

**Towards the analysis of Potyvirus VPg
Interacting Protein (PVIP) gene expression
in response to potyvirus infection**

Smriti Nair

A thesis submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the degree of Master of
Applied Science (MAppSc)

2012

School of Applied Science

Supervisor: Dr. Colleen Higgins

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.

(Smriti Nair)

Table of Contents

List of Figures

List of Tables

Acknowledgements

Abstract

Chapter 1

Introduction

1.1 The <i>Potyviridae</i> family	15
1.2 The <i>Potyvirus</i> genus	15
1.3 Functions of viral proteins	16
1.4 Aphid transmission	21
1.5 Potyvirus translation and replication	23
1.6 Host response to infection at mRNA levels	27
1.7 Potyvirus VPg interacting protein (PVIP)	31
1.8 Quantification of host mRNA accumulation	32

Chapter 2

Analysis of primers for amplification of RT-qPCR reference genes

2.1 Introduction	35
2.1.1 Northern hybridisation	35
2.1.2 Expressed sequence tag (EST) analysis	36
2.1.3 Microarray	38
2.1.4 Reverse transcriptase polymerase chain reaction (RT-PCR)	39
2.1.5 RT-qPCR	41
2.1.6 Reference genes for relative quantification	47
2.2 Materials and Methods	49
2.2.1 Total RNA extraction	49
2.2.2 DNase treatment	50
2.2.3 cDNA synthesis	50

2.2.4 Reverse Transcriptase Polymerase chain reaction (RT-PCR)	51
2.2.5 Gradient PCR	51
2.2.6 Primer Design	51
2.2.7 Conformation of PCR products by cloning and sequencing	52
2.2.7.1 PCR product purification	52
2.2.7.2 Ligation	53
2.2.7.3 Transformation	53
2.2.7.4 Colony PCR and sequencing	54
2.3 Result	55
2.3.1 RNA quality	55
2.3.2 RT-PCR of using primers designed by Lilly et al (2011)	55
2.3.2.1 RT-PCR of EF1 α mRNA	56
2.3.2.2 RT-PCR of F-box mRNA	56
2.3.2.3 RT-PCR of the PDF2 mRNA	57
2.3.2.4 RT-PCR of the SAND mRNA	58
2.3.3 Optimising annealing temperature	59
2.3.3.1 Gradient PCR amplification of the EF1 α mRNA	60
2.3.3.2 Gradient PCR amplification of the PDF2 mRNA	62
2.3.3.3 Gradient PCR amplification of the SAND mRNA	63
2.3.3.4 Gradient PCR amplification of F-box mRNA	65
2.3.4 Primer design	67
2.3.4.1 Primer for the EF1 α gene	67
2.3.4.2 Primer design for PDF2 gene	69
2.3.4.3 Primer design for the SAND gene	70
2.3.4.4 Primer design for the F-box gene	72
2.3.5 New primer design	74
2.3.6 Testing new primer sequences	76
2.3.6.1 Analysis of new EF1 α primers	76
2.3.6.2 Analysis of new primer for F-box	78
2.4 Discussion	82

Chapter 3

Validation of reference genes and analysis of PVIP mRNA accumulation

3.1 Introduction	87
3.1.1 Use of RT-qPCR in virology	87
3.1.2 MIQE guidelines	87
3.1.3 Analysis of RT-qPCR data	89
3.1.4 Reference genes for RT-qPCR	90
3.1.5 Validation of reference genes	91
3.1.6 Potyvirus VPg interacting protein	93
3.2 Materials and Methods	95
3.2.1 Plant material	95
3.2.2 Total RNA extraction	95
3.2.3 RT-PCR to test PVIP primers on taro	95
3.2.4 Validation of reference genes using RT-qPCR	95
3.2.5 Determination of PVIP mRNA accumulation in healthy and DsMV infected taro	96
3.2.6 Statistical analysis	97
3.3 Results	98
3.3.1 RNA quality	98
3.3.2 Testing amplification of reference genes at 50°C	98
3.3.3 Validation of reference genes using RT-qPCR	99
3.3.3.1 Melting curve analysis	100
3.3.3.2 Statistical analysis	101
3.3.4 RT-PCR to test PVIP primers on taro	103
3.3.5 Relative quantification of the PVIP mRNA in healthy and DsMV infected taro	104
3.3.5.1 Melting curve analysis	105
3.4 Discussion	107

Chapter 4

General discussion	110
References	114
Appendices	126
Appendix 1	127
Appendix 2	128
Appendix 3	129

List of Figures

Chapter 1

Figure 1. 1: Organisation of potyvirus genome.	16
Figure 1. 2: This figure shows the involvement of various host proteins and functions of viral proteins in the potyviral replication and movement.	17
Figure 1. 3: This model shows the function of viral protein HC in aphid transmission in a non-persistent manner	22
Figure 1.4: A model of the roles of host chaperon HSP70 and co-chaperone CPIP in potyvirus replication	26
Figure 1. 5: Host responses and altered gene expression associated with plant virus infections.	29

Chapter 2

Figure 2. 1: The process of northern blotting and hybridization	36
Figure 2. 2: The figure describes EST sequencing- based approach in understanding transcriptome analysis	37
Figure 2. 3: This figure shows various steps i.e. RNA extraction, cDNA synthesis, probe labelling, hybridisation to microarray chip and image detection in the microarray data analysis	39
Figure 2. 4: The reverse transcriptase-polymerase chain reaction (RT-PCR) technique to determine the presence of a particular type of mRNA.	40
Figure 2. 5: This figure describes variation in the sample DNA amount as the reaction proceeds from exponential phase to plateau phase	41
Figure 2. 6: This figure shows the amplification plot of RT-qPCR	42
Figure 2. 7: The mechanism of the TaqMan probe RT-qPCR assay	44
Figure 2. 8: The mechanism of the SYBR Green RT-qPCR assay	45
Figure 2. 9: This figure outlines the quality control tools that need to be assessed throughout the entire RT-qPCR experiment	46
Figure 2. 10: Agarose gel electrophoresis of total RNA extracted from the leaf tissue of taro, <i>N.benthamiana</i> and <i>A. thaliana</i>	57
Figure 2. 11: RT-PCR using EF1 α F and EF1 α R primers on <i>A. thaliana</i> , taro and <i>N. benthamiana</i>	58
Figure 2. 12: RT-PCR using F-boxF and F-boxR primers on <i>A. thaliana</i> , taro and <i>N.benthamiana</i>	57
Figure 2. 13: RT-PCR using PDF2F and PDF2R primers on <i>A. thaliana</i> , taro and <i>N.benthamiana</i>	58
Figure 2. 14: RT-PCR using SANDF and SANDR primers on <i>A. thaliana</i> , taro and <i>N.benthamiana</i>	59
Figure 2. 15: Gradient PCR using EF1 α F and EF1 α R primers designed by Lilly et al., (2011) on <i>N.benthamiana</i> , taro and <i>A. thaliana</i>	60
Figure 2. 16: RT-PCR at 47°C using EF1 α F and EF1 α R primers on <i>N. benthamiana</i> , taro and <i>A. thaliana</i>	63
Figure 2. 17: Gradient PCR using PDF2 primers PDF2F and PDF2R designed by Lilly et al., (2011) on <i>N.benthamiana</i> , taro and <i>A. thaliana</i>	65

Figure 2. 18: RT-PCR using PDF2F and PDF2R primers on taro, <i>A. thaliana</i> and <i>N.benthamiana</i>	66
Figure 2. 19: Gradient PCR using SANDF and SANDR designed by Lilly et al., (2011) on <i>N.benthamiana</i> , taro and <i>A. thaliana</i>	64
Figure 2. 20: RT-PCR using SANDF and SANDR primers with on <i>N. benthamiana</i> , taro and <i>A. thaliana</i>	68
Figure 2. 21: Gradient PCR using F-box primers F-boxF and F-boxR designed by Lilly et al., (2011) on <i>N. benthamiana</i> , taro and <i>A.thaliana</i>	66
Figure 2. 22: RT-PCR using F-boxF and F-boxR primers on <i>N. benthamiana</i> , taro and <i>A.thaliana</i>	67
Figure 2. 23: The alignment of monocot and dicot species against <i>A. thaliana</i> in the region amplified by the Arabidopsis EF1 α primer sequences	68
Figure 2. 24: The alignment of monocot and dicot species against <i>A. thaliana</i> with region amplified by the Arabidopsis PDF2 primer sequences.	69
Figure 2. 25: The alignment of monocot and dicot species against <i>A. thaliana</i> with region amplified by the Arabidopsis SAND primer sequences.	75
Figure 2. 26: The alignment of monocot and dicot species against <i>A. thaliana</i> with region amplified by the Arabidopsis F-box primer sequences.	77
Figure 2. 27: The alignment of monocot and dicot species against <i>A. thaliana</i> with region amplified by the Arabidopsis EF1 α MD primer sequences.	79
Figure 2. 28: The alignment of monocot and dicot species against <i>A. thaliana</i> with region amplified by the Arabidopsis F-boxM primer sequences.	79
Figure 2. 29: The alignment of monocot and dicot species against <i>A. thaliana</i> with region amplified by the Arabidopsis F-boxD primer sequences.	80
Figure 2. 30: Gradient PCR using EF1 α MD primers taro, <i>N. benthamiana</i> and <i>A.thaliana</i>	77
Figure 2. 31: RT-PCR using EF1 α MDF and EF1 α MDR primers on taro , <i>A. thaliana</i> and <i>N. benthamiana</i>	77
Figure 2. 32: RT-PCR using EF1 α MDF and EF1 α MDR primers on <i>A. thaliana</i> with and without RT and taro	78
Figure 2. 33: RT-PCR at 50°C using (a) F-boxMF and F-boxMR and (b)F-boxDF and F-boxDR primers on taro, and <i>N.benthamiana</i>	79
Figure 2. 34: RT-PCR using (a) F-boxMF and F-boxMR and (b) F-boxDF and F-boxDR primers on <i>N. benthamiana</i> with and without RT, and taro	84
Figure 2. 35: Colony PCR using T7 and SP6 primers on <i>A. thaliana</i> , taro and <i>N.benthamiana</i> .	85

Chapter 3

Figure 3. 1: This figure shows relative and absolute quantification method of RT-qPCR.	89
Figure 3. 2 This figure shows the delta-delta C _t method where ΔC_t defines the difference between the housekeeping gene/ reference gene and gene of interest and $\Delta\Delta C_t$ defines the difference relative to the calibrator	90

Figure 3. 3: Agarose gel electrophoresis of total RNA extracted from the leaf tissue of taro	98
Figure 3. 4: RT-PCR using EF1 α MDF/ EF1 α MDR also F-boxMF/ F-boxMR and F-boxDF/ F-boxDR primers on <i>A. thaliana</i> , taro and <i>N.benthamiana</i>	99
Figure 3. 5: RT-qPCR results comparing mRNA accumulation of EF1 α and F-box genes in healthy and DsMV infected taro leaf tissues.	104
Figure 3. 6: Melting curve analyses from experiment 1 and 2 of all replicate and triplicate healthy and DsMV infected taro samples amplified using F-boxM and EF1 α MD primers.	105
Figure 3. 7: RT-PCR using PVIPF and PVIPR primers on taro, <i>N. benthamiana</i> and <i>A.thaliana</i>	108
Figure 3. 8: RT-qPCR result comparing expression level of PVIP in healthy and DsMV infected taro normalised against EF1 α and F-box reference genes.	109
Figure 3. 9: Melting curve analyses of all replicate and triplicate healthy and DsMV infected taro samples amplified using PVIP primers.	110

List of Tables

Chapter 2

- Table 2. 1: Primer sequences and expected sizes for EF1 α , PDF2, SAND and F-box mRNA from *A. thaliana*. 52
- Table 2. 2: The table below shows the new primer sequence to amplify both EF1 α and F-box gene from both monocot and dicot species 78

Chapter 3

- Table 3. 1: Checklist for author of MIQE at time of manuscript submission, detailed information about individual parameters associated with each step of the RT-qPCR analysis 88
- Table3. 2: PVIP primer sequences and amplicon size 99
- Table 3. 3: The average mean and standard deviation of EF1 α and F-box in healthy and DsMV infected taro for experiment and experiment 2 103
- Table 3. 4: The t-test analysis for EF1 α and F-box between healthy and DsMV infected taro for experiment 1 and experiment 2 to determine any significant variation of EF1 α and F-box within and among the replicates 103

Abstract

Potyvirus is the largest genus in the RNA plant virus family of *Potyviridae*. Potyviruses infect most of the economically important agricultural and ornamental crops; *Dasheen mosaic virus* (DsMV) is a potyvirus which infects edible aroid plants such as taro, particularly in the South Pacific region. These viruses rely on various host plant proteins for their movement and replication. One such plant protein which is known to interact with the viral protein called virus genome linked protein (VPg) is potyvirus VPg interacting protein (PVIP). Dunoyer et al., (2004) suggested role for PVIP in potyvirus movement, in disease development and as an important factor for virus replication. Further, through bioinformatics sequence analysis of homologous *Arabidopsis thaliana* and *Nicotiana benthamiana* PVIP sequences, the PVIP gene was found to be homologous to OBERON 1 and OBERON 2 in *A. thaliana*, suggesting a role for PVIP in meristem maintenance (Saiga et al, 2008; Anand, 2010). This analysis also suggested that PVIP may have a role independent of virus infection, it may be an important factor in plant development. Anand (2010) analysed the effect of abiotic stress on PVIP mRNA accumulation. The PVIP mRNA levels were assessed in leaf tissue of *N. benthamiana* under various dark and light conditions. A decline in the PVIP mRNA accumulation was observed when these plants were placed in continuous dark, suggesting that light induces the expression of PVIP mRNA. This study concluded that PVIP gene is responsive to this type of abiotic stress; however, the responsiveness of PVIP mRNA level to biotic stress such as virus infection is yet to be unravelled.

The aim of this study was to determine the variation of PVIP mRNA accumulation in healthy and DsMV infected taro using reverse transcriptase quantitative polymerase chain reaction (RT-qPCR). To conduct this analysis, the first objective was to identify appropriate reference genes for use in virus infected studies in taro.

There are several methods to conduct gene quantification studies such as northern hybridization, microarray data analysis; among these RT-qPCR has become the most reliable, common and sensitive method for the quantification of gene expression. According to the MIQE guidelines, there are many factors that need to be considered

while conducting RT-qPCR analysis, one such important variable is the use of appropriate reference genes (Bustin et al., 2010). A reference gene is defined as having a stable expression under different experimental treatments (Taylor et al., 2011). Housekeeping genes such as glyceraldehydes-3-phosphate dehydrogenase (GAPDH), ubiquitin (UBQ), actin 18S rRNA have been used as reference genes; however, it has been found that the expression level of many of these housekeeping genes vary under various experimental conditions.

Lilly et al., (2011) assessed the stability of 12 candidate reference genes in virus infected *A. thaliana*. Among these 12 genes studied such as actin, UBQ, only four genes were shown to have stable expression. These four genes were namely, F-box family protein (F-box), SAND family protein, protodermal factor 2 (PDF2) and elongation factor α (EF1 α). These genes were suggested to be suitable reference genes for analysis of all virus infected plants; however, this would need to be tested in other species before using them. Therefore, in this study the expression stability of these four reference genes namely, EF1 α , F-box, SAND and PDF2 in healthy taro and *N. benthamiana* was assessed using RT-PCR. Various parameters such as annealing temperatures were tested using gradient PCR to optimise for efficient amplification of these genes in taro and *N. benthamiana*; however, efficient amplification was not achieved in taro and *N. benthamiana* using these primer pairs. Further, sequence analysis of homologous *A. thaliana* EF1 α , PDF2, F-box and SAND genes against other publicly available monocot and dicot sequences, suggested significant variation within the primer target sequences, particularly in the 3' end of the reverse primers. This may account for the inefficient amplification of these genes in taro and *N. benthamiana*. Therefore, new generic primers, specific to both monocot and dicot species were designed; however, due to limited availability of sequence information for PDF2 and SAND in publicly available databases along with significant variations among monocots and dicots, new generic primers were only designed for the amplification of EF1 α and F-box genes. A EF1 α primer pair suitable for both monocots and dicots was designed. For F-box, two primer pairs were designed one for monocots and one for dicots. The newly designed primers were then tested on taro, *N. benthamiana* and *A. thaliana* using RT-PCR and optimum PCR conditions was obtained.

Next step was to validate the EF1 α and F-box genes as reference gene in virus infected taro using RT-qPCR. The data obtained was analysed statistically to determine any significant variation in the mRNA accumulation of EF1 α and F-box between healthy and DsMV infected taro. Both these genes were shown to have similar and constant expression in healthy and DsMV infected taro. This suggests that these genes could be used as suitable reference gene for gene quantification studies in taro. For other species such as *N. benthamiana* and other monocot and dicots, this would need to be tested and confirmed empirically.

For the analysis of PVIP mRNA accumulation in healthy and DsMV infected taro, PVIP primers designed for *N. benthamiana* were tested on taro and *A. thaliana* using RT-PCR. A very faint product of expected size was observed for both taro and *A. thaliana* using this primer pair; however, it was assumed that the observed amplification would be sufficient for the efficient amplification of PVIP in taro for RT-qPCR analysis. The mRNA accumulation of PVIP was normalised against the EF1 α and F-box reference genes for the RT-qPCR analysis. However, efficient amplification was not achieved from either healthy or DsMV infected taro using the *N. benthamiana* specific PVIP primer pair. Hence, generic monocot and dicot PVIP primers need to be designed and tested for future gene quantification studies in taro and other monocot and dicot species, to determine the variation in the mRNA accumulation of PVIP in virus infected plants.

Chapter 1

General Introduction

1.1 The *Potyviridae* family

The *Potyviridae* is the largest family of RNA plant viruses (Chung et al., 2008), which includes seven genera namely *Potyvirus*, *Tritimovirus*, *Rymovirus*, *Bymovirus*, *Maclurovirus*, *Brambyvirus* and *Ipomovirus* (Mayo et al., 2002). The *Potyviridae* particles are filamentous in structure and contain a positive sense single stranded RNA genome, with all the members except *Bymovirus* having monopartite genomes. *Potyviridae* members are transmitted by a range of vectors such as aphids, white flies, fungus and mites depending on the genus (Shukla, 1994). Many Potyvirids are important crop pathogens and are therefore commercially important (Berger, 2001).

1.2 The *Potyvirus* genus

Potyvirus is the largest of seven genera in the family *Potyviridae* consisting of the largest number of virus species with about 146 recognised members (Fauquet, 2005). Most of these species cause significant loss in agricultural, horticultural and ornamental crops. These viruses are transmitted by aphids in a non-persistent manner.

These viruses constitute an extraordinarily diverse group infecting a large number of host plant species including a broad range of monocotyledon and dicotyledon plants in most climatic regions. The type species of *Potyvirus* is *Potato virus Y* (PVY), which infects potato. Other species include dicot infecting species such as *Turnip mosaic virus* (TuMV), *Soybean mosaic virus* (SMV), *Bean common mosaic virus* (BCMV), and monocot infecting species such as *Wheat streak mosaic virus* (WSMV), *Sugarcane mosaic virus* and *Dasheen mosaic virus* (DsMV) (Berger, 2001). DsMV is of particular importance in the South Pacific region since it causes a major disease of the staple edible aroid plants taro (*Colocasia esculenta*) and tannia (*Xanthosoma sp*) (O'Hair, 1986).

The *Potyvirus* genome structure (as shown in Figure 1.1) consists of a single stranded positive sense RNA molecule of about 10 kb which is encapsidated by approximately 2000 units of viral coat protein (CP) (Riechmann, 1992) forming filamentous virions (Shuklo and Ward, 1989). The genome contains a single open reading frame (ORF) with short non translated regions (NTR) at both ends. The 5'NTR is attached to the viral –genome linked protein (VPg) while the 3'NTR is polyadenylated. Following translation, the genome is processed into ten polyproteins

including protein 1 (P1), the helper component protease (HC-Pro), protein 3 (P3), 6KDa peptide (6K1), the cylindrical inclusion (CI) protein, a second 6KDa (6K2) protein, VPg, the nuclear inclusion 'a' protein (NIa), the nuclear inclusion 'b' protein (NIb) and the coat protein (CP) (Kasschau et al., 1997).

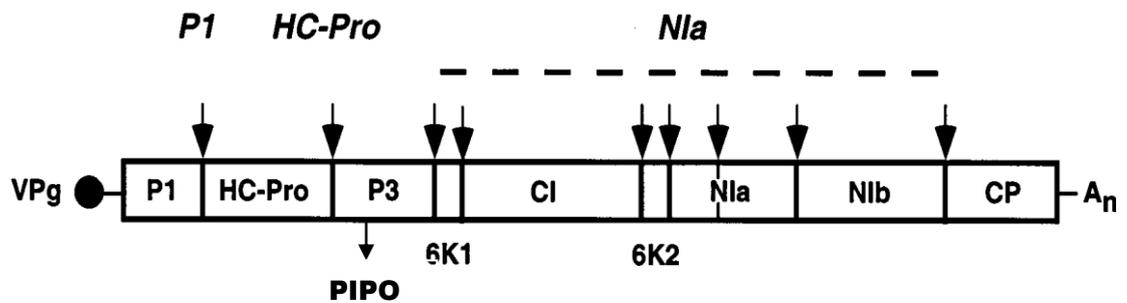


Figure 1. 1: Organisation of potyvirus genome. The vertical lines represent the cleavage sites within the ORF (Riechmann et al., 1992).

In 2010, a second ORF within the P3 coding region but out of frame with the one described above was discovered. This ORF encodes another protein called P3N- Pretty Interesting Potyviridae ORF (PIPO) from a translational frameshift. This second ORF was first identified in TuMV and was then identified by sequence alignment in 48 viruses that represented all genera of *Potyviridae* family (Tatineni et al., 2009).

1.3 Functions of viral proteins

Potyviruses must be transmitted from plant to plant (long distance movement), translated, replicated, and moved systemically within a host plant. All of these processes must be carried out by viral proteins, which in some cases are assisted by host proteins. Because viruses are gene poor, many viral proteins have been shown to be multifunctional. The roles of each potyviral protein are shown in Figure 1.2.

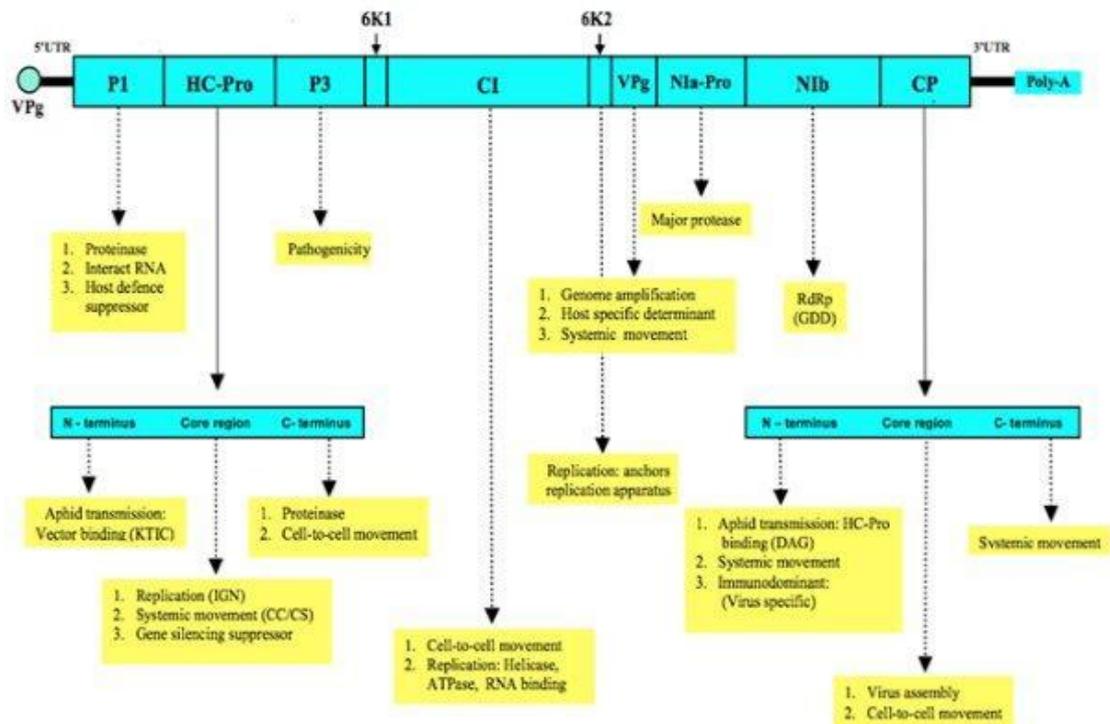


Figure 1. 2: This figure shows the involvement of various host proteins and functions of viral proteins in the potyviral replication and movement (Chang, 2012).

The P1 protein is released from the N-terminal end of the polyprotein by a proteolytic cleavage carried out by the C-terminal serine-type proteinase domain of P1 at a specific site between the P1 and HC-Pro. The cleavage is required for viral infectivity. In a study conducted on *Nicotiana tabacum* upon infection with *Tobacco etch virus* (TEV), it was observed that removal of the entire P1 coding region caused a reduction in the virus accumulation suggesting that P1 contributes to viral multiplication (Arbatova et al., 1998). P1 also interacts with many potyviral proteins that are involved in virus replication, including the HC-Pro (Kasschau and Carrington, 1998). Previously, due to the structural similarity of P1 with movement proteins (MP) of *Tobacco mosaic virus* (TMV) it was expected that P1 would be the potyviral MP; however, no evidence of viral movement function has been shown for P1 (Rajamaki et al., 2005).

The HC-Pro is a multifunctional protein that affects many distinct processes. For Potyviruses, this protein is mainly required for aphid transmission and hence the term “helper component” was coined (Syller, 2006). The N-terminus, central domain and C-terminus contains a cysteine proteinase that acts to cleave the C-terminus of

the HC-Pro from the potyviral polyprotein. The central region of the HC-Pro is also necessary for genome amplification and long-distance movement. Recently, HC-Pro or the central region has been shown that it acts in mediating suppression of post transcriptional gene silencing (PTGS), which is a mechanism used by plants for controlling RNA levels in cells and as a defence mechanism against viral infection (Stenger, French and Gildow 2005).

The P3 protein is the least characterised protein among all other *Potyvirus* proteins due to weak sequence homology among other *Potyvirus* species combined with a lack of identified structural and functional motifs. Mutation studies on the cleavage between P3 and 6K1 in *Plum pox virus* (PPV) showed effects on infection symptoms and variation in the appearance and strength of symptoms during infection. This evidence suggested a role for the P3 protein in virus pathogenicity (Urcuqui-Inchima, Haenni and Bernardi, 2001).

The 6K1 protein is one of the smallest proteins encoded by the potyviral genome. There have been no localisation studies and no reported functions for the 6K1 protein. However, it was suggested that the P3-6K1 region of *Pea seed-borne mosaic virus* (PSbMV) may play an important role in viral replication or cell-to-cell movement (Hjulsager et al., 2006) but the separation of P3 and 6KI in PPV showed no affect on virus viability while deletion of the genomic region encoding the 6K1 protein inhibited proper proteolytic separation of P3 from CI (Riechmann, Cervera and Garcia, 1995). Recently, prokaryotic expression of the *Soybean mosaic virus* (SMV)-P 6K1 protein was analysed by using immunogold labelling to localize the 6K1 protein in *Pinellia ternate* (Kang et al., 2004). The consistent labelling of the 6K1 protein was observed at the regions of cell walls and cell membranes using antiserum raised to 6K1 proteins. Generally, in plant viruses such labelling patterns suggest some role of the protein in viral cell-to-cell movement (Hong et al., 2007); however, such proteins usually contain one more transmembrane domain which was not detected in 6K1 protein (Hong et al., 2007). Therefore, it is still uncertain if the 6K1 protein has a role in cell-to-cell movement.

Studies conducted in *Nicotiana benthamiana* and *N. tabacum* plants upon TEV and *Potato virus A* (PVA) infection involving insertion and deletion of histidine residues (6XHis) into various positions of the membrane bound 6K2 protein suggested that

this protein anchors the viral replication complex to the site of replication on the endoplasmic reticulum (ER) (Spetz and Valkonin, 2004). Further, with the mutation (Gly 2Cys) in the 6K2 adjacent to the 6XHis insert which is placed between Ser1 and Gly2 indicated variation in the infection symptoms. This suggested that 6K2 protein affects viral long-distance movement and symptom induction (Spetz and Valkonin, 2004).

The VPg protein is present at the N- terminus domain of the NIa protein. The VPg is released from the NIa through the action of the viral proteinase called NIa-Pro. During the maturation process, many precursor forms of the viral protein are produced such as VPg-NIa Pro, which has been identified as a soluble protein in the cytoplasm and in the nucleus of infected cells (Schaad et al., 1996). The VPg attaches covalently to the 5' terminus of genomic RNA through a phosphodiester bond with the hydroxyl group of Tyr-62 (Bustamante and Hull, 1998). The VPg has been shown to interact with several proteins both of viral and host origin. For example, it has been suggested that the VPg binds with viral protein HC-Pro to form a bridge between virus particle and aphid mouthparts to enable transmission (Goodfellow, 2011). Similarly, VPg also interacts with host translation factors such as eIF4E, eIF(iso)4E and poly(A) binding protein (PABP) to stimulate the translation initiation of viral RNA (Roudet-Tavert et al., 2007). It has been observed that VPg acts as a cap analogue and interacts with the cap binding initiation factor eIF4E. It has also been suggested that the interaction of VPg with host factors enhances viral movement from cell-to-cell since VPg attaches covalently to viral RNA and so the viruses get transported from cell-to-cell in the form of a viral-protein complex (Jiang et al., 2011).

The NIa protein consists of a C- terminus protease domain and an N- terminus VPg domain. The N-terminal domain also contains a nuclear localisation signal that directs the majority of NIa molecules to the nucleus of infected cells (Schaad, Jensen and Carrington, 1997). The NIa protein catalyzes cleavage of the polyprotein at five positions through a series of autoproteolytic and trans-proteolytic reactions. All of these cleavage sites feature a conserved sequence that is essential for substrate recognition. The NIa proteinase is required for proteolytic maturation of the majority of viral proteins (Verchot, Koonin and Carrington, 1991).

Besides its role in polyprotein processing, it has been observed in TEV using yeast two hybrid (YTH) analysis that NIa interacts with NIB, the RNA-dependent RNA polymerase (RdRp). This interaction involves mainly the NIaPro domain of NIa. This observation suggested that NIa could have an important role during initiation of RNA synthesis by recruiting the NIB polymerase through protein-protein interaction (Daras and Carrington, 1997). The VPg present at the N- terminal region of NIa is covalently attached to the 5' terminus of viral RNA through phosphodiester linkage and this VPg serves an essential role during the initiation steps of viral RNA synthesis, hence suggesting a role for NIa in RNA synthesis as well (Rantalainen et al., 2008).

The NIB is the viral RdRp. The NIB protein is translocated to the nucleus to form nuclear inclusion bodies (Puustinen et al., 2004) as well as being present in the cytoplasm acting as the RdRp. The NIB contains the amino acid triplet Gly-Asp-Asp which has been found to be a universally conserved motif among RdRps (Carrington et al., 1993) and mutation at this motif of TEV NIB led to an inability of the virus to replicate, indicating the role of NIB in viral replication (Puustinen et al., 2004). An YTH analysis of TEV revealed that NIB also interacts with other replication proteins especially with the multifunctional VPg/NIa-Pro and the CI protein (O'Reilly et al., 1997). This was also shown for *Tobacco vein mottling virus* (TVMV) (Hong et al., 1995). Further, by genetic complementation analysis in TEV, the NIB protein was shown to be a *trans*-active protein that could be supplied from outside the context of the TEV polyprotein suggesting that NIB may be recruited to replication complexes either through protein-protein interactions with viral or host factors or by protein-RNA interactions (Puustinen et al., 2004).

The CP is a multifunctional protein and plays an important role in cell-to-cell movement and long distance transport. It contains three domains, namely N- and C-terminal regions which are exposed on the virion surface and a central highly conserved core domain (Rojas et al., 1997). Among these regions, an Asp- Ala- Gly (DAG) motif present in the N-terminal region acts in aphid transmission and in virus accumulation whereas the C-terminal region is involved in long-distance movement. The central domain facilitates virus assembly, cell-to-cell movement and genome amplification (Ullah et al., 2003).

The PIPO protein was first identified in TuMV (Chung et al., 2008) and the analysis conducted using antibodies designed against two synthetic peptides corresponding to amino acids of PIPO indicated that this protein is not expressed independently. Rather it acts as a fusion protein with the N- terminal portion of P3. However, the presence of P3 amino acid sequence in PIPO remains yet to be demonstrated. Further, it was also observed that knocking out PIPO protein expression in TuMV was lethal to the virus in *N. benthamiana* (Wei et al., 2010). A similar study conducted in SMV strains did not show any effect on virus replication instead a knockout of PIPO restricted the movement of resultant mutants in inoculated soybean leaves, suggesting a crucial role of PIPO protein in the movement of SMV (Wen and Hajimorad, 2010). These results indicate that the PIPO protein may be an essential component in the replication cycle of the virus; however, experimental evidence of the PIPO protein in the replication cycle of potyvirus remains unknown.

1.4 Aphid transmission

Viruses adapt different steps of their infection cycle such as protein translation, genome replication, cell-to-cell movement and host- to- host transmission according to the characteristic of their host plant (Dawson et al., 1992). Most plant viruses use a vector for efficient transmission to new hosts. A vector can be defined as a mobile organism that gains access to plant cytoplasm by breaking through the cell wall and membrane using their feeding organs. Examples of such vectors include plant-feeding arthropods particularly insects, parasitic fungi and root nematodes (Graff and Brault, 2008).

Potyriviruses are transmitted through aphids in a non-persistent manner (Pirone and Perre, 2002). The vectors of non-persistent viruses can transmit immediately after an infective feed and remains infective for only a short period of time. The stylets of aphids become contaminated with virus during penetration in infected tissue and only this fraction of the virus is of significance in disease transmission. Some of the virus may also be ingested but those viruses get inactivated by the salivary fluids and so these viruses play no part in disease transmission (Blanc et al., 2002).

Aphids can acquire viruses at all steps of their feeding process, early during intracellular probes within cytoplasm of epidermal or mesophyll cells to later in the

vascular system. Aphids can acquire and inoculate any viral taxon within plants (Ng et al., 2004).

The HC-Pro component of Potyviruses aids them in binding to the aphid mouthparts, thereby facilitating transmission (Syller, 2006). It has been suggested from evidence obtained from transmission electron microscopy that the HC-Pro acts as a bifunctional molecule of which one domain binds to the virus CP and the other binds to the vector mouthparts (Figure 1.3). For a successful transmission, the N-terminal region of the viral CP must interact with the HC-Pro which in turn binds to a specific receptor within the aphid stylet (Ullah et al., 2003).

The interaction between HC-Pro and stylets of aphid vectors was shown to depend on the presence of a peptidyl domain, a four residue motif present at the N-terminal region of the HC-Pro suggesting that these proteins could interact with an unknown aphid receptor (Maia et al., 1996). Another motif called the PTK motif present at the C-terminal end of the HC-Pro was observed to be necessary for efficient interaction between the HC-Pro and Asp-Ala-Gly the DAG domain present at the C-terminus of the CP (Plisson et al., 2003). This evidence indicated that HC-Pro binds to both viral CP and to a putative receptor within the aphid stylets, forming a molecule bridge between the virus and the vector during non-persistent transmission process (Brault et al., 2008).

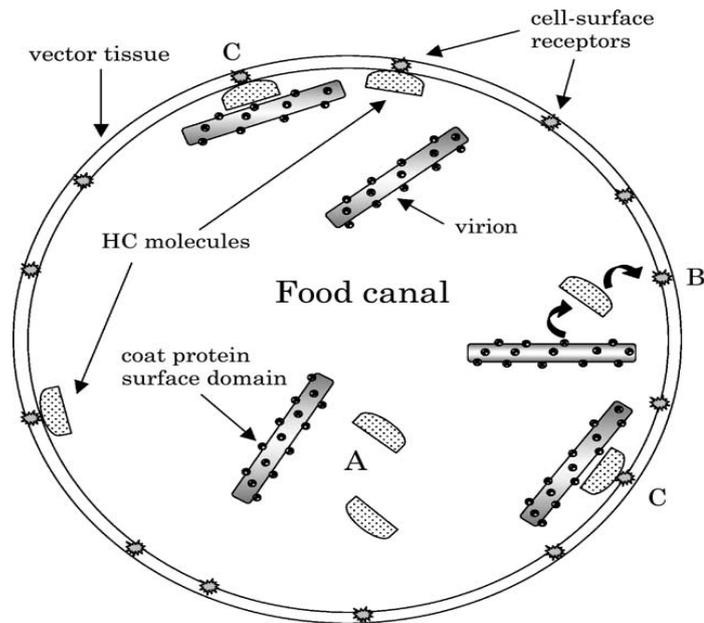


Figure 1. 3: This model shows the function of viral protein HC-Pro in aphid transmission in a non-persistent manner. The HC-Pro acts as a bridge in attaching the viral particles to the cuticle of the maxillary food canal and foregut of aphid vectors. (A) Shows free virions and HC molecules acquired by aphid. (B) Shows the linking of HC to a specific receptor on the vector tissue followed by binding of virus to HC. (C) Shows HC-virion complex bound to the vector stylet (Syller, 2006).

1.5 Potyvirus translation and replication

Once the virion has entered into a host cell, for the viral genome to be replicated, it must be unencapsidated to expose the viral RNA. Once this occurs, the genome must then be translated by the host cell and these translation products together with host proteins can replicate the RNA genome (Figure 1.4).

As described earlier, following translation, the viral proteins are cleaved from the polyprotein either by autocleavage of P1 and HC-Pro, or through cleavage by the NIa protein. Some of these proteins, either as discrete proteins, or in complexes with others, or even as cleavage intermediates, have various roles in replication.

It has been observed that host translation machinery components such as elongation factor (EF1 α) and initiation factors eIF4E and eIF(iso)4E play crucial roles in the virus replication (Blackwell et al., 1997). The two important features that are necessary for the translation of cellular mRNAs are: the 5' m⁷G(5')ppp(5')N cap and the 3' poly(A) tail. The interaction of eIF4E with the cap structure and the interaction of PABP with the poly(A) tail facilitates translation initiation. The large adaptor protein, eIF4G which binds eIF4E and PABP forms a closed loop complex to bring the mRNA ends into proximity (Wilkie et al., 2003). In plants, isoforms of

eIF4E namely eIF4E and eIF(iso)4E and isoforms of eIF4G namely eIF4G and eIF(iso)4G have been found but potyviruses generally need only one specific eIF4E isoform to multiply in a specific host (Goodfellow, 2011).

The potyvirus RNA lacks a 5' cap structure but it has been observed in TuMV using YTH analysis that the VPg at the 5' end of potyviral RNA binds with the host translation factors i.e. eIF4E and eIF(iso)4E (Leonard et al., 2000). The VPg has been considered to be equivalent to a mRNA 5' cap and indeed the VPg of TuMV (amongst other potyviruses) has been shown to interact with eIF4E and eIF(iso)4E (Miyoshi et al., 2006). Mutations in the VPg domain that are necessary for interaction with eIF(iso)4E reduced TuMV infectivity suggesting that VPg has a role in recruiting translation factors; however, there isn't any direct evidence that eIF4E/eIFiso4E interaction with VPg participates in cap dependent translation as described above for cellular mRNAs (Jiang et al., 2011). VPg may have a role in directing cap-independent translation since VPgs have been shown to inhibit translation of capped mRNAs *in vitro* but stimulates translation of uncapped mRNAs *in vitro* (Eskelin et al., 2011). Further, TEV 5'UTR when linked to the coding region of β -glucuronidase (GUS) stimulated cap-independent translation (Carrington et al., 1990) suggesting that TEV 5'UTR promotes translation in a cap-independent manner which may involve binding of proteins or ribosomes to internal sites within the UTR. However, this feature is not a specific characteristic of all positive stranded viral mRNAs since similar fusion of GUS with 5'UTR from *Turnip yellow mosaic* or *Black beetle virus* genomes failed to stimulate translation (Carrington et al., 1990). While it is unclear as what is the exact mechanism for translation initiation of potyviruses, it is clear the interaction of the viral protein VPg with host factors is an essential component.

Following translation, the viral genome must then be replicated. The replication of positive-sense RNA viruses such as *Potyviruses* occurs in two steps both of which are catalysed by the viral RdRp N1b. In the first step, minus strand is synthesised which acts as a template to direct synthesis of genomic RNAs and in the second step, positive sense genomic RNA molecules are synthesised (Whitham et al., 2004). It has been suggested that both 5' and 3' NTRs play a crucial role in viral replication but the RdRp is the major polypeptide that catalyses the synthesis of RNA strands from both negative and positive strand templates of viral RNA. The initiation of RNA synthesis occurs either by primer-independent or primer-dependent initiation.

In the majority of positive-sense RNA viruses, RdRp (i.e. NIB) initiates primer-dependent RNA synthesis wherein a single nucleotide provides a 3' hydroxyl group to initiate the reaction. The interaction of NIa with the NIB stimulates polymerase activity. In other positive-sense RNA viruses such as Polioviruses, the VPg has been shown to interact with the viral 3'-end sequences to prime the negative strand RNA synthesis suggesting that VPg could also be a putative primer for Potyvirus replication (Miyoshi et al., 2006).

Replication occurs in replication complexes (RC) and requires a membranous environment. The source organelle for membranes can be the endoplasmic reticulum (ER), mitochondria or lysosomes (Schaad et al., 1997). In potyviruses, there is evidence that suggests the 6K2 protein acts as a transmembrane anchor (Spetz et al., 2004). In TuMV, it has been observed that 6K2-VPg-Pro polyprotein intermediate induces vesicle formation from ER membranes in plants. These vesicles contain at least three host translation factors along with viral replication proteins suggesting that replication and translation could occur inside same vesicle (Rantalainen et al., 2009).

Since the potyviral genome is positive sense, it must be replicated into a negative sense orientation which then acts as the template for positive sense genome synthesis. The RdRP cannot make nucleic acids *de novo*, it requires a primer-type structure to initiate synthesis (Teycheney et al., 2000). It has been observed that the uridylylated VPg may act as the primer for the viral RNA synthesis. Recombinant *Pepper vein banding virus* (PVBV) NIB and VPg were over expressed in *Escherchia coli* and it was observed that the NIB, which has been found to be active as polymerase, uridylylated VPg (Gruez et al., 2008). The occurrence of uridylation of VPg even in the absence of poly(A) tail as template suggests that this uridylation of VPg is template independent. From the VPg deletion analysis at the N-terminus and C-terminus showed that the N-terminal domain 38 amino acid residues are necessary for the uridylation activity. Further, it was identified using site-directed mutagenesis that the Tyr 66 of PVBV VPg is the amino acid residue which gets uridylylated by the NIB (Gruez et al., 2008).

From the YTH analysis in cucumber leaves it was observed that the RdRp of *Zucchini yellow mosaic potyvirus* (ZYMV) interacts with PABP. Further, deletion

analysis confirmed that the C-terminus of the PABP is necessary for interaction with the RdRp (Dufresne et al., 2008). This study suggested a for host protein PABP in the potyviral infection process. A similar study conducted in tomato with *Pepino mosaic virus* (PepMV) CP showed interaction of PepMV CP with the HSP70 protein (Mathioudakis et al., 2012). HSP has been proposed to have a function in regulating the amount of CP present in the viral replicase complex (VRC) at the early stage of replication (Figure 1.4). The CP is produced through the cleavage of the polyprotein precursor. CP has been found to be inhibitory to potyvirus replication at the early stage of infection. Therefore, a J-domain protein called CPIP and HSP70 together have been shown to facilitate degradation of CP to prevent early cessation of replication (Nagy et al., 2011). Apart from the interaction of CP with HSP70, a YTH analysis conducted on *N. benthamiana* with *Potato virus X* (PVX) showed that PVX CP interacts with 26S proteasome which is normally involved in ubiquitination and has been suggested to have possible role in membrane degradation, possibly to regulate viral phloem unloading (Mathioudakis et al., 2012).

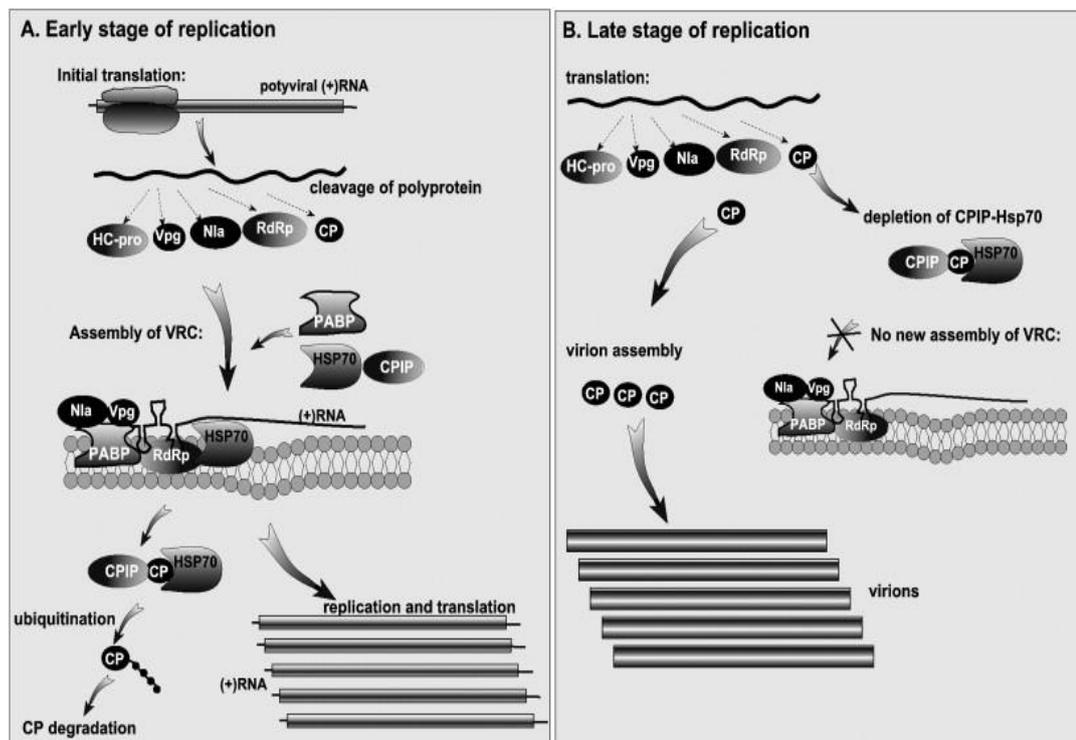


Figure 1. 4: A model of the roles of host chaperones HSP70 and co-chaperone CPIP in potyvirus replication. The interaction between HSP70 and CPIP causes ubiquitination and degradation of CP. Further, as the amount of CP increases, CPIP-HSP70 interaction reduces leading to the shutdown of replication and translation (Nagy et al., 2011)

Further studies conducted on *Arabidopsis thaliana* to identify cellular interactors of TuMV RdRp using tandem affinity purification showed the interaction of HSP70 and PABP with TuMV RdRp *in vitro* (Dufresne et al., 2008). It was also observed that 6K-VPg-Pro which induces the formation of cytoplasmic vesicles and favours the viral replication complex, redirected the RdRp polymerase and HSP70 to large endoplasmic reticulum- derived vesicles where replication has been suggested to occur. Thus it was concluded that HSP70 and PABP could be important components for the replicase complex and could have an important role in the regulation of potyviral RdRp functions (Dufresne et al., 2008).

While many host proteins having a role in virus replication have been identified, it is likely that some are still to be identified.

1.6 Host responses to infection at mRNA levels

Infection by virus constitutes one form of biotic stress. Plants are known to respond to biotic stress by changing gene expression patterns with several studies having been done to learn which genes changes occur in response to viral infection. These types of studies have lead to a better understanding of how the host may be involved in the replication process, or how they may be trying to defend themselves from infection.

Viruses exert negative effects by influencing the metabolism of the host but this mechanism is poorly understood for plant viruses.

Wang and Maule (1995), used *in situ* hybridisation to study host mRNA responses in pea embryo infected with PSbMV. Nine pea genes were analysed, five of which encode pea seed storage proteins that is legumin, vicilin and convicilin, and four genes that encode isoform of granule-bound starch synthase (GBSI and GBSII) together with large and small subunits of adenosine diphosphate glucose phosphorylase (ADPGP) that are involved in starch biosynthesis. These selected genes were involved in two different host metabolic pathways with different steady state amount of transcripts. It was observed in this study that in the section of

uninfected tissues, host transcripts showed uniform distribution in the cortical tissue of the cotyledon with increased expression of seed storage proteins in the peripheral and surface layers of the cotyledon. In contrast, in the section of infected embryo, a band of cotyledonary cells with nearly detectable host gene transcripts were observed regardless of the amount of transcript accumulation of these genes in uninfected cells. In the cells, where the amount of PSbMV negative sense RNA had declined, transcripts of the nine host genes accumulation were similar to the transcript of those genes found in uninfected cells of the same cotyledon section. It was shown that the cells deficient in host transcripts were most active in viral replication.

Since this study was detecting mRNA rather than protein, these results indicated that the major effect exerted by viruses on host protein is the shut off of either transcription and or translation of host mRNA.

Although the above mentioned analysis indicated that the PSbMV infection inhibited host gene expression, further studies using the same system identified genes whose expression appeared to be induced by PSbMV infection were conducted to identify any effect of viral in the induction of host gene expression and these changes were confined to specific cell types within the cotyledon (Aranda et al., 1996). Change in lipoxygenase, host chaperone HSP70 and polyubiquitin expression were analysed in PSbMV infected pea cotyledon. It was observed that the PSbMV infection induced accumulation of HSP70 and polyubiquitin mRNA while the expression of lipoxygenase was down-regulated upon virus infection. It is unknown that, following heat treatment, polyubiquitin and HSP expression are both induced (Aranda et al., 1996). Heat induced stress causes the degradation of host mRNAs against which HSP possibly polyubiquitin RNAs are protected (Aranda et al., 1996). Since many of the features of the PSbMV infection mimics those resulting from heat stress, this induction in HSP expression upon viral infection suggests an effect of viral replication on host protein expression. Once the infection has passed, the expression of HSP70 declined and failed to recover up to the normal level as seen in uninfected tissue; however, polyubiquitin expression rapidly recovered to normal levels once the viral replication had stopped (Aranda et al., 1996). In contrast sustained induction of HSP70 was observed in *Arabidopsis* leaves, infected with *Tobamovirus* infection (Whitham et al., 2003). This variation could be explained if

viral infection had different consequences on host gene expression in *Arabidopsis* leaves versus pea cotyledons. It could also be due to the fact that *Arabidopsis* mRNA was extracted from total leaf tissue and so the sample may have contained a mixture of infected and uninfected cells.

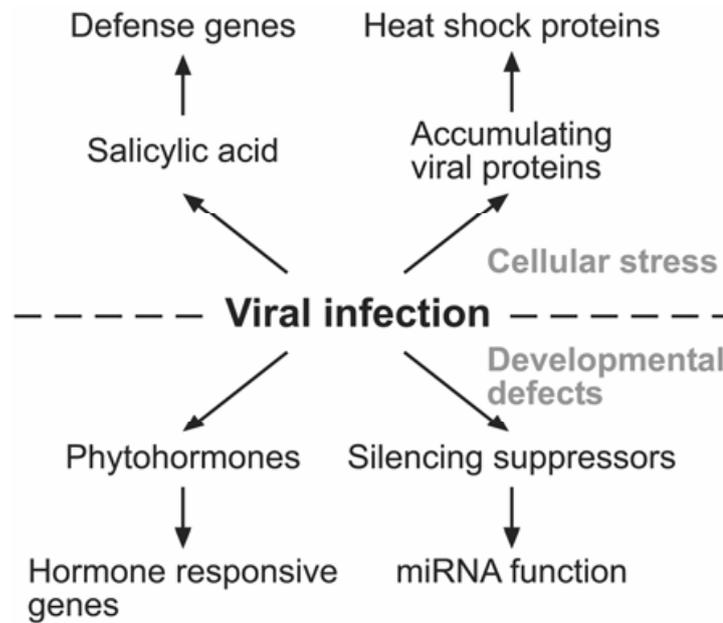


Figure 1. 5: Host responses and altered gene expression associated with plant virus infections. Compatible interactions between hosts and viral pathogens lead to the expression of suites of defense-related and heat shock genes. The expression of these genes is controlled by signalling pathways in the plants associated with initiation of defense responses and by other cellular stress response pathways. Viral infections can also disrupt the functions of regulatory small RNAs, such as micro RNAs, and phytohormone signaling leading to developmental defects (Whitham et al., 2006).

The observation that PSbMV exerted host shutoff and heat shock protein induction throughout embryo development suggests that this is a general mechanism and it is not peculiar to the specialised cell within the cotyledon.

Yang et al. (2007) studied *Arabidopsis* gene responses in response to TuMV infection. To conduct this analysis, a microdissection approach was used since it was believed that the use of whole leaves might mask interesting gene expression changes thus making them unable to detect. Four zones were defined in the leaf based on the average size of infection foci. The accumulation of TuMV occurred in a decreasing gradient radiating from zone 0 to zone 2 with no evidence of replication in zone 3. The host genes which were upregulated included those associated with host defense and stress response such as ribosomal proteins (RPs) whereas

downregulated genes included these involved in chloroplast function, sulphate utilisation and cell wall expansion (Yang et al., 2007). The increased expression of RPs upon TuMV infection may be due to plants trying to compensate for the loss in order to synthesise its own protein since RPs are involved in protein synthesis or it could be the mechanism of TuMV to increase the capacity of cells to synthesise viral proteins. On the other hand, the decrease in the expression of chloroplast, sulphate and cell wall extensibility genes suggests a possible mechanism in the symptom development such as chlorosis and stunted growth in TuMV infected *A. thaliana* plants (Yang et al., 2007).

To analyse virus-general and virus-specific variations in the expression of genes associated with different defense or stress responses, high-density oligonucleotide probe microarray technology was used and the leaves of *Arabidopsis* were inoculated with five different RNA viruses; such as TuMV, PVX, *Cucumber mosaic virus* (CMV) and *Cauliflower mosaic virus* (CaMV). This analysis showed induction of both virus specific and general changes in host gene expression among 114 genes such as HSP70, growth regulator like protein, putative protein kinase, putative DNA binding protein and other unclassified putative protein analysed (Whitham et al., 2003). The co-ordinated induction of stress and defense associated regulatory genes such as PAD4 suggests that various RNA viruses either specifically or non-specifically elicit general plant stress responses which lead to increased accumulation of stress and defense associated mRNA transcripts.

As a defense mechanism by host plants, RNA silencing has been shown to be involved in adaptive defense against various RNA and DNA viruses (Chen et al., 2004). This antiviral response is triggered by virus specific double stranded RNA intermediates produced during genome replication, by transcription or through recognition of viral RNAs by host RdRp. However, as a counter defense the potyviral protein HC-Pro has been shown to suppress the RNA silencing at various steps in the pathway (Chen et al., 2004).

There is still much information lacking regarding the effects of each viral protein in triggering host responses and the signalling network that are involved. The mRNA studies will provide new insights into how plants respond to viral infection.

1.7 Potyvirus VPg interacting protein (PVIP)

As discussed above there are many host proteins that have been shown to interact with viral proteins to support viral replication. There are also many host proteins that have been shown to interact with viral proteins whose function is unclear. One such host protein has been shown to interact with VPg and is called Potyvirus VPg interacting protein (PVIP).

PVIP was initially identified from pea (Dunoyer et al., 2004) and is homologous to small gene family in *A. thaliana*, with its counterparts in *N. benthamiana*. VPg was first identified attached to the RNA genome structure of polioviruses (Schein et al., 2006). Using the YTH system, PVIP of *A. thaliana* was shown to interact with VPg of many potyviruses, such as TuMV. The amino acid sequence analysis conducted on PVIP related sequences from *A. thaliana* and *Oryza sativa* identified the presence of “Plant Homeo Domain (PHD) zinc finger-like” domain in all PVIPs suggesting a role of PVIP for VPg in the nucleus. This domain consists of series of Hys and Cys residues that have roles in transcriptional regulation (Dunoyer et al., 2004). Further, through mutational analysis of two genes of PVIP namely PVIP1 and PVIP2 in Arabidopsis RNAi plants, functional redundancy of these two genes were observed suggesting a possibility of the presence of these two genes in other plant species as well; however, involvement of these genes in the virus movement was not observed. This suggested formation of PVIP1 and PVIP2 heterodimer structure. This study also identified the first 66 amino acids of VPg to be important for the interaction of VPg with the PVIP by analysing VPg sequences between TuMV, *Lettuce mosaic virus* (LMV), TEV and PSbMV. However, TEV VPg was shown not to interact with PVIP, raising speculations about any significant role of PVIP in virus movement.

Although, Dunoyer et al., (2004) suggested a role for PVIP in the virus movement this has not been confirmed.

Later Anand (2010) showed PVIP is homologous to OBERON1 and OBERON2 in Arabidopsis. These genes have a role in meristem maintenance and have a zinc ion binding domain (Saiga et al., 2008).

PVIP has been shown to be expressed throughout healthy tissue of *N. benthamiana* (Anand, 2010), and so has a role that is independent of virus infection. Recently, the

effect on PVIP mRNA level upon abiotic response was analysed by assessing the PVIP mRNA accumulation in the leaf tissue of *N. benthamiana* under various dark and light conditions (Anand, 2010). It was observed that there was a decrease in the PVIP mRNA accumulation when these plants were placed in continuous dark, suggesting that either lack of light represses expression, or light induces the expression of PVIP mRNA; however, no study has been conducted so far to analyse the responsiveness of PVIP mRNA level to biotic stress such as virus infection.

1.8 Quantification of host mRNA accumulation

Quantification of host mRNA accumulation is of great importance when trying to understand responses to virus infection. Currently, quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) is the method of choice for this type of study. How this analysis is carried out will be described in more detail in chapter 2; however, to use this technique effectively, it is necessary to use reference genes to normalise the accumulation of the target mRNAs (Becker et al., 2010). The choice of appropriate reference genes has been problematic. For a gene to be suitable as a reference gene, its mRNA accumulation must not change from one treatment to another (Radonic et al., 2004). For virus infection studies, the expression of reference genes must not change between the healthy and infected states. The genes for actin, ubiquitin and 18S have been used as reference genes (Jain et al., 2006), but have been shown to be unsuitable (Radonic et al., 2004).

Lilly et al (2011) recognised the need for reference genes that are appropriate for virus infection studies. This study was conducted in *A. thaliana* infected with TuMV and other viruses, involved analysis of twelve candidate reference genes; F-box family protein (F-box), Sand family protein (SAND), Protodermal factor 2 (PDF2), actin, β - tubulin (TUB), ubiquitin (UBQ10) and EF1 α . Among these genes tested, four of them, namely F-box, SAND, PDF2 and EF1 α showed most stable accumulation against all other twelve genes tested. It was suggested that these four genes would be suitable as reference genes for virus infection studies; hence genes were analysed only in Arabidopsis. Their usefulness in other species requires assessment.

In this study; however, the transcript accumulation of these four reference genes in healthy and virus- infected *N. benthamiana* and taro was studied.

Although, *A. thaliana* has become the model plant for a wide range of studies in plants, its usefulness as model for studying virus infection of plants is somewhat limited. *N. benthamiana* is the preferred model plant for virus infection studies (Goodin et al., 2008) and therefore, the usefulness of these reference genes in this species needs to be assessed. Both *N. benthamiana* and *A. thaliana* are dicot species. Many commercially important species are monocots; for example maize, rice and wheat which are *Poaceae*. Taro is also a monocot which belongs to *Araceae* family, which is also an important staple food crop in the South Pacific and is therefore an important crop plant to study.

The aim of this study was firstly to determine if F-box, SAND, PDF2 and EF1 α are appropriate reference genes for use in virus infection studies in taro and *N. benthamiana*; and secondly, to analyse the variation in the transcript accumulation of the host protein PVIP between healthy and potyvirus infected taro.

Chapter 2

Analysis of primers for amplification of RT-qPCR reference genes

2.1 Introduction

Gene expression analysis has helped us increase our understanding about metabolic and signalling pathways that underlie environmental response and development (Zhu et al., 2005). There are many techniques available for the quantification and detection of mRNA accumulation. Some of the widely used methods include northern hybridisation, expressed sequence tag (EST) analysis, proteomics, microarray data analysis, reverse transcriptase polymerase chain reaction (RT-PCR), RNase protection assay and Serial analysis of gene expression (SAGE) (Stanton, 2001).

2.1.1 Northern hybridisation

Northern hybridisation is a method used to detect the presence of RNA, RNA degradation, RNA splicing and size of specific RNAs within a population of cells. The methods in northern hybridisation include extraction of RNA from a given sample which is fractioned by electrophoresis that gets immobilised onto a membrane which then gets hybridised with a labelled complementary nucleic acid probe (Figure 2.1). The RNA is separated on a denaturing gel to ensure separation based on size. The probes used for hybridisation can be DNA, RNA or an oligonucleotide of at least 25 oligonucleotides complementary to target sequences (Valoczi et al., 2004). Hybridisation is detected by autoradiography on X-ray film of phosphor imaging if the probe is labelled with a radioactive label, else if the probe is labelled with a non-radioactive chemiluminescent label then colour detection is used, where the substrate is converted to a coloured product (Trayhurn et al., 1995). Northern blots generally detect one RNA sequence at a time, giving the size of that RNA when compared to size standards. Thus, splice variants can be detected. Also, probes with partial complementation can be used to detect targets, allowing cross species detection (Pall et al., 2007). Differences in gene expression can be detected and if combined with densitometry, can provide semi-quantitative measures of RNA (Gavin and Wagner, 2002). Northern hybridisation; however, is not truly quantitative and is time consuming and does not allow efficient quantification of mRNA. It is also a less sensitive method compared to RT-qPCR (Parker and Barnes, 1999).

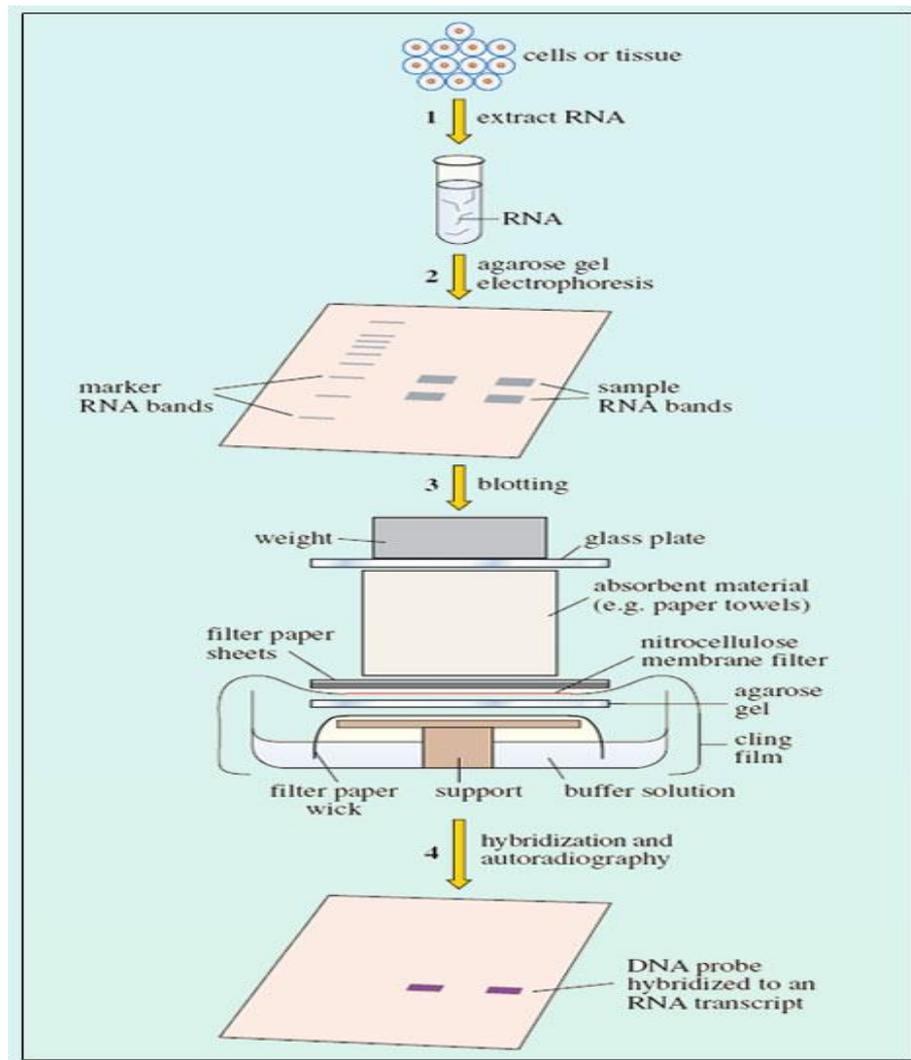


Figure 2. 1: The process of northern blotting and hybridization. (1) The total RNA from tissues is isolated. (2) RNA molecules are separated based on the size by gel electrophoresis and standard markers of known size are included to determine the size of RNA. (3) The gel is blotted onto a filter membrane and an absorbent paper is placed around gel to form a seal. (4) The blot is then hybridized using a radioactively labeled DNA probe to identify mRNA and the using autoradiography, labeled probe attached to the target RNA is observed

(http://open.jorum.ac.uk/xmlui/bitstream/handle/123456789/956/Items/S377_1_section7.html).

2.1.2 Expressed sequence tag (EST) analysis

ESTs are short, unedited, randomly selected single DNA sequencing reads obtained from complementary DNA (cDNA) clone libraries constructed from a given tissue (Figure 2.2). EST sequences are deposited in a special sub database of GenBank. Usually, an EST sequence is approximately 400-500 bases long. Sequencing a large number of these clones allows study of gene expression and transcripts in that particular tissue; it also provides an overview of tissues with active genes under

those conditions. It has been used for transcriptome analysis, gene discovery and for the confirmation of *in silico* gene prediction from genome sequence analysis (Page and Minocha, 2004). Several overlapping ESTs are required to create a sequence for the full length of an mRNA. Thus, EST libraries must be sequenced to considerable depth to ensure such contigs can be determined (Parkinson and Blaxter, 2009).

EST sequences should also be free from contaminations occurring from the use of linkers, adaptors and genomic DNA but there is no standard for cleaning EST sequences sets (Nagaraj, Gasser and Ranganathan, 2007). If these sequences are not cleaned from contaminants, the final Unigene collection will harbour the contamination and the assembly of EST sequences will not be optimal. There are chances that the sequence variations due to genomic variation and RNA editing caused from Single nucleotide polymorphism (SNP) might not be identified efficiently due to variation in sequences from sequencing artefacts (Nagaraj Gasser and Ranganathan, 2007).

Also, the cost of EST sequencing has been very high, making it difficult for many research groups to use this technique. Many EST sequences have been deposited in the public domain; however, there is still much data that is held in private databases (Bouck and Vision, 2007). The cost issue may be overcome with the new sequencing technologies such as Illumina or pyrosequencing.

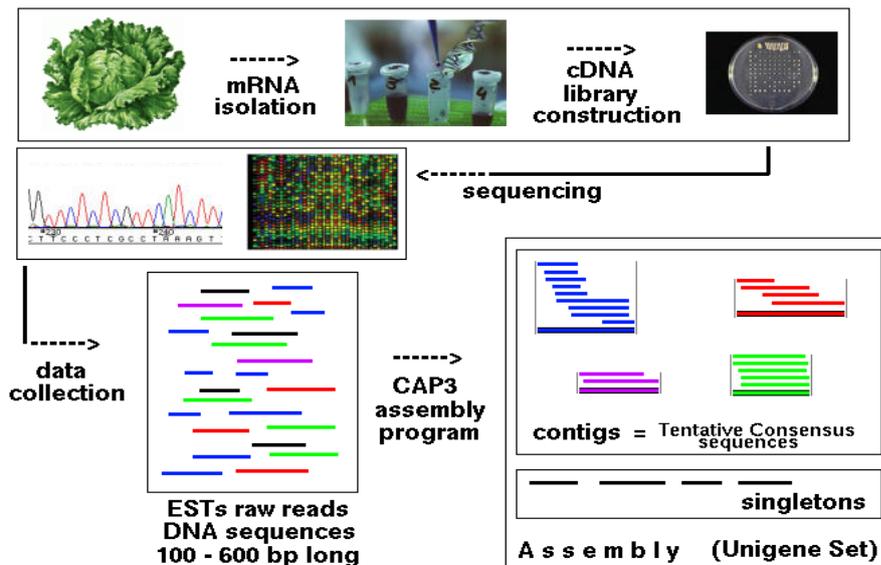


Figure 2. 2: The figure describes sequencing- based approach in understanding transcriptome analysis where, ESTs are generated from the cDNA library and an assembly of contigs is developed (http://www.atgc.org/GP_Ref/presentation/slide_08.html)

2.1.3 Microarray

Microarray technology allows analysis of expression patterns of hundreds to thousands of genes at a time. There is a variety of microarray platforms that have been designed such as glass slides or chips spotted with oligonucleotides representing specific gene coding regions (Figure 2.3). For example, all the genes of an organism may be represented on the slide or chip (Stanton, 2001). Purified RNA from two different slates are fluorescently or radioactively labelled and hybridised to the chip. After thorough washing, the raw data is obtained through laser scanning or by autoradiographic imaging (Huber et al., 2003). The data is then analysed through various statistical methods.

The differences in expression between the two different biological states are compared by the single or merged fluorescence colour of each spot on a single chip (Kauffmann, Gentleman and Huber, 2009). There are a large number of microarray platforms available and experimental design must be carefully considered to assure the expression profiles that are developed are meaningful. Replication and normalisation of individual hybridisation data presents some issues on biological and technical replicates must be carefully considered. Also, standardisation of

platforms, assay protocols and methods of analysis has been difficult. This makes it difficult to compare one experiment to another (Quackenbush, 2002).

The results obtained from microarray analysis need to be verified since possible errors could be incorporated during the manufacture of the chips which might eventually lead to the fidelity of the DNA fragments immobilised to the microarray surface (Huber et al., 2003). Often, RT-qPCR is used as the verification technique for the data obtained from microarray analysis.

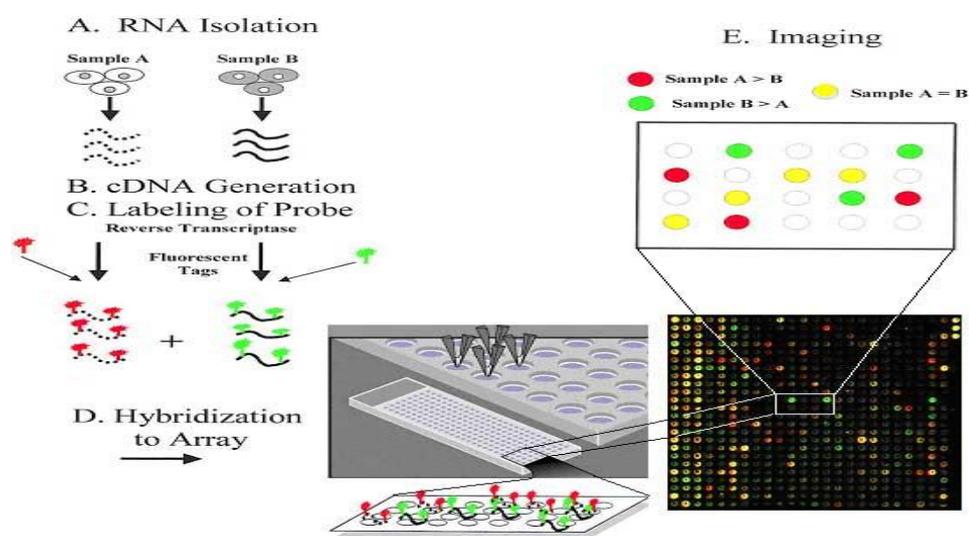


Figure 2. 3: This figure shows various steps i.e. RNA extraction, cDNA synthesis, probe labelling, hybridisation to microarray chip and image detection in the microarray data analysis (www.fastal.com/~renkwitz/microarray-chips.htm).

2.1.4 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

RT-PCR is considered to be a sensitive and semi-quantitative method to detect mRNA levels allowing detection of genes with low expression (Page and Minocha, 2004). First strand cDNA is synthesised from mRNA using reverse transcriptase (RT) enzyme (Figure 2.4). This step is generally primed with an oligo dT primer that will anneal to the poly A tail of the mRNA. Alternatively, cDNA synthesis can be primed with a gene specific primer. The mRNA now bound to the cDNA is degraded with RNase H so that the single stranded cDNA can now act as a template for PCR. The PCR steps include a denaturation step, generally at 94-95°C followed by primer annealing to the cDNA template and then the extension of primer using a

thermostable polymerase enzyme such as *Taq* at 70-75°C. A new DNA strand complementary to the template cDNA strand is synthesised. The now double stranded DNA acts as template for the next cycle. At every PCR cycle, the amount of product is doubled at the end of each cycling (Freeman et al., 1999). Agarose gel electrophoresis is conducted to verify the size and presence of the PCR product.

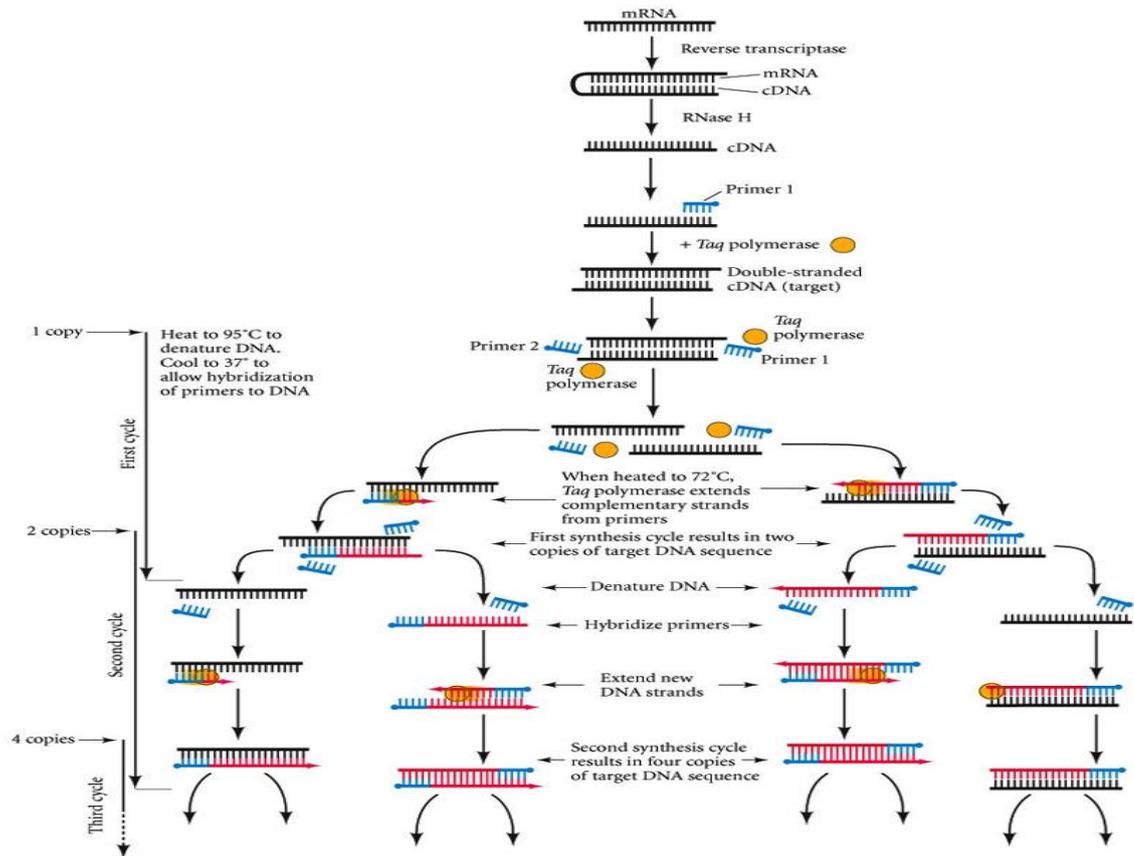


Figure 2. 4: The reverse transcriptase-polymerase chain reaction (RT-PCR) technique to determine the presence of a particular type of mRNA. First, the mRNA is converted to double-stranded cDNA using the enzymes reverse transcriptase and RNase H. The second strand synthesis is completed using thermostable DNA polymerase from *Thermus aquaticus* (*Taq* polymerase). This “target” DNA gets denatured and two sets of primers are added. When *Taq* polymerase is added to the denatured DNA, each strand synthesizes its complement (Gilbert, 2010)

Since it involves the use of cDNA synthesis, DNA amplification, the accuracy of quantification depends on the efficiency across these steps, especially the RT and cDNA amplification (Marone et al., 2001). While RT-PCR can be used to detect if a gene is expressed or not, it is not sufficiently quantitative to allow changes in gene expression to be quantified. It is of interest to know how much mRNA template is present in one sample compared with another; RT-PCR cannot determine this

accurately. If two samples with the same amount of mRNA are used as template, the amount of PCR product produced will vary (Figure 2.5). Thus, this method is semi-quantitative at best (Heid et al., 1996).

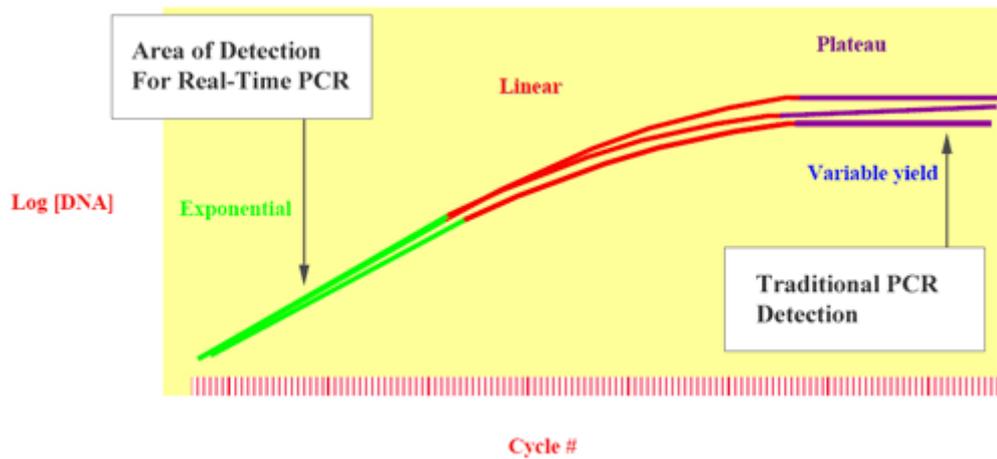


Figure 2. 5: This figure describes variation in the sample DNA amount as the reaction proceeds from exponential phase to plateau phase. Even though the starting DNA amount of the three replicate samples were same initially at the exponential phase, variation in the PCR product was observed for the same set of samples in the plateau phase. This variation could be due to variation in reaction kinetics (<http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/real-time-pcr/real-time-pcr-vs-traditional-pcr.html>).

2.1.5 Reverse Transcriptase- quantitative Polymerase Chain Reaction (RT-qPCR)

Recently, a range of new techniques for mRNA accumulation analysis have been developed such as microarrays and RT-qPCR. Earlier, RT-PCR was used as the most sensitive method for gene expression studies but recently RT-qPCR has become the most widely used method for such studies (Heid et al., 1996). This method allows us to detect the PCR product accumulation as the reaction progresses. The amplicon quantity is proportionate to the initial amount of template which enables the amount of a target sequence to be determined quantitatively (Mestdagh et al., 2009). RT-qPCR has now become the most common, reliable and standard method for in-depth analysis of gene expression due to its tremendous sensitivity, high-sequence specific processing and large dynamic ranges (Becker et al., 2010).

RT-qPCR allows the detection of a target sequence as it is amplified. It includes four main phases namely, linear phase, exponential phase, log-linear phase and the

plateau phase (Figure 2.6). The linear phase is the initial phase when the fluorescence emission does not reach above the background. In the exponential phase, the fluorescence reaches a threshold value, where cycles at which the fluorescence passes the threshold is called the cycle threshold (C_t). The PCR products are achieved in the log-linear phase and in an ideal reaction, the target DNA sequence gets doubled every cycle (Yuan et al., 2006). During the plateau phase, the availability of reaction components gets limited and hence the fluorescence intensity plateaus (Figure 2.6).

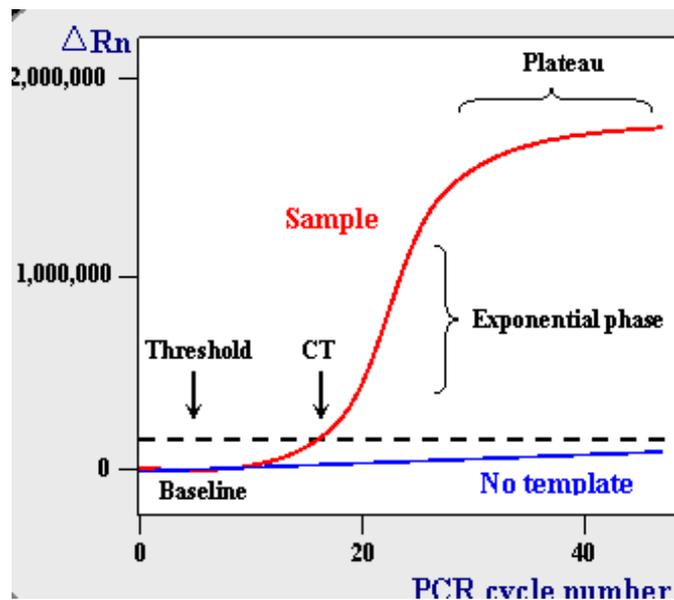


Figure 2. 6: This figure shows the amplification plot of RT-qPCR, where the baseline defines the PCR cycle at which the fluorescent signal emitted by the PCR product is below the limits of detection of the instrument, ΔRn defines the increment in the fluorescent signal as the reaction proceeds. The threshold is an arbitrary line set above the baseline and any signal detected above the threshold line is considered as real signal. C_t represents the PCR cycle number at which the fluorescence is greater than the threshold (<http://www.ncbi.nlm.nih.gov/projects/genome/probe/doc/TechQPCR.shtml>).

The RT-qPCR technique could be a one step or two-step reaction. In a two-step reaction, the extracted mRNA from a sample is converted into cDNA, which is then used in a separate PCR. In a one step reaction, cDNA synthesis and PCR are carried out in one tube (Shaw et al., 2007).

RT-qPCR not only monitors DNA synthesis during the PCR, it can also evaluate the melting point of the product at the end of the amplification. PCR products of a particular primer pair should ideally have similar melting temperatures unless there occurs any contamination, mispriming or primer-dimer artefact in the reaction

(VanGuilder, Vrana and Freeman, 2008). The quantification graph does not differentiate between one DNA and another, therefore melting curve analysis acts as a quality check (Schmittgen and Zakrajsek, 2000).

Different types of RT-qPCR techniques are available such as TaqMan RT-qPCR, molecular beacon and SYBR Green (Taneyhill and Adams, 2008). The TaqMan assay involves the use of a quencher (Q) and a reporter (R) molecule gets attached to the 5' and 3' ends of a probe respectively (Figure 2.7). During the annealing phase the probe binds to the template DNA between the primers, also anneal to the DNA. *Taq* polymerase then adds nucleotides to the 3' end of each primer and uses its 5'→3' exonuclease activity to displace the probe with nucleotides extending the primer. The Q molecule becomes separated from the R and allows the R molecule to emit fluorescence (Nolan, Hands and Bustin, 2006). The amount of fluorescence is recorded. Since the labelled probe is in excess, there is probe to bind to each DNA molecule synthesised. As the reaction proceeds, more PCR product binds more probes which in turn give off more fluorescence as it is degraded. In this way, the amount of fluorescence increases as the reaction proceeds until the plateau phase (Malinen et al., 2003). Because a sequence specific probe is used in addition to the primers, greater specificity can be achieved. Probes that detect single nucleotide polymorphism can be developed so the technique can be used for genotyping as well as quantification (Pfaffl and Berg, 2010).

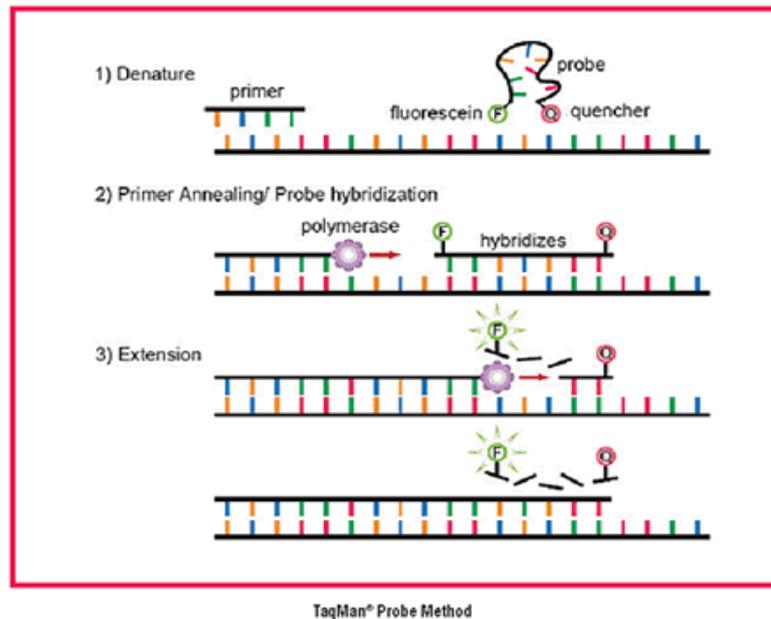


Figure 2. 7: The mechanism of the TaqMan probe RT-qPCR assay (<http://www.foodsafetywatch.com>).

The SYBR Green assay on the other hand exhibits increments in fluorescence intensity upon binding of SYBR Green to double-stranded amplified product (Figure 2.8). Thus, at the end of each PCR cycle SYBR Green binds to the double stranded PCR product (Lekanne Deprez et al., 2002). Upon denaturation at the beginning of the next cycle, the dye falls away from the DNA. Once the next round of amplification is completed, the dye binds to the double stranded DNA again. The amount of fluorescence at the end of each of each cycle is recorded, and this increases as the cycles proceed since more and more product is being synthesised, until the plateau phase is reached. The use of intercalating dyes such as SYBR Green assay has become the most commonly used RT-qPCR method since it is an inexpensive and sequence unspecific method (Pfaffl, 2001).

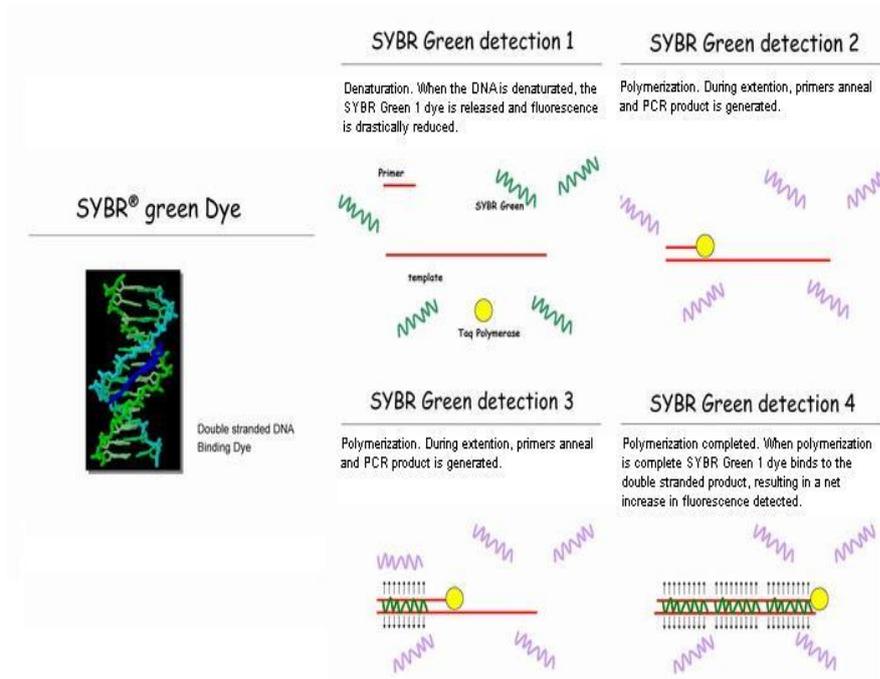


Figure 2. 8: The mechanism of the SYBR Green RT-qPCR assay (<http://www.services.ifom-ieo-campus.it/Pcr/methodology.php>).

Many experiments have used RT-qPCR to quantify the mRNA accumulation variation between two biological groups. For example, the relative amount of a specific mRNA between a healthy and diseased sample can be measured (Kulcheski et al., 2011).

The accuracy of RT-qPCR to measure mRNA accumulation depends on a number of variables such as the amount of starting RNA, RNA quality, amplification efficiency and selection of appropriate reference gene (Ritz et al., 2009). However, no two RNA samples will be exactly the same with regards to RNA quality and concentration. Further, pipetting error means that no two samples will have exactly the same amount of RNA. Many approaches have been used for carrying out RT-qPCR experiments, making it difficult to compare one study to another. In order to standardise RT-qPCR experimental design, guidelines were published recommending minimum information for publication of quantitative real time PCR (MIQE) (Bustin et al., 2009). These guidelines are explained further in detail in chapter 3 section 3.1.1.

Even though RT-qPCR is the most sensitive method for the quantification of gene expression, quantification errors are easily compounded by any of the above

mentioned variables such as RNA quality and reference genes (Mestdagh et al., 2009). If not rectified, such errors can lead to misinterpretation of results and can compromise quantification accuracy (Figure 2.9). These errors can be minimized by normalising the mRNA values to an internal reference RNA (Chang et al., 2010). The amount of reference mRNA can be normalised between samples for the accurate measure of the amount of mRNA expressed from the gene of interest. The normalisation process is also necessary for correcting non-specific variation caused by discrepancy in the initial amount of RNA, the difference in sample quality, nucleic acid recovery efficiency, and variation in cDNA synthesis as well as pipetting errors that can affect the quantification result (Jain, 2009). Hence, selection of appropriate reference genes is critical for RT-qPCR analysis (Coulson et al., 2008).

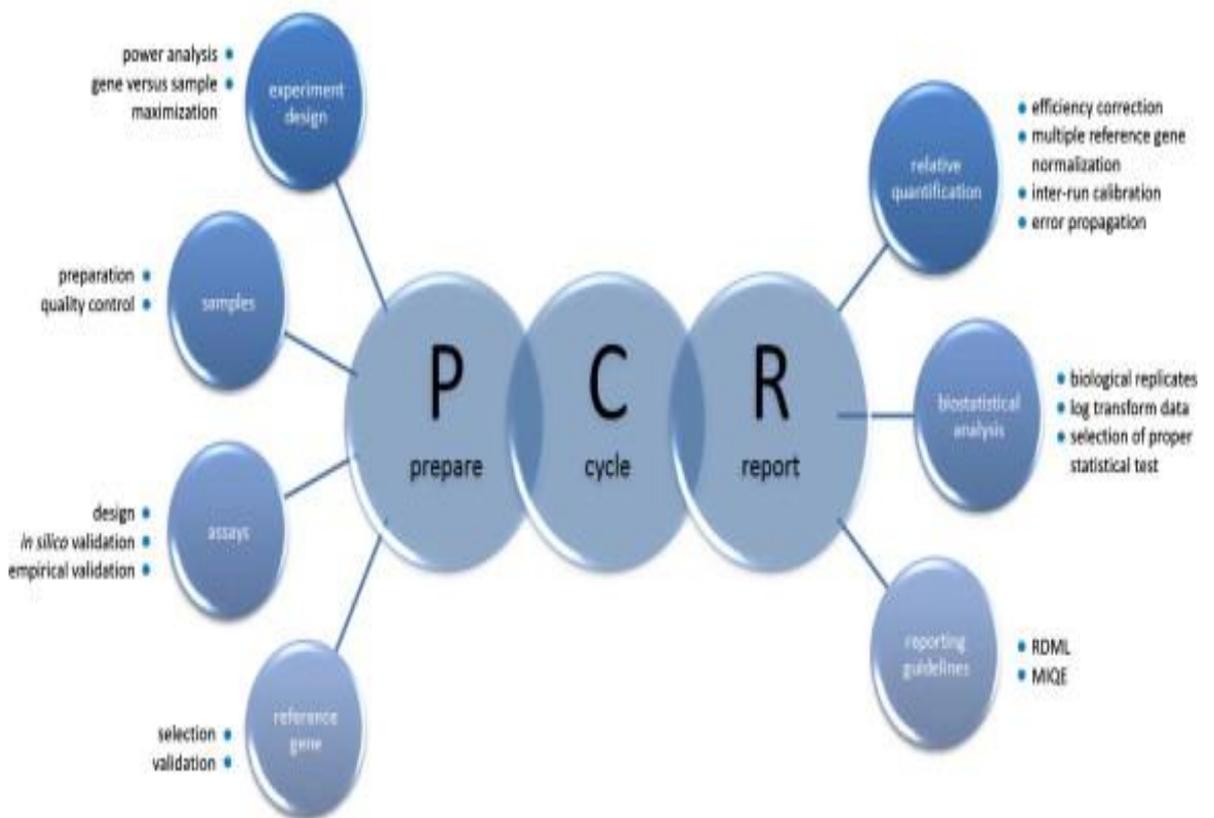


Figure 2. 9: This figure outlines the quality control tools that need to be assessed throughout the entire RT-qPCR experiment (Derveaux, Vandesompele and Hellemans, 2010).

RT-qPCR analysis for the quantification of mRNA accumulation can be done either by absolute or relative analysis methods (Livak and Schmittgen, 2001). The absolute

quantification method determines the input copy number of the transcript by relating the PCR signal to a standard curve of serially diluted standards of known concentration. The absolute quantification is usually used to correlate the viral copy number with a disease (Bustin, 2000). The relative quantification method determines the amount of a target sequence relative to some reference group such as internal control or reference gene (Stankovic and Corfas, 2003). The relative quantification method is used to measure accumulation of mRNA in response to a treatment and is considered to be an adequate method to analyse gene expression changes. The relative quantification method is also much faster and less expensive as compared to absolute quantification method (Livak and Schmittgen, 2001). In this study, the relative quantification method was used.

2.1.6 Reference genes for relative quantification

A reference gene can be defined as a gene that has stable expression among various conditions such as developmental stages and under different experimental treatments. Also, it should not be associated with any pseudogenes in order to avoid amplification of secondary products or genomic DNA (Tong et al., 2009). The reference gene is presumed to be a representation of the cDNA concentration and quality in each sample and subject to the same errors during cDNA preparation as the target genes (Moreno, Gruissem and Vanderschuren, 2011).

Housekeeping genes have traditionally been used as single reference genes in plant sciences. It has been assumed that housekeeping genes have stable expression in the cell and have therefore been accepted for normalisation without the need for any experimental evidence for their stability (Taylor et al., 2010). More recently, the need to validate the stability of reference genes has been recognised (Huggett et al, 2005). A reference gene is now defined as a gene that has been experimentally found to be stably expressed in a given species under a given experimental condition and thus considered suitable for quantitative analysis under that specific condition (Kulcheski et al., 2010).

Some of the most frequently used reference genes for gene expression analysis include housekeeping genes such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ubiquitin (UBQ), actin, cyclophilins and 18S ribosomal RNA (rRNA) which are involved in basic cellular processes of plants (Cordoba et al., 2011).

Even though the above mentioned housekeeping genes have been assumed to have uniform expression, the transcript levels of these genes have been found to differ considerably under various experimental conditions, hence making them unsuitable as reference gene for gene expression analysis (Radonic et al., 2004). For example, it has been found that many of the frequently used housekeeping genes such as GAPDH, ubiquitin, TUB, actin and 18S rRNA under various stress and growth-arresting conditions in *A. thaliana* showed unstable expression levels. This suggests that these genes are not suitable for normalizing gene expression in plants such as *A. thaliana* when sensitive technique such as RT-qPCR was used (Czechowski et al., 2005).

However, many studies have been conducted to evaluate the uniform expression of different housekeeping genes under specific environmental conditions and in different plant species (Jain et al., 2006). One such example assessed the stability of ten most commonly used housekeeping genes such as UBQ5, UBQ10, GAPDH, TUB and 18s rRNA were tested in rice seedlings grown under various environmental conditions of stress and hormone treatment. Among them EF1 α and UBQ5 were identified as appropriate reference genes with the most uniform expression levels (Lovdal and Lillo, 2009).

Another recent study conducted by Lilly et al., (2011) assessed the stability of candidate reference genes in virus infected *A. thaliana*. Twelve candidate reference genes were assessed in plants infected with TuMV including F-box family protein (F-box), SAND, PDF2, Actin, TUB, UBQ10 and EF1 α . Among these genes tested, four of them, namely F-box, SAND, PDF2 and EF1 α showed the most stable accumulation against all other twelve genes tested. It was suggested that the genes identified by Lilly et al., (2011) could be used as reference genes for the analysis of all virus infected plants, although this would need to be tested in other species.

The aim of this part of the study was therefore to test the expression stability of the four reference genes: EF1 α ; F-box; SAND; and PDF2 in healthy taro and *N. benthamiana* using RT-qPCR.

2.2 Materials and Methods

2.2.1 Total RNA extraction

Total RNA was extracted from the leaf tissue of either healthy or virus infected taro (*C. esculenta*), *N. benthamiana* and *A. thaliana* using a Spectrum Plant Total RNA Kit (Sigma Aldrich) following the manufacturer's instructions. Approximately 100 mg of leaf tissue from three plants each were used for extraction. These tissues were ground in liquid nitrogen to form a powder. Lysis buffer containing 2-mercaptoethanol (500 μ l) was added to 100 mg of tissue powder and vortexed vigorously for at least 30 seconds. The samples were then incubated at 56°C for 3-5 minutes followed by a centrifugation at 14,000 rpm for 3 minutes. The lysate supernatant was transferred to a filtration column which had been placed in a 2 ml collection tube. The tubes were then centrifuged at 14,000 rpm for 1 minute and the clarified flow-through lysate was saved. Binding solution (500 μ l) was added to the clarified lysate and mixed thoroughly by pipetting at least five times. The 700 μ l of the reaction mixture was transferred to a binding column seated in a 2 ml collection tube. The samples were then centrifuged at 14,000 rpm for 1 minute. The flow-through liquid was discarded and the column was returned to the collection tube. This step was repeated with the remaining lysate. Wash solution 1 (300 μ l) was transferred into the binding column and centrifuged at 14,000 rpm for 1 minute. The flow-through liquid was decanted and the column was placed back into the collection tube.

On-column DNase treatment was also carried out to remove genomic DNA. For each digestion, 10 μ l of DNase 1 with 70 μ l of DNase digestion buffer was mixed together and all 80 μ l was transferred to the centre of the filter inside the binding column. This was incubated at room temperature for 15 minutes.

Wash solution 1 (500 μ l) was added again to the binding column and centrifuged at 14,000 rpm for 1 minute. The flow-through liquid was poured out and the column was returned to the collection tube. Wash solution 2 (500 μ l) containing ethanol was pipetted to the column and centrifuged at 14,000 rpm for 30 seconds. The flow-

through liquid was decanted and the column was placed back into the collection tube. Wash solution 2 (500 μ l) was added again into the column and centrifuged at 14,000 rpm for 30 seconds. The flow-through liquid was again decanted and the column was returned to the collection tube. The column was dried by centrifuging the tubes at 14,000 rpm for 1 minute. The column was then placed into a new 2 ml collection tube and 50 μ l of elution solution was added into the column. The tubes were incubated at room temperature for 1 minute followed by centrifugation at 14,000rpm for 1 minute. Purified total RNA was present in the flow-through eluate which was stored in aliquots at -80°C .

The absorbance at 260 nm and 280 nm using a NanoVue Spectrophotometer were used to measure the RNA concentration and purity, respectively. The quality of RNA was verified by means of agarose gel electrophoresis of approximately 800 ng of each sample. A 1% agarose gel was run in 1X TAE buffer for 90 minutes at 65 V. A 100 bp molecular weight marker (Dnature) was used for comparison.

2.2.2 DNase treatment

The extraction of total RNA was followed by a further DNase treatment to eliminate potential genomic DNA in the sample. Total RNA (400 ng) was combined with 1 μ l of 10 X DNase1 reaction buffer and 1 μ l of DNase1, Amp Grade (Sigma Aldrich) and 10 μ l of DNase and RNase free water. The mixture was incubated at room temperature for 15 minutes after which 0.5 μ l of 25 mM EDTA was added. The reaction was then stopped by at 65°C for 10 minutes.

2.2.3 cDNA synthesis

First strand cDNA was synthesised from DNase treated total RNA using a qScript Flex cDNA Synthesis Kit (Quanta Biosciences). The reaction was made up of 400 ng of total RNA, 1 μ l of nuclease free water, 2 μ l of GSP Enhancer and 2 μ l of VdT primer which were incubated at 65°C for 5 minutes followed by an immediate transfer to 42°C . 5 X qScript Flex Reaction Mix (4 μ l) along with 1 μ l of qScript Reverse Transcriptase was added to make the final volume to 20 μ l after which the

reaction were incubated as follows: 60 minutes at 42°C, 85°C for 5 minutes and final hold at 4°C. The cDNA was then stored at -20°C.

2.2.4 Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)

PCR was carried out on first strand cDNA with each reaction containing 6.25 µl of 5 X Go Taq Green master mix (Promega), 1 µl of cDNA, 0.7 µl each of 10 µM forward and reverse primers (Table 2.1). Each reaction was made upto 12.5 µl with a no template control reaction was made as above with the cDNA replaced by nuclease free water. PCR was carried out under the following conditions: initial denaturation at 94°C for 5 minutes followed by 35 cycles with denaturation at 94°C for 30 seconds then annealing at 60°C for 30 seconds and extension at 72°C for 30 seconds with a final extension at 72°C for 5 minutes.

PCR reaction (12.5 µl) was run on a 2% agarose gel with 1X TBE at 70 V for 75 minutes using 350 ng of marker 100 bp DNA ladder (Dnature) in order to confirm amplification.

2.2.5 Gradient PCR

Temperature gradient PCR was carried out to optimise the primer annealing temperature. The reactions were set up as described in section 2.2.4. Gradient PCR was carried out under the following conditions: initial denaturation at 94°C for 5 minutes followed by 35 cycles with denaturation at 94°C for 30 seconds then annealing at a temperature range of 40°C to 60°C for 30 seconds and extension at 72°C for 30 seconds with a final extension at 72°C for 5 minutes.

Amplification was verified as described in section 2.2.4.

2.2.6 Primer design

Primers designed by Lilly et al., (2011) for *A. thaliana* EF1 α , F-box, PDF2 and SAND genes (Table 2.1) were tested for their ability to amplify their target genes in *A. thaliana*, *N. benthamiana* and taro. RT-PCR conditions used were those described

in section 2.2.4 with an annealing temperature of 60°C as used by Lilly et al., (2011).

Table 2. 1: Primer sequences, melting temperature (T_m) and expected sizes for EF1 α , PDF2, SAND and F-box mRNA from *A. thaliana*.

Gene	Species	Forward primer	T_m (°C)	Reverse primer	T_m (°C)	Product size
EF1 α	<i>A. thaliana</i>	5'CACCACTGGAG GTTTTGAGG3'	70	5'TGGAGTATTTG GGGGTGGT3'	68	137 bp
PDF2	<i>A. thaliana</i>	5'TCATTCCGATA GTCGACCAAG3'	69	5'AAAGGCGATTC ACCTCACATA3'	65	104 bp
SAND	<i>A. thaliana</i>	5'GTTGGGTCACA CCAGATTGT3'	69	5'GCTCCTTGCAA GAACACTTCA3'	69	127 bp
F-box	<i>A. thaliana</i>	5'GGCTGAGAGG TTCGAGTGTT3'	70	5'GGCTGTTGCAT GACTGAAGA3'	68	140 bp

Primers targeting EF1 α and F-box genes from monocot and dicots were designed. The nucleotide sequence of each Arabidopsis gene was used as a query sequence in a BLASTn search (Johnson et al., 2008) against the GenBank database held at the National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov). Monocot and dicot sequences with E value of 0, a maximum score of 500 and 70% sequence identity were downloaded. Regions of similarity between the selected sequences and the Arabidopsis query sequences were identified by multiple sequence alignment using ClustalW (Thompson et al., 1994) in Geneious Pro 5.4.4 (Drummond et al, 2011). The Arabidopsis primer sequences were redesigned. New primer sequences were designed close to the binding sites for the Arabidopsis primers. Sequences were chosen as primers that targeted very closely the regions targeted by Lilly's primers. Where possible, primers were designed that were conserved between monocots and dicots and where it was not possible to achieve such a design, primers specific to monocot or dicots were designed. One pair of primers, called eEF1 α MDF and eEF1 α MDR, was designed that was expected to amplify the eEF1 α mRNA from both monocot and dicot plant species. For F-Box, two primer pairs were designed, one for amplifying the F-Box sequence from dicots (F-BoxDF/F-BoxDR) and one from monocots ((F-BoxMF/F-BoxMR). BLASTn analysis of each new primer sequence verified their specificity. These primers are shown on p78.

2.2.7 Conformation of PCR products by cloning and sequencing

2.2.7.1 PCR product purification

DNA Binding buffer (90 μ l) was added to each of the RT-PCR products obtained from taro, *N. benthamiana* and *A. thaliana* using EF1 α MDF/EF1 α MDR, F-boxMF/F-boxMR or F-boxDF/F-boxDR described in section 2.2.4). The samples were then loaded into 2 ml tubes containing ZymoSpin column (from Zymo Research) followed by centrifugation at 15,000 rpm for 30 seconds. Wash buffer (200 μ l) was pipetted into the columns and centrifuged at 15,000 rpm for 30 seconds. This step repeated and the columns were transferred into new 1.5 ml tubes. Nuclease free water (10 μ l) was added into the column. The tubes were then incubated at room temperature for 1 minute followed by centrifugation at 15,000 rpm for 1 minute. Purified DNA was present in the eluate.

The absorbance at 230 nm and 260 nm using a NanoVue Spectrophotometer were used to measure the DNA concentration.

2.2.7.2 Ligation

The RT-PCR products obtained from taro, *N. benthamiana* and *A. thaliana* using EF1 α MD, F-boxM, and F-boxD primer pairs (as described in section 2.2.4) were used to set up the ligation reactions. The reaction set up included standard reaction, positive control and background control. All these reaction set up included 5 μ l of 2 X Rapid ligation buffer (from Promega), 1 μ l of 50 ng pGEM-T Easy vector, 1 μ l of T4 DNA Ligase (3 Weiss units/ μ l), and 7.5 ng of purified RT-PCR product. 2 μ l of control insert DNA was added into the positive control. Each reaction was made up to 10 μ l using nuclease free water. These reactions were then incubated for an hour at room temperature.

2.2.7.3 Transformation

2 µl of each of the ligation reactions were transferred into fresh 1.5 ml tubes kept on ice and a control reaction was set up using 0.1 ng of uncut plasmid. Further, 50 µl of *E.coli* JM109 high efficiency competent cells from Promega, was added into each of these ligation reactions and 100 µl of cells were added into the control tube with uncut plasmid. All these tubes were then kept at 42°C for 45-50 seconds, followed by an immediate incubation on ice for 2 minutes. Super optimal broth with catabolite repression (SOC) medium kept at room temperature was added (950 µl) into each of the ligation reactions and 900 µl of the SOC medium was added into the control tube. These tubes were then incubated at 37°C for 1.5 hours with constant shaking at ~150 rpm. 100 µl of each of the ligation reactions was then plated onto duplicate LB/Ampicillin/Isopropyl β- D- 1 thiogalactopyranoside (IPTG)/ 5- bromo- 4- chloro- indolyl- β- D- galactopyranoside (X-Gal) (with 0.5 mM IPTG and 100 µg/ml ampicillin) plates along with the plating out of 1:10 dilution of the uncut DNA with the SOC medium. All these plates were then incubated overnight at 37°C to allow the cells to grow.

2.2.7.3 Colony PCR and sequencing

White colonies were selected and colony PCR on each of those selected colonies was done. Each colony was plated onto a master LB/Ampicillin/IPTG/X-Gal plate with a yellow pipette tip. Plates were then incubated overnight at 37°C and then placed at 4°C. Remaining cell on the tip were diluted in 5 µl of sterile water which was used as template along with 1.4 µl each of 10 µM forward and reverse primers (T7 and SP6, respectively) and 12.5 µl of 5 X Go Taq Green master mix (Promega). Each reaction was made up to 25 µl with nuclease free water. A no template control reaction was made as above with the template replaced by nuclease free water. PCR was carried out under the following conditions: initial denaturation at 95°C for 5 minutes followed by 35 cycles with denaturation at 95°C for 1 minute then annealing at 55°C for 1 minute and extension at 68°C for 1 minute with a final extension at 68°C for 10 minutes. Amplification was verified as described in section 2.2.4.

All the selected sequences were then sent for sequencing to Waikato DNA Sequencing Facility, University of Waikato, Hamilton. The sequencing results are presented in appendix 2.

2.3 Results

2.3.1 RNA Quality

The quality of the RNA is essential to the overall success of the gene expression analysis. Spectrophotometry analysis revealed that the A_{260}/A_{280} absorbance ratio of all RNA samples ranged from 2-2.2 indicating that the RNA was free of contaminating protein, DNA and other polyphenolic substances.

The integrity of total RNA extracted from the leaf tissue of taro, *N. benthamiana* and *A. thaliana* was assessed by electrophoresis (Figure 2.10). It shows the presence of the 23S and 18S rRNA bands with the background smear of mRNA. This pattern was typical of all RNA extracted and indicated the RNA was intact and suitable for RT-PCR.

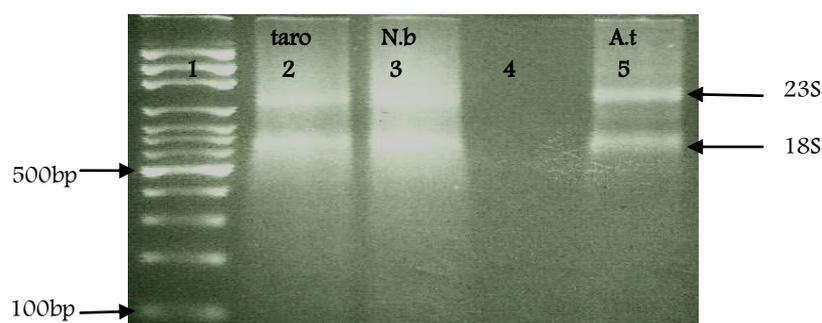


Figure 2. 10: Agarose gel electrophoresis of total RNA extracted from the leaf tissue of taro (Lane 2), *N. benthamiana* (Lane 3) and *A. thaliana* (Lane 5). Lane 1 contained 350 ng of 100 bp marker. Lane 4 contained no sample.

2.3.2 RT-PCR of using primers designed by Lilly et al., (2011)

The primers designed by Lilly et al., (2011) for the Arabidopsis genes $EF1\alpha$, F-box, PDF2 and SAND were tested for their ability to amplify products from *N. benthamiana* and taro.

2.3.2.1 RT-PCR of EF1 α mRNA

RT-PCR amplification of EF1 α mRNA was carried out by using EF1 α F and EF1 α R primers and at an annealing temperature of 60°C as recommended by Lilly et al., (2011). The primers were expected to amplify a product of 137 bp.

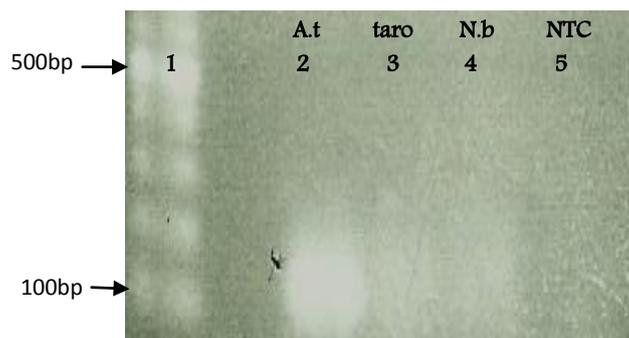


Figure 2. 11: RT-PCR using EF1 α F and EF1 α R primers on *A. thaliana* (Lane 2), taro (Lane 3), *N. benthamiana* (Lane 4) and a no template control (NTC), (Lane 5). Lane 1 contained 350 ng of 100 bp marker.

Absence of bands for the no template control suggested no contamination in the experimental set up (Figure 2.11, Lane 5). A product of the expected size was amplified from *A. thaliana* (Figure 2.11, Lane 2).

Faint bands of the expected size were seen for taro and *N. benthamiana* (Figure 2.11, Lane 3 and Lane 4, respectively). This suggested further optimisation of the PCR conditions would be needed to improve the amplification of EF1 α primers in taro and *N. benthamiana*. The annealing temperature of 60°C was probably too high for the Arabidopsis primers to anneal efficiently in other plant species.

2.3.2.2 RT-PCR of the F-box mRNA

RT-PCR amplification of F-box mRNA was carried out by using F-boxF and F-boxR primers at an annealing temperature of 60°C as recommended by Lilly et al., (2011). The primers were expected to amplify a product of 140 bp.

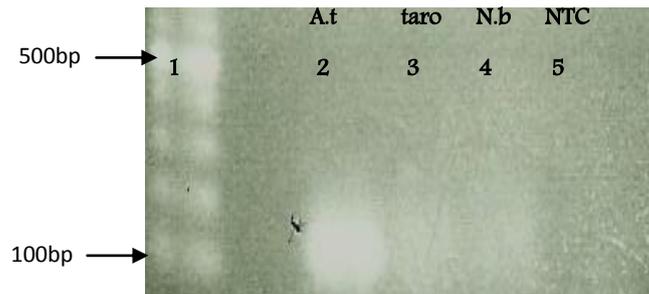


Figure 2. 12: RT-PCR using F-boxF and F-boxR primers on *A. thaliana* (Lane 2), taro (Lane 3), *N. benthamiana* (Lane 4) and a no template control (NTC) (Lane 5). Lane1 contained 350 ng of 100 bp marker.

Absence of bands for the no template control suggested no contamination in the experimental set up (Figure 2.12, Lane 5). A product of the expected size was amplified from *A. thaliana* (Figure 2.12, Lane 2).

Faint bands of the expected size were seen for taro and *N. benthamiana* (Figure 2.12, Lane 3 and Lane 4, respectively). This suggested further optimisation of annealing temperature to improve the amplification of F-box primers in taro and *N. benthamiana*. Again, the annealing temperature of 60°C may have been too high for these primers to work efficiently in different species.

2.3.2.3 RT-PCR of the PDF2 mRNA

RT-PCR amplification of PDF2 mRNA was carried out by using PDF2F and PDF2R primers at an annealing temperature of 60°C as recommended by Lilly et al., (2011). The primers were expected to amplify a product of 104 bp.

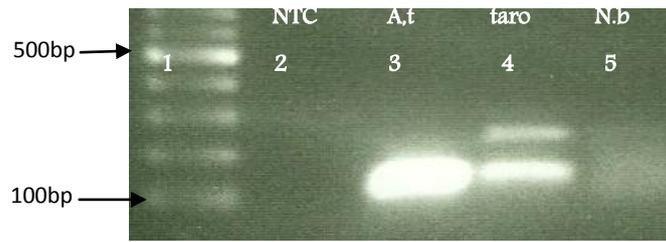


Figure 2. 13: RT-PCR using PDF2F and PDF2R primers with a no template control (NTC) (Lane 2), on *A. thaliana* (Lane 3), taro (Lane 4), *N. benthamiana* (Lane 5). Lane1 contained 350 ng of 100 bp marker.

Absence of bands with no template control indicates no contamination in the experimental set up (Figure 2.13, Lane 2).

PDF2 gene was efficiently amplified only from the template *A. thaliana* giving a product of the expected size (Figure 2.13, Lane 3). Very faint bands of 104 bp were observed for taro and *N. benthamiana* templates (Figure 2.13, Lane 4 and Lane 5, respectively). The presence of a non-specific secondary band was also seen for taro (Figure 2.13, Lane 4) suggesting inefficient amplification of either a non-specific band or contaminating gDNA.

The reduced amount of product suggested that the amplification efficiency from taro and *N. benthamiana* was reduced compared with *A. thaliana*. This may have been due to inappropriate annealing temperature being used. The annealing temperature of 60°C was quite high for amplification from these species (Hwang et al., 2003).

2.3.2.4 RT-PCR for the SAND mRNA

RT-PCR amplification of SAND mRNA was carried out by using SANDF and SANDR primers at an annealing temperature of 60°C as recommended by Lilly et al., (2011). The primers were expected to amplify a product of 127 bp.



Figure 2. 14: RT-PCR using SANDF and SANDR primers on *A. thaliana* (Lane 2), taro (Lane 3), *N. benthamiana* (Lane 4) and a no template control (NTC) (Lane 5). Lane 1 contained 350 ng of 100 bp marker.

Absence of bands for the no template control suggested no contamination in the experimental set up (Figure 2.14, Lane 5). A product of the expected size was amplified from *A. thaliana* (Figure 2.14, Lane 2).

Faint bands of the expected size were seen for taro and *N. benthamiana* (Figure 2.14, Lane 3 and Lane 4, respectively). This suggested further optimisation of the annealing temperature was needed to improve the amplification by SAND primers in taro and *N. benthamiana*.

2.3.3 Optimising Annealing Temperature

Since the primers were designed against *A. thaliana* genes the reduced amplification of similar sized fragments from taro and *N. benthamiana* suggested that the PCR conditions needed to be optimised. This is normally achieved by altering various parameters such as template concentration, primer concentration, reducing PCR cycles and by gradually increasing or decreasing annealing temperature. Among these parameters, selection of the optimum annealing temperature is the most critical component for improving specificity of a PCR reaction (Rychlik, Spencer and Rhoades, 1990). Particularly, as in this case, the primers were designed against Arabidopsis and being used in different species to amplify the orthologous mRNAs. In this situation, the stringency of annealing generally needs to be reduced to allow primer binding (Ishii and Fukui, 2001).

Temperature gradient PCR was carried out to empirically determine an optimal annealing temperature for each primer set using the *N. benthamiana* and taro

templates. Amplification of the *A. thaliana* template was carried out at 60°C as a positive control. From the gradient PCR analysis, optimum annealing temperatures for each of the gene was selected and RT-PCR at their respective chosen temperature was conducted to verify the results obtained from gradient PCR.

2.3.3.1 Gradient PCR amplification of the EF1 α mRNA

The total volume of gradient PCR reaction (12.5 μ l) was run on a 2% agarose gel with 1X TBE at 70 V for 75 minutes to verify the results obtained from gradient PCR.

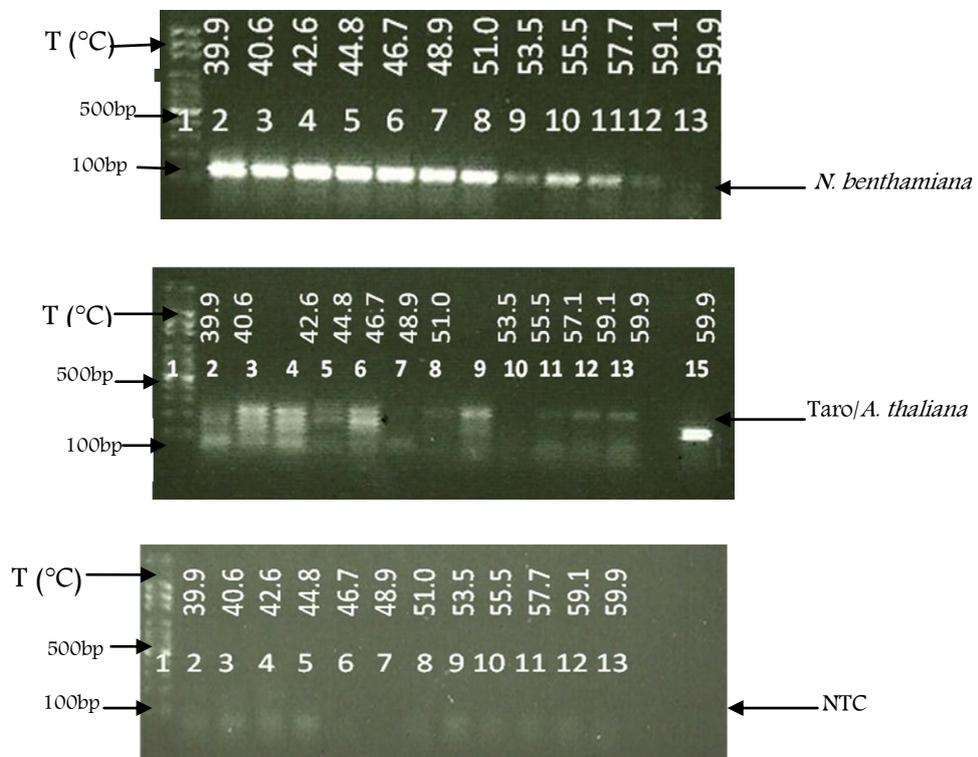


Figure 2. 15: Gradient PCR using EF1 α F and EF1 α R primers designed by Lilly et al., (2011) on *N. benthamiana*, taro and *A. thaliana* (Lanes 2-13, Lanes 2-13 and Lane 15, respectively) and no template control (NTC) (Lanes 2-13). The temperatures tested were 39.9°C (Lane 2), 40.6°C (Lane 3), 42.6°C (Lane 4), 44.8°C (Lane 5), 46.7°C (Lane 6), 48.9°C (Lane 7), 51.0°C (Lane 8), 53.5°C (Lane 9), 55.5°C (Lane 10), 57.7°C (Lane 11), 59.1°C (Lane 12) and 59.9°C (Lane 15).

Absence of bands with no template control indicates no contamination in the experimental set up (Figure 2.15, Lanes 2-15). Amplification of the expected product of 137 bp product for EF1 α was observed with the *A. thaliana* template (Figure 2.15, Lane 15) indicating the reaction set up were suitable for amplification.

Amplification of a 137 bp product was also observed for the *N. benthamiana* template (Figure 2.15, Lane 3-9) at each temperature tested between 40°-60°C with less product amplified at higher temperatures. This indicated that the optimum temperature was 40-50°C.

In contrast, multiple products were observed for taro at all temperatures with inconsistent amplification, although these were reduced at 57.7°-59.1°C (Figure 2.15, Lanes 11-13). No temperature tested appeared optimal for taro since non-specific secondary bands were observed with at most temperatures (Figure 2.15, Lanes 2-13). Thus the non-specific bands observed in taro were likely amplification products from non-specific binding of the Arabidopsis primers to the taro cDNA. It could not be assumed as a genomic DNA since a DNase treatment was done to all the taro RNA samples also, a +RT and –RT RT-PCR was conducted to verify the presence of any genomic DNA in the sample (data not shown).

From the gradient PCR analysis, the annealing temperature of 47°C (Lanes 6) was selected to test further. RT-PCR was conducted for each template to verify the results obtained from gradient PCR. Figure 2.16, shows that a 137 bp size product was amplified from *A. thaliana* at this temperature.

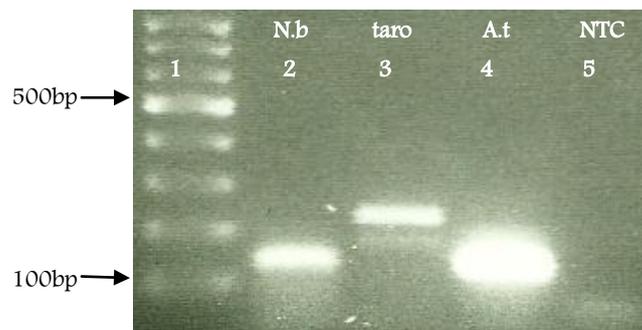


Figure 2. 16: RT-PCR at 47°C using EF1 α F and EF1 α R primers on *N. benthamiana* (Lane 2), taro (Lane 3), *A. thaliana* (Lane 4), and a no template control (Lane 5). Lane 1 contained 350 ng of 100 bp marker.

Figure 2.16 showed that amplification of the expected 137 bp product occurred for *A. thaliana* (Lane 4) and *N. benthamiana* (Lane 2). However for taro (Lane 3), products of 200 bp and 237 bp were amplified, suggesting the primers have not annealed as expected. The no template control showed no amplification as expected

(Lane 5). Thus, the Arabidopsis EF1 α primers appear suitable for use with *N. benthamiana* at 47°C, but are unsuitable for use with taro.

2.3.3.2 Gradient PCR amplification of the PDF2 mRNA

The expected product of 104 bp for PDF2 was observed for the *A. thaliana* template (Figure 2.17, Lane 15) indicating the reaction set up was suitable for amplification. Amplification of a 104 bp product was also observed for the *N. benthamiana* template (Figure 2.17, Lanes 2-13 upper panel) at each temperature tested between 40°-60°C. However, a secondary non-specific product was also amplified at all temperatures. In contrast, the expected sized product was observed for taro (Figure 2.17, Lanes 3-14 middle panel) at all temperatures tested between 40-50°C with efficient amplification. This band was not sharp and there may have been more than one product that has not resolved well in this gel. Absence of bands with the no template control indicated no contamination in the experimental set up (Figure 2.17, Lanes 2-15 lower panel). The optimum annealing temperature of 49°C was selected and tested.

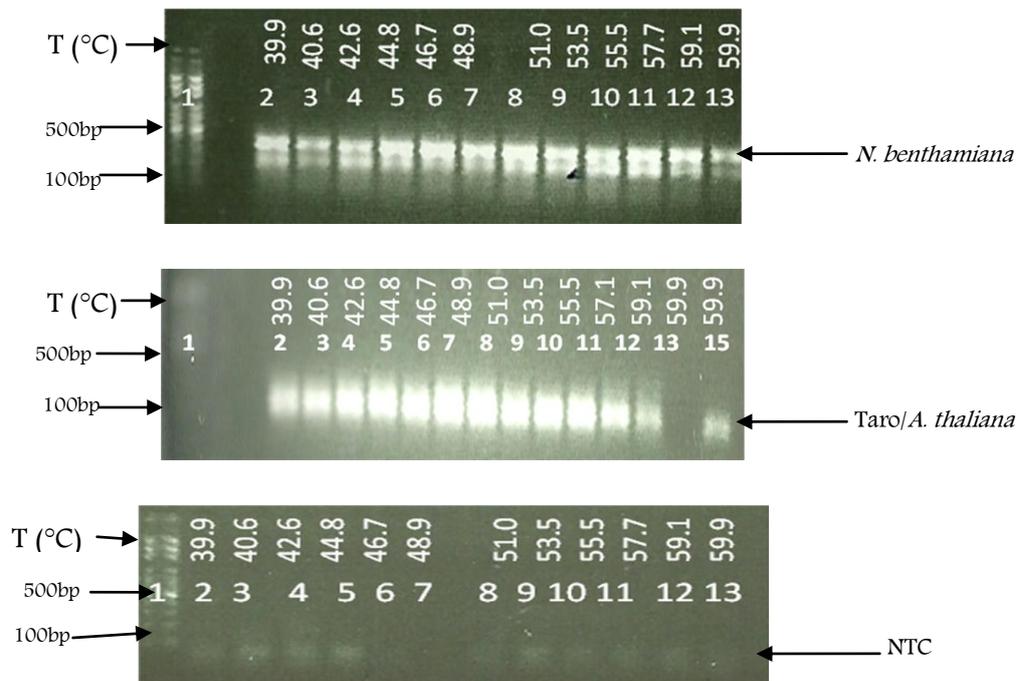


Figure 2. 17: Gradient PCR using PDF2 primers PDF2F and PDF2R designed by Lilly et al., (2011) on *N. benthamiana*, taro, *A. thaliana* (upper panel Lanes 2-13, Lanes 2-13 middle panel and Lane 15 middle panel, respectively) and no template control (NTC) (Lane 2). The temperatures tested were 39.9°C (Lane 2), 40.6°C (Lane 3), 42.6°C (Lane 4), 44.8°C (Lane 5), 46.7°C (Lane 6), 48.9°C (Lane 7), 51.0°C (Lane 8), 53.5°C (Lane 9), 55.5°C (Lane 10), 57.7°C (Lane 11), 59.1°C (Lane 12) and 59.9°C (Lane 15).

Amplification at 49°C gave rise to the expected 104 bp product from *A. thaliana* (Figure 2.18, Lane 4 middle panel).

Absence of bands for the no template control suggested no contamination in the experimental set up (Figure 2.18, Lane 2). However, secondary bands along with a band of the expected size were seen for taro and *N. benthamiana* (Figure 2.18, Lane 3 and Lane 5 upper panel, respectively). This suggesting the primer annealed non-specifically to the cDNA indicated optimisation of annealing temperature to improve the amplification of PDF2 primers in taro and *N. benthamiana*. Thus, the Arabidopsis PDF2 primers do not appear useful under these conditions for use with *N. benthamiana* and taro.

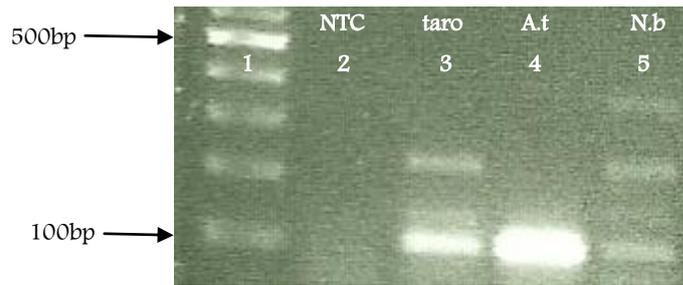


Figure 2. 18: RT-PCR using PDF2F and PDF2R primers with a no template control (Lane 2), on taro (Lane3), *A. thaliana* (Lane 4), and *N. benthamiana* (Lane 5). Lane1 contained 350 ng of 100 bp marker.

2.3.3.3 Gradient PCR amplification of the SAND mRNA

The expected product of 127 bp band was observed with *A. thaliana* template (Figure 2.19, Lane 15) indicating the reaction set up were suitable for amplification. Amplification of a 127 bp product was also observed for taro template (Figure 2.19, Lanes 2-14) at each temperature tested between 40°-60°C with secondary products amplified at some temperatures, particularly at higher temperatures. In contrast, multiple products were observed for *N. benthamiana* at 40-47°C temperature with no amplification at higher temperatures (Figure 2.19, Lanes 2-14). For both taro and *N. benthamiana* no temperature tested appeared optimal. Non-specific secondary bands were also observed with taro template (Figure 2.15, Lane 10-13). Absence of bands with no template control indicated no contamination in the experimental set up (Figure 2.15, Lane 2-13). Thus non-specific bands arose from non-specific annealing of the Arabidopsis primers to the taro and *N. benthamiana* cDNA templates.

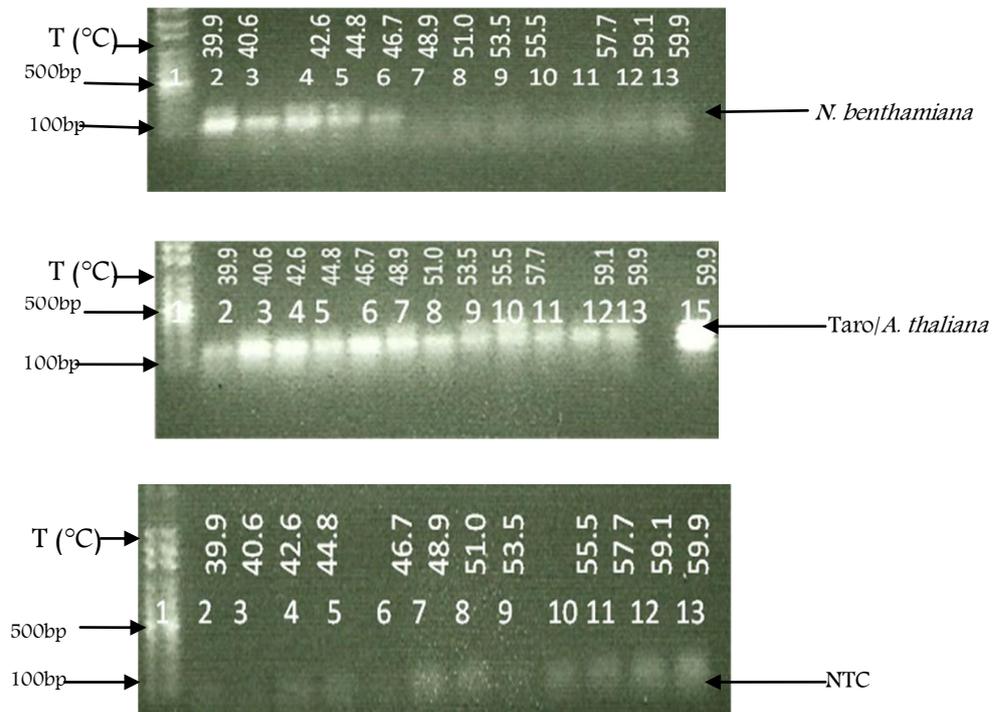


Figure 2. 19: Gradient PCR using SANDF and SANDR designed by Lilly et al., (2011) on *N. benthamiana*, taro, *A. thaliana* (Lanes 2-14, Lanes 2-13 and Lane 15, respectively) and no template control (Lane 2-13). The temperatures tested were 39.9°C (Lane 2), 40.6°C (Lane 3), 42.6°C (Lane 4), 44.8°C (Lane 5), 46.7°C (Lane 6), 48.9°C (Lane 7), 51.0°C (Lane 8), 53.5°C (Lane 9), 55.5°C (Lane 10), 57.7°C (Lane 11), 59.1°C (Lane 12) and 59.9°C (Lane 15).

While no one temperature was identified as optimal for amplification of SAND, 42°C was selected as a compromised temperature between amplification of the specific and non-specific products and a follow up PCR was carried out at 42°C that is a relatively low annealing temperature to confirm the gradient PCR findings.

A product of the expected 127 bp for SAND was amplified at 42°C from *A. thaliana* (Figure 2.20, Lane 4). Faint bands of the expected size were seen for both taro and *N. benthamiana* (Figure 2.20, Lane 3 and Lane 2 respectively). A secondary band (hard to see in figure but it was observed on original photo) of 236 bp was also seen with *N. benthamiana* (Figure 2.20, Lane 2). Absence of bands for the no template control suggested no contamination in the experimental set up (Figure 2.20, Lane 5). This suggested further optimisation of annealing temperature to improve the amplification of SAND primers in taro and *N. benthamiana*. Under the PCR conditions used, no annealing temperature appeared suitable for the use of Arabidopsis SAND primers

with taro and *N. benthamiana*. Other PCR parameters such as magnesium and primer concentration would need to be tested to determine the suitability of these primers.

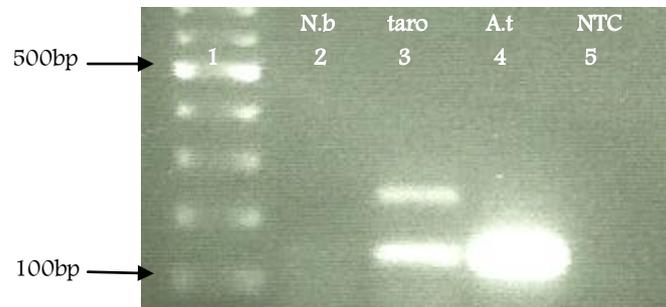


Figure 2. 20: RT-PCR using SANDF and SANDR primers with on *N. benthamiana* (Lane 2), taro (Lane 3), *A. thaliana* (Lane 4) and a no template control (Lane 5). Lane 1 contained 350 ng of 100 bp marker.

2.3.3.4 Gradient PCR amplification of F-box mRNA

The expected product of 140 bp band was observed with *A. thaliana* template (Figure 2.21, Lane 15) indicating the reaction set up was suitable for amplification. Amplification of a 140 bp product was also observed for taro template (Figure 2.21, Lanes 3-9) along with non-specific product or primer-dimer. Multiple products were observed for *N. benthamiana* at all temperatures with inconsistent amplification. No temperature tested appeared optimal. Non-specific secondary bands observed with taro template (Figure 2.21, Lanes 5-14) were likely amplification products from non-specific binding of the Arabidopsis primers to the taro cDNA. Absence of bands with no template control indicated no contamination in the experimental set up (Figure 2.21, Lanes 2-15). Thus, extra products were the result of non-specific binding of the primers to the template.

The total volume of gradient PCR reaction (12.5 μ l) was run on a 2% agarose gel with 1X TBE at 70 V for 75 minutes using 350ng of marker i.e. 100 bp DNA ladder to verify the results obtained from gradient PCR.

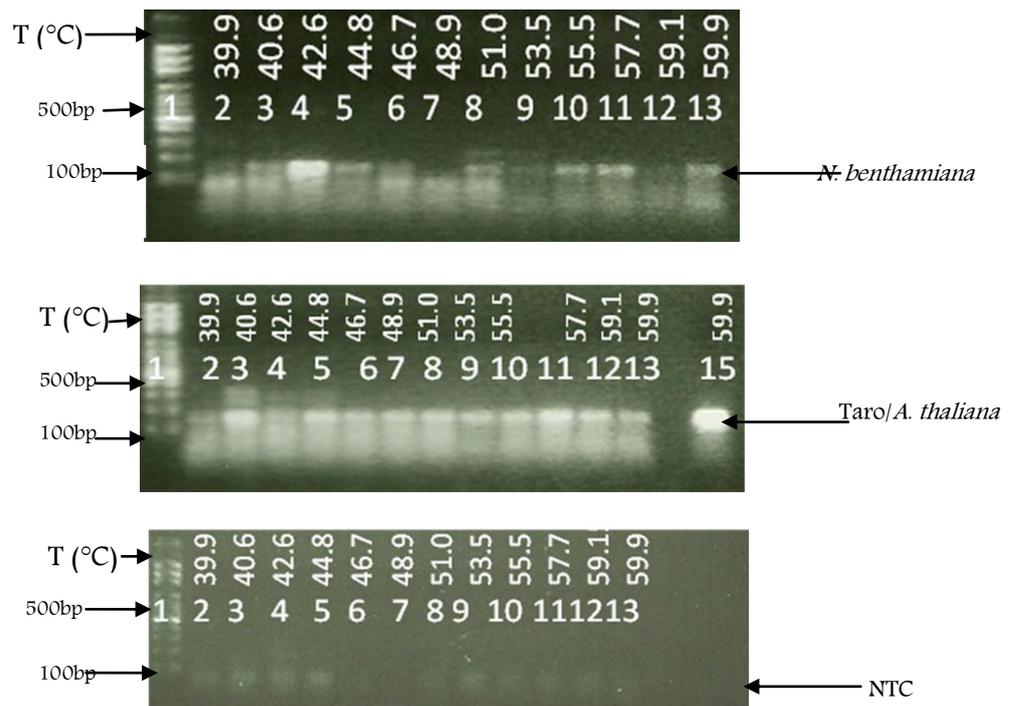


Figure 2. 21: Gradient PCR using F-box primers F-boxF and F-boxR designed by Lilly et al., (2011) on *N. benthamiana*, taro, *A. thaliana* (Lanes 2-13, Lanes 2-13 and Lane 15, respectively) and no template control (Lanes 2-15). The temperatures tested were 39.9°C (Lane 2), 40.6° C (Lane 3), 42.6°C (Lane 4), 44.8°C (Lane 5), 46.7°C (Lane 6), 48.9°C (Lane 7), 51.0°C (Lane 8), 53.5°C (Lane 9), 55.5°C (Lane 10), 57.7°C (Lane 11), 59.1°C (Lane 12) and 59.9°C (Lane 15).

From the gradient PCR analysis, annealing temperature of 42°C and 58°C were selected and for further testing. Around 42°C appeared optimal for amplification from *N.bethamiana* while 58°C appeared optimal for taro.

Amplification at 42°C and 58°C gave rise to the expected 140 bp product from *A. thaliana* (Figure 2.22, Lane 4). Faint bands of the expected size were seen for taro and *N. benthamiana* (Figure 2.22, Lane 3 and Lane 2, respectively). Absence of bands for the no template control suggested no contamination in the experimental set up (Figure 2.22, Lane 5). This suggested that the primers annealed specifically, but not as efficiently for Arabidopsis. Further optimisation of the annealing temperatures may improve the amplification of F-box primers in taro and *N. benthamiana*.

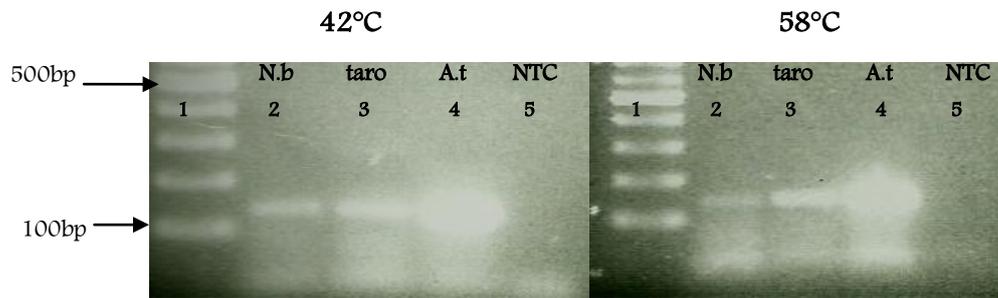


Figure 2. 22: RT-PCR using F-boxF and F-boxR primers on *N. benthamiana* (Lane 2), taro (Lane 3), *A. thaliana* (Lane 4), and a no template control (Lane 5). Lane1 contained 350 ng of 100 bp marker.

2.3.4 Primer design

The above analysis showed that the primers for each of the four candidate reference genes as designed by Lilly et al, (2011) are not useful for amplifying these genes from all species. These primers gave inconsistent results when tested on taro and *N. benthamiana*. If these genes are to be used as reference genes for all virus infection studies, it would be useful to have primers that can either bind to these genes from all plants (i.e. monocots and dicots) or to all dicots or all monocots. Thus, attempts were made to design such universal primers for each of these four reference genes.

A BLASTn search was conducted using each of the *A. thaliana* reference genes as the query sequence. Based on parameters such as the identity percentage of more than 70%, homologues from monocot and dicot species sequences were selected. All sequences including the Arabidopsis query sequence were then aligned using Geneious 5.4.4 (Drummond et al., 2011). Regions of homology close to the expected annealing sites for Lilly et als primers were chosen for primer design.

2.3.4.1 Primer design for the EF1 α gene

The binding sequences for the Arabidopsis EF1 α forward and reverse primers were compared with the corresponding sequences from each of the downloaded homologues. Homologous sequences were identified from the dicots such as *Solanum lycopersicum*, *N. tabacum* and the monocots such as *O. sativa* and *Zea mays*.

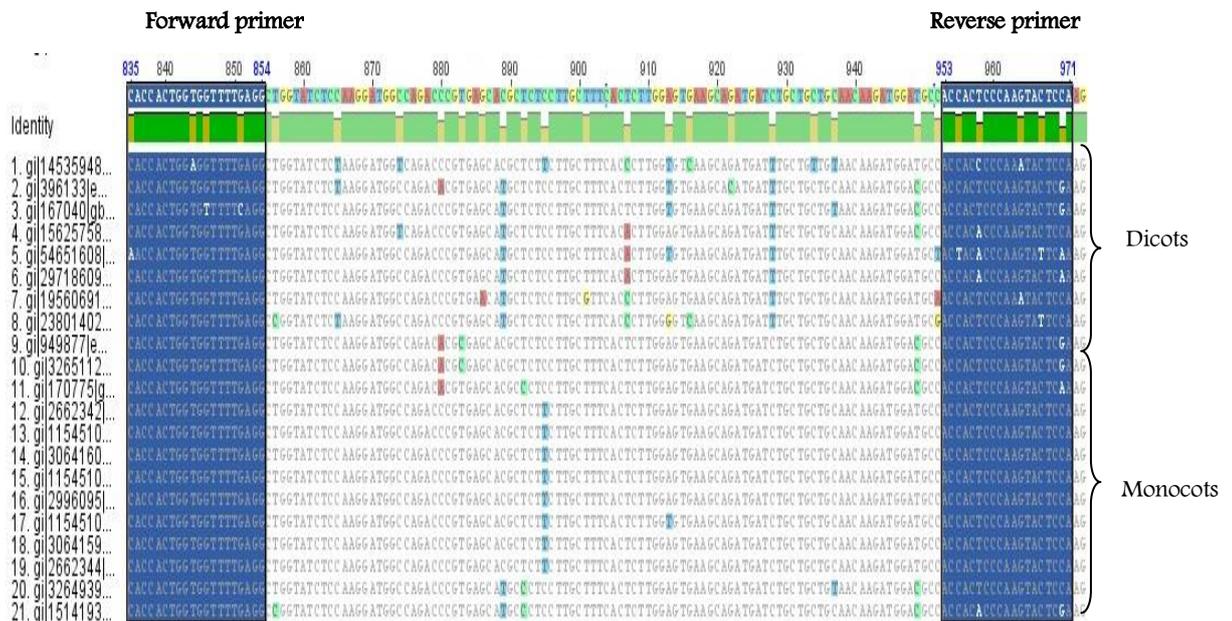


Figure 2. 23: The alignment of monocot and dicot species against *A. thaliana* in the region amplified by the Arabidopsis EF1 α primer sequences. The alignment of selected dicot and monocot species against Arabidopsis in the region amplified by the EF1 α primers designed by Lilly et al., (2011). Arabidopsis is the top sequence, followed by other dicot and monocot species as indicated. The blue highlighted regions show the homology between all species with the EF1 α F (left) and EF1 α R (right) annealing sites. Other highlighted nucleotides show where each species diverges from others.

The region where the 5' EF1 α F primer was expected to bind showed high conservation between monocots and dicots (Figure 2.23). In contrast, the binding region of the 3' EF1 α R primer showed most sequence variation amongst dicots. Amplification with EF1 α F and EF1 α R primers gave the expected product from *N. benthamiana* but not taro. From the alignments in figure 2.23, the reverse would have been expected since there was greater sequence divergence in the EF1 α R primer annealing site in dicots than monocots. No sequence for taro EF1 α was available for comparison, rather only *Poaceae* sequences were available. It is possible that monocots from other families may have varying sequences for one or both EF1 α primers.

For *N. benthamiana*, annotated sequence for EF1 α was not available at the time of writing; however, sequence from the close relative *N. tabacum* was included. In the alignment in figure 2.23, *N. tabacum* sequence is shown directly underneath the

Arabidopsis sequence. It can be seen that these regions showed high homology with Arabidopsis, thus, EF1 α was amplified from *N. benthamiana*.

2.3.4.2 Primer design for the PDF2 gene

The binding sequences for the Arabidopsis PDF2 forward and reverse primers were compared with the corresponding sequences from each of the downloaded homologues. Homologous sequences were identified from the dicots such as *Solanum lycopersicum*, *Populus trichocarpa* and the monocots such as *Sorghum bicolor*, *O.sativa* (Figure 2.2).

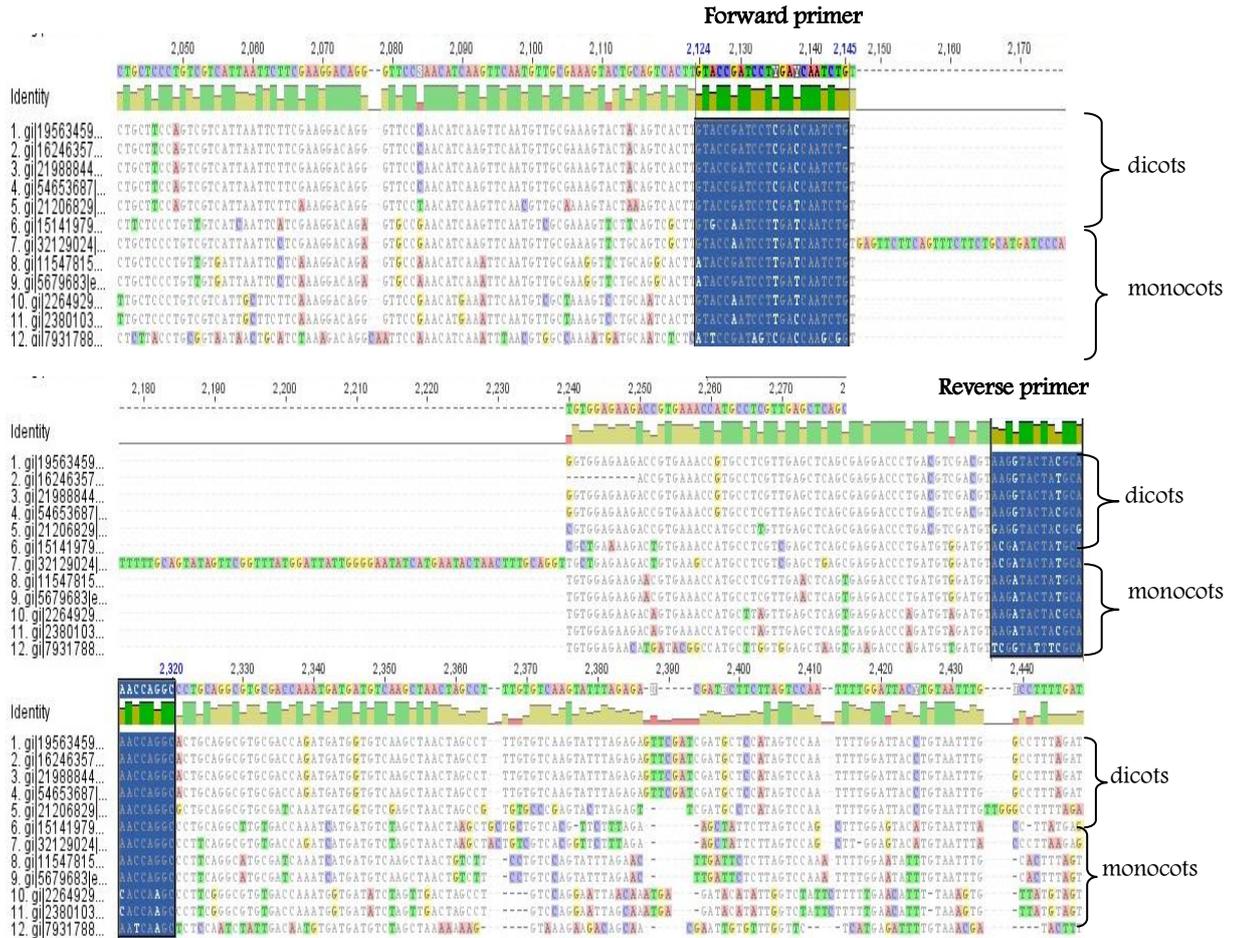


Figure 2. 24: The alignment of monocot and dicot species against *A. thaliana* with region amplified by the Arabidopsis PDF2 primer sequences. The alignment of selected dicot and monocot species against Arabidopsis in the region amplified by the PDF2 primers designed by Lilly et al., (2011). Arabidopsis is the top sequence, followed by other dicot and monocot species as indicated. The blue highlighted regions show the homology between all species with the PDF2F (left) and PDF2R (right) annealing sites. Other highlighted nucleotides show where each sequence diverges from the others.

Amplification with PDF2F and PDF2R primers gave the expected size product from *A. thaliana* but not for *N. benthamiana* and taro. There were non-specific secondary products observed for both taro and *N. benthamiana*. From the alignments in figure 2.24, efficient amplification would have been expected for *N. benthamiana* since there was greater sequence similarity in the PDF2F and PDF2R primer annealing site in dicots than monocots. No sequence for taro PDF2 was available for comparison, rather only *Poaceae* sequences were available. It is possible that monocots from other families may have varying sequences for one or both PDF2 primers.

For *N. benthamiana*, annotated sequence for PDF2 was not available at the time of writing; however, sequence from the close relative *Solanaceae* was included. In the

alignment in figure 2.24, *S. lycopersicum* sequence is shown directly underneath the Arabidopsis sequence. It can be seen that these regions showed high homology with Arabidopsis towards 3' PDF2F primer annealing site but not much similarity at the 3' PDF2R primer annealing site, this may account for the inefficient amplification observed in *N. benthamiana* using PDF2 primer pair.

New primer sequences need to be designed towards more conserved regions of monocot and dicot sequences to attain efficient amplification using RT-PCR; however, because of the sequence variation at either side of primer binding sites along with the lack of conservation among monocot in this region, it became impossible to design a primer that would work on both monocots and dicots.

2.3.4.3 Primer design for the SAND gene

The binding sequences for the Arabidopsis SAND forward and reverse primers were compared with the corresponding sequences from each of the downloaded homologues. Few homologous sequences were identified in GenBank. Homologous sequences were identified from the dicots such as *P. trichocarpa*, *S. lycopersicum* and the monocots such as *Zea mays* and *O. sativa*.

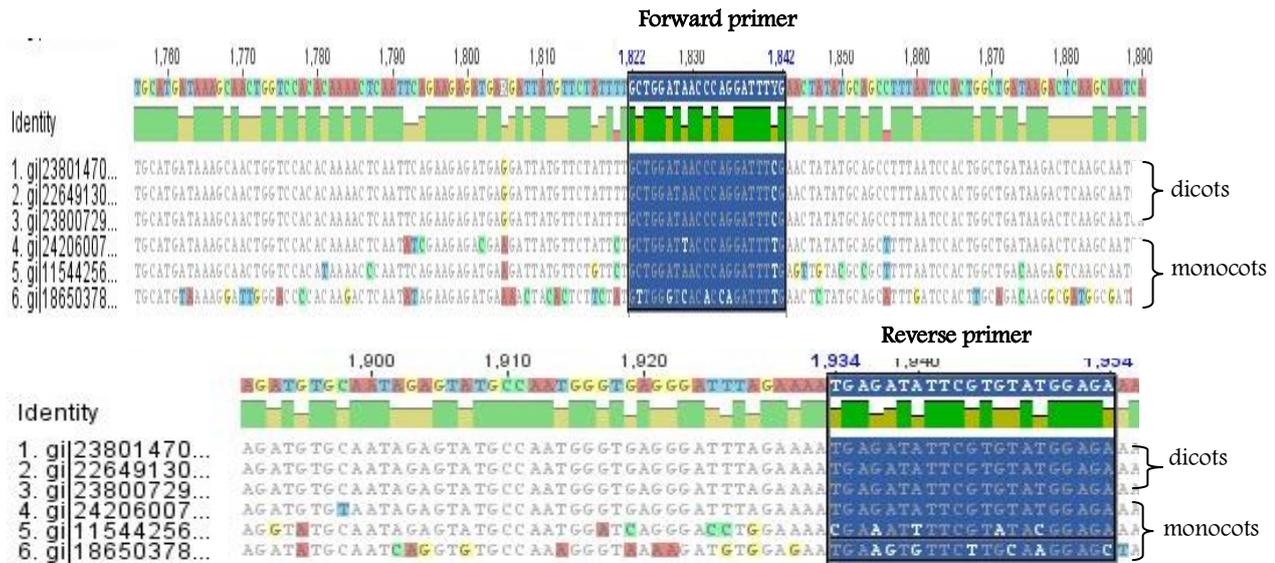


Figure 2. 25: The alignment of monocot and dicot species against *A. thaliana* with region amplified by the Arabidopsis SAND primer sequences. The alignment of selected dicot and monocot species against Arabidopsis in the region amplified by the SAND primers designed by Lilly et al., (2011). Arabidopsis is the top sequence, followed by other dicot and monocot species as indicated. The blue highlighted regions show the homology between all species with the SANDF (left) and SANDR (right) annealing sites. Other highlighted nucleotides show where each sequence diverges from the others.

Amplification with SANDF and SANDR primers gave the expected size product from *N. benthamiana* but not for taro. There were non-specific secondary products observed for both taro and *N. benthamiana*. From the alignments in figure 2.25, efficient amplification would have been expected for *N. benthamiana* since there was greater sequence similarity in the PDF2R primer annealing site in dicots than monocots. No sequence for taro SAND was available for comparison, rather only *Poaceae* sequences were available. It is possible that monocots from other families may have varying sequences for one or both SAND primers.

For *N. benthamiana*, annotated sequence for SAND was not available at the time of writing; however, sequence from the close relative *Solanaceae* was included. In the alignment in figure 2.25, *S. lycopersicum* sequence is shown directly underneath the Arabidopsis sequence. It can be seen that these regions showed high homology with Arabidopsis 3' SANDR primer annealing site, this may account for the inefficient amplification observed in *N. benthamiana* using SAND primer pair.

The alignment of both monocot and dicot sequences against the *A. thaliana* SAND sequence at the region of forward and reverse primer sequences showed these regions to have greater variation with monocot than dicot species.

This analysis suggested that the primer sequences for the SAND gene may not be optimum sequences for the amplification of SAND from in both taro and *N. benthamiana* as well as from both monocot and dicot species. However, due to the limited availability of both monocot and dicot homologues it was difficult to predict primer sequences for efficient amplification of SAND gene from both monocot and dicot species. Figure 2.25, shows that the monocot sequences is quite divergent around these primer binding sites. Further, the position of the intron in Arabidopsis and the lack of conservation among monocots in this region made it impossible to design a primer that would work on monocot and dicots. Thus, design of new primer was not presumed. It can be seen; however, that dicot sequences show greater conservation, therefore dicot primers may be designed.

2.3.4.4 Primer design for the F-box gene

The binding sequences for the Arabidopsis F-box forward and reverse primers were compared with the corresponding sequences from each of the downloaded homologues. Homologous sequences were identified from dicots such as *Sorghum bicolor*, *P. thrichocarpa* and the monocots such as *O. sativa* and *Zea mays*.

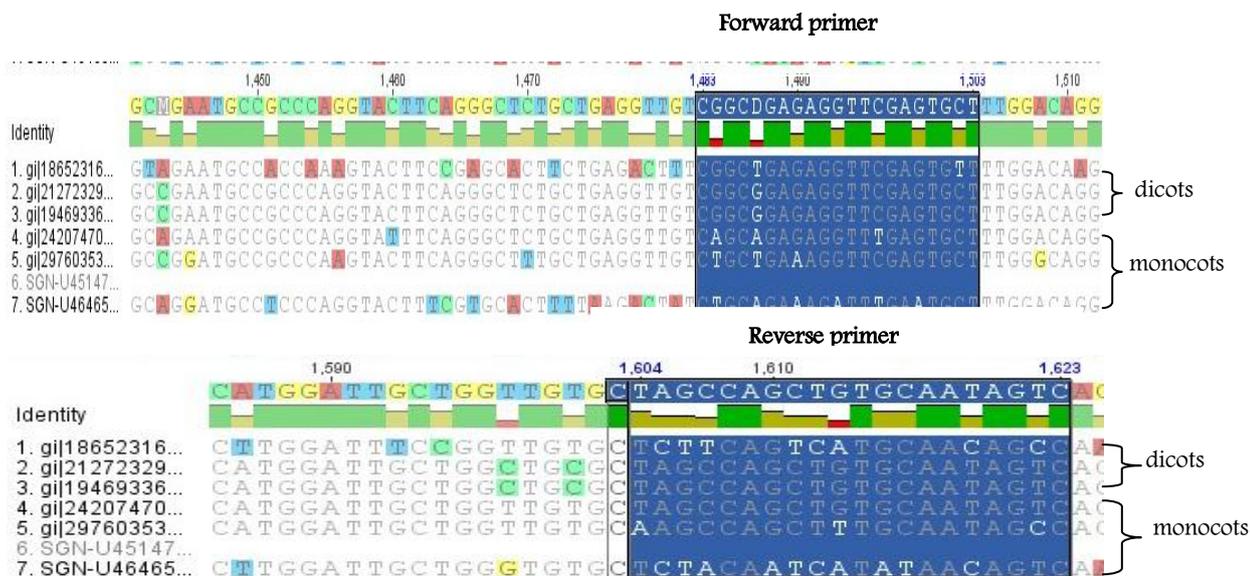


Figure 2. 26: The alignment of monocot and dicot species against *A. thaliana* with region amplified by the Arabidopsis F-box primer sequences. The alignment of selected dicot and monocot species against Arabidopsis in the region amplified by the F-box primers designed by Lilly et al., (2011). Arabidopsis is the top sequence, followed by other dicot and monocot species as indicated. The blue highlighted regions show the homology between all species with the F-boxF (left) and F-boxR (right) annealing sites. Other highlighted nucleotides show where each sequence diverges from the others.

The sequence alignment analysis showed that the primer sequences designed by Lilly et al., (2011) for amplification of genes *EF1 α* , *F-box*, *SAND* and *PDF2* provided insights as to why they were largely unsuitable for amplification from other species. This was primarily due to variation at the 3' end of the reverse primer sequence when compared to the reference sequences of *A. thaliana* for each of these genes. These results supported the RT-PCR results on taro and *N. benthamiana* using these primers which showed inefficient amplification of these genes along with non-specific secondary bands. Thus, new primer sequences would need to be developed for amplification of these genes from other host species.

The alignment of monocot sequences against the *A. thaliana* F-box sequence at the region of forward and reverse primer sequence depicts variation along the length of each region particularly in monocot species (Figure 2.26). In particular, the 3' end of the F-boxR primer annealing site shows variation. While a band of the expected size was amplified, this was not efficient since less product was amplified from approximately the same amount of cDNA. While the Arabidopsis primers appeared to anneal approximately, this annealing was not optimal, suggesting the need for new primers.

2.3.5 New primer design

New primer sequences were designed for the amplification of the EF1 α and F-box reference genes from both monocot and dicot species (Table 2.2). The conserved regions suitable for primer design were identified from the nucleotide sequence alignments of monocot and dicot species sequences against the reference gene sequence from *A. thaliana* (Figure 2.27 and 2.28). Primers could be designed for EF1 α (EF1 α MDF and MDR) that were predicted to amplify the EF1 α gene from both monocot and dicot species (Figure 2.27), whereas, for F-box dicot specific (F-boxDF and DR) and monocot (F-boxMF and MR) specific primers could be designed (Figure 2.28 and 2.29, respectively). There were many criteria that were taken into consideration while designing new primer such as the nucleotide length was kept in the range between 16-20 base pairs, the GC content was kept below 60%, primers were checked for the formation of complementarity, the product length was tried to keep below 300 bp for it to be used efficiently in RT-qPCR analysis and the melting temperatures for forward and reverse primers were kept close. The new primer sequences were designed close to the primer sequences designed by Lilly et al (2011) for the amplification of genes EF1 α and F-box from *A. thaliana*, by excluding the intron spanning regions.

Table 2. 2: The table below shows the new primer sequence to amplify both EF1 α and F-box gene from both monocot and dicot species.

Gene	Species	Forward primer	Tm (°C)	Reverse primer	Tm (°C)	Product size (bp)
EF1 α	Monocots and Dicots	5'ATCATTGACTCC ACCACTGG 3'	56.7	5'GCCTTGGAG TACTTGGG 3'	52.1	151
F-boxM	Monocots (M)	5'GAGTGCTTTGGA CAGGA 3'	50.7	5'CCAGCAATC CATGACCA 3'	55.4	100
F-boxD	Dicots (D)	5'TGAATGCTTTGG ACAGG 3'	51.8	5'CGGCAATCC AAGACC 3'	51.9	100

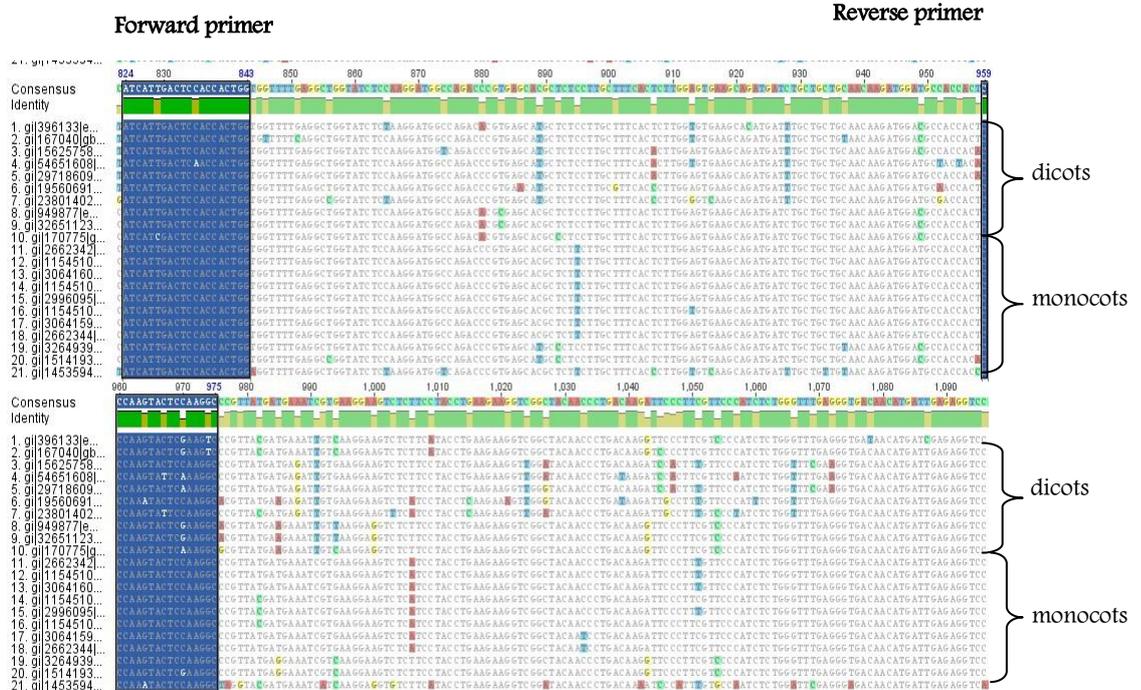


Figure 2. 27: The alignment of monocot and dicot species against *A. thaliana* with region amplified by the Arabidopsis EF1 α MD primer sequences. Arabidopsis is the top sequence, followed by other dicot and monocot species as indicated. The blue highlighted regions show the homology between all species with the EF1 α MDF (left) and EF1 α MDR (right) annealing sites. Other highlighted nucleotides show where each sequence diverges from the others.

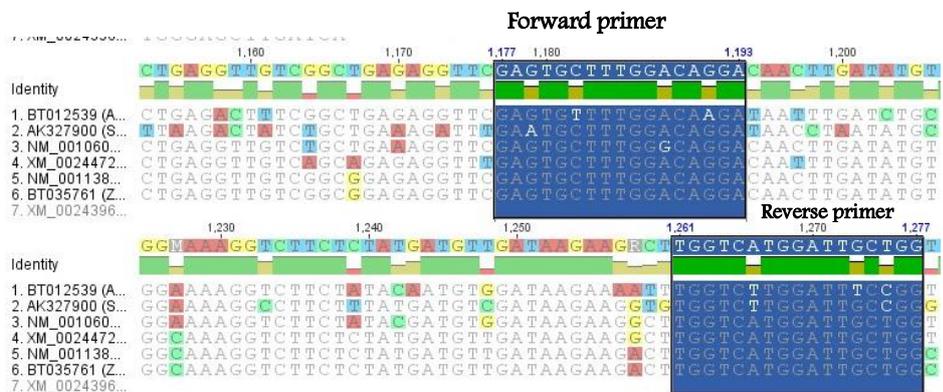


Figure 2. 28: The alignment of monocot species against *A. thaliana* with region amplified by the Arabidopsis F-boxM primer sequences. Arabidopsis is the top sequence, followed by other monocot species as indicated. The blue highlighted regions show the homology between all species with the F-boxF (left) and F-boxR (right) annealing sites. Other highlighted nucleotides show where each sequence diverges from the others.

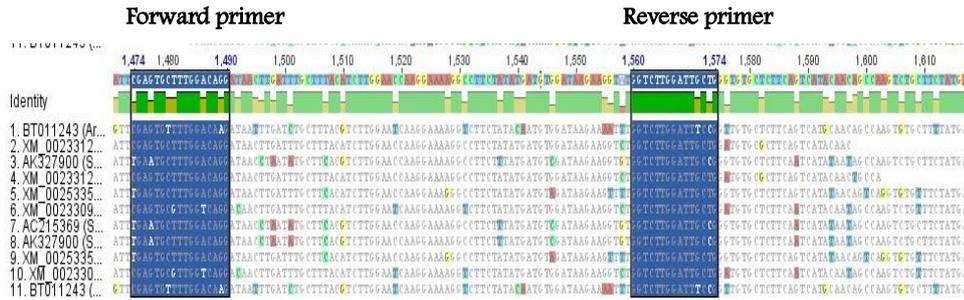


Figure 2. 29: The alignment of dicot species against *A. thaliana* with region amplified by the Arabidopsis F-boxD primer sequences. Arabidopsis is the top sequence, followed by other monocot species as indicated. The blue highlighted regions show the homology between all species with the F-boxDF (left) and F-boxDR (right) annealing sites. Other highlighted nucleotides show where each sequence diverges from the others.

2.3.6 Testing new primer sequences

2.3.6.1 Analysis of new EF1 α primers

Temperature gradient PCR was carried out to empirically determine the optimal annealing temperature to amplify EF1 α gene using the new primers. The EF1 α MDF and EF1 α MDR primers were expected to amplify a 151 bp product from both monocot and dicot plant species. From the gradient PCR analysis, optimum annealing temperature was selected and RT-PCR was conducted at the same temperature to verify the results obtained from gradient PCR. For *N. benthamiana*, two products of 151 bp and 240 bp were amplified at 39.9-48.9°C, with little to no amplification at higher temperatures higher than 49°C (Figure 2.30, Lanes 2-13). The lower band was a very minor product at 46.7-48.9°C. For taro, a single PCR product of 151 bp was observed at all temperatures tested, with the greatest amount of product observed at 39.9-48.9°C (Figure 2.30, Lanes 2-13). A product of the same size i.e. 151 bp was observed in *A. thaliana* at 59.9°C while no products were observed with no template control suggesting absence of contamination in the experimental set up (Figure 2.26, Lane 2-15). From this result it was concluded that 45°C would be a suitable annealing temperature for these primers since the product of the expected size of 151 bp was obtained from taro, *N. benthamiana* and *A. thaliana* at this temperature.

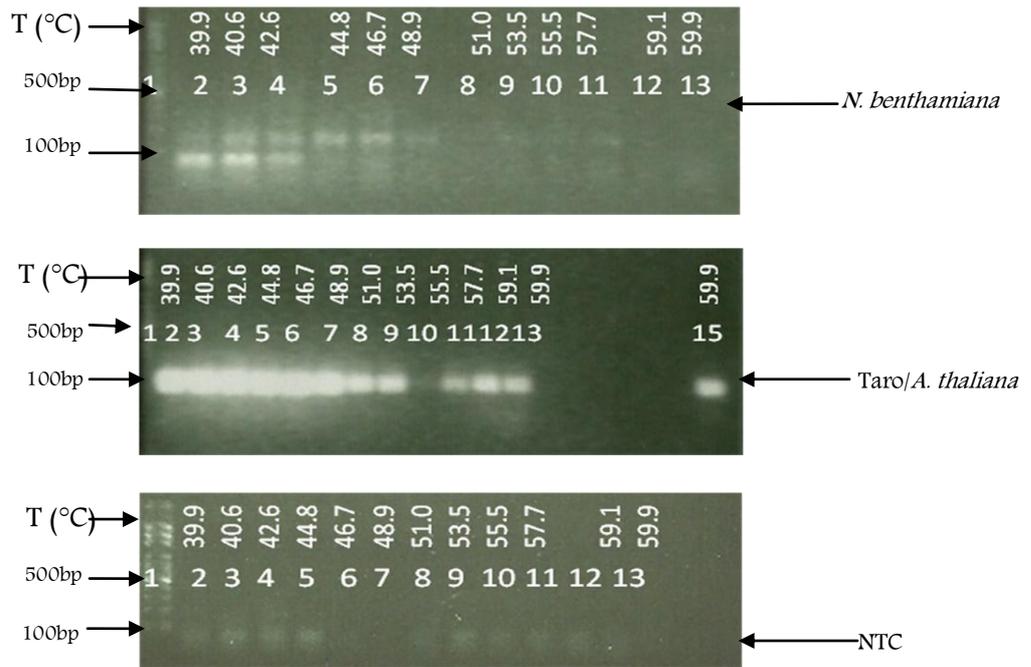


Figure 2. 30: Gradient PCR using EF1 α MD primers taro, *N. benthamiana*, *A. thaliana* (Lanes 2-13, Lanes 2-13 and Lane 15) and no template control (Lane 2-15). The temperatures tested were 39.9°C (Lane2), 40.6°C (Lane3), 42.6°C (Lane4), 44.8°C (Lane5), 46.7°C (Lane6), 48.9°C (Lane7), 51.0°C (Lane8), 53.5°C (Lane9), 55.5°C (Lane10), 57.7°C (Lane11), 59.1°C (Lane12) and 59.9°C (Lane15).

The EF1 α gene was efficiently amplified from taro and *N. benthamiana* cDNA template (Figure 2.31, Lane 2 and Lane 5, respectively). The expected size product of 151 bp was also seen with *A. thaliana* cDNA template along with a secondary product of size 250 bp (Figure 2.31, Lane 3). Absence of band with no template control suggests no contamination (Figure 2.31, Lane 6).

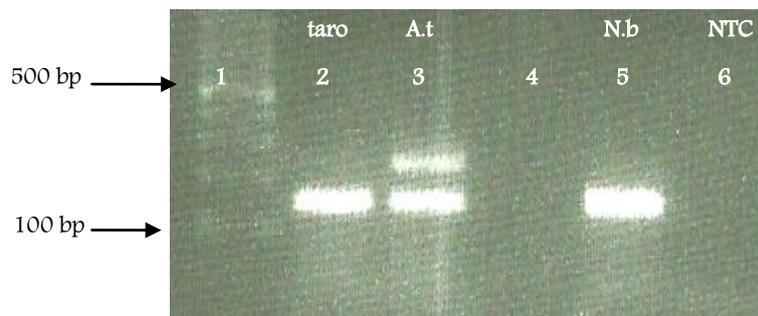


Figure 2. 31: RT-PCR using EF1 α MDF and EF1 α MDR primers on taro (Lane 2), *A. thaliana* (Lane 3), *N. benthamiana* (Lane 5) and a no template control (Lane 6). Lane1 contained 350 ng of 100 bp marker.

For *A. thaliana* an additional product of 250 bp was also observed (Figure 2.31, Lane 3), probably from contaminating genomic DNA since these primers flank intron. Amplification from genomic DNA would be expected to give a product of 250 bp.

To confirm this, RT-PCR was carried out with and without RT. Any contaminating genomic DNA would be available as a template in the absence of RT. Figure 2.32 shows the amplification of expected size 151 bp product from taro (Lane 4) with RT and no amplification of product from taro without RT (Lane 5); however the amplification of the expected size 151 bp product was observed with RT in Arabidopsis (Lane 2) and a product of approximately 250 bp was obtained without RT (Lane 3). Since the primers anneal to sequences flanking an intron in Arabidopsis (and possibly in other species), the expected sized product from Arabidopsis genomic DNA was 250 bp. For other species, the exact size of the product would be unknown unless the genomic sequence is available. Thus, it would be necessary to ensure any contaminating genomic DNA is removed before these primers are used in RT-qPCR.

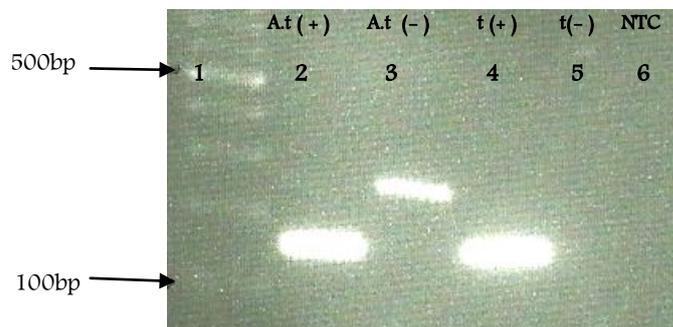


Figure 2. 32: RT-PCR using EF1 α MDF and EF1 α MDR primers on *A. thaliana* with and without RT (Lane 2 and 3 respectively), taro (Lane 3 and 4 respectively) and a no template control (Lane 6). Lane 1 contained 350 ng of 100 bp marker.

Further, sequencing confirmed that the sequences amplified from both *N. benthamiana* and taro were indeed EF1 α .

2.3.6.2 Analysis of new primer for F-box

The new monocot specific and dicot specific F-box primers were designed since universal primers for F-box could not be designed due to the significant variation within the homologous dicot and monocot species. Amplification of each template was carried out at 50°C to verify that this is a suitable annealing temperature for the

new F-box primers since the melting temperatures (T_m) for new monocot specific and dicot specific primers were 50°C and 51°C respectively. Figure 2.33 shows that the expected product of 100 bp was amplified by each specific primer pair in the correct plant type.

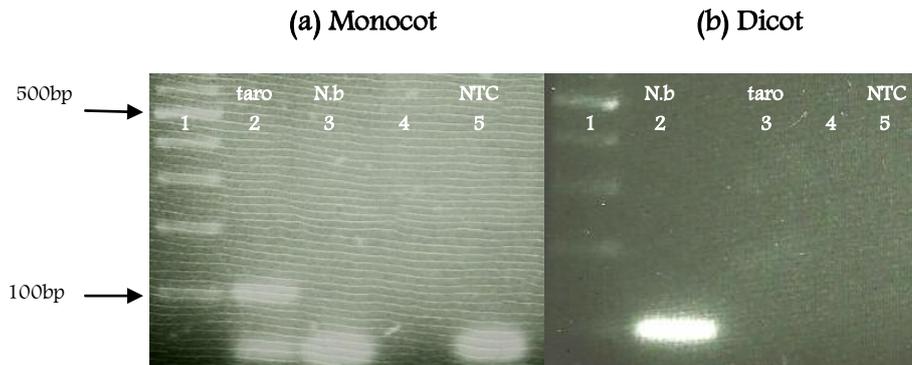


Figure 2. 33: RT-PCR at 50°C using (a) F-boxMF and F-boxMR and (b)F-boxDF and F-boxDR primers on taro (Lane 2 and 3, respectively), and *N. benthamiana* (Lane 2 and 3, respectively) and a no template control (Lane 5). Lane1 contained 350 ng of 100 bp marker. No samples were present in each Lane 4.

Amplification of the expected products for F-box was observed with both taro templates (Figure 2.33 (a), Lane 2) and *N. benthamiana* (Figure 2.33 (b), Lane 2) No product was amplified from *N. benthamiana* template (Figure 2.33 (a), Lane 3) using monocot specific primer set F-boxMF and F-boxMR. Similarly, no product was amplified from taro template (Figure 2.33 (b), Lane 3) using dicot specific primer set F-boxDF and F-boxDR. An absence of band was noticed with no template control suggesting no contamination in the experimental set up (Figure 2.33 (a) and (b), Lane 5). Since specific amplification was observed using an annealing temperature of 50°C, the PCR conditions described are useful for the detection of F-box cDNA from taro and *N. benthamiana*. These conditions may be suitable for detection of F-box mRNA from other monocot and dicot species but would need to be tested.

Neither pair of F-box primers span a predicted intron, therefore RT-PCR was carried out to ensure the product observed above were amplified from cDNA templates rather than contaminating genomic DNA. RT-PCR was carried out for the amplification of F-box gene using the monocot primer pair (F-boxMF and F-boxMR)

as well as the dicot primer pair (F-boxDF and F-box DR) with cDNA templates synthesised with and without RT from taro and *A. thaliana*. The primers were expected to amplify a product of 100 bp with RT.

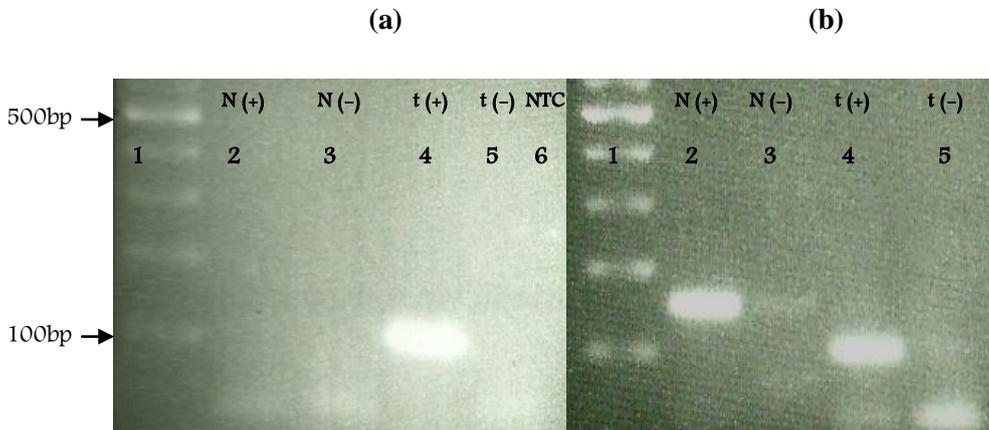


Figure 2. 34: RT-PCR using (a) F-boxMF and F-boxMR and (b) F-boxDF and F-boxDR primers on *N. benthamiana* with and without RT (Lane 2 and 3), taro (Lane 4 and 5, respectively) and a no template control (Lane 6). Lane1 contained 350 ng of 100 bp marker

The expected size product of 100 bp was observed for taro using the monocot specific primers F-boxMF and F-boxMR (Figure 2.34 (a), Lane 4) and with *N. benthamiana* using the dicot specific primers F-boxDF and F-boxDR (Figure 2.34 (b), Lane 2) with the cDNA template synthesized using RT enzyme. No band was observed for either taro template (Figure 2.34 (a), lane 5) or *N. benthamiana* (Figure 2.34 (b), Lane 5) without RT enzyme. Absence of any product was noticed with no template control suggesting no contamination in the experimental set up (Figure 2.34 (a), lane 6). These results suggested that no genomic DNA contamination was present and hence these primers could be used for the amplification of this mRNA in taro or *N. benthamiana* using RT-qPCR analysis.

Since the F-box primers do not span an intron, it is recommended that RNA extractions include a DNase step to remove contaminating genomic DNA using them for RT-qPCR analysis.

The products amplified from taro, *N. benthamiana* and *A. thaliana* using newly designed monocot and dicot specific EF1 α MD, F-boxM and F-boxD primers were verified by cloning the colony products. The colony PCR products for taro,

N. benthamiana and *A. thaliana* using T7 and SP6 primers were shown to be of the expected sizes of 300 bp and 250 bp with the inserts amplified from taro, *N. benthamiana* and *A. thaliana* using EF1 α MD, F-boxM and F-boxD primer pairs (Figure 2.35). Further, sequencing confirmed that EF1 α and F-box sequences were amplified from taro, *N. benthamiana* and *A. thaliana*.

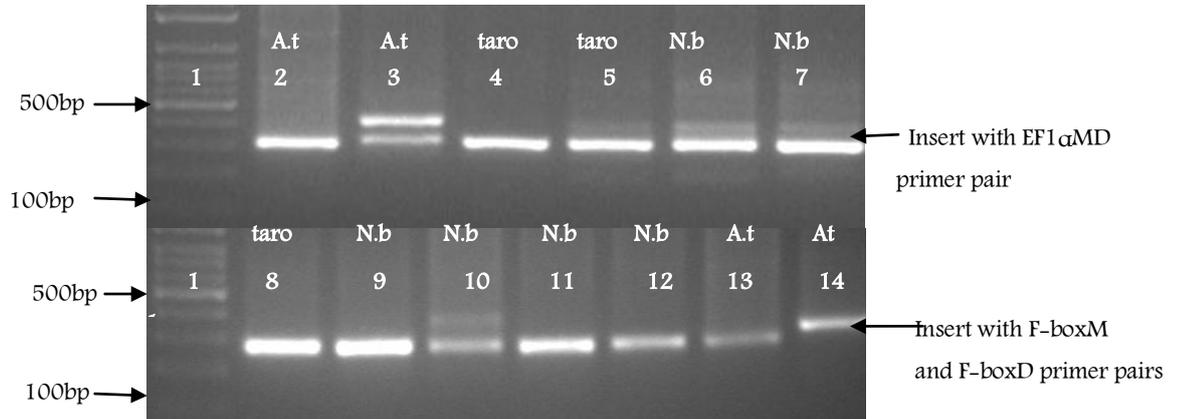


Figure 2. 35: Colony PCR using T7 and SP6 primers on *A. thaliana* (Lanes 2, 3 and 14), taro (Lanes 4, 5, and 8) and *N. benthamiana* (Lanes 6, 7 and lanes 9-12). Lane 1 contained 350 ng of 100 bp.

2.4. Discussion

The analysis of gene expression patterns has been critical to identifying genes and their roles in disease pathology. Various techniques that have been used for gene expression analysis include northern blotting, semi quantitative RT-PCR, RNase protection assays etc. However, the value of these techniques is severely constrained by their limitations and requirements. RT-qPCR is considered to be one of the most powerful tools for gene expression analysis due to its higher specificity, sensitivity and broad quantification range (Feng, et al., 2012).

To quantify the accumulation of mRNA of target genes using RT-qPCR, a selection of optimal reference genes is very important to obtain reliable and accurate data. They also allow us to detect the presence of any amplification inhibitors which could be used to recognize false negative results (Martanez, et al., 2011). Appropriate reference genes also allow inherent variability due to RNA extraction, cDNA synthesis and PCR efficiency to be compensated for (Thomas, Barbeyron and Mtchel, 2011).

The mRNA accumulation from a reference gene should remain constant under various experimental treatments used (Thellin et al., 2009). There have been many identified common reference genes that are used irrespective of their instability and variance. Examples of such widely used reference genes in plants include 18S rRNA, EF1 α , UBQ, actin, GAPDH and α -tubulin (TUA). However, many of these reference genes may not be stably expressed under all conditions. But most of the reference genes identified thus far have shown to vary in expression under certain conditions (Czechowski et al., 2005).

Lilly et al (2011) identified four candidate reference genes for use in RT-qPCR assays to study the effect of virus infection on plant host gene expression. The mRNA accumulation for the four genes, EF1 α , F-box, PDF2 and SAND, was found to be invariant between healthy and virus infected *A. thaliana* (Lilly et al., 2011). *A. thaliana* is considered to be the most commonly used model plant for plant biology due to its small genome size as well for its similarities in the genome organization between closely related species (Meinke et al., 1998). Lilly et al., (2011) also

suggested that the primer sequences provided for the amplification of these reference genes in TuMV infected *A. thaliana* could be beneficial for the selection and design of primer sets targeting similar genes in other plant species. While *A. thaliana* is a suitable model for most types of plant studies, it is less suitable as a model for understanding virus infection (Constantin et al., 2004). It is not infected by many viruses and is often hard to infected *N. benthamiana* is considered more appropriate for virus studies as it is easily infected by many viruses (Goodin et al., 2008). Further, Arabidopsis is a dicot species and may not represent monocot species. Thus, this study focussed on developing primers suitable for amplifying these reference genes in *N. benthamiana* and a monocot species, taro.

The similarity of primer sequences designed by Lilly et al., (2011) against Arabidopsis was tested on taro and *N. benthamiana* using RT-PCR. It has been suggested that reference genes selected on the basis of results obtained for one species require independent experimental validation and verification in the new context (Czechowski et al., 2005). Using 60°C as the annealing temperature as reported by Lilly et al (2011) for these primers, amplification efficiencies were low for all four reference genes in taro and *N. benthamiana* compared to *A. thaliana*. The yield of an RT-PCR reaction product depends on many parameters, one of which is the annealing temperature. The success of RT-PCR amplification depends mainly on the specificity with which a primer anneals to its target sequence (Hwang et al., 2003). In order to obtain an optimum annealing temperature for each of the four candidate reference genes, a gradient PCR was conducted. The gradient PCR technique allows to empirical determination of the optimal annealing temperature using the least number of steps and in one experiment (Erlich, 1989).

During gradient PCR, a temperature gradient of range 40°-60°C was built up across the thermoblock. The best annealing temperature for the amplification of each of the four candidate reference gene was suggested and RT-PCR at these selected temperatures were done to verify the results obtained from gradient PCR. Although the amplification of all candidate reference genes were improved for taro and *N. benthamiana* templates, the efficiency of amplification still did not appear to be as high when compared with the *A. thaliana* template. While standard RT-PCR is not quantitative and the primers consistently gave significant products for approximately the same amount of template, suggested that the Arabidopsis primers did not anneal

very efficiently to the alternative templates. Therefore, the primer sequences for each of the four reference genes were compared to homologous sequences from publicly available monocot and dicot sequences to determine the sequence variation at each binding site.

Sequence comparison showed that variation existed within the primer target sequences, particularly in the 3' end of the reverse primers and particularly for monocots. This sequence variation may account for the inefficient amplification; if the 3' end of a primer does not base pair with its target sequence very efficiently, then the polymerase may not extend the primer. Also, if there is sequence variation between the target sequence and the primer, the primer may bind elsewhere leading to the formation of non-specific products.

A new primer pair was designed for EF1 α mRNA. This primer pair amplified a 151 bp product from both *N. benthamiana* and taro, under the same PCR conditions. These primers flank an intron in known dicot and monocot sequences so any genomic DNA contamination of the RNA sample could be easily detected by the amplification of a product larger than 151 bp.

For F-box, a dicot specific primer pair amplified the expected product from *N. benthamiana* and *A. thaliana* but not from taro. Conversely, the monocot primer pair was specific to taro. Each primer pair amplified 100 bp product under the same conditions, allowing their use in one experiment.

Primers specific to PDF2 and SAND were not developed in this study since the purpose behind new primer design was to design universal primers for the amplification of these genes from both monocot and dicot species. This could not be achieved due to the significant variations among homologous dicot and in particular monocot sequences. Also, the limited availability of homologous sequences for these genes from other monocot and dicot sequences made it difficult to design primers.

Primers specific to EF1 α and F-box can amplify their target sequences from *A. thaliana*, *N. benthamiana* and taro. These primers would need to be tested against a broader panel of plant species to determine their general usefulness. Universal primers for PDF2 and SAND also need to be developed. Also, while Lilly et al.,

(2011) showed these genes were invariant in virus-infected Arabidopsis, they need to be validated in other virus infected plant species. This is the topic of chapter 3.

Chapter 3

Validation of reference genes and analysis of PVIP mRNA accumulation

3.1. Introduction

Due to its ability to detect and measure small amounts of nucleic acids in a wide range of samples, the fluorescence-based technique RT-qPCR has become the method of choice for nucleic acid quantification in molecular diagnostics, agriculture, life science and medicine. Its practical simplicity along with speed, sensitivity and specificity has made it the benchmark for the quantification of nucleic acid (Bustin et al., 2009).

3.1.1 Use of RT-qPCR in virology

In virology, RT-qPCR is used for disease diagnosis by measuring the presence/absence of a specific virus as well as how much virus is present in a sample (Mascia et al., 2010). Further, the TaqMan assay, with different specific probes, allows the detection of multiple viruses simultaneously (Coulson et al., 2008). RT-qPCR is also used to quantify accumulation of specific host mRNAs in response to infection. This type of study allows greater understanding of how an infection occurs as well as how a host may defend itself (Nolan et al., 2006).

3.1.2 MIQE guidelines

Many studies have been reported using RT-qPCR with a wide variety of experimental design and analysis approaches. It has been difficult to meaningfully compare one study to another. In an effort to assist scientists to produce consistent high quality data from RT-qPCR experiments and to standardise the way experiments are carried out, the MIQE guidelines have been developed (Bustin et al., 2009). This covers the aspects of sample acquisition, assay design, validation and data analysis (Taylor et al., 2010). Some of the detailed essential technical information required for an assessment of RT-qPCR under the MIQE guidelines includes sample source, its concentration, extraction method, accession number of target sequence, primer sequences, amplicon size, RT-PCR reagents, protocols, positive and negative controls, statistical justification of data and normalisation (Bustin et al., 2010). These guidelines are given in table 3.1.

Table 3. 1: Checklist for author of MIQE at time of manuscript submission, detailed information about individual parameters associated with each step of the RT-qPCR analysis (Bustin et al., 2010).

Sample/ template	Details
Source	If cancer, was biopsy screened for adjacent normal tissue
Method of preservation	Liquid N2/RNA later/formalin
Storage time	If using samples older than 6 months
Handling	Fresh/frozen/formalin
Extraction method	Trizol/columns
RNA DNA free	Intron-spanning primers/no RT control
Concentration	Nanodrop/ribogreen/microfluidics
RNA integrity	Microfluidics/3'5' assay
Inhibition free	Method of testing
Assay optimisation/validation	Details
Accession number	For example, RefSeq XX_1234567
Amplicon details	Exon location, amplicon size
Primer sequence	Even if previously published
Probe sequence	Identify locked nucleic acid (LNA) or other substitutes
<i>In silico</i>	BLAST/Primer-BLAST/m-fold
Empirical	Primer concentration/annealing temperature
Priming conditions	Oligo-dT/random/combination/target-specific
PCR efficiency	Dilution curve
Linear dynamic range	Spanning unknown targets
Limits of detection	Limit of detection (LOD) /accurate quantification
Intra-assay variation	Copy numbers not C_q
RT-PCR	Details
Protocols	Detailed description, concentrations, volumes
Reagents	Supplier, lot number
Duplicate RT	ΔC_q
NTC	C_q and melt curves
No amplification control (NAC)	ΔC_q
Positive control	Beginning end of qPCR
	Inter-run calibration
Data analysis	Details
Specialist software	Eg, QBAsEPlus
Statistical justification	Eg, Biological replicates
Transparent validated normalisation	Eg, GeNorm summary

The aim of MIQE is to establish a framework within which to conduct RT-qPCR experiments and to provide guidelines for reviewers and editors to use in the evaluation of the technical quality of submitted manuscripts. Accordingly, investigations that use this widely applied methodology will produce data that are more consistent, more compatible and more reliable.

For an accurate quantification of transcript accumulation of a gene using RT-qPCR analysis, it is necessary to use appropriate stably expressed reference genes for the normalisation in order to overcome any variation in the result from sample variation, variation in RNA integrity, variation in cDNA and RT efficiency (Mestdagh et al., 2009). Another method to overcome this error is by normalising to total RNA, but

this requires a reliable RNA quantification method and it also fails to take into account the variability of RT (Becker et al., 2010).

3.1.3 Analysis of RT-qPCR data

There are two different methods to analyse RT-qPCR data: absolute and relative quantification. The absolute quantification method determines the input copy number of the transcript of the gene of interest by relating the PCR signal to a standard curve (Figure 3.1). Relative quantification on the other hand (Figure 3.1) determines the transcript of the gene of interest relative to some reference group such as an internal control or reference gene (Livak and Schmittgen, 2001).

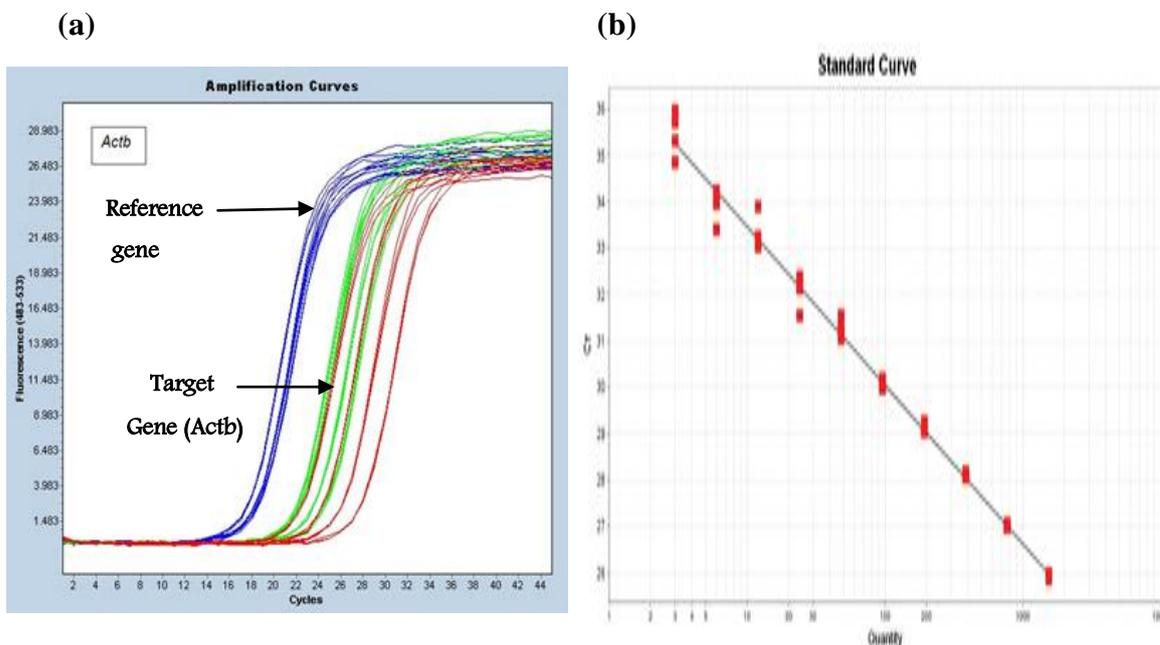


Figure 3. 1: This figure shows relative and absolute quantification method of RT-qPCR. (a) Standard curve plot for the accurate quantification of an unknown sample using a standard curve which represents log transformed C_t values against the concentration of known sample. (b) Relative amplification plot between cycle number (C_t) and fluorescence emitted for a given sample (example, Actb) to quantify its mRNA accumulation (<http://www.roche-applied-science.com>; <http://www.genomics.agilent.com>).

The delta-delta C_t method ($\Delta\Delta C_t$) is an easy and convenient way to analyse the relative change in gene expression using RT-qPCR (VanGuilder, Vrana and Freeman, 2008). In this method, ΔC_t defines the difference in threshold cycles for target and reference (Figure 3.3). The amount of target, normalised to a reference

gene and relative to a calibrator which is an external control which could be an housekeeping gene or an untreated sample, is given by-

$$\text{Change in mRNA accumulation} = 2^{-\Delta\Delta C_t}$$

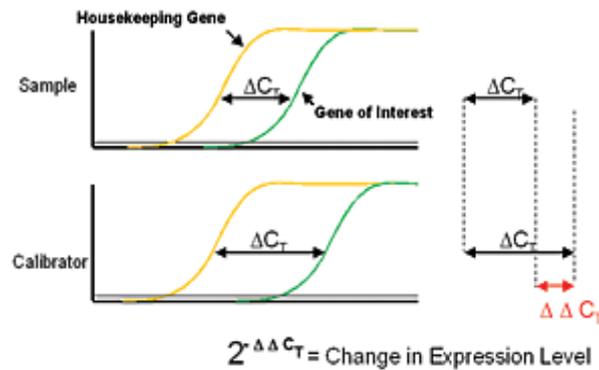


Figure 3. 2 This figure shows the delta-delta C_t method where ΔC_t defines the difference between the housekeeping gene/ reference gene and gene of interest and $\Delta\Delta C_t$ defines the difference relative to the calibrator
<http://www.eppendorf.de/int/index.php?l=131&action=products&contentid=101&sitemap=2.5.1>.

For the delta-delta C_t method calculation to be valid, the amplification efficiencies of the target and reference gene must be approximately equal (VanGuilder, Vrana and Freeman, 2008). This method is used to determine any fold change in gene expression normalised to a reference gene and relative to the untreated control. For the untreated control sample, $\Delta\Delta C_t$ value equals one, so the fold change in gene expression relative to untreated control equals one, by definition. Therefore, for treated samples, evaluation of $2^{-\Delta\Delta C_t}$ indicates the fold change in gene expression relative to the untreated control (Livak and Schmittgen, 2001).

3.1.4 Reference genes for RT-qPCR

A major factor when using RT-qPCR is the selection of appropriate reference genes since any variation in the expression of the reference gene between sample groups will reduce the sensitivity of the assay to identify any change in the expression of a gene of interest. Artificial changes may also be identified (Popovici et al., 2009). As discussed in chapter 2, housekeeping genes such as GAPDH, β -actin, 18S ribosomal

RNA (18S rRNA) have been most commonly used as reference genes. It has been assumed that the expression levels of these housekeeping genes remain constant in response to the experimental conditions (Watson et al., 2007). However, recent reports suggest that the expression of these genes can vary under various experimental conditions making them inappropriate for their use as normalisers (Czechowski et al., 2005). In practicality, no cellular gene maintains constant expression levels under all conditions and hence the evaluation of an appropriate reference gene to normalise RT-qPCR data is an essential requirement when designing a RT-qPCR experiment using new experimental conditions (Peirson, Butler and Foster, 2003).

Particularly, it is critical to use appropriate reference genes when studying the effects of virus infection on host cell gene expression since viruses can interfere with host cell pathways, the components of which the traditional housekeeping genes often encode. These include cell cycle, DNA replication and transcription as well as metabolism in order to support their replication cycle (Watson et al., 2007).

It has been outlined that an appropriate reference gene is the one that is (D'haene and Hellemans, 2010) stably accumulated across all samples and whose transcript levels remain constant under any given experimental treatment, they should not be associated with any pseudogenes to avoid genomic DNA amplification, they should be able to reflect amplification variation in RNA quality, quantity and cDNA synthesis. These genes should also be equivalent in accumulation with that of the target gene transcript and these genes should accumulate in moderate amounts (i.e. C_t within the range 15-30) (Stahlberg et al., 2004).

3.1.5 Validation of reference genes

Reference genes must be validated to ensure they met all the criteria outlined above. Validating candidate reference genes is critical and difficult since the stability of a reference gene's transcript accumulation level must be obtained by comparison with other reference genes. There have been many statistical algorithms developed such as geNorm, NormFinder and BestKeeper to process raw RT-qPCR data to identify the best reference genes to be used under given experimental conditions (Jain et al.,

2006). These algorithms evaluate the relative expression stability of multiple candidate reference genes and have been developed to address the circular problem that for an evaluation of the expression stability of a candidate reference gene, another reference gene is needed for its normalisation (Dheda et al., 2004). This software also considers the relative amplification efficiency of the reference genes.

It has been suggested that for accurate data normalisation, more than one reference gene should be used. This normalisation strategy is based on the geometric averaging of the expression level of multiple reference genes and ranking them based on their stability value, (M). The M value for a gene is expressed as the average pair-wise variation of the log transformed standard deviation of a given gene transcript accumulation ratio compared with the ratio of all other candidate reference genes. The gene with the lowest M value is considered to be the most stable gene (Wicks et al., 2006).

Surprisingly, validation of reference genes in plants has gained very little attention. RT-qPCR to monitor gene expression in plants has been underused even though it could be used to differentiate between the expression of closely related genes, and to quantify transcript levels of very weakly expressed genes (Gutierrez et al., 2008). Czechowski et al (2005) assessed 22 genes of *Arabidopsis* including protein phosphatase 2A subunit, ubiquitin-conjugating enzyme and coatamer subunit as reference genes for RT-qPCR. These genes were identified as novel reference genes for normalisation and quantification of transcript levels in *Arabidopsis* and were found to have stable expression throughout development and under a range of experimental conditions when compared to other traditional reference genes such as GAPDH, UBQ, actin and TUB (Czechowski et al., 2005).

Studies have also been carried out on non-model species. For example, in chicory (*Cichorium intybus*) seven candidate reference genes including GAPDH, EF1 α , 18S rRNA, ACT, histone 3 (H3), TUA and nicotinamide adenine dinucleotide dehydrogenase (NADHP) were analysed for their expression stability in root and leaf tissues. It was concluded from this study that ACT and EF1 α showed the greatest stability in their expression levels and hence could be used as reference genes for normalisation of gene expression in chicory leaf and root tissues (Maroufi, Van Bockstaele and De Loose, 2010). A similar study conducted on wheat, in which 32

genes including 12 commonly used housekeeping genes and 20 other genes from 24 different plant samples including different tissues, development stages and temperature stress were analysed. Among all these genes tested, genes encoding a protein disulfide isomerase (PDI) - like protein showed the most stability (Paolacci et al., 2009).

No analysis of reference genes for RT-qPCR has been carried out on taro despite its economical importance as a staple starch crop in tropical and subtropical regions of the world (Ravi, Aked and Balagopalan, 1996). In particular, no such analysis has been done to find suitable reference genes for studying virus infected taro. DsMV is a potyvirus that infects taro, causing significant economic losses (Zettler and Hartman, 1987). Thus, further analysis of DsMV infected taro is needed to better understand this disease.

3.1.6 Potyvirus VPg interacting protein

As discussed in chapter 1, PVIP is a host protein that has been shown in a variety of plant species to interact with the potyviral VPg protein. A previous study showed that PVIP in *N. benthamiana* is homologous to the OBERON 1 and 2 proteins of Arabidopsis (Anand, 2010) which have a role in the maintenance of apical meristems (Saiga et al., 2008). The PVIP mRNA is present in several tissues, in light grown *N. benthamiana* plants, namely stems, leaves and roots (Anand, 2010). Therefore, this protein has a role that is independent of virus infection. Further, the accumulation of PVIP in each of these tissues appears to be affected by changes in light conditions (Anand, 2010). The abundance of PVIP mRNA was low in leaf tissues that were exposed to the dark as compared to the leaf tissues exposed to standard day/ night cycles (Anand, 2010). Thus, the PVIP gene appears to be responsive to this type of abiotic stress. What is not known is if the PVIP gene is responsive to biotic stress such as virus infection.

Since the RT-PCR conditions for the reference genes EF1 α and F-box were more consistent for taro than for *N. benthamiana* (to be shown in the Results section), this study focussed on taro as the host plant.

The aims of this study were firstly to validate EF1 α and F-box reference genes to be used for mRNA quantification studies in healthy and virus infected taro. Secondly, to compare the transcript accumulation of the host PVIP between healthy and DsMV infected taro.

3.2. Materials and Methods

3.2.1 Plant material

The healthy and DsMV infected taro leaves were collected from taro plants kept in the glasshouse at the School of Biological Science, University of Auckland.

3.2.2 Total RNA extraction

Total RNA was extracted from healthy and virus infected taro leaves using the method described in section 2.2.1.

3.2.3 RT-PCR to test PVIP primers on taro

RT-PCR was carried out as described in section 2.2.4. The forward and reverse primers for PVIP had been designed against the *N.benthamiana* mRNA. The sequence for these primers, PVIPF and PVIPR, are given in table 3.2.

Table 3. 2: PVIP primer sequences and amplicon size

Gene	Species	Forward primer (PVIPF)	Tm (°C)	Reverse primer (PVIPR)	Tm (°C)	Product size
PVIP	<i>N. benthamiana</i>	5'GCACCAAAA GAGTTCGTGTC 3'	50	5'TCCCAGATGAT GTACAGCAA3'	50	100 bp

3.2.4 Validation of reference genes using RT-qPCR

The mRNA levels of EF1 α and F-box genes in healthy and DsMV infected taro leaves were analysed using one step real time PCR analysis. The reactions were carried out in 96 well Eppendorf twin.tec real time PCR plates covered with Applied Biosystems optical adhesive film.

Each reaction consisted of 5 μ l of the one step SYBR Green Master Mix (Quanta Biosciences), 0.2 μ l of 10 μ M forward and reverse primers, 0.2 μ l of the qScript

One-step RT enzyme, 1 μ l (~110 ng) of the RNA template and 3.4 μ l of nuclease free water. The primers used were EF1 α MDF/ EF1 α MR and F-boxMF/ F-boxMR. The sequences for these are shown in table 2.2. The total reaction volume was 10 μ l and each sample was analysed as three biological replicates each in triplicate. Negative controls were also analysed in triplicate. The PCR cycling conditions were- initial denaturation at 95°C for 2 minutes followed by 35 cycles with denaturation at 94°C for 30 seconds then annealing at 50°C for 30 seconds and extension at 72°C for 30 seconds with a final extension at 72°C for 5 minutes.

An additional melting curve cycle of 50°C for 10 minutes was also included to confirm that all the products amplified were of the same size and to differentiate desired PCR products from primer-dimers. RT-qPCR was conducted in an Eppendorf Realplex gradient PCR machine.

3.2.5 Determination of PVIP mRNA accumulation in healthy and DsMV infected taro

Relative quantification of the PVIP mRNA in healthy and DsMV infected taro leaves was analysed using one step real time PCR analysis using both the reference genes EF1 α and F-box. The reaction conditions were the same as discussed in section 3.2.4.

Healthy taro was used as the calibrator; hence the relative amounts of PVIP mRNA in DsMV infected taro leaf tissue was measured with respect to healthy taro leaf tissue.

The default value for the threshold line was accepted and delta-delta C_t values were recorded by the Realplex machine itself by selecting the parameter for relative quantification. Raw C_t values are given in appendix 3.

3.2.6 Statistical analysis

In order to confirm the RT-qPCR result statistically, Minitab 15 was used to conduct the t-test analysis.

3.3 Results

3.3.1 RNA quality

The integrity of total RNA extracted from the leaf tissue of healthy and DsMV infected taro was assessed by electrophoresis. Figure 3.3, shows the presence of 28S and 18S rRNA bands with the background smear of mRNA. This pattern indicated the RNA was intact and suitable for RT-qPCR.

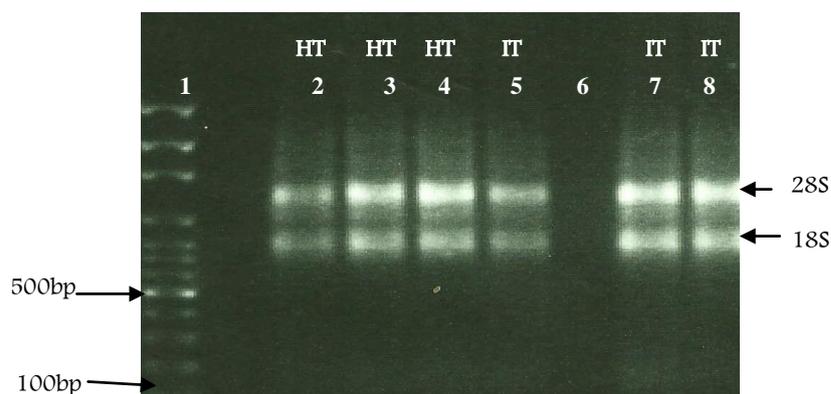


Figure 3. 3: Agarose gel electrophoresis of total RNA extracted from the leaf tissue of taro (Lanes 2-8) where HT stands for healthy taro and IT stands for infected taro. The 28S and 18S rRNA are indicated. Lane 1 contains 350 ng of 100 bp marker.

3.3.2 Testing amplification of reference genes at 50°C

For an optimum reference gene, the transcript level of the gene must remain constant between treated and untreated samples. Accumulation of EF1 α and F-box mRNA in healthy and DsMV infected taro leaves was compared. While optimal annealing temperatures were determined by RT-qPCR for each gene as described in chapter 2, for these genes to be useful, amplification had to occur at the same annealing temperature as the gene of interest that is PVIP. The annealing temperature of 50°C was therefore tested by RT-PCR before using this temperature in quantitative experiments. Optimum amplification of the EF1 α and F-box genes was observed for the taro (Figure 3.5, lanes 4 and 6, respectively). Similar results were seen for *A. thaliana* and *N. benthamiana* (Figure 3.4, lanes 2 and 3, respectively) using the EF1 α primer set. However, a very faint band of the expected size product of 100 bp with the dicot specific F-box primer was noticed for the *A. thaliana* and *N. benthamiana* (Figure 3.4, lanes 5 and 7, respectively). Thus, the EF1 α MDF/MDR and F-

boxMF/MR primer pairs could be used at the 50°C annealing temperature, but F-boxDF/DR primer could not be used at the same temperature. Therefore, analysis of PVIP mRNA accumulation in virus infected plants, focussed only on taro.

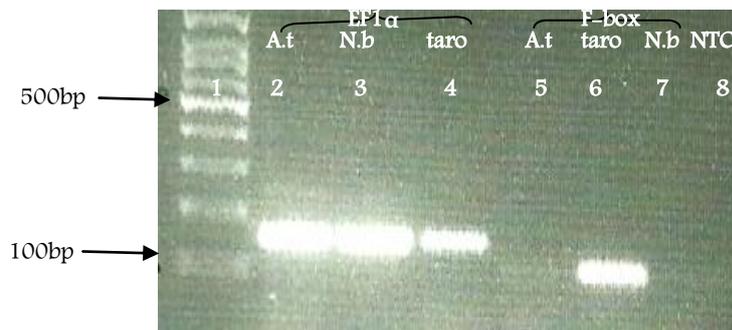


Figure 3. 4: RT-PCR using EF1 α MDF/ EF1 α MDR also F-boxMF/ F-boxMR and F-boxDF/ F-boxDR primers on *A. thaliana* (Lanes 2 and 5, respectively), taro (Lanes 4 and 6, respectively), *N. benthamiana* (Lanes 3 and 7, respectively) and a no template control (Lane 8). Lane1 contained 350 ng of 100 bp marker.

3.3.3 Validation of reference genes using RT-qPCR

It can be observed from Figure 3.5 that the accumulation levels of EF1 α (23-24) and F-box (21-22) genes in taro were within the specified range of C_t values (15-30) (raw C_t values are given in appendix 2). No significant difference in the mRNA accumulation of EF1 α and F-box gene was observed between healthy and DsMV infected taro across all replicates and triplicates in experiment 1; however, differences in the C_t values in both EF1 α and F-box between healthy and DsMV infected taro were observed in experiment 2. Although this difference was not expected, this variation could have been due to pipetting error. These reactions were set up by manual pipetting since no liquid handling robot was available. Manually pipetting small volumes introduce error. While master mixes were used to minimise pipetting error, given the small total volume of 10 μ l for each reaction, it is likely that some error was introduced through pipetting. It could also be due to the fact that these two experiments were conducted separately using different RNA preparations from taro. The C_t value for the no template control was recorded as 32. This is quite a high C_t value, suggesting detection of a signal very late in the reaction, suggesting absence of contamination

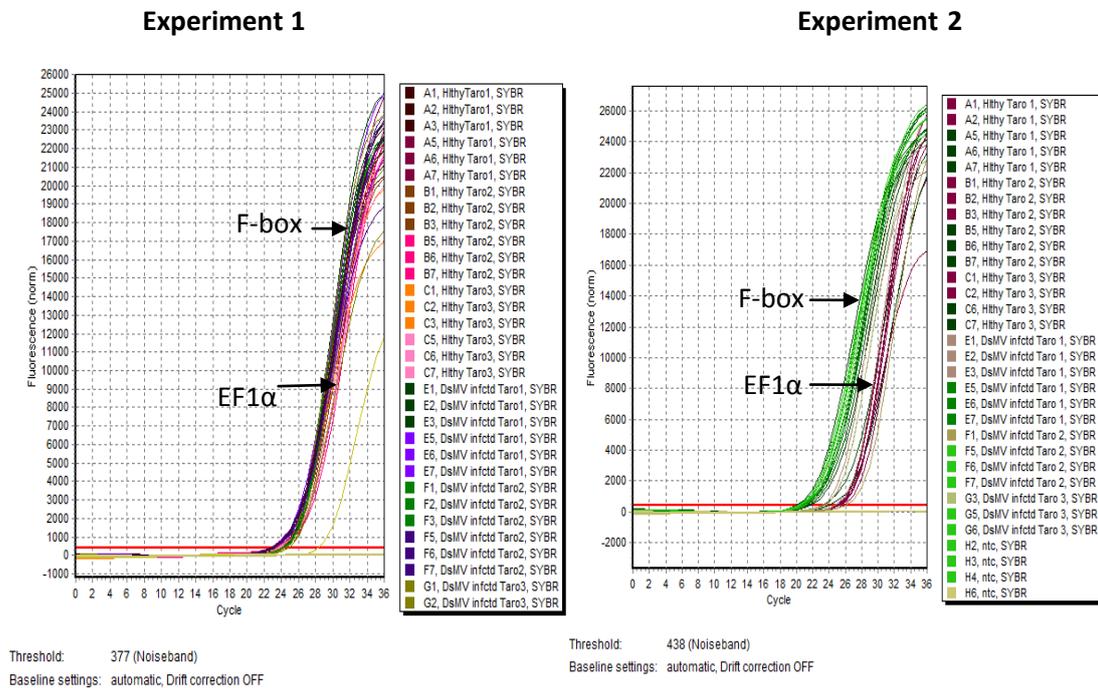


Figure 3. 1: RT-qPCR results comparing mRNA accumulation of EF1 α and F-box genes in healthy and DsMV infected taro leaf tissues. In the experiment 1 graph, dark brown /dark pink represents healthy taro with F-box primers and green/purple represents DsMV infected taro with F-box primers, while brown/pink represents healthy taro with EF1 α primers and yellow/dark blue represent DsMV infected taro with EF1 α primers. Similarly, in the experiment 2 graph, all brown represents healthy taro with F-box primers and dark pink/light pink for DsMV infected taro with F-box primers. All the light green/purple represents healthy taro with EF1 α primers while all the dark green represents DsMV infected taro with EF1 α primers. The yellow lines and light blue lines in experiment 1 and 2, respectively represents the no template control. For both the experiments, the amplification for no template control was observed below the threshold line.

3.3.3.1 Melting curve analysis

The melting curve analysis showed that single PCR products were amplified for both EF1 α and F-box mRNAs from healthy and DsMV infected taro across all replicates. The peaks below the threshold lines indicate the presence of primer-dimer since their melting temperatures were different from that of the desired products. Agarose gel electrophoresis confirmed that these products were the correct size (Figure 3.6). Thus, the specific products of the correct size were amplified from both healthy and DsMV infected taro leaves.

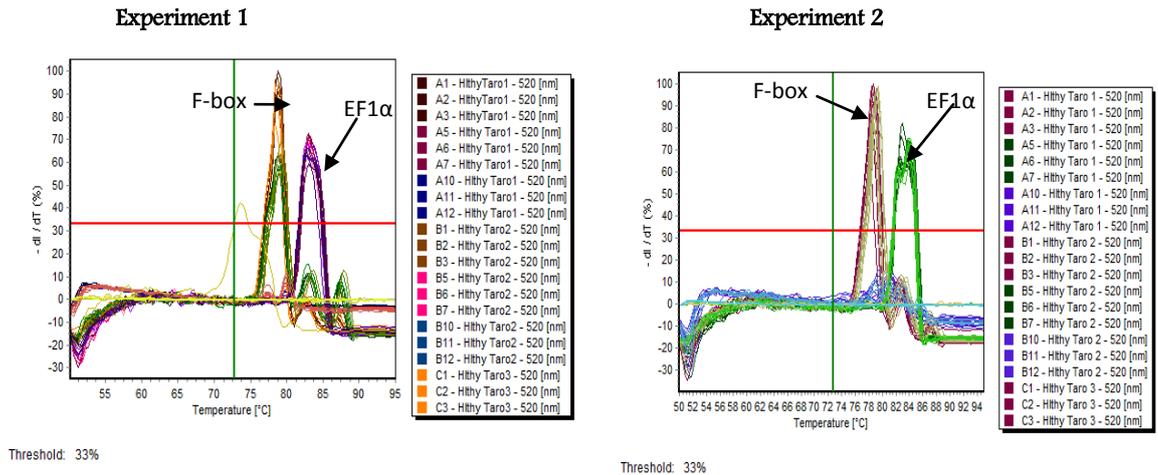


Figure 3. 1: Melting curve analyses from experiment 1 and 2 of all replicate and triplicate healthy and DsMV infected taro samples amplified using F-boxM and EF1 α MD primers.

Further, lack of contamination, non-specific product or primer-dimer from the RT-qPCR analysis of EF1 α and F-box in healthy and DsMV infected taro was verified using agarose gel electrophoresis (data not shown).

3.3.3.2 Statistical analysis

To determine any significant variation in the C_t values between the replicates in healthy and infected taro mean C_t values and standard deviations of these means for EF1 α and F-box mRNA accumulation within an experiment were calculated (Table 3.3). It was observed in this analysis that the standard deviation for both EF1 α and F-box in both healthy and infected tissue was less than one in experiment 1. However, in experiment 2 the standard deviation for EF1 α was close to one in healthy tissue, and for F-box in both healthy and infected tissue. This suggested significant variation in the mRNA quantification within the replicates in the experiment 2, particularly in healthy tissue. This analysis also showed about a cycle difference in the average mean between healthy and infected tissue for EF1 α within experiment 1 and 2; 23.39 vs 22.83 and 21.61 vs 20.63 for healthy vs infected in experiments 1 and 2, respectively. However, for F-box this difference was only present within experiment 2; 24.24 vs 24.20 and 25.37 vs 24.28 for healthy vs infected in experiments 1 and 2, respectively. Biologically, this one cycle difference between

healthy and infected tissue reflects a two- fold difference in the amount of starting template but given the reactions were set up manually this difference was likely due to pipetting error.

Further, t-test analysis was conducted to determine if there was any significant variation of EF1 α and F-box mRNA accumulation between healthy and infected taro both within and between experiments (Table 3.4). The null hypothesis was assumed to be that no significant variation in mRNA accumulation for EF1 α and F-box would occur between healthy and infected taro within an experiment or between experiments.

In this analysis the p- value for EF1 α between healthy and infected taro in both the experiments 1 and 2 was greater than the assumed level of significance of 0.05 (p= 0.266 and 0.11, respectively) suggesting there was no significant variation in the mRNA accumulation of EF1 α between healthy and infected taro within an experiment. However, the p- value for EF1 α between experiments 1 and 2 for both healthy and infected tissue (p= 0.005 and 0.006, respectively) was observed to be less than one, suggesting significant variation in the mRNA accumulation of EF1 α between experiments 1 and 2. This variation was again probably due to pipetting errors.

For F-box, the p- value between healthy and infected taro in the experiments 1 (p= 0.62) was greater than 0.05, suggesting no significant variation in the mRNA accumulation of F-box between healthy and infected taro in experiment 1; however, the p- value for F-box between healthy and infected tissue in experiment 2 (p-value= 0.0002) was noted to be less than 0.05 suggesting significant variation in the mRNA accumulation of F-box between healthy and infected tissue within experiment 2. The t-test analysis conducted to determine any significant variation in the mRNA accumulation of F-box between experiments 1 and 2 showed a p-value of 0.003 for healthy tissue. The observed p-value was less than the assumed p-value of 0.05, suggesting significant variation in the F-box mRNA accumulation in healthy tissue between experiments. However, the p-value for F-box between experiments 1 and 2 in infected tissue (p= 0.64) was observed to be greater than 0.05, suggesting no significant variation in the mRNA accumulation of F-box in infected tissue between experiments. This inconsistency in the mRNA accumulation of both the genes, EF1 α

and F-box result between experiments could be due to the removal of outliers which lead to reduction in the number of samples that were analysed.

Hence from the t-test analysis it could be concluded that there was no significant variation observed in the mRNA accumulation of EF1 α and F-box between healthy and infected tissue within an experiment but there was significant variation in both healthy and infected tissue between experiments. However, for F-box no significant variation was observed in the mRNA accumulation between experiments in infected tissue. This inconsistency in the result observed particularly for F-box could be due to pipetting error since these small volume reactions were set up by manual pipetting, while master mixes were used to minimise pipetting errors, given the small total volume of 10 μ l for each reaction it is likely that some error was introduced through pipetting.

Table 3. 3: The average mean and standard deviation of EF1 α and F-box in healthy and DsMV infected taro for experiment 1 and experiment 2.

	Experiment 1		Experiment 2	
	Healthy (av \pm sd)	Infected (av \pm sd)	Healthy (av \pm sd)	Infected (av \pm sd)
EF1α	23.39 \pm 0.31	22.83 \pm 0.13	21.61 \pm 1.13	20.63 \pm 0.38
F-box	24.24 \pm 0.31	24.20 \pm 0.31	25.37 \pm 0.75	24.28 \pm 0.87

Table 3. 1: The t-test analysis for EF1 α and F-box between healthy and DsMV infected taro for experiment 1 and experiment 2 to determine any significant variation of EF1 α and F-box within and among the replicates.

	EF1 α (p- value)	F-box (p- value)
Healthy vs Infected (exp 1)	0.266	0.62
Healthy vs Infected (exp 2)	0.11	0.0002
Healthy vs Healthy (exp 1)	0.005	0.003
Infected vs Infected (exp 2)	0.006	0.64

3.3.4 RT-PCR to test PVIP primers on taro

RT-PCR amplification of the PVIP mRNA was carried out using PVIPF and PVIPR primers (Table 3.1) at an annealing temperature of 50°C. These primers had been

designed against *N. benthamiana* PVIP mRNA and were tested on taro and *A. thaliana*. The primers were expected to amplify a product of 100 bp.

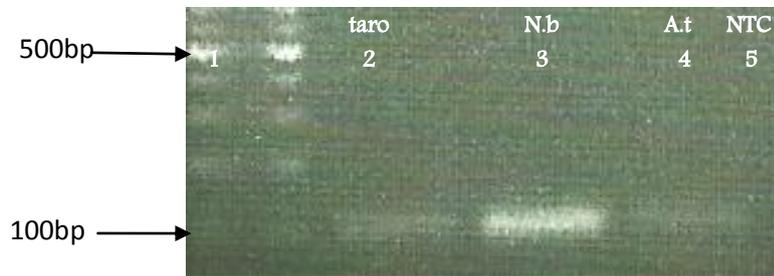


Figure 3. 1: RT-PCR using PVIPF and PVIPR primers on taro (Lane 2), *N. benthamiana* (Lane 3), *A. thaliana* (Lane 4) and a no template control (Lane 5). Lane 1 contained 350 ng of 100 bp marker.

A product of the expected size was amplified from *N. benthamiana* Figure 3.7 (Lane 3); however, only a faint band was observed with taro (Lane 2) and *A. thaliana* (Lane 4) suggesting inefficient amplification. Absence of bands for the no template control suggested no contamination in the experimental set up (Lane 5). While the amplification from taro and *A. thaliana* was faint, a product of the correct size was observed. Since this could be seen with the naked eye, it was thought that this level of amplification would be detectable by RT-qPCR.

3.3.5 Relative quantification of the PVIP mRNA in healthy and DsMV infected taro

RT-qPCR was conducted to assess the PVIP mRNA accumulation in healthy and DsMV infected taro, for which the PVIP mRNA accumulation was normalised against the mRNA accumulation of reference genes *EF1 α* and *F-box* in healthy and DsMV infected taro. It was observed in the Figure 3.8 that there was no significant amplification of the PVIP mRNA from healthy and DsMV infected taro across all replicates and triplicates. The C_t values ranged between 30-33 suggesting either non-specific or inefficient product amplification. This agrees with the RT-PCR result obtained for taro using PVIP primers as shown in Figure 3.7. A very faint band of the expected size was seen on an agarose gel for taro; clearly this was insufficient amplification for detection by RT-qPCR. This suggests the PVIP primers from *N.*

benthamiana are inappropriate for amplification in taro and by inference from the gel in Figure 3.7, in *A. thaliana* also. Either a gradient PCR needs to be carried out to determine if the primers can work at a different temperature (since 50°C was optimal for *N. benthamiana*) or, a new primer for PVIP specific to taro needs to be designed.

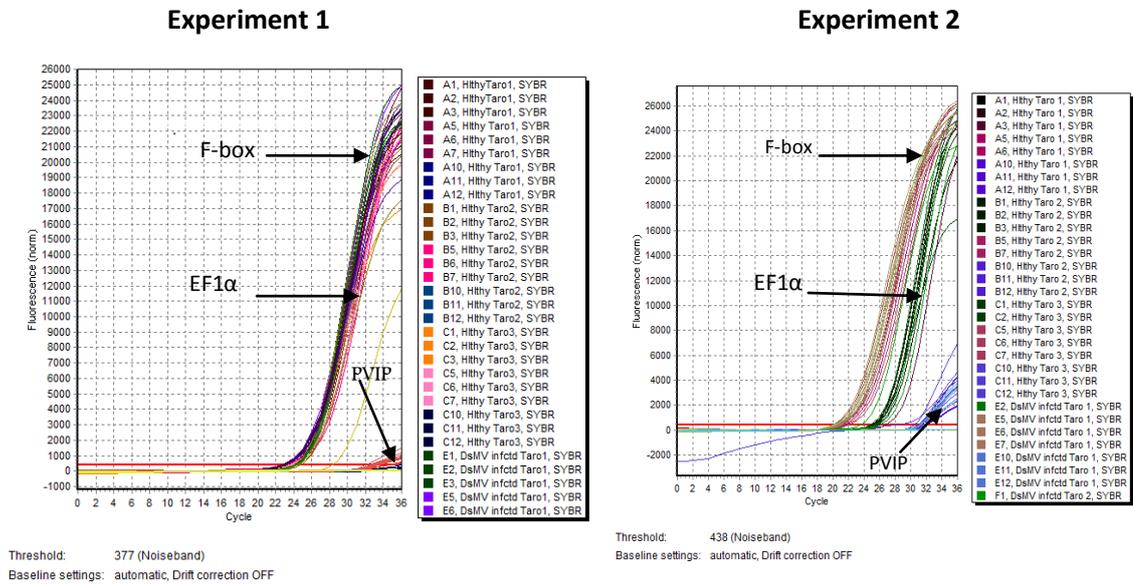
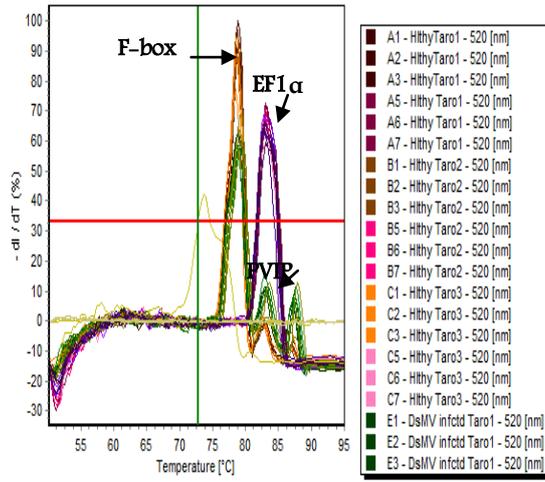


Figure 3. 1: RT-qPCR result comparing expression level of PVIP in healthy and DsMV infected taro normalised against EF1 α and F-box reference genes.

3.3.5.1 Melting curve analysis

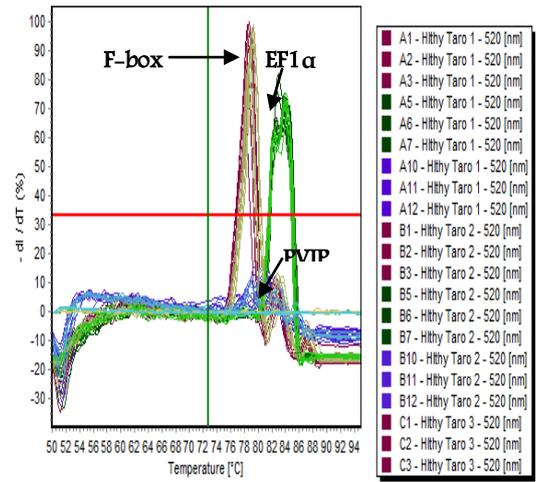
The melting curve analysis showed that single PCR products were amplified using the PVIP primers from healthy and DsMV infected taro across all replicates (Figure 3.9); however, the peaks in both the experiments were below the threshold lines indicating inefficient amplification of the PVIP mRNA from both healthy and DsMV infected taro, suggesting new primer design for the PVIP specific to taro for an efficient determination of PVIP mRNA accumulation in healthy and DsMV infected taro.

Experiment 1



Threshold: 33%

Experiment 2



Threshold: 33%

Figure 3. 1: Melting curve analyses of all replicate and triplicate healthy and DsMV infected taro samples amplified using PVIP primers.

3.4 Discussion

RT-qPCR has now become the most common and standard method for in-depth analysis of gene expression studies due to its increased sensitivity, high sequence specificity and large dynamic range (VanGuilder, Vrana and Freeman, 2008). There are many variables that need to be considered as described in MIQE guidelines, when conducting gene expression analysis using RT-qPCR including amount of RNA, RNA quality and amplification efficiency (Bustin et al., 2009). The major factor among all of this is the use of appropriate reference genes. An optimum reference gene defines a gene that is experimentally found to be stably expressed in a given species under a given experimental condition and is considered to be suitable for quantitative analysis under respective conditions (Coulson et al., 2008).

The aim of this study was to validate the two candidate reference genes EF1 α and F-box to be used as reference genes for mRNA quantification studies in healthy and DsMV infected taro. These two reference genes were identified as stable reference genes for normalisation of transcripts from TuMV and other virus infected *A. thaliana* among 12 other candidate reference genes tested (Lilly et al., 2011). It was assumed that these genes could be used as suitable candidates for use in other plant species since the primer pairs designed to target EF1 α and F-box genes in TuMV infected *A. thaliana* showed greater similarity to genes of other species especially *Nicotiana* species. Also, *A. thaliana* being a model plant it was suggested that these reference genes might be used as appropriate reference genes in other monocot and dicot plants once validated experimentally.

The expression stabilities of EF1 α and F-box were analysed statistically using their C_t values to determine any significant variation between healthy and DsMV infected taro across all replicates and triplicates. Based on the statistical analysis, the mRNA accumulation of both EF1 α and F-box were observed as stable within experiment 1; however, instability for both the genes was observed within experiment 2 and between experiments 1 and 2. This instability reflected the fact that outliers were removed from the data analysis, reducing the number of samples included in both experiments. These outliers were probably the result of pipetting error since small

volumes were pipetted manually. Further, one cycle variation in the C_t value of EF1 α and F-box between healthy and infected tissue average mean within experiments 1 and 2 would biologically reflect a two-fold difference in the template amount. But this difference is probably within the error of the experimental set up, rather than indicating instability. Thus, both EF1 α and F-box appears more stable between healthy and DsMV infected taro. It would be of value to compare mRNA accumulation of EF1 α and F-box in healthy and DsMV infected taro once more, particularly with a liquid handling robot to reduce pipeting errors.

Validation of EF1 α and F-box in other monocots should be carried out, as well as in dicots other than *A. thaliana*. In particular, primers have been designed in this study that amplifies these mRNAs from *N. benthamiana*. This species needs to be studied, particularly as it is an important model species for virus infection studies.

Quantification studies conducted on other monocot species, rice and wheat, showed that EF1 α was a stably expressed reference gene among ten most frequently used housekeeping genes such as UBG5, GAPDH, TUB and 18S rRNA in seedlings grown under different environmental conditions (Jain et al., 2006). Similarly, F-box was also identified as highly expressed gene among 15 other commonly used housekeeping genes in soybean (dicot) (Bilgin et al., 2008). Thus, these genes may be useful for a range of experimental conditions.

This study also attempted to analyse variation in the mRNA accumulation level of the PVIP gene between healthy and DsMV infected taro by normalising against the EF1 α and F-box reference genes. However, efficient amplification was not achieved from either healthy or DsMV infected taro using PVIP primers that had been designed for *N. benthamiana*. The use of these primers needs to be optimised. Alternatively, primers suitable for use specifically with monocots need to be designed. Further, the *N. benthamiana* primers did not amplify PVIP from *A. thaliana*, therefore generic dicot primers should also be designed. These primers would then need to be tested and conditions for their use need to be optimised.

Chapter 4

General discussion

PVIP has been shown to interact *in vitro* with the VPg protein attached at the 5' end of the potyviral genome (Dunoyer et al., 2004). The mRNA levels for this gene were shown to change upon changes in light, indicating this is gene responsive to abiotic stress (Anand, 2010). Given that the protein appears to interact with the virus, it was of interest to determine if it is also responsive to biotic stress specifically virus infection. The method of choice for measuring changes in mRNA accumulation is RT-qPCR, a technique which relies on appropriate reference genes for accurate quantitation. RT-qPCR has become the most common method of choice for measuring gene expression levels of mRNA. However, the accuracy of the result is greatly dependant on proper data normalisation. There are numerous variables in an RT-qPCR which needs to be optimised in order to differentiate experimentally induced variation from true biological changes. One such important variable is the use of appropriate reference gene (Mestdagh, et al., 2009) .

Lilly et al., (2011) identified four reference genes namely EF1 α , PDF2, F-box and SAND, which appeared appropriate for virus infection studies of Arabidopsis. It was also of interest to determine if these genes are suitable for other host plant species.

Traditionally, housekeeping genes such as GAPDH, TUA, UBQ were used as reference genes for quantification studies since they were believed to have stable expression levels across different treatments and tissues. However, later it was identified that these traditional housekeeping genes were not suitable for normalising gene expression in plants such as potato and *A. thaliana* (DeJong et al., 2007). According to the MIQE guidelines, a reference gene is a gene that remains stable over the course of an experiment. Normalisation of a gene of interest to the reference gene is necessary to avoid errors in the result caused by variation in sample amount, RNA integrity, cDNA and RT enzyme (Bustin, et al., 2009).

Thus, the aims of this research were two-fold: firstly, to confirm the suitability of these four reference genes in species other than Arabidopsis, and secondly to use these genes to quantify changes in PVIP mRNA accumulation in virus infected plants.

For the reference genes identified by Lilly et al., (2011) to be of general use for virus infection studies they would need to be validated in each plant host species being tested. For many plant species there is little sequence information available for

these genes, making species specific primer design challenging. Primers that are universal for dicot and / or monocot plants would be very useful for validating and then using these reference genes. Thus, primers were designed for EF1 α and F-box mRNA which may be universal primers. One EF1 α primer pair was designed that has the potential to amplify from both monocot and dicots plants. This primer pair could amplify EF1 α from the monocot taro and the dicot *N. benthamiana* and Arabidopsis. Two primer pairs were designed for F-box mRNA: one for monocot plant species and one for dicot species. The monocot primer pair only amplified from taro, while the dicot primer pair only amplified from *N. benthamiana* and Arabidopsis. These primers have the potential for use as universal primers, but would need to be tested against a broader panel of plant species. No universal primers were designed for PDF2 or SAND because regions of significant conservation could not be identified among homologous sequences of monocot and dicot species. Also a limited number of homologous monocot and dicot sequences were available in the publicly domain. Sequences comparisons showed that it may be possible to design SAND and PDF2 primers for dicot species, since the available homologous showed high conservation in the regions analysed. It may be more difficult to develop primers for monocot species. Primers would need to be developed for these genes; this may become possible with the continued sequencing of plant genomes.

Lilly et al., (2011) showed these four genes were stable between healthy and virus infected Arabidopsis and Jain et al., (2006), also found EF1 α and UBQ5 to be stably expressed in rice seedlings grown under various environmental conditions of stress and hormone treatments among the ten most frequently used housekeeping genes such as 18s rRNA, UBQ5, UBQ10, GAPDH, TUB and EF1 α analysed. Another study involved the analysis of six candidate reference genes including EF1 α , H3, ubiquitin-conjugating enzyme E2 in 442 different perennial ryegrass (*Lolium perenne* L.) samples collected from field and laboratory, grown under wide range of experimental conditions showed EF1 α as the stably expressed reference gene among all other six candidate reference genes analysed (Lee et al., 2010). However, EF1 α might not be a suitable reference gene for gene quantification analysis of all known plant virus infections. For instance, EF1 α was shown to have a fivefold change in gene expression level in Arabidopsis protoplasts upon *Plum pox virus* (PPV)

infection as well as genes related to protein synthesis, storage proteins, soluble sugar and chloroplast analysed (Babu et al, 2008) (Babu, Griffiths, Huang, & Wang, 2008). In this study, a general trend in the down regulation of proteins introduced in protein synthesis and translation was observed. EF1 α gene is a translation elongation factor; this change in expression stability may be due to the fact that EF1 α gene is encoded by a four member multigene family; different members exhibit differential expression (Lee et al., 2010).

F-box gene has also been shown to have varying expression stability. It was found that the F-box gene expression was increased in an analysis conducted on soya bean, where F-box was ranked as the most stable reference gene among 15 other housekeeping genes analysed such as TUB and GAPDH (Bilgin et al., 2008). However, similar to other housekeeping genes such as EF1 α , F-box was also shown to have decrease in expression stability when tested among other nine housekeeping genes such as 18S rRNA, actin and TUB in *Gossypium hirsutum* (Artico et al., 2010).

In this study, both EF1 α and F-box appear to be suitable as reference gene for virus infected taro.

Given taro is a monocot and Arabidopsis is a dicot, it is possible that the stability of EF1 α and F-box genes will hold for all plant species although this would need to be confirmed in different species infected with different viruses. For example, *N. benthamiana* is an important model species for virus infection studies, therefore it would be very important to validate these genes in this species.

RT-PCR amplification showed PVIP primers designed for *N. benthamiana* could amplify PVIP from taro inefficiently. It was thought this would be sufficient for detection by RT-qPCR since this method is considered to be the more sensitive technique. However, amplification of PVIP in RT-qPCR was not detectable until very late, suggesting this method was no more sensitive than non quantitative PCR. PVIP primers for amplification from monocot species would need to be developed to determine if this gene is responsive to virus infection. Indeed, using the same techniques as used for the reference genes, monocot and dicot specific primers could be designed to determine if this gene is responsive to infection in all species that are infected by a variety of potyvirids.

In summary primers for EF1 α and F-box that may have universal application to plant virus studies have been developed. A study has also been initiated for assessing changes in PVIP mRNA accumulation between healthy and virus challenged plants. Further, other techniques such as microarray or transcriptome sequencing analysis could also be conducted to verify any variations in the PVIP mRNA accumulation between healthy and virus infected plant leaf tissue.

References

- Anand, N. (2010) Characterisation of the Potyvirus VPg Interacting Protein (PVIP) gene from *Nicotiana benthamiana*. School of Applied Science. Auckland, AUT University. Thesis.
- Adams, M. J., Antoniw, J. F., & Fauquet, C. M. (2005). Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Archives of Virology*, *150*(3), 459-479. doi:10.1007/s00705-004-0440-6
- Aranda, M. A., Escaler, M., Wang, D., & Maule, A. J. (1996). Induction of HSP70 and polyubiquitin expression associated with plant virus replication. *Proceedings of the National Academy of Sciences*, *93*(26), 15289-15293.
- Arbatova, J., Lehto, K., Pehu, E., & Pehu, T. (1998). Localization of the P1 protein of potato Y potyvirus in association with cytoplasmic inclusion bodies and in the cytoplasm of infected cells. *Journal of General Virology*, *79*(10), 2319- 2323.
- Artico, S., Nardeli, S. M., Brilhante, O., Grossi-de-Sa, M. F., & Alves-Ferreira, M. (2010). Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biology*, *10*(1), 49-61. doi:10.1186/1471-2229-10-49
- Babu, M., Griffiths, J., Huang, T. S., & Wang, A. (2008). Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to *Plum pox virus* infection. *BMC Genomics*, *9*(1), 325-332. doi:10.1186/1471-2164-9-325
- Becker, C., Hammerle-Fickinger, A., Riedmaier, I., & Pfaffl, M. W. (2010). mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, *50*(4), 237-243. doi:10.1016/j.ymeth.2010.01.010
- Becker, C., Hammerle-Fickinger, A., Riedmaier, I., & Pfaffl, M. W. (2010). mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, *50*(4), 237-243. doi:10.1016/j.ymeth.2010.01.010
- Bilgin, D. D., Stacey, G., Benitez, M., Clough, S. J., Thibivilliers, S., Libault, M., et al. (2008). Identification of four soybean reference genes for gene expression normalization. *The Plant Genome*, *1*(1), 44-54. doi:10.3835/plantgenome2008.02.0091
- Bilgin, D. D., Stacey, G., Benitez, M., Clough, S. J., Thibivilliers, S., Libault, M., et al. (2008). Identification of four soybean reference genes for gene expression normalization. *The Plant Genome*, *1*(1), 44-54. doi:10.3835/plantgenome2008.02.0091
- Blackwell, J. L., & Brinton, M. A. (1997). Translation elongation factor-1 alpha interacts with the 3'stem-loop region of West Nile virus genomic RNA. *Journal of Virology*, *71*(9), 6433-6444.
- Bouck, A. M. Y., & Vision, T. (2007). The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology*, *16*(5), 907-924. doi:10.1111/j.1365-294X.2006.03195
- Bustamante, P. I., & Hull, R. (1998). Plant virus gene expression strategies. *Electronic Journal of Biotechnology*, *1*(2), 13-24.
- Bustin, S., Beaulieu, J. F., Huggett, J., Jaggi, R., Kibenge, F., Olsvik, P., et al. (2010). MIQE precis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology*, *11*(1), 74-79. doi:10.1186/1471-2199-11-74
- Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology*, *25*(2), 169-193. doi:10.1677/jme.0.0250169

- Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., et al. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, *55*(4), 611-622. doi:10.1373/clinchem.2008.112797
- Carrington, J. C., & Freed, D. D. (1990). Cap-independent enhancement of translation by a plant potyvirus 5'nontranslated region. *Journal of Virology*, *64*(4), 1590-1597.
- Carrington, J. C., Haldeman, R., Dolja, V. V., & Restrepo-Hartwig, M. A. (1993). Internal cleavage and trans-proteolytic activities of the VPg-proteinase (NIa) of tobacco etch potyvirus in vivo. *Journal of Virology*, *67*(12), 6995-7000.
- Chang, K., Mestdagh, P., Vandesompele, J., Kerin, M., & Miller, N. (2010). MicroRNA expression profiling to identify and validate reference genes for relative quantification in colorectal cancer. *BMC Cancer*, *10*(1), 173-185. doi:10.1186/1471-2407-10-173
- Chang (2012). Temporal and spatial analysis of *Dasheen Mosaic Virus*. School of Applied Science. Auckland, AUT University. Thesis.
- Chen, J., Li, W. X., Xie, D., Peng, J. R., & Ding, S. W. (2004). Viral virulence protein suppresses RNA silencing-mediated defense but upregulates the role of microRNA in host gene expression. *The Plant Cell Online*, *16*(5), 1302-1313. doi:10.1105/tpc.018986
- Chung, B. Y. W., Miller, W. A., Atkins, J. F., & Firth, A. E. (2008). An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences*, *105*(15), 5897-5902. doi:10.1073/pnas.0800468105
- Constantin, G. D., Krath, B. N., MacFarlane, S. A., Nicolaisen, M., Elisabeth Johansen, I., & Lund, O. S. (2004). Virus-induced gene silencing as a tool for functional genomics in a legume species. *The Plant Journal*, *40*(4), 622-631. doi:10.1111/j.1365-313X.2004.02233
- Cordoba, E. M., Die, J. V., Gonzaez-Verdejo, C. I., Nadal, S., & Roman, B. (2011). Selection of reference genes in "*Hedysarum coronarium*" under various stresses and stages of development. *Analytical Biochemistry*, *409*(2), 236-243. doi:10.1016/j.ab.2010.10.031
- Coulson, D., Brockbank, S., Quinn, J., Murphy, S., Ravid, R., Irvine, G. B., et al. (2008). Identification of valid reference genes for the normalisation of RT-qPCR gene expression data in human brain tissue. *BMC Molecular Biology*, *9*(1), 46-57. doi:10.1186/1471-2199-9-46
- Coux, O., Tanaka, K., & Goldberg, A. L. (1996). Structure and functions of the 20S and 26S proteasomes. *Annual Review of Biochemistry*, *65*(1), 801-847. doi:10.1146/annurev.bi.65.070196.004101
- Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K., & Scheible, W. R. (2005). Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiology*, *139*(1), 5-17. doi:10.1104/pp.105.063743
- Daras, J. A., & Carrington, J. C. (1997). RNA binding activity of NIa proteinase of Tobacco etch potyvirus. *Virology*, *237*(2), 327-336. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1006/viro.1997.8802
- Dawson, W. O., & Hilf, M. E. (1992). Host-range determinants of plant viruses. *Annual Review of Plant Biology*, *43*(1), 527-555.
- De Jonge, H. J. M., Fehrmann, R. S. N., De Bont, E. S. J. M., Hofstra, R. M. W., Gerbens, F., Kamps, W. A., et al. (2007). Evidence based selection of housekeeping genes. *PloS one*, *2*(9), 292-297.

- Derveaux, S., Vandesomepele, J., & Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods*, 50(4), 227-230. doi:10.1016/j.ymeth.2009.11.001
- D'haene, B., & Hellemans, J. (2010). The importance of quality control during qPCR data analysis. *International Drug Discovery*, 1, 18-24.
- Dheda, K., Huggett, J. F., Bustin, S. A., Johnson, M. A., Rook, G., & Zumla, A. (2004). Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques*, 37(1), 112-119.
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A., et al.(2011). Geneious v 5.4. <http://www.geneious.com/>
- Dufresne, P. J., Thivierge, K., Cotton, S., Beauchemin, C., Ide, C., Ubalijoro, E., et al. (2008). Heat shock 70 protein interaction with *Turnip mosaic virus* RNA-dependent RNA polymerase within virus-induced membrane vesicles. *Virology*, 374(1), 217-227. doi:10.1016/j.virol.2007.12.014
- Dunoyer, P., Thomas, C., Harrison, S., Revers, F., & Maule, A. (2004). A cysteine-rich plant protein potentiates potyvirus movement through an interaction with the virus genome-linked protein VPg. *Journal of Virology*, 78(5), 2301-2309. doi:10.1128/JVI.78.5.2301-2309.2004
- Erlich, H. A. (1989). *PCR technology. principles and applications for DNA amplification*: Stockton press.
- Eskelin, K., Hafren, A., Rantalainen, K. I., & Makinen, K. (2011). Potyviral VPg enhances viral RNA translation and inhibits reporter mRNA translation in planta. *Journal of Virology*, 85(17), 9210-9221. doi:10.1128/JVI.00052-11
- Feng, H., Huang, X., Zhang, Q., Wei, G., Wang, X., & Kang, Z. (2012). Selection of suitable inner reference genes for relative quantification expression of microRNA in wheat. *Plant Physiology and Biochemistry*, 151, 116-122. doi:10.1016/j.plaphy.2011.10.010
- Freeman, W. M., Walker, S. J., & Vrana, K. E. (1999). Quantitative RT-PCR: pitfalls and potential. *Biotechniques*, 26, 112-125.
- Gavin, T. P., & Wagner, P. D. (2002). Attenuation of the exercise-induced increase in skeletal muscle Flt-1 mRNA by nitric oxide synthase inhibition. *Acta Physiologica Scandinavica*, 175(3), 201-209. doi:10.1046/j.1365-201X.2002.00987
- Gilbert, S. F. (2010). *Reverse-Transcriptase Polymerase Chain Reaction (RT-PCR)* (9 ed.), Sinauer Associates.
- Goodfellow, I. (2011). The genome-linked protein VPg of vertebrate viruses-a multifaceted protein. *Current Opinion in Virology*, 1(5), 355-362. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/j.coviro.2011.09.003
- Goodin, M. M., Zaitlin, D., Naidu, R. A., & Lommel, S. A. (2008). *Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions. *Molecular Plant-Microbe Interactions*, 21(8), 1015-1026. doi:10.1094/MPMI-21-8-1015
- Gruez, A., Selisko, B., Roberts, M., Bricogne, G., Bussetta, C., Jabafi, I., et al. (2008). The crystal structure of coxsackievirus B3 RNA-dependent RNA polymerase in complex with its protein primer VPg confirms the existence of a second VPg binding site on Picornaviridae polymerases. *Journal of Virology*, 82(19), 9577-9590. doi:10.1128/JVI.00631-08
- Gutierrez, L., Mauriat, M., Guanin, S., Pelloux, J., Lefebvre, J. F., Louvet, R., et al. (2008). The lack of a systematic validation of reference genes: a serious

- pitfall undervalued in reverse transcription-polymerase chain reaction (RT-qPCR) analysis in plants. *Plant Biotechnology Journal*, 6(6), 609-618. doi:10.1111/j.1467-7652.2008.00346
- Hampton, R. Y., Gardner, R. G., & Rine, J. (1996). Role of 26S proteasome and HRD genes in the degradation of 3-hydroxy-3-methylglutaryl-CoA reductase, an integral endoplasmic reticulum membrane protein. *Molecular Biology of the Cell*, 7(12), 2029-2044.
- Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. *Genome Research*, 6(10), 986-994. doi:10.1101/gr.6.10.986
- Hjulsager, C. K., Olsen, B. S., Jensen, D. M. K., Cordea, M. I., Krath, B. N., Johansen, I. E., et al. (2006). Multiple determinants in the coding region of *Pea seed-borne mosaic virus* P3 are involved in virulence against *sbm-2* resistance. *Virology*, 355(1), 52-61.
- Hong, X. Y., Chen, J., Shi, Y. H., & Chen, J. P. (2007). The "6K1" protein of a strain of *Soybean mosaic virus* localizes to the cell periphery. *Archives of Virology*, 152(8), 1547-1551.
- Hong, Y., Levay, K., Murphy, J. F., Klein, P. G., Shaw, J. G., & Hunt, A. G. (1995). A potyvirus polymerase interacts with the viral coat protein and VPg in yeast cells. *Virology*, 214(1), 159-166. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1006/viro.1995.9944
- Huber, W., Von Heydebreck, A., Suelmann, H., Poustka, A., & Vingron, M. (2003). Parameter estimation for the calibration and variance stabilization of microarray data. *Statistical Applications in Genetics and Molecular Biology*, 2(1), 1008-1030.
- Huggett, J., Dheda, K., Bustin, S., & Zumla, A. (2005). Real-time RT-qPCR normalisation; strategies and considerations. *Genes and Immunity*, 6(4), 279-284. doi:10.1038/sj.gene.6364190
- Hwang, I. T., Kim, Y. J., Kim, S. H., Kwak, C. I., Gu, Y. Y., & Chun, J. Y. (2003). Annealing control primer system for improving specificity of PCR amplification. *Biotechniques*, 35(6), 1180-1191.
- Ishii, K., & Fukui, M. (2001). Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Applied and Environmental Microbiology*, 67(8), 3753-3755. doi:10.1128/AEM.67.8.3753-3755.2001
- Jain, M. (2009). Genome-wide identification of novel internal control genes for normalization of gene expression during various stages of development in rice. *Plant Science*, 176(5), 702-706. doi:10.1016/j.plantsci.2009.02.001
- Jain, M., Nijhawan, A., Tyagi, A. K., & Khurana, J. P. (2006). Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, 345(2), 646-651. doi:10.1016/j.bbrc.2006.04.140
- Jain, M., Nijhawan, A., Tyagi, A. K., & Khurana, J. P. (2006). Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, 345(2), 646-651. doi:10.1016/j.bbrc.2006.04.140
- Jiang, J., & Lalibertae, J. F. (2011). The genome-linked protein VPg of plant viruses—a protein with many partners. *Current Opinion in Virology*, 1(5), 347-354. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/j.coviro.2011.09.010

- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST : a better web interface. *Nucleic Acid Research*, 36(1), W5-W9. doi:10.1093/nar/gkn201
- Kang, S. H., Lim, W. S., & Kim, K. H. (2004). A protein interaction map of soybean mosaic virus strain G7H based on the yeast two-hybrid system. *Molecules and Cells*, 18(1), 122-126.
- Kasschau, K. D., & Carrington, J. C. (1998). A counterdefensive strategy of plant viruses: suppression of posttranscriptional gene silencing. *Cell*, 95(4), 461-470.
- Kasschau, K. D., Cronin, S., & Carrington, J. C. (1997). Genome amplification and long-distance movement functions associated with the central domain of tobacco etch potyvirus helper component-proteinase. *Virology*, 228(2), 251-262.
- Kauffmann, A., Gentleman, R., & Huber, W. (2009). arrayQualityMetrics-a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25(3), 415-416. doi:10.1093/bioinformatics/btn647
- Khan, M. A., Miyoshi, H., Ray, S., Natsuaki, T., Suehiro, N., & Goss, D. J. (2006). Interaction of genome-linked protein (VPg) of turnip mosaic virus with wheat germ translation initiation factors eIFiso4E and eIFiso4F. *Journal of Biological Chemistry*, 281(38), 28002-28010. doi:10.1074/jbc.M605479200
- Kulcheski, F., de Oliveira, L., Molina, L., Almerao, M., Rodrigues, F., Marcolino, J., et al. (2011). Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics*, 12(1), 307-324. doi:10.1186/1471-2164-12-307
- Kulcheski, F. R., Marcelino-Guimaraes, F. C., Nepomuceno, A. L., Abdelnoor, R. V., & Margis, R. (2010). The use of microRNAs as reference genes for quantitative polymerase chain reaction in soybean. *Analytical Biochemistry*, 406(2), 185-192. doi:10.1016/j.ab.2010.07.020
- Lee, J. M., Roche, J. R., Donaghy, D. J., Thrush, A., & Sathish, P. (2010). Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (*Lolium perenne* L.). *BMC Molecular Biology*, 11(1), 8-22. doi:10.1186/1471-2199-11-8
- Lekanne Deprez, R. H., Fijnvandraat, A. C., Ruijter, J. M., & Moorman, A. F. M. (2002). Sensitivity and accuracy of quantitative real-time polymerase chain reaction using SYBR green I depends on cDNA synthesis conditions. *Analytical Biochemistry*, 307(1), 63-69. doi:10.1016/S0003-2697(02)00021-0
- Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M. G., & Laliberte, J. F. (2000). Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. *Journal of Virology*, 74(17), 7730-7737. doi:10.1128/JVI.74.17.7730-7737.2000
- Lilly, S. T., Drummond, R. S. M., Pearson, M. N., & MacDiarmid, R. M. (2011). Identification and validation of reference genes for normalization of transcripts from virus-infected *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions*, 24(3), 294-304. doi:10.1094/MPMI-10-10-0236
- Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $^{-\Delta\Delta CT}$ method. *Methods*, 25(4), 402-408. doi:10.1006/meth.2001.1262
- Lovdal, T., & Lillo, C. (2009). Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. *Analytical Biochemistry*, 387(2), 238-242. doi:10.1016/j.ab.2009.01.024

- Maia, I. G., Haenni, A. L., & Bernardi, F. (1996). Potyviral HC-Pro: a multifunctional protein. *Journal of General Virology*, *77*(7), 1335-1341.
- Malinen, E., Kassinen, A., Rinttila, T., & Palva, A. (2003). Comparison of real-time PCR with SYBR Green I or 5'-nuclease assays and dot-blot hybridization with rDNA-targeted oligonucleotide probes in quantification of selected faecal bacteria. *Microbiology*, *149*(1), 269-277. doi:10.1099/mic.0.25975-0
- Marone, M., Mozzetti, S., De Ritis, D., Pierelli, L., & Scambia, G. (2001). Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample. *Biological Procedures Online*, *3*(1), 19-25. doi:10.1251/bpo20
- Maroufi, A., Van Bockstaele, E., & De Loose, M. (2010). Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Molecular Biology*, *11*(1), 15. doi:10.1186/1471-2199-11-15
- Martanez, N., Cruz Martin, M., Herrero, A., Fernandez, M., Alvarez, M. A., & Ladero, V. (2011). qPCR as a powerful tool for microbial food spoilage quantification: Significance for food quality. *Trends in Food Science & Technology*, *22*(7), 367-376.
- Mascia, T., Santovito, E., Gallitelli, D., & Cillo, F. (2010). Evaluation of reference genes for quantitative reverse-transcription polymerase chain reaction normalization in infected tomato plants. *Molecular Plant Pathology*, *11*(6), 805-816. doi:10.1111/j.1364-3703.2010.00646
- Mathioudakis, M. M., Veiga, R., Ghita, M., Tsikou, D., Medina, V., Canto, T., et al. (2012). Pepino mosaic virus capsid protein interacts with a tomato heat shock protein cognate 70. *Virus Research*, *163*(1), 28-39. doi:10.1016/j.virusres.2011.08.007
- Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., & Koornneef, M. (1998). *Arabidopsis thaliana*: a model plant for genome analysis. *Science*, *282*(5389), 662-682. doi:10.1126/science.282.5389.662
- Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., et al. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biology*, *10*(6), R641-610. doi:10.1186/gb-2009-10-6-r64
- Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., et al. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biology*, *10*(6), R64.61-64.10. doi:10.1186/gb-2009-10-6-r64
- Moreno, I., Gruissem, W., & Vanderschuren, H. (2011). Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. *Journal of Virological Methods*, *177*(1), 49-54. doi:10.1016/j.jviromet.2011.06.013
- Morozova, O., Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual Review of Genomics and Human Genetics*, *10*, 135-151. doi:10.1146/annurev-genom-082908-145957
- Nagaraj, S. H., Gasser, R. B., & Ranganathan, S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in Bioinformatics*, *8*(1), 6-21. doi:10.1093/bib/bbl015
- Nagy, P. D., Wang, R. Y., Pogany, J., Hafren, A., & Makinen, K. (2011). Emerging picture of host chaperone and cyclophilin roles in RNA virus replication. *Virology*, *411*(2), 374-382. doi:10.1016/j.virol.2010.12.061

- Ng, J. C. K., & Perry, K. L. (2004). Transmission of plant viruses by aphid vectors. *Molecular Plant Pathology*, 5(5), 505-511. doi:10.1111/j.1364-3703.2004.00240.x
- Nolan, T., Hands, R. E., & Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature protocols*, 1(3), 1559-1582. doi:10.1038/nprot.2006.236
- O'Hair, S. K., & Asokan, M. P. (1986). Edible aroids: botany and horticulture. *Horticultural Reviews*, 8(1), 43-99. doi:10.1002/9781118060810.ch2
- O'Reilly, E. K., Paul, J. D., & Kao, C. C. (1997). Analysis of the interaction of viral RNA replication proteins by using the yeast two-hybrid assay. *Journal of Virology*, 71(10), 7526-7532.
- Page, A. F., & Minocha, S. C. (2004). Analysis of gene expression in transgenic plants. *Methods in Molecular Biology*, 286, 291-312.
- Palacios, I., Drucker, M., Blanc, S., Leite, S., Moreno, A., & Fereres, A. (2002). Cauliflower mosaic virus is preferentially acquired from the phloem by its aphid vectors. *Journal of General Virology*, 83(12), 3163-3171.
- Pall, G. S., Codony-Servat, C., Byrne, J., Ritchie, L., & Hamilton, A. (2007). Carbodiimide-mediated cross-linking of RNA to nylon membranes improves the detection of siRNA, miRNA and piRNA by northern blot. *Nucleic Acids Research*, 35(8), 60-71. doi:10.1093/nar/gkm112
- Paolacci, A., Tanzarella, O., Porceddu, E., & Ciaffi, M. (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. *BMC Molecular Biology*, 10(1), 11-38. doi:10.1186/1471-2199-10-11
- Parker, R. M. C., & Barnes, N. M. (1999). mRNA: detection by in situ and northern hybridization. *Methods in Molecular Biology*, 106, 247-283. doi:10.1385/0-89603-530-1:247
- Parkinson, J., & Blaxter, M. (2009). Expressed sequence tags: an overview. *Methods Molecular Biology*, 533, 1-12. doi:10.1007/978-1-60327-136-3_1
- Peirson, S. N., Butler, J. N., & Foster, R. G. (2003). Experimental validation of novel and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Research*, 31(14), 73-79. doi:10.1093/nar/gng073
- Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research*, 29(9), 2003-2007. doi:10.1093/nar/29.9.e45
- Pfaffl, M. W., & Berg, W. (2010). Guest editor's introduction: the ongoing evolution of qPCR. *Methods*, 50(4), 215-336.
- Pirone, T. P., & Perry, K. L. (2002). Aphids: non-persistent transmission. *Advances in Botanical Research*, 36, 1-19. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/S0065-2296(02)36056-7
- Plisson, C., Drucker, M., Blanc, S., German-Retana, S., Le Gall, O., Thomas, D., et al. (2003). Structural characterization of HC-Pro, a plant virus multifunctional protein. *Journal of Biological Chemistry*, 278(26), 23753-23761. doi:10.1074/jbc.M302512200
- Popovici, V., Goldstein, D. R., Antonov, J., Jaggi, R., Delorenzi, M., & Wirapati, P. (2009). Selecting control genes for RT-qPCR using public microarray data. *BMC Bioinformatics*, 10(1), 42-52. doi:10.1186/1471-2105-10-42
- Puustinen, P., & Makinen, K. (2004). Uridylylation of the potyvirus VPg by viral replicase NIb correlates with the nucleotide binding capacity of VPg. *Journal of Biological Chemistry*, 279(37), 38103-38110.

- Quackenbush, J. (2002). Microarray data normalization and transformation. *Nature Genetics*, 32(18), 496-501.
- Radonic, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, 313(4), 856-862. doi:10.1016/j.ymeth.2010.01.010
- Radonic, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, 313(4), 856-862. doi:10.1016/j.bbrc.2003.11.177
- Rajamaki, M. L., Kelloniemi, J., Alminaitte, A., Kekarainen, T., Rabenstein, F., & Valkonen, J. (2005). A novel insertion site inside the potyvirus P1 cistron allows expression of heterologous proteins and suggests some P1 functions. *Virology*, 342(1), 88-101.
- Rantalainen, K. I., Uversky, V. N., Permi, P., Kalkkinen, N., Dunker, A. K., & Makinen, K. (2008). *Potato virus A* genome-linked protein VPg is an intrinsically disordered molten globule-like protein with a hydrophobic core. *Virology*, 377(2), 280-288. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/j.virol.2008.04.025
- Ravi, V., Aked, J., & Balagopalan, C. (1996). Review on tropical root and tuber crops I. Storage methods and quality changes. *Critical Reviews in Food Science & Nutrition*, 36(7), 661-709. doi:10.1080/10408399609527744
- Riechmann, J. L., Cervera, M. T., & Garcia, J. A. (1995). Processing of the plum pox virus polyprotein at the P3-6K1 junction is not required for virus viability. *Journal of General Virology*, 76(4), 951-956.
- Riechmann, J. L., Lain, S., & Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology*, 73(1), 1-16.
- Riechmann, J. L., Lain, S., & Garcia, J. A. (1992). Highlights and prospects of potyvirus molecular biology. *Journal of General Virology*, 73(1), 1-16.
- Ritz, M., Garenaux, A., Berge, M., & Federighi, M. (2009). Determination of rpoA as the most suitable internal control to study stress response in *C. jejuni* by RT-qPCR and application to oxidative stress. *Journal of Microbiological Methods*, 76(2), 196-200. doi:10.1016/j.mimet.2008.10.014
- Rojas, M. R., Zerbini, F. M., Allison, R. F., Gilbertson, R. L., & Lucas, W. J. (1997). Capsid protein and helper component-proteinase function as potyvirus cell-to-cell movement proteins. *Virology*, 237(2), 283-295. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1006/viro.1997.8777
- Roudet-Tavert, G., Michon, T., Walter, J., Delaunay, T., Redondo, E., & Le Gall, O. (2007). Central domain of a potyvirus VPg is involved in the interaction with the host translation initiation factor eIF4E and the viral protein HcPro. *Journal of General Virology*, 88(3), 1029-1033. doi:10.1099/vir.0.82501-0
- Rychlik, W., Spencer, W. J., & Rhoads, R. E. (1990). Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Research*, 18(21), 6409-6412. doi:10.1093/nar/18.21.6409
- Saiga, S., Furumizu, C., Yokoyama, R., Kurata, T., Sato, S., Kato, T., et al. (2008). The Arabidopsis OBERON1 and OBERON2 genes encode plant homeodomain finger proteins and are required for apical meristem maintenance. *Development*, 135(10), 1751-1759. doi:10.1242/dev.014993
- Schaad, M. C., Haldeman-Cahill, R., Cronin, S., & Carrington, J. C. (1996). Analysis of the VPg-proteinase (NIa) encoded by *Tobacco etch potyvirus*:

- effects of mutations on subcellular transport, proteolytic processing, and genome amplification. *Journal of Virology*, 70(10), 7039-7048.
- Schaad, M. C., Jensen, P. E., & Carrington, J. C. (1997). Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. *The EMBO Journal*, 16(13), 4049-4059.
- Schaad, M. C., Jensen, P. E., & Carrington, J. C. (1997). Formation of plant RNA virus replication complexes on membranes: role of an endoplasmic reticulum-targeted viral protein. *The EMBO journal*, 16(13), 4049-4059. doi:10.1093/emboj/16.13.4049
- Schmittgen, T. D., & Zakrajsek, B. A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *Journal of Biochemical and Biophysical Methods*, 46(1), 69-81. doi:10.1016/S0165-022X(00)00129-9
- Shaw, A. E., Reid, S. M., Ebert, K., Hutchings, G. H., Ferris, N. P., & King, D. P. (2007). Implementation of a one-step real-time RT-PCR protocol for diagnosis of foot-and-mouth disease. *Journal of Virological Methods*, 143(1), 81-85. doi:10.1016/j.jviromet.2007.02.009
- Shuklo, D. D., & Ward, C. W. (1989). Structure of Potyvirus Coat Proteins and its Application in the Taxonomy of the *Potyvirus* Group. *Advances in Virus Research*, 36, 273-312.
- Spetz, C., & Valkonen, J. P. T. (2004). Potyviral 6K2 protein long-distance movement and symptom-induction functions are independent and host-specific. *Molecular Plant-Microbe Interactions*, 17(5), 502-510.
- Stahlberg, A., Hakansson, J., Xian, X., Semb, H., & Kubista, M. (2004). Properties of the reverse transcription reaction in mRNA quantification. *Clinical Chemistry*, 50(3), 509-515. doi:10.1373/clinchem.2003.026161
- Stankovic, K. M., & Corfas, G. (2003). Real-time quantitative RT-PCR for low-abundance transcripts in the inner ear: analysis of neurotrophic factor expression. *Hearing Research*, 185(1-2), 97-108. doi:10.1016/S0378-5955(03)00298-3
- Stanton, L. W. (2001). Methods to profile gene expression. *Trends in Cardiovascular Medicine*, 11(2), 49-54. doi:10.1016/S1050-1738(01)00085-8
- Stenger, D. C., French, R., & Gildow, F. E. (2005). Complete deletion of Wheat streak mosaic virus HC-Pro: a null mutant is viable for systemic infection. *Journal of Virology*, 79(18), 12077-12080.
- Syller, J. (2006). The roles and mechanisms of helper component proteins encoded by potyviruses and caulimoviruses. *Physiological and Molecular Plant Pathology*, 67(3), 119-130.
- Taneyhill, L. A., & Adams, M. S. (2008). Investigating Regulatory Factors and Their DNA Binding Affinities Through Real Time Quantitative PCR (RT-qPCR) and Chromatin Immunoprecipitation (ChIP) Assays. *Methods in Cell Biology*, 87(2), 367-389. doi:10.1016/S0091-679X(08)00219-7
- Tatineni, S., Ziems, A. D., Wegulo, S. N., & French, R. (2009). Triticum mosaic virus: A distinct member of the family Potyviridae with an unusually long leader sequence. *Phytopathology*, 99(8), 943-950.
- Taylor, S., Wakem, M., Dijkman, G., Alsarraj, M., & Nguyen, M. (2010). A practical approach to RT-qPCR--Publishing data that conform to the MIQE guidelines. *Methods*, 50(4), 81-85. doi:10.1016/j.ymeth.2010.01.005

- Teycheney, P. Y., Aaziz, R., Dinant, S., Salanki, K., Tourneur, C., Balazs, E., et al. (2000). Synthesis of (-)-strand RNA from the 3'untranslated region of plant viral genomes expressed in transgenic plants upon infection with related viruses. *Journal of General Virology*, *81*(4), 1121-1126.
- Thellin, O., ElMoualij, B., Heinen, E., & Zorzi, W. (2009). A decade of improvements in quantification of gene expression and internal standard selection. *Biotechnology Advances*, *27*(4), 323-333. doi:10.1016/j.biotechadv.2009.01.010
- Thomas, F., Barbeyron, T., & Michel, G. (2011). Evaluation of reference genes for real-time quantitative PCR in the marine flavobacterium *Zobellia galactanivorans*. *Journal of Microbiological Methods*, *84*(1), 61-66. doi:10.1016/j.mimet.2010.10.016
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Research*, *22*(22), 4673-4680.
- Tong, Z., Gao, Z., Wang, F., Zhou, J., & Zhang, Z. (2009). Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Molecular Biology*, *10*(1), 71-84. doi:10.1186/1471-2199-10-71
- Trayhurn, P., Duncan, J. S., Nestor, A., Thomas, M. E. A., Eastmond, N. C., & Rayner, D. V. (1995). Rapid chemiluminescent detection of mRNAs on Northern blots with digoxigenin end-labelled oligonucleotides. *Electrophoresis*, *16*(1), 341-344. doi:10.1002/elps.1150160157
- Ullah, Z., Chai, B., Hammar, S., Raccach, B., Gal-On, A., & Grumet, R. (2003). Effect of substitution of the amino termini of coat proteins of distinct potyvirus species on viral infectivity and host specificity. *Physiological and Molecular Plant Pathology*, *63*(3), 129-139. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/j.pmpp.2003.11.001
- Urcuqui-Inchima, S., Haenni, A. L., & Bernardi, F. (2001). Potyvirus proteins: a wealth of functions. *Virus Research*, *74*(1-2), 157-175. doi:http://dx.doi.org/10.1016/S0168-1702(01)00220-9
- Valoczi, A., Hornyik, C., Varga, N., Burgyan, J., Kauppinen, S., & Havelda, Z. (2004). Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Research*, *32*(22), 175-182. doi:10.1093/nar/gnh171
- VanGuilder, H. D., Vrana, K. E., & Freeman, W. M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, *44*(5), 619-627. doi:10.2144/000112776
- Verchot, J., Koonin, E. V., & Carrington, J. C. (1991). The 35-kDa protein from the N-terminus of the potyviral polyprotein functions as a third virus-encoded proteinase. *Virology*, *185*(2), 527-535. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/0042-6822(91)90522-D
- Wang, D., & Maule, A. J. (1995). A model for seed transmission of a plant virus: Genetic and structural analyses of pea embryo invasion by *Pea seed-borne mosaic virus*. *The Plant Cell Online*, *6*(6), 777-787. doi:10.1105/tpc.6.6.777
- Watson, S., Mercier, S., Bye, C., Wilkinson, J., Cunningham, A. L., & Harman, A. N. (2007). Determination of suitable housekeeping genes for normalisation of quantitative real time PCR analysis of cells infected with human immunodeficiency virus and herpes viruses. *Virology Journal*, *4*(1), 130-135. doi:10.1186/1743-422X-4-130

- Wei, T., Zhang, C., Hong, J., Xiong, R., Kasschau, K. D., Zhou, X., et al. (2010). Formation of complexes at plasmodesmata for potyvirus intercellular movement is mediated by the viral protein P3N-PIPO. *PLoS Pathogens*, 6(6), e1000962. doi:doi:10.1371/journal.ppat.1000962
- Wen, R. H., & Hajimorad, M. R. (2010). Mutational analysis of the putative pipo of soybean mosaic virus suggests disruption of PIPO protein impedes movement. *Virology*, 400(1), 1-7. doi:http://dx.doi.org.ezproxy.aut.ac.nz/10.1016/j.virol.2010.01.022
- Whitham, S. A., Quan, S., Chang, H. S., Cooper, B., Estes, B., Zhu, T., et al. (2003). Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *The Plant Journal*, 33(2), 271-283. doi:10.1046/j.1365-313X.2003.01625
- Whitham, S. A., & Wang, Y. (2004). Roles for host factors in plant viral pathogenicity. *Current Opinion in Plant Biology*, 7(4), 365-371. doi:10.1016/j.pbi.2004.04.006
- Wicks, J., Haitchi, H. M., Holgate, S. T., Davies, D. E., & Powell, R. M. (2006). Enhanced upregulation of smooth muscle related transcripts by TGF β 2 in asthmatic (myo) fibroblasts. *Thorax*, 61(4), 313-319. doi:10.1136/thx.2005.05000
- Wilkie, G. S., Dickson, K. S., & Gray, N. K. (2003). Regulation of mRNA translation by 5'-and 3'-UTR-binding factors. *Trends in Biochemical Sciences*, 28(4), 182-188. doi:10.1016/S0968-0004(03)00051-3
- Yuan, J., Reed, A., Chen, F., & Stewart, C. N. (2006). Statistical analysis of real-time PCR data. *BMC Bioinformatics*, 7(1), 85-97. doi:10.1186/1471-2105-7-85
- Zettler, F. W., & Hartman, R. D. (1987). Dasheen mosaic virus as a pathogen of cultivated aroids and control of the virus by tissue culture. *Plant Disease*, 71(11), 958-963.
- Zhu, L., & Altmann, S. W. (2005). mRNA and 18S-RNA coapplication-reverse transcription for quantitative gene expression analysis. *Analytical Biochemistry*, 345(1), 102-109. doi:10.1016/j.ab.2005.07.028
- Ziegler-Graff, V., & Brault, V. (2008). Role of vector-transmission proteins. *Methods in Molecular Biology*, 451(1), 81-96. doi:10.1007/978-1-59745-102-4_6

Appendices

Appendix 1

Gene	ID code	Species
EF1 α	145359481	<i>A. thaliana</i>
EF1 α	396133	<i>N. tabacum</i>
EF1 α	167040	<i>N. tabacum</i>
EF1 α	156257584	<i>S.lycopersicum</i>
EF1 α	54651608	<i>S.lycopersicum</i>
EF1 α	29718609	<i>N.paniculata</i>
EF1 α	19560691	<i>S.tuberosum</i>
EF1 α	23801402	<i>L.esculenta</i>
EF1 α	949877	<i>S.lycopersicum</i>
EF1 α	3265112	<i>O.sativa</i>
EF1 α	170775	<i>O.sativa</i>
EF1 α	2662342	<i>Zea mays</i>
EF1 α	1154510	<i>O.sativa</i>
EF1 α	3064160	<i>Barley</i>
EF1 α	115410	<i>Saccharum officinarum</i>
EF1 α	2996095	<i>H.vulgare</i>
EF1 α	1154510	<i>H.vulgare</i>
EF1 α	3064159	<i>Saccharum hybrid</i>
EF1 α	2662344	<i>Zea mays</i>
EF1 α	3264939	<i>Zea mays</i>
EF1 α	1514193	<i>Zea mays</i>
PDF2	19563459	<i>A. thaliana</i>
PDF2	16246357	<i>Zea mays</i>
PP2A	21988824	<i>N. tabacum</i>
PDF2	54653687	<i>S.lycopersicum</i>
PDF2	21206829	<i>P.trichocarpa</i>
PDF2	15141979	<i>Glycine max</i>
PDF2	32129024	<i>O.staiva</i>
PDF2	11547815	<i>Zea mays</i>
PDF2	5679683	<i>A. thaliana</i>
PDF2	2264929	<i>Zea mays</i>
SAND	2380103	<i>A. thaliana</i>
SAND	23801470	<i>Zea mays</i>
SAND	22649130	<i>O.sativa</i>
SAND	23800729	<i>Zea mays</i>
SAND	24206007	<i>Sorgum bicolor</i>
SAND	11544256	<i>Populus trichocarpa</i>
SAND	18650378	<i>A. thaliana</i>
F-box	18652316	<i>A. thaliana</i>
F-box	21272329	<i>S.lycopersicum</i>
F-box	19469336	<i>Zea mays</i>
F-box	24207470	<i>O.sativa</i>
F-box	29760353	<i>Zea mays</i>
F-box	U451474	<i>N. tabacum</i>
F-box	U464651	<i>N. tabacum</i>

Appendix 2

Clone ID	Primer pair	Sequence
AE26	EF1 α MD	5'GCCTTGGAGTACTTGGGGGTGGTGGCATCCATCTTGTCCAACATGAAATCCTCAGCTT GAATGAAAGGTGAATGCTTTCACACCGGGTGTGAAGGCCTGAACCTGATGAGAGATA CGGATACCTTGTTCAGCATGATGTAATCTGCGAGA 3'
AE27	EF1 α MD	5'GCCTGAATCCTTGGGGGTGATTTCAATGTTGATGTGCCAGTGATGTAGTCCTGATAA TCACTATTGCATTGCCTGCCGCTGCAGGTGCACCATATGGCAGAGCTCCCAACGCGTT GGATGCATAGCTTGTGTATTCTTTAATGTCACT 3'
TE16	EF1 α MD	5'ATCATTGACTCCACCACTGGTGGTTTTGAGGCCGGTATCTCGAAGGATGGACAGACCCG TGAGCATGCTCTGCTTGCCTTACCCTTGGAGTCAAGCACATGATCTGCTGCTGCAACAA GATGGATGCCACCACCCAAGTACTCCAAGG 3'
TE19	EF1 α MD	5'GCCTTGGATACTTGGGGGTGGTTTTTCCACCTTGATGCAGCAGCAGATCCTCAGCTTGAC TCCAGGGGTGATTGCAATCACACTAGGCTCACAGGTCAAGTCCCTCCTTCTACATACCGAC CTCAGCCCCCCTGTTGACTCAATGATA 3'
NE1	EF1 α MD	5'ATCATTGACTCCACCACTGGTGGTTTTGAAGCTGGTATCTCTAAGGATGGACAGACCCGT GAACATGCATTGCTTGTCTTTCACCTTGGTGTCAAGCAAATGATTTGCTGCTGTAACAAG ATGGATGCTACCACCCCAAGTACTCCAAGG 3'
NE6	EF1 α MD	5'ATCATTGACTCCACCACTGGTGGTTTTGAAGCTGGTATTTCCAAGGATGGTCAAGACCCGT GAGCACGCATTGCTTGTCTTTCACCTTGGTGTCAAGCAAATGATTTGCTGCTGCAACAAG ATGGATGCTACCACCCCAAGTACTCCAAGG 3'
TF7	F-boxM	5' CAGCAATCCATGACCATACTTTCTTTTCAACATCAAACAGAAGGCCCTTACCTTGGTTCC ATGATGTGAAACATATTAAGTTATCCTGTCCAAAGCACT 3'
TF9	F-boxM	5'CCAGCAATCCATGACCATACTTTCTTTTCAACATCAAACAGAAGGCCCTTACCTTGGTTCC ATGATGTGAAACATATTAAGTTATCCTGTCCAAAGCACT 3'
NF12	F-boxD	5'CACTAGTGGTTCGGCAATCCAAGACCACACCTTCTTATCGACATCATAAAGAAGGCCTTT GCCTTGGTTCCAAGATGTGAAGCATATTAGGTTATCCTGT 3'
NF15	F-boxD	5'GAATGCTTTGGACAGGATAACCTAATATGCTTACATCTTGAACCAAGGCAAAGGCCTT CTTTATGATGTCGATAAGAAGGTGTGGTCTTGGATTGCCG 3'
AF24	F-boxD	5'GGCTGAGAGGTTGAGTGTGGTGGACAAGATAATTTGATCTGCTTACGTCTTGAATCA AGGAAAAGTCTTCTATACAATGTGGATAAAAAAATTTGGTCTT 3'

Appendix 3

Experiment 1

Quantification SYBR

Pos	Name	Ct SYBR	Ct Mean SYBR	Ct Dev. SYBR	Amount SYBR	Target SYBR	
7	A1	HlthyTaro1	23.78	24.16	0.32	-	f-boxmf
7	A2	HlthyTaro1	24.37	24.16	0.32	-	f-boxmf
7	A3	HlthyTaro1	24.32	24.16	0.32	-	f-boxmf
7	A5	Hlthy Taro1	23.22	23.50	0.25	-	ef1mdf
7	A6	Hlthy Taro1	23.67	23.50	0.25	-	ef1mdf
7	A7	Hlthy Taro1	23.61	23.50	0.25	-	ef1mdf
7	A10	Hlthy Taro1				-	pvip
7	A11	Hlthy Taro1				-	pvip
7	A12	Hlthy Taro1				-	pvip
7	B1	Hlthy Taro2	24.35	24.43	0.31	-	f-boxmf
7	B2	Hlthy Taro2	24.17	24.43	0.31	-	f-boxmf
7	B3	Hlthy Taro2	24.77	24.43	0.31	-	f-boxmf
7	B5	Hlthy Taro2	23.62	23.39	0.46	-	ef1mdf
7	B6	Hlthy Taro2	22.87	23.39	0.46	-	ef1mdf
7	B7	Hlthy Taro2	23.70	23.39	0.46	-	ef1mdf
7	B10	Hlthy Taro2				-	pvip
7	B11	Hlthy Taro2				-	pvip
7	B12	Hlthy Taro2	34.67			-	pvip
7	C1	Hlthy Taro3	23.78	24.13	0.34	-	f-boxmf
7	C2	Hlthy Taro3	24.15	24.13	0.34	-	f-boxmf
7	C3	Hlthy Taro3	24.47	24.13	0.34	-	f-boxmf
7	C5	Hlthy Taro3	22.92	23.30	0.33	-	ef1mdf
7	C6	Hlthy Taro3	23.53	23.30	0.33	-	ef1mdf
7	C7	Hlthy Taro3	23.45	23.30	0.33	-	ef1mdf
7	C10	Hlthy Taro3				-	pvip
7	C11	Hlthy Taro3				-	pvip
7	C12	Hlthy Taro3				-	pvip
7	E1	DeMV Infctd Taro	23.86	24.13	0.27	-	f-boxmf
7	E2	DeMV Infctd Taro	24.11	24.13	0.27	-	f-boxmf
7	E3	DeMV Infctd Taro	24.41	24.13	0.27	-	f-boxmf
7	E5	DeMV Infctd Taro	22.69	22.88	0.17	-	ef1mdf
7	E6	DeMV Infctd Taro	22.93	22.88	0.17	-	ef1mdf
7	E7	DeMV Infctd Taro	23.01	22.88	0.17	-	ef1mdf
7	E10	DeMV Infctd Taro	34.67	34.00	0.77	-	pvip
7	E11	DeMV Infctd Taro	34.18	34.00	0.77	-	pvip
7	E12	DeMV Infctd Taro	33.16	34.00	0.77	-	pvip

	F1	DsMV Infctd Taro	23.84	24.12	0.25	-	f-boxmf
	F2	DsMV Infctd Taro	24.23	24.12	0.25	-	f-boxmf
	F3	DsMV Infctd Taro	24.29	24.12	0.25	-	f-boxmf
	F5	DsMV Infctd Taro	23.03	22.84	0.19	-	ef1mdf
	F6	DsMV Infctd Taro	22.84	22.84	0.19	-	ef1mdf
	F7	DsMV Infctd Taro	22.64	22.84	0.19	-	ef1mdf
	F10	DsMV Infctd Taro	33.02	33.01	0.34	-	pvip
	F11	DsMV Infctd Taro	33.35	33.01	0.34	-	pvip
	F12	DsMV Infctd Taro	32.67	33.01	0.34	-	pvip
	G1	DsMV Infctd Taro	23.87	24.37	0.45	-	f-boxmf
	G2	DsMV Infctd Taro	24.75	24.37	0.45	-	f-boxmf
	G3	DsMV Infctd Taro	24.50	24.37	0.45	-	f-boxmf
	G5	DsMV Infctd Taro	22.77	22.80	0.04	-	ef1mdf
	G6	DsMV Infctd Taro	22.79	22.80	0.04	-	ef1mdf
	G7	DsMV Infctd Taro	22.84	22.80	0.04	-	ef1mdf
	G10	DsMV Infctd Taro	32.93	32.48	0.45	-	pvip
	G11	DsMV Infctd Taro	32.49	32.48	0.45	-	pvip
	G12	DsMV Infctd Taro	32.03	32.48	0.45	-	pvip
	H2	ntc	28.38			-	f-boxmf
	H3	ntc	-			-	f-boxmf
	H4	ntc	-			-	f-boxmf
	H6	ntc	-			-	ef1mdf
	H7	ntc	-			-	ef1mdf
	H8	ntc	-			-	ef1mdf
	H10	ntc	-			-	pvip
	H11	ntc	-			-	pvip
	H12	ntc	-			-	pvip

Experiment 2

Quantification SYBR

Pos	Name	Ct SYBR	Ct Mean SYBR	Ct Dev. SYBR	Amount SYBR	Target SYBR
A1	Hlthy Taro 1	25.87	25.95	0.90	-	f-boxmf
A2	Hlthy Taro 1	25.10	25.95	0.90	-	f-boxmf
A3	Hlthy Taro 1	26.88	25.95	0.90	-	f-boxmf
A5	Hlthy Taro 1	21.13	22.00	1.19	-	ef1mdf
A6	Hlthy Taro 1	21.51	22.00	1.19	-	ef1mdf
A7	Hlthy Taro 1	23.36	22.00	1.19	-	ef1mdf
A10	Hlthy Taro 1	31.94	31.70	0.49	-	pvip
A11	Hlthy Taro 1	32.03	31.70	0.49	-	pvip
A12	Hlthy Taro 1	31.14	31.70	0.49	-	pvip
B1	Hlthy Taro 2	25.38	25.27	0.26	-	f-boxmf
B2	Hlthy Taro 2	25.45	25.27	0.26	-	f-boxmf
B3	Hlthy Taro 2	24.96	25.27	0.26	-	f-boxmf
B5	Hlthy Taro 2	20.96	21.82	2.04	-	ef1mdf
B6	Hlthy Taro 2	24.15	21.82	2.04	-	ef1mdf
B7	Hlthy Taro 2	20.36	21.82	2.04	-	ef1mdf
B10	Hlthy Taro 2	31.53	31.29	0.25	-	pvip
B11	Hlthy Taro 2	31.29	31.29	0.25	-	pvip
B12	Hlthy Taro 2	31.04	31.29	0.25	-	pvip
C1	Hlthy Taro 3	25.40	24.92	0.76	-	f-boxmf
C2	Hlthy Taro 3	25.32	24.92	0.76	-	f-boxmf
C3	Hlthy Taro 3	24.05	24.92	0.76	-	f-boxmf
C5	Hlthy Taro 3	22.10	21.34	1.09	-	ef1mdf
C6	Hlthy Taro 3	21.83	21.34	1.09	-	ef1mdf
C7	Hlthy Taro 3	20.09	21.34	1.09	-	ef1mdf
C10	Hlthy Taro 3	30.99	30.08	1.23	-	pvip
C11	Hlthy Taro 3	28.68	30.08	1.23	-	pvip
C12	Hlthy Taro 3	30.59	30.08	1.23	-	pvip
E1	DsMV Infctd Taro	23.31	23.93	0.81	-	f-boxmf
E2	DsMV Infctd Taro	24.84	23.93	0.81	-	f-boxmf
E3	DsMV Infctd Taro	23.63	23.93	0.81	-	f-boxmf
E5	DsMV Infctd Taro	20.91	21.00	0.22	-	ef1mdf
E6	DsMV Infctd Taro	20.83	21.00	0.22	-	ef1mdf
E7	DsMV Infctd Taro	21.24	21.00	0.22	-	ef1mdf
E10	DsMV Infctd Taro	31.30	31.58	0.25	-	pvip
E11	DsMV Infctd Taro	31.73	31.58	0.25	-	pvip
E12	DsMV Infctd Taro	31.71	31.58	0.25	-	pvip

Pos	Name	Ct SYBR	Ct Mean SYBR	Ct Dev. SYBR	Amount SYBR	Target SYBR
2	F1	26.12	25.12	1.27	-	f-boxmf
2	F2	25.56	25.12	1.27	-	f-boxmf
2	F3	23.69	25.12	1.27	-	f-boxmf
2	F5	20.94	20.68	0.23	-	ef1mdf
2	F6	20.48	20.68	0.23	-	ef1mdf
2	F7	20.64	20.68	0.23	-	ef1mdf
2	F10	31.79	31.52	0.24	-	pvip
2	F11	31.45	31.52	0.24	-	pvip
2	F12	31.34	31.52	0.24	-	pvip
2	G1	23.65	23.82	0.34	-	f-boxmf
2	G2	24.21	23.82	0.34	-	f-boxmf
2	G3	23.59	23.82	0.34	-	f-boxmf
2	G5	20.34	20.22	0.21	-	ef1mdf
2	G6	20.35	20.22	0.21	-	ef1mdf
2	G7	19.98	20.22	0.21	-	ef1mdf
2	G10	31.32	31.22	0.09	-	pvip
2	G11	31.15	31.22	0.09	-	pvip
2	G12	31.20	31.22	0.09	-	pvip
-	H2	-	-	-	-	f-boxmf
-	H3	-	-	-	-	f-boxmf
-	H4	-	-	-	-	f-boxmf
-	H6	-	-	-	-	ef1mdf
-	H7	-	-	-	-	ef1mdf
-	H8	-	-	-	-	ef1mdf
-	H10	-	-	-	-	pvip
-	H11	-	-	-	-	pvip
-	H12	-	-	-	-	pvip