

ABSTRACT

THAMMARAT, PHANIT. The Regulation of *Nicotiana benthamiana* Gene Expression at the Early Stages of *Red clover necrotic mosaic virus* Infection. (Under the direction of Steven A. Lommel and Dahlia M. Nielsen.)

The success of plant viruses as pathogens depends on their ability to recruit host factors to support their propagation whilst defeating host defense mechanisms early in the infection process. While the virus reprograms plant physiological processes during this critical early period, many of the events and interactions taking place remain unknown. In this thesis, I utilized *Nicotiana benthamiana* (Nb, a model host for most plant viruses) and *Red clover necrotic mosaic virus* (RCNMV, a typical plus strand RNA plant virus) to elucidate these early events in the host-virus interaction. The Nb unigene collection, a custom microarray representing 13,413 host genes (~38% of the Nb transcriptome) was created for this study along with a corresponding functional annotation resource. The performance of the Nb array demonstrated its potential usage as a genomic tool in a genome wide study.

This thesis is the first study to examine the host transcriptome at the earliest stages of the virus infection process of 2, 6, 12 and 24 hours post-inoculation (hpi). The 1,654 host genes exhibited differential expression at an FDR cutoff of 0.01. This global snapshot of gene expression revealed that host genes are significantly down-regulated at 2, 6 and 24 hpi and significantly up-regulated at 12 hpi. This suggests that (i) host gene expression was suppressed as early as 2 hpi, and (ii) one infection cycle within the primary infected cell takes 12-24 hours. Gene set enrichment analysis (GSEA) revealed that RCNMV affected the following key host functions/pathways: defense, translation, photosynthesis and chloroplast-related functions, carbon fixation (Calvin cycle), metabolism, multidrug and toxin extrusion (MATE) transporter, cell wall-associated functions and protein kinases.

Four host genes representing various plant biological functions were transiently silenced or over-expressed to further functionally analyze their impact(s) on the RCNMV infection process. The genes chosen were: 1) NbSAR (systemic acquired resistance), the only host gene in the microarray study that was significantly regulated across all four time points (down regulated in the first 6 hours followed by up regulation), 2) NbWRKY (putative transcription factor) which was up-regulated (10 fold) at 24 hpi, 3) NbBTF3 (basal transcription factor 3) which was up-regulated (1.5 fold) at 6 hpi and 4) NbSI (C-8, 7- sterol isomerase) which was up-regulated (1.2 fold) at 12 hpi.

Pre-silencing of NbSAR decreased the replication levels of RCNMV RNAs suggesting that a certain NbSAR level may be required for RCNMV replication. NbWRKY is a novel WRKY gene that was first identified and characterized in this thesis. The putative NbWRKY polypeptide (325 amino acids in length) is most similar to potato StWRKY6, possessing a single WRKY DNA binding domain, four β -stranded sheet structure and a nuclear localization signal which was confirmed by cellular localization studies with a GFP fusion. NbWRKY may possess dual functionality in regulating the defense response: shifting its function between both a positive and negative regulator of disease resistance depending on the duration of the RCNMV infection. NbBTF3 was predominantly localized to the nucleus suggestive of its transcription regulator function. NbSI was first isolated and characterized in this thesis. It is an integral membrane protein containing five α -helical transmembrane domains and predominantly localizes along cellular, nuclear and endoplasmic reticulum membranes. Functional assays suggested a potential transition of function from promoting membrane proliferation (RCNMV replication) to acting as a precursor of brassinosteroid (modulating host defense).

The genome-wide Nb screen in this thesis reveals the changes in early host gene expression patterns in response to RCNMV. This work forms the foundation for future studies into significant host factors/pathways that determine the fate of virus survival in an inhospitable cellular environment.

© Copyright 2014 by Phanit Thammarat

All Rights Reserved

The Regulation of *Nicotiana benthamiana* Gene Expression at the Early Stages of
Red clover necrotic mosaic virus Infection

by
Phanit Thammarat

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Functional Genomics

Raleigh, North Carolina

2014

APPROVED BY:

Dr. Steven A. Lommel
Committee Co-Chair

Dr. Dahlia M. Nielsen
Committee Co-Chair

Dr. Robert G. Franks

Dr. Heike W. Sederoff

DEDICATION

I dedicate this work to all human curiosity in science.

BIOGRAPHY

Phanit Thammarat was born and raised in Khon Kaen, Thailand. She graduated with a Bachelor of Pharmaceutical Sciences from Khon Kaen University in 2005. She decided to take a chance to pursue her love of science in the Functional Genomics program at North Carolina State University in 2007.

ACKNOWLEDGMENTS

I would like to thank Dr. Steven Lommel, my advisor, for giving me this wonderful opportunity to pursue my curiosity of science. I also thank Dr. Tim Sit for his patience and holding my hand through parts of the project. I thank all of my committee members for giving me feedback in this project. I would like to thank Dr. Eva Johannes at the Center for Molecular Imaging Facility for assistance with all the GFP imaging in this thesis, and Dr. Jonathan Horowitz and Shannon Chiera for assistance with the IVIS Lumina imager. I would like to thank Dr. Skylar Marvel for his empathy. Lastly, I thank my parents for distracting me and trying to convince me to quit and go to medical school.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xii

Chapter 1

Literature review	1
Host defense against plant RNA viruses.....	2
Introduction.....	2
Systemic acquired resistance, cell death and the hypersensitive reaction	5
Lipoxygenases.....	10
Mitogen-activated protein kinases (MAPKs)	14
An involvement of WRKY transcription factors in a regulation of MAPKs	17
Experimental models	19
<i>N. benthamiana</i>	19
RCNMV	22
DNA microarray and its application to study plant-virus interactions	24
DNA microarray limitations	25
DNA microarray applications in plant-virus interactions.....	26
Dissertation contributions and thesis organization	28
References.....	30

Chapter 2

Analysis of expressed sequence tags in <i>Nicotiana benthamiana</i> defines a unigene set comprising 13,000 genes	41
Abstract.....	42
Introduction.....	43
Materials and Methods.....	49
Result and discussion.....	52
<i>N. benthamiana</i> ESTs and unigene assembly	52

Potential highly-expressed <i>N. benthamiana</i> genes from standard, normalized, and subtracted EST libraries	54
Functional analysis of <i>N. benthamiana</i> unigenes	54
Construction of the <i>N. benthamiana</i> microarray.....	56
Conclusion	58
References.....	68

Chapter 3

<i>N. benthamiana</i> genome-wide screen reveals early changes in host gene expression patterns in response to <i>Red clover necrotic mosaic virus</i> infection	75
Abstract.....	76
Introduction.....	78
Experimental designs	88
Whole plant model, time point experimental design and rationale	88
Microarray experimental design, statistical analysis and rationale	92
Microarray confirmation experimental design, statistical analysis and rationale.....	109
Result and discussion.....	116
Microarray data quality control and reliability	116
Host gene expression cumulatively fluctuated during the first 24 hours of RCNMV infection	123
Switch mode of host gene expression at 12 hour RCNMV infection.....	125
Systemic acquired resistance gene (SAR) expression was modulated over 24 hours RCNMV infection.....	130
Functional analysis to determine the impacts of an early RCNMV infection on host physiological process	132
Host defense responses	134
Host multidrug and toxin extrusion (MATE) transporter	136
Host translation process and machinery	139
Host photosynthesis process and chloroplast-related functions.....	143
Host carbon fixation and Calvin cycle.....	149
Host protein kinases	157
Host cell wall	160

Tracking RCNMV RNA profile using the microarray	165
Microarray confirmation	168
Conclusion	171
Materials and methods	176
References	188

Chapter 4

<i>Red clover necrotic mosaic virus</i> down-regulates <i>Nicotiana benthamiana</i> systemic acquired resistance (NbSAR) gene expression in the first 6 hours of infection, followed by an up-regulation	200
Abstract	201
Introduction	202
Result and discussion	206
RCNMV suppressed <i>N. benthamiana</i> defense-related function in the first 6 hours of infection followed by a de-suppression	206
<i>N. benthamiana</i> systemic acquired resistance (NbSAR or Nb12974) is the only host gene identified that was modulated across 2, 6, 12 and 24 hours post RCNMV inoculation	208
A functional annotation of Nb12974 to NbSAR	209
Nb12974 sequence analysis	209
A prediction of the full-length cDNA sequence of Nb12974	210
SAR8.2 phylogenetic analysis	212
Pre-silenced NbSAR (Nb12974) promoted RCNMV RNA-1 and RNA-2 replication at 24 hpi and increased RCNMV R1SG1-sGFP accumulation	214
Conclusion	215
Materials and Methods	231
References	236

Chapter 5

Modulation of <i>N. benthamiana</i> WRKY gene expression early in a <i>Red clover necrotic mosaic virus</i> infection implies its involvement in plant immunity	239
Abstract	240
Introduction	241
Result and discussion	245
RCNMV up-regulated <i>N. benthamiana</i> WRKY TF expression 12 hours after the initiation of infection	245
A functional annotation of Nb04875 as a WRKY TF using a bioinformatic/computational approach	247
Nb04875 is a WRKY homolog	247
Putative Nb04875 protein structure is composed of a four β -stranded sheet	249
Putative Nb04875 protein contains the N-terminal signal peptide that enables its translocation to the nucleus	252
Construction of the full-length NbWRKY cDNA	256
The putative NbWRKY promoter contains W-boxes, indicating an auto/cross-regulation	258
NbWRKY - sGFP fusion protein localizes to the nucleus	259
The functional impacts of NbWRKY on RCNMV replication	262
Pre-silenced NbWRKY promoted RCNMV replication at 24 hpi and increased RCNMV R1SG1-sGFP accumulation	262
Pre-overexpressed NbWRKY did not affect RCNMV replication	264
Conclusion	268
Materials and Methods	296
References	308

Chapter 6

Altered regulation of <i>Nicotiana benthamiana</i> basal transcription factor 3 (NbBTF3) expression early in a <i>Red clover necrotic mosaic virus</i> infection	314
Abstract	315
Introduction	316
Results and discussion	320

RCNMV infection up-regulated <i>N. benthamiana</i> basal transcription factor 3 (NbBTF3) gene expression at 6 hpi	320
NbBTF3 (Nb09394) sequence analysis and functional annotation	322
Silencing of the NbBTF3 gene was reversed 12 hours after RCNMV infection.....	322
Pre-silenced NbBTF3 promoted RCNMV replication in the first 12 hours	327
Pre-overexpressed NbBTF3 suppressed RCNMV replication in the first 12 hours	328
NbBTF3 - sGFP fusion protein localizes to the nucleus.....	329
Conclusions.....	330
Materials and Methods.....	351
Reference	357

Chapter 7

Involvement of <i>Nicotiana benthamiana</i> sterol biosynthesis in the early stages of <i>Red clover necrotic mosaic virus</i> infection.....	361
Abstract.....	362
Introduction.....	363
Result and Discussion	366
RCNMV up-regulated <i>N. benthamiana</i> C-8, 7-sterol isomerase (NbSI) gene expression at 12 hpi	366
NbSI (Nb02987) sequence analysis and functional annotation	367
Pre-silenced NbSI suppressed RCNMV replication in the first 6 hours, followed by a promotion after 12 hours.....	369
Pre-overexpressed NbSI suppressed RCNMV replication in the first 12 hours.....	371
NbSI is an integral membrane.....	372
NbSI - sGFP fusion protein localizes to cellular, nuclear, and ER membranes	372
NbSI is predicted to localize to the plasma membrane and ER membrane	374
NbSI protein is predicted to contain five α -helical transmembrane domains.....	374
Conclusions.....	376
Materials and Methods.....	396
References.....	402

LIST OF TABLES

Chapter 2

Table 1	Sources and numbers of <i>N. benthamiana</i> EST sequences used for unigene assembly.....	60
Table 2	Potential highly-expressed <i>N. benthamiana</i> genes from standard, normalized, and subtracted EST libraries	60

Chapter 3

Table 1	The structure of microarray data analysis using a 2 x 4 factorial design.....	108
Table 2	Degree of freedom (df) for microarray data statistical analysis	108
Table 3	KEGG functional annotation of <i>N. benthamiana</i> host genes in "carbon fixation in photosynthetic organisms (KO 00710)"	154
Table 4	Microarray confirmed by qRT-PCR analysis	169
Table 5	Differential host gene expression data confirmed by qRT-PCR analysis.....	170
Table 6	Top 20 GSEA result at 2 hours post RCNMV inoculation.....	182
Table 7	Top 20 GSEA result at 6 hours post RCNMV inoculation.....	183
Table 8	Top 20 GSEA result at 12 hours post RCNMV inoculation.....	184
Table 9	Top 20 GSEA result at 24 hours post RCNMV inoculation.....	185
Table 10	Oligonucleotide information used in the real-time PCR analysis Chapter 3	186

Chapter 4

Table 1	Numbers of <i>N. benthamiana</i> microarray genes assigned to the defense-related functional categories based on GO database	216
Table 2	Publicly available gene members in the SAR8.2 family	229
Table 3	Oligonucleotide information used in Chapter 4.....	230

Chapter 5

Table 1	<i>N. benthamiana</i> WRKY (transcription factor) unigene annotation information	270
Table 2	Nb04875 has the top BLAST hits to WRKY transcription factors from various plant species	273
Table 3	WRKY (transcription factor) homolog sequence information	274
Table 4	Oligonucleotide information used in Chapter 5	295

Chapter 6

Table 1	Nb09394 is highly homologous to a basal transcription factor 3 (BTF3) from various <i>Solanaceae</i> plants	333
Table 2	<i>N. benthamiana</i> RNAi-related genes	348
Table 3	Oligonucleotide information used in Chapter 6	350

Chapter 7

Table 1	Nb02987 is highly homologous to sterol isomerase from various plant species	379
Table 2	Oligonucleotide information used in Chapter 7	395

LIST OF FIGURES

Chapter 2

Figure 1	Transient silenced <i>PDS</i> (phytoene desaturase) gene in <i>N. benthamiana</i> using TRV-VIGS system.....	61
Figure 2	<i>N. benthamiana</i> healthy and RCNMV-infected plants.....	61
Figure 3	Distribution of <i>N. benthamiana</i> unigene length.....	62
Figure 4	Distribution of <i>E</i> -values of top BLAST hit <i>N. benthamiana</i> unigenes searched against GenBank nr database	62
Figure 5	Distribution of percent sequence similarity of top BLAST hit <i>N. benthamiana</i> unigenes searched against GenBank nr database.....	63
Figure 6	Distribution of organism species of top BLAST hit <i>N. benthamiana</i> unigenes searched against GenBank nr database	64
Figure 7	GO distribution for "biological process" category assigned to <i>N. benthamiana</i> unigenes	64
Figure 8	GO distribution for "molecular function" category assigned to <i>N. benthamiana</i> unigenes	65
Figure 9	GO distribution for "cellular compartment" category assigned to <i>N. benthamiana</i> unigenes	65
Figure 10	The sources of expressed sequence tag (EST) collection for creating a custom <i>N. benthamiana</i> microarray	66
Figure 11	A layout design of a custom <i>N. benthamiana</i> microarray	67
Figure 12	Blast2GO graphical user interphase (GUI).....	68

Chapter 3

Figure 1	<i>N. benthamiana</i> healthy and RCNMV-infected symptoms	86
Figure 2	RCNMV virion and genome organization.....	87
Figure 3	Microarray experimental design diagram	104
Figure 4	P-value distributions shows the effects of treatment, time, and duration of infection on the modulation host gene expression in the early stages of RCNMV infection.....	105
Figure 5	Microarray fold change distributions display the differential expressions of 13,413 host genes at 2, 6, 12, and 24 hours post inoculation (hpi).....	106

Figure 6	Microarray data distribution and quality control before and after Loess normalization	118
Figure 7	Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 2 hpi.....	119
Figure 8	Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 6 hpi.....	120
Figure 9	Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 12 hpi.....	121
Figure 10	Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 24 hpi.....	122
Figure 11	Volcano plots of the microarray data of 13,413 host genes shows the cumulative changes in gene expression at 2, 6, 12, and 24 hours post inoculation.....	124
Figure 12	Total numbers of significantly differentially expressed host genes from the microarray data analysis at 2, 6, 12, and 24 hours post inoculation, with an FDR cutoff of 0.01	125
Figure 13	Hierarchical clustering the differential microarray expression of 13,413 host genes at 2, 6, 12, and 24 hours post inoculation (hpi) reveals a reversal of regulation at 12 hours (a switch mode effect).....	129
Figure 14	Venn diagram displays the microarray data of the overlapped host genes that were significantly differentially modulated at 2, 6, 12, and 24 hours post inoculation (hpi) at an FDR cutoff of 0.01	131
Figure 15	GSEA enrichment plots display host defense-related functions that were highly correlated to RCNMV infection at different time points.....	135
Figure 16	GSEA enrichment plots display host multidrug and toxin extrusion (MATE)-related functions that were highly correlated to RCNMV infection at different time points.....	138
Figure 17	GSEA enrichment plots display host ribosome/translation-related functions that were highly correlated to RCNMV infection at different time points	140
Figure 18	GSEA enrichment plots display host chloroplast/photosynthesis-related functions that were highly correlated to RCNMV infection at different time points.....	145
Figure 19	GSEA enrichment plots display host Calvin cycle-related function (GO19253) that was highly correlated to RCNMV infection at different time points.....	151

Figure 20	KEGG diagram of "carbon fixation in photosynthetic organisms (KO 00710)" and microarray expression data	152
Figure 21	Hierarchical clustering of the microarray expression data of host carbon fixation- and RuBisCO- related genes	153
Figure 22	GSEA enrichment plots display host protein kinase-related functions that were highly correlated to RCNMV infection at different time points.....	159
Figure 23	GSEA enrichment plots display host cell wall-related functions that were highly correlated to RCNMV infection at different time points.....	163
Figure 24	Functional categories (FunCat scheme) of the significantly differentially expressed host genes at different time points of study (2, 6, 12, and 24 hours post inoculation (hpi)) from microarray data analysis with an FDR cutoff of 0.01.....	164
Figure 25	Tracking RCNMV RNA-1 profiles from the microarray data.....	166
Figure 26	Tracking RCNMV RNA-2 profiles from the microarray data.....	167

Chapter 4

Figure 1	Host defense-related functions were highly enriched and correlated to an early RCNMV infection at time points 2, 6, 12 and 24 hpi	217
Figure 2	Nb12974 (NbSAR) was the only host gene in the microarray study that was modulated across all four times points of study at 2, 6, 12, and 24 hours-post RCNMV inoculation.....	219
Figure 3	NbSAR (Nb12974) gene expression was suppressed in <i>N. benthamiana</i> by using TRV-VIGS vector system	220
Figure 4	Pre-silenced NbSAR (Nb12974) affected RCNMV RNA accumulation at 3, 6, and 24 hours-post inoculation (hpi), but not at 12 hpi.....	221
Figure 5	Pre-silenced NbSAR (Nb12974) promoted RCNMV GFP accumulation at 3 days post inoculation.....	222
Figure 6	BLASTN information and alignment between Nb12974 and NbSAR8.2m nucleotide sequences.....	223
Figure 7	BLASTX information and alignment between Nb12974 and NbSAR8.2m amino acid sequences.....	224
Figure 8	The predicted amino acid sequence of Nb12974 using ExPASy translation tool and standard genetic code.....	224
Figure 9	Signature motif elucidation of a signal peptide at 5' cDNA end of all <i>N. tabacum</i> SAR8.2 gene members.....	225

Figure 10	Amino acid sequence comparison between NtSAR8.2m and predFLNbNtSAR8.2m using BLASTP alignment	225
Figure 11	Amino acid sequence comparison between Nb12974 and the two predicted isoforms of full-length Nb12974 (predFLNb12974_1 and predFLNb12974_2) using BLASTP alignment	225
Figure 12	Nucleotide sequence comparison between the two predicted isoforms of full-length Nb12974 cDNA (predFLNb12974_1 and predFLNb12974_2) using BLASTN alignment	226
Figure 13	Amino acid sequence comparison between the predicted full-length cDNA of <i>N. benthamiana</i> SAR homologs using BLASTP alignment	226
Figure 14	Multiple alignments of the predicted full-length cDNA sequences of <i>N. benthamiana</i> SAR8.2 homologs and all publicly available gene members in the SAR8.2 family	227
Figure 15	Phylogenetic analysis of all publicly available gene members in the SAR8.2 family	228

Chapter 5

Figure 1	Hierarchical clustering the differential microarray expression data of <i>N. benthamiana</i> WRKY genes at 2, 6, 12 and 24 hours post RCNMV inoculation (hpi).....	272
Figure 2	WRKY homolog multiple sequence alignment identifies WRKY DNA binding domain (WRKYGQK, blue box), zinc finger motif (Cx7Cx23HxC, brown box), and a nuclear localization signal (NLS) (KRRK, green box) in Nb04875 and NbWRKY.....	275
Figure 3	A phylogenetic tree of WRKY homologs indicates StWRKY6 gene is the closest homolog to Nb04875 and NbWRKY (red label)	276
Figure 4	NbWRKY and Nb04875 both contain a nuclear localization signal (NLS).....	277
Figure 5	Nb04875 has a top BLAST hit to potato StWRKY6.....	278
Figure 6	Nb04875 nucleotide sequence (940 bp) is conceptually translated to 313 amino acids by using ExPASy translate tool and standard genetic code.....	279
Figure 7	Amino acid sequence comparison between the predicted full-length Nb04875 cDNA (predFLNbWRKY) and the obtained clone (NbWRKY).....	280
Figure 8	Amino acid sequence comparison between the obtained clone NbWRKY and the unigene Nb04875	280
Figure 9	Use of AtWRKY1 to predict NbWRKY three-dimensional protein structure	281

Figure 10	NbWRKY and Nb04875 three dimensional protein structure model prediction using SWISS-MODEL.....	283
Figure 11	NbWRKY promoter architecture predicted by PLACE bioinformatic tool	284
Figure 12	Nuclear localization of NbWRKY protein	287
Figure 13	NbWRKY gene expression was suppressed in <i>N. benthamiana</i> by using TRV-VIGS vector system	289
Figure 14	Pre-silenced NbWRKY affected RCNMV RNA accumulation at 3, 6 and 24 hours-post inoculation (hpi), but not at 12 hpi.....	290
Figure 15	Pre-silenced NbWRKY promoted RCNMV GFP accumulation at 3 days post inoculation.....	291
Figure 16	NbWRKY gene was overexpressed in <i>N. benthamiana</i> by using PZP212 and pRTL2 vector system.....	292
Figure 17	Pre-overexpressed NbWRKY did not affect RCNMV RNA accumulation at 6, 12 and 24 hpi.....	293
Figure 18	<i>N. benthamiana</i> plant size selection for a gene over-expression study (A) and a gene silencing study (B).....	294

Chapter 6

Figure 1	Host NbBTF3 (Nb09394) gene expression was up-regulated at 6 hours post RCNMV inoculation (hpi)	332
Figure 2	Nb09394 has a complete coding sequence (CDS) and its encoded peptide has a 100% residue identity to <i>N. benthamiana</i> BTF3 GenBank ID ABE01085.1, with E-value = 3×10^{-67}	333
Figure 3	Nb09394 sequence features and its encoded peptide.....	334
Figure 4	The gene silencing of NbBTF3 was reversed 12 hours post RCNMV inoculation.....	335
Figure 5	Leaf chlorosis phenotype produced by silencing of NbBTF3 gene in <i>N. benthamiana</i>	337
Figure 6	Pre-silenced NbBTF3 affected RCNMV RNA-1 and RNA-2 accumulation at sub 24 hours	342
Figure 7	Pre-silenced NbBTF3 suppressed RCNMV GFP accumulation > 2 fold at 3 days post inoculation.....	343
Figure 8	NbBTF3 gene was overexpressed in <i>N. benthamiana</i> by using PZP212 and pRTL2 vector system.....	344

Figure 9	Pre-overexpressed NbBTF3 reduced RCNMV RNA accumulation at 6 hours of infection	345
Figure 10	Nuclear localization of NbBTF3 protein	346
Figure 11	Heat map displays the differential expression of <i>N. benthamiana</i> RNAi-related genes in an early RCNMV infection from the microarray study	349

Chapter 7

Figure 1	Host NbSI (Nb02987) gene expression was up-regulated at 12 hours post RCNMV inoculation (hpi)	378
Figure 2	Nb02987 has a partial coding sequence (CDS) based on its alignment to the full-length sterol isomerase peptide sequences from tomato (SlySt, GenBank ID XP_004242445.1) and potato (StSI, GenBank ID NP_001275362.1)	379
Figure 3	Amino acid sequence comparison between Nb02987 and its full-length protein (NbSI)	380
Figure 4	Amino acid sequence comparison between Nb02987, a full-length <i>N. benthamiana</i> sterol isomerase (NbSI), a full-length tomato sterol isomerase (SlySI, Genbank ID XP_004242445.1) and a full-length potato sterol isomerase (StSI, GenBank ID NP_001275362.1)	380
Figure 5	NbSI gene expression was suppressed in <i>N. benthamiana</i> using TRV-VIGS vector system	381
Figure 6	Pre-silenced NbSI suppressed RCNMV RNA-1 and RNA-2 accumulation in the first 6 hours, followed by a promotion at 24 hpi	382
Figure 7	Pre-silenced NbSI promoted RCNMV GFP accumulation > 4 fold at 3 days post inoculation	383
Figure 8	NbSI gene was overexpressed in <i>N. benthamiana</i> by using PZP212 and pRTL2 expression vector system	384
Figure 9	Pre-overexpressed NbSI suppressed RCNMV RNA accumulation in the first 12 hours of infection	385
Figure 10	Membrane localization of NbSI protein	386
Figure 11	NbSI protein is predicted to contain five α -helical transmembrane domains	389
Figure 12	NbSI does not contain a signal peptide	390
Figure 13	Sterol-specific biosynthetic pathway and Brassinosteroid (BR)-specific biosynthetic pathway	391
Figure 14	The sterol biosynthetic pathway distinctions in animals, fungi, higher plants and algae	393

Figure 15 Heat map displays the modulation of *N. benthamiana* sterol biosynthesis-related gene expression at the early stages of the RCNMV infection from the microarray study (Chapter 3)394

Chapter 1

Literature review

Chapter summary

This literature review is focused on three major topics; (i) host defense against plant RNA viruses, (ii) experimental models: *N. benthamiana* (a model plant host) and RCNMV (a model virus), and (iii) DNA microarray and applications in plant-virus interaction. I also include details about my dissertation contributions and chapter organization.

Host defense against plant RNA viruses

Introduction

Plants, like animals, are continuously exposed to pathogen attacks. However, most plants are normally able to resist pathogen invasion because they possess an innate immunity that activate suites of elaborate and effective defense mechanisms to prevent, protect, and rapidly limit damage owing to pathogen infection [1]. Some pathogens are restricted to certain parts of the plant (e.g. the inoculated leaf) whereas others are able to spread throughout the plant causing a systemic infection. For this reason, plant defense mechanisms have evolved the ability to not only protect plants from pathogens at the primary site of infection, but also throughout the entire plant [2]. Some defense mechanisms are already activated in the plants without requiring prior exposure to pathogen stimulation, whereas others are induced upon the perception of pathogen signals. Plant defense also includes physical barriers and production of toxic compounds [3].

There are approximately 450 species of plant-pathogenic viruses and the majority of these viruses have a positive-sense single-stranded ((+)ss) RNA genome. Plant viruses cause diseases in most, if not all, species of both monocot and dicot plants, but most viruses have a restricted host range limited to several species or genera of plants. A plant will likely recover from a virus infection if its innate immunity is able to arrest and defeat the virus. However, if the virus infection circumvents host defense mechanisms, a systemic infection will result and lead to a compromised or dead plant [3].

Viruses cause many important plant diseases and are responsible for large losses in crop production world-wide [4]. Infected plants show a range of symptoms depending on the type of virus but often these symptoms are leaf yellowing (either of the whole leaf or in a pattern of stripes or blotches), leaf distortion (e.g. curling), and other growth distortion (e.g. stunting of whole plant, abnormalities in flower or fruit formation). For example, *Lettuce mosaic virus* ((+)ss RNA) causes yellow mosaic symptom on lettuce, *Grapevine fanleaf virus* ((+)ss RNA) causes yellow vein-banding on grapevine, *Tomato bushy stunt virus* ((+)ss RNA) causes fruit distortion on eggplant, and *Citrus psorosis virus* ((-)ss RNA) causes bark scaling. Some plant virus diseases cause only mild visible symptoms, but others result in destructive symptoms and plant death. With industrial-scale farming, some plant virus diseases result in severely reduced crop yields and quality, and in some instances complete regional crop failure [5]. It is essential to understand plant virus pathogenesis at the molecular level in order to develop effective control strategies for plant virus diseases.

Currently there is no effective chemical application that controls plant virus infections at the field level. The major chemical or input based control strategy for reducing and preventing the spread of plant virus diseases is exclusively targeted against plant virus vectors (often an insect), either by the use of chemical (pesticide) or biological methods. However, this strategy is only effective if a) viruses are transmitted by insect vectors, and b) the vectors succumb to a pesticide-sprayed crop before the viruses are transmitted. The most effective approach by far to control plant viruses relies on the use of virus resistant cultivars [6, 7]. Plant virus resistance genes allow plants to detect virus infection and launch effective defense responses. Traditional breeding techniques extensively employ virus resistance genes

for development of virus-resistant lines. Intensive characterization of these resistance genes and how they trigger defense are now taking place and hundreds of naturally occurring genes have been reported from studies of monocots, dicots, and the plant model *Arabidopsis*. The isolation and characterization of resistance genes have resulted in an understanding of defense pathways that are critical in determining the outcome of plant virus infection. If a new transgenic resistance line is able to confer a substantial degree of virus resistance, then novel resistance genes could potentially be utilized to resolve problems of plant virus disease and provide security for large-scaled farming of sustainable foods [6].

This leads to the goal of my dissertation, which is focused on the interaction between plant RNA viruses and the host immune system. My work is centered on an effort to expand the knowledge base of (i) how the plant host reprograms physiological processes during a healthy state to deliver a highly sophisticated defense response directed against the viral invader and (ii) how the virus infection circumvents the host immune system resulting in a robust infection.

Plants employ different strategies to launch their defense responses against a viral attack. This review is focused on four different aspects of the defense strategies: (i) systemic acquired resistance, cell death and hypersensitive reaction, (ii) lipoxygenases, (iii) mitogen-activated protein kinases (MAPKs) and (iv) involvement of WRKY transcription factors regulating MAPKs.

Systemic acquired resistance, cell death and the hypersensitive reaction

Systemic acquired resistance (SAR) is the whole-plant resistance response that occurs following an earlier localized exposure to a pathogen. SAR can be induced by a wide range of plant pathogens, especially (but not only) those that cause tissue necrosis. The death of locally infected cells is a common cause that triggers SAR response. This cell death could be caused by a part of hypersensitive response (HR) or a symptom of disease [2, 8].

HR induces rapid localized cell death at the point of pathogen entry [8]. Cell death is caused by a loss of membrane integrity and the generation of lipid peroxides and reactive oxygen species (ROS). This rapid cell death of virus infected cells is a generally effective strategy to restrict the growth and spread of pathogens to prevent the infection from moving to other parts of the plant. The HR in plants is analogous to the programmed cell death (PCD) found in animal innate immune systems [9]. HR commonly precedes a slower systemic (whole plant) response, which ultimately leads to systemic acquired resistance (SAR). The activation of SAR results in the development of a broad-spectrum, systemic resistance. SAR is important for plants to resist disease, as well as to recover from disease once formed. SAR can be distinguished from other disease resistance responses by both the spectrum of pathogen protection and the associated changes in gene expression [10].

HR is initiated after a gene-for-gene recognition between a resistance gene encoded plant resistance protein and a pathogen microbial elicitor. Recognition is postulated to result from the interaction of plant resistance gene (*R*) product with a corresponding pathogen avirulence gene (*Avr*) product [3]. The interaction is genetically endowed with the capacity

for mutual recognition. Plants producing a specific *R* gene product are resistant towards a pathogen that produces the corresponding *Avr* gene product. *R* genes that mediate HR encode products that function as receptors for recognition of specific pathogen *Avr* gene products. The interaction between the *R* gene product and the *Avr* gene product activates signal transduction pathways that lead to the expression of resistance responses [11]. The primary response is programmed cell death leading to HR, which then activates a distal resistance called SAR. The SAR induced by HR then prevents pathogen invasion of the uninfected parts of the plant. Consequently, the development of disease is controlled by a matching pair of plant *R* gene and microbial *Avr* gene. If either the plant or pathogen partner lacks a functional product of the corresponding gene pair, then resistance is not triggered and the plant becomes diseased and systemically infected.

Plant *R* genes encode cell surface pattern-recognition receptors (PRR) to recognize conserved microbial signatures. The first PRR identified in plants was the rice XA21 gene, whose protein confers resistance to the Gram-negative bacterial pathogen *Xanthomonas oryzae* pv. *oryzae* [12]. Plants also carry immune receptors that recognize highly variable pathogen effectors. These include the NBS-LRR class of proteins. The protein products of the NBS-LRR resistance (*R*) genes contain a nucleotide binding site (NBS) and a leucine rich repeat (LRR). This NBS-LRR class includes the *N* resistance gene of tobacco against *Tobacco mosaic virus* (TMV) [11].

SAR is associated with the induction of a wide range of pathogenesis-related (PR) genes and the activation of SAR requires the accumulation of endogenous salicylic acid

(SA). SA participates in the local and systemic response, but SAR does not require long-distance translocation of SA. The pathogen-induced SA signal activates a molecular signal transduction pathway that is identified by a gene called *NIMI*, *NPRI* or *SAII* (three names for the same gene) in the model genetic system *Arabidopsis thaliana* [10].

The SAR signal transduction pathway appears to determine the outcome of pathogen infection. When SAR is activated, a normally compatible plant-pathogen interaction (host susceptible; one in which disease is the normal outcome) can be converted into an incompatible one (host resistant) [13, 14]. Conversely, when the SAR pathway is incapacitated, a normally incompatible interaction becomes compatible [14, 15].

(i) Genes associated with systemic acquired resistance

A gene is classified as a SAR marker gene if a) its expression is modulated tightly with the onset of SAR, and b) the presence and activity of its encoded protein presence correlates tightly with maintenance of whole-plant resistance state. Analysis of SAR proteins showed that many belong to the class of pathogenesis-related (PR) proteins, which originally were identified as novel proteins accumulating after TMV infection of tobacco leaves [16, 17]. In tobacco, the set of SAR markers consists of at least nine families comprising acidic forms of PR-1 (PR-1a, PR-1b, and PR-1c), β -1,3-glucanase (PR-2a, PR-2b, and PR-2c), class II chitinase (PR-3a and PR-3b, also called PRQ), hevein-like protein (PR-4a and PR-4b), thaumatin-like protein (PR-5a and PR-5b), acidic and basic isoforms of class III chitinase, an extracellular β -1,3-glucanase (PR-Q'), and the basic isoform of PR-1 [18]. A basic protein

family called SAR 8.2 that is induced during the onset of SAR but which shows a pattern of gene expression distinct from that of the other SAR genes has also been described [18, 19]. In *Arabidopsis*, the SAR marker genes are PR-1, PR-2, and PR-5 [13]. SAR has also been observed in monocotyledon species. A number of dicot SAR genes homologs have been identified in monocots. For example, homologs of the PR-1 family from *N. tabacum* cv. Xanthi-nc were detected in maize and barley [20, 21]. The genes encoding these SAR marker proteins have been used extensively to evaluate the onset of SAR [13, 18].

(ii) Systemic acquired resistance gene expression varies by plant species and pathogen

SAR gene products are secondary plant metabolites which are not normally produced when plants are in a healthy state. However, under pathogen attack, the expression of these defense genes rapidly increases. Studies have shown that plants contain more than one type of SAR gene such as observed in tobacco, but not all SAR genes are activated during a pathogen attack because of the high energy cost to do so. Consequently, plants selectively induce specific subsets of SAR genes. This selection depends on the type of pathogen as well as the type of plant species. For example, in cucumber, acidic PR-1 is weakly expressed, whereas in tobacco and *Arabidopsis*, acidic PR-1 is the predominant SAR-related protein [10]. Such species-specific differences may reflect different evolutionary or breeding constraints that have selected for the most effective SAR response against the particular suite of pathogens to which an individual species is subject [22].

(iii) Salicylic acid accumulation increases during systemic acquired resistance activation in virus-infected plants

Salicylic acid (SA) is an important component in the signal transduction pathway leading to SAR against the entire spectrum of pathogen: bacteria, fungi, and viruses [10, 23]. The most common trigger for SA synthesis is an HR, which results from the interaction of a host resistance gene product with a specific pathogen elicitor [24]. One of the best-studied model systems for investigation of HR and the induction of SA is the interaction between *Tobacco mosaic virus* (TMV) and tobacco plants possessing the *N* gene. The *N* gene encodes a receptor protein able to recognize specific sequences in the 126-kDa TMV replicase protein [25, 26]. This recognition event leads to the HR, in which the virus is restricted to the tissue immediately surrounding necrotic lesions that form at the viral inoculation sites. The HR is followed by an increase in SA levels in the plant leading to SAR, which is manifested by the production of fewer, smaller necrotic lesions in response to a second challenge with TMV [8, 27-29]. An accumulation of SA is also required to prevent spread of the virus beyond the immediate vicinity of the HR lesions [30]. However, exogenous application of SA or the synthetic resistance-inducing chemical benzo(1,2,3) thiadiazole-7-carbothioic acid methyl ester (BTH) can also induce some degree of resistance to viruses even in plants that do not possess a corresponding resistance gene and that, therefore, would in normal circumstances be completely susceptible [31-33].

SA induces a range of defense genes, most notably those encoding the pathogenesis-related (PR) proteins, several of which have been shown to possess antifungal or antibacterial

properties [34]. It is unclear how SA mediates resistance to viruses. Pretreatment of TMV-susceptible (nn genotype) tobacco tissue with SA reduced the levels of viral RNAs and viral coat protein accumulating after inoculation with TMV [35]. It was found that the ratio of genomic RNA to coat protein (CP) mRNA and the ratio of plus- to minus-sense RNAs were affected by SA treatment [35]. These findings suggested that SA induces interference with the activity of the TMV RNA-directed RNA polymerase (RdRp). At the whole plant level this resistance to the virus was manifested as a delay in the onset of systemic TMV disease. Furthermore, it was also shown that salicylhydroxamic acid (SHAM), an inhibitor of the plant mitochondrial alternative oxidase (AOX) [36-39], antagonized the SA-induced delay in TMV disease in TMV susceptible tobacco, as well as SA-induced acquired resistance in tobacco plants possessing the *N* gene [35]. However, SHAM treatment did not prevent the induction of the PR-1 PR protein or resistance to a fungal and a bacterial pathogen [35]. These results suggest that the defensive signal transduction pathway separates downstream of SA into two branches. It is proposed that one branch leads to induction of resistance to TMV and is SHAM sensitive (and potentially requires AOX activity to function), while the other pathway (SHAM insensitive) leads to the induction of antifungal and antibacterial mechanisms [35].

Lipoxygenases

Lipoxygenases (LOXs) are enzymes widely distributed in plants and animals. In plants, LOX gene expression is associated with a number of developmental stages, and is induced by a pathogen challenge [40-42]. Upon a plant-pathogen interaction, LOXs generate

the products that are either directly involved in plant defense response or are used as precursors in subsequent reaction pathways to synthesize other products involved in the defense response. These products can be divided into three groups according to their defense function strategies: i) hydroperoxides, the highly reactive molecules, that could give rise to free radicals that contributes to the initiation of hypersensitive reaction by promoting localized cell death, ii) defense signaling molecules, such as jasmonic acid and its methyl ester, that can trigger defense gene expression and amplify the initial defense response, and iii) antimicrobial compounds, such as 2-trans-hexanal, that constitutes a direct defense against pathogen attack [43]. In animals, LOXs are involved in the arachidonic acid cascade that contributes to a serial conversion of arachidonic acid (polyunsaturated fatty acids) into eicosanoids known as leukotrienes and lipoxins. These products, in turn, regulate physiopathological processes, including the immune system, inflammatory reactions, apoptotic cell death, and the response to infectious pathogens [44]. The involvement of LOXs in the immune system suggests a common defense strategy that animals and plants may share via LOXs pathway to fight against pathogens.

Lipoxygenases (LOXs) are non-heme dioxygenases that catalyze the addition of oxygen molecules (also known as oxidation or oxygenation) to polyunsaturated fatty acids containing a (cis, cis)-1, 4-pentadiene system to yield an unsaturated fatty acid hydroperoxides [45]. These hydroperoxides are highly reactive and can easily breakdown to form free radicals, or they can be rapidly metabolized by serving as the precursors in the subsequent biosynthetic pathways that transform them into series of oxylipins or octadecanoids in plants, or eicosanoids in animals [44-46]. Plant oxylipins play a pivotal role

in defense toward a pathogen attack. The best-characterized oxylipins in plant defense response is jasmonic acid which is a signaling molecule that modulates the expression of defense genes, resulting in systemic acquired resistance (SAR) [45, 47, 48]. In tobacco, LOX gene expression is induced and jasmonic acid is produced in the very first hours after elicitor treatment. Using an antisense strategy, a specific inducible LOX was shown to be essential for the resistance of tobacco to fungal infection (*Phytophthora parasitica*). A linear relationship was found between the extent of LOX suppression and the size of the lesion caused by the fungus. Indeed, suppression of this activity was sufficient to turn a resistant phenotype into a susceptible one [43, 49].

The first key rate-limiting step in oxylipins biosynthesis essentially depends on the action of LOXs. In plants, oxylipins are mainly derived from linoleic acid (18:2) and α -linolenic acid (18:3). These two polyunsaturated fatty acids contains (cis, cis)-1, 4-pentadiene system that act as the substrates of LOX activities [45, 46]. Jasmonic acid is originally derived from α -linolenic acid that undergoes oxidation in LOX enzymatic reaction [48]. LOXs can add oxygen to either end of these pentadiene substrates with high stereospecificity. So, LOXs catalyzes the oxidation of these two substrates into either 9(*S*)- or 13(*S*)-hydroperoxy derivatives or both depending on the specific isoforms of LOXs. Most plants appear to produce multiple isoforms of LOX to fulfill these putative functions. In general, LOX isoforms can be classified as 9- and 13-LOX according to the position of oxygen that are catalyzed into substrates. For example, in *Arabidopsis thaliana*, six predicted lipoxygenase genes are present in its genome, two genes encoding for 9-LOXs and four genes encoding for 13-LOXs [50].

Different LOX isoforms are expressed in a temporal and spatial specific manner within a single plant species. This expression pattern is an oxylipin production control strategy that makes one oxylipin biosynthetic pathway favorable over others. The activities of LOX isoforms in different subcellular regions yield different pools of hydroperoxides which then serve as the precursors in distinct oxylipins biosynthetic pathway. As a result, this distinction leads to a production of specific oxylipin derivatives that have different physiological functions. The relative concentration, compartmentalization, and localization of LOX isoforms and enzymes involved in individual oxylipin biosynthetic pathways affect the competition among these various oxylipin biosynthetic pathways, resulting in yielding a multitude of oxylipins. Typically, oxylipins are not constitutively synthesized or stored in tissue but are formed on demand. Changes in oxylipin profiles were observed throughout plant developmental process and upon pathogen attack. In- depth analyses have shown that the composition of oxylipins is different among given organelles, tissue, plants and species. The difference in oxylipin profiles to those respective entities is referred as “oxylipin signature”[46, 51].

An examination of the virus invasion front in immature pea embryos infected with *Pea seed-borne mosaic virus* (PSbMV), (+)ss RNA virus in family *Potyviridae*, demonstrated that a down-regulation of two lipoxygenase genes was the early response to virus infection. This study used an in situ hybridization technique to monitor the changes in an accumulation of viral transcripts as well as host gene transcripts upon the active viral replication process taking place in immature pea embryos (cotyledon). They found that the

down-regulation of lipoxygenase genes occurred coordinately with the onset of virus replication and the up-regulation of HSP70 (heat shock protein) and polyubiquitin [52-54].

Mitogen-activated protein kinases (MAPKs)

Signal transduction is initiated by the complex protein-protein interactions between ligands, receptors, and kinases. Mitogen-activated protein kinase (MAPK) cascades are the major components in signal transduction that facilitate a downstream conversion of signals from extracellular stimuli after being perceived by receptors into intracellular responses. This results in eliciting physiological changes in response to particular stimuli. MAPKs are found in eukaryotes only. Studies in plants indicate that MAPK cascades are vital to fundamental physiological functions involved in hormone responses, cell cycle regulation, abiotic stress signaling (temperature, drought, salinity, osmolarity, UV irradiation, ozone, and reactive oxygen species), and anti-pathogen responses. Activation of MAPK cascades confer resistance to bacterial, fungal, and viral pathogens, suggesting that signaling events initiated by diverse pathogens converge into the conserved MAPK cascades [55, 56].

The basic assembly of a MAPK cascade is a three-kinase module (MAPKKK-MAPKK-MAPK) that is conserved in all eukaryotes [55, 56]. MAPK, the last kinase in the cascade, is activated by dual phosphorylation of the Thr and Tyr residues in the tripeptide motif (Thr-Xaa-Tyr, where Xaa could be Glu, Gly, Pro or Asp) located in the activation loop (T-loop) between subdomains VII and VIII of the kinase catalytic domain. This dual phosphorylation is mediated by a MAPK kinase (MAPKK), which, in turn, is activated by a

MAPKK kinase (MAPKKK). There are multiple members of each of the three tiers of kinases in a cell, which contribute to the specificity of the transmitted signal and its outcome. The MAPKKK are the most divergent group of kinases of the three kinases in MAPK cascades [55]. Several unrelated kinases can function as MAPKKKs to initiate the MAPK cascade for even a single MAPK. This is one of the mechanisms by which different stimuli coverage onto one MAPK. The redundancy of MAPK, MAPKK, and MAPKKK leads to two fundamental problems, 1) understanding which upstream kinases regulate which downstream kinases, and 2) identification of the downstream substrates that are targeted by these MAPK cascades.

A whole-genome survey of MAPK genes and the genes encoding their upstream regulators (MAPKKs and MAPKKKs) is currently available in *Arabidopsis*. By using highly conserved signature motifs in searching for homology, at least 23 MAPKs and 20 MAPKK genes were predicted in the *Arabidopsis* genome. The sequences of putative MAPKKK genes are more divergent than those other members of the MAPK cascade. More than 80 genes have been identified as the potential candidates for MAPKKKs in *Arabidopsis* through a comprehensive database search [55-57]. However, the true functions of these computationally predicted MAPKs, MAPKKs, and MAPKKKs are still required to be further investigated and confirmed by a combination of genetic, biochemical and physiological assays.

The activated MAPK phosphorylates a variety of cytoplasmic and nuclear substrates. The phosphorylated or dephosphosphrylated state of these substrates tremendously affects

their localization, functionality (active/inactive form), stability, and transcript levels [58]. Despite extensive accumulation of evidence that shows importance of MAPK in basal defense and R-gene mediated resistance, the mechanism through which MAPKs and MAPK substrates transduce the signals are largely unknown. Analysis of MAPKs, MAPK substrates and phosphorylation networks is presumably important to determine underlined pathways and outcomes of the signal transduction induced during pathogen attack. The high-density *Arabidopsis* protein microarray was used to identify potential MAPK substrates and predict phosphorylation networks [59]. This microarray contained 2,158 protein candidates. The array was screened with 10 different MAPK probes to determine their phosphorylation targets. These probes were pre-activated/phosphorylated in the plant by specific MAPKKs. The finding revealed 570 putative MAPK phosphorylation substrates, with an average of 128 substrates per activated MAPK. These substrates were enriched in transcription factors involved in the regulation of development, defense, and stress responses. The transcription factors, acting as MAPK substrates found in this array screen, were a subset of the WRKY and TGA transcription factor superfamilies [59].

Several distinct MAPK cascades have been implicated in the regulation of plant disease resistance, either positively or negatively [55-57, 60-62]. Infection of resistant tobacco plants carrying the *N* resistance gene against TMV lead to the activation of two tobacco MAPKs, salicylic acid-induced protein kinase (NtSIPK) and wounding-induced protein kinase (NtWIPK) [63]. NtSIPK and NtWIPK were also activated during tobacco Cf-9-dependent resistant against fungal *Cladosporium fulvum*-derived elicitor Avr9 [64]. The activation of NtSIPK and NtWIPK by TMV and *C. fulvum* Avr9 is considered gene-for-gene

specific, suggesting their role(s) in disease resistance [63, 64]. Gain of function studies showed that these two tobacco MAPKs induced the expression of defense genes and cause HR cell death. In contrast, loss of function studies revealed negative regulation of diseases resistance. These gain and loss of function studies supported the evidence that NtSIPK and NtWIPK are involved in the positive regulation of plant defense responses [56, 65]. Both pathogen infection and wounding induce NtWIPK expression systemically, suggesting that NtWIPK participates in systemic response to these stimuli [63, 66]. More direct evidence for the role of NtSIPK and NtWIPK in HR cell death came from an over-expression study of NtMEK2, the common upstream MAPKK shared between NtSIPK and NtWIPK [67]. A constitutively over-expressed mutant of NtMEK2 under the control of a steroid-inducible promoter was able to activate NtSIPK and NtWIPK functions. The magnitude and kinetic of NtSIPK and NtWIPK activation by over-expressed NtMEK2 are similar to those induced by pathogens [63, 68]. Within 12 hours of the induction of over-expressed NtMEK2, HR-like cell death was visible and defense gene expression was induced in the absence of pathogens [67]. Parallel studies of three other tobacco MAPKKs failed to activate NtSIPK and NtWIPK, and importantly failed to induce cell death. This finding suggests the specificity of the NtMEK2-SIPK cascade and NtMEK2-WIPK cascade involved in HR cell death and defense gene activation [67].

An involvement of WRKY transcription factors in a regulation of MAPKs.

MAPK signaling cascades do not directly induce defense gene expression but do so indirectly through interaction with WRKY itself or WRKY repressor. The WRKY proteins

are a superfamily of plant transcription factors. These WRKY transcription factors are important in transcription initiation of many defense-related genes. *N. benthamiana* WRKY8 transcription factor is a physiological substrate of SIPK, NTF4, and WIPK. These three kinases are MAPKs, the final kinases in MAPK cascade, and are characterized as pathogen-responsive MAPKs. WRKY8 is normally in its inactive state (dephosphorylated form). Under a pathogen attack, an activated WRKY8 binds to and initiates a transcription of defense-related genes, such as 3-hydroxy-3-methylglutaryl CoA reductase 2 and NADP-malic enzyme. Phosphorylated WRKY8 by any of SIPK, NTF4, or WIPK increased its DNA binding activity to the cognate W-box sequence on its target genes and increased its transactivation activity. As a result, the transcription of the defense-related/ WRKY8-targeted genes was induced and yielded a resistance to a pathogen invasion [69].

A specific WRKY repressor is phosphorylated and inactivated by a flagellin MAPK cascade upon pathogen infection [57]. Flagellin is a highly conserved component of bacterial flagella that functions as an elicitor in gene-for-gene interaction inducing plant defense response. The plant flagellin receptor is known as FLS2, a leucine-rich-repeat (LRR) receptor kinase. Flagellin activates the FLS2 receptor, leading to the activation of the MAPK cascade. A complete flagellin MAPK cascade is composed of MEKK1 (MAPKKK), MKK4/MKK5 (MAPKK), and MPK3/MPK6 (MAPK). The *Arabidopsis* flagellin MAPK cascade is similar to the activation of the tobacco MAPK cascade by fungal elicitor, as well as during Avr9-Cf-9 and TMV-N gene-for-gene interaction [63, 64, 70].

Based on homology comparison, *Arabidopsis* MPK3 and MPK6 are orthologous to tobacco NtWIPK and NtSIPK respectively. It has been found that the *Arabidopsis* WRKY22/WRKY29 repressor is a final target or substrate of MPK3/MPK6. The WRKY repressor inhibits WRKY transcription factor activity. Upon flagellin challenge, a MAPK cascade inactivates the WRKY repressor by phosphorylation, resulting in an actively functional WRKY transcription factor that induces downstream defense gene transcription, ultimately leading to disease resistance. This supports the evidence that the components of the flagellin MAPK cascade (MEKK1, MKK4/MKK5, and MPK3/MPK6) are key mediators of *Arabidopsis* innate immunity [57].

Experimental models

This dissertation is designed to study the interaction between host and plant RNA virus. To elucidate the events in the host-virus interaction, I utilized *N. benthamiana* (a model host for most plant viruses) and *Red clover necrotic mosaic virus* (RCNMV, a typical plus strand RNA plant virus) as the experimental models. This section provides general information of these experimental models.

N. benthamiana

N. benthamiana is a dicot plant, a close relative to tobacco, and a species in the genus *Nicotiana*, family *Solanaceae*. It is indigenous to Australia, more specifically the Northern Territory and Western Australia, and it originally arose as a result of the hybridization

between the two progenitor species of *N. debneyi* and *N. suaveolens* [71-73]. Next to *Arabidopsis thaliana*, *N. benthamiana* is the prominent model organism for conducting research in the field of plant biology and plant pathology. It is an excellent experimental plant subject and the best known model plant for performing agroinfiltration and virus-induced gene silencing (VIGS) to conduct plant functional assays, either transient gene knockdown or transient protein expression studies [74, 75]. *N. benthamiana* is also an efficient and viable alternative platform for the industrial production of recombinant proteins, particularly in the field of pharmaceutical biotechnology [76-78]. *N. benthamiana* is susceptible to a wide range of plant pathogens (bacteria, oomycetes, fungi, and viruses), making this plant species a cornerstone host for host-pathogen interaction studies, in a particular context of the host immunity and defense signaling [71]. In comparison to *Arabidopsis thaliana*, *N. benthamiana* is susceptible to a wider range of pathogens and its life-cycle is longer allowing for productive pathogen infections.

Although *Arabidopsis* is the most used host in plant-virus studies and its genome is very well characterized, it is not nearly as susceptible to viruses. *N. benthamiana* is a more suitable host model for my dissertation due mainly to its susceptibility to > 500 plant viruses, even to those that are typically restricted to monocot hosts and, most surprisingly, even to several animal viruses [79]. *N. benthamiana* is highly capable of (i) allowing plant viruses to establish a local infection, and (ii) facilitating viral cell-to-cell movement leading to a systemic infection. Therefore *N. benthamiana* is a local and systemic host for many plant viruses. In this dissertation, we studied virus-host interaction by challenging *N. benthamiana* with RCNMV. *N. benthamiana* is highly susceptible to RCNMV. The RCNMV-infected *N.*

benthamiana exhibits local lesions on the inoculated leaves at 2-3 days after inoculation. The green fluorescent protein expressed from a modified RCNMV can be detected under the microscope as early as 8 hours after inoculation, within individually infected cells. RCNMV systemic symptoms appear 3-4 days after inoculation, which include leaf curling, growth stunting, and characteristic necrotic-mosaic (Chapter 3, Figure 1).

The *N. benthamiana* genome (~ 3 Gbp) is nearly 20 fold larger than the *A. thaliana* genome (~ 135 Mbp) [80-82]. *N. benthamiana* is an amphiploid species with $n = 19$; whereas, *A. thaliana* is a diploid species with $n=5$. Genome sequencing is a breakthrough technology that provides a stepping stone to understanding the genetic composition of an organism and the function of genes at a larger scale. The first *N. benthamiana* draft genome was recently published in 2012. This *N. benthamiana* genome draft has 63-fold coverage and is available on Sol Genomic Network for BLAST search and for downloading to a local server [83]. The Sol Genomics Network (SGN) website [84] provides approximately 16,000 unigenes derived from this genome draft.

Although *N. benthamiana* is an excellent plant model in many research areas, the genome-wide study and functional analysis of this plant species is likely far more difficult. The complexity of the *N. benthamiana* genome structure and organization, together with a shortage of other genomic database sources leads to a lack of essential genomic information to enable the use of *N. benthamiana* as a plant model for genome-wide study. Therefore, one of the major contributions of my dissertation was to create a genomic tool (a custom microarray) and data resource of *N. benthamiana* for the scientific community.

RCNMV

RCNMV is a plant RNA virus that was first described infecting red clover in Czechoslovakia in 1967 (Musil and Matisova, 1967). It has been shown to infect a wide range of experimental hosts including tobacco, cucumber, and cowpea.

RCNMV is in the genus *Dianthovirus*, and family *Tombusviridae*. RCNMV has a bipartite genome structure consisting of two positive-sense single-stranded RNAs (Chapter 3 Figure 2): RNA-1 (3.9 kb) and RNA-2 (1.45 kb), which are packaged together into virions composed of 180 identical capsid protein subunits [85-88]. RNA-1 contains three open reading frames (ORF): (i) p27 encoding a 27-kDa functionally unknown protein, (ii) p88 (expressed as a -1 ribosomal frameshift of p27) encoding the 88-kDa viral polymerase protein (contains an RNA-dependent RNA polymerase motif (RdRp)) [89-91], and (iii) p37 (expressed from subgenomic RNA) encoding the 37-kDa viral capsid protein [92]. Whereas RNA-2 is a monocistronic RNA (one ORF) which encodes the 35-kDa viral movement protein required for virus cell-to-cell movement (local infection) in plants and infection of whole plants (systemic infection) [87, 93]. Although RNA-2 is not required for the replication of RNA-1 in protoplasts [94-96], a 34-nucleotide (nt) sequence in RNA-2 is required for transcription of subgenomic RNA from RNA-1 [97]. The addition of (i) a modified GTP (guanosine triphosphate) or any cap analogue (m^7GpppA , m^7GpppG) to the 5' end (cap), and (ii) poly (A) (adenine nucleotide) to 3' end (tail) of the premature-mRNA is the important canonical feature in the translation of many eukaryotic mRNAs [98]. Both RCNMV RNA-1 and RNA-2 lack a 5' cap structure and their 3' ends are not polyadenylated

[99]. However, the uncapped in vitro transcripts of RCNMV RNA-1 and RNA-2 show a comparable infectivity to that of capped transcripts in *N. benthamiana* [100] and in cowpea protoplasts [99]. The high infectivity observed in the uncapped RCNMV RNAs implies that this virus utilizes a cap-independent mechanism to translate viral proteins.

While the RCNMV genome and function is well characterized for at least the past 30 years, very little is known about how RCNMV interacts with the plant host. To date, no host resistance (*R*) genes have been described to recognize or target RCNMV during infection. RCNMV is known host tissue necrosis in many hosts. It is likely that the plant host must contain *R* genes that recognize RCNMV proteins. RCNMV has two known suppressors of host gene silencing. The first characterized is a composite of the presence of p27, p88, and a replication competent RNA-2 [101], and the second is the RNA-2 encoded movement protein [102]. RCNMV replication process takes place on the remodeled endoplasmic reticulum (ER) membrane. The Lommel laboratory previously found that RCNMV replication proteins (p27 and p88) accumulate at the ER membrane in *N. benthamiana* host and cause membrane restructuring, rearranging and proliferation [103]. Yet, there are no reports about the interaction between RCNMV and host factors/proteins associated with ER. Replication of ss(+) RNA virus is a complex process that involves numerous interactions among viral RNA, viral proteins, host proteins, and host membrane lipids. The movement of RCNMV to adjacent cells occurs through a plasmodesmata channels that connect cells. However, this intercellular transport channel is dimension restricted and so is too small for virions or even free viral DNA or RNA strands to move through. An assumption is that RCNMV requires an interaction with host proteins in order to increase the plasmodesmata size exclusion limit

(SEL) [94, 104, 105]. The most relevant information that may help understand the interaction events between RCNMV and host comes from the studies previously performed on its close relative *Tomato bushy stunt virus* (TBSV), especially the studies from the Nagy group [106, 107]. However, the Nagy group used yeast (*Saccharomyces cerevisiae*) as a model host to study the interaction between plant virus and host [108]. Genome-wide screening of yeast libraries led to the identification of a number of essential and nonessential host factors affecting TBSV RNA replication [109, 110]. The identification of yeast genes revealed that they might affect TBSV RNA recombination by altering (i) the ratio of the two viral replication proteins, (ii) the stability of the viral RNA, and/or (iii) the replicatability of the recombinant viral RNAs [110].

DNA microarray and its application to study plant-virus interactions

DNA microarrays are commonly known as DNA chips. They are used to measure the expression levels of large numbers of genes simultaneously. A DNA microarray is composed of a collection of specific DNA sequences, known as probes (or reporters or oligos) attached to a solid surface. The core principle behind microarrays is the hybridization between a probe and its corresponding target gene. The probe is made of a complementary sequence to its target gene. The probe can be a short section of a gene sequence; however, it is required to specifically hybridize to its corresponding target gene. Only the probe that hybridizes to its corresponding target gene generates a signal. The signal intensity depends upon the abundance of the target gene present in the sample [111, 112]. DNA microarrays can

therefore be used to measure changes in expression levels of genes of interest under different experimental conditions; for example, changes in the expression of a gene in a healthy and infected state. DNA microarrays provide the ability to study the expression of thousands of genes simultaneously in a single experiment.

DNA microarray limitations

Although DNA microarrays are a powerful and high throughput technology, they do have limitations when being utilized in genome-wide studies.

1. Microarrays measure only the steady-state mRNA abundance and do not measure the rate of transcription.
2. Microarrays do not measure protein abundance or protein activity levels. The correlation between mRNA measured by the microarray and its corresponding protein level may not be correlated due to post-transcriptional or post-translational regulation.
3. DNA microarray technology is not able to monitor expression of genes that are differentially expressed under certain criteria; genes that are expressed transiently, at low levels, or in a small number of cells. Careful experimental design is required for genes that exhibit these expression patterns.
4. DNA microarrays may not be able to distinguish between gene family members with high sequence similarities.

5. DNA microarrays have a capacity limitation regarding the number of probes that can be printed on a chip; therefore, typically only portions of a genome are represented on a typical chip.

6. Possible errors exist in array printing, probe sequence, probe self-hybridization, sample handling, RNA sample preparation, and hybridization procedure.

7. Statistical methods used in microarray data analysis affect conclusions. There has not been a standardization of analyses. With the same microarray data set, different statistical analyses therefore can lead to different conclusions.

8. Creation of a custom DNA microarray may require prior knowledge of genomic information (such as expressed sequence tags) of the organism of interest. This makes microarray technology more difficult to use with non-model organisms that do not have extensive genomic information available. The alternative for this issue is the use of a cDNA microarray [112]. This type of microarray does not require prior knowledge of genome under study in that clones isolated directly from cDNA libraries are spotted onto the array.

DNA microarray applications in plant-virus interactions

Plant-pathogen interactions and the resulting downstream signaling are extremely complex and dynamic, and therefore are difficult to monitor with traditional genetic and biochemical methods which focus on studying a few host genes at a time. With the advent of large-scale genomic sequencing and EST (expressed sequence tag) projects in plants and the

development of DNA microarray technology, it is now possible to simultaneously monitor the expression of thousands of plant host genes at a time, in some cases all of the plant host genes in a genome, upon a pathogen attack. Host transcriptome profiling based on microarray technology opens up tremendous opportunities to examine differential host gene expression between healthy and infected plant samples, potentially leading to the discovery of new host factors, which can then lead to identification of host metabolic/defense/signaling pathways that are associated with pathogen attack [111, 113]. Although further validation is needed, the correlative results between host genes and pathogen invasion provide a vast amount of information that can guide hypothesis-driven research to elucidate the molecular mechanisms involved in transcriptional regulation and signaling networks in plant hosts.

Microarray analysis is becoming a standard method for the study of plant-virus interactions. The majority of these microarray studies are with *Arabidopsis* [114-119] and they have revealed a remarkable complexity in host response to plant viruses that vary depending on the source of host tissue, duration of infection, virus strain, and host specificity. Comparative *Arabidopsis* microarray studies have also identified a number of common host factors induced by different plant viruses. Although *Arabidopsis* has been widely used to study host-virus interactions, it is not always an ideal host for most plant viruses that require a specific host to exhibit infected symptom. In this respect, *N. benthamiana* offers an alternative model host to study plant-virus interaction. Two previous studies have successfully utilized a potato cDNA array to probe the modulation of *N. benthamiana* gene expression by negative-sense, enveloped RNA viruses (*Sonchus yellow net virus* (SYNV) and *Impatiens necrotic spot virus* (INSV)) [120]; and positive sense RNA viruses (*Plum pox*

potyvirus (PPV), *Tomato ringspot nepovirus* (ToRSV) , and *Prunus necrotic ringspot ilarvirus* (PNRSV)) [121]. The development of a commercially available *N. benthamiana* microarray was one of the contributions of my dissertation, the details of which are described in chapter 2.

Dissertation contributions and thesis organization

How viruses modulate host physiological processes in cells, tissues and whole plants to favor their infectivity, while defeating plant host immune system remain one of the gaps in our knowledge of plant virology. We are particularly interested in determining how host physiological processes and host immune system changes correspond to each stage of a virus infection cycle.

The work in this dissertation is a unique plant-virus interaction study in two significant ways; (i) we employed a high-throughput technology to study the expression flux of thousands of host factors simultaneously upon infection and (ii) this study was designed to capture the very early events of host-virus interaction from 1 to 24 hours.

This dissertation is ultimately aimed to increase our understanding of not only the interaction between *N. benthamiana* and RCNMV, but also to expand the general knowledge base of the involvement of host factors to other ss(+)-RNA viruses. This dissertation makes three important contributions.

1: Genomic information resource and a high-throughput tool for *N. benthamiana* was generated.

N. benthamiana genomic data and a microarray were developed not only for use as a high-throughput tool to study host-virus interaction in this dissertation but also to provide a resource to the scientific community (chapter 2). We created a unigene collection, a gene expression microarray, and a functional annotation for *N. benthamiana*. Our array design is the first microarray layout of *N. benthamiana* gene expression (exon) array and also the first commercial *N. benthamiana* microarray available.

2: The global gene expression pattern of *N. benthamiana* at the early stages of viral infection was studied in order to determine changes in host physiological processes and defense responses.

The custom *N. benthamiana* microarray created from contribution 1 was utilized to study the regulation of host gene expression at the early stages of RCNMV infection. Chapter 3 describes the profiling of host gene expression changes at 2, 6, 12, and 24 hours post inoculation.

3: Plant host genes that affect viral replication at the early stages of and RCNMV infection were identified and partially characterized.

Four host genes representing various plant biological functions were selected from the microarray study (chapter 3) to further functionally assay their impact(s) on RCNMV

replication. The genes chosen were: 1) NbSAR (systemic acquired resistance, chapter 4), 2) NbWRKY (putative WRKY transcription factor, chapter 5), 3) NbBTF3 (basal transcription factor 3, chapter 6) and 4) NbSI (C-8, 7- sterol isomerase, chapter 7). NbWRKY and NbSI are identified, isolated, and characterized for the first time in this dissertation. Therefore, the work in this dissertation is the first report of these two novel genes.

References

1. Jones JD, Dangl JL: **The plant immune system.** *Nature* 2006, **444**(7117):323-329.
2. Durrant W, Dong X: **Systemic acquired resistance.** *Annu Rev Phytopathol* 2004, **42**:185-209.
3. Soosaar JL, Burch-Smith TM, Dinesh-Kumar SP: **Mechanisms of plant resistance to viruses.** *Nature Reviews Microbiology* 2005, **3**(10):789-798.
4. Agrios G: **Plant diseases caused by viruses.** *Plant Pathology Fifth Edition Elsevier Academia Press* 2005:724-820.
5. Legg J, Fauquet C: **Cassava mosaic geminiviruses in Africa.** *Plant molecular biology* 2004, **56**(4):585-599.
6. Fraser R: **The genetics of plant-virus interactions: implications for plant breeding.** *Euphytica* 1992, **63**:175-185.
7. Kang B-C, Yeam I, Jahn MM: **Genetics of plant virus resistance.** *Annu Rev Phytopathol* 2005, **43**:581-621.
8. Ross AF: **Systemic acquired resistance induced by localized virus infections in plants.** *Virology* 1961, **14**(3):340-358.
9. Heath MIC: **Apoptosis, programmed cell death and the hypersensitive response.** *European Journal of Plant Pathology* 1998, **104**(2):117-124.
10. Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner H-Y, Hunt MD: **Systemic acquired resistance.** *The Plant Cell* 1996, **8**(10):1809.

11. Baker B, Zambryski P, Staskawicz B, Dinesh-Kumar S: **Signaling in plant-microbe interactions.** *Science* 1997, **276**(5313):726-733.
12. Song W-Y, Wang G-L, Chen L-L, Kim H-S, Pi L-Y, Holsten T, Gardner J, Wang B, Zhai W-X, Zhu L-H: **A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21.** *Science* 1995, **270**(5243):1804-1806.
13. Uknes S, Mauch-Mani B, Moyer M, Potter S, Williams S, Dincher S, Chandler D, Slusarenko A, Ward E, Ryals J: **Acquired resistance in Arabidopsis.** *The Plant Cell Online* 1992, **4**(6):645-656.
14. Mauch-Mani B, Slusarenko AJ: **Production of salicylic acid precursors is a major function of phenylalanine ammonia-lyase in the resistance of Arabidopsis to Peronospora parasitica.** *The Plant Cell Online* 1996, **8**(2):203-212.
15. Delaney TP, Uknes S, Vernooij B, Friedrich L, Weymann K, Negrotto D, Gaffney T, Gut-Rella M, Kessmann H, Ward E: **A central role of salicylic acid in plant disease resistance.** *Science* 1994, **266**(5188):1247-1250.
16. Van Loon L, Van Kammen A: **Polyacrylamide disc electrophoresis of the soluble leaf proteins from Nicotiana tabacum var. 'Samsun' and 'Samsun NN': II. Changes in protein constitution after infection with tobacco mosaic virus.** *Virology* 1970, **40**(2):199-211.
17. Van Loon L: **Pathogenesis-related proteins.** *Plant molecular biology* 1985, **4**(2):111-116.
18. Ward ER, Uknes SJ, Williams SC, Dincher SS, Wiederhold DL, Alexander DC, Ahl-Goy P, Metraux J-P, Ryals JA: **Coordinate gene activity in response to agents that induce systemic acquired resistance.** *The Plant Cell Online* 1991, **3**(10):1085-1094.
19. Alexander D, Stinson J, Pear J, Glascock C, Ward E, Goodman RM, Ryals J: **A new multigene family inducible by tobacco mosaic virus or salicylic acid in tobacco.** *Molecular plant-microbe interactions* 1992, **5**(6):513-515.
20. White R, Rybicki E, Von Wechmar M, Dekker J, Antoniw J: **Detection of PR 1-type proteins in Amaranthaceae, Chenopodiaceae, Graminae and Solanaceae by immunoelectroblotting.** *Journal of General Virology* 1987, **68**(7):2043-2048.
21. Nasser W, de Tapia M, Kauffmann S, Montasser-Kouhsari S, Burkard Gr: **Identification and characterization of maize pathogenesis-related proteins. Four maize PR proteins are chitinases.** *Plant molecular biology* 1988, **11**(4):529-538.

22. Kessmann H, Staub T, Hofmann C, Maetzke T, Herzog Jr, Ward E, Uknes S, Ryals J: **Induction of systemic acquired disease resistance in plants by chemicals.** *Annual review of phytopathology* 1994, **32**(1):439-459.
23. Durner Jr, Shah J, Klessig DF: **Salicylic acid and disease resistance in plants.** *Trends in plant science* 1997, **2**(7):266-274.
24. Staskawicz BJ, Ausubel FM, Baker BJ, Ellis JG, Jones JD: **Molecular genetics of plant disease resistance.** *SCIENCE-NEW YORK THEN WASHINGTON-* 1995:661-661.
25. Whitham S, Dinesh-Kumar S, Choi D, Hehl R, Corr C, Baker B: **The product of the tobacco mosaic virus resistance gene *N*: Similarity to toll and the interleukin-1 receptor.** *Cell* 1994, **78**(6):1101-1115.
26. Padgett HS, Watanabe Y, Beachy RN: **Identification of the TMV replicase sequence that activates the *N* gene-mediated hypersensitive response.** *Molecular plant-microbe interactions* 1997, **10**(6):709-715.
27. Malamy J, Carr JP, Klessig DF, Raskin I: **Salicylic acid: a likely endogenous signal in the resistance response of tobacco to viral infection.** *Science* 1990, **250**(4983):1002-1004.
28. Metraux J, Signer H, Ryals J, Ward E, Wyss-Benz M, Gaudin J, Raschdorf K, Schmid E, Blum W, Inverardi B: **Increase in salicylic acid at the onset of systemic acquired resistance in cucumber.** *Science* 1990, **250**(4983):1004-1006.
29. Ross AF: **Localized acquired resistance to plant virus infection in hypersensitive hosts.** *Virology* 1961, **14**(3):329-339.
30. Mur LA, Bi YM, Darby RM, Firek S, Draper J: **Compromising early salicylic acid accumulation delays the hypersensitive response and increases viral dispersal during lesion establishment in TMV-infected tobacco.** *The Plant Journal* 1997, **12**(5):1113-1126.
31. White R, Antoniw J, Carr J, Woods R: **The effects of aspirin and polyacrylic acid on the multiplication and spread of TMV in different cultivars of tobacco with and without the *N* gene.** *Journal of Phytopathology* 1983, **107**(3):224-232.
32. Van Huijsduijnen RH, Alblas S, De Rijk R, Bol J: **Induction by salicylic acid of pathogenesis-related proteins and resistance to alfalfa mosaic virus infection in various plant species.** *Journal of General Virology* 1986, **67**(10):2135-2143.

33. Friedrich L, Lawton K, Ruess W, Masner P, Specker N, Rella MG, Meier B, Dincher S, Staub T, Uknes S: **A benzothiadiazole derivative induces systemic acquired resistance in tobacco.** *The Plant Journal* 1996, **10**(1):61-70.
34. Bowles DJ: **Defense-related proteins in higher plants.** *Annual review of biochemistry* 1990, **59**(1):873-907.
35. Chivasa S, Murphy AM, Naylor M, Carr JP: **Salicylic acid interferes with tobacco mosaic virus replication via a novel salicylhydroxamic acid-sensitive mechanism.** *The Plant Cell Online* 1997, **9**(4):547-557.
36. Laties GG: **The cyanide-resistant, alternative path in higher plant respiration.** *Annual Review of Plant Physiology* 1982, **33**(1):519-555.
37. Raskin I: **Role of salicylic acid in plants.** *Annual review of plant biology* 1992, **43**(1):439-463.
38. McIntosh L: **Molecular biology of the alternative oxidase.** *Plant Physiology* 1994, **105**(3):781.
39. Day D, Whelan J, Millar A, Siedow J, Wiskich J: **Regulation of the alternative oxidase in plants and fungi.** *Functional plant biology* 1995, **22**(3):497-509.
40. Porta H, Rocha-Sosa M: **Plant lipoxygenases. Physiological and molecular features.** *Plant Physiology* 2002, **130**(1):15-21.
41. Siedow JN: **Plant lipoxygenase: structure and function.** *Annual review of plant biology* 1991, **42**(1):145-188.
42. Kolomiets MV, Hannapel DJ, Chen H, Tymeson M, Gladon RJ: **Lipoxygenase is involved in the control of potato tuber development.** *The Plant Cell Online* 2001, **13**(3):613-626.
43. Veronesi C, Rickauer M, Fournier J, Pouenat M-L, Esquerre M-T: **Lipoxygenase gene expression in the tobacco-Phytophthora parasitica nicotianae interaction.** *Plant Physiology* 1996, **112**(3):997-1004.
44. Kuhn H, Thiele BJ: **The diversity of the lipoxygenase family: Many sequence data but little information on biological significance.** *FEBS letters* 1999, **449**(1):7-11.
45. Feussner I, Wasternack C: **The lipoxygenase pathway.** *Annual review of plant biology* 2002, **53**(1):275-297.
46. Blee E: **Impact of phyto-oxylipins in plant defense.** *Trends in plant science* 2002, **7**(7):315-322.

47. Creelman RA, Mullet JE: **Biosynthesis and action of jasmonates in plants.** *Annual review of plant biology* 1997, **48**(1):355-381.
48. Vick BA, Zimmerman DC: **The biosynthesis of jasmonic acid: a physiological role for plant lipoxygenase.** *Biochemical and biophysical research communications* 1983, **111**(2):470-477.
49. Rance I, Fournier J, Esquerre-Tugaye M-T: **The incompatible interaction between *Phytophthora parasitica* var. *nicotianae* race 0 and tobacco is suppressed in transgenic plants expressing antisense lipoxygenase sequences.** *Proceedings of the National Academy of Sciences* 1998, **95**(11):6554-6559.
50. Bannenberg G, MartÃ-nez M, Hamberg M, Castresana C: **Diversity of the enzymatic activity in the lipoxygenase gene family of *Arabidopsis thaliana*.** *Lipids* 2009, **44**(2):85-95.
51. Weber H, Vick BA, Farmer EE: **Dinor-oxo-phytodienoic acid: a new hexadecanoid signal in the jasmonate family.** *Proceedings of the National Academy of Sciences* 1997, **94**(19):10473-10478.
52. Aranda M, Maule A: **Virus-induced host gene shutoff in animals and plants.** *Virology* 1998, **243**(2):261-267.
53. Escaler M, Aranda MA, Thomas CL, Maule AJ: **Pea embryonic tissues show common responses to the replication of a wide range of viruses.** *Virology* 2000, **267**(2):318-325.
54. Aranda MA, Escaler M, Wang D, Maule AJ: **Induction of HSP70 and polyubiquitin expression associated with plant virus replication.** *Proceedings of the National Academy of Sciences* 1996, **93**(26):15289-15293.
55. Tena G, Asai T, Chiu W-L, Sheen J: **Plant mitogen-activated protein kinase signaling cascades.** *Current opinion in plant biology* 2001, **4**(5):392-400.
56. Zhang S, Klessig DF: **MAPK cascades in plant defense signaling.** *Trends in plant science* 2001, **6**(11):520-527.
57. Asai T, Tena G, Plotnikova J, Willmann MR, Chiu W-L, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J: **MAP kinase signalling cascade in *Arabidopsis* innate immunity.** *Nature* 2002, **415**(6875):977-983.
58. Whitmarsh AJ: **Regulation of gene transcription by mitogen-activated protein kinase signaling pathways.** *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 2007, **1773**(8):1285-1298.

59. Popescu SC, Popescu GV, Bachan S, Zhang Z, Gerstein M, Snyder M, Dinesh-Kumar SP: **MAPK target networks in Arabidopsis thaliana revealed using functional protein microarrays.** *Genes & Development* 2009, **23**(1):80-92.
60. Frye CA, Tang D, Innes RW: **Negative regulation of defense responses in plants by a conserved MAPKK kinase.** *Proceedings of the National Academy of Sciences* 2001, **98**(1):373-378.
61. Innes RW: **Mapping out the roles of MAP kinases in plant defense.** *Trends in plant science* 2001, **6**(9):392-394.
62. Petersen M, Brodersen P, Naested H, Andreasson E, Lindhart U, Johansen B, Nielsen HB, Lacy M, Austin MJ, Parker JE: **Arabidopsis MAP Kinase 4 Negatively Regulates Systemic Acquired Resistance.** *Cell* 2000, **103**(7):1111-1120.
63. Zhang S, Klessig DF: **Resistance gene N-mediated de novo synthesis and activation of a tobacco mitogen-activated protein kinase by tobacco mosaic virus infection.** *Proceedings of the National Academy of Sciences* 1998, **95**(13):7433-7438.
64. Romeis T, Piedras P, Zhang S, Klessig DF, Hirt H, Jones JD: **Rapid Avr9- and Cf-9-dependent activation of MAP kinases in tobacco cell cultures and leaves: Convergence of resistance gene, elicitor, wound, and salicylate responses.** *The Plant Cell Online* 1999, **11**(2):273-287.
65. Liu Y, Jin H, Yang KY, Kim CY, Baker B, Zhang S: **Interaction between two mitogen-activated protein kinases during tobacco defense signaling.** *The Plant Journal* 2003, **34**(2):149-160.
66. Seo S, Okamoto M, Seto H, Ishizuka K, Sano H, Ohashi Y: **Tobacco MAP kinase: a possible mediator in wound signal transduction pathways.** *Science* 1995, **270**(5244):1988-1992.
67. Yang K-Y, Liu Y, Zhang S: **Activation of a mitogen-activated protein kinase pathway is involved in disease resistance in tobacco.** *Proceedings of the National Academy of Sciences* 2001, **98**(2):741-746.
68. Zhang S, Liu Y, Klessig DF: **Multiple levels of tobacco WIPK activation during the induction of cell death by fungal elicitors.** *The Plant Journal* 2000, **23**(3):339-347.
69. Ishihama N, Yamada R, Yoshioka M, Katou S, Yoshioka H: **Phosphorylation of the Nicotiana benthamiana WRKY8 transcription factor by MAPK functions in the defense response.** *The Plant Cell Online*, **23**(3):1153-1170.

70. Zhang S, Du H, Klessig DF: **Activation of the tobacco SIP kinase by both a cell wall-derived carbohydrate elicitor and purified proteinaceous elicitors from *Phytophthora* spp.** *The Plant Cell Online* 1998, **10**(3):435-449.
71. Goodwin MM, Zaitlin D, Naidu RA, Lommel SA: ***Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions.** *Molecular plant-microbe interactions* 2008, **21**(8):1015-1026.
72. Clemente T: ***Nicotiana* (*Nicotiana tabacum*, *Nicotiana benthamiana*).** In: *Agrobacterium Protocols*. Springer; 2006: 143-154.
73. Goodspeed TH, Wheeler H, Hutchison P: **The genus *Nicotiana*: origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics**, vol. 16: Chronica Botanica Company Waltham; 1954.
74. Senthil-Kumar M, Mysore KS: **New dimensions for VIGS in plant functional genomics.** *Trends in plant science*, **16**(12):656-665.
75. Wydro M, Kozubek E, Lehmann Pa: **Optimization of transient *Agrobacterium*-mediated gene expression system in leaves of *Nicotiana benthamiana*.** *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION*- 2006, **53**(2):289.
76. Gleba Y, Klimyuk V, Marillonnet S: **Magniffection-a new platform for expressing recombinant vaccines in plants.** *Vaccine* 2005, **23**(17):2042-2048.
77. Matoba N, Davis KR, Palmer KE: **Recombinant protein expression in *Nicotiana*.** In: *Plant Chromosome Engineering*. Springer: 199-219.
78. Strasser R, Stadlmann J, Schahs M, Stiegler G, Quendler H, Mach L, Glossl J, Weterings K, Pabst M, Steinkellner H: **Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure.** *Plant biotechnology journal* 2008, **6**(4):392-402.
79. Dasgupta R, Garcia BH, Goodman RM: **Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes.** *Proceedings of the National Academy of Sciences* 2001, **98**(9):4910-4915.
80. Bennett MD, Leitch IJ: **Nuclear DNA amounts in angiosperms.** *Annals of Botany* 1995, **76**(2):113-176.
81. Bennett MD, Leitch IJ, Price HJ, Johnston JS: **Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* genome initiative estimate of ~125 Mb.** *Annals of Botany* 2003, **91**(5):547-557.

82. Bennett M, Leitch I: **Nuclear DNA amounts in angiosperms: progress, problems and prospects.** *Annals of Botany* 2005, **95**(1):45-90.
83. Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB: **A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research.** *Molecular plant-microbe interactions* 2012, **25**(12):1523-1530.
84. Bombarely A, Menda N, Teclé IY, Buels RM, Strickler S, Fischer-York T, Pujar A, Leto J, Gosselin J, Mueller LA: **The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl.** *Nucleic acids research*, **39**(suppl 1):D1149-D1155.
85. Gould A, Francki R, Hatta T, Hollings M: **The bipartite genome of red clover necrotic mosaic virus.** *Virology* 1981, **108**(2):499-506.
86. Xiong Z, Lommel S: **The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1.** *Virology* 1989, **171**(2):543-554.
87. Lommel S, Weston-Fina M, Xiong Z, Lomonosoff GP: **The nucleotide sequence and gene organization of red clover necrotic mosaic virus RNA-2.** *Nucleic acids research* 1988, **16**(17):8587-8602.
88. Hollings M, Stone OM: **Red clover necrotic mosaic virus.** *CMI/AAB Descriptions of Plant Viruses* 1977, **181**.
89. Kim K-H, Lommel SA: **Sequence element required for efficient⁺ 1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus.** *Virology* 1998, **250**(1):50-59.
90. Kim K, Lommel S: **Identification and analysis of the site of-1 ribosomal frameshifting in red clover necrotic mosaic virus.** *Virology* 1994, **200**(2):574-582.
91. Koonin EV: **The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses.** *J Gen Virol* 1991, **72**(Pt 9):2197-2206.
92. Zavriev S, Hickey C, Lommel S: **Mapping of the red clover necrotic mosaic virus subgenomic RNA.** *Virology* 1996, **216**(2):407-410.
93. Xiong Z, Kim K, Giesman-Cookmeyer D, Lommel S: **The roles of the red clover necrotic mosaic virus capsid and cell-to-cell movement proteins in systemic infection.** *Virology* 1993, **192**(1):27-32.
94. Fujiwara T, Giesman-Cookmeyer D, Ding B, Lommel SA, Lucas WJ: **Cell-to-cell trafficking of macromolecules through plasmodesmata potentiated by the red clover necrotic mosaic virus movement protein.** *The Plant Cell Online* 1993, **5**(12):1783-1794.

95. Osman T, Buck K: **Replication of red clover necrotic mosaic virus RNA in cowpea protoplasts: RNA 1 replicates independently of RNA 2.** *Journal of General Virology* 1987, **68**(2):289-296.
96. Paje-Manalo L, Lommel S: **Independent replication of red clover necrotic mosaic virus RNA-1 in electroporated host and nonhost *Nicotiana* species protoplasts.** *Phytopathology* 1989, **79**(4):457-461.
97. Sit TL, Vaewhongs AA, Lommel SA: **RNA-mediated trans-activation of transcription from a viral RNA.** *Science* 1998, **281**(5378):829-832.
98. Wilkie GS, Dickson KS, Gray NK: **Regulation of mRNA translation by 5'- and 3'-UTR-binding factors.** *Trends in biochemical sciences* 2003, **28**(4):182-188.
99. Mizumoto H, Tatsuta M, Kaido M, Mise K, Okuno T: **Cap-independent translational enhancement by the 3' untranslated region of *Red clover necrotic mosaic virus* RNA1.** *Journal of Virology* 2003, **77**(22):12113-12121.
100. Xiong Z, Lommel SA: **Red clover necrotic mosaic virus infectious transcripts synthesized *in Vitro*.** *Virology* 1991, **182**(1):388-392.
101. Takeda A, Tsukuda M, Mizumoto H, Okamoto K, Kaido M, Mise K, Okuno T: **A plant RNA virus suppresses RNA silencing through viral RNA replication.** *The EMBO journal* 2005, **24**(17):3147-3157.
102. Powers JG, Sit TL, Heinsohn C, George CG, Kim K-H, Lommel SA: **The Red clover necrotic mosaic virus RNA-2 encoded movement protein is a second suppressor of RNA silencing.** *Virology* 2008, **381**(2):277-286.
103. Turner KA, Sit TL, Callaway AS, Allen NS, Lommel SA: **Red clover necrotic mosaic virus replication proteins accumulate at the endoplasmic reticulum.** *Virology* 2004, **320**(2):276-290.
104. Wolf S, LUCAS WJ, DEOM CM, BEACHY RN: **Movement protein of *Tobacco mosaic virus* modifies plasmodesmatal size exclusion limit.** *Science* 1989, **246**(4928):377-379.
105. Derrick PM, Barker H, Oparika KJ: **Increase in Plasmodesmatal Permeability during Cell-to-Cell Spread of *Tobacco Rattle Virus* from Individually Inoculated Cells.** *The Plant Cell Online* 1992, **4**(11):1405-1412.
106. Nagy PD, Pogany J: **The dependence of viral RNA replication on co-opted host factors.** *Nature Reviews Microbiology*, **10**(2):137-149.

107. Nagy PD, Pogany J: **Global Genomics and Proteomics Approaches to Identify Host Factors as Targets to Induce Resistance Against *Tomato Bushy Stunt Virus*.** *Advances in virus research*, **76**:123-177.
108. Pogany J, Panavas T, Serviene E, Nawaz-ul-Rehman MS, D Nagy P: **A High-Throughput Approach for Studying Virus Replication in Yeast.** *Current protocols in microbiology*:16J. 11.11-16J. 11.15.
109. Jiang Y, Serviene E, Gal J, Panavas T, Nagy PD: **Identification of essential host factors affecting tombusvirus RNA replication based on the yeast Tet promoters Hughes Collection.** *Journal of Virology* 2006, **80**(15):7394-7404.
110. Serviene E, Jiang Y, Cheng C-P, Baker J, Nagy PD: **Screening of the yeast yTHC collection identifies essential host factors affecting tombusvirus RNA recombination.** *Journal of Virology* 2006, **80**(3):1231-1241.
111. Brown PO, Botstein D: **Exploring the new world of the genome with DNA microarrays.** *Nature genetics* 1999, **21**:33-37.
112. Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM: **Expression profiling using cDNA microarrays.** *Nature genetics* 1999, **21**:10-14.
113. Harmer SL, Kay SA: **Microarrays: determining the balance of cellular transcription.** *The Plant Cell Online* 2000, **12**(5):613-615.
114. Golem S, Culver JN: **Tobacco mosaic virus induced alterations in the gene expression profile of *Arabidopsis thaliana*.** *Molecular plant-microbe interactions* 2003, **16**(8):681-688.
115. Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang H-S, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA: **Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses.** *The Plant Cell Online* 2002, **14**(3):559-574.
116. Huang Z, Yeakley JM, Garcia EW, Holdridge JD, Fan J-B, Whitham SA: **Salicylic acid-dependent expression of host genes in compatible *Arabidopsis*-virus interactions.** *Plant Physiology* 2005, **137**(3):1147-1159.
117. Marathe R, Guan Z, Anandalakshmi R, Zhao H, Dinesh-Kumar S: **Study of *Arabidopsis thaliana* resistome in response to Cucumber mosaic virus infection using whole genome microarray.** *Plant molecular biology* 2004, **55**(4):501-520.
118. Trinks D, Rajeswaran R, Shivaprasad P, Akbergenov R, Oakeley EJ, Veluthambi K, Hohn T, Pooggin MM: **Suppression of RNA silencing by a geminivirus nuclear**

- protein, AC2, correlates with transactivation of host genes. *Journal of Virology* 2005, **79**(4):2517-2527.**
119. Whitham SA, Quan S, Chang HS, Cooper B, Estes B, Zhu T, Wang X, Hou YM: **Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants.** *The Plant Journal* 2003, **33**(2):271-283.
120. Senthil G, Liu H, Puram V, Clark A, Stromberg A, Goodin M: **Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses.** *Journal of General Virology* 2005, **86**(9):2615-2625.
121. Dardick C: **Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses.** *Molecular plant-microbe interactions* 2007, **20**(8):1004-1017.

Chapter 2

Analysis of expressed sequence tags in *Nicotiana benthamiana* defines a unigene set comprising 13,000 genes

Chapter summary

Herein, I describe the development of a *N. benthamiana* 13 K unigene collection, a microarray, and functional annotation of the unigenes. This unigene collection covers approximately 38% of the *N. benthamiana* genome. This report constitutes the first microarray of *N. benthamiana* gene expression (exon) array and also the first commercial *N. benthamiana* microarray available from Nimblegen Company. The performance of this *N. benthamiana* microarray is illustrated in the host-RCNMV interaction study which is described in Chapter 3.

Abstract

Because of the complexity of the *Nicotiana benthamiana* genome and its estimated size at 3 Gb, sequencing of the expressed portion provides an alternative approach to rapidly study gene expression. Here we report our collection of *N. benthamiana* expressed sequence tags (ESTs). We sequenced 12,000 ESTs from three libraries: a standard (expanding leaf tissue), a normalized (green aerial part of the plant), and a subtracted cDNA (a library of healthy against a library of *Colletotrichum destructivum*-infected *N. benthamiana*) libraries. The expression analysis shows that, as expected, RuBisCO- and photosynthesis- related genes are highly expressed among these three libraries. Our sequenced ESTs were combined with publicly available ESTs, producing a total set of 41,700 *N. benthamiana* ESTs which was then condensed down to 13,014 unigenes. Homology searches against GenBank non-redundant protein sequences (nr) database using BLASTX and *E*-value cut off of 1×10^{-5} showed that 66% of the unigenes had at least one hit. A majority of homologs contributed to cell growth, enzymes, metabolism, cell communication, defense, and transcription regulation. We identified 261 unigenes that are involved in transcription. The most redundant transcription factor-related genes belong to AP2/ERF, WRKY and GRAS families. To enable a genome-wide study in *N. benthamiana*, we developed the first ready-to-use 13K unigene – based *N. benthamiana* microarray. The array probes are 60-oligonucleotides long. The array contains 9 probes per unigene and 3 technical replicates per probe printed on the 385K-formatted array. Currently, a custom *N. benthamiana* microarray and its unigene collection,

array design, sequences, and functional annotation information are available through a request from the Lommel laboratory and Nimblegen Company.

Keywords; *Nicotiana*, *Nicotiana benthamiana*, ESTs, unigenes, microarray, gene expression

Introduction

N. benthamiana is a dicot plant, a close relative to tobacco, and a species in the genus *Nicotiana*, family *Solanaceae*. It is indigenous to Australia, more specifically the Northern Territory and Western Australia, and it originally arose as a result of the hybridization between the two progenitor species of *N. debneyi* and *N. suaveolens* [1-3]. Next to *Arabidopsis thaliana*, *N. benthamiana* is the prominent model organism for conducting research in the field of plant biology and plant pathology. It is an excellent experimental plant subject and the best known model plant for performing agroinfiltration and virus-induced gene silencing (VIGS) to conduct plant functional assays, either transient gene knockdown or transient protein expression studies [4, 5]. Here, we demonstrate the phenotype of silenced *PDS* gene (phytoene desaturase) in *N. benthamiana* using TRV (*Tobacco rattle virus*)-VIGS method in Figure 1. The suppression of the *PDS* gene exhibits a visible leaf bleaching symptom at day 7-10 after agroinfiltration (*Agrobacterium tumefaciens*). The phenotype is systemic due to the spreading of the *PDS* gene silencing signal to the upper and younger non- agro-infiltrated leaves from the agroinfiltrated leaf. *N. benthamiana* is also an efficient and a viable alternative platform for the industrial production of recombinant proteins, particularly in the field of pharmaceutical biotechnology [6-8]. *N. benthamiana* is

susceptible to a wide range of plant pathogens (bacteria, oomycetes, fungi, and viruses), making this plant species a cornerstone host for host-pathogen interaction studies, [1]. In comparison to *Arabidopsis thaliana*, *N. benthamiana* is susceptible to a wider range of pathogens and its life-cycle is longer allowing for productive pathogen infections.

N. benthamiana is the most widely used host in plant virology studies due mainly to its susceptibility to more than 500 plant viruses, even to those that are typically restricted to monocot hosts and, most surprisingly, even to several animal viruses [9]. *N. benthamiana* is highly capable of (i) allowing plant viruses to establish a local infection, and (ii) facilitating viral cell-to-cell movement leading to a systemic infection. Therefore *N. benthamiana* is a local and systemic host for many plant viruses. Here, we demonstrate a viral susceptibility of *N. benthamiana* in Figure 2. Using, *Red clover necrotic mosaic virus*, (RCNMV) infected *N. benthamiana* exhibit local lesions on the inoculated leaves at 2-3 days after inoculation. The GFP protein expressed from a modified RCNMV can be detected under the microscope as early as 8 hours after inoculation, within individually infected cells. RCNMV systemic symptoms appear 3-4 days after inoculation, which are leaf curling, growth stunting, and characteristic necrotic-mosaic.

Although *N. benthamiana* is an excellent plant model in many research areas, the genome-wide study and functional analysis of this plant species is likely far more difficult, largely due to the complexity of its genome structure and organization. The *N. benthamiana* genome (~ 3 Gb) is nearly 20 fold larger than the *A. thaliana* genome (~ 135 Mbp) [10-12]. *N. benthamiana* is an amphiploid species with $n = 19$; whereas, *A. thaliana* is a diploid

species with $n=5$. Genome sequencing is the cornerstone of high-throughput functional analysis and genomic study. It is fundamental to understanding the genetic composition of an organism. The first *N. benthamiana* draft genome was recently published in 2012. This *N. benthamiana* genome draft is a 63-fold coverage and is available on Sol Genomic Network for BLAST search and for downloading to local server [13]. However, because of the size and the complexity of the *N. benthamiana* genome structure and organization, together with a shortage of other sources of genomic database, it is likely that a comprehensive sequence and assembly, functional annotation, and gene/protein predictions will remain computationally and experimentally challenging. This complication leads to a lack of essential genomic information to enable the use of *N. benthamiana* as a plant model for genome-wide study.

In this project, we aim to provide an alternative approach to enable an immediate and productive high-throughput gene function and pathway studies of *N. benthamiana*. While awaiting the comprehensive genome information, a random sequencing of gene transcripts and expressed sequence tags (ESTs) has been recognized as a simple and efficient method to identify expressed genes in an organism. ESTs are short ~ 200-800 nucleotide, unedited, randomly selected single-pass sequence reads derived from cDNA (complementary DNA) libraries. ESTs and cDNAs provide direct evidence for all the sample transcripts and they are currently one of the most important resources for transcriptome exploration [14]. In 1991, ESTs were used as a primary source for human gene discovery [15]. Thereafter, for at least two decades, high-throughput EST sequencing has become a valuable, efficient, and rapid method for gene and pathway discovery in plants, human, animals, and microorganisms [15-23]. The sources of the cDNA libraries that are used for EST sequencing can come from

various tissue types which could be derived from different organs and different developmental stages, as well as, different environmental and stress exposure. Certain plant genes are tightly regulated both temporally and spatially [21, 24]. In addition, certain genes are expressed only under specific biotic or abiotic stresses [25-28]. The use of various sources to obtain ESTs increases the representation of events that are able to induce the activation of transcription of the majority of the organism's genes to be expressed highly enough to be detected and sequenced during the sequencing process. On the other hand, with a use of a particular source to produce ESTs, an investigation of the ESTs could help identify the highly expressed genes under the particular experimental treatment. This comes from the assumption that when sampling is truly random from these different tissue types, the frequency of any given EST is related to differentials in gene expression on the conditions in which the corresponding gene is acting [23, 29-31]. So, an analysis of ESTs could provide an approach to identify genes that are significantly expressed in response to a particular biological milieu. The collection of ESTs also provides valuable preliminary information for genome studies. When aligning ESTs to the genome using physical mapping, it provides a resource to ensure the genome assembly and the estimation of gene numbers is accurate as well as distinguishing authentic genes from pseudogenes. ESTs were first used to construct maps of the human genome [32]. EST data mining is a powerful tool to help refine predicted transcript, SNP (single-nucleotide polymorphism) characterization, and splicing events for those predicted genes, which leads to a prediction of their corresponding gene products and ultimately their functions [14, 33-40]. Global, multi-tissue EST projects have been reported for *Arabidopsis* [41], rice [31], soybean [42], potato [24], and *Populus* [43]. To date, no

report exists of global, multi-tissue EST analysis from *N. benthamiana*. With an urgency to enable the use of *N. benthamiana* in many research areas, particularly as a plant model in genome-wide study, we propose here the examination of *N. benthamiana* ESTs, unigene assembly, and functional annotation.

Single gene-orientated studies emphasize the study of one particular gene or a few genes at a time. This type of study is practical when research questions are focused on a certain aspect of the gene function. Despite genes being well-characterized in any single gene-orientated study; information obtained from this type of study limits an ability to resolve two fundamental research problems: (i) how an individual gene is inter-connected to other genes in biological networks; and (ii) how biological networks are comprehensively altered during the experimental challenges. One gene could be involved in multiple metabolic pathways. For this reason, high-throughput approaches are required to be implemented in order to tackle these two issues. In this project, we are particularly interested in the application microarray technology to develop the first *N. benthamiana* microarray as a ready-to-use genomic tool. In recent years, the number of available plant genomic tools has grown greatly. In fact, a search of the Affymetrix® website shows that not only is a whole genome array of *A. thaliana* offered, but there are now over 10 different plant arrays of agronomically important crops readily available for expression analysis. However, at the time of this study, a *N. benthamiana* array was not commercially available.

Gene expression microarray is a very innovative high-throughput technology. It is prepared by a high-speed robotic printing of the complementary DNAs on glass which were

used to measure a quantitative expression of the corresponding genes. Microarray was developed to monitor the expression of many genes in parallel. In other words, it allows the host transcriptomic profile to be collectively and simultaneously gathered from individual genes in a single shot of study [44]. This pushes a boundary of our curiosity from the perspective of single gene study to many thousand gene study at one time of experiment. Bioinformatics and genomic databases are necessary tools to manipulating large datasets from the high-throughput experiment, including a dataset from microarray experiment. They extend an ability to extract significant and meaningful information from a massive genomic data to predict how the whole network of an organism's physiological process behaves in response to the experimental challenges. A collection of differential gene expressions are dissected to construct the predicted models for metabolic pathway networks, gene-regulatory network (protein-DNA interaction such as transcription factor and its cognate nucleotide sequences), and interactome network (protein-protein interactions) that are impacted by the experimental challenges.

Previous genome-wide studies using *N. benthamiana* as an experimental plant model have employed the heterologous microarray made from *Solanum tuberosum* (potato) for hybridization [24, 45, 46]. Potato is a close relative to *N. benthamiana* in the family *Solanaceae*. In spite of their close genetic relatedness [47, 48], the hybridization capacity of cDNA synthesized from *N. benthamiana* total RNA to the potato microarray is only 75% [45]. Here we provide a ready-to-use genomic tool of *N. benthamiana* for a the scientific community. We collaborated with Nimblegen Company to develop the first gene expression microarray of *N. benthamiana* by utilizing our EST and unigenes collection, and we present

here the design of our custom microarray. The validation and performance test of our custom microarray is thoroughly described in Chapter 3.

We hope that our unigene collection, annotation information, and microarray presented in this work could serve as the promising powerful genomic resource and tool for genome-wide study in *N. benthamiana*.

Materials and Methods

1) Sources of EST sequences

A collection of *N. benthamiana* EST sequences were compiled from six distinct libraries and a summary shown in Table 1.

(i) A standard library prepared from expanding leaf tissue of *N. benthamiana*, cultivar Berkeley. The standard library was composed of ESTs from expanded leaf tissue made commercially by Amplicon Express, Inc (Amplicon Express, Inc., Pullman, WA). Total RNA was extracted and the poly A fraction isolated on oligo-dT column. The mRNA was reverse transcribed and subsequently double stranded. The poly A double stranded cDNA ranging in size from 400-4000 bp was directionally cloned into the EcoR I – Xho I restriction site of pBluescript S/K⁺ vector for sequencing.

(ii) A normalized library created in Dr. Andy Maule's laboratory at the John Innes Centre (Norwich, UK) from the green aerial parts of the plant. HAP columns were used to normalize the double stranded cDNA three times.

(iii) A subtracted library donated by Dr. Paul Goodwin from University of Guelph (Guelph, Canada). This library is a subtraction from a library of healthy *N. benthamiana* against a library of *Colletotrichum destructivum*-infected *N. benthamiana*.

(iv) A US library.

(v) A TIGR *N. benthamiana* EST libraries (The Institute for Genomic Research)

The EST sequences were derived from a search of TIGR database. The TIGR library combined *N. benthamiana* ESTs from various tissue types that were exposed to different abiotic and biotic stresses. RNA was isolated from *N. benthamiana* tissues that include callus, roots from liquid culture grown plants, heat-stressed leaves (38 °C, 3 hr and 6 hr), cold-stressed leaves (5 °C 3 hr, and 6hr), and pathogen challenged leaves (*Pseudomonas syringae* pv tomato 12 hr; *Xanthomonas campestris* pv campestris 12 hr, 18 hr; *Pseudomonas syringae* pv phaseolicola 18 hr, and *Xanthomonas campestris* pv vesicatoria 18 hr). RNA was isolated from these tissues and pooled in approximately equal molar amounts.

Complementary DNA was cloned into pCMVSPORT6.1 vector through EcoRI and NotI for sequencing.

(vi) A GenBank *N. benthamiana* EST libraries

2) Sequence processing pipeline

The EST clones from libraries i, ii, and iii were sequenced by the Tobacco Genome Initiative (TGI, North Carolina State University, Raleigh, NC). Prepared DNA was sequenced on an ABI3700 DNA sequencer (Applied Biosystems, Foster City, CA). The EST

clones from library (iv) were sent to Agencourt for sequencing. After eliminating all poor quality sequences, the sequenced *N. benthamiana* ESTs were vector trimmed using a PEARL script created by the TGI. Vector trimmed sequences were run through RepeatMasker to screen for “interspersed repeats and low complexity DNA sequences” [49, 50]. These ESTs were then combined with the remaining *N. benthamiana* sequences from TIGR and GenBank to be collated into unigenes.

3) Contig/Unigene assembly

The EST sequences that passed the sequence processing pipeline were combined and assembled into contigs using the default parameter setting of the CAP3 program [51]. Incorporation of ESTs into a contig requires a minimum of a 40 bp overlap and at least 95% sequence identity.

4) Functional annotation of unigene sequences

Blast2GO software [52] was applied to assign functional information to the unigenes (Figure 12). The unigene sequences were compared to the GenBank nr database using BLASTX [53] and only similarities with expected value (E) smaller or equal to 1×10^{-5} were further classified into categories according to gene ontology (GO) scheme. With the GO system, gene functions are classified into three categories: cellular component, biological process and molecular function [54]. In addition, all unigenes were searched against KEGG

[55] and InterProScan [56] databases to assign their protein functions, and their involvement in enzymatic reactions and metabolic pathways.

5) The construction of *N. benthamiana* microarray

The set of unigene sequences were submitted to Nimblegen Systems, Inc (Madison, WI) to construct a custom gene expression microarray (array design ID 1911 and 1871). The array design is based on Nimblegen's proprietary method.

Result and discussion

***N. benthamiana* ESTs and unigene assembly**

Messenger RNAs (mRNAs) in the cell represent copies from the expressed genes. As RNA cannot be cloned directly, they are reverse transcribed to double-stranded cDNA using reverse transcriptase. The resultant cDNAs were then cloned to make libraries representing a set of transcribed genes of the original cells or tissues. Subsequently, these cDNA clones are randomly sequenced from both directions in a single-pass run with no validation or full-length sequencing to obtain 5' to 3' chromatogram ESTs. These resultant ESTs are pre-processed into high-quality ESTs wherein they are screened for sequence repeats, contaminants, and low-complexity sequences, which are eliminated. The collection of EST sequences in this project represent a wide range of tissue types from different plant developmental stages and organs (liquid culture grown plant, expanding leaves, and root), as well as, biotic and abiotic challenges (heat-stressed, cold-stressed, and pathogen challenged

tissues). All EST sequences from six EST libraries were combined to generate a total of 41,700 *N. benthamiana* EST sequences in the collection. A majority of TGI and TIGR sequences is available in the GenBank EST database [57] under the library name “LIBEST_015697 SAL_US N. Benthome” [58] and “LIBEST_015055 *N. benthamiana* mixed tissue cDNA library, normalized, full-length” [59] respectively. All EST sequences in the collection were passed through a sequence processing pipeline for a quality check prior to generating a unigene assembly. The set of 41,700 EST sequences were clustered and the assembly approach yielded 5,948 contigs with an average length of 963 bp (a range between 45 bp and 6447 bp), and 7066 singletons with an average length of 752 bp (a range between 9 bp and 6429 bp). The combined contigs and singletons produced a total of 13,014 unique sequences with an average length of 752 bp (a range between 9 bp and 6447 bp). The distribution of unigene length is illustrated in Figure 3. The average length of our ESTs is considered highly informative [14]. Each unigene was oriented from the 5' to 3' direction using three criteria: (i) an orientation according to a BLASTX hit against GenBank nr database with *E*-value cut off of less than 1×10^{-5} , (ii) unigenes with no BLAST hit orientated according to a sequencing direction of the underlying ESTs, and (iii) the remaining sequences were screened for polyadenylation tail to identify 3' end of a transcript.

Potential highly-expressed *N. benthamiana* genes from standard, normalized, and subtracted EST libraries

We analyzed sequence similarity and functional annotation of 3888 clones that were derived from three EST libraries: standard library, subtracted library, and normalized library. Expression patterns across these libraries were discerned using BLASTClust software [60]. BLASTClust is a program within the standalone BLAST package. It was used to cluster sequence similarity of either protein or nucleotide sequences. The program begins with pairwise matches and places a sequence in a cluster if the sequence matches at least one sequence already in the cluster. In the case of nucleotide sequences, the Megablast algorithm was used to compute the pairwise matches. The functional annotation of each cluster was obtained by BLASTX with *E*-value cut off of 1×10^{-5} . Considering the frequency of EST clones among these three libraries, the potential highly-expressed genes are reported in Table 2. Proteins that belong to ribulose biphosphate carboxylase (RuBisCO) family were found to be the most highly-expressed gene and the most highly-correlated gene among standard, normalized, and subtracted libraries.

Functional analysis of *N. benthamiana* unigenes

Homology searches of 13,014 *N. benthamiana* unigenes against GenBank nr database [57] using BLASTX and *E*-value cut off of less than 1×10^{-5} [53] showed that 66% of the unigenes have at least one hit to an existing sequence. Of all the sequences that have BLAST

hits, 76% have an *E*-value between 1×10^{-100} to 1×10^{-10} (Figure 4). The average sequence identity between unigene queries and their top BLAST hits is 82% (Figure 5).

We analyzed functional annotation using a top hit from each unigene. A distribution of top hit organism species is demonstrated in Figure 6. A majority of homologous proteins are from grape (*Vitis vinifera*), castor oil plant (*Ricinus communis*), and black cottonwood (*Populus trichocarpa*). However, as would be expected, the most frequent top hit plant species are in the *Solanaceae* family. About 24% of *N. benthamiana* unigenes are homologous to genes from other *Solanaceae* plants, 2.3% to *A. thaliana*, and 2% to soybean; whereas, 0.6% to rice and 0.3% to corn. A majority of the homologous protein functions contributed to cell growth, enzymes, cell communication, defense, and transcription regulation. We identified 261 unigenes that are involved in the transcription regulatory network. The most redundant transcription factor-related genes belong to AP2/ERF, WRKY and GRAS transcription factor superfamilies. Of the 13,014 input unigene sequences, 7,841 returned at least one GO hit, which is about 60% of the starting dataset. This return is not as high as we expected. However, given only 66% of unigenes had a BLAST hit, we believe that the GO hit result demonstrates that our search method is efficient. The 27,254 gene ontology terms for three principle GO categories, biological process, molecular function and cellular component were assigned to 4,493, 5,457, and 4,537 unigenes respectively. Approximately 60% of the biological process hits fell into the anabolic/catabolic process category which includes metabolic process of carbohydrate, lipid, protein, DNA, and secondary metabolites (Figure 7). About 25% of the hits related to cellular communication, signal transduction, and stress response. While only 10% involved in cell growth and

maintenance. The largest proportion of molecular function (Figure 8), about 46%, corresponded to binding activities (chromatin, DNA, RNA, nucleotide, protein, ion, and metal binding); while catalytic activities (peptidase, hydrolase, transferase, kinase, and phosphatase) were almost equally represented by 37% of the hits in molecular function. Under the cellular component (Figure 9), 75% of annotated unigenes were predicted to localize in or were subcomponent of intracellular membrane-bounded organelle (plastid, mitochondria, vacuole, endoplasmic reticulum, Golgi apparatus, endosome, and nucleus). About 7% and 4% were part of plasma membrane and cell wall respectively.

Construction of the *N. benthamiana* microarray

A total of 13,014 unique sequences were submitted to NimbleGen to develop a custom gene expression microarray. With NimbleGen's proprietary method, the submitted 13,014 *N. benthamiana* unigenes assembled by CAP3 were reduced to 12,738 unigenes. Given that *N. benthamiana* is widely used in plant virology study and host-virus interaction study, we decided to incorporate plant orthologs of genes from animals and insects into our array design. These ortholog genes were chosen because they were previously found to interact with other RNA viruses. Including these orthologs on the array provides for wider range of potential interacting plant host factors. The source of our EST collection for *N. benthamiana* is illustrated in Figure 10. Nimblegen used ArrayScribe software package to arrange the probes on the array. The array was designed for 60-mer oligonucleotide probes format. The longer 60-mer probe gives a stronger signal, and higher sensitivity than the

shorter 24-mer probe [61]. Each unigene was represented by nine different probes, repeated three times, for a grand total of 362,205 probes representing the total of 12,738 *N. benthamiana* unigenes and 677 orthologs printed on a 385K array platform (Figure 10 and 11).

The process for creating the array, as described by Nimblegen, is as follows. All possible probes 24 base pairs in length are generated by “walking” the length of the sequence one base pair at a time. The probes were evaluated for uniqueness in two different ways. The first evaluation technique, a simple frequency count, involved counting the number of times each probe appears in the genome of interest. A more thorough, second way of evaluating the probes was a three step process. First, each probe was compared to the target genome and the probe was given a pass/fail score based on the level of uniqueness. The purpose of this scan is to avoid long runs of homopolymers or sequences that are overly GC or AT rich. Second, a self-annealing score for each oligonucleotide was calculated by comparing the oligonucleotide to its reverse complement. Any probe that was greater than 60% self-complementary was eliminated. Third, probes that require a large number of cycles to synthesize were eliminated. The scores from the evaluation were combined and rank scored. The highest scoring probe was selected and the rank scores are recalculated. Neighboring oligonucleotide scores were decreased as oligonucleotides were selected in an effort to pick evenly spaced probes.

Conclusion

We constructed an alternative platform for a transcriptome analysis in *N. benthamiana* using an EST collection, a unigene collection, and gene expression microarray technology. The design of the oligonucleotide probes printed on the microarray was based on EST analysis and unigene assembly. The array represented about 13,000 *N. benthamiana* unigenes derived from more than 40,000 ESTs. At the time of study, the GenBank EST database contained 56,102 sequences for *N. benthamiana* [57]. We believe that the coverage of our array is estimated to be approximately 70% of available *N. benthamiana* ESTs and 38% of the transcriptome. It is not surprising that the coverage of our unigene collection to the entire transcriptome is low because it must be noted that EST sequencing is subject to sampling bias, leading to under-representation of rare transcripts, often accounting for 60% of an organism's genes [14, 62]. The important factors contributing to this issue is the source of cDNA libraries used to obtain ESTs.

However, functional annotation indicates that our unigene collection represents a very wide-range of plant functions and pathways from primary/secondary metabolism, transportation, molecular function (transcription, translation), signal transduction/cellular communication, to plant defense systems, suggesting a good representation of the majority of important expressed genes in the *N. benthamiana* genome. The broad coverage of these annotated functions of the genes on our microarray affirms that our microarray can be used to study a variety of experimental challenges, including but not limited to host-pathogen interaction studies. Currently, our custom *N. benthamiana* microarray is commercially

available from Nimblegen. A use of the array and the information of our *N. benthamiana* unigenes, functional annotation, array design, and sequence are available through a request from the Lommel laboratory.

The validation and application of our custom microarray is thoroughly described in Chapter 3. We applied the array to study a modulation of *N. benthamiana* gene expression at the early stages of *Red clover necrotic mosaic virus* infection. We found that the microarray yields almost 100% hybridization capacity. We are able to identify significant differential gene expression and important host factors/pathways at sub 24 hours after inoculation. A study by Kim et al. (2011) has also shown that the microarray and functional annotation data were able to determine a change in gene expression level of *N. benthamiana* in a bacterial pathogen interaction (*Pectobacterium carotovorum*) [63]. We anticipate that the unigene collection and microarray presented in this work could ultimately contribute to the scientific community as an alternative high-throughput tool and genomic resource that can help enable the use of *N. benthamiana* as a plant model for genome wide studies.

Acknowledgement

We wish to thank Dr. Robin Buell (TIGR), Dr. Paul Goodwin (University of Guelph, Canada), and Dr. Andy Maule (John Innes Centre, UK) for their contribution of EST sequences. This work was supported by Philip Morris foundation.

Table 1 Sources and numbers of *N. benthamiana* EST sequences used for unigene assembly

Source	Contact	ESTs	Reference
Tobacco Genome Initiative (TGI), NC State University, Raleigh, USA	Dr. Steven Lommel (steve_lommel@ncsu.edu)	12,000	[58]
John Innes Centre, Norwich, UK	Dr. Andy Maule (andy.maule@bbsrc.ac.uk)	6,700	[64]
University of Guelph, Guelph, Canada	Dr. Paul Goodwin	1,000	[65]
The Institute for Genomic Research (TIGR)	Dr. Robin Buell	18,000	[59]
GenBank		4,000	[57]
Total		41,700	

Table 2 Potential highly-expressed *N. benthamiana* genes from standard, normalized, and subtracted EST libraries

Rank	Top Hit Description	Accession number	No. of clones	% frequency	E-value
1	Ribulose biphosphate carboxylase [<i>N. plumbaginifolia</i>]	AAA34110.1	702	18%	2.00E-95
2	Ribulose biphosphate carboxylase small chain 8B [<i>N. plumbaginifolia</i>]	P26573	660	17%	9.00E-96
3	Chlorophyll a/b-binding protein type I precursor [<i>S. lycopersicum</i>]	S16294	120	3%	1.00E-82
4	Light harvesting chlorophyll a/b-binding protein [<i>N. sylvestris</i>]	BAA25392.1	107	2.8%	1.00E-118
5	Light harvesting chlorophyll a/b-binding protein [<i>N. sylvestris</i>]	BAA25389.1	105	2.8%	1.00E-125
6	Light harvesting chlorophyll a/b-binding protein [<i>N. sylvestris</i>]	BAA25391.1	105	2.8%	1.00E-121
7	Photosystem I reaction center subunit II [<i>N. sylvestris</i>]	P29302	83	2%	4.00E-99
8	Metallothionein-like protein type 2 [<i>N. tabacum</i>]	CAC12823.1	79	2%	2.00E-26
9	Glyceraldehyde 3-phosphate dehydrogenase A [<i>N. tabacum</i>]	P09043	68	1.7%	1.00E-126
10	Cab50 protein precursor (chlorophyll A-B binding protein) [<i>N. tabacum</i>]	CAA36956.1	67	1.7%	1.00E-113
11	Ribulose biphosphate carboxylase small chain [<i>N. tabacum</i>]	P00866	66	1.7%	6.00E-94

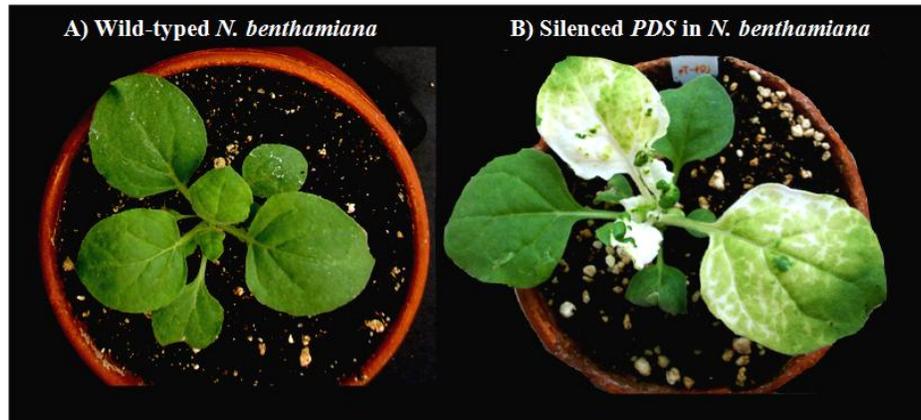


Figure 1 Transient silenced *PDS* (phytoene desaturase) gene in *N. benthamiana* using TRV-VIGS system

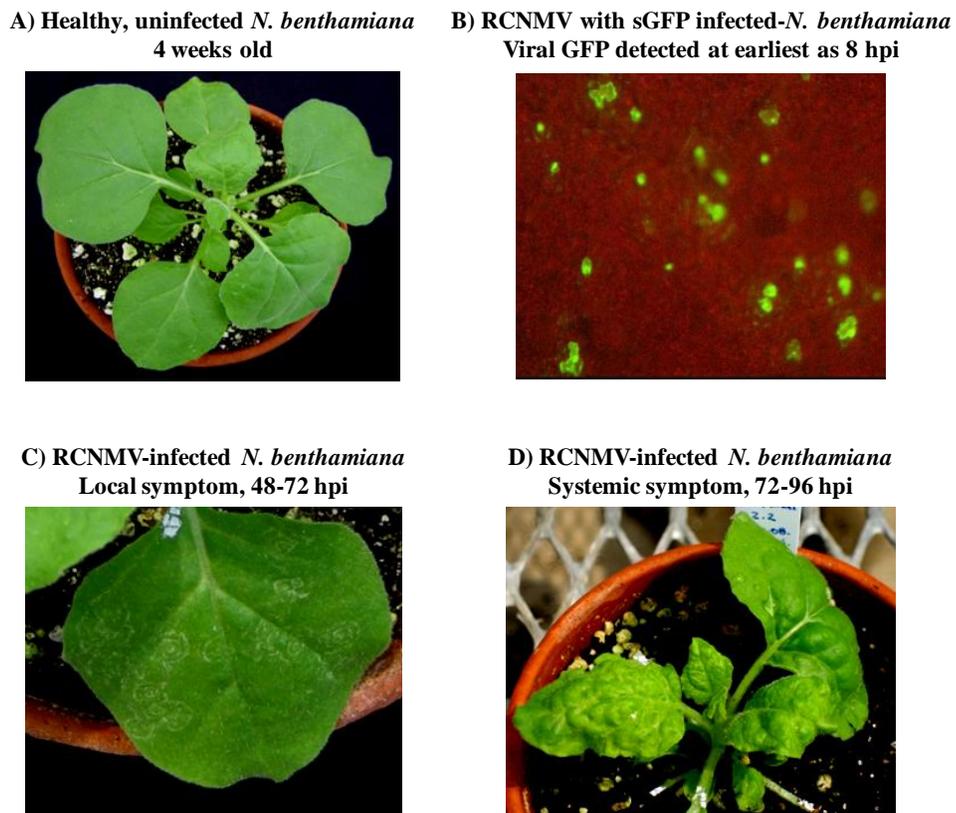


Figure 2 *N. benthamiana* healthy and RCNMV-infected plants

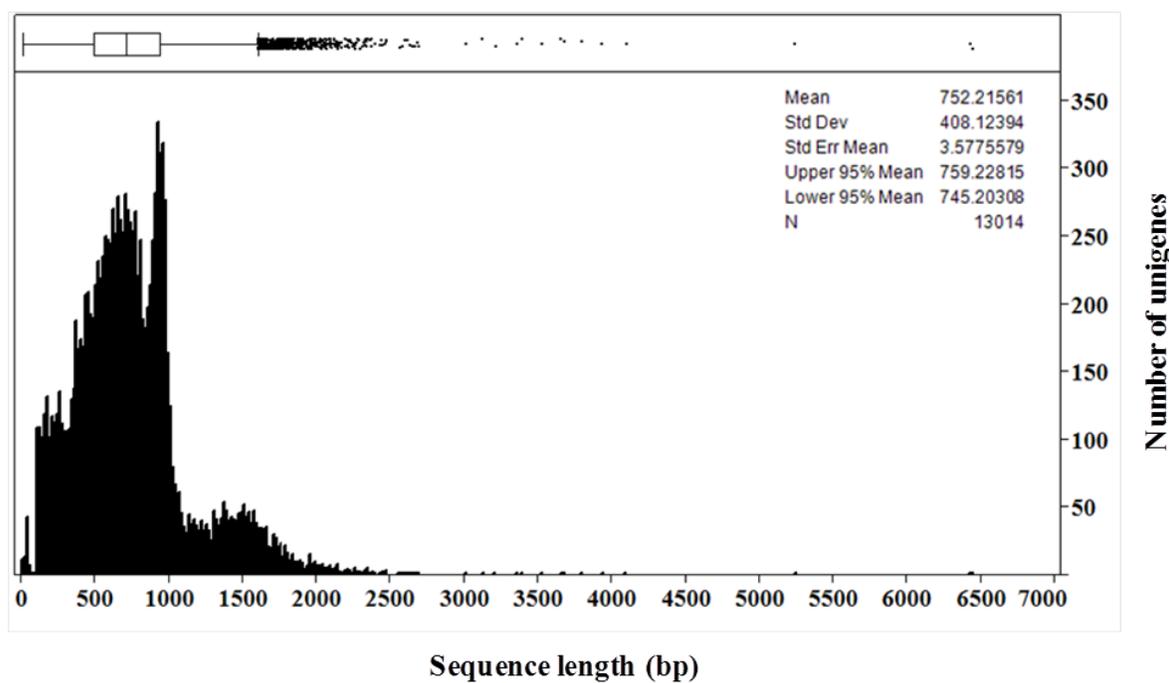


Figure 3 Distribution of *N. benthamiana* unigene length

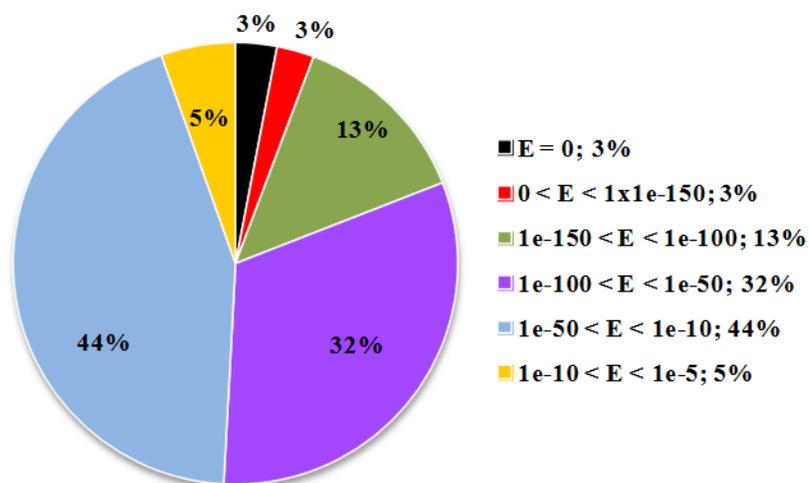


Figure 4 Distribution of *E*-values of top BLAST hit *N. benthamiana* unigenes searched against GenBank nr database

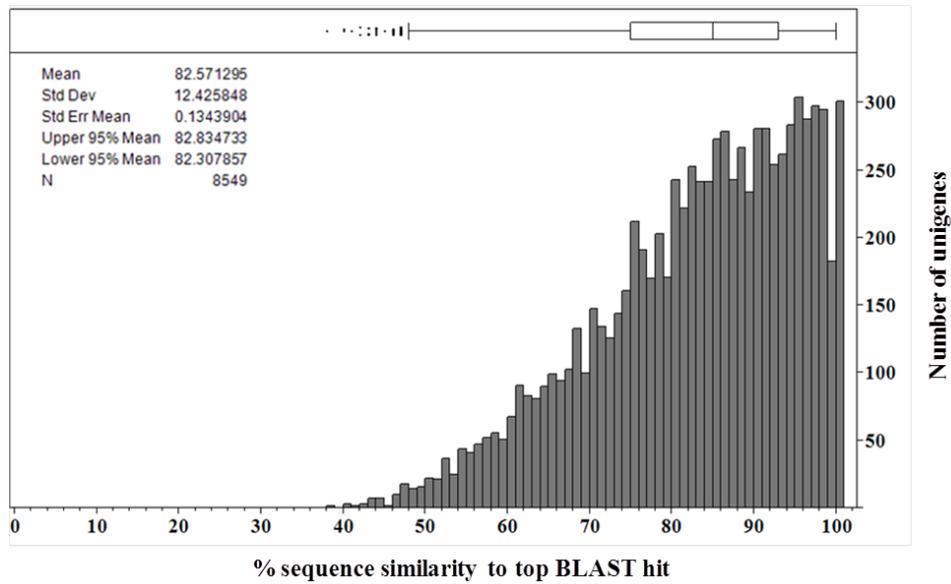


Figure 5 Distribution of percent sequence similarity of top BLAST hit *N. benthamiana* unigenes searched against GenBank nr database

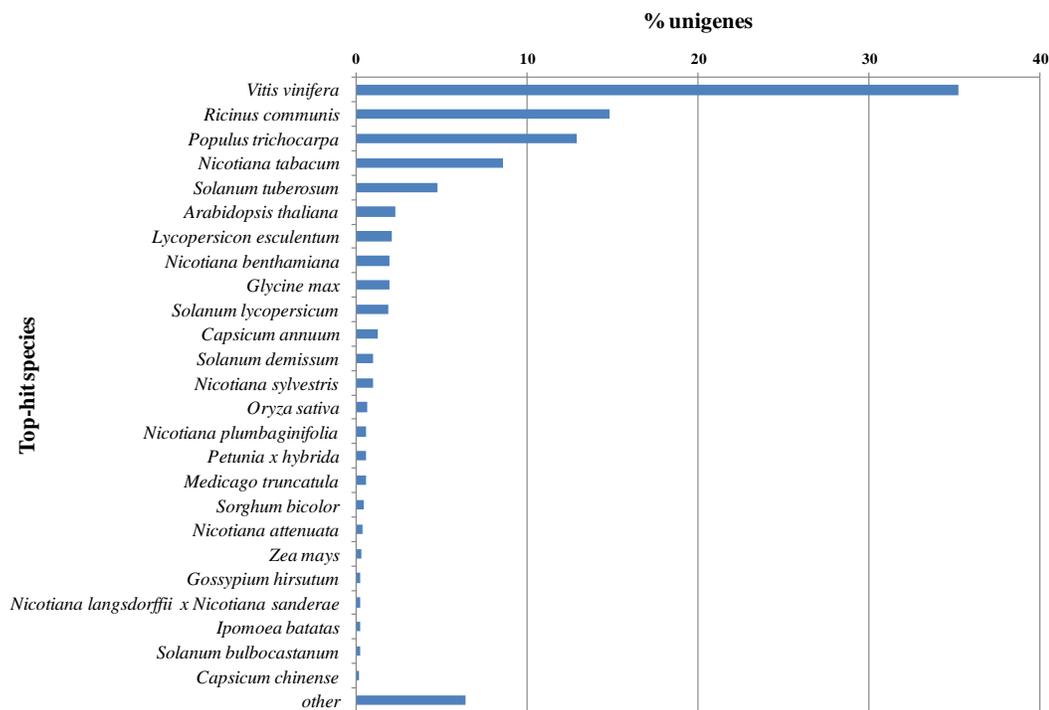


Figure 6 Distribution of organism species of top BLAST hit *N. benthamiana* unigenes searched against GenBank nr database

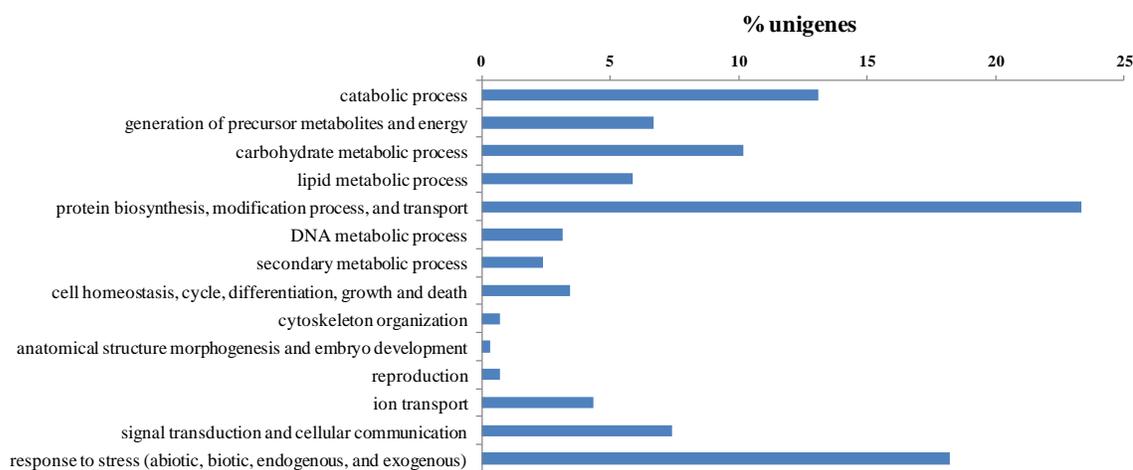


Figure 7 GO distribution for “biological process” category assigned to *N. benthamiana* unigenes

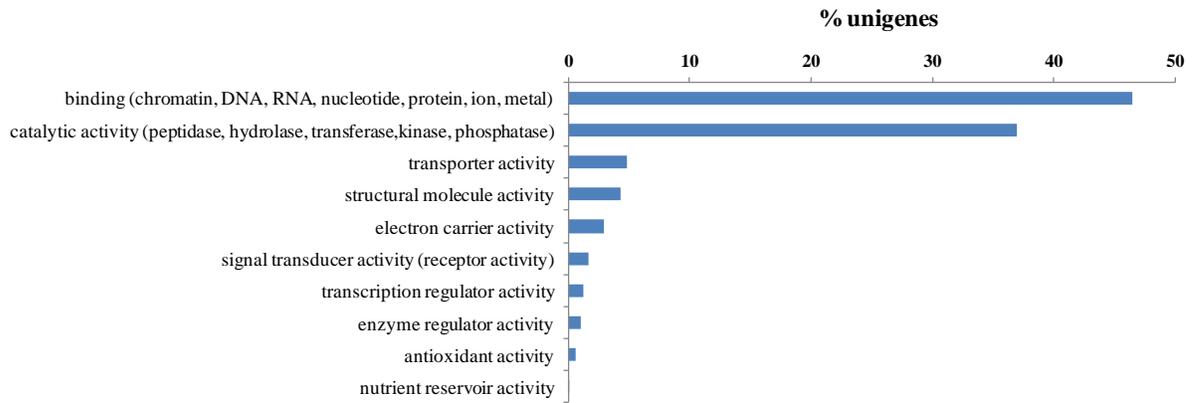


Figure 8 GO distribution for “molecular function” category assigned to *N. benthamiana* unigenes

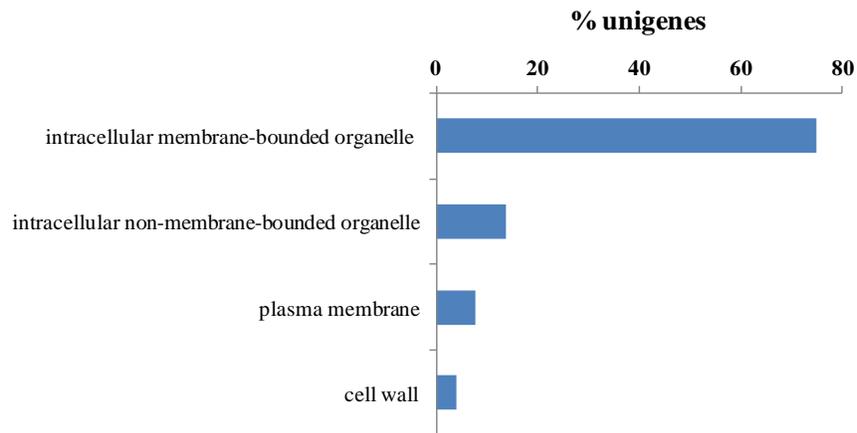


Figure 9 GO distribution for “cellular compartment” category assigned to *N. benthamiana* unigenes

The sources of ESTs collection for *N. benthamiana* microarray

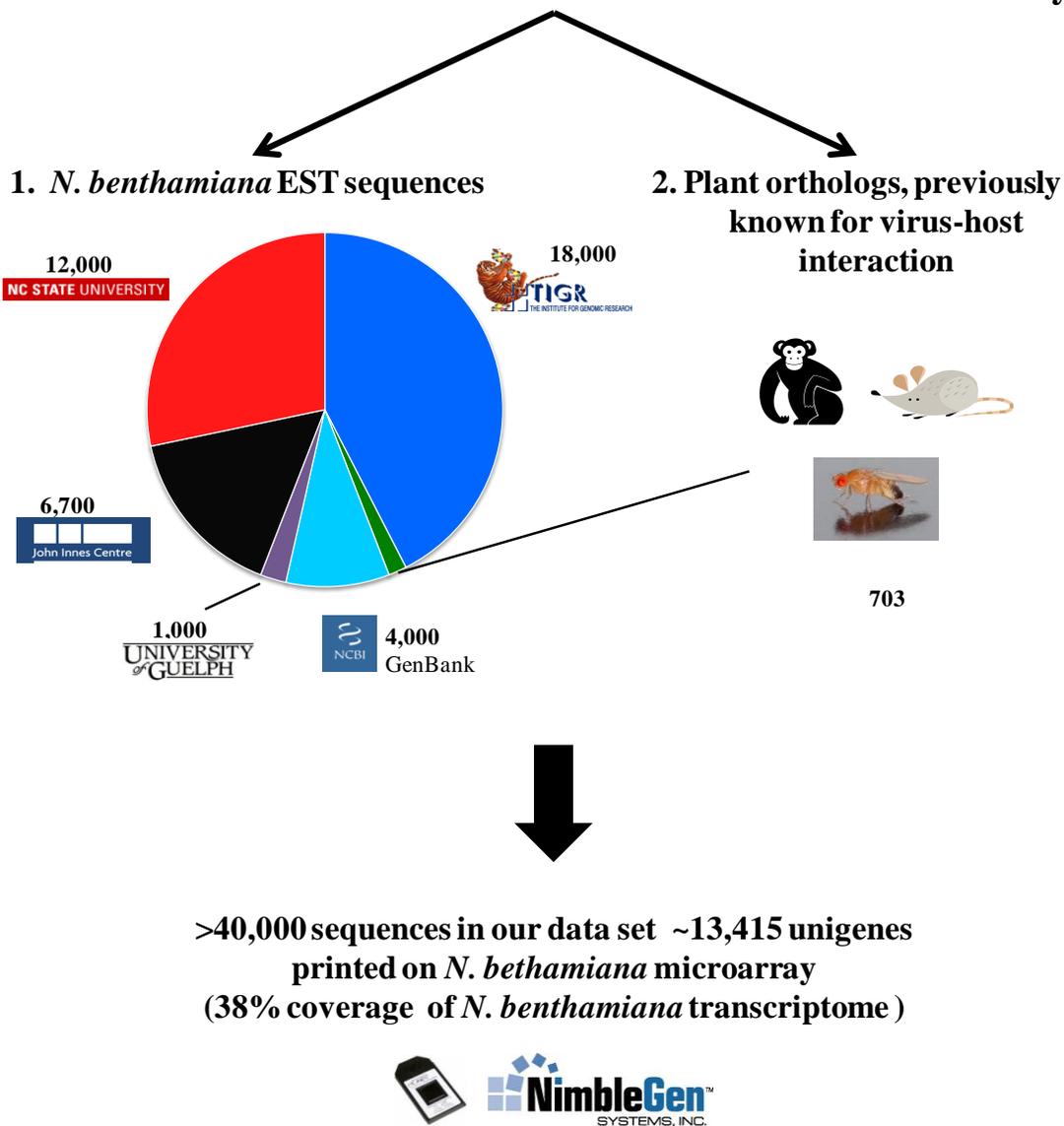


Figure 10 The sources of expressed sequence tag (EST) collection for creating a custom *N. benthamiana* microarray

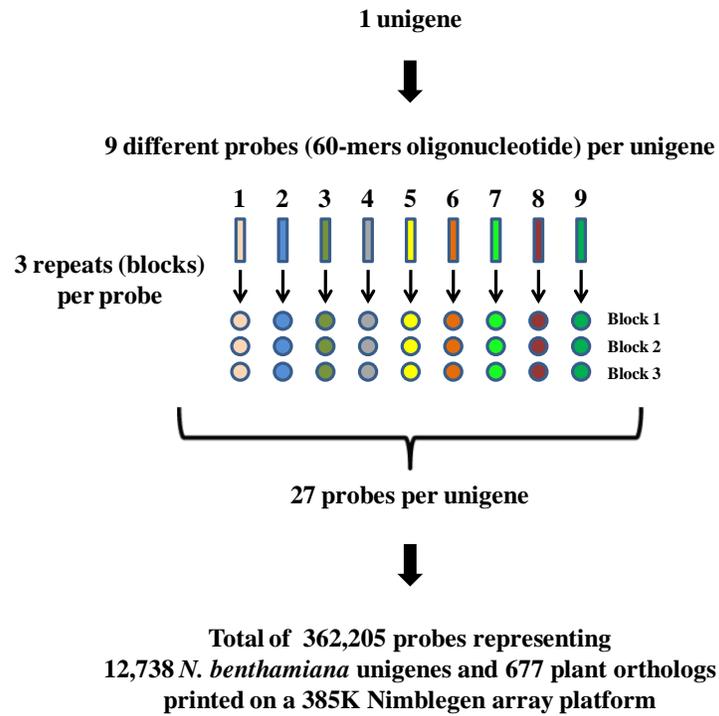


Figure 11 A layout design of a custom *N. benthamiana* microarray

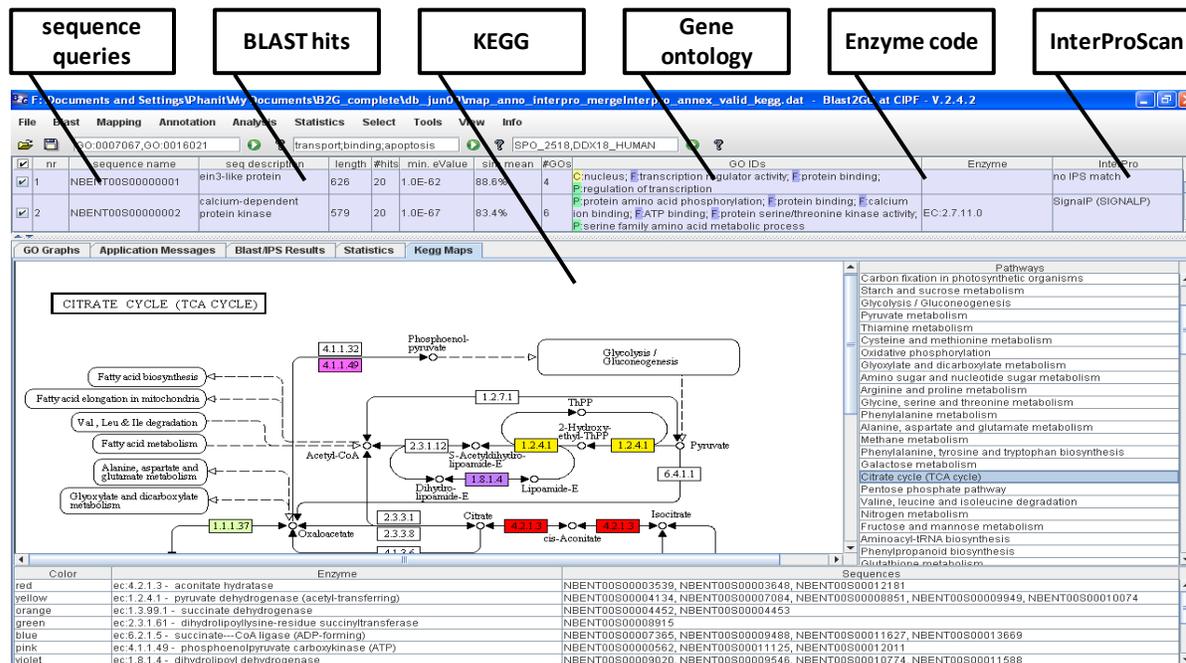


Figure 12 Blast2GO graphical user interphase (GUI)

References

1. Goodwin MM, Zaitlin D, Naidu RA, Lommel SA: *Nicotiana benthamiana: its history and future as a model for plant-pathogen interactions*. *Molecular plant-microbe interactions* 2008, **21**(8):1015-1026.
2. Clemente T: *Nicotiana (Nicotiana tabacum, Nicotiana benthamiana)*. In: *Agrobacterium Protocols*. Springer; 2006: 143-154.
3. Goodspeed TH, Wheeler H, Hutchison P: *The genus Nicotiana: origins, relationships and evolution of its species in the light of their distribution, morphology and cytogenetics*, vol. 16: Chronica Botanica Company Waltham; 1954.
4. Senthil-Kumar M, Mysore KS: *New dimensions for VIGS in plant functional genomics*. *Trends in plant science*, **16**(12):656-665.
5. Wydro M, Kozubek E, Lehmann Pa: *Optimization of transient Agrobacterium-mediated gene expression system in leaves of Nicotiana benthamiana*. *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION*- 2006, **53**(2):289.

6. Gleba Y, Klimyuk V, Marillonnet S: **Magniffection-a new platform for expressing recombinant vaccines in plants.** *Vaccine* 2005, **23**(17):2042-2048.
7. Matoba N, Davis KR, Palmer KE: **Recombinant protein expression in *Nicotiana*.** In: *Plant Chromosome Engineering*. Springer: 199-219.
8. Strasser R, Stadlmann J, Schahs M, Stiegler G, Quendler H, Mach L, Glossl J, Weterings K, Pabst M, Steinkellner H: **Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure.** *Plant biotechnology journal* 2008, **6**(4):392-402.
9. Dasgupta R, Garcia BH, Goodman RM: **Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes.** *Proceedings of the National Academy of Sciences* 2001, **98**(9):4910-4915.
10. Bennett MD, Leitch IJ: **Nuclear DNA amounts in angiosperms.** *Annals of Botany* 1995, **76**(2):113-176.
11. Bennett MD, Leitch IJ, Price HJ, Johnston JS: **Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* genome initiative estimate of ~125 Mb.** *Annals of Botany* 2003, **91**(5):547-557.
12. Bennett M, Leitch I: **Nuclear DNA amounts in angiosperms: progress, problems and prospects.** *Annals of Botany* 2005, **95**(1):45-90.
13. Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB: **A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research.** *Molecular plant-microbe interactions* 2012, **25**(12):1523-1530.
14. Nagaraj SH, Gasser RB, Ranganathan S: **A hitchhiker's guide to expressed sequence tag (EST) analysis.** *Briefings in bioinformatics* 2007, **8**(1):6-21.
15. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF: **Complementary DNA sequencing: expressed sequence tags and human genome project.** *Science* 1991, **252**(5013):1651-1656.
16. Putney SD, Herlihy WC, Schimmel P: **A new troponin T and cDNA clones for 13 different muscle proteins, found by shotgun sequencing.** *Nature* 1983, **302**(5910):718-721.
17. Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R: **Gene discovery in the wood-forming tissues**

- of poplar: analysis of 5,692 expressed sequence tags.** *Proceedings of the National Academy of Sciences* 1998, **95**(22):13330-13335.
18. Hillier L, Lennon G, Becker M, Bonaldo MF, Chiapelli B, Chissoe S, Dietrich N, DuBuque T, Favello A, Gish W: **Generation and analysis of 280,000 human expressed sequence tags.** *Genome research* 1996, **6**(9):807-828.
 19. Marra M, Hillier L, Kucaba T, Allen M, Barstead R, Beck C, Blistain A, Bonaldo M, Bowers Y, Bowles L: **An encyclopedia of mouse genes.** *Nature genetics* 1999, **21**(2):191-194.
 20. Jung J, Park H-W, Hahn Y, Hur C-G, In D, Chung H-J, Liu J, Choi D-W: **Discovery of genes for ginsenoside biosynthesis by analysis of ginseng expressed sequence tags.** *Plant cell reports* 2003, **22**(3):224-230.
 21. White JA, Todd J, Newman T, Focks N, Girke T, de IlÃ¡rduya OMÌn, Jaworski JG, Ohlrogge JB, Benning C: **A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil.** *Plant Physiology* 2000, **124**(4):1582-1594.
 22. Ohlrogge J, Benning C: **Unraveling plant metabolism by EST analysis.** *Current opinion in plant biology* 2000, **3**(3):224-228.
 23. Adams MD, Kerlavage AR, Fleischmann RD, Fuldner RA, Bult CJ, Lee NH, Kirkness EF, Weinstock KG, Gocayne JD, White O: **Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence.** *Nature* 1995, **377**(6547 Suppl):3-174.
 24. Kloosterman B, De Koeijer D, Griffiths R, Flinn B, Steuernagel B, Scholz U, Sonnewald S, Sonnewald U, Bryan GJ, Prat S: **Genes driving potato tuber initiation and growth: identification based on transcriptional changes using the POCI array.** *Functional & integrative genomics* 2008, **8**(4):329-340.
 25. Tecsi LI, Smith AM, Maule AJ, Leegood RC: **A spatial analysis of physiological changes associated with infection of cotyledons of marrow plants with cucumber mosaic virus.** *Plant Physiology* 1996, **111**(4):975-985.
 26. Havelda Z, Maule AJ: **Complex spatial responses to cucumber mosaic virus infection in susceptible Cucurbita pepo cotyledons.** *The Plant Cell Online* 2000, **12**(10):1975-1985.
 27. Lokko Y, Anderson J, Rudd S, Raji A, Horvath D, Mikel M, Kim R, Liu L, Hernandez A, Dixon A: **Characterization of an 18,166 EST dataset for cassava (Manihot esculenta Crantz) enriched for drought-responsive genes.** *Plant cell reports* 2007, **26**(9):1605-1618.

28. Wang Z-l, Li P-h, Fredricksen M, Gong Z-z, Kim C, Zhang C, Bohnert HJ, Zhu J-K, Bressan RA, Hasegawa PM: **Expressed sequence tags from *Thellungiella halophila*, a new model to study plant salt-tolerance.** *Plant Science* 2004, **166**(3):609-616.
29. Manger ID, Hehl A, Parmley S, Sibley LD, Marra M, Hillier L, Waterston R, Boothroyd JC: **Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes.** *Infection and immunity* 1998, **66**(4):1632-1637.
30. Tanabe K, Nakagomi S, Kiryu-Seo S, Namikawa K, Imai Y, Ochi T, Tohyama M, Kiyama H: **Expressed-sequence-tag approach to identify differentially expressed genes following peripheral nerve axotomy.** *Molecular brain research* 1999, **64**(1):34-40.
31. Ewing RM, Kahla AB, Poirot O, Lopez F, Audic Sp, Claverie J-M: **Large-scale statistical analyses of rice ESTs reveal correlated patterns of gene expression.** *Genome research* 1999, **9**(10):950-959.
32. Wilcox AS, Khan AS, Hopkins JA, Sikela JM: **Use of 3'untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome.** *Nucleic acids research* 1991, **19**(8):1837-1843.
33. Jongeneel CV: **Searching the expressed sequence tag (EST) databases: panning for genes.** *Briefings in bioinformatics* 2000, **1**(1):76-92.
34. Dong Q, Kroiss L, Oakley FD, Wang B-B, Brendel V: **Comparative EST analyses in plant systems.** *Methods in enzymology* 2005, **395**:400-418.
35. Rudd S: **Expressed sequence tags: alternative or complement to whole genome sequences?** *Trends in plant science* 2003, **8**(7):321-329.
36. Eyraas E, Caccamo M, Curwen V, Clamp M: **ESTGenes: alternative splicing from ESTs in Ensembl.** *Genome research* 2004, **14**(5):976-987.
37. Nelson RT, Grant D, Shoemaker RC: **ESTminer: a suite of programs for gene and allele identification.** *Bioinformatics* 2005, **21**(5):691-693.
38. Usuka J, Zhu W, Brendel V: **Optimal spliced alignment of homologous cDNA to a genomic DNA template.** *Bioinformatics* 2000, **16**(3):203-211.
39. Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Muller WE, Wetter T, Suhai S: **Using the miraEST assembler for reliable and automated mRNA transcript**

- assembly and SNP detection in sequenced ESTs.** *Genome research* 2004, **14**(6):1147-1159.
40. Krause A, Haas SA, Coward E, Vingron M: **SYSTEMS, GeneNest, SpliceNest: exploring sequence space from genome to protein.** *Nucleic acids research* 2002, **30**(1):299-300.
 41. Delseny M, Cooke R, Raynal M, Grellet F: **The *Arabidopsis thaliana* cDNA sequencing projects.** *FEBS letters* 1997, **403**(3):221-224.
 42. Shoemaker R, Keim P, Vodkin L, Retzel E, Clifton SW, Waterston R, Smoller D, Coryell V, Khanna A, Erpelding J: **A compilation of soybean ESTs: generation and analysis.** *Genome* 2002, **45**(2):329-338.
 43. Sterky F, Bhalerao RR, Unneberg P, Segerman B, Nilsson P, Brunner AM, Charbonnel-Campaa L, Lindvall JJ, Tandre K, Strauss SH: **A *Populus* EST resource for plant functional genomics.** *Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**(38):13951-13956.
 44. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**(5235):467-470.
 45. Senthil G, Liu H, Puram V, Clark A, Stromberg A, Goodin M: **Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses.** *Journal of General Virology* 2005, **86**(9):2615-2625.
 46. Dardick C: **Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses.** *Molecular plant-microbe interactions* 2007, **20**(8):1004-1017.
 47. D'arcy W: **The classification of the Solanaceae.** Hawkes, J, G., Lester, R, N., Skelding, A, D ed (s) *The biology and taxonomy of the Solanaceae* London, Academic Press for the Linnean Society 1979:3-47.
 48. Borisjuk N, Borisjuk L, Petjuch G, Hemleben V: **Comparison of nuclear ribosomal RNA genes among *Solanum* species and other Solanaceae.** *Genome* 1994, **37**(2):271-279.
 49. Smit AF, Green P: **RepeatMasker.** 1996.
 50. Tarailo-Graovac M, Chen N: **Using RepeatMasker to identify repetitive elements in genomic sequences.** *Current Protocols in Bioinformatics* 2009:4.10. 11-14.10. 14.

51. Huang X, Madan A: **CAP3: A DNA sequence assembly program.** *Genome research* 1999, **9**(9):868-877.
52. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674-3676.
53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
54. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**(1):25-29.
55. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28**(1):27-30.
56. Zdobnov EM, Apweiler R: **InterProScan- an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**(9):847-848.
57. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2008, **36**(suppl 1):D25-D30.
58. Opperman CH, Lommel SA, Burke M, Feulner G, Carlson J, George C, Gove S, Houfek TD, Jefferys SR, Kalat S *et al*: **Tobacco Genome Initiative (TGI) *Nicotiana benthamiana* ESTs, Lib name:LIBEST_015697 SAL_US N.Bentheme.** *GenBank EST database* 2004.
59. Buell CR, Hart A, Zismann V, Karamycheva SA, Day B, Staskawicz B, Jin H, Baker B: **Generation of EST sequences from *Nicotiana benthamiana*, Lib Name: LIBEST_015055 *Nicotiana benthamiana* mixed tissue cDNA library, normalized, full-length.** *GenBank EST database* 2003.
60. Dondoshansky I, Wolf Y: **Blastclust (ncbi software development toolkit).** *NCBI, Bethesda, Md* 2002.
61. Buck MJ, Lieb JD: **ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments.** *Genomics* 2004, **83**(3):349-360.
62. Bonaldo MdF, Lennon G, Soares MB: **Normalization and subtraction: two approaches to facilitate gene discovery.** *Genome research* 1996, **6**(9):791-806.
63. Kim H-S, Thammarat P, Lommel SA, Hogan CS, Charkowski AO: ***Pectobacterium carotovorum* elicits plant cell death with DspE/F but the *P. carotovorum* DspE**

- does not suppress callose or induce expression of plant genes early in plant-microbe interactions.** *Molecular plant-microbe interactions* 2011, **24**(7):773-786.
64. Debout M, Bennett G, Williams A, Page A, Angell S, Clarke J, Maule A: ***Nicotiana benthamiana* ESTs, Lib name: LIBEST_020736 j001.** *GenBank EST database* 2006.
65. Shan XC, Goodwin PH: **Identification of a SAR8. 2 gene in the susceptible host response of *Nicotiana benthamiana* to *Colletotrichum orbiculare*.** *Functional plant biology* 2005, **32**(3):259-266.

Chapter 3

***N. benthamiana* genome-wide screen reveals early changes in host gene expression patterns in response to *Red clover necrotic mosaic virus* infection**

Chapter summary

In this chapter, I report the first genome-wide analysis of the modulation of *N. benthamiana* gene expression early in the RCNMV infection process. This is also the first host transcriptome profile studied as early as 2 hours after RCNMV infection. Key host genes, functions, and pathways that contribute to the RCNMV infection cycle are successfully identified. I also describe how host physiological processes as well as host immune system changes correspond to a particular stage of RCNMV infection (2, 6, 12, and 24 hpi). The microarray data of host-RCNMV interaction dissected in this chapter was used as preliminary information for the subsequent functional studies of host genes that are described in the chapters that follow.

Abstract

Plant viruses are minimalist pathogens that are required to co-opt host factors during the early period to produce a successful infection. Many events and interactions during this critical early period where the virus recruits host factors and reprograms plant physiological processes remain unknown. We utilized *Nicotiana benthamiana* (a model host for most plant viruses) and *Red clover necrotic mosaic virus* (RCNMV, a typical plus strand RNA plant virus) to elucidate these early events in the virus-host interaction. A custom microarray comprising 362,205 probes representing a total of 13 K *N. benthamiana* unigenes (equivalent to an estimated coverage of ~38% of the transcriptome) was developed for this study. Host transcriptome profiles were analyzed at four different time points of 2, 6, 12 and 24 hours post-inoculation (hpi). Microarray data was subjected to statistical analysis and 1,654 genes exhibited differential expression at FDR cutoff of 0.01. Selected genes displaying significant changes in expression levels from different metabolic pathways were validated by qRT-PCR analysis. With the use of a different set of plant samples for the microarray validation, we screened 60 genes and found that the expressions of 16 genes were significantly modulated and agreed with the microarray data. The global snapshot of microarray genes revealed that a majority of host genes was significantly down-regulated at 2, 6 and 24 hpi and significantly up-regulated at 12 hpi. This implies that host gene expression was shutdown early within 6 hours of RCNMV infection but then was recovered at 12 hours. And by 24 hours, the same host expression pattern reappeared again. This finding leads to a conclusion that the complete infection cycle within the primary infected cell and the viral movement to adjacent

cells takes place at the critical time point between 12 and 24 hours. Host systemic acquired resistance (NbSAR8.2a) was significantly modulated across all four time points of the study. Host SAR gene expression was down-regulated in the first 6 hours, but followed by an up-regulation after 12 hours. The regulation switch suggests an allowance of RCNMV to presumably establish its primary infection. The hypothesis is that a down regulation of host SAR gene expression in the first 6 hours is driven by RCNMV, but once an initial infection is successful, an increasing host SAR gene expression was driven by a response from host to defend itself from further viral intrusion to other parts of the plant (systemic infection). Gene set enrichment analysis (GSEA) revealed that RCNMV infection affected the following key host functions/pathways: defense, translation, photosynthesis and chloroplast-related functions, carbon fixation (Calvin cycle), metabolism, multidrug and toxin extrusion (MATE) transporter, cell wall-associated functions, and protein kinases. GSEA revealed that host defense functions were suppressed in the first 6 hours of RCNMV infection. During this host defense functions suppression, the host apoptosis pathway appeared to be activated as a presumable result of a hypersensitive reaction in order to confine a primal infected area and so prevent further viral invasion. We conclude that host apoptosis is an immediate defense response of *N. benthamiana* against RCNMV infection as early as 2 hours; whereas, host defense gene expression is a delayed defense response. Host functions involved in chloroplast and Calvin cycle were significantly correlated to RCNMV infection across all stages of the infection cycle. A delay in Calvin cycle in the first 6 hours of RCNMV infection and a suppression of chloroplast-related functions until 12 hours indicates that RCNMV prevented host from being able to perform carbon fixation, and as a result, RCNMV

indirectly decreased plant biosyntheses of organic compounds, importantly carbohydrate biosyntheses.

Keywords

Plant-virus interaction, virus-host interaction, microarray, *Nicotiana*, RCNMV, Red clover necrotic mosaic virus, plant virus

Introduction

How viruses modulate host physiological processes upon cells, tissues and whole plants to reconstruct host environment to favor their infectivity while defeating plant host immune system remains one of the gaps in our knowledge of plant virology. Virus particles (known as virions) basically consist of two components; (i) genetic material in the form of either RNA or DNA, and (ii) coat proteins or capsid proteins that protect viral genetic material. Most plant viruses encode only 3-4 viral proteins due to their limited genetic makeup. These viral proteins are mostly viral replication proteins, coat protein, and movement protein [1, 2]. Due to a lack of other proteins or factors necessary for biological processes, importantly for viral translation, replication and movement, plant viruses absolutely rely on utilizing host factors/machineries to produce a successful infection [3-7]. However, plants are known to fight against invading viruses by using two different defense strategies; innate and adaptive immunity. A thorough review of the plant immune system is described in Chapter 1. Plant innate immunity involves plant resistance (*R*) genes, which encode specific resistance proteins that either directly recognize specific viral proteins (also known as *Avr* proteins), or indirectly recognize a modulation of host physiological processes

by viruses. In either case, this recognition leads to a hypersensitive reaction followed by an activation of host programmed cell death and host defense gene expression, resulting in an inhibition of viral infection in primary infected cell and a prevention of viral spreading to other parts of the plant [8]. Plant adaptive immunity involves a defense mechanism known as RNA silencing that tightly relies on the recognition of viral double-stranded RNA (dsRNA) generated during the viral replication process, leading to specific targeting of viral RNA for degradation [9-11].

The field of plant defense strategies directed against viruses has been studied extensively over the past decades and has led to the observation that viruses have evolved a special counter defense repertoire that allows them to escape from host defense responses long enough to allow a productive infection. However, it is important to note that virus-host interaction requires balance for successful infections as illustrated by tobacco mosaic virus (TMV). TMV suppresses RNA silencing via one of its replication proteins, the 126-kDa protein [12]; however, its movement proteins antagonize the activity of this 126-kDa protein, resulting in promoting RNA silencing and so a reduction of viral infection [13]. This counter interact is preventing fatal effect to host while viruses are still able to maintain their infectivity.

After the initial infection event, viruses remain within cytoplasm of primary infected cells. Neighboring host cells are infected only after virus progenies passage through plasmodesmata (intercellular channel) [14, 15]. The outcome of the local infection event and the progressive spread of the virus to other parts of the plant is the appearance of physical

symptoms, either local symptoms or systemic symptoms; for instance, leaf yellowing, chlorosis, leaf curling, cell/tissue necrotic pattern, ring spot, and growth stunting [16]. These physical symptoms represent the sum of the impact of host defense response and host physiological processes changed at the cellular level. The causes that drive the changes in host physiology and structure during virus attack comes from two sources; first is a host-driven change and second is a virus-driven change. However, these two driving forces are difficult to distinguish. The assumption is that the modulation that is driven by the host is mostly a trigger of host immunity system for host survival. Whereas, the modulation that is driven by virus is an attempt to manipulate host pathways and outcompete host factors over the host's own substrates in order to exploit host cellular resources for viral survival and reproduction.

Virus movement to adjacent cells is observed within a few hours after first infection. In our study, we found that virus takes between 12 and 24 hours to complete the infection process in a primary infected cell. Since one infection cycle takes as short period as 2 to 24 hours, dissecting the biochemical impact of virus replication at the host cellular level during this time period is very critical in order to identify key host factors responsible for each stage of infection. The earliest time point study of host-virus interaction is mostly after 24 hours, which by that time, the virus has already completed an initial infection in primary infected cells and then moved to adjacent cells to initiate another infection cycle; therefore, the changes in host physiological processes no longer reflect their involvement at each stage of infection due to asynchronous infection among infected cells. The main hypothesis of our study is that the changes in host gene expression at sub 24 hours-time frame are the results of

a particular stage of virus infection cycle. We carefully planned out a time course experiment to capture all the early events that happen to the host at 2, 6, 12, and 24 hours of infection.

We are particularly interested in determining how host physiological processes and host immune system changes correspond to each stage of the virus infection cycle. However, we did not directly measure host biochemical changes. Instead, we measured the host transcriptome profile and therefore, this entire study is based on the assumption that host transcriptome changes are the reflection of host physiological process changes. The study in this chapter is one of the few first reports of host-virus interaction at the earliest time point of infection at 2 hours and is also the first study that uses genome-wide analysis to study host gene changes at sub 24 hours time frame of virus infection. We used *Red clover necrotic mosaic virus* (RCNMV) and *Nicotiana benthamiana* as our virus-host interaction model system. Our main objective is to identify key host factors/ functions/ pathways that are essentially modulated by a particular stage of the virus infection cycle. Rather studying one host gene at a time, we utilized a microarray technology to enable us to study many thousands of host genes simultaneously. The microarray chip design was developed in our lab based on 13,000 *N. benthamiana* unigenes, covering about 38% of the transcriptome. Our microarray is the first *N. benthamiana* microarray and it is also the only one that is commercially available through a request from us and the Nimblegen Company. A thorough detail of the creation of the *N. benthamiana* microarray is described in Chapter 2.

N. benthamiana is a dicotyledon plant, a close relative to tobacco, and a species of genus *Nicotiana* in the *Solanaceae* family. Its habitat is indigenous to Australia [17] and it

originally arose from the hybridization between the two progenitor species of *N. debneyi* and *N. suaveolens* [18]. The first *N. benthamiana* genome draft was recently published in 2012. This *N. benthamiana* genome draft is a 63-fold coverage and is available on Sol Genomic Network for BLAST search and for downloading to local server [19]. The *N. benthamiana* genome (~ 3 Gbp) is nearly 20 fold larger than the *Arabidopsis thaliana* genome (~ 135 Mbp) [20-22]. *N. benthamiana* is an amphiploid species with $n = 19$; whereas, *A. thaliana* is a diploid species with $n=5$. *N. benthamiana* is currently becoming more popular in scientific communities around the world as a prominent model organism for conducting research in the field of plant biology and plant pathology. It is an excellent experimental plant subject and the best known model plant for performing agroinfiltration and virus-induced gene silencing (VIGS) to conduct plant functional assays, either transient gene knockdown studies or transient protein expression studies [23, 24]. It is also an efficient and a viable alternative platform for the industrial production of recombinant proteins, particularly in the field of pharmaceutical biotechnology [25-27]. *N. benthamiana* is susceptible to a wide range of plant pathogens (bacteria, oomycetes, fungi and viruses), making this plant species a cornerstone for plant pathogen study and host-pathogen interaction study in the particular context of host immunity and defense signaling [17]. By a comparison to *Arabidopsis*, *N. benthamiana* is susceptible to a broader range of pathogens. *N. benthamiana* is the most widely used host in plant virology studies due mainly to its susceptibility to > 500 plant viruses, even to those that are typically restricted to monocot hosts and, most surprisingly, even to several animal viruses [28]. *N. benthamiana* is also highly susceptible to RCNMV and capable of causing a systemic infection (Figure 1). The local lesions of RCNMV-infected

N. benthamiana appear 2-3 days after inoculation as the necrotic ring spots on inoculated leaves. And RCNMV systemic symptoms appear 3-4 days after inoculation, at which time the virus has spread from the inoculated leaves (primary infected leaves) through the plant vascular system to distal leaves, and on which the systemic symptoms develop as leaf curling, growth stunting, and a characteristic mosaic. Altogether, *N. benthamiana* represents a suitable model host for our study. We speculate that the identified key host factors from *N. benthamiana* interacting with RCNMV in our study may shed light on understanding why this plant species possesses the unique trait of being susceptible to such a vast number of diverse plant viruses.

RCNMV is a plant virus in the genus *Dianthovirus*, and family *Tombusviridae*. RCNMV has a bipartite genome structure consisting of two positive-sense single-stranded RNAs (Figure 2): RNA-1 (3.9 kb) and RNA-2 (1.45 kb), which are packaged together into virions composed of 180 identical capsid protein subunits [29-32]. RNA-1 contains three open reading frames (ORF): (i) p27 encoding a 27-kDa functionally unknown protein, (ii) p88 (expressed as a -1 ribosomal frameshift of p27) encoding the 88-kDa viral polymerase protein (contains an RNA-dependent RNA polymerase motif (RdRp)) [33-35], and (iii) p37 (expressed from subgenomic RNA) encoding the 37-kDa viral capsid protein [36]. Whereas RNA-2 is a monocistronic RNA (one ORF) which encodes the 35-kDa viral movement protein required for virus cell-to-cell movement (local infection) in plants and infection of whole plants (systemic infection) [31, 37]. Although RNA-2 is not required for the replication of RNA-1 in protoplasts [38-40], a 34-nucleotide (nt) sequence in RNA-2 is required for transcription of subgenomic RNA from RNA-1 [41]. The addition of (i) a

modified GTP (guanosine triphosphate) or any cap analogue (m^7GpppA , m^7GpppG) to the 5' end (cap), and (ii) poly (A) (adenine nucleotide) to 3' end (tail) of the premature-mRNA is the important canonical feature in the translation of many eukaryotic mRNAs [42]. However, the uncapped *in vitro* transcripts of RCNMV RNA-1 and RNA-2 show a comparable infectivity to that of capped transcripts in *N. benthamiana* [43] and in cowpea protoplasts [44]. The high infectivity observed in uncapped RCNMV RNAs implies that this virus utilizes a cap-independent mechanism to translate viral proteins.

While the RCNMV genome and function is well characterized for at least the past 30 years, very little is known about how RCNMV interacts with the plant host. To date, no host resistance (*R*) genes have been described to recognize or target RCNMV during infection. RCNMV is known by a symptom to cause host tissue necrosis. It is likely that the plant host must contain *R* genes that recognize RCNMV proteins. RCNMV is found to be capable of suppressing host silencing defense mechanism by a use of its silencing suppressor capability through either one of these conditions; (i) the presence of p27, p88, and a replication competent RNA-2 [45], or (ii) in the presence of RCNMV movement protein [46]. A redundancy of genes encoding silencing suppressor proteins found in RNA-1 and RNA-2 suggests RCNMV is evolutionarily competent to defeat host adaptive immune system. RCNMV replication process takes place on the remodeled endoplasmic reticulum (ER) membrane. Our lab previously found that RCNMV replication proteins (p27 and p88) accumulate at the ER membrane in *N. benthamiana* host and cause membrane restructuring, rearranging and proliferation [47]. Yet, there are no reports about the interaction between RCNMV and host factors/proteins associated with ER. Replication of ss(+) RNA virus is a

complex process that involves numerous interactions among viral RNA, viral proteins, host proteins, and host membrane lipids. The movement of RCNMV to adjacent cells occurs through a plasmodesmata channel that connects cells. However, this intercellular transport channel is dimension restricted and so is too small for any virions or even free viral DNA or RNA strands to move through. An assumption is that RCNMV requires an interaction with host proteins in order to increase plasmodesmata size exclusion limit (SEL) [14, 38, 48]. The most relevant information that may help understand the interaction events between RCNMV and host comes from the studies previously performed on its close relative *Tomato bushy stunt virus* (TBSV), especially the studies from the Nagy group [6, 49]. However, the Nagy group used yeast (*Saccharomyces cerevisiae*) as a model host to study the interaction between plant virus and host [50]. Genome-wide screening of yeast libraries led to the identification of a number of essential and nonessential host factors affecting TBSV RNA replication [51, 52]. The identification of yeast genes revealed that they might affect TBSV RNA recombination by altering (i) the ratio of the two viral replication proteins, (ii) the stability of the viral RNA, and/or (iii) the replicatability of the recombinant viral RNAs [52].

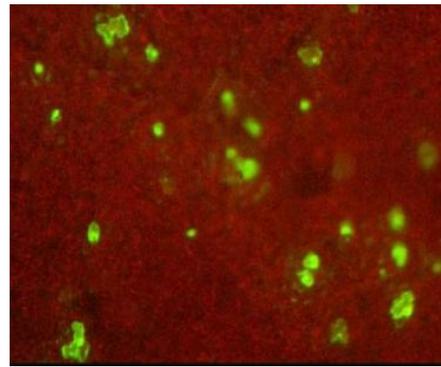
The work in this chapter is ultimately aimed at increasing an understanding of not only host-RCNMV interactions, but also to expand the knowledge base of the involvement of host factors to other ss(+)RNA viruses. This contribution comes from an assumption that the interaction between host factors and the changes in host physiological processes might be analogous among other various ss(+)RNA plant viruses in spite of their diverse genome organization and gene expression strategies. Most of the previously identified host factors are conserved among other plant species, suggesting that ss(+)RNA viruses might selectively

target these conserved host genes/functions/pathways to produce a robust infection as opposed to plant species-specific factors. Such strategies could help viruses expand their host range. In addition, the studies on host factors could provide an insight into their normal cellular function, thus promoting an understanding of the basic biology of plant physiology and why certain plant species are susceptible to viruses.

**A) Healthy, uninfected *N. benthamiana*
4 weeks old**



**B) RCNMV with sGFP infected-*N. benthamiana*
Viral GFP detected at earliest as 8 hpi**



**C) RCNMV-infected *N. benthamiana*
Local symptom, 48-72 hpi**

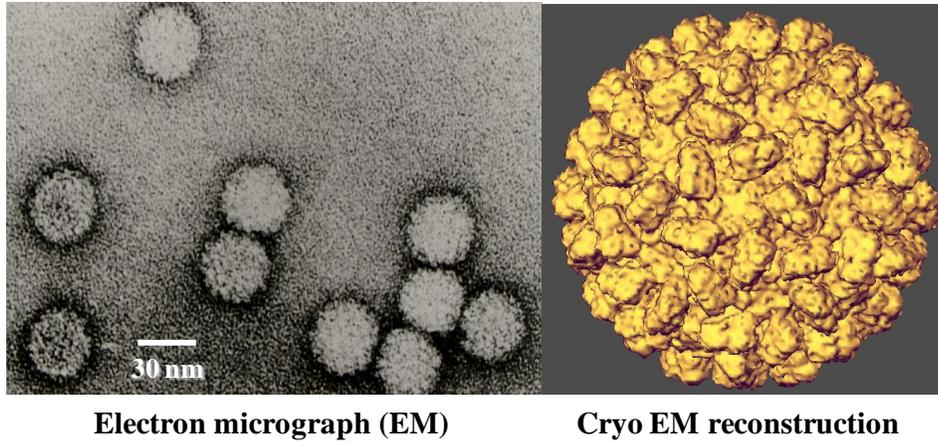


**D) RCNMV-infected *N. benthamiana*
Systemic symptom, 72-96 hpi**



Figure 1 *N. benthamiana* healthy and RCNMV-infected symptoms

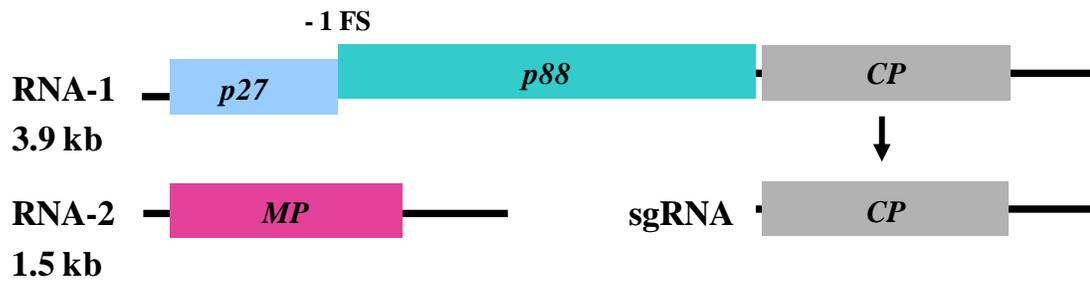
RCNMV virion image



Electron micrograph (EM)

Cryo EM reconstruction

RCNMV bi-partite genome



RCNMV RNA-1 with sGFP reporter gene



Figure 2 RCNMV virion and genome organization

Experimental designs

Here I describe the rationale behind our experimental design in three sections; (i) whole plant model and time point experiment, (ii) microarray experiment, and (iii) microarray confirmation experiment.

Section 1 Whole plant model, time point experimental design and rationale

Virus- infected tissues are biologically complex. This section discusses how several components contribute to this complexity. Each infected leaf consists of multiple cell types ranging from epidermis, mesophyll (palisade and spongy), guard cell, trichome, stoma, to vascular tissues (xylem and phloem) [53]. An infected leaf consists of multiple states of virus infection largely due to viral cell-to-cell movement. In any one infected cell, virus may just start be starting its first step of the replication cycle; whereas, virus may have already completed its replication cycle and assembly in another cell and be ready to migrate to the neighboring cells to begin another infection cycle. Any inoculation method (methods of how viruses are delivered into plant cells) produces a restricted number of initially infected cells, therefore, different inoculation methods yield different viral quantities (viral titer) in the inoculated leaf. Viral inoculum is usually delivered to a host leaf by a mechanical inoculation (rubbing with carborundum abrasive dust or scratching with needle), a needleless-syringe infiltration (*Agrobacterium*-mediated inoculation) or soaking method. There are no inoculation methods reported in the literature that are 100 percent effective, even when using viruses known to be highly infectious. Some cells are able to be infected right away while other cells are not. An infection rate of 60 to 80 per cent has usually been regarded as

extremely satisfactory [54]. Virus-free zone on the inoculated leaf is sometimes observed as green islands which are the areas that are resistant to a subsequent virus invasion due to an activation of host defense response [55, 56]. Therefore, any available inoculation method that attempts to deliver virus inoculum into “a plant tissue based-model” is considered to be biased, unless the inoculations are performed on “a single plant cell-based system” such as protoplasts.

The protoplast system offers an attractive alternative; however, lack of cell wall and neighboring cells changes a naturally-occurring baseline of host cell physiological processes [57], affecting how host cells interact with viruses and limiting the ability to study how viruses modulate host gene expression to benefit their movement and how host cells communicate with each other to defend themselves from further invasion into the remaining healthy host cells. Thus, we decided to use a whole-plant system in this study in order to mimic the natural state of the host interacting with viruses.

Given the complexity of host leaf/tissue and critical timing of viral movement, regardless of any biased inoculation method, two important problems must be resolved; (a) an inconsistent infection throughout the inoculated leaf, resulting in only specific zones containing higher virus titer than the other zones; and (b) asynchronous replication states, resulting in a mix of various replication states among the infected cells within the same leaf. These two issues have shaped the experimental design in this chapter. There are no ideal plant tissue-based experimental systems available to absolutely conquer these two

fundamental problems; however, the cautious and refined experimental design and statistical analysis offer a possible approach to reduce this noisy background.

In our lab, we employ a mechanical inoculation technique of rubbing viral inoculum with carborundum abrasive and inoculation buffer; however, we are not able to infect all the cells on the inoculated leaf. As seen in Figure 1, RCNMV with sGFP reporter genes (green from viral sGFP protein) is visible in only a small number of host cells on the inoculated leaf; whereas majority of host cells do not contain the virus (red from chlorophyll). The uninfected cells are expected to cause a noise background or dilutional effect in our virus-host interaction study. In the experimental design, we also carefully planned out the time course experiment in order to increase the synchronous infection states among infected cells within the same leaf.

If leaf samples are harvested after viral cell-to-cell movement, it is likely that RNA samples will be a mix of various viral replication states. This is due to asynchronous infection with respect to viral multiplication and movement. Therefore, the time course experiment was applied to the study system in this chapter. The hypothesis is that if the infected leaf is harvested and host cell physiological process is stopped at the precise time point prior to viral movement taking place (24 hours after inoculation), viruses in a majority of infected cells throughout the entire leaf are still in the same replication state and any significantly differential host gene expression is modulated as a result of how the plant responds to the particular state of viral replication. Viral cell-to-cell movement is generally an early event in the infection process, occurring in 4 hours for *Tobacco rattle virus* (TRV) in

N. clevelandii, and 5 hours for *Tobacco mosaic virus* (TMV) in *N. tabacum* and the level of movement protein produced in the first few hours is low [48, 58]. *Tobacco etch virus* (TEV) expressing the uidA reporter gene produces the lesion on host leaf which expands at a rate of 2 hours per one host cell [59]. In our study, we speculate that RCNMV takes about 12-24 hours to complete infection in a primary infected cell, and we found that RNA-2 transcript which encodes movement protein is not increased until after 12 hours (Figure 26). Since viral movement can take as short as a few hours up to 24 hours, the selection of time point of study at 2, 6, 12, and 24 hours post inoculation (hpi) was designed to harvest leaf samples and halt host physiological process in order to achieve the maximum infection synchronization, and to extend the study to cover the information of how host gene expression behaves after viruses complete the infection process in the primary infected cell, and move to the neighboring cells. Our hypothesis is that the stopping of virus infection at four different time points at 2, 6, 12, and 24 hours provides the maximum synchronization of RCNMV infection between infected cells. This leads to the other hypothesis that the impact of host gene modulation is affected by the duration of infection, not solely the infection or time of the day. The details of these hypothesis testing is discussed in section 2 and the result is illustrated in Figure 4.

The nature of the leaf architecture tremendously affects host genes whose expression is tightly controlled by temporal and spatial regulation: (1) certain genes that are profoundly altered at specific developmental stages in specific cell types or in specific zones on the leaf, and (2) certain genes that have the opposite expression pattern among different cell types. In addition, we found that a younger *N. benthamiana* leaf is more susceptible to RCNMV

infection. Total RNA ideally should be isolated from a whole leaf with the same developmental stage to avoid these problems, and the hypothesis is that if these genes are not modulated during virus infection, their expression values are cancelled out or yield a small differential expression during a subtraction between infected-tissue sample and control. In our experimental design, we were able to extract total RNA from a leaf with the same age; however, we were not able to use a whole leaf due to limitations of the RNA isolation kit (QiagenTM). Therefore, if tissue used for a total RNA extraction is cut from different parts of a leaf and contains different cell types, a false positive signal may be unintentionally picked up from a differential expression of host genes that are affected by the leaf architecture, rather than by viral infection.

A differential expression of host genes is computed from a subtraction of RCNMV-inoculated leaf and mock-inoculated leaf (control using inoculation buffer). The data presented in this chapter represents an overview of host transcription profiling changes at four different durations of viral infectivity. The time points or durations of study are assumed to correspond to different states of the viral replication cycle. The major hypothesis of this chapter is the extent of up- or down-regulation of host gene expression is biologically correlated with respect to the particular state of virus replication cycle.

Section 2 Microarray experimental design, statistical analysis and rationale

Details of how host biological networks and host factors underline the establishment of successful virus infections in plants by influencing viral replication, translation,

movement, and suppression of plant host defenses remains poorly understood. Single gene-orientated studies emphasize the study of one particular host gene or a few host genes at a time. This type of study is practical when research questions are focused on a certain aspect of host gene function. Despite host genes being well-characterized in any single gene-orientated studies, information obtained from this type of study limits an ability to resolve two fundamental research problems: (i) how an individual host gene is connected to other host genes in biological networks; and (ii) how host biological networks are comprehensively altered during virus infection. One plant gene could be involved in various metabolic pathways. For this reason, high-throughput approaches are required to be implemented in host-virus interaction studies in order to tackle these two issues.

Gene expression microarray is a very powerful high-throughput technology. It is prepared by a high-speed robotic printing of the complementary DNAs on glass which were used for quantitative expression measurements of the corresponding genes. Microarray was developed to monitor the expression of many genes in parallel. In other words, it allows the host transcriptomic profile to be collectively and simultaneously gathered from individual genes in a single shot study [60]. This pushes the boundary of our curiosity from the perspective of single genes studied to many thousands of genes studied with a single experiment. Bioinformatics and genomic databases are necessary tools for high-throughput data manipulation. They extend an ability to extract significant and meaningful information from a massive amount of genomic data to predict how the whole network of host physiological processes behaves in response to virus infection. A collection of differential gene expressions are dissected to construct the predicted models for metabolic pathway

networks, gene-regulatory network (protein-DNA interaction such as transcription factor and its cognate nucleotide sequences), and interactome network (protein-protein interactions) that are impacted by virus infection. In our microarray study, the overall data mining was extracted from many different genomic databases, including GenBank [61], GEO (Gene Expression Omnibus) [62], SGN (Sol Genomic Network) [63], GO (Gene Ontology) [64], InterProScan [65], TAIR (The Arabidopsis Information Resource) [66], KEGG [67], and FunCatDB (Functional Catalogue Database) from MIPS (The Munich Information Center for Protein Sequences) [68, 69]. The ultimate goal of the microarray experiment in plant-virus interaction study is to determine host pathways that are most affected by virus infection. This could possibly lead to an identification of host factors that determine the fate of virus survival in the host cell and whether or not the entire plant commits to this decision. Some host genes are driven by the plant host itself to mainly trigger its defense mechanism; conversely, some host genes are driven by the virus to benefit its own survival and produce a successful infection.

At the time of this study, there were no commercially available microarrays for *N. benthamiana*. So we decided to develop a custom microarray to be used in our virus-host interaction study. The microarray was derived from *N. benthamiana* expressed sequence tags and unigenes. The creation of the *N. benthamiana* microarray is thoroughly described in Chapter 2. Our microarray is a 13K unigene-based *N. benthamiana*; therefore, the expression of 13,413 host genes were simultaneously monitored and collectively gathered at one time of study. Four different time points after virus inoculation was selected to study host gene expression. The hypothesis is that the duration of virus infection elicits the unique changes in

host gene expression to suit a particular stage of the viral infection cycle. The microarray experiment was designed by a previous graduate student [70]. The simple diagram that demonstrates our microarray experimental design is shown in Figure 3. The weakness of this microarray experimental design is the use of only single biological-pooled samples to represent one time point of study for either mock- or virus- treated samples. Therefore, at any time point of study, there are no biological replicates in our plant-virus interaction microarray study; instead, there are four microarray technical replicates. The experiment was performed in one day and so the microarray sample was pooled from plants performed within the same day. As a result, our microarray experimental design does not have samples from other days of experimentation. In other words, this design does not contain samples that possess day-to-day variation. Biological replicates from independent studies performed at different times or with different sets of plants or reagents are important for biological research to test the experimental reliability with respect to the reproducibility of the result; therefore, this microarray experimental design may not be appropriate for plant-virus interaction study [71]. From my understanding, this experimental design was not originally intended to be used for plant-virus interaction study; instead, it was aimed to test the repeatability and reliability of microarray hybridization (with a use of 4 array technical replicates for each biological pooled sample) [72] because our *N. benthamiana* microarray is the first microarray for this plant species and so it was used the first time in this project.

Despite our microarray data presented here (either mock or virus at each time point of study) being obtained from a one day experiment and one biological sample pooled from four different plants within the same day of experiment, it still contains a meaningful result that

can be used as preliminary information for further study. Given the weakness of the microarray experimental design, we employed a statistical analysis of the microarray data to validate two issues: (1) time point experimental design to test if the state of the virus infection cycle is synchronized between infected cells; and (2) microarray experimental design to test if our microarray data is amenable for host-virus interaction studies. The major hypothesis of the microarray experiment is that host gene expression is modulated by duration of virus infection. The condition of mock or virus infection treated on the plant is called “treatment”. This hypothesis is focused on the interaction effect between time and treatment (together, equal to duration of infection). Statistical concept states that if host gene expression is significantly modulated by duration of infection, it means that the effect of treatment to host gene expression depends on the effect of time; therefore, treatment (factor 1) and time (factor 2) are the factors that contribute to the interaction effect which is a duration of infection (factor 3). However, it is not known which of these three factors significantly affects the changes in host gene expression. Therefore, microarray statistical analysis must be designed to distinguish these three effects in order to answer three important research questions in this experiment. These three questions are:

- 1) is host gene expression modulated by only virus infection (factor 1)?
- 2) is host gene expression modulated by only the time of day (factor 2)?
- 3) is host gene expression modulated by duration of infection (factor 3)?

Microarray statistical analysis was designed to implement “treatment factor”, time factor” and “(treatment x time) factor” in a statistical model which is illustrated as an effect model shown in equation 1. An expression of each host gene is a response observed from the

combination of treatment and time. The microarray experiment is a two-way factorial experiment that contains two categorical factors, time and treatment. The microarray data structure of individual host genes is a 2 x 4 factorial design, representing 2 levels of treatments and 4 levels of time points, as shown in Table 1. The expression data of each host gene came from 4 array replicates of mock treatment and 4 array replicates of virus treatment at each time point of study. Four time points were selected to monitor host gene expression; therefore, each gene had a total of 32 data points, which is a multiplication product derived from (2 treatments x 4 time points x 4 array replicates) equal 32. The degree of freedom (df) for the analysis of each host gene is shown in Table 2 [73-75].

Our microarray statistical model is shown in equation 1. The primary objective of the microarray statistical model is to test the hypotheses that corresponds to “treatment factor”, “time factor”, and “duration of infection factor (interaction effect between treatment and time)” in order to determine if host gene expression is significantly associated with any of these three factors. The hypotheses that are being tested in a statistical model are collectively called “model hypothesis”. Details of these three factors (effects) and their corresponding model hypothesis statements are described below.

1) **Effect of treatment** (mock (m) and virus infection (v))

Hypothesis statements are

H_0 : host gene expression is not associated with treatment.

H_a : host gene expression is associated with treatment.

2) **Effect of time** (2, 6, 12, and 24 hours post inoculation (hpi))

Hypothesis statements are

H_0 : host gene expression is not associated with the time of day.

H_a : host gene expression is associated with the time of day.

3) **Interaction effect of treatment and time** (mock at 2 hpi (2m), virus infection at 2 hpi (2v), mock at 6 hpi (6m), virus infection at 6 hpi (6v), mock at 12 hpi (12m), virus infection at 12 hpi (12v), mock at 24 hpi (24m), and virus infection at 24 hpi (24v))

Hypothesis statements are

H_0 : host gene expression is not associated with duration of infection (treatment and time).

H_a : host gene expression is associated with duration of infection (treatment and time).

The *N. benthamiana* microarray contains probes representing 13,413 host genes and 2 RCNMV genomic RNAs (RNA-1 and RNA-2); therefore, the microarray statistical model had to be iterated for 13,415 times to analyze the expression of all the genes on the array. In other words, individual model hypothesis analyzes data from a single gene data at a time. Therefore, by analyzing one gene at a time, the overall 13,415 modeled hypothesis tests were conducted for 13,415 microarray genes. Individual model hypothesis determines three important p-values which are used to examine whether or not, any of these three effects (“treatment”, “time”, and “duration of infection”) is significantly associated with the expression of the gene being analyzed in the model. The frequency of all p-values derived from one particular effect of the model hypothesis tests is called “p-value distribution”. An expression value of any given gene is called “a response” and expression of all 13,415

microarray genes is called “overall responses”. The pattern of p-value distribution derived from any effect that is being analyzed in the model is used to determine if the effect is significantly associated with the overall responses.

If a trend of p-value distribution is mostly clustered closer to zero, effect being tested has a significant association with the overall responses. But if a trend is closer to one, it is likely that it does not have a significant association with the overall responses. In the case of an ambiguous trend, it is still in question whether or not this effect has a significant association with the overall responses. Therefore, individual p-value distributions derived from either “treatment”, “time”, or “duration of infection” provides useful information to determine its association with the overall expressions of 13,413 host genes and 2 RCNMV genomic RNAs (RNA-1 and RNA-2). This strategy is used to draw a preliminary summation about the association of the three effects with the overall changes in the host transcriptomic profile in an early virus infection. In other words, this technique is useful to test if our time course experimental design is appropriate to be used for RCNMV-host interaction studies in the early infection process. It is necessary that our time course experimental design (four different time points) is able to distinguish the effect of modulation of host gene by RCNMV at different durations of infection. The p-distribution graphs of these three effects (“treatment”, “time”, and “duration of infection”) were illustrated in Figure 4. The distribution graphs of “treatment” (Figure 4A) and “time” (Figure 4B) are close to zero suggesting that these two parameters affect the expression of host genes. However, as seen again in the p-value distribution of “duration of infection” (Figure 4C), the trend of the p-value from all 13,415 microarray genes is also accumulated close to zero. The strong

association between host gene expression and duration of RCNMV infection confirms that there is an interaction effect between “treatment” and “time” to host gene expression and viral gene expression. Thus, our time course experimental design is appropriate and acceptable for studying RCNMV-host interactions early in the infection process. In addition, it confirms that our microarray experimental design as well as our microarray data from this RCNMV-host interaction study at 2, 6, 12, and 24 hpi could be used as preliminary data to examine the modulation of host gene expression at different duration of infection. Our microarray statistical analysis supports the conclusion that the alteration of host gene expression is influenced by a particular stage of viral infection. Further analysis of the p-value distributions indicates inflation, especially for “time” and “duration of infection” (Figure 4, B and C). This needs to be addressed, for example by using permutation testing to obtain more accurate p-values, before claiming any significant associations. Instead of correcting the p-values to claim significance, an exploratory approach was used where the increased chance of false positives due to inflation was acceptable.

All mean expression data, of any one host gene, retrieved from any combination of “treatment and time” were compared in a pair-wise manner. This pair-wise comparison is a differential gene expression between the pair. Pair-wise means that each comparison looks at a difference (a subtraction) between the means of two populations, and each population mean is obtained from a data group under a combination of “treatment and time”. Four time points (2, 6, 12, and 24 hpi) and two treatments (mock (m) and virus (v)) were selected in the microarray experiment; therefore, all combinations of “treatment and time” yield 8 data groups with respect to 8 population means (2m, 6m, 12m, 24m, 2v, 6v, 12v, and 24v). These

8 population means were compared in a pair-wise fashion and so there are 28 pairs for each host gene. Therefore, there are 375,620 pairs corresponding to 375,620 independent hypothesis tests for all 13,415 microarray genes. The general statements of these hypotheses are

H_0 : gene expression between the pair are not different ($\mu_1 = \mu_2$), and

H_a : gene expression between the pair are different ($\mu_1 \neq \mu_2$).

Statistical concept states that if multiple pair-wise comparisons or multiple hypothesis tests are conducted in any statistical analysis, the consequence is an increased chance of making a type I error [76-78]. When one rejects H_0 because the calculated p-value is less than a p-value cutoff, one accepts the possibility of making a false rejection of H_0 when H_0 is true. The false rejection in this decision making is known as a type I error. The p-value cutoff, known as “ α ”, is used to define the possibility of a type I error that one is willing to commit upon hypothesis testing. If $\alpha = 0.05$, there is a 5% chance of committing a type I error. With each comparison using “ α ” in the hypothesis test, there is a chance of making a type I error on each comparison. Therefore, the more comparison pairs are tested, the more likely a type I error occurs at least once. If each pair-wise comparison test is accounted on $\alpha = 0.01$, the total of pairwise comparisons is 375,620 and so the probability of committing a type I error at least once in the microarray multiple pair-wise comparison tests is $1 - (1 - 0.01)^{375620} \sim 1$, or a ~100% chance of making a type I error at least once [79]. Therefore, a type I error definitely occurs in this microarray multiple pair-wise comparison. In any multiple comparison, there are two different types of errors, the comparisonwise error rate (a

probability of committing a type I error for a single pair-wise comparison), and the experimentwise error rate (a probability of committing a type I error somewhere in the analysis). Altogether, it leads me to focus on a control strategy of the experimentwise type I error. The strategy that I used to tackle this problem is reducing “ α ” or a p-value threshold for rejecting H_0 in each pair-wise comparison test. Bonferroni correction is the simplest method for adjusting p-value cutoff that for any individual test, H_0 is rejected if calculated p-value $\leq (\alpha/m)$, when m is the number of pair-wise comparison tests [79, 80]. However, Bonferroni correction is extremely conservative in the sense that while it reduces the number of false positives, it also reduces the number of true discoveries; thus it has very little power to find differences between means. The other method that I used is False Discovery Rate (FDR), which is designed to control the expected proportion of incorrectly rejected H_0 [78]. The FDR also reduces the p-value threshold; however, it controls the number of false discoveries in those tests that result in a discovery. It is less conservative than Bonferroni and has a greater power to find truly significant results. One way to look at the difference between these two approaches is that a Bonferroni adjusted p-value cutoff of 0.05 implies 5% of all multiple pair-wise comparison tests will result in false positives; whereas, an FDR adjusted p-value cutoff of 0.05 implies that 5% of significant tests will result in false positives. Bonferroni adjusts the p-value threshold solely based on the number of pair-wise comparison tests; on the other hand, FDR takes the characteristic of p-value distribution derived from each p-value calculated from each pair-wise comparison test into account to adjust the p-value threshold. Bonferroni is more conservative in that it requires extra evidence to reject H_0 for pair-wise comparisons; while, FDR is more sensitive because it is

able to detect the significance of a smaller change in gene expression. The fold change distribution (comparing host gene expression between virus and mock at each time point of study) of all 13,413 host genes spotted on the microarray is shown as \log_2 scale in Figure 5. The expression of all genes was changed in a narrow range of about 2-fold in either direction. This is quite expected because the experimental system using a whole-plant model is expected to cause a dilution effect on host gene expression as described in the section 1(whole plant model, time point experimental design and rationale). For this reason, FDR was more preferred in my microarray statistical analysis in order to determine a significantly differential gene expression between mock- and virus- inoculated plants at different time points of study.

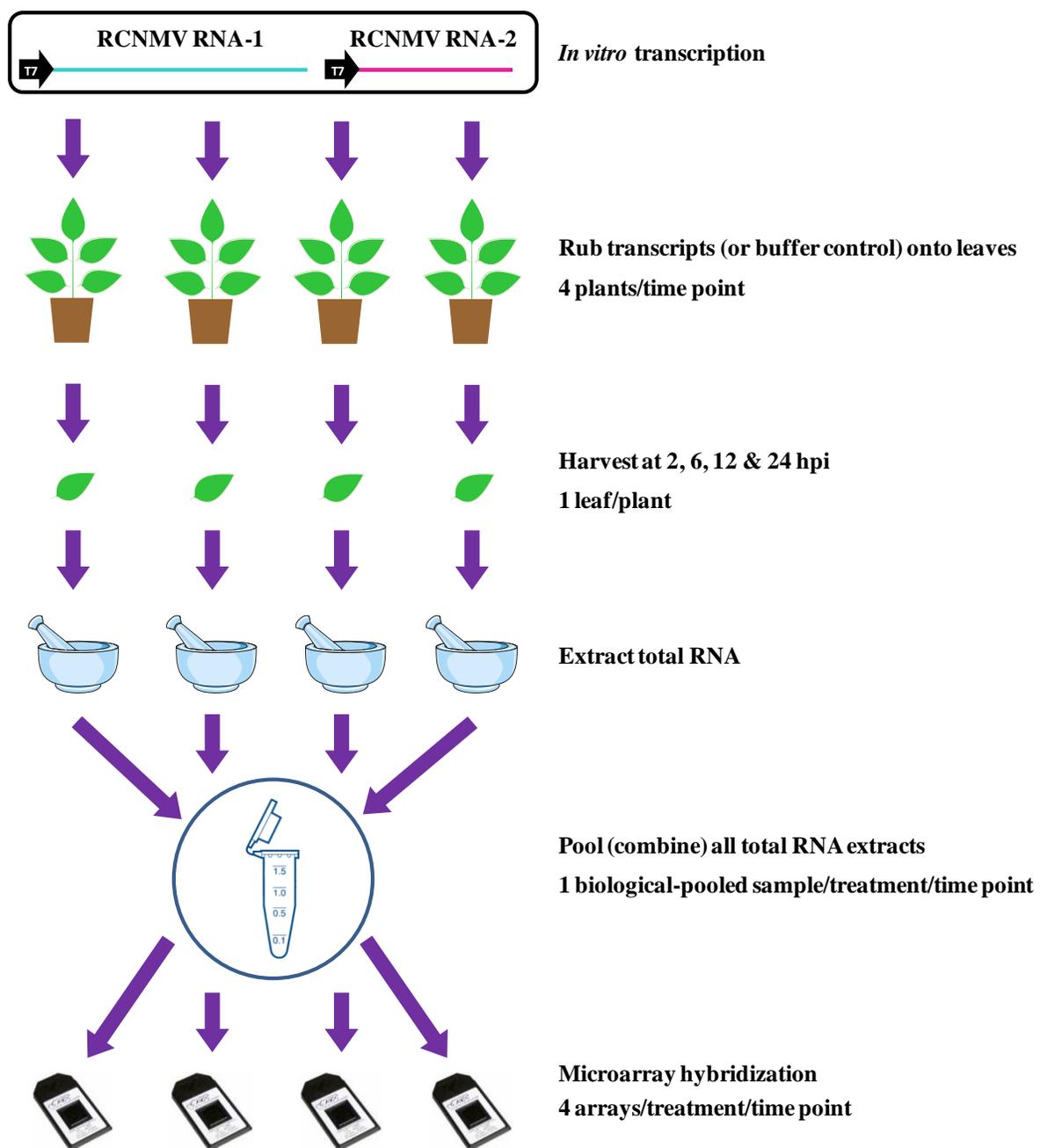


Figure 3 Microarray experimental design diagram

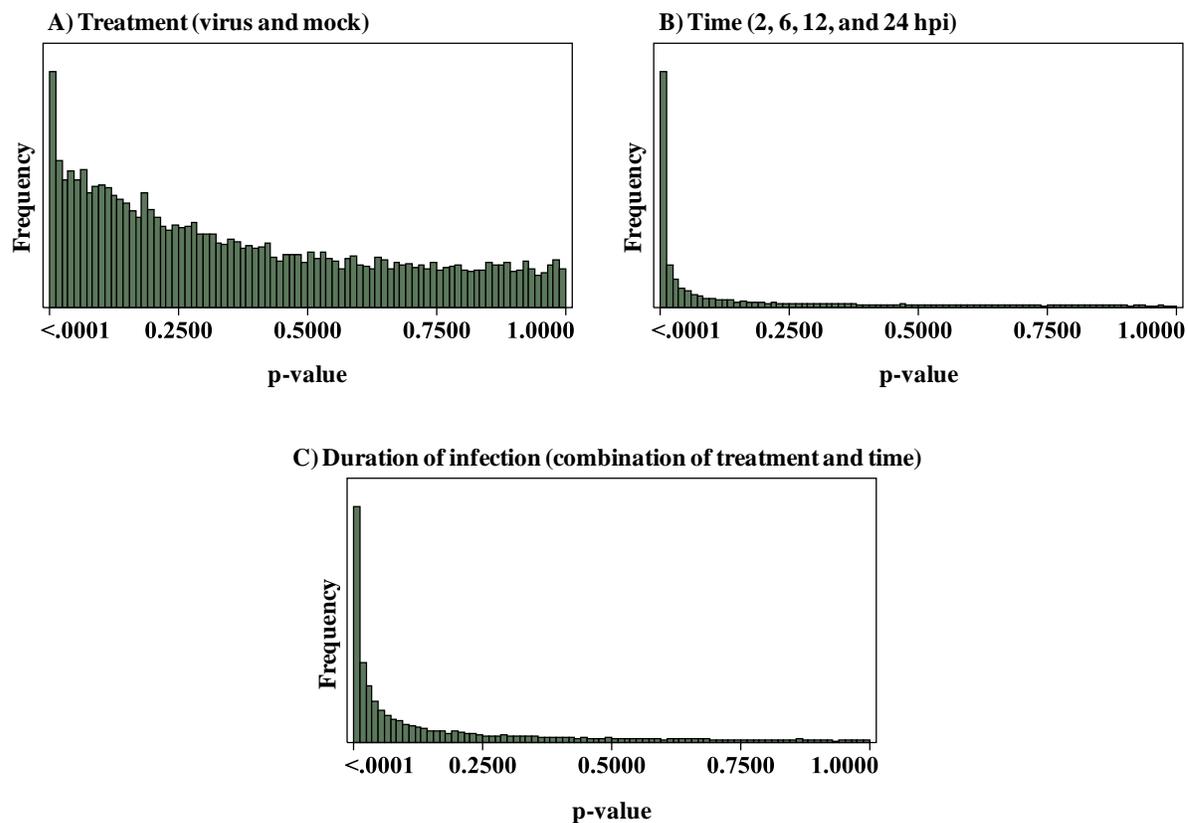


Figure 4 P-value distributions shows the effects of treatment, time, and duration of infection on the modulation host gene expression in the early stages of RCNMV infection. The p-value distribution displayed in each effect: A) treatment, B) time, and C) duration of infection were derived from the microarray statistical analysis utilizing the effect model (equation 1) and ANOVA (Table 1 and 2).

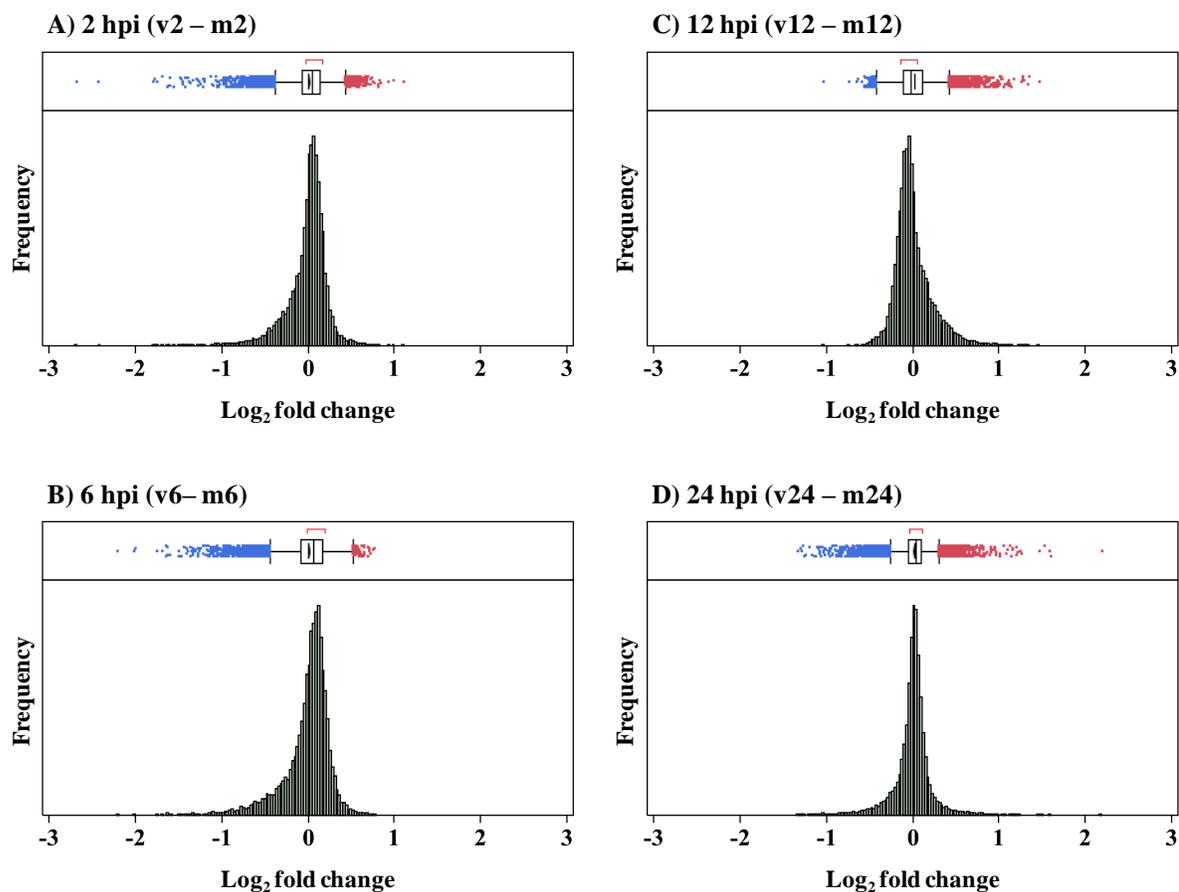


Figure 5 Microarray fold change distributions display the differential expressions of 13,413 host genes at 2, 6, 12, and 24 hours post inoculation (hpi). A fold change of each host gene was computed from a subtraction between microarray expression data of RCNMV-inoculated plants (v) and mocks (m) at each time point of study.

Equation 1 Statistical model for microarray data analysis

The model is illustrated as an effect model.

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Notation

y_{ijk} : the k th host gene expression data point observed from the combination of the i th level of “treatment factor” and the j th level of “time factor”

μ : reference level (intercept)

α_i : effect of the i th level of “treatment factor”

β_j : effect of the j th level of “time factor”

$(\alpha\beta)_{ij}$: interaction effect between the i th level of “treatment factor” and the j th level of “time factor”

ε_{ijk} : the k th random error from the combination of the i th level of “treatment factor” and the j th level of “time factor”

Table 1 The structure of microarray data analysis using a 2 x 4 factorial design

		Time (hours post inoculation; hpi)			
		2 hpi	6 hpi	12 hpi	24 hpi
Treatment	mock	4 arrays	4 arrays	4 arrays	4 arrays
	virus	4 arrays	4 arrays	4 arrays	4 arrays
Total = 32 arrays					

Table 2 Degree of freedom (df) for microarray data statistical analysis

Sources	Description	Formula for df calculation	df
Treatment	Mock, virus (p = 2)	$p - 1$	1
Time	2, 6, 12, 24 hpi (q = 4)	$q - 1$	3
Treatment*time	-	$(p - 1) \times (q - 1)$	3
Error	-	$n_T - (p \times q)$	24
Total	32 arrays ($n_T=32$)	$n_T - 1$	31

Section 3 Microarray confirmation experimental design, statistical analysis and rationale

The plant samples for our microarray hybridization were no longer available and given the weakness of limited biological replicates in the microarray experimental design as described in section 2, I decided to re-design the experiment for microarray confirmation. Although this design is aimed to confirm microarray data, the main goal is to increase the number of biological replicates and to reduce the experimental errors.

No matter how carefully the experiment is designed and conducted, there are always experimental errors in the measurement of host gene expression. The experimental errors consisted of both systematic error and random error [73]. In the microarray confirmation experiment, I am interested in only the effect of virus infection on host gene expression; therefore, any other effects on host gene expression are considered experimental errors. Plant maintenance environment, personnel, viral inoculum preparation, and inoculation procedure contribute to the random error. This type of error is reduced by implementing statistical analysis in the hypothesis testing. The statistical analysis is based on the assumption that random error is truly random and scattered around the mean value; as a result, the arithmetic mean of the random error is expected to be zero [73]. My microarray confirmation experimental design is based on how statistical methods were applied to reduce random error. Therefore, the statistical analyses were planned out before the actual experiments were carried out. The pre-planned statistical methods were used to determine numbers of plant samples, RNA sample handling, and day of experiment.

The word “random” indicates that random error is caused by inherently unpredictable fluctuation (variation) in an experimental system [73]. There are two variables which contributed to random error that are required to be removed from our microarray confirmation experimental design. These two variables are (i) day-to-day variation and (ii) plant-to-plant variation within the same day of experimentation. These two variables influence plant gene expression throughout the period of study. The hypothesis of the microarray confirmation experiment is that host gene expression is solely modulated by virus infection. This experiment is aimed to measure changes in host gene expression at sub 24 hours of virus infection. Prior to being able to determine such measurements, these two variables (day-to-day variation and plant-to-plant variation within the same day of experimentation) have to be eliminated. I am only interested in the effect of virus infection on host gene expression. By eliminating these two variables, I can now focus only on differential host gene expression caused by the effect of virus infectivity. Therefore, it is necessary to design experiments and perform statistical analyses that are able to distinguish between the effect of virus infection and the effect of these two variables on changes in host gene expression.

Time points of study are at 2, 6, 12, and 24 hpi. Experiments for all time points started around the same time at 7-8 am so that plants were in the same environment. Host gene expression was monitored over 24 hours in this time point experiment. This 24 hour period included 16-hour daylight (from 5 am to 9 pm) and 8-hour dark (from 9 pm to 5 am). The nature of plant physiological processes relies strictly on day time and night time [81-83]. Certain plant metabolic pathways are shut down during day time but turned on during night

time, and vice versa. Since plants were maintained at a green house facility (NCSU, Method Road, Raleigh, NC), light-dark period was mainly controlled by the natural sunlight and fluorescent light system. The fluorescent light system was programmed to detect the natural sunlight intensity. Any cloudy days or any cloudy periods where the natural sunlight is insufficient during a 16-hour daylight setup, triggered the fluorescent lighting system to be automatically turned on. However, plant signal transductions in certain light-regulated metabolic pathways are influenced drastically within less than a second. Therefore, any fluctuation due to light-dark period throughout 24 hours of study easily alters how plant physiological processes behave, resulting in the variation of experiments on a day-to-day basis. This variation is not the main effect that I am focused on in this experiment; therefore, it needs to be eliminated. The assumption is that there is no variation between the days of experiments. Meaning, no matter what day the experiment is carried out, the result should be the same. Therefore, four independent studies performed on four different days, a week apart were assigned in my microarray confirmation experiment to test this assumption. Performing an independent study on a different day offers more refined results and higher repeatability of the experiment.

The major goal of microarray confirmation statistical analysis is to compare host gene expression between mock plants and virus-infected plants at one specific time point after inoculation. I am not interested in comparing host gene expression between time points. In other words, my statistical analysis for microarray confirmation is to compare two population means. One population mean is host gene expression data from mock samples and the other is host gene expression data from virus-infected plants at one specific time point of study.

Statistical methods that are typically used to compare the differences between two population means are student's t-test (equation 2) and matched pair t-test (equation 3) [73]. Usage of these two methods depends on the nature or the assumption of samples. Student's t-test is based on the assumption that samples are independent from each other. On the other hand, matched pair t-test is based on the assumption that samples are dependent on each other, so-called paired data.

I performed four independent studies on four different days in my microarray confirmation experiment. I use matched pair t-test (equation 3) to test my hypothesis (equation 4) and my statistical model is shown in (equation 5). The assumptions of my plant samples used in the microarray confirmation experiment are (i) no variation between days of study and (ii) mock samples are the baseline controls for virus-infected samples within the same day of study. In other words, the control mock plant performed on that day is used as a baseline for only that particular day. It cannot be used as a baseline for another day's experiments. Therefore, mock and virus infected samples collected from each day are considered dependent or paired samples. This is like blood pressure taken from the same patient before and after treatment. But in my experiment, ideally, host gene expression was taken from the same plant if inoculated with buffer or if inoculated with virus. The data from mock plants on that day is match-paired control subject of the data from virus-infected plant performed on the same day. Matched pair t-test is considered a more sensitive test than student's t-test. If there is a true difference between the pairs without the influence of the day of experiment, the matched pair t-test is more likely to pick this up. If student's t-test is used instead of matched pair t-test, day-to-day variation will be taken into account as a main effect

to cause the changes in host gene expression. Therefore, matched pair t-test (equation 3) is more appropriate to be implemented in my hypothesis testing than student's t-test (equation 2) in order to reduce the effect of day-to-day variation in host gene expression. Four independent experiments were performed on four different days. Therefore, four match-paired data sets by day of experiment were used to test my hypothesis (equation 4) that the changes in host gene expression are modulated solely by the effect of virus infection.

Four independent studies are performed on four different days to obtain plant samples for the microarray confirmation experiment. Different sets of plants were prepared specifically for different days of experimentation. Plant-to-plant variation within the same day of experiment is unavoidable. On the day of experiment, I decided to randomly select plants that are different sizes to perform the experiment. This strategy is based on the assumption that the expression values retrieved from wide-ranged plant sizes represent wide-ranged expression values as well. Therefore, the data retrieved from this assumption covers possible expression values influenced by virus infection. I do this to hopefully increase experimental repeatability or reproducibility. Standard deviation indicates the precision of measurements [73]. The wide-ranged expression values yield a larger standard deviation within the same day of experiment. However, this standard deviation is not important for me because I am not interested in the precision of expression values retrieved from the same day experiment. I am more interested in if I am able to have the same or similar result when I performed the experiment on a different day with a different plant set. In other words, I am interested in the variation of host expression values between days more than the variation of host gene expression between each plant within a single day. Therefore, my assumption is no

variation between plants on the same day of experiment. I draw this assumption from the condition that plants were prepared from the same original seed stocks and maintained under the same conditions throughout the experiment.

In any independent study, mock controls were performed on four different plants and viruses were also inoculated on four different plants. I used a sample pooling strategy to physically combine RNA extracts that were obtained from the same day of experiment together in order to reduce the variation due to plants as well as other variations that occurred on that day (such as the position of plant on the bench, water consumption, light intensity). The expression value from the sample pooled within the same day is a more reproducible measurement. Meaning, no matter what plant size used in the experiment, there is a possibility that the significant outcome will fall into this range. Consequently, four independent studies yield four biological replicates for mock and four biological replicates for virus infection, resulting in four match-paired data sets for my statistical analysis.

The main-driven hypothesis of microarray confirmation experiment is that if day-to-day variation and plant-to-plant variation within the same day of experiment are removed, the changes in host gene expression are modulated only by the effect of virus infection and the result should be repeatable.

Equation 2 Student's t-test

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma(\bar{x}_1 - \bar{x}_2)}$$

If the population variances are equal then the information from both samples is combined to give the “pooled estimate of the standard deviation”.

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

The standard error for the sampling distribution of the differences in sample mean is

$$\sigma(\bar{x}_1 - \bar{x}_2) = \hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{with} \quad df = n_1 + n_2 - 2$$

If the population variances are not equal, then the standard error is

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad \text{with} \quad df = \min(n_1 - 1, n_2 - 1)$$

Equation 3 Matched pair t-test

$$t = \frac{\frac{\sum d}{N}}{\sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{N}}{N(N-1)}}$$

d = difference between matched scores

N = number of *pairs* of scores

df = N (number of pairs) – 1

Equation 4 Hypothesis testing in microarray confirmation experiment

μ_1 is population mean of host gene expression from mock samples

μ_2 is population mean of host gene expression from virus-infected samples

Null hypothesis; two population means are not different.

$$H_0; \mu_1 = \mu_2$$

Alternative hypothesis; two population means are different.

$$H_a; \mu_1 \neq \mu_2$$

Equation 5 Statistical model for microarray confirmation data analysis

The model is illustrated as simple linear regression with an indicator variable.

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i$$

Notation

y_i : the i th host gene expression data point observed from the effect of “treatment factor”

β_0 : reference level (intercept)

β_1 : slope

x_1 : the effect of “treatment factor (categorical variable)”

$$x_1 = \begin{cases} 0 & \text{if plant is mock control} \\ 1 & \text{if plant is inoculated with virus} \end{cases}$$

ε_i : the i th random error

Result and discussion**1. Microarray data quality control and reliability**

The distribution of microarray data was visualized by a probability density function and a box plot (Figure 6). The probability density function was estimated by applying kernel

density estimation to smooth the data set [84]. These visualization techniques provided an overall image of how gene expression data was distributed across all arrays. I found that microarray raw data had an inconsistency distribution among its array replicates (Figure 6A and 6B). The differences in data distribution can cause bias and lead to a false interpretation when gene expression was compared between mock- and virus- inoculated plants. I used the Loess normalization method to correct this problem [85]. And as a result, the normalized microarray data was more concordantly distributed across array replicates (Figure 6C and 6D). To assess data reliability, normalized microarray data were compared across all four replicates that were derived from within the same time point and treatment. Pearson correlation coefficient was calculated between these array replicate set. The variation due to the differences between array chips was evaluated based on Pearson correlation coefficient. Data correlation was visualized by a group scattering plot technique (Figures 7 to 10). I found that Pearson correlation coefficients of all combinations of time point and treatment was more than 0.9. The high correlation observed among array technical replicates indicated that the difference due to array hybridization were not significant. This suggests that our custom microarray chip and the hybridization protocol suggested by Nimblegen are reliable.

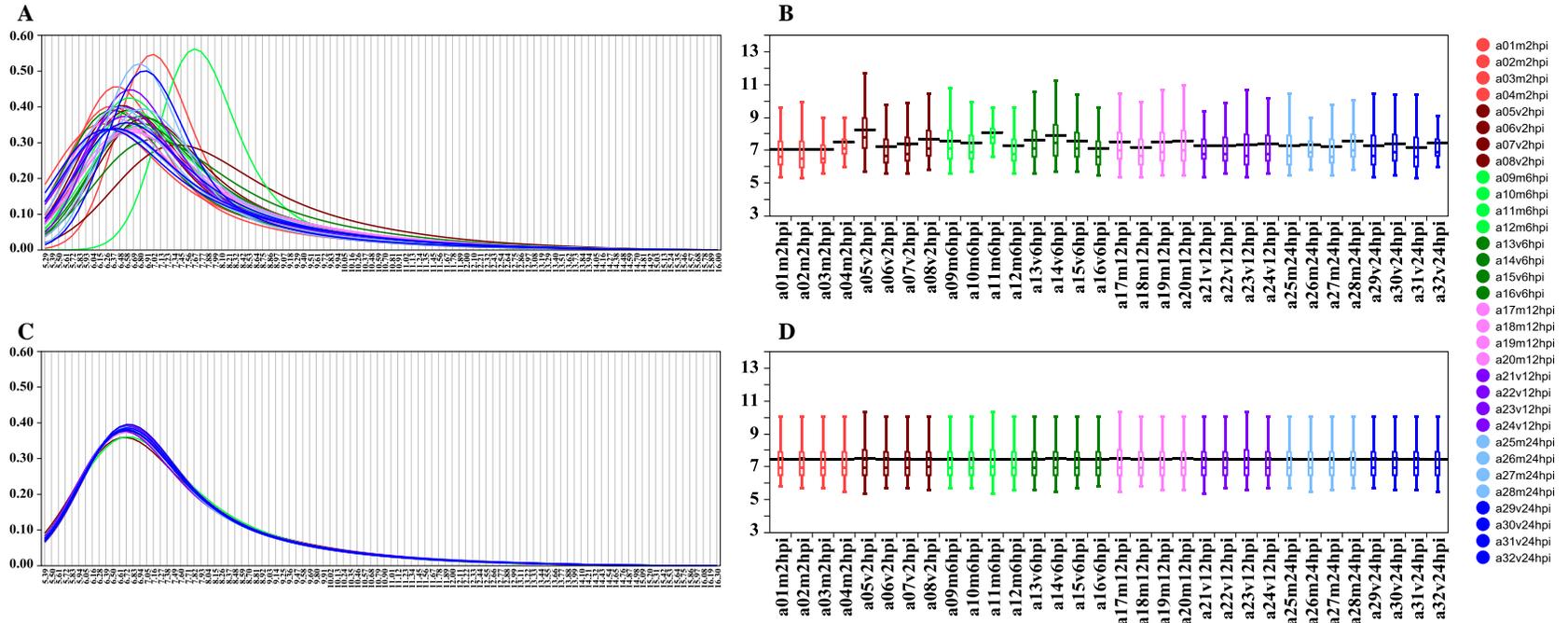


Figure 6 Microarray data distribution and quality control before and after Loess normalization. These Figures are A) probability density function of raw microarray data, B) box plot of raw microarray data, C) probability density function of normalized microarray data, and D) box plot of normalized microarray data.

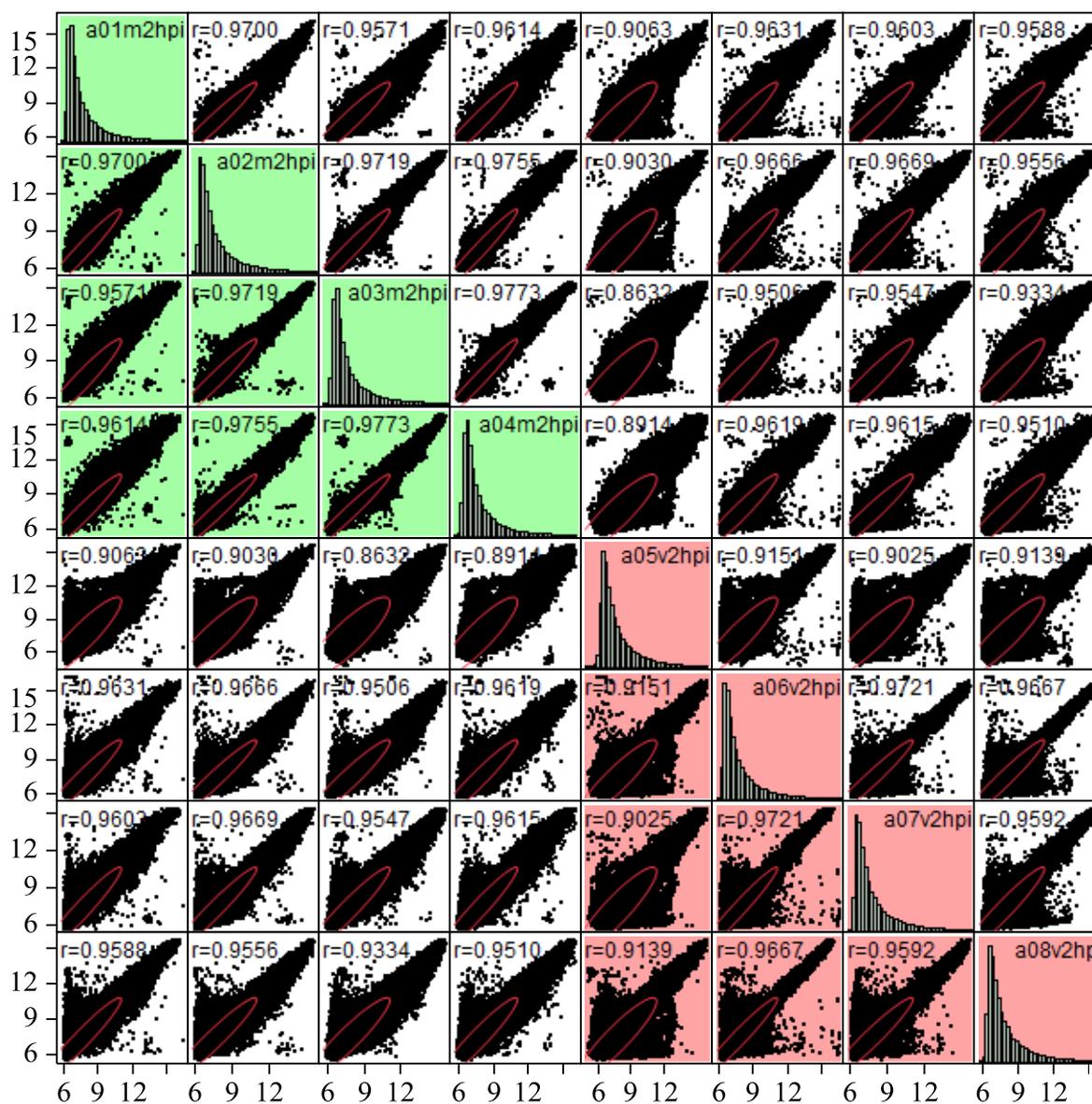


Figure 7 Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 2 hpi

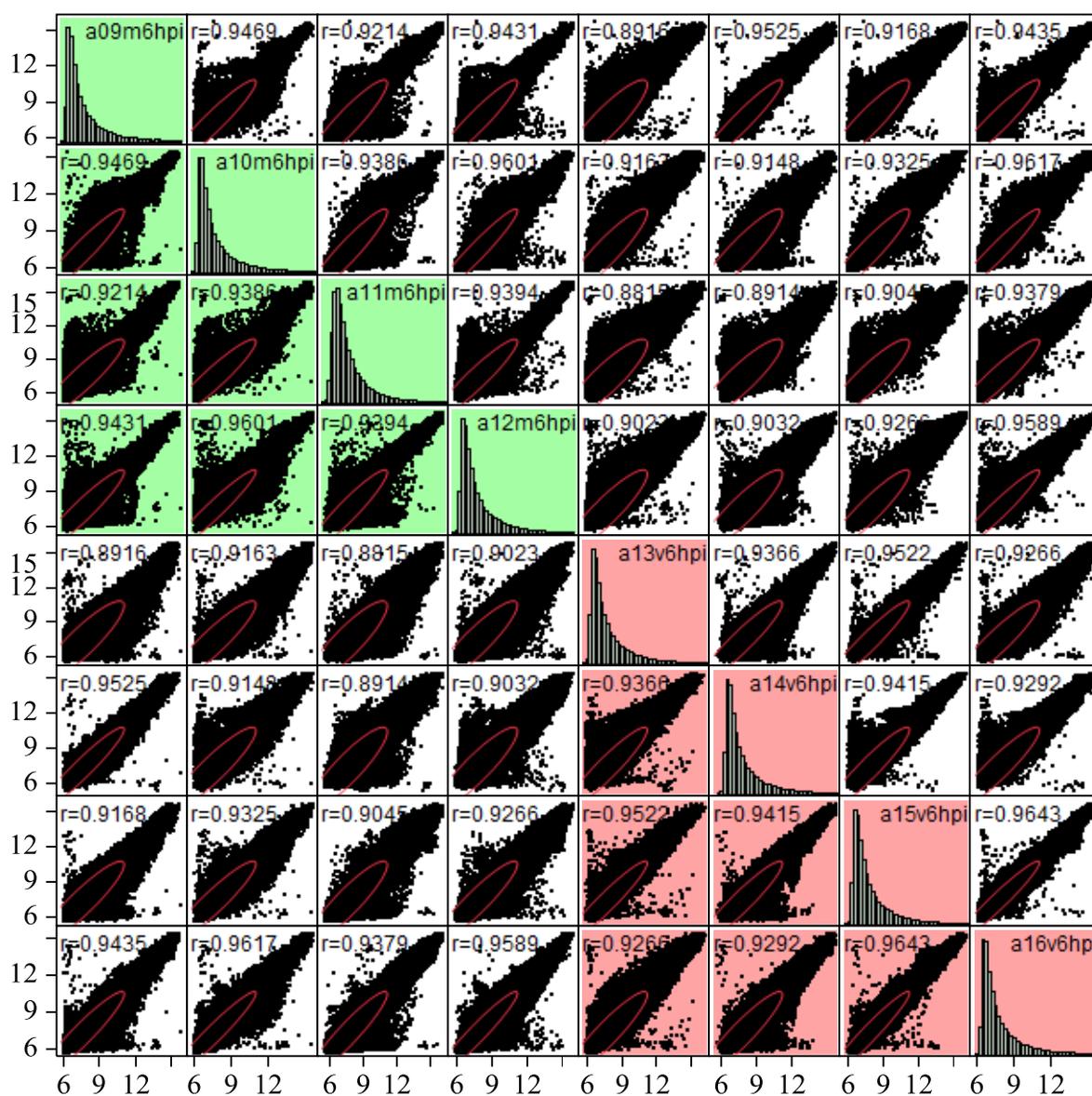


Figure 8 Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 6 hpi

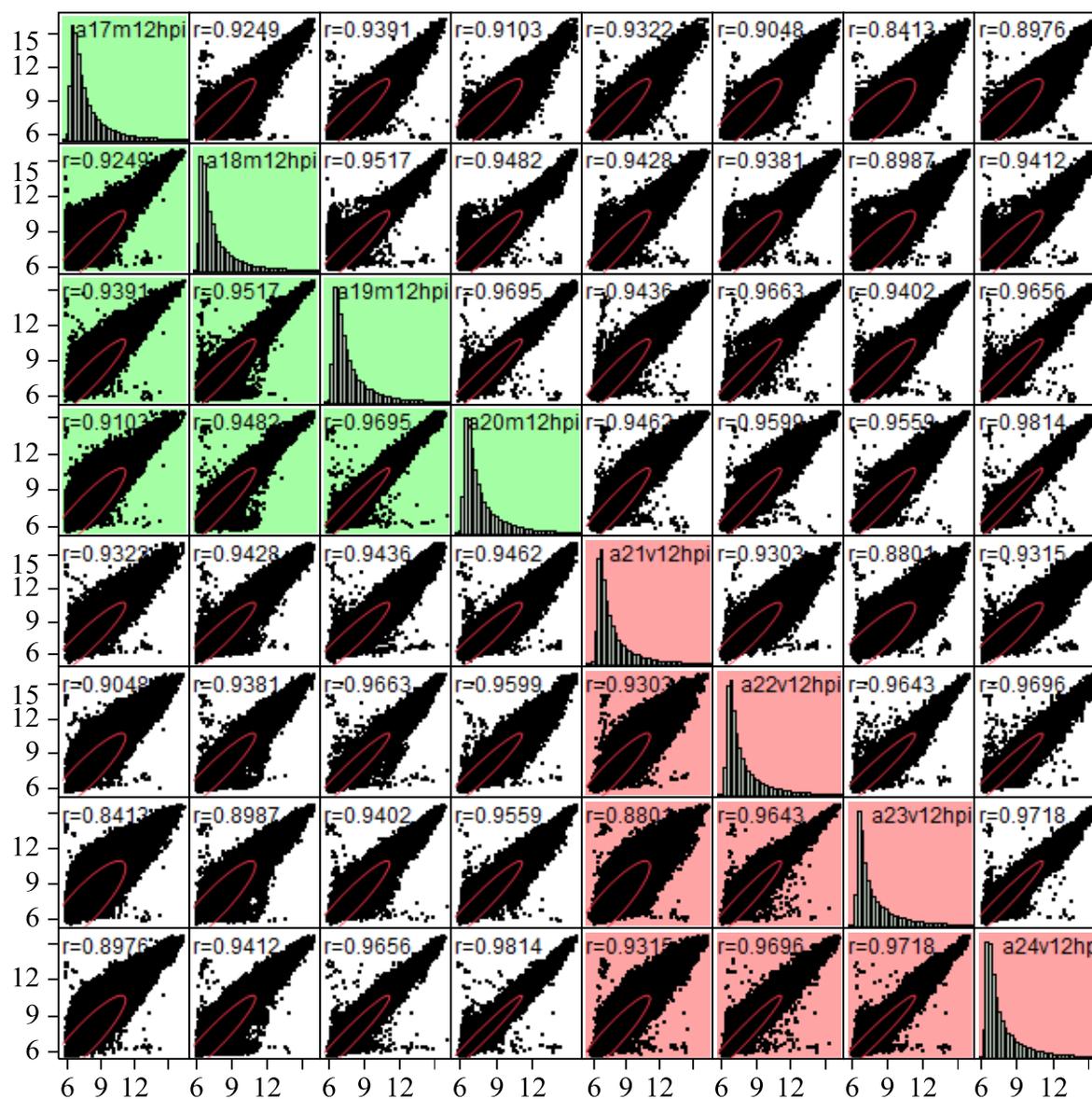


Figure 9 Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 12 hpi

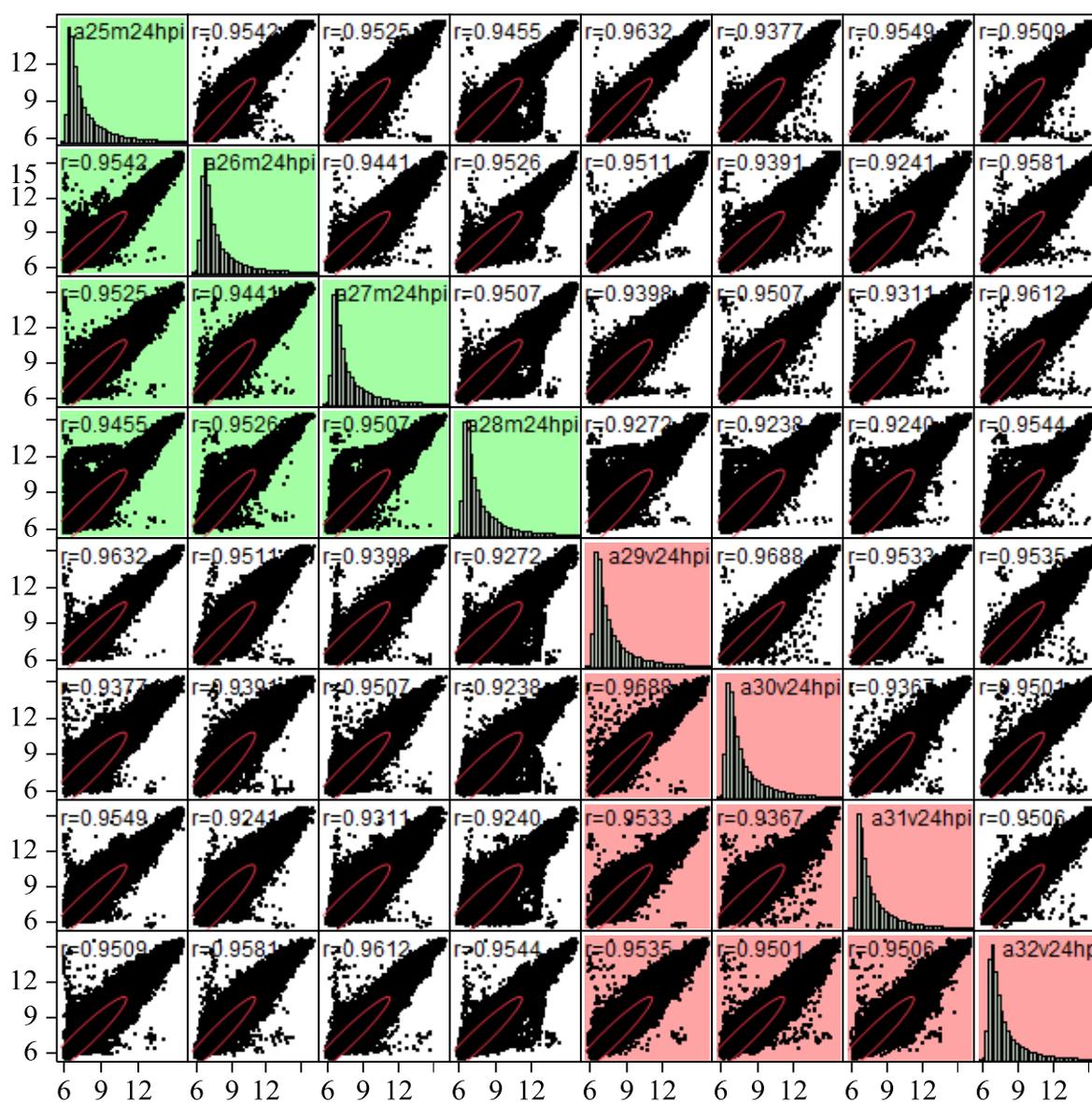


Figure 10 Correlation coefficients of the normalized microarray data derived from 4 array replicates of mocks (green) and 4 array replicates of RCNMV-inoculated plants (pink) at 24 hpi

2. Host gene expression cumulatively fluctuated during the first 24 hours of RCNMV infection

I monitored host transcription profile during the first 24 hours of RCNMV infection. Host genes were studied at four different time points; 2, 6, 12, and 24 hours post virus inoculation (hpi). I found that host gene expression fluctuated at all time points of study as shown in volcano plots (Figure 11). All 13,413 host genes exhibited a unique expression pattern during the first 24 hours of RCNMV infection. Within the first 2 hours, the majority of host genes asserted a dramatic down-regulation pattern. But then, at 6 hpi, the expression gradually shifted to the opposite direction. By time point 12 hpi, an up-rising expression of the majority of host genes revealed a strong reciprocal modulation, an opposing impact to the first 6 hours. This up-rising modulation persisted until time point 24 hpi. The changes in host gene expression pattern is more obvious and confirmed when only significantly differentially expressed genes were dissected as displayed in Figure 12. When I count the number of host genes that are either up- or down-regulated and whose significance is based on an FDR cutoff of 0.01, I found that 84% of host genes were down-regulated at 2 hpi, but then the number of up-regulated genes started to increase to 35% at 6 hpi. And at 12 hpi, 95% of host genes were up-regulated. The same gene expression pattern re-appeared at time point 24 hpi. This host gene expression pattern is not only observed in this overall significant gene counting method (Figure 12), but also gene counting in functional categories (Figure 24). This suggests that the RCNMV infection process in a primary infected cell took between 12 to 24 hours. After 24 hours of infection, virus started its movement from the primary infected cell to a neighboring cell to initiate a new infection cycle.

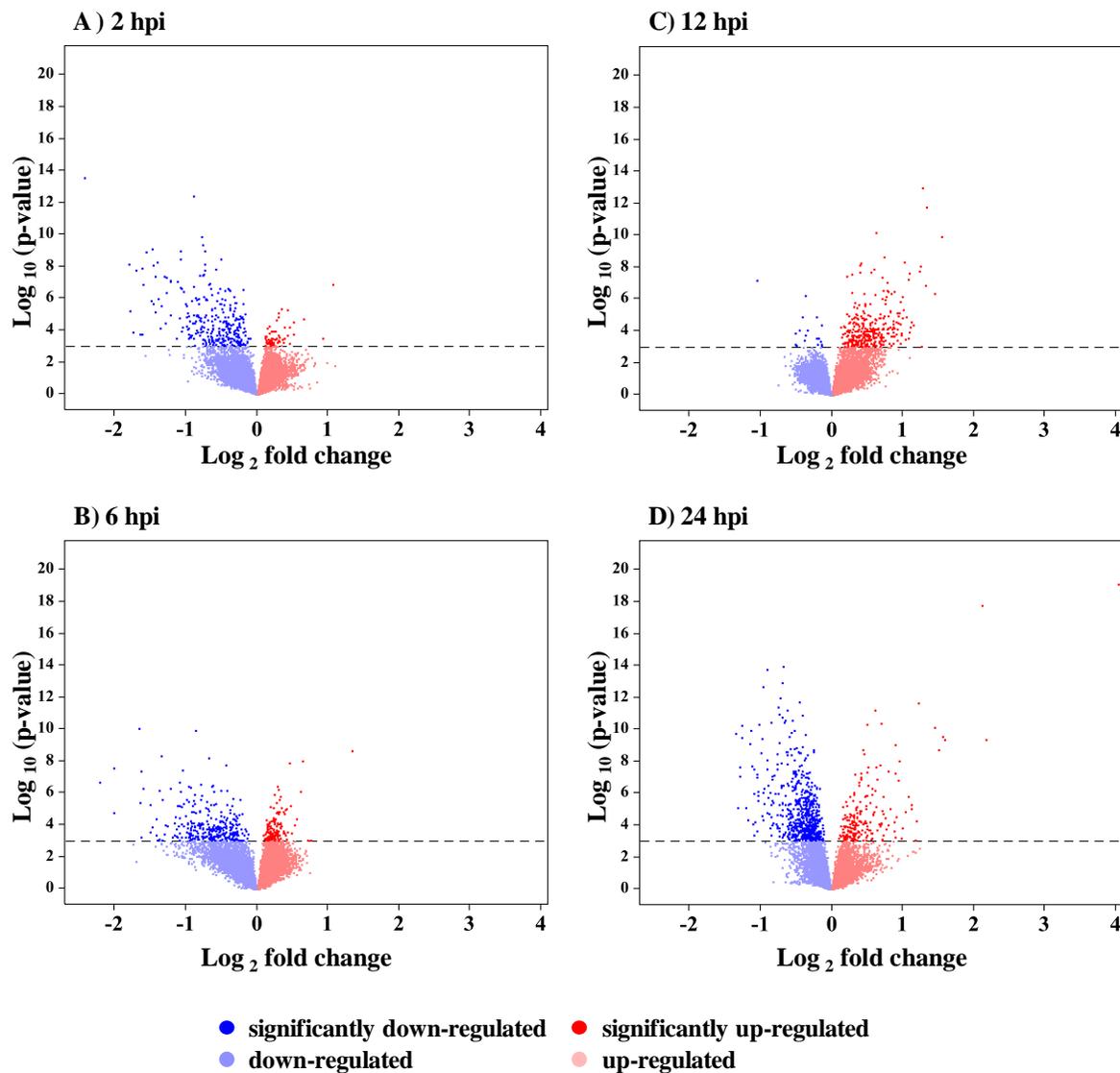


Figure 11 Volcano plots of the microarray data of 13,413 host genes shows the cumulative changes in gene expression at 2, 6, 12, and 24 hours post inoculation. A fold change of each host gene was computed from a subtraction between microarray expression data of RCNMV-inoculated plants (v) and mocks (m) at each time point of study. The dashed line is an FDR cutoff ($-\log_{10}(p) = 3.04$) of 0.01.

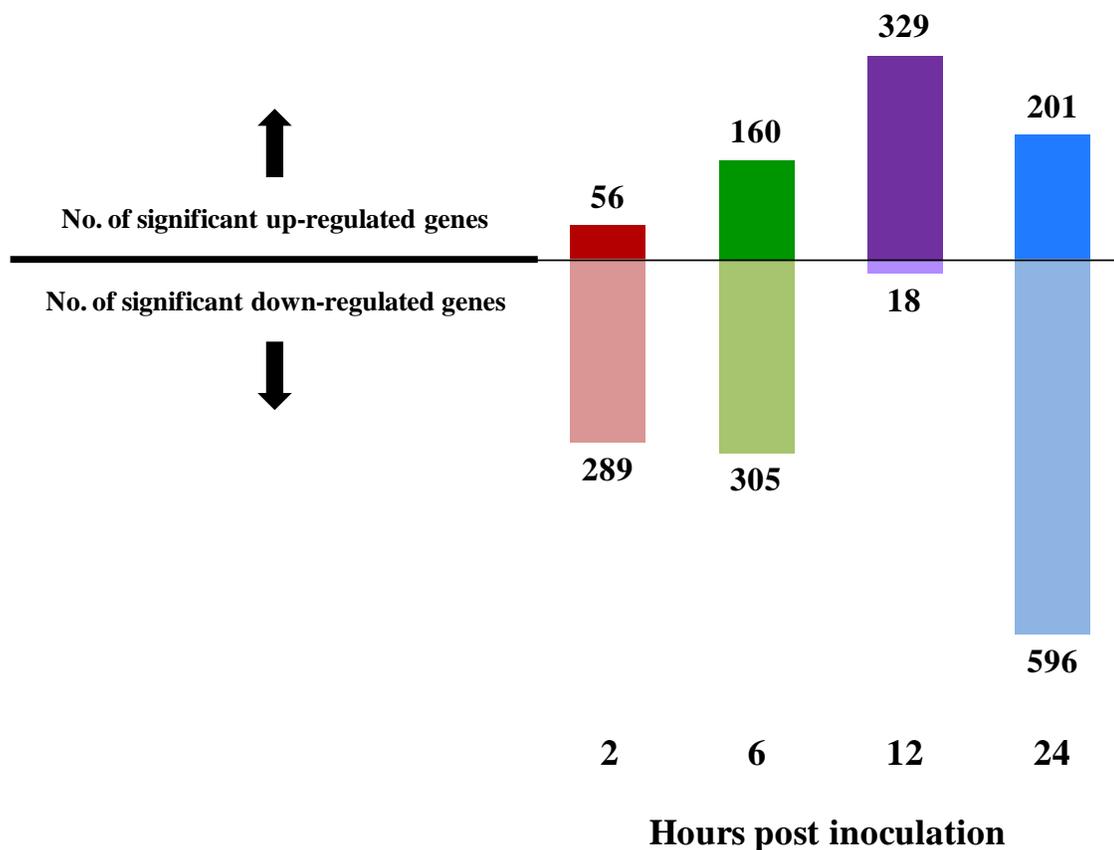


Figure 12 Total numbers of significantly differentially expressed host genes from the microarray data analysis at 2, 6, 12, and 24 hours post inoculation, with an FDR cutoff of 0.01. A differential expression of each host gene was computed from a subtraction between microarray expression data of RCNMV-inoculated plants (v) and mocks (m) at each time point of study.

3. Switch mode of host gene expression at 12 hour RCNMV infection

I used hierarchical clustering method [86, 87] to study how each host gene behaved over 24 hour RCNMV infection. Host gene expression, gene by gene at a time, was analyzed across four time points of study at 2, 6, 12, and 24 hpi. This method grouped host genes that had similar expression behavior into the same cluster, and so clusters collectively revealed an

overall host gene expression pattern (Figure 13) [87]. Most genes maintained the same expression behavior from 2 to 6 hpi but then dramatically shifted its expression to the other direction at 12 hpi. The switched modulation of host gene expression by RCNMV infection was particularly prominent at 12 hpi, and thereafter subsided by 24 hpi. The switch mode event at 12 and 24 hpi was also concurrent with a dramatic accumulation of RCNMV RNA-1 and RNA-2 transcripts after 12 hours (Figure 25 and 26).

Evidence from 2) and 3) support the main finding that the majority of host genes are shutdown at the earliest state, approximately within 6 hours, of RCNMV infection; but then host gene expression appears to recover after 12 hours. This plant host gene shutdown was reminiscent of commonly observed phenomenon in animal virus infection, known as host-gene “shut-off”. Animal system shut-off is achieved at diverse points in host gene expression, and has been widely interpreted as a mechanism to place virus gene expression at a competitive advantage over the expression of host genes [88]. A similar decline in host gene expression was also found to be modulated by other plant viruses. Most important works of these plant host gene shut-off studies were done by Andrew Maule’s research group (John Innes Centre, Norwich, UK), using a technique called spatial analysis [88, 89]. This technique was developed to measure host gene transcripts at the advancing infection fronts between infected and uninfected cells/areas using in situ hybridization (to detect host transcripts), immunocytochemistry (to detect viral protein) and tissue blotting technique. With this spatial analysis, they found host gene shut-off in pea host by *Pea seed borne mosaic potyvirus* (PSbMV) [90], and in marrow host (*Cucurbita pepo*) by *Cucumber mosaic virus* (CMV) [53]. However, this spatial analysis can only be applied to study one host gene

at a time. So far, in plant viruses, the study presented in this chapter is the first report of whole plant transcriptome shut-off at the earliest time point of RCNMV infection as within 6 hours, using a temporal (time course study) rather than a spatial analysis. I observed that the coordinate response in the depletion of host transcripts covers a wide-ranged loss of host metabolism functions (Figure 24). This host shut-off phenomenon could potentially cause plant death. But why does RCNMV induce such a harsh pressure on host? Could it be a strategy that RCNMV employs to take over host translation machineries for viral protein production? In the case of *Turnip mosaic virus* (Potyviridae), it has been proposed that the binding of the viral protein known as virus genome-linked protein (VPg) to host eIFiso4E (eukaryotic initiation factor involved in the initiation phase of eukaryotic translation) competitively inhibits the expression of host genes [91]. This interaction is a proposed mechanism that *Turnip mosaic virus* employs to decrease host mRNA translation so that it can outcompete host essential translation factors. Messenger RNAs not recruited for translation are usually rapidly degraded [92]. In my study, the simultaneous depletion of many host transcripts by RCNMV might suggest that active RNA degradation plays a part in suppressing the expression of host genes.

I assume that a reduction of host transcripts leads to a reduction of host translation of their own proteins. I speculate that the induction of host transcript degradation is a central strategy that RCNMV employs to manipulate and shift host translation machineries to produce viral protein. I note that our experiment only measured host changes at the transcript level, not the translation level. To test this hypothesis, protein quantification and biochemical assays are necessary.

If host gene shut-off is modulated by virus at the risky expense to potentially cause plant cell death so that virus can successfully produce viral replication protein to initiate replication process, how do some host infected cells survive after viral egress? In my study, interestingly, I found that the plant host was allowed to recover at 12 hours as illustrated by the “switch mode” effect. This indicates the transient nature of the host shut-off which explains how host cells are able to survive and maintain their function after the virus replication phase is complete.

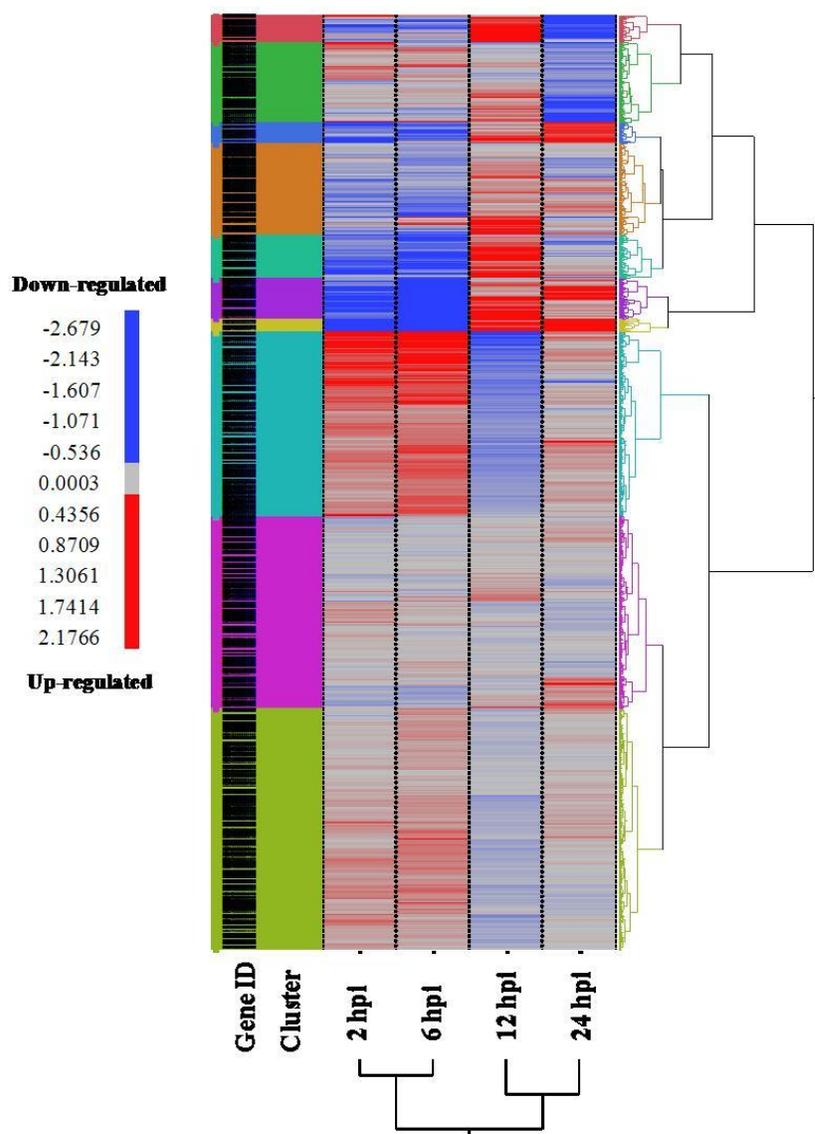


Figure 13 Hierarchical clustering the differential microarray expression of 13,413 host genes at 2, 6, 12, and 24 hours post inoculation (hpi) reveals a reversal of regulation at 12 hours (a switch mode effect). A differential expression of each host gene was computed from a subtraction between microarray expression data of RCNMV-inoculated plants (v) and mocks (m) at each time point of study.

4. Systemic acquired resistance gene (SAR) expression was modulated over 24 hours RCNMV infection

I used a Venn diagram to illustrate host genes that were significantly differentially expressed overlapping to other time points of RCNMV infection. In this diagram, I found only one host gene that was significantly modulated across all four time points of RCNMV infection (Figure 14). This one host gene is a *N. benthamiana* systemic acquired resistance gene 8.2m (NbSAR8.2m) [93]. This finding suggests that SAR is a key host gene that RCNMV battles throughout the course of its infection.

With microarray and qRT-PCR data (Figure 14), I found that NbSAR 8.2m was significantly down-regulated in the first 6 hours, followed by a significant up-regulation after 12 hours. The report in this chapter is the first study that shows host SAR gene suppression at the earliest state of viral infection. I hypothesize that SAR was down-regulated in the first 6 hours of infection to presumably allow the establishment of virus in the primary infected cell. Once the virus produces a successful infection, the plant started to recognize the damage and started sending out the defense signal to the healthy cells which could be neighboring cells or cells located at the distal distance from the primary infected cell; as a result, SAR expression started to up-regulated after 12 hour of infection.

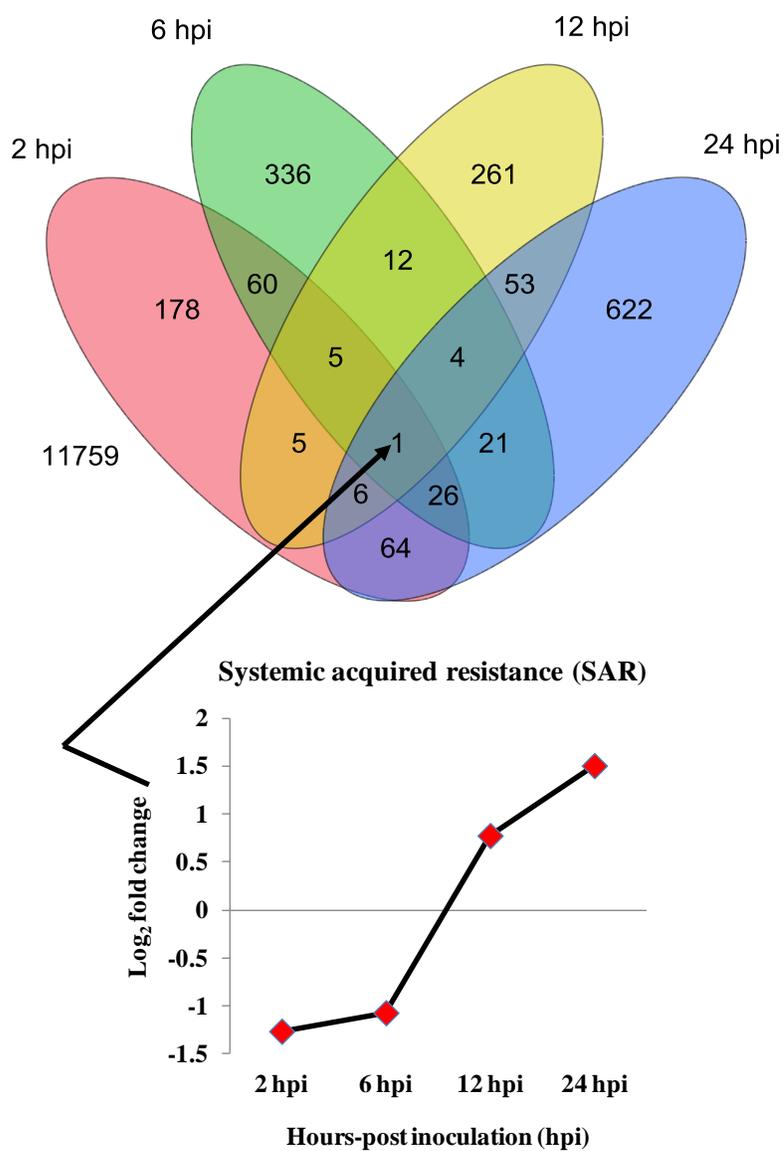


Figure 14 Venn diagram displays the microarray data of the overlapped host genes that were significantly differentially modulated at 2, 6, 12, and 24 hours post inoculation (hpi) at an FDR cutoff of 0.01

5. Functional analysis to determine the impacts of an early RCNMV infection on host physiological process

This study focused on patterns of expression among different functional categories of host genes in order to identify key host biological pathways or host gene functions that were significantly correlated with a course of RCNMV infection (duration of infection at 2, 6, 12, and 24 hours post inoculation). *N. benthamiana* genes spotted on microarray were assigned functionalities by searching their sequence similarity against GenBank [61], Gene Ontology (GO) [64], MIPS functional catalogue (FunCat) [68], and KEGG [67] databases. Sequence alignment method was based on the BLAST algorithm (basic local alignment search tool) and E-value cutoff at 1×10^{-5} [94]. Although these genomic databases were powerful and convenient for functional annotation search, each had its own limitation; for instance, available organism genomic data, latest information, number of known annotations, annotation sources (did the annotation sources come from a computational analysis or an experimental analysis?). For these different limitations, more than one database was used in our microarray functional annotation.

A direct counting number of significant up- or down- regulated genes that were in the same functional categories was the most straightforward approach to identify a correlation between host functions and virus infection event (Figure 24). However, upon my initial gene functional analysis, it became apparent that existing functional terms that were assigned on our microarray genes, in many cases, were incomplete, very general, and offered little and ambiguous insight into biological mechanisms. Most genes were assigned to major or general

functional groups rather than specific functions. This was due to a database limitation. It was not clear whether any of these terms represented significant enrichments in virus-infected plants. Therefore, a direct gene counting and use of these functional terms to perform a comparable functional analysis among host genes that were spotted on the array could cause a bias and false interpretation. I used GSEA (Gene Set Enrichment Analysis) [95] to minimize this potential problem. GSEA is a statistical-based method and it provides more advantages than a simple counting number of significant genes in each functional category. It took a consideration of all available functional terms of all genes spotted on the microarray (whether or not they were significantly regulated in an early RCNMV infection) and their associated t-statistic values (derived from a differential expression analysis between virus- and mock- inoculated plants) into a functional analysis. GSEA applied a permutation statistical method to compute an enrichment score of each host gene function based on the t-statistic values of the gene members that were assigned to that particular function in order to define if the function being analyzed was significantly enriched and correlated in virus-infected plants [95]. Therefore, a major difference between GSEA and a simple gene counting was that all genes and all functional terms that were associated with them were analyzed, rather than only significantly regulated genes. And as a result, GSEA provided overall information of how host pathways (or functions) were statistically significantly changed and modulated in an early RCNMV infection. GO terms were used to conduct GSEA analysis. The resulting top 20 significant GO terms, both positive and negative correlation enriched at each time point of RCNMV infection (Table 6 to 9), served as the

main context to interpret how the host interacted with RCNMV at the different stages of the viral life cycle in this study, as described in the following detail.

5.1) Host defense responses

Host defense response (GO 42742), particularly fighting against a bacterial infection, was negatively correlated to RCNMV infection in the first 6 hours (2 and 6 hpi), but then was positively correlated at 12 hpi (Figure 15). This indicated that host defense gene expression was suppressed in the first 6 hours of RCNMV infection but then de-suppressed at 12 hpi. The report in this chapter is the first study that shows host defense suppression at the earliest state of viral infection of 2 hours. Decreasing host defense gene expression in the early 6 hours was likely driven by RCNMV infection, which was expected in a susceptible *N. benthamiana* host.

The early host defense gene suppression may play a crucial role for a successful virus infection. However, during this host defense gene suppression, apoptosis-related host activities were increased (Figure 15), suggesting that the host utilized the apoptosis pathway (GO 6915) to immediately confine the infected area. This finding lead to my belief that host apoptosis was a primary and acute host defense response in *N. benthamiana* against RCNMV infection at the earliest time point of 6 hpi. However, in a longer defense response, the host shifted its defense strategy to increasing defense gene expression (GO 42742) which is more efficient than apoptosis and does not cause host cell death.

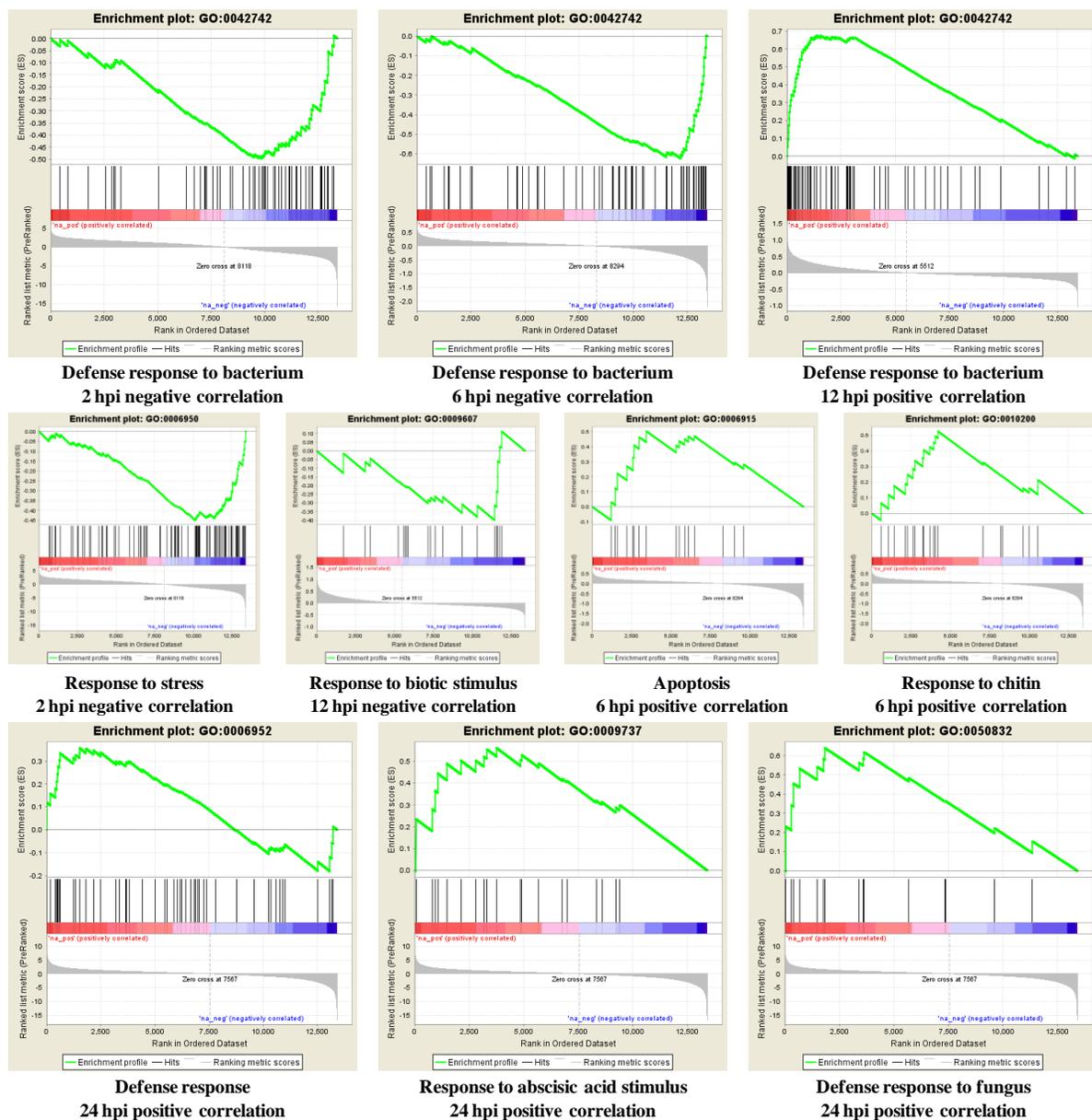


Figure 15 GSEA enrichment plots display host defense-related functions that were highly correlated to RCNMV infection at different time points

5.2) Host multidrug and toxin extrusion (MATE) transporter

MATE transporters are found in prokaryotes and eukaryotes and represent conserved protein families [96, 97]. In bacteria, MATE transporter is a major cause of multidrug resistance (MDR). Mammalian MATE transporters enhance the efflux of cytotoxic drugs from cells [98, 99]; in addition, they're involved in the immune system and in apoptosis [100]. Plants have the largest number of *MDR* genes. *Arabidopsis* possesses at least 58 known *MDR* genes related to animal *MDR* genes [96]. *MDR*-like genes of *Arabidopsis* have been found to be involved in the detoxification of endogenous secondary metabolites and xenobiotics [101]. In addition, they are required for auxin transport and auxin-mediated development. One of these is *AtMDR1* which was shown to be induced by the hormone auxin [102]. Other than detoxification and plant developmental roles, plant MATE-like proteins are also involved in pathogen resistance. *Arabidopsis* EDS5 (ENHANCED DISEASE SUSCEPTIBILITY 5), a homolog of the MATE transporter family, is an essential component of salicylic acid-dependent signaling for disease resistance [103]. Salicylic acid (SA) is synthesized in chloroplasts [104, 105]. Recently, it was found that EDS5 is a MATE-like SA transporter located at the chloroplast envelope. Mutants with impaired EDS5 function are unable to export SA from the chloroplast to the cytoplasm and so SA is trapped in the chloroplast, leading to an inhibition of SA synthesis which might be explained by a possible auto negative feedback. Thus export of salicylic acid from chloroplasts requires EDS5 transporter [106].

With so many diverse protein isoforms attributed to the products of *MDR* genes, it remains an experimental challenge to characterize functions of the *MDR*-like genes present in the *Arabidopsis* genome. And so far, no report has shown an involvement of MATE transporter in any plant viral infection. My functional analysis found that an activity of the *N. benthamiana* homolog of MATE protein was significantly enriched and increased in the first 6 hours of RCNMV infection (Figure 16). The transcript level of the MATE-like gene was also immediately accumulated after 2 hours of RCNMV infection but not significantly until after 6 hours. This leads to an assumption that an increase in MATE activities and its transcript accumulation are involved in viral resistance, possibly through a mode of action similar to *Arabidopsis* EDS5 as a facilitator of SA transport.

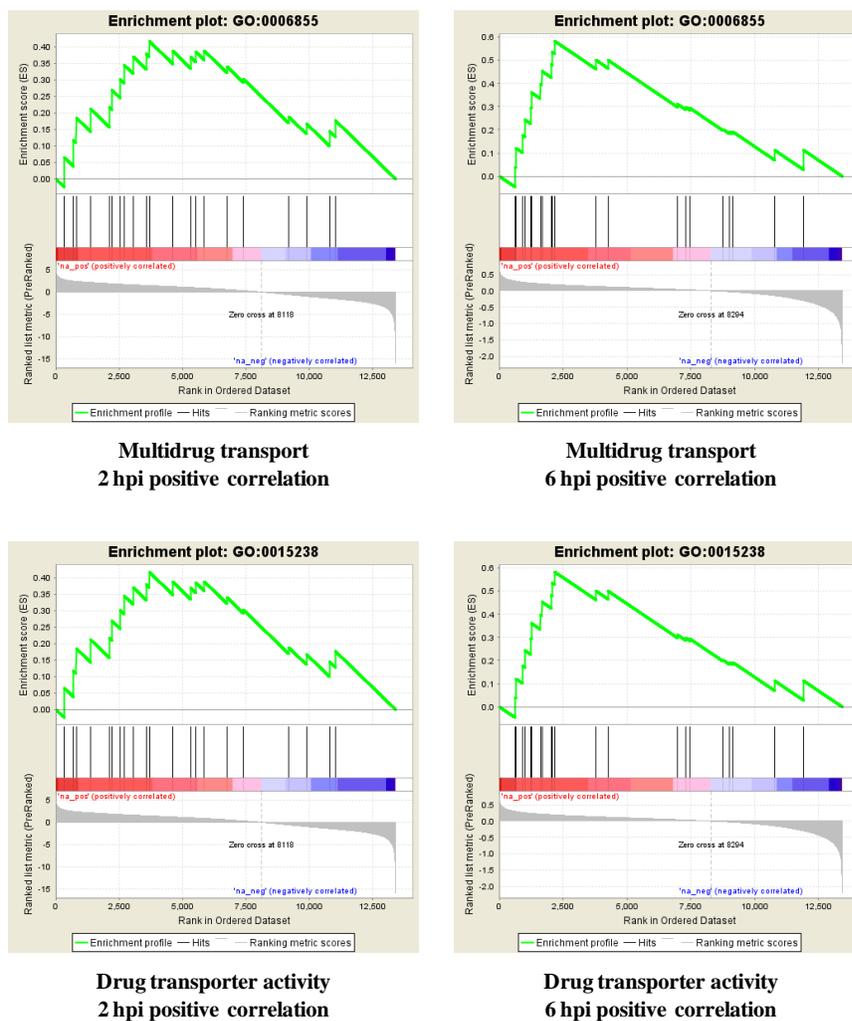
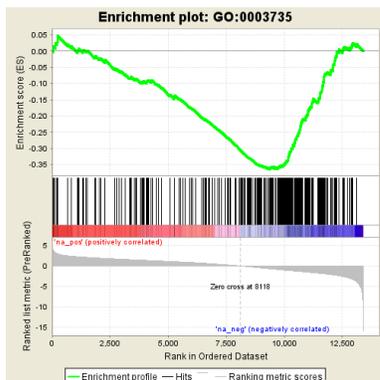


Figure 16 GSEA enrichment plots display host multidrug and toxin extrusion (MATE)-related functions that were highly correlated to RCNMV infection at different time points

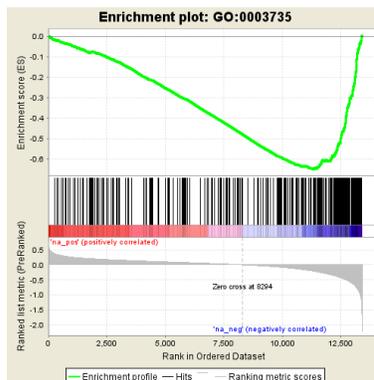
5.3) Host translation process and machinery

RCNMV is a positive sense, single stranded RNA plant virus and it does not carry any replication proteins as part of the virion. Therefore, RCNMV is required to first translate replication proteins from its own genome to initiate the viral replication process. RCNMV does not have its own translation machinery. My hypothesis is that RCNMV co-opted the host translation machinery at the very early time points of infection to facilitate its translation of replication proteins. As expected, GO terms related to host translation machinery (GO3735, GO5840, GO6412, GO6414, GO6448, GO22625, and GO42254) were altered during an early RCNMV infection. However, overall host translation was decreased in the first 6 hours of RCNMV infection (Figure 17). An increase in host translation elongation and ribosome biogenesis was not observed until 12 hours of RCNMV infection. The delay in overall host translation may implicate the host gene shut-off phenomenon that was discussed earlier in the results and discussion (2) and (3).

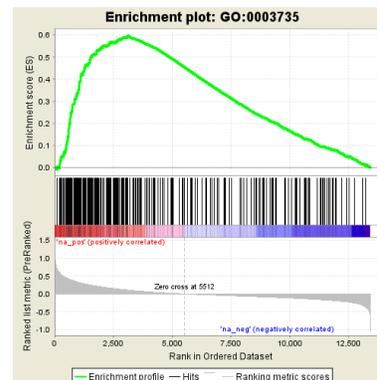
Figure 17 GSEA enrichment plots display host ribosome/translation-related functions that were highly correlated to RCNMV infection at different time points



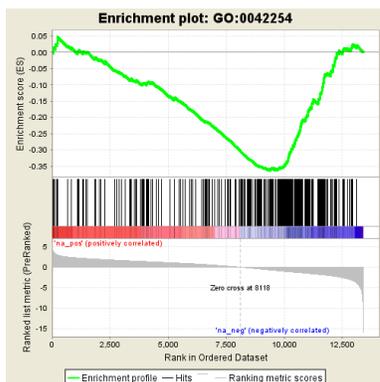
**Structural constituent of ribosome
2 hpi negative correlation**



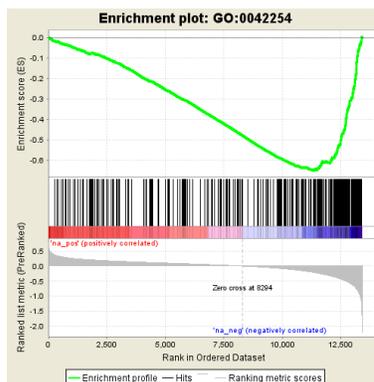
**Structural constituent of ribosome
6 hpi negative correlation**



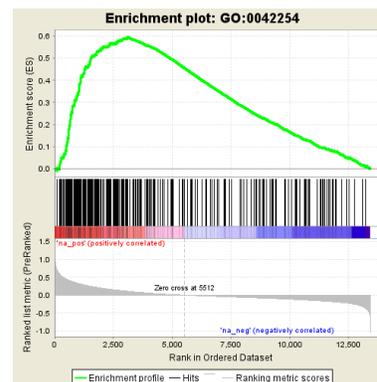
**Structural constituent of ribosome
12 hpi positive correlation**



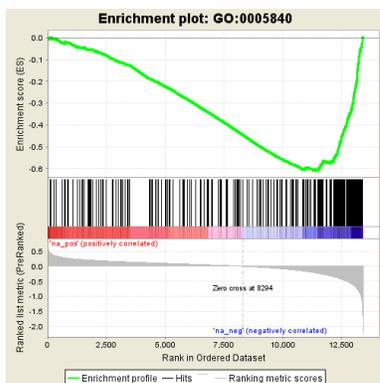
**Ribosome biogenesis
2 hpi negative correlation**



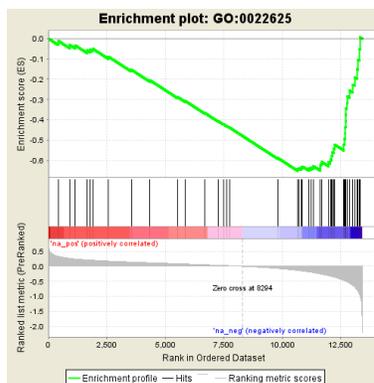
**Ribosome biogenesis
6 hpi negative correlation**



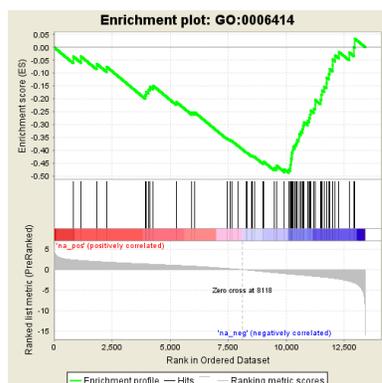
**Ribosome biogenesis
12 hpi positive correlation**



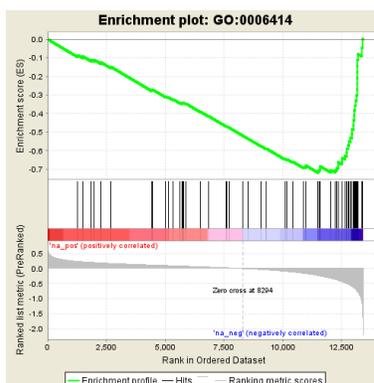
**Ribosome
6 hpi negative correlation**



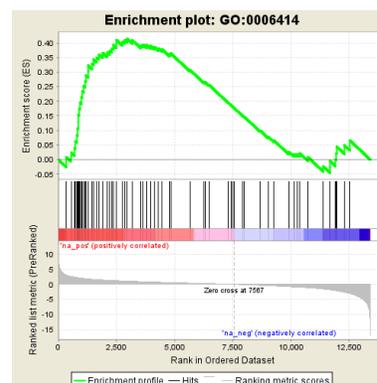
**Cytosolic large ribosomal subunit
6 hpi negative correlation**



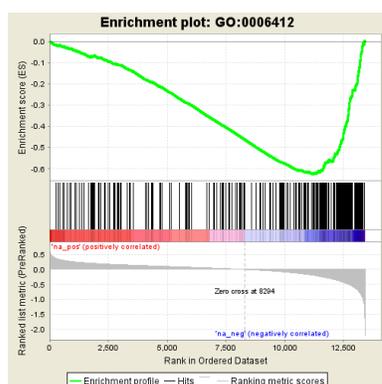
Translation elongation
2 hpi negative correlation



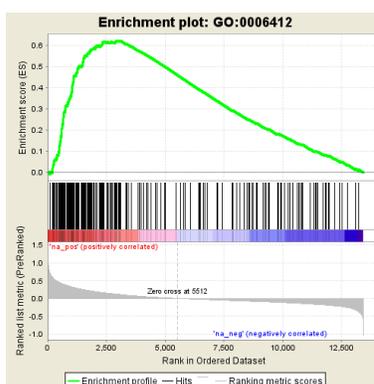
Translation elongation
6 hpi negative correlation



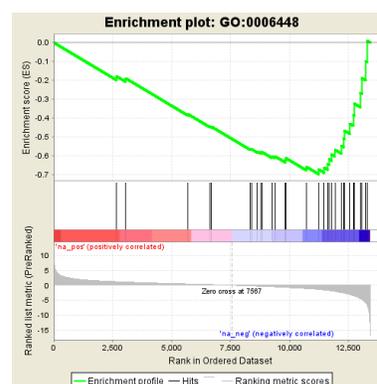
Translation elongation
24 hpi positive correlation



Translation
6 hpi negative correlation



Translation
12 hpi positive correlation



Regulation of translation
24 hpi negative correlation

5.4) Host photosynthesis process and chloroplast-related functions

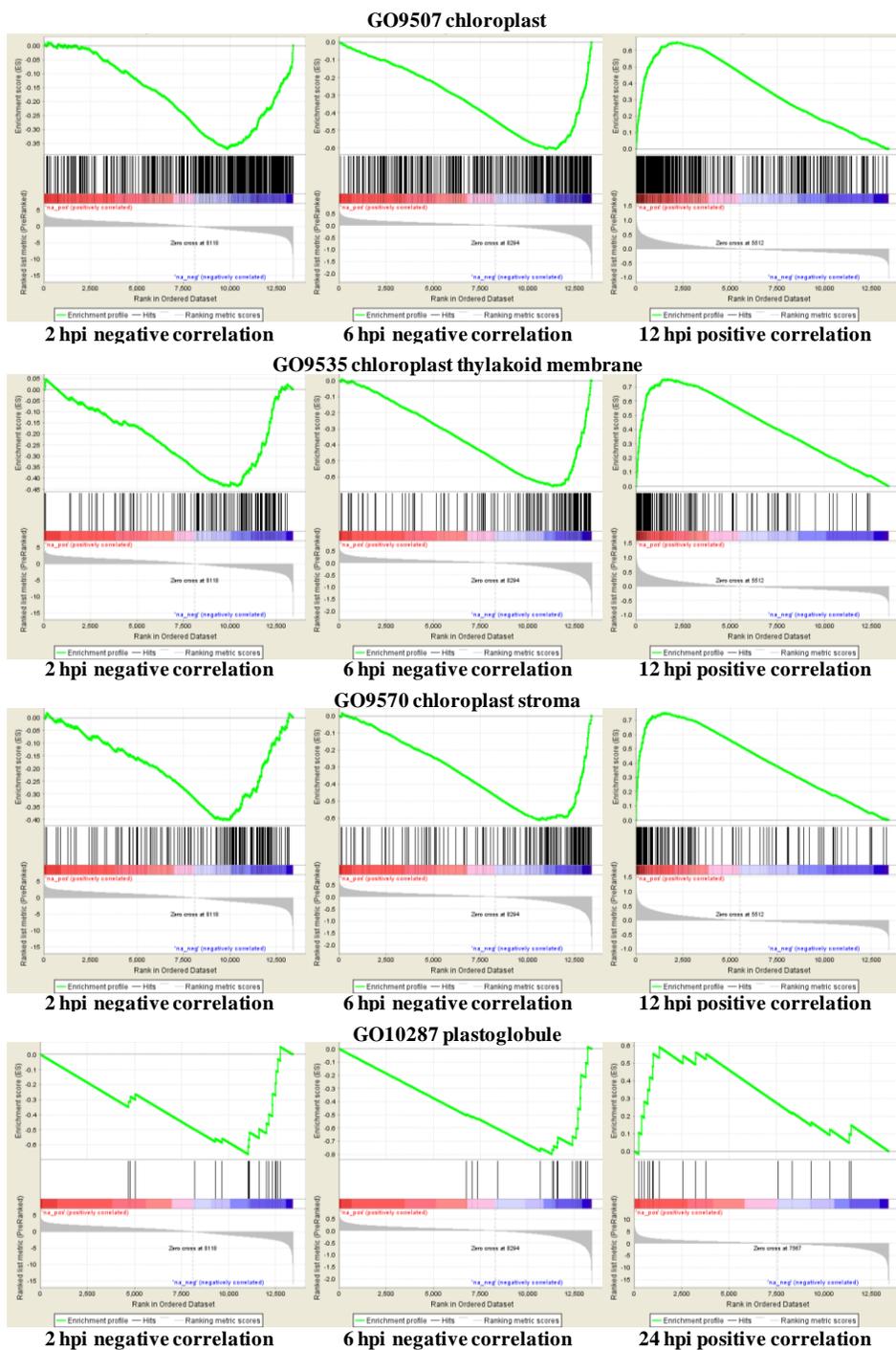
Several plant metabolic processes take place within chloroplasts, including photosynthesis, carbon fixation, amino acid biosynthesis, fatty acid biosynthesis, and carbohydrate metabolism [107]. Chloroplasts also have a critical role in plant immunity as a site for the production of salicylic acid (SA) and jasmonic acid (JA), two important mediators of plant immune system [104-106].

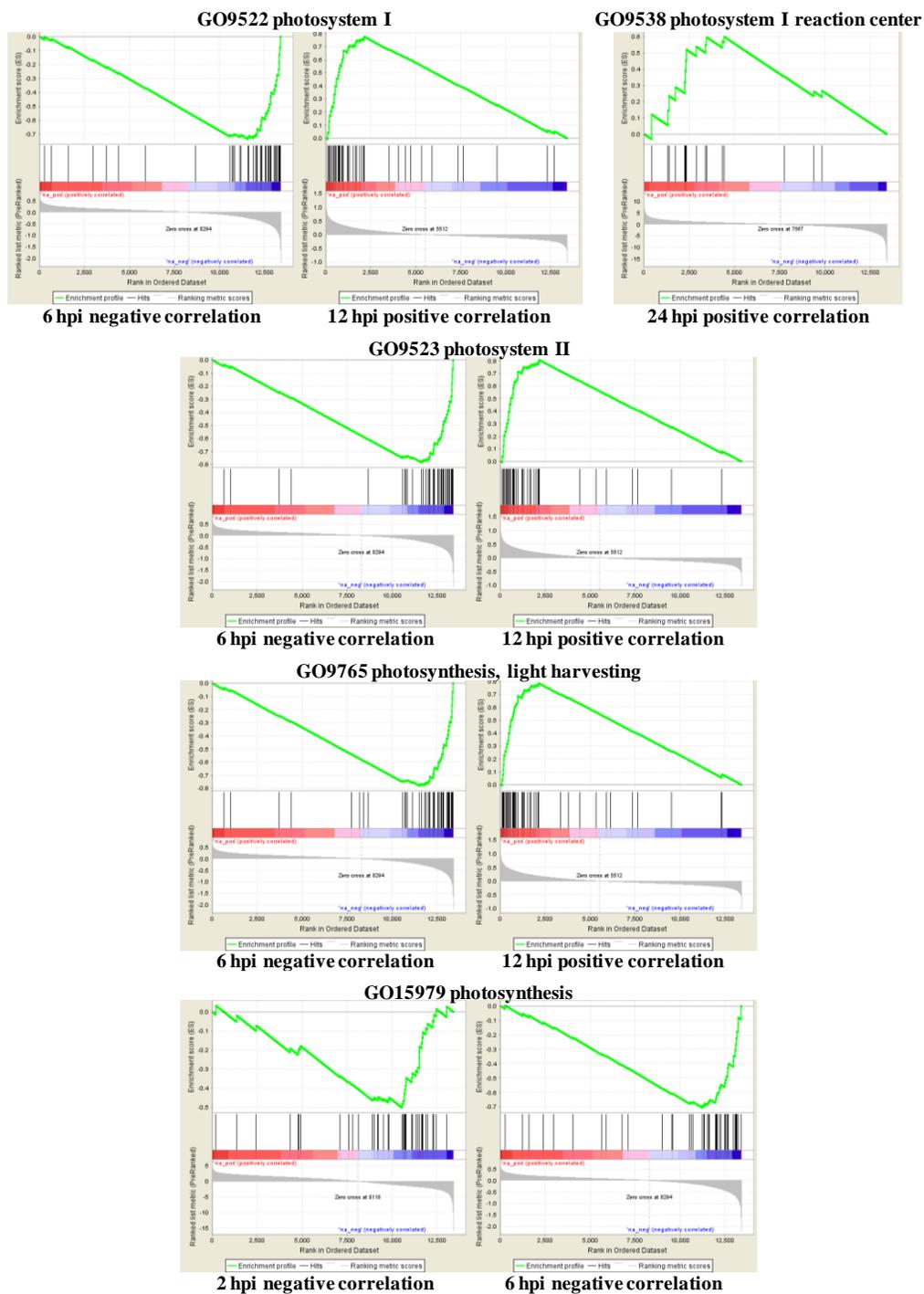
Functional analysis using GSEA revealed that host functions that take place in the chloroplast were highly enriched and significantly correlated with an early RCNMV infection (Figure 18). Interestingly, a majority of the host functions in the top GSEA results enriched across all four time points of study are related to chloroplast activities. Host activities occurring in chloroplast thylakoid membrane and stroma were decreased in the first 6 hours of RCNMV infection, but were later increased at 12 hours. Likewise, photosynthesis and light harvesting activities were suppressed in the first 6 hours but were then increased at 12 hours. Photosystems I and II activities were both down at 6 hpi but followed by up-regulation after 12 hours. And photosystem I maintained its increasing activity until 24 hours. Surprisingly, an increase in photorespiration (GO9853) was observed at 24 hpi. The photorespiratory pathway is one major source of H₂O₂. Through production of these highly reactive oxygen species, photorespiration makes a key contribution to cellular redox homeostasis. In doing so, it influences several signaling pathways, in particular, those that govern plant defense response and programmed cell death [108, 109], suggesting that its

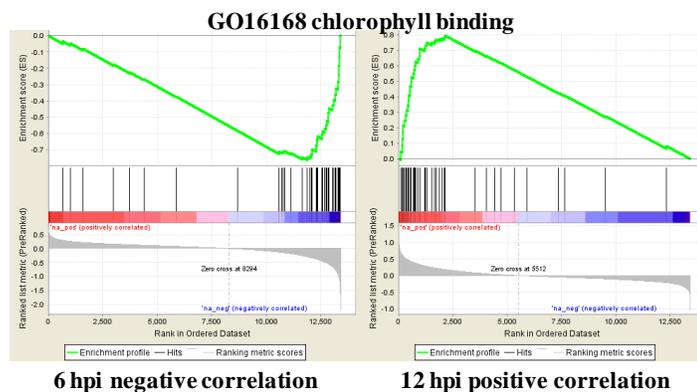
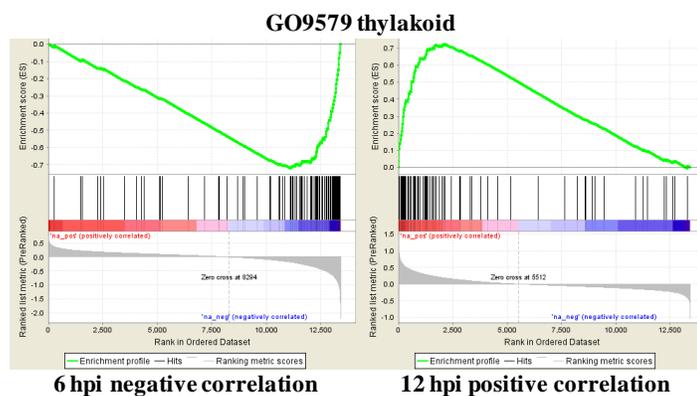
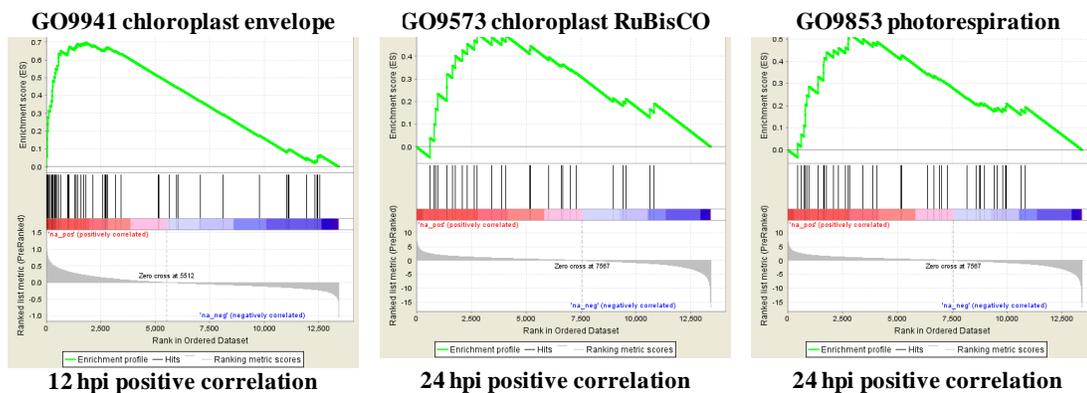
increasing activity at 24 hours of RCNMV infection may take a part in host defense response during the movement of virus to the adjacent cells.

Overall, host physiological processes that occurred in the chloroplasts were suppressed as early as 2 hpi and the suppression continued until 6 hours. Due to the fact that the chloroplast is a center for many primary metabolite syntheses as well as an important plant organelle for producing defense signaling molecules, the sudden shutdown of chloroplast activities by RCNMV implied that RCNMV hijacked chloroplasts as one of the major targets at the very early stage of the infection process. This impact could cause dramatic changes in plant physiology to ultimately transform the host cell environment to favor the viral infection process.

Figure 18 GSEA enrichment plots display host chloroplast/photosynthesis-related functions that were highly correlated to RCNMV infection at different time points







5.5) Host carbon fixation and Calvin cycle

The Calvin cycle, also known as the reductive pentose-phosphate pathway or C3 cycle, is a series of biochemical redox reactions that results in a conversion of carbon dioxide into triose phosphate (TP) known as glyceraldehydes-3-phosphate (G3P) which is subsequently used as a substrate to synthesize organic molecules; for instance, carbohydrates such as glucose, sucrose, cellulose, starch, etc. The Calvin cycle occurs in the stroma, which is the area of the chloroplast surrounding the thylakoid membranes [107]. GSEA revealed that host functions involved in the Calvin cycle were highly enriched and significantly correlated to RCNMV infection across all stages of the infection process (2, 6, 12, and 24 hpi) (Figure 19).

The Calvin cycle was down-regulated in the first 6 hours of RCNMV infection, but then was up-regulated after 12 hours. GSEA also indicated that host function related to RuBisCO (GO 9573), an important enzyme in the Calvin cycle, was increased at 24 hpi (Figure 18). With the use of a KEGG diagram of “carbon fixation in photosynthetic organisms (KO 00710)”, I was able to look at the microarray expression data of individual host genes that encode enzymes in the carbon fixation pathway (Figure 20). I found that the majority of these host genes were down-regulated in the first 6 hours, followed by the up-regulation at 12 hours. I also performed hierarchical clustering of the microarray expression data of these host genes (Figure 21) [87]. According to the clustering results, I found that the expression of host genes encoding the enzymes in the carbon fixation were dramatically down-regulated in the first 6 hours and up-regulated at 12 hours, but then a mixed direction

of expressions were seen at 24 hours, indicating the reappearance of host gene shut down at 24 hours. The expressions of RuBisCO-related genes were also down-regulated in the first 6 hours; however, they were able to maintain their up-regulation from 12 to 24 hours. With these consistent results, the delay in the Calvin cycle was an indicator that RCNMV prevented the host from being able to perform carbon fixation, and as a result, RCNMV might indirectly decrease plant biosyntheses of organic compounds, importantly carbohydrate biosyntheses.

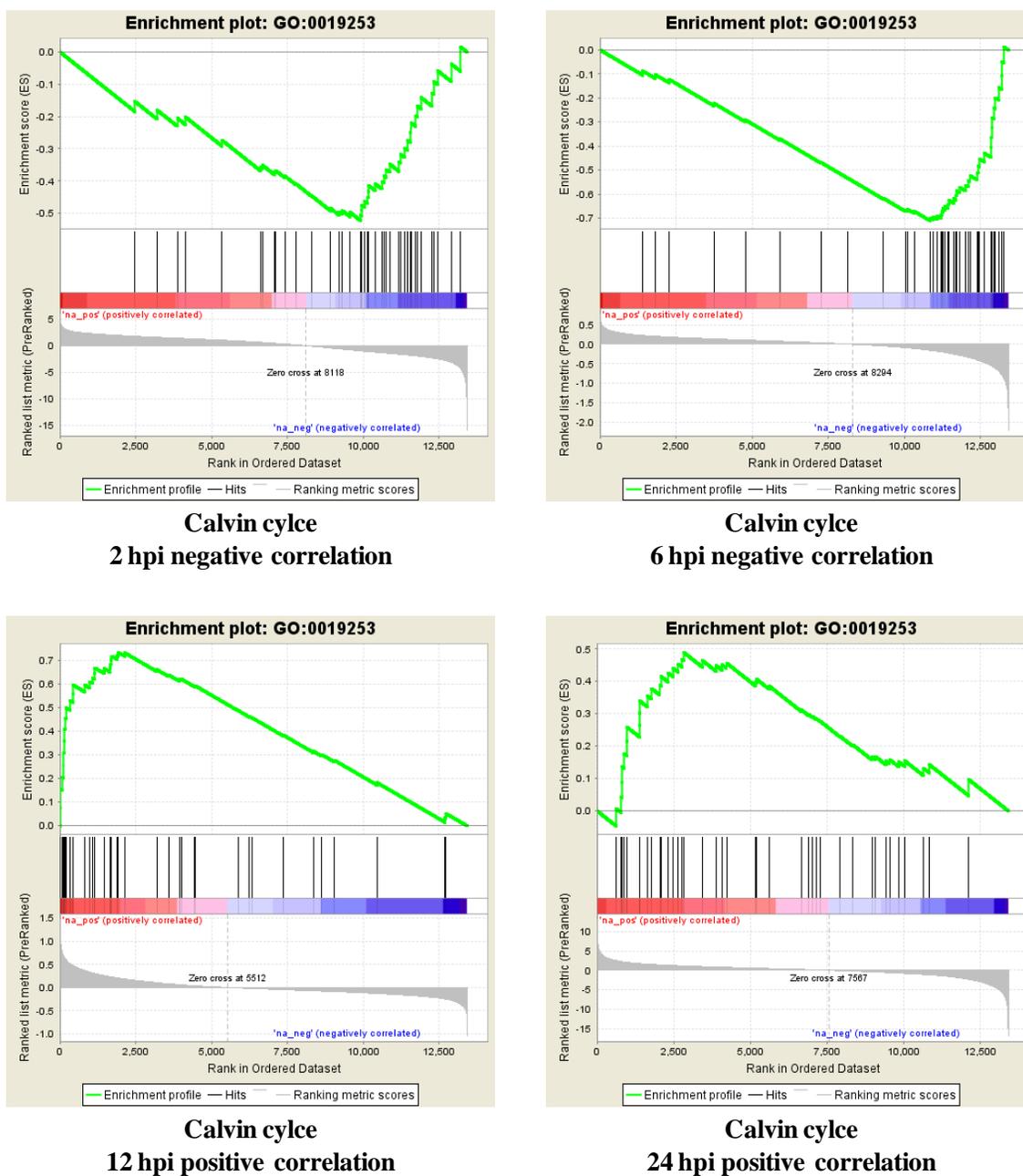


Figure 19 GSEA enrichment plots display host Calvin cycle-related function (GO19253) that was highly correlated to RCNMV infection at different time points

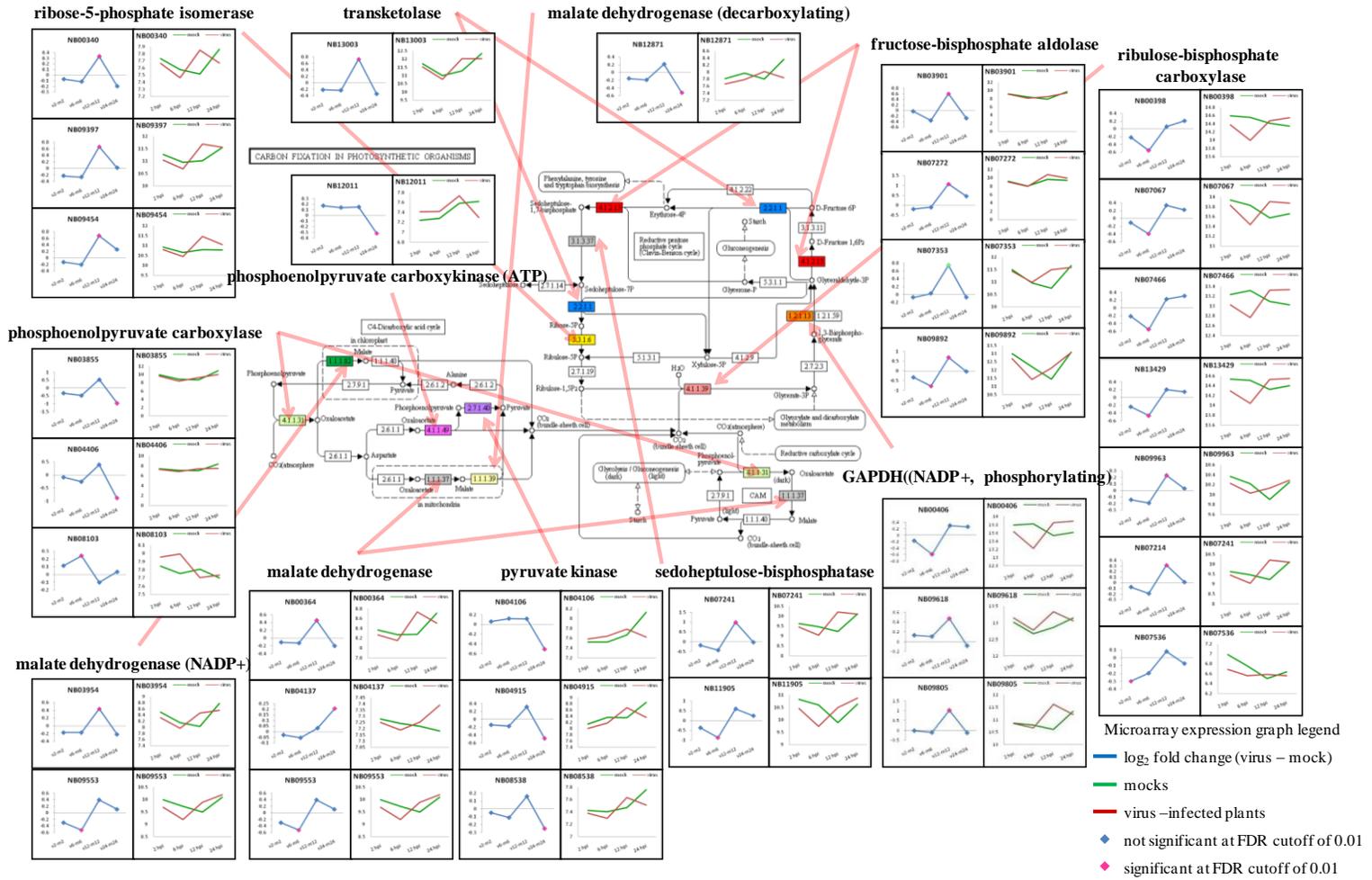
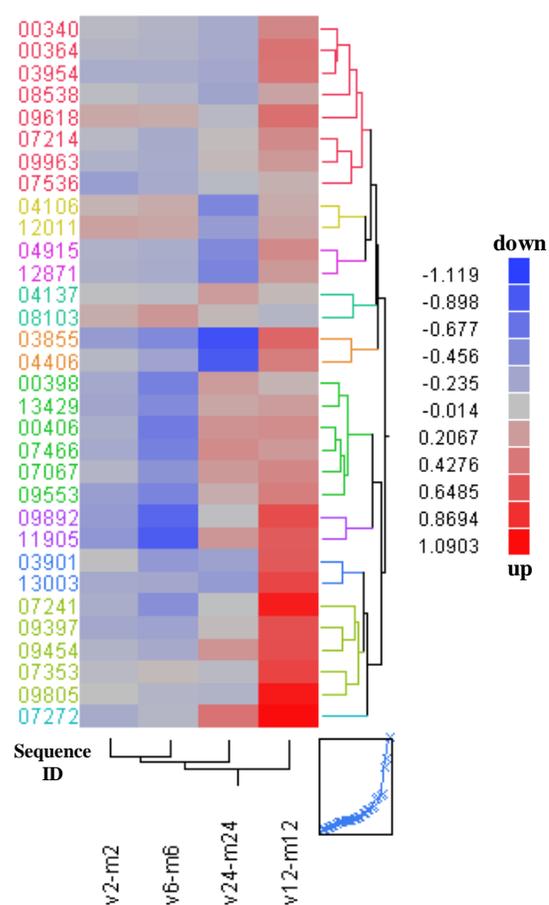


Figure 20 KEGG diagram of “carbon fixation in photosynthetic organisms (KO 00710)” and microarray expression data

A) Carbon fixation-related genes



B) RuBisCO (ribulose-bisphosphate carboxylase)

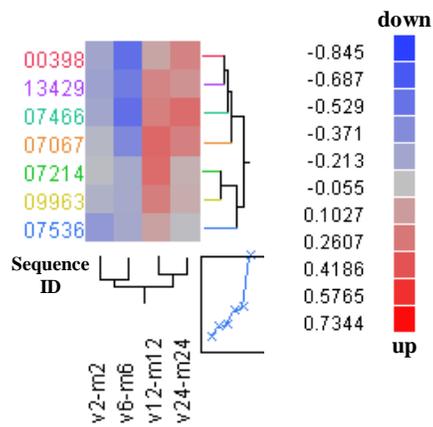


Figure 21 Hierarchical clustering of the microarray expression data of host carbon fixation- and RuBisCO- related genes. The function annotation of these host genes are described in Table 3.

Table 3 KEGG functional annotation of *N. benthamiana* host genes in “carbon fixation in photosynthetic organisms (KO 00710)”. The microarray expression data of these host genes and the carbon fixation KEGG diagram are displayed in Figures 20 and 21.

Sequence ID	Enzyme name	Enzyme ID
340	ribose-5-phosphate isomerase	5.3.1.6
364	malate dehydrogenase	1.1.1.37
398	ribulose-bisphosphate carboxylase	4.1.1.39
406	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)	1.2.1.13
3855	phosphoenolpyruvate carboxylase	4.1.1.31
3901	fructose-bisphosphate aldolase	4.1.2.13
3954	malate dehydrogenase (NADP+)	1.1.1.82
4106	pyruvate kinase	2.7.1.40
4137	malate dehydrogenase	1.1.1.37
4406	phosphoenolpyruvate carboxylase	4.1.1.31
4915	pyruvate kinase	2.7.1.40
7067	ribulose-bisphosphate carboxylase	4.1.1.39
7214	ribulose-bisphosphate carboxylase	4.1.1.39
7241	sedoheptulose-bisphosphatase	3.1.3.37
7272	fructose-bisphosphate aldolase	4.1.2.13
7353	fructose-bisphosphate aldolase	4.1.2.13
7466	ribulose-bisphosphate carboxylase	4.1.1.39
7536	ribulose-bisphosphate carboxylase	4.1.1.39
8103	phosphoenolpyruvate carboxylase	4.1.1.31
8538	pyruvate kinase	2.7.1.40
9397	ribose-5-phosphate isomerase	5.3.1.6
9454	ribose-5-phosphate isomerase	5.3.1.6
9553	malate dehydrogenase (NADP+)	1.1.1.82
9553	malate dehydrogenase	1.1.1.37
9618	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)	1.2.1.13
9805	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)	1.2.1.13
9892	fructose-bisphosphate aldolase	4.1.2.13
9963	ribulose-bisphosphate carboxylase	4.1.1.39
11905	sedoheptulose-bisphosphatase	3.1.3.37
12011	phosphoenolpyruvate carboxykinase (ATP)	4.1.1.49
12871	malate dehydrogenase (decarboxylating)	1.1.1.39
13003	transketolase	2.2.1.1
13429	ribulose-bisphosphate carboxylase	4.1.1.39

Results from 5.4) and 5.5) strongly suggests the idea that the host activities occurring in the chloroplast are altered by an early RCNMV infection. I should make an important note that my study in this chapter does not investigate the development, structure, or genetic materials of the host chloroplast itself, but this study is rather focused on the expression level of host genes (i) whose functions are related to, and (ii) whose expressed proteins are localized in the chloroplast. Therefore, the report in this chapter is the first observation of host transcript levels showing the modulation of host chloroplast-related genes by an early virus infection.

The disruption in host photosynthesis is also observed in other plant viruses [110-119]. TMV is a model virus that is widely used to study its impact on host photosynthesis. There are two major impacts on host chloroplasts that are caused by plant viruses: (1) host chloroplast development and (2) host chloroplast functions. The most dramatic and perhaps the most fundamental effect of virus infection is the shutdown of the synthesis of chloroplast ribosomal RNAs and of chloroplast proteins. This phenomenon could account for the reduction of chloroplast numbers in virus-infected leaves [111, 120]. We have not yet investigated such an effect in RCNMV-host interactions. The viral symptoms described as mosaics, green islands, and yellows (or chlorosis) are all attributed largely due to the disruption/aberration of chloroplast structure, function, and/or development [120]. RCNMV causes visible mosaic patterns and yellowing at 3-4 days after inoculation (Figure 1). The examination of host chloroplast development during RCNMV-host interaction will be our next challenge. In this chapter, my study demonstrates a dramatic modulation effect of the

expression of host genes related to photosynthetic processes early in the RCNMV infection at sub 24 hours.

In TMV, marked reduction in host photosynthetic activity arises early after inoculation in association with reduced size and number of chloroplasts, reduced chlorophyll content and low efficiency of CO₂ fixation in chloroplasts [110, 115]. The impairment of chloroplast function is often coupled with a presence of viral coat protein in these organelles. There is strong evidence suggesting a close association between TMV and host chloroplasts. With the chloroplast isolation technique [121], both TMV coat protein and TMV RNA have been isolated from Percoll-purified chloroplasts [120, 122]. TMV coat proteins are found associated with chloroplasts and thylakoid membranes of infected leaves [118]. The coat protein of a TMV strain inducing chlorosis was detected inside chloroplasts 3 days after infection and thereafter accumulated at a rapid rate, first in the stroma and then in the thylakoid membranes. On the other hand, the coat proteins of a TMV strain that caused mild symptoms accumulated in chloroplast to lower levels and little coat protein was associated with thylakoids. *In vivo* and *in vitro* measurements of electron transport revealed that photosystem II activity was inhibited in plants infected with an aggressive TMV strain while no reduction was observed in plants infected with the mild strain [112]. Photochemical activities of the whole electron transport chain in isolated chloroplast dramatically declined with the progression of symptoms [114]. As TMV multiplies in the plant, the disease becomes visible as numerous light and dark green islands creating a clear-cut mosaic. The virus content is higher in light spots and it interferes there with chloroplast development and functions [123, 124]. Furthermore, TMV coat protein shares a common antigenic determinant

(an immunological cross-reactivity) and some amino acid sequence homology with the large subunit of host RuBisCO; however, the significance of this observation is unknown [125].

However, contrary to the association between TMV and host chloroplasts, no evidence so far shows the correlation of the host chloroplast and host chloroplast-related activities to RCNMV replication, RCNMV RNAs and RCNMV proteins (viral replication proteins, coat proteins, movement proteins). In our lab, we previously described that RCNMV replication occurs on the endoplasmic reticulum (ER) membranes. We know that RCNMV replication proteins accumulate on ER membranes [47]. Therefore, hypothetically, it is unlikely that host chloroplast activity alteration were caused by RCNMV accumulation on the chloroplast. However, more works is needed to test this hypothesis. We do not yet observe any RCNMV component accumulation on chloroplast, or have determined if there is any chloroplast-localization signaling motif in the RCNMV genome. And since the work in this chapter is only focused on host transcript measurement, further studies are necessary to confirm the biochemical impact of RCNMV on host chloroplast activities. The investigation of the parameters of photosynthesis; for instance ,sensitivity of chloroplasts to inhibitors, activity of RuBisCO, rates of cyclic and non-cyclic photophosphorylation, products of CO₂ fixation, should be applied to evaluate the impact of RCNMV on host photosynthesis.

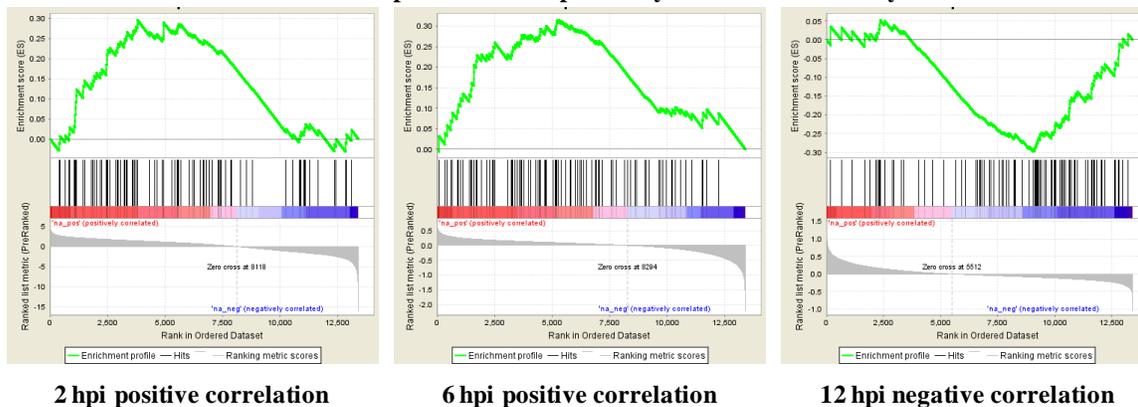
5.6) Host protein kinases

In plants, protein phosphorylations are catalyzed by protein kinases. Plant protein kinases have been implicated in response to many signal transductions, including light,

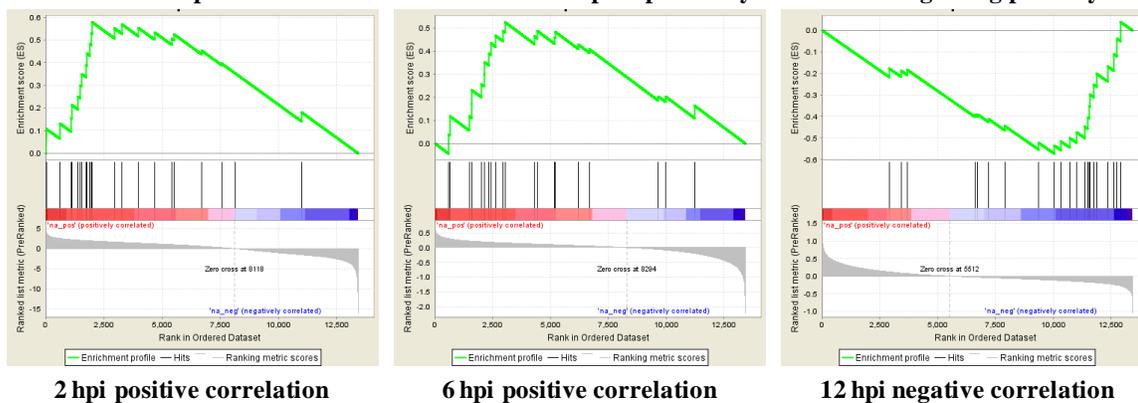
hormones, temperature stress, and nutrient deprivation. Protein kinases also play a central role in signaling during pathogen recognition and the subsequent activation of plant defense mechanisms. With this diversity of functions, there is a broad array of different protein kinases, collectively making up a large superfamily of homologous proteins [126-129].

Our enrichment study found that host homologs of protein kinase functions (GO4713, GO7169, and GO4674) were positively correlated (up-regulation) to RCNMV infection in the first 2 and 6 hours of infection; but later, they were negatively correlated (down-regulation) at 12 hours (Figure 22). An interesting note is that the switch of regulation of host protein kinase expression is against the direction of the majority of host genes, which were down-regulated in the first 2 and 6 hours following by an up-regulation at 12 hours (Figure 11, 12, and 13). Protein kinases govern the regulation of host defense gene expression [128]; however, host defense gene expressions was not increased until 12 hours of infection (Figure 15). This led to my belief that an increasing host protein kinase function in the first 6 hours of infection did not affect host defense gene expression until later on, after 12 hours.

Enrichment plots: GO4713 protein tyrosine kinase activity



Enrichment plots: GO7169 transmembrane receptor protein tyrosine kinase signaling pathway



Enrichment plots: GO4674 protein serine/threonine kinase activity

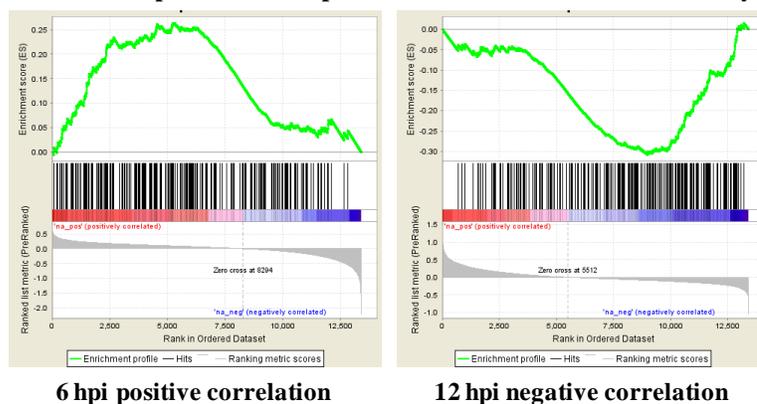


Figure 22 GSEA enrichment plots display host protein kinase-related functions that were highly correlated to RCNMV infection at different time points

5.7) Host cell wall

Cell walls are an important feature of plant cells that provide shape to different plant cell types needed to form tissue and organs [130]. Plant cell walls also form the interface barrier to adjacent cells; therefore, intercellular communication and trafficking tightly depends on cell wall composition and structure. There are two well-characterized pathways known for plant cell-to-cell communication and transportation: apoplastic pathway (through cell) and symplastic pathway (through cytoplasm) [107]. An intercellular trafficking through a symplastic pathway is facilitated by a channel between cells, known as plasmodesmata [38, 48]. Because of their surface location and function as a barrier between cells, plant cell walls play an important role in plant-microbe interactions, including defense response against potential pathogens [131]. Plant cell walls are the first barrier that most pathogens encounter and need to overcome in order to establish a primary infection. They are then also the first barrier that most pathogen progenies need to conquer during their migration to other parts of the plant in order to produce a systemic infection. The most characteristic component found in all plant cell walls is cellulose.

My functional analysis revealed that host functions related to cell wall activities, structural constitutive of cell walls, and cell wall biogenesis (GO5199, GO9505, and GO42546) were significantly increased in the first 6 hours of RCNMV infection, followed by a significant decrease at 24 hours (Figure 23). This result implies that host cell wall activities are modulated as early as 2 hours. The immediate host response of increasing cell wall activities are likely driven by the host itself in order to build a stronger barrier to protect the

plant from the external stress stimuli, particularly viruses in our study. We also observed that the transcript level of the cellulose synthase gene was increased within 6 hours, but this microarray data is not significant (FDR cut off of $\alpha = 0.01$). We did not observe an enrichment of host cell wall activities at 12 hours. We speculate that their reappearance at 24 hours suggests that their function as a barrier between adjacent cells was once again disturbed during the movement of virus from cell-to-cell.

Functional analysis also revealed that host functions related to the apoplastic pathway (GO48046) was significantly activated at 12 hours of RCNMV infection. RCNMV is known to travel through plasmodesmata [38, 132], which is a trafficking channel in the symplastic pathway. If RCNMV is only involved in the symplastic pathway, why was the host apoplastic pathway altered during RCNMV infection? The explanation may lie in the context of the pathogen induced-oxidative burst occurring in the cell wall, which then leads to an apoplast alkalization. The apoplastic alkalization is an important factor supporting disease resistance towards plant pathogens. Previous studies found that massive apoplastic alkalization was induced by fungal elicitors from powdery mildew fungus in barley [133].

During a pathogen attack, local oxidative stress occurred and generated hydrogen peroxide and superoxide anion which can then diffuse through the apoplast and transport a warning signal to neighboring cells. In addition, such an oxidative burst can cause a local alkalization of the apoplast which then travels within minutes to the rest of the plant body via xylem and triggers systemic acquired resistance (SAR) [134, 135]. Our microarray and qRT-PCR analysis confirmed that *N. benthamiana* SAR gene expression was significantly

down-regulated in the first 6 hours of RCNMV infection, but followed by an up-regulation after 12 hours. With this concurrent observation, an increase in host apoplast activities at 12 hours of RCNMV infection suggests they are responsible for the onset of SAR activation at 12 hours.

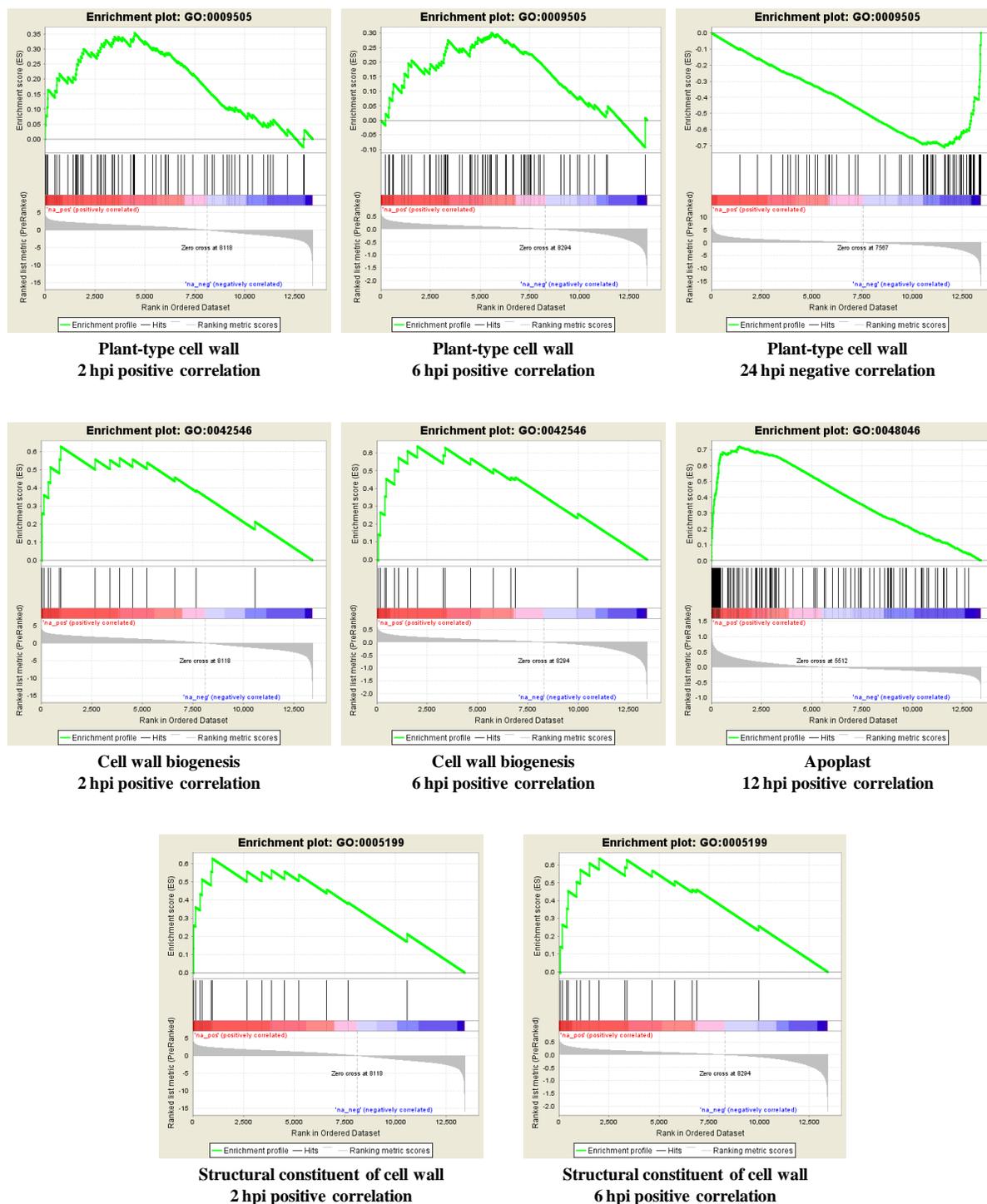


Figure 23 GSEA enrichment plots display host cell wall-related functions that were highly correlated to RCNMV infection at different time points

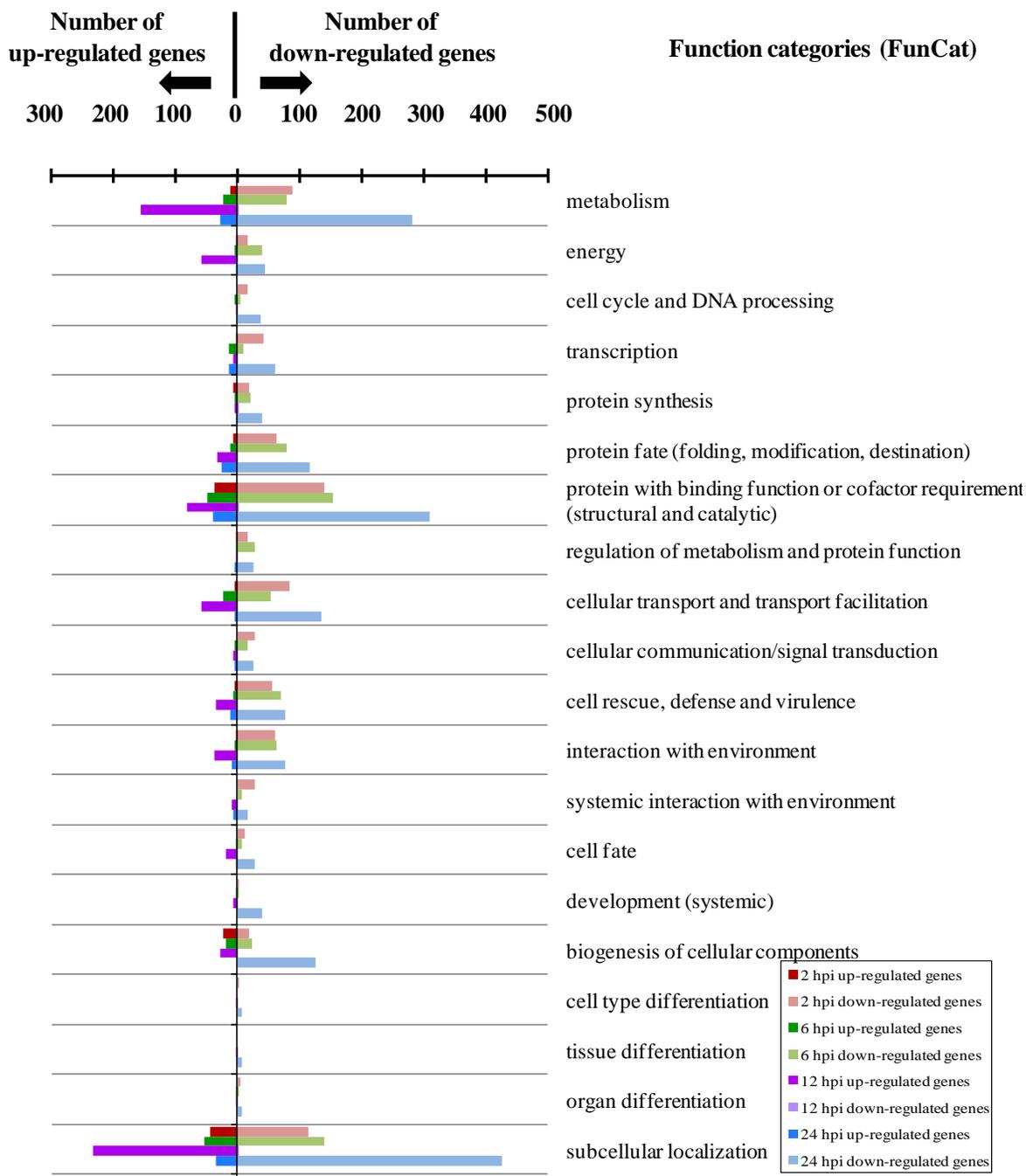


Figure 24 Functional categories (FunCat scheme) of the significantly differentially expressed host genes at different time points of study (2, 6, 12, and 24 hours post inoculation (hpi)) from microarray data analysis with an FDR cutoff of 0.01.

6. Tracking RCNMV RNA profile using the microarray

We incorporated RCNMV probes on our microarray in order to monitor the progress of the RCNMV replication process and RNA accumulation, both RNA-1 and RNA-2. We designed 9 probes spanning across the RCNMV genome. RNA-1 was probed from position 2489 to 3426 (5'-3' direction) which covers the subgenomic RNA encoding the 37-kDa capsid protein (Figure 2 and 25). And RNA-2 was probed from position 172 to 1300 (5'-3' direction) encoding the 35-kDa movement protein (Figure 2 and 26). Surprisingly, both RNA-1 and RNA-2 accumulation was not changed much in the first 12 hours. But the accumulation of both RCNMV RNAs was rapidly increased after 12 hours. The small accumulation of RCNMV RNA may be due to the effect of host defense mechanisms. However, host defense gene expression was not observed until after 12 hours. I assume that the degradation of RCNMV RNA may be largely due to the activated host silencing mechanism, rather than other host defense mechanisms. Although RCNMV is able to encode viral silencing suppressor from RNA-1 and RNA-2, the accumulation of viral RNA is delayed until 12 hours. We know that RCNMV movement protein encoding from RNA-2 is translated in the late infection process, for the purpose of viral movement to adjacent cells. This leads to my speculation that viral silencing suppressor encoding from RNA-2 may also be translated in the late infection process and may play a critical role in the survival of RCNMV RNA progeny in the secondary infected cells. We do not know the mechanism of the RCNMV silencing suppressor yet. However, the expression of RCNMV silencing suppressor may be reflected in the correlation between the dramatic increase in RCNMV RNA accumulation and the sudden up-regulation of host gene expression described as “a

switch mode effect” observed on the microarray at 12 hours (Figure 11, 12, and 13). I hypothesize that viral silencing suppressors expressed during 12 hours of infection may play a part in viral RNA accumulation and the host gene expression switch.

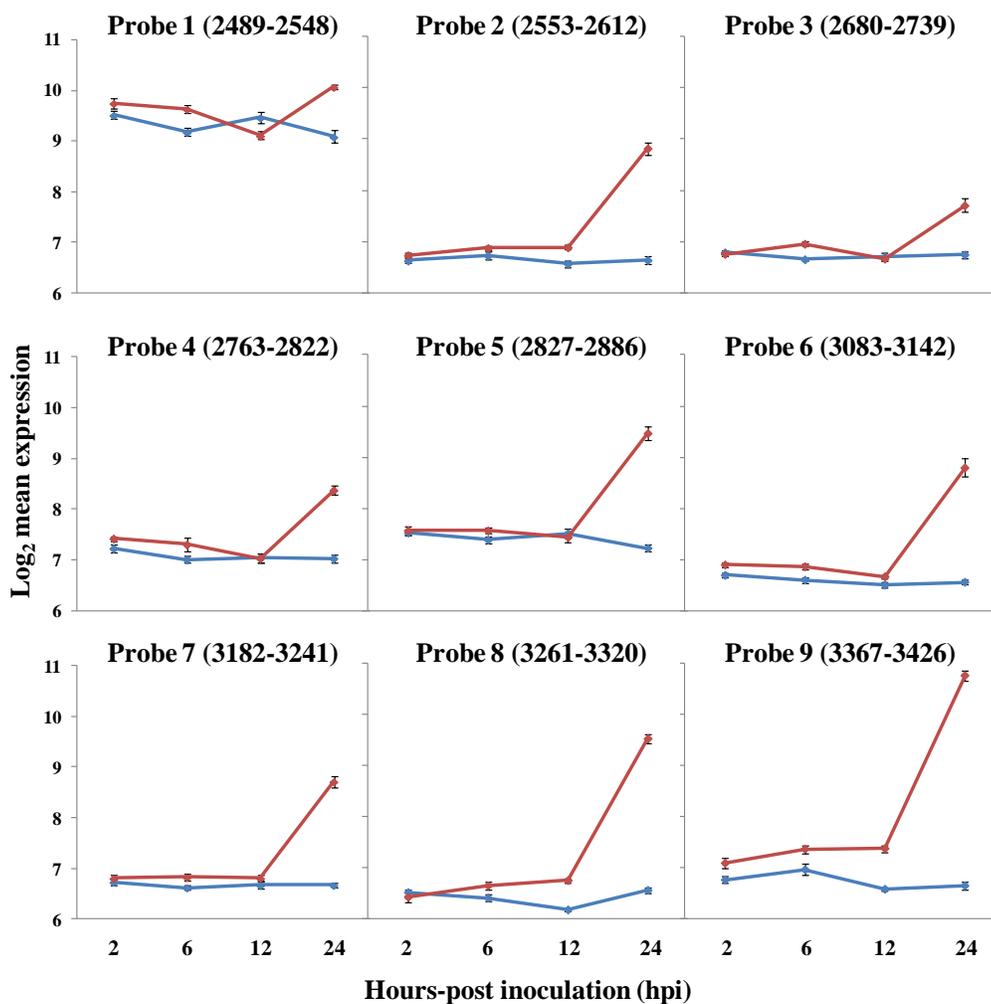


Figure 25 Tracking RCNMV RNA-1 profiles from the microarray data. Number in a parenthesis is a probe position (5'-3') on RCNMV RNA-1 complete sequence from GenBank ID: J04357.1. These microarray probes span from position 2489 to 3426 which covers a subgenomic RNA region (2428-3447) that encodes the 37-kDa capsid protein. Red line is the expression data from RCNMV-inoculated plants and blue line is the expression data from mocks.

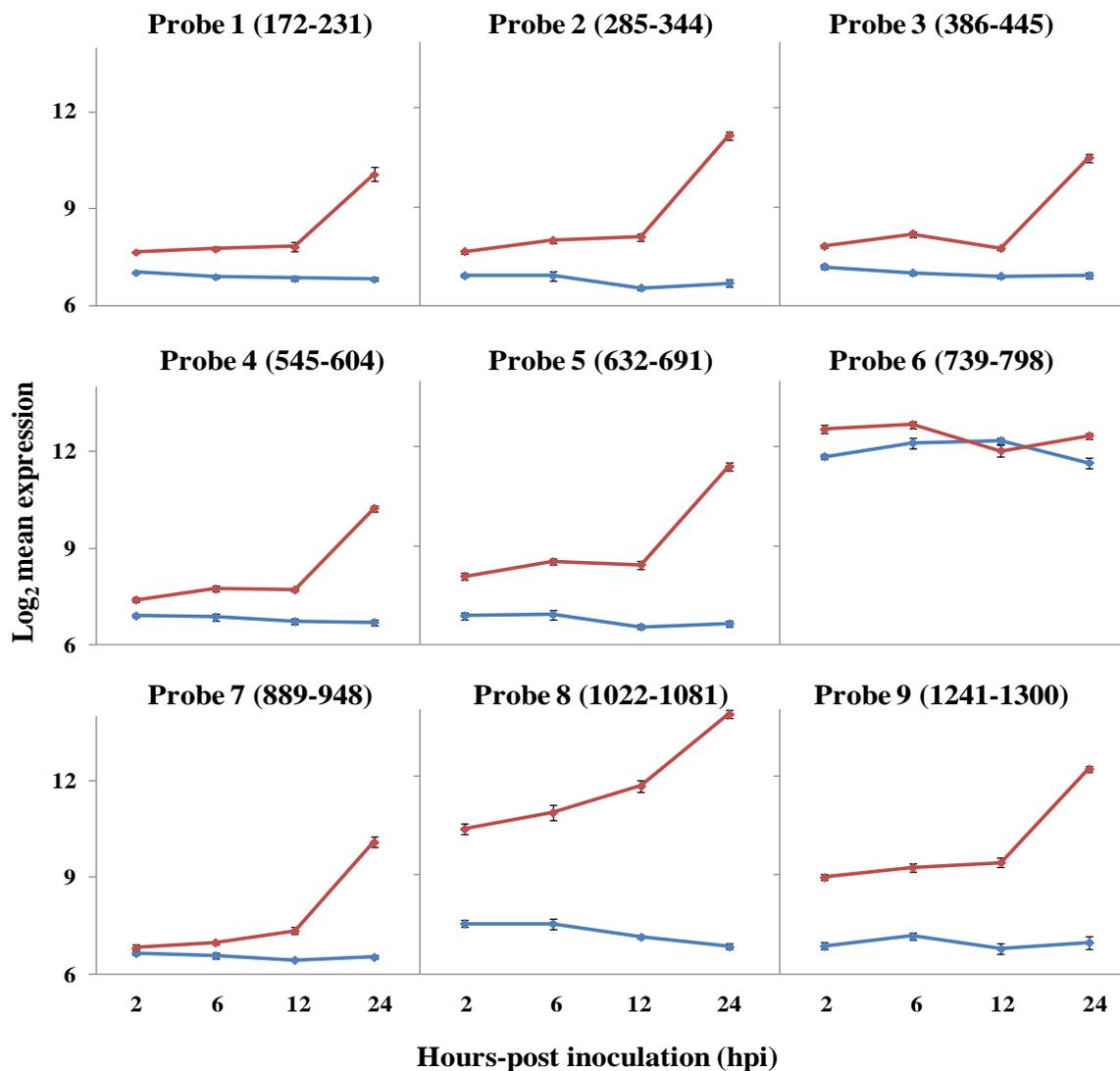


Figure 26 Tracking RCNMV RNA-2 profiles from the microarray data. Number in a parenthesis is a probe position (5'-3') on RCNMV RNA-2 complete sequence from GenBank ID: X08021.1. These microarray probes span from position 172 to 1300 which covers a region (80-1033) that encodes the 35-kDa movement protein. Red line is the expression data from RCNMV-inoculated plants and blue line is the expression data from mocks.

7. Microarray confirmation

Candidate genes were selected to validate the microarray expression data. I selected these candidate genes using three criteria: (i) host genes that were previously known to be involved in plant-virus interaction, (ii) host genes that are important to plant physiology, and (iii) host genes that showed dramatic changes between mock and virus-infected plants. I used qRT-PCR method to validate the candidate genes [136], comparing expression profiles between mock and virus-infected plants. The 60 candidate genes were screened over 24 hours of RCNMV infection. I found that 16 genes had a significant differential expression at $\alpha = 0.05$. The expression profile of these 16 genes is shown in Table 4 and 5. I found that the majority of the selected genes were up-regulated. More than half of the genes are from time point 24. An interesting observation here is that SAR was significantly modulated at all time points of infection. It was down-regulated at time point 2 and 6, but then up-regulated at 12 and 24. Selected candidates were further assayed functionally by either transient gene silencing or over-expression studies in order to determine their effect(s) on RCNMV replication. Details of these functional assays will be described in the chapters that follow.

Table 4 Microarray confirmed by qRT-PCR analysis. The differential host gene expression displayed here has a p-value < 0.1. The red arrow shows host gene up-regulation and the blue arrow shows host gene down regulation. The qRT-PCR expression data is displayed in Table 5.

Gene	BLAST annotation	2 hpi	6 hpi	12 hpi	24 hpi
1	Sucrose transporter type 1	↑			
2	Basal transcription factor 3		↑		
3	Δ-9-desaturase		↑		
4	Fatty oxidation complex α subunit		↑		
5	Hexokinase		↓		
6	C-8,7 sterol isomerase			↑	
7	C-4 -sterol methyl oxidase 2			↑	
8	Succinate dehydrogenase			↑	
9	Ribose-5-phosphate isomerase			↓	
10	26S proteasome regulatory subunit				↑
11	Heat shock protein, putative				↑
12	Pyruvate kinase				↑
13	WRKY transcription factor 6			↑	↑
14	Lipoxygenase 3		↓		↑
15	Mitogen-activated protein kinase 3		↓	↑	↑
16	Systemic acquired resistance 8.2m	↓	↓	↑	↑

Table 5 Differential host gene expression data confirmed by qRT-PCR analysis. The primer set used in this analysis is shown in Table 10. A fold change was computed by a comparison between mocks and RCNMV-infected plants within the same time point of study. A statistical analysis was performed using a match-paired t-test.

Gene	ID and primer set	GenBank BLAST Annotation	hpi	Fold change	Fold change \pm standard error	p-value
1	NB12361	Sucrose transporter type 1	2	1.16	1.14 - 1.19	0.02
2	NB09394	Basal transcription factor 3	6	1.36	1.25 - 1.48	0.04
3	NB01195	Δ -9-desaturase	6	1.78	1.67 - 1.89	0.00
4	NB08626	Fatty oxidation complex α subunit	6	1.30	1.22 - 1.40	0.03
5	NB08500	Hexokinase	6	0.56	0.48 - 0.65	0.03
6	NB02987	C-8,7 sterol isomerase	12	1.23	1.17 - 1.29	0.02
7	NB06928	C-4-sterol methyl oxidase 2	12	1.60	1.41 - 1.82	0.03
8	NB04453	Succinate dehydrogenase	12	1.16	1.12 - 1.21	0.03
9	NB09397	Ribose-5-phosphate isomerase	12	0.86	0.83 - 0.90	0.04
10	NB05721	26s proteasome regulatory subunit	24	2.29	2.20 - 2.38	0.00
11	NB02945	Heat shock protein, putative	24	1.50	1.33 - 1.69	0.05
12	NB08538	Pyruvate kinase	24	1.63	1.58 - 1.69	0.00
13	NB04875	WRKY transcription factor 6	12	1.36	1.22 - 1.52	0.06
			24	10.13	8.16 - 12.56	0.00
14	NB10747	Lipoxygenase 3	6	0.56	0.46 - 0.69	0.10
			24	1.83	1.68 - 1.98	0.01
15	NB06980	Mitogen-activated protein kinase 3	6	0.81	0.75 - 0.87	0.10
			12	1.68	1.45 - 1.95	0.07
			24	13.67	12.71 - 14.72	0.00
16	NB12974	Systemic acquired resistance 8.2m	2	0.69	0.61 - 0.78	0.09
			6	0.82	0.77 - 0.87	0.05
			12	1.88	1.66 - 2.13	0.02
			24	1.57	1.43 - 1.72	0.02

Conclusion

With our time point experimental design and the use of microarray technology, the report presented in this chapter was able to thoroughly describe the temporal changes in the host transcriptomic profiles in relation to an early RCNMV infection. We believe that our sub 24 hour-time course experiment successfully provides the maximum synchronization of infection between infected cells at the different time points of this study. Thus, microarray data obtained from this time course study presents the specific modulation of host gene expression by the particular stage of infection. By comparing changes in host gene expression between RCNMV-inoculated and mock-inoculated plants at 2, 6, 12, and 24 hours post inoculation, it is possible to draw some general parallels between the observed alteration of host gene expression and the stage of virus infection. The majority of host genes exhibited a sudden shutoff as early as 2 hours. The host gene shutoff continued until 6 hours; but then at 12 hours, the majority of host genes expressed a switch mode to up-regulation. Consistent with these host gene expression patterns, RCNMV RNA-1 and RNA-2 transcript accumulation was not increased until after 12 hours. Host defense gene function was repressed in the first 6 hours. One might expect a dramatic increase in viral transcript accumulation within this first 6 hours due to the host defense suppression. But we do not find the accumulation of RCNMV RNA-1 and RNA-2 in the first 12 hours. In other words, RCNMV was not able to maintain its transcript progeny and it barely survived in the first 12 hours. Our results support the idea that the repression of host defenses by RCNMV is presumably to only allow the establishment of viral infection in the primary infected cells. It

is unlikely that the repression of host defense genes influences viral transcript accumulation in the first 6 hours. So what inhibits or delays viral transcript accumulation? I speculate that the exclusive selection of certain host defense strategies at the earliest time periods is the determinant of the fate of the infected host cell. Given our observation of the increased host apoptosis at 6 hours, together with the known highly reactive plant RNA silencing as a plant immune response against viruses [9], I believe that *N. benthamiana* preferably utilizes host programmed cell death and possibly RNA silencing over the deployment of transcriptional-dependent defenses to destroy RCNMV and so this may cause a delay of viral transcript accumulation in the first 12 hours.

Host genes associated with chloroplast, photosynthesis (photosystem I and II), the Calvin cycle, and RuBisCO are repressed in the first 6 hours followed by activation after 12 hours. Different analyses revealed that the suppression of host chloroplast-related functions and the reduction of host photosynthetic rate are the common impacts caused by plant viruses, especially plant viruses that cause leaf yellowing and mosaic symptom [110, 112, 115, 116, 118-120]. Reiner and Beachy found a strong correlation between TMV coat protein associated inside tobacco chloroplasts and the reduction of host photosystem II activity [112]. Most reports of the interaction between virus and host photosynthesis/chloroplast utilized TMV and tobacco as the experimental host model. Recently, more research has utilized *N. benthamiana*. PMMoV and PaMMoV (pepper and paprika mild mottle viruses) cause a reduction of host photosynthetic electron transport in photosystem II in both symptomatic and asymptomatic leaves of *N. benthamiana* [118]. Dardick [137] reported the down-regulation pattern of chloroplast-related genes in his

genome wide study of the interaction between either ToRSV or PPV and *N. benthamiana* with the use of a potato microarray. Taken together with the mosaic symptoms of RCNMV visible at 3-4 days after infection, our results add to the body of knowledge that RCNMV may not only affect host activities occurring in the chloroplast but also host chloroplast development.

Host translation, ribosome biogenesis and cytosolic large ribosomal subunits were also suppressed in the first 6 hours. This result leads to the assumption that RCNMV outcompetes the host translation machinery in the very early stages of infection (as early as 2 hours) in order to use them for viral protein production. This competitiveness is very sudden as expected due largely to the necessity of RCNMV translation occurring prior to the replication process. The virus outcompeting the host for the translation machinery may cause a synergistic impact to host shutoff resulting in not only the sudden host shutoff at the transcript level but also at the protein level.

In the opposite to the host gene shutoff phenomenon in the first 6 hours, host protein kinase was positively correlated to RCNMV infection in the first 6 hours but negatively correlated after 12 hours. Protein kinases play a major role in signal transduction against biotic stress [127, 128]. The increase host protein kinase transcription demonstrates the canonical central strategy that a plant host employs to activate a signal transduction cascade like a “light switch” during the intervention of RCNMV, perhaps in order to activate host defense gene expression. Host cell wall biogenesis was also increased in the first 6 hours,

possibly due to its function as a barrier protecting the plant from external stress and a barrier between adjacent cells.

We believe that the alteration of host gene expression at the different time points of this study is a reflection of the modulation at the different stages of RCNMV infection. RCNMV necessarily recruits or co-opts host factors throughout the course of the infection process in order to produce a robust infection. Our study reveals that the crucial strategy that RCNMV employ is to reprogram host gene expression to accommodate its survival and reproduction. The modulation of host gene expression by RCNMV is a dynamic process; RCNMV alters the expression of a certain suite of host genes/functions/pathways to suit a particular state of the infection process.

Whilst our experimental data provides significant insight into the biology of host changes during the attack by RCNMV, our study in this chapter only investigates the host changes at the transcriptional level. Future studies are expected to establish: (i) the biochemical impact at protein and functional levels of the modulation of host gene expression by a virus, (ii) the spatial, in addition to the temporal, relationship between host gene expression and sites of virus accumulation (either viral RNA or proteins) at the organelle level within an infected cell. For instance, I found a strong modulation of host chloroplast-related genes by RCNMV. Reinero and Beachy [112] found that the level of TMV coat protein inside chloroplasts seemed to be correlated with the severity of chlorotic symptoms. They proposed that a large accumulation of TMV coat protein inside chloroplasts may affect photosynthesis in virus-infected plants by inhibiting photosystem II. The

localization study of virus at the organelle level may shed light on the strategy that RCNMV employs to manipulate host chloroplast-related activities. Moreover, when coupled with facile genetic modification methods for functional assays; for instance, gene silencing, gene over expression, transgenic plant, the report in this chapter provides a valuable foundation for studies aimed at identifying host genes that are required for viral replication and movement. Identification of such host genes may provide targets from which novel strategies can be developed for controlling related viruses that threaten crops and agricultural stability.

In addition to providing insight into the interaction between host and RCNMV presented in this chapter, I have demonstrated that our custom microarray is a powerful technology for transcriptional profiling of *N. benthamiana* genes. Previous genome wide studies using *N. benthamiana* plants as an experimental plant model have employed the heterologous microarray made from *Solanum tuberosum* (potato) for hybridization [137, 138]. Potato is a close relative to *N. benthamiana* in the family *Solanaceae*. In spite of their close genetic relatedness [139, 140], the hybridization capacity of cDNA synthesized from *N. benthamiana* total RNA to the potato microarray is only 75% [138]. Thus, I hope that our microarray could ultimately be our contribution to the scientific community as an alternative high-throughput tool for the genome wide study of *N. benthamiana* in the future.

Selected genes from our microarray validation studies in this chapter are further functionally assayed in the chapters that follow. With our interest to explore the interaction between various host functions and RCNMV, we selected candidate genes from three different groups of host functions: defense, transcription factor, and metabolism. These

selected genes were either transiently silenced or over-expressed in order to determine their effects on RCNMV infection.

Materials and methods

1) *N. benthamiana* microarray design and functional annotation

A thorough detail of *N. benthamiana* microarray creation is described in Chapter 2. The microarray is composed of 13,415 *N. benthamiana* unigenes representing an estimated 38% of the transcriptome. The *N. benthamiana* sequences composing the array were compiled from five distinct EST libraries: expanding leaf tissue manufactured by Amplicon Express, Inc. (Pullman, WA, U.S.A.); green aerial parts of the plant created in A. Maule's laboratory at the John Innes Centre (Norwich, U.K.); a subtraction library of healthy plants against *Colletotrichum destructivum*-infected *N. benthamiana* donated by P. Goodwin from the University of Guelph (Guelph, ON Canada); EST sequences derived from a search of TIGR; and GenBank databases. The unigene set was collated from these 40,000 individual ESTs and the sequences were submitted to NimbleGen Systems, Inc. (Roche NimbleGen, Madison, WI, U.S.A.) to customize a oligonucleotide-based microarray which was designed to consist of 60-mer probes, and each EST consisted of nine different probes, repeated three times, on the array [70]. Our array is commercially available through a request from us and Nimblegen.

The Blast2GO [141] software was applied to assign the functionality information of genes represented on this array. The sequences were subjected to an NR database of

GenBank search using BLASTX [94] and only similarities with expected value (E) smaller or equal to 1×10^{-5} were further classified into categories according to a gene ontology scheme (GO); cellular component, biological process, and molecular function [64]. In addition to GO, we also had annotation information from KEGG [67] and MIPS functional catalogue (FunCat) [68].

2) Plants, RCNMV inoculum preparation and inoculation procedure

N. benthamiana was a plant model used in this experiment. Details of plant maintenance, RCNMV inoculum preparation and inoculation procedure were described in Chapter 5 materials and methods. Briefly, RCNMV inoculum was prepared in a 110 μl volume per 1 plant. This 110 μl inoculum was a mixture of 1 μl *in vitro* T7 RNA-1 transcript and 1 μl *in vitro* T7 RNA-2 transcript and 108 μl inoculation buffer (10 mM sodium diphosphate, pH 7.2). Four leaves per 1 plant were inoculated with either RCNMV inoculum or buffer. 27 μl of the *in vitro* RCNMV transcript mixture (or inoculation buffer for mock) was pipetted onto each leaf and mechanically rubbed with carborundum (abrasive). Inoculated plants were maintained in a temperature and light controlled environment at 18-22°C, 16 hour-light and 8 hour-dark period at the NCSU greenhouses (Method Road).

3) Microarray experiment

The microarray experiment was previously described by a former graduate student [70]. Briefly, four plants and four leaves per 1 plant were inoculated with either RCNMV inoculum or buffer. One leaf from each inoculated plant was harvested at 2, 6, 12, and 24 hpi.

Each harvested leaf was immediately submerged and stored in RNALater (Ambion™) at 4° until RNA extraction took place. Total RNA was isolated from 100 mg tissue per leaf according to an RNeasy Plant Mini Kit (Qiagen™). Isolated total RNA for each timepoint/treatment combination was pooled and DNase treated with Turbo DNA-free (Ambion™) to remove any contaminating DNA. Purified RNA was quantified using a spectrophotometer (Bio-Rad Spec3000), separated on a 1% gel to check for RNA integrity and sent on dry ice to Nimblegen for microarray hybridization. Each RNA sample was hybridized on an independent array.

4) Microarray data analysis

The statistical software packages JMP genomics 5 and SAS 9.2 were used to conduct the microarray data analysis. Microarray raw data was pre-processed with a \log_2 transformation and Loess normalization prior to a differential gene expression analysis. The quality control of the microarray data before and after being processed was evaluated by two approaches: a distribution analysis, and a correlation and grouped scatterplot.

A differential gene expression analysis is a comparison of host gene expression between virus- and mock- inoculated plants. The ANOVA row-by-row modeling was used for a differential gene expression analysis of all the genes spotted on the array. The model fixed effects are “condition”, “hpi”, and “condition*hpi”. And a model LSmean effect is “condition*hpi”. Condition in this context is a treatment: virus- and mock- inoculation. The multiple testing correction FDR cutoff of 0.01 was used to define a significance of a

differential gene expression. The result of the differential gene expression analysis was visualized using Venn diagram, volcano plot, and hierarchical clustering.

5) Microarray confirmation experiment and data analysis

Three plants and four leaves per 1 plant were inoculated with either RCNMV inoculum or buffer (mock). Two leaves per 1 inoculated plant were harvested at 2, 6, 12, and 24 hpi. To precisely stop plant physiological process at the time point of the study, 100 mg tissue was cut from the harvested leaf, placed in 1.5 ml microcentrifuge tube containing 3 3-mm glass beads, immediately submerged in liquid nitrogen and kept in -80 °C until a total RNA extraction.

Total RNA extraction, total RNA cleanup, RNA quantification by real-time PCR (qRT-PCR), and real-time PCR data analysis are described in Chapter 5 materials and methods. Briefly, the total RNA extraction was performed accordingly to the TRIzol protocol (Invitrogen™). The total RNA extracts isolated from the leaves harvested within the same day of the experimentation were pooled based on the time point of study and the treatment to represent a biological replicate. Total RNA extracts destined for qRT-PCR analysis were treated with Turbo DNA-free (Ambion™) according to the manufacturer's protocol to remove any residual input DNA. First strand cDNA synthesis was performed using a DNA-free total RNA (equivalent to ~1 µg total RNA).

SYBR Green-based quantitative real-time PCR (qRT-PCR) method was used in this study. The qRT-PCR reaction was prepared accordingly to FastStart Universal SYBR Green

Master ROX protocol (RocheTM). The first strand cDNA equivalent to 40 ng total RNA was used to prepare qRT-PCR reaction. The information of primers used in this qRT-PCR analysis is shown in Table 10. The qRT-PCR reactions were setup in a 384 well plate and placed in ABI7900 HT Fast real-time PCR system (Applied BiosystemsTM). The Applied Biosystems SDS software version 2.4 was used to monitor and dissect real-time PCR data. Three qRT-PCR reactions were prepared per 1 biological replicate. All PCR products are less than 200 base pairs. The cycling condition was set as follows: 1 cycle at 95°C for 10 min, and 40 cycles at 95°C for 15 sec /55°C for 1 min. An association analysis was also performed in order to test if there were any non-specific PCR products formed. The condition for the association analysis was set as follows: 1 cycle at 95°C for 15 sec, 60°C for 15 sec and 95°C for 15 sec.

A relative quantification, $2^{-\Delta\Delta C_t}$ method (equation shown in Chapter 5 materials and methods) [136] was used to analyze the real-time PCR data. Expression of the target gene was normalized against the expression of a reference gene. Actin and 18s rRNA were used as the reference genes to normalize the target gene that was performed within the same plate.

Four independent studies were performed on four different days of the experimentation. Therefore, there were four biological replicates per treatment per time point of study. The matched-pair t-test statistical analysis was used to examine the real-time PCR data. A pair data is determined by the day of the experimentation; therefore, four matched pairs were used to compute a differential gene expression at one time point of study.

6) Gene set enrichment analysis (GSEA)

GSEA software [95] was applied in the functional analysis in order to test an enrichment of host gene functions, that whether or not they are correlated to RCNMV infection. This method is focused on analyzing a set of genes that have the same functionality, rather than analyzing one gene at a time. GO (Gene Ontology) was used to classified host gene function in this study. Our microarray consists of 13,413 host genes that represents 1,931 unique GO terms. Therefore, GSEA was used to analyze the association between these 1,931 GO terms and RCNMV infection. A pre-rank method was used to organize our input files prior to uploading into GSEA software. The 1,931 GO terms and their corresponding host genes were used to create “GeneSetData.gmt” input file. Host genes and their corresponding t-statistic values were used to create “Tstat.rnk” input file. The t-statistic value of each host gene was obtained from the microarray data analysis between mock- and virus- inoculated plants at each time point of study. These two input files were then subjected to GSEA version 2.0 software. The GSEA analysis was set at 1,000 permutations. The resulting top 20 of the positively enriched GO terms and the resulting top 20 of the negatively enriched GO terms were used to interpret their correlation to RCNMV infection. The data from each time point of study (2, 6, 12, and 24 hpi) was run on GSEA software separately.

Table 6 Top 20 GSEA result at 2 hours post RCNMV inoculation

Top 20 positive correlation		Top 20 negative correlation	
GO ID	Description	GO ID	Description
3964	RNA-directed DNA polymerase activity	3735	structural constituent of ribosome
3993	acid phosphatase activity	6118	electron transport
4713	protein tyrosine kinase activity	6414	translational elongation
4872	receptor activity	6525	arginine metabolic process
5199	structural constituent of cell wall	6950	response to stress
5529	sugar binding	7264	small GTPase mediated signal transduction
6278	RNA-dependent DNA replication	9507	chloroplast
6771	riboflavin metabolic process	9535	chloroplast thylakoid membrane
6813	potassium ion transport	9570	chloroplast stroma
6855	multidrug transport	10287	plastoglobule
7169	transmembrane receptor protein tyrosine kinase signaling pathway	15035	protein disulfide oxidoreductase activity
9505	plant-type cell wall	15979	photosynthesis
9987	cellular process	19253	reductive pentose-phosphate cycle
15238	drug transporter activity	19482	beta-alanine metabolic process
15297	antiporter activity	22900	electron transport chain
19497	hexachlorocyclohexane metabolic process	42254	ribosome biogenesis
35091	phosphoinositide binding	42742	defense response to bacterium
42546	cell wall biogenesis	45271	respiratory chain complex I
44238	primary metabolic process	46487	glyoxylate metabolic process
44464	cell part	51246	regulation of protein metabolic process

Table 7 Top 20 GSEA result at 6 hours post RCNMV inoculation

Top 20 positive correlation		Top 20 negative correlation	
GO ID	Description	GO ID	Description
3700	transcription factor activity	287	magnesium ion binding
4190	aspartic-type endopeptidase activity	3735	structural constituent of ribosome
4674	protein serine/threonine kinase activity	5840	ribosome
4713	protein tyrosine kinase activity	6412	translation
4872	receptor activity	6414	translational elongation
5199	structural constituent of cell wall	9507	chloroplast
5667	transcription factor complex	9522	photosystem I
6855	multidrug transport	9523	photosystem II
6915	apoptosis	9535	chloroplast thylakoid membrane
7169	transmembrane receptor protein tyrosine kinase signaling pathway	9570	chloroplast stroma
7275	multicellular organismal development	9579	thylakoid
9117	nucleotide metabolic process	9765	photosynthesis, light harvesting
9505	plant-type cell wall	10287	plastoglobule
10200	response to chitin	15979	photosynthesis
15238	drug transporter activity	16168	chlorophyll binding
15297	antiporter activity	18298	protein-chromophore linkage
34960	cellular biopolymer metabolic process	19253	reductive pentose-phosphate cycle
35091	phosphoinositide binding	22625	cytosolic large ribosomal subunit
42546	cell wall biogenesis	42254	ribosome biogenesis
44238	primary metabolic process	42742	defense response to bacterium

Table 8 Top 20 GSEA result at 12 hours post RCNMV inoculation

Top 20 positive correlation		Top 20 negative correlation	
GO ID	Description	GO ID	Description
287	magnesium ion binding	3700	transcription factor activity
786	nucleosome	3964	RNA-directed DNA polymerase activity
3735	structural constituent of ribosome	4674	protein serine/threonine kinase activity
6096	glycolysis	4713	protein tyrosine kinase activity
6412	translation	5576	extracellular region
9507	chloroplast	5667	transcription factor complex
9522	photosystem I	6278	RNA-dependent DNA replication
9523	photosystem II	6355	regulation of transcription, DNA-dependent
9535	chloroplast thylakoid membrane	6468	protein amino acid phosphorylation
9570	chloroplast stroma	6813	potassium ion transport
9579	thylakoid	7169	transmembrane receptor protein tyrosine kinase signaling pathway
9765	photosynthesis, light harvesting	7275	multicellular organismal development
9941	chloroplast envelope	9069	serine family amino acid metabolic process
15986	ATP synthesis coupled proton transport	9607	response to biotic stimulus
16168	chlorophyll binding	16042	lipid catabolic process
18298	protein-chromophore linkage	34960	cellular biopolymer metabolic process
19253	reductive pentose-phosphate cycle	35091	phosphoinositide binding
42254	ribosome biogenesis	43565	sequence-specific DNA binding
42742	defense response to bacterium	44238	primary metabolic process
48046	apoplast	46983	protein dimerization activity

Table 9 Top 20 GSEA result at 24 hours post RCNMV inoculation

Top 20 positive correlation		Top 20 negative correlation	
GO ID	Description	GO ID	Description
272	polysaccharide catabolic process	166	nucleotide binding
6414	translational elongation	786	nucleosome
6952	defense response	3676	nucleic acid binding
8080	N-acetyltransferase activity	3779	actin binding
9538	photosystem I reaction center	3924	GTPase activity
9573	chloroplast ribulose biphosphate carboxylase complex	4190	aspartic-type endopeptidase activity
9737	response to abscisic acid stimulus	5215	transporter activity
9853	photorespiration	5730	nucleolus
10287	plastoglobule	5783	endoplasmic reticulum
15035	protein disulfide oxidoreductase activity	5886	plasma membrane
16579	protein deubiquitination	5982	starch metabolic process
16984	ribulose-biphosphate carboxylase activity	5985	sucrose metabolic process
16998	cell wall catabolic process	6012	galactose metabolic process
19253	reductive pentose-phosphate cycle	6334	nucleosome assembly
19941	modification-dependent protein catabolic process	6448	regulation of translational elongation
30001	metal ion transport	6508	proteolysis
45271	respiratory chain complex I	7010	cytoskeleton organization
46872	metal ion binding	9505	plant-type cell wall
50832	defense response to fungus	43169	cation binding
51246	regulation of protein metabolic process	50660	FAD binding

Table 10 Oligonucleotide information used in the real-time PCR analysis Chapter 3

Primer set name		Sequence information
NB01195	Forward	ACACCACTTGCGTTTGAGTA
	Reverse	CGAAGTTGAACCTCAAACAA
NB02945	Forward	CTGACCCCTTTTGCTCATCA
	Reverse	GCCATGCAAATTAGCCCTAA
NB02987	Forward	ATATTCCTCTCACAGTCACAC
	Reverse	GTAAAGACAAAATACCCCTC
NB04453	Forward	ACCTCCAAAGACAACCAGCA
	Reverse	GAGAAAAGTCGGCGAAAATG
NB04875	Forward	CGTTGCCGGAAAAGTCATCA
	Reverse	CGGAAAGTGTTTCGGAAAATGA
NB05721	Forward	CTTTTCAGACAAGATGACTCAAGAC
	Reverse	TAGTGCCATCGTCAACCTCA
NB06928	Forward	TCTTCTTGTCTGGACTCCCG
	Reverse	GCGTCCCATGAATTTAAAG
NB06980	Forward	TTCCTTCGGTTTTAACTCATGG
	Reverse	ACGCATTCGCGATTTTCTTA
NB08500	Forward	TATAGGTTATCTTCCGTTTCG
	Reverse	CAAGATGAGGATGGTAAAGA

Table 10 Continued

Primer set name		Sequence information
NB08538	Forward	ATATCATTCTGGATGTGTGAGC
	Reverse	GATCCTAAGCTTGGTGGTTC
NB08626	Forward	CTTCGTCTCTTCCTCAGTTCC
	Reverse	GGAACAGAACAGCAGAGTCG
NB09394	Forward	ACATTGAAAGTAAGGGATGC
	Reverse	ATTTGAAGAAGTTGGCTGAG
NB09397	Forward	CACGAGGCAACCCTCCATTC
	Reverse	TGGCGGAGAGTAGAAGGGGA
NB10747	Forward	CCACCTCTCATGCGTCGATTAA
	Reverse	TCGTCTGGGTGAATGAGTCGA
NB12361	Forward	ACAAACACACAGTAGCTAGTTC
	Reverse	GGAGTCAATAAAGAAAGCTG
NB12974	Forward	AACTAGCTGATGCAAGGGAGA
	Reverse	AACATCTGCAAGCCAAACAA
Actin	Forward	GTGACCTCACTGATAGTTTGA
	Reverse	TACAGAAGAGCTGGTCTTTG
18s rRNA	Forward	GCGACGCATCATTCAAATTC
	Reverse	TCCGGAATCGAACCCCTAATTC

References

1. Matthews R: **Plant virology**: Elsevier; 1991.
2. Hull R: **Plant virology**: Gulf Professional Publishing; 2001.
3. Maule A, Leh V, Lederer C: **The dialogue between viruses and hosts in compatible interactions**. *Current opinion in plant biology* 2002, **5**(4):279-284.
4. Kushner DB, Lindenbach BD, Grdzlishvili VZ, Noueir AO, Paul SM, Ahlquist P: **Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus**. *Proceedings of the National Academy of Sciences* 2003, **100**(26):15764-15769.
5. Ahlquist P, Noueir AO, Lee W-M, Kushner DB, Dye BT: **Host factors in positive-strand RNA virus genome replication**. *Journal of Virology* 2003, **77**(15):8181-8186.
6. Nagy PD, Pogany J: **The dependence of viral RNA replication on co-opted host factors**. *Nature Reviews Microbiology*, **10**(2):137-149.
7. Boevink P, Oparka KJ: **Virus-host interactions during movement processes**. *Plant Physiology* 2005, **138**(4):1815-1821.
8. van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL: **Structure and function of resistance proteins in solanaceous plants**. *Annu Rev Phytopathol* 2007, **45**:43-72.
9. Voinnet O: **RNA silencing as a plant immune system against viruses**. *TRENDS in Genetics* 2001, **17**(8):449-459.
10. Voinnet O: **Induction and suppression of RNA silencing: insights from viral infections**. *Nature Reviews Genetics* 2005, **6**(3):206-220.
11. Ding S-W, Voinnet O: **Antiviral immunity directed by small RNAs**. *Cell* 2007, **130**(3):413-426.
12. Ding XS, Liu J, Cheng N-H, Folimonov A, Hou Y-M, Bao Y, Katagi C, Carter SA, Nelson RS: **The Tobacco mosaic virus 126-kDa protein associated with virus replication and movement suppresses RNA silencing**. *Molecular plant-microbe interactions* 2004, **17**(6):583-592.

13. Vogler H, Kwon M-O, Dang V, Sambade A, Fasler M, Ashby J, Heinlein M: **Tobacco mosaic virus movement protein enhances the spread of RNA silencing.** *PLoS pathogens* 2008, **4**(4):e1000038.
14. Wolf S, LUCAS WJ, DEOM CM, BEACHY RN: **Movement protein of Tobacco mosaic virus modifies plasmodesmatal size exclusion limit.** *Science* 1989, **246**(4928):377-379.
15. Carrington JC, Kasschau KD, Mahajan SK, Schaad MC: **Cell-to-cell and long-distance transport of viruses in plants.** *The Plant Cell* 1996, **8**(10):1669.
16. Agrios G: **Plant diseases caused by viruses.** *Plant Pathology Fifth Edition Elsevier Academia Press* 2005:724-820.
17. Goodwin MM, Zaitlin D, Naidu RA, Lommel SA: **Nicotiana benthamiana: its history and future as a model for plant-pathogen interactions.** *Molecular plant-microbe interactions* 2008, **21**(8):1015-1026.
18. Clemente T: **Nicotiana (Nicotiana tobaccum, Nicotiana benthamiana).** In: *Agrobacterium Protocols.* Springer; 2006: 143-154.
19. Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB: **A draft genome sequence of Nicotiana benthamiana to enhance molecular plant-microbe biology research.** *Molecular plant-microbe interactions* 2012, **25**(12):1523-1530.
20. Bennett MD, Leitch IJ: **Nuclear DNA amounts in angiosperms.** *Annals of Botany* 1995, **76**(2):113-176.
21. Bennett MD, Leitch IJ, Price HJ, Johnston JS: **Comparisons with Caenorhabditis (~100 Mb) and Drosophila (~175 Mb) using flow cytometry show genome size in Arabidopsis to be ~157 Mb and thus ~25% larger than the Arabidopsis genome initiative estimate of ~125 Mb.** *Annals of Botany* 2003, **91**(5):547-557.
22. Bennett M, Leitch I: **Nuclear DNA amounts in angiosperms: progress, problems and prospects.** *Annals of Botany* 2005, **95**(1):45-90.
23. Senthil-Kumar M, Mysore KS: **New dimensions for VIGS in plant functional genomics.** *Trends in plant science*, **16**(12):656-665.
24. Wydro M, Kozubek E, Lehmann Pa: **Optimization of transient Agrobacterium-mediated gene expression system in leaves of Nicotiana benthamiana.** *ACTA BIOCHIMICA POLONICA-ENGLISH EDITION-* 2006, **53**(2):289.
25. Gleba Y, Klimyuk V, Marillonnet S: **Magniffection-a new platform for expressing recombinant vaccines in plants.** *Vaccine* 2005, **23**(17):2042-2048.

26. Matoba N, Davis KR, Palmer KE: **Recombinant protein expression in *Nicotiana***. In: *Plant Chromosome Engineering*. Springer: 199-219.
27. Strasser R, Stadlmann J, Schahs M, Stiegler G, Quendler H, Mach L, Glossl J, Weterings K, Pabst M, Steinkellner H: **Generation of glyco-engineered *Nicotiana benthamiana* for the production of monoclonal antibodies with a homogeneous human-like N-glycan structure**. *Plant biotechnology journal* 2008, **6**(4):392-402.
28. Dasgupta R, Garcia BH, Goodman RM: **Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes**. *Proceedings of the National Academy of Sciences* 2001, **98**(9):4910-4915.
29. Gould A, Francki R, Hatta T, Hollings M: **The bipartite genome of red clover necrotic mosaic virus**. *Virology* 1981, **108**(2):499-506.
30. Xiong Z, Lommel S: **The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1**. *Virology* 1989, **171**(2):543-554.
31. Lommel S, Weston-Fina M, Xiong Z, Lomonosoff GP: **The nucleotide sequence and gene organization of red clover necrotic mosaic virus RNA-2**. *Nucleic acids research* 1988, **16**(17):8587-8602.
32. Hollings M, Stone OM: **Red clover necrotic mosaic virus**. *CMI/AAB Descriptions of Plant Viruses* 1977, **181**.
33. Kim K-H, Lommel SA: **Sequence element required for efficient ⁺1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus**. *Virology* 1998, **250**(1):50-59.
34. Kim K, Lommel S: **Identification and analysis of the site of ⁺1 ribosomal frameshifting in red clover necrotic mosaic virus**. *Virology* 1994, **200**(2):574-582.
35. Koonin EV: **The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses**. *J Gen Virol* 1991, **72**(Pt 9):2197-2206.
36. Zavriev S, Hickey C, Lommel S: **Mapping of the red clover necrotic mosaic virus subgenomic RNA**. *Virology* 1996, **216**(2):407-410.
37. Xiong Z, Kim K, Giesman-Cookmeyer D, Lommel S: **The roles of the red clover necrotic mosaic virus capsid and cell-to-cell movement proteins in systemic infection**. *Virology* 1993, **192**(1):27-32.
38. Fujiwara T, Giesman-Cookmeyer D, Ding B, Lommel SA, Lucas WJ: **Cell-to-cell trafficking of macromolecules through plasmodesmata potentiated by the red**

- clover necrotic mosaic virus movement protein.** *The Plant Cell Online* 1993, **5**(12):1783-1794.
39. Osman T, Buck K: **Replication of red clover necrotic mosaic virus RNA in cowpea protoplasts: RNA 1 replicates independently of RNA 2.** *Journal of General Virology* 1987, **68**(2):289-296.
 40. Paje-Manalo L, Lommel S: **Independent replication of red clover necrotic mosaic virus RNA-1 in electroporated host and nonhost *Nicotiana* species protoplasts.** *Phytopathology* 1989, **79**(4):457-461.
 41. Sit TL, Vaewhongs AA, Lommel SA: **RNA-mediated trans-activation of transcription from a viral RNA.** *Science* 1998, **281**(5378):829-832.
 42. Wilkie GS, Dickson KS, Gray NK: **Regulation of mRNA translation by 5'- and 3'-UTR-binding factors.** *Trends in biochemical sciences* 2003, **28**(4):182-188.
 43. Xiong Z, Lommel SA: **Red clover necrotic mosaic virus infectious transcripts synthesized *in Vitro*.** *Virology* 1991, **182**(1):388-392.
 44. Mizumoto H, Tatsuta M, Kaido M, Mise K, Okuno T: **Cap-independent translational enhancement by the 3' untranslated region of *Red clover necrotic mosaic virus* RNA1.** *Journal of Virology* 2003, **77**(22):12113-12121.
 45. Takeda A, Tsukuda M, Mizumoto H, Okamoto K, Kaido M, Mise K, Okuno T: **A plant RNA virus suppresses RNA silencing through viral RNA replication.** *The EMBO journal* 2005, **24**(17):3147-3157.
 46. Powers JG, Sit TL, Heinsohn C, George CG, Kim K-H, Lommel SA: **The Red clover necrotic mosaic virus RNA-2 encoded movement protein is a second suppressor of RNA silencing.** *Virology* 2008, **381**(2):277-286.
 47. Turner KA, Sit TL, Callaway AS, Allen NS, Lommel SA: **Red clover necrotic mosaic virus replication proteins accumulate at the endoplasmic reticulum.** *Virology* 2004, **320**(2):276-290.
 48. Derrick PM, Barker H, Oparka KJ: **Increase in Plasmodesmatal Permeability during Cell-to-Cell Spread of *Tobacco Rattle Virus* from Individually Inoculated Cells.** *The Plant Cell Online* 1992, **4**(11):1405-1412.
 49. Nagy PD, Pogany J: **Global Genomics and Proteomics Approaches to Identify Host Factors as Targets to Induce Resistance Against *Tomato Bushy Stunt Virus*.** *Advances in virus research*, **76**:123-177.

50. Pogany J, Panavas T, Serviene E, Nawaz-ul-Rehman MS, D Nagy P: **A High-Throughput Approach for Studying Virus Replication in Yeast.** *Current protocols in microbiology*:16J. 11.11-16J. 11.15.
51. Jiang Y, Serviene E, Gal J, Panavas T, Nagy PD: **Identification of essential host factors affecting tombusvirus RNA replication based on the yeast Tet promoters Hughes Collection.** *Journal of Virology* 2006, **80**(15):7394-7404.
52. Serviene E, Jiang Y, Cheng C-P, Baker J, Nagy PD: **Screening of the yeast yTHC collection identifies essential host factors affecting tombusvirus RNA recombination.** *Journal of Virology* 2006, **80**(3):1231-1241.
53. Havelda Z, Maule AJ: **Complex spatial responses to cucumber mosaic virus infection in susceptible Cucurbita pepo cotyledons.** *The Plant Cell Online* 2000, **12**(10):1975-1985.
54. Samuel G: **Some experiments on inoculating methods with plant viruses, and on local lesions.** *Annals of Applied Biology* 1931, **18**(4):494-507.
55. Atkinson P, Matthews R: **On the origin of dark green tissue in tobacco leaves infected with Tobacco mosaic virus.** *Virology* 1970, **40**(2):344-356.
56. Xie Z, Fan B, Chen C, Chen Z: **An important role of an inducible RNA-dependent RNA polymerase in plant antiviral defense.** *Proceedings of the National Academy of Sciences* 2001, **98**(11):6516-6521.
57. Cocking EC: **Plant cell protoplasts-isolation and development.** *Annual Review of Plant Physiology* 1972, **23**(1):29-50.
58. Fannin F, Shaw JG: **Evidence for concurrent spread of Tobacco mosaic virus from infected epidermal cells to neighboring epidermal and mesophyll cells.** *Plant Science* 1987, **51**(2):305-310.
59. Dolja VV, McBride HJ, Carrington JC: **Tagging of plant potyvirus replication and movement by insertion of beta-glucuronidase into the viral polyprotein.** *Proceedings of the National Academy of Sciences* 1992, **89**(21):10208-10212.
60. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270**(5235):467-470.
61. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2008, **36**(suppl 1):D25-D30.

62. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R: **NCBI GEO: mining tens of millions of expression profiles- database and tools update.** *Nucleic acids research* 2007, **35**(suppl 1):D760-D765.
63. Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y: **The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond.** *Plant Physiology* 2005, **138**(3):1310-1317.
64. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**(1):25-29.
65. Zdobnov EM, Apweiler R: **InterProScan- an integration platform for the signature-recognition methods in InterPro.** *Bioinformatics* 2001, **17**(9):847-848.
66. Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M: **The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community.** *Nucleic acids research* 2003, **31**(1):224-228.
67. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic acids research* 2000, **28**(1):27-30.
68. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, Tetko I, Guldener U, Mannhaupt G, Munsterkötter M: **The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes.** *Nucleic acids research* 2004, **32**(18):5539-5545.
69. Mewes H-W, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkötter M, Rudd S, Weil B: **MIPS: a database for genomes and protein sequences.** *Nucleic acids research* 2002, **30**(1):31-34.
70. Feulner GE: **The Modulation of *Nicotiana benthamiana* gene expression by Red clover necrotic mosaic virus.** *North Carolina State University Electronic Theses and Dissertations* 2008.
71. Churchill GA: **Fundamentals of experimental design for cDNA microarrays.** *Nature genetics* 2002, **32**:490-495.
72. Lee M-LT, Kuo FC, Whitmore G, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.** *Proceedings of the National Academy of Sciences* 2000, **97**(18):9834-9839.

73. Quinn GP, Keough MJ: **Experimental design and data analysis for biologists:** Cambridge University Press; 2002.
74. Kathleen Kerr M, A CHURCHILL G: **Statistical design and the analysis of gene expression microarray data.** *Genetical research* 2001, **77**(02):123-128.
75. Kerr MK, Martin M, Churchill GA: **Analysis of variance for gene expression microarray data.** *Journal of computational biology* 2000, **7**(6):819-837.
76. Hochberg Y, Tamhane AC: **Multiple comparison procedures:** John Wiley & Sons, Inc.; 1987.
77. Dunnett CW: **A multiple comparison procedure for comparing several treatments with a control.** *Journal of the American Statistical Association* 1955, **50**(272):1096-1121.
78. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *Journal of the Royal Statistical Society Series B (Methodological)* 1995:289-300.
79. Bland JM, Altman DG: **Multiple significance tests: the Bonferroni method.** *Bmj* 1995, **310**(6973):170.
80. Hochberg Y: **A sharper Bonferroni procedure for multiple tests of significance.** *Biometrika* 1988, **75**(4):800-802.
81. Hazen SP, Naef F, Quisel T, Gendron JM, Chen H, Ecker JR, Borevitz JO, Kay SA: **Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays.** *Genome Biol* 2009, **10**(2):R17.
82. McClung CR: **Plant circadian rhythms.** *The Plant Cell Online* 2006, **18**(4):792-803.
83. Dodd AN, Salathia N, Hall A, KÃ©ve E, TÃ©th Rk, Nagy F, Hibberd JM, Millar AJ, Webb AA: **Plant circadian clocks increase photosynthesis, growth, survival, and competitive advantage.** *Science* 2005, **309**(5734):630-633.
84. Parzen E: **On estimation of a probability density function and mode.** *Annals of mathematical statistics* 1962, **33**(3):1065-1076.
85. Smyth GK, Speed T: **Normalization of cDNA microarray data.** *methods* 2003, **31**(4):265-273.
86. Quackenbush J: **Computational analysis of microarray data.** *Nature Reviews Genetics* 2001, **2**(6):418-427.

87. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proceedings of the National Academy of Sciences* 1998, **95**(25):14863-14868.
88. Aranda M, Maule A: **Virus-induced host gene shutoff in animals and plants.** *Virology* 1998, **243**(2):261-267.
89. Maule A, Escaler M, Aranda M: **Programmed responses to virus replication in plants.** *Molecular plant pathology* 2000, **1**(1):9-15.
90. Wang D, Maule AJ: **Inhibition of host gene expression associated with plant virus replication.** *Science* 1995, **267**(5195):229-231.
91. Leonard S, Plante D, Wittmann S, Daigneault N, Fortin MG, Laliberte J-F: **Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity.** *Journal of Virology* 2000, **74**(17):7730-7737.
92. Gallie DR: **Translational control of cellular and viral mRNAs.** In: *Post-Transcriptional Control of Gene Expression in Plants*. Springer; 1996: 145-158.
93. Shan XC, Goodwin PH: **Identification of a SAR8. 2 gene in the susceptible host response of *Nicotiana benthamiana* to *Colletotrichum orbiculare*.** *Functional plant biology* 2005, **32**(3):259-266.
94. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
95. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**(43):15545-15550.
96. Omote H, Hiasa M, Matsumoto T, Otsuka M, Moriyama Y: **The MATE proteins as fundamental transporters of metabolic and xenobiotic organic cations.** *Trends in pharmacological sciences* 2006, **27**(11):587-593.
97. Kuroda T, Tsuchiya T: **Multidrug efflux transporters in the MATE family.** *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics* 2009, **1794**(5):763-768.
98. Ambudkar SV, Dey S, Hrycyna CA, Ramachandra M, Pastan I, Gottesman MM: **Biochemical, cellular, and pharmacological aspects of the multidrug transporter 1.** *Annual review of pharmacology and toxicology* 1999, **39**(1):361-398.

99. Roepe PD: **What is the precise role of human MDR 1 protein in chemotherapeutic drug resistance.** *Current pharmaceutical design* 2000, **6**(3):241-260.
100. Johnstone RW, Ruefli AA, Smyth MJ: **Multiple physiological functions for multidrug transporter P-glycoprotein?** *Trends in biochemical sciences* 2000, **25**(1):1-6.
101. Li L, He Z, Pandey GK, Tsuchiya T, Luan S: **Functional cloning and characterization of a plant efflux carrier for multidrug and heavy metal detoxification.** *Journal of biological chemistry* 2002, **277**(7):5360-5368.
102. Noh B, Murphy AS, Spalding EP: **Multidrug resistance-like genes of Arabidopsis required for auxin transport and auxin-mediated development.** *The Plant Cell Online* 2001, **13**(11):2441-2454.
103. Nawrath C, Heck S, Parinthewong N, Metraux J-P: **EDS5, an essential component of salicylic acid-dependent signaling for disease resistance in Arabidopsis, is a member of the MATE transporter family.** *The Plant Cell Online* 2002, **14**(1):275-286.
104. Fragniere C, Serrano M, Abou-Mansour E, Metraux J-P, L'Haridon F: **Salicylic acid and its location in response to biotic and abiotic stress.** *FEBS letters*, **585**(12):1847-1852.
105. Strawn MA, Marr SK, Inoue K, Inada N, Zubieta C, Wildermuth MC: **Arabidopsis isochorismate synthase functional in pathogen-induced salicylate biosynthesis exhibits properties consistent with a role in diverse stress responses.** *Journal of biological chemistry* 2007, **282**(8):5919-5933.
106. Serrano M, Wang B, Aryal B, Garcion C, Abou-Mansour E, Heck S, Geisler M, Mauch F, Nawrath C, Metraux J-P: **Export of Salicylic Acid from the Chloroplast Requires the Multidrug and Toxin Extrusion-Like Transporter EDS5.** *Plant Physiology* 2013, **162**(4):1815-1821.
107. Taiz L: **Zeiger, 2006.** *Plant Physiology 4th ed Massachusetts: Sinauer Associates Inc.*
108. Levine A, Tenhaken R, Dixon R, Lamb C: **H₂O₂ from the oxidative burst orchestrates the plant hypersensitive disease resistance response.** *Cell* 1994, **79**(4):583-593.
109. Foyer CH, Bloom AJ, Queval G, Noctor G: **Photorespiratory metabolism: genes, mutants, energetics, and redox signaling.** *Annual review of plant biology* 2009, **60**:455-484.

110. Tu J, Ford R, Krass C: **Comparisons of chloroplasts and photosynthetic rates of plants infected and not infected by maize dwarf mosaic virus.** *Phytopathology* 1968, **58**(3):285-&.
111. Magyarosy AC, Buchanan BB, Schurmann P: **Effect of a systemic virus infection on chloroplast function and structure.** *Virology* 1973, **55**(2):426-438.
112. Reinero A, Beachy RN: **Reduced photosystem II activity and accumulation of viral coat protein in chloroplasts of leaves infected with Tobacco mosaic virus.** *Plant Physiology* 1989, **89**(1):111-116.
113. Seo S, Okamoto M, Iwai T, Iwano M, Fukui K, Isogai A, Nakajima N, Ohashi Y: **Reduced levels of chloroplast FtsH protein in Tobacco mosaic virus-infected tobacco leaves accelerate the hypersensitive reaction.** *The Plant Cell Online* 2000, **12**(6):917-932.
114. Wilhelmova N, Prochazkova D, Sindelarova M, Sindelar L: **Photosynthesis in leaves of *Nicotiana tabacum* L. infected with Tobacco mosaic virus.** *Photosynthetica* 2005, **43**(4):597-602.
115. Jensen S: **Photosynthesis respiration and other physiological relationships in barley infected with barley yellow dwarf virus.** *Phytopathology* 1968, **58**(2):204-&.
116. Van Kooten O, Meurs C, Van Loon L: **Photosynthetic electron transport in tobacco leaves infected with Tobacco mosaic virus.** *Physiologia Plantarum* 1990, **80**(3):446-452.
117. Balachandran S, Hurry V, Kelley S, Osmond C, Robinson S, Rohozinski J, Seaton G, Sims D: **Concepts of plant biotic stress. Some insights into the stress physiology of virus-infected plants, from the perspective of photosynthesis.** *Physiologia Plantarum* 1997, **100**(2):203-213.
118. Rahoutei J, GarcÃ-aâ€ • Luque I, BarÃ³n M: **Inhibition of photosynthesis by viral infection: effect on PSII structure and function.** *Physiologia Plantarum* 2000, **110**(2):286-292.
119. Guo D-P, Guo Y-P, Zhao J-P, Liu H, Peng Y, Wang Q-M, Chen J-S, Rao G-Z: **Photosynthetic rate and chlorophyll fluorescence in leaves of stem mustard *Brassica juncea* var. *tsatsai* after Turnip mosaic virus infection.** *Plant Science* 2005, **168**(1):57-63.
120. Zaitlin M, Hull R: **Plant virus-host interactions.** *Annual Review of Plant Physiology* 1987, **38**(1):291-315.

121. Price CA, Reardon EM: **Isolation of chloroplasts for protein synthesis from spinach and *Euglena gracilis* by centrifugation in silica soils.** *Methods in chloroplast molecular biology*/M Edelman, RB Hallick, and NH Chua, editors 1982.
122. Reinero A, Beachy RN: **Association of TMV coat protein with chloroplast membranes in virus-infected leaves.** *Plant molecular biology* 1986, **6**(5):291-301.
123. Goodman RN, Király Zn, Wood KR: **The biochemistry and physiology of plant disease:** University of Missouri Press; 1986.
124. Sindelarova M, Sindelar L: **TMV-RNA biosynthesis in the light-green and dark-green regions of tobacco leaves.** *Biologia plantarum* 2004, **48**(3):419-423.
125. Dietzgen RG, Zaitlin M: **Tobacco mosaic virus coat protein and the large subunit of the host protein ribulose-1, 5-bisphosphate carboxylase share a common antigenic determinant.** *Virology* 1986, **155**(1):262-266.
126. Lehti-Shiu MD, Shiu S-H: **Diversity, classification and function of the plant protein kinase superfamily.** *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**(1602):2619-2639.
127. Stone JM, Walker JC: **Plant protein kinase families and signal transduction.** *Plant Physiology* 1995, **108**(2):451-457.
128. Romeis T: **Protein kinases in the plant defence response.** *Current opinion in plant biology* 2001, **4**(5):407-414.
129. Tena G, Asai T, Chiu W-L, Sheen J: **Plant mitogen-activated protein kinase signaling cascades.** *Current opinion in plant biology* 2001, **4**(5):392-400.
130. Keegstra K: **Plant cell walls.** *Plant Physiology*, **154**(2):483-486.
131. Vorwerk S, Somerville S, Somerville C: **The role of plant cell wall polysaccharide composition in disease resistance.** *Trends in plant science* 2004, **9**(4):203-209.
132. Xoconostle-Cazares B, Xiang Y, Ruiz-Medrano R, Wang H-L, Monzer J, Yoo B-C, McFarland K, Franceschi VR, Lucas WJ: **Plant paralog to viral movement protein that potentiates transport of mRNA into the phloem.** *Science* 1999, **283**(5398):94-98.
133. Felle HH, Herrmann A, Hanstein S, Hückelhoven R, Kogel K-H: **Apoplasmic pH signaling in barley leaves attacked by the powdery mildew fungus *Blumeria graminis* f. sp. *hordei*.** *Molecular plant-microbe interactions* 2004, **17**(1):118-123.

134. Felle H, Herrmann A, Huckelhoven R, Kogel K-H: **Root-to-shoot signalling: apoplastic alkalization, a general stress response and defence factor in barley (*Hordeum vulgare*)**. *Protoplasma* 2005, **227**(1):17-24.
135. Wojtaszek P: **Oxidative burst: an early plant response to pathogen infection**. *Biochem J* 1997, **322**:681-692.
136. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR**. *Nucleic acids research* 2001, **29**(9):e45-e45.
137. Dardick C: **Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses**. *Molecular plant-microbe interactions* 2007, **20**(8):1004-1017.
138. Senthil G, Liu H, Puram V, Clark A, Stromberg A, Goodin M: **Specific and common changes in *Nicotiana benthamiana* gene expression in response to infection by enveloped viruses**. *Journal of General Virology* 2005, **86**(9):2615-2625.
139. D'arcy W: **The classification of the Solanaceae**. Hawkes, J, G., Lester, R, N., Skelding, A, D ed (s) *The biology and taxonomy of the Solanaceae* London, Academic Press for the Linnean Society 1979:3-47.
140. Borisjuk N, Borisjuk L, Petjuch G, Hemleben V: **Comparison of nuclear ribosomal RNA genes among *Solanum* species and other Solanaceae**. *Genome* 1994, **37**(2):271-279.
141. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research**. *Bioinformatics* 2005, **21**(18):3674-3676.

Chapter 4

***Red clover necrotic mosaic virus* down-regulates *Nicotiana benthamiana* systemic acquired resistance (NbSAR) gene expression in the first 6 hours of infection, followed by an up-regulation**

Chapter summary

In this chapter, I report a study on the regulation of *Nicotiana benthamiana* defense gene expression in an early *Red clover necrotic mosaic virus* (RCNMV) infection. The microarray study previously described in Chapter 3 revealed that Nb12974 was the only host gene and the only defense-related gene that was significantly modulated across all four time points sub 24 hours post RCNMV infection (FDR cutoff of 0.01). This study is focused on a further characterization of the Nb12974 gene. I used a bioinformatics tool to perform Nb12974 sequence and phylogenetic analysis, and its full-length cDNA sequence prediction. This computational analysis leads to the conclusion that Nb12974 is a gene member of the SAR8.2 family (systemic acquired resistance) and is therefore assigned the name NbSAR based on this functionality. The host gene NbSAR was also further functionally assayed for its effect on RCNMV infection by a transient gene silencing study in *N. benthamiana*. This functional assay suggests that the modulation of NbSAR gene expression plays a critical role in allowing RCNMV to establish an infection in the primary infected cell.

Abstract

Plant viruses are minimalist pathogens that are not only required to recruit host factors but also to manipulate the host defense system during the critical early period in order to produce a successful infection. This study is focused on a global examination of host defense responses in an early virus infection by utilizing a custom *Nicotiana benthamiana* array and *Red clover necrotic mosaic virus* (RCNMV). Host transcriptome profiles were analyzed at four different time points; 2, 6, 12 and 24 hours post RCNMV inoculation (hpi). The microarray study revealed that the majority of host genes were significantly down-regulated at 2, 6, and 24 hpi but up-regulated at 12 hpi (FDR cutoff of 0.01), suggesting that a single replication cycle within the primary infected cell takes 12-24 hours. In the first 6 hours, host defense-related functions were suppressed, but de-suppressed after 12 hpi. My hypothesis is that host defense systems were suppressed until after an initial round of infection occurred. An interesting observation in the microarray study is that a host systemic acquired resistance gene (NbSAR or Nb12974) was the only host gene and the only host defense-related gene that was significantly modulated across all four time points of the study (FDR cutoff of 0.01). Transiently silenced NbSAR expression study suggests that a down regulation of host SAR gene expression in the first 6 hours is driven by RCNMV, but once an initial infection is successful, an increase in host SAR gene expression was driven by an attempt from the host to defend itself from a further viral intrusion to other parts of the plant (systemic infection).

Keywords; SAR, systemic acquired resistance, *Nicotiana benthamiana*, RCNMV

Introduction

In spite of the extensive number of studies on host defense in response to virus infections, these previous reports are likely a summation of a few host defense genes studied individually. Moreover, these reports were focused on studying host defense at later points in the viral infection process. As such, there is still a knowledge gap in understanding the fundamental behavior of the host defense system as a whole at the early stages of a viral infection. Why is this issue so important? The underlying assumption of this project is that the immediate modulation of expression of host defense responses early in the infection process is a key and critical determinant of viral survival and reproduction; allowing the virus to establish a successful local infection leads to a subsequent robust systemic infection. This project is aimed at capturing the global co-regulation pattern of host defense genes early in the infection process. Thus, we designed our time-course study based on the onset of viral cell-to-cell movement. The study in this chapter is the first report of a genome-wide examination of host defense-related genes during the early stages of virus infection.

The selection of the experimental host model is a critical feature in our project. How viruses accomplish infection solely relies on their ability to modulate the host defense system. *N. benthamiana* is widely used as a model host in many plant virology studies since it is highly susceptible to a wide range of plant viruses (500+ plant viruses, even to those that are typically restricted to monocotyledonous hosts) and, most surprisingly, even to several animal viruses [1]. *N. benthamiana* is highly capable of (i) allowing plant viruses to establish a local infection, and (ii) facilitating a viral cell-to-cell movement leading to a

systemic infection. *N. benthamiana* is also susceptible to a broad array of other plant pathogens, including bacteria, oomycetes and fungi, making this plant species a cornerstone universal host for plant pathogen and host-pathogen interaction studies [2]. In comparison to the more commonly used model plant *Arabidopsis thaliana*, *N. benthamiana* is susceptible to a broader range of pathogens. It is likely that these pathogens may share a common strategy to hijack the *N. benthamiana* host defense system. We hypothesize that *N. benthamiana* possesses a specific suite of host defense genes that is vulnerable to a viral infection. In our laboratory, we are particularly interested in studying the interaction between *Red clover necrotic mosaic virus* (RCNMV, a typical plus strand RNA plant virus) and *N. benthamiana* as a host, as described in Chapter 3. We previously found that RCNMV takes about 12-24 hours to complete its infection in the primary infected cells. Therefore, we used *N. benthamiana* and RCNMV as the experimental host-virus combination to study the modulation of the host defense system by a virus early within the infection process (24 hours).

We are interested in studying the expression pattern of host defense genes that are co-regulated by virus infection. Therefore, in this project, rather than studying one host defense gene at a time, we utilized our custom *N. benthamiana* microarray to enable a high-throughput measurement of the differential expression of hundreds of host defense genes at 2, 6, 12, and 24 hours post RCNMV inoculation (hpi). Our microarray is the first commercially available *N. benthamiana* microarray. The development of microarray design was based on our 13,000 *N. benthamiana* unigene collection, covering about 38% of the transcriptome. This custom microarray has a capacity to measure the expression of at least

170 host defense-related genes, including those involving in the innate and adaptive immunities. The construction of the *N. benthamiana* microarray is thoroughly described in Chapter 2.

Our microarray study reveals the unique observation that Nb12974 is the only host gene that was modulated across all four time points of the study at 2, 6, 12, and 24 hours post RCNMV inoculation. Nb12974 has the highest similarity to NbSAR8.2m which is a gene member of the SAR8.2 family. Therefore, this study is focused on the further characterization of Nb12974. The main hypothesis is that if Nb12974 is an essential host defense factor involved early in the RCNMV infection process, the genetic modulation of this gene prior to RCNMV inoculation would affect viral infectivity. To test this hypothesis, expression of Nb12974 was transiently suppressed in *N. benthamiana* prior to RCNMV inoculation. The associated impact of the genetically pre-modulated host Nb12974 gene to viral infection efficiency is indirectly assessed by two different quantifying methods: (i) a quantitative measurement of viral RNA accumulation by qRT-PCR analysis, and (ii) a quantitative measurement of viral GFP accumulation by IVIS fluorescence imaging assay.

Systemic Acquired Resistance (SAR)

Systemic acquired resistance (SAR) is characterized by an enhanced state of a long-lasting resistance against infection by a wide range of pathogens, extending to plant tissues distant from the initial infection site [3, 4]. The term SAR8.2 was initially assigned to several related genes with unknown functions that were highly expressed during screening of upper,

non-inoculated *N. tabacum* cv. Xanthi leaves that exhibited SAR characteristics following the development of the hypersensitive reaction response in lower leaves inoculated with *Tobacco mosaic virus* (TMV) [5]. The initial discovery of increased SAR8.2 gene expression in plants undergoing SAR implies the association of this gene family to a systemic defense response. The increased SAR8.2 gene expression is also associated with other stimuli, including various environmental stresses, compatible plant host-pathogen interactions, as well as agents inducing SAR characteristics such as salicylic acid (SA)[5-11]. The common features of SAR8.2 protein structure are: (i) a signal peptide at N-terminus indicating secretion from rough endoplasmic reticulum (ER), (ii) a cysteine-rich C-terminus, and (iii) being highly basic with a molecular weight ranging from 3.0 to 9.4 kDa [12]. Members of the SAR8.2 family are found in Solanaceae plants (Table 2), including 15 members identified in *N. tabacum* (tobacco) and 4 members in *Capsicum annuum* L. (pepper).

Thus far, NbSAR8.2m is the first and only SAR8.2 homolog identified from *N. benthamiana* [12]. The discovery of NbSAR8.2m in *N. benthamiana* came from a PCR-based cDNA subtraction library constructed from mRNA of *N. benthamiana* Domin. healthy against infected with a plant pathogenic fungi, *Colletotrichum orbiculare* [12]. Despite the identification of NbSAR8.2m from the original report as a partial coding sequence (CDS), the phylogenetic tree analysis shows a relatively high similarity between NbSAR8.2m and both the coding and 3'-UTR of NtSAR8.2m from *N. tabacum*. Expression of NbSAR8.2m occurred in healthy plants but was gradually increased from 0 to 4 days post *C. orbiculare* infection with a highest up-regulating peak of 8-fold induction at day 4, and the magnitude subsided afterward. Virus-induced gene silencing of NbSAR8.2m conducted on *N.*

benthamiana reduced its expression and resulted in the development of disease symptoms 24 hours earlier than control plants, indicating that NbSAR8.2m affects the length of the biotrophic fungal infection phase.

Result and discussion

RCNMV suppressed *N. benthamiana* defense-related function in the first 6 hours of infection followed by a de-suppression

Host defense response early events in an RCNMV infection were globally examined by utilizing our custom *N. benthamiana* array. The microarray chip is a high-throughput technology that enables the monitoring of transcription changes of many thousands of host genes simultaneously. Therefore, instead of studying only one host defense gene, this study is aimed at observing the expression flux of many defense genes simultaneously. This study is the first report of genome-wide study of host defense-related genes early during a virus infection. A thorough detail of the construction of the *N. benthamiana* microarray is described in Chapter 2. Our custom microarray represents 13,014 host genes with a coverage of approximately 38% of the *N. benthamiana* transcriptome. About 60% (7,841 genes) of the microarray genes have at least one GO annotation (Gene Ontology database). And 27,254 GO terms were assigned to these 7,841 microarray genes. Approximately 2% (170 genes) of the microarray genes with GO annotation (7,841 genes) are assigned to “defense-related function categories”. These 170 defense-related genes are classified into six different GO

categories (Table 1). The majority of the defense-related genes fall into the “defense response to bacterium category (GO ID 42742)”.

Host transcriptome profiles and host gene function (represented by GO term) were analyzed at four different time points of 2, 6, 12 and 24 hpi. All 27,254 GO terms assigned to our microarray genes and their associated t-statistic values were inputted into Gene Set Enrichment Analysis (GSEA) in order to determine which host functions are significantly enriched early in an RCNMV infection. The t-statistic value is derived from a differential gene expression comparing virus-infected samples to mocks at each point of study.

The GSEA result (Figure 1) reveals that host defense response function (GO ID 42742) was negatively correlated to RCNMV infection in the first 6 hours (2 and 6 hpi), but then was positively correlated at 12 hpi. This indicates that host defense gene expression was suppressed in the first 6 hours of the RCNMV infection but then de-suppressed at 12 hpi. Decreasing host defense gene expression within the first 6 hours was likely driven by the RCNMV infection, which was expected in the susceptible *N. benthamiana* host.

Early host defense gene suppression may play a crucial role for a successful viral infection. However, during this host defense gene suppression, apoptosis-related host activities were increased (Figure 1) suggesting that the host utilized an apoptosis pathway (GO 6915) to immediately confine the infected area. This finding leads to the belief that host apoptosis is a primary and acute host defense response in *N. benthamiana* against RCNMV infection at the early stages 6 hpi. However, in a longer defense response, the host shifted its

defense strategy to increasing defense gene expression (GO 42742) which is more efficient than apoptosis and does not cause host cell death.

An interesting observation is that all six defense-related GO terms assigned to our microarray genes (Table 1) are all significantly enriched in the top 20 GSEA result for either negative correlation or positive correlation to an early RCNMV infection (Figure 1). This suggests that the ability to immediately modulate the host defense response is an essential strategy that RCNMV employs early in order to produce a successful infection.

***N. benthamiana* systemic acquired resistance (NbSAR or Nb12974) is the only host gene identified that was modulated across 2, 6, 12 and 24 hours post RCNMV inoculation**

A genome-wide study of the modulation of *N. benthamiana* host gene expression early in an RCNMV infection is thoroughly described in Chapter 3. Host transcriptome profiles were analyzed at four different time points of 2, 6, 12 and 24 hours post-inoculation (hpi). Microarray data was subjected to statistical analysis and 1,654 genes exhibited differential expression at an FDR cutoff of 0.01. This microarray analysis revealed that Nb12974 is the only host gene that was significantly modulated across all four time points of study between 2 and 24 hpi at an FDR cutoff of 0.01 (Figure 2). With a confirmation from microarray and qRT-PCR analyses, the expression of Nb12974 was down-regulated in the first 6 hours of an RCNMV infection followed by an up-regulation (Figure 2).

A functional annotation of Nb12974 to NbSAR

i) Nb12974 sequence analysis

The unigene Nb12974 sequence was searched against the GenBank nucleotide collection (nr/nt) and non-redundant protein sequences (nr) databases by using BLASTN and BLASTX algorithms, respectively [13, 14]. The BLASTN result showed that the Nb12974 nucleotide sequence is matched to NbSAR8.2m (accession number AY644732) with 99% identity, and an E-value = 7×10^{-170} (Figure 6). Also, the BLASTX result shows that the Nb12974 amino acid sequence is matched to NbSAR8.2m (accession number AAV65395) with 100% identity, and an E-value = 5×10^{-11} (Figure 7). This leads to the conclusion that Nb12974 is a gene member of the SAR8.2 family and Nb12974 is possibly an NbSAR8.2m gene [12]. An interesting observation is that a BLAST search of the Nb12974 sequence against the GO database (Gene Ontology [15]) does not return a result, suggesting that none of the sequences in the GO database has a significant similarity match to Nb12974.

NbSAR8.2m was originally cloned and identified by Shan & Goodwin [12]. NbSAR8.2m is the only SAR gene that has been discovered to date in *N. benthamiana*. In spite of its partial coding sequence (Figure 8), it was assigned the name NbSAR8.2m due to its high similarity to NtSAR8.2m from tobacco. However, a BLAST search of NbSAR8.2m and NtSAR8.2m against the *N. benthamiana* genome from the SolGenomics network returned a different result. This observation was quite unexpected because the difference in results indicates that there is another gene that has a closer similarity to NtSAR8.2m than the gene that was originally assigned the name NbSAR8.2m. This leads to the conclusion that the

gene sequence that was originally assigned as NbSAR8.2m may not actually be NbSAR8.2m. Thus while Nb12974 is the same gene that Shan & Goodwin assigned to NbSAR8.2m, Nb12974 may not be NbSAR8.2m. This result is discussed in the next section.

ii) A prediction of the full-length cDNA sequence of Nb12974

Nb12974 is 621 base pairs long. Using the ExPASy translation tool [16, 17], a standard genetic code, and its highest similarity to NbSAR8.2m, a translation frame of Nb12974 was chosen from frame +2 in the 5' to 3' direction. This translation frame yields a peptide of 60 residues in length (Figure 8). The early translational termination coupled with a poly (A) tail at the 3'-end suggests that the Nb12974 sequence is an mRNA segment containing the complete 3' region of the CDS (coding sequence) and the 3' UTR (untranslated region) of approximately 437 nucleotides. However, the lack of a start codon (ATG) and a translation starting from frame +2 indicates that the Nb12974 sequence does not contain a complete CDS at the 5'-end. In this section, I will focus on how to use a bioinformatics tool to predict potential sequence information for a complete CDS at the 5' end of Nb12974.

A comparison of all publicly available genes in the SAR8.2 family, 15 members from *N. tabacum* (tobacco) and 4 members from *C. annuum* (pepper) (Table 2), reveals that the size of the full-length SAR8.2 cDNA ranges from 162 to 333 base pairs with approximately 54 to 111 amino acids. All 15 members of the NtSAR8.2 family from tobacco contain a signature motif "MFSKT" at the 5' cDNA end (Figure 9). According to the rule of von

Heijne (1983) [18], this motif is the first five amino acids of the signal peptide at the N-terminus, indicating secretion of the protein from the rough endoplasmic reticulum (ER). Given the close genetic relationship to *N. tabacum*, my assumption is that a SAR8.2 homolog in *N. benthamiana* may also contain a similar signature motif at its 5' cDNA end.

With this information, the multiple alignments of NtSAR8.2m and Nb12974 to the *N. benthamiana* draft genome (SolGenomic network database [19]) will presumably yield the potential full-length cDNAs of *N. benthamiana* SAR8.2 homologs. In spite of their high similarity (71% amino acid identity with E-value = 2×10^{-26}), the NtSAR8.2m and Nb12974 nucleotide sequences are mapped to different regions of the *N. benthamiana* genome.

The BLASTN prediction using the BLOSUM62 matrix alignment score shows that NtSAR8.2m from position 1 to 115 (5' to 3' direction) was 96% identity matched to a genome number Niben.v0.3.Scf25076466 and NtSAR8.2m from position 114 to 203 (5' to 3' direction) was 85% identity matched to a genome number Niben.v0.3.Scf24801001. Whereas, Nb12974 was 100% identity matched to two different regions on the same genomic scaffold number Niben.v0.4.2.Scf6278. These two different regions on the same genomic segment may imply the isoform property of this gene.

A full-length cDNA was predicted from an assembly of the matched exons that are discontinuously dispersed on the genomic hit regions. The discontinuous genomic DNA segments or exons representing a high similarity to NtSAR8.2 and Nb12974 are likely the result of RNA splicing events [20] and so it may offer the splicing information of SAR8.2 genes.

The predicted full-length cDNA matched to NtSAR8.2m is 67 amino acids and assigned the name to “predFLNbNtSAR8.2m”. The BLASTP alignment between NtSAR8.2m and predFLNbNtSAR8.2m shows 84% similarity with an E-value = 2×10^{-38} (Figure 10).

The predicted full-length cDNA match to Nb12974 appears to have two isoforms which are assigned the names to “predFLNb12974_1” for isoform 1 and “predFLNb12974_2” for isoform 2. Both isoforms are translated into peptides 65 residues in length with 100% similarity to Nb12974 (E-value of 4×10^{-42} and 3×10^{-42} for isoforms 1 and 2, respectively; Figure 11). These two isoforms contain a single nucleotide difference at residue 9 (Figure 12) resulting in cysteine (isoform 1) or tryptophan (isoform 2) (Figure 11). Recently, a survey of the *N. benthamiana* genome identified 164,595 single nucleotide polymorphisms (SNP), representing a polymorphism rate of 0.007% throughout the genome [21]. The occurrence of these two potential isoforms of the same SAR gene in *N. benthamiana* may reflect its SNP characteristics; in this case, there are two alleles occurring in the coding region.

iii) SAR8.2 phylogenetic analysis

The predicted amino acid sequences of predFLNbNtSAR8.2m, predFLNb12974_1, and predFLNb12974_2, along with all available SAR8.2 genes (Table 2) were then used in a phylogenetic analysis in order to re-confirm the identification previously performed by Shan & Goodwin [12]. Phylogenetic analysis (Figure 15), using the *hdrC2* gene from

Methanocaldococcus jannaschii as a control, reveals that predFLNbNtSAR8.2m, predFLNb12974_1, and predFLNb12974_2 are closely related to NtSAR8.2m. However, predFLNbNtSAR8.2m is closer to NtSAR8.2m than the other two genes (Figure 15). The BLASTP alignment shows that predFLNbNtSAR8.2m is 97 % identical to predFLNb12974_1 and predFLNb12974_2 with E-value 2×10^{-31} and 1×10^{-31} , respectively (Figure 13).

It is likely that predFLNbNtSAR8.2m is the NbSAR8.2m homolog found in *N. benthamiana*. Whereas predFLNb12974_1 and predFLNb12974_2 (the predicted isoforms of full-length Nb12974 cDNA) are likely not the homologs of NtSAR8.2m. The computational evidence suggests that predFLNb12974_1 and predFLNb12974_2 may be novel gene members of the SAR8.2 family found in *N. benthamiana*. In this study, Nb12974 is therefore assigned its name and functionality to “NbSAR”.

The complete coding region covering the entire open reading frame (ORF) from start to stop codon is essential for clone construction and gene over-expression studies. The study in this chapter only presents the bioinformatics and computational evidence of the potential full-length sequence of the SAR8.2 homolog in *N. benthamiana*. Further work will be required to retrieve the actual clone and verify the full-length sequence. Other than using bioinformatics tools, the common alternative experimental method employed to retrieve the full-length cDNA sequence information from the 5' terminus is 5'RACE (rapid amplification of cDNA ends).

Pre-silenced NbSAR (Nb12974) promoted RCNMV RNA-1 and RNA-2 replication at 24 hpi and increased RCNMV R1SG1-sGFP accumulation

NbSAR (or Nb12974) gene expression was suppressed in *N. benthamiana* by utilizing the TRV1:TRV2-NbSAR VIGS vector system [22] (Figure 3). NbSAR gene was pre-silenced 10 days prior to RCNMV inoculation. The expression of pre-silenced NbSAR significantly remained suppressed > 2 fold throughout the 72 hours of RCNMV infection ($\alpha = 0.05$). The pre-silenced NbSAR affected RNA-1 and RNA-2 replication (Figure 4) as well as sGFP accumulation from R1SG1 (Figure 5). The RNA-1 POL, RNA-1 CP and RNA-2 MP RNA levels were significantly increased > 2 fold at 24 hpi ($\alpha = 0.1$). Also, R1SG1-sGFP accumulation was significantly increased > 3 fold after 3 days ($\alpha = 0.05$). However, the pre-silenced NbSAR significantly decreased RNA-1 POL and CP RNA accumulation > 2 fold in the first 6 hours of RCNMV infection ($\alpha = 0.1$). At 12 hpi, both RNA-1 and RNA-2 RNA levels were not affected by the pre-silenced NbSAR. An interesting observation is that RNA-2 RNA level was significantly increased > 2 fold at 3 hpi, followed by decrease > 2 fold at 6 hpi ($\alpha = 0.1$).

This result suggests that the pre-silenced NbSAR does not promote RCNMV replication until the late infection phase at 24 hpi. However, it promotes RNA-2 replication as early as 3 hpi. In a wild-type RCNMV infection of *N. benthamiana*, microarray data showed that RNA-2 was not accumulated until after 12 hours (Chapter 3, Figure 21). RNA-2 is encoded for a viral movement protein which is presumably not required until viral cell-to-cell movement occurs at 12 hpi (?). Also, RNA-2 is encoded a viral suppressor of RNA

silencing (VSR) [23]. Taken together, the sudden increase in RNA-2 MP at 3 hpi by the pre-silenced NbSAR suggests that NbSAR in a wild-type RCNMV infection of *N. benthamiana* may likely take part in regulating the expression of the viral silencing suppressor, rather than viral movement.

Conclusion

RCNMV down-regulated a subset of *N. benthamiana* defense-related genes in the first 6 hours of infection followed by up-regulation. During this early host defense gene suppression, RCNMV increased host apoptosis function as a primary and acute host defense response in *N. benthamiana* to immediately confine the infected area. However, in a longer defense response, the host shifted its defense strategy to increasing defense gene expression. The temporal microarray study revealed that Nb12974 is the only host gene on the array that was modulated across all four time points of study at 2, 6, 12, and 24 hpi. Nb12974 was previously known as NbSAR8.2m; however, a bioinformatics prediction indicates that Nb12974 is not NbSAR8.2m and therefore, it may be a novel gene of the SAR8.2 family in *N. benthamiana*. In this study, Nb12974 is assigned a name and functionality to “NbSAR”.

The alteration of NbSAR gene expression early in an RCNMV infection demonstrates its contribution to a successful virus infection. Microarray and qRT-PCR analysis indicated that NbSAR expression in a wild-type RCNMV infected *N. benthamiana* was down-regulated in the first 6 hours post infection, followed by an up-regulation. The early down-regulation of NbSAR in a wild-type RCNMV infected host was to presumably allow the

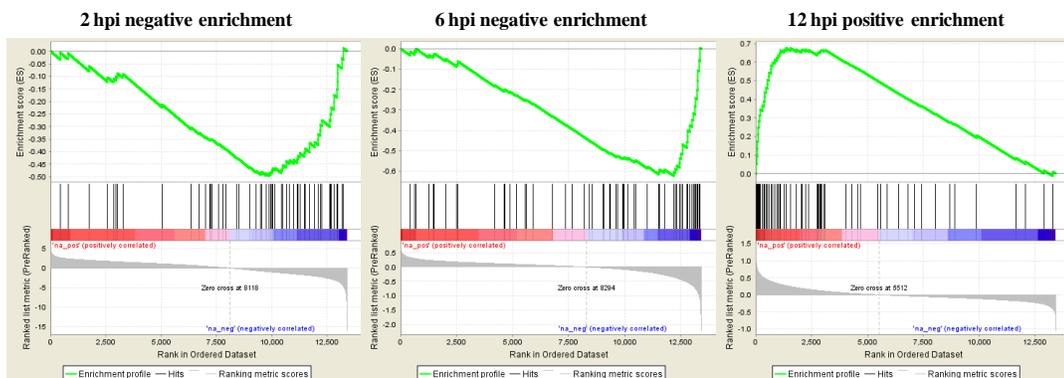
establishment of the viral infection in the primary infected cells within 12 hours. But once the virus completes the infection, NbSAR expression was allowed to be increased after 12 hours to engage in a host defense mechanism. It is likely that the early down-regulation of NbSAR in the first 6 hours was driven by RCNMV in order to not only disarm host defense but also to presumably helped promote the expression of the viral silencing suppressor. This assumption is drawn from the result that the pre-silenced NbSAR promoted RNA-2 replication as early as 3 hpi. On the other hand, the late up-regulation of NbSAR after 6 hpi in a wild-type RCNMV infected host was likely driven by the host itself in order to increase systemic resistance toward a further viral invasion. This latter assumption is consistent with the results from silencing experiments showing that the pre-silenced NbSAR did promote both RCNMV RNA-1 and RNA-2 replications until 24 hpi. The counter interaction driven by the host may actually benefit the virus in terms of preventing a fatal effect to the host while viruses are still able to maintain their infectivity.

Table 1 Numbers of *N. benthamiana* microarray genes assigned to the defense-related functional categories based on GO database. A total of 170 *N. benthamiana* microarray genes are assigned to the GO defense-related categories.

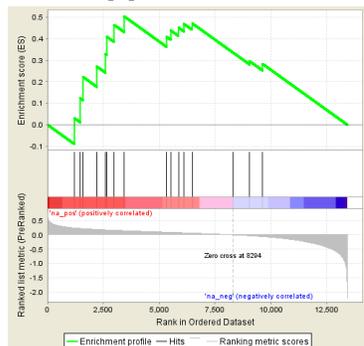
GO ID	GO term	% of all genes with defense-related function (no. of genes in the assigned GO term)
6915	apoptosis	9% (16)
6952	defense response	30% (50)
9737	response to abscisic acid stimulus	11% (19)
42742	defense response to bacterium	37% (63)
50832	defense response to fungus	9% (15)
10200	response to chitin	12% (20)

Figure 1 Host defense-related functions were highly enriched and correlated to an early RCNMV infection at time points 2, 6, 12 and 24 hpi. Host function enrichment is analyzed by using GSEA (Gene Set Enrichment Analysis) software. The GSEA enrichment plots demonstrate if the GO term is positively or negatively enriched and correlated to RCNMV infection. The defense-related GO terms displayed in this figure are in the top 20 result of either negatively or positively enriched GO terms at each time point of study. GSEA was performed using the microarray data and all 1,931 GO terms that are assigned to the microarray genes. These GO terms are pre-ranked by their associated t-statistic values. The t-statistic values was derived from a microarray differential gene expression data comparing between virus-infected plants and mocks at each point of study.

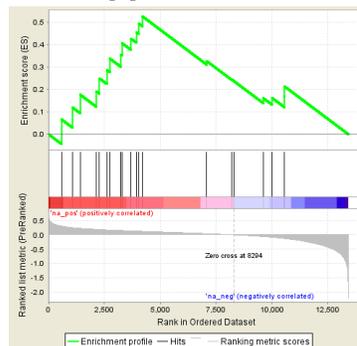
GO 42742: defense response to bacterium



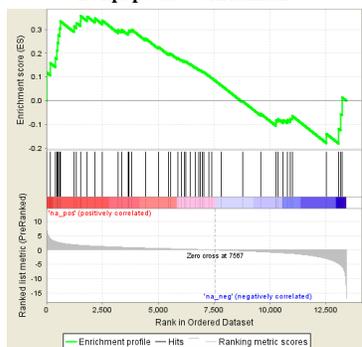
GO 6915: apoptosis
6 hpi positive enrichment



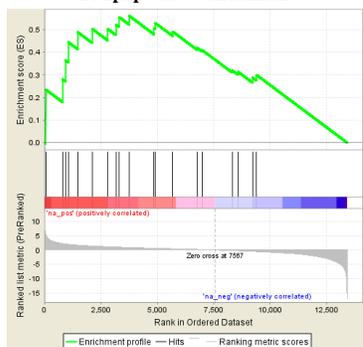
GO 10200: response to chitin
6 hpi positive enrichment



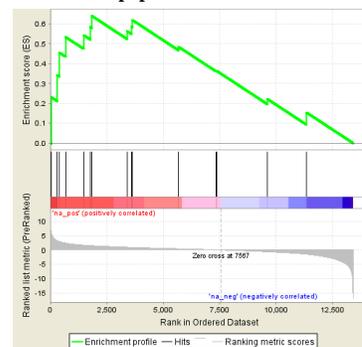
GO 6952: defense response
24 hpi positive enrichment



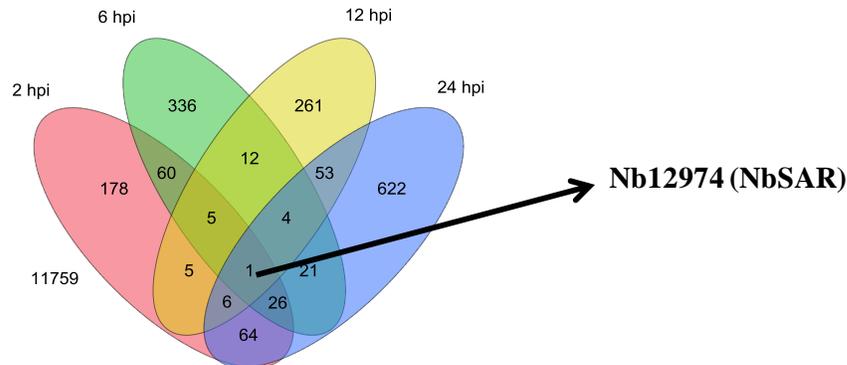
GO 9737: response to abscisic acid stimulus
24 hpi positive enrichment



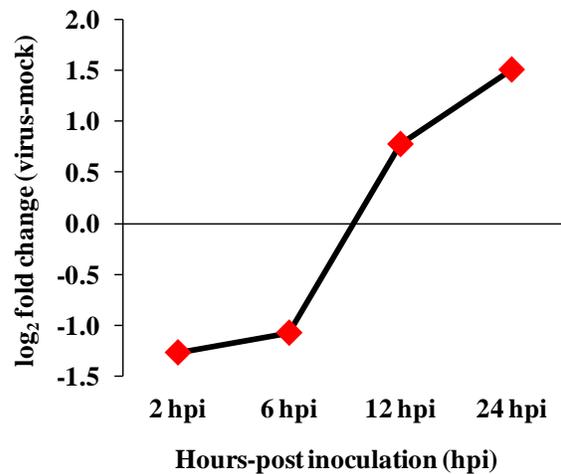
GO 50832: defense response to fungus
24 hpi positive enrichment



A) Venn diagram of microarray data



B) Microarray data of Nb12974



C) qRT-PCR analysis of Nb12974

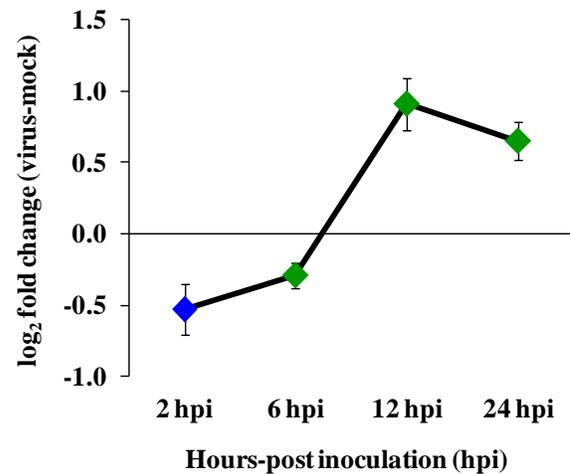


Figure 2 Nb12974 (NbSAR) was the only host gene in the microarray study that was modulated across all four times points of study at 2, 6, 12, and 24 hours-post RCNMV inoculation. A) Venn diagram of microarray data displaying an overlap of host genes that were a significantly differentially expressed with at an FDR cut off of 0.01 across other time points of study. B) Microarray data of Nb12974 (red marker indicates significance at FDR cutoff of 0.01). C) qRT-PCR analysis of Nb12974 (blue marker indicates significance at $\alpha = 0.1$ and green marker indicates significance at $\alpha = 0.05$).

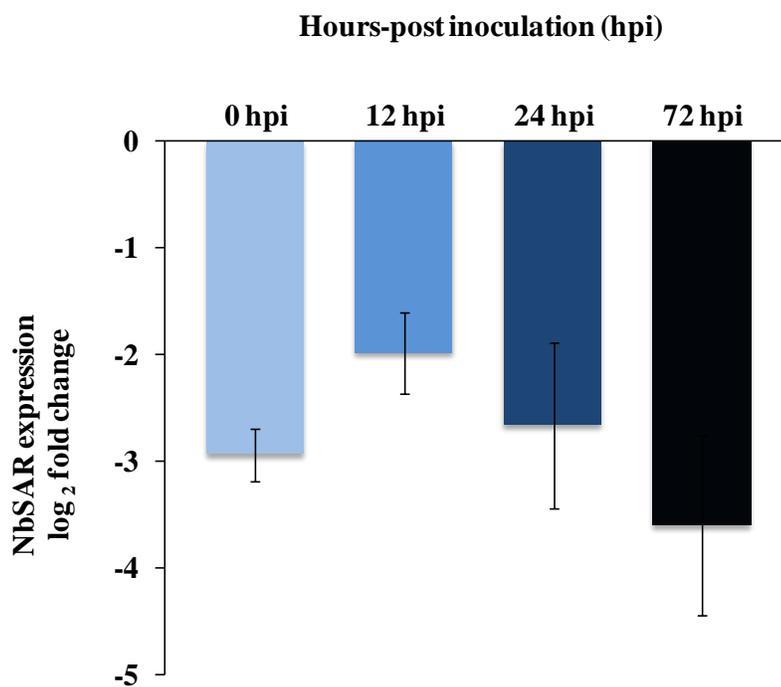


Figure 3 NbSAR (Nb12974) gene expression was suppressed in *N. benthamiana* by using TRV-VIGS vector system. The silencing vectors were delivered into plant cells by an agroinfiltration method using 1 ml needleless syringe. NbSAR was pre-silenced by the silencing vectors TRV1:TRV2-NbSAR for 10 days prior to RCNMV inoculation. The pre-silenced NbSAR plants were inoculated with *in vitro* T7 RNA-1 transcript and *in vitro* T7 RNA-2 transcript, and then harvested between 0 and 72 hours-post RCNMV inoculation (hpi). The qRT-PCR analysis showed that the expression of pre-silenced NbSAR significantly remained suppressed over 72 hpi ($\alpha = 0.05$). Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis.

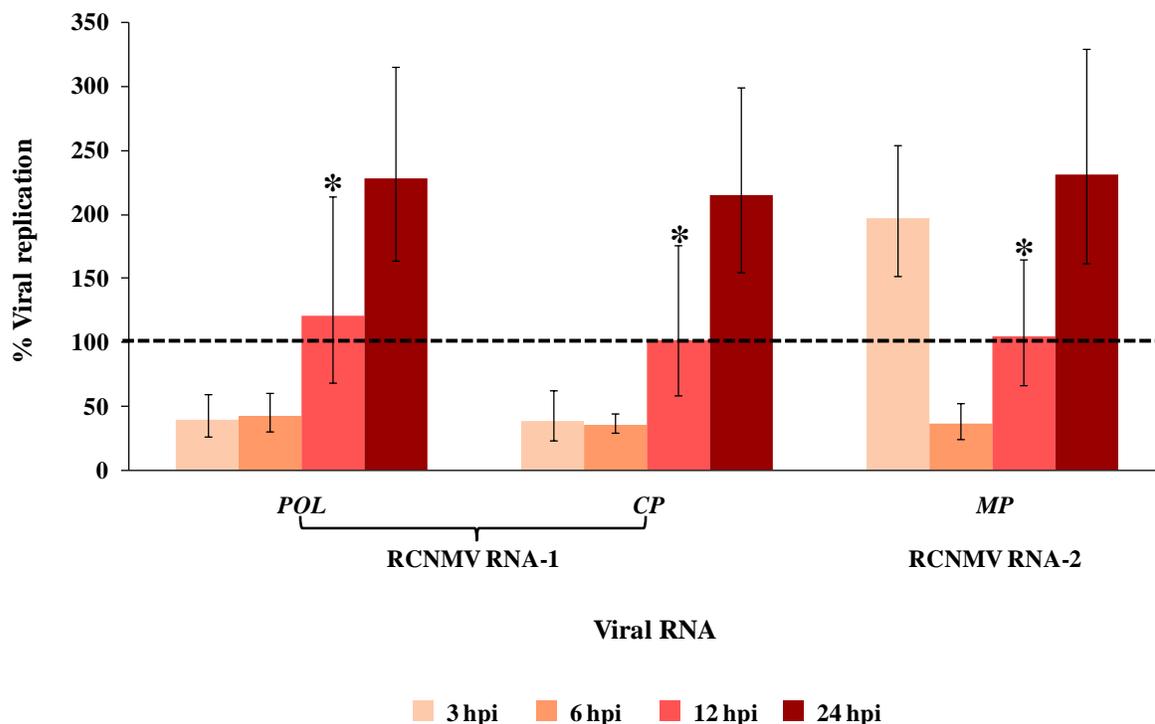


Figure 4 Pre-silenced NbSAR (Nb12974) affected RCNMV RNA accumulation at 3, 6, and 24 hours-post inoculation (hpi), but not at 12 hpi. The expression of NbSAR was transiently suppressed by TRV1:TRV2-NbSAR (TRV-VIGS system) 10 days prior to RCNMV inoculation. The pre-silenced NbSAR plants were inoculated with *in vitro* T7 RNA-1 transcript and *in vitro* T7 RNA-2 transcript, and then harvested between 3 and 24 hours after RCNMV inoculation. RCNMV RNAs were measured by qRT-PCR analysis. The viral RNA accumulation was computed by a relative quantification method, and displayed as % viral replication. Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. All data shown in this figure was significant at $\alpha = 0.1$, except the data with an asterisk symbol (*).

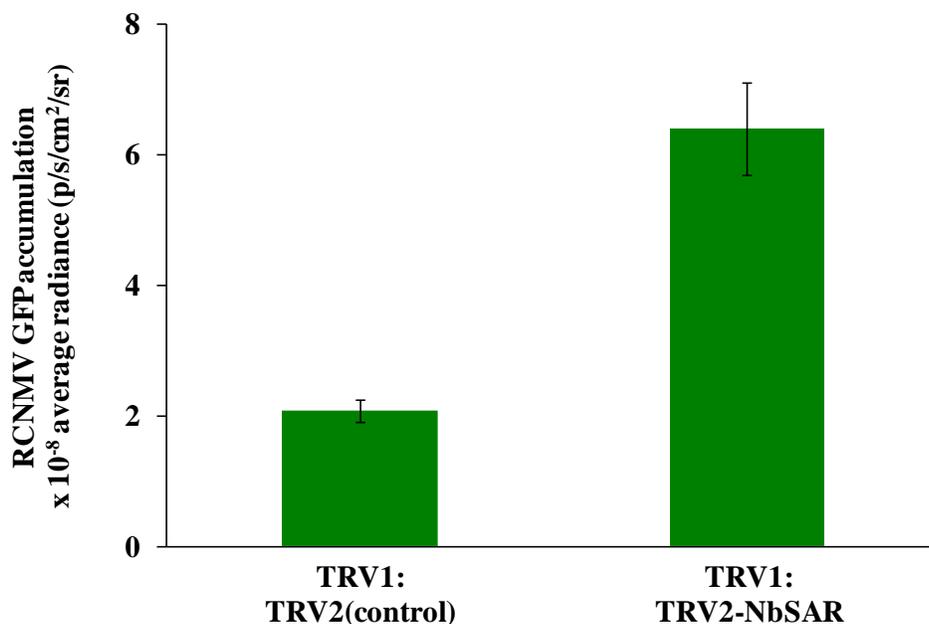


Figure 5 Pre-silenced NbSAR (Nb12974) promoted RCNMV GFP accumulation at 3 days post inoculation. The expression of NbSAR was transiently suppressed using TRV-VIGS system 10 days prior to RCNMV inoculation. The pre-silenced NbSAR plants were inoculated with *in vitro* T7 R1SG1 transcript and *in vitro* T7 RNA-2 transcript, and then harvested at 3 days after RCNMV inoculation. Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis ($\alpha = 0.05$).

Unigene ID (Query)	BLASTN hit	Accession number	Query coverage	E-value	Identity	Gaps	Strand
Nb12974 621 bp	NbSAR8.2m (<i>N. benthamiana</i>) 331 bp, partial CDS	AY644732	53%	7×10^{-170}	330/331 (99%)	0/331 (0%)	+/+

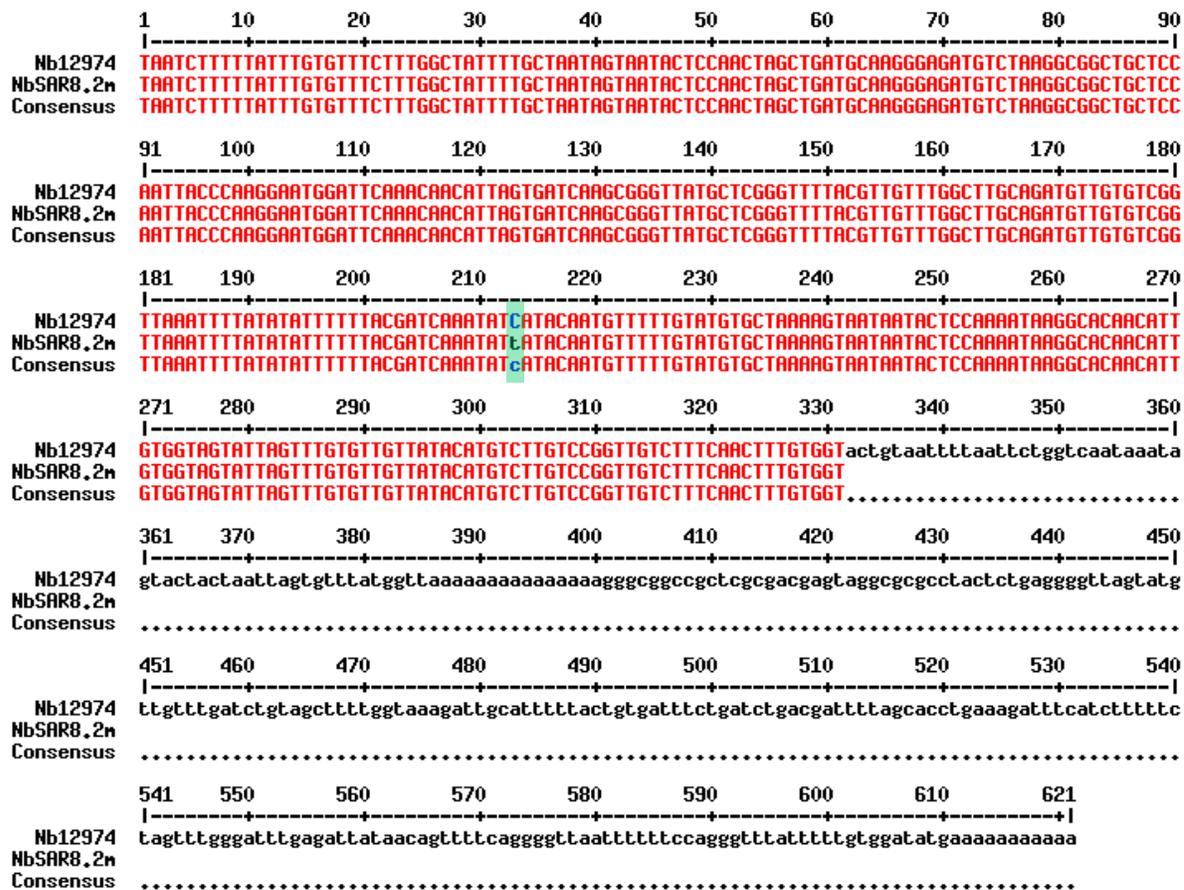


Figure 6 BLASTN information and alignment between Nb12974 and NbSAR8.2m nucleotide sequences. A green box shows one nucleotide difference between Nb12974 and NbSAR8.2m.

Unigene ID (Query)	BLASTX hit	Accession number	Query coverage	E-value	Identity	Gaps	Frame
Nb12974 60 aa	NbSAR8.2m (<i>N. benthamiana</i>) 60 aa, partial	AAV65395	100%	5×10^{-11}	60/60 (100%)	0/60 (0%)	+2

1 10 20 30 40 50 60
 |-----+-----+-----+-----+-----+-----+-----|
NB12974 **NLFICVSLAILLIYILQLADAREMSKAAPITQGMDSNNISDQAGYARVLRCLACRCCVG**
NbSAR8.2m **NLFICVSLAILLIYILQLADAREMSKAAPITQGMDSNNISDQAGYARVLRCLACRCCVG**
Consensus **NLFICVSLAILLIYILQLADAREMSKAAPITQGMDSNNISDQAGYARVLRCLACRCCVG**

Figure 7 BLASTX information and alignment between Nb12974 and NbSAR8.2m amino acid sequences.

```

taatctttttatattgtgtttctttggctatatttgctaataactccaactagctgat
  N L F I C V S L A I L L I V I L Q L A D
gcaagggagatgtctaaggcggtgctccaattaccaaggaatggattcaacaacatt
  A R E M S K A A A P I T Q G M D S N N I
agtgatcaagcgggttatgctcgggttttacggttggcttgcagatggtgtgctggt
  S D Q A G Y A R V L R C L A C R C C V G
taaattttatataatttttacgatcaaatacatacaatgttttgtatgtgctaaaagt
- I L Y I F Y D Q I S Y N V F V C A K S
aataatactccaaaataaggcacaacattgtggtagtattagtttgtggtttatacatg
  N N T P K - G T T L W - Y - F V L L Y M
tcttgccgggttgtctttcaactttgtggtactgtaatttttaattctgggtcaataaatag
  S C P V V F Q L C G T V I L I L V N K -
tactactaattagtgtttatggttaaaaaaaaaaaaaaaaaagggcggcgcgctcgcgacgagt
  Y Y - L V F M V K K K K G R P L A T S
agggcgcgcctactctgaggggttagtatggttggatctgtagcttttggtaaagattg
  R R A Y S E G L V C C L I C S F W - R L
catttttactgtgatttctgatctgacgatttttagcacctgaaagatttcatctttttct
  H F Y C D F - S D D F S T - K I S S F S
agtttgggatttgagattataacagttttcaggggttaattttttccaggggtttattttt
  S L G F E I I T V F R G - F F P G F I F
gtggatatgaaaaaaaaaaaaa
  V D M K K K

```

Figure 8 The predicted amino acid sequence of Nb12974 using ExPASy translation tool and standard genetic code. A translation starts from a direction 5' to 3' of Nb12974 and frame +2. A green box indicates a stop codon. Blue hi-lighted residues represent the selected translational frame for Nb12974, yielding a peptide of 60 residues in length. Nb12974 is a partial coding sequence with incomplete coding at its 5' end.

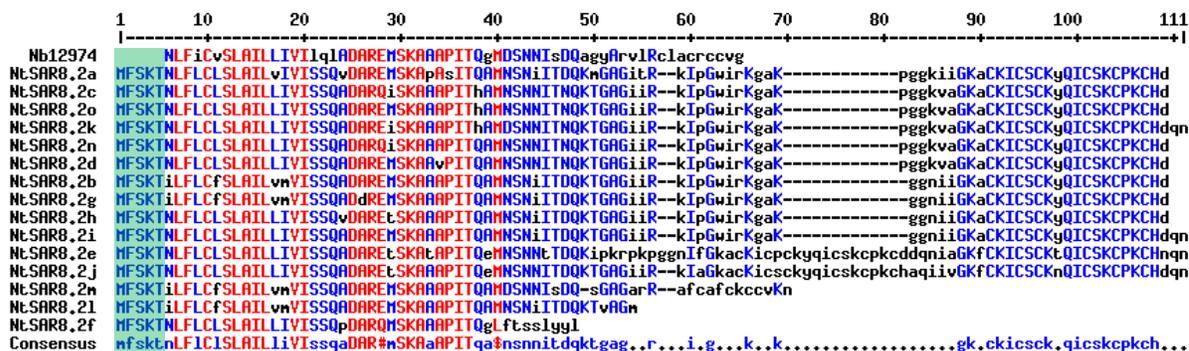


Figure 9 Signature motif elucidation of a signal peptide at 5' cDNA end of all *N. tabacum* SAR8.2 gene members. A green box shows the signature motif “MFSKT” at the 5' end.

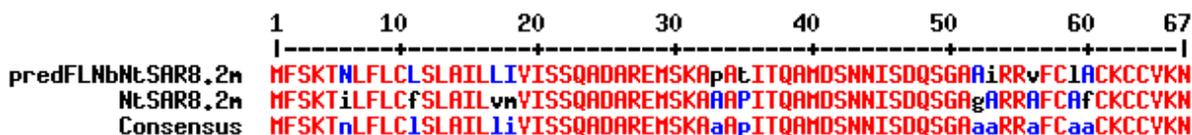


Figure 10 Amino acid sequence comparison between NtSAR8.2m and predFLNbNtSAR8.2m using BLASTP alignment.



Figure 11 Amino acid sequence comparison between Nb12974 and the two predicted isoforms of full-length Nb12974 (predFLNb12974_1 and predFLNb12974_2) using BLASTP alignment. A green box indicates one amino acid difference between the two predicted isoforms of full-length Nb12974 cDNA.

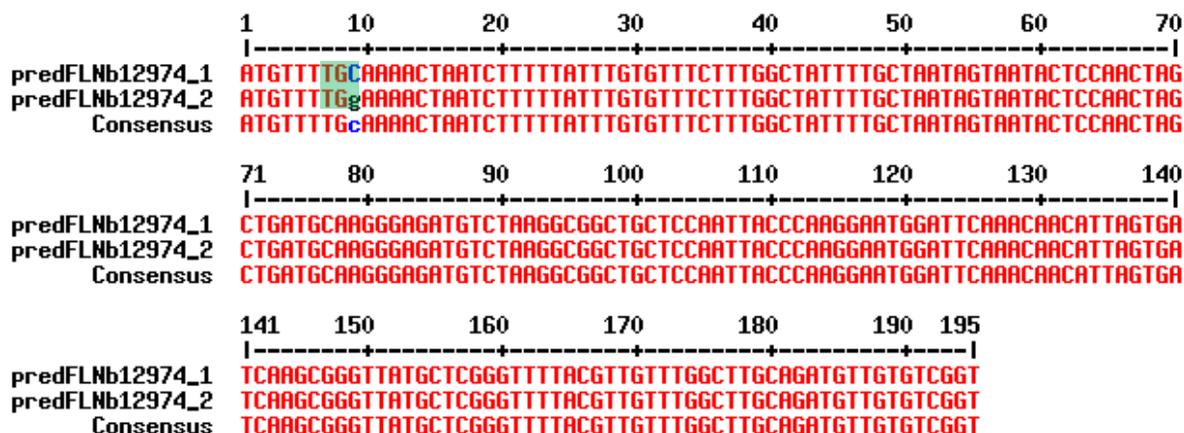


Figure 12 Nucleotide sequence comparison between the two predicted isoforms of full-length Nb12974 cDNA (predFLNb12974_1 and predFLNb12974_2) using BLASTN alignment. A green box shows two different codons encoding two different amino acids as shown in Figure 11.

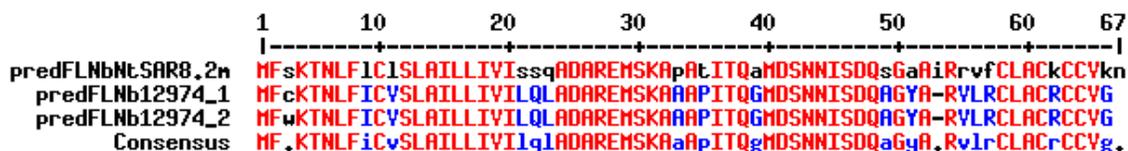


Figure 13 Amino acid sequence comparison between the predicted full-length cDNA of *N. benthamiana* SAR homologs using BLASTP alignment.

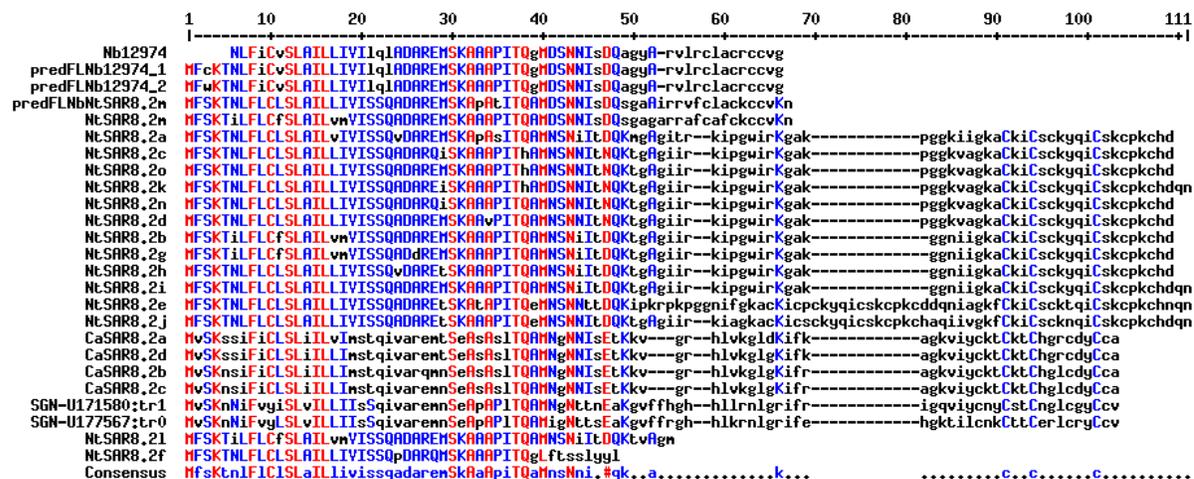


Figure 14 Multiple alignments of the predicted full-length cDNA sequences of *N. benthamiana* SAR8.2 homologs and all publicly available gene members in the SAR8.2 family. Details of these genes are described in Table 2.

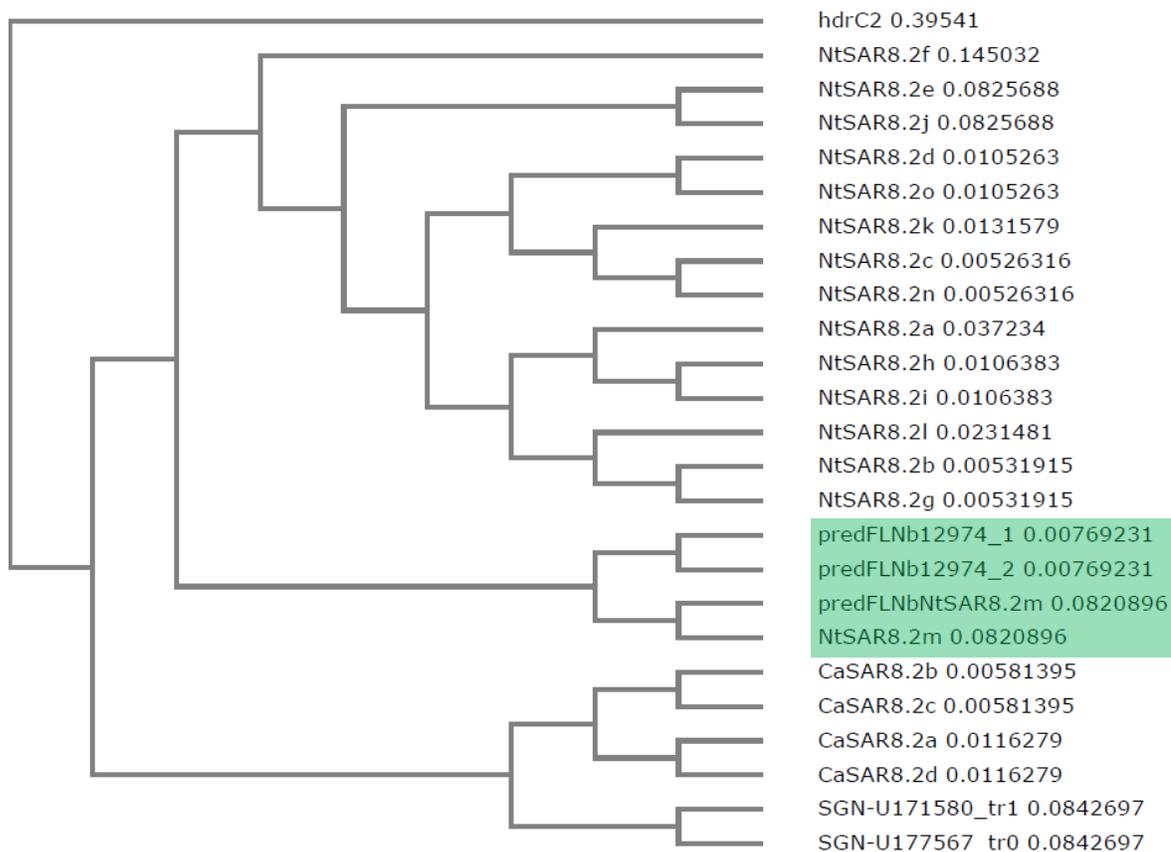


Figure 15 Phylogenetic analysis of all publicly available gene members in the SAR8.2 family. The phylogenetic tree is constructed by utilizing ClustalW2-phylogeny software with UPGMA clustering method and the *hdrC2* gene as a control. *HdrC2* is a heterodisulfide reductase subunit *hdrC2* from a thermophilic methanogenic archaea, *Methanocaldococcus jannaschii*. The genes used in this phylogenetic analysis are described in Table 2. The green box denotes the relationship between the tobacco NtSAR8.2m gene and the predicted *N. benthamiana* SAR genes (predFLNb12974_1, predFLNb12974_2, and predFLNbNtSAR8.2m).

Table 2 Publicly available gene members in the SAR8.2 family. This sequence information is used to construct multiple alignment analysis in Figure 9 and 14 and the SAR8.2 phylogenetic tree in Figure 15.

Plant species and cultivar	Gene name	GenBank accession number		Coding sequence (CDS)	Ref.
		DNA	Protein		
<i>N. tabacum</i> Xanthi nc	NtSAR8.2a	M97194 (584 bp)	AAA34117 (95 aa)	Complete	[5]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2b	M97359 (502 bp)	AAA34118 (94 aa)	Complete	[5]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2c	M97360 (560 bp)	AAA34119 (95 aa)	Complete	[5]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2d	M97361 (529 bp)	AAA34120 (95 aa)	Complete	[5]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2e	M97362 (607 bp)	AAA34121 (111 aa)	Complete	[5]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2f	U64807 (1865 bp)	NA	Complete	[24]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2g	U64810 (1864 bp)	AAF23850 (94 aa)	Complete	[24]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2h	U64808 (1881 bp)	AAF23848 (94 aa)	Complete	[24]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2i	U64811 (1940 bp)	AAF23851 (96 aa)	Complete	[24]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2j	U64812 (1394 bp)	AAF23852 (109 aa)	Complete	[24]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2k	U64814 (2006 bp)	AAF23854 (97 aa)	Complete	[24]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2l	U96152 (488 bp)	AAB53799 (54 aa)	Complete	[8, 25]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2m	U89604 (201 bp)	AAB49767 (67 aa)	Complete	[8, 25]
<i>N. tabacum</i> Samsun NN	NtSAR8.2n	AB040145 (483 bp)	BAB13706 (95 aa)	Complete	[11]
<i>N. tabacum</i> Xanthi nc	NtSAR8.2o	U64815 (2013 bp)	AAF23855 (95 aa)	Complete	[24]
<i>C. annuum</i> Hanbyul	CaSAR8.2a	AF112868 (580 bp)	AAF18935 (86 aa)	Complete	[10]
<i>C. annuum</i> Hanbyul	CaSAR8.2b	AF313765 (626 bp)	AAL16782 (86 aa)	Complete	[10]
<i>C. annuum</i> Hanbyul	CaSAR8.2c	AF313766 (450 bp)	AAL16783 (86 aa)	Complete	[10]
<i>C. annuum</i> Bugang	CaSAR8.2d	AF327570 (591 bp)	AAL56986 (86 aa)	Complete	[26]
Lycopersicon Combined 2 <i>S. lycopersicum</i> <i>S. habrochaites</i> <i>S. pennellii</i>	UNIGENES match to AAL16783	SGN-U171580			[27]
<i>S. tuberosum</i> build 2	UNIGENES match to AAL16783	SGN-U177567			[27]
<i>Methanocaldococcus jannaschii</i>	hdrC2	U67530,old L77117, new	AAB98869	complete	[28]

Table 3 Oligonucleotide information used in Chapter 4.

Name	Sequence information
Nb12974F	AACTAGCTGATGCAAGGGAGA
Nb12974R	AACATCTGCAAGCCAAACAA
GS12974F	AATTTCTAGAGCTCGTAATACTCCAAGCTGATGC
GS12974R	AATTTCTAGAGCTCACATCTGCAAGCCAAACAAC
RNA1-POLF	TCACAAGGGTCAAATTCTCAAATCCT
RNA1-POLR	TGCTGCTTTTTTGGTATAACTTCCTCTT
RNA1-CPF	ATGTCTTCAAAGCTCCCAA
RNA1-CPR	CGCTCATGACTAACTGGGTA
RNA2F	CGCGTCTGATTGAGTTGGAAGTA
RNA2R	CTGCCTTGATGCTCGACAGTA
TRV2F	TTACTCAAGGAAGCACGATGAGC
TRV2R	GAACCGTAGTTTAATGTCTTCGGG
ActF	GTGACCTCACTGATAGTTTGA
ActR	TACAGAAGAGCTGGTCTTTG

Materials and Methods

1) Plants, RCNMV inoculum preparation and inoculation procedure

N. benthamiana was a plant model used in this experiment. Details of plant maintenance, RCNMV inoculum preparation and inoculation procedure were described in Chapter 5 materials and methods. Briefly, RCNMV inoculum was prepared in a 110 μ l volume per 1 plant. This 110 μ l inoculum was a mixture of 1 μ l *in vitro* T7 RNA-1 transcript (or 1 μ l *in vitro* T7 R1SG1 transcript), 1 μ l *in vitro* T7 RNA-2 transcript and 108 μ l inoculation buffer (10 mM sodium diphosphate, pH 7.2). Four leaves per 1 plant were inoculated with either RCNMV inoculum or buffer. 27 μ l of the RCNMV transcript mixture (or inoculation buffer) was pipetted onto each leaf and mechanically rubbed with carborundum (abrasive). Inoculated plants were maintained in a temperature and light controlled environment at 18-22°C, 16 hour-light and 8 hour-dark period at the NCSU greenhouses (Method Road).

2) Plasmid DNA constructs

The silencing construct TRV2-NbSAR utilizing the TRV VIGS system [29] was designed to down regulate NbSAR expression in *N. benthamiana*. A construction of the silencing construct is described in Chapter 5 materials and methods. Briefly, a 132 bp fragment of the NbSAR gene was amplified accordingly to OneTaq DNA polymerase protocol (NEB™), with a primer set of GS12974F and GS12974R (Table 3) to generate the GS-NbSAR. The PCR conditions for amplifying GS-NbSAR were as follows: 1) a denaturation cycle of 94°C for 30 sec, 2) 30 amplification cycles of 94°C for 30 sec, 55°C for

30 sec and 68°C for 1 min, and 3) a final extension cycle of 68°C for 5 min. A ligation between GS-NbSAR and TRV2 was performed via *SacI* restriction site. A colony PCR technique using the primer set TRV2F (forward primer) and TRV2R (reverse primer) (Table 3) was used to identify TRV2 clones that contained the GS-NbSAR insert. A positive plasmid was also confirmed by digesting with *SacI*. The resultant construct TRV2-NbSAR was used for down regulating NbSAR expression in *N. benthamiana* for a gene silencing study.

3) *Agrobacterium* preparation and infiltration

Agrobacterium was prepared accordingly to a detail in Chapter 5 materials and methods. Briefly, the silencing constructs TRV1, TRV2-NbSAR were transformed into *Agrobacterium tumefaciens* strain C58C1 using a BioRad electroporator. Transformed cells were plated onto LB agar plates containing rifampicin, gentamicin and kanamycin and then incubated at 28°C for 48 hours. Individual colonies were inoculated into 2 ml LB broth cultures with the appropriate antibiotics and incubated at 28°C for 20 hours with shaking at 270 rpm. From these initial cultures 250 µl was used to inoculate 5 ml LB broth cultures with the appropriate antibiotics and 40µM acetosyringone/ 10 mM MES, pH 5.6. These cultures were similarly incubated at 28°C for 20 hours with shaking at 270 rpm. Cultures were subsequently pelleted, resuspended in 10 mM MgCl₂/ 10 mM MES, pH 5.6/ 200 µM acetosyringone to the OD₆₀₀ reading ~1 and then incubated at room temperature for at least 3 hr prior to syringe infiltration into plants.

The *Agrobacterium* cultures of TRV1 and TRV2-NbSAR (or TRV2 for the control) were mixed at a 1:1 ratio immediately prior to infiltration into plants. Agroinfiltration was performed on 2-3 week old (Chapter 5 Panel B in Figure 18) *N. benthamiana* plants with two leaves per plant being infiltrated by using a 1 ml needleless syringe on the abxial side of the leaf. Plants were maintained in a temperature and light controlled environment at 26°C, 16 hour-light and 8 hour-dark period at NCSU greenhouses (Method Road).

4) Total RNA extraction and cleanup

Four plants and two leaves per 1 plant were used to represent the test condition (silencing of NbSAR) as well as the control (see above). Total RNA extracts isolated from the same plant were pooled to represent one biological replicate (4 biological replicates for the test condition and 4 biological replicates for controls). Leaf samples were collected 10 days after agroinfiltration as well as various time points after RCNMV transcript inoculation.

A total RNA extraction and cleanup were performed accordingly to details in Chapter 5 materials and methods. Briefly, total RNA was extracted from 100 mg *N. benthamiana* leaf tissue samples according to the TRIzol protocol (InvitrogenTM). And total RNA extracts destined for qRT-PCR analysis were treated with Turbo DNA-free (AmbionTM) according to the manufacturer's protocol to remove any residual input DNA. The samples were stored at -20°C.

5) RNA quantification by real-time PCR (qRT-PCR)

First strand cDNA synthesis and quantitative real-time PCR (qRT-PCR) preparation were performed according to details in Chapter 5 material and methods. Briefly, First strand cDNA was synthesized by using a DNA-free total RNA (equivalent to ~1 µg total RNA) as a template. SYBR Green-based quantitative real-time PCR (qRT-PCR) method was used in this study. The qRT-PCR reaction was prepared accordingly to FastStart Universal SYBR Green Master ROX protocol (Roche™). The first strand cDNA equivalent to 40 ng total RNA was used to prepare qRT-PCR reaction. The qRT-PCR reactions were setup in a 384 well plate and placed in ABI7900 HT Fast real-time PCR system (Applied Biosystems™). The Applied Biosystems SDS software version 2.4 was used to monitor and dissect real-time PCR data. Three qRT-PCR reactions were prepared per 1 biological replicate. All PCR products are less than 200 base pairs. The cycling condition was set as follows: 1 cycle at 95°C for 10 min, and 40 cycles at 95°C for 15 sec /55°C for 1 min. An association analysis was also performed in order to test if there were any non-specific PCR products formed. The condition for the association analysis was set as follows: 1 cycle at 95°C for 15 sec, 60°C for 15 sec and 95°C for 15 sec.

NbSAR transcript levels were assayed using primers Nb12974F and Nb12974R (Table 3). RCNMV RNA levels were assayed with 2 primer pairs for RNA-1 and 1 primer pair for RNA-2: 1) RNA1-POLF and RNA1-POLR_for probing RNA-1 at the polymerase (RNA-1 POL), 2) RNA1-CPF and RNA1-CPR_for probing RNA-1 at the coat protein (RNA-1 CP) and 3) RNA2F and RNA2R_for probing RNA-2 at the movement protein (RNA-2 MP).

6) qRT-PCR data analysis

A relative quantification, $2^{-\Delta\Delta C_t}$ method (equation shown in Chapter 5 materials and methods) [30] was used to analyze the real-time PCR data. Expression of the target gene was normalized against the expression of a reference gene. Actin was used as the reference genes to normalize the target gene that was performed within the same plate. Student's t-test statistical analysis was used to examine the real-time PCR data.

7) RCNMV sGFP quantification

N. benthamiana plants were inoculated with a combination of *in vitro* T7 R1SG1 and RNA-2 transcripts. Ten plants and two leaves per plant were used for each test condition and each control. Leaf samples were harvested 3 days post inoculation. The GFP fluorescence intensity was analyzed by the IVIS imaging system and subjected to Student's t-test statistical analysis. The IVIS Lumina System (Xenogen Corporation, Alameda, CA) is capable of quantifying photon emission from a variety of sources. A CCD camera measured and recorded photon emission data which was then incorporated into Living Image Software (Xenogen Corp.) for further analysis. Whole leaves were placed under the CCD camera and measurements were taken with a GFP excitation filter, and exposure time of 1 second.

References

1. Dasgupta R, Garcia BH, Goodman RM: **Systemic spread of an RNA insect virus in plants expressing plant viral movement protein genes.** *Proceedings of the National Academy of Sciences* 2001, **98**(9):4910-4915.
2. Goodwin MM, Zaitlin D, Naidu RA, Lommel SA: ***Nicotiana benthamiana*: its history and future as a model for plant-pathogen interactions.** *Molecular plant-microbe interactions* 2008, **21**(8):1015-1026.
3. Ross AF: **Systemic acquired resistance induced by localized virus infections in plants.** *Virology* 1961, **14**(3):340-358.
4. Kiefer IW, Slusarenko AJ: **The pattern of systemic acquired resistance induction within the *Arabidopsis* rosette in relation to the pattern of translocation.** *Plant Physiology* 2003, **132**(2):840-847.
5. Ward ER, Uknes SJ, Williams SC, Dincher SS, Wiederhold DL, Alexander DC, Ahl-Goy P, Metraux J-P, Ryals JA: **Coordinate gene activity in response to agents that induce systemic acquired resistance.** *The Plant Cell Online* 1991, **3**(10):1085-1094.
6. Friedrich L, Lawton K, Ruess W, Masner P, Specker N, Rella MG, Meier B, Dincher S, Staub T, Uknes S: **A benzothiadiazole derivative induces systemic acquired resistance in tobacco.** *The Plant Journal* 1996, **10**(1):61-70.
7. Herbers K, Meuwly P, Metraux J-P, Sonnewald U: **Salicylic acid-independent induction of pathogenesis-related protein transcripts by sugars is dependent on leaf developmental stage.** *FEBS letters* 1996, **397**(2):239-244.
8. Guo A, Salih G, Klessig DF: **Activation of a diverse set of genes during the tobacco resistance response to TMV is independent of salicylic acid; induction of a subset is also ethylene independent.** *The Plant Journal* 2000, **21**(5):409-418.
9. Lee G-J, Shin R, Park C-J, Yoo TH, Paek K-H: **Induction of a pepper cDNA encoding SAR8. 2 protein during the resistance response to tobacco mosaic virus.** *Molecules & Cells* 2001, **12**(2).
10. Lee S, Hwang B: **Identification of the pepper SAR8. 2 gene as a molecular marker for pathogen infection, abiotic elicitors and environmental stresses in *Capsicum annuum*.** *Planta* 2003, **216**(3):387-396.
11. Takemoto D, Noriyuki D, Kawakita K: **Characterization of elicitor-inducible tobacco genes isolated by differential hybridization.** *Journal of General Plant Pathology* 2001, **67**(2):89-96.

12. Shan XC, Goodwin PH: **Identification of a SAR8. 2 gene in the susceptible host response of *Nicotiana benthamiana* to *Colletotrichum orbiculare*.** *Functional plant biology* 2005, **32**(3):259-266.
13. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
14. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2008, **36**(suppl 1):D25-D30.
15. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**(1):25-29.
16. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, Duvaud Sv, Flegel V, Fortier A, Gasteiger E: **ExPASy: SIB bioinformatics resource portal.** *Nucleic acids research*, **40**(W1):W597-W603.
17. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExPASy: the proteomics server for in-depth protein knowledge and analysis.** *Nucleic acids research* 2003, **31**(13):3784-3788.
18. Heijne G: **Patterns of amino acids near signal-sequence cleavage sites.** *European journal of biochemistry* 1983, **133**(1):17-21.
19. Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y: **The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond.** *Plant Physiology* 2005, **138**(3):1310-1317.
20. Reddy AS: **Alternative splicing of pre-messenger RNAs in plants in the genomic era.** *Annu Rev Plant Biol* 2007, **58**:267-294.
21. Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB: **A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research.** *Molecular plant-microbe interactions* 2012, **25**(12):1523-1530.
22. Bachan S, Dinesh-Kumar SP: **Tobacco rattle virus (TRV)-based virus-induced gene silencing.** In: *Antiviral Resistance in Plants*. Springer: 83-92.
23. Powers JG, Sit TL, Heinsohn C, George CG, Kim K-H, Lommel SA: **The Red clover necrotic mosaic virus RNA-2 encoded movement protein is a second suppressor of RNA silencing.** *Virology* 2008, **381**(2):277-286.
24. Moraes MG, Goodman RM: **Structure and expression of tobacco Sar8.2 gene family.** *GenBank NCBI* 1996.

25. Guo A, Klessig DF: **Activation of a diverse set of genes during the tobacco resistance response to TMV is independent of salicylic acid and distinct from the induction of PR-1 genes.** *GenBank NCBI* 1997.
26. Lee G-J, Shin R, Park C-J, Paek K-H: ***Capsicum annuum* putative SAR8.2 mRNA, induced by Tobacco mosaic virus.** *GenBank NCBI* 2000.
27. Bombarely A, Menda N, Teclé IY, Buels RM, Strickler S, Fischer-York T, Pujar A, Leto J, Gosselin J, Mueller LA: **The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl.** *Nucleic acids research*, **39**(suppl 1):D1149-D1155.
28. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD: **Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*.** *Science* 1996, **273**(5278):1058-1073.
29. Dinesh-Kumar SP, Anandalakshmi R, Marathe R, Schiff M, Liu Y: **Virus-induced gene silencing.** In: *Plant Functional Genomics*. Springer; 2003: 287-293.
30. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic acids research* 2001, **29**(9):e45-e45.

Chapter 5

Modulation of *N. benthamiana* WRKY gene expression early in a *Red clover necrotic mosaic virus* infection implies its involvement in plant immunity

Chapter summary

In this chapter, I report a study on the regulation of *N. benthamiana* WRKY gene expression early in a *Red clover necrotic mosaic virus* (RCNMV) infection. The microarray study described in Chapter 3 reveals that three *N. benthamiana* WRKY genes are significantly modulated in response to RCNMV infection (FDR cutoff of 0.01). Among these three WRKY genes, the unigene Nb04875 was selected for a functional characterization. The study in this chapter is the first and only report that thoroughly analyzes the characteristics of Nb04875. The cumulative computational evidence from the sequence, protein structure and subcellular localization analyses led to the conclusion that Nb04875 is a novel *N. benthamiana* WRKY gene. The GFP tagging Nb04875 study also confirms its nuclear localization. Nb04875 was also further functionally assayed for its effects on RCNMV infection by both transient gene silencing and transient gene overexpression studies in *N. benthamiana*. These functional assays suggest that Nb04875 plays a critical role in the regulation of early host defenses towards an RCNMV infection.

Abstract

Transcriptional re-programming is a key step of the plant defense response to pathogen recognition. Transcriptional profiling of *N. benthamiana* early in the RCNMV infection process reveals that three host WRKY genes were significantly modulated at prior to 24 hours post inoculation (hpi) at an FDR cutoff of 0.01. Among these three WRKY genes, the unigene Nb04875 was further functionally characterized due to its 10 fold up-regulation at 24 hpi. Its full-length cDNA of 975 bp encoded a putative polypeptide consisting of 325 amino acids. The putative Nb04875 translation product contained a single WRKY DNA binding domain which included the conserved sequence “WRKYGQK” at its N-terminus, and an atypical zinc finger motif at its C-terminus. It has the highest similarity to StWRKY6, a member of the WRKY transcription factor (TF) family in potato. A 3D protein structure prediction indicated that Nb04875 consists of four β -stranded sheets. It also contains a potential nuclear localization signal (NLS) at its N-terminus. These features lead to the conclusion that Nb04875 is a novel gene belonging to the WRKY TF family. Thus, its full-length protein is assigned the name NbWRKY. A C-terminally tagged NbWRKY with GFP also confirmed its nuclear localization. NbWRKY (Nb04875) was further functionally assayed for its effects on RCNMV infection by both transient gene silencing and gene overexpression studies in *N. benthamiana*. These functional assays suggest that NbWRKY possesses dual functionality in regulating the defense response: shifting its function between both a positive and negative regulator of disease resistance depending on the duration of the RCNMV infection. A bioinformatics prediction indicates that the promoter region of

NbWRKY contains at least three putative W-box elements, suggesting that the regulation of NbWRKY transcription is under the control of either auto- or cross- regulation, and may be responsive early in the RCNMV infection process.

Keywords; WRKY, *Nicotiana benthamiana*, defense, RCNMV

Introduction

Plants are constantly exposed to a plethora of pathogens in their natural habitat. Adapting to such changes in order to survive and reproduce requires a great degree of phenotypic and genotypic plasticity that is mainly determined by the plant's genome or a plant's genetic makeup. How plants are able to integrate the multitude of different defense strategies that enable them to respond rapidly, effectively and properly under any given phytopathogen-induced stress condition remains a fundamental question in plant-pathogen interaction studies. What has become apparent, however, is that plants are capable of extensive reprogramming of their transcriptome in a highly dynamic and temporal manner, which is mainly achieved through an exclusive enforcement of a network of various transcription factors (TFs). This responsive regulation leads to an adaptive plasticity of plants during exposure to highly variable environmental stresses. WRKY TFs are a large family of regulatory proteins forming and executing such a network [1]. They are involved in various plant processes but most notably in coping with stresses induced by a diverse array of pathogen attack [1, 2]. A large body of circumstantial evidence has accumulated during the past twenty years implicating WRKY factors in transcriptional reprogramming during plant

immune responses [3-5]. The majority of the 74 *Arabidopsis* WRKY genes are transcriptionally inducible upon pathogen infection and other defense-related stimuli [6-8]. In this study, I will restrict my attention to investigating the immunity role of WRKY TFs in plant-virus infections.

Plant adaptability early in the virus infection period is a critical determinant in deciding the fate of plant survival and reproduction. Such rapid adaptability requires an effective strategy to re-program plant physiological processes to protect the plant from a viral attack and prevent subsequent further viral invasion. What strategy the plant employs to rapidly manipulate such processes to launch an immediate immune system against a virus remains unknown. To investigate this strategy during a plant-virus interaction, I utilized *Nicotiana benthamiana* (a model host for most plant viruses) and *Red clover necrotic mosaic virus* (RCNMV, a typical plus strand RNA plant virus). A custom microarray representing a total of 13,415 unigenes (equivalent to an estimated coverage of ~38% of the *N. benthamiana* transcriptome) was developed to simultaneously monitor the modulation of expression of thousands of host genes early in an RCNMV infection. The construction of the *N. benthamiana* microarray was previously described in Chapter 2. Host transcription profiles were analyzed at 2, 6, 12 and 24 hours post RCNMV inoculation (hpi). The microarray analysis described in Chapter 3 reveals that three WRKY genes were significantly modulated early in an RCNMV infection at an FDR cutoff of 0.01. Among these three genes, the unigene Nb04875 was selected for a functional characterization in this chapter due to the rapid accumulation of its transcripts and the high degree of modulation after 12 hours. The central hypothesis in this chapter is that the alteration of host WRKY gene expression is

driven by the plant host in order to rapidly engage the regulation of host defense gene expression to induce a disease resistance toward an early RCNMV infection.

What is WRKY TF?

The defining feature of WRKY TFs are their unique DNA binding domain, known as the WRKY DNA binding domain after the almost invariant WRKY amino acid sequence at the N-terminus [9]. The WRKY DNA binding domain appears to exist exclusively in plants, in one or two copies in a superfamily of plant TFs [4]. WRKY TFs are involved in the regulation of various physiological processes that are unique to plants, including pathogen defense [1, 10], response to abiotic stress (wounding, drought, cold adaptation and heat-induced chilling tolerance) [11-14], hormone signaling [15], secondary metabolism [16], senescence [13, 17, 18] and trichome development [19].

The WRKY DNA binding domain is 60 amino acids in length that is defined by the conserved invariant amino acid sequence “**WRKYGQK**” at its N-terminus, together with a novel zinc-finger-like motif (zinc binding motif) at the C-terminus [20]. This atypical zinc-finger motif occurs in two different amino acid sequence motifs: **CX_{4,5}CX₂₂₋₂₃HXH** and **CX₇CX₂₃HXC**. In 2005, Yamasaki *et al.* reported the first solution structure of a WRKY domain [21]. They found that the WRKY domain consisted of four or five β -strands, with the zinc coordinating Cys/His residues forming a zinc-binding pocket. Of these four β -strands, the N-terminal strand contains the WRKY sequence that binds DNA and the other three strands form a novel zinc finger structure. They also demonstrated that the **WRKYGQK**

residues corresponding to the N-terminal β -strand: (i) enables extensive hydrophobic interactions, contributing to the structural stability of the β -sheet, and (ii) partly protrudes from one surface of the protein, thereby enabling access to the major DNA groove and contacts with the DNA. It was proposed that the five-stranded β -sheet containing the WRKYGQK motif makes contact with an approximately 6-bp region of the DNA, which is consistent with the length of the W box, the binding sites of most known WRKY proteins.

In 2012, the same research group [22] reported the NMR solution structure of a *Arabidopsis* WRKY4 domain complexed with the W box binding site. This was the first structure of the WRKY-DNA complex. They concluded that the apolar contacts by residues in the conserved WRKYGQK motif with thymine methyl groups were important in recognition of the W-box sequence. They discovered that a four-stranded β -sheet of the WRKY domain enters the major groove of the DNA in a nearly perpendicular manner to the DNA helical axis. This atypical mode of interaction between the WRKY DNA binding domain and its DNA target was assigned the term B-wedge.

What is the target of WRKY TF?

The WRKY proteins control transcription of the target genes by binding to the promoter regions that contain the specific DNA element **(T)(T)TGAC(C/T)**, termed the W-box [3, 23, 24]. The invariant “**TGAC**” core is essential for function and WRKY binding. It has long been clear that the conservation of the WRKY domain in WRKY proteins is mirrored by a remarkable conservation of the W-box. Gel shift experiments, random binding

site selection, yeast one-hybrid screens and co-transfection assays performed on many different WRKY proteins confirmed that the W-box is the minimal consensus element required for specific DNA binding. The W-box mediates transcriptional response to pathogen-derived elicitors [3, 25] and is present in the promoters of many plant genes that are associated with defense [26].

Result and discussion

1) RCNMV up-regulated *N. benthamiana* WRKY TF expression 12 hours after the initiation of infection

Our custom *N. benthamiana* microarray includes 24 unigenes that have a BLAST hit to WRKY TF but only 19 of these contain a conserved WRKY-DNA binding domain (Table 1). The functionalities of these unigenes were therefore computationally assigned as WRKY TFs. The microarray experiment described in Chapter 3 showed that three *N. benthamiana* WRKY-like unigenes were significantly modulated early during an RCNMV infection at sub 24 hours (FDR cutoff of 0.01) (Figure 1). Of all these WRKY-like genes, the expression of the unigene Nb04875 had the highest modulation magnitude during RCNMV infection. The qRT-PCR analysis also confirmed that Nb04875 was significantly up-regulated greater than 10 fold at 24 hours post RCNMV inoculation compared to the mock inoculated control plants (p-value = 0.00). The up-regulation of this WRKY-like gene was likely driven by the plant host in order to fight against the RCNMV invasion. Nb04875 contains a highly conserved WRKY domain. WRKY proteins are thought to play important roles in plant defense

responses to pathogens [1]. The hypothesis that will be tested here is that the plant host increases Nb04875 expression to engage in transcriptional reprogramming of other defense genes, possibly inducing host defense gene expression. A delayed modulation of this WRKY-like unigene may reflect host recognition of only late events in the RCNMV infection cycle of the primary infected cells. The stimulant could be the late encoded viral proteins, either movement or capsid protein.

The unigene Nb04875 has not been documented elsewhere in the literature or in any public genomic databases. The evidence in this report points toward the conclusion that Nb04875 is a novel gene belonging to the WRKY TF family in *N. benthamiana*. Given its unknown functions, this unigene required a substantial amount of analysis for authoritative identification. In the next section, I describe a preliminary analysis of Nb04875, including sequence analysis, WRKY DNA binding domain and zinc finger motif identification, computational protein structural study, full-length cDNA sequence prediction, protein subcellular localization and a promoter analysis. Also, the unigene Nb04875 is further functionally assayed by both gene silencing and gene overexpression studies in order to determine its impact on RCNMV replication.

2) A functional annotation of Nb04875 as a WRKY TF using a bioinformatic/computational approach

2.1) Nb04875 is a WRKY homolog

The unigene Nb04875 sequence was searched against the GenBank database [27] using the BLASTX method [28]. The homologous search revealed that the top hits matched to Nb04875 were WRKY TFs from various plant species such as potato, tobacco, cocoa, cucumber, Barbados nut, castor oil plant, mulberry, grapevine and cotton (Table 2). The multiple alignments comparing Nb04875 to the experimentally verified WRKY TFs support the assumption that Nb04875 contains a relatively strong match to a conserved WRKY domain (Figure 2). The conserved invariant amino acid sequence “**WRKYGQK**” was found at the N-terminus of Nb04875 while a novel zinc finger motif with the amino acid sequence “**CX₇CX₂₃HXC**” was identified at its C-terminus (Figure 2).

The unigene Nb04875 nucleotide and amino acid sequence has the highest similarity to the StWRKY6 gene from potato (*Solanum tuberosum*; 66 % amino acid identity and E-value = 4×10^{-115} ; Table 2 and Figure 5). Also, the phylogenetic tree analysis demonstrated in Figure 3 indicates that Nb04875 is most closely related to StWRKY6.

A BLAST search against the plant TF database v3.0 (PlantTFDB) [29] also revealed that Nb04875 possesses the highest similarity to WRKY proteins from *Solanaceae* plants, with the top 4 hits to *N. tabacum* (tobacco), *S. tuberosum* (potato), *S. lycopersicum* (tomato) and *Capsicum annuum* (pepper). Nb04875 is also similar to *Arabidopsis thaliana* WRKY70 (AT3G56400.1) with an E-value = 4×10^{-40} . This accumulation of evidence leads one to

speculate that Nb04875 functions as a WRKY TF. However, the homology search and multiple alignments of nucleotide and amino acid sequence approaches only provides the sequence similarity information between compared sequences. These bioinformatics approaches do not take into consideration similarities in protein structure.

To annotate an unknown protein as a WRKY TF by using BLAST and multiple alignment approaches may not be appropriate because only particular regions are conserved among WRKY homologs and so whatever is matched to these conserved regions may lead to the false positive assumption that the query sequence is a WRKY TF. In other words, the query sequence containing the conserved WRKY domain and zinc finger is not necessarily a WRKY TF. Why? The functionality of TFs strictly relies on the proper protein conformation. Therefore, if the protein structure of the query sequence containing the conserved WRKY domain and zinc finger is unable to fold in the proper conformation, it is likely that it could not function as a WRKY TF.

Protein structure is an important feature that defines WRKY functionality. In order to be able to perform their biological/molecular function, WRKY proteins are required to fold into a more specific spatial conformation, driven by a number of interactions between amino acid residues. The proper conformation of WRKY TFs is defined by two characteristics of WRKY proteins: (i) their active or inactive states, that is whether or not their DNA binding domain is exposed or accessible to bind/interact with their target DNAs and (ii) their subcellular localization, that all TFs including WRKY TFs are destined for the cellular nucleus.

The nuclear localization signal (NLS) [30, 31] is an important element of the WRKY protein structure that allows the protein to be imported from the cytosol into the nucleus where its molecular functions occur – binding to/interacting with its target DNAs and regulating the transcription of its targets, either activation or repression of transcription. NLS is characterized as a signal patch which is made up of amino acid residues that are distant to one another in the primary sequence, but come into close proximity in the tertiary structure of the folded protein.

Although the non-conserved regions of the WRKY TF do not contribute to the functionality of the WRKY TF, these regions are an essential component that enables proper folding of the WRKY TF. The multiple alignments between Nb04875 and WRKY homologs indicate that the region that is outside of the conserved DNA binding domain and zinc finger are highly variable. Even with its best hit to StWRKY, only 66% amino acid identity is found to Nb04875. Therefore, in order to confirm the functional annotation of the unigene Nb04875 as a WRKY TF, I also conducted Nb04875 protein structural analysis to determine its: (i) three dimensional structure, and (ii) nuclear localization signaling capability.

2.2) Putative Nb04875 protein structure is composed of a four β -stranded sheet

The unigene Nb04875 protein structure was predicted by using SWISS-MODEL server [32]. The SWISS-MODEL is a web interface that applies homology modeling methods to build a three dimensional (3D) model for a protein of unknown structure from the known experimental structures. Homolog modeling relies on evolutionarily related structures

(templates), utilizing evolutionary information between the potentially related proteins, to generate a protein tertiary and quaternary structural model for the protein of interest (target).

The Nb04875 amino acid sequence (target) was searched against the SWISS-MODEL template library (STML). However, the unigene Nb04875 is originally derived from an expressed sequence tag (EST) so it is a nucleotide sequence. Therefore, the unigene Nb04875 was first putatively translated to an amino acid sequence using ExPASy software [33] with a standard genetic code prior to its structural analysis. Since the unigene Nb04875 sequence is not a complete coding sequence (CDS), it is unlikely that the positions of the start and stop codons are known. Nb04875 sequence structure is discussed later in section 2.4. ExPASy software translated the Nb04875 nucleotide sequence into an amino acid sequence using six frame translations from both 5'-3' direction and 3'-5' direction. The putative Nb04875 amino acid sequence that was used for structural analysis was selected from the translation frame that included the conserved WRKY domain and zinc finger motif.

The putative Nb04875 amino acid sequence is best hit to the protein structure of *A. thaliana* WRKY1 (2ayd.1.A) with 42.2 % sequence identity. Therefore, the *A. thaliana* WRKY1 protein structure was used as a template to build a predicted 3D model for Nb04875 (Figures 9 and 10). The protein structure of *A. thaliana* WRKY1 (2ayd.1.A) was previously well-characterized using an X-ray diffraction method with 1.60 angstroms resolution [34]. The 3D protein structure of *A. thaliana* WRKY1 in Figure 10 shows that: (i) its tertiary structure contains five β -stranded sheets with a zinc binding pocket, (ii) the conserved sequence “WRKYGQK” is located on β_1 , (iii) a Zn^{2+} ion interacts with the zinc finger motif

at its amino acid positions C40, C45, H69, and H71 (C₂H₂ is a typical zinc finger motif), and (iv) the DNA-binding residues are located on β_2 and β_3 .

With the use of the *A. thaliana* WRKY1 template, the predicted 3D model revealed that the Nb04875 amino acid sequence is able to fold into a tertiary structure containing a four-stranded β sheet (Figures 9 and 10). The model-template alignment shows that a conserved DNA binding domain of the “WRKYGQK” sequence from both models (Nb04875 or NbWRKY) and template (AtWRKY1) is located on the first β strand, and the zinc finger motif is also located at the same position in the alignment (Figure 9) as well as in the 3D model (Figure 10).

The model quality is assessed with a QMEAN Z-score which relates the predicted model to the reference structure solved by X-ray crystallography. The QMEAN Z-score is an estimate of the “degree of nativeness” of the structural features (geometrical features: pairwise atomic distances, torsion angles, solvent accessibility) observed in an entire predicted model by describing the likelihood that a predicted model is of comparable quality to the high-resolution experimental structures. The GMQE (Global Model Quality Estimation) is another estimation of model quality which combines properties from the target-template alignment. The resulting GMQE score is expressed as a number between zero and one, reflecting the expected accuracy of a model built with that alignment and template. Higher numbers indicate higher reliability.

The predicted 3D model of Nb04875 has a QMEAN Z-score of -2.91 and a GMQE estimate of 0.12, indicating a relatively low quality and low reliability model. This is not a

surprising result due to the fact that the entire model and template possesses only 42% sequence identity. Despite the low quality score observed for the entire predicted model, the local quality estimation suggests otherwise as shown in Figures 9 and 10. Predicted local similarity between template and target is estimated as a number between zero and one; the higher number indicates the higher quality of the predicted structure. Figure 9 shows that certain regions of Nb04875 have a higher estimation of similarity to the template, suggesting a high quality predicted protein structure of Nb04875. In Figure 9, the alignment between model and template also reveals that the residues of Nb04875 that have a high quality score (blue shaded) are particularly located in the predicted β strand structure, confirming that these residues potentially drive the interactions to enable protein folding into β strand structure. The predicted 3D model of Nb04875 in Figure 10 shows that the majority of residues contributing to β strand structure have a high quality score estimation (blue shade). The predicted 3D model of Nb04875 confirms its potential ability to fold into the four β stranded-sheet; therefore, this piece of information is strong evidence for the assumption that the Nb04875 functions as a WRKY TF.

2.3) Putative Nb04875 protein contains the N-terminal signal peptide that enables its translocation to the nucleus

Subcellular localization provides information on the location of the mature protein in the cell. The functions of a protein are essentially dependent on its subcellular localization. Therefore, the ability to predict subcellular localization directly from the protein sequence is useful to inferring the potential role of a protein of unknown function.

The subcellular localization of the putative Nb04875 product was predicted using a variety of software applications to avoid bias. Their differences lie in the methods (algorithms) and databases that are used to compute the prediction. Individual prediction tools implement their own algorithms and rules, and use different benchmark data (standard data that is used to train algorithms) to determine their abilities and reliabilities at predicting protein subcellular localization. Some algorithms are better than others at predicting subcellular localization to particular localization sites (subcellular sites/compartments). These prediction tools also have their own databases (or libraries) that are used to predict the localization of the query protein. The databases from the individual prediction tools are also different in three important features: (i) organism types (ranging from prokaryotes, eukaryotes, animal, plant, bacteria, yeast, to specific species), (ii) number of proteins or motifs with experimentally verified subcellular localizations, and (iii) number of subcellular localization sites (for instance, some software analyzes the query protein localization for up to 12 subcellular compartment sites).

A number of reports have noted that sequence similarity is useful in predicting subcellular localization such that one can accurately infer the subcellular compartment of a protein if one can find close homologs with an experimentally verified localization [35]. Some software applications rely strongly on sequence similarity for predicting the subcellular localization. However, the accuracy of the prediction depends on the certain similarity threshold. Studies indicate that this homology search approach performs well down to 30% sequence similarity and the performance deteriorates considerably for sequences sharing lower sequence identity [36]. Given a low sequence similarity between Nb04875 and other

WRKY homologs (% identity ranging from 37% to 66% as shown in Table 2), utilizing software that depends solely on sequence similarity to predict a subcellular localization may lead to a false positive prediction.

The subcellular localization of the putative Nb04875 product was predicted by utilizing four different software tools: 1) PSORT [37], 2) WoLF PSORT [38], 3) Plant-mPLOC [39], and 4) CELLO [36]. WoLF PSORT is an extension of the PSORT tool where WoLF PSORT incorporates UniProt and the Gene ontology database into the methodology. PSORT and WoLF PSORT both include homology search (sequence comparison against their libraries) in their methods. However, they exclude highly homologous protein sequences in predicting protein localization; they instead emphasize mainly on the non-homologous proteins that have similar localization motifs to the query protein to predict its subcellular localization. The result from WoLF PSORT provides the so-called nearest neighbors which are proteins with a similar PSORT score (localization score), but with low homology to the query protein. The Plant-mPLOC tool also applies a homology-based approach to predict protein subcellular localization. However, the advantages of this tool are: (i) its algorithm is specifically designed to handle predicting plant protein localization, and (ii) it provides the prediction for up to 12 subcellular localization sites. CELLO is the only protein localization prediction tool used in my analysis that does not rely on homology search; therefore its performance remains relatively unaffected by sequence similarity.

The top results from PSORT, WoLF PSORT, CELLO, and Plant-mPLOC software collectively affirm that the putative Nb04875 protein localizes to the nucleus. WoLF PSORT

indicates that Nb04875 has the highest similar localization feature (nearest neighbors) to the nucleus localization of proteins from *A. thaliana*: PLETHORA TF (At5g65510.1), AP2 domain TF (At3g54990.1), and WRKY TF 29 (WR29_ARATH). These *A. thaliana* proteins have between 14% and 15% amino acid sequence identity to Nb04875.

In addition to its protein subcellular localization prediction, the putative Nb04875 product was further analyzed for its nuclear importing abilities. As would be expected from TFs, WRKY proteins contain an NLS. The two most common NLSs are: (i) monopartite signals, which are short stretches enriched in basic amino acids [31], and (ii) bipartite signals, which are composed of two short basic stretches separated by a spacer [40]. The WoLF PSORT, CELLO, and Plant-mPLOC programs did not pinpoint an exact NLS motif on Nb04875. However, the PSORT tool revealed that the monopartite NLS pattern with the consensus four residue sequence K-K/R-X-K/R [31] was detected in the putative Nb04875 translation product as K-R-R-K (Figure 4). The multiple alignments in Figure 2 demonstrated that the monopartite NLS pattern of “K-R-R-K” was also identified in many WRKY gene homologs from other plant species. PSORT did not find any bipartite signals for nuclear targeting in the Nb04875 sequence.

In summary, the bioinformatic prediction produced from homology search, multiple alignments, three dimensional protein structure, and localization prediction point to the conclusion that Nb04875 is highly similar to WRKY homologs found in other plant species. The putative Nb04875 protein is capable of folding into a four β -stranded sheet protein structure. Nb04875 contains a conserved WRKY DNA binding domain, which includes the

conserved invariant amino acid sequence “WRKYGQK” at its N-terminus, together with a novel zinc-finger-like motif (zinc binding motif) at the C-terminus. Nb04875 also contains an NLS consensus sequence at its N-terminus that enables protein translocation to the nucleus. These findings lead to my belief that the putative Nb04875 product is a *N. benthamiana* WRKY TF.

To date, there have been no studies reporting the experimental characterization of the WRKY TF genes from *N. benthamiana*. Although this study does not directly investigate Nb04875 functionality as a TF, it is the first report of the computational analysis of Nb04875 as a novel *N. benthamiana* WRKY TF. To further confirm the function of the putative Nb04875 product as a WRKY TF, it is necessary to identify its biochemical characteristics, especially its ability to interact with its DNA target (DNA-protein interaction). The most common experimental approaches would be DNA foot printing or gel-shift assays.

2.4) Construction of the full-length NbWRKY cDNA

The unigene Nb04875 sequence is 940 bp in length which translates to 313 amino acids using ExPASy translation software [33] and a standard genetic code (Figure 6). The sequence of unigene Nb04875 does not contain a complete open reading frame (ORF): it appears to be a partial coding sequence (CDS) with a complete 5' cDNA end and a 12 bp 5' untranslated region (UTR), but lacking a complete 3' cDNA end. The 5'UTR is not conserved among WRKY homologs, therefore, the prediction of the start codon position in the Nb04875 sequence was assumed by computational evidence. The full-length Nb04875

cDNA was predicted from an alignment of the unigene sequence to the *N. benthamiana* genome (Sol Genomics Network [41]). The 3' cDNA terminal sequence was predicted from the first stop codon that occurs on the genome at the 3' end of the unigene sequence. This may not be the actual complete 3' cDNA end. However, the sequence prediction of the full-length cDNA provides the necessary information required to design primers to obtain the clone that has the longest sequence of the potential full-length Nb04875 cDNA. This predicted full-length Nb04875 cDNA is assigned the name predFLNbWRKY. I successfully designed the primers (primer set of FL04875F and FL04875R as shown in Table 4) and cloned this predicted full-length Nb04875 cDNA using total RNA extracted from an RCNMV-infected *N. benthamiana* leaf at 3 days post inoculation. The obtained clone was named NbWRKY.

The sequence alignment comparison between the predicted full-length cDNA (predFLNbWRKY) and the obtained clone (NbWRKY) is shown in Figure 7. The NbWRKY clone contains 325 amino acids and displays 90% identity to the predicted full-length cDNA. However, the NbWRKY clone is 99% identical to the Nb04875 unigene, with only one amino acid difference from arginine (R) to lysine (K) at position 162 (Figure 8). The side chain structures of these two amino acids are both basic polar with a positive charge. This amino acid (R → K) is located on the third β strand (β_3) of the four- β -stranded sheet of the predicted 3D WRKY protein structure as shown in Figure 10A and B. The predicted 3D structures of both Nb04875 and NbWRKY are very similar. Therefore, this amino acid difference (R → K) is unlikely to affect the protein folding.

2.5) The putative NbWRKY promoter contains W-boxes, indicating an auto/cross-regulation

In this study, I utilized bioinformatic tools to analyze the promoter region of NbWRKY. The promoter analysis of NbWRKY was performed on the nucleotide sequence retrieved from its genome data (*N. benthamiana* genome draft from Sol Genomics Network [41]). A BLAST search [28] of the full-length NbWRKY nucleotide sequence against the draft *N. benthamiana* genome database reveals 99% similarity with only one mismatched nucleotide to the database scaffold ID Niben.v0.4.2.Scf16974, hit region between 10802-12451. The search result also suggests that NbWRKY mRNA is a constituent product from a splicing process retrieved from three exons (Figure 11). The 1,500 bp nucleotide sequence upstream (Figure 11) from a predicted start codon of the NbWRKY gene retrieved from this scaffold was therefore used for its computational promoter analysis.

The core promoter is the minimal portion of the promoter required to properly initiate transcription. One of the important elements in the core promoter experimentally identified in both plants and animals is a TATA box feature [42]. This study utilized the PLACE bioinformatic tool [43] to computationally predict the promoter architecture of the NbWRKY gene. PLACE is a plant promoter prediction tool and a database of plant *cis*-regulatory DNA elements. Using PLACE, two TATA box elements are predicted at positions -743 (TATAAAA) and -330 (TATATAA) upstream from the predicted NbWRKY start codon (Figure 11). An interesting observation was the presence of three W-box regulatory elements at positions -1357 (TTGAC), -1059 (TTGAC), and -579 (TGACT) upstream from the

NbWRKY start codon (Figure 11). This finding is not uncommon for WRKY promoters. Indeed, multiple studies in *A. thaliana* have revealed that many WRKY genes that are themselves responsive to pathogenic stimuli typically contain numerous W box elements within their promoters [1, 8]. There are multiple reports demonstrating the interaction of WRKY TFs with either their own promoters or those of other family members [18, 24, 44, 45]. This suggests that several WRKY genes enriched for W boxes are extensively engaged under a direct positive or negative control by WRKY TFs via specific feed-back mechanisms (auto/cross regulation). The computational identification of W-box elements in the putative NbWRKY promoter leads to the conclusion that NbWRKY transcription is potentially regulated either by itself or by other WRKY TFs. Such a WRKY interactome may ensure the fast, efficient, and specific defense signal amplification upon a pathogen attack. In addition, it may also allow for a tighter control in limiting the extent of defense responses via negative feedback mechanisms.

3) NbWRKY - sGFP fusion protein localizes to the nucleus

Green fluorescent protein (GFP) was used as a visual protein tag to identify the subcellular localization [46] of the full-length NbWRKY protein in this study. The GFP was fused to the C-terminus of the full-length NbWRKY protein. This experimental design ensures that the translation will have to start from the full-length NbWRKY first and then the GFP, which is from the N-terminus to the C-terminus of the NbWRKY- GFP construct. A study in HEK293T cell (human cell line) indicated that the N-terminal tagging with GFP adversely affects the protein localization in reverse transfection assays, whereas tagging with

GFP at the C-terminus is generally better in preserving the localization of the native protein [47]. The study showed that all C-terminal fusion proteins localized to cellular compartments in accordance with previous studies and/or bioinformatic predictions [47]. The limitations of N-terminal or C-terminal tagging with GFP have not yet been thoroughly investigated in plant cells. However, with this concern, the study in this section utilized an experimental design to tag the full-length NbWRKY at its C-terminus. The GFP used in this study is a synthetic copy of GFP (sGFP) which was codon optimized for human expression and further modified with the addition of 6xHis to the N-terminus GFP structure in order to improve its stability [48].

The confocal images of NbWRKY-sGFP expressed in plant cells demonstrate that the full-length NbWRKY proteins are intensively localized inside the nucleus (Figure 12). This experimental evidence confirms the computational prediction regarding the NbWRKY nuclear localization that is discussed in section 2.3. However, the confocal images of sGFP (control) also show the typical characteristic localization of sGFP inside the nucleus. Many studies also confirm the nuclear translocation of sGFP as a result of GFP's relatively small size (~27 kDa), which allows for passive diffusion through the nuclear pores [49]. The size exclusion limit of nuclear pore complexes is approximately >60 kDa [50-52]. Any particle with a mass less than 60 kDa may passively diffuse into the nucleus. The natural nuclear localization of sGFP and the use of a GFP fusion product with a total molecular mass less than 60 kDa could hamper a conclusion regarding true nuclear localization. The mass of NbWRKY and the fusion protein NbWRKY-sGFP was predicted using a bioinformatic tool (ExPASy [33]). The computational molecular mass estimation of NbWRKY is 37 kDa and

NbWRKY-sGFP is 64 kDa, suggesting that the molecular mass of the NbWRKY-sGFP fusion product exceeds the limit for passive transport into the nucleus. This finding leads to the conclusion that the NLS computationally identified in NbWRKY truly facilitates the translocation of sGFP to the nucleus.

However, this initial study does not include an organelle staining/dye experiment for a specific subcellular localization comparison. The nuclear stain/dye study is therefore further required to precisely examine the nuclear localization of NbWRKY. A commonly used nuclear staining dye to use would be DAPI due to its strong binding to A-T rich regions in DNA. DAPI can pass through an intact cell membrane; therefore, it can be used to stain both live and fixed cells [53]. When bound to double-stranded DNA, DAPI has an absorption maximum at a wavelength 358 nm (ultraviolet) and its emission maximum at 461 nm (blue). Therefore, for fluorescence microscopy, DAPI is excited with ultraviolet light and is detected through a blue/cyan filter. DAPI's blue emission is practical for using with a multiple fluorescent staining in a single sample. There is some fluorescence overlap between DAPI and GFP but the effect is small. In a future study, the use of DAPI staining of the leaf tissue sample containing the expressed NbWRKY-sGFP will allow the more precise analysis of the nuclear localization of NbWRKY protein.

4) The functional impacts of NbWRKY on RCNMV replication

4.1) Pre-silenced NbWRKY promoted RCNMV replication at 24 hpi and increased RCNMV R1SG1-sGFP accumulation

The NbWRKY gene was significantly down-regulated > 8 fold ($\alpha = 0.05$) in *N. benthamiana* by TRV1:TRV2-NbWRKY VIGS vector system [54]. The silencing constructs TRV1:TRV2-NbWRKY were delivered to plant cells using an agroinfiltration method. The NbWRKY gene was pre-silenced 10 days prior to RCNMV inoculation.

The expression of the pre-silenced WRKY gene remained down-regulated throughout 72 hours of RCNMV infection (Figure 13). The pre-silenced NbWRKY affected RCNMV RNA-1 and RNA-2 replication (Figure 14) as well as R1SG1-sGFP accumulation (Figure 15). The RCNMV RNA-1, the CP subgenomic RNA and RNA-2 levels were significantly increased > 4 fold at 24 hpi ($\alpha = 0.1$). Also, R1SG1-sGFP accumulation was significantly increased > 3 fold after 3 days ($\alpha = 0.05$). However, the pre-silenced NbWRKY significantly decreased RNA-1 and its subgenomic RNA accumulation > 5 fold in the first 6 hours of RCNMV infection ($\alpha = 0.1$). At 12 hpi, both RNA-1 and RNA-2 levels were not affected by the pre-silenced NbWRKY. An interesting observation is that RNA-2 was significantly increased > 3 fold at 3 hpi, followed by decreasing > 5 fold at 6 hpi ($\alpha = 0.1$). This result suggests that the pre-silenced NbWRKY does not promote RCNMV replication until a late infection phase at 24 hpi; however, it promotes RNA-2 replication as early as 3 hpi.

The pre-silenced NbWRKY affected RCNMV replication and progeny RNA accumulation in a pattern similar to the effect caused by the pre-silenced NbSAR but with a

greater magnitude. RCNMV genomic and sub-genomic accumulation was increased 4 fold by the pre-silenced NbWRKY (Figure 14) compared to 2 fold by the pre-silenced NbSAR at 24 hpi (Chapter 4 Figure 4). This evidence leads to the conclusion that the pre-silenced NbWRKY produces a higher positive impact on RCNMV replication than the pre-silenced NbSAR.

The WRKY TFs are involved in regulating expression of many defense genes [55]. Also, individual WRKY TFs transcriptionally reprogram expression of more than one defense gene at a time. Given the possibility of NbWRKY functioning as a transcription regulator of more than one host defense gene, the pre-silenced NbWRKY could presumably govern host defense system on a larger scale than the pre-silenced NbSAR. This may explain why the pre-silenced NbWRKY produces a higher impact on RCNMV replication than the pre-silenced NbSAR. Another assumption is that NbWRKY and NbSAR expression is co-regulated at the early stage of a RCNMV infection. I speculate that NbWRKY gene expression is correlated to NbSAR gene expression and NbSAR gene expression may be under the control of NbWRKY TF. This speculation is drawn from the evidence that AtWRKY53 and AtWRKY70 from *A. thaliana* both positively modulate SAR [56].

Loss-of-function and gain-of-function studies in *Arabidopsis* have been pivotal in demonstrating that WRKY TFs control a complex defense response network as both positive and negative regulators [1]. The transient gene silencing study in this report demonstrates that the pre-silenced NbWRKY in *N. benthamiana* enhances host susceptibility toward RCNMV infection early (after 12 hpi), suggesting that the wild-typed NbWRKY functions as

a positive regulator of resistance against an RCNMV infection early. Considering that TF is an upstream regulator of gene expression, it suggests that a wild-type *N. benthamiana* host increases NbWRKY gene expression at 24 hpi in order to rapidly activate the expression of a wide range of defense-related genes for generating defense proteins or molecules which subsequently augment the launching of defense efforts against an RCNMV infection.

4.2) Pre-overexpressed NbWRKY did not affect RCNMV replication

In wild-type *N. benthamiana*, RCNMV infection triggers the activation of WRKY gene expression at 24 hpi. The pre-silenced NbWRKY increased RCNMV replication at 24 hpi. My hypothesis is that NbWRKY is a positive regulator of host defense against RCNMV infection and the increased NbWRKY expression subsequently induces host defense gene expression. The increased NbWRKY expression is therefore driven by the host itself to launch the defense response against RCNMV. To test this hypothesis, I have also performed the NbWRKY gene overexpression experiment and studied its impact on RCNMV replication.

The NbWRKY gene was overexpressed in *N. benthamiana* with the PZP212-NbWRKY overexpression construct which was derived from the pRTL2 expression cassette and PZP212 binary vector system. PZP212-NbWRKY was delivered to plant cells along with the RNA silencing suppressor (PZP212-HCPro) using agroinfiltration method. Two days after agroinfiltration, the WRKY expression in leaves agroinfiltrated with PZP212-NbWRKY/PZP212-HCpro was significantly increased greater than 40 fold compared to the

control leaves which are agroinfiltrated with only PZP212-HCPro ($\alpha = 0.05$) (Figure 16). The NbWRKY gene was overexpressed for two days prior to RCNMV inoculation.

Pre-silenced NbWRKY increases host vulnerability to RCNMV infection. Therefore, one would expect that the overexpression of NbWRKY would enhance resistance toward RCNMV infection. However, it appears that overexpression of NbWRKY did not significantly affect RCNMV replication and progeny RNA accumulation between 3 to 24 hours post RCNMV inoculation (Figure 17) ($\alpha = 0.1$). This finding is the opposite of the assumed outcome based on the results from the NbWRKY gene silencing study. Several reasons contribute to this unexpected observation. The full-length NbWRKY protein used in this study may not properly function as a TF inside the host cells. This could be due to an improper protein conformation that was previously discussed in section 2.1. Despite the prediction that this full-length NbWRKY protein would be capable of folding into the proper conformation based on its similarity to AtWRKY1, this study does not experimentally confirm that the NbWRKY protein is able to bind/interact with its target DNAs and initiate their transcription processes. Although this study lacks the biochemical evidence of NbWRKY functioning as a TF, the GFP fusion tagging NbWRKY demonstrates that the protein product of full-length NbWRKY used in this overexpression study does localize inside the nucleus as previously discussed in section 3.

If, on the other hand, the full-length NbWRKY protein is able to function properly as a TF, an alternative scenario may be that NbWRKY possesses opposing functions in regulating resistance toward RCNMV infection in wild-type *N. benthamiana*. This

assumption is drawn from the evidence that WRKY TFs from *A. thaliana* are involved in transcriptional reprogramming of plant defense gene expression in both directions, that is, as both positive and negative regulator, dependent on the invading pathogens. Dual functionality in defense signaling was observed for AtWRKY53 where mutants showed delayed symptom development against *Ralstonia solanacearum* while displaying increased susceptibility toward *Pseudomonas syringae* [57, 58]. Dual functionality was also suggested for AtWRKY41 with plants overexpressing AtWRKY41 showing enhanced resistance toward virulent *Pseudomonas* but decreased resistance toward *Erwinia carotovora* [59].

The pre-silenced NbWRKY increases host resistance toward RCNMV infection in the first 6 hours but dramatically decreases host resistance after 12 hours (Figure 14). This implies that NbWRKY may possess a dual regulatory function of the resistance toward RCNMV infection. The wild-typed NbWRKY may act as a negative regulator of the resistance toward RCNMV infection in the first 6 hours then shift its function to a positive regulator after 12 hours. This time-dependent response suggests that the dual functionality of NbWRKY may associate with the onset or duration of RCNMV infection.

Given functioning as negative regulator of resistance in the first 6 hours in wild-type *N. benthamiana*, the overexpression of NbWRKY may already prime the negative regulation of resistance and so decreasing a defense response prior to RCNMV infection. Thus, one would expect an increasing accumulation of RCNMV RNA in the first 6 hours. However, NbWRKY studied in this project is not the only naturally occurring WRKY TF in *N. benthamiana*. The WRKY TF family is well known for its structural and functional

redundancy regulating plant defense response upon pathogen attack [56, 60]. Plants evolved this ability to ensure the proper function of essential upper layer transcriptional regulatory network. If NbWRKY is a prime negative regulator of the defense surveillance system, a continuously overexpressed NbWRKY gene will dramatically suppress plant immunity and place it in a very vulnerable state to any pathogen attack. To avoid this unexpected circumstance in the WRKY regulatory network, the redundancy of WRKY genes may compensate or overcome the consequences of NbWRKY overexpression by increasing other defense strategies. Therefore, the overexpressed NbWRKY in this study would not likely affect RCNMV RNA accumulation. Manipulation of WRKY gene expression by RCNMV may partly reveal the existence of redundancy within the WRKY TF family in *N. benthamiana* as reinforcement to the essential regulatory functions. In the case of the redundant WRKY genes that compensate when other WRKY gene functions are disturbed, gene mutation or gene downregulation experiments of more than one WRKY gene may prove to be more effective at manipulating host resistance to pathogen attack.

The functional redundancy among structurally related WRKY family members has been observed in *A. thaliana*. For instance, AtWRKY53 and AtWRKY70 are both a positive modulator of SAR [56]. AtWRKY18, AtWRKY40, and AtWRKY60 form both homocomplexes and heterocomplexes and DNA binding activities were shifted depending on which WRKY proteins were present in these complexes [60]. Single *WRKY* mutant exhibits no or small alterations in response to *Pseudomonas syringae* and *Botrytis cinerea*. However, *wrky18 wrky40* and *wrky18 wrky60* double mutants and the *wrky18 wrky40 wrky60* triple mutant were substantially more resistant to *P. syringae* but more susceptible to *B. cinerea*

than wild-type plants. Thus, the three WRKY proteins have partially redundant roles in plant responses to the two distinct types of pathogens, with WRKY18 playing a more important role than the other two.

Conclusion

This study demonstrates that Nb04875, assigned the name NbWRKY, is a novel WRKY gene found in *N. benthamiana*. In a wild-type *N. benthamiana*, microarray and qRT-PCR analysis indicated that NbWRKY expression was not increased until very late in an RCNMV infection at 24 hpi. It was previously found in Chapter 3 that at 24 hours RCNMV has completed its replication cycle in an initially infected cell and has begun second rounds of replication in neighboring infected cells. My hypothesis is that increasing NbWRKY expression is driven by the plant host, specifically, very late in the virus infection at 24 hours, in order to defend itself against RCNMV infection. However, the plant host does not make an effort to increase NbWRKY before 12 hours of RCNMV infection based on the assumption that NbWRKY has dual functionality as both a negative regulator of host defense in the first 6 hours and a positive regulator after 12 hours. Increasing expression of NbWRKY prior to 12 hours of infection does not likely benefit plant well-being but only costs energy without profit in return as seen in the overexpression of NbWRKY experiment where over production of NbWRKY did not affect early RCNMV infection. This implies that a coordinated modulation of positive- and negative- acting host defense factors is an essential feature to

facilitate the proper amplitude and duration of plant defense response toward RCNMV attack.

Some key questions remain to be addressed in the future investigations of NbWRKY and its impact on RCNMV replication. These key questions are: (i) what are the exact gene targets that are transcriptionally regulated by NbWRKY?, (ii) how are NbWRKY genes themselves regulated?, (iii) which cellular/nuclear component(s) does NbWRKY interact with during regulating resistance toward RCNMV infection?, and (iv) what other *N. benthamiana* WRKY genes/proteins might be co-regulated with NbWRKY during RCNMV infection, and what is the relationship/association between these co-regulated WRKY genes? The promoter analysis of the NbWRKY gene described in this study demonstrates that NbWRKY possibly contains at least three W-boxes that could potentially interact with WRKY TFs (Figure 11). This analysis is not experimentally confirmed but the computational evidence suggests that the transcription of NbWRKY may be regulated by itself in an auto-feedback loop mechanism or by other WRKY TFs as a cross regulation.

Table 1 *N. benthamiana* WRKY (transcription factor) unigene annotation information. The *N. benthamiana* unigene nucleotide sequences are searched against GenBank non-redundant protein sequence database (nr) using a BLASTX method. The conserved WRKY DNA-binding domain is identified by a search of the unigene sequences against GenBank conserved domain database (CDD).

Unigene ID	Annotation description	Conserved WRKY domain (smart00774, pfam03106)*	GenBank accession number	E-value
Nb00460	WRKY TF30 (<i>Capsicum annuum</i>)	no	gi 209978913	7 X 10 ⁻¹⁶
Nb03337	predicted WRKY TF13-like isoform X2 (<i>Citrus sinensis</i>)	no	gi 568819978	4 X 10 ⁻⁹
Nb04425	WRKY TF (<i>Capsicum annuum</i>)	no	gi 254762128	8 X 10 ⁻¹²
Nb04426	WRKY TF (<i>Capsicum annuum</i>)	yes	gi 254762128	1 X 10 ⁻¹⁴⁰
Nb04875	WRKY TF6 (<i>Solanum tuberosum</i>)	yes	gi 156118328	4 X 10 ⁻¹¹⁵
Nb05135	putative WRKY TF1a (<i>Coffea arabica</i>)	no	gi 86155943	6 X 10 ⁻⁶⁴
Nb05865	predicted WRKY TF72-like (<i>Solanum tuberosum</i>)	yes	gi 565381839	2 X 10 ⁻⁸⁹
Nb05926	DNA-binding protein 2 (<i>Nicotiana tabacum</i>)	yes	gi 4322940	1 X 10 ⁻¹¹⁶
Nb06607	predicted WRKY TF22-like (<i>Solanum tuberosum</i>)	yes	gi 565345089	8 X 10 ⁻²⁶
Nb06938	WRKY1 (<i>Nicotiana benthamiana</i>)	yes	gi 47176940	2 X 10 ⁻⁸²
Nb06939	WRKY3 (<i>Nicotiana benthamiana</i>)	yes	gi 47176938	5 X 10 ⁻⁴⁹
Nb08344	predicted WRKY TF4-like isoform X2 (<i>Solanum tuberosum</i>)	yes	gi 565366990	1 X 10 ⁻⁹⁹
Nb09029	predicted WRKY TF32-like (<i>Solanum tuberosum</i>)	yes	gi 565353448	1 X 10 ⁻⁵⁷
Nb09156	WRKY TF, WIZZ (wound-induced leucine zipper zinc finger) (<i>Nicotiana tabacum</i>)	no	gi 6472585	5 X 10 ⁻⁵⁹
Nb10179	WRKY TF2 (<i>Nicotiana benthamiana</i>)	yes	gi 45239442	9 X 10 ⁻¹¹⁶
Nb10180	DNA-binding protein 2 (<i>Nicotiana tabacum</i>)	yes	gi 4322940	7 X 10 ⁻¹⁵³
Nb10247	predicted WRKY TF55-like (<i>Solanum lycopersicum</i>)	yes	gi 460390958	2 X 10 ⁻¹²⁴
Nb10311	predicted WRKY TF22-like (<i>Solanum tuberosum</i>)	yes	gi 565345089	6 X 10 ⁻⁶²
Nb10418	predicted WRKY TF6-like (<i>Solanum tuberosum</i>)	yes	gi 565397153	4 X 10 ⁻¹⁷⁵
Nb10967	WRKY TF (<i>Capsicum annuum</i>)	no	gi 254762128	7 X 10 ⁻³⁷

Table 1 Continued

Unigene ID	Annotation description	Conserved WRKY domain (smart00774, pfam03106)*	GenBank accession number	E-value
Nb11052	WRKY TF2 (<i>Capsicum annuum</i>)	yes	gi 164666156	0.00
Nb11159	predicted WRKY TF57-like isoform 1 (<i>Solanum lycopersicum</i>)	yes	gi 460386766	4 X 10 ⁻¹³⁰
Nb11752	predicted WRKY TF48-like (<i>Solanum tuberosum</i>)	yes	gi 565375607	3 X 10 ⁻²⁵
Nb12332	predicted WRKY TF69-like isoform X1 (<i>Solanum tuberosum</i>)	yes	gi 565364358	5 X 10 ⁻¹³⁵

* The accession numbers “smart00774 and pfam03106” are assigned to proteins containing WRKY DNA-binding domain.

TF = transcription factor

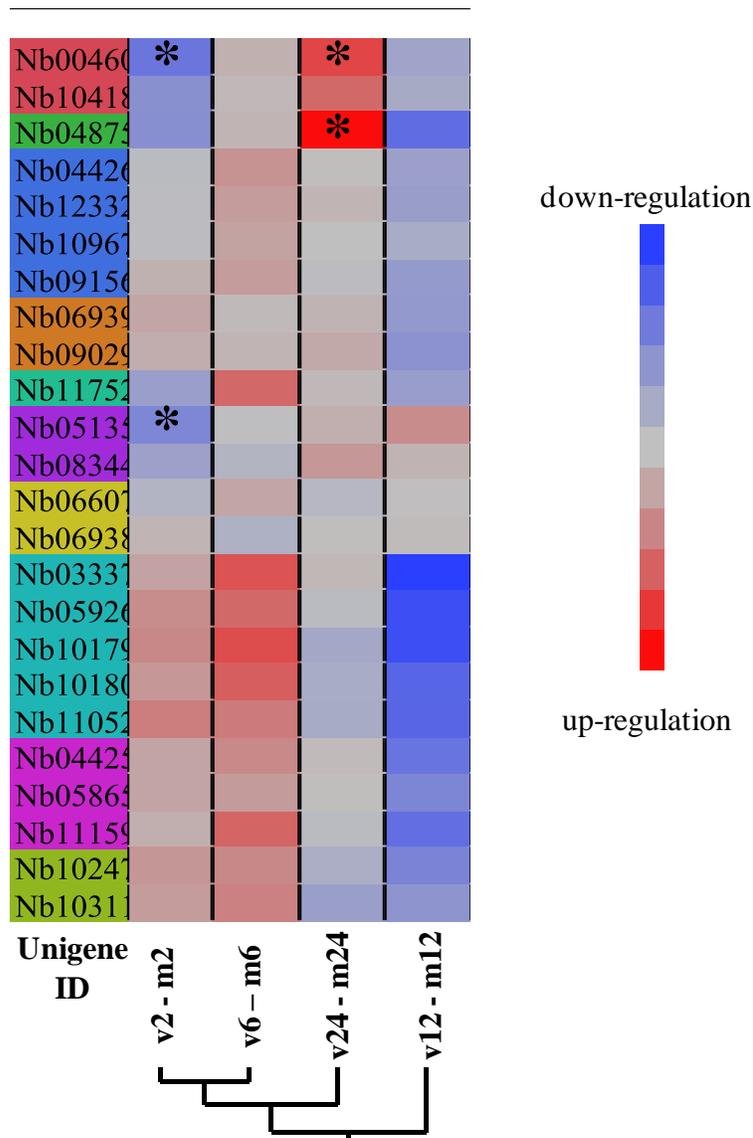


Figure 1 Hierarchical clustering the differential microarray expression data of *N. benthamiana* WRKY genes at 2, 6, 12 and 24 hours post RCNMV inoculation (hpi). A differential expression of each host WRKY gene was computed from a subtraction between microarray expression data of RCNMV-inoculated plants (v) and mocks (m) at each time point of study. An asterisk (*) indicates a significant differential host gene expression at FDR cutoff of 0.01. Unigene ID is colored according to its cluster. The host WRKY gene expression patterns at 2, 6 and 24 hpi appear to be clustered in the same expression group according to the duration of infection. This clustering result suggests that the WRKY gene expression pattern at 12 hpi is left out. The distinct WRKY gene expression pattern at 12 hpi is likely a consequence of RCNMV migrating to neighboring cells, which was previously discussed in Chapter 3.

Table 2 Nb04875 has the top BLAST hits to WRKY transcription factors from various plant species. The unigene sequence Nb04875 was searched against the GenBank non-redundant protein sequence (nr) database using BLASTX.

Species name	BLASTX hit description	GenBank Accession number	Query coverage	E-value	Identity
<i>Solanum tuberosum</i> (potato)	StWRKY6	NP_001275414	96%	4×10^{-115}	66%
<i>Nicotiana tabacum</i> (tobacco)	DNA-binding protein 4	AAF61864	60%	3×10^{-72}	63%
<i>Theobroma cacao</i> (cocoa)	WRKY70, putative isoform 1	XP_007033512	91%	1×10^{-58}	42%
<i>Cucumis sativus</i> (cucumber)	WRKY70	NP_001267577	64%	2×10^{-55}	48%
<i>Jatropha curcas</i> (Barbados nut)	WRKY56	AGQ04250	93%	1×10^{-54}	41%
<i>Ricinus communis</i> (castor oil plant)	WRKY transcription factor, putative	XP_002528697	85%	1×10^{-54}	42%
<i>Morus notabilis</i> (mulberry tree)	putative WRKY transcription factor 70	EXB74394.1	63%	6×10^{-54}	44%
<i>Vitis pseudoreticulata</i> (grapevine)	WRKY transcription factor	ACY69975.1	75%	1×10^{-51}	41%
<i>Gossypium hirsutum</i> (cotton)	WRKY transcription factor 5	AGV75930.1	89%	9×10^{-49}	37%
<i>Gossypium hirsutum</i> (cotton)	WRKY transcription factor 31	AIE43830.1	89%	4×10^{-48}	37%

Table 3 WRKY (transcription factor) homolog sequence information. These genes were used in the multiple sequence alignments in Figure 2 in order to identify WRKY DNA binding domains and zinc finger motifs in Nb04875 and NbWRKY. This sequence information was retrieved from a BLAST result of Nb04875 and NbWRKY against the GenBank non-redundant protein sequence database (nr) and conserved domain database (CDD).

Species name	GenBank accession ID	Description
<i>Arabidopsis thaliana</i>	gi 56966912	Solution structure of C-terminal WRKY domain of Atwrky4
<i>Arabidopsis thaliana</i>	gi 18400580	Putative WRKY transcription factor 17
<i>Arabidopsis thaliana</i>	gi 237769813	Putative disease resistance protein , WRKY DNA binding domain
<i>Arabidopsis thaliana</i>	gi 118137307	Crystal structure Of C-terminal WRKY domain of Atwrky1, SA-induced and partially Npr1-dependent transcription factor
<i>Arabidopsis thaliana</i>	gi 29839678	WRKY DNA-binding protein 46
<i>Arabidopsis thaliana</i>	gi 20978793	WRKY DNA-binding protein 39
<i>Arabidopsis thaliana</i>	gi 29839684	WRKY DNA-binding protein 29
<i>Arabidopsis thaliana</i>	gi 29839672	WRKY DNA-binding protein 14
<i>Arabidopsis thaliana</i>	gi 29839444	WRKY DNA-binding protein 22
<i>Arabidopsis thaliana</i>	gi 20978782	WRKY DNA-binding protein 18
<i>Arabidopsis thaliana</i>	gi 29839551	WRKY DNA-binding protein 41
<i>Avena sativa</i>	gi 4894963	DNA-binding protein WRKY3
<i>Avena sativa</i>	gi 75215322	DNA-binding protein WRKY3
<i>Cucumis sativus</i>	gi 525507067	Probable WRKY transcription factor 70-like
<i>Nicotiana benthamiana</i>	gi 110354649	WRKY cDNA clone
<i>Nicotiana tabacum</i>	gi 7406997	DNA-binding protein 4
<i>Nicotiana tabacum</i>	gi 14530681	WRKY DNA-binding protein
<i>Nicotiana tabacum</i>	gi 75213273	NtWRKY2
<i>Oryza sativa</i> (Indica)	gi 218193913	Hypothetical protein OsI_13911, WRKY DNA binding domain
<i>Oryza sativa</i> (Indica)	gi 125550704	Hypothetical protein OsI_18309, WRKY DNA binding domain
<i>Oryza sativa</i> (Japonica)	gi 46394316	WRKY transcription factor 31
<i>Oryza sativa</i> (Japonica)	gi 57863827	Hypothetical protein WRKY DNA binding domain
<i>Oryza sativa</i> (Japonica)	gi 297607012	Hypothetical protein Os07g0273700, WRKY DNA binding domain
<i>Solanum tuberosum</i>	gi 156118328	WRKY transcription factor 6 or StWRKY6
<i>Theobroma cacao</i>	gi 590653760	WRKY DNA-binding protein 70, putative isoform 1

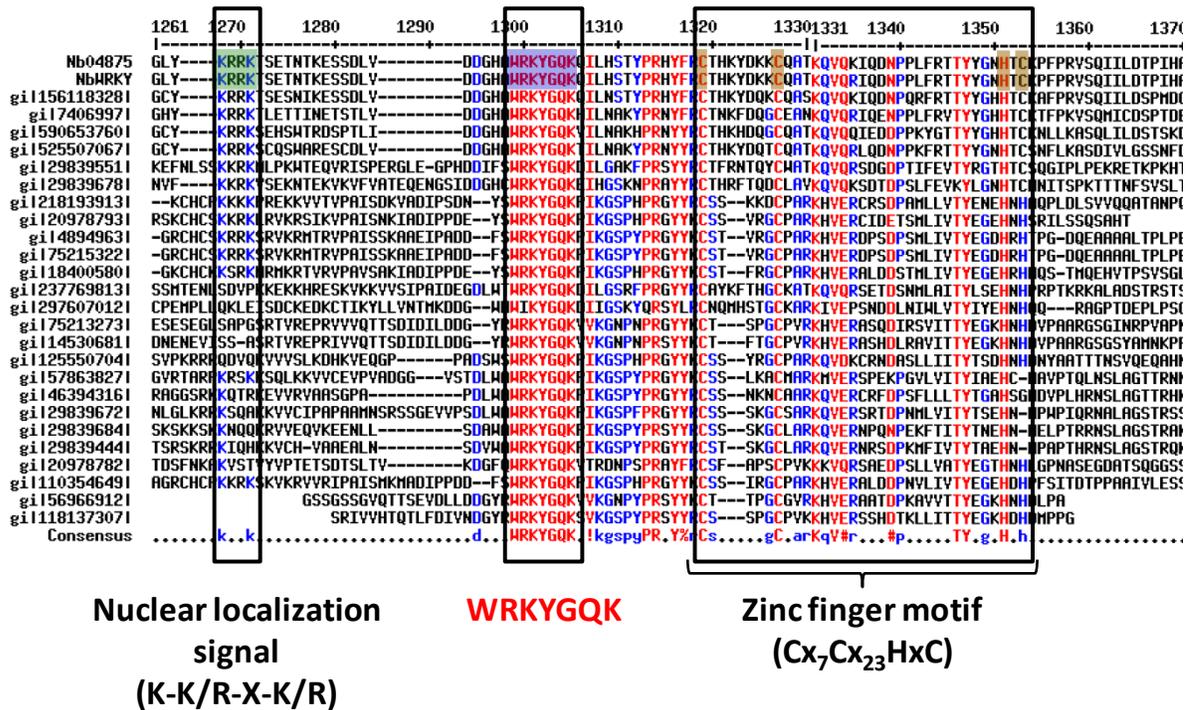


Figure 2 WRKY homolog multiple sequence alignment identifies WRKY DNA binding domain (WRKYGQK, blue box), zinc finger motif (Cx₇Cx₂₃HxC, brown box), and a nuclear localization signal (NLS) (KRRK, green box) in Nb04875 and NbWRKY. The WRKY homolog sequence information is described in Table 3.

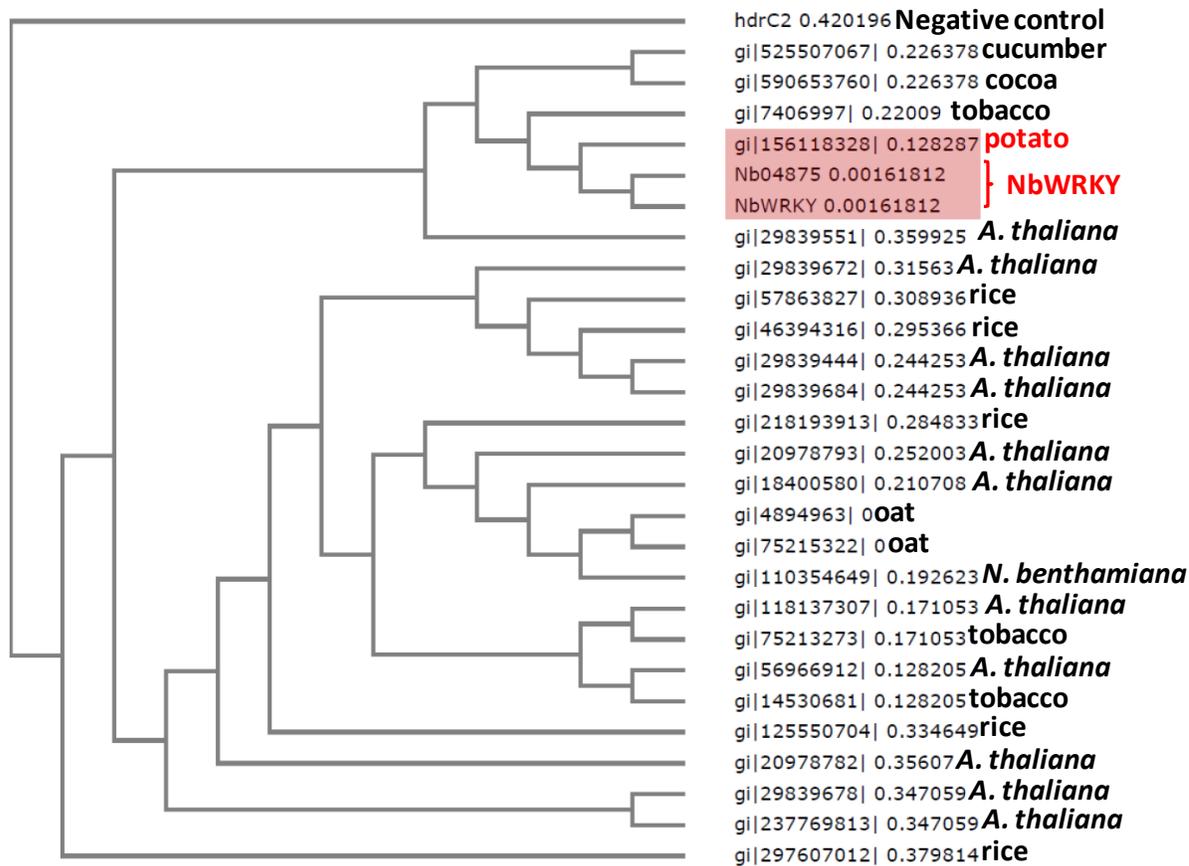


Figure 3 A phylogenetic tree of WRKY homologs indicates StWRKY6 gene is the closest homolog to Nb04875 and NbWRKY (red label). The phylogenetic tree was constructed by ClustalW2-Phylogeny web application using UPGMA method and hdrC2 as a control. Homolog information and accession number used in this analysis is described in Table 3.

```

1      10      20      30      40      50      60      70
|-----|-----|-----|-----|-----|-----|
NbWRKY  MESPLPEKSSADLKRAIDGLIRGQELTRRLKEIIEKPLGGEVATIMAENLIGKIMNSFSETLSVIN
Nb04875 t.khnMESPLPEKSSADLKRAIDGLIRGQELTRRLKEIIEKPLGGEVATIMAENLIGKIMNSFSETLSVIN
Consensus  ...MESPLPEKSSADLKRAIDGLIRGQELTRRLKEIIEKPLGGEVATIMAENLIGKIMNSFSETLSVIN

71     80     90     100    110    120    130    140
|-----|-----|-----|-----|-----|-----|
NbWRKY  SGESDEDTAEVKSSSEDSSGCKSTSFKDRRGLYKRRKTSETNTKESDLVDDGHAWRKYGQKQILHSTYP
Nb04875 SGESDEDTAEVKSSSEDSSGCKSTSFKDRRGLYKRRKTSETNTKESDLVDDGHAWRKYGQKQILHSTYP
Consensus  SGESDEDTAEVKSSSEDSSGCKSTSFKDRRGLYKRRKTSETNTKESDLVDDGHAWRKYGQKQILHSTYP

141    150    160    170    180    190    200    210
|-----|-----|-----|-----|-----|-----|
NbWRKY  RHYFRCTHKYDKKCQATKQVQrIQDNPPFLRTTYGNIHTCKPFPRVSIILDTPIHADSSILLCFDHNNN
Nb04875 RHYFRCTHKYDKKCQATKQVQkIQDNPPFLRTTYGNIHTCKPFPRVSIILDTPIHADSSILLCFDHNNN
Consensus  RHYFRCTHKYDKKCQATKQVQrIQDNPPFLRTTYGNIHTCKPFPRVSIILDTPIHADSSILLCFDHNNN

211    220    230    240    250    260    270    280
|-----|-----|-----|-----|-----|-----|
NbWRKY  NNYSSVQANHNHYQPYDIPTFPSIKQETKEEVFQRSCTYYPKIEYQNQSSNSDYFLQANDVHLSTPAYFE
Nb04875 NNYSSVQANHNHYQPYDIPTFPSIKQETKEEVFQRSCTYYPKIEYQNQSSNSDYFLQANDVHLSTPAYFE
Consensus  NNYSSVQANHNHYQPYDIPTFPSIKQETKEEVFQRSCTYYPKIEYQNQSSNSDYFLQANDVHLSTPAYFE

281    290    300    310    320    328
|-----|-----|-----|-----|-----|
NbWRKY  ASGDHMAAALSPDVISSGVYSSRTTSTONLEIDfdfeeslwnfegyns
Nb04875 ASGDHMAAALSPDVISSGVYSSRTTSTONLEID
Consensus  ASGDHMAAALSPDVISSGVYSSRTTSTONLEID.....

```

Figure 4 NbWRKY and Nb04875 both contain a nuclear localization signal (NLS). The NLS prediction was performed by the PSORT bioinformatic tool. The monopartite NLS pattern of the consensus four residue sequence K-K/R-X-K/R was detected in the conceptually translated NbWRKY and Nb04875 as K-R-R-K (green box).

Unigene ID (Query)	Unigene assigned name	BLASTX hit	Accession number	Query coverage	E-value	Identity
Nb04875 940 bp (313 aa), partial CDS	NbWRKY	StWRKY6 (<i>S. tuberosum</i>) 1036 bp (293 aa) complete CDS	NP_001275414	96%	4×10^{-115}	66%

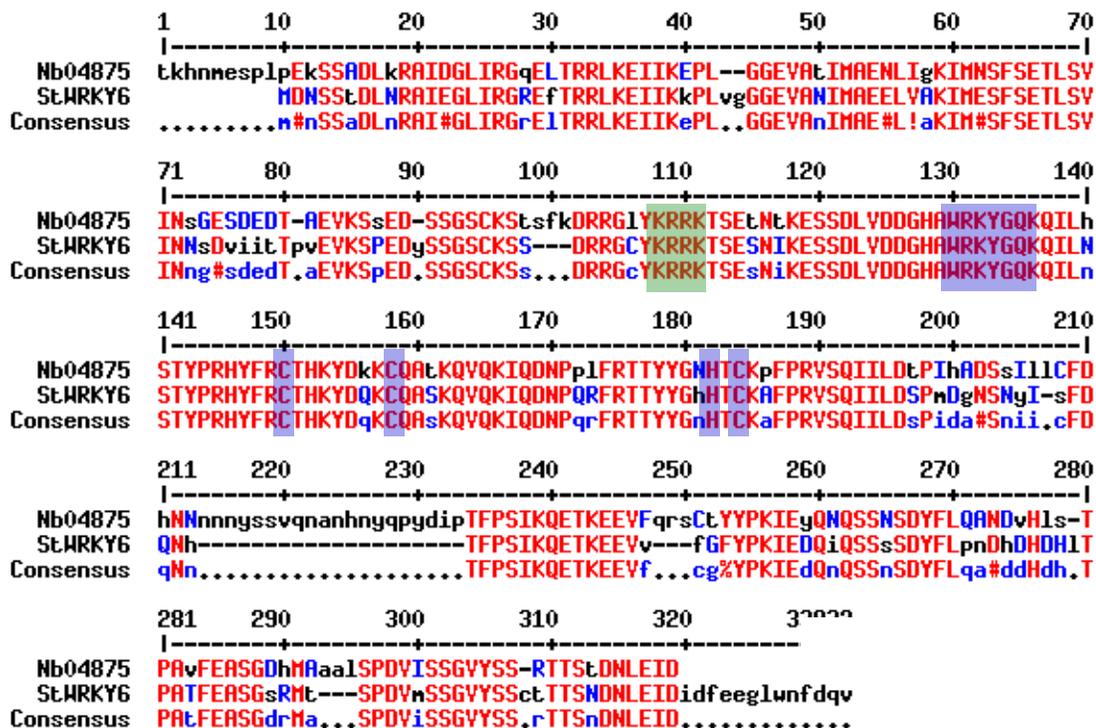


Figure 5 Nb04875 has a top BLAST hit to potato StWRKY6. The unigene sequence Nb04875 was searched against the GenBank non-redundant protein sequence (nr) database using BLASTX. The unigene Nb04875 has 66% sequence identity to StWRKY6 (potato). WRKY DNA binding domain is identified with a blue box. A nuclear localization signal (NLS) is identified by a green box.

acgaaacacaac**atg**gagtctccggtgccgaaaagtcacagctgatctgaaaaggca
 T K H N **M** E S P L P E K S S A D L K R A
 atcgacgggttaattcgtggccaggaattaacgcgccgattaaaagagataatcaaggaa
 I D G L I R G Q E L T R R L K E I I K E
 cctcttggtggtgaagttgcaaccattatggctgagaatttaattggaaaaattatgaat
 P L G G E V A T I M A E N L I G K I M N
 tcattttccgaaacactttccgtcataaactccggcgagtcggatgaggataccgccgaa
 S F S E T L S V I N S G E S D E D T A E
 gttaagtcgtcgggaagattctagtggaagttgcaagagtacatcattcaaagatcgaaga
 V K S S E D S S G S C K S T S F K D R R
 ggattgtaca**agagaaggaaa**acttcagaacaaacacaaaagaatcctcagatttggtg
 G L Y **K R R K** T S E T N T K E S S D L V
 gatgatggtcatgct**tgagaaaatatggacaaaaa**caaatcctccattccacttatcca
 D D G H A **W R K Y G Q K** Q I L H S T Y P
 aggcaactatthtagg**tgc**actcataaatatgataaaaa**tgt**caagcaaccaagcaggtg
 R H Y F R **C** T H K Y D K K **C** Q A T K Q V
 cagaaaattcaagacaatccaccactattccggacaacatactatggaaat**cacaca****tgt**
 Q K I Q D N P P L F R T T Y Y G N **H T C**
 aaactttccctagagtttctcaataatthttggataccctattcatgcccattcctct
 K P F P R V S Q I I L D T P I H A D S S
 attctactttgtttgatcacaacaacaataattactcttcagtccaaaacgcta
 I L L C F D H N N N N N Y S S V Q N A N
 cataattaccagccttatgatattcctacatttccctcaataaaacaggaaaccaagag
 H N Y Q P Y D I P T F P S I K Q E T K E
 gaagtattccaaagatcatgtacttactatccaaaaatagaatatcaaaaccaatcatca
 E V F Q R S C T Y Y P K I E Y Q N Q S S
 aactctgattatthttctacaagccaatgatgttcatctgagtactccggcagatthttaa
 N S D Y F L Q A N D V H L S T P A V F E
 gcctccggcgaccacatggcggcggcattgtcgccggatgtcatatcatctggggtctat
 A S G D H M A A A L S P D V I S S G V Y
 tcttctcgtacgactagtagataatcttgagattgatt
 S S R T T S T D N L E I D

Figure 6 Nb04875 nucleotide sequence (940 bp) is conceptually translated to 313 amino acids by using ExPASy translate tool and standard genetic code. The translation starts at frame 1 in the 5' to 3' direction. This translation frame was chosen because it includes a predicted start codon (M, red), a predicted NLS (KRRK, green), a predicted WRKY DNA binding domain (WRKYGQK, blue) and zinc finger motif (CX₇CX₂₃HXC, pink and italicized). The structure of the unigene Nb04875 is only a partial coding sequence missing the 3' cDNA end.

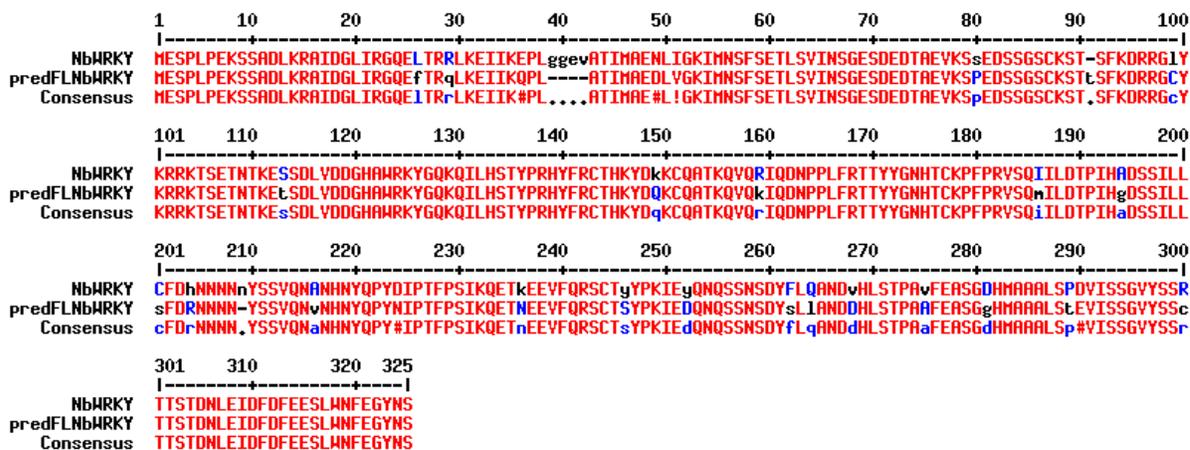


Figure 7 Amino acid sequence comparison between the predicted full-length Nb04875 cDNA (predFLNbWRKY) and the obtained clone (NbWRKY). The prediction for predFLNbWRKY was derived from the *N. benthamiana* genomic draft (Sol Genomic Network database).

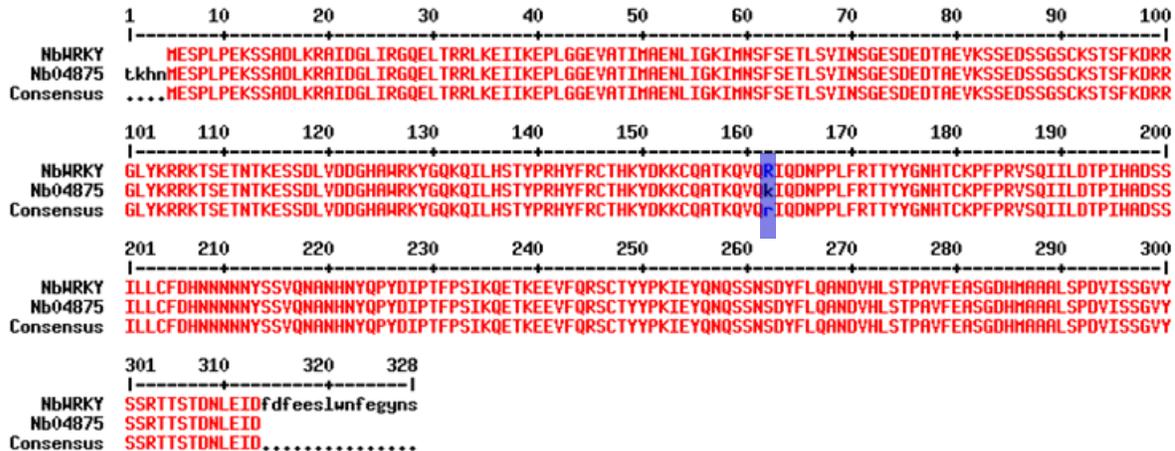
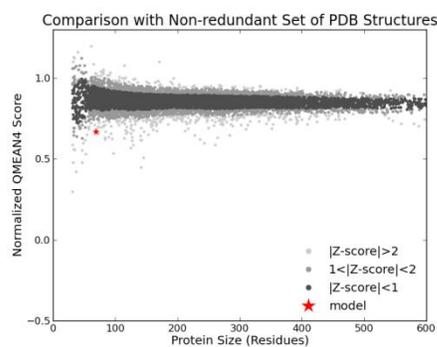
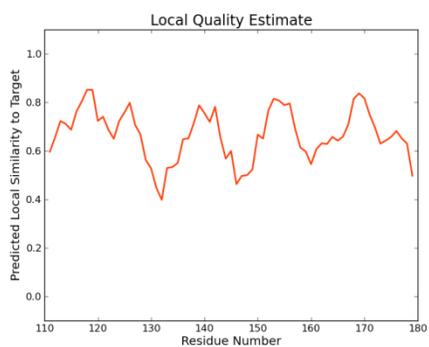
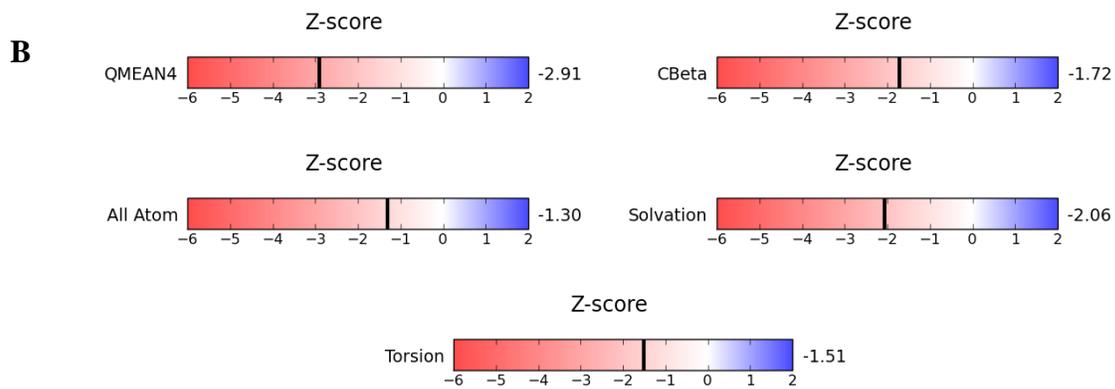
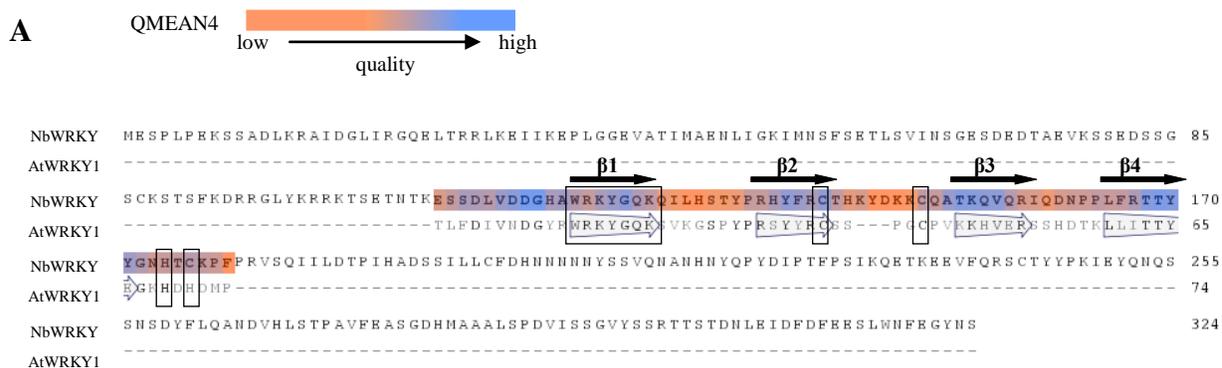


Figure 8 Amino acid sequence comparison between the obtained clone NbWRKY and the unigene Nb04875. NbWRKY clone has a 99% identity to the Nb04875 unigene, with only one amino acid difference from arginine (R) to lysine (K) (blue box).

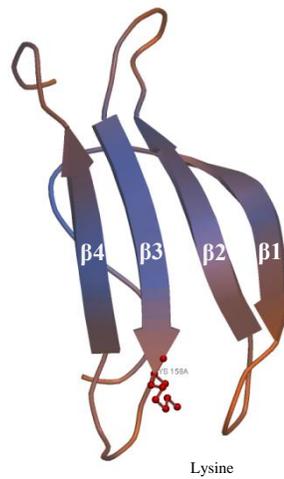
Figure 9 Use of AtWRKY1 to predict NbWRKY three-dimensional protein structure. NbWRKY protein structure was predicted using the SWISS-MODEL web application. AtWRKY1 (2ayd.1.A) was used as the model template to predict NbWRKY protein structure since it shared the highest homology with 42.2% sequence identity. A) NbWRKY tertiary structure prediction. A sequence alignment comparison between NbWRKY and AtWRKY1 revealed that NbWRKY is capable of folding into a tertiary structure consisting of four β strands (arrow). The DNA binding domain with sequence pattern “WRKYGQK” is identified in the first β strand of AtWRKY1 and NbWRKY. QMEAN4 is model quality estimation. The blue shade shows a higher prediction quality and the orange shade is a lower prediction quality. The box in β 1 indicates a conserved WRKY DNA binding domain motif. B) NbWRKY model quality. An overall estimation of NbWRKY model quality is based on the reference template AtWRKY1. Higher numbers (toward the blue region) indicates a higher model quality.



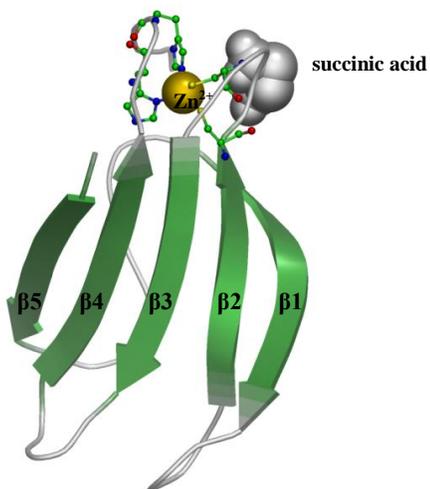
A) NbWRKY 3D model



B) Nb04875 3D model



C) AtWRKY1 (2ayd.1A) 3D template model



D) AtWRKY4 (1wj2.1) 3D template model

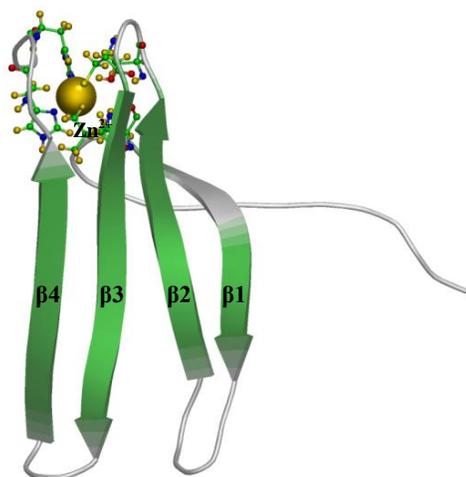
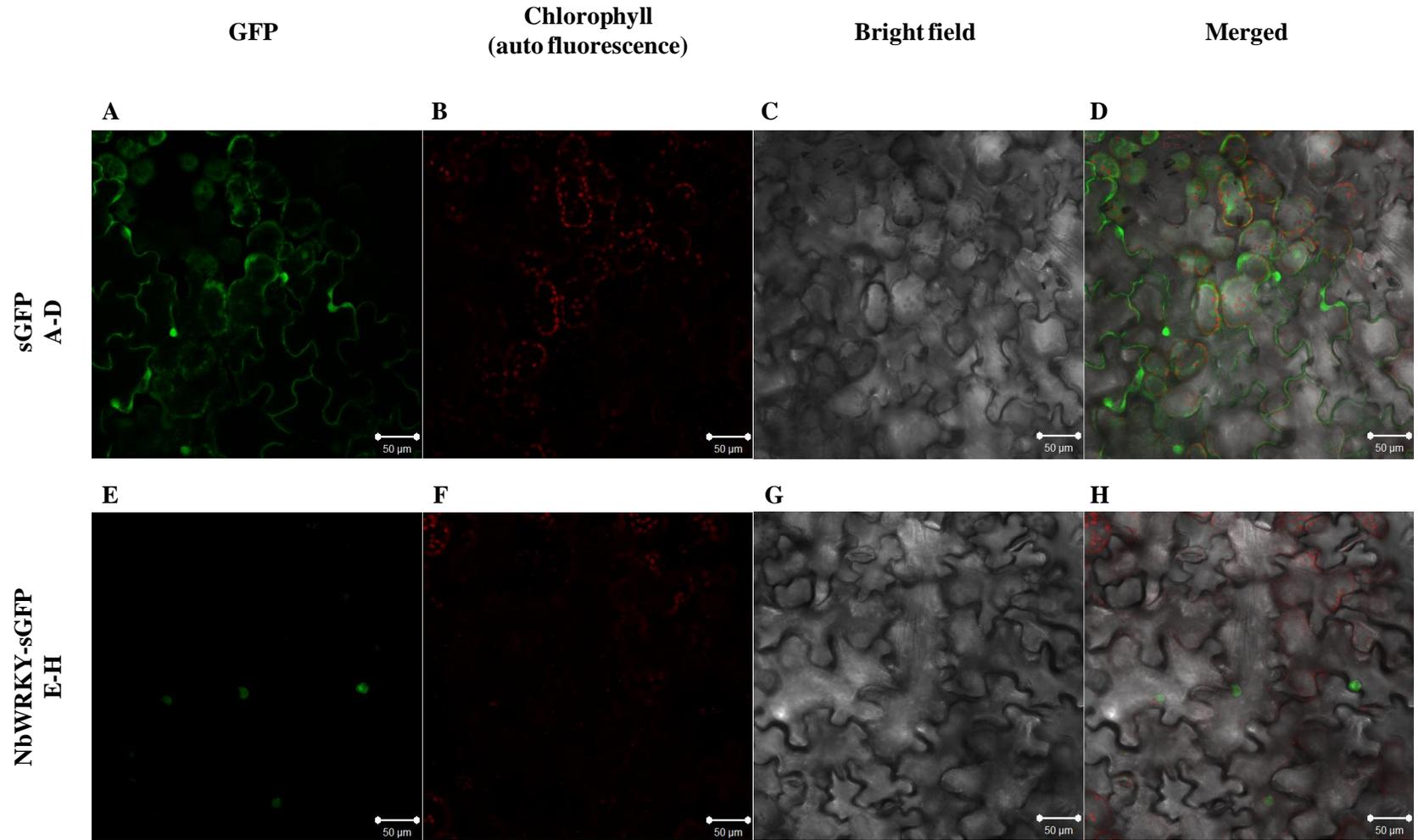


Figure 10 NbWRKY and Nb04875 three dimensional protein structure model prediction using SWISS-MODEL. The highest structure match to NbWRKY (A) and Nb04875 (B) is AtWRKY1 (C) and AtWRKY4 (D). One amino acid difference between NbWRKY (A) and Nb04875 (B) from arginine to lysine had no effect on protein folding.

Figure 11 NbWRKY promoter architecture predicted by PLACE bioinformatic tool. Promoter analysis was performed on the 1,500 bp region upstream from the predicted start codon of the NbWRKY gene. The upstream nucleotide sequence was retrieved from the *N. benthamiana* genomic draft (Sol Genomic Network database), scaffold ID Niben.v0.4.2.Scf16974 (NbWRKY hit region between 10802-12451). Three W-box elements are identified in green text. Two TATA box elements are identified in red text. Three exons are identified with blue shading. Predicted start and stop codons are underlined.

Figure 12 Nuclear localization of NbWRKY protein. A modified green fluorescent protein derivative (sGFP) was used as a visual tag to identify the subcellular localization of the full-length NbWRKY protein. The full-length NbWRKY protein was fused at its C-terminus to the sGFP tagging protein to generate NbWRKY-sGFP fusion protein. The localization constructs were agroinfiltrated into *N. benthamiana* leaves and the photographs were taken 2 days after agroinfiltration with a confocal microscope. The images were scanned using three different filters (green = GFP, red = chloroplast auto fluorescence, and white/black = bright field). (A-D) shows a typical pattern of sGFP subcellular localization (control), images taken from a leaf agroinfiltrated with a mixture of PZP212-sGFP and PZP212-HCPro. A merged image of sGFP (D) shows that a fluorescence signal was accumulated inside the nucleus and cytoplasm. However, sGFP does not appear to localize to the chloroplast. (E-H) shows a characteristic pattern of NbWRKY-sGFP subcellular localization, images taken from a leaf agroinfiltrated with a mixture of PZP212-NbWRKY-sGFP and PZP212-HCPro. A merged image of NbWRKY-sGFP (H) indicates that a fluorescence signal was most concentrated inside the nucleus.



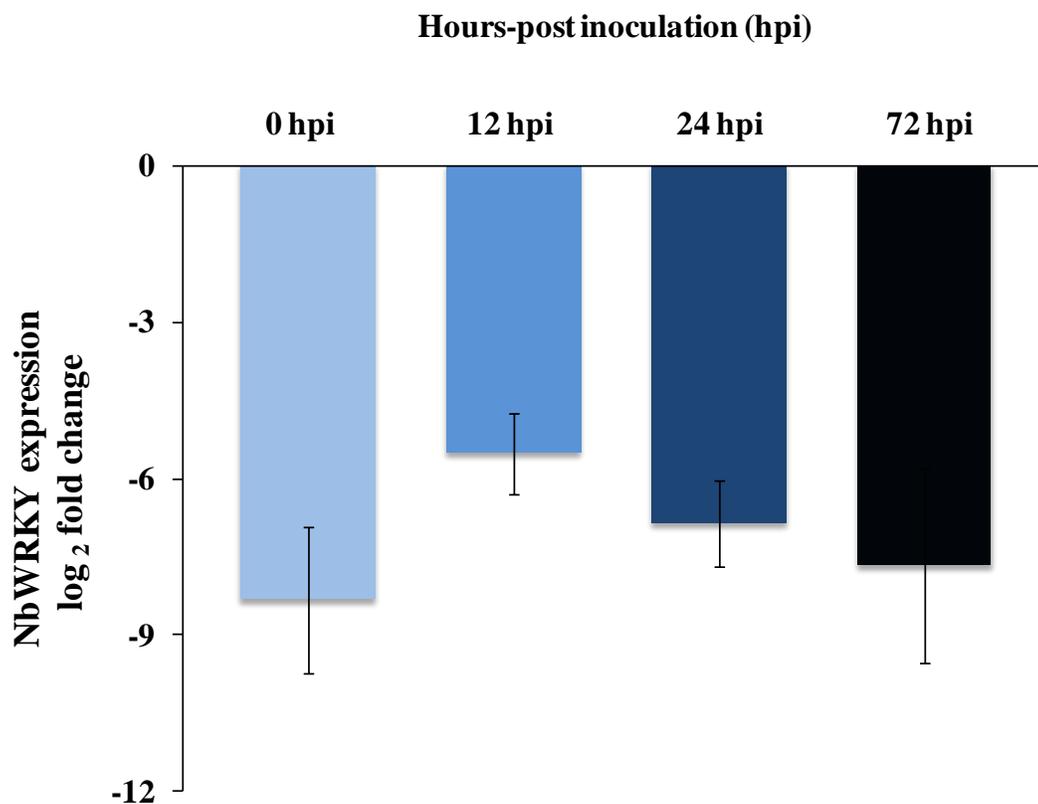


Figure 13 NbWRKY gene expression was suppressed in *N. benthamiana* by using TRV-VIGS vector system. The silencing vector TRV1:TRV2-NbWRKY was delivered into plant cells by an agroinfiltration method. NbWRKY was pre-silenced 10 days prior to RCNMV inoculation. The pre-silenced NbWRKY plants were inoculated with RCNMV RNA-1/RNA-2, and then harvested between 0 and 72 hours-post RCNMV inoculation (hpi). The qRT-PCR analysis showed that the expression of pre-silenced NbWRKY remained significantly repressed over 72 hpi ($\alpha = 0.05$). Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis.

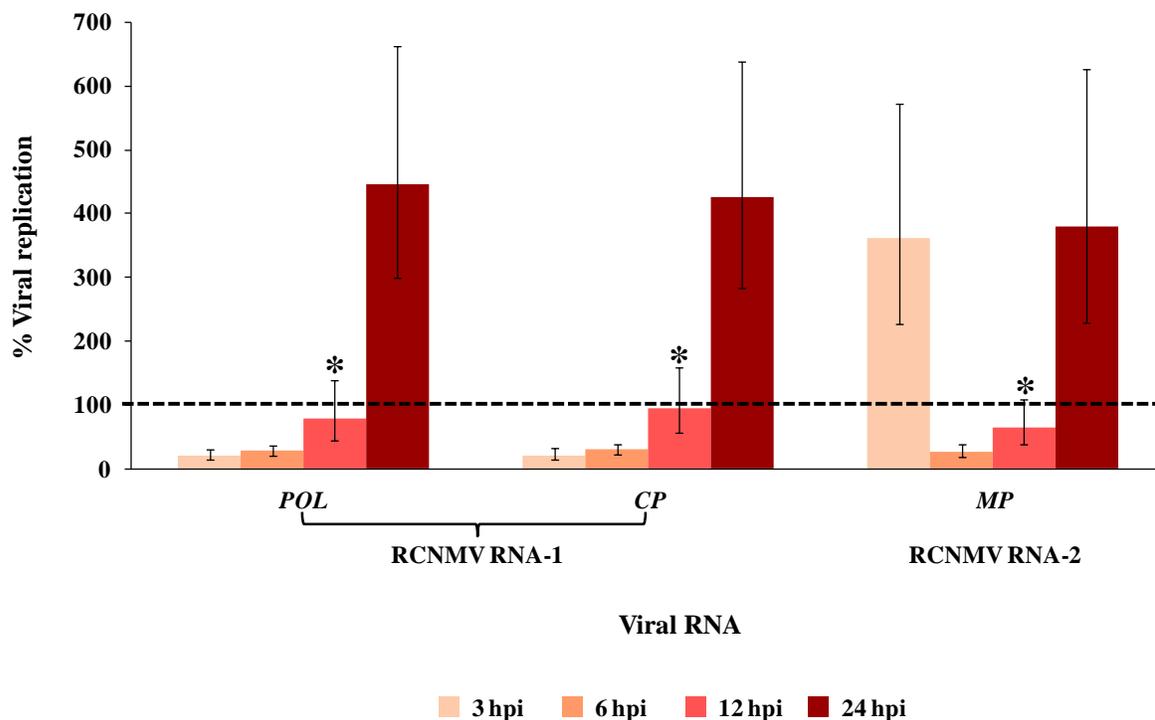


Figure 14 Pre-silenced NbWRKY affected RCNMV RNA accumulation at 3, 6 and 24 hours-post inoculation (hpi), but not at 12 hpi. The expression of NbWRKY was transiently knocked down by TRV1:TRV2-NbWRKY (TRV-VIGS system) 10 days prior to RCNMV inoculation. The pre-silenced NbWRKY plants were inoculated with RCNMV RNA-1/RNA-2, and then harvested between 3 and 24 hours after RCNMV inoculation. RCNMV RNA levels were measured by using qRT-PCR analysis of three regions of the genome: 1) within the p88 coding region of RNA-1 (POL), 2) within the CP coding region of RNA-1 (CP) and 3) within the MP coding region of RNA-2 (MP). The viral RNA accumulation was computed by a relative quantification method and displayed as % viral replication. Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. All data shown in this figure was significant at $\alpha = 0.1$, except the data with an asterisk symbol (*).

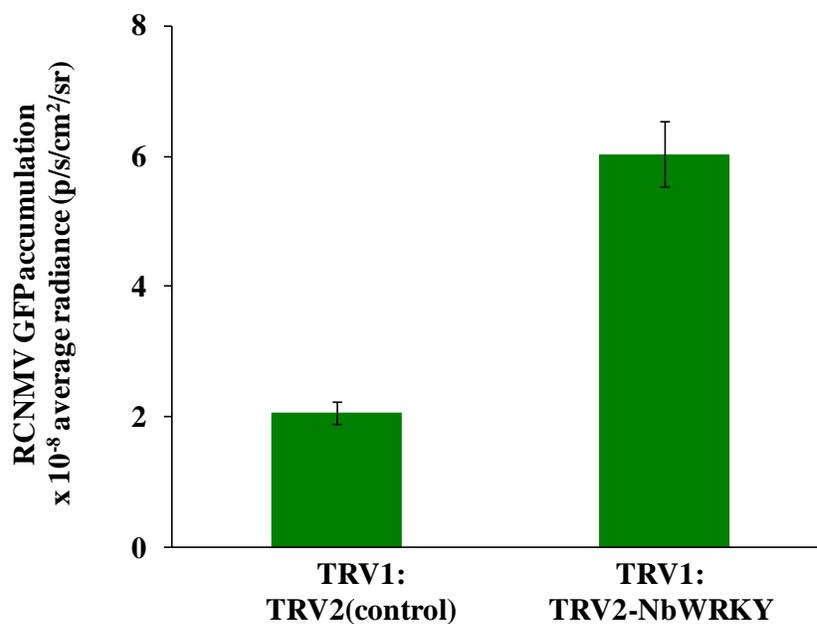


Figure 15 Pre-silenced NbWRKY promoted RCNMV GFP accumulation at 3 days post inoculation. The expression of NbWRKY was transiently knocked down using the TRV-VIGS system 10 days prior to RCNMV inoculation. The pre-silenced NbWRKY plants were inoculated with RCNMV R1SG1/RNA-2, and then harvested at 3 days after RCNMV inoculation. Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis ($\alpha = 0.05$).

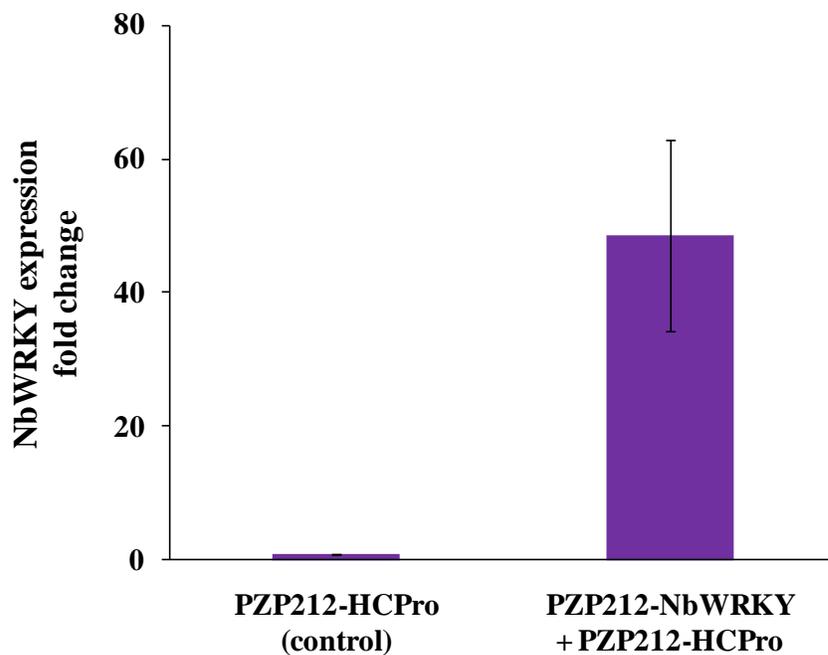


Figure 16 NbWRKY gene was overexpressed in *N. benthamiana* by using PZP212 and pRTL2 vector system. The overexpression construct PZP212-NbWRKY was delivered into plant cells along with the RNA silencing suppressor construct PZP212-HCPro by an agroinfiltration method to constitutively express NbWRKY. The leaf samples were harvested 2 days after agroinfiltration. Plants agroinfiltrated with only PZP212-HCPro were used as the baseline controls in a student's t-test statistical analysis. The qRT-PCR analysis showed that the expression of NbWRKY gene was significantly increased greater than 40 fold ($\alpha = 0.05$).

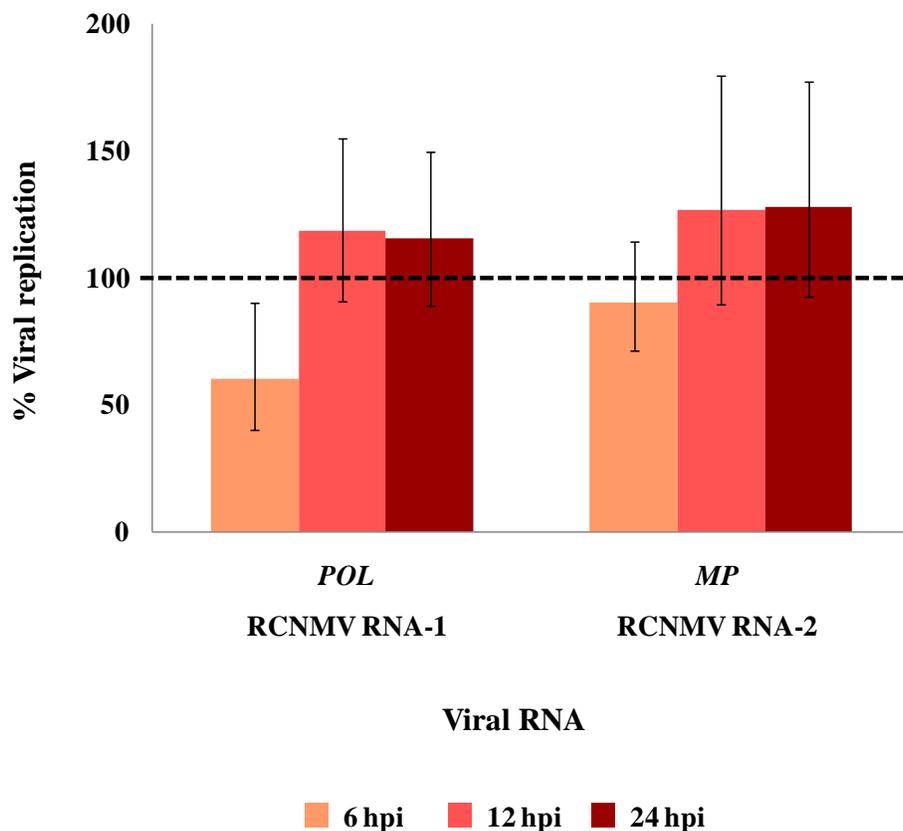


Figure 17 Pre-overexpressed NbWRKY did not affect RCNMV RNA accumulation at 6, 12 and 24 hpi. The expression of NbWRKY was transiently overexpressed by agroinfiltrated PZP212-NbWRKY along with the RNA silencing suppressor construct PZP212-HCPro 2 days prior to RCNMV inoculation. The pre-overexpressed NbWRKY plants were inoculated with RCNMV RNA-1/RNA-2, and then harvested between 6 and 24 hours after RCNMV inoculation. RCNMV RNAs were measured by using qRT-PCR analysis. The viral RNA accumulation was computed by a relative quantification method, and displayed as % viral replication. Plants agroinfiltrated with only PZP212-HCPro 2 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. All data shown in this figure was not significant at $\alpha = 0.1$.



Figure 18 *N. benthamiana* plant size selection for a gene over-expression study (A) and a gene silencing study (B). The plant size is compared to a US ten-cent coin (0.7 inch diameter).

Table 4 Oligonucleotide information used in Chapter 5.

Name	Sequence information
Nb04875F	CGTTGCCGGAAAAGTCATCA
Nb04875R	CGGAAAGTGTTTCGGAAAATGA
GS04875F	AATTTCTAGAGCTCAATCGACGGGTTAATTCGTG
GS04875R	AATTTCTAGAGCTCGCCTTGGATAAGTGGAATGG
FL04875F	AATTCCATGGGATCGATGGAGTCTCCGTTGCCGGAAAAGTC
FL04875R	AATTTCTAGATTACATCGATGCAGAATTGTACCCTTCAAATTCC
3'sGFP/XbaI	GCTCTAGACGCGTTACTTGTACAGCTCGTCC
RNA1-POLF	TCACAAGGGTCAAATTCTCAAATCCT
RNA1-POLR	TGCTGCTTTTTTGGTATAACTTCCTCTT
RNA1-CPF	ATGTCTTCAAAGCTCCCAA
RNA1-CPR	CGCTCATGACTAACTGGGTA
RNA2F	CGCGTCTGATTGAGTTGGAAGTA
RNA2R	CTGCCTTGATGCTCGACAGTA
TRV2F	TACTCAAGGAAGCACGATGAGC
TRV2R	GAACCGTAGTTTAATGTCTTCGGG
TEVLeader80	CGAATCTCAAGCAATCAAGCATTC
35sTERM	ATAAGAACCCTAATTCCC
ActF	GTGACCTCACTGATAGTTTGA
ActR	TACAGAAGAGCTGGTCTTTG

Materials and Methods

1) Plants

N. benthamiana seeds were sowed in a tray and the two-week old seedlings were transplanted to individual pots. Seedlings and transplants were maintained in a temperature and light controlled environment at 26°C, 16 hour-light and 8 hour-dark period at NCSU greenhouses (Method Road).

2) Plasmid DNA constructs

RC169, R1SG1 and RC2 are plasmid DNA constructs containing the infectious cDNAs of RCNMV RNA-1, RCNMV RNA-1 with the sGFP in place of the CP and RCNMV RNA-2, respectively. All three plasmids were transformed into competent *Escherichia coli* DH5 α in order to increase their copy number. The bacterial transformation procedure is as follows: 10 μ l of competent bacterial cells were gently mixed with 1 μ l plasmid DNA (at least 100 ng) in 0.5 ml microcentrifuge tubes, heat shocked at 42°C for 1 min and immediately snap cooled on ice. 100 μ l of LB broth containing 10 mM glucose was added to each tube and the transformants were plated onto LB agar containing ampicillin (100 μ g/ml) and incubated at 37°C for 16-18 hours. Isolated colonies were picked from the LB agar plates and inoculated into 3 ml LB broth containing ampicillin (75 μ g/ml) in a 14 ml culture tube (FalconTM). The culture tubes were placed at an angled position in order to increase the surface area for oxygen exposure. The bacterial cultures were incubated at 37°C with shaking at 270 rpm for 16-18 hours. Plasmid DNA was isolated from the transformants using QIAprep Spin Miniprep Kit (QiagenTM).

The silencing construct TRV2-NbWRKY utilizing the TRV VIGS system [61] was designed to down regulate NbWRKY expression in *N. benthamiana*. A 365 bp fragment of the NbWRKY gene was amplified from cDNA using OneTaq DNA polymerase [New England Biolabs (NEBTM), Ipswich, MA] according to the manufacturer's protocol to generate construct GS-NbWRKY. First strand cDNA was synthesized using random primers (PromegaTM) and total RNA extracted from RCNMV-infected plant tissue. The PCR primers used were GS04875F and GS04875R (Table 4). The PCR conditions for amplifying GS-NbWRKY were as follows: 1) an initial denaturation cycle of 94°C for 30 sec, 2) 30 amplification cycles of 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min, and 3) a final extension cycle of 68°C for 5 min. The PCR product was electrophoresed on a 1% agarose gel and isolated using the QIAquick Gel Extraction Kit (QiagenTM). The gel purified amplification product was then cloned into the pGEM T-Easy vector (PromegaTM) to produce construct pGEM T-Easy-GS-NbWRKY. The antibiotic used for pGEM T-Easy vector was ampicillin. The resultant construct was submitted for sequencing to confirm its completeness. The pGEM T-Easy-GS-NbWRKY was digested with restriction enzyme *SacI* (NEB) to release the insert which was gel electrophoresed and purified. This isolated product was then cloned into the TRV2 vector that was previously linearized with *SacI* and treated with Antarctic Phosphatase (NEBTM). This treatment is required to remove -PO₄³⁻ from the 5' end of *SacI*-TRV2 to prevent vector self-ligation (preventing the formation of 3'-5' phosphodiester bonds) and so decreasing vector background. The ligation between GS-NbWRKY fragment and TRV2 was performed according to the T4 DNA ligase protocol (NEBTM). The ligation mixture was transformed into competent *E. coli* DH5 α and plated onto

LB agar plates containing kanamycin. A colony PCR technique using the primer set TRV2F (forward primer) and TRV2R (reverse primer) (Table 4) was used to identify TRV2 clones that contained the GS-NbWRKY insert. A positive plasmid was also confirmed by digesting with *SacI*. The resultant construct TRV2-NbWRKY was used for down regulating NbWRKY expression in *N. benthamiana* for a gene silencing study.

The PZP212-NbWRKY construct was designed for overexpression of NbWRKY in *N. benthamiana* and is based on the pRTL2 expression cassette and the PZP212 binary vector. A full-length NbWRKY DNA (975 bp) was obtained via PCR using the same starting source material and conditions as mentioned above for the silencing construct with the following modifications: the PCR primers used were FL04875F and FL04875R (Table 4) and the extension cycle at 68°C was performed for 1 min 30 sec rather than the 1 min used previously. The PCR product was gel electrophoresed, purified and cloned into the pGEM T-Easy vector as mentioned previously. The resultant construct pGEM-T-Easy-NbWRKY was verified by sequencing prior to digestion with *NcoI* and *XbaI* restriction enzymes (NEB™). The digested product was gel electrophoresed, isolated, and subsequently cloned into the pRTL2 vector which had been linearized with *NcoI* and *XbaI*. A colony PCR was used to identify the pRTL2 clone that contained a full-length NbWRKY with the primer set of TEVleader80 and 35sTERM (Table). The pRTL2 construct containing the full-length NbWRKY was then verified by sequencing using sequencing primers TEVleader80 and 35sTERM (Table). The resultant plasmid pRTL2-NbWRKY was digested with *HindIII* (NEB™) to release the insert containing the full-length NbWRKY expression cassette. This isolated fragment was subsequently cloned into PZP212 linearized with *HindIII* and treated

with Antarctic Phosphatase (NEBTM). PZP212 is kanamycin resistant. The PZP212-NbWRKY construct was used to constitutively express the putative full-length NbWRKY protein in *N. benthamiana* for a gene over-expression study.

The PZP212-NbWRKY-sGFP construct was designed for subcellular localization study of NbWRKY. A full-length NbWRKY DNA was initially cloned into the RNA transcript expression vector pHST2-sGFP through the unique *ClaI* restriction site. The orientation of the clone was confirmed by a restriction analysis and the properly orientated construct was used as a template to amplify the NbWRKY-sGFP product using FL04875F and 3'sGFP/*XbaI* (Table 4). The PCR product was then cleaved with *NcoI* and *XbaI* and ligated into similarly digested pRTL2 to produce construct pRTL2-NbWRKY-sGFP. The nature of the construct was verified by sequencing and the expression cassette containing NbWRKY-sGFP was released by digestion with *HindIII* and subsequently cloned into PZP212 to yield construct PZP212-NbWRKY-sGFP.

3) *In vitro* T7 RNA transcription

Prior to transcription, plasmid DNA was linearized with *SmaI* (NEBTM) in a 100 μ l reaction containing 10 μ g plasmid DNA, 10 μ l 10X NEB Buffer 4, 1 μ l *SmaI* (20 units/ μ l), and dH₂O up to 100 μ l. 0.5 μ l was taken from this reaction and mixed with 4.5 μ l dH₂O to use as a control prior to adding *SmaI*. The restriction reaction and control were incubated at room temperature overnight. The completeness of digestion was verified by gel electrophoresis on a 1% agarose gel with TAE buffer at 115 V for 45-50 minutes. A gel loading sample for the restriction reaction was set at 6 μ l which was a mixture of 0.5 μ l

restriction reaction, 1 μl 6X gel loading dye (0.25% bromophenol blue v/v, 0.25% xylene cyanol FF, 30% glycerol) and 4.5 μl dH_2O . A gel loading sample for the control was also set at 6 μl which was a mixture of 5 μl control and 1 μl 6X gel loading dye. The gel was stained with in 1X TAE buffer containing ethidium bromide (2 $\mu\text{g}/\text{ml}$) for 1 minute and then rinsed with dH_2O for 5-10 minutes. The gel was visualized and photographed under a UV transilluminator. The linearized plasmid DNA was phenol/chloroform extracted to remove the restriction enzyme and buffer and then precipitated with 0.3 M sodium acetate/2.5 volumes 95% ethanol at -20°C . The linearized plasmid DNA was pelleted and resuspended in dH_2O . The concentration and purity of this cleaned, linearized plasmid DNA was assessed by using a NanoDrop spectrophotometer.

In vitro T7 RNA transcription was prepared according to MEGAscript® T7 Transcription Kit (Ambion™). The reaction volume was set at 20 μl total in a 0.5 ml microcentrifuge tube. The reaction components were assembled at room temperature to prevent template precipitation. The reagents were combined in the following order: nuclease-free dH_2O up to 20 μl , 2 μl ATP, 2 μl UTP, 2 μl CTP, 2 μl GTP, 1 μg linearized plasmid DNA template, 2 μl 10X reaction buffer and 2 μl T7 RNA polymerase. The assembled reaction was gently mixed and incubated at 37°C for 2 hours.

The quality of the RNA transcripts obtained from the *in vitro* T7 transcription reaction was assessed by gel electrophoresis on a 1% agarose RNA gel in TAE buffer. The gel loading sample was set at 6 μl which was a mixture of 0.2 μl *in vitro* T7 transcription reaction, 3 μl 2X RNA loading dye (NEB), and 2.8 μl dH_2O . The premixed gel loading

sample was heat shocked at 55°C for 1 min then snap cooled on ice. The 2X RNA loading dye and heating/snap cooling helps prevent formation of RNA secondary structure. The RNA gel was pre-run (prior to sample loading) at 90V for 15 mins. This pre-run method helps remove impurities from the RNA gel. The premixed gel loading sample was then loaded and the RNA gel was run at 90V for 50 min. The gel was stained, destained and photographed as above.

4) RCNMV inoculum preparation and inoculation procedure

RCNMV inoculum was prepared in a 110 µl volume per 1 plant. This 110 µl inoculum was a mixture of 1 µl *in vitro* T7 RNA-1 transcript (or 1 µl *in vitro* T7 R1SG1 transcript), 1 µl *in vitro* T7 RNA-2 transcript and 108 µl inoculation buffer (10 mM sodium diphosphate, pH 7.2). The RCNMV transcripts were mechanically inoculated onto leaves (either healthy or previously agroinfiltrated; see below) by the following method. Firstly, leaves were lightly dusted with carborundum (abrasive) and 27 µl of the RCNMV transcript mixture was pipetted onto each of four leaves per plant. Secondly, the inoculum was mechanically and gently rubbed onto the adaxial side of the leaf. The inoculation was performed using a middle finger for rubbing and the other hand to support the leaf. The inoculated leaves were then rinsed with water to remove excess inoculum. Inoculated plants were maintained in a temperature and light controlled environment at 18-22°C, 16 hour-light and 8 hour-dark period at the NCSU greenhouses (Method Road).

5) *Agrobacterium* preparation and infiltration

Agrobacterium based constructs (the silencing constructs TRV1, TRV2-NbWRKY, the overexpression construct PZP212-NbWRKY and the subcellular localization construct PZP212-NbWRKY-sGFP along with the RNA silencing suppressor construct PZP212-HCPro) were transformed into *Agrobacterium tumefaciens* strain C58C1 using a BioRad electroporator. This strain is resistant to rifampicin and gentamicin. Transformed cells were plated onto LB agar plates containing rifampicin, gentamicin and a vector specific antibiotic (kanamycin for both TRV vectors and PZP212-based vectors) and then incubated at 28°C for 48 hours. Due to rifampicin light sensitivity, the plates were wrapped with foil during the incubation period. Several clones were picked per plate and their identities confirmed by colony PCR as described above. Suitable colonies were cultured in 2 ml LB broth containing all three antibiotics and incubated at 28°C for 20 hours with shaking at 270 rpm. A 250 µl inoculum from this initial culture was then used to inoculate 5 ml LB broth containing 40 µM acetosyringone, 10 mM MES, pH 5.6, and all three antibiotics and then incubated at 28°C for 20 hours with shaking at 270 rpm. This overnight culture was pelleted, resuspended in agroinfiltration buffer (10 mM MgCl₂, 10mM MES, pH 5.6, 200 µM acetosyringone) to an OD₆₀₀ reading ~ 1 and then incubated at room temperature for at least 3 hours prior to infiltration into plants.

For NbWRKY silencing experiments, *Agrobacterium* cultures of TRV1 and TRV2-NbWRKY (or TRV2 for the control) were mixed at a 1:1 ratio immediately prior to infiltration into plants. For NbWRKY overexpression experiments, *Agrobacterium* cultures of PZP212-NbWRKY and the RNA silencing suppressor PZP212-HCPro (or PZP212-HCPro

alone for the control) were mixed at a ratio of 1:1 immediately prior to infiltration into plants. For subcellular localization experiments, *Agrobacterium* cultures of PZP212-NbWRKY-sGFP (or PZP212-sGFP for the control) and the RNA silencing suppressor PZP212-HCPro were mixed at a ratio of 1:1 immediately prior to infiltration into plants. Agroinfiltration was performed on either 2-3 week old (silencing experiments; Panel B in Figure 18) or 3-4 week old (overexpression and subcellular localization experiments; Panel A in Figure 18) *N. benthamiana* plants with two leaves per plant being infiltrated by using a 1 ml needleless syringe on the abxial side of the leaf. Plants agroinfiltrated with a silencing construct were maintained in a temperature and light controlled environment at 26°C, 16 hour-light and 8 hour-dark period at NCSU greenhouses (Method Road). Plants agroinfiltrated with an overexpression construct and a subcellular localization construct were maintained in a temperature and light controlled environment at 18-22°C, 16 hour-light and 8 hour-dark period. Leaves infiltrated for the subcellular localization experiments were harvested 2 days after agroinfiltration and imaged via confocal microscopy (Cellular and Molecular Imaging Facility, NCSU).

6) Total RNA extraction and cleanup

Leaves to be sampled in each experiment consisted of the following: four plants and two leaves per plant were used to represent the test condition (silencing or overexpression of NbWRKY) as well as the respective control (see above). Total RNA extracts isolated from the same plant were pooled to represent one biological replicate (4 biological replicates for the test condition and 4 biological replicates for controls). Leaf samples were collected either

2 days (overexpression experiments) or 10 days (silencing experiments) after agroinfiltration as well as various time points after RCNMV transcript inoculation for both types of experiments.

Total RNA was extracted from 100 mg *N. benthamiana* leaf tissue samples according to the TRIzol protocol (Invitrogen™), with a few modifications in the sample homogenization step. Plant tissue was kept in 1.5 ml microcentrifuge tubes containing 3 mm-glass beads at -80°C. Samples were transferred to a liquid nitrogen bath and snap-frozen for at least 3 minutes. Tissue was homogenized by vigorously shaking the tubes for 20 seconds using a Silamat S5 mixer (Ivoclar Vivadent). 1 ml TRIzol reagent was immediately added to this ground tissue and then vigorously mixed using a vortex. The reaction was incubated at room temperature for 5 minutes in order to facilitate complete cell lysis and then centrifuged at 12,000 rpm for 10 minutes in order to isolate and precipitate foreign contaminant (e.g. dirt, insoluble material). If samples were pooled in the experiment, 700 µl of the supernatant per sample was combined to the pool. Then 1 ml was drawn from the pool and further processed according to the TRIzol™ protocol (Invitrogen™). The final total RNA pellet was resuspended in 50 µl dH₂O. The quantity and quality of the total RNA extract was evaluated by using the NanoDrop spectrophotometer and RNA gel electrophoresis, respectively. All total RNA extracts were stored at -20°C.

Total RNA extracts destined for qRT-PCR analysis were treated with Turbo DNA-free (Ambion™) according to the manufacturer's protocol to remove any residual input DNA. The 50 µl total reaction volume was a mixture of 10 µg total RNA, 5 µl 10X TURBO

DNase buffer, nuclease-free water up to 50 μ l and 1 μ l Turbo DNase I (2 units/ μ l) placed in a 0.5 ml microcentrifuge tube. The reaction was assembled at room temperature and incubated at 37°C for 30 minutes. 5 μ l DNase inactivation reagent was then added to the reaction, gently mixed and incubated at room temperature for 5 minutes with occasional flicking of the tube to redisperse the DNase inactivation reagent. The sample was centrifuged at 13,000 rpm for 1 minute 30 seconds and the supernatant transferred to a new tube and stored at -20°C.

7) RNA quantification by real-time PCR (qRT-PCR)

First strand cDNA synthesis was performed in a 50 μ l total reaction volume consisting of 5 μ l DNA-free total RNA (equivalent to ~1 μ g total RNA), 14.25 μ l dH₂O, 5 μ l 10X GeneAmp® PCR Buffer II (Applied Biosystems™), 11 μ l 25 mM MgCl₂, 10 μ l dNTPs (2.5 mM each dNTP), 2.5 μ l 50 μ M random primers (Promega™), 1 μ l RNase inhibitor (20 units/ μ l, Applied Biosystems™) and 1.25 μ l MultiScribe reverse transcriptase (50 units/ μ l, Applied Biosystems™). The reaction was assembled in 0.2 ml PCR tubes at room temperature. A thermal cycler was used to incubate the cDNA synthesis reaction as follows: 25°C for 10 minutes, 48°C for 30 minutes and 95°C for 5 minutes. The resultant first strand cDNA was stored at -20°C.

SYBR Green-based quantitative real-time PCR (qRT-PCR) method was used in this study. The reaction volume was set at 10 μ l which consisted of 2 μ l first strand cDNA (equivalent to 40 ng total RNA), 4.4 μ l dH₂O, 0.3 μ l forward primer (10 μ M), 0.3 μ l reverse primer (10 μ M) and 5 μ l 2X FastStart Universal SYBR Green Master ROX (Roche™). The Taq polymerase of this SYBR Green master mix contains a hot start property which means

the PCR reaction will not start until the reaction is heated up to 95°C for 10 minutes. The reactions were setup in a 384 well plate (Applied Biosystems™) at room temperature. The plate was centrifuged at 3000 rpm for 2 minutes prior to placement in the ABI7900 HT Fast real-time PCR system (Applied Biosystems™). All PCR products were less than 200 base pairs. Annealing and extension steps were designed to occur in the same temperature range and so these steps were combined. The cycling condition was set as follows: 1 cycle at 95°C for 10 min (Taq polymerase activation) and 40 cycles at 95°C for 15 sec (denaturation)/55°C for 1 min (annealing and extension). An association analysis was also performed in order to test if there were any non-specific PCR products formed. The condition for the association analysis was set as follows: 1 cycle at 95°C for 15 sec, 60°C for 15 sec and 95°C for 15 sec.

NbWRKY transcript levels were assayed using primers Nb04875F and Nb04875R (Table 4). RCNMV RNA levels were assayed with 2 primer pairs for RNA-1 and 1 primer pair for RNA-2 (Table 4): 1) RNA1-POLF and RNA1-POLR for probing RNA-1 at the polymerase (RNA-1 POL), 2) RNA1-CPF and RNA1-CPR for probing RNA-1 at the coat protein (RNA-1 CP) and 3) RNA2F and RNA2R for probing RNA-2 at the movement protein (RNA-2 MP).

8) qRT-PCR data analysis

A relative quantification, $2^{-\Delta\Delta Ct}$ method (equation 1) [62] was used to analyze qRT-PCR data. Expression of the target gene was normalized against the expression of a reference gene. This method was aimed to correct “sample to sample variations” that occurred due to the differences in tissue weight used for extracting total RNA, total RNA extraction

procedure, residual input DNA removal, cDNA synthesis and qRT-PCR reaction preparation. This enabled comparisons of RNA concentrations across different samples. Applied Biosystems SDS software version 2.4 was used to dissect real-time PCR data. Actin (primer set of ActF and ActR (Table 4)) was used as the reference gene to normalize the target gene that was performed within the same plate. Student's t-test statistical analysis was used to examine the real-time PCR data.

Equation 1 Relative quantification (RQ) using $2^{-\Delta\Delta C_t}$ method

$$RQ = 2^{-\Delta\Delta C_t}$$

$$-\Delta\Delta C_t = -(\Delta C_{t(\text{exp})} - \Delta C_{t(\text{control})})$$

$$-\Delta\Delta C_t = -\{[C_{t(\text{target gene})} - C_{t(\text{reference gene})}]_{(\text{exp})} - [C_{t(\text{target gene})} - C_{t(\text{reference gene})}]_{(\text{control})}\}$$

C_t = cycle threshold

exp = experiment

9) RCNMV sGFP quantification

N. benthamiana plants were inoculated with a combination of *in vitro* T7 R1SG1 and RNA-2 transcripts. Ten plants and two leaves per plant were used for each test condition and each control. Leaf samples were harvested 3 days post inoculation. The GFP fluorescence intensity was analyzed by the IVIS imaging system and subjected to student's t-test statistical analysis.

The IVIS Lumina System (Xenogen Corporation, Alameda, CA) is capable of quantifying photon emission from a variety of sources. A CCD camera measured and recorded photon emission data which was then incorporated into Living Image Software (Xenogen Corp.) for further analysis. Whole leaves were placed under the CCD camera and measurements were taken with a GFP excitation filter, and exposure time of 1 second.

References

1. Eulgem T, Somssich IE: **Networks of WRKY transcription factors in defense signaling.** *Current opinion in plant biology* 2007, **10**(4):366-371.
2. Pandey SP, Somssich IE: **The role of WRKY transcription factors in plant immunity.** *Plant Physiology* 2009, **150**(4):1648-1655.
3. Rushton PJ, Torres JT, Parniske M, Wernert P, Hahlbrock K, Somssich I: **Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes.** *The EMBO journal* 1996, **15**(20):5690.
4. Ulker B, Somssich IE: **WRKY transcription factors: from DNA binding towards biological function.** *Current opinion in plant biology* 2004, **7**(5):491-498.
5. Eulgem T: **Regulation of the Arabidopsis defense transcriptome.** *Trends in plant science* 2005, **10**(2):71-78.
6. Kalde M, Barth M, Somssich IE, Lippok B: **Members of the Arabidopsis WRKY group III transcription factors are part of different plant defense signaling pathways.** *Molecular plant-microbe interactions* 2003, **16**(4):295-305.
7. Li J, Brader Gn, Palva ET: **The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense.** *The Plant Cell Online* 2004, **16**(2):319-331.
8. Dong J, Chen C, Chen Z: **Expression profiles of the Arabidopsis WRKY gene superfamily during plant defense response.** *Plant molecular biology* 2003, **51**(1):21-37.

9. Rushton PJ, Somssich IE, Ringler P, Shen QJ: **WRKY transcription factors**. *Trends in plant science*, **15**(5):247-258.
10. Liu X, Bai X, Wang X, Chu C: **OsWRKY71, a rice transcription factor, is involved in rice defense response**. *Journal of plant physiology* 2007, **164**(8):969-979.
11. Huang T, Duman JG: **Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamara***. *Plant molecular biology* 2002, **48**(4):339-350.
12. Rizhsky L, Liang H, Mittler R: **The combined effect of drought stress and heat shock on gene expression in tobacco**. *Plant Physiology* 2002, **130**(3):1143-1151.
13. Robatzek S, Somssich IE: **A new member of the Arabidopsis WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence- related processes**. *The Plant Journal* 2001, **28**(2):123-133.
14. Sanchez-Ballesta MaT, Lluch Y, Gosalbes MaJ, Zacarias L, Granell A, Lafuente MaT: **A survey of genes differentially expressed during long-term heat-induced chilling tolerance in citrus fruit**. *Planta* 2003, **218**(1):65-70.
15. Rushton DL, Tripathi P, Rabara RC, Lin J, Ringler P, Boken AK, Langum TJ, Smidt L, Boomsma DD, Emme NJ: **WRKY transcription factors: key components in abscisic acid signalling**. *Plant biotechnology journal*, **10**(1):2-11.
16. Kato N, Dubouzet E, Kokabu Y, Yoshida S, Taniguchi Y, Dubouzet JG, Yazaki K, Sato F: **Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica***. *Plant and cell physiology* 2007, **48**(1):8-18.
17. Yoshida S: **Molecular regulation of leaf senescence**. *Current opinion in plant biology* 2003, **6**(1):79-84.
18. Robatzek S, Somssich IE: **Targets of AtWRKY6 regulation during plant senescence and pathogen defense**. *Genes & Development* 2002, **16**(9):1139-1149.
19. Johnson CS, Kolevski B, Smyth DR: **TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor**. *The Plant Cell Online* 2002, **14**(6):1359-1375.
20. Eulgem T, Rushton PJ, Robatzek S, Somssich IE: **The WRKY superfamily of plant transcription factors**. *Trends in plant science* 2000, **5**(5):199-206.

21. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T, Tomo Y: **Solution structure of an Arabidopsis WRKY DNA binding domain.** *The Plant Cell Online* 2005, **17**(3):944-956.
22. Yamasaki K, Kigawa T, Watanabe S, Inoue M, Yamasaki T, Seki M, Shinozaki K, Yokoyama S: **Structural basis for sequence-specific DNA recognition by an Arabidopsis WRKY transcription factor.** *Journal of biological chemistry* 2012, **287**(10):7683-7691.
23. de Pater S, Greco V, Pham K, Memelink J, Kijne J: **Characterization of a zinc-dependent transcriptional activator from Arabidopsis.** *Nucleic acids research* 1996, **24**(23):4624-4631.
24. Eulgem T, Rushton PJ, Schmelzer E, Hahlbrock K, Somssich IE: **Early nuclear events in plant defence signalling: rapid gene activation by WRKY transcription factors.** *The EMBO journal* 1999, **18**(17):4689-4699.
25. Fukuda Y, Shinshi H: **Characterization of a novel cis-acting element that is responsive to a fungal elicitor in the promoter of a tobacco class I chitinase gene.** *Plant molecular biology* 1994, **24**(3):485-493.
26. Rushton PJ, Somssich IE: **Transcriptional control of plant genes responsive to pathogens.** *Current opinion in plant biology* 1998, **1**(4):311-315.
27. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2008, **36**(suppl 1):D25-D30.
28. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
29. Jin J, Zhang H, Kong L, Gao G, Luo J: **PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors.** *Nucleic acids research*:gkt1016.
30. Kalderon D, Roberts BL, Richardson WD, Smith AE: **A short amino acid sequence able to specify nuclear location.** *Cell* 1984, **39**(3):499-509.
31. Chelsky D, Ralph R, Jonak G: **Sequence requirements for synthetic peptide-mediated translocation to the nucleus.** *Molecular and cellular biology* 1989, **9**(6):2487-2492.
32. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L: **SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information.** *Nucleic acids research*:gku340.

33. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExpPASy: the proteomics server for in-depth protein knowledge and analysis.** *Nucleic acids research* 2003, **31**(13):3784-3788.
34. Duan M-R, Nan J, Liang Y-H, Mao P, Lu L, Li L, Wei C, Lai L, Li Y, Su X-D: **DNA binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana WRKY1 protein.** *Nucleic acids research* 2007, **35**(4):1145-1154.
35. Nair R, Rost B: **Sequence conserved for subcellular localization.** *Protein Science* 2002, **11**(12):2836-2847.
36. Yu CS, Chen YC, Lu CH, Hwang JK: **Prediction of protein subcellular localization.** *Proteins: Structure, Function, and Bioinformatics* 2006, **64**(3):643-651.
37. Nakai K, Horton P: **PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization.** *Trends in biochemical sciences* 1999, **24**(1):34-35.
38. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier C, Nakai K: **WoLF PSORT: protein localization predictor.** *Nucleic acids research* 2007, **35**(suppl 2):W585-W587.
39. Chou K-C, Shen H-B: **Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization.** *PloS one* 2010, **5**(6):e11335.
40. Makkerh JP, Dingwall C, Laskey RA: **Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids.** *Current Biology* 1996, **6**(8):1025-1027.
41. Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y: **The SOL Genomics Network. A comparative resource for Solanaceae biology and beyond.** *Plant Physiology* 2005, **138**(3):1310-1317.
42. Smale ST, Kadonaga JT: **The RNA polymerase II core promoter.** *Annual review of biochemistry* 2003, **72**(1):449-479.
43. Higo K, Ugawa Y, Iwamoto M, Korenaga T: **Plant cis-acting regulatory DNA elements (PLACE) database: 1999.** *Nucleic acids research* 1999, **27**(1):297-300.
44. Turck F, Zhou A, Somssich IE: **Stimulus-dependent, promoter-specific binding of transcription factor WRKY1 to its native promoter and the defense-related gene PcPR1-1 in parsley.** *The Plant Cell Online* 2004, **16**(10):2573-2585.
45. Cormack RS, Eulgem T, Rushton PJ, Kochner P, Hahlbrock K, Somssich IE: **Leucine zipper-containing WRKY proteins widen the spectrum of immediate**

- early elicitor-induced WRKY transcription factors in parsley.** *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 2002, **1576**(1):92-100.
46. Tsien RY: **The green fluorescent protein.** *Annual review of biochemistry* 1998, **67**(1):509-544.
47. Palmer E, Freeman T: **Investigation into the use of C- and N- terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays.** *Comparative and functional genomics* 2004, **5**(4):342-353.
48. Chiu W-l, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J: **Engineered GFP as a vital reporter in plants.** *Current Biology* 1996, **6**(3):325-330.
49. Seibel NM, Eljouni J, Nalaskowski MM, Hampe W: **Nuclear localization of enhanced green fluorescent protein homomultimers.** *Analytical biochemistry* 2007, **368**(1):95-99.
50. Wang R, Brattain MG: **The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa.** *FEBS letters* 2007, **581**(17):3164-3170.
51. Ribbeck K, Gorlich D: **Kinetic analysis of translocation through nuclear pore complexes.** *The EMBO journal* 2001, **20**(6):1320-1330.
52. Paine PL, Moore LC, Horowitz SB: **Nuclear envelope permeability.** *Nature* 1975, **254**(5496):109-114.
53. Kapuscinski J: **DAPI: a DNA-specific fluorescent probe.** *Biotechnic & Histochemistry* 1995, **70**(5):220-233.
54. Bachan S, Dinesh-Kumar SP: **Tobacco rattle virus (TRV)-based virus-induced gene silencing.** In: *Antiviral Resistance in Plants*. Springer: 83-92.
55. Eulgem T: **Dissecting the WRKY web of plant defense regulators.** *PLoS pathogens* 2006, **2**(11):e126.
56. Wang D, Amornsiripanitch N, Dong X: **A genomic approach to identify regulatory nodes in the transcriptional network of systemic acquired resistance in plants.** *PLoS pathogens* 2006, **2**(11):e123.
57. Murray SL, Ingle RA, Petersen LN, Denby KJ: **Basal resistance against *Pseudomonas syringae* in *Arabidopsis* involves WRKY53 and a protein with homology to a nematode resistance protein.** *Molecular plant-microbe interactions* 2007, **20**(11):1431-1438.

58. Hu J, Barlet X, Deslandes L, Hirsch J, Feng DX, Somssich I, Marco Y: **Transcriptional responses of *Arabidopsis thaliana* during wilt disease caused by the soil-borne phytopathogenic bacterium, *Ralstonia solanacearum*.** *PloS one* 2008, **3**(7):e2589.
59. Higashi K, Ishiga Y, Inagaki Y, Toyoda K, Shiraishi T, Ichinose Y: **Modulation of defense signal transduction by flagellin-induced WRKY41 transcription factor in *Arabidopsis thaliana*.** *Molecular Genetics and Genomics* 2008, **279**(3):303-312.
60. Xu X, Chen C, Fan B, Chen Z: **Physical and functional interactions between pathogen-induced *Arabidopsis* WRKY18, WRKY40, and WRKY60 transcription factors.** *The Plant Cell Online* 2006, **18**(5):1310-1326.
61. Dinesh-Kumar SP, Anandalakshmi R, Marathe R, Schiff M, Liu Y: **Virus-induced gene silencing.** In: *Plant Functional Genomics*. Springer; 2003: 287-293.
62. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic acids research* 2001, **29**(9):e45-e45.

Chapter 6

Altered regulation of *Nicotiana benthamiana* basal transcription factor 3 (NbBTF3) expression early in a *Red clover necrotic mosaic virus* infection

Chapter summary

Herein, I report a study on the regulation of *Nicotiana benthamiana* basal transcription factor 3 (NbBTF3) gene expression early in the *Red clover necrotic mosaic virus* (RCNMV) infection process. The microarray study and qRT-PCR analysis described in Chapter 3 confirmed that the host NbBTF3 gene expression was increased at 6 hours post RCNMV inoculation. An NbBTF3-GFP fusion protein translocated to the nucleus. NbBTF3 was further functionally assayed for its effects on RCNMV infection by both transient gene silencing and gene overexpression in *N. benthamiana*. These functional assays suggest that NbBTF3 expression was up-regulated by the plant host in an attempt to combat the RCNMV infection.

Abstract

Plant viruses are minimalist pathogens that are required to not only co-opt host processes to produce a successful infection, but also to defeat a plant's defense response. The plant's ability to respond to a viral pathogen early in the infection process is critical to its survival. Many of the physiological changes that occur during this critical early phase of the infection remain unknown. The transcriptome from *Red clover necrotic mosaic virus* (RCNMV) infected *Nicotiana benthamiana* (Nb) was examined very early in the infection process utilizing a custom Nb array. A custom microarray representing an estimated coverage of ~38% of the Nb genome was developed. Host transcriptome profiles were analyzed at 2, 6, 12 and 24 hours post-inoculation (hpi) with RCNMV. The vast majority of host genes, whose expression was altered early in the infection process, were significantly down-regulated at 2, 6 and 24 hpi but up-regulated at 12 hpi at an FDR cutoff of 0.01. However, one gene, NbBTF3 (basal transcription factor 3) exhibited the opposite trend with increased expression at 2 and 6 hpi followed by a dramatic suppression at 12 hpi followed by a gradual increase again at 24 hpi. Furthermore, transiently silencing NbBTF3 expression in *N. benthamiana* increased RCNMV RNA accumulation while overexpression decreased RCNMV RNA accumulation. This suggests that early up-regulation of NbBTF3 expression in *N. benthamiana* is an attempt to suppress RCNMV infection. Localization studies revealed that NbBTF3 is predominantly translocated to the nucleus, suggestive of its function as a transcription regulator of host defense gene expression. Surprisingly, transient silencing was reversed by RCNMV infection after 12 hours indicating that RCNMV may actually require

NbBTF3 for a part of the infection process. Collectively, these observations lead to the scenario where NbBTF3 expression is enhanced early in the infection process as an immediate (albeit, ultimately futile) defense response against RCNMV infection which the virus has exploited to the point where NbBTF3 expression actually becomes a crucial part of the RCNMV infection process.

Keywords; BTF3, basal transcription factor 3, *Nicotiana benthamiana*, RCNMV

Introduction

Plant viruses are intracellular obligate pathogens that require various host factors for multiplication and movement while also needing to defeat plant host defense factors [1]. Many events and interactions during the critical early period of a virus infection where the plant responds to the attack remain unknown. This chapter is aimed at studying the earliest initial defense strategy that plants employ to defend against virus infection. To elucidate these early virus-host interaction events, this study uses the ubiquitous plant virus host *Nicotiana benthamiana* and *Red clover necrotic mosaic virus* (RCNMV, a typical plus strand RNA plant virus). A custom microarray representing approximately 13,000 host genes with an estimated coverage of ~38% of the *N. benthamiana* genome was created and used to monitor the differential gene expression pattern for thousands of genes simultaneously. The manufacturing of the microarray was described in Chapter 2. Host transcription profiles were analyzed early in an RCNMV infection at 2, 6, 12 and 24 hours post inoculation (hpi). The microarray study in Chapter 4 revealed that host apoptosis was a primary and acute host

defense response in *N. benthamiana* against RCNMV early in the infection. Apoptosis is a programmed cell death that results in the visible phenotype known as a hypersensitive reaction (HR). Plant HR-induced cell death is known for its “walling off” effect occurring at the initial site of infection to prevent the virus infection from spreading to and infecting neighboring healthy plant cells. HR cell death is associated with viral resistance genes. Other genes that do not belong to these resistance gene clusters may support HR cell death via protein-protein interactions, co-transcriptional activation, and modulating signaling pathways [2-4]. This study is aimed at identifying and characterizing a transcription factor involved in regulating the transcription of HR-related genes. The main assumption underlying this study is that host transcription factors impact HR-induced cell death pathway activation. Host gene Nb09394 (and its putative product NbBTF3 or *N. benthamiana* basal transcription factor 3) was selected from a microarray study (Chapter 3) for further study in this chapter due to its highly up-regulated levels at 6 hpi, its known involvement in co-regulating other HR-defense genes [5] and its initial characterization as a transcription factor [6].

BTF3 (β -NAC) was first identified in HeLa cell extracts as one of the basal or general transcription factors involved in initiation of transcription from class II promoters [6]. The BTF3 transcription factor is required for RNA polymerase II-dependent transcription. However, BTF3 was later found to also be a β -subunit of nascent-polypeptide-associated complex (NAC) that has been implicated in regulating protein localization during translation. NAC is a heterodimer of α (α -NAC) and β (β -NAC) subunits. NAC binds to ribosome-associated nascent polypeptide chains as they emerge from the ribosomal complex [7]. NAC controls targeting of nascent proteins to their correct subcellular locations by preventing

inappropriate interactions with signal recognition particles (SRP) and thereby inhibiting the targeting of non-secretory proteins to the endoplasmic reticulum [8]. NAC is also associated with other processes including stimulating protein import into yeast mitochondria [9]. NAC is a highly conserved protein complex that is present in many organisms including human, mice, yeast (*Saccharomyces cerevisiae*), fruit fly (*Drosophila melanogaster*) and plants [10]. NAC is an abundant cytosolic protein [10]. Consistent with the general roles of NAC in the ribosome, NAC predominantly localizes to the cytosolic region where it is involved in contacting the ribosome and gaining access to nascent polypeptides in mammalian cells. The diverse functions of the α - and β -subunits of NAC indicate that they may simultaneously exist in, or move between, two or more different subcellular locations such as the nucleus, cytosol and mitochondria [11].

Yeast mutants lacking NAC have only slight defects in growth and morphology [12]. NAC mutations in mice and fruit flies exhibit an embryonic lethal phenotype [13, 14]. Human BTF3 protein was found to be decreased in Burkitt's lymphoma and Jurkat T-cells within 6 hours of inducing apoptosis suggesting that human BTF3 is an apoptosis-associated protein and a potential candidate for controlling apoptosis in B-lymphocytes [11]. In *Caenorhabditis elegans*, the inhibitor of cell death-1 gene (ICD-1) that encodes the β -NAC functions in preventing apoptosis. RNA silencing (or RNA interference; RNAi) of ICD-1 induces inappropriate apoptosis, whereas over expression of ICD-1 inhibits apoptosis [15]. This study also reported that ICD-1 protein is localized to the mitochondria, not to the nucleus or cytosol.

In plants, BTF3 or NAC functions are not yet fully understood in terms of ribosome- or non-ribosome associated functions. *Arabidopsis* BTF3 interacts with the translation initiation factor (iso)4E (eIFiso4E) in a yeast two-hybrid system and might be associated with the regulation of translation initiation [16]. Inhibition of rice Osj10gBTF3 (*Oryza sativa* L. subsp. japonica) resulted in plant miniaturization and pollen abortion. The Osj10gBTF3::EGFP (enhanced green fluorescent protein) fusion protein was localized to both the nucleus and the cytoplasmic membrane system [17]. Silencing of the wheat (*Triticum aestivum*) BTF3, TaBTF3, suggests that it is associated with the development of the wheat chloroplast and mitochondria as well as the structure of mesophyll cells [18]. Silencing TaBTF3 also impairs tolerance to freezing and drought stresses. Subcellular localization studies revealed that TaBTF3 was mainly located in the cytoplasm and nucleus [19].

One study demonstrated that *N. benthamiana* BTF3 (NbBTF3) predominantly localizes to the nucleus but not to the cytosol and silenced NbBTF3 plants exhibit an abnormal developmental phenotype [20]. NbBTF3 silencing also represses the transcript levels of some plastid- and mitochondria-encoded genes. It was observed that BTF3 deficiency in *N. benthamiana* plants preferentially affected the development and/or physiology of chloroplasts and mitochondria [20]. NbBTF3 was up-regulated 2 fold in plants challenged with AC2 protein from the begomovirus *Tomato chlorotic mottle virus* [21]. In pepper (*Capsicum annuum*), CaBTF3 was shown to be involved in HR cell death upon *Tobacco mosaic virus* (TMV) infection and could function as a transcription factor in the nucleus by transcriptional regulation of HR-related gene expression [5].

Collectively these observations lead to the main hypothesis of this study that NbBTF3 is associated with HR-mediated cell death or involved in the biotic stress response as a transcription factor in the nucleus early in the RCNMV infection process. In this study, the NbBTF3 expression pattern was analyzed during an RCNMV infection within 24 hpi. Additionally, the subcellular localization of NbBTF3 and the effect of silencing and overexpression of NbBTF3 on RCNMV infection efficiency were also investigated.

Results and discussion

1) RCNMV infection up-regulated *N. benthamiana* basal transcription factor 3 (NbBTF3) gene expression at 6 hpi

The microarray study of the modulation of *N. benthamiana* gene expression early in an RCNMV infection described in Chapter 3 revealed that host NbBTF3 gene expression gradually increased after 2 hours and reached a peak at 6 hours (Figure 1). NbBTF3 gene expression then dramatically subsided at 12 hpi and started rising again at 24 hpi. The microarray study also indicated that RCNMV takes 12-24 hours to complete its infection within a primary infected cell and the first 12 hours of infection is a critical point when RCNMV begins movement to the adjacent cells. The increased expression of NbBTF3 at 6 hpi was also confirmed by quantitative real-time PCR (qRT-PCR) analysis (Figure 1). A dramatic drop in NbBTF3 expression at 12 hpi is likely concordant with the RCNMV infection migrating to the adjacent cells. The rising expression of NbBTF3 at 24 hpi confirms a repeat expression pattern of the first 6-12 hours expression in secondarily infected cells.

Considering that the vast majority of host genes are down regulated within the first 6 hours of an RCNMV infection (described in Chapter 3), the enhanced expression of NbBTF3 during this critical early period is an unexpected observation. A previous study by Su et al., 2012 [5] demonstrated that CaBTF3 and NbBTF3 are associated with the activation of an HR, which is caused by host programmed cell death. Silencing of NbBTF3, a gene nearly homologous to CaBTF3, led to reduced Bax- and Pto-mediated cell death. Responding to a virus infection with an HR triggered cell death is a standard two pronged defense strategy: (i) inhibition of viral spread from the original or primary infected cells via a “walling off effect” and (ii) signaling to establish a longer lasting defense to protect the entire plant against further systemic infection via systemic acquired resistance (SAR). Plants silenced for expression of CaBTF3 exhibited reduced HR cell death and decreased expression of some HR-related genes upon TMV infection, but increased TMV coat protein levels compared with control plants. This led to my speculation that increased NbBTF3 gene expression observed in the first 6 hours of an RCNMV infection is an integral part of the plant defense strategy against RCNMV.

The central hypothesis of this chapter is that altering the potential interaction between the host NbBTF3 gene and RCNMV during the infection process will affect the efficiency of RCNMV replication. To test this hypothesis, NbBTF3 will be transiently silenced and overexpressed to assay its effects on RCNMV replication by: (i) measurement of viral RNA accumulation via quantitative real-time PCR (qRT-PCR) and (ii) measurement of GFP (by IVIS fluorescence imaging assay) produced by an RCNMV construct that expresses GFP in place of the coat protein as an indicator for viral gene expression levels.

2) NbBTF3 (Nb09394) sequence analysis and functional annotation

The *N. benthamiana* NbBTF3 gene is represented by microarray unigene ID (probe ID) Nb09394. The custom microarray was originally created from a collection of 13K *N. benthamiana* unigenes. The details of the unigene collection and microarray construction are thoroughly described in Chapter 2. A homology search of the Nb09394 sequence against the GenBank non-redundant protein sequence (nr) database using BLASTX [22, 23] revealed that the Nb09394 top five BLAST hits were to basal transcription factor 3 (BTF3) from tobacco, pepper, potato and tomato (Table 1). The Nb09394 nucleotide sequence produces a peptide of 160 residues (ExPASy tool [24]) (Figures 2 and 3) [20]. This complete coding sequence information was later used to design a primer for the full-length NbBTF3 cDNA in sections 5 and 6 for gene over expression and subcellular localization studies, respectively.

Nb09394 encodes a peptide that has a 100% residue identity to *N. benthamiana* BTF3 GenBank ID ABE01085.1, with an E-value = 3×10^{-67} (Figure 2). A homology search against GenBank conserved domain database [25], revealed a NAC in Nb09394. Therefore, Nb09394 is assigned the protein name NbBTF3. NbBTF3 is predicted to be 160 residues in length with a molecular weight of 17.35 kDa (ExPASy [24]).

3) Silencing of the NbBTF3 gene was reversed 12 hours after RCNMV infection

N. benthamiana NbBTF3 gene expression was silenced in *N. benthamiana* using the *Tobacco rattle virus* (TRV) derived viral induced gene silencing (VIGS) vector TRV1:TRV2-NbBTF3 system [26, 27]. The silencing constructs TRV1:TRV2-NbBTF3 was

delivered to plant cells using an agroinfiltration method. Silencing of the *N. benthamiana* phytoene desaturase (NbPDS) gene was used as a positive control for this silencing study [28]. NbPDS is a key enzyme in the carotenoid biosynthesis pathway that converts phytoene to zeta-carotene [29]. Silenced NbPDS causes a visible leaf bleaching 5 days after agroinfiltration (Figure 5). Interestingly, silencing of NbBTF3 also causes a leaf chlorosis (leaf yellowing) approximately 5 days after agroinfiltration and the plants are slightly smaller in size compared to the various controls (wild-type, buffer infiltrated and TRV1:TRV2 agroinfiltrated plants; Figure 5). This suggests that NbBTF3 affects plant photosynthesis and growth. Yang et al., 2007 [20] revealed that NbBTF3 was involved in chloroplast development when silenced NbBTF3 produced a severe abnormality in chloroplast morphology which likely explains the leaf chlorosis phenotype.

The NbBTF3 gene was silenced 10 days prior to RCNMV inoculation. Silencing reduced the expression levels of NbBTF3 greater than 2 fold compared to the TRV1:TRV2 negative control (Figure 4). Expression levels of NbBTF3 were then monitored over the first 3 days of an RCNMV infection. Plants were harvested at 0, 12, 24 and 72 hours after RCNMV inoculation with NbBTF3 expression levels assayed by qRT-PCR (Figure 4). NbBTF3 expression remains suppressed 4 fold until 12 hours after RCNMV inoculation. Beginning at 12 hrs and continuing for 72 hours, NbBTF3 silencing was reversed and expression increased 8 fold.

The reversal of NbBTF3 silencing during an RCNMV infection was unexpected and raises the question of whether silencing reversal is unique to NbBTF3 or is this also the

response for a large suite of host genes? It is possible that one or more suppressors of RNA silencing encoded by RCNMV may be partially or completely responsible for the reversal of NbBTF3 silencing upon virus infection [30, 31]. As is widely known, plant viruses carry an array of host gene silencing suppressors (viral suppressors of RNA silencing or VSRs) as a counter defense to the host's defense system (32) [32]. RNA silencing is also important for host mRNA turnover [33, 34] and regulates host gene expression at the post-transcriptional level. If virus suppressors function to disrupt the RNA silencing function, this leads to an unspecific disturbance in host gene expression. Consequently, I hypothesized that the reversal of NbBTF3 silencing resulted from the action of RCNMV encoded VSRs produced during the infection. To test this hypothesis, the expression of NbWRKY (WRKY transcription factor), NbSAR (systemic acquired resistance), and NbSI (sterol isomerase) were transiently silenced in *N. benthamiana* using the TRV VIGS system. These genes were pre-silenced 10 days prior to RCNMV inoculation. The expression of these silenced genes was assayed between 0-72 hours post inoculation (hpi) by using qRT-PCR analysis. The expression levels of NbWRKY, NbSAR and NbSI all remained suppressed through 72 hpi (Figure 4). The fact that the suppression of these three genes was not reversed by an RCNMV infection suggests that the reversal of silencing is not a universal phenomenon but rather specifically targeted to NbBTF3 for its enhanced expression.

The step(s) in the *N. benthamiana* silencing pathway that are affected by the two characterized RCNMV VSRs [30, 31] remains unknown. There are 11 RNAi-related unigenes on the *N. benthamiana* microarray chip. The functional annotation of these RNAi-related unigenes is described in Table 2. Most are annotated into the *N. benthamiana*

argonaute protein group or *N. tabacum* regulators of gene silencing. A heat map displaying the differential gene expression of these RNAi-related host genes from the microarray study described in Chapter 3 is shown in Figure 11. This differential expression is a subtraction of gene expression between RCNMV-infected plants and mocks at 2, 6, 12 and 24 hpi. Only one RNAi-related host gene was significantly decreased at 24 hpi at an FDR cutoff of 0.01. The expressions of the majority of the host RNAi-related genes were unchanged early in the RCNMV infection process. This suggests that RCNMV does not elicit or directly suppress the expression of RNAi pathway components.

This raises the important question of just exactly how does RCNMV prevent the activation of the majority of RNAi-related genes during an infection, given that virus replication triggers host RNAi defense responses [35]? The lack of activation of most RNAi-related host genes implies that RCNMV VSRs either target genes downstream in the RNAi pathway or they don't directly interact with host components per se. To further investigate this assumption, an RCNMV mutant is required that is capable of infection but has either one or both of the suppressor activities abolished. This would seem highly doubtful given that one set of VSRs are the replication proteins (p27 and p88). Thus, more investigation will be necessary to determine: 1) how the RCNMV VSRs manipulate host RNAi-related activity and 2) does RCNMV disruption of the host RNAi pathway have an impact on the RNAi-induced degradation of host mRNA?

If the RCNMV VSRs are not responsible for the reversal of silencing effect seen with NbBTF3, what other explanation is plausible for this observed effect? One alternative

hypothesis is that RCNMV accelerates the rate of NbBTF3 transcription, resulting in a transcription rate that exceeds the degradation rate due to TRV VIGS silencing. This hypothesis is partially based on microarray and qRT-PCR analysis showing that host NbBTF3 gene expression was up-regulated at 2 hours post RCNMV inoculation and reached its highest expression level at 6 hours in wild-type *N. benthamiana* plants (Figure 1). I further hypothesized that the transcription rate exceeds the turnover rate of the silenced NbBTF3 leading to an overall increase in NbBTF3 transcript accumulation upon virus infection. To test this hypothesis, measurement of the NbBTF3 transcription rate is required in the future study. The qRT-PCR analysis is not appropriate for this purpose because it measures a quantity of mRNA at its steady state. A nuclear run on assay is one method to investigate the transcription rate for a gene of interest.

A second alternative hypothesis might be that instead of combating the production of NbBTF3, RCNMV has evolved into actually taking advantage of this defense response gene product by incorporating it into the infection process. The microarray data suggest that RCNMV replication is unaffected by the increase in NbBTF3 expression. However, at 12 hpi, NbBTF3 expression is curtailed, presumably since its expression is no longer beneficial to the RCNMV infection process. Upon re-initiation of infection in neighboring cells, NbBTF3 expression levels are allowed to elevate once again, in a manner beneficial to RCNMV. This alternative temporal hypothesis will be explored further in the next sections.

4) Pre-silenced NbBTF3 promoted RCNMV replication in the first 12 hours

The *N. benthamiana* NbBTF3 gene expression was transiently silenced by the TRV1:TRV2-NbBTF3 VIGS system 10 days prior to RCNMV inoculation. The silencing of NbBTF3 affected RCNMV RNA-1 and RNA2 replication (Figure 6) as well as R1SG1-sGFP accumulation (Figure 7). The silencing of NbBTF3 increased RNA-1 and subgenomic RNA (sgRNA) accumulation in the first 6 hours > 2 fold. The RNA-1 CP RNA increased >2 fold at 6 hpi but appeared to be the first RCNMV RNA to be decreased > 3 fold as early as 12 hpi. By 24 hours, RCNMV RNA-1, CP sgRNA and RNA-2 levels were all dramatically decreased by > 10 fold. Also, R1SG1-sGFP accumulation was decreased > 2 fold after 3 days.

This leads to the assumption that the silencing of NbBTF3 affects RCNMV replication in two different ways depending on the time point in the infection process being examined. The silencing of NbBTF3 presumably promotes RCNMV replication in the first 12 hours, but impedes replication between 12-24 hours as well as decreasing R1SG1-sGFP accumulation at 3 days. These observations might give credence to the second alternative hypothesis as NbBTF3 may be essential to the later stages of the RCNMV infection process. When NbBTF3 expression is silenced, RCNMV replication is enhanced up to 12 hpi by limiting the defense response to RCNMV until NbBTF3 becomes necessary for the rest of the infection process. At this point of the infection, NbBTF3 levels are limited and the RCNMV infection stalls as evidenced by decreasing genomic RNA as well as sgRNA levels. This is also reflected in the decreased production of sGFP from R1SG1 (via a sgRNA) after 3

days as R1SG1 does not replicate to as high a titer as wild-type RNA-1 and may not overcome the initially limiting conditions. Conversely, the increase in RCNMV replication levels in the NbBTF3 silenced plant may not be as positive for RCNMV as it appears due to the nature of the qRT-PCR assay used: qRT-PCR does not discriminate between positive strand vs. negative strand accumulation if the cDNA is synthesized with random primers as was the case for these experiments. Indeed, the increase in RCNMV RNA levels may reflect a blockage in the conversion of negative strand RNA templates to positive strand RNA progeny. To test this presumed temporal requirement for NbBTF3 in the RCNMV infection process, NbBTF3 was overexpressed in *N. benthamiana* to determine if this had a disruptive effect on the RCNMV infection process.

5) Pre-overexpressed NbBTF3 suppressed RCNMV replication in the first 12 hours

The NbBTF3 gene was transiently overexpressed in *N. benthamiana* with the PZP212-NbBTF3 overexpression construct derived from the pRTL2 expression cassette and the PZP212 binary vector. The overexpressed construct PZP212-NbBTF3 was delivered to *N. benthamiana* plants along with a silencing suppressor (PZP212-HCPro) by agroinfiltration. Two days after agroinfiltration, the expression of the NbBTF3 gene was 2 fold greater than the control leaves agroinfiltrated with only PZP212-HCPro (Figure 8).

Overexpression of NbBTF3 decreased RCNMV RNA-1 accumulation by 10 fold at 6 hpi, and 2 fold at 12 and 24 hpi. Also, RNA-2 accumulation was decreased by 3 fold at 6 hpi, and 1.5 fold at 12 hpi. This data supports the idea that increasing the expression of NbBTF3 beyond normal levels disrupts the RCNMV replication process.

6) NbBTF3 - sGFP fusion protein localizes to the nucleus

Full-length NbBTF3 protein was previously found to localize to the nucleus [20]. We now confirm that the full-length NbBTF3 clone used in the NbBTF3 gene overexpression studies in section 5 also targets to and accumulates in the nucleus.

GFP was used as the fluorescent reporter to ascertain the subcellular localization [36] of the full-length NbBTF3 in this study. GFP was fused to the full-length NbBTF3 protein at the C-terminus. This experimental design ensures that the translation will initiate from the full-length NbBTF3 first and then into the GFP. A study in HEK293T cells (human cell line) indicated that N-terminal tagging with GFP adversely affects the protein localization in reverse transfection assays, whereas tagging with GFP at the C-terminus is generally better in preserving the localization of the native protein [37]. The study showed that all C-terminal fusion proteins localized to cellular compartments in accordance with previous studies and/or bioinformatic predictions [37]. The limitation of N-terminal or C-terminal tagging with GFP has not yet been thoroughly investigated in plant cells. However, with this concern, the study in this section utilized an experimental design to tag the full-length NbBTF3 at its C-terminus. The GFP used in this study is a synthetic copy of GFP (sGFP) which was previously modified with the addition of 6xHis to the GFP structure in order to improve its stability and its fluorescence intensity [38].

The confocal images of NbBTF3-sGFP expressed in plant cells demonstrated that the full-length NbBTF3 proteins used in the overexpression study were localized inside the

nucleus (Figure 10). This localization is consistent with the nuclear translocation of NbBTF3 previously reported by Yang et al., 2007 [20].

Conclusions

The wild-type *N. benthamiana*, microarray and qRT-PCR analysis indicated that NbBTF3 expression was significantly increased 1.5 fold at 6 hpi ($\alpha = 0.05$). Interestingly, the microarray study found that NbBTF3 gene expression gradually increased beginning at 2 hpi. The silencing of NbBTF3 initially promoted RCNMV replication 2 fold while the overexpression of NbBTF3 reduced RCNMV RNA-1 10 fold and RNA-2 3 fold. NbBTF3 was previously found to be positively associated with HR-induced programmed cell death [5]. It was assumed that increasing NbBTF3 gene expression in the first 6 hours in a wild-type plant was driven by the host to activate programmed cell death. To test this hypothesis, future experiments are necessary to determine an association between NbBTF3 gene expression, the onset of HR-related gene regulation and the phenotypic cell death in a wild-type and pre-modulated NbBTF3 (both silenced and overexpressed) *N. benthamiana* host during RCNMV infection. Trypan blue is a common vital staining technique that is used for visualizing phenotypic cell death. It selectively colors dead tissues or cells blue. This association study is expected to help determine the strategy of how host NbBTF3 gene is employed to defend the plant host early in the RCNMV infection process.

Subcellular localization confirmed that NbBTF3 is localized to the nucleus. Given its nuclear targeting, it is logical to conclude that NbBTF3 is involved in regulating transcription

of other host genes. NbBTF3 is further predicted to modulate host defense genes early in an RCNMV infection. In addition to functioning as a transcription factor, BTF3 has been implicated in regulating protein localization during translation [7-9]. This suggests that NbBTF3 could aid in protein localization, however, this function would require NbBTF3 to be present in the cytoplasm. Although this study demonstrates the nuclear localization of NbBTF3, the subcellular localization study was performed with uninfected *N. benthamiana* plants. Additional experiments are required to determine the site(s) of NbBTF3 localization in *N. benthamiana* plants upon an RCNMV infection and the possibility of its function in protein translocation.

The most interesting observation in this study was the reversal of NbBTF3 silencing upon RCNMV infection: the expression of NbBTF3 was significantly shifted from a suppression 4 fold to a dramatic increase 8 fold at 12 hpi ($\alpha = 0.05$). If the RCNMV encoded VSRs impact how host genes behave, it should be broad across a cluster of genes and non-specific. However, the reversal effect is specific to only NbBTF3. This leads to my speculation that either the transcription rate of NbBTF3 is faster than the degradation rate under silencing conditions during an RCNMV infection or NbBTF3 is a required host component of the later stages of the RCNMV infection process. To test this hypothesis, additional experiments will be required to either quantitatively measure the transcription rate of NbBTF3 during an RCNMV infection or to perform qRT-PCR assays which are strand specific. These experiments are expected to shed a light on determining if NbBTF3 expression is governed at the synthesis level or the degradation level and whether the observed RCNMV accumulation is indicative of a block in the infection process.

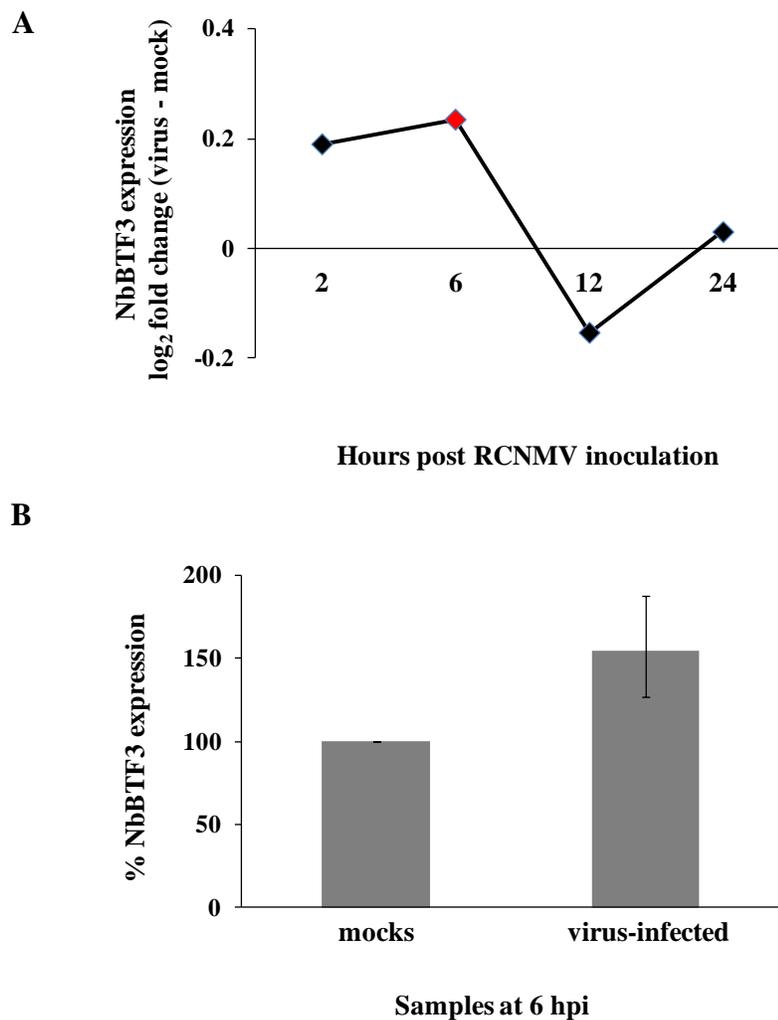


Figure 1 Host NbBTF3 (Nb09394) gene expression was up-regulated at 6 hours post RCNMV inoculation (hpi). A) Microarray analysis indicated that NbBTF3 was significantly up-regulated > 1.18 fold at 6 hpi (FDR cutoff of 0.01) (red label). B) qRT-PCR analysis confirmed that NbBTF3 was significantly up-regulated > 1.5 fold at 6 hpi ($\alpha = 0.05$).

Table 1 Nb09394 is highly homologous to a basal transcription factor 3 (BTF3) from various *Solanaceae* plants. The unigene sequence Nb09394 was searched against GenBank non-redundant protein sequence (nr) database using BLASTX method.

Species name	BLASTX hit description	GenBank Accession number	Query coverage	E-value	Identity
<i>N. benthamiana</i>	BTF3	ABE01085.1	65%	3×10^{-67}	100%
<i>N. tabacum</i> (tobacco)	BTF3-like transcription factor	AII99813.1	48%	3×10^{-65}	100%
<i>Capsicum annuum</i> (pepper)	Putative transcription factor BTF3	ABM55742.1	48%	6×10^{-64}	98%
<i>Solanum tuberosum</i> (potato)	BTF3-like transcription factor	NP_001275047.1	48%	8×10^{-64}	97%
<i>S. lycopersicum</i> (tomato)	BTF3-like transcription factor isoforms 1	XP_004246474.1	48%	4×10^{-63}	97%

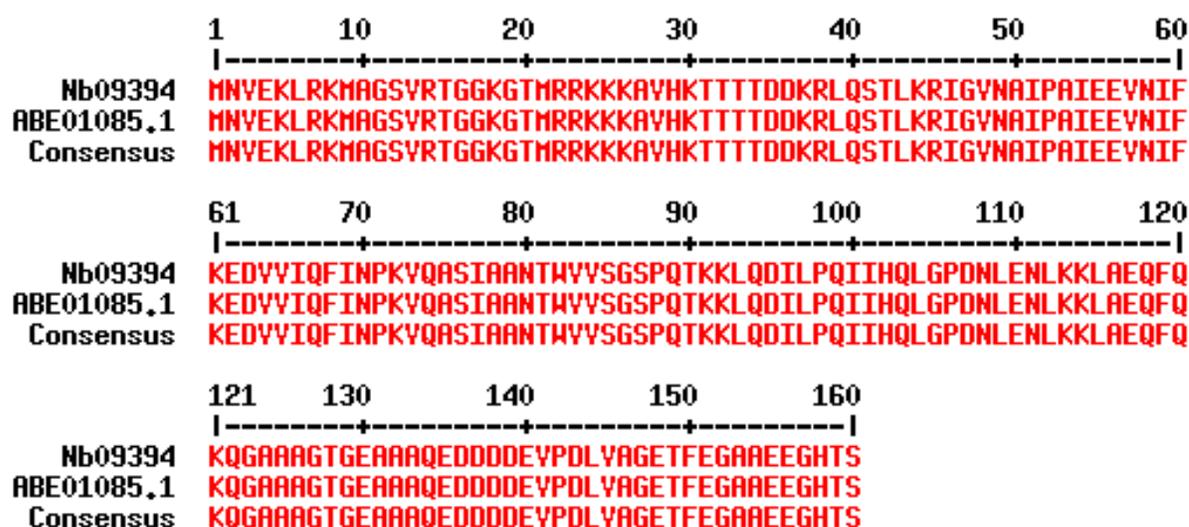


Figure 2 Nb09394 has a complete coding sequence (CDS) and its encoded peptide has a 100% residue identity to *N. benthamiana* BTF3 GenBank ID ABE01085.1, with E-value = 3×10^{-67} .

```

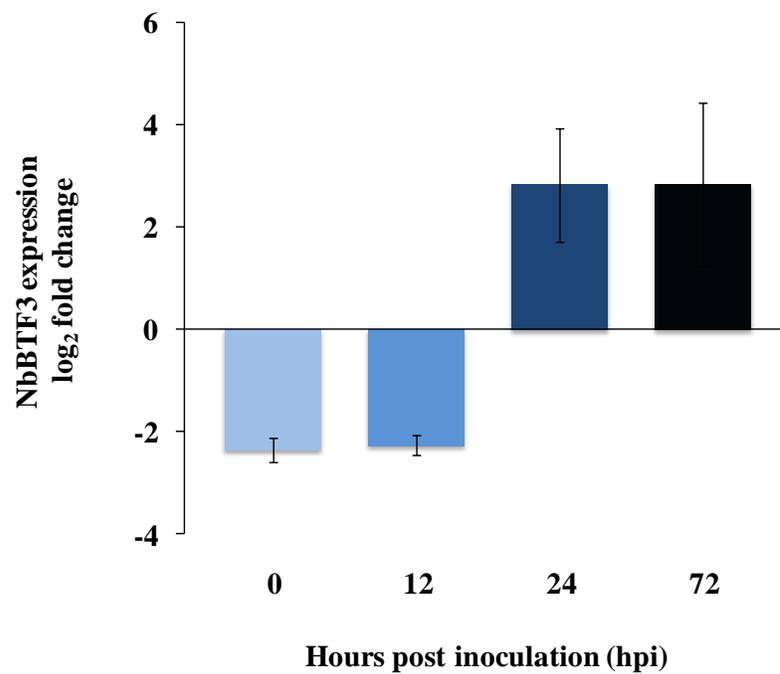
cacgaggctgtctttcattttgcagccctttgcttcccccttcccactgcggttagggt
tttagctcactcttcaagatgaatgtagaaaagctgcgtaagatggccggttcggtcaga
M N V E K L R K M A G S V R
accggtggaaagggtaccatgagaagaagaagaagccgttcacaaaacaactacaacc
T G G K G T M R R K K K A V H K T T T T
gatgacaaaagactacagagcaccttgaaaagaataggtgtcaatgccattcctgctatt
D D K R L Q S T L K R I G V N A I P A I
gaagaagtcaacattttcaaggaggatggttattccaattcattaatcccaaagttcaa
E E V N I F K E D V V I Q F I N P K V Q
gcattctattgctgcaaacacatgggttggtagtggttccccccagacaaagaattgcag
A S I A A N T W V V S G S P Q T K K L Q
gatattccttctcaaattattcaccaattgggtcctgataatttggagaatttgaagaag
D I L P Q I I H Q L G P D N L E N L K K
ttggtgagcagttccagaagcagggtgctgctgcaggtacaggtgaggctgcagcacag
L A E Q F Q K Q G A A A G T G E A A A Q
gaagacgatgatgaagtaccagatctcgtggctgggtgaaacttttgaagggtgctgct
E D D D D E V P D L V A G E T F E G A A
gaggagggtcacacttttgaagtttttgttaaattttgtggcatccttactttcaatg
E E G H T S -
ttttgatgtcaattttacatttttagcttttgagtgatctatttgatgtctatttgtgaca
catctcaaaagaattccttaaagctttttaagtaagctacttttcccttcaaagccaaaa
aaaaaaaaaaaaaa

```

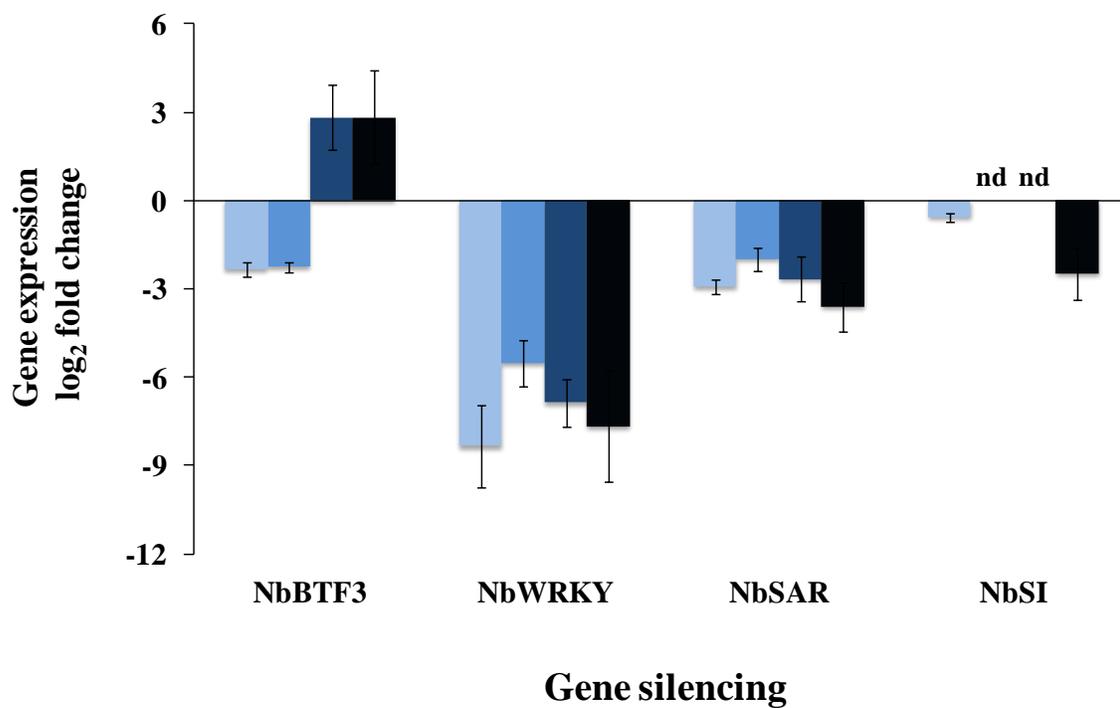
Figure 3 Nb09394 sequence features and its encoded peptide. The Nb09394 nucleotide sequence was translated to yield a peptide of 160 residues by using ExPASy tool and a standard genetic code. The selection of translation frame was based on its similarity to *N. benthamiana* BTF3 GenBank ID ABE01085.1 (Table 1 and Figure 2). The translation starts from 5' to 3' direction of unigene Nb09394 at frame +2. The Nb09394 sequence is 734 bp, containing a complete coding sequence (CDS) or a complete open reading frame (ORF; red), start codon (red box), stop codon (open box), 5' untranslated region (UTR; green), 3' UTR (purple) and 3' poly (A) tail (underlined).

Figure 4 The gene silencing of NbBTF3 was reversed 12 hours post RCNMV inoculation. (A) NbBTF3 gene expression was significantly reduced > 4 fold in *N. benthamiana* by using the TRV-VIGS vector system ($\alpha = 0.05$). The silencing vector TRV1:TRV2-NbBTF3 was delivered into plant cells by the agroinfiltration method using a 1 ml needleless syringe. NbBTF3 was pre-silenced 10 days prior to RCNMV inoculation. The pre-silenced NbBTF3 plants were inoculated with *in vitro* T7 RNA-1 transcript and *in vitro* T7 RNA-2 transcript, and inoculated leaves were harvested between 0 and 72 hours post inoculation (hpi). The qRT-PCR analysis showed that the expression of NbBTF3 remained significantly reduced > 4 fold until 12 hours followed by a significant up-regulation > 8 fold ($\alpha = 0.05$). Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. (B) The reversal of silencing effect after RCNMV inoculation was only seen with NbBTF3 among the genes selected for further analysis (Chapter 4, Chapter 5, and Chapter 7). The expression levels of silenced NbBTF3 were compared to those of other silenced genes during a 72 hour RCNMV infection assay: NbWRKY (transcription factor), NbSAR (systemic acquired resistance) and NbSI (sterol isomerase).

A



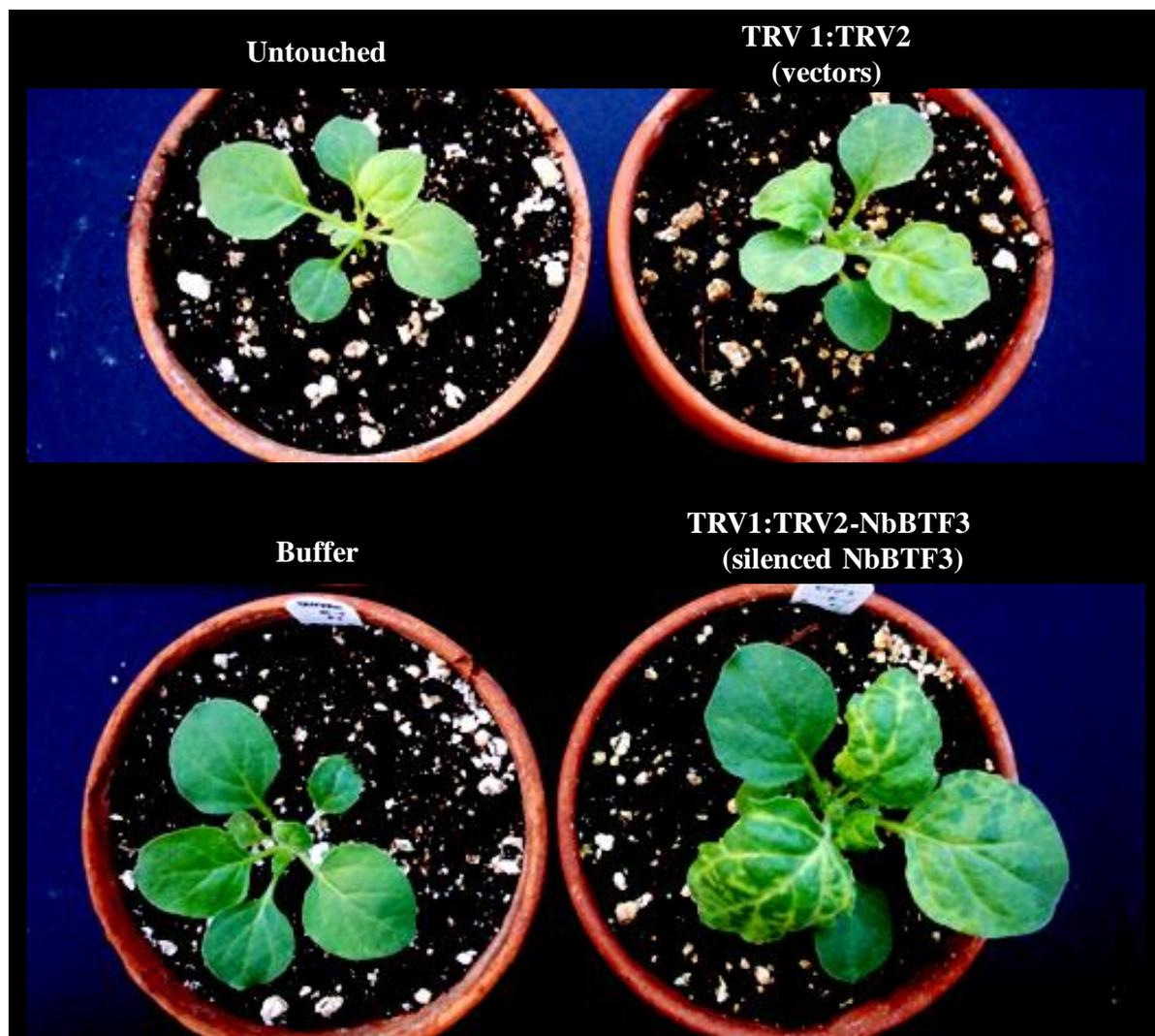
B



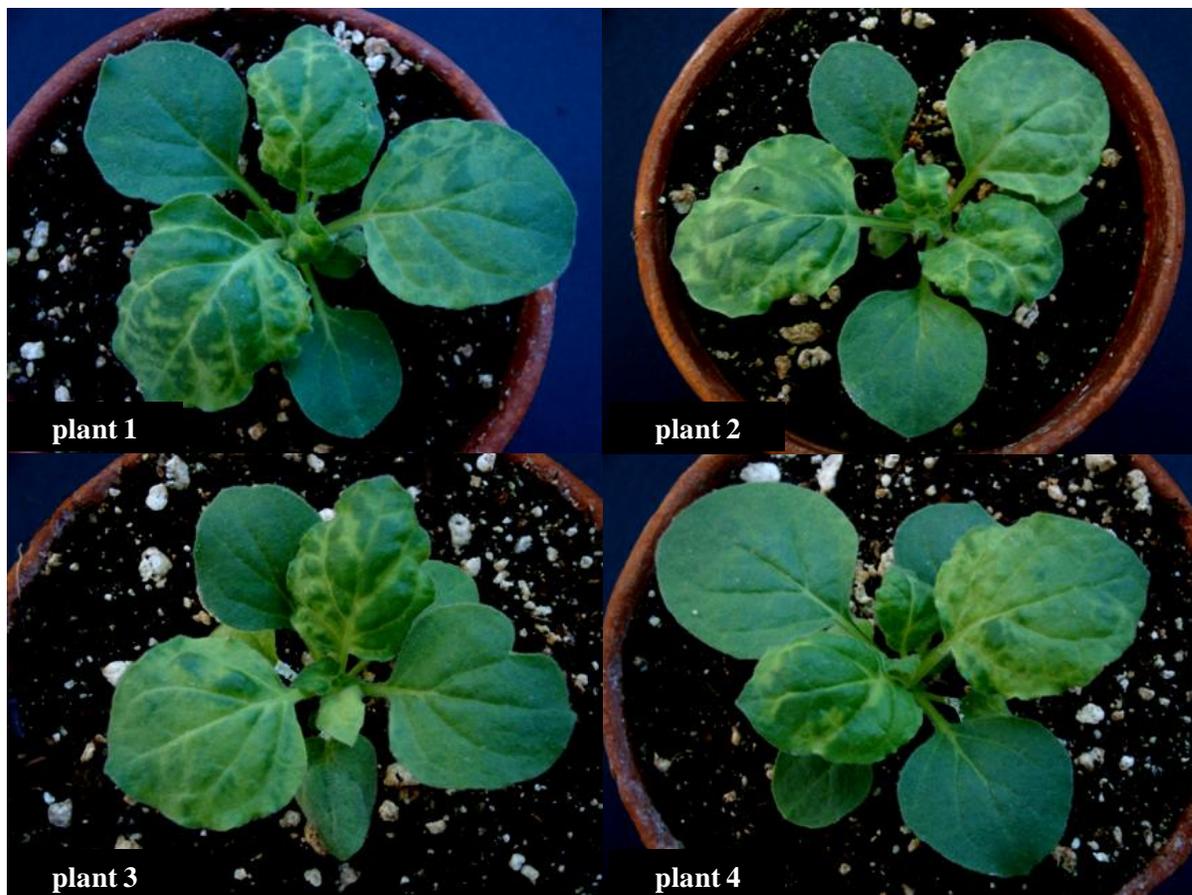
0 hpi 12 hpi 24 hpi 72 hpi nd = no data

Figure 5 Leaf chlorosis phenotype produced by silencing of NbBTF3 gene in *N. benthamiana*. A) Silencing NbBTF3 causes leaf yellowing and chlorosis in *N. benthamiana* as early as 5 days after agroinfiltration. B) Silenced NbBTF3 phenotype occurs consistently. Photographs were taken 10 days after agroinfiltration. C) Transient silencing of *NbPDS* (phytoene desaturase) gene in *N. benthamiana* using TRV-VIGS system produces a leaf bleaching phenotype. D) A closer look at *N. benthamiana* leaf phenotypes produced by different treatments. Leaves were harvested 7 days after exposing to treatments: (D1) untouched, (D2) buffer infiltration, (D3) TRV1:TRV2 agroinfiltration, (D4) TRV1:TRV2-NbPDS agroinfiltration and (D5) TRV1:TRV2-NbBTF3 agroinfiltration. All leaf samples for each treatment were harvested from the same plant. Leaves are arranged from oldest (left) to youngest (right). TRV2 is the backbone VIGS construct with no insert. TRV2-NbBTF3 is the VIGS construct with an NbBTF3 gene fragment insert (300 bp). TRV2-NbPDS is the VIGS construct with an NbPDS gene fragment insert (353 bp).

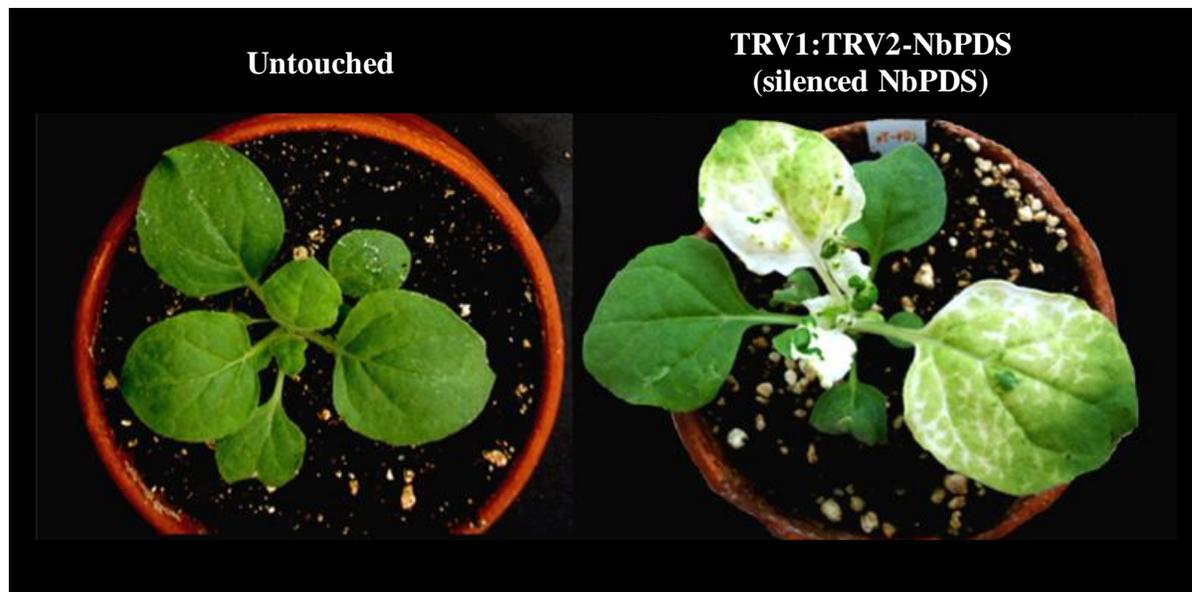
A



B



C



D

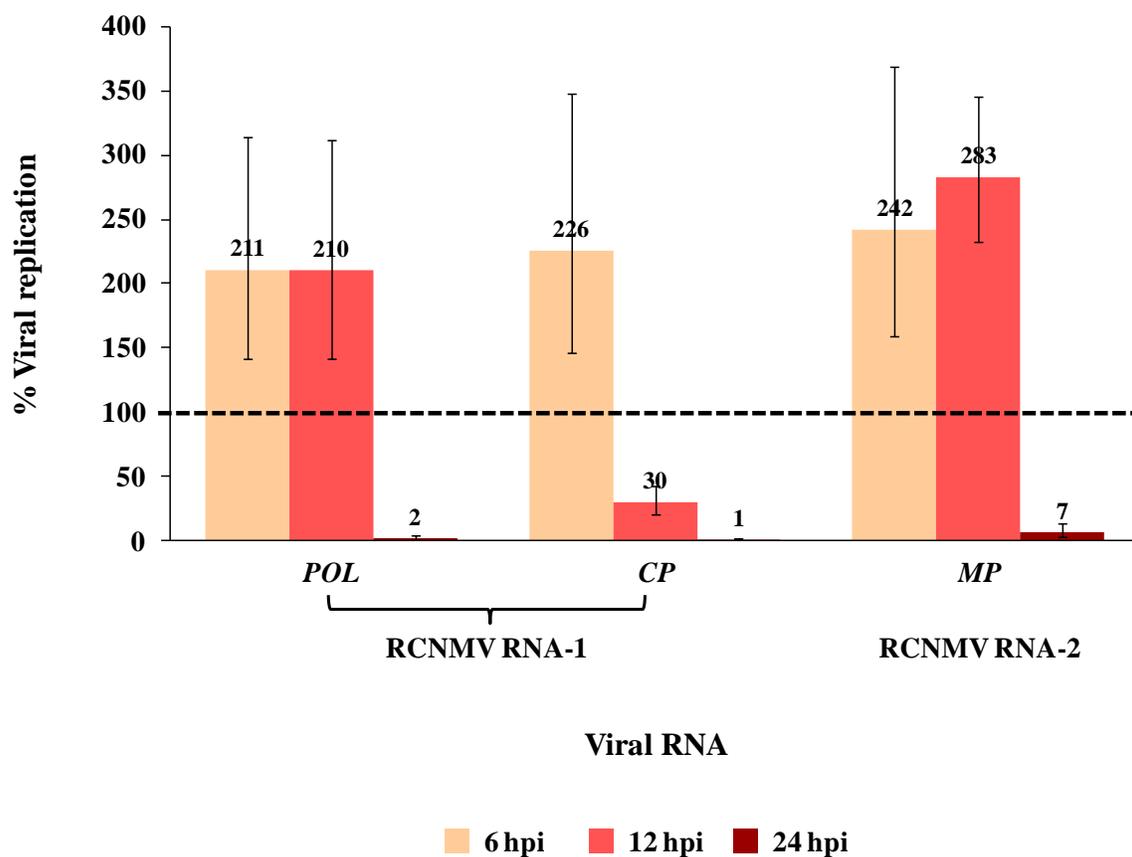


Figure 6 Pre-silenced NbBTF3 affected RCNMV RNA-1 and RNA-2 accumulation at sub 24 hours. The expression of NbBTF3 was transiently silenced by TRV1:TRV2-NbBTF3 (TRV-VIGS system) 10 days prior to RCNMV inoculation. The NbBTF3 silenced plants were inoculated with *in vitro* T7 RNA-1 and RNA-2 transcripts and leaves were harvested between 6 and 24 hours post inoculation. RCNMV RNA levels were measured using qRT-PCR analysis. Viral RNA accumulation was computed by a relative quantification method and displayed as % viral replication. Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. All data shown in this figure was significant at $\alpha = 0.1$.

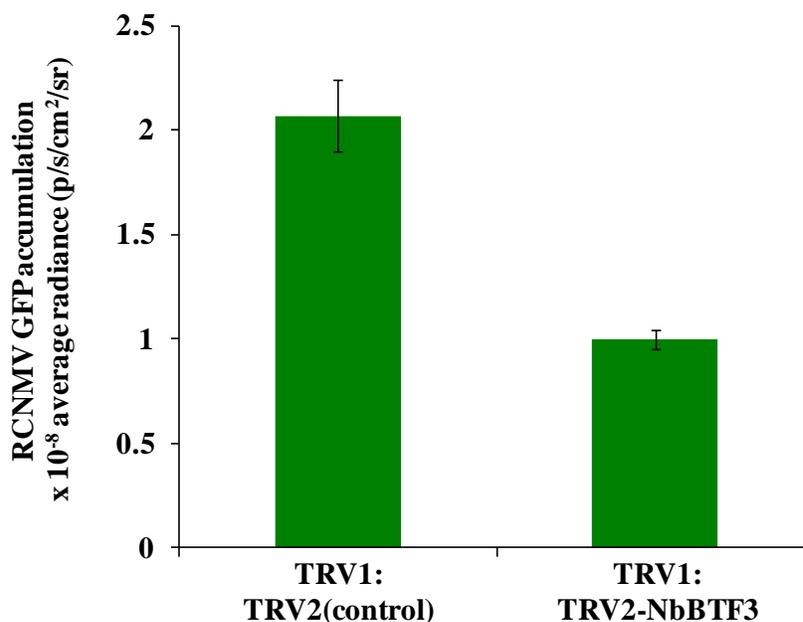


Figure 7 Pre-silenced NbBTF3 suppressed RCNMV GFP accumulation > 2 fold at 3 days post inoculation. The expression of NbBTF3 was transiently knocked down using the TRV-VIGS system 10 days prior to RCNMV inoculation. The pre-silenced NbBTF3 plants were inoculated with *in vitro* T7 R1SG1 transcript and *in vitro* T7 RNA-2 transcript, and inoculated leaves were harvested 3 days post inoculation. Plants agroinfiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis ($\alpha = 0.05$). The GFP quantification was performed with an IVIS imaging system.

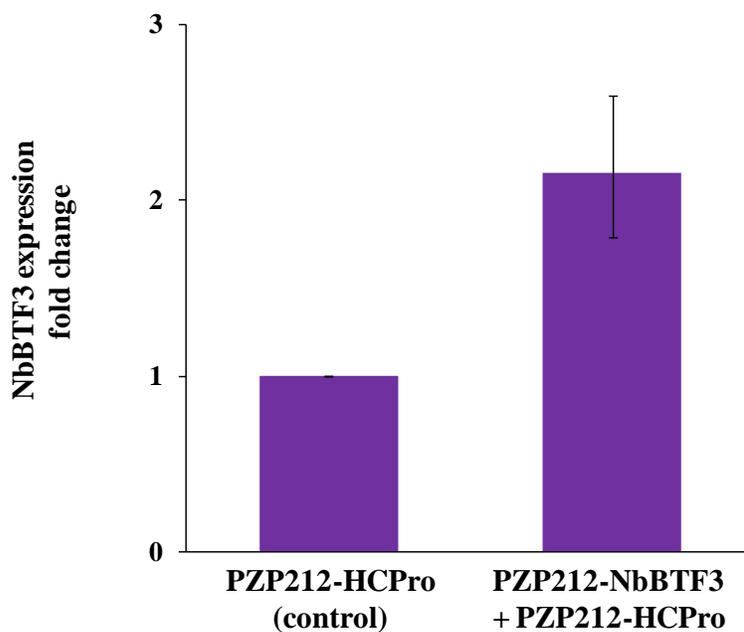


Figure 8 NbBTF3 gene was overexpressed in *N. benthamiana* by using PZP212 and pRTL2 vector system. The overexpression construct PZP212-NbBTF3 was delivered into plant cells along with the RNA silencing suppressor construct PZP212-HCPro by an agroinfiltration method to constitutively express the NbBTF3 gene. The leaf samples were harvested 2 days after agroinfiltration. Plants agroinfiltrated with only PZP212-HCPro were used as the baseline controls in a student's t-test statistical analysis. The qRT-PCR analysis showed that the expression of NbBTF3 gene was significantly increased greater than 2 fold ($\alpha = 0.05$).

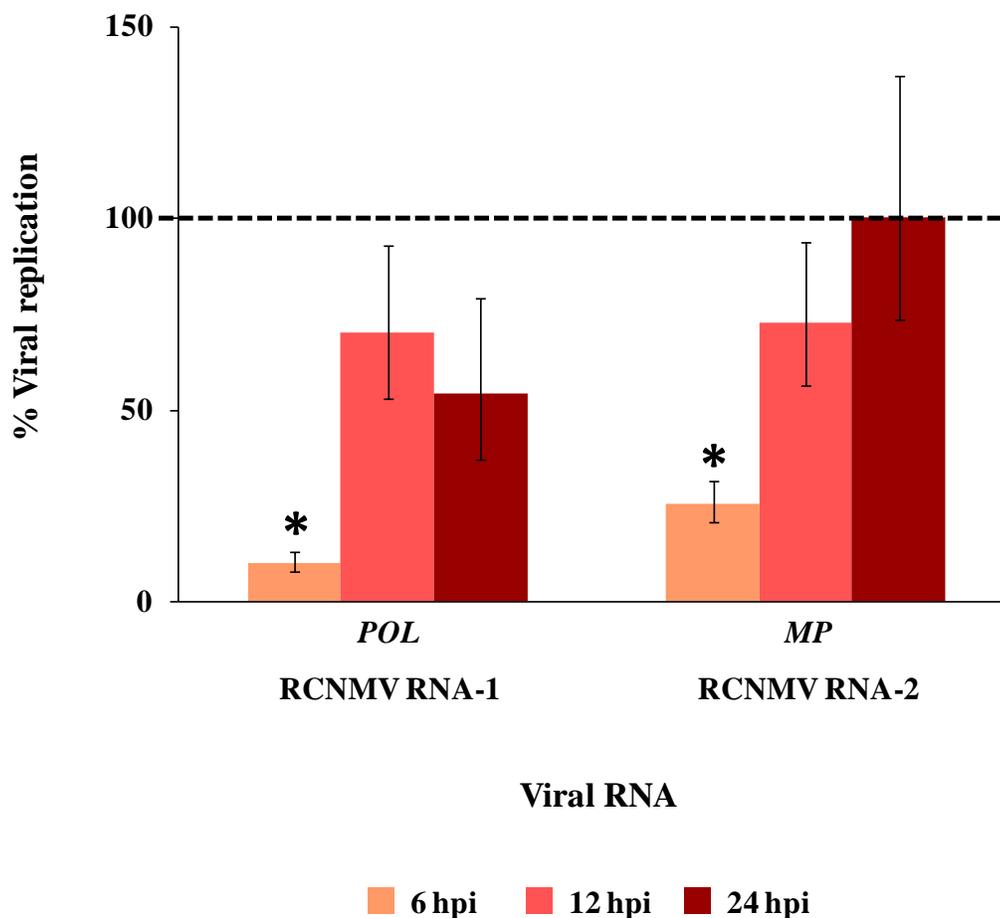


Figure 9 Pre-overexpressed NbBTF3 reduced RCNMV RNA accumulation at 6 hours of infection. The data marked with an asterisk (*) was significant at $\alpha = 0.01$. The expression of NbBTF3 gene was transiently overexpressed by agroinfiltration with PZP212-NbBTF3 (along with the RNA silencing suppressor PZP212-HCPro) 2 days prior to RCNMV inoculation. The pre-overexpressed NbBTF3 plants were inoculated with *in vitro* T7 RNA-1 and RNA-2 transcripts, and the inoculated leaves were harvested between 6 and 24 hours post inoculation (hpi). RCNMV RNAs were measured by using qRT-PCR analysis. Viral RNA accumulation was computed by a relative quantification method and displayed as % viral replication. Plants agroinfiltrated with only PZP212-HCPro 2 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis.

Figure 10 Nuclear localization of NbBTF3 protein. A modified green fluorescent protein derivative (sGFP) was used as a visual reporter to identify the subcellular localization of the full-length NbBTF3 protein. The full-length NbBTF3 protein was fused at its C-terminus to the sGFP tagging protein to generate NbBTF3-sGFP fusion protein. The localization constructs were agroinfiltrated into *N. benthamiana* leaves and the photographs were taken 2 days after agroinfiltration with a confocal microscope. The images were scanned using three different filters (green = GFP, red = chloroplast auto fluorescence, and white/black = bright field). (A-D) shows a typical pattern of sGFP subcellular localization (control), images taken from a leaf agroinfiltrated with a mixture of PZP212-sGFP and PZP212-HCPro. A merged image of sGFP (D) shows that a fluorescence signal was accumulated inside the nucleus and cytoplasm. However, sGFP does not appear to localize to the chloroplast. (E-H) shows a characteristic pattern of NbBTF3-sGFP subcellular localization, images taken from a leaf agroinfiltrated with a mixture of PZP212-NbBTF3-sGFP and PZP212-HCPro. A merged image of NbBTF3-sGFP (H) indicates that the fluorescence signal was exclusively concentrated inside the nucleus.

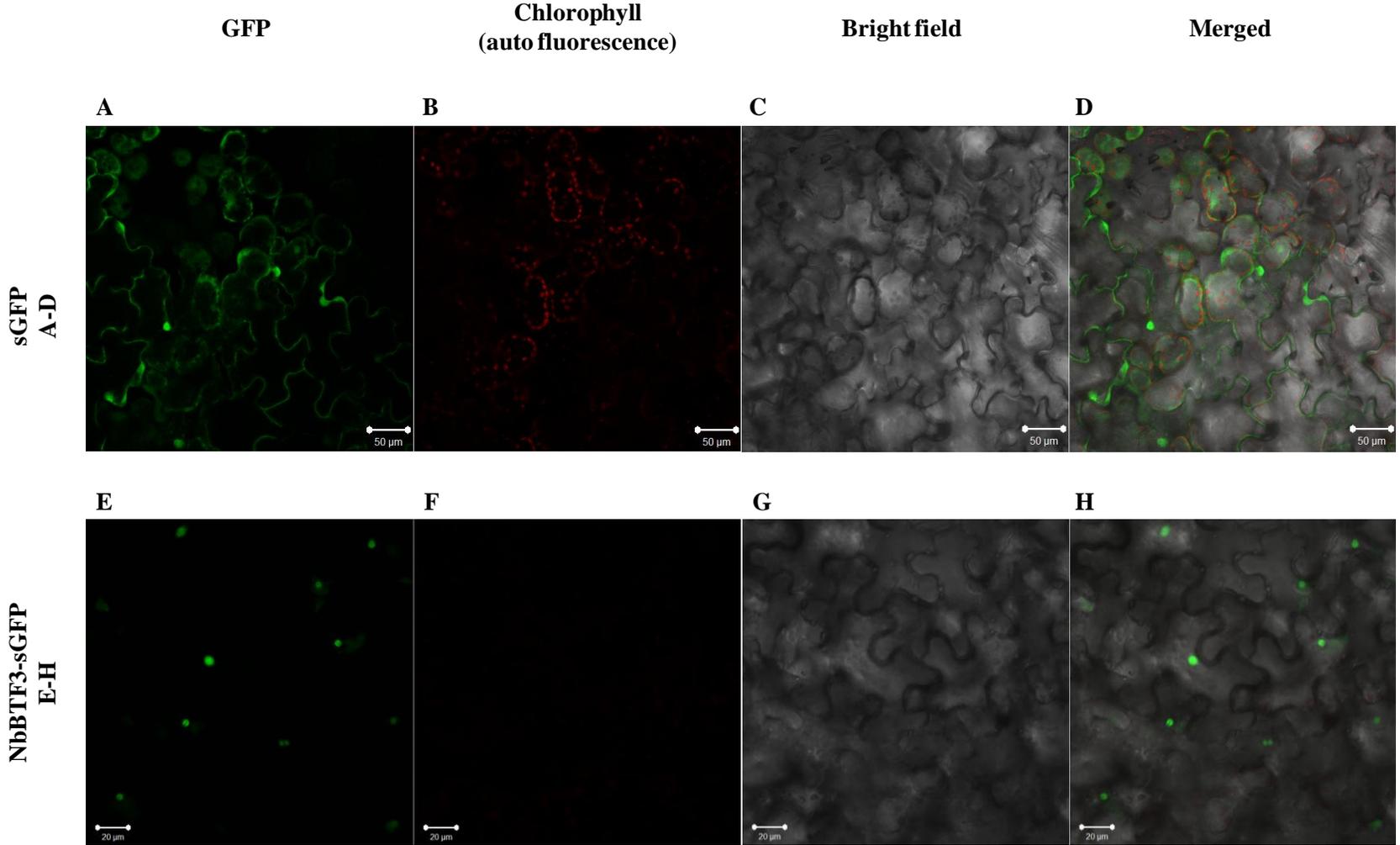


Table 2 *N. benthamiana* RNAi-related genes. These host genes are presented in our custom microarray (Chapter 2). Homology search of these genes were performed against GenBank nr database (non-redundant protein sequences) and GO database (Gene Ontology) using BLASTX.

Gene ID	Function	GenBank accession number	GenBank E-value	GO	GO description
NB00035	argonaute protein group	gi 84688908 gb ABC61503.1 AGO1-2 [<i>Nicotiana benthamiana</i>]	1.39X 10 ⁻⁸⁰	mf	nucleic acid binding
NB03873	argonaute protein group	gi 84688912 gb ABC61505.1 AGO4-2 [<i>Nicotiana benthamiana</i>]	2.74X 10 ⁻¹²²	mf	nucleic acid binding
NB04102	argonaute protein group	gi 84688912 gb ABC61505.1 AGO4-2 [<i>Nicotiana benthamiana</i>]	3.71X 10 ⁻³¹	mf	nucleic acid binding
NB04902	argonaute protein group	gi 84688910 gb ABC61504.1 AGO4-1 [<i>Nicotiana benthamiana</i>]	2.06X 10 ⁻⁹⁰	mf	nucleic acid binding
NB04903	argonaute protein group	gi 84688910 gb ABC61504.1 AGO4-1 [<i>Nicotiana benthamiana</i>]	5.81X 10 ⁻¹⁰³	mf	nucleic acid binding
NB09170	argonaute protein group	gi 84688906 gb ABC61502.1 AGO1-1 [<i>Nicotiana benthamiana</i>]	1.93X 10 ⁻¹³⁴	mf	nucleic acid binding
NB11968	dsRNA-specific nuclease dicer and related ribonuclease	gi 224064945 ref XP_002301611.1 predicted protein [<i>Populus trichocarpa</i>]	3.59X 10 ⁻⁶⁷	mf mf mf bp bp	protein binding RNA binding ribonuclease III activity RNA processing regulation of RNA metabolic process
NB00590	regulator of gene silencing–calmodulin-like protein	gi 12963415 gb AAK11255.1 AF329729_1 regulator of gene silencing [<i>N. tabacum</i>]	2.00X 10 ⁻³⁸	mf	calcium ion binding
NB03631	regulator of gene silencing–calmodulin-like protein	gi 12963415 gb AAK11255.1 AF329729_1 regulator of gene silencing [<i>N. tabacum</i>]	2.47X 10 ⁻⁴³	mf	calcium ion binding
NB10906	regulator of gene silencing–calmodulin-like protein	gi 12963415 gb AAK11255.1 AF329729_1 regulator of gene silencing [<i>N. tabacum</i>]	3.81X 10 ⁻⁶³	mf	calcium ion binding
NB02934	silencing group b protein	gi 255628397 gb ACU14543.1 unknown [<i>Glycine max</i>]	2.12X 10 ⁻²²	mf mf bp	N-acetyltransferase activity protein binding acyl-carrier-protein biosynthetic process

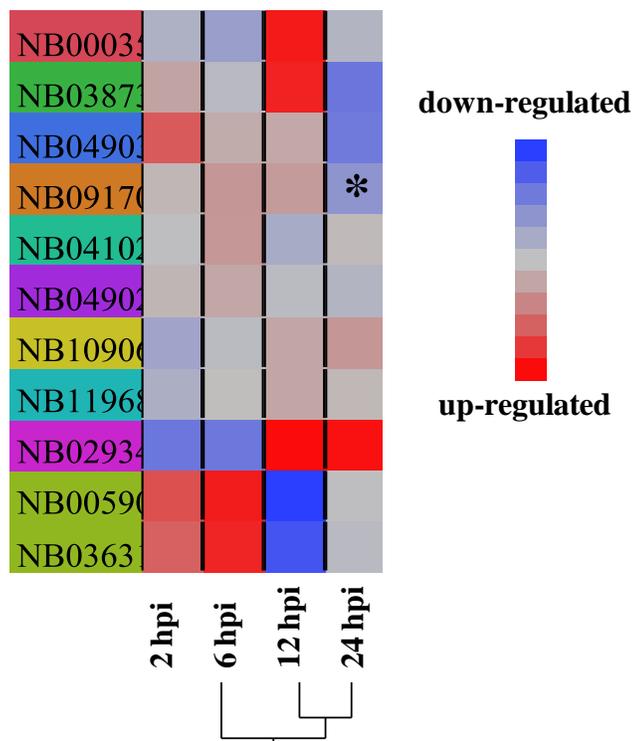


Figure 11 Heat map displays the differential expression of *N. benthamiana* RNAi-related genes in an early RCNMV infection from the microarray study. The differential gene expression is a subtraction between RCNMV-infected plants and mocks at 4 time points of 2, 6, 12 and 24 hours post RCNMV inoculation (hpi). The differential expression was computed from the microarray study previously described in Chapter 3. The asterisk (*) denotes significantly differentially expressed gene at FDR cutoff of 0.01. The functional annotation of these RNAi-related host genes is described in Table 2.

Table 3 Oligonucleotide information used in Chapter 6.

Name	Sequence information
Nb09394F	ACATTGAAAGTAAGGGATGC
Nb09394R	ATTTGAAGAAGTTGGCTGAG
GS09394F	AATTTCTAGAGCTCACATGGGTTGTTAGTGG
GS09394R	AATTTCTAGAGCTCTGAAGGGAAAAGTAGCTTAC
FL09394F	AATTCCATGGGATCGATGAATGTAGAAAAGCTGCG
FL09394R	AATTTCTAGATTACATCGATGCAGAAGTGTGACCCTCCTCAG
3'sGFP/XbaI	GCTCTAGACGCGTACTTGTACAGCTCGTCC
RNA1-POLF	TCACAAGGGTCAAATTCTCAAATCCT
RNA1-POLR	TGCTGCTTTTTGGTATAACTTCCTCTT
RNA1-CPF	ATGTCTTCAAAAGCTCCCAA
RNA1-CPR	CGCTCATGACTAACTGGGTA
RNA2F	CGCGTCTGATTGAGTTGGAAGTA
RNA2R	CTGCCTTGATGCTCGACAGTA
TRV2F	TACTCAAGGAAGCACGATGAGC
TRV2R	GAACCGTAGTTTAATGTCTTCGGG
TEVLeader80	CGAATCTCAAGCAATCAAGCATTC
35sTERM	ATAAGAACCCTAATTCCC
ActF	GTGACCTCACTGATAGTTTGA
ActR	TACAGAAGAGCTGGTCTTTG

Materials and Methods

1) Plants, RCNMV inoculum preparation and inoculation procedure

N. benthamiana was a plant model used in this experiment. Details of plant maintenance, RCNMV inoculum preparation and inoculation procedure were described in Chapter 5 materials and methods. Briefly, RCNMV inoculum was prepared in a 110 μ l volume per 1 plant. This 110 μ l inoculum was a mixture of 1 μ l *in vitro* T7 RNA-1 transcript (or 1 μ l *in vitro* T7 R1SG1 transcript), 1 μ l *in vitro* T7 RNA-2 transcript and 108 μ l inoculation buffer (10 mM sodium diphosphate, pH 7.2). Four leaves per 1 plant were inoculated with either RCNMV inoculum or buffer. 27 μ l of the RCNMV transcript mixture (or inoculation buffer) was pipetted onto each leaf and mechanically rubbed with carborundum (abrasive). Inoculated plants were maintained in a temperature and light controlled environment at 18-22°C, 16 hour-light and 8 hour-dark period at the NCSU greenhouses (Method Road).

2) Plasmid DNA constructs

A construction of the silencing construct, the overexpression construct, and the subcellular localization construct are performed accordingly to detail in Chapter 5. Following is a brief detail and the modification.

The silencing construct TRV2-NbBTF3 utilizing the TRV VIGS system [27] was designed to down regulate NbBTF3 expression in *N. benthamiana*. A 300 bp fragment of the NbBTF3 gene was amplified accordingly to OneTaq DNA polymerase protocol (NEBTM), with a primer set of GS09394F and GS09394 R (Table 3) to generate the GS-NbBTF3. The

PCR conditions for amplifying GS-NbBTF3 were as follows: 1) a denaturation cycle of 94°C for 30 sec, 2) 30 amplification cycles of 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min, and 3) a final extension cycle of 68°C for 5 min. A ligation between GS-NbBTF3 and TRV2 was performed via *SacI* restriction site. A colony PCR technique using the primer set of TRV2F and TRV2R (Table 3) was used to identify TRV2 clones that contained the GS-NbBTF3 insert. A positive plasmid was also confirmed by digesting with *SacI*. The resultant construct TRV2-NbBTF3 was used for down regulating NbBTF3 expression in *N. benthamiana* for a gene silencing study.

The PZP212-NbBTF3 construct was designed for overexpression of NbBTF in *N. benthamiana* and is based on the pRTL2 expression cassette and the PZP212 binary vector. A full-length NbBTF3 DNA (450 bp) was amplified accordingly to OneTaq DNA polymerase protocol (NEB™), with a primer set of FL09394F and FL09394R (Table 3) and the extension cycle at 68°C for 1 min 30 sec. A full-length NbBTF3 DNA was firstly cloned into pGEM T-Easy in order to increase a copy number. A ligation between a full-length NbBTF3 DNA and pRTL2 was performed via *NcoI* and *XbaI* restriction sites. The pRTL2 construct containing the full-length NbBTF3 was then verified by sequencing using sequencing primers TEVleader80 and 35sTERM (Table 3). The resultant plasmid pRTL2-NbBTF3 was digested with *HindIII* to release the insert containing the full-length NbBTF3 expression cassette. This isolated fragment was subsequently cloned into PZP212 via *HindIII* restriction site. The PZP212-NbBTF3 construct was used to constitutively express the putative full-length NbBTF3 protein in *N. benthamiana* for a gene over-expression study.

The PZP212-NbBTF3-sGFP construct was designed for subcellular localization study of NbBTF3. A full-length NbBTF3 DNA was initially cloned into the RNA transcript expression vector pHST2-sGFP through the unique *ClaI* restriction site. The orientation of the clone was confirmed by a restriction analysis and the properly orientated construct was used as a template to amplify the NbBTF3-sGFP product using a primer set of FL09394F and 3'sGFP/*XbaI* (Table 3). The PCR product was then cleaved with *NcoI* and *XbaI* and ligated into similarly digested pRTL2 to produce construct pRTL2-NbBTF3-sGFP. The nature of the construct was verified by sequencing and the expression cassette containing NbBTF3-sGFP was released by digestion with *HindIII* and subsequently cloned into PZP212 to yield construct PZP212-NbBTF3-sGFP.

3) *Agrobacterium* preparation and infiltration

Agrobacterium was prepared accordingly to a detail in Chapter 5 materials and methods. Briefly, *Agrobacterium* based constructs were transformed into *Agrobacterium tumefaciens* strain C58C1. Individual colonies were inoculated into 2 ml LB broth cultures with the appropriate antibiotics and incubated at 28°C for 20 hr with shaking. From these initial cultures 250 µl was used to inoculate 5 ml LB broth cultures with the appropriate antibiotics and 40µM acetosyringone/10mM MES, Ph 5.6. These cultures were similarly incubated at 28°C for 20 hr with shaking. Cultures were subsequently pelleted, resuspended in 10mM MgCl₂/10mM MES, pH 5.6/200µM acetosyringone to the appropriate OD₆₀₀ reading and incubated at room temperature for at least 3 hr prior to syringe infiltration into plants.

For NbBTF3 silencing experiments, *Agrobacterium* cultures of TRV1 and TRV2-NbBTF3 (or TRV2 for the control) were mixed at a 1:1 ratio immediately prior to infiltration into plants. For NbBTF3 overexpression experiments, *Agrobacterium* cultures of PZP212-NbBTF3 and the RNA silencing suppressor PZP212-HCPro (or PZP212-HCPro alone for the control) were mixed at a ratio of 1:1 immediately prior to infiltration into plants. For subcellular localization experiments, *Agrobacterium* cultures of PZP212-NbBTF3-sGFP (or PZP212-sGFP for the control) and the RNA silencing suppressor PZP212-HCPro were mixed at a ratio of 1:1 immediately prior to infiltration into plants. Leaves infiltrated for the subcellular localization experiments were harvested 2 days after agroinfiltration and imaged via confocal microscopy (Cellular and Molecular Imaging Facility, NCSU).

4) Total RNA extraction and cleanup

Leaves to be sampled in each experiment consisted of the following: four plants and two leaves per plant were used to represent the test condition (silencing or overexpression of NbBTF3) as well as the respective control (see above). Total RNA extracts isolated from the same plant were pooled to represent one biological replicate (4 biological replicates for the test condition and 4 biological replicates for controls). Leaf samples were collected either 2 days (overexpression experiments) or 10 days (silencing experiments) after agroinfiltration as well as various time points after RCNMV transcript inoculation for both types of experiments.

A total RNA extraction and cleanup were performed accordingly to details in Chapter 5 materials and methods. Briefly, total RNA was extracted from 100 mg *N. benthamiana* leaf

tissue samples according to the TRIzol protocol (Invitrogen™). And total RNA extracts destined for qRT-PCR analysis were treated with Turbo DNA-free (Ambion™) according to the manufacturer's protocol to remove any residual input DNA. The samples were stored at -20°C.

5) RNA quantification by real-time PCR (qRT-PCR)

First strand cDNA synthesis and quantitative real-time PCR (qRT-PCR) preparation were performed according to details in Chapter 5 material and methods. Briefly, First strand cDNA was synthesized by using a DNA-free total RNA (equivalent to ~1 µg total RNA) as a template. SYBR Green-based quantitative real-time PCR (qRT-PCR) method was used in this study. The qRT-PCR reaction was prepared accordingly to FastStart Universal SYBR Green Master ROX protocol (Roche™). The first strand cDNA equivalent to 40 ng total RNA was used to prepare qRT-PCR reaction. The qRT-PCR reactions were setup in a 384 well plate and placed in ABI7900 HT Fast real-time PCR system (Applied Biosystems™). The Applied Biosystems SDS software version 2.4 was used to monitor and dissect real-time PCR data. Three qRT-PCR reactions were prepared per 1 biological replicate. All PCR products are less than 200 base pairs. The cycling condition was set as follows: 1 cycle at 95°C for 10 min, and 40 cycles at 95°C for 15 sec /55°C for 1 min. An association analysis was also performed in order to test if there were any non-specific PCR products formed. The condition for the association analysis was set as follows: 1 cycle at 95°C for 15 sec, 60°C for 15 sec and 95°C for 15 sec.

NbBTF3 transcript levels were assayed using primers Nb09394F and Nb09394R (Table 3). RCNMV RNA levels were assayed with 2 primer pairs for RNA-1 and 1 primer pair for RNA-2 (Table 3): 1) RNA1-POLF and RNA1-POLR for probing RNA-1 at the polymerase (RNA-1 POL), 2) RNA1-CPF and RNA1-CPR for probing RNA-1 at the coat protein (RNA-1 CP) and 3) RNA2F and RNA2R for probing RNA-2 at the movement protein (RNA-2 MP).

6) qRT-PCR data analysis

A relative quantification, $2^{-\Delta\Delta C_t}$ method (equation shown in Chapter 5 materials and methods) [39] was used to analyze the real-time PCR data. Expression of the target gene was normalized against the expression of a reference gene. Actin was used as the reference genes to normalize the target gene that was performed within the same plate. Student's t-test statistical analysis was used to examine the real-time PCR data.

7) RCNMV sGFP quantification

N. benthamiana plants were inoculated with a combination of *in vitro* T7 R1SG1 and RNA-2 transcripts. Ten plants and two leaves per plant were used for each test condition and each control. Leaf samples were harvested 3 days post inoculation. The GFP fluorescence intensity was analyzed by the IVIS imaging system and subjected to Student's t-test statistical analysis. The IVIS Lumina System (Xenogen Corporation, Alameda, CA) is capable of quantifying photon emission from a variety of sources. A CCD camera measured and recorded photon emission data which was then incorporated into Living Image Software

(Xenogen Corp.) for further analysis. Whole leaves were placed under the CCD camera and measurements were taken with a GFP excitation filter, and exposure time of 1 second.

Reference

1. Dangl JL, Jones JD: **Plant pathogens and integrated defence responses to infection.** *Nature* 2001, **411**(6839):826-833.
2. Kachroo P, Chandra-Shekara A, Klessig DF: **Plant signal transduction and defense against viral pathogens.** *Advances in virus research* 2006, **66**:161-191.
3. Soosaar JL, Burch-Smith TM, Dinesh-Kumar SP: **Mechanisms of plant resistance to viruses.** *Nature Reviews Microbiology* 2005, **3**(10):789-798.
4. Loebenstein G: **Local lesions and induced resistance.** *Advances in virus research* 2009, **75**:73-117.
5. Huh SU, Kim K-J, Paek K-H: **Capsicum annuum basic transcription factor 3 (CaBTF3) regulates transcription of pathogenesis-related genes during hypersensitive response upon Tobacco mosaic virus infection.** *Biochemical and biophysical research communications* 2012, **417**(2):910-917.
6. Zheng X-M, Moncollin V, Egly J-M, Chambon P: **A general transcription factor forms a stable complex with RNA polymerase B (II).** *Cell* 1987, **50**(3):361-368.
7. Wiedmann B, Sakai H, Davis TA, Wiedmann M: **A protein complex required for signal-sequence-specific sorting and translocation.** 1994.
8. Luring B, Sakai H, Kreibich G, Wiedmann M: **Nascent polypeptide-associated complex protein prevents mistargeting of nascent chains to the endoplasmic reticulum.** *Proceedings of the National Academy of Sciences* 1995, **92**(12):5411-5415.
9. Funfschilling U, Rospert S: **Nascent Polypeptide-associated Complex Stimulates Protein Import into Yeast Mitochondria.** *Molecular biology of the cell* 1999, **10**(10):3289-3299.
10. Rospert S, Dubaquié Y, Gautschi M: **Nascent-polypeptide-associated complex.** *Cellular and Molecular Life Sciences CMLS* 2002, **59**(10):1632-1639.

11. Brockstedt E, Otto A, Rickers A, Bommert K, Wittmann-Liebold B: **Preparative high-resolution two-dimensional electrophoresis enables the identification of RNA polymerase B transcription factor 3 as an apoptosis-associated protein in the human BL60-2 Burkitt lymphoma cell line.** *Journal of protein chemistry* 1999, **18**(2):225-231.
12. George R, Beddoe T, Landl K, Lithgow T: **The yeast nascent polypeptide-associated complex initiates protein targeting to mitochondria in vivo.** *Proceedings of the National Academy of Sciences* 1998, **95**(5):2296-2301.
13. Deng JM, Behringer RR: **An insertional mutation in the BTF3 transcription factor gene leads to an early post implantation lethality in mice.** *Transgenic research* 1995, **4**(4):264-269.
14. Markesich DC, Gajewski KM, Nazimiec ME, Beckingham K: **bicaudal encodes the Drosophila beta NAC homolog, a component of the ribosomal translational machinery.** *Development* 2000, **127**(3):559-572.
15. Bloss TA, Witze ES, Rothman JH: **Suppression of CED-3-independent apoptosis by mitochondrial Î²NAC in Caenorhabditis elegans.** *Nature* 2003, **424**(6952):1066-1071.
16. Freire MA: **Translation initiation factor (iso) 4E interacts with BTF3, the beta subunit of the nascent polypeptide-associated complex.** *Gene* 2005, **345**(2):271-277.
17. Wang Y, Zhang X, Lu S, Wang M, Wang L, Wang W, Cao F, Chen H, Wang J, Zhang J: **Inhibition of a basal transcription factor 3-like gene Osj10gBTF3 in rice results in significant plant miniaturization and typical pollen abortion.** *Plant and cell physiology* 2012, **53**(12):2073-2089.
18. Ma H-Z, Liu G-Q, Li C-W, Kang G-Z, Guo T-C: **Identification of the TaBTF3 gene in wheat (Triticum aestivum L.) and the effect of its silencing on wheat chloroplast, mitochondria and mesophyll cell development.** *Biochemical and biophysical research communications* 2012, **426**(4):608-614.
19. Kang G, Ma H, Liu G, Han Q, Li C, Guo T: **Silencing of TaBTF3 gene impairs tolerance to freezing and drought stresses in wheat.** *Molecular Genetics and Genomics* 2013, **288**(11):591-599.
20. Yang K-S, Kim H-S, Jin U-H, Lee SS, Park J-A, Lim YP, Pai H-S: **Silencing of NbBTF3 results in developmental defects and disturbed gene expression in chloroplasts and mitochondria of higher plants.** *Planta* 2007, **225**(6):1459-1469.

21. Carmo LST, Resende RO, Silva LP, Ribeiro SGA, Mehta A: **Identification of host proteins modulated by the virulence factor AC2 of Tomato chlorotic mottle virus in *Nicotiana benthamiana*.** *Proteomics* 2013, **13**(12-13):1947-1960.
22. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2008, **36**(suppl 1):D25-D30.
23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
24. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExpPASy: the proteomics server for in-depth protein knowledge and analysis.** *Nucleic acids research* 2003, **31**(13):3784-3788.
25. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR: **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic acids research*, **39**(suppl 1):D225-D229.
26. Ratcliff F, Martin-Hernandez AM, Baulcombe DC: **Technical advance: Tobacco rattle virus as a vector for analysis of gene function by silencing.** *The Plant Journal* 2001, **25**(2):237-245.
27. Dinesh-Kumar SP, Anandalakshmi R, Marathe R, Schiff M, Liu Y: **Virus-induced gene silencing.** In: *Plant Functional Genomics*. Springer; 2003: 287-293.
28. Senthil-Kumar M, Hema R, Anand A, Kang L, Udayakumar M, Mysore KS: **A systematic study to determine the extent of gene silencing in *Nicotiana benthamiana* and other Solanaceae species when heterologous gene sequences are used for virus-induced gene silencing.** *New phytologist* 2007, **176**(4):782-791.
29. Cunningham Jr F, Gantt E: **Genes and enzymes of carotenoid biosynthesis in plants.** *Annual review of plant biology* 1998, **49**(1):557-583.
30. Takeda A, Tsukuda M, Mizumoto H, Okamoto K, Kaido M, Mise K, Okuno T: **A plant RNA virus suppresses RNA silencing through viral RNA replication.** *The EMBO journal* 2005, **24**(17):3147-3157.
31. Powers JG, Sit TL, Heinsohn C, George CG, Kim K-H, Lommel SA: **The Red clover necrotic mosaic virus RNA-2 encoded movement protein is a second suppressor of RNA silencing.** *Virology* 2008, **381**(2):277-286.
32. Voinnet O: **Induction and suppression of RNA silencing: insights from viral infections.** *Nature Reviews Genetics* 2005, **6**(3):206-220.

33. Hannon GJ: **RNA interference**. *Nature* 2002, **418**(6894):244-251.
34. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R: **Control of translation and mRNA degradation by miRNAs and siRNAs**. *Genes & Development* 2006, **20**(5):515-524.
35. Voinnet O: **RNA silencing as a plant immune system against viruses**. *TRENDS in Genetics* 2001, **17**(8):449-459.
36. Tsien RY: **The green fluorescent protein**. *Annual review of biochemistry* 1998, **67**(1):509-544.
37. Palmer E, Freeman T: **Investigation into the use of C- and N- terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays**. *Comparative and functional genomics* 2004, **5**(4):342-353.
38. Chiu W-l, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J: **Engineered GFP as a vital reporter in plants**. *Current Biology* 1996, **6**(3):325-330.
39. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR**. *Nucleic acids research* 2001, **29**(9):e45-e45.

Chapter 7

Involvement of *Nicotiana benthamiana* sterol biosynthesis in the early stages of *Red clover necrotic mosaic virus* infection

Chapter summary

Herein, I report on the regulation of *Nicotiana benthamiana* C-8, 7-sterol isomerase (NbSI) gene expression at the early stages of *Red clover necrotic mosaic virus* (RCNMV) infection. This constitutes the first report of a sterol isomerase gene in *N. benthamiana*. Both computational and experimental methods were employed to analyze the NbSI sequence structure, functional annotation and subcellular localization. The data leads to the conclusion that NbSI is a novel C-8, 7-sterol isomerase gene from *N. benthamiana*. Microarray and qRT-PCR analysis described in Chapter 3 reveals that host NbSI was up-regulated at 12 hours post RCNMV inoculation. The NbSI gene was further functionally assayed for its effects on RCNMV infection by both transient gene silencing and gene overexpression studies in *N. benthamiana*. These functional assays suggest a potential dual effect of NbSI on RCNMV replication: its early involvement in membrane proliferation (presumably to the benefit of RCNMV) followed by a switch to being a precursor in the plant defense pathway (presumably a detriment to RCNMV).

Abstract

Replication is a critical step in the plant viral infection cycle. A common feature among positive-stranded RNA plant viruses is a requirement for their replication in association with host membranes. How plant viruses rearrange host membranes and recruit host membrane components to promote membrane proliferation for their replication to take place remains unknown. A genome-wide screening in the model plant host *N. benthamiana* indicated that the sterol biosynthesis genes C-4 sterol methyl oxidase2 (SMO2) and C-8, 7-sterol isomerase (NbSI) were both significantly up-regulated at 12 hours post RCNMV inoculation at an FDR cutoff of 0.01. This is the first report of the full-length cDNA sequence of C-8, 7-sterol isomerase from *N. benthamiana*. Sequence analysis and a subcellular localization study indicated that NbSI is an integral membrane protein containing five α -helical transmembrane domains while NbSI fusions with sGFP predominantly localized along cellular, nuclear and endoplasmic reticulum membranes. To gain insight into the role of sterols in RCNMV replication, NbSI was further functionally assayed by both transient gene silencing and overexpression studies in *N. benthamiana*. Silencing of NbSI in *N. benthamiana* suppressed RCNMV replication in the first 6 hours, but promoted RCNMV replication at 24 hours as well as increased R1SG1-sGFP accumulation at 3 days. Overexpression of NbSI in *N. benthamiana* inhibited RCNMV replication in the first 12 hours. These functional assay results suggest a potential transition in NbSI function as a component of membranes (promoting membrane proliferation for RCNMV replication) to acting as a precursor of Brassinosteroid synthesis (modulating host defense system). The role

switching could affect RCNMV replication in an opposite direction, suggesting a so-called antagonistic effect.

Keywords; *Nicotiana benthamiana*, RCNMV, viral replication, sterol isomerase, sterol, lipid

Introduction

Replication of most if not all plus-stranded RNA plant viruses occurs on host intracellular membranes or host organellar membranes. The formation of spherules, consisting of host lipid membranes invaginated and viral replication proteins as well as recruited host proteins, has been observed in infected host cells for several plus-stranded RNA viruses. These virus-induced spherules serve as the sites of viral replication for facilitating the building of viral factories, synthesis and accumulation of viral replication proteins and providing protection against cellular nucleases, proteases and cellular defenses including RNAi [1-5]. Therefore, for replication to occur, these plant viruses need to hi-jack host intracellular/organellar membranes as well as host membrane component synthesis to affect remodeling of the appropriate host membrane into a site for viral replication. These virus-host interaction events are hypothesized to: 1) occur at the early stages of virus infection and 2) result in the induction of host membrane proliferation that requires the production of more host membrane components and consequently more lipid biosynthesis. To study these events, we utilized *Nicotiana benthamiana* (a model host for most plant viruses) and *Red clover necrotic mosaic virus* (RCNMV, a typical plus-stranded RNA plant virus). A previous study in the Lommel laboratory found that RCNMV replication is

associated with the endoplasmic reticulum (ER) membranes [5]. We were able to determine the co-localization of RCNMV replication proteins (p27 and p88) to the ER with subsequent membrane restructuring and proliferation. Both replication proteins co-localized to the cortical and cytoplasmic ER and were associated with invaginations of the nuclear envelope.

The goal of this project is to examine the modulation of *N. benthamiana* lipid biosynthesis-related gene expression at the early stages of an RCNMV infection. Rather than study one host gene at a time, we employed microarray technology to monitor the expression of thousands of host genes simultaneously over time. A custom *N. benthamiana* microarray was developed. This microarray is composed of 362,205 probes with a capacity of 9 probes per gene and 3 replicates per probe. The microarray represents a total of 13,415 unigenes, equivalent to an estimated coverage of ~38% of the *N. benthamiana* transcriptome. The creation of this custom microarray is described in Chapter 2.

Host transcription profiles were analyzed at 2, 6, 12 and 24 hours post- RCNMV inoculation (hpi). The microarray study described in Chapter 3 suggests that host sterol biosynthesis may be associated with RCNMV replication. Sterols are ubiquitous and essential membrane components in all eukaryotes, affecting many membrane functions. Sterols regulate membrane rigidity, fluidity and permeability by interacting with other lipids and proteins within the membranes [6-8]. The *N. benthamiana* microarray contains 7 known genes involved in sterol biosynthesis. Among these seven genes, two were significantly up-regulated at 12 hpi at an FDR cutoff of 0.01 (Figure 15). One of the genes is the *N. benthamiana* C-4 sterol methyl oxidase 2 (SMO2; GenBank accession number AAQ83692.1)

[9]. The other gene is microarray gene ID Nb02987 which is computationally annotated to C-8, 7 sterol isomerase (NbSI) due to a strong sequence similarity to sterol isomerases from other plant species (Table 1). Nb02987 has never been documented in any public database. This study is the first report of this gene, its full-length cDNA encoding a C-8, 7 sterol isomerase from *N. benthamiana*, its sequence analysis and subcellular localization.

C-4 sterol methyl oxidase (SMO) and sterol isomerase (SI) are key enzymes in sterol biosynthesis (Figures 13 and 14) [9]. Sterol biosynthesis differs in several steps in animals, fungi and plants, but the removal of two methyl groups at the C-4 position by SMO and a catalysis of isomerization of the C-8 double bond to the C-7 position within the steroid B ring by SI are critical and rate limiting [9]. The silencing of SMO1 and SMO2 genes in *N. benthamiana* reduced *Tomato bushy stunt virus* (TBSV) RNA accumulation but had a lesser inhibitory effect on *Tobacco mosaic virus* [10]. TBSV is a closely related virus to RCNMV with both residing in the *Tombusviridae* family. The up-regulated SMO2 at 12 hours post RCNMV inoculation supports the idea that SMO2 may be an important host factor in sterol biosynthesis contributing to the membrane proliferation that positively correlates with TBSV replication.

At this time, there has been no report of an association between SI and virus infection. Due to sterol functioning as a membrane component, my first hypothesis is that the increased expression of NbSI is driven by RCNMV in order to increase production of sterol for promoting membrane proliferation. Although most studies focus on the role of plant sterol functioning as a membrane component [6, 7, 11], sterol also serves as a precursor in

Brassinosteroid biosynthesis (Figure 13) [12]. Brassinosteroids (BRs) are plant hormones that are essential in plant development [13, 14]. A remarkable feature of BRs is their potential to increase resistance in plants to a wide spectrum of stresses such as low and high temperatures, drought, high salt and pathogen attack [15-18]. Studies of the BR signaling pathway and gene-regulating properties indicate that there is cross-talk between BRs and other hormones, including those with established roles in plant defense responses such as abscisic acid, jasmonic acid and ethylene [19]. Given that sterol can function as a precursor of BRs, this leads to my second hypothesis in this project that a modulation of host sterol biosynthesis at the early stages of RCNMV infection is not driven by virus for promoting membrane proliferation, but rather is driven by the plant host in order to launch a defense response.

Result and Discussion

1) RCNMV up-regulated *N. benthamiana* C-8, 7-sterol isomerase (NbSI) gene expression at 12 hpi

The microarray study of the modulation of *N. benthamiana* gene expression early in an RCNMV infection described in Chapter 3 revealed that host NbSI gene expression was not altered in the first 6 hours of the RCNMV infection. However, it was significantly up-regulated at 12 hours (Figure 1) at an FDR cutoff of 0.01. The expression was subsequently down-regulated at 24 hours post RCNMV inoculation (hpi). The microarray study also indicated that RCNMV takes about 12-24 hours to complete its infection in the primary

infected cell with the first 12 hours being critical as RCNMV starts its movement to the neighboring cells. An up-regulation of NbSI expression at 12 hpi is likely concordant with RCNMV migrating to the cells adjacent to the primary infected cell. Also, a dramatic down-regulation of NbSI expression at 24 hpi suggests that the plant host may be attempting to re-adjust NbSI expression levels. The increased expression of host NbSI at 12 hpi was also confirmed by qRT-PCR analysis (Figure 1).

The central hypothesis in this chapter posits that if there is an interaction between NbSI and RCNMV replication components, pre-modulation of this host gene (either up or down-regulation) will affect RCNMV replication efficiency. To test this hypothesis, the impact of altering NbSI gene expression on RCNMV replication was examined by utilizing transient gene silencing and gene overexpression assays. The associated impact of the pre-modulated host NbSI gene to viral infection efficiency was indirectly assessed by two different methods: 1) a quantitative measurement of viral RNA accumulation by quantitative real-time PCR (qRT-PCR) analysis and 2) a quantitative measurement of viral GFP accumulation produced by R1SG1 (the RCNMV RNA-1 construct that contains the green fluorescent protein in place of the coat protein coding region) via IVIS fluorescence imaging assay.

2) NbSI (Nb02987) sequence analysis and functional annotation

The host NbSI gene is represented by the microarray unigene ID (probe ID) Nb02987. The custom microarray was created from 13 *K. N. benthamiana* unigenes. The

detail of the unigene collection and microarray construction is thoroughly described in Chapter 2. The unigene Nb02987 from *N. benthamiana* has not been characterized or documented in any public databases. Therefore, this is the first study that thoroughly analyzes its sequence, functional annotation and subcellular localization.

A homology search of the Nb02987 sequence against the GenBank non-redundant protein sequence (nr) database using BLASTX [20, 21] reveals that the Nb02987 top five BLAST hits are to C-8, 7-sterol isomerase from tomato, potato, grape, cocoa and corn (Table 1).

A search against GenBank Conserved Domain Database (CDD) [22] showed that the Nb02987 protein is most similar to the Emopamil binding protein (EBP) gene. EBP is also known as 3 β -hydroxysteroid- Δ 8, Δ 7-isomerase. This enzyme is responsible for one of the final steps in the production of cholesterol. EBP is conserved in *Arabidopsis* [23], humans [24] and mice [25]. EBP is an integral membrane protein of the endoplasmic reticulum, cellular membrane and other intracellular membrane-bounded organelles. EBP shares structural features with bacterial and eukaryotic drug transport proteins and contains four putative transmembrane segments. GO annotations [26] related to EBP include transmembrane signaling receptor activity and C-8 sterol isomerase activity. Human sterol isomerase, a homologue of mouse EBP, is suggested not only to play a role in cholesterol biosynthesis, but also to affect lipoprotein internalization. In humans, mutations of EBP are known to cause the genetic disorder chondrodysplasia punctata (CDPX2) [24, 25]. This

syndrome is lethal in most males, and affected females display asymmetric hyperkeratotic skin and skeletal abnormalities.

The Nb02987 nucleotide sequence encodes a peptide of 191 residues determined by the ExPASy tool [27] (Figure 2). The selection of the correct translation frame was based on the translation product's similarity to sterol isomerase from tomato (SlySt, Genbank ID XP_004242445.1) and potato (StSI, GenBank ID NP_001275362.1) [28].

A full-length Nb02987 cDNA sequence was predicted based on a strong sequence similarity to tomato and potato sterol isomerase (Figures 3 and 4). The predicted full-length sequence information was later used to design a primer set to clone/construct a full-length NbSI protein for an overexpression study (Section 4) as well as the subcellular localization study (Section 5).

3) Pre-silenced NbSI suppressed RCNMV replication in the first 6 hours, followed by a promotion after 12 hours

Host NbSI gene expression was reduced in *N. benthamiana* by the TRV1:TRV2-NbSI VIGS vector system [29, 30]. The silencing constructs TRV1:TRV2-NbSI were delivered to plant cells using an agro-infiltration method. The host NbSI gene was transiently silenced in *N. benthamiana* 10 days prior to RCNMV inoculation. The pre-silenced NbSI plants were inoculated with full-length infectious RNA transcripts from linearized plasmids of RC169 and RC2 and harvested at 0 and 72 hours post RCNMV inoculation (hpi). The qRT-PCR

analysis showed that the expression of silenced NbSI remained reduced relative to control at 72 hpi (Figure 5).

Silencing of NbSI affected RCNMV RNA-1 and RNA2 replication (Figure 6). The silenced NbSI decreased RNA-1 POL, RNA-1 CP and RNA-2 MP RNA accumulation within the first 6 hours (Figure 6). RNA-POL and RNA1-CP levels were both suppressed as early as 3 hpi whereas; RNA2-MP was not suppressed until 6 hpi. Interestingly, neither RNA-1 nor RNA-2 levels were affected at 12 hpi. However, RNA-1 and RNA-2 levels were all increased at 24 hpi. The production of sGFP from R1SG1-sGFP was increased after 3 days (Figure 7).

These observations lead to the assumption that silencing of NbSI may affect RCNMV replication in two different ways depending on the progress of the infection. The silenced NbSI presumably suppressed RCNMV replication in the first 6 hours, but promoted RCNMV replication at 24 hours as well as increased sGFP accumulation from R1SG1-sGFP at 3 days post inoculation.

The transition from suppression to promotion of RCNMV replication could indicate a switch in the role of NbSI from functioning as a component of membranes to acting as a precursor of Brassinosteroid synthesis which modulates the plant defense system. This leads to further speculation that sterol isomerase may have a dual effect on RCNMV replication. To test this hypothesis, a gene overexpression assay was used to examine the impact of overexpression of NbSI on RCNMV replication efficiency in the next section.

4) Pre-overexpressed NbSI suppressed RCNMV replication in the first 12 hours

The NbSI gene was overexpressed in *N. benthamiana* by using the PZP212-NbSI construct derived from the pRTL2 expression cassette and the PZP212 binary vector. The overexpression construct PZP212-NbSI was delivered to plant cells along with the RNA silencing suppressor construct (PZP212-HCPro) using the agro-infiltration method. At day 2 after agro-infiltration, the expression of the NbSI gene in leaves agro-infiltrated with PZP212-NbSI/PZP212-HCPro was significantly increased > 600 fold in comparison to the control leaves agro-infiltrated with only PZP212-HCPro at $\alpha = 0.05$ (Figure 8).

The NbSI gene was overexpressed for two days prior to RCNMV inoculation. The pre-overexpressed NbSI suppressed RCNMV replication in the first 12 hours as indicated by the qRT-PCR results utilizing the RNA-1 POL and RNA-2 MP primer sets (Figure 9). Interestingly, overexpression of NbSI gene caused a visible cell death 2 days after agroinfiltration. The observed cell death is the likely reason for suppression of viral RNA accumulation as early as 3 hpi. Likewise, overexpression of the fusion protein NbSI-sGFP also displayed visible cell death. Microscopic examination confirmed a cell death-like phenotype for both unfused and sGFP-fused NbSI. To further investigate this phenomenon, a common vital staining technique such as Trypan blue should be used in future studies. This staining technique will selectively color only dead cells.

The results of this overexpression study did not support the idea of NbSI functioning as a membrane component during the first 6 hours of RCNMV infection in the NbSI silenced

plants. However, it likely reaffirms the functional role of NbSI as an inducer of the defense response after 12 hours of RCNMV infection.

5) NbSI is an integral membrane

5.1) NbSI - sGFP fusion protein localizes to cellular, nuclear and ER membranes

Green fluorescent protein (GFP) was used as a fluorescence tag to identify the subcellular localization [31] of the full-length NbSI protein. GFP was fused to the C-terminus of the full-length NbSI protein. This experimental design ensures that the translation will have to start from the full-length NbSI first followed by the GFP in the NbSI- GFP construct. A study in HEK293T cells (human cell line) indicated that N-terminal tagging with GFP adversely affects the protein localization in reverse transfection assays, whereas tagging with GFP at the C-terminal is generally better in preserving the localization of the native protein [32]. This study further showed that all C-terminal fusion proteins localized to cellular compartments in accordance with previous studies and/or bioinformatic predictions [32]. The limitation of N-terminal or C-terminal tagging with GFP has not been thoroughly investigated in plant cells. With this concern, I utilized an NbSI C-terminal GFP fusion for subcellular localization studies. The GFP used in this study is a synthetic copy of GFP (sGFP) which was previously modified by codon optimization for expression in humans and maize along with the addition of 6xHis to the GFP N-terminus in order to improve its stability and its fluorescence intensity [33].

The free sGFP control confocal images exhibited typical sGFP localization, accumulating along the cellular membrane and inside the nucleus (Figure 10). Many studies have previously confirmed nuclear translocation of GFP as a result of its relatively small size (~27 kDa), which allows for passive diffusion through the nuclear pores [34]. The size exclusion limit of nuclear pore complexes is approximately >60 kDa [35-37]. Any protein with a mass less than 60 kDa may passively diffuse into the nucleus.

The mass of NbSI and the NbSI-sGFP fusion protein was predicted using a bioinformatic tool (ExpASY [27]). The computed molecular mass estimation for NbSI is 25.41 kDa and NbSI-sGFP is 52.31 kDa. Given the molecular mass of the NbSI-sGFP fusion product is within the limits for passive transport into the nucleus, one would expect NbSI-sGFP localization inside the nucleus, similar to free sGFP. However, the localization experiment demonstrated that NbSI-sGFP did not accumulate within the nucleus, but rather predominantly accumulated along the nuclear cellular membranes (Figure 10). A closer look at NbSI-sGFP in Figure 10 (J and K) also suggested accumulation on or at the endoplasmic reticulum (ER) as indicated by the fluorescent network or reticulate pattern.

This localization observation leads to the speculation that NbSI may be a transmembrane protein or an integral membrane protein. In the next section, bioinformatic tools were utilized to predict the presence of subcellular localization signals in the NbSI protein structure.

5.2) NbSI is predicted to localize to the plasma membrane and ER membrane

Three bioinformatic tools were utilized to predict the NbSI subcellular localization. These three bioinformatic tools used different algorithms to predict protein localization and used different benchmark datasets for training their algorithm. CELLO [38] confidently predicted that NbSI localizes to the plasma membrane with a weaker prediction for the ER. WoLFPSORT [39] has a first prediction that NbSI is localized to the ER with a score 5.5 and a second prediction to the plasma membrane with a score 5.0. Plant-mPLoc [40] predicted that NbSI is localized to the ER.

5.3) NbSI protein is predicted to contain five α -helical transmembrane domains

A transmembrane domain is any three-dimensional protein structure which is thermodynamically stable in a membrane. The ability to fold into a transmembrane domain structure is an important characteristic of transmembrane proteins. These proteins are also known as integral proteins. Transmembrane domains can be composed of a single transmembrane α -helix, a stable complex of several transmembrane α -helices or a transmembrane β -barrel. In this study, the bioinformatic tool, TMHMM server v.2.2 [41, 42], was used to predict the presence of one or more transmembrane domains in NbSI. This program has been recognized as one of the best for identifying transmembrane domains [43]. TMHMM predicted that NbSI contains five α -helical transmembrane domains. The graphic result (Figure 11) shows the probability of the residue location. This detail helps predict which residues reside within protein segments that face inside, outside or are transmembrane

in nature. A weakness of this prediction is that it cannot distinguish a signal anchor α -helical structure (target protein localization to membrane) from a signal peptide α -helical structure (target protein localization to ER). This is the primary limitation of TMHMM: predicted transmembrane segments in N-terminal regions sometimes turnout to be signal peptides. This could lead to a false interpretation of whether the protein of interest is an integral ER membrane protein or a secretory protein that is destined to the ER for its secretory pathway.

A signal peptide can be either an ER signal peptide or a secretory signal peptide. It directs the protein across the ER membrane in eukaryotes in order to facilitate secretion. A signal anchor is a transmembrane helix that functions much like a signal peptide since it is recognized by the Signal Recognition Particle (SRP) and inserted into the translocon; but instead of being cleaved and degraded, it remains in the membrane and anchors the protein to it [44, 45].

My assumption is that the α -helical structure that was identified in NbSI is a characteristic feature of the transmembrane signal anchor. Another bioinformatic tool, SignalP 4.1 server, was employed to help discriminate signal peptides from signal anchors [46]. SignalP 4.1 suggests that NbSI contains a transmembrane domain. The graphic result (Figure 12) from SignalP 4.1 shows no cleavage site found in the first N-terminal 70 amino acids therefore indicating that NbSI has no signal peptide.

The cumulative computational evidence supports the conclusion that NbSI is an integral protein that contains a signal anchor of five α -helical transmembrane domains. This

reaffirms the reasoning behind the observed localization of NbSI along cellular, nuclear and ER membranes.

Conclusions

This study demonstrates that the unigene Nb02987, assigned the name NbSI, is a novel C-8, 7-sterol isomerase found in *N. benthamiana*. NbSI is an integral membrane protein predicted to have five α -helical transmembrane domains, which predominantly localizes to cellular, nuclear and ER membranes. *N. benthamiana* microarray expression analysis and qRT-PCR analysis indicated that NbSI expression was increased at 12 hours post RCNMV inoculation. Pre-silenced NbSI suppressed RCNMV accumulation in the first 6 hours, but promoted RCNMV accumulation at 24 hours as well as increased sGFP accumulation from R1SG1-sGFP at 3 days post-inoculation. Pre-overexpressed NbSI inhibited RCNMV replication in the first 12 hours. My hypothesis is that an increased NbSI expression in a wild-type host is driven by the virus in order to promote membrane proliferation for its replication. However, the excessive sterol accumulation may disturb Brassinosteroid biosynthesis resulting in an activation of host defense. Brassinosteroid hormone treatment has been shown to enhance a broad range of disease resistance in rice and tobacco, but the molecular mechanism remains unknown [15]. Wild-type tobacco treated with Brassinolide exhibited enhanced resistance to the viral pathogen *Tobacco mosaic virus* (TMV), the bacterial pathogen *Pseudomonas syringae* pv. *tabaci* (*Pst*) and fungal pathogens of the *Oidium* sp.

Some key questions remain and need to be answered to determine the role of NbSI and sterol biosynthesis on RCNMV replication. These key questions are: 1) how does inhibition of sterol biosynthesis affect RCNMV replication? (this will determine if sterol molecules are essential host factors for RCNMV replication), 2) are the levels of Brassinosteroid hormones or key intermediates in Brassinosteroid synthesis altered at the early stages of an RCNMV infection? and 3) are the amounts of sterols within membranes altered during RCNMV infection? In contrast to animals and fungi (which contain only one major sterol), plant cells synthesize a complex array of sterol mixtures in which sitosterol, stigmasterol and 24-methylcholesterol often predominate. Sitosterol and 24-methylcholesterol regulate membrane fluidity and permeability in a similar manner to cholesterol in mammalian cell membranes. In contrast, stigmasterol might be specifically required for cell proliferation [6, 7]. Future experimentation is required to determine which plant sterols are essential host factors contributing to membrane proliferation for viral replication.

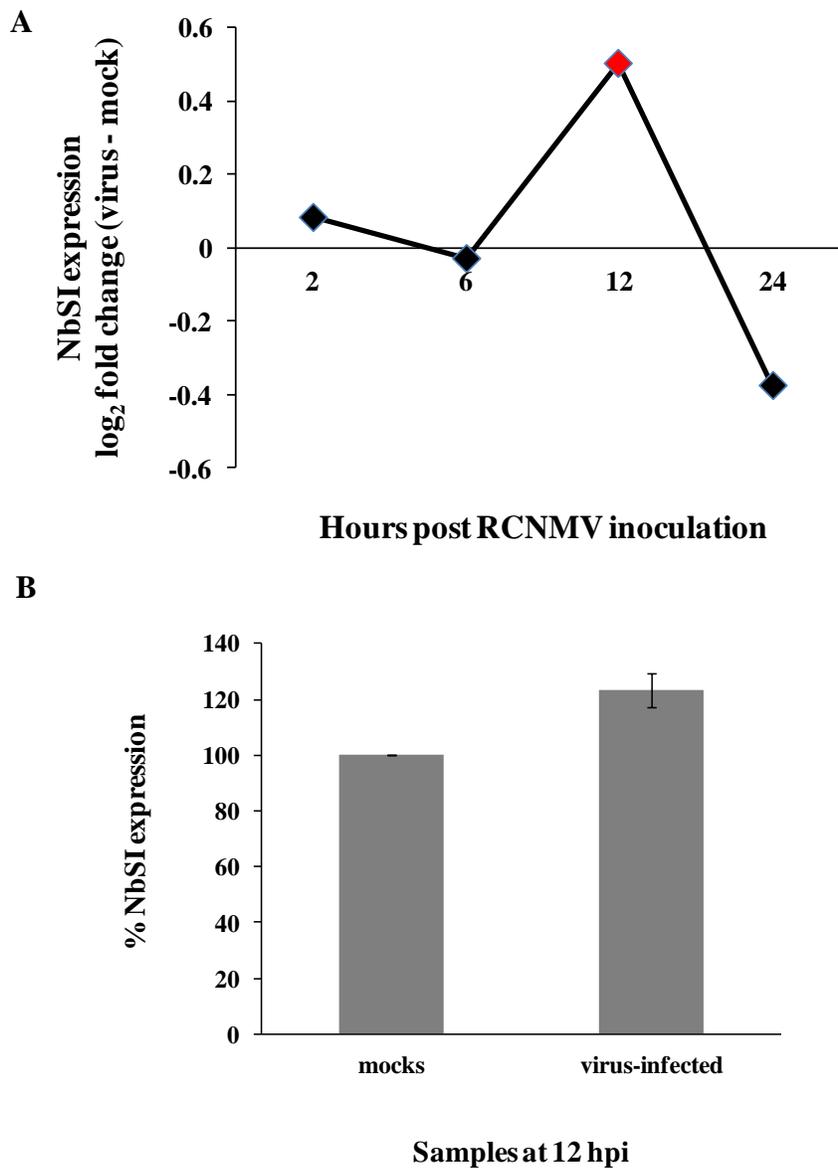


Figure 1 Host NbSI (Nb02987) gene expression was up-regulated at 12 hours post RCNMV inoculation (hpi). A) The microarray study indicated that host NbSI was significantly up-regulated at 12 hpi (FDR cutoff of 0.01) (red label). B) qRT-PCR analysis confirmed that host NbSI was significantly up-regulated at 12 hpi ($\alpha = 0.05$).

Table 1 Nb02987 is highly homologous to sterol isomerase from various plant species. The unigene sequence Nb02987 was searched against the GenBank non-redundant protein sequence (nr) database using BLASTX method.

Species name	BLASTX hit description	GenBank Accession number	Query coverage	E-value	Identity
<i>Solanum lycopersicum</i> (tomato)	3-beta-hydroxysteroid- $\Delta(8)$, $\Delta(7)$ -isomerase	XP_004242445.1 (complete CDS)	83%	2×10^{-105}	87%
<i>Solanum tuberosum</i> (potato)	C-8,7 sterol isomerase	NP_001275362.1 (complete CDS)	83%	1×10^{-104}	86%
<i>Vitis vinifera</i> (grape)	3-beta-hydroxysteroid- $\Delta(8)$, $\Delta(7)$ -isomerase isoform 2	XP_002285453.1 (complete CDS)	72%	2×10^{-82}	72%
<i>Theobroma cacao</i> (cocoa)	C-8,7 sterol isomerase isoform 1	XP_007050006.1 (complete CDS)	72%	7×10^{-76}	68%
<i>Zea mays</i> (corn)	sterol-8,7-isomerase	NP_001105846.1	82%	4×10^{-63}	64%

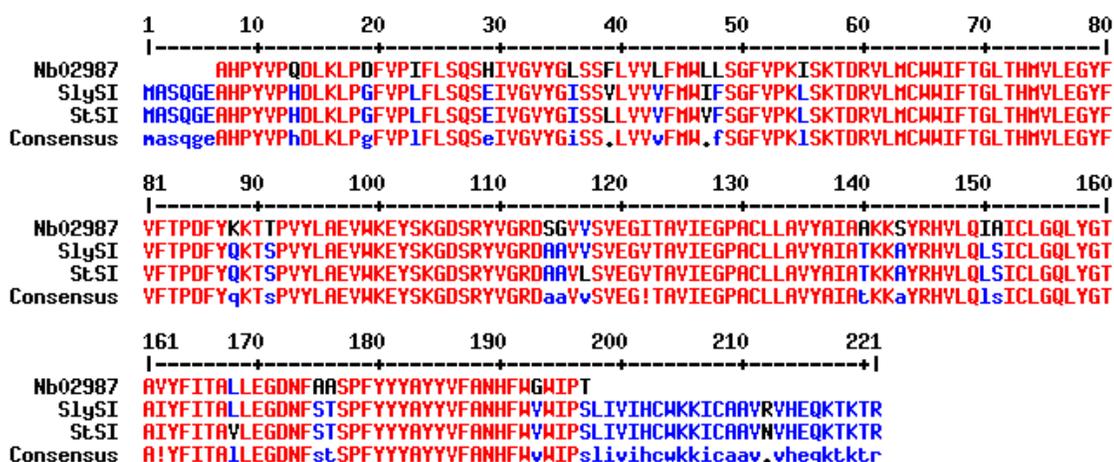


Figure 2 Nb02987 has a partial coding sequence (CDS) based on its alignment to the full-length sterol isomerase peptide sequences from tomato (SlySt, GenBank ID XP_004242445.1) and potato (StSI, GenBank ID NP_001275362.1).

```

1      10      20      30      40      50      60      70      80
|-----|-----|-----|-----|-----|-----|-----|-----|
Nb02987      AHPYVPQDLKLPDFVPIFLSQSHIVGVYGLSSFLVYLFMALLSGFVPKISKTDORVLMCHWIFITGLTHMVLEGYF
NbSI        MASQGEAHPYVPQDLRLPDFVPIFLSQSHIVGVYGLSSFLVYLFMALLSGFVPKISKTDORVLMCRWIFITGLTHMVLEGYF
Consensus    .....AHPYVPQDLrLPDFVPIFLSQSHIVGVYGLSSFLVYLFMALLSGFVPKISKTDORVLMCrWIFITGLTHMVLEGYF

81      90      100     110     120     130     140     150     160
|-----|-----|-----|-----|-----|-----|-----|-----|
Nb02987      VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDSGVVSVVEGITAVIEGPACLLAVYAIARAKKSYRHLVQIATICLGQLYGT
NbSI        VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDSGVVSVVEGITAVIEGPACLLAVYAIARAKKSYRHLVQIATICLGQLYGT
Consensus    VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDSGVVSVVEGITAVIEGPACLLAVYAIARAKKSYRHLVQIATICLGQLYGT

161     170     180     190     200     210     221
|-----|-----|-----|-----|-----|-----|
Nb02987      RYVFITALLEGDNFAASPFYYYAYVVFANHFHGHIPIT
NbSI        RYVFITALLEGDNFAASPFYYYAYVVFANHFHWIPI TLIVIHCKKKICAAVKVHEQKTKTR
Consensus    RYVFITALLEGDNFAASPFYYYAYVVFANHFHGHIPIT.....

```

Figure 3 Amino acid sequence comparison between Nb02987 and its full-length protein (NbSI).

```

1      10      20      30      40      50      60      70      80
|-----|-----|-----|-----|-----|-----|-----|-----|
Nb02987      AHPYVPQDLKLPDFVPIFLSQSHIVGVYGLSSFLVYLFMALLSGFVPKISKTDORVLMCHWIFITGLTHMVLEGYF
NbSI        MASQGEAHPYVPQDLRLPDFVPIFLSQSHIVGVYGLSSFLVYLFMALLSGFVPKISKTDORVLMCRWIFITGLTHMVLEGYF
SlySI       MASQGEAHPYVPQDLKLPDFVPIFLSQSEIVGVYGLSSVLYVVFHWIFSGFVPKISKTDORVLMCHWIFITGLTHMVLEGYF
StSI        MASQGEAHPYVPQDLKLPDFVPIFLSQSEIVGVYGLSSLLVYVFMWVFSGFVPKISKTDORVLMCHWIFITGLTHMVLEGYF
Consensus    masqgeAHPYVPqDLkLPdFVPiFLSQShIVGVYGLSSfLVYlFMWllSGFVPKiSKTDORVLMChWIFITGLTHMVLEGYF

81      90      100     110     120     130     140     150     160
|-----|-----|-----|-----|-----|-----|-----|-----|
Nb02987      VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDSGVVSVVEGITAVIEGPACLLAVYAIARAKKSYRHLVQIATICLGQLYGT
NbSI        VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDSGVVSVVEGITAVIEGPACLLAVYAIARAKKSYRHLVQIATICLGQLYGT
SlySI       VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDRAAVSVVEGVTAVIEGPACLLAVYAIATKKAYRHLVQLSICLGQLYGT
StSI        VFTPDFYKKTTPVYLAEVNKEYSKGDSRYVGRDRAAVSVVEGVTAVIEGPACLLAVYAIATKKAYRHLVQLSICLGQLYGT
Consensus    VFTPDFYkKtTpVYLAEVNKEYSKGDSRYVGRDsgVsvVEG!TAVIEGPACLLAVYAIARaKkSYRHLVQiaICLGQLYGT

161     170     180     190     200     210     221
|-----|-----|-----|-----|-----|-----|
Nb02987      RYVFITALLEGDNFAASPFYYYAYVVFANHFHGHIPIT
NbSI        RYVFITALLEGDNFAASPFYYYAYVVFANHFHWIPI TLIVIHCKKKICAAVKVHEQKTKTR
SlySI       RIYFITALLEGDNFSTSPFYYYAYVVFANHFHWIPI SLIVIHCKKKICAAVRVHEQKTKTR
StSI        RIYFITAVLEGDNFSTSPFYYYAYVVFANHFHWIPI SLIVIHCKKKICAAVNVHEQKTKTR
Consensus    R!YFITALLEGDNFaaSPFYYYAYVVFANHFHwIPItIivihcukkiCaav.vheqtktr

```

Figure 4 Amino acid sequence comparison between Nb02987, a full-length *N. benthamiana* sterol isomerase (NbSI), a full-length tomato sterol isomerase (SlySI, Genbank ID XP_004242445.1) and a full-length potato sterol isomerase (StSI, GenBank ID NP_001275362.1).

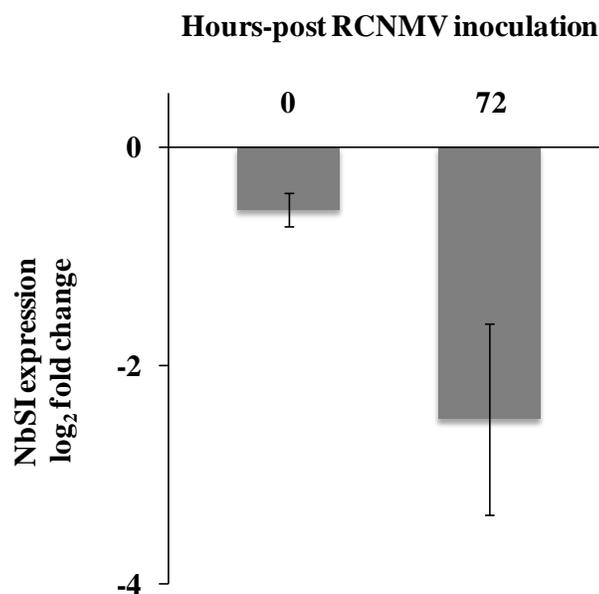


Figure 5 NbSI gene expression was suppressed in *N. benthamiana* using TRV-VIGS vector system. The silencing vector TRV1:TRV2-NbSI was delivered into plant cells by an agro-infiltration method. NbSI was pre-silenced 10 days prior to RCNMV inoculation. The pre-silenced NbSI plants were inoculated with RNA transcripts from plasmids RC169 and RC2, and harvested at 0 and 72 hours post RCNMV inoculation (hpi). The qRT-PCR analysis showed that the expression of NbSI significantly remained suppressed until 72 hours ($\alpha = 0.05$). Plants agro-infiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis.

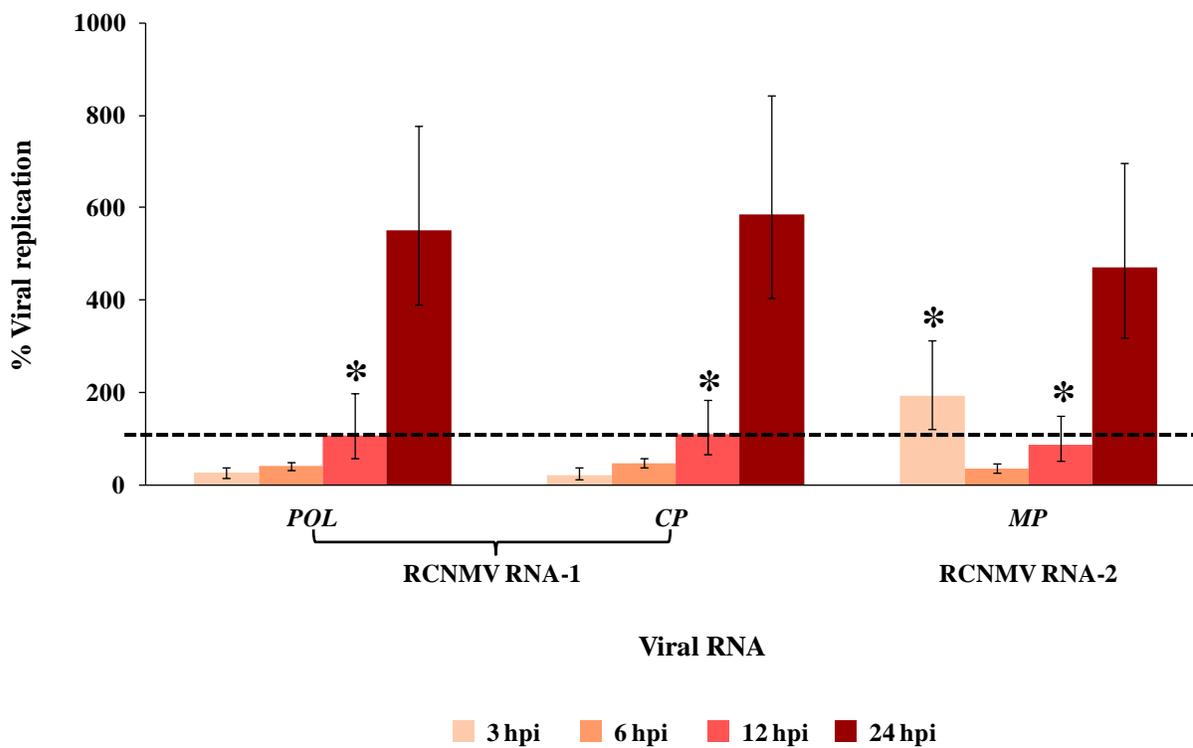


Figure 6 Pre-silenced NbSI suppressed RCNMV RNA-1 and RNA-2 accumulation in the first 6 hours, followed by a promotion at 24 hpi. The expression of NbSI was transiently silenced by TRV1:TRV2-NbSI (TRV-VIGS system) 10 days prior to RCNMV inoculation. The pre-silenced NbSI plants were inoculated with RNA transcripts from RC169 and RC2, and then harvested between 3 and 24 hours after RCNMV inoculation. RCNMV RNA levels were measured using qRT-PCR analysis. The viral RNA accumulation was computed by a relative quantification method and displayed as % viral replication. Plants agro-infiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. All data shown in this figure was significant at $\alpha = 0.1$, except the data with an asterisk symbol (*).

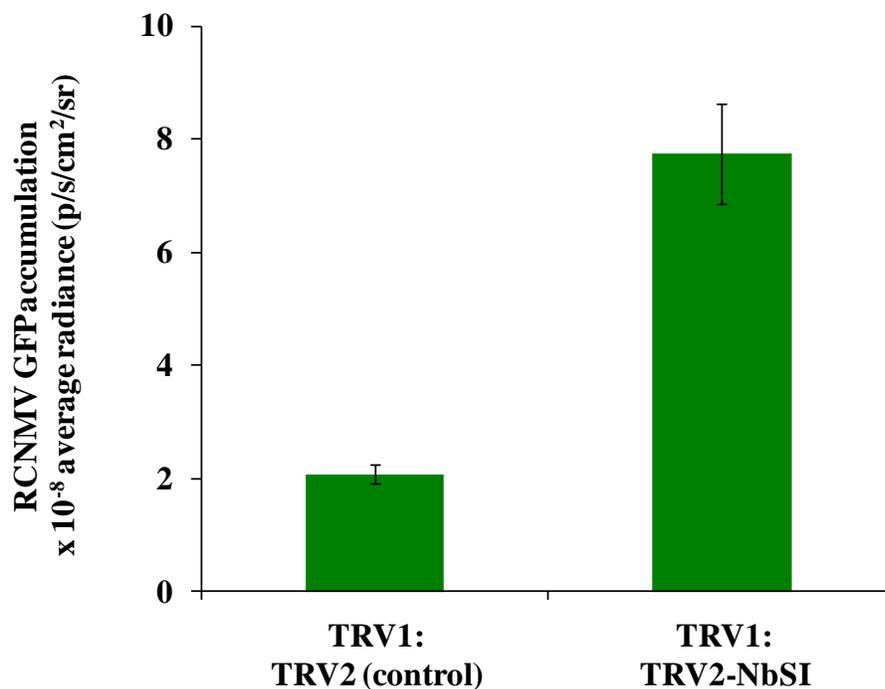


Figure 7 Pre-silenced NbSI promoted RCNMV GFP accumulation > 4 fold at 3 days post inoculation. The expression of NbSI was transiently silenced using the TRV-VIGS system 10 days prior to RCNMV inoculation. The pre-silenced NbSI plants were inoculated with RNA transcripts from R1SG1 and RC2, and then harvested 3 days after inoculation. Plants agro-infiltrated with TRV1:TRV2 10 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis ($\alpha = 0.05$). The sGFP quantification was performed with an IVIS imaging system.

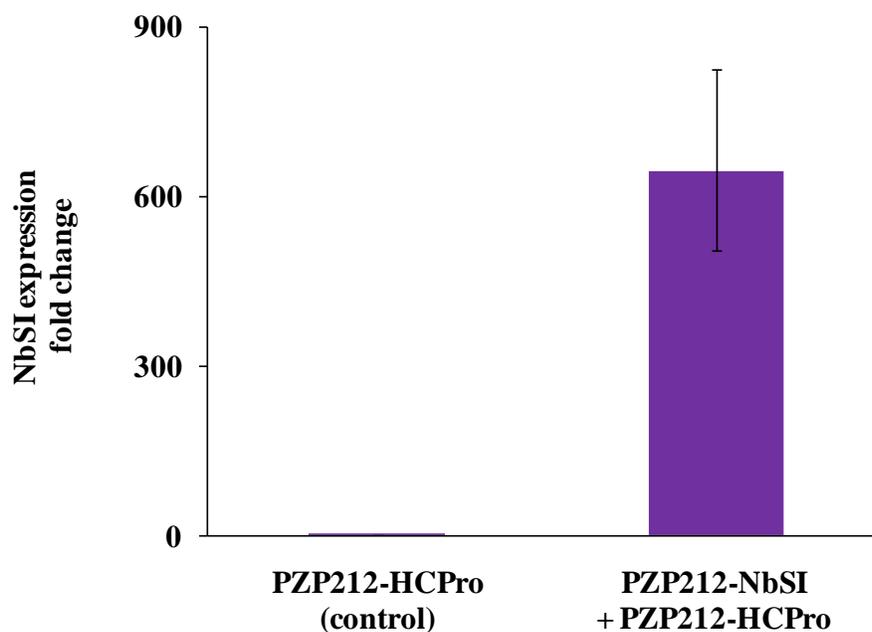


Figure 8 NbSI gene was overexpressed in *N. benthamiana* by using PZP212 and pRTL2 expression vector system. The overexpression construct PZP212-NbSI was delivered into plant cells along with the RNA silencing suppressor construct PZP212-HCPro by an agroinfiltration method to constitutively express the NbSI gene. The leaf samples were harvested 2 days after agroinfiltration. Plants agroinfiltrated with only PZP212-HCPro were used as the baseline controls in a student's t-test statistical analysis. The qRT-PCR analysis showed that the expression of NbSI gene was significantly increased > 600 fold ($\alpha = 0.01$).

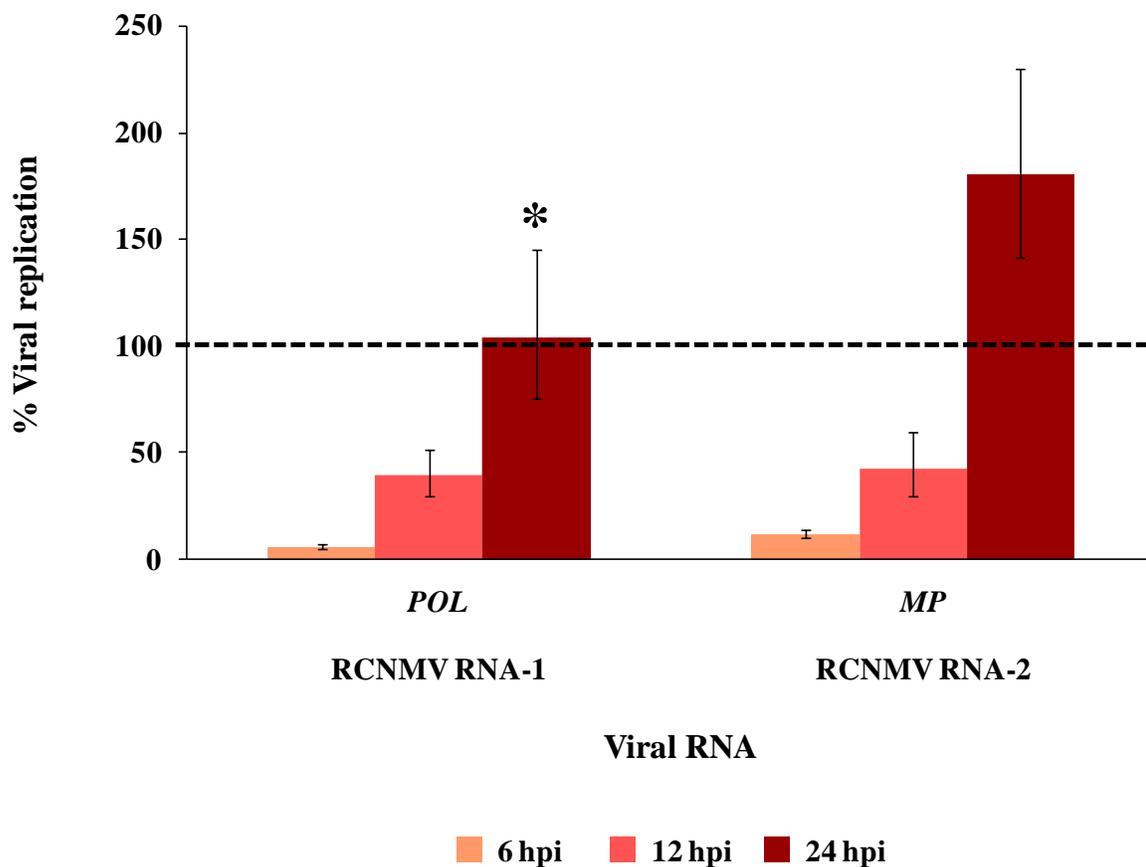


Figure 9 Pre-overexpressed NbSI suppressed RCNMV RNA accumulation in the first 12 hours of infection. The expression of NbSI was transiently overexpressed by agro-infiltration with PZP212-NbSI along with the RNA silencing suppressor (PZP212-HCPro) 2 days prior to RCNMV inoculation. The pre-overexpressed NbSI plants were inoculated with RNA transcripts derived from RC169 and RC2, and then harvested between 6 and 24 hours post RCNMV inoculation (hpi). RCNMV RNA levels were measured via qRT-PCR analysis. The viral RNA accumulation was computed by a relative quantification method, and displayed as % viral replication. Plants agro-infiltrated with only PZP212-HCPro 2 days prior to RCNMV inoculation were used as the baseline controls in a student's t-test statistical analysis. All data shown in this figure was significant at $\alpha = 0.05$, except the data with an asterisk symbol (*).

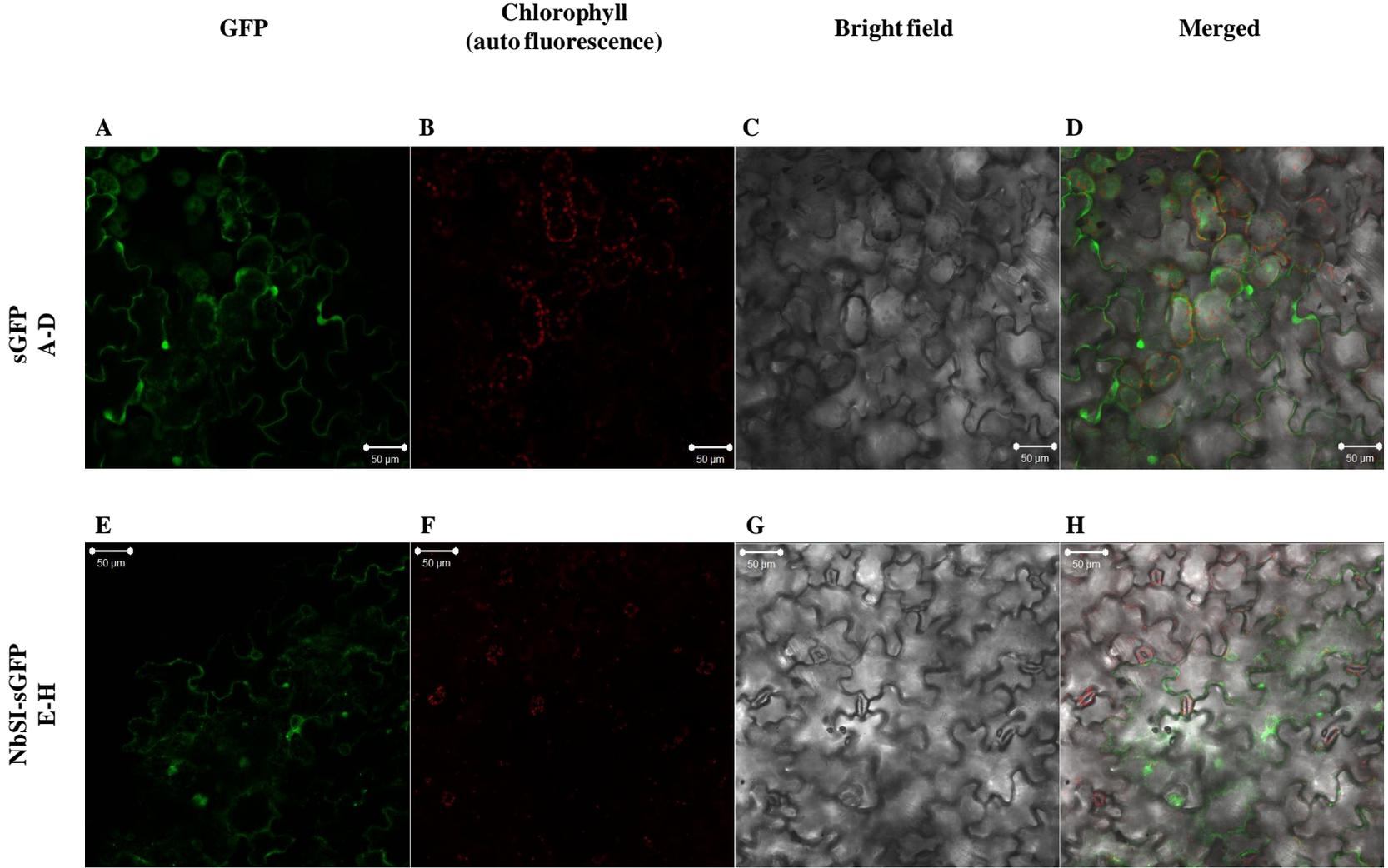
Figure 10 Membrane localization of NbSI protein. A modified green fluorescent protein (sGFP) was used as a visual reporter to identify the subcellular localization of the full-length NbSI protein. The full-length NbSI protein was fused at its C-terminus to the sGFP tagging protein to generate NbSI-sGFP fusion protein. The localization constructs were agroinfiltrated into *N. benthamiana* leaves and the photographs were taken 2 days after agroinfiltration with a confocal microscope. The images were scanned using three different filters (green = GFP, red = chloroplast auto fluorescence, and white/black = bright field).

(A-D) shows the typical pattern of sGFP subcellular localization (control). Images were taken from a leaf agro-infiltrated with a mixture of PZP212-sGFP and PZP212-HCPro. A merged image of sGFP (D) shows that the fluorescence signal accumulated inside the nucleus, cytoplasm, and along a cellular membrane. However, sGFP does not appear to localize to the chloroplast.

(E-H) shows the characteristic pattern of NbSI-sGFP subcellular localization. Images were taken from a leaf agro-infiltrated with a mixture of PZP212-NbSI-sGFP and PZP212-HCPro (silencing suppressor). The merged image of NbSI-sGFP (H) indicates that a fluorescence signal was exclusively concentrated along a cellular membrane, a nuclear membrane and the endoplasmic reticulum membrane.

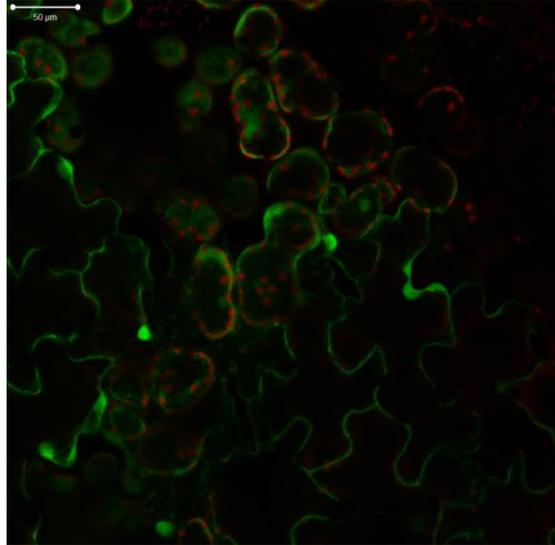
(I) shows a larger image of sGFP, confirming its localization in the nucleus, cytoplasm and along cellular membranes.

(J and K) show images of NbSI-sGFP, confirming its localization along the cellular membrane, the nuclear membrane, and the endoplasmic reticulum membrane.

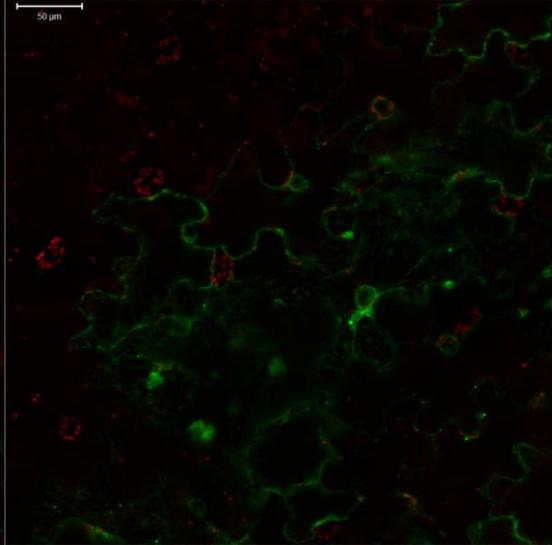


Overlay image between GFP and chloroplast autoflorescence filters

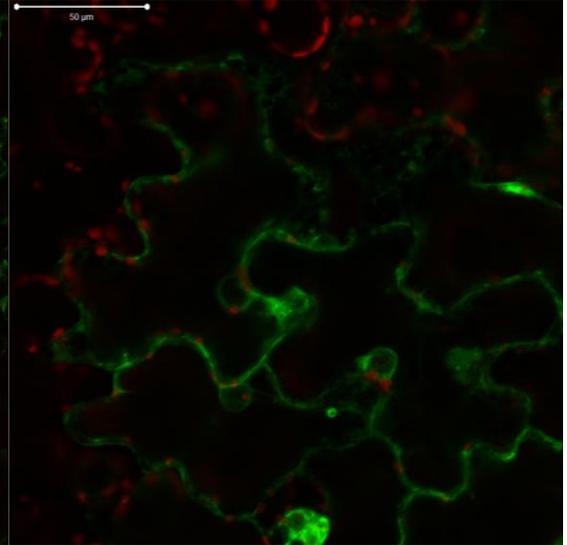
D) sGFP



J) NbSI-sGFP



K) NbSI-sGFP



A)

Residue position from N- to C- terminus of NbSI	Protein segment
1-29	outside
30-52	transmembrane helix
53-60	inside
61-83	transmembrane helix
84-117	outside
118-140	transmembrane helix
141-146	inside
147-169	transmembrane helix
170-178	outside
179-201	transmembrane helix
202-224	inside

B)

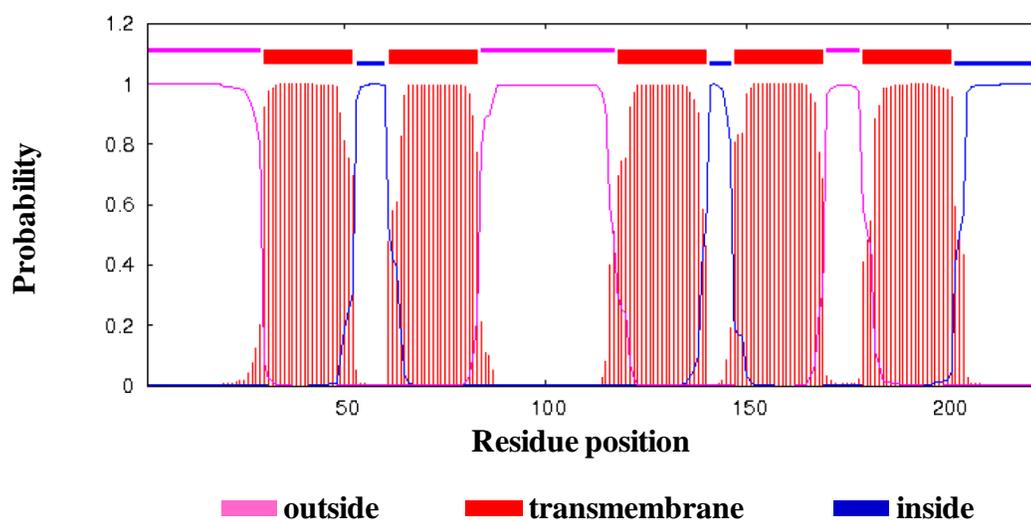


Figure 11 NbSI protein is predicted to contain five α -helical transmembrane domains. The prediction was performed with the TMHMM2.0 bioinformatic tool. (A) and (B) TMHMM predicts the orientation of NbSI residues (segment oriented inside, outside or transmembrane).

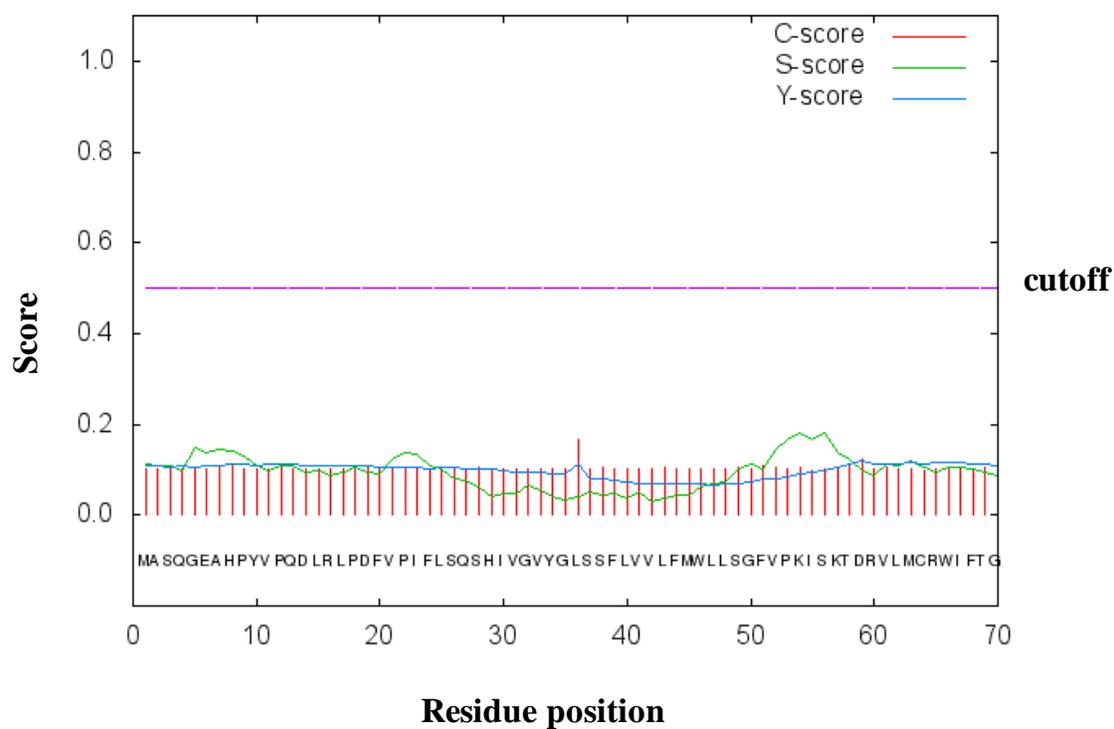


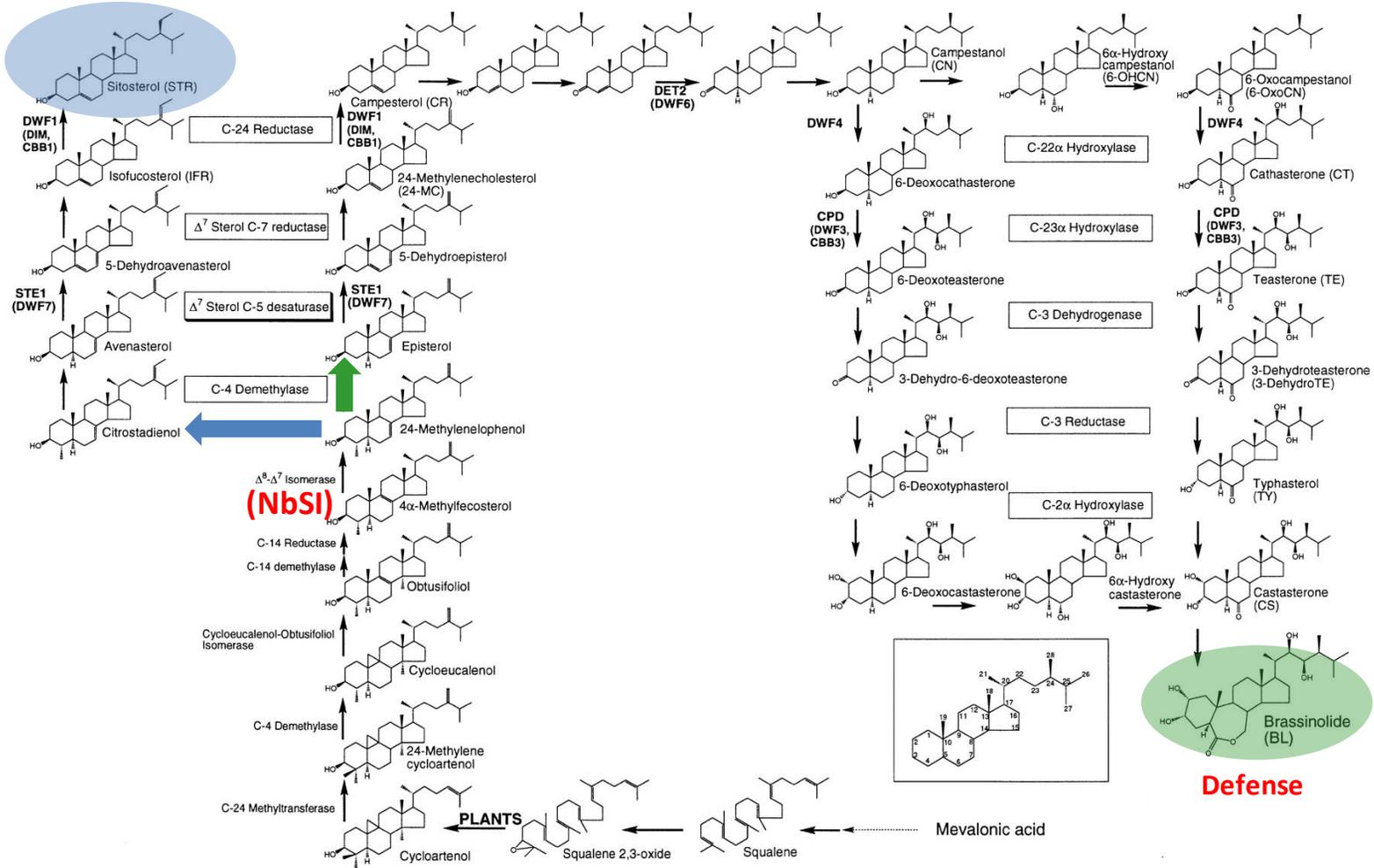
Figure 12 NbSI does not contain a signal peptide. The prediction was performed by using SignalP-4.1 bioinformatic tool. The SignalP-4.1 did not find any putative cleavage site or any putative signal peptide in the first 70 amino acids at the N-terminus of the NbSI protein.

Figure 13 Sterol-specific biosynthetic pathway and Brassinosteroid (BR)-specific biosynthetic pathway [12]. Red label indicates a potential function of Nb02987 (NbSI) as a C-8, 7 sterol isomerase. Common names for the compounds are labeled and proposed enzymes involved in each reaction are boxed and labeled. Genes identified by mutants are marked. The acronyms for some compounds are in parentheses. In the inset, the carbon atoms of the sterol core rings and side chain are numbered.

Sterol-Specific pathway

Brassinosteroid (BR) -Specific pathway

Lipid membrane



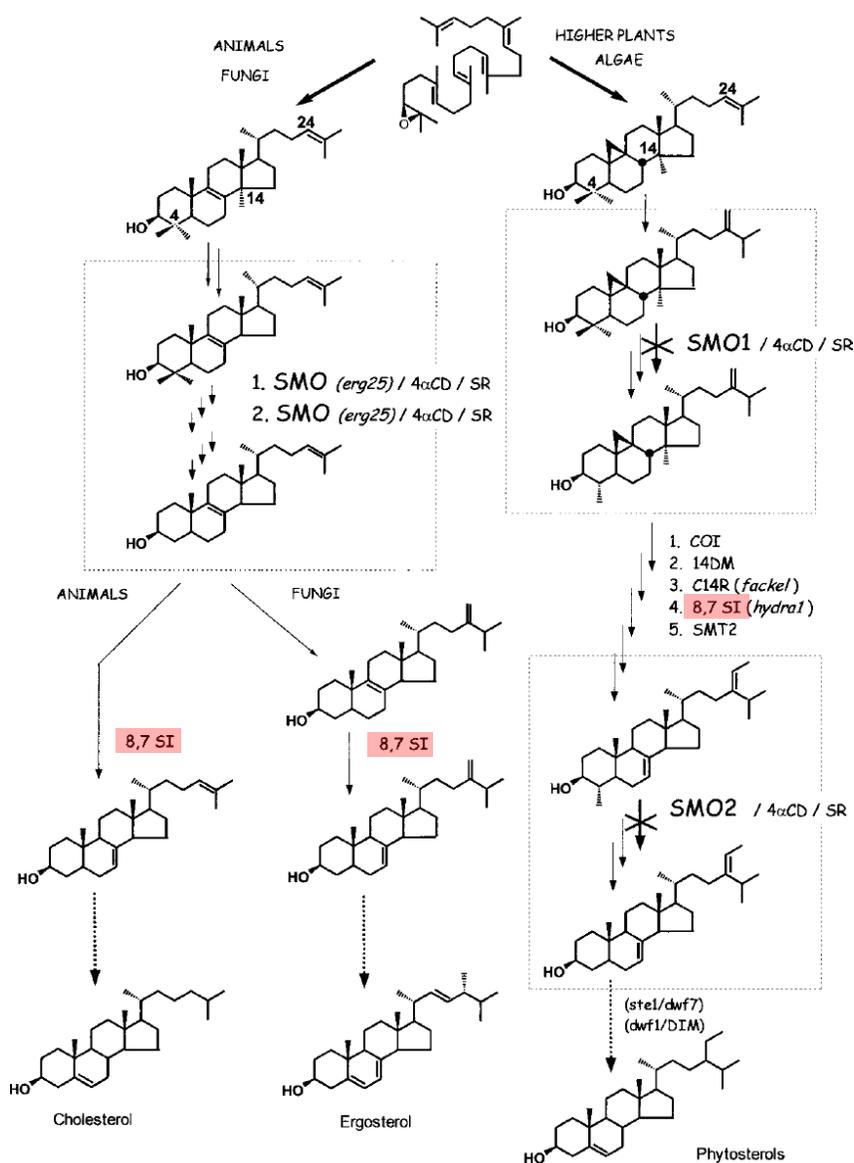


Figure 14 The sterol biosynthetic pathway distinctions in animals, fungi, higher plants and algae [9]. Red box indicates a potential function of Nb02987 (NbSI) as C-8, 7 sterol isomerase in different organisms. Full arrows represent distinct enzymes and arrows with broken lines represent several biosynthetic steps. The boxes represent C-4-demethylation multienzymic complexes, including SMO, 4 α -CD and SR. Abbreviations: COI, cycloeucaleanol isomerase; 14DM, obtusifoliol 14-demethylase; C14R, $\Delta^{8,14}$ -sterol C-14 reductase; 8,7 SI, Δ^8 -sterol 8,7-isomerase; SMT2, sterol methyltransferase 2.

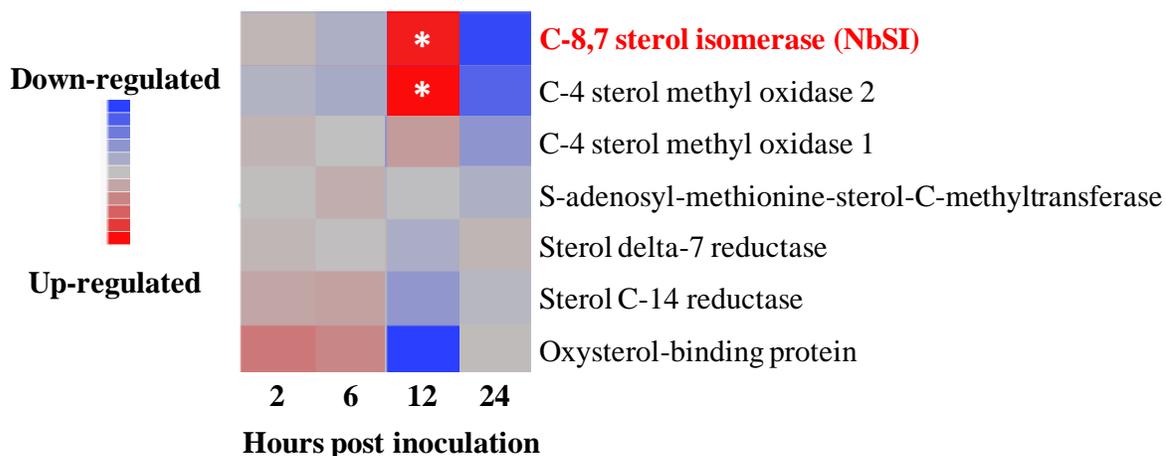


Figure 15 Heat map displays the modulation of *N. benthamiana* sterol biosynthesis-related gene expression at the early stages of the RCNMV infection from the microarray study (Chapter 3). A differential gene expression was computed from a subtraction between RCNMV-infected plants and mocks at 2, 6, 12 and 24 hours-post inoculation (hpi). Asterisk (*) indicates a significant differentially expressed host gene at FDR cutoff of 0.01. Nb02987 (NbSI) was computationally annotated with the function of C-8, 7 sterol isomerase (red label).

Table 2 Oligonucleotide information used in Chapter 7.

Name	Sequence information
Nb02987F	ATATTCCTCTCACAGTCACAC
Nb02987R	GTAAAGACAAAATACCCCTC
GS02987F	AATTTCTAGAGCTCGACTCAGATTCGCTATTACAG
GS02987R	AATTTCTAGAGCTCAGTAGCCACATGAACAAGAC
FL02987F	AATTCATGGGATCGATGGCAAGCCAGGGAGAAGC
FL02987R	AATTTCTAGATTACATCGATGCGCGGGTTTTGGTCTTCTGCTCG
3'sGFP/XbaI	GCTCTAGACGCGTACTTGTACAGCTCGTCC
RNA1-POLF	TCACAAGGGTCAAATTCTCAAATCCT
RNA1-POLR	TGCTGCTTTTTGGTATAACTTCCTCTT
RNA1-CPF	ATGTCTTCAAAGCTCCCAA
RNA1-CPR	CGCTCATGACTAACTGGGTA
RNA2F	CGCGTCTGATTGAGTTGGAAGTA
RNA2R	CTGCCTTGATGCTCGACAGTA
TRV2F	TACTCAAGGAAGCACGATGAGC
TRV2R	GAACCGTAGTTTAATGTCTTCGGG
TEVLeader80	CGAATCTCAAGCAATCAAGCATTC
35sTERM	ATAAGAACCCTAATTCCC
ActF	GTGACCTCACTGATAGTTTGA
ActR	TACAGAAGAGCTGGTCTTTG

Materials and Methods

1) Plants, RCNMV inoculum preparation and inoculation procedure

N. benthamiana was a plant model used in this experiment. Details of plant maintenance, RCNMV inoculum preparation and inoculation procedure were described in Chapter 5 materials and methods. Briefly, RCNMV inoculum was prepared in a 110 μ l volume per 1 plant. This 110 μ l inoculum was a mixture of 1 μ l *in vitro* T7 RNA-1 transcript (or 1 μ l *in vitro* T7 R1SG1 transcript), 1 μ l *in vitro* T7 RNA-2 transcript and 108 μ l inoculation buffer (10 mM sodium diphosphate, pH 7.2). Four leaves per 1 plant were inoculated with either RCNMV inoculum or buffer. 27 μ l of the RCNMV transcript mixture (or inoculation buffer) was pipetted onto each leaf and mechanically rubbed with carborundum (abrasive). Inoculated plants were maintained in a temperature and light controlled environment at 18-22°C, 16 hour-light and 8 hour-dark period at the NCSU greenhouses (Method Road).

2) Plasmid DNA constructs

A construction of the silencing construct, the overexpression construct, and the subcellular localization construct are performed accordingly to detail in Chapter 5 materials and methods. Following is a brief detail and the modification.

The silencing construct TRV2-NbSI utilizing the TRV VIGS system [30] was designed to down regulate NbSI expression in *N. benthamiana*. A 196 bp fragment of the NbSI gene was amplified accordingly to OneTaq DNA polymerase protocol (NEBTM), with a

primer set of GS02987F and GS02987R (Table 2) to generate the GS-NbSI. The PCR conditions for amplifying GS-NbSI were as follows: 1) a denaturation cycle of 94°C for 30 sec, 2) 30 amplification cycles of 94°C for 30 sec, 55°C for 30 sec and 68°C for 1 min, and 3) a final extension cycle of 68°C for 5 min. A ligation between GS-NbSI and TRV2 was performed via *SacI* restriction site. A colony PCR technique using the primer set of TRV2F and TRV2R (Table 2) was used to identify TRV2 clones that contained the GS-NbSI insert. A positive plasmid was also confirmed by digesting with *SacI*. The resultant construct TRV2-NbSI was used for down regulating NbSI expression in *N. benthamiana* for a gene silencing study.

The PZP212-NbSI construct was designed for overexpression of NbSI in *N. benthamiana* and is based on the pRTL2 expression cassette and the PZP212 binary vector. A full-length NbSI DNA (663 bp) was amplified accordingly to OneTaq DNA polymerase protocol (NEB™), with a primer set of FL02987F and FL02987R (Table 2) and the extension cycle at 68°C for 1 min 30 sec. A full-length NbSI DNA was firstly cloned into pGEM T-Easy in order to increase a copy number. A ligation between a full-length NbSI DNA and pRTL2 was performed via *NcoI* and *XbaI* restriction sites. The pRTL2 construct containing the full-length NbSI was then verified by sequencing using sequencing primers TEVleader80 and 35sTERM (Table 2). The resultant plasmid pRTL2-NbSI was digested with *HindIII* to release the insert containing the full-length NbSI expression cassette. This isolated fragment was subsequently cloned into PZP212 via *HindIII* restriction site. The PZP212-NbSI construct was used to constitutively express the putative full-length NbSI protein in *N. benthamiana* for a gene overexpression study.

The PZP212-NbSI-sGFP construct was designed for subcellular localization study of NbSI. A full-length NbSI DNA was initially cloned into the RNA transcript expression vector pHST2-sGFP through the unique *ClaI* restriction site. The orientation of the clone was confirmed by a restriction analysis and the properly orientated construct was used as a template to amplify the NbSI-sGFP product using a primer set of FL02987F and 3'sGFP/*XbaI*(Table 2). The PCR product was then cleaved with *NcoI* and *XbaI* and ligated into similarly digested pRTL2 to produce construct pRTL2-NbSI-sGFP. The nature of the construct was verified by sequencing and the expression cassette containing NbSI-sGFP was released by digestion with *HindIII* and subsequently cloned into PZP212 to yield construct PZP212-NbSI-sGFP.

3) *Agrobacterium* preparation and infiltration

Agrobacterium was prepared accordingly to a detail in Chapter 5 materials and methods. Briefly, *Agrobacterium* based constructs were transformed into *Agrobacterium tumefaciens* strain C58C1. Individual colonies were inoculated into 2 ml LB broth cultures with the appropriate antibiotics and incubated at 28°C for 20 hr with shaking. From these initial cultures 250 µl was used to inoculate 5 ml LB broth cultures with the appropriate antibiotics and 40µM acetosyringone/10mM MES, Ph 5.6. These cultures were similarly incubated at 28°C for 20 hr with shaking. Cultures were subsequently pelleted, resuspended in 10mM MgCl₂/10mM MES, pH 5.6/200µM acetosyringone to the appropriate OD600 reading and incubated at room temperature for at least 3 hr prior to syringe infiltration into plants.

For NbSI silencing experiments, *Agrobacterium* cultures of TRV1 and TRV2-NbSI (or TRV2 for the control) were mixed at a 1:1 ratio immediately prior to infiltration into plants. For NbSI overexpression experiments, *Agrobacterium* cultures of PZP212-NbSI and the RNA silencing suppressor PZP212-HCPro (or PZP212-HCPro alone for the control) were mixed at a ratio of 1:1 immediately prior to infiltration into plants. For subcellular localization experiments, *Agrobacterium* cultures of PZP212-NbSI-sGFP (or PZP212-sGFP for the control) and the RNA silencing suppressor PZP212-HCPro were mixed at a ratio of 1:1 immediately prior to infiltration into plants. Leaves infiltrated for the subcellular localization experiments were harvested 2 days after agroinfiltration and imaged via confocal microscopy (Cellular and Molecular Imaging Facility, NCSU).

4) Total RNA extraction and cleanup

Leaves to be sampled in each experiment consisted of the following: four plants and two leaves per plant were used to represent the test condition (silencing or overexpression of NbSI) as well as the respective control (see above). Total RNA extracts isolated from the same plant were pooled to represent one biological replicate (4 biological replicates for the test condition and 4 biological replicates for controls). Leaf samples were collected either 2 days (overexpression experiments) or 10 days (silencing experiments) after agroinfiltration as well as various time points after RCNMV transcript inoculation for both types of experiments.

A total RNA extraction and cleanup were performed accordingly to details in Chapter 5 materials and methods. Briefly, total RNA was extracted from 100 mg *N. benthamiana* leaf

tissue samples according to the TRIzol protocol (Invitrogen™). And total RNA extracts destined for qRT-PCR analysis were treated with Turbo DNA-free (Ambion™) according to the manufacturer's protocol to remove any residual input DNA. The samples were stored at -20°C.

5) RNA quantification by real-time PCR (qRT-PCR)

First strand cDNA synthesis and quantitative real-time PCR (qRT-PCR) preparation were performed according to details in Chapter 5 material and methods. Briefly, First strand cDNA was synthesized by using a DNA-free total RNA (equivalent to ~1 µg total RNA) as a template. SYBR Green-based quantitative real-time PCR (qRT-PCR) method was used in this study. The qRT-PCR reaction was prepared accordingly to FastStart Universal SYBR Green Master ROX protocol (Roche™). The first strand cDNA equivalent to 40 ng total RNA was used to prepare qRT-PCR reaction. The qRT-PCR reactions were setup in a 384 well plate and placed in ABI7900 HT Fast real-time PCR system (Applied Biosystems™). The Applied Biosystems SDS software version 2.4 was used to monitor and dissect real-time PCR data. Three qRT-PCR reactions were prepared per 1 biological replicate. All PCR products are less than 200 base pairs. The cycling condition was set as follows: 1 cycle at 95°C for 10 min, and 40 cycles at 95°C for 15 sec /55°C for 1 min. An association analysis was also performed in order to test if there were any non-specific PCR products formed. The condition for the association analysis was set as follows: 1 cycle at 95°C for 15 sec, 60°C for 15 sec and 95°C for 15 sec.

NbSI transcript levels were assayed using primers Nb02987F and Nb02987R (Table 2). RCNMV RNA levels were assayed with 2 primer pairs for RNA-1 and 1 primer pair for RNA-2 (Table 2): 1) RNA1-POLF and RNA1-POLR for probing RNA-1 at the polymerase (RNA-1 POL), 2) RNA1-CPF and RNA1-CPR for probing RNA-1 at the coat protein (RNA-1 CP) and 3) RNA2F and RNA2R for probing RNA-2 at the movement protein (RNA-2 MP).

6) qRT-PCR data analysis

A relative quantification, $2^{-\Delta\Delta C_t}$ method (equation shown in Chapter 5 materials and methods) [47] was used to analyze the real-time PCR data. Expression of the target gene was normalized against the expression of a reference gene. Actin was used as the reference genes to normalize the target gene that was performed within the same plate. Student's t-test statistical analysis was used to examine the real-time PCR data.

7) RCNMV sGFP quantification

N. benthamiana plants were inoculated with a combination of *in vitro* T7 R1SG1 and RNA-2 transcripts. Ten plants and two leaves per plant were used for each test condition and each control. Leaf samples were harvested 3 days post inoculation. The GFP fluorescence intensity was analyzed by the IVIS imaging system and subjected to Student's t-test statistical analysis. The IVIS Lumina System (Xenogen Corporation, Alameda, CA) is capable of quantifying photon emission from a variety of sources. A CCD camera measured and recorded photon emission data which was then incorporated into Living Image Software

(Xenogen Corp.) for further analysis. Whole leaves were placed under the CCD camera and measurements were taken with a GFP excitation filter, and exposure time of 1 second.

References

1. Ahlquist P, Noueir AO, Lee W-M, Kushner DB, Dye BT: **Host factors in positive-strand RNA virus genome replication.** *Journal of Virology* 2003, **77**(15):8181-8186.
2. Denison MR: **Seeking membranes: positive-strand RNA virus replication complexes.** *PLoS biology* 2008, **6**(10):e270.
3. Novoa RR, Calderita G, Arranz Ro, Fontana J, Granzow H, Risco C: **Virus factories: associations of cell organelles for viral replication and morphogenesis.** *Biology of the Cell* 2005, **97**(2):147-172.
4. Pogany J, Stork J, Li Z, Nagy PD: **In vitro assembly of the Tomato bushy stunt virus replicase requires the host Heat shock protein 70.** *Proceedings of the National Academy of Sciences* 2008, **105**(50):19956-19961.
5. Turner KA, Sit TL, Callaway AS, Allen NS, Lommel SA: **Red clover necrotic mosaic virus replication proteins accumulate at the endoplasmic reticulum.** *Virology* 2004, **320**(2):276-290.
6. Hartmann M-A: **Plant sterols and the membrane environment.** *Trends in plant science* 1998, **3**(5):170-175.
7. Bloch KE: **Sterol, Structure and Membrane Function.** *Critical reviews in biochemistry and molecular biology* 1983, **14**(1):47-92.
8. Grunwald C: **Plant sterols.** *Annual Review of Plant Physiology* 1975, **26**(1):209-236.
9. Darnet S, Rahier A: **Plant sterol biosynthesis: identification of two distinct families of sterol 4alpha-methyl oxidases.** *Biochem J* 2004, **378**:889-898.
10. Sharma M, Sasvari Z, Nagy PD: **Inhibition of sterol biosynthesis reduces tombusvirus replication in yeast and plants.** *Journal of Virology*, **84**(5):2270-2281.
11. Boutte Y, Grebe M: **Cellular processes relying on sterol function in plants.** *Current opinion in plant biology* 2009, **12**(6):705-713.

12. Choe S, Noguchi T, Fujioka S, Takatsuto S, Tissier CP, Gregory BD, Ross AS, Tanaka A, Yoshida S, Tax FE: **The Arabidopsis dwf7/stel1 mutant is defective in the delta-7 sterol C-5 desaturation step leading to brassinosteroid biosynthesis.** *The Plant Cell Online* 1999, **11**(2):207-221.
13. Clouse SD, Langford M, McMorris TC: **A brassinosteroid-insensitive mutant in Arabidopsis thaliana exhibits multiple defects in growth and development.** *Plant Physiology* 1996, **111**(3):671-678.
14. Yamamuro C, Ihara Y, Wu X, Noguchi T, Fujioka S, Takatsuto S, Ashikari M, Kitano H, Matsuoka M: **Loss of function of a rice brassinosteroid insensitive1 homolog prevents internode elongation and bending of the lamina joint.** *The Plant Cell Online* 2000, **12**(9):1591-1605.
15. Nakashita H, Yasuda M, Nitta T, Asami T, Fujioka S, Arai Y, Sekimata K, Takatsuto S, Yamaguchi I, Yoshida S: **Brassinosteroid functions in a broad range of disease resistance in tobacco and rice.** *The Plant Journal* 2003, **33**(5):887-898.
16. Belkhadir Y, Jaillais Y, Epple P, Balsemão-Pires E, Dangl JL, Chory J: **Brassinosteroids modulate the efficiency of plant immune responses to microbe-associated molecular patterns.** *Proceedings of the National Academy of Sciences*, **109**(1):297-302.
17. Krishna P: **Brassinosteroid-mediated stress responses.** *Journal of Plant Growth Regulation* 2003, **22**(4):289-297.
18. Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nurnberger T, Jones JD, Felix G, Boller T: **A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence.** *Nature* 2007, **448**(7152):497-500.
19. Robert-Seilaniantz A, Grant M, Jones JD: **Hormone crosstalk in plant disease and defense: more than just jasmonate-salicylate antagonism.** *Annual review of phytopathology*, **49**:317-343.
20. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL: **GenBank.** *Nucleic acids research* 2008, **36**(suppl 1):D25-D30.
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
22. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR: **CDD: a Conserved Domain Database for the functional annotation of proteins.** *Nucleic acids research*, **39**(suppl 1):D225-D229.

23. Grebenok RJ, Ohnmeiss TE, Yamamoto A, Huntley ED, Galbraith DW, Della Penna D: **Isolation and characterization of an Arabidopsis thaliana C-8, 7 sterol isomerase: functional and structural similarities to mammalian C-8, 7 sterol isomerase/emopamil-binding protein.** *Plant molecular biology* 1998, **38**(5):807-815.
24. Has C, Bruckner-Tuderman L, Muller D, Floeth M, Folkers E, Donnai D, Traupe H: **The Conradi-Hunermann-Happle syndrome (CDPX2) and emopamil binding protein: novel mutations, and somatic and gonadal mosaicism.** *Human molecular genetics* 2000, **9**(13):1951-1955.
25. Derry JM, Gormally E, Means GD, Zhao W, Meindl A, Kelley RI, Boyd Y, Herman GE: **Mutations in a delta 8 - delta 7 sterol isomerase in the tattered mouse and X-linked dominant chondrodysplasia punctata.** *Nature genetics* 1999, **22**(3):286-290.
26. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT: **Gene Ontology: tool for the unification of biology.** *Nature genetics* 2000, **25**(1):25-29.
27. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A: **ExPASy: the proteomics server for in-depth protein knowledge and analysis.** *Nucleic acids research* 2003, **31**(13):3784-3788.
28. HongBin N, RunE B, DeZhi D, Jun Y: **Cloning and expression of C-8, 7 sterol isomerase gene (StSI1) cDNA from potato.** *Acta Horticulturae Sinica* 2009, **36**(4):521-526.
29. Ratcliff F, Martin-Hernandez AM, Baulcombe DC: **Technical advance: Tobacco rattle virus as a vector for analysis of gene function by silencing.** *The Plant Journal* 2001, **25**(2):237-245.
30. Dinesh-Kumar SP, Anandalakshmi R, Marathe R, Schiff M, Liu Y: **Virus-induced gene silencing.** In: *Plant Functional Genomics*. Springer; 2003: 287-293.
31. Tsien RY: **The green fluorescent protein.** *Annual review of biochemistry* 1998, **67**(1):509-544.
32. Palmer E, Freeman T: **Investigation into the use of C- and N- terminal GFP fusion proteins for subcellular localization studies using reverse transfection microarrays.** *Comparative and functional genomics* 2004, **5**(4):342-353.
33. Chiu W-l, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J: **Engineered GFP as a vital reporter in plants.** *Current Biology* 1996, **6**(3):325-330.

34. Seibel NM, Eljouni J, Nalaskowski MM, Hampe W: **Nuclear localization of enhanced green fluorescent protein homomultimers.** *Analytical biochemistry* 2007, **368**(1):95-99.
35. Wang R, Brattain MG: **The maximal size of protein to diffuse through the nuclear pore is larger than 60kDa.** *FEBS letters* 2007, **581**(17):3164-3170.
36. Ribbeck K, Gorlich D: **Kinetic analysis of translocation through nuclear pore complexes.** *The EMBO journal* 2001, **20**(6):1320-1330.
37. Paine PL, Moore LC, Horowitz SB: **Nuclear envelope permeability.** *Nature* 1975, **254**(5496):109-114.
38. Yu CS, Chen YC, Lu CH, Hwang JK: **Prediction of protein subcellular localization.** *Proteins: Structure, Function, and Bioinformatics* 2006, **64**(3):643-651.
39. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier C, Nakai K: **WoLF PSORT: protein localization predictor.** *Nucleic acids research* 2007, **35**(suppl 2):W585-W587.
40. Chou K-C, Shen H-B: **Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization.** *PloS one* 2010, **5**(6):e11335.
41. Krogh A, Larsson Br, Von Heijne G, Sonnhammer EL: **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *Journal of molecular biology* 2001, **305**(3):567-580.
42. Sonnhammer EL, Von Heijne G, Krogh A: **A hidden Markov model for predicting transmembrane helices in protein sequences.** In: *Ismb: 1998*; 1998: 175-182.
43. Moller S, Croning MD, Apweiler R: **Evaluation of methods for the prediction of membrane spanning regions.** *Bioinformatics* 2001, **17**(7):646-653.
44. Heijne G: **Signal Peptides.** *eLS* 1990.
45. Blazquez M, Shennan KI: **Protein Secretory Pathways.** *eLS* 2006.
46. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nature methods* 2011, **8**(10):785-786.
47. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic acids research* 2001, **29**(9):e45-e45.