ABSTRACT

FEULNER, GRETCHEN. The Modulation of *Nicotiana benthamiana* Gene Expression by *Red Clover Necrotic Mosaic Virus*. (Under the direction of Dr. Steven A. Lommel.)

Nicotiana benthamiana, a member of the Solanaceae family, is an amphidiploid species with n=19 chromosomes. It is the premier host for the maintenance and study of plant virus/host interactions. Not only is it susceptible to a large number of monocot and dicot infecting RNA plant viruses, but it has also been shown to serve as a host for an RNA insect virus. In collaboration with NimbleGen Systems, Inc (Madison, WI), we developed and validated a microarray from *N. benthamiana* expressed sequence tags (ESTs). The array is composed of oligonucleotides (60mers) corresponding to 13,014 *N. benthamiana* ESTs, equivalent to an estimated coverage of approximately 38% of the *N. benthamiana* transcriptome.

The majority of previous studies that have investigated global plant host gene expression changes during a viral infection have used the Affymetrix *Arabidopsis thaliana* gene chip. Additionally, only one of these studies has considered changes that occur within the first 24 hours after inoculation. We are interested in the changes that occur in *N. benthamiana* early in the infection process by an RNA plant virus. Our goal was to study these early changes by challenging *N. benthamiana* with *Red clover necrotic mosaic virus* (RCNMV). RCNMV, a member of the *Dianthovirus* genus, is a bipartite, single stranded, positive sense RNA virus. Total RNA, from mock and RCNMV infected *N. benthamiana* was isolated from leaves at 2, 6, 12 and 24 hours post inoculation

(hpi) and hybridized to the *N. benthamiana* microarrays. Statistical analysis determined a total of 1,775 ESTs exhibited differential expression for at least one timepoint in response to infection by RCNMV. As determined in other array experiments employing plant viruses, the preponderance of these genes are related to those known or predicted to be involved in pathogen defense. Key host pathways affected by the infection, as determined by Gene Ontology functional classifications, include metabolism, transport, membrane/cell wall associated, and defense-related genes. We hypothesize that the observed suppression and induction of these genes relates directly to the host factors required by the virus at each stage of its life cycle. This is the first report of a *N. benthamiana* microarray used to monitor host gene expression profiles upon virus infection at a sub 24 hpi timeframe.

The Modulation of *Nicotiana benthamiana* Gene Expression by *Red clover necrotic mosaic virus*

by

Gretchen Elizabeth Feulner

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

GENETICS

Raleigh, North Carolina

2007

APPROVED BY:

Steven A. Lommel Chair of Advisory Committee	Gregory C Gibson Advisory Committee
Barbara Sherry Advisory Committee	Jose M. Alonso Advisory Committee

BIOGRAPHY

Gretchen Elizabeth Feulner was born on September 26, 1972 in Mt. Kisco, NY. She grew up in Greensboro, NC and graduated from Page High School in 1990. Bound and determined to get at least a day's drive away from Greensboro, she left for Vanderbilt University, where despite the Commodores best efforts to persuade her otherwise, her love for college sports grew. Graduating with a B.S. in Mathematics in 1994, she headed to Raleigh, enrolling in the Mathematics department at North Carolina State University. Finishing her M.S. in Applied Mathematics in 1996, she moved back to Greensboro to work for Novartis Crop Systems as a systems engineer. After three and a half years of on-call duty and 3AM pages regarding downed computer systems, she realized the database administrator life was not for her and decided to head back to school one last time. She returned to NCSU, joining the Genetics Department in 2000. There she began to work with Dr. Steven Lommel in his plant virology lab, and undertook a project that would make use of her mathematics and computing background. While at North Carolina State University, she met an amazing person, Chris Halweg, and they were married April 29, 2006. She is looking forward to starting the next phase of her life with her wonderful husband and their two psychotic cats in their soon to be new house.

ACKNOWLEDGEMENTS

I have to thank my parents, Dick and Debbie Feulner, for supporting my decision to go back to graduate school. Without their encouragement, I might still be answering pages in the wee hours of the morning. I also need to thank my brother and sister-in-law, Brett and Helen Feulner, for giving me a “science-free” place to hang out my first year back in Raleigh and for showing me the proper way to tailgate.

There are many people at North Carolina State University that I must thank. I is imperative that I thank Dr. Steve Lommel, my advisor, for taking a chance on a graduate student who had no molecular biology experience. I also must thank Dr. Tim Sit, who not only held my hand as I learned the ropes at the bench, but counseled me as I made way through this project and life outside the lab. I need to thank current and former Lommel lab members, especially Dr. Katherine Turner-Schneider who shared the trials and tribulations of graduate school and marriage to a Midwesterner. Various members of the Tobacco Genome Initiative have contributed to sequencing efforts and data compilation. Jason Osborne of the Statistics Department helped with statistical analysis. I would be remiss if I did not thank the martini group – Katie Jordan, Lisa Sanders and Rhonda Wilson, for the much needed Monday night stress breaks.

Last, but certainly not least, I must thank the man that is my rock, who encouraged and supported me when I wanted to quit, but also provided the much needed kick in the rear when I needed it most, my husband, Chris.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	activation domain
AMV	*Alfalfa mosaic virus*
BLAST	Basic local alignment search tool
BMV	*Brome mosaic virus*
BYDV	*Barley yellow dwarf*
CMV	*Cauliflower mosaic virus*
CP	capsid protein
Ct	threshold cycle
DBD	DNA binding domain
DNA	deoxyribonucleic acid
EST	expressed sequence tag
FDR	False Discovery Rate
GFP	green fluorescent protein
GO	gene ontology
HSP	heat shock protein
MAS	maskless array synthesizer
MIPS	Munich Information Center for Protein Sequences
MP	movement protein
NR	non-redundant
ORMV	*Oilseed rape tobamovirus*
PD	plasmodesmata
PME	pectin methylesterase
PNSRV	*Prunus necrotic ringspot virus*
PPV	*Plum pox virus*
PVX	*Potato virus X*
qRT-PCR	real-time quantitative polymerase chain reaction
rRNA	ribosomal ribonucleic acid
RCNMV	*Red clover necrotic mosaic virus*
RNA	ribonucleic acid
SEL	size exclusion limit
TA	trans-activator
TBSV	*Tomato bushy stunt virus*
TE	translational enhancer
TGI	Tobacco Genome Initiative
TMV	*Tobacco mosaic virus*
ToRSV	*Tomato ringspot virus*
TuMV	*Turnip mosaic virus*
TVCV	*Turnip vein clearing virus*
TYMV	*Turnip yellow mosaic virus*
UTR	untranslated region
Y2H	yeast two-hybrid
YKO	yeast knock out

Chapter 1

A review of the life cycle of positive-sense, single-stranded RNA plant viruses

A review of the life cycle of positive-sense, single-stranded RNA plant viruses

Plants are attacked by a wide variety of microbial pathogens (bacteria, fungi, nematodes and viruses) and the success or failure of these interactions is predetermined by the genetic makeup of each participant. A defense/counter-defense arms race has driven the co-evolution of both the host as well as any given pathogen. As a result, genes critical for pathogenicity are the same ones that trigger a host defense response. This has led to the proposed gene-for-gene mechanism to describe these interactions (Flor, 1955). Classical genetic studies of plant/pathogen interactions have focused primarily on either host resistance genes or pathogen virulence determinants. Initially, research was concentrated on the relatively less complex genomes of the microbial pathogens. Recent technological advances have brought about more intensive investigation into the host genetic background involved in plant/microbe interactions (Whitham *et al.*, 2006). While many plant/microbe combinations have been identified, research has primarily focused on a limited number of model systems for each class of pathogen.

Tomato, *Lycopersicon esculentum* serves as a host for each class of microbial pathogen and demonstrates the broad spectrum of interactions capable by any one particular host. The fungal pathogen *Cladosporium fulvum* (leaf mold) encodes the avirulence determinants AVR4 and AVR9 which interact with the resistance genes *Cf-4* and *Cf-9* (Joosten *et al*., 1994; and Van den Ackerveken *et al*., 1992). The bacterial pathogen *Pseudomonas syringae* (bacterial speck) encodes *avrPto* for the corresponding tomato resistance gene *Pto* (Ronald *et al.*, 1992 and

Pitblado *et al.*, 1983, Martin *et al*., 1993). While an avirulence factor has not yet been determined for the nematode pathogen *Meloidogyne incognita* (root knot), the *Mi* gene, originally from *Lycopersicon peruvianum* (wild tomato) confers resistance to this pathogen (reviewed in Williamson, 1998). The *Tm-1* gene confers resistance to *Tobacco mosaic virus* (TMV; Pelham, 1966, 1972). The targets of this viral resistance gene are the TMV encoded replicase subunits (126 kDa and 183 kDa; Ishikawa *et al*., 1986). Viral pathogens (predominantly single stranded positive sense RNA) differ from the other microbes by being minimalist obligate intracellular parasites (typically encoding only four proteins) requiring numerous host factors merely for replication while also interacting with host factors at each stage of the viral life cycle. Research into viral/host interactions has attempted to identify such host factors.

For single stranded positive sense RNA viruses, infection begins upon entry into a suitable host cell. Since the virion is composed of RNA and capsid protein (CP) subunits, other viral proteins must be translated from the genomic RNA in any newly infected cell. Before translation can begin, the RNA must be liberated from the virion in a process known as disassembly. *In vitro* studies of TMV disassembly have shown that host ribosomes bind to the free 5' terminus of the genomic RNA to begin translation of the 126/183 kDa open reading frames (ORFs). These ribosomes generate enough force to dislodge CP subunits from the encapsidated RNA in a process termed co-translational disassembly (Wilson, 1984).

Translation of cellular mRNAs, which contain a 5' cap structure and a 3' poly(A) tail, relies on its 5'-3' circularization mediated by a variety of host proteins. In

order for single stranded plus sense RNA viruses to mimic this process, the circularization mechanism employed to produce this interaction is specified by the genome structure. Some viruses, such as *Potato virus X* (PVX), mimic cellular mRNAs exactly by containing a 5’ cap structure and a 3’ poly(A) tail (ref). One third of the positive sense RNA plant viruses have a 5’ cap but do not have a 3’ poly(A) tail (Dreher and Miller, 2006). *Alfalfa mosaic virus* (AMV) RNAs contain 3’ hairpin structures which bind CP. The AMV CP has been shown to interact with eIF4G/eIFiso4G which would promote circularization of the genomic RNAs performing a function similar to that of poly(A) binding protein (Krab *et al.*, 2005). In the case of *Turnip yellow mosaic virus* (TYMV), the tRNA-like structure (TLS) at the 3’ end functionally replaces the 3’ poly(A) tail (Matsuda and Dreher, 2004). One fourth of the positive sense RNA plant viruses have a VPg at the 5’ terminus instead of a cap structure (Dreher and Miller, 2006). The VPg of *Turnip mosaic virus* (TuMV) has been shown to interact directly with *Arabidopsis thaliana* eIF(iso) 4E in a manner analogous to the 5’ cap structure of cellular mRNAs (Wittmann *et al.*, 1997). Leonard *et al.* (2000) took this relationship a step further and showed that binding of VPg to eIF(iso)4E is necessary for TuMV replication. Some viruses do not have well defined terminal structures and instead utilize a direct RNA-RNA interaction to achieve circularization. *Barley yellow dwarf virus* (BYDV) is lacking typical 5’ and 3’ terminal structures. Instead there is a 109 nucleotide translational enhancer (TE) within the 3’ untranslated region (UTR) which has been shown to have a cruciform secondary structure (Wang *et al.*, 1997 and Guo *et al.*, 2000). Subsequent studies have shown

that one of the stem loops from the TE interacts with a stem loop in the 5'UTR to circularize the genome for translation (Guo *et al.*, 2001).

The first proteins directly translated from the viral genomes are typically the virally-encoded replicase subunits. These minimal components must be supplemented with host factors to form the complete and functional replicase complex. This is illustrated most clearly by the studies on *Brome mosaic virus* (BMV) and *Tomato bushy stunt virus* (TBSV) replication in yeast single-gene deletion libraries. The yeast *DED1* gene encodes for a DEAD-box RNA helicase and is required for translation initiation in yeast. *ded1i* mutant yeast inhibited viral RNA synthesis by limiting translation of BMV RNA2 which encodes one of the viral polymerase subunits (Noueiry *et al.*, 2000). Additionally *lsm1* mutant yeast suppressed BMV RNA3 replication by affecting the ability of BMV 1a replicase subunits to recruit RNA3 templates into a replication complex (Diez, *et al.*, 2000).

Genetic screens of the model plant *A. thaliana* have resulted in the identification of additional host factors required for viral replication. TOM1 and TOM3 have parallel and essential roles in the multiplication of a crucifer-infecting tobamovirus (TMV-Cg). In fact, the double mutant, tom1-1 tom3-1, has been shown to completely suppress replication of TMV-Cg in *Arabidopsis* (Yamanaka *et al.*, 2002).

Protoplast infection studies of positive-sense RNA viruses have illustrated the time frame for the aforementioned stages of the viral life cycle. The most important steps in this cycle occur within the first 24 hours after inoculation. Accumulation of viral RNA (both genomic and subgenomic) in *Nicotiana benthamiana* protoplasts

was detected as early as 2 hours post transfection with TBSV transcripts (Qiu and Scholthof, 2001). This implies that translation and replication of TBSV RNA was initiated prior to the 2 hour sampling timepoint. These observations as well as those for other RNA viruses determine sampling time points for future studies looking at early events in the viral life cycle.

After replication and late gene expression (e.g. CP) have occurred, assembly and/or movement of the viral genome takes place depending on the infected cell type/location. Movement of the virus to uninfected cells is termed either local or systemic. For local movement, the virus moves through plasmodesmata (PD), a series of intercellular channels connecting the contents of each cell. PD are highly dynamic structures that are able to modulate their pore size or size exclusion limit (SEL) to control the flow of macromolecules between cells (Epel, 1994; Boyko *et al.*, 2000**)**. The MPs of many viruses, including *Red clover necrotic mosaic virus* (RCNMV), *Cucumber mosaic virus* (CMV) and PVX, have been shown to modify the SEL of PD (Ding *et al.*, 1995, Fujiwara *et al.*, 1993, Vaquero *et al.*, 1994). It has been observed that TMV movement protein (MP) increases the SEL from 1 to 10 kDa (Wolf *et al.*, 1989) or even 20 kDa (Waigmann *et al.*, 1994). Angell *et al.* (1996) confirmed that an increase in the SEL of *N. clevelandii* is associated with the movement of PVX.

In terms of host/pathogen interactions during movement, the first interactions occur intracellularly. In the case of TMV, the 30K MP binds to components of the cytoskeleton (namely actin, myosin and tubulin) to effect transport of viral replication complexes to the plasmodesmata (Beachy and Heinlein, 2000). While MPs have

been demonstrated to localize to the PD and modify their SELs, no direct interaction between an MP and a distinct PD protein has yet been identified. Infectious, fluorescently labeled TMV clones illustrated the movement of the viral replication complex (MP, viral replicase, and genomic RNA) across the plasmodesmata into adjacent uninfected cells (Kawakami *et al.*, 2004). Researchers, separately utilizing both affinity chromatography and blot overlay binding assays, identified a 33-kDa *N. tabacum* cell wall protein with strong binding affinity to the TMV MP (Chen *et al.*, 2000; Dorokhov *et al.*, 1999). Through sequence analysis, the protein was determined to be pectin methylesterase (PME), a ubiquitous plant enzyme responsible for modifiying pectin. These findings imply that PME plays a role in the cell-to-cell movement of TMV. Affinity chromatography, two-dimensional electrophoresis and HPLC were employed to isolate *N. tabacum* proteins bound to TMV MP (Chen at al., 2005). Calreticulin, a protein known to accumulate in the PD, was identified and subsequent transgenic studies showed its overexpression limited local movement.

Far-Western blotting was used to screen an *N. benthamiana* cDNA library to investigate host factors that interact with the BMV MP (Kaido *et al.*, 2007). A protein, NbNACa1, was identified and determined to regulate localization to the PD, thus influencing cell-to-cell movement of BMV.

Systemic movement occurs through the plant vasculature, a network of phloem companion cell/sieve element complexes, and involves long distances. The spread of the virus through the vascular system is a more passive process, with virions moving with the flow of nutrients (reviewed by Leisner and Howell, 1993).

While many host proteins have been shown to interact with MPs for local movement, host proteins involved in systemic movement have not been as well characterized. However, PME, the protein first isolated in screens for local TMV movement, was also shown to play a role in the systemic spread of TMV by influencing the egress of the virions from the vasculature of systemically infected leaves (Chen and Citovsky, 2003).

Many of the above findings have employed shotgun techniques to identify host factors that interact with viral proteins. These techniques have involved direct protein-protein assays (yeast two-hybrid assay), gain/loss of function assays in a genetically more amenable heterologous host (viral replication in yeast) and indirectly through profiling of host mRNA expression levels (microarray). The yeast two-hybrid (Y2H) assay is designed to identify direct protein-protein interactions by *in vivo* reconstitution of an active yeast transcription complex. Yeast transcription factors have two domains, a DNA binding domain (DBD) for promoter recognition and an activation domain (AD) for transcriptional activation. The viral protein to be studied is cloned as a fusion to the DBD while the host expression library is cloned as fusions to the AD. These constructs are co-transformed into yeast cells. If the viral protein interacts with a host protein, expression of various reporter genes is induced that allow growth on selective media. PME in tomato and Valencia oranges was shown to interact with TMV MP and the *Turnip vein clearing virus* (TVCV) MP through the Y2H assay (Chen *et al.*, 2000). Y2H assays also confirmed an interaction between the BMV MP and barley HCP1, an oxidoreductase protein (Okinaka *et al.*, 2003).

In contrast to the Y2H system, the yeast itself has been used as a heterologous host to support the replication of BMV and TBSV. This approach takes advantage of pre-existing mutant yeast libraries to systematically identify yeast genes that play a role in virus viability. Once identified, plant homologues of these yeast proteins can then be isolated and further characterized for their direct role *in planta*. For both BMV and TBSV, yeast cells are transformed with plasmids expressing the corresponding replicase subunits and transfected with either a replication competent RNA (BMV RNA3) or a construct that expresses a replication competent RNA (TBSV defective-interfering RNA; DI RNA). In the case of BMV, a reporter gene in place of the CP is only expressed if replication of RNA3 and subsequent transcription of RNA4 (subgenomic RNA for the reporter gene) is supported (Janda and Ahlquist, 1993). In the case of TBSV, the DI RNA is expressed from its plasmid vector at very low levels and replication is assayed by an increase in DI RNA levels above background (Panavas and Nagy, 2003). Either system is then used to screen yeast single-gene-knockout (YKO) libraries for their effect(s) on viral replication. Almost 100 different host genes influenced BMV replication by at least 3 fold when using a yeast diploid deletion series (Research Genetics, Huntsville, AL; Kushner *et al.*, 2003). Not surprisingly, this study confirmed the importance of three previously identified genes (*LSM1*, *LSM6*, *PAT1*) shown to affect BMV translation. (Diez *et al.*, 2000; Noueiry *et al.*, 2003). Using the YKO deletion series (Open Biosystems, Huntsville, AL), 96 genes were identified that influenced the replication of TBSV with 20% of these genes having an effect on transport/targeting (Panavas et al, 2005). Comparison of the BMV and TBSV studies

revealed that only four genes were conserved between both systems (three in the ubiquitin pathway, one a transcriptional regulator).

While both yeast-based techniques examined direct protein/protein interactions, oligonucleotide microarrays can be used to indirectly monitor global changes in host mRNA expression upon viral infection at a given point in time. Oligonucleotide microarrays consist of a series of DNA oligonucleotides (>20,000) representing expressed sequences of a particular host that are synthesized directly onto a glass slide as individual spots. The oligonucleotides can very in length from 25-30 mers (short) to 50-60 mers (long). Total RNA isolated from each treatment is then fluorescently labeled and separately hybridized to an array. Differences in hybridization profiles between treatments are compared to identify statistically significant changes in mRNA expression levels.

To date six studies have been published examining the relationship between a positive sense RNA virus and its plant host. Of the six studies, four use the Affymetrix Arabidopsis (ATH1) GeneChip (Santa Clara, CA). The other two studies use custom synthesized cDNA EST arrays (*Solanum tuberosum* and *A. thaliana*) (Dardick, 2007; Golem and Culver, 2003).

The first study published was not solely focused on the virus/host interaction per se but rather using viral infection as one of many stressor treatments employed to examine the expression levels of 402 known and putative transcription factor genes (Chen *et al.*, 2002). The second study published appears to extract data from the first array experiment with a more narrow focus on the systemic effects of viral infection (Whitham *et al.*, 2003). The researchers challenged *A. thaliana* with one of

five different viruses: CMV, *Oilseed rape tobamovirus* (ORMV), PVX, TuMV and TVCV. RNA from mock and viral-infected leaves was collected at 1, 2, 4 and 5 days post-inoculation (dpi) and hybridized to the Affymetrix Arabidopsis ATH1 GeneChip at the same timepoints. Attention was focused on the 114 genes that showed differential expression at a given time point in response to all five viruses. As time after inoculation increased, so did the number of induced genes. Based on the Munich Information Center for Protein Sequences (MIPS) classification scheme, genes associated with cell rescue, defense, death and aging, comprise one-third of the significant genes, the largest such grouping. An interesting possible cause and effect relationship was revealed by a group of five heat shock proteins (HSP) that were all upregulated but in a differential manner. Gene expression was induced within 1 dpi for TVCV and ORMV, while the other three viruses did not elicit a heat shock response until 2 dpi, suggesting that tobamoviruses specifically elicit HSP gene expression via a common elicitation mechanism.

Golem and Culver (2003) took a different approach and investigated changes in gene expression within both the inoculated and systemically infected leaves. They utilized the virus/host combination of TMV and the Shahdara ecotype of *A. thaliana* which supports more rapid viral accumulation than other *A. thaliana* ecotypes. cDNA arrays were hybridized with either RNA from inoculated leaves at 4 dpi or systemically infected leaves 14 dpi. As with the previous study, as time post inoculation increased, so did the number of genes showing altered expression: 23 genes at 4 dpi and 53 genes at 14 dpi, representing a total of 68 distinct genes. On inoculated leaves, all genes were induced, while in systemic leaves, only 18 were

induced. Ten of the 53 systemic leaf genes represent metabolic functions and 9 represent stress and transcription functions. Unlike earlier studies, the largest category of genes represented was unclassified with no known function (at the time of publication). The low number of total genes implies that TMV does not have a great effect on *Arabidopsis* gene expression.

A more extensive study of *A. thaliana* gene expression patterns in response to infection by CMV, strain Y (CMV-Y) was undertaken to examine the RCY1 mediated resistance response (Marathe *et al.*, 2004). Gene expression profiles were measured at 3, 5, and 18 hours post-inoculation. 444 differentially expressed genes were discovered of which 106 were induced within the first 18 hours. Overall, many of the 444 genes were related to defense and cell death while 25% had no known function. One trend observed was that very few genes changed their expression profile over the limited course of the experiment.

One of the limitations of the aforementioned studies is the use of RNA extracted from a mixture of cells at various stages of viral infection. This can lead to possible dilution of expression caused by sampling at different disease and/or developmental stages. The only study to combine both the temporal and spatial effects of a viral infection utilized the Affymetrix *Arabidopsis* ATH1 GeneChip probed with RNA extracted from TuMV-GFP infected leaf tissue (Yang *et al.*, 2007). This approach yielded new suites of genes not previously identified as being involved in the response to viral infection including the induction of a large number of ribosomal and protein synthesis genes. Interestingly, defense and stress related genes did not show a great change in expression levels in advance of the viral infection front. This

report illustrates the necessity to incorporate a spatial component in future expression studies.

The most recently published study is a comparative study where a potato cDNA array was probed with mRNA from upper non-inoculated *N. benthamiana* leaves 14 dpi by one of three stone fruit viruses: *Plum pox virus* (PPV; *Potyviridae*), *Tomato ringspot virus* (ToRSV; *Nepoviridae*) and *Prunus necrotic ringspot virus* (PNSRV; *Bromoviridae*; Dardick, 2007). The viruses chosen vary in genome organization and symptomatology. A potato cDNA array was employed because an *N. benthamiana* array was not available and potato is a close Solanceous relative of *N. benthamiana*. The hybridization results from all three viruses were pooled to generate a total set of 1,438 statistically significant unique genes. The number of genes per virus correlated with the severity of the symptoms observed with 1,082 genes for ToRSV, 744 for PPV and 89 for PNRSV. Several of the same genes were altered by each virus and most of the changes observed were of the same trend (repressed or induced). Discounting the unknown category, chloroplast related genes make up the largest category of differentially expressed genes (mostly repressed) followed by protein synthesis (mostly induced) and stress/pathogenesis related genes (mostly induced). This work verifies the association between *N. benthamiana* symptomatology and the suite of genes regulated during the viral infection process.

As seen from the numerous aforementioned studies, *A. thaliana* (a member of the *Brassicae* and thus closely related to cabbage, kale and broccoli) is the preferred host for plant genetic research due to the following advantages: i) *A.*

thaliana, was the first genome of a flowering plant to be sequenced (Arabidopsis Genome Initiative 2000) due to its small genome size (119Mbp, 27,029 protein coding genes, > 65% non coding genes; www.arabidopsis.org, Taft et al., 2007), ii) it posseses a short 6 week generation time (from seed to seed), iii) the plant itself has a small physical size, Despite these advantages, the major drawback to viral studies with *A. thaliana* is the limited number of viruses that can infect *A. thaliana*. Thus, any findings may not have wide applicability to other viral/host combinations. Many of the reports looked at later time points (after symptoms have already appeared) which would indicate the occurrence of multiple infections/re-infections as the virus spreads throughout the inoculated leaf. The varying host gene expression profiles would then become overlapping due to the variety of viral infection stages encountered in a particular leaf.

For these reasons, we sought an alternative viral/host system that was more representative of plant virology research. *N. benthamiana* is the preferred host for plant/viral interaction studies due to its near universal susceptibility to viral infection, despite the presence of a fully functional RNA-based antiviral defense system. *N. benthamiana* is a member of the *Solanaceae*, a group that includes several agronomically important crops such as potato, tomato and tobacco. The base chromosome number for members of the *Solanaceae* is 12, but amphiploidy is common among the *Solanaceae. N. benthamiana* (2n=38) is believed to be the result of a cross between *N. suaveolens* (2n=32) and *N. debneyi* (2n=48) followed by chromosome loss. The predicted genome size is between 3.2 and 3.8 Gbp with 99% of it non-coding DNA. While we don't have an exact number of protein coding

regions, we estimate that there are 34,000 unique transcripts in the *N. benthamiana* genome. The high percentage of non-coding DNA and the high number of repetitive elements makes the sequencing of and gene finding within *N. benthamiana* very difficult. In spite of these drawbacks, we have assembled over 13,000 singletons and contigs for use in a custom oligonucleotide microarray of *N. benthamiana*. In spite of the difficulty presented by its cumbersome genome (at least 27X the size of *A. thaliana*), we felt that it was important to develop this tool for use in plant virology studies.

For the viral pathogen, we have chosen RCNMV, a single-strand plus sense member of the *Tombusviridae*. Some of the advantages to using RCNMV include its ease of manipulation, relatively high titer and ability to monitor infection with a GFP expressing construct. This combination of RCNMV on *N. benthamiana* will be more indicative of virus/host pairings typically studied in other plant virology labs.

RCNMV is a member of the *Dianthovirus* genus, the only genus within the *Tombusviridae* to possess a segmented genome. The RCNMV genome consists of two positive sense single stranded RNAs, encoding a total of four ORFs (Figure 1). RNA-1 has three ORFs encoding a 27 kDa subunit (p27), an 88 kDa -1 ribosomal frameshifting readthrough product of p27 (p88; Xiong *et al.*, 1993) and the 37 kDa CP (Xiong and Lommel, 1989). RNA-2 encodes the 35 kDa MP required for cell-to-cell movement (Lommel *et al.*, 1988). RNA-1 can replicate independently but cannot move from the initially inoculated cells without RNA-2 (Osman *et al.*, 1987; Paje-Manalo and Lommel, 1989).

Both genomic RNAs lack well defined terminal structures and rely on cap-independent translation. A TE element, similar to the one described for BYDV, was identified within the 3'UTR of RNA-1 (Mizumoto *et al.*, 2003). However, the translational mechanism employed by RNA-2 is different due to the lack of a discernible TE element (Mizumoto *et al.*, 2006). p27 and p88 represent the viral replicase subunits with p88 containing the core polymerase catalytic site of RNA-dependent RNA polymerases (Xiong and Lommel, 1989). Both p27 and p88 have been detected in purifed replicase extracts from infected tissue (Bates *et al*., 1995.) and shown to localize to the endoplasmic reticulum suggesting that these membranes are the site of RCNMV replication (Turner *et al.*, 2004). Viral RNA accumulation is temperature dependent as illustrated by the difference between the Canadian and Australian strains. The Australian strain supports replication at both 17°C and 25°C while the Canadian strain only replicates at 17°C (Mizumoto *et al.*, 2002). This temperature specificity was shown to reside within the 3'UTR of RNA-1.

CP is expressed from a subgenomic RNA (sgRNA) that is synthesized via a proposed premature termination mechanism (Sit *et al.*, 1998, White, 2002). This mechanism involves a direct RNA-RNA interaction between the *trans*-activator (TA) element on RNA-2 and the TA binding site within the subgenomic promoter on RNA-1 (Sit *et al.*, 1998).

N. benthamiana and *N. clevelandii* are typical hosts for infectivity studies. Local symptoms appear in 2-3 dpi and systemic spread is observed at 3-4 dpi. Symptoms include local necrotic ringspot lesions and systemically infected leaves display a mosaic pattern, leaf curling and overall stunting of growth (Figure 2).

Studies using RCNMV tagged with green fluorescent protein (GFP) have shown cell-to-cell movement of the viral RNA as early as twelve hours after inoculation (T.L. Sit & S.A. Lommel, unpublished data). Additionally, Vaewhongs and Lommel (1995) showed that virion formation is required for long distance movement. For RCNMV, two populations of virions have been identified: one type consisting of RNA-1 and RNA-2 and a second type containing multiple copies of only RNA-2 (Basnayake *et al*., 2005). It is thought that genomic RNA moves through the vascular system in the form of a virion as a way to protect the genomic RNA from degradation by nucleases.

Very few studies to date have tackled the elucidation of global changes in host gene expression during infection by a positive sense RNA plant virus. The majority utilized an *A. thaliana* infecting virus and the commercially available Affymetrix *Arabidopsis thaliana* GeneChip. There have been two published comparative studies in which a *Solanum tuberosum* cDNA microarray was probed with RNA extracted from virally infected *N. benthamiana* tissue with only one using a positive sense RNA virus as the pathogen.

In this thesis we created an *N. benthamiana* based oligonucleotide microarray and hybridize it with RNA from RCNMV infected *N. benthamiana.* The intent of our study is to examine earlier infection time points than previous studies to obtain a more definitive sampling of host gene expression profiles. Over 13,000 uniquely expressed sequences or approximately 38% of the *N. benthamiana* transcriptome, will be included on the array. Nimblegen (Madison, WI) was chosen to manufacture this array because their Maskless Array Synthesizer (MAS), utilizing virtual masks

and synthesizing oligonucleotides directly onto glass slides, enables them to include almost 400,000 spots on each array. This maximizes the ability to include several spots and repeats for each sequence ensuring high confidence in hybridization results. Our study will look at the events taking place in a localized infection. mRNA abundance in mock and virally inoculated plants will be compared at four time points within the first 24 hours after inoculation. After hybridization, an annotation using GenBank data, a functional assignment using the GeneOntology (GO) classification system and a statistical analysis will be performed. Our goal is to get an overall picture of changes in host gene expression as an initial survey for a core set of *N. benthamiana* genes that are universally expressed upon positive strand RNA plant virus infections. It is our belief that this experiment will offer new insights into the interaction between host and viral pathogen, which will expand our knowledge of *N. benthamiana* as it has not been previously studied to this extent.

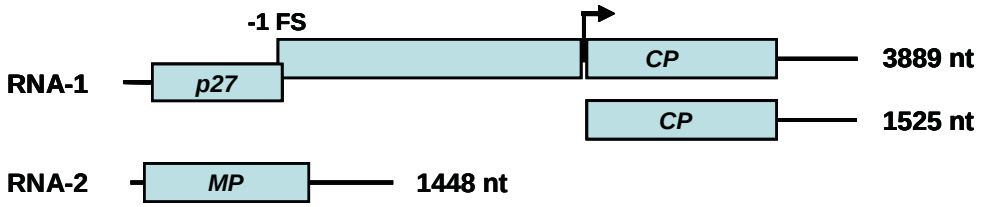

Figure 1-1. RCNMV Genome Organization
A schematic representation of RNA-1 and RNA-2. Encoded proteins and ORFs are shown. Arrow depicts subgenomic promoter of CP.

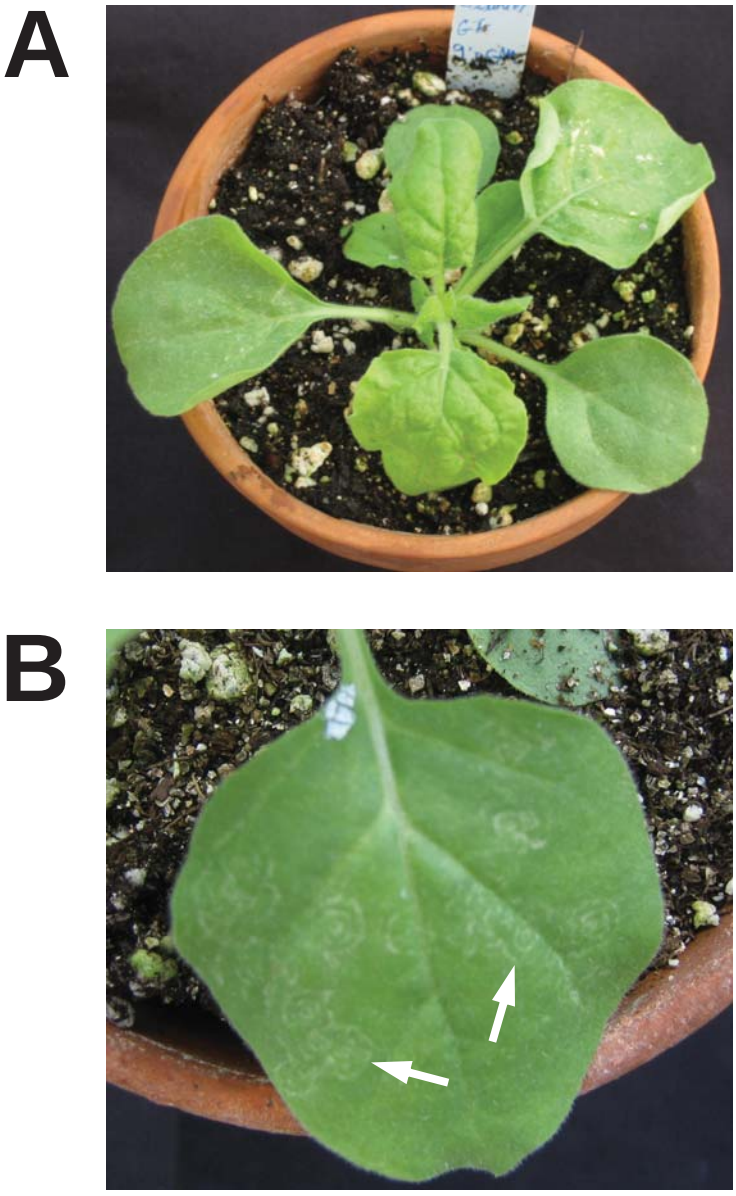

Figure 1-2. RCNMV Infected *Nicotiana benthamiana*
(A) Systemic symptoms including leaf curling. **(B)** Local symptoms characterized by necrotic ringspot lesions (arrows).

References

Angell, S.M., Davies, C., Baulcombe, D.C. 1996. Cell-to-cell movement of *Potato virus X* is associated with a change in the size-exclusion limit of plasmodesmata in trichome cells of *Nicotiana clevelandii*. Virology 216, 197-201.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408,796-815.

Basnayake, V.R., Sit, T.L., Lommel, S.A. 2006. The genomic RNA packaging scheme of *Red clover necrotic mosaic virus*. Virology 345, 532-539.

Bates, H.J., Farjah, M., Osman, T.A., Buck, K.W. 1995. Isolation and characterization of an RNA-dependent RNA polymerase from *Nicotiana clevelandii* plants infected with Red clover necrotic mosaic dianthovirus. Journal of General Virology 76, 1483-1491.

Beachy, R.N. and Heinlein, M. 2000. Role of P30 in Replication and spread of TMV. Traffic 1, 540-544.

Boyko, V., Feralli, J., Ashby, J., Schellenbaum, P., Heinlein, M. 2000. Function of microtubules in intercellular transport of plant virus RNA. Nature Cell Biology 2, 826-832.

Chen, M.H., Citovsky, V. 2003. Systemic movement of a tobamovirus requires host cell pectin methylesterase. Plant Journal 35, 386-392.

Chen, M.H., Sheng, J., Hind, G., Handa, A.K., Citovsky, C. 2000. Interaction between the tobacco mosaic virus movement protein and host cell pectin methylesterases is required for viral cell-to-cell movement. The EMBO Journal 19, 913-920.

Chen, M.H., Tian, G.W., Gafni, Y. and Citovsky, V. 2005. Effects of calreticulim on viral cell-to-cell movement. Plant Physiology 138, 1866-1876.

Chen, W., Provart, N.J., Glazebrook, J., Katagiri, F., Chang, H.S., Eulgem, T., Mauch, F., Luan, S., Zou, G., Whitham, S.A., Budworth, P.R., Tao, Y., Xie, Z., Chen, X., Lam, S., Kreps, J.A., Harper, J.F., Si-Ammourt, A., Mauch-Mani, B., Heinlein, M., Kobayashi, K., Hohn, T., Dangl, J.L., Wang, X., Zhu, T. 2002. Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. The Plant Cell 14, 559-574.

Dardick, C. 2007. Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses. Molecular Plant-Microbe Interactions 20, 1004-1017.

Diez, J., Ishikawa, M., Kaidom M., Ahlquist, P. 2000. Identification and characterization of a host protein required for efficient template selection in viral RNA replication. Proceedings of the National Academy of Science, USA 97, 3913-3918.

Ding, B., Li, Q., Nguyen, L., Palukaitis, P., Lucas, W.J. 1995. *Cucumber mosaic virus* 3a protein potentiates cell-to-cell trafficking of CMV RNA in tobacco plants. Virology 207, 345-353.

Dorokhov, Y.L., Makinen, K., Yu, O., Merits, A., Saarinen, J., Kalkkinen, N., Atabekov, J.G. and Saarma, M. 1999. A novel function for a ubiquitous plant enzyme pectin methylesterase: the host-cell receptor for the *Tobacco mosaic virus* movement protein. FEBS Letters 461, 223-228.

Dreher, T.W., Miller, W.A. 2006. Translational control in positive strand RNA plant viruses. Virology 344, 185-197/

Epel, B.L. 1994. Plasmodesmata: composition, structure and trafficking. Plant Molecular Biology 26, 1343-1356.

Flor, H.H. 1955. Host-parasite interactions in flax rust – its genetics and other implications. Phytopathology 45, 680-685.

Fujiwara, T., Geisman-Cookmeyer, D., Ding, B., Lommel, S.A., Lucas, W.J. 1993. Cell-to-cell trafficking of macromolecules through plasmodesmata potentiated by the *Red clover necrotic mosaic virus* movement protein. Plant Cell 5, 1783-1794.

Golem S. and Culver, J.N. 2003. *Tobacco mosaic virus* induced alterations in the gene expression profile of Arabidopsis thaliana. Molecular Plant-Microbe Interactions 16, 681-688.

Guo, L., Allen, E., Miller, W.A. 2000. Structure and function of a cap-independent translation element that functions in either the 3’ or the 5’ untranslated region. RNA 6, 1808-1820.

Guo, L., Allen, E.M., Miller, W.A. 2001. Base-pairing between untranslated regions facilitates translation of uncapped, nonpolyadenylated viral RNA. Molecular Cell 7, 1103-1109.

Ishikawa, M., Meshi, T., Motoyoshi, F., Takamatsu, N., Okada, Y. 1986. In vitro mutagenesis of the putative replicase genes of *Tobacco mosaic virus*. Nucleic Acids Research 14, 8291-8305.

Janda, M., Ahlquist, P. 1993. RNA-dependent replication, transcription, and persistence of brome mosaic virus RNA replicons in S. cerevisiae. Cell 72, 961-970.

Joosten, M.H., Cozijnsen, T.J., De Wit, P.J. 1994. Host resistance to a fungal tomato pathogen lost by a single base-pair change in an avirulence gene. Nature 367, 384-386.

Kaido, M., Inoue, Y., Takeda, Y., Sugiyama, K., Takeda, A., Mori, M., Tamai, A., Meshi, T., Okuno, T. and Mise, K. 2007. Downregulation of the *NbNACa1* gene encoding a movement-protein-interacting protein reduces cell-to-cell movement of *Brome mosaic virus* in *Nicotiana benthamiana*. Molecular Plant-Microbe Interactions 20, 671-681.

Kawakami, S., Watanabe, Y., Beachy, R.N. 2004. *Tobacco mosaic virus* infection spreads cell to cell as intact replication complexes. Proceedings of the National Academy of Science, USA 101, 6291-6296.

Krab, I.M., Caldwell, C., Gallie, D.R., Bol, J.F. 2005. Coat protein enhances translational efficiency of *Alfalfa mosaic virus* RNAs and interacts with the eIF4G component of initiation factor eIF4F. Journal of General Virology 86, 1841-1849.

Kushner, D.B., Lindenbach, B.D., Grdzelishvili, V.Z., Noueiry, A.O., Paul, S.M., Ahlquist, P. 2003. Systematic, genome-wide identification of host genes affecting replication of a positive-sense RNA virus. Proceedings of the National Academy of Science, USA 100, 15764-15769.

Leisner, S.M., Howell, S.H. 1993. Long-distance movement of viruses in plants. Trends in Microbiology 1, 314-317.

Leonard, S., Plante, D., Wittmann, S., Daigneault, N., Fortin, M.G., Laliberte, J.F. 2000. Complex formation between potyvirus VPg and translation eukaryotic initiation factor 4E correlates with virus infectivity. Journal of Virology 74, 7730-7737.

Lommel S.A., Weston-Fina, M., Xiong, Z., Lomonossoff, G.P. 1988. The nucleotide sequence and genome organization of *Red clover necrotic mosaic virus* RNA-2. Nucleic Acids Research 16, 8587-8602.

Marathe, R., Guan, Z., Anandalakshmi, R., Zhao, H., Dinesh-Kumar, S.P. 2004. Study of *Arabidopsis thaliana* resistome in response to *Cucumber mosaic virus* infection using whole genome microarray. Plant Molecular Biology 55, 501-520.

Martin, G.B., Brommonschenkel, S.H., Chunwongse, J., Frary, A., Ganal, M.W., Spivey, R., Wu, T., Earle, E.D., Tanksley, S.D. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science 262, 1432-1436.

Matsuda, D., Dreher, T.W. 2004. The tRNA-like structure of *Turnip yellow mosaic virus* is a 3'-translational enhancer. Virology 321, 36-46.

Mizumoto, H., Hikichi, Y., Okuno, T. 2002. The 3'-untranslated region of RNA1 as a primary determinant of temperature sensitivity of *Red clover necrotic mosaic virus* Canadian strain. Virology 293, 320-327.

Mizumoto, H., Iwakawa H.O., Kaido, M., Mise, K., Okuno, T. 2006. Cap-independent translation mechanism of *Red clover necrotic mosaic virus* RNA2 differs from that of RNA1 and is linked to RNA replication. Journal of Virology 80, 3781-3791.

Mizumoto, H. Tatsuta, M., Kaido, M., Mise, K., Okuno, T. 2003. Cap-independent translational enhancement by the 3' untranslated region of Red clover necrotic mosaic virus RNA1. Journal of Virology 77, 12113-12121.

Noueiry, A.O., Chen, J., Ahlquist, P. 2000. A mutant allele of essential, general transalation initiation factor DED1 selectively inihibits translation of a viral mRNA. Proceedings of the National Academy of Science, USA 97, 12985-12990.

Noueiry, A.O., Diez, J., Falk, S.P., Chen, J., Ahlquist, P. 2003. Yeast Lsm1p-7p/Pat1p deadenylation-dependent mRNA decapping factors are required for *Brome mosaic virus* genomic RNA translation. Molecular and Cellular Biology 23, 4094-4106.

Okinaka, Y., Mise, K., Okuno, T., Furusawa, I. 2003. Characterization of a novel barley protein, HCP1, that interacts with the *Brome mosaic virus* coat protein. Molecular Plant-Microbe Interactions 16, 352-359.

Osman, T.A.M., Buck, K.W. 1987. Replication of *Red clover necrotic mosaic virus* RNA in cowpea protoplasts: RNA-1 replicates independently of RNA-2. Journal of General Virology 68, 289-296.

Paje-Manalo, L.L., Lommel, S.A. 1989. Independent replication of *Red clover necrotic mosaic virus* RNA-1 in electroporated host and nonhost *Nicotiana* species protoplasts. Phytopathology 79, 457-461.

Panavas T., Nagy, P.D. 2003. Yeast as a model host to study replication and recombination of defective interfering RNA of *Tomato bushy stunt virus*. Virology 314, 315-325.

Panavas, T., Serviene, E., Brasher, J., Nagy, P.D. 2005. Yeast genome-wide screen reveals dissimilar sets of host genes affecting replication of RNA viruses. Proceedings of the National Academy of Sciences, USA 102, 7326-7331.

Pelham, J. 1966. Resistance in tomato to *Tobacco mosaic virus*. Euphytica 15, 258-267.

Pelham, J. 1972. Strain-genotype interaction of *Tobacco mosaic virus* in tomato. Annals of Applied Biology 71, 219-228.

Pitblado, R.E. and MacNeill, B.H. 1983. Genetic basis of resistance to *Pseudomonas synringae* pv. tomato in field tomatoes. Canadian Journal of Plant Pathology 5, 251-255.

Qiu, W. and Scholthof, H.B. 2001. Effects of inactivation of the coat protein and movement genes of Tomato bushy stunt virus on early accumulation of genomic and subgenomic RNAs. Journal of General Virology 82, 3107-3114.

Ronald, P.C., Salmeron, J.M. Carland, F.M. and Staskawicz, B.J. 1992. The clones avirulence gene *avrPto* induces disease resistance in tomato cultivars containing the *Pto* resistance gene. Journal of Bacteriology 174, 1604-1611.

Selling, B.H., Allison, R.F., Kaesberg, P. 1990. Genomic RNA of an insect virus directs synthesis of infectious virions in plants. Proceedings of the National Academy of Sciences, USA 87, 434-438.

Sit, T.L., Vaewhongs, A.A., Lommel, S.A. 1998. RNA-mediated trans-activation of transcription from a viral RNA. Science 281, 829-832.

Taft, R.J., Pheasant, M., Mattick, J.S. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays 29, 288-299.

Turner, K.A., Sit, T.L., Callaway, A.S., Allen, N.S., Lommel, S.A. 2004. *Red clover necrotic mosaic virus* replication proteins accumulate at the endoplasmic reticulum. Virology 320, 276-290.

Vaewhongs, A.A., Lommel, S.A. 1995. Virion formation is required for the long-distance movement of *Red clover necrotic mosaic virus* in movement protein transgenic plants. Virology 212, 607-613.

Van den Ackerveken, G.F., Van Kan, J.A., De Wit, P.J. 1992. Molecular analysis of the avirulence gene avr9 of the fungal tomato pathogen *Cladosporium fulvum* fully supports the gene-for-gene hypothesis. Plant Journal 2, 359-366.

Vaquero, C., Turner, A.P., Demangeat, G., Sanz, A., Serra, M.T., Roberts, K., Garcia-Luque, I. 1994. The 3a protein from *Cucumber mosaic virus* increase the gating capacity of plasmodesmata in transgenic tobacco plants. Journal of General Virology 75, 3193-3197.

Waigmann, E., Lucas, W.J., Citovsky, V., Zambryski, P. 1994. Direct functional assay for *Tobacco mosaic virus* cell-to-cell movement protein and identification of a domain involved in increasing plasmodesmal permeability. Proceedings of the National Academy of Sciences, USA 91, 1433-1437.

Wang, S., Browning, K.S., Miller, W.A. 1997. A viral sequence in the 3'-untranslated region mmics a 5' cap in facilitating translation of uncapped mRNA. The EMBO Journal 16, 4107-4116.

White, K.A. 2002. The premature termination model: a possible third mechanism for subgenomic mRNA transcription in (+)-strand RNA viruses. Virology 304, 147-154.

Whitham, S.A., Quan, S., Chang, H.S., Cooper, B., Estes, B., Xhu, T., Wang, X., Hou, Y.M. 2003. Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. The Plant Journal 33, 271-283.

Whitham, S.A., Yang, C., Goodin, M.M. 2006. Global impact: elucidating plant response to viral infection. Molecular Plant-Microbe Interactions 19, 1207-1215.

Wittmann, S., Chatel, H., Fortin, M.G., Laliberte, J.F. 1997. Interaction of the viral protein genome linked of *Turnip mosaic potyvirus* with the translational eukaryotic initiation factor (iso) 4E of *Arabidopsis thaliana* using the yeast two-hybrid system. Virology 234, 84-92.

Williamson, V.M. 1998. Root-knot nematode resistance genes in tomato and their potential for future use. Annual Reviews of Phytopathology 36, 277-293.

Wilson, T.M.A. 1984. Cotranslational disassembly of *Tobacco mosaic virus in vitro.* Virology 137, 255-265.

Wolf, S., Deom, C.M., Beachy, R.N., Lucas, W.J. 1989. Movement protein of tobacco mosaic virus modifies plasmodesmatal size exclusion limit. Science 246, 377-379.

Xiong, Z., Kim, K.H., Kendall, T.L., Lommel, S.A. 1993. Synthesis of the putative *Red clover necrotic mosaic virus* RNA polymerase by ribosomal frameshifting *in vitro*. Virology 193, 213-221.

Xiong Z. and Lommel, S.A. 1989. The complete nucleotide sequence and genome organization of *Red clover necrotic mosaic virus* RNA-1. Virology 171, 543-554.

Yamanaka, T., Imai, T., Satoh, R., Kawashima, A., Takahashi, M., Tomita, K., Kubota, K., Meshi, T., Naito, S., Ishikawa, M. 2002. Complete inhibition of tobamovirus multiplication by simulataneous mutations in two homologous host genes. Journal of Virology 76, 2491-2497.

Yang, C., Guo, R., Jie, F., Nettleton, D., Peng, J., Carr, T., Yeakley, J.M., Fan, J.B., Whitham, S.A. 2007. Spatial analysis of *Arabidopsis thaliana* gene expression in response to *Turnip mosaic virus* infection. Molecular Plant-Microbe Interactions 20, 358-370.

Chapter 2

Creation of a *Nicotiana benthamiana* microarray

Background

Nicotiana benthamiana, a species in the genus *Nicotiana*, is native to Australia, more specifically the Northern Territory and Western Australia (Goodspeed, 1954). It is thought to be the result of a hybridization between *N. debneyi* and *N. suaveolens*. *N. benthamiana* is an amphiploid species, n=19, with a genome size estimated between 3.2 to 3.8 Gbp (uncited). *N. benthamiana* is often used as a model host for numerous fungal and bacterial pathogens and is the model system for the study of gene silencing in plants. Additionally, *N. benthamiana* is the model host for viral-plant host interaction studies. Not only is it susceptible to a large number of plant viruses, including many that are typically restricted to monocots, but it has also been shown to be susceptible to some animal viruses. Our lab's field of interest, plant virology, encompasses many areas of study, including, but not limited to gene silencing, host interactions, replication, gene expression and cell biology. In particular, we are interested in the interaction between host and virus. It is for these reasons that we chose to use *N*. *benthamiana* as the host in our host-viral interaction study. We hypothesize that a large number of host genes are affected during the course of a viral infection, either directly or indirectly. With a project of this size, we felt a microarray-based approach would be the most practical approach to study this interaction.

In recent years, the number of plant genomic tools available has grown greatly. In fact, a search of the Affymetrix® website shows us that not only is a whole genome array of *Arabidopsis thaliana* offered, but there are now over 10 different plant arrays of agronomically important crops readily available for expression

analysis. However, at this time study a *N. benthamiana* array was not commercially available, nor had the genome been sequenced and annotated. While the importance and significance of the various plant microarrays readily available is not in question, for plant virologists, perhaps one of the most desired tools , a microarray of *Nicotiana benthamiana*, was not available. Plant virologists have made do with whatever microarray was available at the time, the most popular being the *Arabidopsis thalaiana* array. However the simplicity and size of the *Arabidopsis* genome along with the fact that is serves as host to only a few viruses in comparison to *N. benthamiana*, would make the availability of a *N. benthamiana* a breakthrough of sorts for the plant virology community

It was our goal to create a *N. benthamiana* microarray not only for use in our study, but to serve as a tool for other plant virologists in their own studies. When this project began, the intent was to create a cDNA microarray. However, as work on this project progressed, new technologies became available and a collaboration with Nimblegen Systems Inc. (Madison, WI) was formed to create an oligonucleotide microarray. One big advantage of the spotted oligonucleotide microarray is that only sequences, not biological samples are needed to create it. Nimblegen has the technology to make custom oligonucleotide microarrays with a capacity of almost 400,000 spots per array. With this technology, we were no longer limited to using the ESTs that we could sequence, but we could expand the scope of our project to include any *N. benthamiana* sequences we could find and any other sequences that we were interested in. An oligonucleotide array would, in the long run, give us more flexibility, better coverage of the genome, and save valuable time by avoiding the

labor intensive process of creating a cDNA array *de novo*. It was these facts that persuaded us to employ the Nimblegen system, rather than create our own array.

Materials and Methods

EST libraries and sequencing efforts

Three different *Nicotiana benthamiana* EST libraries, one standard, one normalized and one subtracted, were sequenced as a part of this project. The standard library was composed of ESTs from expanded leaf tissue made commercially by Amplicon Express, Inc (Amplicon Express, Inc., Pullman, WA). The normalized library was made from the green aerial parts of *N. benthamiana*, by Dr. Andy Maule's laboratory at the John Innes Centre, Norwich, UK (Soares et al., 1994). HAP columns were used to normalize the double stranded cDNA three times. The last EST library, a subtraction library against the tobacco pathogen, *Collectotrichum destructivum*, was donated to the project by Dr. Paul Goodwin, of the University of Guelph, Guelph, Canada. A large portion of these three libraries were sequenced at the Genome Research Lab (GRL) on Centennial Campus, at North Carolina State University, Raleigh, NC. Approximately 7,000 clones from the Amplicon Express library and 900 clones from the Canadian library were sequenced at the GRL. Prepped DNA was sequenced on an ABI3700 (Applied Biosystems, Foster City, CA). To increase the number of sequences we had to work with, we sent 4,992 clones from the US library to Agencourt for sequencing. In addition to sequencing part of the JIC library at the GRL, we had access to all other *N. benthamiana* ESTs that Andy Maule's group had sequenced at the JIC.

Other sequences

Since we were no longer limited to clones we had sequenced, we wanted to use as many *N. benthamiana* sequences as there were available. Dr. Robin Buell of The Institute for Genomic Research (TIGR) had undertaken a similar *N. benthamiana* EST sequencing project. Dr.Buell's group, as a part of a Solanaceae Genomics project, sequenced almost 19,000 *Nicotiana benthamiana* ESTs and deposited the sequences into GenBank. At the same time we performed a thorough search of GenBank which yielded another 4,000 *Nicotiana benthamiana* EST sequences. In total, over 22,000 *Nicotiana benthamiana* EST sequences were pulled from GenBank to be included on this microarray.

There is a hypothesis within the virology community that RNA viruses, regardless of the host, be it plant, animal or prokaryote, regulate the expression of a common core suite of host genes, altered to allow the RNA virus to replicate and establish an infection. A search of the literature was performed looking for host-pathogen interaction studies with a focus on positive sense RNA viruses infecting an animal or plant host (Bigger CB et al, 2001, Labrada L et al, 2002, Saha S and Rangarajan PN, 2003, Shih SR et al, 2004, Bigger CB et al, 2004, Kapadia SB and Cisari FV, 2005, and Panavas T et al, 2005). Any non-immune system genes mentioned as being differentially expressed were considered a gene of interest (GOI). We pulled the sequence of these GOI's, ran them through Basic Local Alignment Search Tool (BLAST) against the GenBank plant database to obtain plant orthologs and these orthologs, were included in the sequence set submitted to

Nimblegen (Altshcul et al. 1990). In some cases the top two or three orthologs were pulled to be included on the array. Additionally, we had access to a set of genes from Dr. Bob Kelly's laboratory at NCSU. Dr. Kelly's laboratory had a data set from *Drosophila melanogaster* infected with Sindbis virus that showed differential expression during the infection. The *Drosophila* ESTs were processed in a similar fashion to what has been described above. A total of 703 non–*Nicotiana benthamiana* plant orthologs oligonucleotides were to be included on the array.

. Lastly, we included the sequence of the challenge virus, *Red clover necrotic mosaic virus* (RCNMV), RNA1 and RNA2, in the data set sent to Nimblegen to be incorporated into the microarray. Knowing ahead of time that our experiments would focus on the interaction between RCNMV and *N. benthamiana*, we felt that these two sequences would serve as a valuable internal control to check on the progression of the viral infection.

Sequence processing

After eliminating all poor quality sequences, the in-house sequenced *Nicotiana benthamiana* ESTs were vector trimmed using a script created in-house by programmers in the Tobacco Genome Initiative (TGI). Vector trimmed sequences were run through RepeatMasker to screen for "interspersed repeats and low complexity DNA sequences" (Smit, et al. 1996). These ESTs were combined with the remaining *N. benthamiana* sequences from TIGR and GenBank, and assembled using Cap3 (Huang and Madan, 1999). The end result was 13,014 unique sequences, 5,948 contigs and 7,066 singlets. The assembled sequences were

submitted to Nimblegen along with the previously mentioned control genes and 703 plant orthologs. It is important to note that the 703 plant orthologs and the two control genes were not included in the clustering process. In total, 13,719 sequences were submitted to Nimblegen for inclusion on the *N. benthamiana* microarray.

Array design

Nimblegen Systems Inc. was selected to create our custom *N. benthamiana* microarray because at the time they had the unique ability to spot almost 400,000 oligonucleotides per array. Given the choice of short (24mers) or long (60mers) oligonucleotides, we chose to use the longer 60mer oligonucleotides. We felt that the longer oligonucleotides would allow for better coverage, give a stronger signal and have increased sensitivity (Buck and Lieb, 2004). 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 are evaluated for uniqueness in two different ways. The first evaluation technique, a simple frequency count, involves counting the number of times each probe appears in the genome of interest. A more thorough, second way of evaluating the probes is a three step process. First, each probe is compared to the target genome and the probe is 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 oligo is calculated by comparing the oligo to its reverse compliment. Any probe that is greater than 60% self-complimentary is eliminated. Third, probes that require a

large number of cycles to synthesize are eliminated. The scores from the evaluation are combined and rank scored. The highest scoring probe is selected and the rank scores are recalculated. Neighboring oligo scores are decreased as oligonucleotides are selected in an effort to pick evenly spaced probes.

Nimblegen uses ArrayScribe (Nimblegen Systems Inc.) software to arrange the probes on the array. Due to the large capacity of Nimblegen's system, we requested that each EST represented on the array be repeated as many times on the array as possible. This resulted in nine different probes per EST, repeated three times, on the array. After data processing by Nimblegen, 13,415 ESTs of the original 13,719 ESTs submitted were included on the array.

Sequence analysis

We are working with a genome that has not been annotated. Thus, we did not know the function of any of the *N. benthamiana* ESTs. After receiving from Nimblegen a final gene list as to the ESTs included on the array, we needed to determine to the function of these ESTs. It was decided that the simplest, most straight forward way to determine function was to hand annotate the sequences by taking the top BLAST hit against the plant database in GenBank for each EST, with preference given to hits from a dicot and if possible a member of the Solanaceous family. A top hit with a low e-value score ($<1e^{-05}$), was thrown out. In total, 4,911 ESTs had either a low score or no hit at all and 8,708 ESTs (64%) had significant similarity to an existing sequence. We now had a starting point for a functional

analysis based on our microarray experimental results, which is to be discussed in Chapter 3.

While the manual hand annotation was a starting point, in order to further elucidate a function for each assembled sequence, it was determined that a more accurate way of categorizing the genes on the array would be to use a Gene Ontology (GO) identifier to describe gene function (www.geneontology.org). With the GO system, gene classification is broken down into three categories: molecular function, cellular component and biological process. Members of the TGI have an algorithm in place that assigns GO categories to genes based on sequence homology to known, annotated genes in GenBank (Diener et al., 2004). As a part of the algorithm, input sequences were BLAST searched against the non-redundant (NR) protein database and previously annotated GO terms from the NR proteins were transferred to our sequences based on an e-value (Expect value) limit. Of the over 13,000 input sequences, 6,325 returned at least 1 GO hit, which is almost half of our starting data set. we were able to get some sort of GO term for 6325 sequences, approximately 46%. This return rate is not as high as we had hoped for, but with a data set of this size, we felt it was more than adequate for the data analysis that we would eventually do, especially when looking at the percentage of differentially expressed genes that returned hits, which will be discussed in chapter 3. In comparison, our simple BLAST seach only had a return rate of 64%.

The three principle GO categories, biological process, cellular component and molecular function had 6,195, 6,214 and 6,139 hits respectively. Approximately 90% of the biological process hits fell into the physiological process category which

includes cell growth/maintenance, metabolism, photosynthesis and death (Figure 2-1A). Under the cellular component, the category cell, which is composed of cell fraction (a generic term for parts of cells prepared by disruptive biochemical techniques; www.geneontology.org), membrane and intracellular, contributed to 88% of its hits (Figure 2-1B). The largest proportion of molecular function sequences (Figure 2-1C) corresponded to catalytic activity, which includes but is not limited to hydrolase, kinase, oxidoreductase and transferase activities, while binding was almost equally represented by 33% of the hits in molecular function.

Conclusion

We constructed a *Nicotiana benthamiana* long oligonucleotide microarray with over 13,000 EST sequences from a number of EST libraries representing several tissue types. Not surprisingly, almost half of the ESTs have no known or predicted function. We believe that this microarray, which is now available to the plant research community as a whole through Nimblegen, has the potential to serve not only as a valuable tool for plant virologists, but for many plant pathologists and plant biologists.

Acknowledgements

We thank Aimee Salstead for help in sequencing the JIC and Canadian cDNA libraries and Mark Burke, Steve Diener and Dr. Betsy Scholl for computational help.

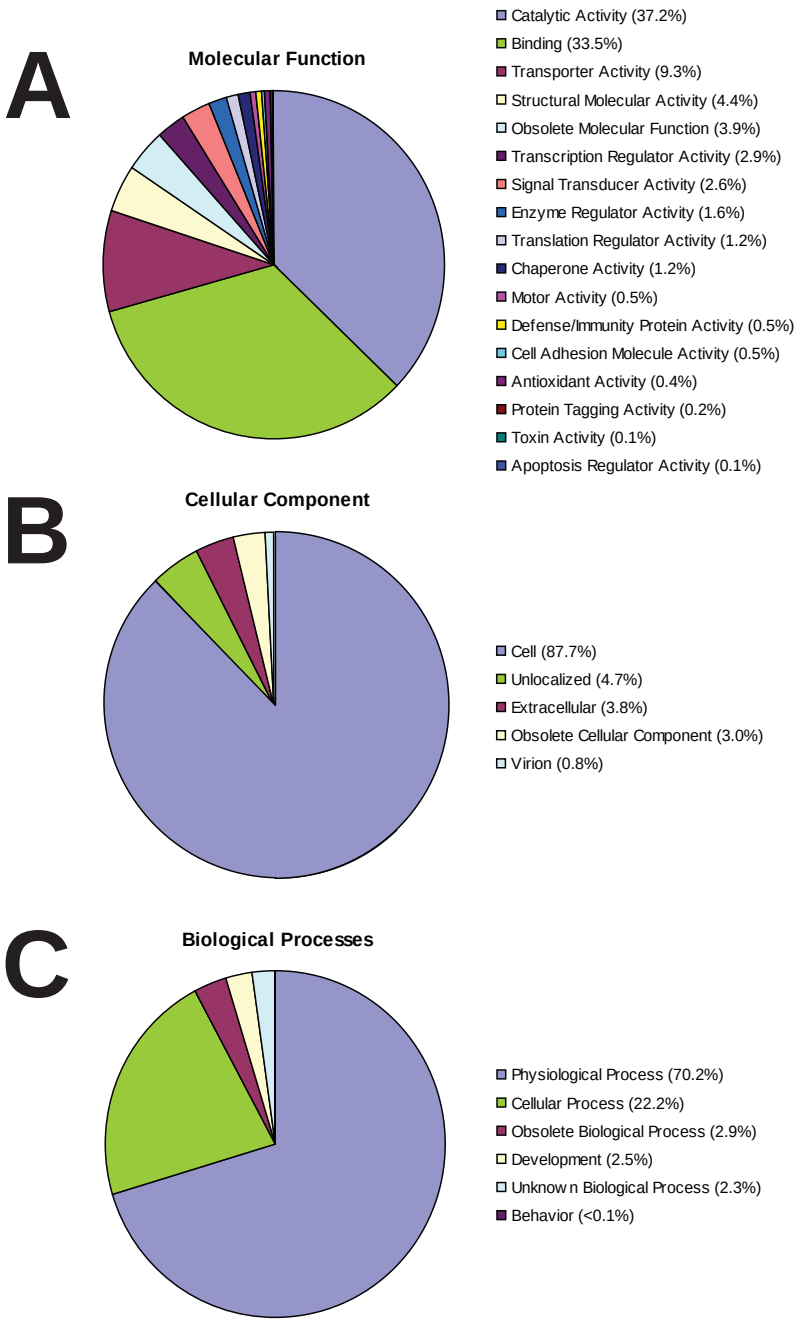

Figure 2-1. Gene Ontology (GO) Classification of All ESTs Represented on *N. benthamiana* Microarray
Each pie chart represents one of three head categories: **(A)** molecular function **(B)** cellular component and **(C)** biological processes.

References

Altschul SF, Gish W, Miller W, Myers EW Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology 215: 403-10.

Bigger, CB, Brasky, KM and Lanford, RE. 2001. DNA microarray analysis of chimpanzee liver during acute resolving Hepatitis C Virus Infection. Journal of Virology, 75 (15): 7059-7066.

Bigger, CB, Guerra, B, Brasky, KM, Hubbard, G, Beard, MR, Luxon, BA, Lemon, SM and Lanford, RE. 2004. Intrahepatic gene expression during chronic hepatitis C virus infection in chimpanzees. Journal of Virology, 78(24):13779-13792.

Buck, MJ and Lieb, JD. 2004. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics, 83:349-360.

Diener SE, Dunn-Coleman N, Foreman P, Houfek TD, Teunissen PJM, van Solingen P, Dankmeyer L, Mitchell TK, Ward M, and Dean RA. 2004. Characterization of the protein processing and secretion pathways in a comprehensive set of expressed sequence tags from *Trichoderma reesei*. FEMS Microbiology Letters 230 (2), 275–282.

Goodspeed, Thomas Harper. 1954. **The Genus Nicotiana**, Chronica Botanica Company, Waltham, MA.

Huang, X. and Madan, A. 1999. CAP3: A DNA sequence assembly program. Genome Research 9: 868-877.

Kapadia, SB and Chisari, FV. 2005. Hepatitis C virus RNA replication is regulated by host geranylgeranylation and fatty acids. Proceedings of the National Academy of Sciences, USA, 102(7):2561-2566**.**

Labrada, L, Liang, XH, Zheng, W, Johnston, C and Levine, B. 2002. Age-dependent resistance to lethal alphavirus encephalitis in mice: analysis of gene expression in the central nervous system and identification of a novel interferon-inducible protective gene, mouse ISG12. Journal of Virology, 76(22): 11688-11703.

Panavas, T, Serviene, E, Brasher, J and Nagy, PD. 2005. Yeast genome-wide screen reveals dissimilar sets of host genes affecting replication of RNA viruses. Proceedings of the National Academy of Sciences, USA 102(20):7326-7331.

Saha, S and Rangarajan, PN. 2003. Common host genes are activated in mouse brain by Japanese encephalitis and rabies viruses. Journal of General Virology, 84: 1729-1735.

Shih, SR, Stollar, V, Lin, JY, Chang, SC, Chen, GW and Li, ML. 2004. Identification of genes involved in the host response to enterovirus 71 infection. Journal of Neurovirology, 10:293-304.

Smit, AFA, Hubley, R and Green, P. RepeatMasker Open 3.0 1996-2004. http://www.repeatmasker.org.

Soares, MB, Bonaldo, MF, Jelene, P, Su. L, Lawton, L, Efstratiadis, A. 1994. Construction and characterization of a normalized cDNA library. Proceedings of the National Academy of Sciences, USA 91: 9228-9232.

Chapter 3

Early changes in host gene expression induced by *Red Clover Necrotic Mosaic Virus*

Early changes in host gene expression induced by *Red Clover Necrotic Mosaic Virus*

INTRODUCTION

Plants are susceptible to a variety of microbial pathogens and classical genetic studies of the plant/pathogen interaction focused primarily on the host defense response (resistance genes, virulence determinants). If the pathogen is a virus, an obligate, intracellular parasite, virologists expect an expression change in not just defense genes, but other host factors as well. This dependence on plant host factors is necessary for its survival and distinguishes the virus from other microbial pathogens. Host factors are essential for all stages of the virus life cycle, including but not limited to translation, replication and movement. For instance, the *Alfalfa mosaic* virus (AMV) capsid protein (CP) interacts with the host eIF4G/eIFiso4G to promote circularization of the viral genome, which is essential for translation (Krab *et al*., 2005). *Tobacco mosaic virus*, strain Cg (TMV-Cg), cannot replicate in *Arabidopsis thaliana* without the presence of transmembrane proteins TOM1 and TOM3 (Yamanaka *et al*., 2002). Pectin methylesterase, a *Nicotiana tabacum* cell wall protein, is necessary for both cell-to-cell and systemic movement of TMV (Chen and Citovsky, 2003; Chen *et al*., 2000; Dorokhov *et al*., 1999). These examples represent only a handful of the host factors documented to interact with the invading positive sense RNA virus

Traditionally, host/virus interaction studies have had a very narrow scope, either investigating an individual host gene or a small group of host genes with a

focus on a previously implicated pathway. Identification of host factors is a slow process when a gene-by-gene approach is adopted, but recent technological advances have sped the discovery of previously unidentified host genes affected by the viral infection. Both the yeast two-hybrid assay and the yeast replication system represent a means to comprehensively search for direct host/virus interactions with an effect on either binding viral proteins or on viral replication. Each of these assays have identified almost 100 host genes (unique to each virus) with a role in the of replication *Brome mosaic virus* and *Tomato bushy stunt virus* (Panavas *et al*., 2005; Kushner *et al*., 2003).

A third way to study host/virus interactions is through the use of DNA oligonucleotide microarrays, which provide the ability to simultaneously scan thousands of genes for host factors affected by the virus. The volume of potential candidate host genes is such that it may even be possible to break down the results into smaller, more manageable sets of information based on the function of the regulated genes. Information gleaned from microarray experiments can be used in follow-up studies to determine if the influence of the virus is the result of a direct or indirect interaction. The downside to microarrays experiments is that any array data should be confirmed, using one of several mRNA quantitation techniques, the most common being quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR; Dallas et al., 2005).

To this day, the majority of the published oligonucleotide experiments studying the interaction between plant and positive sense RNA virus have used the *A. thaliana* GeneChip (Affymetrix, Santa Clara, CA). These studies have

represented a wide variety of interests including *A. thaliana* resistome, the differences between inoculated and systemically infected leaves, temporal effects of a viral infection and even comparisons betweens multiple viruses, and only one of these studies has considered the effects within the first 24 hours. However, these studies are somewhat limited in their applicability to other plant/virus systems because of the narrow viral host range of *A. thaliana*. *Nicotiana benthamiana* is the model host plant for plant/virus interaction despite a much larger genome, a longer generation time and a lack of annotation, for a simple reason of its high susceptibility to a vast range of viruses despite an intact defense system. *Nicotiana benthamiana* can act as a host to *Red clover necrotic mosaic virus* (RCNMV), a well characterized, bipartite, single stranded, positive sense RNA plant virus. Approximately 3 days after infection local necrotic ringspot lesions are observed and systemically infected leaves display a mosaic pattern, leaf curling and overall stunting of growth, seen 4-5 days after inoculation.

In collaboration with Nimblegen Systems, Inc (Madison, WI), we developed a custom oligonucleotide microarray representing over 13,000 *N. benthamiana* ESTs (which, for simplicity, will be referred to as genes throughout this paper), covering an approximated 38% of the transcriptome. We used this array to examine the changes in *N. benthamiana* gene expression induced by RCNMV. Through comparision of the gene expression profiles from early times to the RCNMV life cycle, we have hypothesized why the observed changes have occurred. For simplicity, from this point forward, ESTs will be referred to as genes

MATERIALS AND METHODS

Inoculation procedure

Viral RNA transcripts of RCNMV RNA-1 and RNA-2 were prepared from Sma1 linearized templates as previously described (Xiong and Lommel, 1991) using a T7 Megascript kit (Ambion, Austin, TX). Viral inoculum master mix consisted of 1 µl each of RNA-1 and RNA-2 transcripts and 108 µl of 10mM sodium phosphate buffer (pH 7.0). Mock inoculum was a straight 10mM phosphate buffer. To inoculate, 4 leaves on each of 40 plants were dusted with carborundum and 25 µl of inoculum per leaf was applied for rub inoculation. After inoculation, plants were kept under standard greenhouse conditions (18-20^{o} C) until leaves were harvested.

RNA isolation

One leaf from each inoculated *N. benthamiana* was harvested: 1 untreated, wild-type plant at time 0, mock inoculated plants at 2, 6, 12, 24 hours post inoculation (hpi) and viral inoculated plants at 1, 2, 6, 12 and 24 hpi. In total, 40 leaves were collected. Each harvested leaf was immediately submerged and stored in RNALater (Ambion) at 4^{o} until RNA extraction took place. Total RNA was isolated from 100mg of tissue per leaf using an RNeasy Plant Mini Kit (Qiagen, Valencia, CA) per manufacturer instructions. Isolated total RNA for each timepoint/treatment combination was pooled and DNase treated with Turbo DNA-free (Ambion) to remove any contaminating DNA that may have been left behind. 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 further processing.

cDNA labeling and microarray hybridization

The custom *N. benthamiana* oligonucleotide microarray was designed by Nimblegen (as described in Chapter 2) and manufactured using their proprietary maskless photolithographic technology (Nuwaysir et al., 2002; Albert et al., 2003; Singh-Gasson et al., 1999). All labeling and hybridization experiments were performed by Nimblegen (Nuwaysir EF et al., 2002). Each RNA sample was hybridized on 4 independent arrays (40 microarrays total).

Statistical analysis/Data processing

Microarray scanning, spot finding raw data collection and quantile normalization (Bolstad et al., 2003) was carried out by Nimblegen. On the recommendation of Nimblegen, normalized data was used for our analysis. The three replicate values for each EST represented on the microarray were averaged and then $\log_2$ transformed to give one value per gene. A mixed model analysis was performed using JMP® Genomics (version 6.0.0, SAS) to test for significant differences between mock and viral treatments at the four overlapping timepoints (2, 6, 12, 24 hpi). Originally, a strict Bonferroni correction of 0.01 was used, but data was re-analyzed utilizing the less stringent False Discovery Rate (FDR) cutoff of 0.01 (Benjamini and Hochberg, 1995) .

Functional annotation of significant genes

Differentially expressed genes were initially annotated by taking the top hit from a manual BLASTX against the GenBank plant database and assigning the value to the sequence in question. This hit was assigned to the query sequence only if its e-value was smaller than 1e-05. If there was more than 1 hit with the same score, preferential treatment was first given to values from Solaneaceous plants, followed by those values from dicotelydons. If the top e-value did not meet the threshold value of if the BLAST search returned a value of "no significant similarity found" a null value was assigned to the sequence. A second, more powerful functional analysis was recently performed and it made use of a series of algorithms developed by Diener et al. (2004). The scripts assigns Gene Ontology (GO; www.geneontology.org) descriptors to our EST sequences based on their homology to known, annotated sequences in the NR protein database. These GO descriptors were used throughout the rest of our studies.

Real time qRT-PCR

qRT-PCR was performed on total RNA (DNase treated) from mock and RCNMV infected leaves harvested at 2, 6, 12, and 24 hpi using an ABI Prism 7900HT sequence detector (Applied Biosystems Inc., Foster City, CA). The only difference between this RNA and the RNA used for array hybridization is that plants were grown at a different time of year. For each sample, 40 ng of total RNA was reverse transcribed for each sample and the master mix for each individual reaction consisted of: 0.85 µl Rnase-free water, 2.2 µl $MgCl_2$ (25mM), 1 µl 10X TaqMan RT

buffer, 2 µl dNTPs (2.5mM each), 0.5 µl random hexamers (50mM), 0.2 µl Rnase Inhibitor (20U/mL), 0.25 Multiscribe RTase (50U/mL) using the following conditions: 25°C for 10 minutes, 48°C for 30 minutes, 5 minutes at 95°C. Gene specific Taqman (Applied Biosystems) primers were designed using ABI FileBuilder 3.0 (Applied Biosystems, Inc.) to yield products varying in length from 57 to 105 base pairs (Table 3-1). Two *N. benthamiana* genes (NBENTH001584, NEBENTH006442) were tested for use as an endogenous control for relative quantification. It was determined the 18s rRNA was a stronger endogenous control for the relative quantification studies. Synthezied cDNA was amplified using the custom TaqMan GeneExpression Assays (see chart for primers) and each reaction was performed in triplicate. Each 20 µl reaction consisted of: 10 µl of TaqMan 2X Universal PCR Master Mix, 1 µl of Taqman Gene Expression Assay (20X) and 7 µl of water and 2 µl of cDNA template from the reverse transcription reaction (40 ng total). Thermocycler parameters on the ABI Prism 7900HT were 95°C for 10 minutes followed by 95°C for 15 sec and 60°C for 1 min for 50 cycles.

Northern blot analysis

Several genes tested in the real-time PCR reactions were chosen for further expression analysis using Northern blots, using the same batch of RNA isolated for the real-rime experiments. To synthesize probes, total RNA was reverse transcribed in a 20 µl reaction as follows: 2 µl total RNA, 1 µl of 3' primer (100pmol/µl), 11 µl of dH_20 heated at 70°C for 10 minutes, solution was chilled on ice, and 2 µl Stratascript 10X buffer, 1 µl 10mMdNTPs, 2 µl 0.01M DTT and 1 µl of Superscript

RT (Stratagene, Cedar Creek, TX) were added and the solution was incubated at 42^{o}C for 50 min and followed by a 15 min enzyme denaturation at 70^{o}C. Product cDNA was amplified via PCR using GoTaq (Promega, Madison, WI) with forward and reverse primers designed to ensure nearly complete coverage of the EST sequence in question. PCR master mix consisted of: 10 µl of 5X buffer, 4 µl 10mM dNTPs, 1.5 µl of 5' and 3' primer (10pmol/µl), 1 µl template, 0.25 µl GoTaq and 31.75 µl dH_20. PCR product was run on a 1% TAE gel and gel purified using a Qiaquick Gel Extraction kit (Qiagen). Product DNA was used as the probe and in subsequent transcription reactions to produce a positive control.

To synthesize the positive control, the purified DNA product was transcribed using MEGAScript T7 kit (Ambion) with the same forward primer used for PCR, but the primer also included a T7 promoter at the 5' end. 1% TBE gels were loaded with RNA ladder, the gene specific positive control (from the transcription reaction), and total RNA collected at the timepoint in question from mock and viral infected plants. Each gel was electrophoresed for 90 minutes at 90V, was stained with ethidium bromide and transferred to a nylon membrane for 4 hours using a 10mM NaOH solution. The Amersham Rediprime II Random Prime Labelling System (GE Healthcare, Piscataway, NJ) was used to label the DNA probe. Membranes were pre-hybridized for 10 minutes, hybridized with labeled probed overnight, washed twice for 20 minutes using a 5% SDS solution and washed twice using a 1% SDS solution, all at 65^{o}. The membranes were exposed to either film or a phospo screen overnight (at a minimum) and developed the next day.

RESULTS

Microarray analysis/cumulative changes

The goal of this study was to identify early changes in host gene expression induced by a positive sense RNA virus through the use of the first *N. benthamiana* oligonucleotide microarray. Because visible symptoms are not observed prior to 3 days post inoculation, it was not expected that symptoms of RCNMV infection on the inoculated plants would be observed, nor were they. Since infectious transcripts (raw RNA) were used which are immediately translatable, as opposed to whole virions, it was expected changes in *N. benthamiana* gene expression would occur soon after virus entry. To identify these genes, the simple effect of treatment was tested using a mixed model and compared the mock to the viral infected at each of the 4 timepoints. Comparing mock to viral will cancel out any host changes attributed specifically to the stress of the inoculation procedure. A Bonferroni cutoff of 0.01 yielded 227 differentially expressed genes, less than 2% of the array. It was determined that this was not enough data for a complete characterization of the effects of virus infection Data was re-analyzed using a FDR of 0.01 (which estimates the 1% of the identified genes are false positives) which is less stringent than the Bonferroni cutoff. A total of 1775 genes, 13% of the array, were considered statistically significant for at least 1 of the 4 timepoints tested. The breakdown is as follows: at 2 hours – 366 supressed, 20 induced, at 6 hours – 347 supressed, 393 induced, at 12 hours – 13 supressed, 256 induced, and at 24 hours – 491 supressed, 188 induced (Figure 3-1). Of the 1775 genes, 269 were significant at more than one time point (Figure 3-2). Furthermore, 181 (of the 269) genes showed

a change in direction of response (from induced to supressed or vice versa) over the course of the experiment, with the majority of the direction changes taking place at the two later timepoints, 12 or 24 hpi.

Positive controls

As a positive control, the sequences of RNA-1 and RNA-2 of RCNMV were included on the microarray. RNA-2 was differentially expressed (induced) at all 4 timepoints and RNA-1 was differentially expressed at 6 and 24 hpi (induced). While RNA-1 expression was not statistically significant at 2 and 12 hpi, the direction of expression was induced.

Functional classification of differentially expressed genes

The genome of *N. benthamiana* has not been fully annotated and functional information for any of the ESTs included on the array was not available. The sequence of each differentially expressed gene was BLAST'd against the plant protein database (of GenBank) in an effort to collect some information on gene function and there were hits for 1353 of the 1775 genes (76%). This annotation gave a starting point in an attempt to discover patterns of gene expression based on function. These results were found to be somewhat unreliable and difficult to work with because there are no standards for annotating a submitted sequence and it was difficult to consistently functionally categorize the genes based on the returned annotation value.

After it was determined that the manual BLAST was not an ideal way to annotate our sequences, a search commenced for other available tools that would aid in the discernment of gene function. GoFigure, a web-based application that assigns GeneOntology (GO; www.geneontology.org) descriptors to submitted sequences based on homology to known sequences (Khan *et al*., 2003), was chosen for use in annotation of the ESTs. While this tool returned the desired GO data, its drawbacks included the manual upload of each sequence individually, output files had to be altered to contain all necessary information before they could all be merged together, and a high rate of returning null values. GoFigure simply required too much input time for very little output.

A final annotation was made using a tool researchers working with the Tobacco Genome Initiative (TGI) had developed and implemented. A series of algorithms assigned GO identifiers to the submitted sequences based on their similarity to previously annotated genes. This method of annotation performed similarly to GoFigure, but was capable of processing a large number of sequences and it returned a more user friendly output. Of the 1775 sequences submitted, 1059 had a hit to a known ontology, 395 had a hit to an unknown ontology and 321 did not have a hit to any ontology. Even though the return rate of 60% was less than the 76% return rate from the manual BLAST (its criteria for returning values was more stringent), this classification scheme was immediately implemented. A breakdown of the differentially expressed genes categorized based on the three head ontologies (cellular component, biological process, molecular function) is found in Figure 3-3. It is important to note that a gene may be located in more than one cellular

component, participate in more than one biological process or perform more than one molecular function, so it is highly probable a sequence will be assigned more than one GO term.

Gene trends, by category

At 2 hpi, 95% of the differentially expressed genes are supressed, with the largest groups of supressed genes representing metabolism (87), membrane/cell wall associated genes (81), transport (60), defense (49) and transcription (20) (Figure 3-4). At 6 hpi, the number of supressed and induced are split almost evenly, which is mimicked in the transport (49 supressed, 41 induced) and membrane/cell wall (50 supressed, 76 induced) categories. At this point defense, nucleotide/protein biosynthesis, transcription and translation associated genes are more greatly induced than supressed. Halfway through the sampling timeline, at 12 hpi, host genes are overwhelmingly induced (256 induced, 13 supressed) with defense, metabolism, membrane/cell wall associated and transport again representing the majority of the differentially expressed genes. After 24 hours, supression of host genes has rebounded, outnumbering induced genes almost 2 to 1. Again these results are best represented by defense (128 supressed, 22 induced), membrane/cell wall associated (149 supressed, 20 induced), metabolism (192 supressed, 35 induced), transcription (25 supressed, 8 induced) and transport (97 supressed, 20 induced).

The metabolism category is primarily comprised of protein, nucleotide, carbohydrate and amino acid metabolism, but also includes smaller categories (but

is not limited to) such as alcohol, lipid, phosphorous and organic acid metabolism. Hydrolases, oxidoreductases, transferases, plant hormones and heat shock proteins were included in the defense category as all have been implicated in some sort of defense response.

Animal host homologs

As stated in Chapter 2, approximately 400 animal genes (after further processing by Nimblegen) which have shown differential expression in similar plant/virus array studies were included on the microarray. These animal genes were BLAST'd against GenBank's plant database and the plant homologues were included on the N. benthamiana microarray. A total of 28 unique genes were differentially expressed: at 2 hpi - 2 induced, 7 supressed, 6 hpi -12 induced, 7 supressed, 12 hpi – 2 induced, 1 supressed and 24 hpi – 1 induced, 7 supressed (Figure 3-5). In comparing the trends seen to the entire data set of differentially expressed genes (Figure 3-1), similar trends are observed between 2, 6, and 12 hpi. However at 24 hpi, the animal host genes appear to have a greater proportion of genes induced than supressed by RCNMV, which is the opposite of the trend seen with the total set at 24 hpi. We cannot however, compare these expression levels and direction trends to the original animal host/virus studies they came from since our study evaluated expression within the first 24 hours of infection and that is not the case with the studies from which they came.

Validation of array data by real-time qRT-PCR

To verify the microarray data, qRT-PCR was used on a different batch of RNA than what was used to hybridize the array, although preparation was the same. Two genes (NBENTH001584, NBENTH006442) that showed baseline expression (similar non-significant expression profiles across all 4 timepoints) were tested for use as our positive control. Neither were detected after 40 rounds of amplification and both were dismissed as potential positive controls. Based on the suggestion of Applied Biosystems, the 18s rRNA was used as the internal endogenous control for the relative quantification studies. A minimum of two genes per timepoint were selected for confirmation. Initially 14 genes were chosen, based proposed functions from the original BLAST annotation (with a focus primarily on defense related genes), representing both induction and supression in response to RCNMV: 2 hpi – NBENTH006980, NBENTH010593, 6 hpi – NBENTH002945, NBENTH013208, 12 hpi – NBENTH009724, NBENTH011675, NBENTH006848, 24 hpi – NBENTH010658, NBENTH011163, NBENTH011453, NBENTH000509, NBENTH007463, NBENTH008320. Even though we were interested in each gene at a specific time point, the expression level at all 4 timepoints was tested. Expression levels were analyzed using the SDS 2.2.2 software (Applied Biosystems). One of the 14 genes chosen, NBENTH009724, was not included in our analysis due to its extremely high threshold cycle (Ct) values (>35). Threshold cycle is dependent on the copy number of the starting template indicating the relative abundance of the gene in question (Mutch *et al*., 2001). The 9 genes (NBENTH006980, 010593, 002945, 011675, 006848, 010658, 011439, 11453, 008320) do not show the same

direction of gene expression (induction/supression) as the microarray while 4 genes (NBENTH011163, 000509, 007463, 013208) confirm the microarray results. The problem with these results is that the differences in expression level seen via the qRT-PCR are very small. The only genes that show appreciable differences are NBENTH006980 and NBENTH011675. NBENTH008320 had shown over an 8 fold difference in expression on the microarray, but the difference using qRT-PCR is minute. In general, the qRT-PCR shows that the viral infection does not have a huge effect when compared to the mock infection.

Validation using Northern blots

Six genes from the qRT-PCR pool of genes were chosen for subsequent Northern blot studies (NBENTH000509, 002945, 006848, 006980, 007463, 008320). We first tried to reverse transcribe and PCR amplify each gene using total RNA. Three (NBENTH002945, 007463, 008320) were easily amplified. Numerous tests were performed using a variety of RNA starting concentrations and thermal cycler parameters on the other three genes (NBENTH000509, NB6848, NB6980), but they either did not amplify at the expected size or they simply did not amplify. The three successfully amplified genes were used for our Northern blot experiments. An RNA gel was run, checked for quality of RNA and equal concentration, and transferred successfully to a nylon membrane. After hybridization with dCTP, the same results for each gene were observed: signal for the positive control was strong, but there was no signal detected from the genes of interest, only faint non-specific binding. After many unsuccessful hybridizations, Northerns were re-run using the

MicroPoly(A)Purist™ Kit (to remove rRNA and recover mRNA for use; Ambion) and the NorthernMax®-Gly Kit and BrightStar®-Plus (Ambion) positively charged membranes in an effort to increase sensitivity. The end results were the same, successful detection of the positive control, but no signal from either the mock or the viral RNA samples, only non-specific bands.

Second attempt at validation using real-time RT-PCR and Northern blotting

Because the microarray results had not been successfully confirmed with greater than 50% efficiency, using either qRT-PCR or Northern blots, a second attempt to confirm the microarray results was made. Three new genes were chosen for validation, NBENTH000030, NBENTH000238, NBENTH006303, based on fold change as opposed to function. A third batch of RNA was isolated for these assays. We were interested in NBENTH000030 and NBENTH006303 at 2 and 24 hpi, and NBENTH000238 at only 2 hpi. Results at 2 hpi for all 3 genes did not match the microarray results, but at 24 hpi both genes of interest did match the microarray results. Since qRT-PCR did not conclusively correlate the microarray results, we also attempted Northern blotting for these 3 genes. Amplification of NBENTH000030 and NBENTH000238 was unsuccessful and there was poor amplification of NBENTH006303, so it was not advantageous to continue with the Northern blotting.

DISCUSSION

Experimental optimization

It was important to eliminate as much biological variation as possible in the total RNA used for hybridization of our custom *N. benthamiana* microarrays because this type of experiment will simultaneously quantify thousands of host transcripts. The inoculation system in place in our lab is well established and has been optimized over a 15 year period to ensure maximum infectivity. Some labs choose to work with whole virions but we have found that infectious *in vitro* RCNMV transcripts produce a more uniform infection than whole virions. The quasispecies nature of virions, in general, combined with the population distribution of RCNMV virions (2:1 mixture of virions containing both RNA-1 and RNA-2 vs. RNA-2 only virions; Basnayake *et al*., 2006) makes it more difficult to produce a clean, consistent infection. The amount of RNA used in the inoculation experiments can be standardized when *in vitro* transcripts are used and depending on the nature of the study, the ability to alter the ratios of RNA-1 and RNA-2 does exist.

4-5 week old *N. benthamiana* plants were chosen for inoculation studies as these were determined from past infection studies to be optimal for RCNMV infection. Leaves older than this stage will have a mature, developed defense system and are thus, less susceptible to infection. Leaves younger than this stage are rather small and have not yet completed the sink to source transition and are thus limiting to virus spread across the leaf (Roberts *et al*., 1997). An effort was made to ensure plants exhibiting the same physiological characteristics were inoculated by choosing similar sized leaves.

During the RNA isolation procedure an attempt was made to eliminate variation as much experimental variation as possible. As stated earlier, 4 leaves/plant and 4 plants per treatment/timepoint combination were inoculated. We chose 1 of the 4 inoculated leaves from each plant from which to isolate RNA. We pooled the four RNA samples for each treatment to eliminate or reduce any biological variation between plants that may have remained after our extraction procedure.

The optimal time of year for RCNMV infectivity on *N. benthamiana* is during the spring and fall. Due to timing constraints, all inoculations could not be optimized. RNAs used for microarray hybridizations by Nimblegen were isolated from plants infected in June. Almost all, if not all, of the RNA was sent to Nimblegen for processing. RNA used for the primary validation studies, was isolated from plants infected in October, introducing an unwanted and unintentional source for variation. Even if there had been enough RNA leftover from the June inoculations, reproducibility issues would have necessitated harvesting RNA from other inoculations for subsequent validation studies. While the Phytotron facility at NC State does provide a controlled environment for plant growth, it has never been optimal for RCNMV infectivity studies (possibly due to lighting?).

Implications of observed gene expression profiles

RCNMV, like all viruses, is an obligate parasite and thus relies on the active respiration, protein synthesis and metabolic pathways of their hosts for survival. The nature of the experimental design precludes the discrimination of RNA source. Thus,

any observed RNA profiles are a combination of transcripts expressed in infected and neighbouring non-infected cells. Additionally, microarray assays cannot distinguish between changes in transcription vs. changes in RNA degradation. Translation of RCNMV can begin as soon as the infectious transcripts enter the cell. At 2 hpi we observe a predominantly down-regulated response by *N. benthamiana* which may represent basal non-specific efforts by *N. benthamiana* to shut down the site of infection produced by any given pathogen. Limitation of access to necessary host factors required for translation and replication would prevent viral infection prior to the onset of an active defense response such as the hypersensitive response. Included in this non-specific type of general response are the decrease in expression of carbohydrate, nucleotide and protein metabolic genes with smaller representation from phosphorous, lipid, and amino acid metabolic pathway genes. It should be noted again that the infection was initiated with viral transcripts. If whole virions had been used as the inoculum, an added disassembly step would have been required prior to translation. The presence of whole virions/viral CP might have triggered a different host response and/or altered the timing of the observed expression profile. In the case of AMV, CP needs to be included in the inoculum, either as intact virions or provided from a co-inoculated sgRNA from which it can be translated, to establish infection (Bol *et al*., 1971).

Any early non-specific attempts by *N. benthamiana* to prevent the spread of RCNMV has failed by 6 hpi. By this time point, the virus is actively translating and replicating and to do so, has taken over host processes for its own benefit resulting in increased levels of host gene expression. RCNMV replication causes

endoplasmic reticulum (ER) membrane proliferation (Turner et al., 2004) which would possibly explain the increase in expression of cell wall/membrane associated proteins as well as those for metabolism and nucleotide/protein biosysnthesis. Movement protein (MP) is an early translation product of RCNMV, while CP is a late translation product. Studies utilizing a GFP reporter in place of the CP detected movement at 12 hpi (T. Sit, personal communication). Thus, it is reasonable to conclude that wild-type RCNMV movement has already started before the 12 hpi time point. As a result, the newly infected cells would be repeating some of the gene expression changes observed at 2 hpi such as the general down-regulation of gene expression.

As infected cells communicate with non-infected cells regarding the RCNMV attack, signals are sent through the plant and defense genes are expected to be up-regulated. This movement of metabolites and defense responses would also necessitate an increase in transport mechanisms. Interestingly, Takeda *et al*. (2005) have shown that RCNMV requires *dcl1* (a host gene involved in RNA silencing) for replication. This co-opting of a host defense gene may be a general countermeasure employed by viruses to circumvent host defenses.

A dramatic drop in the number of differentially expressed genes that are suppressed is the most striking characteristic of the 12 hpi sampling. While the categories expressing significantly upregulated genes are the same as observed at 6 hpi, there is minimal overlap between specific genes. At this time point, the RCNMV life cycle is progressing towards completion with virion formation which would sequester viral RNAs from participating in any further rounds of

translation/replication. We are uncertain of the implications that this stage of the virus life cycle would have on host gene expression if we assume that viral RNA is being encapsidated.

As seen in Fig. 3-1, there is a radical switch from predominantly up regulation in a few key pathways at 12 hpi to predominantly down regulation at 24 hpi. The profiles observed at 24 hpi mimic those at 2 hpi, except the magnitude of differentially expressed genes has almost doubled. At this stage of the viral life cycle, there is a build up of virus titer in the initially infected cells and the initiation of infection into neighboring cells. It is assumed that secondary and tertiary infection of cells by RCNMV has started, resulting in a reoccurrence of the expression profile observed at 2 hpi in addition to the differential expression of previously non-affected genes which are probably the result of increased viral movement.

Possible overlap of animal and plant host factors

Based on the ability of yeast cells to support the replication of both plant and animal infecting viruses, it can be postulated that viruses, which infect hosts from different kingdoms, utilize common host components during their life cycles. One of the goals of this project was to test this hypothesis. After a survey of the literature describing studies investigating host response to a viral infection in animals through the use of microarrays, plant homologs of these cited animal host genes were included. Only 28 of the over 400 genes representing these plant homologs included on our *N. benthamiana* microarray were differentially expressed. The function of these plant homologs is wide ranging (Table 3-2), including a putative development

protein, nucleolin, a histone, and not surprisingly a wound induced kinase. With such a variety of functions, we were not able to form any distinct groupings of these genes. The function of a gene in an animal system will not necessarily translate to the same function in a plant system. Also, we took homologs based on sequence similarity, not functionally similarity. At this point, we can only conclude that our plant/animal hypothesis has not been disproven by our results and more rigorous testing should be performed before dismissing this hypothesis.

Validation studies

Today's standards for scientific publication almost always require that microarray studies include independent mRNA quantitation studies such as RNAse protection assays, Northern blotting or qRT-PCR, because as a high-throughput technology they tend to be error prone. qRT-PCR is considered to be the "gold standard" due to its higher sensitivity, requirement for less input RNA, rapid quantitation and upscalability. The results from our qRT-PCR studies are not uncommon. Previous studies have reported a range of 13-16% up to 27% poor correlation (Dallas et al., 2005; Morey et al., 2006), however the extremes we observed (67% of the tested genes did not correlate) have not been reported in the literature. Liu and Slininger (2007) expressed concern over normalization techniques which use of housekeeping genes as the reference sample for qRT-PCR, but this is thought not to be an issue with the use of 18s rRNA as a reference sample due to the tight, overlapping amplification plots observed. The inconsistencies we observed

could be a result of the use of different controls for normalization in the two compared assays (microarray and qRT-PCR; Liu and Slininger, 2007).

Additionally, it is believed the number of genes, and subsequently experiments, that fail to correlate is greater than that which is reported. In our survey of similar microarray studies, 1 did not include validation experiments and 2 stood out due to their minimal validation data. In the spatial analysis of the response to a viral infection, researchers simply state that qRT-PCR assays of 2 genes (out of 500 that were differentially expressed) confirmed the trends seen in the microarray experiments (Yang et al., 2007). In Dardick's study (2007) of stone fruit viruses, he chose 38 genes for validation studies, but before performing qRT-PCR, he designed primers for and tested for first strand synthesis. Only after he successfully amplified cDNA fragments for 11 of the 38 genes did he begin his real-time experiments. Perhaps if we had taken this approach in narrowing down our candidate gene list, we could have had a higher correlation rate with our qRT-PCR assays.

Since the qRT-PCR results did not have a high rate of correlation to the microarray results, we attempted to amplify and hybridize 9 different *N. benthamiana* genes via Northern blot analysis. We were able to successfully PCR amplify and synthesize DNA and RNA for use as probe and positive control for 3 of the 9 genes, but attempts to hybridize failed. Even after we selected for mRNA through the use of a Poly(A) Purist kit (Ambion) and increased binding specificity with the NorthernMax-Gly kit (Ambion). We hypothesize the lack of signal should be attributed to the fact that these genes are low expressors. Over 50% of the processed sequences used to create our microarray came from normalized libraries, which reduces the frequency

of repeated sequences and enriches for rare and unique genes, lending strength to our hypothesis.

A final consideration as to the low correlation rates is the RNA used for microarray hybridization was collected in June, a suboptimal month for infectivity, while RNA used for Northern analysis and qRT-PCR was collected in October. If baseline *N. benthamiana* gene expression is altered by the season in which it is grown, as it is suspected, another source of variation has inadvertently been introduced into the experimental system, which could affect the quantitative expression assays. At this point, it is difficult to pinpoint one specific reason for a lack of correlation between the microarray data and either the Northern blots or the qRT-PCR.

Conclusions

We have not only created the first *N. benthamiana* oligonucleotide microarray, but we are only the second group to use a microarray for the study of early effects of the plant/virus interaction using a single stranded, positive sense RNA virus. Here we hypothesize as to the reasons why we see the changes we do at 2 and 6 hpi and we are able to relate observed changes at 12 and 24 hpi to our current knowledge of plant/virus interactions and RCNMV. While we have not yet been able to successfully validate our microarray data to our desired standards, we believe that our study has uncovered potential information that can easily be explored with future studies using tools already available.

Future work

The microarray data set genereated in this study can be further analyzed statistically to determine the effect across/between timepoints. The sequence data can be annotated using updated GO information available in GenBank, possibly allowing researchers to study parts of or entire pathways in *N. benthamiana* affected by the viral infection that can be compared to information already available for *A. thaliana*. Furthermore, comparisons can be made to time=0 mock inoculated plants to determine the affect of the inoculation procedure on host gene expression.

RCNMV is a highly flexible system lending itself to many potential future studies. Comparison of RCNMV RNA-1 co-inoculated with a movement deficient RNA-2 mutant vs. wild-type RNA-2 will allow us to examine the effect of MP expression on host gene expression. RNA-1 with the RNA-2 mutant will not spread beyond the initially infected cell and thus, we would not expect to see the same profiles at later time points. The availability of transgenic *N. benthamiana* plants constitutively expressing the RCNMV MP protein (MP+ plants) allows for the study of RNA-1 only infections uncoupling the effects of movement from replication. This will also allow us to examine the regulatory effect of RNA-2 on RNA-1 functions since RNA-2 is known to directly interact with RNA-1 via the *trans*-activator element (Sit et al., 1998). The lack of a replicating RNA-2 in these types of experiments would preclude CP production (late gene expression) and thus, virion formation. It will be interesting to observe the changes in host gene expression if late viral gene expression is eliminated. Co-inoculation with a mutant RNA-2 that does not produce MP will allow CP production and encapsidation (Vaewhongs and Lommel, 1995)

uncoupled from the direct translation of RNA-2 which is believed to be required for RNA-2 replication (Mizumoto et al., 2006).

One obvious experiment would be the use of virions as the source of incoulum. As mentioned previously, virions would add an extra layer of complexity to the study, namely the disassembly step which may or may not involve host components. Additionally, the affect of CP on host gene expression (eg. Defense responses) during the initial infection can be examined. While we would expect the changes evidenced in the transcript infected plants to come at later time points since disassembly of the virion needs to occur before translation can begin, that may not be the case.

Acknowledgements

We would like to thank Dr. Jason Osborne for his assistance in designing the statistical model used, Steve Diener for his help with the GO annotation, Drs. Chris Halweg, Daniel Menendez and Mike Resnick of NIEHS for advice and equipment use in the real-time qRT-PCR experiments.

Table 3-1. qRT-PCR Primers

Nimblegen ID/In house ID	Orientation	Sequence
NBenth000030/SAL_AGN012xa24f1	forward	CACTGCTATGCATGGAGGTCAT
	reverse	TGGCATCCTGAATGTTGAGGTTT
NBenth000238/SAL_AGN017e03f1	forward	GGCTCCAAACCTCAACATTCAG
	reverse	GCTGTGGTTAGCATGATTTCCAA
NBenth000509/SAL_CAN004xa11f1	forward	GGTGATGATACTGCCCGTAGGA
	reverse	GGAACCACCAGTAGTTTCATGAGAA
NBenth001584/SAL_UKX185xg23f1	forward	TCACTCGAATAACGGACTAAAGATTTCTTTT
	reverse	GCGCGCCTACTCACTAGTT
NBenth002945/SAL_UKX092xo17f1	forward	ACCCCTCGATCTGTCATAGCT
	reverse	CATCGCTGCAGGTGCAT
NBenth006303/gi|39860806|	forward	TCCAGCAGATAAGGCAAATGAAATCA
	reverse	TCTGTTTGGTGTTGTTGTTGTTGTTT
NBenth006442/gi|39859344|	forward	GCCTCATGCTATAATCCATAGAGATTTGA
	reverse	CTGTTGTGCAATTTTACTGAGTCCAA
NBenth006848/gi|39851555|	forward	GGAACATGTGGTTGCAATTCAAGAT
	reverse	GGAAACCCTTCTTCGCCTTAAAGTA
NBenth006980/gi|27374987|	forward	CGCAAATGTCCTTCATAGAGATCTCA
	reverse	GCCTAGCAAGACCAAAGTCACATAT
NBenth007463/Contig397	forward	TGGCGCAGATTCAGGTTCAG
	reverse	GCCGCCACACCATTTCC
NBenth008320/Contig1254	forward	CCATTCGGCTCTGGAAGAAGA
	reverse	TGTGCGACCACCAAACGAA
NBenth009724/Contig2658	forward	GGGCCGAGGGTCGAAA
	reverse	GGTTTCTTCTTGTGAGGTACTTTTAGTTCAA
NBenth010593/Contig3527	forward	GCAATCATTGTCTCCGTCTTCCT
	reverse	GCCAGAGGAGAGTATCGAAACC
NBenth010658/Contig3592	forward	TCCAAGAGCTACAGGGTGTTCT
	reverse	CCATTAGCTGCAGGTGAAATACCAA
NBenth011163/Contig4097	forward	GGCCAAGGATCGTAATGATGAGA
	reverse	CATCTCCCCATGGAAGTCTACTATCT
NBenth011439/Contig4373	forward	GCCATGCTGCAGCTTGATC
	reverse	CTTATGGCTGTGCACAAGCA
NBenth011453/Contig4387	forward	ACCACCGACTCGCGAATTAC
	reverse	CGTACGATTCAACGACGATGGT
NBenth011675/Contig4609	forward	ATGGTTACACACAATATGAGGCTGAA
	reverse	GACATACATTTCCTTGTGATGCTCATTT
NBenth013208/gi|21325919|	forward	CAACACTTCTCAGTTTGGGCCTAATA
	reverse	CGACTTTGGTATGATGTCCTTCCA

Table 3-2. Animal Genes Showing Differential Expression and Their Plant Homologs

Original Gene	Organism	Function	Plant Homolog
DID2	*S. cerevisae*	vacuolar sorting protein	*N. benthamiana* putative development protien
YNL321W	*S. cerevisae*	Putative cation exchanger	*A. thaliana* calcium antiporter, CAX2
AAF47024	*D. melanogaster*	no known function	*A. thaliana* exostosin family protein
AAF50251	*D. melanogaster*	no known function	*Brassica nigra* PSR9
NP_733453	*D. melanogaster*	transposable element, isoform D	*N. tabacum* NNF1, tobacco nuclolin
J01430	*Neurospora crassa*	cytochrome oxidase III	*Brassica napus* mitochondrial DNA
M13755	*Homo sapiens*	Human interferon-induced 17-kDa	*N. tabacum* Ubi U4
AB000409	*Homo sapiens*	MNK1, a new MAP kinase-activated protein kinase	*N. tabacum* calcium dependent kinase
X66401	*Homo sapiens*	TAP1, TAP2, LMP2, LMP7 and DOB, major histocompatibility complex	*A. thaliana* proteosome subunit
J03909	*Homo sapiens*	gamma-interferon-inducible protein precursor	*A. thaliana* GILT family protein
M15796	*Homo sapiens*	cyclin protein	*N. tabacum* proliferating cell nuclear antigen
M13444	*Mus musculus*	alpha-tubulin isotype M-alpha-4 mRNA	*N. tabacum* alpha tubulin
M35131	*Mus musculus*	neurofilament component (NF-H) mRNA	*N. tabacum* NNF1, tobacco nuclolin
U50631	*Mus musculus*	heat-responsive protein (HRP12)	*A. thaliana* AT3g20390
M21574	*Homo sapiens*	Human platelet-derived growth factor receptor alpha (PDGFRA)	*N. benthamiana* NbWIK
J04596	*Mus musculus*	platelet-derived growth factor-inducible KC protein	*N. silvestris* mRNA for the small subunit of RuBisCo
U49350	*Mus musculus*	CTP synthetase mRNA	Malus x domestica CTP synthase 1a mRNA
M36084	*Mus musculus*	cytosolic malate dehydrogenase (cMDHase) gene	*N. tabacum* malate dehydrogenase
L10244	*Mus musculus*	spermidine/spermine N1-acetyltransferase (SSAT) mRNA	*N. tabacum* S-adenosylmethionine decarboxylase gene
U34277	*Mus musculus*	PAF acetylhydrolase mRNA	*A. thaliana* putative pectinacetylesterase (At3g62060)
U29539	*Mus musculus*	retinoic acid-inducible E3 protein mRNA	Potato gene for granule-bound starch synthase
U36220	*Mus musculus*	FK506 binding protein 51 mRNA	*N. benthamiana* P58IPK mRNA
J03482	*Mus musculus*	histone H1	*Cicer arietinum* histone H1
M22432	*Mus musculus*	protein synthesis elongation factor Tu (eEF-Tu, eEf-1-alpha)	*N. tabacum* mRNA for elongation factor-1 alpha
AI369284	*Homo sapiens*	EST - permidine/spermine N1-acetyltransferase (SSAT) mRNA	*Santalum album* proline rich protein mRNA
AA664180	*Homo sapiens*	EST - PLASMA GLUTATHIONE PEROXIDASE PRECURSOR	*N. tabacum* mRNA for glutathione peroxidase NtEIG-C08
AA121387	*Homo sapiens*	EST -similar to contains element MER22 repetitive element	*N. alata* arabinogalactan protein (Agp4) gene
AA411619	*Homo sapiens*	EST - no function given	*N. tabacum* mRNA for cysteine proteinase precursor

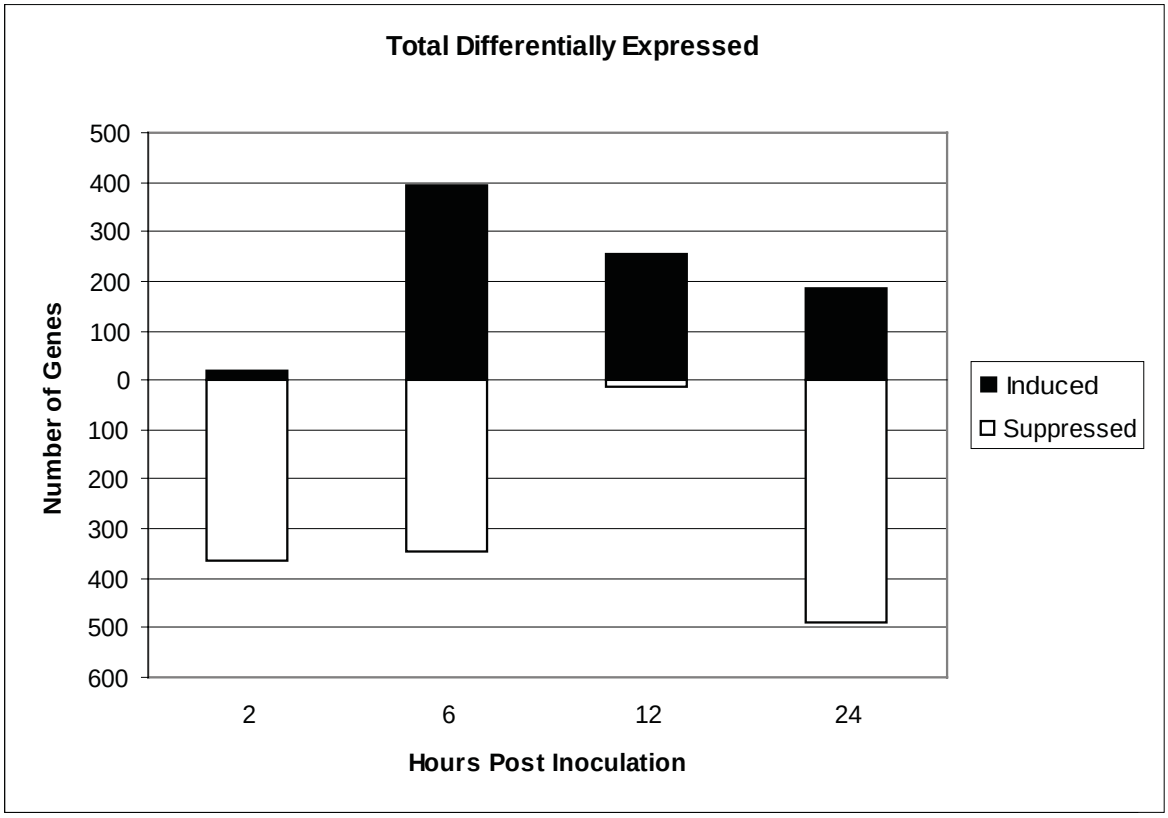

Figure 3-1. Summary of Differentially Expressed Genes Over Time
Bars represent the total number of induced and suppressed genes following inoculation by Red Clover Necrotic Mosaic Virus. (FDR=0.01)

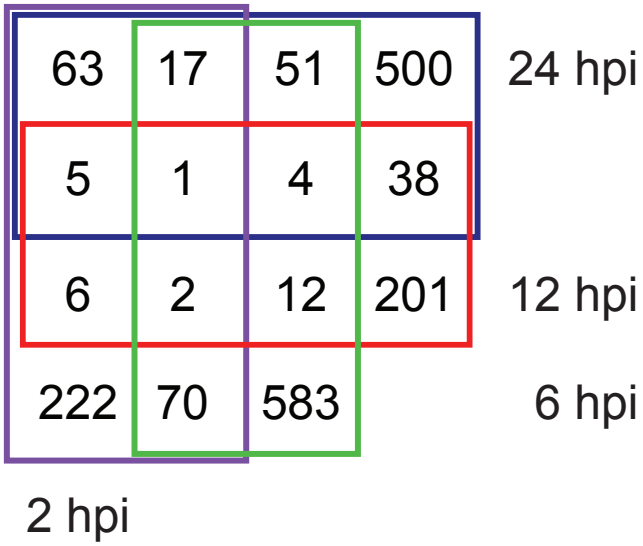

Figure 3-2. Venn Diagram Representing the Overlap of Gene Expression at All Four Timepoints

The purple box contains all genes differentially expressed at 2 hours, the green at 6 hours, the red at 12 hours and the blue at 24 hours. Values in the non-overlapping sections represent differentially expressed genes unique to that particular timepoint, and values within overlapping regions represent genes expressed at more than one time point. For example, 63 in far upper left corner means that 63 of the same genes were significant at 2 and 24 hpi and only 1 gene was differentially expressed at all 4 timepoints compared.

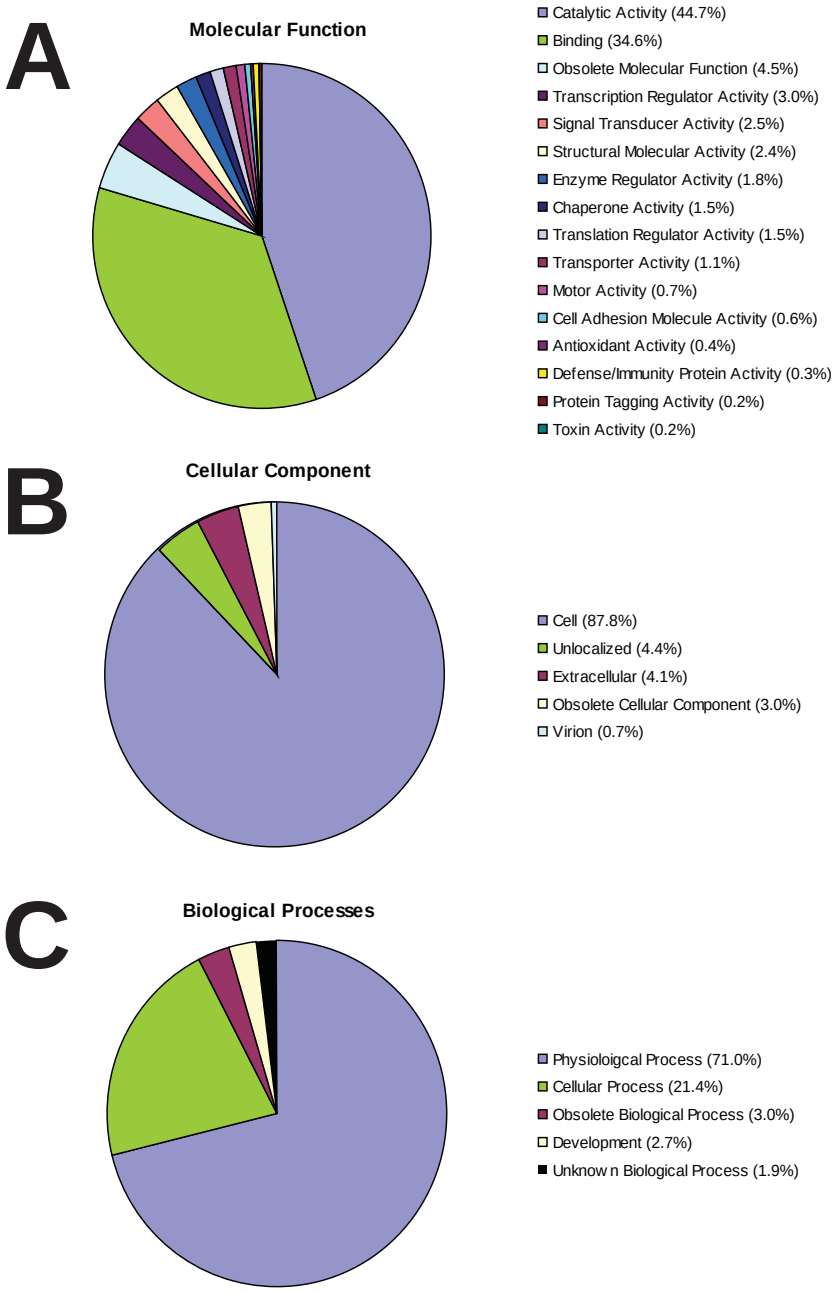

Figure 3-3. Distribution of Statistically Significant Genes According to Gene Ontology (GO) Classification Scheme
Each pie chart represents one of three head categories: **(A)** molecular function **(B)** cellular component and **(C)** biological processes.

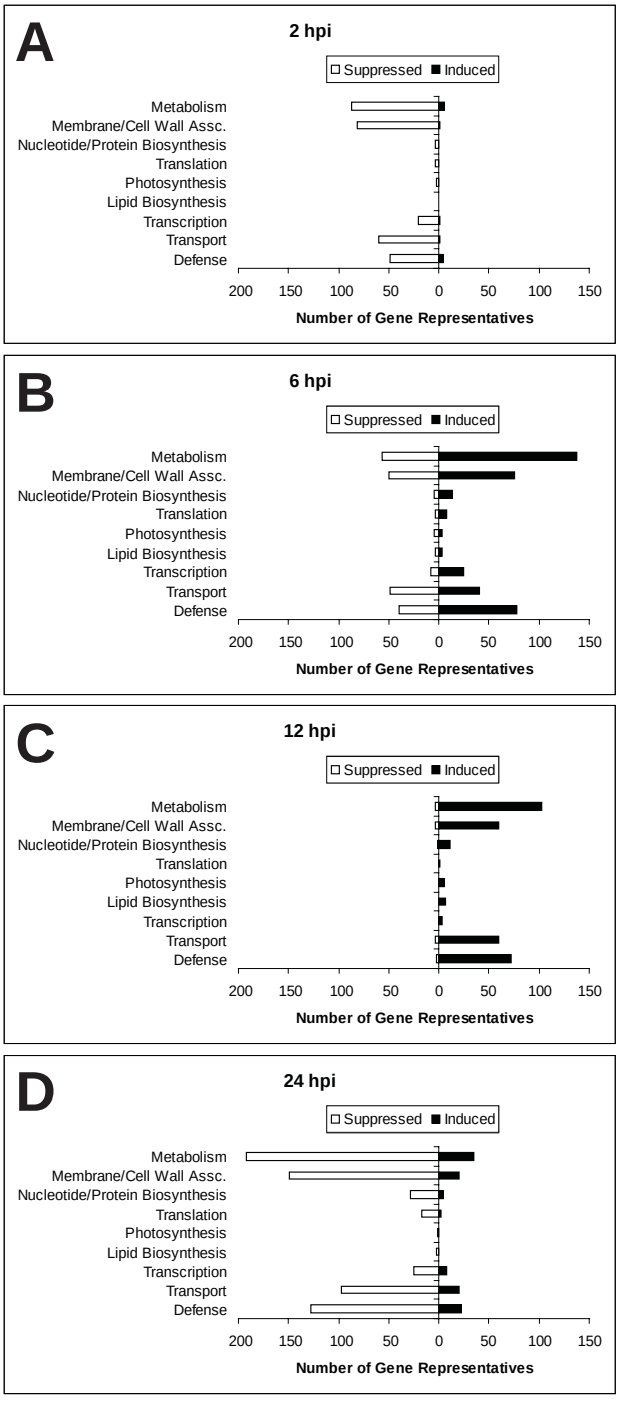

Figure 3-4. Impact of RCNMV on Top Functional Categories
Genes were assigned to categories based on keyword searches with geneontology results. Bars represent the number of genes induced or suppressed by RCNMV at 2 hpi (A), 6 hpi (B), 12 hpi (C), or 24 hpi (D).

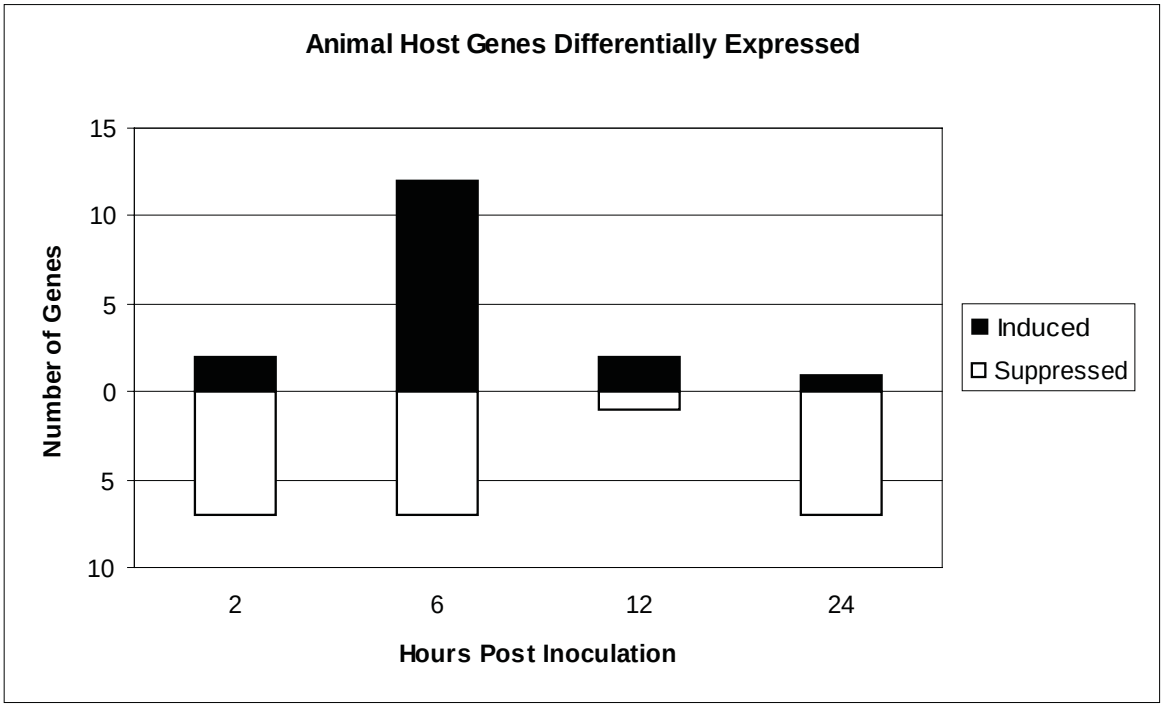

Figure 3-5. Animal Host Genes Differentially Expressed by RCNMV
Expression profiles of the animal host genes included on the microarray that were considered to be statistically significant.

REFERENCES

Albert, T.J., Norton, J., Ott, M., Richmond, T., Nuwaysir, K., Nuwaysir, E.F., Stengele, K.P., Green, R.D. 2003. Light directed 5'->3' synthesis of complex oligonucleotide microarrays. Nucleic Acids Research 31, e35.

Basnayake, V.R., Sit, T.L., Lommel, S.A. 2006. The genomic RNA packaging scheme of *Red clover necrotic mosaic virus*. Virology 345, 532-539.

Benjamini, Y., Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Ser. B 57, 289-300.

Bol, J.F., Van Vloten-Doting, L., Jaspars, E.M.J. 1971. A functional equivalence of top component a RNA and coat protein in the initiation of infection by alfalfa mosaic virus. Virology 46, 73-85.

Bolstad B, Irizarry R, Astrand M, Speed T 2003 A comparison of normalization methods for high density oligonucleotide array data based on bias and variance. Bioinformatics 19, 185-193

Chen, M.H., Citovsky, V. 2003. Systemic movement of a tobamovirus requires host cell pectin methylesterase. Plant Journal 35, 386-392.

Chen, M.H., Sheng, J., Hind, G., Handa, A.K., Citovsky, C. 2000. Interaction between the tobacco mosaic virus movement protein and host cell pectin methylesterases is required for viral cell-to-cell movement. The EMBO Journal 19, 913-920.

Dallas, P.B., Gottardo, N.G., Firth, M.J., Beesley, A.H., Hoffman, K., Terry, P.A., Freitas, J.R., Boag, J.M., Cummings, A.J., Kees, U.R. 2005. Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR – how well do they correlate? BMC Genomics 6, 59.

Dardick, C. 2007. Comparative expression profiling of *Nicotiana benthamiana* leaves systemically infected with three fruit tree viruses. Molecular Plant-Microbe Interactions 20, 1004-1017.

Diener S.E., Dunn-Coleman N., Foreman P., Houfek T.D., Teunissen P.J.M., van Solingen P., Dankmeyer L., Mitchell T.K., Ward M. and Dean R.A.. (2004) Characterization of the protein processing and secretion pathways in a comprehensive set of expressed sequence tags from Trichoderma reesei. FEMS Microbilogy Letters 230, 275-282.

Dorokhov, Y.L., Makinen, K., Yu, O., Merits, A., Saarinen, J., Kalkkinen, N., Atabekov, J.G. and Saarma, M. 1999. A novel function for a ubiquitous plant enzyme

pectin methylesterase: the host-cell receptor for the *Tobacco mosaic virus* movement protein. FEBS Letters 461, 223-228.
Khan, S., Situ, G., Decker, K., Schmidt, C.J. 2003. GoFigure: Automated Gene Ontology annotation. Bioinformatics 19, 2484-2485.

Kushner, D.B., Lindenbach, B.D., Grdzelishvili, V.Z., Noueiry, A.O., Paul, S.M., Ahlquist, P. 2003. Systematic, genome-wide identification of host genes affecting replication of a positive-sense RNA virus. Proceedings of the National Academy of Science, USA 100, 15764-15769.

Liu, Z.L., Slininger, P.J. 2007. Universal external RNA controls for microbial gene expression analysis using microarray and qRT-PCR. Journal of Microbiological Methods 68, 486-496.

Morey, J.S., Ryan, J.C., Van Dolah, F.M. 2006. Microarray validation: factors influencing correlation between oligonucleotide microarrays and real-time PCR. Biological Procedures Online 8, 175-193.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., McCormick, M., Norton, J., Pollock, T., Sumwalt, T., Butcher, L., Porter, D., Molla, M., Hall, C., Blattner, F., Sussman, M.R., Wallace, R.L., Cerrina, F., Green, R.D. 2002. Gene expression analysis using oligonucleotide arrays produced by Maskless photolithography. Genome Research 12, 1749-1755.

Panavas, T., Serviene, E., Brasher, J., Nagy, P.D. 2005. Yeast genome-wide screen reveals dissimilar sets of host genes affecting replication of RNA viruses. Proceedings of the National Academy of Sciences, USA 102, 7326-7331.

Roberts, A.G., Santa Cruz, S., Roberts, I.M., Prior, D.A.M., Turgeon, R., Oparka, K.J. 1997. Phloem unloading in sink leaves of Nicotiana benthamiana: comparison of a fluorescent solute with a fluorescent virus. The Plant Cell 9, 1381-1396.

Singh-Gasson, S. Green, R.D., Yue, Y., Nelson, C., Blattner, F., Sussman, M.R., Cerrina, F. 1999. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nature Biotechnology 17, 974-978.

Takeda, A. Tsukuda, M., Mizumoto, H., Okamato, K., Kaido, M., Mise, K., Okuno, T. 2005. A plant RNA virus suppresses RNA silencing through viral RNA replication. The RMBO Journal 24, 3147-3157.

Turner, K.A., Sit, T.L., Callaway, A.S., Allen, N.S., Lommel, S.A. 2004. *Red clover necrotic mosaic virus* replication proteins accumulate at the endoplasmic reticulum. Virology 320, 276-290.

Vaewhongs, A.A., Lommel, S.A. 1995. Virion formation is required for the long-distance movement of *Red clover necrotic mosaic virus* in movement protein transgenic plants. Virology 212, 607-613.

Xiong Z. and Lommel, S.A. 1989. The complete nucleotide sequence and genome organization of *Red clover necrotic mosaic virus* RNA-1. Virology 171, 543-554.

Yamanaka, T., Imai, T., Satoh, R., Kawashima, A., Takahashi, M., Tomita, K., Kubota, K., Meshi, T., Naito, S., Ishikawa, M. 2002. Complete inhibition of tobamovirus multiplication by simulataneous mutations in two homologous host genes. Journal of Virology 76, 2491-2497.

Yang, C., Guo, R., Jie, F., Nettleton, D., Peng, J., Carr, T., Yeakley, J.M., Fan, J.B., Whitham, S.A. 2007. Spatial analysis of *Arabidopsis thaliana* gene expression in response to *Turnip mosaic virus* infection. Molecular Plant-Microbe Interactions 20, 358-370.