Eighteen inoculated plants were each sampled at 0, 7, 14 dpi, as well as five mock inoculated plants. Visual inspection for LNYV-like symptoms was conducted and leaves suspected to be infected were sampled. If no symptoms were present, plants were randomly chosen to fully sample up to 18 plants at each time point. Following RNA extraction, if the sample RNA concentration was < 70 ng/μl or the A₂₆₀ / A₂₈₀ ratio was less than 2, the samples were not used in an RT-PCR (Table 3.6). RT-PCR analysis of the presence of LNYV showed that no samples were infected with the virus (Figure 3.27). Some RT-PCRs were repeated with an increased amount of RNA (500 ng) but no LNYV product was observed (Figure 3.27g)

Table 3.6: Summary of inoculum, sample day, plant ID and whether the sample tested positive for Hv19 by RT-PCR. ‘-ve’ denotes sample tested negative for infection.

Inoculum	Day sampled	Plant sample #	Infected
LNYV-Hv19_057	0	1	Not tested
	0	2	-ve
	0	6	-ve
	0	7	-ve
	0	10	Not tested
	0	11	Not tested
	0	14	-ve
	0	16	-ve
Buffer	0	Mock 1	-ve
	0	Mock 2	-ve
	0	Mock 3	Not tested
	0	Mock 4	Not tested
	0	Mock 5	-ve
	7	31	-ve

LNYV- Hv19_057	7	32	-ve
	7	33	-ve
	7	34	-ve
	7	35	-ve
	7	36	-ve
	7	37	-ve
	7	39	-ve
	7	41	-ve
	7	42	-ve
	7	43	-ve
	7	44	-ve
	7	45	-ve
	7	95	-ve
	Buffer	7	Mock 8
7		Mock 19	-ve
7		Mock 23	-ve
LNYV- Hv19_057	14	50	-ve
	14	65	-ve
	14	76	-ve
	14	77	-ve
	14	78	-ve
	14	79	-ve
	14	80	-ve
	14	81	-ve
	14	82	-ve
	14	83	-ve
	14	84	-ve
	14	85	-ve
	14	86	-ve
	14	87	-ve
	14	88	-ve
	14	89	-ve
14	90	-ve	
Buffer	14	Mock 21	-ve
	14	Mock 24	-ve
	14	Mock 25	-ve

3.3.3.2 Confirmation of infection of Hv19 inoculated *N. glutinosa* by one step RT-PCR

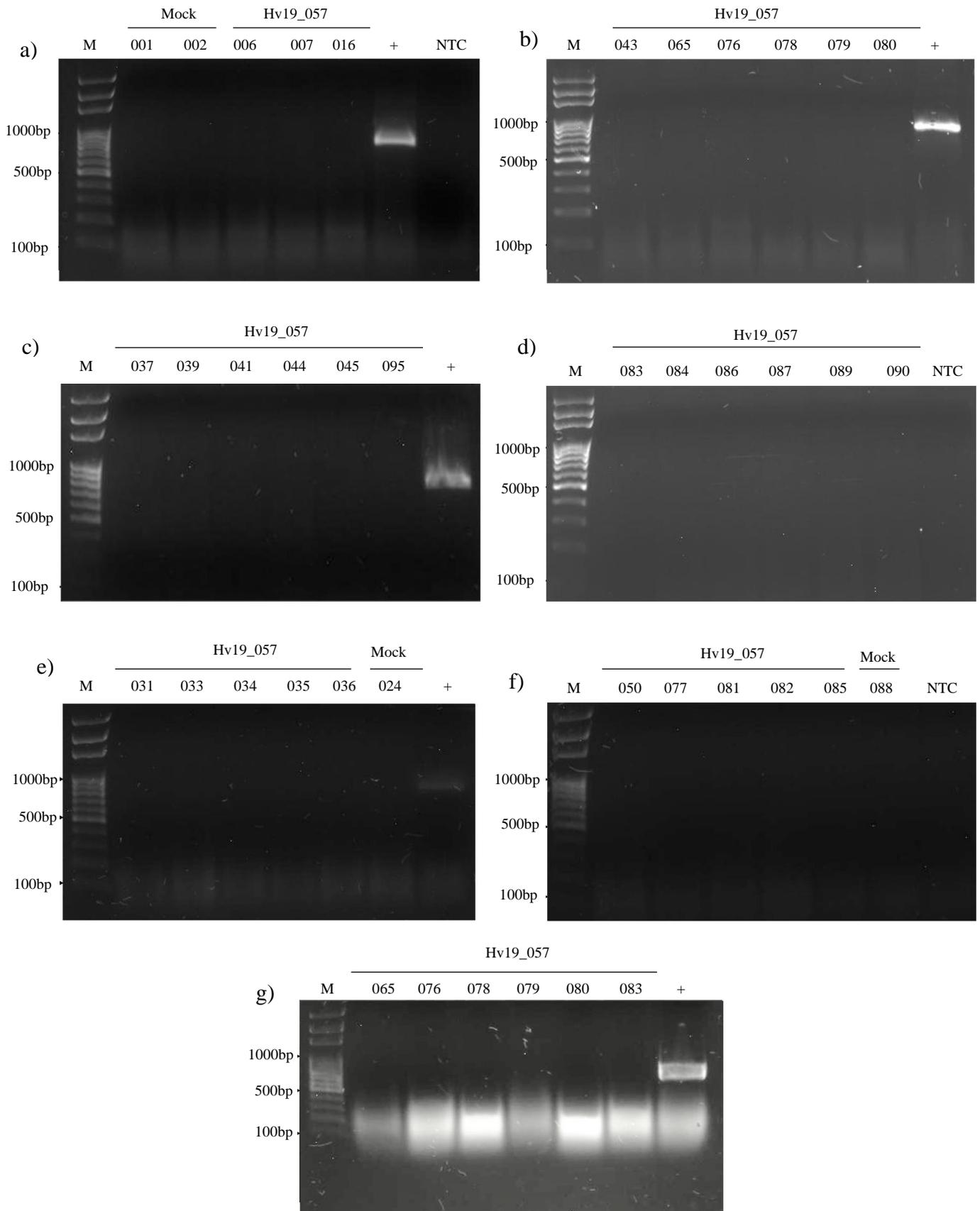


Figure 3.27: Detection of LNYV-Hv19_057 by RT-PCR using the LNYV440F/1185R primer combination with an annealing temperature of 50°C. Samples in gels a) and b) were run in the same RT-PCR reaction, as were samples in gels c) and d), and also e) and f). Gel g) has select samples run with 500 ng template RNA instead of 300 ng M: 100bp DNA Solis Biodyne ladder, NTC: No template control.

Since no LNYV could be detected in plants that were sampled at 0, 7 and 14 dpi, plants were allowed to continue growing until 21 and 28 dpi. However, no symptoms were observed at days 21 and 28. It was decided not to use reagents to extract the RNA from these samples nor test them for confirmation of LNYV infection. To date in this project, no asymptomatic samples had been found to be LNYV positive and the extraction and testing of samples in this case was deemed a waste of time and resources, as it was unlikely enough samples would be obtained to fulfil the requirements for the qPCR projects.

3.3.3.3 Changes to experimental design

After two growth experiments of not being able to obtain the necessary number of LNYV infected biological replicates at 14, 21 and 28 dpi, and due to no symptoms being observable at 0, 1 and 7 dpi, it was decided to amend the experimental design for the RT-qPCR experiment.

The number of time points was reduced to testing samples obtained only from day 28 dpi, and the biological samples to be used were the LNYV infected samples and mock inoculated samples from the subgroup 2 (Hv19) infection rate study in Section 3.3.1.3.2. In addition, it was decided to use the positive LNYV subgroup I infected (Hv14) samples and mock inoculated samples from the subgroup I infection rate study described in Section 3.3.1.3.1. This would allow the gene expression between subgroups to be compared, as well as between infected and uninfected. Though the size of the study is far smaller than initially conceived, it may still provide an indication of changes in expression of the genes of interest, as well as if the candidate reference genes are suitable for use as reference genes in future RT-qPCR studies. A summary of the samples used is present in Table 3.7.

Table 3.7: Summary of Hv14 and Hv19 and uninfected leaf material used in RT-qPCR experiments

Inoculum	LNYV Subgroup
Hv14_011	1
Hv14_019	1
Hv14_023	1
Hv14_029	1
Hv19_036	2
Hv19_043	2
Hv19_054	2
Hv19_057	2
Hv19_059	2
Hv19_060	2
Hv14_Mock_01	-
Hv14_Mock_02	-
Hv14_Mock_03	-
Hv19_Mock_01	-
Hv19_Mock_02	-
Hv19_Mock_03	-

3.3.4 Analysis of *N. glutinosa* gene expression following infection by LNYV subgroups I & II

3.3.4.1 Experimental design

Table 3.8 shows the sample layout used in 96 well plates for the RT-qPCR assays. Each gene was tested for separately (columns 1-9) in triplicate, in two replicate plants (rows A - C and D - F). Each primer for each gene was also tested in duplicate without template (NTC, rows G and H). The NTC assay for *CPK3* is shown in column 1, for *SGS3* in column 2, and so on along to *SAND* in column 8.

A standard sample was included, in triplicate, on every plate to allow plate-to-plate comparisons. This standard was Mock_037, an LNYV mock inoculated *N. glutinosa* plant sampled at 28 dpi.

Table 3.8: RT-qPCR plate design for the LNYV gene expression study. NTC refers to the samples with no template for the primers. ‘-’ represents empty wells on the plates.

		1	2	3	4	5	6	7	8	9	10	11	12
Biological sample 1	A	<i>CPK3</i>	<i>SGS3</i>	<i>WRKY70</i>	<i>Actin</i>	<i>EF1α</i>	<i>Ntubc2</i>	<i>PP2A</i>	<i>SAND</i>	Standard	-	-	-
	B	<i>CPK3</i>	<i>SGS3</i>	<i>WRKY70</i>	<i>Actin</i>	<i>EF1α</i>	<i>Ntubc2</i>	<i>PP2A</i>	<i>SAND</i>	Standard	-	-	-
	C	<i>CPK3</i>	<i>SGS3</i>	<i>WRKY70</i>	<i>Actin</i>	<i>EF1α</i>	<i>Ntubc2</i>	<i>PP2A</i>	<i>SAND</i>	Standard	-	-	-
Biological sample 2	D	<i>CPK3</i>	<i>SGS3</i>	<i>WRKY70</i>	<i>Actin</i>	<i>EF1α</i>	<i>Ntubc2</i>	<i>PP2A</i>	<i>SAND</i>	-	-	-	-
	E	<i>CPK3</i>	<i>SGS3</i>	<i>WRKY70</i>	<i>Actin</i>	<i>EF1α</i>	<i>Ntubc2</i>	<i>PP2A</i>	<i>SAND</i>	-	-	-	-
	F	<i>CPK3</i>	<i>SGS3</i>	<i>WRKY70</i>	<i>Actin</i>	<i>EF1α</i>	<i>Ntubc2</i>	<i>PP2A</i>	<i>SAND</i>	-	-	-	-
	G	NTC	NTC	NTC	NTC	NTC	NTC	NTC	NTC	NTC	-	-	-
H	NTC	NTC	NTC	NTC	NTC	NTC	NTC	NTC	NTC	-	-	-	-

3.3.4.2 High resolution melt curve analysis of product amplification

At the end of each RT-qPCR, a HRM step was included to confirm amplification of one specific product for each pair of primers. A melt curve analysis was generated per primer pair per run to confirm this. While only the melt curve analyses from the first RT-qPCR are included here, the same profiles were observed for every RT-qPCR carried out. Examples for each primer pair for the candidate reference genes are presented in Figure 3.28 and Figure 3.30.

3.3.4.2.1 HRM analysis of candidate reference genes

For the candidate reference genes, with the exception of *EF1 α* , discussed in the following section, each primer pair generated the expected one specific product (Figure 3.28). *Actin*, *Ntubc2*, *PP2A* and *SAND* had melting temperatures of approximately 82.2°C to 82.6°C, 79.5°C, 79.3°C and 77.1°C, respectively.

In the first RT-qPCR run with optimised conditions based on the amendments made at the end of the previous chapter, the melt curve analysis suggested that *EF1 α* did not amplify one specific product, as multiple peaks were presented in the data (Figure 3.28b). The qPCR product was run on a 1.5% agarose gel using electrophoresis and with the exception of *EF1 α* only one product was identified for each of the six samples tested (Figure 3.29).

For *EF1 α* , the HRM profile suggested amplification of two products, one with a melting temperature of approximately 80°C and one of approximately 84°C. The product with the T_m of 80°C appeared to be more abundant since it gave a higher relative fluorescence value. This would suggest the other product is larger, and therefore unlikely to be primer dimer, and probably a different sequence. The gel in Figure 3.29 does not show amplification of a larger product. Therefore, the product is not larger, or has amplified to a level below the sensitivity of the gel.

The multiple peaks on the melt curve analysis maybe be the result of the primers binding to different isoforms of the same gene (Camacho Londoño and Philipp 2016). In *A. thaliana* and *N. attenuata* multiple isoforms of *EF1 α* with similar structures have been reported, therefore it may be possible there are multiple isoforms of *EF1 α* in *N. glutinosa*. This would mean that *EF1 α* primers may be binding to slightly different sequences and, hence, the reason for multiple HRM peaks and the lack of a second product on the gel. As a result of this, the *EF1 α* primer pair was no longer considered to be suitable for the study and was not included in subsequent RT-qPCR experiments for analysis.

For *Actin*, whilst the HRM profile suggested amplification of one product, with the melt curves predominantly overlapping, it did appear that there may be a range of temperatures over which the single product would melt, ranging from 82.2°C to 82.6°C. Previous agarose gels (Figure

2.26) confirmed *Actin* only amplified one product, and so the range of temperatures across 0.5°C suggested by the melt curves could be as a result of the difference in GC content between the two primers used resulting in multiple melting domains in the product (Dwight et al., 2011). It was decided to test the *Actin* primer as a candidate reference gene using these primers, but potentially redesigning the primers in the future to avoid these multiple melting domains in the product may be necessary to ensure the most robust data is obtained.

Therefore, the remaining candidate reference genes were *Actin*, *Ntubc2*, *PP2A* and *SAND*.

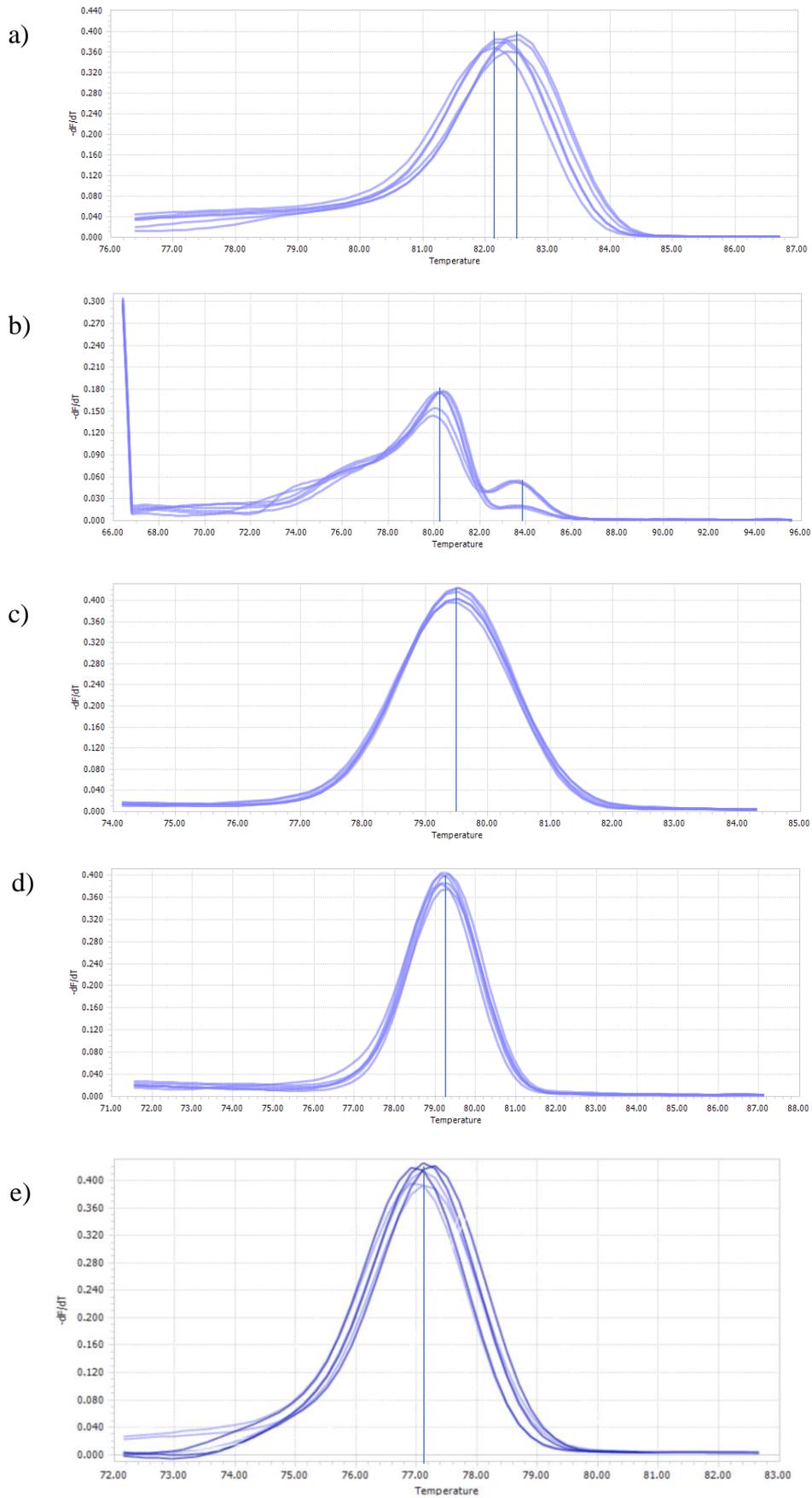


Figure 3.28: a) *Actin*, b) *EF1 α* , c) *Ntubc2*, d) *PP2A* e) *SAND* melt curve analysis. Vertical blue lines indicate the presence of an amplicon at a particular temperature.

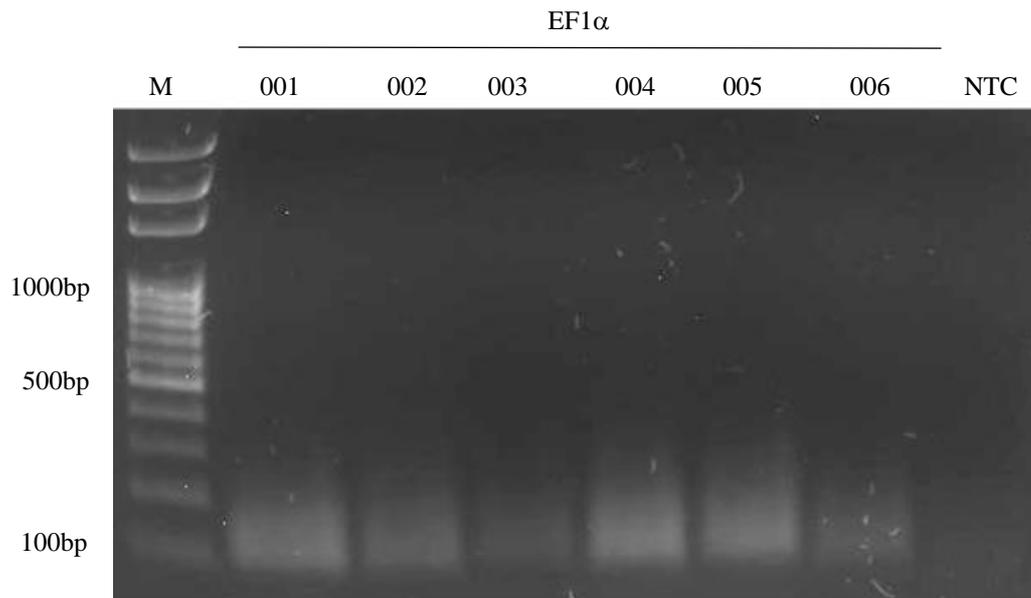


Figure 3.29: Analysis of RT-PCR by agarose gel electrophoresis of EF1 α primer with Hv19 and Hv14 infected samples to confirm amplification of single products. M: 100bp DNA Solis Biodyne ladder, NTC: No template control

3.3.4.2.2 HRM analysis of GOIs

HRM analysis of the RT-qPCR products amplified using the GOI primer pairs are shown in Figure 3.30. For the candidate target genes, each primer pair generated the expected one specific product. *CPK3*, *SGS3* and *WRKY70* had melting temperatures of approximately 79.8°C, 80.3°C and 78.9°C, respectively. Thus, all three genes were included in the RT-qPCR analyses.

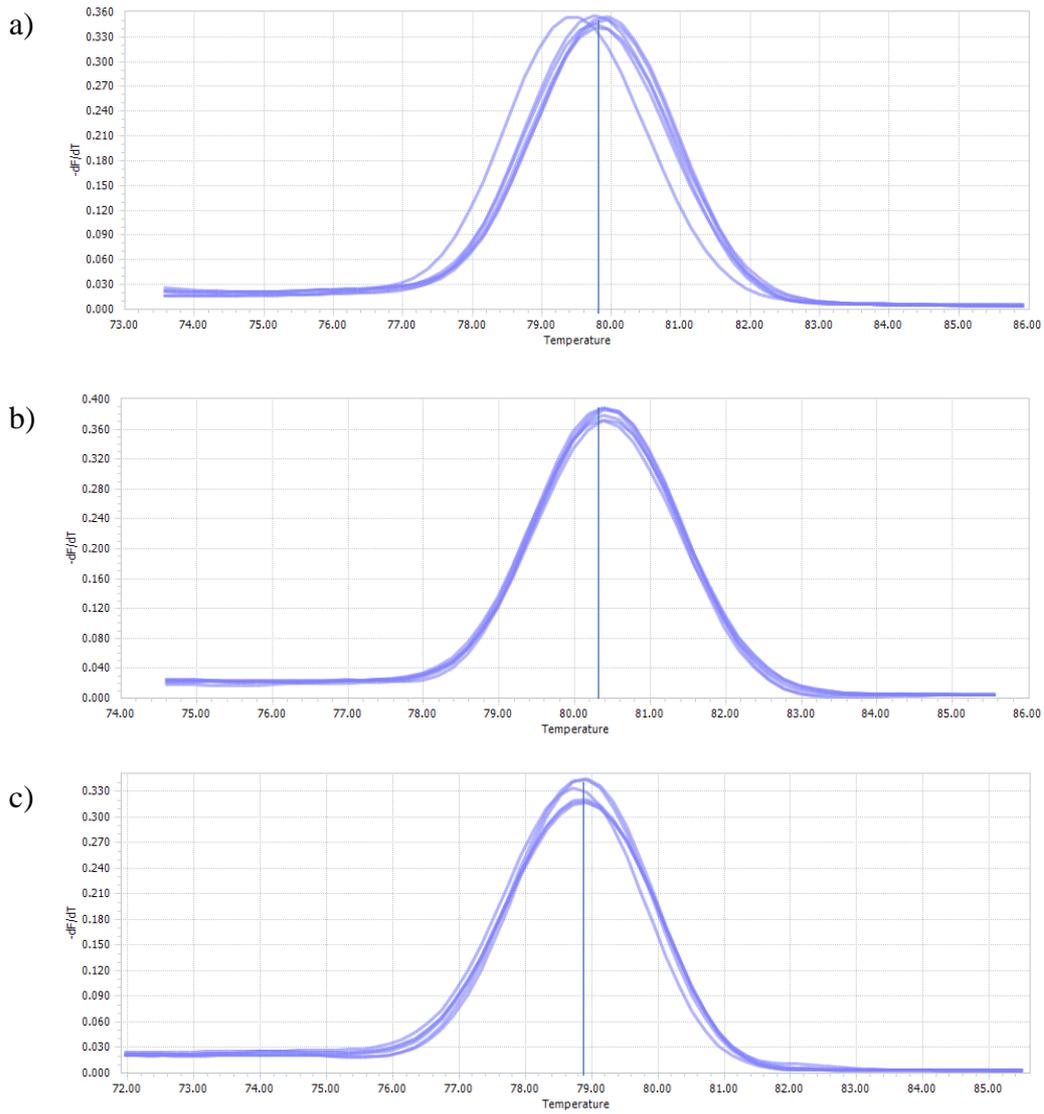


Figure 3.30: a) *CPK3* b) *SGS3* c) *WRKY70* melt curve analysis. Vertical blue lines indicate the presence of an amplicon at a particular temperature.

3.3.4.3 PCR amplification efficiencies of genes across all plates

Linear regression analysis using LinRegPCR 11.1 (Ruijter et al. 2009) allowed the amplification efficiencies of each gene to be determined across all plates. Generally, values above 1.8 are considered acceptable, indicating an amplification efficiency of at least 90% (Ruijter et al. 2009). Efficiency values obtained for all the genes were above 1.8, indicating acceptable amplification efficiencies (Table 3.9).

Table 3.9: Table showing the PCR amplification efficiency of all of the genes tested by RT-qPCR in this research

Gene	Amplification efficiency
Actin	1.843
Ntubc2	1.936
PP2A	1.942
SAND	1.868
CPK3	1.913
SGS3	1.895
WRKY70	1.925

3.3.4.4 Outliers

3.3.4.4.1 Standards

The raw Cq data values were obtained for all sample wells in all plates. Values where amplification had failed were removed from the analyses. From the remaining data, the Cq values of the samples and standards from all plates were analysed to identify any outlier values to remove them to obtain a more valid and representative dataset. Figure 3.31 shows the box plot analysis; no outliers were identified.

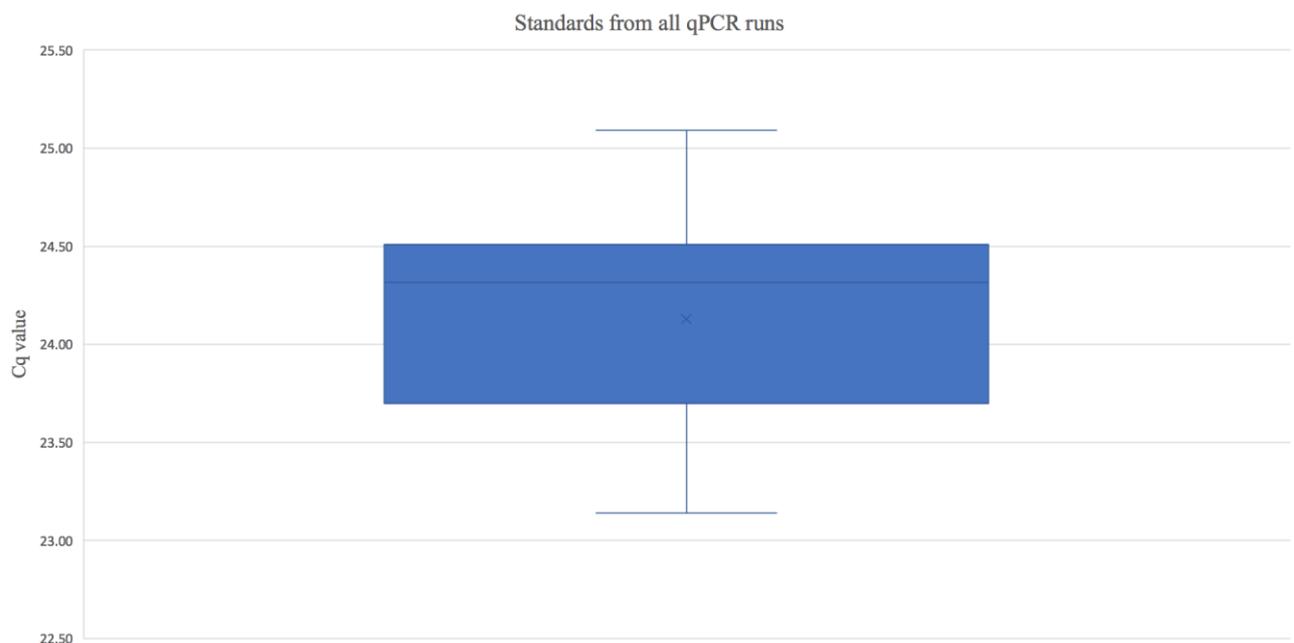


Figure 3.31: Boxplot showing the distribution of Cq values of the standards, a sample identified as Mock_037, run across all plates in the qPCR experiments

The difference between individual averaged standard plate values to the mean of the standards from all plates was calculated. The differences were applied to the Cq values of the non-standard samples on each plate to normalize all of the data. This allowed plate to plate comparisons to be made, giving the opportunity to determine the relative expression of the different genes in uninfected and infected samples from different experiments run at different times.

3.3.4.4.2 Candidate reference gene normalised data

The Cq values from the different biological replicates were grouped based on candidate reference gene and experimental condition (uninfected and infected) and subgroup (I or II), and again box plots were created to determine outlier values in the normalised data (Figure 3.32). Outlier values were removed from the subsequent analysis to obtain a more representative data set.

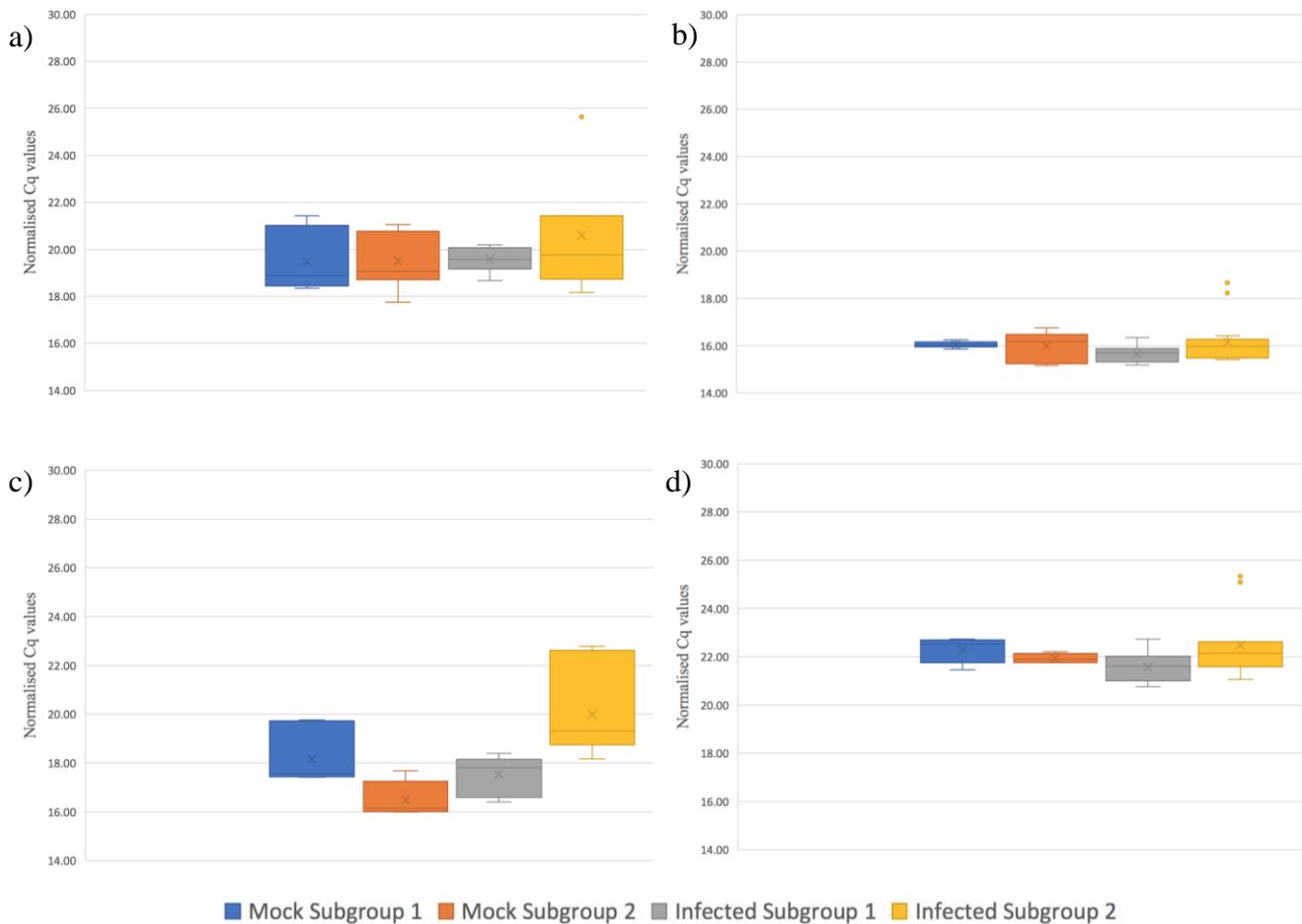


Figure 3.32: Boxplots showing the distribution of Cq values of a) *Actin*, b) *Ntubc2*, c) *PP2A*, d) *SAND* under different experimental conditions across all plates in the qPCR experiments. Outliers are visible as filled in dots, and were removed from the subsequent analysis.

The distribution of normalised Cq values for *Actin* was consistent across the four experimental groups, with the mean of three of the four groups being approximately cycle 19.5, with the subgroup II infected mean a Cq value of approximately 21 (Figure 3.32a). However, the range of values in some groups such as the mock subgroup I and the LNYV infected subgroup II was spread over three cycles (~18.5 – ~21.5), suggesting that the accumulation of *Actin* mRNA is likely to be unstable as that number of cycles indicates a large difference in transcript numbers. An outlier was removed from the infected subgroup II data for subsequent analysis.

The distribution of Cq values for *Ntubc2* was consistent across the four experimental groups, with the mean of all groups being approximately cycle 15.8-16.1 (Figure 3.32b). The range of values in the groups was low, ranging from ~15.2 – ~16.2, indicating *Ntubc2* mRNA accumulation is likely to be stable between treatments. Two outliers were removed from the infected subgroup II data for subsequent analysis.

Analysis of the normalised Cq values for *PP2A* showed inconsistent distribution across all four experimental groups, with the means being at least 0.5 cycles different from one another (Figure 3.32c). The range of normalised Cq values across the groups was spread over six cycles (~16 – ~22.5), indicating the expression of this gene is likely to be unstable, as that number of cycles suggests a large difference in transcript accumulation. No outliers were removed for subsequent analysis.

The distribution of Cq values for *SAND* were consistent across the four experimental groups, with the means of all four groups being placed between ~21.8 to ~22.6 (Figure 3.32d). The range of values in the groups was low, ranging from ~21 – ~22.5, indicating *SAND* accumulation is likely to be stable between treatments. Two outliers were removed from the infected subgroup II data for subsequent analysis.

A boxplot was constructed using the normalised data, with outliers removed, to determine the likely most to least stable genes, visually, before processing the data using geNorm and BestKeeper (Figure 3.33). Corroborating the information obtained from the graphs in the previous section, combining both mock and infected samples, *Ntubc2* and *SAND* appeared to be the most stable since the range of Cq values was lowest, followed by *Actin* then *PP2A*. An outlier was removed from the *Actin* data before further analysis.

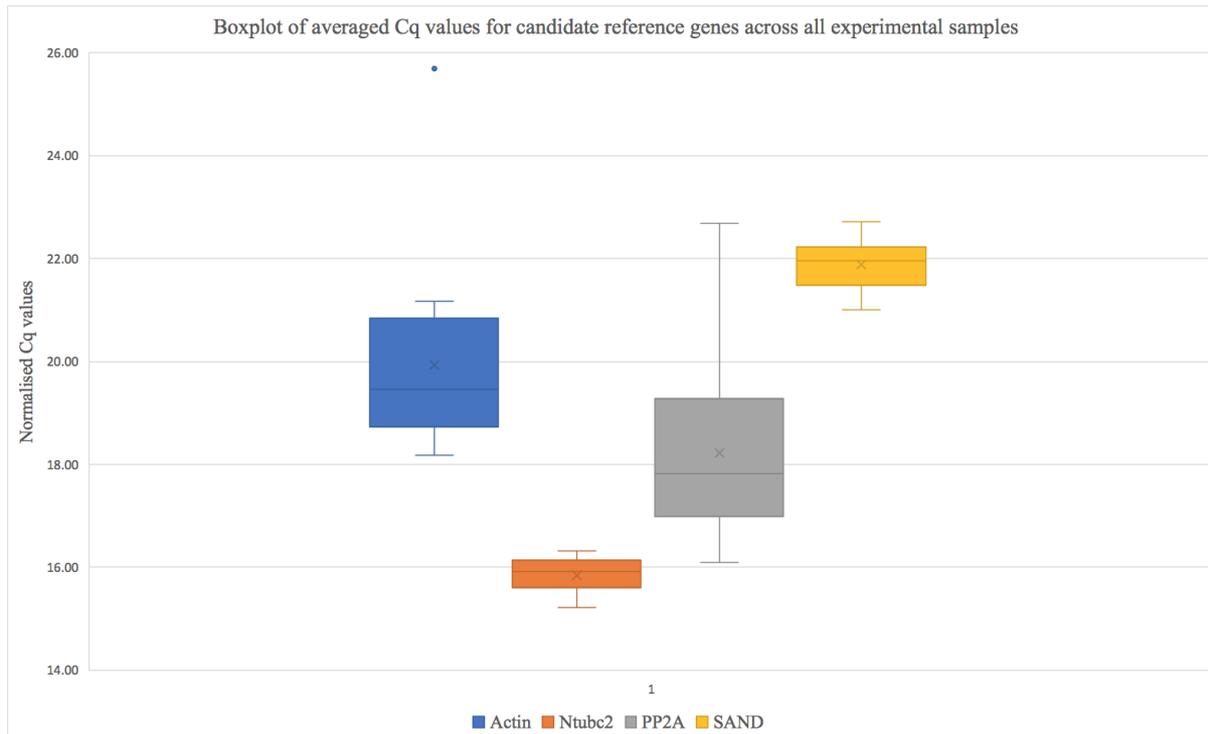


Figure 3.33: Boxplot of averaged Cq values for all candidate reference genes across all experimental samples. Outliers are visible as filled in dots.

3.3.4.4.3 GOI normalised data

The Cq values from the different biological replicates were grouped based on GOI and experimental condition (uninfected and infected) and subgroup (I or II), and again box plots were created to determine outlier values in the normalised data (Figure 3.34). Outlier values were removed from the subsequent analysis, in order to obtain a more representative data set.

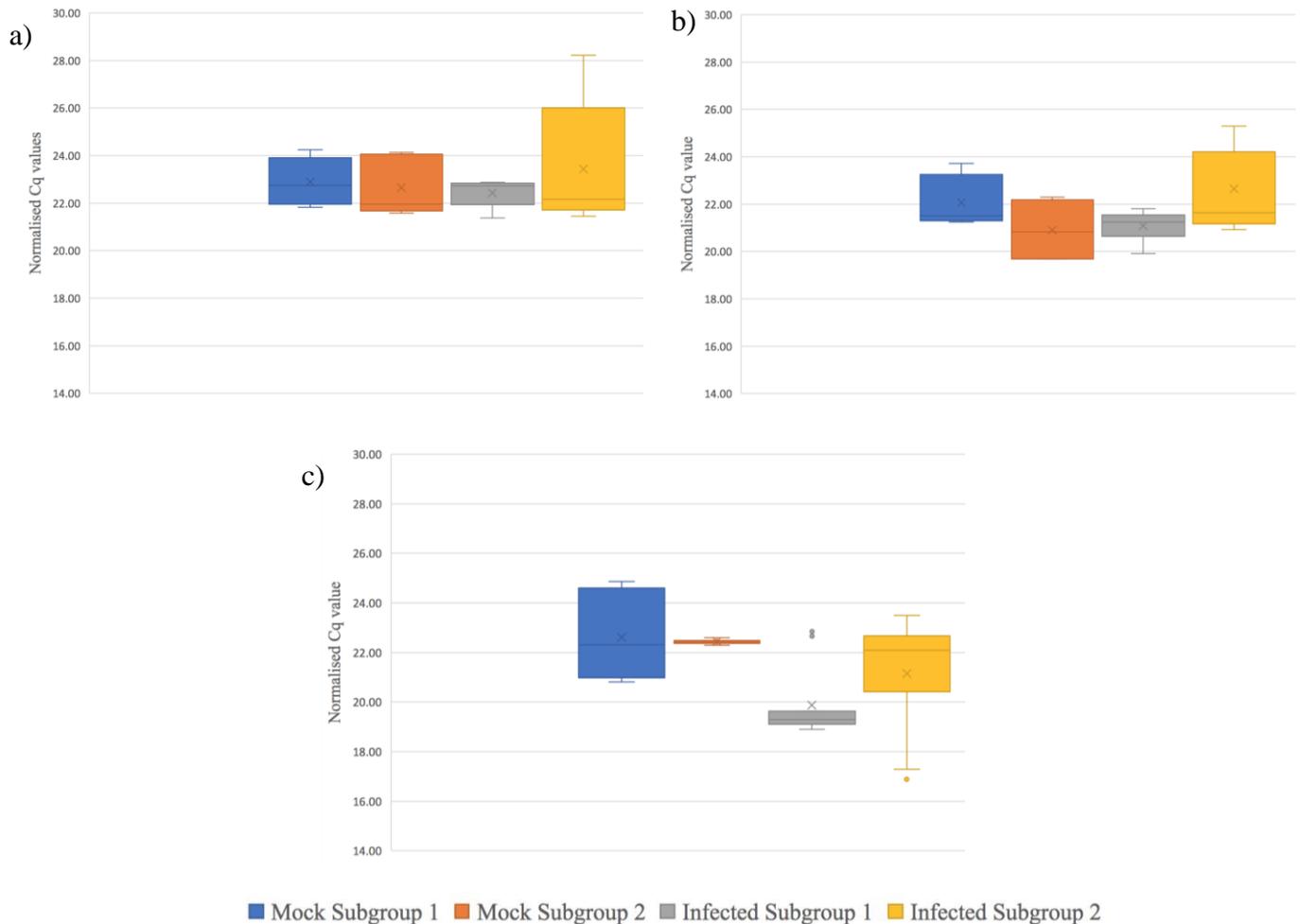


Figure 3.34: Boxplots showing the distribution of Cq values of a) *CPK3* b) *SGS3* c) *WRKY70* under different experimental conditions across all plates in the qPCR experiments.

Figure 3.34a shows the distribution of normalised Cq values for *CPK3* was consistent across the mock inoculated groups, with the means placed between ~22.5 to ~23.5 and the range between ~22 and ~24. However, there was variability in the infected groups, with a narrow range in subgroup I from ~22 – ~22.6 compared to a larger range between ~22 - ~26. This could indicate variable expression upon infection by different subgroups of LNYV, compared to the most stably identified reference genes. No outliers were removed from the dataset.

The distribution of Cq values for *SGS3* varied across the four experimental groups, with the means of the four groups being placed between ~21 to ~23 (Figure 3.34b). The range of values in the groups was larger, ranging from ~21.9 – ~24), meaning that there could be some variation in gene expression between the different conditions, though not necessarily by a large margin, once compared to the most stably identified reference genes. There is not a lot of overlap in the mock samples across the two subgroups. There appears to be a wider range of values in subgroup II than subgroup I as well. No outliers were removed from the dataset.

The distribution of Cq values for *WRKY70* varied across the four experimental groups, with the means of the four groups being placed between ~19.4 to ~22.5 (Figure 3.34c). The range of values in the groups was larger, ranging from ~19.0 – ~24.6), meaning that there could be some variation in gene expression between the different conditions, though not necessarily by a large margin, once compared to the most stably identified reference genes. There is a large range in the subgroup I mock samples but subgroup II mock samples were within a more limited range. Differences in the mean values between subgroup I and subgroup II infections suggests there would be significant differences in expression between these. After outliers were removed there was no overlap in Cq values between subgroup I and the other experimental conditions.

3.3.4.5 Analysis of candidate reference genes

3.3.4.5.1 GeNorm

Using the geNorm algorithm (Vandesompele et al. 2009), the normalised, outlier removed Cq data from the amplification of the candidate reference genes were processed to rank the genes from the least to most stable. The geNorm software generated this information as two graphs (Figure 3.35 and Figure 3.36). Average expression stability values in the form of an M value was generated for each of the four candidate genes. These M values were 1.6 for *Actin*, 0.6 for *PP2A* and 0.24 for both *Ntubc2* and *SAND*. Expression values <1.5 means a gene can be considered stable and be used as a reference gene (Baek et al. 2017; Lilly et al. 2011; Liu et al. 2012; Schmidt and Delaney 2010). Ranked, these values place *Ntubc2* as the most stable gene, followed by *SAND*, *PP2A* and *Actin*, as shown in Figure 3.35.

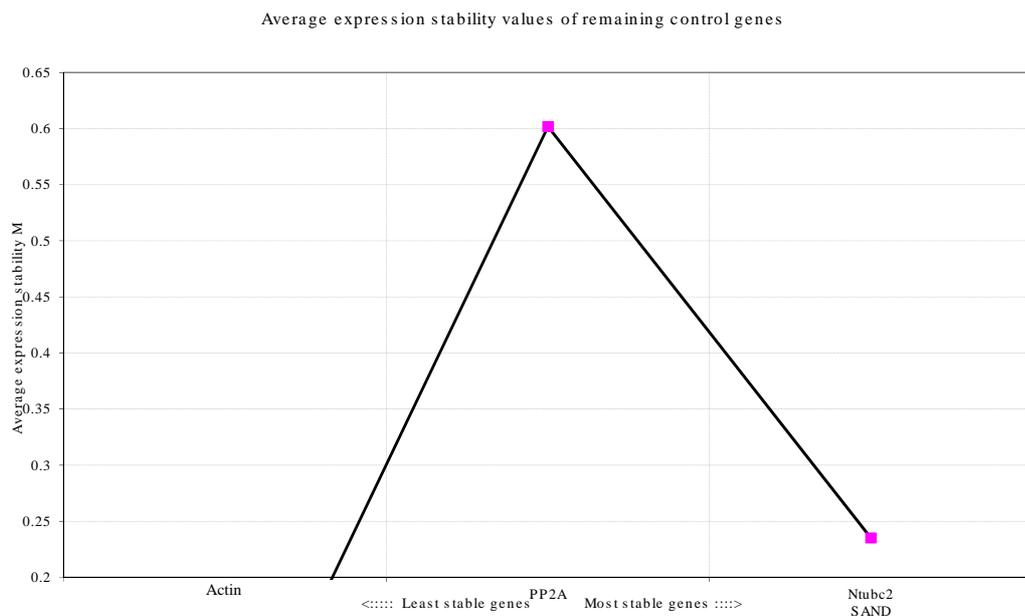


Figure 3.35: GeNorm generated graph showing stability of candidate reference genes from least stable on the left of the graph to most stable on the right

Pairwise variation values were also generated for different combinations of the four reference genes. The GeNorm algorithm removes the least stable gene, in this case *Actin*, and gives a remaining confidence value in using the remaining genes as reference genes. The pairwise value generated in this study was 0.172 on the basis of using the three reference genes *Ntubc2*, *PP2A* and *SAND* (Figure 3.36). The use of *Ntubc2*, *SAND* and *PP2A* fall just outside the

accepted 0.15 value (Ling and Salvaterra 2011; St-Pierre et al. 2017), meaning additional reference genes are necessary in future experiments in order to have a fully valid qPCR experiment with enough reference genes.

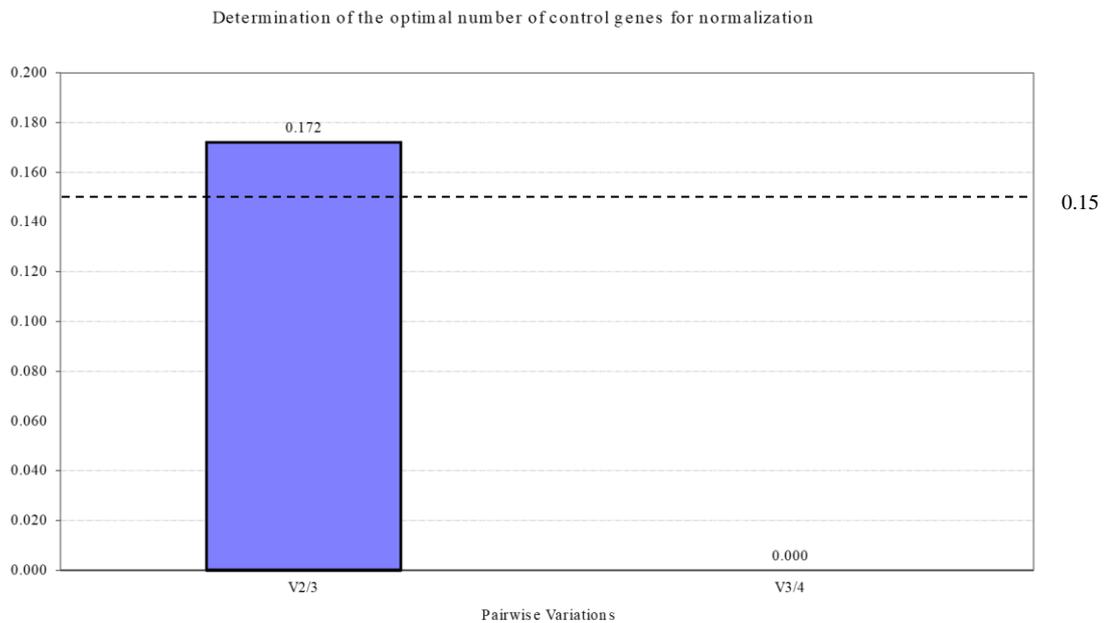


Figure 3.36: GeNorm generated graph of pairwise variation for the candidate reference genes analysed in this research.

3.3.4.5.2 BestKeeper

Multiple sources of validation using different software and algorithms are often utilised to confirm that conclusions drawn utilising one method of analysis can be corroborated by another (Baek et al. 2017). The normalized, outlier removed Cq values from the different biological replicates across the different conditions were also analysed using BestKeeper software (Pfaffl et al. 2004). The values returned suggested the most to least stable reference genes were in the order *Ntubc2*, *SAND*, *Actin*, *PP2A* (Table 3.10). However, the confidence levels for *Ntubc2* and *SAND* were > 0.05 , therefore, further experimental evidence may be necessary in the future to fully corroborate this.

Table 3.10: BestKeeper generated values for the four candidate reference genes, showing the gene, number of samples, coefficient of correlation and p-value for each.

Gene	Ntubc2	SAND	Actin	PP2A
Samples (n)	12	12	14	12
Standard deviation	0.22	0.46	0.84	0.94
Coefficient of correlation (r)	0.358	0.571	0.644	0.922
p-value	0.254	0.052	0.013	0.001

3.3.4.6 Analysis of GOIs

Using *SAND* and *Ntubc2* as reference genes, the normalised, outlier removed Cq values from the GOIs were first averaged for the different biological replicates across the plates (Table 3.11). Using the $2^{-\Delta\Delta Cq}$ calculation, the relative differences between the expression levels of the GOIs and the reference genes were determined (Table 3.12 and Table 3.13).

According to these values, it would indicate that *CPK3* is upregulated by approximately 12% in response to a subgroup I infection, downregulated by 32% in a subgroup II infection, and is expressed 37% higher in response to LNYV-subgroup 1 than subgroup 2. *SGS3* is upregulated by approximately 49% in response to a subgroup I infection, downregulated by 79% in a subgroup II infection, and is expressed 266% higher in response to LNYV-subgroup 1 than subgroup II. Finally, *WRKY70* is upregulated by approximately 685% in response to a subgroup I infection, upregulated by 8% in a subgroup II infection, and is expressed 720% higher in response to LNYV-subgroup 1 than subgroup 2.

Table 3.11: Average Cq, ΔCq and $\Delta\Delta Cq$ values between reference genes and GOI under different experimental conditions.

	Average Cq			Ntubc2 as reference gene						SAND as reference gene					
	SI	SII	Mock	SI ΔCq	SII ΔCq	Mock ΔCq	$\Delta\Delta Cq$ SI/Mock	$\Delta\Delta Cq$ SII/Mock	$\Delta\Delta Cq$ SI/SII	SI ΔCq	SII ΔCq	Mock ΔCq	$\Delta\Delta Cq$ SI/Mock	$\Delta\Delta Cq$ SII/Mock	$\Delta\Delta Cq$ SI/SII
CPK3	22.42	22.90	22.77	6.63	7.05	6.75	-0.12	0.3	-0.42	0.84	1.02	0.65	0.19	0.37	-0.18
SGS3	21.10	22.65	21.49	5.31	6.8	5.47	-0.16	1.33	-1.49	-0.48	0.77	-0.63	0.15	1.4	-1.25
WRKY70	19.24	22.09	22.52	3.45	6.24	6.5	-3.05	-0.26	-2.79	-2.34	0.21	0.4	-2.74	-0.19	-2.55
Ntubc2	15.79	15.85	16.02												
SAND	21.58	21.88	22.12												

Table 3.12: Averaged $\Delta\Delta Cq$ values of GOI under different experimental conditions against *SAND* and *Ntubc2* reference genes in uninfected and LNYV subgroup I and II infected *N. glutinosa*.

	Average $\Delta\Delta Cq$ SI/Mock	Average $\Delta\Delta Cq$ SII/Mock	Average $\Delta\Delta Cq$ SI/SII
CPK3	0.035	0.335	-0.3
SGS3	-0.005	1.365	-1.37
WRKY70	-2.895	-0.225	-2.67

Table 3.13: $2^{-\Delta\Delta Cq}$ values for each of the three GOIs against *SAND* and *Ntubc2* reference genes in uninfected and LNYV subgroup I and II infected *N. glutinosa*.

	Subgroup I vs. Uninfected	Subgroup 2 vs. Uninfected	Subgroup I vs. Subgroup 2
CPK3	0.976	0.793	1.231
SGS3	1.004	0.388	2.585
WRKY70	7.438	1.169	6.364

3.4 Discussion

This is the first study conducted to try to determine a set of reference genes for gene expression studies in *N. glutinosa*, as well as examining the gene expression of a group of target genes in response to LNYV infection in this model plant species. Whilst the methods used in this experiment were based on existing techniques, it is important to review the processes utilised against the results to determine how future studies can develop on the research conducted here.

3.4.1 Inoculation of *N. glutinosa* using infected LNYV samples

No studies were present in the literature that outlined the success rate of the mechanical inoculation of LNYV into *N. glutinosa* plants, with research only reporting whether infection had been established at all as part of other studies. Existing protocols for mechanical inoculation were followed from previous studies, and the success rate of obtaining positive infection varied widely, ranging from 3% to 26.6% for Hv19 subgroup II isolates and 0% to 13.3% for subgroup I isolates. It is not known if the individual isolates have specific infection rates or whether the variation identified was down to other factors such as the inoculation procedure and sampling. This would be an interesting area of future research if repetition of the same isolates were used as inoculum and may help elucidate how effective the different subgroups are at infecting hosts, which may explain the distribution of the different subgroups in the environment.

The literature was inconsistent as to the optimum day to inoculate young *N. glutinosa* plants with LNYV to establish infection. Some experiments suggested as few as 8 days to 12 days, though more papers suggested regions closer to 21 days (Crowley 1967; Dietzgen et al. 1989; Randles and Coleman 1970; Wang et al. 2015). However, there was a higher rate of mortality for plants inoculated at this day in the experiments run in this research, therefore, this was amended to 28 days. However, the later plants are inoculated, the more difficult it may be for the virus to establish infection, as plant defence mechanisms may have developed sufficiently to counteract the invading pathogen (Roossinck 2015). Whilst mechanically inoculating at 28 days with a finger rub remained difficult due to the relatively small size of the leaves, higher rates of infection were obtained at this time. Therefore, 28 day plants were used for this

research. Future studies could investigate the optimal date for inoculation for different LNYV subgroup isolates, as this could potentially be different for each establishing infection.

Several other factors in this research could explain the variance in infection rate. For LNYV and other enveloped plant viruses, freezing at -80°C can have an effect on the recovery of infectious viruses, meaning the infectivity of the virus may be reduced using thawed positive compared to fresh leaf material. However, the scarcity of fresh LNYV samples from the field made this unavoidable for this particular study. Similarly, it is also important to consider that LNYV concentration in *L. sativa*, is generally lower than in *N. glutinosa*, and therefore when changing host species or when recovering virus from frozen samples, it is important to allow a build-up of virus titre by sequential inoculations to eventually obtain a far higher infection rate; however the timescale of this project and difficulties obtaining sufficient positive samples meant this was not achieved in this study, an area which should be improved upon if further research were to occur.

3.4.2 Sampling of LNYV inoculated LNYV samples

Sampling of the LNYV inoculated plants was conducted initially at 21 dpi; however, this yielded low infection rates in this research. This could have been because not all samples were tested using a molecular test to establish LNYV infection, but principally because no LNYV-like symptoms were identified during visual inspection of the plants. Previous researchers have reported mild symptoms being present on plants as early as 8-12 dpi; however, this was not the case in this study, with the earliest symptoms being identified at day 14. More obvious symptoms were present at day 21 and were most easily identified at day 28. Visual identification was approximately 75% successful. Infections were principally identified by the mottled dark green and light green appearance (for example areas indicated by red arrows in Figure 3.15a-f). These symptoms appear to be the most identifiable in *N. glutinosa* samples and can aid future researchers trying to visually identify infections. The size of infected plants compared to mock inoculated plants was not an accurate way of determining infection in this research, as plants were not necessarily smaller between uninoculated and infected plant samples. Symptoms were similar between subgroup I and subgroup II isolates and no distinctive difference was noted between the two that would aid subgroup identification as reported by Higgins et al., 2016. Further work needs to be done to identify LNYV isolates that

give severe symptoms on *N. glutinosa*, which previously has only been successful twice in Australia; one was the original isolate with a second being a garlic isolate identified in the 1980's. This would make it easier to identify infected plants visually.

Sampling of leaf material was planned to be identical from sample to sample, with three non-inoculated systemic leaves being sampled at the point of sampling. However, when sampling plants at 0 and 7 dpi, this was not always possible due to the small nature of the plants and the delayed growth of some of the plants due to the mechanical inoculations. Future research may have to determine an alternative way to sample at these time points if they are of interest.

3.4.3 RNA extraction methods

Extraction of quality RNA is a crucial step for a successful study utilising qPCR, with the necessity to remove genomic DNA to avoid false positives and contamination. Throughout this research, principally commercially available kits were utilised in order to extract the RNA from the infected or uninfected leaf material. These are well established kits that are suitable for obtaining good concentration and quality RNA such as Spectrum™ Plant Total RNA kit (Sigma-Aldrich, St Louis, MO, USA) or RNeasy Plant Mini Kit (Qiagen). Low quality RNA was likely due to too much leaf material being utilised in the extraction. The majority of the samples extracted had both an acceptable nucleic acid concentration and integrity. Disadvantages of commercial kits are the cost as well as having fewer extraction columns in the kits compared to the rest of the reagents, and whilst additional columns could be purchased separately, often a lot of reagents remain in the kits when other key components run out. As a result, protocols to reuse the nucleic acid filtration and binding columns were researched, as well as a protocol to clean the Eppendorf tubes which are specifically designed to house the extraction columns. The reuse protocol for the columns stated that the method was suitable for 10-15 reuses of the columns (Nicosia et al. 2010). However, it was determined that the heating step in a water bath warped the nucleic acid columns after a couple of applications, and so this finding was not confirmed in this research. However, there was no observable reduction in concentration or integrity of RNA by reusing the columns once, allowing the uses of a commercial kit to be doubled and allowing more of the remaining reagents to be utilised. Subsequent research can look at developing this protocol further to obtain further uses of the kit, providing quality RNA can still be maintained. This would allow researchers with limited

budgets to extend the uses of a commercial kit. Whilst alternative methods exist that don't rely on the use of commercial kits, such as a CTAB extraction, use of these were predominantly outside the scope of this research. A high throughput RNA extraction protocol would be beneficial for extracting large volumes of RNA from different samples from different time points, as the extraction method utilised in this research took several hours for a low volume of samples, so was relatively time consuming.

3.4.4 RNA integrity analysis using agarose gel electrophoresis

RNA integrity was measured using a Nanovue spectrophotometer for the entirety of the research. Previous research has suggested that the spectrophotometer may overestimate the concentration of the RNA in a sample as it may not be able to specifically differentiate between DNA and RNA (Ajithkumar 2018). The majority of samples extracted and analysed had an integrity of above 2 and a concentration above 100 ng/ μ l so were deemed acceptable for subsequent analysis, and any samples that were lower were likely to have been due to selecting an area of the leaf with a low virus titer or due to errors undertaking the RNA extraction, and not due to overestimation by the spectrophotometer.

3.4.5 Infection rate studies

In this research, there was an observable difference in the number of subgroup II isolates from mechanical inoculation compared to subgroup I isolates. Before the gene expression study, the inoculation and plant growth conditions were optimised and the same experiments were conducted between subgroup I and subgroup II isolates, though a wide variance in infection rate was obtained; 0%-13.3% for subgroup I and 26.6% in subgroup II . No papers had been previously found to report subgroup specific infection rates, and further experimentation would be necessary to support these values. Limited isolates of subgroup II were available so it is unknown if the subgroup II values applies to all isolates or just the Hv19 strain, and whether the same would be seen in other model plants.

3.4.6 Amplification of reference genes and GOI using specifically designed primers in a gene expression assay

The reference genes and GOIs identified in Chapter 2 as being suitable for study were amplified using RT-qPCR from multiple samples of both LNYV infected and uninfected *N. glutinosa* leaf material. A commercial one step RT-qPCR kit was utilised for this and, as per the MIQE guidelines, sufficient biological and technical replicates were included run on each plate in order to obtain statistically meaningful results, and to minimise experimental errors and variation between samples. Manual pipetting was conducted for the RT-qPCR assays and it may be beneficial for future researchers to utilise a robot for pipetting to minimise pipetting errors further.

There was insufficient time to redesign the *EFl α* , and *WRKY26* primers, though the multiple sequence alignments created in this study could be used to design alternative primers in subsequent studies if these genes remain of interest to other researchers, as the expression of these genes, if able to be elucidated may provide further information regarding the LNYV virus and how the different subgroups operate.

The volume of quality RNA necessary to conduct the RT-qPCR experiments was a limiting factor in this research due to the difficulty in obtaining infected LNYV samples. As the experimental design was changed late in the process to only use samples at 28 dpi, enough biological replicates were grown and infected for this small study. For larger time dependent studies, it may be beneficial to utilise a two-step RT-qPCR reaction, first synthesising complementary DNA before use in a qPCR experiment. This would allow lower amounts of RNA to be used, and for the cDNA to be tested with more primer pairs on the same sample and obtain more information about more genes in the same experiment if there are more GOI or reference genes to be tested in the same experiment.

3.4.6.1 Functional analysis of gene expression changes in reference genes and GOI

3.4.6.1.1 Reference genes

Ntubc2 and *SAND* were identified as having stable expression between subgroup I and II LNYV infected and uninfected *N. glutinosa*. From the information in Table 2.1, *Ntubc2* is

thought to be responsible for encoding a phosphorylation enzyme that is responsible for posttranslational modification of proteins. This may impact the functionality of a phosphorylated protein, and the *SAND* gene encodes a membrane protein with a role in vesicle trafficking and endocytosis. It can be suggested that neither the establishment of infection nor replication of the virus of either of the LNYV subgroups impacts these processes via the alteration of expression of these genes. However, it cannot be definitively said that the virus does not target these processes via alternative genes with similar functionality, that is, another phosphorylation enzyme or another protein associated with the cellular membrane or vesicle trafficking.

PDF2 and *Actin* were identified as not having stable expression across the different experimental conditions, making them unsuitable for acting as reference genes in an RT-qPCR study. Therefore, it can be asserted that LNYV may have an impact on the expression of these genes between uninfected and infected conditions. Table 2.1 shows that *Actin* codes for a component of the cytoskeleton which is responsible for cell motility and signalling processes, whereas *PDF2* encodes a regulatory subunit of a protein phosphatase enzyme that targets the amino acids serine and threonine. Whilst specific studies would have to be conducted to investigate this, it could be hypothesised that the virus impacts *Actin* and *PDF2*, which may result in signalling processes, cell motility and phosphorylation of other proteins being impacted that allow the new virion particles to be constructed or for the virus to spread to other cells. The expression of these genes could be studied in a time dependent manner to determine at what point this may occur in the life cycle of the virus; however, not enough data are available to confirm these hypotheses in this research.

3.4.6.1.2 GOI

From the gene expression values obtained in the qPCR assay, possible plant host pathways LNYV impacts and variances generated in these pathways in utilising different LNYV subgroups can begin to be understood. However, it is important to remember that these findings only reflect events at 28 dpi. It would be important to understand how these genes are influenced in a time dependent manner to fully understand the progression of viral infection in a host.

CPK3 is a multifunctional signalling protein involved in signal transduction in molecular pathways related to stomatal movement and adaptation to environmental factors such as drought, salt and cold stress (Arimura and Sawasaki 2010; Valmonte 2016). It has been identified as being an upstream component in the transcriptional activation of the plant defensin gene *PDF1.2*, a marker of jasmonate-dependent defence responses (Brown et al. 2003). In this study, it was slightly downregulated in both LNYV subgroup I and subgroup II infections compared to uninfected samples, by approximately 3% and 21% respectively. Expression was approximately 23% higher in subgroup I infections compared to subgroup II. Higher expression levels of *CPK3* in subgroup I compared to subgroup II would suggest that there would be increased activation of the downstream pathways that the protein activates, including the jasmonate-dependent defence responses, which may lead to increased suppression of LNYV being able to successfully establish an infection in a plant host. This would lead to lower rates of infection in subgroup I compared to subgroup II, which is corroborated within the varying infection rates identified earlier within this research, and also with the current prevalence of the virus within the environment found today, particularly in Australia.

SGS3 has a role in natural resistance. There was no difference in *SGS3* gene expression between subgroup I infections compared to uninfected samples. There was downregulation in subgroup II infections by approximately 62%, and expression was approximately 260% lower in subgroup II infections compared to subgroup I infections. This level of reduction may lead to decreased resistance to viral infection, and therefore mean subgroup II isolates are more efficient at infecting a plant host than subgroup I isolates. This possibly could be a mechanism that supports the hypothesised proposed by Higgins et al (2016) that subgroup II is more efficient at infecting hosts than subgroup I, and responsible for the current geographical spread and wider dispersal of the virus in the environment compared to subgroup I.

WRKY70 is a transcription factor at the junction of SA & JA dependent defence pathways and therefore has a role in natural resistance (Li et al. 2004). Results from previous research indicated that *WRKY70* may play a critical role in suppressing virus multiplication (Ando et al. 2013) and plants overexpressing *WRKY70* showed increased resistance to bacteria (Li et al. 2006). In the RT-qPCR experiments conducted within this research, *WRKY70* was upregulated by approximately 740% in subgroup I infections compared to uninfected samples, by approximately 16% in subgroup II infections, and therefore being expressed 636% higher in subgroup I infections compared to subgroup II. The significantly higher expression levels of

WRKY70 in subgroup I infections than both subgroup II and uninfected samples may indicate lower infection efficiency by these isolates, again supporting the hypothesis made by Higgins et al (2016) that subgroup II has the ability to infect more efficiently than subgroup I.

The difference observed between responses to each LNYV subgroup suggests each subgroup impacts *N. glutinosa* differently. However, this is a very limited study only focussing on one time point in the life cycle of the virus, and repeated studies would be necessary to confirm this as well as further examination of where in the host plant the proteins these genes code for operate to specifically link the gene expression patterns seen in this study to the virus itself, or whether they are just intermediate steps in a larger pathway of changes elicited by the virus.

3.5 Summary

Figure 3.37 below outlines the experiments described in this Chapter and a brief summary of the findings, as well as modifications to the experimental design that occurred throughout the course of the research.

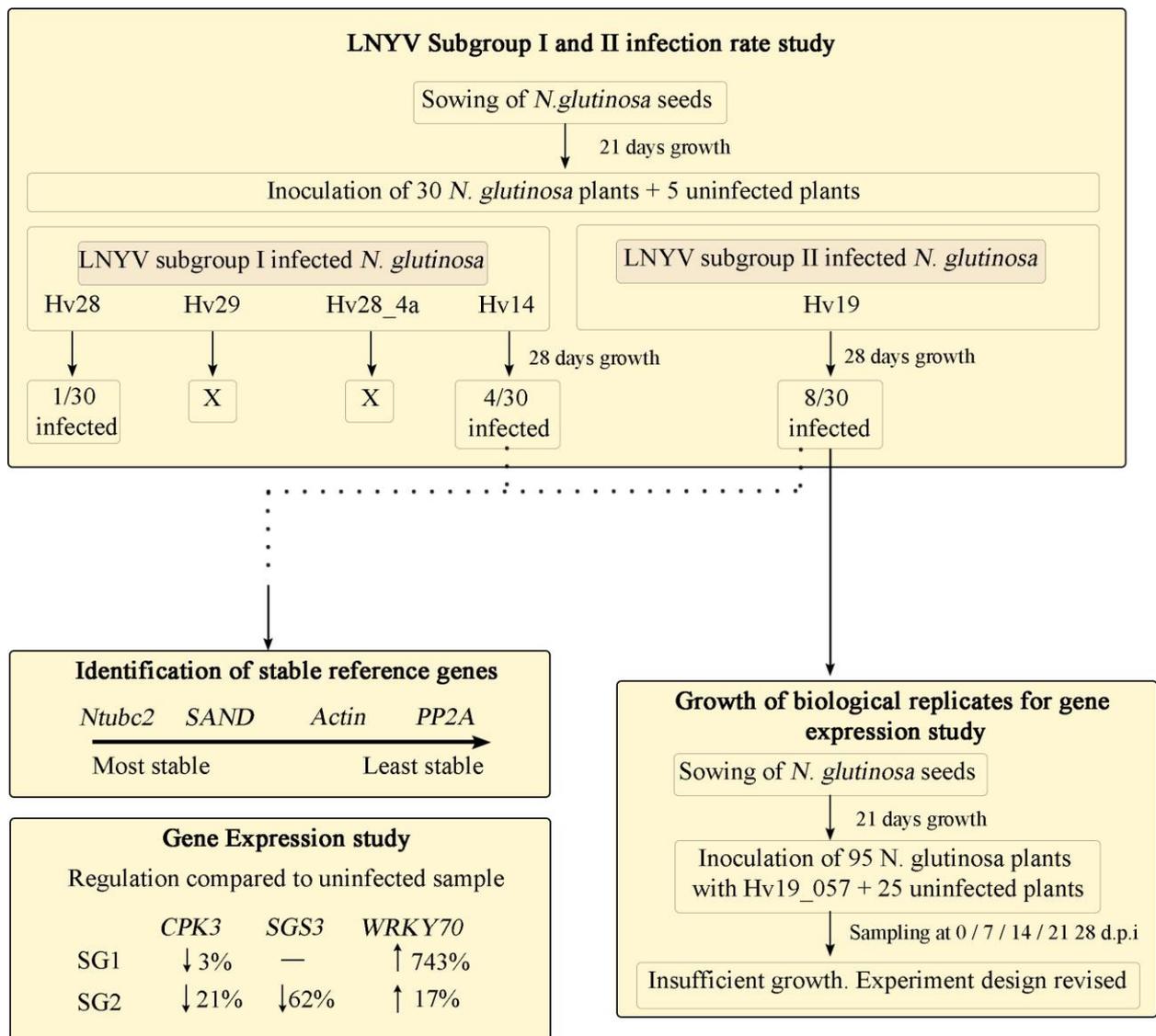


Figure 3.37: Flowchart of processes undertaken in this chapter, the samples grown and utilised and the outcomes of the gene expression studies. “X” denotes studies where no LNyV positive infections were obtained.

Chapter 4

General Discussion

4.1 Final Discussion

Lettuce necrotic yellows virus is the type species of the genus *Cytorhabdovirus* that principally infects lettuce crops in Australia and New Zealand, causing a range of symptoms in infected plants (Dietzgen et al. 2007). To date, there have been no studies published that examine the possible genetic basis for the host plant's response to infection by LNYV. Previous research identified two subgroups of the virus, subgroup I and subgroup II, based on differences in the N gene sequences of isolates of this virus (Callaghan and Dietzgen 2005). It is also unknown if the different subgroups cause different biological changes in a host plant upon infection. Identification of any difference in gene expression changes in response to infection between the subgroups may help explain the current distribution of the virus in the environment; the apparent extinction of subgroup I in Australia and the rapid radiation of subgroup II. The hypothesis suggested by Higgins et al (2016), that subgroup II isolates may have a more efficient means of establishing infection was utilised as a background framework for the study, and the experiments to determine if this could be valid were designed utilising *N. glutinosa* as a model host.

Previously it was shown to be difficult to mechanically infect *L. sativa* with LNYV (Dietzgen et al. 2007). For a gene expression study, the amount of infected leaf material and number of biological replicates necessary to obtain meaningful data meant that it was necessary to inoculate the virus into the model host plant, *N. glutinosa*. *N. glutinosa* can be utilised to discover changes in mRNA accumulation in response to the virus that then can be assessed in other hosts. Currently, no full genome is available for *N. glutinosa*, therefore, it was necessary to infer the likely structure of target genes by examining sequences of orthologous genes in related plant species for primer design.

Viruses are likely to have an impact on multiple, widespread pathways involving many changes in gene expression in a host plant when establishing an infection (Boualem et al. 2016; Caarls et al. 2015; Hernández et al. 2016). General sets of responses have been identified in the literature that occur in many plant hosts in response to viral infection (Senthil et al. 2005; Wang et al. 2015), broadly divided in cellular stress and developmental defects. Transcriptomic studies from microarray assays have previously been conducted on *Solanaceae* plant hosts to determine genes that are significantly up or down regulated in response to viral infection (Chen

et al. 2017; Senthil et al. 2005; Whitham et al. 2003). Whilst large amounts of data were obtained in these studies about differential expression of genes in different host plants, it is necessary to conduct specific virus-plant host studies to corroborate whether this would be supported when looking at LNYV infection in *N. glutinosa*. Due to technical availability, budget and timescale, it was decided to focus on a minimal set of target genes to study for this research. The GOIs that were identified as being likely to have significant, observable differences in gene expression in a plant virus infection based on studies into similar viruses in related plant hosts were *CPK3*, *SGS3*, *WRKY26* and *WRKY70*. These genes had also been identified as specifically having roles in plant defence mechanisms, either directly or indirectly, in previous research (Ando et al. 2013; Jiang et al. 2017; Li et al. 2004; Phukan et al. 2016; Senthil et al. 2005), and functional information had been published in gene ontology databases for them.

Currently, no full genome has been published for *N. glutinosa*, therefore limited sequence data exists, including that for the GOIs. It was necessary to obtain existing sequence data from related *Nicotiana* species to try and determine the likely intron/exon structure of the genes via a multiple sequence alignment in order to be able to design primers to amplify the gene, specifically around an intron/exon junction to avoid amplification of genomic DNA. Primers were then designed based on criteria for robust primer design from previously published literature. To date, this is the first published information regarding the possible sequence of these genes in *N. glutinosa*, and fully sequencing the plant's genome would corroborate if the alignments constructed in this research are valid.

From researching existing literature, no validated set of reference genes have been reported that can be utilised in a RT-qPCR study for *N. glutinosa*. Previous researchers have suggested that it is necessary to have a specific group of reference genes for the particular organism being studied (Baek et al. 2017; Kozera and Rapacz 2013; Liu et al. 2012). Related host plants that did have validated reference genes reported in the literature were studied and a candidate set of genes were identified that could possibly act as stable reference genes for *N. glutinosa*. These were *Actin*, *EF1 α* , *F-BOX*, *L23*, *Ntubc2*, *PP2A*, *PDF2* and *Ubiquitin*. It was necessary to also design primers utilising the same method as the genes of interest to amplify the genes and determine their stability across multiple experimental conditions.

Existing databases with genomic and mRNA data were searched for all of the aforementioned genes and as much available sequence data was obtained from related *Nicotiana* species to generate the multiple sequence alignments. Whilst this approach does not give a guarantee that the sequences entirely match the actual gene sequences in *N. glutinosa*, short of sequencing the entire genome, it was the most robust method available and a high degree of conservation in the coding regions between the different *Solanaceae* species was observed, therefore, the *N. glutinosa* sequence was likely to be similar. If problems with particular sequences were identified, the research into those genes did not continue. For example, it was not possible to design a primer pair to amplify *F-BOX*, *L23* and *Ubiquitin*; *F-BOX* had no identifiable intron, thus there was no way to be confident that genomic DNA would not be amplified, variation across the *Nicotiana* sequences in the *L23* multiple sequence alignment meant no site was identifiable for primers to bind to across all *Nicotiana* species. This meant the possibility of them binding to the unknown *N. glutinosa* sequences was diminished. No suitable primer pair could be designed for *Ubiquitin* that had characteristics suitable for use in a RT-qPCR assay, as designed primers had to share similar characteristics in order to run all of the samples and primers across similar cycling conditions. Additional sequence data or fully sequencing the *N. glutinosa* genome may allow for these genes to be examined and primers to be designed, and the genes could still be considered as candidate reference genes in future studies. Sequencing of the individual products of those genes that primers were designed for is also an option to possibly determine the gene structure of the reference genes and GOIs, though the products would be small.

The designed primers were tested on total RNA extracted from LNYV infected and uninfected *N. glutinosa* leaf material and, with the exception of *PDF2* and *WRKY26*, all amplified the correct sized product when tested in an end point one step RT-PCR (Sections 2.3.11 and 2.3.12). Each primer pair amplified one product, with the exception of the primer pair designed to amplify *EF1 α* . Therefore, it was decided to only test the four remaining candidate reference genes and three GOIs. Again, additional sequence data may enable the design of primer pairs that generate specific products and allow for *EF1 α* to be considered as a candidate reference gene, or for the changes in *WRKY26* between uninfected and infected plants to be assessed.

Previous studies had outlined methods by which to inoculate *N. glutinosa* plants from LNYV infected *L. sativa* plants (Dietzgen et al. 2007; Stubbs and Grogan 1963). Whilst these methods

were followed, large variance was seen in terms of the number of infections established, ranging from 0% to 26% across the different inoculation experiments. Multiple different sources of inoculum were used from different sources of *L. sativa*, suggesting that not all isolates are as infective as one another. Also, repeating the use of the same inoculum gave no guarantee that the same infection rate would be observed using a separate batch of plants. Further, the infection rate did not improve by using *N. glutinosa* leaf material freshly infected with LNYV as inoculum. With relatively low infection rates, particularly with subgroup I isolates, limitations in available space meant that it was difficult to grow sufficient leaf material to conduct the originally planned study of determining the gene expression patterns of the GOIs in uninfected and LNYV infected samples at six different time points. No observable symptoms were detected before 14 days. Whilst mechanical inoculation is a well-established method for inoculating LNYV into *N. glutinosa*, due to the amount of positively infected samples necessary for a RT-qPCR experiment with multiple time points, alternative means of inoculation may be of interest to future researchers to explore. Also, identifying a severe isolate for each subgroup would help with symptom identification. After insufficient LNYV infected *N. glutinosa* plant material was grown over the different time points, it was decided to utilise subgroup I and subgroup II infected material sampled 28 days post inoculation in order to conduct the gene expression assay, testing both the GOIs and reference genes together, despite usually determining stable reference genes in advance of running a gene expression study.

The RT-qPCR assays and analyses were carried out using the MIQE guidelines as a basis for the experimental design to try and attain a valid set of data. Each RT-qPCR plate was run with plate standards to allow between plate comparison. Normalisation was necessary to remove experimental variation, which cannot be completely removed from the process. After removal of outliers and normalising the data and subjecting the candidate reference genes to analysis using several different established pieces of software utilising different algorithms, it was determined that the most stable reference genes to use for comparison were *SAND* and *Ntubc2*, followed by *PP2A* and *Actin*. Whilst M values obtained for all of the genes were within acceptable limits, pairwise variation values suggested that additional reference genes are necessary for future RT-qPCR experiments. In addition to this, confidence levels using BestKeeper and GeNorm were above the 0.05 level meaning that further experimental work may be necessary to confirm these findings. However, for the purposes of this research, and as no other set of reference genes exist for *N. glutinosa*, the top two most stable reference genes, *Ntubc2* and *SAND* were utilised for relative quantification the expression of the three GOIs.

It was determined that *SGS3* and *CPK3* were downregulated by varying amounts in a LNYV subgroup I & II infection, and *WRKY70* was upregulated in response to both subgroups, though to a far higher degree in subgroup I. All three genes have been identified as having multifunctional roles, including roles in plant defence pathways as outlined in Section 1.5. Previously, none of these genes have been solely attributed to causing a particular host response when its expression level changed as part of a viral infection; they are one component of a larger molecular response pathway (Arimura and Sawasaki 2010; Jiang et al. 2017; Phukan et al. 2016; Senthil et al. 2005). However, previously published information has outlined that the GOI's are affiliated with particular processes, the JA signal pathway for example with *CPK3* and *WRKY70* (Caarls et al. 2015; Li et al. 2006; Li et al. 2004). The expression differences observed between subgroups suggest there may be variance in the extent to which they impact these pathways, and how these possible changes may explain the distribution of the two subgroups in the environment. Across the three GOIs, the differences in expression levels indicated that the subgroup I isolates may possibly cause defence mechanisms to be more active in a host plant than if a subgroup II isolate caused the infection. If extrapolated further, this would likely mean that subgroup II could be more successful at establishing infections in a plant host than subgroup I, an idea supported by the infection rate study within this research whereby successful subgroup II infections were more frequent than subgroup I infections. It is not clear from this research specifically how subgroup II isolates would have a selective advantage over subgroup I isolates, as only host responses were being analysed, but identifying that the virus may impact particular pathways will help future research by identifying specific components of the LNYV that may be capable of targeting or interacting with these pathways. Though limited, the findings in this research support the hypothesis that subgroup II may have more efficient infection methods compared to subgroup I, and may begin to explain the current distribution of the viral isolates found in Australia and New Zealand. Further analysis of these genes at different time points, or further analysis of alternate genes, perhaps those associated with similar molecular pathways as the GOIs in this research, may bring further elucidation to this theory. Also, as currently only a complete genome is available for subgroup I of LNYV, once a complete genome is available for subgroup II, they should be compared to study the differences that may impact virus-host interactions

A portion of this research also involved testing a previously published but not highly cited method for the cleaning of commercial nucleic acid columns in order to utilise them multiple

times in order to keep research costs down. The work here supported the original findings; however, the number of times a column could be reused was found to be once rather than the value of fifteen published in the paper (Nicosia et al. 2010). Over the course of this research, nucleic acid extraction columns were reused once with confidence that there was no leftover cellular material from previous experiments and the columns continued to work with high efficiency. Modifications to the method may allow for further uses if further research is undertaken.

It is important to consider that the gene expression changes observed in this research apply solely to the model plant *N. glutinosa*, and therefore may not be observable in *L. sativa*. However, due to the difficulty in mechanically infecting *L. sativa*, it may be difficult to replicate the same study utilising this important commercial host of the virus. Additional research could utilise other assays such as a microarray or RNA-Seq to confirm the findings in this experiment. Additional genes may also be identified that may help develop a fuller picture of gene regulatory networks that LNYV impacts upon to establish infection, and differences in these between the subgroups. By elucidating this information, it may be possible to determine how the genetic differences between the subgroups manifests, and provide greater understanding of the mechanisms of cytorhabdoviruses, and more broadly, rhabdoviruses.

Conclusion

In conclusion, this research has begun to formulate some of the components necessary to conduct future RT-qPCR experiments utilising *N. glutinosa* as a host plant. Two stable reference genes, *Ntubc2* and *SAND*, were identified as being suitable, however, further experimentation with more biological samples is necessary to confirm the findings in these experiments, and also additional reference genes need to be researched, have primers designed, tested and validated to have enough to adequately fulfil the MIQE guidelines. Ideally, an additional three reference genes would be identified and have primers designed for future RT-qPCR experiments.

In addition, several genes of interest have had their expression levels determined across uninfected and LNYV subgroup I and II infected conditions to determine, on a small scale, the possible pathways that LNYV may influence when it manages to establish infection in a host plant. *CPK3* and *SGS3* were downregulated and *WRKY70* was upregulated under LNYV

infected conditions compared to mock samples, and all to a greater degree in subgroup I infections compared to subgroup II infections. This suggests that subgroup I isolates may induce changes in the *N. glutinosa* that causes the establishment of infection to be more effectively prevented. These limited findings support the hypothesis that subgroup II isolates may have more efficient infection mechanisms than subgroup I isolates, though further work is necessary to corroborate this.

References

- Abrahamian PE, Seblani R, Sobh H, Abou-Jawdah Y. 2013. Detection and quantitation of two cucurbit criniviruses in mixed infection by real-time RT-PCR. *Journal of Virological Methods*. 193(2):320-326.
- Ahne W, Kurath G, Winton JR. 1998. A ribonuclease protection assay can distinguish spring viremia of carp virus from pike fry rhabdovirus. *Bulletin of the European Association of Fish Pathologists*. 18(6):220-224.
- Ajithkumar P. 2018. Diagnosis and genome analysis of lettuce necrotic yellows virus subgroups [Master of Science]. AUT.
- Alexander MM, Cilia M. 2016. A molecular tug-of-war: Global plant proteome changes during viral infection. *Current Plant Biology*. 5:13-24.
- Ammar E-D, Tsai C-W, Whitfield AE, Redinbaugh MG, Hogenhout SA. 2009. Cellular and molecular aspects of rhabdovirus interactions with insect and plant hosts. *Annual Review of Entomology*. 54(1):447-468.
- Anderson GJ, Cipolla C, Kennedy RT. 2011. Western blotting using capillary electrophoresis. *Analytical Chemistry*. 83(4):1350-1355.
- Ando S, Obinata A, Takahashi H. 2013. WRKY70 interacting with rcyl disease resistance protein is required for resistance to cucumber mosaic virus in *Arabidopsis thaliana*. *Physiological and Molecular Plant Pathology*. 85.
- Arimura G-i, Sawasaki T. 2010. Arabidopsis CPK3 plays extensive roles in various biological and environmental responses. *Plant Signaling & Behavior*. 5(10):1263-1265.
- Assenberg R, Delmas O, Morin B, Graham SC, De Lamballerie X, Laubert C, Coutard B, Grimes JM, Neyts J, Owens RJ et al. 2010. Genomics and structure/function studies of rhabdoviridae proteins involved in replication and transcription. *Antiviral Research*. 87(2):149-161.
- Atkinson NJ, Urwin PE. 2012. The interaction of plant biotic and abiotic stresses: From genes to the field. *Journal of Experimental Botany*. 63(10):3523-3543.
- Baek E, Yoon JY, Palukaitis P. 2017. Validation of reference genes for quantifying changes in gene expression in virus-infected tobacco. *Virology*. 510:29-39.
- Bao X, Roossinck MJ. 2013. A life history view of mutualistic viral symbioses: Quantity or quality for cooperation? *Current Opinion in Microbiology*. 16(4):514-518.
- Ben-Ami F, Mouton F, Ebert D. 2008. The effects of multiple infections on the expression and evolution of virulence in a daphnia-endoparasite system. *Evolution*. 62(7):1700-1711.
- Ben-Ami F, Routtu J. 2013. The expression and evolution of virulence in multiple infections - the role of specificity, relative virulence and relative dose. *Evolutionary Biology*. 13(97):1-11.

- Benito EP, Prins T, van Kan JA. 1996. Application of differential display RT-PCR to the analysis of gene expression in a plant-fungus interaction. *Plant Molecular Biology*. 32(5):947-957.
- Bivalkar-Mehla S, Vakharia J, Mehla R, Abreha M, Kanwar JR, Tikoo A, Chauhan A. 2011. Viral RNA silencing suppressors (rss): Novel strategy of viruses to ablate the host RNA interference (RNAi) defense system. *Virus Research*. 155(1):1-9.
- Bose J, Kloesener MH, Schulte RD. 2016. Multiple-genotype infections and their complex effect on virulence. *Zoology*. 119(4):339-349.
- Boualem A, Dogimont C, Bendahmane A. 2016. The battle for survival between viruses and their host plants. *Current Opinion in Virology*. 17:32-38.
- Brault V, Uzest M, Monsion B, Jacquot E, Blanc S. 2010. Aphids as transport devices for plant viruses. *Comptes Rendus Biologies*. 333(6-7):524-538.
- Broeders S, Huber I, Gorohmann L, Berben G, Taverniers I, Mazzara M, Roosens N, Morisset D. 2014. Guidelines for the validation of qualitative real-time PCR methods. *Trends in Food Science & Technology*. 37:115-126.
- Brown RL, Kazan K, McGrath KC, Maclean DJ, Manners JM. 2003. A role for the GCC-box in jasmonate-mediated activation of the *pdf1.2* gene of *Arabidopsis*. *Plant Physiology*. 132(2):1020-1032.
- Bubici G, Carluccio AV, Cillo F, Stabolone L. 2014. Virus-induced gene silencing of pectin methylesterase protects *Nicotiana benthamiana* from lethal symptoms caused by tobacco mosaic virus. *European Journal of Plant Pathology*. 141(2):339-347.
- Bustin S, Huggett J. 2017. qPCR primer design revisited. *Biomolecular Detection and Quantification*. 14:19-28.
- Bustin SA, Benes V, Garson J, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley G et al. 2013. The need for transparency and good practices in the qPCR literature. *Nature Methods*. 2013/11/01 ed. p. 1063-1067.
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL et al. 2009. The MIQE guidelines: Minimum information for publication of quantitative real-time pcr experiments. *Clinical Chemistry*. 55(4):611-622.
- Caarls L, Pieterse CM, Van Wees SC. 2015. How salicylic acid takes transcriptional control over jasmonic acid signaling. *Frontiers in Plant Science*. 6:170.
- Callaghan B, Dietzgen RG. 2005. Nucleocapsid gene variability reveals two subgroups of lettuce necrotic yellows virus. *Archives of Virology*. 150(8):1661-1667.
- Camacho Londoño J, Philipp SE. 2016. A reliable method for quantification of splice variants using RT-qPCR. *BMC Molecular Biology*. 17:8-8.

- Campbell PR. 2003. Genetic modification of lettuce for resistance to lettuce necrotic yellows virus [PhD Thesis]. The University of Queensland.
- Chambers TC, Crowley NC, Francki RIB. 1965. Localization of lettuce necrotic yellows virus in host leaf tissue. *Virology*. 27:320-328.
- Chavez-Calvillo G, Contreras-Paredes CA, Mora-Macias J, Noa-Carrazana JC, Serrano-Rubio AA, Dinkova TD, Carrillo-Tripp M, Silva-Rosales L. 2016. Antagonism or synergism between papaya ringspot virus and papaya mosaic virus in *Carica papaya* is determined by their order of infection. *Virology*. 489:179-191.
- Chen L, Fei C, Zhu L, Xu Z, Zou W, Yang T, Lin H, Xi D. 2017. RNA-seq approach to analysis of gene expression profiles in dark green islands and light green tissues of cucumber mosaic virus-infected *Nicotiana tabacum*. *PLoS One*. 12(5):e0175391.
- Chen Z, Tan JY, Wen Y, Niu S, Wong SM. 2012. A game-theoretic model of interactions between Hibiscus latent Singapore virus and tobacco mosaic virus. *PLoS One*. 7(5):e37007.
- Chu PWG, Francki RIB. 1982. Detection of lettuce necrotic yellows virus by an enzyme-linked immunosorbent assay in plant hosts and the insect vector. *Annals of Applied Biology*. 100:149-156.
- Coutts BA, Thomas-Carroll ML, Jones RAC. 2004. Analysing spatial patterns of spread of lettuce necrotic yellows virus and lettuce big-vein disease in lettuce field plantings. *Annals of Applied Biology*. 145:339-343.
- Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K. 2011. Effects of abiotic stress on plants: A systems biology perspective. *BMC Plant Virology*. 11(163):1-14.
- Crowley NC. 1967. Factors affecting the local lesion response of *Nicotiana glutinosa* to lettuce necrotic yellows virus. *Virus Res*. 31:107-113.
- D'Ippolito S, Vankova R, Joosten MH, Casalongue CA, Fiol DF. 2016. Knocking down expression of the auxin-amidohydrolase IAR3 alters defense responses in Solanaceae family plants. *Plant Science*. 253:31-39.
- Dacheux L, Berthet N, Dissard G, Holmes EC, Delmas O, Larrous F, Guigon G, Dickinson P, Faye O, Sall AA et al. 2010. Application of broad-spectrum resequencing microarray for genotyping rhabdoviruses. *Journal of Virology*. 84(18):9557-9574.
- Die JV, Roman B. 2012. RNA quality assessment: A view from plant qPCR studies. *Journal of Experimental Botany*. 63(17):6069-6077.
- Dietzgen RG, Callaghan B, Campbell PR. 2007. Biology and genomics of lettuce necrotic yellows virus. *Plant Viruses*. 1(1):85-92.
- Dietzgen RG, Callaghan B, Wetzel T, Dale JL. 2006. Completion of the genome sequence of lettuce necrotic yellows virus, type species of the genus *Cytorhabdovirus*. *Virus Research*. 118(1-2):16-22.

- Dietzgen RG, Francki RIB. 1988. Analysis of lettuce necrotic yellows virus structural proteins with monoclonal antibodies and concanavalin A. *Virology*. 166:486-494.
- Dietzgen RG, Hunter BG, Francki RIB, Jackson AO. 1989. Cloning of lettuce necrotic yellows virus RNA and identification of virus-specific polyadenylated RNAs in infected *Nicotiana glutinosa* leaves. *Journal of General Virology*. 70:2299-2307.
- Dietzgen RG, Kondo H, Goodin MM, Kurath G, Vasilakis N. 2017. The family rhabdoviridae: Mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research*. 227:158-170.
- Dwight Z, Palais R, Wittwer CT. 2011. uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. *Bioinformatics*. 27(7):1019-1020.
- Edwards KD, Bombarely A, Story GW, Allen F, Mueller LA, Coates SA, Jones L. 2010. Toba: An atlas of tobacco gene expression from seed to senescence. *BMC Genomics*. 11(1):142.
- Edwards KD, Fernandez-Pozo N, Drake-Stowe K, Humphry M, Evans AD, Bombarely A, Allen F, Hurst R, White B, Kernodle SP et al. 2017. A reference genome for *Nicotiana tabacum* enables map-based cloning of homologous loci implicated in nitrogen utilization efficiency. *BMC Genomics*. 18(1):448.
- El-Wahab A. 2012. Transmission efficiency of lettuce mosaic virus (LMV) by different aphid species and new aphid vectors in Egypt. *Academic Journal of Entomology*. 5(3):158-163.
- Elena SF, Bedhomme S, Carrasco P, Cuevas JM, de la Iglesia F, Lafforgue G, Lalic J, Prosper A, Tromas N, Zwart MP. 2011. The evolutionary genetics of emerging plant RNA viruses. *The American Phytopathological Society*. 24(3):287-293.
- Elena SF, Bernet GP, Carrasco JL. 2014. The games plant viruses play. *Current Opinion in Virology*. 8:62-67.
- Fletcher J, Walker M, Davidson H, Paull S, Palmer A. 2017. Outdoor lettuce virus disease project 2016-2018 year 1. Plant and Food Research. [Accessed 2018 May 01] <https://www.freshvegetables.co.nz/assets/Lettuce-virus-disease-project-Report-John-Fletcher-PFR-2016-2018-Year-1-FINAL.pdf>
- Fletcher JD, France CM, Butler RC. 2005. Virus surveys of lettuce crops and management of lettuce big-vein disease in New Zealand. *New Zealand Plant Protection*. 58:239-244. [Access 2018 July 11] <https://www.freshvegetables.co.nz/assets/Uploads/16208-John-Fletcher-Outdoor-lettuce-virus-disease-project-year-2-FINAL.pdf>
- Fletcher JD, Zhang Y, Kean AM, Davidson MM. 2018. Outdoor lettuce virus disease project, year 2 report, 2018. Plant and Food Research.

- Flint SJ, Enquist LW, Krug RM, Racaniello VR, Skalka AM. 2015. Principles of Virology: Molecular biology, pathogenesis and control. Herndon: ASM Press.
- Fregene M, Matsumura H, Akano A, Dixon A, Terauchi R. 2004. Serial analysis of gene expression (SAGE) of host-plant resistance to the cassava mosaic disease (CMD). *Plant Molecular Biology*. 56(4):563-571.
- Fry PR, Close RC, Procter CH, Sunde R. 1973. Lettuce necrotic yellows virus in New Zealand. *New Zealand Journal of Agricultural Research*. 16(1):143-146.
- Gao R, Liu P, Wong S-M. 2012. Identification of a plant viral RNA genome in the nucleus. *PLOS ONE*. 7(11).
- Garbutt J, Bonsall MB, Wright DJ, Raymond B. 2011. Antagonistic competition moderates virulence in *Bacillus thuringiensis*. *Ecology Letters*. 14(8):765-772.
- Garrett KA, Nita M, De Wolf ED, Esker PD, Gomez-Montano L, Sparks AH. 2016. Chapter 21 - plant pathogens as indicators of climate change. In: Letcher TM, editor. *Climate Change* (second edition). Boston: Elsevier. p. 325-338.
- Ghorbani A, Izadpanah K, Dietzgen RG. 2018. Gene expression and population polymorphism of maize Iranian mosaic virus in *Zea mays*, and intracellular localization and interactions of viral N, P, and M proteins in *Nicotiana benthamiana*. *Virus Genes*. 54(2):290-296.
- Ghosh R, Gilda JE, Gomes AV. 2014. The necessity of and strategies for improving confidence in the accuracy of western blots. *Expert Review of Proteomics*. 11(5):549-560.
- Goralski M, Sobieszczanska P, Obrepalska-Stepulowska A, Swiercz A, Zmienko A, Figlerowicz M. 2016. A gene expression microarray for *Nicotiana benthamiana* based on de novo transcriptome sequence assembly. *Plant Methods*. 12:28.
- Groen AK. 2001. The pros and cons of gene expression analysis by microarrays. *Journal of Hepatology*. 35:295-296.
- Haas BJ, Zody MC. 2010. Advancing RNA-seq analysis. *Nature Biotechnology*. 28:421.
- Harkin DP, Hall PA. 2000. Measuring a cell's response to stress: The p53 pathway. *Genome Biology*. 1(1):1-5
- Harrison BD, Crowley NC. 1965. Properties and structure of lettuce necrotic yellows virus. *Virology*. 26:297-310.
- Hashimoto Y, Macri D, Srivastava I, McPherson C, Felberbaum R, Post P, Cox M. 2017. Complete study demonstrating the absence of rhabdovirus in a distinct SF9 cell line. *PLoS ONE*. 12(4):e0175633.
- Havelda Z, Vrallyay v, Vlczi A, Burgyn J. 2008. Plant virus infection-induced persistent host gene down regulation in systemically infected leaves. *The Plant Journal*. 55(2):278-288.

- Hernández JA, Gullner G, Clemente-Moreno MJ, Künstler A, Juhász C, Díaz-Vivancos P, Király L. 2016. Oxidative stress and antioxidative responses in plant–virus interactions. *Physiological and Molecular Plant Pathology*. 94:134-148.
- Higgins CM, Chang WL, Khan S, Tang J, Elliott C, Dietzgen RG. 2016. Diversity and evolutionary history of lettuce necrotic yellows virus in Australia and New Zealand. *Archives of Virology*. 161(2):269-277.
- Hily JM, Poulicard N, Mora MA, Pagan I, Garcia-Arenal F. 2016. Environment and host genotype determine the outcome of a plant-virus interaction: From antagonism to mutualism. *New Phytologist*. 209(2):812-822.
- Horticulture New Zealand. 2017. New Zealand domestic vegetable production: The growing story. Horticulture New Zealand.
- Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, Cao S, Liu Y. 2012. Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Molecular Genetics and Genomics*. 287(6):495-513.
- Huber D, Voith von Voithenberg L, Kaigala GV. 2018. Fluorescence *in situ* hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro and Nano Engineering*. 1:15-24.
- Hull R. *Plant Virology (Fifth Edition)*. Boston: Academic press, 2014. [accessed 2018 Jul 02 <http://www.sciencedirect.com/science/book/9780123848710>]
- International Committee on Taxonomy Virus: Master species list 2017. 2018. [accessed 2018 Nov 18]. <https://talk.ictvonline.org/files/master-species-lists/m/msl/7185>.
- Incarbone M, Dunoyer P. 2013. RNA silencing and its suppression: Novel insights from in planta analyses. *Trends in Plant Science*. 18(7):382-392.
- Irwin JAG, Jackson KJ. 1977. Safflower diseases in Queensland. *Queensland Agricultural Journal*. 103(6):516-520.
- Jackson AO, Dietzgen RG, Goodin MM, Bragg JN, Deng M. 2005. Biology of plant rhabdoviruses. *Annual Review of Phytopathology*. 43:623-660.
- Jaksik R, Iwanaszko M, Rzeszowska-Wolny J, Kimmel M. 2015. Microarray experiments and factors which affect their reliability. *Biology Direct*. 10:46.
- Jensen EC. 2012. The basics of western blotting. *The Anatomical Record*. 295(3):369-371.
- Jeong J-J, Ju H-J, Noh J. 2014. A review of detection methods for the plant viruses. *Research in Plant Disease*. 20(3):173-181.
- Jiang J, Ma S, Ye N, Jiang M, Cao J, Zhang J. 2017. WRKY transcription factors in plant responses to stresses. *Journal of Integrated Plant Biology*. 59(2):86-101.

- Jones RAC. 2006. Control of plant virus diseases. *Advances in Virus Research*. Academic Press. p. 205-244. [Accessed 2018 May 14]
https://www.researchgate.net/publication/6766750_Control_of_Plant_Virus_Diseases
- Jones RAC. 2012. Influence of climate change on plant disease infections and epidemics caused by viruses and bacteria. *CAB Reviews*. 7(22):1-31.
- Jones RAC. 2014. Plant virus ecology and epidemiology: Historical perspectives, recent progress and future prospects. *Annals of Applied Biology*. 164(3):320-347.
- Jovel J, Walker M, Sanfaçon H. 2007. Recovery of *Nicotiana benthamiana* plants from a necrotic response induced by a nepovirus is associated with RNA silencing but not with reduced virus titer. *Journal of Virology*. 81(22):12285-12297.
- Kamitani M, Nagano AJ, Honjo MN, Kudoh H. 2016. RNA-Seq reveals virus-virus and virus-plant interactions in nature. *FEMS Microbiology Ecology*. 92(11):1-11.
- Kaur N, Hasegawa DK, Ling KS, Wintermantel WM. 2016. Application of genomics for understanding plant virus-insect vector interactions and insect vector control. *Phytopathology*. 106(10):1213-1222.
- Kesanakurti P, Belton M, Saeed H, Rast H, Boyes I, Rott M. 2016. Screening for plant viruses by next generation sequencing using a modified double strand rna extraction protocol with an internal amplification control. *Journal of Virological Methods*. 236:35-40.
- Khraiweh B, Zhu JK, Zhu J. 2012. Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochim Biophys Acta*. 1819(2):137-148.
- Kliot A, Kontsedalov S, Lebedev G, Brumin M, Cathrin PB, Marubayashi JM, Skaljac M, Belausov E, Czosnek H, Ghanim M. 2014. Fluorescence *in situ* hybridizations (FISH) for the localization of viruses and endosymbiotic bacteria in plant and insect tissues. *Journal of Visualized Experiments (JoVE)*. (84):51030.
- Koonin EV. 2010. The wonder world of microbial viruses. *Expert Review of Anti-infective Therapy*. 8(10):1097-1099.
- Kormelink R, Garcia ML, Goodin M, Sasaya T, Haenni AL. 2011. Negative-strand RNA viruses: The plant-infecting counterparts. *Virus Research*. 162(1-2):184-202.
- Kozera B, Rapacz M. 2013. Reference genes in real-time PCR. *Journal of Applied Genetics*. 54(4):391-406.
- Lawrence SD, Novak NG, Ju CJ-T, Cooke JEK. 2008. Potato, *Solanum tuberosum*, defense against Colorado potato beetle, *Leptinotarsa decemlineata*: Microarray gene expression profiling of potato by Colorado potato beetle regurgitant treatment of wounded leaves. *Journal of Chemical Ecology*. 34(8):1013-1025.
- Lebeda A, Křístková E, Kitner M, Mieslerová B, Jemelková M, Pink DAC. 2013. Wild *Lactuca* species, their genetic diversity, resistance to diseases and pests, and exploitation in lettuce breeding. *European Journal of Plant Pathology*. 138(3):597-640.

- Li J, Brader G, Kariola T, Palva ET. 2006. WRKY70 modulates the selection of signaling pathways in plant defense. *The Plant Journal*. 46(3):477-491.
- Li J, Brader G, Palva ET. 2004. The WRKY70 transcription factor: A node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *The Plant Cell*. 16(2):319-331.
- Liu Y, Zhenzhen D, Wang H, Zhang S, Cao M, Wang X. 2018. Identification and characterization of Wheat Yellow Striate Virus, a novel leafhopper-transmitted nucleorhabdovirus infecting wheat. *Frontiers of Microbiology*. 9: 468
- Lilly ST, Drummond RSM, Pearson MN, MacDiarmid RM. 2011. Identification and validation of reference genes for normalization of transcripts from virus-infected *Arabidopsis thaliana*. *The American Phytopathological Society*. 24(3):294-304.
- Ling D, Salvaterra PM. 2011. Robust RT-qPCR data normalization: Validation and selection of internal reference genes during post-experimental data analysis. *PloS one*. 6(3):e17762-e17762.
- Liu D, Shi L, Han C, Yu J, Li D, Zhang Y. 2012. Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time pcr. *PLoS One*. 7(9):e46451.
- Liu W, Zhao X, Zhang P, Mar TT, Liu Y, Zhang Z, Han C, Wang X. 2013. A one step real-time RT-PCR assay for the quantitation of wheat yellow mosaic virus (WYMV). *Virology Journal*. 10:173-173.
- Liu Y, Du Z, Wang H, Zhang S, Cao M, Wang X. 2018. Identification and characterization of wheat yellow striate virus, a novel leafhopper-transmitted nucleorhabdovirus infecting wheat. *Frontiers in Microbiology*. 9(468).
- Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2CT method. *Methods*. 25:402-408.
- Lu J, Du ZX, Kong J, Chen LN, Qiu YH, Li GF, Meng XH, Zhu SF. 2012. Transcriptome analysis of *Nicotiana tabacum* infected by cucumber mosaic virus during systemic symptom development. *PLoS One*. 7(8):e43447.
- Ma X, Nicole M-C, Metegnier L-V, Hong N, Wang G, Moffett P. 2014. Different roles for RNA silencing and RNA processing components in virus recovery and virus-induced gene silencing in plants. *Journal of Experimental Botany*. 66(3):919-932.
- Mahmood T, Yang P-C. 2012. Western blot: Technique, theory, and trouble shooting. *North American Journal of Medical Sciences*. 4(9):429-434.
- Maneechoat P, Takeshita M, Uenoyama M, Nakatsukasa M, Kuroda A, Furuya N, Tsuchiya K. 2015. A single amino acid at N-terminal region of the 2b protein of cucumber mosaic virus strain M1 has a pivotal role in virus attenuation. *Virus Research*. 197:67-74.
- Mann KS, Bejerman N, Johnson KN, Dietzgen RG. 2016a. Cytorhabdovirus p3 genes encode 30k-like cell-to-cell movement proteins. *Virology*. 489:20-33.

- Mann KS, Dietzgen RG. 2014. Plant rhabdoviruses: New insights and research needs in the interplay of negative-strand RNA viruses with plant and insect hosts. *Archives of Virology*. 159(8):1889-1900.
- Mann KS, Dietzgen RG. 2017. Functional analysis of a weak viral RNA silencing suppressor using two GFP variants as silencing inducers. *Journal of Virological Methods*. 239:50-57.
- Mann KS, Johnson KN, Carroll BJ, Dietzgen RG. 2016b. Cytorhabdovirus P protein suppresses RISC-mediated cleavage and RNA silencing amplification *in planta*. *Virology*. 490:27-40.
- Mann KS, Johnson KN, Dietzgen RG. 2015. Cytorhabdovirus phosphoprotein shows RNA silencing suppressor activity in plants, but not in insect cells. *Virology*. 476:413-418.
- Martin DK. 1983. Studies on the main weed host and aphid vector of lettuce necrotic yellows virus in South Australia. Oxford: Blackwell Scientific Publications. p. 211-219.
- Martin KM, Dietzgen RG, Wang R, Goodin MM. 2012. Lettuce necrotic yellows cytorhabdovirus protein localization and interaction map, and comparison with nucleorhabdoviruses. *Journal of General Virology*. 93(4):906-914.
- Martinelli F, Scalenghe R, Davino S, Panno S, Scuderi G, Ruisi P, Villa P, Stroppiana D, Boschetti M, Goulart LR et al. 2014. Advanced methods of plant disease detection. A review. *Agronomy for Sustainable Development*. 35(1):1-25.
- McLean GD, Wolanski BS, Francki RIB. 1971. Serological analysis of lettuce necrotic yellows virus preparations by immunodiffusion. *Virology*. 42:480-487.
- Miura T, Scharf M. 2010. Molecular basis underlying caste differentiation in termites. p. 211-253.
- Moore S, Payton P, Wright M, Tanksley S, Giovannoni J. 2005. Utilization of tomato microarrays for comparative gene expression analysis in the Solanaceae. *Journal of Experimental Botany*. 56(421):2885-2895.
- Moustafa K, Cross JM. 2016. Genetic approaches to study plant responses to environmental stresses: An overview. *Biology*. 5(2).
- Muthamilarasan M, Prasad M. 2013. Plant innate immunity: An updated insight into defense mechanism. *Journal of Biosciences*. 38(2):433-449.
- Nicosia A, Tagliavia M, Costa S. 2010. Regeneration of total RNA purification silica-based columns. *Biomedical Chromatography*. 24(12):1263-1264.
- Noris E, Miozzi L. 2015. Real-time PCR protocols for the quantification of the Begomovirus tomato yellow leaf curl Sardinia virus in tomato plants and in its insect vector. *Methods in Molecular Biology*. 1236:61-72.

- Ozsolak F, Milos PM. 2011. RNA sequencing: Advances, challenges and opportunities. *Nature Reviews Genetics*. 12(2):87-98.
- Pabinger S, Rodiger S, Kriegner A, Vierlinger K, Weinhausel A. 2014. A survey of tools for the analysis of quantitative PCR (qPCR) data. *Biomolecular Detection and Quantification*. 1(1):23-33.
- Paudel DB, Ghoshal B, Jossey S, Ludman M, Fatyol K, Sanfaçon H. 2018. Expression and antiviral function of Argonaute 2 in *Nicotiana benthamiana* plants infected with two isolates of tomato ringspot virus with varying degrees of virulence. 524:127-139
- Pecman A, Kutnjak D, Gutiérrez-Aguirre I, Adams I, Fox A, Boonham N, Ravnikar M. 2017. Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Frontiers in Microbiology*. 8:1998.
- Perez-Canamas M, Blanco-Perez M, Forment J, Hernandez C. 2017. *Nicotiana benthamiana* plants asymptotically infected by Pelargonium line pattern virus show unusually high accumulation of viral small RNAs that is neither associated with DCL induction nor RDR6 activity. *Virology*. 501:136-146.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*. 29(9):40-46.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: Bestkeeper – excel-based tool using pair-wise correlations. *Biotechnology Letters*. 26(6):509-515.
- Qu Y, Boutjdir M. 2007. Rnase protection assay for quantifying gene expression levels. In: Zhang J, Rokosh G, editors. *Cardiac Gene Expression: Methods and Protocols*. Totowa, NJ: Humana Press. p. 145-158.
- Ragozzino A, Alioto D, Iengo C, Iego C. 1989. The yellowing virus and mycoplasma diseases of lettuce in Campania and Latium regions. *Rivista di Patologia Vegetale*. 25(1):15-19.
- Randles JW, Carver M. 1971. Epidemiology of lettuce necrotic yellows virus in South Australia. II. Distribution of virus, host plants, and vectors. *Australian Journal of Agricultural Research*. 22(2).
- Randles JW, Coleman DF. 1970. Loss of ribosomes in *Nicotiana glutinosa* L. infected with lettuce necrotic yellows virus. *Virology*. 41:459-464.
- Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. 2011. Improving RNA-seq expression estimates by correcting for fragment bias. *Genome Biology*. 12(3):R22.
- Rodríguez A, Rodríguez M, Córdoba JJ, Andrade MJ. 2015. Design of primers and probes for quantitative real-time PCR methods. In: Basu C, editor. *PCR primer design*. New York, NY: Springer New York. p. 31-56.
- Ronholm J, Naseri N, Petronella N, Pagotto F. 2016. Navigating microbiological food safety in the era of whole-genome sequencing. *Clinical Microbiology Reviews*. 29:109-116.

- Roossinck MJ. 2015. Plants, viruses and the environment - ecology and mutualism. *Virology*. 479-480:271-277.
- Roossinck MJ. 2017. Deep sequencing for discovery and evolutionary analysis of plant viruses. *Virus Research*. 239:82-86.
- Rubio-Huertos M, Garcia-Hidalgo F. 1982. A rhabdovirus resembling lettuce necrotic yellows from lettuce in Spain. *Phytopathology*. 103:232-238.
- Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, Moorman AFM. 2009. Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Research*. 37(6):e45-e45.
- Ruiz-Villalba A, van Pelt-Verkuil E, Gunst QD, Ruijter JM, van den Hoff MJ. 2017. Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR). *Biomolecular Detection and Quantification*. 14:7-18.
- Salvaudon L, De Moraes CM, Mescher MC. 2013. Outcomes of co-infection by two potyviruses: Implications for the evolution of manipulative strategies. *Proceedings of the Royal Society B: Biological Sciences*. 280(1756):1-9.
- Sanfacon H. 2017. Grand challenge in plant virology: Understanding the impact of plant viruses in model plants, in agricultural crops, and in complex ecosystems. *Frontiers in Microbiology*. 8:860.
- Schmidt GW, Delaney SK. 2010. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Molecular Genetics and Genomics*. 283(3):233-241.
- Senthil G, Liu H, Puram VG, Clark A, Stromberg A, Goodin MM. 2005. Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses. *Journal of General Virology*. 86(Pt 9):2615-2625.
- Shargil D, Zemach H, Belausov E, Lachman O, Kamenetsky R, Dombrovsky A. 2015. Development of a fluorescent *in situ* hybridization (FISH) technique for visualizing CGMMV in plant tissues. *Journal of Virological Methods*. 223:55-60.
- St-Pierre J, Grégoire J-C, Vaillancourt C. 2017. A simple method to assess group difference in RT-qPCR reference gene selection using GeNorm: The case of the placental sex. *Scientific Reports*. 7(1):16923-16923.
- Stubbs L, Grogan R. 1963. Necrotic yellows: A newly recognized virus disease of lettuce. *Australian Journal of Agricultural Research*. 14(4):439-459.
- Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. 2015. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*. 3:9-16.
- Sward RJ. 1990. Lettuce necrotic yellows rhabdovirus and other viruses infecting garlic. *Australasian Plant Pathology*. 19(2):46-51.

- Syller J. 2012. Facilitative and antagonistic interactions between plant viruses in mixed infections. *Molecular Plant Pathology*. 13(2):204-216.
- Syller J, Grupa A. 2016. Antagonistic within-host interactions between plant viruses: Molecular basis and impact on viral and host fitness. *Molecular Plant Pathology*. 17(5):769-782.
- Tarasov KV, Brugh SA, Tarasova YS, Boheler KR. 2007. Serial analysis of gene expression (SAGE): A useful tool to analyze the cardiac transcriptome. *Methods in Molecular Biology*. 366:41-59.
- Taylor S, Wakem M, Dijkman G, Alsarnaj M, Nguyen M. 2009. A practical approach to RT-qPCR - publishing data that conform to the MIQE guidelines. *Methods*. 50(4):1-5.
- Tollenaere C, Susi H, Laine AL. 2016. Evolutionary and epidemiological implications of multiple infection in plants. *Trends in Plant Science*. 21(1):80-90.
- Toriyama S, Peters D. 1981. Differentiation between broccoli necrotic yellows virus and lettuce necrotic yellows virus by their transcriptase activities. *Journal of General Virology*. 56(59-66).
- Tsai C-W, Redinbaugh MG, Willie KJ, Reed S, Goodin M, Hogenhout SA. 2005. Complete genome sequence and in planta subcellular localization of maize fine streak virus proteins. *Journal of virology*. 79(9):5304-5314.
- Valmonte GR. 2016. Calcium-dependent protein kinases: Understanding functional diversification and specificity of plant signalling hubs using the most conserved members [PhD]. Auckland University of Technology.
- Vandesompele J, Kubista M, Pfaffl MW. 2009. Reference gene validation software for improved normalization. *Real-time PCR: Current Technology and Applications*. p. 1-22.
- Varma A. 1993. Integrated management of plant viral diseases. *Ciba Foundation Symposium*. 177:140-155. [Accessed 2018 April 19]. <https://www.ncbi.nlm.nih.gov/pubmed/8149818>
- Vorwerk S, Sonntag K, Blaich R, Forneck A. 2008. Application of current *in situ* hybridization techniques for grape phylloxera. *Journal of Grapevine Research*. 47(2):113-116.
- Wacker MJ, Godard MP. 2005. Analysis of one-step and two-step real-time RT-PCR using Superscript III. *Journal of Biomolecular Techniques*. 16:266-271.
- Walker PJ, Blasdell KR, Calisher CH, Dietzgen RG, Kondo H, Kurath G, Longdon B, Stone DM, Tesh RB, Tordo N et al. 2018. Rhabdoviridae. *Journal of General Virology*. 99:447-447.

- Walker PJ, Dietzgen RG, Joubert DA, Blasdell KR. 2011. Rhabdovirus accessory genes. *Virus Research*. 162(1-2):110-125.
- Wang H, Wang J, Xie Y, Fu Z, Wei T, Zhang X-F. 2018. Development of leafhopper cell culture to trace the early infection process of a nucleorhabdovirus, rice yellow stunt virus, in insect vector cells. *Virology Journal*. 15(1):72.
- Wang K-D, Empleo R, Nguyen TTV, Moffett P, Sacco MA. 2015. Elicitation of hypersensitive responses in *Nicotiana glutinosa* by the suppressor of RNA silencing protein p0 from poleroviruses. *Molecular Plant Pathology*. 16(5):435-448.
- Wang Z, Gerstein M, Snyder M. 2009. RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*. 10(1):57-63.
- Wetzel T, Dietzgen RG, Geering ADW, Dale JL. 1994. Analysis of the nucleocapsid gene of lettuce necrotic yellows rhabdovirus. *Virology*. 202:1054-1057.
- White EJ, Venter M, Hiten NF, Burger JT. 2008. Modified cetyltrimethylammonium bromide method improves robustness and versatility: The benchmark for plant RNA extraction. *Journal of Biotechnology*. 3(11):1424-1428.
- Whitham SA, Quan S, Chang H, Cooper B, Estes B, Zhu T, Wang X, Hou YM. 2003. Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *The Plant Journal*. 33(2):271-283.
- Whitham SA, Yang C, Goodin MM. 2006. Global impact: Elucidating plant responses to viral infection. *The American Phytopathological Society*. 19(11):1207-1215.
- Wieczorek P, Obrepalska-Stęplowska A. 2013. Multiplex RT-PCR reaction for simultaneous detection of tomato torrado virus and pepino mosaic virus co-infecting *Solanum lycopersicum*. *Journal of Plant Protection Research*. 53(3):289-294.
- Willie K, and Stewart LR. 2017. Complete Genome Sequence of a New Maize-Associated Cytorhabdovirus. *Genome Announcements*. 5(31):e00591-17.
- Wu G, Zheng G, Hu Q, Ma M, Li M, Sun X, Yan F, Qing L. 2018. NS3 protein from rice stripe virus affects the expression of endogenous genes in *Nicotiana benthamiana*. *Virology*. 15(1):105.
- Yamamoto M, Wakatsuki T, Hada A, Ryo A. 2001. Use of serial analysis of gene expression (SAGE) technology. *Journal of Immunological Methods*. 250(1-2):45-66.
- Yang X, Huang J, Liu C, Chen B, Zhang T, Zhou G. 2017. Rice stripe mosaic virus, a novel cytorhabdovirus infecting rice via leafhopper transmission. *Frontiers in Microbiology*. 7(2140).