Two newly discovered open reading frames in the genomes of two cucurbit-infecting tobamoviruses

By

SANDRA VU

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Plant Pathology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Bryce W. Falk, Chair

Richard Bostock

Joanne Emerson

Committee in Charge

2021

Table of Contents

ACKNOWLEDGEMENTS

I would like to thank the entire Falk lab, past and present members, for your support during the course of this work. Thank you all so much for listening to me talk relentlessly about my wild western blots, my confocal confusion, and my problem primer designs. Thank you for your words of encouragement and gentle nudges in the right direction. I genuinely appreciate your expertise, time, and patience. I would especially like to thank Dr. Jun Jiang. Jun, you are always working on a million projects at once, but somehow you always find time to help me and everyone else in our lab. Thank you for showing me the wonders of the confocal, sending me papers, and not laughing too hard at my western blots.

To my committee members, Dr. Joanne Emerson and Dr. Rick Bostock: you two are writing and editing rockstars. Thank you for always responding with kindness to my emails and all your help to make this happen.

To my major advisor, Dr. Bryce Falk: thank you for taking a chance on me. I have learned a ton in this lab about science and collaborating/communicating, maybe more than I wanted! I am a better scientist, presenter, and overall person for having worked here.

To my Plant Pathology family: we may not all be working on the same things, but we were definitely all in this together. Thank you for all the happy hours, dinners, lunches, and coffee dates to keep me sane.

And last, but not least: thank you to Andrew Martinez. We started off strong in plant pathology, and you've for sure committed to the bit. Thank you for your endless optimism and encouragement.

ABSTRACT

Two small ORFs were identified through the use of ORF predictors in genomes of two cucurbit-infecting tobamoviruses: Cucumber green mottle mosaic virus (CGMMV) and Watermelon green mottle mosaic virus (WGMMV). One of these ORFs (ORFH) is found in both viruses, while the other ORF (ORFN) is only found in WGMMV. In addition to these two viruses, similar ORFs can also be found in several but not all other cucurbit-infecting tobamoviruses. ORFH overlaps the ORF encoding the 30kDa of CGMMV and WGMMV, but in the +1 reading frame register. Transient expression of ORFH fused to GFP showed fluorescence localization to the endoplasmic reticulum in *Nicotiana benthamiana* epidermal cells. Truncation mutations of ORFH in CGMMV do not affect the virus's ability to infect systemically in *Nicotiana bethamiana* plants, or two melon cultivars Laredo and Expedition. ORFN is located in WGMMV, overlapping the 126 and 183 kDa protein ORFs, also in the +1 reading frame register. The predicted ORFN protein contains a bi-partite nuclear localization signal, which when disrupted changes the nuclear localization to a chloroplast localization. Western blots with an ORFN protein specific antibody show a larger protein than the size predicted for the ORFN protein, suggesting a fusion protein or a protein aggregate. Mutations introduced into the ORFN resulted in mutant WGMMV infections which showed less severe symptoms in *N. benthamiana* and cucurbit plants, but no changes in viral titer for wildtype and WGMMV mutants were observed. Although it has not yet been confirmed whether these predicted proteins are produced during infection, the presence of these ORFs in more than one virus genome suggests that they are not just the result of prediction errors or artifacts within genome assemblies.

INTRODUCTION

The genus *Tobamovirus,* family *Virgaviridae*, is composed of 37 members (Lefkowitz et al. 2018). Tobamoviruses have rod-shaped virions of ca. 18 X 300 nm and contain positive-sense single-stranded RNA genomes of ca. 6,400 nucleotides (Gibbs et al. 2010). Tobamoviruses are transmitted to plants via mechanical wounding. No specific biological vector is known for any virus in this genus. The Tobamovirus genome encodes at least 4 proteins: a 124-132 kDa protein associated with methyl/guanylyl transferases and helicases, a readthrough 181-189 kDa protein with motifs common to RNA dependent RNA polymerases (RdRps), a 28-31 kDa protein for cell-to-cell movement (movement protein, MP), and a 17-18 kDa capsid protein (CP) (Okada 1999). Various Tobamoviruses can infect a wide range of hosts, including plants in the Brassicaceae, Cucurbitaceae, and Solanaceae families (Dawson and Hilf 1992).

Among the cucurbit-infecting tobamoviruses are Cucumber green mottle mosaic virus (CGMMV) and Watermelon green mottle mosaic virus (WGMMV) (Lefkowitz et al. 2018). These two tobamoviruses have a genomic RNA and two subgenomic RNAs that are separately encapsidated by the CP resulting in three classes of virions, with only the largest being infectious. CGMMV, first discovered in 1935 in England (Ainsworth 1935), is the more studied of the two viruses with 112 published genomes on NCBI. WGMMV on the other hand has 6 complete genomes, with the first publication on this virus in 2018 (Cheng et al. 2019). Interestingly, both viruses are also able to be transmitted through virus contaminated seeds, a trait that not all tobamoviruses possess. Young leaves on infected plants will develop mottle and mosaic symptoms, but the main issue in infected cucurbit plants is that the fruits are not suitable for sale as they are often malformed and mottled (Tian et al. 2014). In severe cases, plants may collapse before even producing fruit. Some varieties of cucurbits may also be asymptomatically infected.

Because of the ease of its transmission, CGMMV is a quarantined pest by USDA-APHIS. Regardless of this classification, incidences of CGMMV have still occurred in the United States, specifically California (Tian et al. 2014), where more than 16 million cwt of melons are produced for a net value of 356 million dollars (CDFA 2019), without including squash and pumpkin production. In order to trace the origin of these outbreaks, virus isolates from infected plants were collected and sequenced and compared to other sequences from around the world. During the process of this mass sequencing effort, a different tobamovrius was discovered in a mix-infected bitter melon sample. Comparison of the sequence of this virus with the virus database on NCBI revealed it to be WGMMV (Pitman et al. 2019), which had previously only been found in Taiwan (Cheng et al. 2019). WGMMV has a similar host range, symptom morphology, and genome structure to CGMMV, making it very difficult to visually distinguish the two viruses. Therefore, we characterized it further.

With the influx of new CGMMV and WGMMV isolates, as well as the improvement of sequence analyzing programs, two small ORFs were identified within their genomes. These ORFs are not found in other tobamoviruses, raising our interest as to their potential importance. Comparison of all known CGMMV and WGMMV genome sequences showed that these small ORFs existed across all assemblies and isolates of their respective viruses. The work presented in this thesis attempts to determine whether these ORFs in fact encode real proteins, and to assess their potential role(s) in plant infection.

Chapter 1 - ORFH: A novel small open reading frame found within the genome of Cucumber green mottle mosaic virus, but not other tobamoviruses

ABSTRACT

Cucumber green mottle mosaic virus (CGMMV; genus *Tobamovirus*) has recently re-emerged as an important virus affecting cucurbit production in many parts of the world. Although many of the CGMMV ORFs and encoded proteins have been previously identified and well-characterized, a novel small ORF was identified in the genome of CGMMV. Comparison of 50 CGMMV genome sequences showed that this ORF was found within the genomes of all of the CGMMV isolates and for all, the predicted amino acid sequence was identical. This ORF (ORFH) overlaps the ORF encoding the 30 kDa movement protein of CGMMV in the +1 reading frame register. Transient expression of the ORFH encoded protein fused to GFP on both the N and C termini showed that the fusion proteins localized to the endoplasmic reticulum in *Nicotiana benthamiana* epidermal cells. Mutations of this ORF introduced into the CGMMV genome with little to no disruption to the 30 kDa movement protein did not affect the virus's ability to systemically infect *N. benthamiana* plants or two tested cucurbit plants, *Cucumis melo* var. Laredo and *Cucumis melo* var. Expedition. However, CGMMV mutants containing truncations of ORF H were less fit than wild-type CGMMV in *Chenopodium amaranticolor* plants. Subcellular fractionation and western blot analysis failed to unequivocally confirm the ORFH encoded protein from CGMMV-infected *N. benthamiana* and cucurbit plants.

INTRODUCTION

Cucumber green mottle mosaic virus (CGMMV) was first described in 1935 in England (Ainsworth 1935). Since its description CGMMV has been detected in many Eurasian countries before reaching North America and California in 2013 (Tian et al. 2014). Like all viruses in the genus *Tobamovirus*, CGMMV is not spread plant-to-plant by specific vectors but via mechanical wounding. However, CGMMV is seedborne in cucurbits, and this has likely resulted in its worldwide distribution. Because California imports cucurbit seeds from many different countries, care must be taken to prevent new CGMMV introductions.

Recent work has suggested that nucleotide sequence and bioinformatics analysis of CGMMV isolates may help determine the origins of different CGMMV isolates (Pitman, unpublished). Bioinformatic/phylogenetic comparison of these sequences to those of CGMMV isolates from other locations could help identify the country of origin for different CGMMV isolates. Another benefit of the rapidly accumulating virus genome sequences is that by comparing them in detail, novel features can sometimes be identified. One example is the discovery of previously unidentified open reading frames (ORFs), such as in the case of Tobacco mosaic virus (TMV) and the identification of ORF6 (Erokhina et al., 2017).

Detailed comparison of genomic sequences for more than 50 CGMMV isolates showed that there was one small ORF overlapping the 30 kDa protein ORF in the +1 reading frame register, named ORFH. This ORF was not found in the genomes of other tobamoviruses. In this chapter, ORFH was identified by a combination of next generation sequencing (NGS) and bioinformatics. ORFH's putatively encoded protein was examined by western blot and confocal microscopy analyses. The possible function of this protein was studied in in plants by creating mutants of ORFH with little disruption to the overlapping 30 kDa protein.

MATERIALS AND METHODS

CGMMV isolation

Most of the CGMMV sequences analyzed in this study were from ongoing research by Tera Pitman. However, I received CGMMV-contaminated seeds and recovered new CGMMV isolates which were used for sequence comparisons here. CGMMV-contaminated seeds were received from several different sources and seed extracts were used to inoculate *N. benthamiana* plants in order to recover the virus. One g of seeds from these seedlots was homogenized using a coffee grinder. From each sample, 0.1 g was mixed with 10 mL Paul's buffer (0.3 M potassium phosphate buffer, 1 mM EDTA, 5 mM DIECA, 5 mM sodium thioglycate) to inoculate *N. benthamiana* plants. After 14 days, 0.1 g tissue from the plants was collected and homogenized with a tissue homogenizer using 2.3 mm silica/zirconia beads before RNA was extracted using Trizol life reagent (Life Technologies). cDNA was generated from this RNA using iScript cDNA Kit (BioRad), which was then used in PCR using primers TP_CGMMV CP Fwd and TP_CGMMV CP Rvrs (Table 5) to check for the presence of CGMMV. Tissue from plants that tested positive for CGMMV was subsequently used to prepare virions for RNA isolation and sequencing.

Sequencing Sample Preparation

Twenty g of tissue was collected from CGMMV positive *N. benthamiana* plants. Tissue was homogenized in 0.1 M phosphate buffer with 0.001% β-mercaptoethanol and filtered through cheese cloth to remove larger particulates. The extract was mixed with equal volumes of *n*-butanol and chloroform and stirred for 5 minutes on ice before being centrifuged for 10 minutes at 8,000 g. The top aqueous phase was recovered. Sodium chloride was added to the recovered aqueous phase to give a final concentration of 0.2 M, and polyethylene glycol 6000 added to 4% (w/v) (Van Kammen 1967). This solution was stirred on ice for 30-45 minutes, then was centrifuged for 30 minutes at 8,000 g. The precipitated virion pellet was resuspended in 2 mL of 50 mM pH 7 Tris buffer.

The resuspended virions' RNA was extracted using Trizol life reagent (Life Technologies) according to the manufacturer's specifications. Samples were analyzed by spectrophotometry to check concentration and to check for potential phenol contamination, then sent to Amaryllis Nucleics (Oakland, CA) for library preparation and Illumina sequencing. Five samples were processed in this way.

Alignment of sequences and identification of ORF

Using CGMMV accession NC_001801 (Ugaki, 1991) as a reference genome, the Burrows-Wheeler Alignment tool (BWA, version 0.7.17 (Li, 2009)) was used to align reads from the five samples received from Amaryllis to the reference with bwa aln, using default settings with the following exception: –n 4. Unaligned reads were filtered out before using RNAspades (Bankevich, Nurk et al., 2012) to assemble contigs for the full CGMMV genome sequence, using the default settings.

Antibody development and western blot analysis of the putative ORFH protein

The ORFH sequence was sent to BioMatik (https://www.biomatik.com/) to predict potential antigenic sites in the putative protein and to generate an antibody to be used for western blot analysis. Only one potential antigenic site was predicted according to their program to use for antibody production (Figure 5). Lyophilized antibodies from two different rabbits were received from the manufacturer and resuspended according to package instructions. Aliquots of this stock solution were made to avoid freeze-thaw cycles.

Tris-Tricine gels (13%) (Schägger 2006) were used to separate proteins isolated from CGMMV-infected plants by using a subcellular fractionation procedure (Donald, Zhou, and Jackson 1993) with an uninfected plant as a control. Four g of *N. benthamiana* leaves were homogenized using

liquid nitrogen and a mortar and pestle before 6 mL of protein extraction buffer (400 mM sucrose, 100 mM Tris, 10 mM KCl, 5 mM MgCl2, 10% glycerol, 10 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF) pH 7.5) was added to the homogenized tissue. This extract was filtered through a nylon mesh (Miracloth, Calbiochem). The filtrate was centrifuged at 1,000 g for 10 minutes to pellet nuclei, mitochondria, and chloroplasts. The aqueous phase from this centrifugation was centrifuged again, this time at 30,000 g for 45 minutes to separate the remaining disrupted cellular membranes from the cytosol.

Twenty µL of the protein samples were loaded into the gels and stacked in a 4% acrylamide stacking gel at 30V until the dye from the sample buffer reached the resolving gel. The voltage was then increased to 150V to separate the proteins in the resolving gel. The proteins were wet transferred to a 0.45 µm nitrocellulose membrane at 100V for 1.5 hours. Membranes were blocked for 1 hour with 7% nonfat dry milk in TBST (Tris buffered saline with 0.1 % Tween 20), then probed with the BioMatik generated antibody ORFH-AB at 4 µg/mL overnight in the blocking solution before being washed. Secondary antibodies Goat Anti-Rabbit IgG (H + L)-HRP Conjugate (BioRad) were applied in a 1:3,000 ratio for 1 hour before doing a final wash. The membrane was treated with Pierce ECL Western Blotting Substrate and photographed using BioRad's ChemiDoc.

Generation of constructs for ORF H: GFP fusion proteins to be used for intracellular localization studies

RNA was extracted from a CGMMV (SJ isolate) infected a *Nicotiana benthamiana* plant using the standard TRIzol (Thermo Fisher) protocol after 0.1 g of plant tissue was homogenized for 15 seconds in a tissue homogenizer with 2.3 mm silica/zirconia beads. After the extraction, RNA samples were eluted with 100 µL nuclease free water. cDNA was generated using iScript cDNA Kit (BioRad) following the protocol provided with the kit. Transient expression vector pEAQ containing the eGFP gene sequence was used to amplify GFP.

The ORFH and GFP coding sequences were amplified using CloneAmp HiFi (Takara) with primers (Table 1) and reaction mix (Table 2), following the manufacturer's protocols. The PCR products were separated on a 2% agarose gel in 0.5x TAE buffer at 100V for 30 minutes. The DNA bands were cut from the gel and extracted using Zymoclean Gel DNA Recovery Kit (Zymo), then corresponding DNA fragments were ligated into plasmid pEAQ (linearized with NruI and StuI; New England Biolabs) with NEBuilder HiFi DNA Assembly (New England Biolabs). Circularized plasmids were transformed into *E. coli* (strain DH5α) to increase copy number before being purified using the QIAprep Spin Miniprep Kit (QIAGEN). The sequences were then confirmed using a combination of restriction enzyme digestion (BamHI-HF and PacI; New England Biolabs) and by Sanger sequencing using primers pEAQ-seq-F and pEAQ-seq-R (Table 3) flanking the insert region. After the correct sequences were verified, the plasmids were transformed into *A. tumefaciens* (strain GV3101).

Leaf infiltration and visualization of subcellular localization of fusion proteins

Agrobacterium tumefaciens cells containing the fusion protein plasmids as well as the plasmids for two subcellular markers (endoplasmic reticulum (ER) and nucleus)(Nelson, Cai, and Nebenführ 2007) were grown in Luria-Bertani (LB) liquid media containing their corresponding antibiotics (Jiang, 2021) for roughly two days at 28°C. The cells were pelleted by centrifugation to remove them from the media and resuspended in a modified induction buffer (10 mM MgCl2 containing 100 µM acetosyringone) to an optical density (OD) of 0.02 at 600 nm in a spectrophotometer. Cultures were incubated in the modified induction buffer at room temperature for at least 4 hours before young, wild-type *N. benthamiana* plants were infiltrated (Li 2011). After two days, the leaves were checked by confocal microscopy for fluorescent protein expression and localization.

Mutant CGMMV construction and infection in cucurbits

Mutations were introduced into the ORFH coding region of the CGMMV genome in attempts to eliminate its expression in plants without affecting the overlapping 30 kDa ORF or its encoded protein. Using RNA for CGMMV isolate SJ as a template, primers F5'CGMMV, RmidCGMMV, FmidCGMMV, and R3'CGMMV (Table 4) were designed to introduce two nucleotide changes per mutant, resulting in two amino acid changes in ORF H putative protein with either no disruption to the 30 kDa protein, or with changes to similar amino acids (e.g. hydrophobic to hydrophobic) (Table 11). These primers were used in conjunction with primers corresponding to the 5' and 3' ends of the CGMMV genomic RNA in RT-PCR reactions to create amplicons for two 'halves' of the genomic RNA, mapped in Figure 1. These were then ligated together into the plasmid pJL89 with the NEBuilder HiFi assembly kit (NEB). The mutated infectious clones were then transformed into *E. coli*, plated on kanamycin plates, and incubated at 37°C overnight.

Ten colonies were chosen from each plate and grown in 3 mL LB media containing kanamycin overnight at 37°C. The cells were pelleted, then plasmids were extracted with a commercial miniprep kit (Qiagen) and sent for sequencing to confirm the mutations. Plasmids that maintained the mutations were transformed into *A. tumefaciens* and used to infiltrate wild-type *N. benthamiana* plants. *A. tumefaciens* containing the plasmids for the mutants (M1, M2, and M3) as well as a wild-type infectious clone of CGMMV were grown in LB media for two days, after which the bacteria were pelleted by centrifugation and resuspended in induction buffer (10 mM $MgCl_2$ containing 10 mM MES and 100 µM acetosyringone) to an optical density (OD) of 1 at 600nm in the spectrophotometer. Cultures were incubated in induction buffer for at least 4 hours before being infiltrated into leaves of young, wild-type *N. benthamiana* plants. These plants were maintained in a greenhouse for 10 days, after which infection was evaluated by RT-PCR. cDNA was generated using the iScript cDNA Kit (BioRad). This cDNA was used in the subsequent PCR

with specific CGMMV coat protein primers TP_CGMMV CP Fwd and TP_CGMMV CP Rvrs (Table 5) (Pitman, unpublished) with GoTaq Flexi (Promega) following the manufacturer's protocol.

After infection was confirmed, the mutations were confirmed by amplifying the ORFH coding sequence and the surrounding regions via RT-PCR, followed by Sanger sequencing. Primers which flank ORFH on the CGMMV genome, OOH-F and OOH-R (Table 6) were used to both amplify and sequence the RT-PCR products. Once it was established that the mutations were maintained, the *N. benthamiana* plants were used to obtain inoculum to inoculate plants of two cucurbit cultivars: Expedition and Laredo melons (HM Clause). Ten plants of each variety were used per each mutant and wildtype CGMMV. The plants were inoculated after the emergence of their first true leaf.

For the inoculation, 0.5 g of infected tissue was ground in 10 mL Paul's inoculation buffer, then this extract was gently brushed onto the carborundum-coated cotyledons of the cucurbit plants. The plants were maintained in the greenhouse for 10 days before samples were collected from the uppermost leaves. RNA was purified from these samples using TRIzol and cDNA was generated using iScript cDNA Kit, which was subsequently used for Taqman qPCR using CGMMV specific primers and probe (Table 7), as well as primers and probe for cytochrome oxidase (COX) as a housekeeping gene. The qPCR assay was performed using iTaq Universal Probes Supermix (BioRad) following their mix protocol (Table 8). All results were analyzed using BioRad's CFX Maestro program. Samples were also spot checked for mutation maintenance by using the flanking primers OOH-F, OOH-R (Table 6) by RT-PCR and Sanger sequencing.

Mutants versus wild-type CGMMV in systemic infections

In order to test the competitiveness of the CGMMV ORFH mutants against the wild-type CGMMV, mutants M1, M2, M3, wild-type CGMMV, and healthy *N. benthamiana* tissue were collected and

used to co-inoculate more *N. benthamiana* plants. A Bioreba extraction bag containing 10 mL Paul's buffer was used to homogenize 0.1 g of each type of infected or healthy plant (Table 9).

After homogenization, three new *N. benthamiana* plants were inoculated per mix. These plants were grown in a greenhouse for 10 days to allow for establishing systemic infections. Then, 0.1 g tissue from these plants was harvested into Bioreba bags containing 10 mL Paul's buffer. Four 10-fold dilutions were made from this original homogenate for a total of five different inoculum concentrations per sample. These five inoculum concentrations were used to inoculate *Chenopodium amaranticolor* leaves following the diagram illustrated in Figure 2.

After local lesions appeared on the leaves, individual lesions were excised and homogenized with zirconia/silica beads and RNA extracted with TRIzol reagent. cDNA was generated with the iScript cDNA Kit, then the ORFH sequence was amplified using the flanking primers OOH-F and OOH-R. The resulting amplicons were extracted by phenol:chloroform and alcohol precipitated (Oliner, Kinzler, and Vogelstein 1993) analyzed by Sanger sequencing. Sequences were compared against the wild-type virus to see which of the two isolates was more prevalent in the infection.

RESULTS

BLAST and Snapgene sequence comparisons identified a novel small ORF (ORFH) in CGMMV

Four new 6421-6424 nucleotide (nt) CGMMV genome sequences were assembled. Snapgene was used to compare the sequences by confirming that known ORFs had appropriate start and stop codons, as well as confirming that the 5' and 3' nucleotides were conserved. In doing so, a small ORF (ORFH) was observed in the +1 reading frame register, overlapping the 30 kDa movement protein (Figure 3, bottom). Because observation of small ORFs can be commonly attributed to assembly artifacts, other isolates of CGMMV were also compared using Snapgene for the appearance of this ORF. Upon review, it was found that all 107 CGMMV isolates published in Genbank and 37 unpublished CGMMV isolate sequences generated in our lab contained ORFH. It was then hypothesized that ORFH could potentially be a cucurbit-infecting tobamovirus-specific ORF. To assess this further, one representative sequence from all 38 tobamoviruses was uploaded into Snapgene to follow the same analysis that led to the discovery of ORFH (Table 10). Examples of the analysis are shown in Figure 3, where Tobacco mosaic virus (TMV) and Kyuri green mottle mosaic virus (KGMMV) have their ORFs compared to CGMMV's ORFs, showing the lack of an ORFH-like ORF.

The genomes of only two other tobamoviruses contained a similarly positioned, small ORF with similar sequences to ORFH from CGMMV: Cucumber mottle virus (CMoV) and Watermelon green mottle mosaic virus (WGMMV). These similar small ORFs (Figure 4) are also located overlapping the 30 kDa movement protein ORF. Comparison of the deduced amino acid sequences for these small ORFs among CGMMVs and the other cucurbit-infecting tobamoviruses showed absolute conservation among all CGMMV isolates, and a high degree of amino acid similarity for CMoV and WGMMV (Figure 4). The presence of these ORFs in all CGMMV isolates and the high degree of nucleotide sequence (and deduced amino acid) conservation strongly suggests that this region

of the CGMMV genome is under strong selection pressure and that ORFH may not just be an artifact of sequencing/sequence assembly. The translated amino acid sequence of this genomic region is also conserved among two other cucurbit-infecting tobamoviruses (Figure 4), with 48% of the amino acids being the same across the three viruses (46/94 amino acids).

Western blot analysis

In order to determine if the CGMMV ORF H putative protein was detectable in CGMMV-infected plants, the ORFH coding sequence was sent to BioMatik for analysis and generation of polyclonal antibodies. Only one potential antigenic site was predicted (Figure 5). Antibodies from two rabbits were obtained and used for serological analysis. The peptide used to generate the antibodies was also received from Biomatik as a positive control for assays.

Because the putative ORF H protein could be present in only small amounts, subcellular fractionation of CGMMV-infected plant tissues was used in attempts to increase the potential for detection the ORFH protein. Leaf tissue from healthy and 10-day systemically infected *N. benthamiana* plants (CGMMV isolate San Joaquin) was separated into three fractions: CW (cell wall containing fraction), P1 (containing nuclei, mitochondria, and chloroplasts), and P30 (containing the disrupted membranes and cytosol).

In the first experiment a faint signal in the P1 fraction from CGMMV infected tissue was obtained and was not seen in healthy controls (Figure 6). The estimated size of this protein is ~20 kDa. However, attempts to concentrate the protein and reproduce these results have been unsuccessful. Tissue of varying ages (7, 10, and 14 days post inoculation) from the native host (cucurbits) as well as from *N. benthamiana* was used for protein extraction to attempt to recapture the protein. Membranes with a smaller pore size (0.22 micron) were also used in attempts to get a stronger signal from the protein. Denaturing buffers of different strengths and non-denaturing

buffers were used to try to determine if the antibody could bind in either the native or denatured states without success. RIPA protein extraction buffer (Holden and Horton 2009) was used as an aggressive membrane disruption buffer after using the transmembrane predictor from TMHMM (Sonnhammer, von Heijne, and Krogh 1998) to show that the putative ORF H protein had many predicted transmembrane domains (Figure 7). None of these efforts allowed for unequivocal detection of the ORF H-encoded protein.

ORFH:GFP fusion proteins

Another approach used to characterize the putative ORF H protein in plant tissues was to examine intracellular localization of the ORF H protein:GFP fusions (ORFH:GFP) by using confocal microscopy. To attempt to identify ORF H's localization in cells, the GFP coding sequence was added to ORFH so that the ORF H protein would be fused to GFP at either the N' or C' terminus. The engineered plasmids (ORFH:GFP or GFP:ORFH) were transformed into *A. tumefaciens*, which was then used to infiltrate wild-type *N. benthamiana* plants.

When examining the infiltrated leaf tissues by using confocal microscopy, the ORFH:GFP fusion proteins appeared to localize to the endoplasmic reticulum (ER). The web-like structure of the ER appeared disrupted, forming aggregates that are dispersed around the periphery of the cell and around the nucleus. GFP fused to the N-terminus of the ORF H protein gave a brighter fluorescent signal than GFP fused to the C-terminus of the ORF H protein (Figure 8). The ER localization was confirmed by co-expressing an mCherry ER marker with GFP:ORFH. When the ER marker was expressed by itself, the web-like structure of the ER can be seen inside the cell as well as along the periphery (Figure 8).

Again, when *A. tumefaciens* harboring the plasmid containing ORFH was infiltrated into cells, the ER (tagged with mCherry) was disrupted and formed aggregates that center around the nucleus

(Figure 10). Cells that express only mCherry:ER do not show these aggregates and maintain their web-like structure around the nucleus and throughout the cell (Fig. 8).

In order to confirm that the ORF H protein affected only the ER and not the nucleus, cells were co-infiltrated with *A. tumefaciens* expressing ORF H tagged with GFP and a red fluorescent protein (mCherry) that will localize to the nucleus. Cells that were co-expressing ORFH:GFP and nuclear mCherry were visualized by confocal microscopy. The red and green fluorescence did not overlap, which means the ORF H protein does not enter the nucleus. While the ORF H fusion protein continued to form aggregates around the nucleus, the nucleus itself was unaffected (Figure 11).

Mutant CGMMV infections in cucurbits

Mutations to introduce stop codons were engineered into the ORFH coding sequence so as to prevent its expression in plants. Care was taken to introduce mutations that did not change the amino acids encoded by the overlapping CGMMV 30 kDa ORF (Table 11), thus creating three truncated ORFH mutants to study the effect of this ORF and protein on CGMMV infection. *Nicotiana benthamiana* plants were inoculated with the wildtype and CGMMV mutants. Then, RT-qPCR was used to compare wildtype and mutant CGMMV accumulation in systemically infected cucurbit plants. Statistical analysis showed that all the mutants showed similar accumulation in 10 plants of each variety, and none were different from the wild-type CGMMV (Figure 12). Symptoms developed at the same rate and severity in the mutant and wild type infected plants (Figure 12). Infected *N. benthamiana* plants are shown because infected cucurbit plants display less obvious symptoms, even within the wild-type infection.

To see if the ORF H protein affected the transmissibility of CGMMV, tissue was collected from these infected plants and used to rub inoculate *N. benthamiana* plants. The mutants were passed

on this way for six generations. Tissue was collected at 10 days post inoculation at every passage, and RT-PCR using primers OOH-F and OOH-R were used to amplify the mutated regions. The amplicons were then sequenced to see if the mutants had reverted to the wild-type virus. The mutants were successfully passed six times from *N. benthamiana* to *N. benthamiana* without reversion to the wild type.

Mutants versus Wild-type CGMMV in Systemic Infections

Because the above experiments did not show obvious differences in accumulation between the CGMMV wildtype and mutants, a competition experiment was performed. For this experiment, *N. benthamiana* plants were co-inoculated with combinations of the CGMMV mutants and wildtype CGMMV. Then systemically infected tissues were collected and dilutions of inocula were used to inoculate *C. quinoa* plants. Individual local lesions were recovered and analyzed to determine if the lesion resulted from the wildtype or mutant CGMMVs. During the double-infection, plants were monitored for symptoms compared to singly infected *N. benthamiana* plants. Symptoms were similar across all infections. *Chenopodium amaranticolor* plants were similarly observed during their localized infection for differences in morphology of the local lesions. The leaves that were rubbed with the initial inoculum (0.1 g each of mutant and wild-type in 10 mL Paul's buffer) were heavily infected showing abundant local lesions; so many that some leaves did not survive the inoculation. Other than the initially inoculated leaves, there was no difference in lesion size or abundance between the singly-infected plants and the double-infected plants. However, the number of mutants that managed to survive the double infection with the wild-type virus was low (<10% for all three mutants). Table 11 shows the number of local lesions isolated and extracted from each combination, as well as the number of the lesions for wildtype and CGMMV mutants.

DISCUSSION

In this chapter, the ORFH sequence was identified from all CGMMV isolates in databases and those sequenced by our lab, as well as two other cucurbit-infecting tobamoviruses. Mutations of this region of the genome did not seem to affect symptom severity or the ability of the mutant viruses to infect plants of the model system (in *N. benthamiana*) or of native hosts (cucurbits). Disruption of the ER membranes was observed during transient expression of ORF H with and without GFP.

Efforts to identify the existence of ORF H-encoded protein by western blot provided inconclusive results. Because it was possible that the ORF H protein was produced in too low abundance in the plant to be detected, isolation of virions for *in vitro* translation were attempted. However, the sequence of the ORF H protein contained no lysines, and the reagents obtained for the *in vitro* translation were not effective. In the future, the same protocol should be used to try to obtain radioactive-labelled proteins (labelling methionines instead of lysines; data not shown).

Another approach to finding ORF H can be using nondenaturing methods to detect the protein with the antibodies generated. ORF H is predicted to contain many transmembrane and helical domains (Figure 7), which could fold and aggregate. Because western blot assays in this thesis were performed with SDS-PAGE gels that denature the protein, it is possible that the antibody cannot recognize and bind to the now-denatured protein. Using a non-denaturing PAGE gel could be a potential solution to this issue. In the event that a nondenaturing PAGE gel cannot be used, the ELISA assay can be another way to detect the protein, although this method will not help to determine the size, only the presence.

Somehow this ORF, or at least this region of the CGMMV genome nucleotide sequence, is important, but the reason was not discovered during this study. Future work should expand upon

virus infection in host plants and possibly look at virion structure by electron microscopy to compare the differences between the wild type and the mutants, as was done for the localization of ORF6 in the wild type infection of Tobacco mosaic virus (Erokhina et al. 2017). Another aspect that could be informative is to look for the presence of inclusion bodies in both types of infections. Since the ORF H fusion proteins interfered with ER membranes, it might stand to reason that these membranes can/are used in the formation of inclusion bodies in the CGMMV-infected plant. Disrupting ORF H could mean that these inclusion bodies are formed at a lower rate, which would not necessarily interfere with infection by the mutants but could explain their lowered efficacy when compared to the infection of wild-type CGMMV.

TABLES CHAPTER 1

	Primer Name	Sequence	Amplified Gene
N-terminal GFP	N-GFP-F1	gtatattctgcccaaattcgCGATGGTGAGCAAGGGCG	GFP
	N-GFP-R1	ACGAGGGATTCTTGTACAGCTCGTCCATGCC	GFP
	N-GFP-ORFH-F2	GCTGTACAAGAATCCCTCGTGCCTGTCAAG	ORFH
	N-GFP-ORFH-R2	aatgaaaccagagttaaaggcctttaCAAATAAAGACCAAGGATC	ORFH
C-terminal GFP	C-GFP-ORFH-F1	gtatattctgcccaaattcgcgATGAATCCCTCGTGCC	ORFH
	C-GFP-ORFH-R1	CCTTGCTCACCAAATAAAGACCAAGGATCGC	ORFH
	C-GFP-F2	TCTTTATTTGGTGAGCAAGGGCGAGGAG	GFP
	C-GFP-R2	aatgaaaccagagttaaaggcctttaCTTGTACAGCTCGTCC	GFP

Table 1. Primers for generation of GFP fusion protein constructs, categorized by the terminus fused to ORFH GFP. N-terminal fusions have GFP at the amino terminus, while C-terminal fusions have GFP at the carboxy terminus of the ORFH predicted protein. Capital letters correspond to the nucleotide sequence coding for the protein of interest, lowercase letters to the plasmid sequence.

Reagent	Concentration (uM)	uL/rxn
CloneAmp HiFi Premix	-	25
Primer F	10	1.5
Primer R	10	1.5
Water	-	19
cDNA/DNA	-	3
	Total rxn volume	50

Table 2. PCR mix for amplifying the ORFH and GFP coding sequences. Final concentrations were derived from Takara protocols for CloneAmp Hi-Fi.

Primer Name	Sequence
pEAQ-seq-F	gcttctgtatattctgcc
pEAQ-seq-R	gcacaccgaataacagtaaattc

Table 3. Sequencing primers for pEAQ plasmid containing the sequence for ORFH and GFP coding sequences. Primers flank the insertion site of pEAQ,

Primer Name	Seqeunce
F5'CGMMV	GTTTTAATTTTTATAATTAAA
RmidCGMMV	GTTACGGCATCGAAATCATTCAG
FmidCGMMV	CTGAATGATTTCGATGCCCTAAC
R3'CGMMV	TGGGCCCCTACCCGGGGAAAG

Table 4. Primer names and sequences used to generate RT-PCR products corresponding to the two halves of the CGMMV genomic RNA. Colors coordinate with Figure 1 to show approximately where the primers bind on the genome.

Primer Name	Sequence
TP_CGMMV CP Fwd	CTT ACA ATC CGA TCA CAC CTA G
TP_CGMMV CP Rvrs	CTA AGC TTT CGA GGT GGT AGC

Table 5. Primers from Pitman (unpublished) used in PCR to detect CGMMV in wild-type and mutant CGMMV infections.

Primer Name	Sequence
Outside ORFH F (OOH-F)	GTT GAG GAG TTT CGC ATA TCT
Outside ORFH R (OOH-R)	CTG GGA TCT CTT GCC ATT AC

Table 6. Primers flanking the ORFH sequence in the CGMMV genome, used to amplify and sequence ORFH

Primer Name	Sequence
Cox F	CGT CGC ATT CCA GAT TAT CCA
Cox R	CAA CTA CGG ATA TAT AAG RRC CRR AAC TG
Cox Probe	HEX-AGG GCA TTC CAT CCA GCG TAA GCA-3BHQ_1
CGMMV 5968 F	ACG CTT TCC TCA ACG GT
CGMMV 6094 R	GCG TTA AGC GAC TCA GCA
CGMMV 6043 probe	Cy5-TCA TTG AGG TTG TAG ATC CTA GCA ATC C-TAO

Table 7. Primers and probes used for RT-qPCR. Cox primers and probe are for the reference gene, cytochrome oxidase. Primers are labelled with the orientation in which they will amplify: F for forward, R for reverse. Probe sequences have the fluorophore at the 5' end (HEX and Cy5) and the quencher at the 3' end (3BH 1).

Reagent	x1
iTaq Supermix	10 uL
CGMMV 5968 F (100uM)	0.18 uL
CGMMV 6094 R (100 uM)	0.18uL
CGMMV 6043 probe (100uM)	0.05 uL
Cox F (100uM)	0.18 uL
Cox R (100uM)	0.18uL
Cox Probe (100uM)	0.05 uL
dH2O	9.18
cDNA	1 uL
Total Reaction Volume: 20 uL	

Table 8. qPCR master mix protocol

Addition 1	Addition 2
SJ	M1
	M2
	M3
Healthy NB	SJ
	M1
	M2
	M3

Table 9. Table of inoculations: first inoculation is addition 1, second inoculation is addition 2.

Species	Exemplar accession number	Virus abbreviation	ORFH?
Bell pepper mottle virus	DQ355023	BPMV	NO
Brugmansia mild mottle virus	AM398436	BrMMV	NO
Cactus mild mottle virus	EU043335	CMMoV	NO
Clitoria yellow mottle virus	JN566124	CliYMV	NO
Cucumber fruit mottle mosaic virus	AF321057	CFMMV	NO
Cucumber green mottle mosaic virus	D12505	CGMMV	YES
Cucumber mottle virus	AB261167	CMoV	YES
Frangipani mosaic virus	HM026454	FrMV	NO
Hibiscus latent Fort Pierce virus	AB917427	HLFPV	NO
Hibiscus latent Singapore virus	AF395898	HLSV	NO
Kyuri green mottle mosaic virus	AJ295948	KGMMV	NO
Maracuja mosaic virus	DQ356949	MarMV	NO
Obuda pepper virus	D13438	ObPV	NO
Odontoglossum ringspot virus	X82130	ORSV	NO
Opuntia chlorotic ringspot virus	N/A	SOV	N/A
Paprika mild mottle virus	AB089381	PaMMV	NO
Passion fruit mosaic virus	HQ389540	PFMV	NO
Pepper mild mottle virus	M81413	PMMoV	NO
Plumeria mosaic virus	KJ395757	PluMV	NO
Rattail cactus necrosis-associated virus	JF729471	RCNaV	NO
Rehmannia mosaic virus	EF375551	RheMV	NO
Ribgrass mosaic virus	HQ667979	RMV	NO
Streptocarpus flower break virus	AM040955	SFBV	NO
Sunn-hemp mosaic virus	U47034; J02413	SHMV	NO
Tobacco latent virus	AY137775	TLV1	NO
Tobacco mild green mosaic virus	M34077	TMGMV	NO
Tobacco mosaic virus	V01408	TMV	NO
Tomato brown rugose fruit virus	KT383474	TBRFV	NO
Tomato mosaic virus	AF332868	ToMV	NO
Tomato mottle mosaic virus	KF477193	ToMMV	NO
Tropical soda apple mosaic virus	KU659022	TSAMV	NO
Turnip vein-clearing virus	U03387	TVCV	NO
Ullucus mild mottle virus	N/A	UMMV	N/A
Watermelon green mottle mosaic virus	MK070867.1	WGMMV	YES
Wasabi mottle virus	AB017503	WMoV	NO
Yellow tailflower mild mottle virus	KF495564	YTMMV	NO
Youcai mosaic virus	U30944	YoMV	NO
Zucchini green mottle mosaic virus	AJ295949	ZGMMV	NO

Table 10. A table of tobamoviruses and their accession numbers used for the ORFH-like sequence search. Tobamoviruses containing similar ORFH sequences are highlighted in green.

	ORFH		30kDa	
	1	2	1	2
Mutant 1	ASN -> PHE	SER -> STOP	HIS -> ASN	HIS -> ASN
Mutant 2	LEU -> STOP	TRP -> STOP	VAL -> VAL	LEU -> LEU
Mutant 3	LEU -> STOP	SER -> STOP	VAL -> VAL	PHE -> LEU

Table 11. Mutations were engineered to introduce stop codons in the ORFH encoded protein, and the corresponding amino acid change in the 30 kDa protein are shown. Changes made in the 30 kDa protein are to a similar amino acid (i.e. a polar amino acid like histidine is changed to another polar amino acid, asparagine).

Mutant	**# mutants/# WT CGMMV**	**% mutant**
M1	5/54	9.3
M2	0/58	0
M3	2/62	3.2

Table 12. Local lesion tests: number of mutant lesions over the number of wild-type CGMMV lesions.

FIGURES CHAPTER 1

Figure 1. Diagram of primer set locations on the CGMMV genome. Green arrows are for amplification of nucleotides 1-5006, red arrows are for nucleotides 4976-6422.

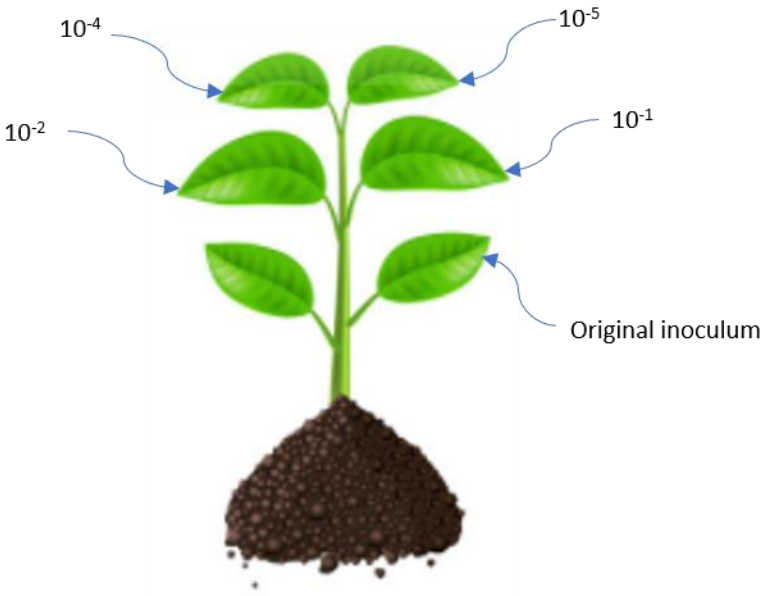

Figure 2. Inoculation scheme for C. amaranticolor *plants to isolate local lesions. The newer leaves have a more dilute inoculum than the older leaves.*

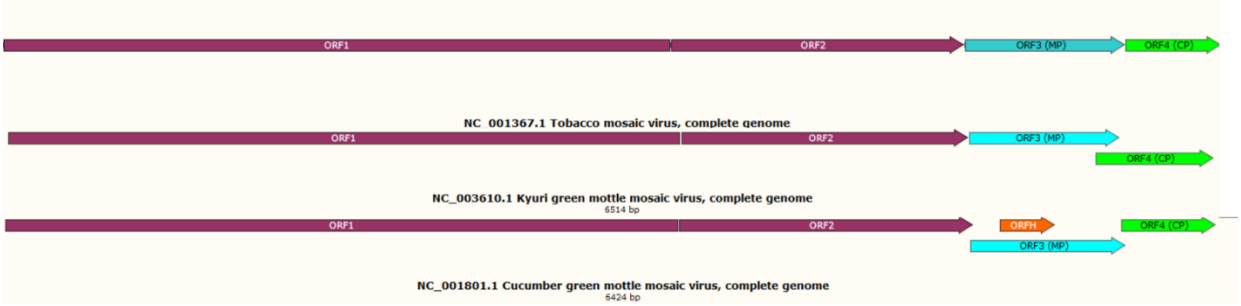

Figure 3. Comparison of the genome organizations and predicted ORFs of three tobamoviruses: Tobacco mosaic virus (top), Kyuri green mottle mosaic virus (middle), and Cucumber green mottle mosaic virus (bottom). Only Cucumber green mottle mosaic virus contains the ORF of interest (shown in orange). Different colors indicate individual ORFs.

```
CGMMV       MNPSCLSSCCVVLILQNTFMSHCWALWFLVYGTYRNPVGVVLLLLWLTQGCILLQRELYA 60
WGMMV       MSLFSLSSCCVVLILLNTSMSLCWALLFLVFGQFQKDVQVVLLLVWLTHVWNASLRVLCV 60
CMoV        MSLFSLSSCCVVLILLNTSMSLCWALLFLVFGQFQKDAQVVLLLVWLTHVWNVLLRALCV 60
            *.  .********** ** ** **** ***:* ::: . *****:***:      * * .

CGMMV       NFQLPPPSANSLLGSYLIILSWLRMPFAILGLYL*  94
WGMMV       SFLYPRPSESSTCDSFLTIPLLLQMLLVILGLYL*  94
CMoV        NFPCPRLSESSMCASYPIILSLLQMLLAILGLYL*  94
            .*  *  * .*   *:  *   *:* :.*******
```

Figure 4. Amino acid sequence alignment of small putative proteins from CGMMV, WGMMV, and CMoV with Clustal Omega.

MNPSCLSSCCVVLILQNTFMSHCWALWFLVYGTYRNPVGVVLLLLWLTQGCILLQRELYANFQLPPPSANSLLGSYLIILSWLRMPFAILGLYL*

Figure 5. Protein sequence of ORF H with antigenic site in red

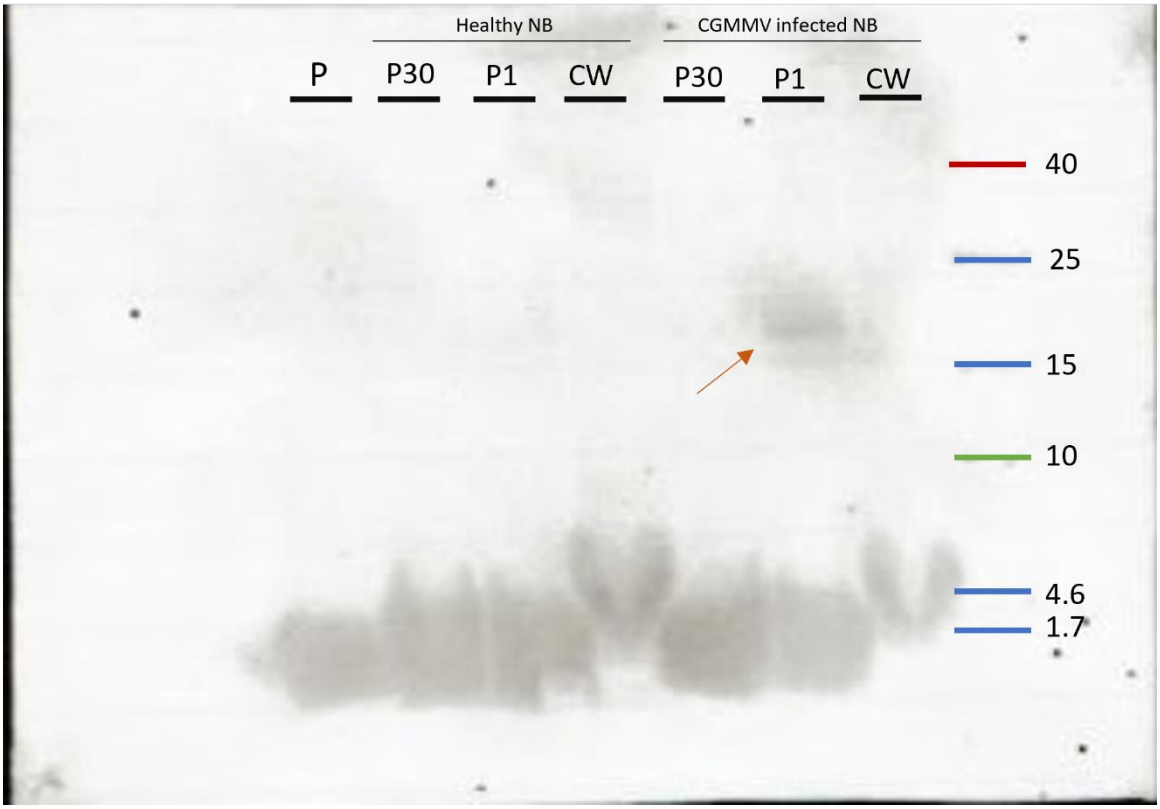

Figure 6. Western blot of subcellular fractionations. From left to right: (P) peptide from Biomatik, P30, P1, and CW from healthy plants followed by P30, P1, and CW from CGMMV-infected plants. The protein size marker ladder is indicated at the right. The band of interest is indicated by the arrow.

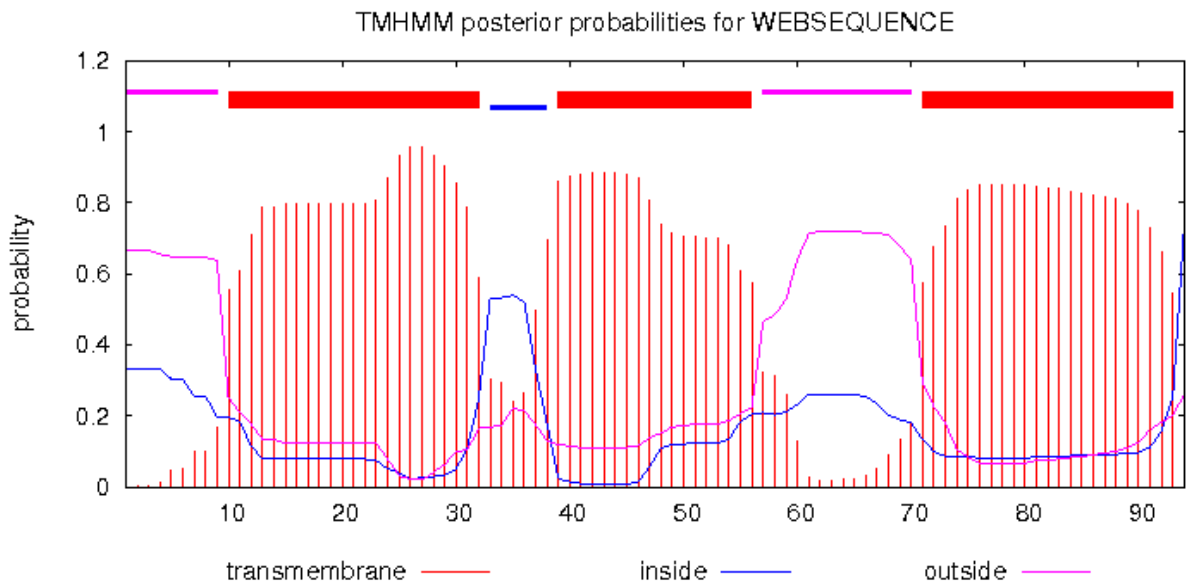

Figure 7. Transmembrane site predictor TMHMM indicates that there are three predicted transmembrane domains within the ORF H predicted protein.

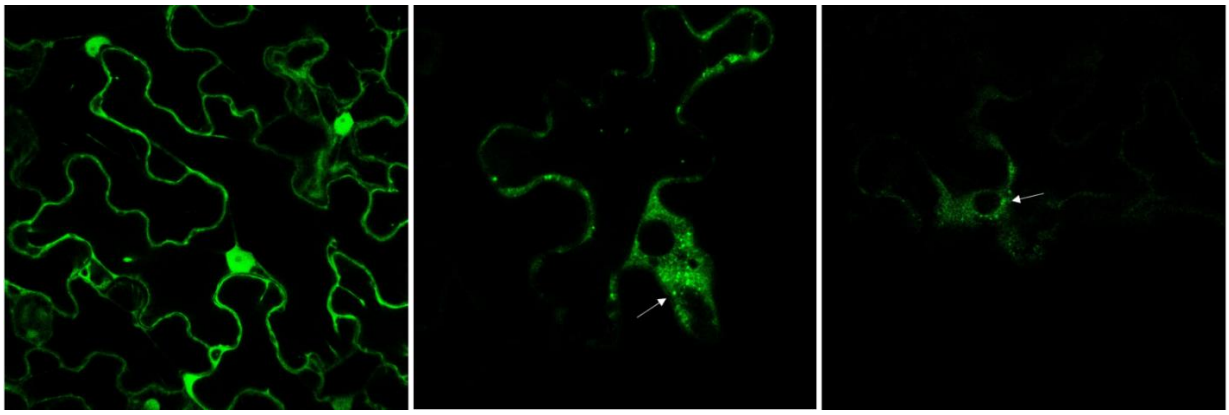

Figure 8. From left to right: Free GFP, GFP:ORFH (N-terminal fusion), ORFH:GFP (C-terminal fusion). Arrows indicate the aggregates formed by the ORF H/GFP fusion protein.

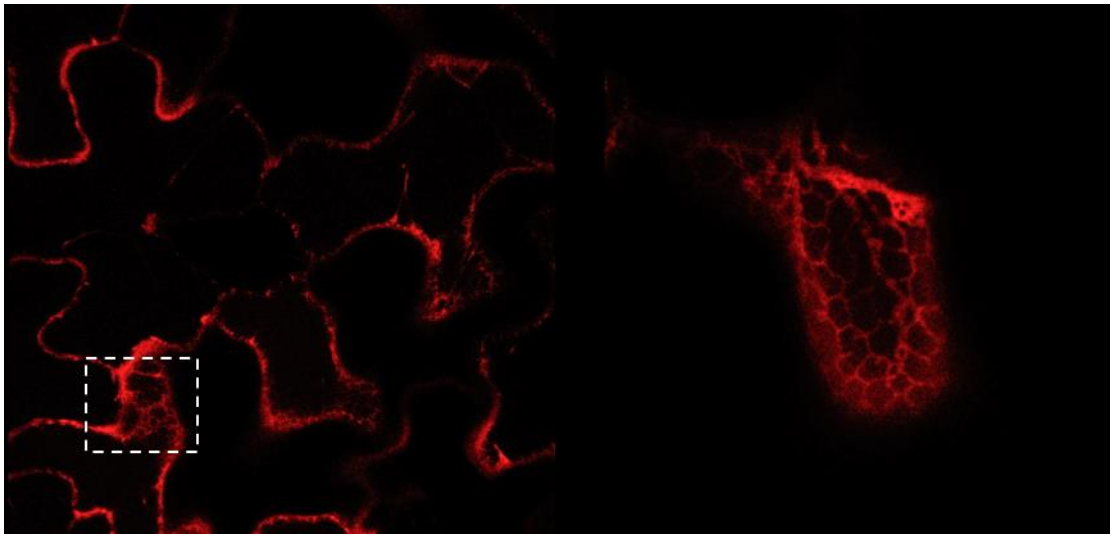

Figure 9. Left: cells infiltrated with A. tumefaciens *expressing the mCherry ER marker. Right: zoomed in image of the dashed square to show ER webbing without the addition of the ORF H protein*

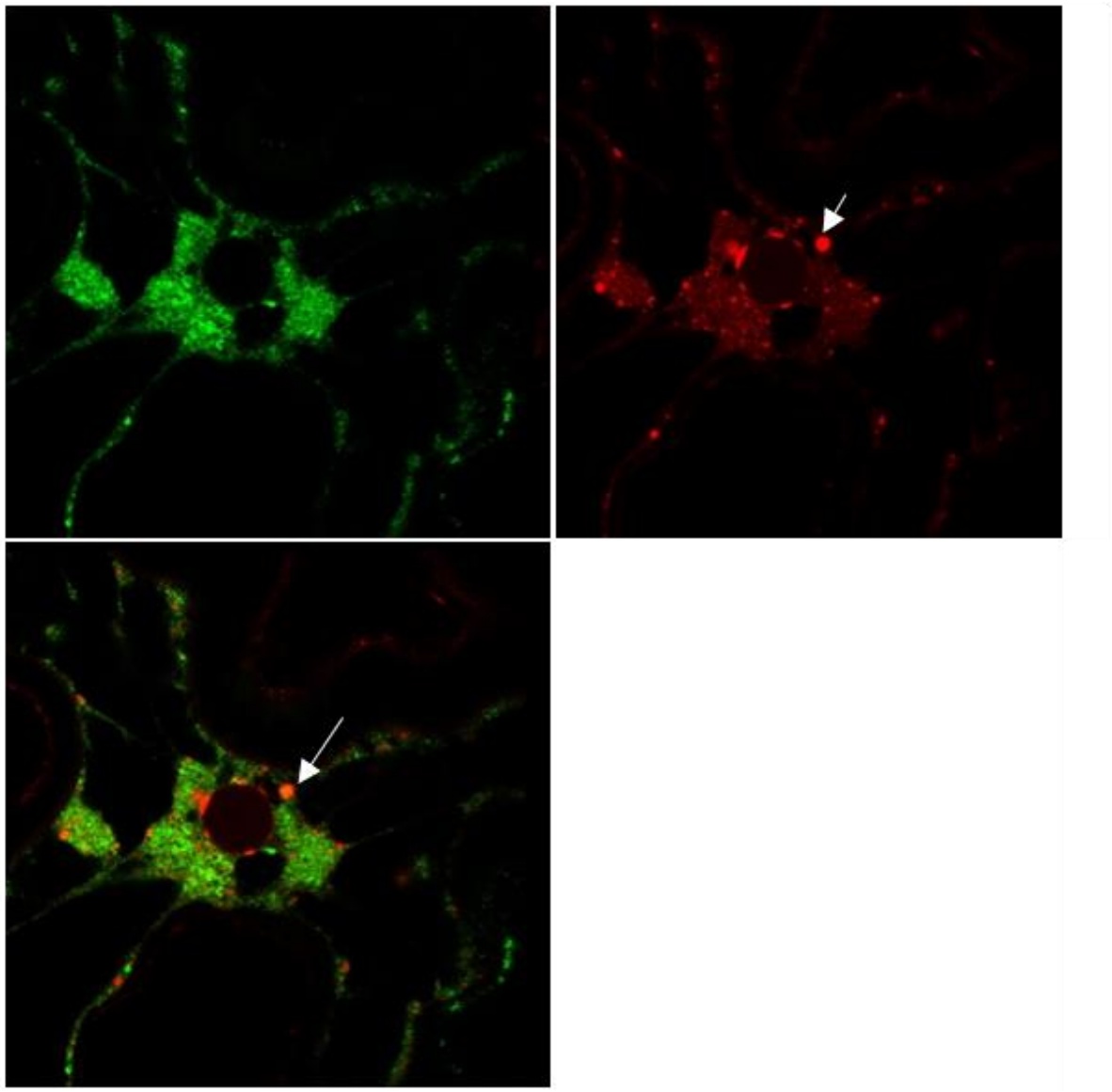

Figure 10. mCherry tagged ER with GFP:ORFH. Top left: GFP:ORFH. Top right: ER tagged with mCherry. Bottom left: ORFH:GFP and mCherry:ER overlayed. Arrows indicate the ER aggregates formed by the addition of ORFH.

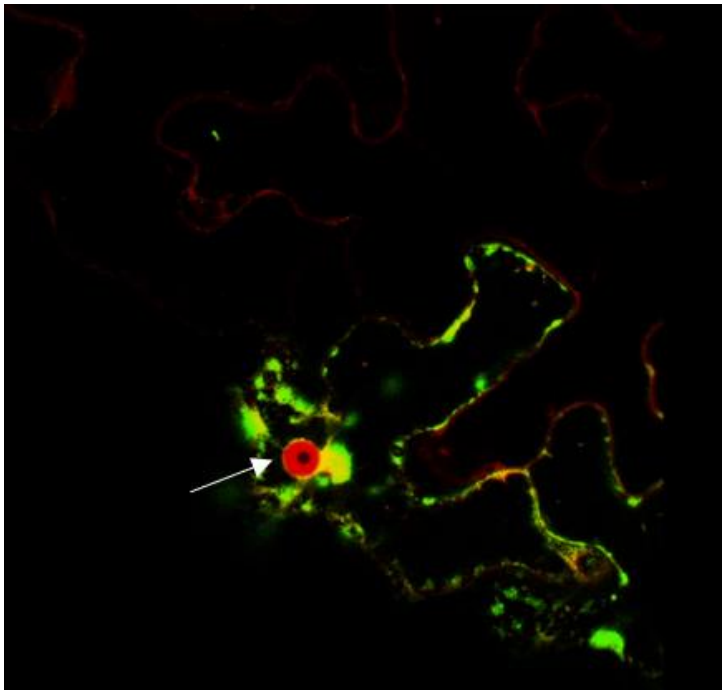

Figure 11. mCherry nucleus (red) and GFP:ORF H (green). The arrow indicates the nucleus, which remains unaffected by the GFP:ORF H fusion protein.

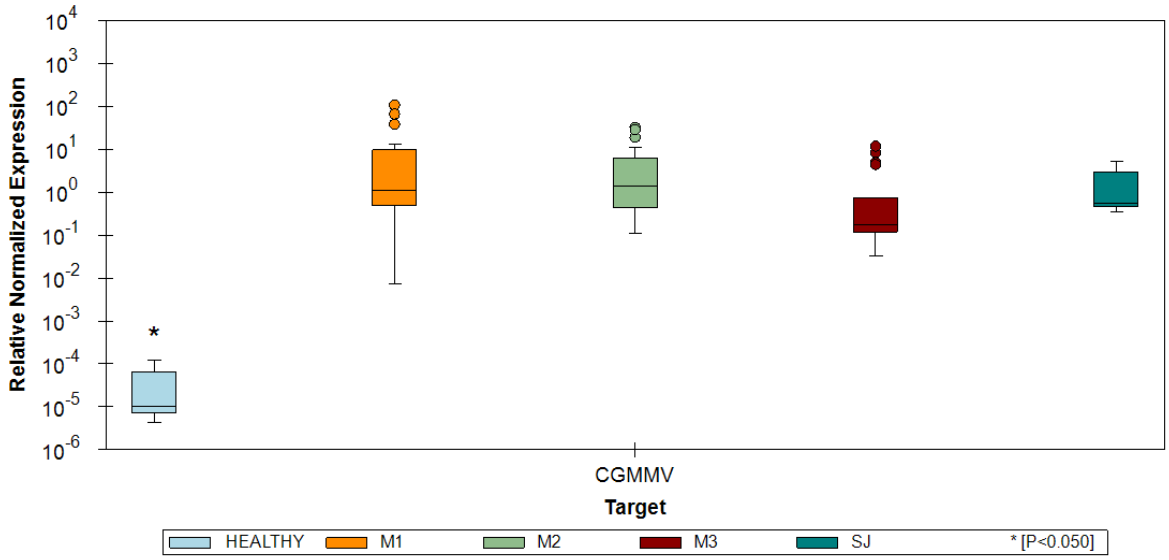

Figure 12. Chart of relative normalized accumulation of wild-type CGMMV compared to accumulation of the mutants in cucurbit plants. The only significant difference in viral accumulation was between the healthy control and CGMMV infected plants (indicated by asterisk, P≤??, by what statistical test?).

Figure 13. CGMMV mutants that have been rub-inoculated into N. benthamiana: (a) healthy, (b) mutant 1, (c) mutant 2, (d) mutant 3, (e) wild-type CGMMV. Pictures were taken 10 days post rub inoculation.

CHAPTER II - ORFN is a small ORF overlapping the 126 and 183 ORF of Watermelon green mottle mosaic virus

ABSTRACT

A small ORF (ORFN) was identified through the use of ORF predictors in the genome of the cucurbit-infecting *Tobamovirus* Watermelon green mottle mosaic virus (WGMMV). ORFN overlaps the 126 and 183 kDa protein ORFs in the +1 reading frame register. The predicted ORF N protein contains a bi-partite nuclear localization signal and transient expression of the ORF N protein fused with GFP showed nuclear localization in *Nicotiana benthamiana* plants. When either or both nuclear localization signals were disrupted by mutational analysis, the mutant proteins showed chloroplast localization. Western blots with an ORF N protein specific antibody show a larger protein than the size predicted for the ORF N protein, suggesting a fusion protein or a protein aggregate. Mutations introduced into the ORFN resulted in mutant WGMMV infections which showed less severe symptoms in *N. benthamiana* and cucurbit plants, but no changes in viral titer for wildtype and WGMMV mutants were observed.

INTRODUCTION

Watermelon green mottle mosaic virus (WGMMV) is a relatively newly described virus in the genus *Tobamovirus,* first identified in 2018 in Taiwan after crinkling and mosaic symptoms were observed on plants in watermelon fields (Cheng et al. 2019). The genome sequence (accession number MH837097.1) from this first report was published in NCBI. Since the publication of this sequence, more WGMMV isolates were quickly discovered (Li et al, 2020) (Pitman et al. 2019). The first reported incidence of WGMMV in California was in opo gourds *(Lagenaria siceraria*) found at a Santa Clara farmer's market in 2017 (Pitman et al. 2019).

Because the original inoculum source of WGMMV in California is not known, WGMMV was isolated from different samples, including seed lots and field samples (Pitman, unpublished), and compared at the genome nucleotide sequence level in attempts to assess overall genetic diversity and phylogeny in order to try to determine if all the isolates were highly similar and likely originated from the same place. During the course of analyzing genome sequences for these different WGMMV isolates, a small ORF (ORFN) overlapping the 126 and 183 kDa protein ORFs was discovered. ORFN was observed to be conserved in all WGMMV isolates. Other viruses in the genus *Tobamovirus* (tobamoviruses) do not have this ORF, thus arousing our curiosity as to its potential significance.

Since WGMMV is a relatively newly discovered tobamovirus, there are few published full-length sequences to use for comparison and analysis. However, it is important to note that even though there are only six sequences so far, all of them contain the ORFN sequence. This suggests that this is a highly conserved region of the WGMMV genome and the ORF may be a real "gene", lending evidence to the existence of an ORFN-encoded protein. In this work I characterized the putative ORF N protein and assessed its role in plant infection.

MATERIALS AND METHODS

Analysis of the 6 WGMMV published genomes and identification of ORFN

Various cucurbit seed lots were received from seed companies in California in order to isolate Cucumber green mottle mosaic virus (CGMMV) by rub inoculation onto *Nicotiana benthamiana* plants. In the process of doing so, it was noticed that some plants had slightly different symptoms than the others. Using degenerate tobamovirus primers (Li et al. 2018) to perform RT-PCR led to the discovery of a mixed infection of CGMMV and WGMMV. Because both of these viruses are tobamoviruses, virions of both were isolated simultaneously from the doubly-infected plants by using the same virion precipitation protocol (C. Sun 1972). The resulting virions were used for RNA extraction by using TRIzol (Life Technologies) and RNAs were visualized on 1% 1xHEPES agarose gels to confirm the size and integrity of the genomic RNA, as well as the presence of the two subgenomic RNAs (Figure 1). The resulting RNAs were quantified by nanodrop spectrophotometry and submitted to Amaryllis Nucleics (Oakland, CA) for library preparation and Illumina sequencing. Next Generation Sequencing (NGS) data derived from virion RNA samples were received from Amaryllis. *De novo* genome assemblies using the program RNASPADES (Bushmanova et al. 2019) from the mix infected samples for CGMMV and WGMMV yielded three new complete WGMMV genomes (accession numbers MT184947.1, MT184946.1, MT184945.1) to add to the three that existed on Genbank (accession numbers MN971594.1, MH837097.1, MK070867.1). The total of six genomes were further analyzed using Snapgene to identify ORFs.

Western Blots with ORF N protein antibodies

The deduced amino acid sequence of the putative ORF N protein was submitted to Biomatik (Wilminton, Delaware), where two antigenic sites were predicted using an in-house program (Figure 8). Peptides to both sites were synthesized and used to produce two polyclonal antibodies from rabbits for use in serological analysis.

Proteins were separated using 13% acrylamide tris-tricine SDS-PAGE gels (Schägger 2006) using proteins prepared by subcellular fractionation (Holden and Horton 2009) of wild-type WGMMV infected *Nicotiana benthamiana* plants, and *N. benthamiana* plants agroinfiltrated to express the green fluorescent protein (GFP)-fused with the ORF N protein as a positive control.

Twenty µL of the fractionation protein extractions were loaded into the polyacrylamide gels. Samples were stacked in the 4% stacking gel at 30V until the loading dye reached the resolving gel, after which the voltage was increased to 150V to separate the proteins by size. The separated proteins were then wet transferred to a 0.22-micron nitrocellulose membrane at 100V for 45 minutes for western blot analysis. Both of the ORFN putative protein antibodies were tested on these samples in different concentrations (1 µg/mL, 2 µg/mL, 3 µg/mL, and 4 µg/mL) to determine if they could detect the putative ORF N protein, as well as to identify what concentrations to use for subsequent western blots. Before the application of the primary antibodies, membranes were incubated in 7% milk/Tris buffered saline with 0.3 % Tween 20 (TBST) blocking solution for one hour. The membranes were incubated with these antibodies at different concentrations in the same 7% milk/ TBST blocking solution overnight.

The following day, the membranes were washed three times with TBST for 5 minutes. Secondary antibody Goat Anti-Rabbit IgG (H + L)-HRP Conjugate (BioRad) was added at a dilution of 1:3000 to the membranes, again in a 7% milk/TBST buffer for one hour. Again, the membranes were washed three times with TBST. The washed membranes were incubated in Pierce ECL Western Blotting Substrate before they were exposed to x-ray film to confirm the size and existence of the protein.

Expression of the ORF N protein fused with GFP

Using the NEBuilder program (https://nebuilder.neb.com/), overlapping primers (Table 1) were designed to fuse the ORFN coding sequence to that of GFP such that products could be ligated into the transient expression plasmid vector pEAQ. Amino and C'-terminal fusion proteins were generated and later compared (Figure 2). GFP and ORFN coding sequences were amplified using Takara's CloneAmp HiFi polymerase (Table 2). The DNAs generated from these PCR reactions were analyzed by electrophoresis on 2% agarose gels in 0.5x TAE buffer at 100V for 30 minutes. The DNAs of expected size were excised from the gel and purified using the Zymoclean Gel DNA Recovery Kit (Zymo). Corresponding DNA fragments were then ligated into plasmid pEAQ (linearized with NruI and StuI (New England Biolabs)) with NEBuilder HiFi DNA Assembly (New England Biolabs). Circularized plasmids were transformed into *Escherichia coli* (strain DH5α) to increase copy number. Plasmids were then purified using QIAprep Spin Miniprep Kit (QIAGEN). The plasmids were analyzed with a combination of restriction enzyme digestion (BamHI-HF and PacI (New England Biolabs)) followed by gel electrophoresis, and by Sanger sequencing using primers flanking the insert region (Table 3).

After the correct sequences were verified, plasmids were transformed into *Agrobacterium tumefaciens* (strain GV3101). *Agrobacterium tumefaciens* containing the fusion protein vectors were grown in Luria-Bertani (LB) liquid media (Bertani 2004) containing their corresponding antibiotics for approximately two days. The *A. tumefaciens* cells were pelleted by centrifugation and resuspended in a modified induction buffer (10 mM $MgCl_2$ containing 100 µM acetosyringone) to an optical density (OD) of 0.02 at 600 nm in a spectrophotometer. Cultures were incubated in the modified induction buffer at room temperature for at least four hours before young, wild-type *N. benthamiana* plants were infiltrated (Li 2011). After two days, the leaves were checked by confocal microscopy for fluorescent protein expression and localization.

Creation of WGMMV infectious clones and infectious mutants

An infectious clone of WGMMV was constructed by using primers (Table 4) and RT-PCR to generate two halves of the WGMMV genome (as cDNA) for ease of cloning (Figure 3). Two primers corresponding to approximately the middle of the WGMMV genome were generated to overlap so that they could be ligated together using NEBuilder HiFi DNA Assembly (New England Biolabs) after following the protocol described previously to clone sequences encoding the fusion proteins. The full-length cDNA of the WGMMV genome was then ligated into vector pJL89 using the same DNA assembly kit.

Plasmids with the full-length cDNA were transformed into *E. coli* DH5α to increase copy number for further analysis. DH5α containing the plasmids were spread onto LB plates containing kanamycin. Ten colonies were chosen from each plate and grown in 3 mL LB media containing kanamycin overnight at 37°C. The cells were pelleted by centrifugation, then plasmids were purified with the Qiagen miniprep kit, and sent for sequencing to confirm the sequence of the full-length cDNA. Plasmids that contained the correct and full sequence were transformed into *A. tumefaciens* and used to infiltrate wild-type *N. benthamiana* plants to test for infectivity.

Using the same method of overlapping primers described above, mutants were created for ORFN in WGMMV. Because ORFN overlaps the 126kD ORF, care was taken to ensure that changes in ORFN would not affect the 126 kDa protein amino acid sequence. Primers used to create these mutants are listed in Table 5. The same protocols for amplification, ligation, and transformation were used to create the mutants: start, M1, M2, and double, which is a combination of both the M1 and M2 mutations in the same full-length WGMMV genome (Table 7). Plasmids containing these mutants were checked by sequencing, then transformed into *A. tumefaciens* for infiltration of *N. benthamiana* plants.

After successfully establishing infections with the wildtype and WGMMV mutants, RT-PCR and Sanger sequencing were used to confirm that the intended sequences were maintained for each construct. Tissues from the infected *N. benthamiana* plants were used to inoculate plants of the *Cucumis melo* varieties Sumter and Expedition. Ten plants from each variety were inoculated after the emergence of their cotyledons. Ten days after inoculation, 0.1 g tissue was collected from the newest leaves of each plant and homogenized. RNA was purified from these samples using TRIzol and cDNA was generated using iScript cDNA Kit, which was subsequently used to perform SYBR qPCR using the following WGMMV specific primers and cytochrome oxidase primers as a housekeeping gene (Table 6).

Using a combination of the mutant primers described in Table 1 and Table 5, mutated ORFN sequences except for the start mutation were fused to the GFP coding sequence in the transient expression vector pEAQ, created for the rapid, high yield expression of recombinant proteins via Cowpea Mosaic Virus's expression system (Peyret and Lomonossoff 2013). This resulted so that GFP was fused to either the amino or carboxy terminal of the ORF N protein (Figure 5).

RESULTS

Nucleotide sequence analysis of the six WGMMV genomes and identification of the ORFN sequence

After *de novo* assembly of WGMMV genome sequences, ORFs were compared using Snapgene. A small ORF overlapping the ORF encoding the 126 kDa protein in the +1 reading frame was observed. This ORF, referred to as ORFN, was found in all three newly assembled genomes (accession numbers MT184947.1, MT184946.1, MT184945.1) and in all three available WGMMV genomes on Genbank (accession numbers MN971594.1, MH837097.1, MK070867.1), suggesting that it could be a true ORF and not an assembly artifact. ORFN-like ORFs were searched for in all tobamoviruses, referencing NCBI's database of tobamovirus representative genomes (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=12234). Of the 34 sequences analyzed, no ORFN-like sequences were found. The ORFN putative protein sequence is highly conserved in all WGMMV isolates, having only seven amino acids different among them (92% identity).

Expression of ORFN:GFP fusion proteins in N. benthamiana plants

In attempts to observe the localization of the ORF N protein in cells of *N. benthamiana* plants, ORFN:GFP fusion proteins were expressed in leaves of *N. benthamiana* plants. Agroinfiltration of *N. benthamiana* plants was performed as described and leaf tissues were examined by confocal microscopy. GFP was fused to either the N- or C-terminus of the ORF N protein to ensure that it would not interfere with any localization signals at either terminus. When examining the infiltrated leaf tissues by confocal microscopy, the fusion proteins were seen localizing to the nuclei (Figure 7). In order to confirm that these globular structures were nuclei and not chloroplasts, the chloroplasts' autofluorescing nature was used to visualize them. This confirmed that the ORFN:GFP fusion proteins were indeed localizing to the nucleus and not to chloroplasts. These fusion proteins were then used as positive controls in subsequent western blots.

Western Blots with ORF N antibody

Two antigenic regions were reported (Figure 8), and from these regions peptides were synthesized to generate antibodies for the ORF N protein. Both antibodies were used at concentrations from 1 µg/mL to 4 µg/mL in western blots to optimize concentrations for later western blots. Samples included the GFP fusions as a positive control. One of the antibodies failed to give a positive reaction, regardless of concentration, while the other showed positive reactions at 4 µg/mL. Other methods, such as attempting different blocking substrates (7% nonfat dry milk in 1x TBST, TSW buffer (10 mM Tris-HCl, pH 7.4, 0.9% NaCl, 0.25% gelatin, 0.02% SDS, 0.1% Triton X-100)), a different membrane (PVDF instead of nitrocellulose), and varying transfer times from the SDS-PAGE gel to the membrane (15-45 minutes) were used to identify optimal conditions for binding of the antibody, but none of the methods tried were successful.

The antibody ORFN-AB was used to probe subcellular fractionation extracts from healthy and WGMMV-infected *N. benthamiana* plants (Holden and Horton 2009) and from plants expressing the ORFN:GFP fusion protein, used as a positive control. Subcellular fractions were used in the hopes of concentrating the small proteins as well as supporting the results obtained by confocal microscopy. The samples were separated on a 13% acrylamide tris-tricine gel (Schägger 2006) and transferred to 0.22 micron nitrocellulose membrane, blocked for 30 minutes with a 7% milk and TBST solution before being probed with the antibody overnight. The membrane was washed in the morning before being probed with the secondary antibody. The proteins were detected using Pierce chemiluminescent substrate and x-ray films. Interestingly, the protein of the expected size for this ORF (~12 kDa) was not present in any of the cellular fractions. However, a larger protein of ~35 kDa was present in the fraction that contains the most nuclei (P1). This could be due to the aggregation of ORF N proteins, or it could be a fusion with another viral protein. The ORFN:GFP fusion protein is the predicted size of ~37 kDa.

Creation of infectious clones and infectious mutants

Because ORFN:GFP fusion protein localized to the nucleus, the amino acid sequence of the ORF N protein was analyzed using a nuclear localization signal predictor, cNLS Mapper (Kosugi et al. 2009). A predicted nuclear localization signal was observed in the center of the protein (Fig. 9) and was also predicted to be a bi-partite signal (Figure 11). Using this prediction, four mutants were created: M1, M2, M3, and D. Arginines (R) and lysines (K) from the nuclear localization signal sequence were mutated to other amino acids without changing the amino acid sequence of the 126 kD protein, as arginines and lysines are the amino acids most commonly responsible for the nuclear localization (Makkerh, Dingwall, and Laskey 1996). The resulting amino acid sequences are shown in Table 7.

These four mutants, and wildtype WGMMV, were then used to inoculate *N. benthamiana* plants. Mutants M1 and M2 were easily transmissible by agroinoculation. However, M3 and D were not able to systemically infect *N. benthamiana* via agroinfiltration. Since it is sometimes difficult to overcome plants' RNA interference (RNAi) defenses when inoculating via agroinfiltration, p19 silencing suppressor can be used to interfere with RNAi defenses and allow a more robust infection. In this case, the addition of p19, a silencing suppressor (Lakatos et al. 2004) allowed for establishing systemic infections for mutants M3 and D. These were confirmed by conventional PCR. Interestingly, after the successful infection of *N. benthamiana* by agroinfiltration, all four mutants were able to be rub inoculated into new plant hosts, including *Cucumis melo.*

In order to test the infection efficacy of the mutant WGMMVs in the native host, ten plants each of Sumter and Expedition melon were rub inoculated with infected *N. benthamiana* tissue to simulate a natural infection. Tissue was collected ten days post inoculation, RNA extracted, cDNA synthesized, and qPCR performed. Symptoms recorded at this time showed a difference in

severity: the wild-type infected melon plants showed more severe leave mottling and crinkling than did plants infected with the mutants (Figure 12). Between the mutants there were no definitive, consistent differences in the symptoms observed (Figure 13). In addition to comparing symptoms, the relative levels of WGMMV RNA were compared between plants infected by wildtype and mutant WGMMVs. Quantification of the WGMMV RNA copy number in these plants showed that there was no significant difference in the accumulation when wildtype and mutant WGMMV constructs were compared (Figure 13).

Expression of mutants with GFP

In order to determine if the nuclear localization of the ORF N protein was disrupted by the mutations, mutant ORF N proteins were fused to GFP, and localization in agroinfiltrated *N. benthamiana* plants was assessed by confocal microscopy. These analyses showed that when GFP was fused to the N-terminus of mutant ORF N proteins, the fusion protein behaved much like free GFP, localization was dispersed in the cell. However, when GFP was fused to the C-terminus of the mutant ORF N proteins, it is clear that the localization changed from the nucleus to the chloroplasts (Figure 14).

DISCUSSION

The work presented here showed the identification of a novel, small ORF in all six known WGMMV isolates. Although only WGMMV sequences were discussed, representatives from all 34 tobamoviruses were analyzed using the same program, Snapgene, to search for ORFN-like sequences. Some contained ORFs in similar places, but when the amino acid sequences were compared to that of the ORF N putative protein, they were too different (<50% amino acid identity) to be considered ORF N homologs. Furthermore, analysis of these other tobamovirus protein sequences also showed no nuclear localization signal as shown here for the ORF N protein.

Because the three new sequences of WGMMV were derived from mixed infections with CGMMV, going back to review old CGMMV Illumina sequencing data might yield more WGMMV isolates to look for variations in ORFN. Discovering more isolates that contain ORFN and determining how conserved this protein/nucleotide sequence is within these isolates will help strengthen the argument that this protein does indeed exist.

In addition to searching for ORFN and ORF N-like proteins *in silico*, the western blot for ORFN should also be optimized. Initial western blots with the antibody showed that the protein band for ORF N is larger than expected (~37 kDa, the same size as the ORFN:GFP fusion protein). Whether or not this is the true size of ORFN should be further studied. If the protein band is distinct enough in the SDS PAGE gel after it is stained, it can be extracted and sent for sequencing via mass spectrometry.

ORF N has been shown to localize to the nucleus (Figure 6), but whether it interacts with other viral proteins has yet to be seen. By creating fusion proteins with the characterized proteins of WGMMV and a fluorescing protein such as mCherry, the confocal microscope can be used to visualize any interactions between the larger proteins of WGMMV and the ORF N protein.

Whether or not the ORFN and its putative encoded protein are biologically relevant (i.e. newly discovered gene and its protein) was not definitively demonstrated here. However, ORF N:GFP fusion proteins were found to localize in nuclei of plant cells and this was predicted by analysis of the amino acid sequence. Furthermore, WGMMV ORF N mutants showed less severe symptoms than wild type WGMMV, suggesting that the ORF N protein may play a role in symptom determination. The necessity of including the silencing suppressor p19 along with mutants M3 and D during agroinfiltration, but not in the subsequent rub inoculations, suggests that mutating the “tail” end of the NLS was detrimental to the virus’ ability to overcome the initial silencing by the plant.

The plethora of new virus genome sequences and their analysis using modern bioinformatics programs has already led to the discovery of previously unrecognized genes and encoded proteins. One example is the PIPO protein of potyviruses, where a small ORF was discovered in the genome of Turnip mosaic virus using the software package MLOGD. This small ORF is embedded within the region encoding P3 of the polyprotein, but translated in the +2 reading frame (Chung et al. 2008). Similarly, the small protein V3 was discovered in the Tomato yellow leaf curl virus (TYLCV), and functions as an RNA silencing suppressor (Gong et al. 2021). It seems highly probable that continued analyses will lead to discovering more proteins. Thus, it is not necessarily surprising that ORFN was discovered here, but as can be seen, demonstrating a role in virus biology is not simple or straightforward.

TABLES CHAPTER 2

	Primer Name	Sequence	Amplified Gene
C-terminal GFP	C-WORF-F1	ATA TTC TGC CCA AAT TCG CGA TGA CAG CAT ACA GAA GTA TGT CAA C	ORFN
	C-WORF-R1	CCT TGC TCA CGT AAG GCC GGA CCG ATG TC	ORFN
	C-WORF-GFP-F1	CCG GCC TTA CGT GAG CAA GGG CGA GGA G	GFP
	C-GFP-R2	aat gaa acc aga gtt aaa ggc ctt taC TTG TAC AGC TCG TCC	GFP
N-terminal GFP	N-WORF-F1	GCT GTA CAA GAC AGC ATA CAG AAG TAT GTC AAC AAG	ORFN
	N-WORF-R1	ACC AGA GTT AAA GGC CTT TAG TAA GGC CGG ACC GAT GTC	ORFN
	N-GFP-F1	gta tat tct gcc caa att cgC GAT GGT GAG CAA GGG CG	GFP
	N-WORF-GFP-R1	TGT ATG CTG TCT TGT ACA GCT CGT CCA TGC	GFP

Table 13. Sequences of primers used to generate constructs for transient expression of GFP and ORFN fusion proteins. N and C indicate which end of the protein that the GFP sequence is connected to, with N being the free amino group terminus and C being the free carboxyl group terminus. Capital letters represent the protein coding sequences, while lowercase letters sequences are from the plasmid pEAQ. F and R in the primer names are for forward and reverse primers, respectively.

Reagent	Concentration (µM)	µL/rxn
CloneAmp HiFi Premix	-	25
Primer F	10	1.5
Primer R	10	1.5
Water	-	19
cDNA/DNA	-	3
	Total rxn volume	50

Table 14. CloneAmp Hifi Premix used to amplify the ORFN sequence for cloning into plasmid pEAQ.

Primer Name	Sequence
pEAQ-seq-F	gcttctgtatattctgcc
pEAQ-seq-R	gcacaccgaataacagtaaattc

Table 15. Primers used to sequence fusion construct plasmids

Primer Name	Sequence
IC-F1	GTT CAT TTC ATT TGG AGA GGG TAT TTC TTT CAC TAC AAC AAC
IC-R1	CAT CTA TCG TAA TAG AAT TGC ATG TCA GAT ATA G
IC-F2	CAA TTC TAT TAC GAT AGA TGT CTT CCC GGT AAT TCC TTT GTG CTT AAT G
IC-R2	TGG AGA TGC CAT GCC GAC CCT GGG CCC CTA CCC GGG GA

Table 16. Primers used to amplify the full WGMMV genome for the infectious clone (IC).

Primer Name	Sequence
START-F	GTG ATG TAT CAC GAC AGC ATA CAG AAG
START-R	CTT CTG TAT GCT GTC GTG ATA CAT CAC
M1-F	GGC ACG TTC TGG GGG CGG GGA TTT GTG CT
M1-R	CAG GTC AAA GCA ATC CAA AGG TAT CAT TCT G
M2-F	CAA ATC CCC GCC CCC AGA ACG TGC C
M2-R	CAG AAT GAT ACC TTT GGA TTG CTT TGA CCT G

Table 17. Overlapping primers used to generate the ORF N mutants. These primers were paired with IC-F1 and IC-R2 (end primers) to create the two "halves" of the full-length WGMMV genome (as cDNA).

Primer Name	Seqeunce
Cox F	CGT CGC ATT CCA GAT TAT CCA
Cox R	CAA CTA CGG ATA TAT AAG RRC CRR AAC TG
WGMMV-CP F	GCAGTCGCTGGATACAAGAA
WGMMV-CP R	GACATGCGAACTGACAGAAGA

Table 18. Primers used in RT-QPCR to quantify WGMMV. WGMMV-CP primers bind to the coat protein coding region of WGMMV. Cox primers are for the housekeeping gene, cytochrome oxidase.

MUTANTS	MUTATION	
M1	START CODON	MTAYRSMS
		TTAYRSMS
M2	BEGINNING OF NLS	RIRAPKTCLQR
		QIPAPGTCLQR
M3	END OF NLS	RISKSRRSR
		RISRSKQSK
D	M2+M3	RIRAPKTCLQRPQLPDMPGGSYRISKSRRSR
		QIPAPGTCLQRPQLPDMPGGSYRISRSKQSK

Table 19. Regions that were mutated in the ORF N putative protein. The table shows which amino acids were changed to the new amino acid (in red). Sequences on the top row for each mutant are the original amino acid sequence, while the bottom row is the mutated amino acid change.

FIGURES CHAPTER 2

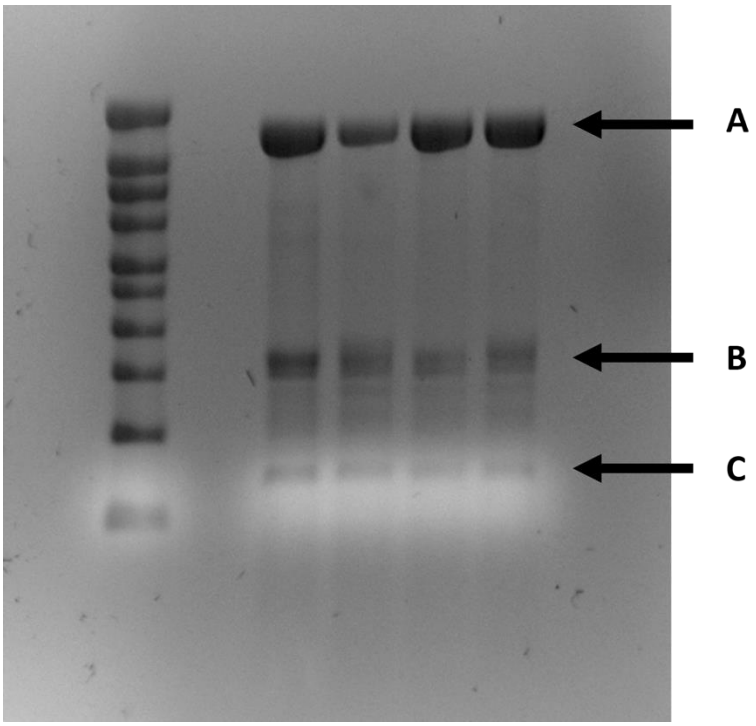

Figure 14. RNA gel to verify the presence and quality of the genomic and subgenomic RNAs of WGMMV. Four samples are displayed next to RNA ladder from Invitrogen. Arrows indicate A) the genomic RNA (6.4 kb), B) subgenomic RNA 1 (~0.8 kb), and C) subgenomic RNA 2 (0.5 kb).

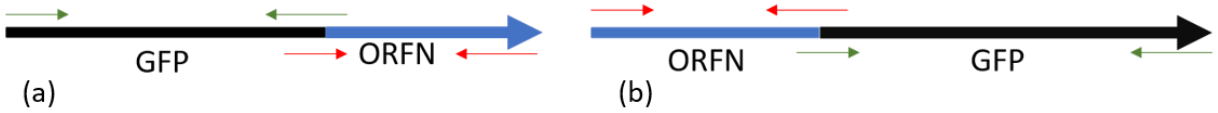

Figure 15. (a) and (b) show where the primers bind and overlap to create constructs for use in transient expression of the fusion proteins. (a) is for the creation of the N-terminal fused GFP, (b) is for the C-terminal fused GFP. Not to scale.

Figure 16. Binding sites for primers to "split" the virus for ease of cloning. Primers IC-F1 and IC-R1(Table 4) bind to the green arrows, while IC-F2 and IC-R2 (Table 4) bind to the red arrows. PCR sizes are ~2.4 kb for green arrows and 4 kb for red arrows. Figure not to scale.

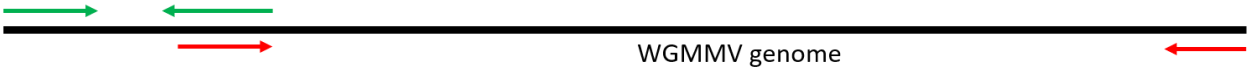

Figure 17. Diagram of where the overlapping and end primers fall. Green arrows indicate primers from Table 5, red arrows indicate primers from Table 4. The overlapping red and green primers on the WGMMV genome are where ORFN is located (nucleotides 483 – 761). Not to scale.

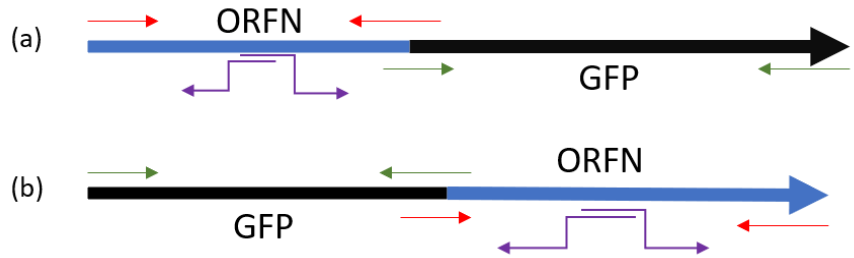

Figure 18. Primer binding sites for mutations. Purple arrows are the mutated binding sites, while the rest of the primers used are from the section describing the fusion of GFP to the original sequence of ORFN. Not to scale.

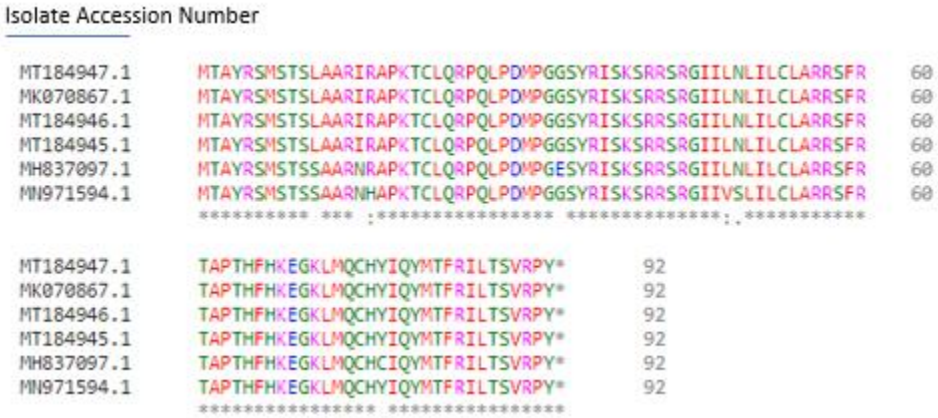

Figure 19. Clustal Omega alignment of ORFN putative proteins found in WGMMV isolates. Asterisks indicate identical amino acids among all proteins.

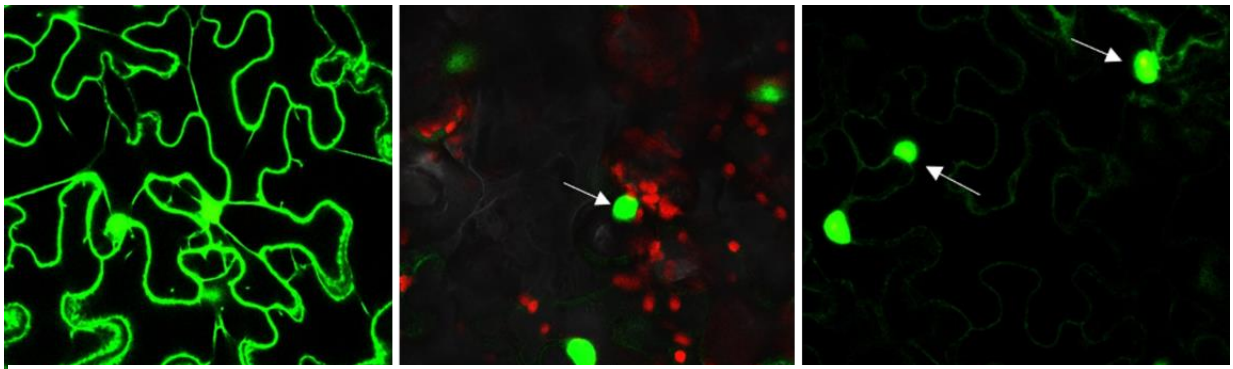

*Figure 20. Confocal images of ORFN:GFP fusion proteins in N. benthamiana leaf epidermal cells. **Left**: free GFP. **Middle**: GFP fused to the C-terminal of ORFN, with arrows indicating the nuclear localization. Chloroplasts are shown in red. **Right**: GFP fused to the N-terminal of ORFN. Arrows indicate nuclear localization.*

ORFN-NO ORFN-AB

MTAYRSMSTSLAARIRAPKTCLQRPQLPDMPGGSYRISKSRRSRGIILNLILCLARRSFRTAPTHFHKEGKLMQCHYIQYMTFRILTSVRPY

Figure 21. Antibody binding sites (shown in red) predicted by BioMatik: ORFN-NO is the antibody that did not bind to any proteins on the western blot, while ORFN-AB was able to bind to the ORFN:GFP fusion protein.

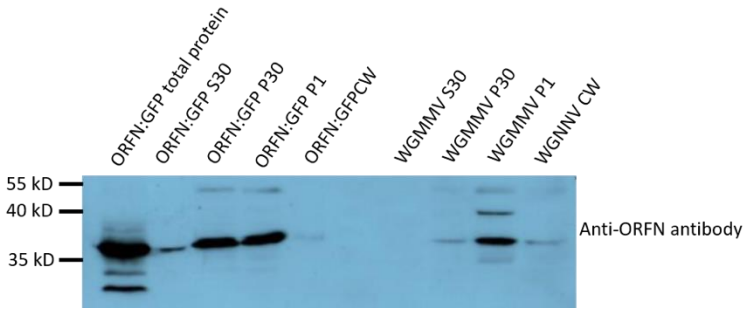

Figure 22. Western blot of proteins prepared by subcellular fractionation. Fusion proteins from infiltrated leaves are at left, proteins from WGMMV-infected plants are at the right. Arrows depict size markers and are shown at left.

Figure 23. Predicted NLS sequence (in red) of the ORF N putative protein.

Figure 24. Mutation locations on the ORF N putative protein sequence. The two parts of the NLS are in red.

Figure 25. Symptom comparisons for wildtype and mutant WGMMV in Expedition melon plants. (a) healthy, (b) wild-type infected, (c) M1, (d) M2, (e) M3, (f) D

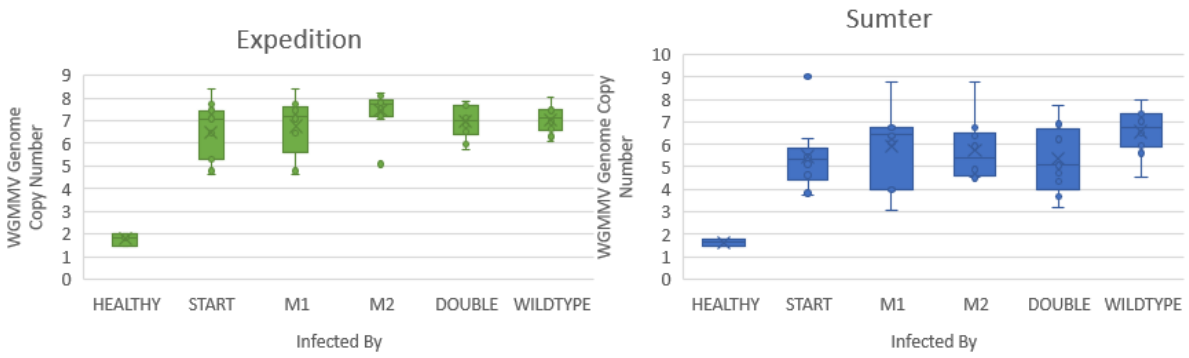

Figure 26. Graphs of WGMMV wildtype and mutant gene copy number between the two different melon varieties Expedition and Sumter. The only significant difference is between the healthy plants and the WGMMV- infected plants.

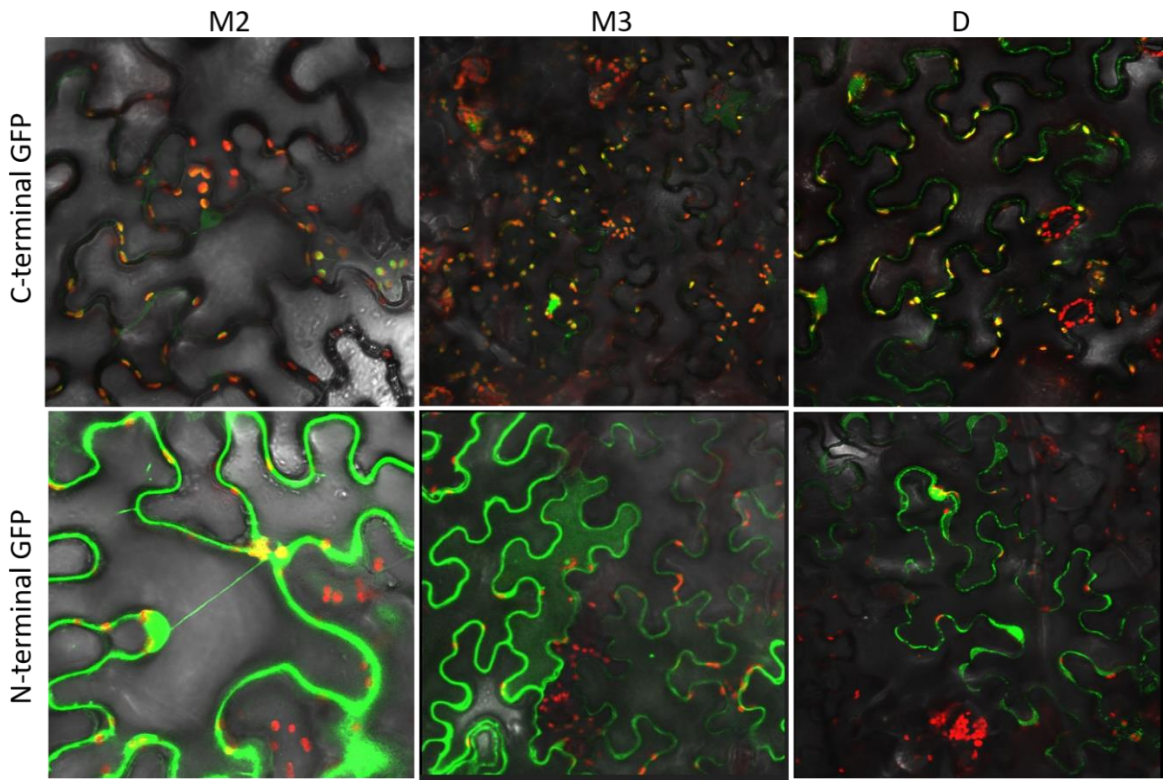

Figure 27. Fusion proteins (GFP and mutated ORF N protein) in N. benthamiana *plant epidermal cells.* ***Top row, left to right****: Carboxy-terminal GFP fused to mutant 2 (M2), mutant 3 (M3), and double mutant (D) of the ORF N protein, respectively, showing the change of localization from the nucleus (Figure 7) to the chloroplasts.* ***Bottom row, left to right****: Amino-terminal GFP fused to mutant 2 (M2), mutant 3 (M3), and double mutant (D) of the ORF N protein, respectively, showing the free-GFP-like behavior of the mutant fusion proteins.*

CONCLUSIONS

Tobamoviruses, specifically Tobacco mosaic virus (TMV), were the first viruses to be identified and partially characterized in 1935 (Stanley 1935). Twenty-five years later, the complete amino acid sequence of the coat protein, then referred to as “the protein” of TMV was published (Tsugita et al. 1960). Since then, TMV has become one of the most intensely studied viruses. Several studies on the proteins of TMV have been published to elucidate their function(s) in virus replication, *in planta* movement, and combatting host defenses. Among the many gene studies that have been reported are interactions between host and viral proteins to study cell-to-cell movement (Chen et al. 2000), as well as creating mutants of viral proteins in order to deter cell-to-cell movement (Bendahmane et al. 2002). By examining these proteins and determining their importance to the virus infection cycle, the TMV replication/translation machinery can be exploited to perform other tasks, such as creating TMV-based vectors for rapid, high-level transient expression of proteins, regardless of the organism of origin (Kagale et al. 2012).

However, the amount of time that has passed between the first discovery of TMV and the current number of studies that are still being published about this virus highlights/suggests the number of unknowns that are remaining. The recent advancements in sequence analysis, bioinformatics, and ORF prediction software have led to the discovery of yet another TMV ORF: ORF6, which encodes a small, 4.8 kDa protein which was determined to be a contributing factor in viral pathogenicity in *N. benthamiana* plants (Canto, MacFarlane, and Palukaitis 2004), then further characterized by subcellular localization in 2017 (Erokhina et al. 2017). Similarly, the *Begomovirus,* Tomato yellow leaf curl virus (TYLCV), was first identified in 1964 (Cohen and Harpaz 1964), and all major proteins required for replication and *in planta* movement were identified in the years closely following TYLCV discovery (Navot et al. 1991). But just last year, TYLCV was also shown to encode a small protein: V3, a 9.3 kDa protein identified as a silencing suppressor (Gong et al. 2021).

Improvements in genome analysis and ORF prediction software, as well as their increasing availability and the rapid accumulation of virus sequences have led to multiple discoveries of small ORFs that were either previously not recognized or overlooked. The identification and characterization of these ORFs and their encoded proteins offer new opportunities to discover previously unrecognized factors facilitating virus:host interactions. In this thesis, I discovered two new ORFs and made progress towards elucidating how their respective encoded proteins interact with the plant host cell. More needs to be done to study the intricacies of how CGMMV and WGMMV interact with their hosts, and how virus-encoded proteins interact with those of the host. I believe that taking a fresh look at small ORFs and encoded proteins offers great potential to further understand virus:host interactions.

REFERENCES

Ainsworth, G. C. 1935. “Mosaic Diseases of the Cucumber.” *Annals of Applied Biology* 22(1):55–67. doi: 10.1111/j.1744-7348.1935.tb07708.x.

Bendahmane, Mohammed, Judit Szécsi, Iju Chen, R. Howard Berg, and Roger N. Beachy. 2002. “Characterization of Mutant Tobacco Mosaic Virus Coat Protein That Interferes with Virus Cell-to-Cell Movement.” *Proceedings of the National Academy of Sciences of the United States of America* 99(6). doi: 10.1073/pnas.062041499.

Bertani, Giuseppe. 2004. “Lysogeny at Mid-Twentieth Century: P1, P2, and Other Experimental Systems.” *Journal of Bacteriology* 186(3).

Bushmanova, Elena, Dmitry Antipov, Alla Lapidus, and Andrey D. Prjibelski. 2019. “RnaSPAdes: A de Novo Transcriptome Assembler and Its Application to RNA-Seq Data.” *GigaScience* 8(9). doi: 10.1093/gigascience/giz100.

C. Sun, M. K. 1972. “Purification and Properties of a Severe Strain of Peanut Mottle Virus.” *Phytopathology* 62(8). doi: 10.1094/phyto-62-832.

Canto, Tomas, Stuart A. MacFarlane, and Peter Palukaitis. 2004. “ORF6 of Tobacco Mosaic Virus Is a Determinant of Viral Pathogenicity in Nicotiana Benthamiana.” *Journal of General Virology* 85(10). doi: 10.1099/vir.0.80270-0.

CDFA. 2019. “California Agricultural Statistics Review 2018-19.” *California Agricultural Statistics Review*.

Chen, Min Huei, Jinsong Sheng, Geoffrey Hind, Avtar K. Handa, and Vitaly Citovsky. 2000. “Interaction between the Tobacco Mosaic Virus Movement Protein and Host Cell Pectin Methylesterases Is Required for Viral Cell-to-Cell Movement.” *EMBO Journal* 19(5). doi: 10.1093/emboj/19.5.913.

Cheng, Ying Huey, Chih Hung Huang, Chung Jan Chang, and Fuh Jyh Jan. 2019. “Identification and Characterisation of Watermelon Green Mottle Mosaic Virus as a New Cucurbit-Infecting Tobamovirus.” *Annals of Applied Biology* 174(1). doi: 10.1111/aab.12467.

Chung, Betty Y. W., W. Allen Miller, John F. Atkins, and Andrew E. Firth. 2008. “An Overlapping Essential Gene in the Potyviridae.” *Proceedings of the National Academy of Sciences of the United States of America* 105(15). doi: 10.1073/pnas.0800468105.

Cohen, S., and I. Harpaz. 1964. “Periodic, Rather than Continual Acquisition of a New Tomato Virus by Its Vector, the Tobacco Whitefly (Bemisia Tabaci Gennadius).” *Entomologia Experimentalis et Applicata* 7(2). doi: 10.1111/j.1570-7458.1964.tb02435.x.

Dawson, William O., and Mark E. Hilf. 1992. “Host-Range Determinants of Plant Viruses.” *Annual Review of Plant Physiology and Plant Molecular Biology* 43(1). doi: 10.1146/annurev.pp.43.060192.002523.

Donald, R. G. K., H. Zhou, and A. O. Jackson. 1993. “Serological Analysis of Barley Stripe Mosaic Virus-Encoded Proteins in Infected Barley.” *Virology* 195(2). doi: 10.1006/viro.1993.1417.

Erokhina, T. N., E. A. Lazareva, K. R. Richert-Pöggeler, E. V. Sheval, A. G. Solovyev, and S. Y. Morozov. 2017. “Subcellular Localization and Detection of Tobacco Mosaic Virus ORF6 Protein by Immunoelectron Microscopy.” *Biochemistry (Moscow)* 82(1). doi: 10.1134/S0006297917010060.

Gibbs, Andian J., D. Fargette, F. García-Arenal, and M. J. Gibbs. 2010. “Time - The Emerging Dimension of Plant Virus Studies.” *Journal of General Virology* 91(1).

Gong, Pan, Huang Tan, Siwen Zhao, Hao Li, Hui Liu, Yu Ma, Xi Zhang, Junjie Rong, Xing Fu, Rosa Lozano-Durán, Fangfang Li, and Xueping Zhou. 2021. “Geminiviruses Encode Additional Small Proteins with Specific Subcellular Localizations and Virulence Function.” *Nature Communications* 12(1). doi: 10.1038/s41467-021-24617-4.

Holden, Paul, and William Horton. 2009. “Crude Subcellular Fractionation of Cultured Mammalian Cell Lines.” *BMC Research Notes* 2. doi: 10.1186/1756-0500-2-243.

Kagale, Sateesh, Shihomi Uzuhashi, Merek Wigness, Tricia Bender, Wen Yang, M. Hossein Borhan, and Kevin Rozwadowski. 2012. “TMV-Gate Vectors: Gateway Compatible Tobacco Mosaic Virus Based Expression Vectors for Functional Analysis of Proteins.” *Scientific Reports* 2. doi: 10.1038/srep00874.

Van Kammen, A. 1967. “Purification and Properties of the Components of Cowpea Mosaic Virus.” *Virology* 31(4). doi: 10.1016/0042-6822(67)90192-4.

Kosugi, Shunichi, Masako Hasebe, Masaru Tomita, and Hiroshi Yanagawa. 2009. “Systematic Identification of Cell Cycle-Dependent Yeast Nucleocytoplasmic Shuttling Proteins by Prediction of Composite Motifs.” *Proceedings of the National Academy of Sciences of the United States of America* 106(25). doi: 10.1073/pnas.0900604106.

Lakatos, Lóránt, György Szittya, Dániel Silhavy, and József Burgyán. 2004. “Molecular Mechanism of RNA Silencing Suppression Mediated by P19 Protein of Tombusviruses.” *EMBO Journal* 23(4). doi: 10.1038/sj.emboj.7600096.

Lefkowitz, Elliot J., Donald M. Dempsey, Robert Curtis Hendrickson, Richard J. Orton, Stuart G. Siddell, and Donald B. Smith. 2018. “Virus Taxonomy: The Database of the International Committee on Taxonomy of Viruses (ICTV).” *Nucleic Acids Research* 46(D1). doi: 10.1093/nar/gkx932.

Li, Xiyan. 2011. “Infiltration of Nicotiana Benthamiana Protocol for Transient Expression via Agrobacterium.” *BIO-PROTOCOL* 1(14). doi: 10.21769/bioprotoc.95.

Li, Yueyue, Guanlin Tan, Pingxiu Lan, Ansheng Zhang, Yong Liu, Ruhui Li, and Fan Li. 2018. “Detection of Tobamoviruses by RT-PCR Using a Novel Pair of Degenerate Primers.” *Journal of Virological Methods* 259. doi: 10.1016/j.jviromet.2018.06.012.

Makkerh, Joe P. S., Colin Dingwall, and Ronald A. Laskey. 1996. “Comparative Mutagenesis of Nuclear Localization Signals Reveals the Importance of Neutral and Acidic Amino Acids.”

Current Biology 6(8). doi: 10.1016/S0960-9822(02)00648-6.

Navot, Nir, Eran Pichersky, Muhammad Zeidan, Dani Zamir, and Henryk Czosnek. 1991. "Tomato Yellow Leaf Curl Virus: A Whitefly-Transmitted Geminivirus with a Single Genomic Component." *Virology* 185(1). doi: 10.1016/0042-6822(91)90763-2.

Nelson, Brook K., Xue Cai, and Andreas Nebenführ. 2007. "A Multicolored Set of in Vivo Organelle Markers for Co-Localization Studies in Arabidopsis and Other Plants." *Plant Journal* 51(6). doi: 10.1111/j.1365-313X.2007.03212.x.

Okada, Y. 1999. "Historical Overview of Research on the Tobacco Mosaic Virus Genome: Genome Organization, Infectivity and Gene Manipulation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 354(1383). doi: 10.1098/rstb.1999.0408.

Oliner, Jonathan D., Kenneth W. Kinzler, and Bert Vogelstein. 1993. "In Vivo Cloning of PCR Products in E.Coli." *Nucleic Acids Research* 21(22). doi: 10.1093/nar/21.22.5192.

Peyret, Hadrien, and George P. Lomonossoff. 2013. "The PEAQ Vector Series: The Easy and Quick Way to Produce Recombinant Proteins in Plants." *Plant Molecular Biology* 83(1–2).

Pitman, T. L., K. G. Posis, T. Tian, C. A. Belanger, Avijit Roy, and B. W. Falk. 2019. "First Report of Watermelon Green Mottle Mosaic Virus in North America." *Plant Disease* 103(12). doi: 10.1094/pdis-02-19-0308-pdn.

Schägger, Hermann. 2006. "Tricine-SDS-PAGE." *Nature Protocols* 1(1). doi: 10.1038/nprot.2006.4.

Sonnhammer, E. L., G. von Heijne, and A. Krogh. 1998. "A Hidden Markov Model for Predicting Transmembrane Helices in Protein Sequences." *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* 6.

Stanley, W. M. 1935. "Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus." *Science* 81(2113).

Tian, T., K. Posis, C. J. Maroon-Lango, V. Mavrodieva, S. Haymes, T. L. Pitman, and B. W. Falk. 2014. "First Report of Cucumber Green Mottle Mosaic Virus on Melon in the United States." *Plant Disease* 98(8). doi: 10.1094/PDIS-02-14-0176-PDN.

Tsugita, A., D. T. Gish, J. Young, H. Fraenkel-Conrat, C. A. Knight, and W. M. Stanley. 1960. "THE COMPLETE AMINO ACID SEQUENCE OF THE PROTEIN OF TOBACCO MOSAIC VIRUS." *Proceedings of the National Academy of Sciences* 46(11). doi: 10.1073/pnas.46.11.1463.