BIOGEOGRAPHY AND GENETIC DIVERSITY OF *CUCUMBER MOSAIC CUCUMOVIRUS* (CMV) AND THE ROLE OF SATELLITE RNA IN SYMPTOM DEVELOPMENT

By

Shahideh Nouri

A dissertation submitted in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

(Plant Pathology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2012

Date of final oral examination: 08/13/2012

The dissertation is approved by the following members of the Final Oral Committee:

Russell L. Groves, Associate Professor, Entomology

Paul G. Ahlquist, Professor, Oncology and Molecular Virology/Plant Pathology

Bryce W. Falk, Professor, Plant Pathology

Amy O. Charkowski, Associate Professor, Plant Pathology

Aurelie Rakotondrafara, Assistant Professor, Plant Pathology

Thomas L. German, Emeritus Professor, Entomology/Plant Pathology

Acknowledgments

"*I shall be telling this with a smile*
Somewhere ages and ages hence:
Two roads diverged in a wood, and I,
I took the one less traveled by,
And that has opened my eyes to the world"

And today is the last day of my journey. The journey that I started five years ago, but its dream took shape when I was a 6 year old girl. At that age, I experienced another world under the microscope of my aunt`s laboratory and when I asked about the name of that strange world, this word came to my ear: "Science".

There were many good and bad moments in my journey shared with many people. My first and deepest gratitude must go to my supervisor, Dr. Russell Groves, not only for all his unfailing advice, support and encouragement, but also for his trust. He always gave me freedom to pursue my ideas in research. He taught me how a supervisor can have the talent to be a supervisor and a friend simultaneously, and to make the lab more environmentally friendly. He taught me to value each and every human being, regardless of gender, race, culture, religion, language and geographic boundary.

Special thanks to my advisory committee members, Dr. Amy Charkowski, Dr. Aurelie Rakotondrafara, Dr. Tom German, Dr. Bryce Falk and Dr. Paul Ahlquist for their helpful suggestions, encouragements and serving as outstanding role models to me embodying down-to-earthness, persistence, hope, open-mindedness and compassion. I owe them my heartfelt appreciation.

It has been my honor to be a member of the Department of Plant Pathology at University of Wisconsin-Madison for these several years. I would like to express my sincere gratitude to all the faculty and staff notably Dr. Murray Clayton, Dr. Tom German, Dr. Paul Ahlquist, Dr. Amy Charkowski, Dr. Nancy Keller, Dr. Dennis Halterman, Dr. John Andrews, Dr. Doug Rouse, Dr. Paul Esker, Dr. Brian Hudelson, Ms. Maya Hayslett and Dr. Caitilyn Allen for teaching me new aspects of Science.

A big appreciation from the bottom of my heart goes to Dr. Carol Groves who laughed when I laughed and cried when I cried. Undoubtedly, I could not finish my journey without her support.

I am so grateful for the opportunity to spend several months at Dr. Bryce Falk`s lab at the University of California-Davis and to do research under his supervision. I really enjoyed my time with colleagues at Falk`s lab.

Appreciation is extended to Walnut Street Greenhouse staff at the University of Wisconsin-Madison for their unlimited assistance.

My friends from different parts of the world were sources of support and joy in these years. The friendships we developed will last a lifetime. Special thanks go to Emily Mueller, Ismael Badillo, Mapi Marquez, Shabnam Dabirshahsahebi, Alejandra Huerta, Ana Cristina Fulladolsa, Anders Huseth, Mahdieh Hosseininasab, Ken Frost, Nasim Jamali, Chen Zhang, Jose Pablo Soto-Arias (JP), Mitra Farnoodian, Erika Yashiro, Lingyun Hao, Saori Amaike, Fanhong Meng, Stemmadenia (Claudio) Passiflora, Maria Gabriela, Fatemeh Mirsharifi, Yu (Monica) Chen and Maryam Mohammadi.

Words cannot express my gratitude to my family. My parents, Minou and Shahrokh. I owe them my whole life. My mother was my first grade teacher who taught me how to read and write but

beyond that, how to live and love. My father who showed me there is no limitation for love. Their love, prayer and sacrifice is my driving force. I wish I could show them how much I love them. My sister and brother, Shadnaz and Shahin. They are two shining stars of my life whose glitter keeps my life bright. My aunts, whose unconditional love was the biggest source of energy of my journey. My lost relatives including my grandparents, uncle and aunts watching my success from above. I wish I could tell you how much I missed you.

And at the end, I should say that everybody pays the price to make her/his dreams come true and the highest price that I paid was leaving my homeland and staying far away for five years. So, I would like to dedicate this work to my Motherland, IRAN.

BIOGEOGRAPHY AND GENETIC DIVERSITY OF *CUCUMBER MOSAIC CUCUMOVIRUS* (CMV) AND THE ROLE OF SATELLITE RNA IN SYMPTOM DEVELOPMENT

Shahideh Nouri

Under the supervision of Associate Professor Russell L. Groves

At the University of Wisconsin-Madison

Snap bean (*Phaseolus vulgaris* L.) and pepper (*Capsicum annuum* L.) crops in the Midwest and Northeast United States have been affected by a complex of plant virus epidemics since 2000. *Cucumber mosaic cucumovirus* (CMV) was consistently recovered from infected plants. The reasons behind these virus outbreaks are not well known. In the current project, the phylogeographic structure and genetic variation of CMV isolates was investigated at the local, state and national levels. The genetic diversity was low in the US and Wisconsin, with the 2b gene possessing the greatest variability. CMV subgroup I was identified as the predominant subgroup in the US and within the state of Wisconsin. Furthermore, both subgroups IA and IB were recognized among the CMV isolates evaluated. In Wisconsin specifically, a host-specific association between genetic subgroups and host plants was observed, such that subgroup IA was the only subgroup obtained from snap bean, whereas the CMV population in pepper was more diverse and consisted of both subgroups. However, no significant association between phylogenetic groups, location or year was observed in the US and Wisconsin. More specifically, CMV subpopulations from different fields within Wisconsin were not genetically differentiated in most instances, while they were different in two hosts. Evidence for genetic exchange through

recombination (6%) and reassortment (10%) was rare in natural populations of CMV in US. Purifying selection was determined as the major natural force on the CMV genome, whereas positively selected sites were detected in the 2a, 2b and CP proteins. In addition, satellite RNA (satRNA) was discovered in CMV field isolates of snap bean for the first time. Sequence comparisons with other available CMV satRNAs illustrated that these are novel satRNAs. Results obtained from a greenhouse bioassay illustrated that this extra RNA exacerbated symptom development and hastened the time of symptom expression in snap bean. In contrast, satRNAs attenuated symptoms in tobacco. In conclusion, our analyses provided a snap shot of the genetic diversity and phylogenetic relationships among representative CMV isolates in the US and Wisconsin. It seems that the recent virus epidemics may not have been due to the invasion of a new CMV subgroup or genotype in snap bean and probably a founder effect has played a role in shaping the CMV structure in this host. The presence of satRNA in the CMV population together with the introduction of a novel vector, the soybean aphid (*Aphis glycines* Matsumura), may be partial explanations for the emergence of these virus outbreaks.

Approved by: Russell L. Groves

13 August 2012

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

Plants provide food to humans and nearly all other non-primary producers, either directly or indirectly. According to the Food and Agriculture Organization (FAO), the world population will rise from 6.8 billion today to 9.1 billion in 2050 - a third more mouths to feed than there are today (FAO 2012). As agriculture struggles to support the rapidly growing global population, plant diseases reduce the production and quality of food. More than 800 million people do not have adequate food and at least 10% of global food production is lost to plant disease (132). Plant pathogens are difficult to control because their populations are variable in time, space, and genotype.

Crops

Snap bean (*Phaseolus vulgaris* L.) and pepper (*Capsicum annuum* L.) are two of the most important, commercially grown vegetables in the United States. According to NASS (Natural Agricultural Statistics Service, 2011) estimates, the US ranks 15th in snap bean production in the world with Wisconsin being the domestic leader in total processing production. In 2011, 176,000 acres of snap bean were planted in US and the crop was valued at $161 M dollars (NASS 2011). According to the FAO, the US ranked sixth in green pepper production in 2009 (http://www.agmrc.org/commodities_products/vegetables/bell_and_chili_peppersprofile.cfm) with more recent estimates in 2011, showing 56,200 and 23,400 acres of bell and chili pepper were planted in US, respectively. The crop was valued at $684 and 146.8 M dollars for bell and chili pepper, respectively. California leads the nation in pepper production (NASS 2011). Processing and fresh pepper production in Wisconsin has been estimated around 1300 acres in 2007 (WI NASS – Census of Agriculture 2007.). However, plant pathogens and insects are major limiting factors for snap bean and pepper production because the quality and size of pods

and fruits are often adversely affected by pest outbreaks and disease. In particular, virus diseases can affect processing crop production dramatically.

Virus complex disease epidemics in the Upper Midwest of the United States

A disease outbreak beginning in 2000, caused by a novel complex of plant viruses, has affected the snap bean processing industry and has limited revenue increases of snap bean in the Midwest and the Northeastern US (New York State) (41, 47, 76, 94). *Cucumber mosaic virus* (CMV), *Alfalfa mosaic virus* (AMV), *Bean common mosaic virus* (BCMV), *Bean yellow mosaic virus* (BYMV) and *Clover yellow mosaic virus* (ClYMV) were viruses documented in affected snap beans and associated with this outbreak. Prior to 2000, all of these viruses were noted to affect processing snap bean, however significant increases in virus incidence were observed in several regions, with a significantly large proportion of infection explained by the prevalence of CMV (41, 43, 47, 76, 125). In New York in 2005, up to 100% of plants in some snap bean fields were infected with CMV (96). Several hypotheses have emerged to explain these virus epidemics. Coincident with the increasing incidence of viruses statewide, the soybean aphid (*Aphis glycines* Matsumura) was first discovered feeding on soybean in Wisconsin in July of 2000. Soybean aphid is native to Southeast Asia and a recent investigation has placed the soybean aphid among the four most efficient vectors of CMV (43). Therefore, one hypothesis which has emerged suggests that CMV and its local vectors have co-existed for a long time without serious disease outbreaks on cultivated plants prior to the introduction of the exotic soybean aphid. Additionally, a significant influence on the occurrence, or emergence, of a novel virus epidemic, may result from changes or variations in the virus population. However, there is currently insufficient molecular-based phylogenetic information to describe bio-geographical differences of CMV isolates in the US. To our knowledge, very few investigations have

demonstrated the genetic structure of the US CMV population or its relationship to disease development. Genetic resistance is by far the most effective way to control plant viruses, and without a comprehensive understanding of the genetic structure of the virus, the durability and success of long-term resistance is in question. A final hypothesis for these recent outbreaks, many of which have occurred around the world, may be due to the presence of an extra RNA molecule (satellite RNA) in the CMV genome (20, 64, 66, 73).

Cucumber mosaic cucumovirus (CMV)

Cucumber mosaic cucumovirus is one of the most economically important plant viruses (102) which can cause severe damage in numerous crop species including cucumbers, melons, squash, peppers, tomatoes, crucifers, lilies and many non-crop weed species. CMV is the type member of the *Cucumovirus* genus in the family *Bromoviridae* (119). CMV was first reported in 1916 as the causal agent of select plant diseases (29). It has the largest host range of any virus, a worldwide distribution and a severe impact on cultivated crops as previously noted (102).

CMV Structure and Genome

Cucumber mosaic cucumovirus is a tripartite, polyhedral virus with a diameter of 29 nm (102). CMV particles are isometric and are composed of a coat protein shell which encapsidates the single-stranded, plus-sense RNA genome. The virions contain 18% RNA and 82% protein. The RNA consists of three genomic RNAs, designated RNA 1 (3.3 kb in length), RNA 2 (3.0 kb), and RNA 3 (2.2 kb), and two subgenomic RNAs, RNA 4 (1.0 kb) and RNA 4A (682 nucleotides), which are transcribed from the 3' portions of RNA 3 and RNA 2, respectively (24,104). The genomic RNAs are packaged in individual particles while the two subgenomic

RNAs are packaged with genomic RNA3. Five proteins designated 1a, 2a, 2b, 3a and 3b are encoded by five open reading frames (ORFs) located on these three RNAs (102).

Encoded proteins by CMV coding regions and their functions

RNA 1 encodes protein 1a (110 KDa) that is necessary for viral replication and contains two functional domains: an N-terminal domain that is a putative methyl-transferase domain involved in the capping of genomic and subgenomic RNAs and a C-terminal proximal domain that is a putative helicase (45, 51, 56, 65).

The 2a protein (98 kDa) is translated directly from RNA 2 containing a number of motifs characteristic of RNA-dependent RNA polymerases (16, 51, 60, 110). The 2a protein interacts with the 1a protein in vivo and in vitro (72) forming the replication complex along with host factors (53). Moreover, CMV 2a sequences, located in two polymerase motifs (D and B), affect systemic infection in bean (*Phaseolus vulgaris* L.) and pea (*Pisum sativum* L.) (69).

RNA 2 also encodes another protein called 2b protein (11-13 KDa) via its subgenomic RNA 4A (24). The 2b ORF overlaps with the C-terminus of the 2a ORF and the 2b protein is considered a multifunctional protein involved in suppression of RNA silencing (14, 23), symptom induction (25) and viral movement (25, 104, 131). The CMV 2b protein has been shown to inhibit the ability of the gene silencing signal to activate gene silencing in distant tissues, as well as interfering with DNA methylation in such tissues (14, 49). The 2b is a nuclear protein (82, 87) and this localization facilitates its ability to inhibit DNA methylation. The CMV 2b protein also inhibits salicylic acid-mediated virus resistance (62).

RNA 3 encodes two proteins, 3a, or movement protein (MP) (28.8 KDa) and 3b, or capsid protein (CP) (24KDa); this latter protein being translated from subgenomic RNA 4 (123). The 3a protein is a nucleic acid-binding protein that is essential for the virus during short and long-

distance movement (11, 17, 135). The 3a protein has been localized to plasmodesmata (9, 141), as well as to large aggregates inside vascular sieve elements (9). The 3b or capsid protein is a structural protein and is also required for virus movement within plants and as well as between plants by aphid vectors (11, 17, 90, 135).

All of the five known genes of CMV are also considered as host range and pathogenicity determinants (104).

CMV Satellite RNA

In addition to the CMV genomic and subgenomic RNAs, some strains of CMV carry an extra RNA molecule known as satellite RNA. The satRNA is a 5`- capped, linear, single-stranded RNA molecule, ranging in size from 332 to 405 nucleotides (38, 127). These small RNAs are dependent on CMV (helper virus) for replication and movement (46). In addition, satRNA can be encapsidated in the CMV virion which allows transmission of satRNA along with CMV by the aphid vectors (46). However, the efficiency of transmission of CMV isolates supporting satRNA is much less than CMV isolates without satRNA, presumably because of the depression of CMV accumulation (33). Although, different ORFs have been found for CMV satRNAs producing protein products in vitro (6, 55, 99), none of them are conserved and translation products have not been found in vivo (116). Also, mutational experiments showed that these ORFs probably are not functional (116). They demonstrate a high secondary structure that is important for their biological properties, sharing little or no sequence homology with CMV (38, 59, 127). This highly structured form could explain stability and infectivity of these molecules (38). In spite of lack of any functional ORF, the presence of this extra RNA can influence symptom expression, ranging from attenuation to increased severity depending on the particular CMV genotype and host plant (20, 38, 44, 64, 66, 73, 127, 137). The presence of satRNA is considered very common

in greenhouse environments, but the frequency of satRNA in field populations of CMV appears to be low (3, 48, 68). Most CMV satRNAs reduce the replication and accumulation of their helper virus, which effectively results in the attenuation of symptoms (8, 50, 58, 67, 78). Attenuated satRNAs can be used as biocontrol agents of their helper virus (CMV), whether expressed as a transgene in the host plants or applied mechanically to the host (38, 116, 139). However, applying attenuated satRNAs in a broad- spectrum manner may increase the risk of disease as there is some evidence suggesting that a benign satRNA can easily mutate to a pathogenic form (103). In some cases, satRNAs that attenuate the symptoms in many hosts can exacerbate symptoms in certain hosts, such as severe systemic chlorosis in tobacco, pepper and tomato (44, 101, 137). The co-occurrence of the satRNA with CMV has been linked to four important historical epidemics in tomato in different parts of the world (20, 64, 66, 73). A specific domain within the 5` half of the molecule inducing chlorosis has been determined (61,75), and specific changes that affected the host specificity have been identified (128, 129, 147). A necrosis-inducing domain has been mapped within the 3` half of the satRNA molecule (21, 75, 86). The pathogenicity mechanism (s) that satRNAs apply to intensify disease symptoms was a big mysterious for many years, but recent studies have brought evidence that RNA silencing mechanisms of the host plants are involved in the pathogenicity of satRNAs (126, 130, 144) and there is a complex competition among host plants, helper viruses and satRNAs in the targeting of the RNA silencing mechanisms. The CMV-satRNA populations in field are reportedly very heterogeneous because of high genetic variability resulting from the accumulation of mutations (4, 112). It has also been shown that CMV satRNA genetic structure and population dynamics are different from the helper virus (CMV) and that satRNA spreads epidemically as a parasite on the population of CMV (3). The spatio-temporal dynamics of the

principal aphid vector(s) and the efficiency of their transmission also play an important role in the invasion of the CMV population by satRNA (34).

Symptoms

Symptoms induced by CMV can vary greatly depending on the host and the age of the plant at the infection time. Most strains induce mosaic (light or dark green) symptoms; however some strains induce other symptoms such as stunting, filiform, epinasty or necrotic lesions (22, 41, 124, 136). Studies to date have shown that all three RNAs of CMV are involved in symptom induction (32, 104). Also, the occurrence of satellite RNA with CMV can further influence the induced symptoms during infection (8, 44, 48, 50, 58, 64, 66, 67, 78, 101, 137). CMV symptoms in affected snap bean often consist of green mottling, leaf curl and blistering. Pod infection and loss is greatest if infection occurs before bloom, because CMV causes flower abortion and abnormal development (63, 143). Infected pods are typically mottled and can be reduced in size (76, 125). In pepper, CMV causes chlorosis in young leaves, whereas oak leaf and ring spot symptoms can be observed in older, infected leaves (138). Pepper fruit may develop ring spots and irregular patches of necrotic tissue affecting the quality of fruit (138).

Variability and Evolution

Studies of host range, symptomatology, immunology, and molecular analysis of the CMV genomic RNAs demonstrate that isolates of this virus are heterogeneous (104).

Genetic diversity and Taxonomy

Early classification divided CMV isolates into two subgroups designated as subgroups I and II based on serology (102, 142), nucleic acid hybridization (100), RT-PCR followed by RFLP (113) and nucleotide sequence identity (102, 118). These two subgroups show approximately

75% nucleotide identity (118). Isolates of subgroup II demonstrate 96% identity, whereas this percentage is estimated at 88% among isolates in subgroup I. This shows that subgroup I is more heterogeneous than subgroup II (104). Analysis of the CP gene and 5` non-translated region (NTR) of the RNA3 has led to further sub-division of subgroup I into subgroups IA and IB, with 92-95% nucleotide identity (117, 118). However, phylogenetic analysis of nucleotide sequences of some CMV strains showed that the estimated trees for various ORFs located on the different RNAs were not congruent and did not completely support the subgrouping indicated by the CP ORF. This indicates that different RNAs may have independent evolutionary histories and reassortment plays an important role in CMV evolution (118). Subgroups IA and II have been documented as worldwide in their distribution (91), while the majority of subgroup IB isolates have principally been restricted to East Asia (118). However, CMV subgroup IB isolates have been reported from Italy (36) and India (10, 83, 132). Recently reported in the US, subgroup IB has now been confirmed from both Hawaii and California (80, 117). There is evidence showing a relationship between some strains of CMV and specific hosts such as legumes (70, 145). Hence, CMV is a heterogenic group with significant variation among isolates documented in different crops and geographic locations, providing the ability to rapidly evolve in unique environments with shifting selection pressures. (39, 74, 80, 81, 115, 118, 134). However, phylogeography studies of CMV natural populations in Spain and California illustrated no correlation of the CMV population structure with geographical region, year, or plant species (35, 80). Thus, a metapopulation structure has been proposed for CMV natural population with local extinction and recolonization by either local or distant isolates (35). It is impossible to predict the type of isolates that will occur in a certain region at a particular time in this type of structure. This limits the use of strategies based on strain-specific resistance.

Variation mechanisms

Several potential mechanisms of CMV variation and evolution have been proposed. Random mutation, recombination and reassortment are most common sources of RNA virus evolution and variability (26, 40, 42, 57, 89).

Mutation

Mutation results in nucleotide sequence differences between the template and daughter strand during virus replication. RNA viruses have high mutation rates because of error-prone RNA replication, large population sizes and short generation times (31); and a consequence of the replication is a swarm of sequence variants around a consensus sequence called a 'quasispecies' (27, 28). Plant RNA viruses show a lower mutation rate than animal RNA viruses (84, 121, 140) and the results of several studies show that plant RNA viruses are highly genetically stable (35, 54, 85, 114) in comparison with animal viruses. This behavior might be due to strong bottlenecks either during colonization of the host plant (52, 79, 120) or during transmission by the arthropod vector (1, 7, 92) or weaker immune-mediated positive selection (40). Although, there is no documented data to illustrate the exact mutation rate of CMV, this plant RNA virus is not an exceptional case (122).

Genetic exchange (Recombination and reassortment)

Genetic exchange by either recombination or by reassortment is another source of genetic variation in CMV populations. Genetic exchange can compensate the fitness losses due to the accumulation of deleterious mutations (94). Recombination is the result of switching of genetic information segments between the nucleotide strands of different genetic variants during the replication. It has been shown that CMV replicase can switch templates in vitro through a copy

choice mechanism, generating recombinant RNAs (71). Recombination in the 5` and 3` nontranslated regions (NTR), between ORFs 3a and 3b (12, 18, 35, 104, 109, 117) and in natural populations containing sat-CMV's (5) have been shown to be sources of variability in the virus population. Recombination may change the biological properties of the virus at the population level, including the appearance of resistance-breaking strains or host range expansion which potentially leads to disease epidemics (77, 88). In natural populations of CMV in Spain, recombinants in the intergenic region of RNA 3, between subgroups IA and IB, were estimated at frequencies of about 0.07. These data, however, indicate that they were selected against and did not become established in CMV populations (35). Reassortment, also called pseudorecombination is another approach by which viruses with a segmented genome can exchange their genomic information. There is direct evidence for the occurrence of reassortment within all three subgroups of CMV (12, 19, 81, 118). However, in natural CMV populations in Spain, reassortants between subgroups IA and IB isolates were rare (4%) and appeared to be selected against (35). Thus, it seems that CMV reassortants or recombinants are not favored under natural conditions. The nature of this disadvantage is not clear, but one hypothesis suggests that in natural populations of CMV, reassortment is selected against during aphid transmission (35, 37).

CMV Epidemiology and Control

Cucumber mosaic cucumovirus is one of the most economically important plant viruses (102) which infects more than 1,000 species including monocots and dicots, as well as herbaceous shrubs and trees in over 70 families of plants (102). Most temperate zones, tropical regions and Mediterranean countries are affected by CMV. CMV has been documented as an endemic virus and the most important virus of some annual crops in Argentina, eastern China, Croatia, France,

Egypt, Greece, Israel, Italy, Japan, Poland, Portugal, Spain, Sweden and in the north east of US. In other countries, CMV ranks second or third in importance (138). Numerous non-crop weed species and crop plants can serve as reservoirs for CMV because of its wide host range (93, 125). These reservoirs can contribute to virus spread to crops at various times throughout the season. CMV can be transmitted through seed in some host species, including weed hosts (2, 15, 30, 97, 98, 102, 146). Seed transmission (7%) of CMV in snap bean (*P. vulgaris*) was reported for the first time with a Spanish isolate in 1974 (13). There is evidence that seed transmission in *P. vulgaris* may be determined by the RNA1, suggesting a relationship between the RNA1-encoded product of the viral replicase complex and seed-transmission (36). CMV can also be acquired and transmitted by more than 80 different aphid vector species in a non-persistent manner (102). During non-persistent transmission, viruses are acquired after short probes (generally less than 60 sec.), retained for short periods of time (a few minutes), and lost due to normal feeding activities. As a non-persistently transmitted virus, CMV does not require a latent period in the vector (111). The rate of transmission can be affected by the virus strain, aphid vector species and by the quality of the host plant (106, 107). Different aphid species can be expected to select different components in virus populations (108). Recent investigations have shown that the four most efficient vectors of CMV include melon/cotton aphid (*Aphis gossypii* Glover), soybean aphid (*Aphis glycines* Matsumura), pea aphid (*Acyrthosiphon pisum* Harris) and yellow clover aphid (*Therioaphis trifolii* Monell) (43). An example of the importance of the arthropod vector in CMV disease development is highlighted by a tomato necrosis epidemic reported in France in the early 1970's. This epidemic was induced by CMV with the co-occurrence of satellite RNA (66). Three similar epidemics in tomato were reported in Spain, Italy and Japan (20, 64, 73), all of which reportedly contained satRNA with a helper (CMV) virus. The origin of these satellite

RNAs was unknown. However, results of a study in Spain suggest that changes in the size and composition of the aphid (vector) populations may have created conditions extremely favorable for CMV transmission, increasing the infection frequency in tomatoes and its spread to epidemic proportions. Since CMV has been very successful in rapidly adapting to new hosts and environments, it is not easy to control. Application of insecticide to control aphid vectors has had minor impact, but it is not a good candidate for a long-term control of virus diseases. Because of the absence of resistance genes in the germplasm of most susceptible crops and CMV's unique epidemiological features, it seems that the only effective, long- term control strategy to limit this virus is through genetic resistance, either conventional or transgenic, or by cross protection with attenuated strains, which often involves the use of attenuated satRNAs. However, fast genetic divergence in the satRNA population, and in the virus itself, may make this approach very challenging.

Objectives of the research

The overall goal of this research was to improve our understanding of the molecular epidemiology of CMV in susceptible crops on a local (state of Wisconsin) and national (United States) scale. The reason(s) behind the recent virus outbreaks in select regions of the US is not well understood. In this dissertation research we investigated the potential for major shifts, or changes in the natural CMV population structure resulting from the introduction of a new genotype or subgroup. The presence of satRNA in CMV field populations was also investigated as a partial reason for the virus epidemics in selected regions. Hence, the following objectives were pursued in this thesis: 1) provide a "snapshot" of the geographic distribution (phylogeography) and the genetic variation of CMV isolates in different locations of the US. Possible genetic variation mechanism (s) is investigated for the representative, US-collected

CMV isolates 2) estimate the genetic variation and phylogenetic relationships of CMV isolates at different geographic locations and two host crops across years in Wisconsin, a region challenged with recent CMV epidemics for over a decade. Genetic differentiation of CMV subpopulations within and among fields, hosts and years is investigated in the state 3) assessment of the presence of satRNA in CMV populations in select regions and the role of this extra RNA molecule in disease symptom development. Taken together, this information will be useful to address possible reasons behind the recent virus outbreaks in select regions of the US and development of more comprehensive control strategies for virus diseases, in particular CMV epidemics.

References

1. Ali, A., Li, H., Schneider, W. L., Sherman, D. J. & Gray, S. (2006). Analysis of genetic bottlenecks during horizontal transmission of *Cucumber mosaic virus*. *J Virol* **80**, 8345–8350.

2. Ali, A. &Kobayashi, M. (2010). Seed transmission of *Cucumber mosaic virus* in pepper. *J Virol methods* **163**, 234–237.

3. Alonso-Prados, J. L., Aranda, M. A., Malpica, J. M., Garcı´a-Arenal, F. & Fraile, A. (1998). Satellite RNA of *cucumber mosaic cucumovirus* spreads epidemically in natural populations of its helper virus. *Phytopathology* **88**, 520–524.

4. Aranda, M.A., Fraile, A. & Garcia-Arenal, F. (1993). Genetic Variability and Evolution of the Satellite RNA of *Cucumber Mosaic Virus* during Natural Epidemics. *J Virol* **67**, 5896-5901.

5. Aranda, M. A., Fraile, A., Dopazo, J., Malpica, J. M. & Garcı´a-Arenal, F. (1997). Contribution of mutation and RNA recombination to the evolution of a plant pathogenic RNA. *J Mol Evol* **44**, 81–88.

6. Avila-Rincon, M. J., Collmer, C. W. & Kaper, J. M. (1986). In vitro translation of cucmoviral satellites. II. CARNA5 from *cucumber mosaic virus* strain S and SP6 transcript of cloned (S) CARNA 5 cDNA produce electrophoretically comigrating protein products. *Virology* **152**, 455-458.

7. Betancourt, M., Fereres, A., Fraile, A. & Garcı´a-Arenal, F. (2008). Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *J Virol* **82**, 12416–12421.

8. Betancourt, M., Fraile, A. & Garcı´a-Arenal, F. (2011). *Cucumber mosaic virus* satellite RNAs that induce similar symptoms in melon plants show large differences in fitness. *J Gen Virol* **92**, 1930–1938.

9. Blackman, M. L., Boevink, P., Santa Cruz, S., Palukaitis, P. & Oparka, K. J. (1998). The movement protein of *cucumber mosaic virus* traffics into sieve elements in minor veins of *Nicotiana clevelandii*. *Plant Cell* **10**, 525–537.

10. Bhat, A. I., Faisal, T. H., Madhubala, R., Hareesh, P. S. & Pant, R. P. (2004). Purification, production of antiserum and development of enzyme linked immunosorbent assay-based diagnosis for *Cucumber mosaic virus* infecting blackpepper (*Piper nigrum* L.). *J Spices Aromatic Crops* **13**, 16‑21.

11. Boccard, F. & Baulcombe, D. (1993). Mutational analysis of cis-acting sequences and gene function in RNA3 of *cucumber mosaic virus*. *Virology* **193**, 563–578.

12. Bonnet, J., Fraile, A., Sacrista´ n, S., Malpica, J. M. & Garcı´a-Arenal, F. (2005). Role of recombination in the evolution of natural populations of *Cucumber mosaic virus*, a tripartite RNA plant virus. *Virology* **332**, 359–368.

13. Bos, L. & Maat, H. Z. (1974). A starin of *Cucumber mosaic virus*, seed-transmitted in beans. *Neth J Plant Path* **80**, 113- 123.

14. Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W. & Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J* **17**, 6739–6746.

15. Brunt, A., Crabtree, K. & Gibbs, A. (1990). Viruses of Tropical Plants: Description and Lists from the VIDE Database. CAB International, Wallingford, UK.

16. Bruenn, J. A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res* **19**, 217–226.

17. Canto, T., Prior, D. A. M., Hellwald, K. H., Oparka, K. J. & Palukaitis, P. (1997). Characterization of *cucumber mosaic virus*. IV. Movement protein and coat protein are both essential for cell-to-cell movement of *cucumber mosaic virus*. *Virology* **237**, 237–248.

18. Chen, Y.-K., Goldback, R. & Prins, M. (2002). Inter- and intramolecular recombination in the *cucumber mosaic virus* genome related to adaptation to alstroemeria. *J Virol* **76**, 4119–4124.

19. Chen, Y., Chen, J., Zhang, H., Tang, X. & Du, Z. 2007. Molecular evidence and sequence analysis of a natural reassortant between *Cucumber mosaic virus* subgroup IA and II strains. *Virus Genes* 35, 405- 413

20. Crescenzi, A., Barbarossa, L., Cillo, F., Di Franco, A., Vovlas, N. & Gallitelli, D. (1993). Role of *cucumber mosaic virus* and its satellite RNA in the etiology of tomato fruit necrosis in Italy. *Arch Virol* **131**, 321–333.

21. Devic, M., Jaegle, M. & Baulcombe, D. (1989). Symptom production on tobacco and tomato is determined by two distinct domains of the satellite RNA of *cucumber mosaic virus* (strain Y). *J Gen Virol* **70**, 2765–2774.

22. Diveki, Z., Salanki, K. & Balazs, E. (2004). The necrotic pathotype of the *cucumber mosaic virus* ns strain is solely determined by amino acid 461 of the 1a protein. *Mol Plant- Microbe Interact* **17**, 837- 845.

23. Diaz-Pendon, J.A., Li, F., Li, W.X. & Ding, S.W. (2007). Suppression of antiviral silencing by *cucumber mosaic virus* 2b protein in *Arabidopsis* is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* 19, 2053-2063.

24. Ding, S. W., Anderson, B. J., Haase, H. R. & Symon, R. H. (1994). New overlapping gene encoded by the *cucumber mosaic virus* genome. *Virology* **198**, 593-601.

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

26. Domingo, E. & Holland, J. I. (1988). High error rates, population equilibrium, and evolution of RNA replication systems. In "RNA Genetics," Vol. Ill, "Variability of RNA Genomes" (E. Domingo, J. J. Holland, and P. Ahlquist, Eds.), pp. 3-36. CRC Press, Boca Raton, FL.

27. Domingo, E., Holland, J., Biebricher, C. & Eigen, M. (1995). Quasispecies: the concept and the word. In Molecular Basis of Virus Evolution, pp. 181-191. Edited by A. J. Gibbs, C. H. K. Calisher and F. Garcia-Arenal. Cambridge: Cambridge University Press.

28. Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151–178.

29. Doolittle, S.P. (1916). A new infectious mosaic disease of cucumber. *Phytopathology* **6**, 145–147.

30. Douine, L., Quite, J. B., Marchoux, G. & Archange, P. (1979). Index of plants susceptible to *cucumber mosaic virus* bibliographical study. *Ann Rev Phytopathol* **11**, 439–475.

31. Drake, J. W. & Holland, J. J. (1999). Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A* **96**, 13910–13913.

32. Du, Z., Chen, F., Zhao, Z., Liao, Q., Palukaitis, P. & Chen, J. (2008). The 2b protein and the C- terminal of the 2a protein of *Cucumber Mosaic Virus* subgroup I strains both play a role in viral RNA accumulation and induction of symptoms. *Virology* **380**, 363- 370.

33. Escriu, F., Fraile, A. & Garcı´a-Arenal, F. (2000a). Evolution of virulence in natural populations of the satellite RNA of *cucumber mosaic virus*. *Phytopathology* **90**, 480–485.

34. Escriu, F., Fraile, A. & Garcı´a-Arenal, F. (2003). The evolution of virulence in a plant virus. *Evolution* **57**, 755–765.

35. Fraile, A., Alonso-Prados, J.L., Aranda, M.A., Bernal, J.J., Malpica, J. & Garcı´a-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *J Virol* **71**, 934–940.

36. Gallitelli, D. (2000). The ecology of *Cucumber mosaic virus* and sustainable agriculture. Virus Res **71**, 9-21.

37. Gallitelli, D., Grieco, F. & Cillo, F. (1997). The potential of a beneficial satellite RNA of *cucumber mosaic virus* to acquire deleterious functions: nature versus greenhouses. In: Balazs, E., Tepfer, M. (Eds.), Virus-Resistant Transgenic Plants: Potential Ecological Impact. Springer-Verlag, Berlin, pp. 100–106.

38. Garcı´a-Arenal, F. & Palukaitis, P. (1999). Structure and functional relationships of satellite RNAs of *cucumber mosaic virus*. *Curr Top Microbiol Immunol* **239**, 37–63.

39. Garcia-Arenal, F., Escriu, F., Aranda, M. A., Alonso-Prados, J. L., Malpica, J. M. & Fraile A. (2000). Molecular epidemiology of *cucmber mosaic virus* and its satellite RNA. *Virus Res* **71**, 1-8.

40. Garcı´a-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* **39**, 157–186.

41. German, T. L., Thompson, A. & Wills, D. K. (2004). Statewide distribution of virus problems on processing beans. In. Proc 2004 Midwest Food Processors Association Processing crops Conference. Feb 23-24, La Cross, WI Department of Horticulture, University of Wisconsin.

42. Gibbs, A., Calisher, C. H. & Garcia-Arenal, F. (1995). Molecular Basis of Viral Evolution. Cambridge University Press, New York, NY

43. Gildow, F. E., Shah, D. A., Sackett, W. M., Butzler, T., Nault, B. A. & Fleischer, S. J. (2008). Transmission efficiency of *Cucumber mosaic virus* by aphids associated with virus epidemics in snap bean. *Phytopathology* **98**, 1233-1241.

44. Gonsalves, D., Provvidenti, R. & Edwards, M. C. (1982). Tomato white leaf: the relation of an apparent satellite RNA and *cucumber mosaic virus*. *Phytopathology* **72**, 1533–38.

45. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1988). A novel super family of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Lett. 235:16–24.

46. Gorden, K.H.J. & Symons, R.H. (1983). Satellite RNA of *cucumber mosaic virus* forms a secondary structure with partial 3`- terminal homologyto genomal RNAs. *Nucleic Acids Res* **11**, 947-960.

47. Grau, C. R., Stevenson, W. R. & Mondjana, A. M. (2002). Viruses causing losses on processing beans in the Midwest, In. Proc 2002 Wisconsin fertilizer, Aglime and Pest Management Conference, 41: 248-256.

48. Grieco, F., Lanave, C. & Gallitelli, D. (1997). Evolutionary dynamics of *cucumber mosaic virus* satellite RNA during natural epidemics in Italy. *Virology* **229**, 166–174.

49. Guo, H. S. & Ding, S. W. (2002). A viral protein inhibits the long range signaling activity of the gene-silencing signal. *EMBO J* **21**, 398–407.

50. Habili, N. & Kaper, J. M. (1981). *Cucumber mosaic virus*-associated RNA 5. VII. Double-stranded form accumulation and disease attenuation in tobacco. *Virology* **112**, 250–261.

51. Habili, N. & Symons, R. H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res* **17**, 9543–9555.

52. Hall, J. S., French, R., Hein, G. L., Morris, J. & Stenger, D. C. (2001). Three distinct mechanisms facilitate genetic isolation of sympatric *Wheat streak mosaic virus* lineages. *Virology* **282**, 230– 236.

53. Hayes, R. J. & Buck, K. W. (1990). Complete replication of a eukaryotic virus RNAin vitro by a purified RNA-dependent RNA polymerase. *Cell* **63**, 363–368.

54. Herra´nz, M. C., Al Rwahnih, M., Sa´nchez-Navarro, J. A., Elena, S. F. & Choueiri, E. (2008). Low genetic variability in the coat and movement proteins of *American plum line pattern virus* isolates from different geographic origins. *Arch Virol* **153**, 367– 373.

55. Hidaka, s., Hanada, K. & Ishikawa, K. (1990). In vitro messenger properties of a satellite RNA of *cucumber mosaic virus*. *J Gen Virol* **71**, 439- 442.

56. Hodgman, T. C. (1988). A new superfamily of replicative proteins. *Nature* **333**, 22–23.

57. Holland, J. J., De La Torre, J. C. & Steinhauer, D. A. (1992). RNA virus population as quasispecies. *Curr Top Microbiol Immunol* **176**, 1- 20.

58. Hou, W., Duan, C., Fang, R., Zhou, X. & Guo, H. (2011). Satellite RNA reduces expression of the 2b suppressor protein resulting in the attenuation of symptoms caused by *Cucumber mosaic virus* infection. *Mol Plant Path* DOI: 10.1111/J.1364-3703.2010.00696.X

59. Hu, C. C., Hsu, Y. H. & Lin, N. S. (2009). Satellite RNAs and Satellite Viruses of Plants. *Viruses* 1, 1325-1350.

60. Ishihama, A. & Barbier, P. (1994). Molecular anatomy of viral RNA-directed RNA polymerase. *Arch Virol* **134**, 235-258.

61. Jaegle, M., Devis, M., Longstaff, M. & Baulcombe, D. (1990). *Cucumber mosaic virus* satellite RNA (Y strain): analysis of sequences with affect yellow mosaic symptoms on tobacco. *J Gen Virol* **71**, 1905- 1912.

62. Ji, L.H. & Ding, S.W. (2001). The suppressor of transgene RNA silencing encoded by *cucumber mosaic virus* interferes with salicylic acid-mediated virus resistance. *Mol Plant-Microbe Interact* **14**, 715–724.

63. Jones, R.A.C., Coutts, B. A., Latham, L. J. & McKirdy, S. J. (2008). *Cucumber mosaic virus* infection of chickpea stands: temporal and spatial patterns of spread and yield limiting potential. *Plant Pathol* **57**, 842-853.

64. Jord´a, C., Alfaro, A., Aranda, M. A., Moriones, E. & Garc´ıa-Arenal, F. (1992). Epidemic of *cucumber mosaic virus* plus satelliteRNA in tomatoes in eastern Spain. Plant Dis **76**, 363–66.

65. Kadare´, G. & Haenni, A. L. (1997). Virus-encoded RNA helicases. *J Virol* **71**, 2583–2590.

66. Kaper, J. M. & Waterworth, H. E. (1977). *Cucumber mosaic virus*- associated RNA 5: Causal agent for tomato necrosis. *Science* **196**, 429- 431.

67. Kaper, J. M. (1982). Rapid synthesis of double-stranded *cucumber mosaic virus*-associated RNA 5: Mechanism controlling viral pathogenesis? *Biochem Biophys Res Commun* **105**, 1014–1022.

68. Kearney, C. M., Zitter, T. A. & Gonsalves, D. (1990). A field survey for serogroups and the satellite RNA of *cucumber mosaic virus*. *Phytopathology* **80**, 1238–1243.

69. Kim, C. H. & Palukaitis, P. (1993). Genomic mapping of B-CMV RNA 2 to determine the domain responsible for the infection of cowpea. *Phytopathology* **83**, 1349–1350.

70. Kim, C. H. & Palukaitis, P. (1997). The plant defense response to *cucumber mosaic virus* in cowpea is elicited by the viral polymerase gene and affects virus accumulation in single cells. EMBO J **16**, 4060–4068.

71. Kim, M. J. & Kao, C. (2001). Factors regulating template switch in vitro by viral RNA dependent RNA polymerases: Implication for RNA-RNA recombination. *Proc Natl Acad Sci USA* **98**, 4972–4977.

72. Kim, S. H., Palukaitis, P. & Park, Y. I. (2002). Phosphorylation of *cucumber mosaic virus* RNA polymerase 2a protein inhibits formation of replicase complex. EMBO J **21**, 2292–2300.

73. Kosaka, Y., Hanada, K., Fukunishi, T. & Tochihara, H. (1989). *Cucumber mosaic virus* tomato necrotic disease in Kyoto prefecture. *Ann Phytopathol Soc Jpn* **55**, 229- 232.

74. Koundal, V., Mohd, Q., Haq, R. & Praveen, S. (2011). Characterization, genetic diversity, and evolutionary link of *Cucumber mosaic virus* strain New Delhi from India. *Biochem Genet* **49**, 25- 38.

75. Kurath, G. & Palukaitis, P. (1989). Satellite RNAs of *cucumber mosaic virus*: Recombinants constructed in vitro reveal independent functional domains for chlorosis and necrosis in tomato. *Mol Plant-Microbe Interact* **2**, 91–96.

76. Larsen, R. C., Miklas, P. N., Eastwell, K. C., Grau, C. R. & Mondjana, A. (2002). A virus disease complex devastating late season snap bean production in the Midwest. *Annu Rep Bean Improv Coop* **45**, 36- 37.

77. Legg, J.P & Thresh, J.M. (2000). *Cassava mosaic virus* disease in East Africa: a dynamic disease in a changing environment. *Virus Res* **71**, 135–49.

78. Liao, Q., Zhu, L., Du, Z., Zeng, R., Feng, J. & Chen, J. (2007). Satellite RNA-mediated Reduction of *Cucumber Mosaic Virus* Genomic RNAs Accumulation in *Nicotiana tabacum.* *Acta Biochim Biophys Sin* **39**, 217–223.

79. Li, H. & Roossinck, M. J. (2004). Genetic bottlenecks reduce population variation in an experimental RNA virus population. J Virol **78**, 10582–10587.

80. Lin, H. X., Rubio, L., Smythe, A., Jiminez, M. & Falk, B. W. (2003). Genetic diversity and biological variation among California isolates of *Cucumber mosaic virus*. J Gen Virol **84**, 249–258.

81. Lin, H.X., Rubio, L., Smythe A. & Falk, B.W. (2004). Molecular population genetics of *cucumber mosaic virus* in California: Evidence for founder effects and reassortments. *J Virol* **78**, 6666- 6675.

82. Lucy, A. P., Guo, H.-S., Li, W. X. & Ding, S. W. (2000). Suppression of post-transcriptional gene silencing by a plant viral protein localized in the nucleus. *EMBO J* **19**, 1672–1680.

83. Madhubala, R., Bhadramurthy, V., Bhat, A. I., Hareesh, P. S., Retheesh, S. T. & Bhai, R. S. (2005). Occurrence of *Cucumber mosaic virus* on vanilla (*Vanilla planifolia* Andrews) in India. *J Biosci* **30**, 339‑350.

84. Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I. & Drake, J. W. (2002). The rate and character of spontaneous mutation in an RNA virus. *Genetics* **162**, 1505–1511.

85. Marco, C. F. & Aranda, M. A. (2005). Genetic diversity of a natural population of Cucurbit yellow stunting disorder virus. *J Gen Virol* **86**, 815–822.

86. Masuta, C. & Takanami, Y. (1989). Determination of sequence and structural requirements for pathogenicity of a *cucumber mosaic virus* satellite RNA (Y-satellite RNA). *Plant Cell* **1**, 1165–1173.

87. Mayers, C. N., Palukaitis, P. & Carr, J. P. (2000). Subcellular distribution analysis of the cucumber mosaic virus 2b protein. *J Gen Virol* **81**, 219–226.

88. Monci, F., Sa´nchez-Campos, S., Navas-Castillo, J. & Moriones, E. (2002). A natural recombinant between the geminiviruses *tomato yellow leaf curl* Sardinia virus and *tomato yellow leaf curl* virus exhibits a novel pathogenic phenotype and is becoming prevalent in Spanish populations. *Virology* **303**, 317–326.

89. Morse, S. S. (1994). The Evolutionary Biology of Viruses. Raven Press, New York, NY.

90. Mossop, D. W. & Francki, R. I. B. (1977). Association of RNA 3 with aphid transmission of *cucumber mosaic virus*. *Virology* **81**, 177–181.

91. Moury, B. (2004). Differential selection of genes of *cucumber mosaic virus* subgroups. *Mol Viol Evol* **21**, 1602-1611.

92. Moury, B., Fabre, F. & Senoussi, R. (2007). Estimation of the number of virus particles transmitted by an insect vector. *Proc Natl Acad Sci U S A* **104**, 17891–17896.

93. Mueller, E. E., Groves, R. L. & Gratton, C. (2012). Crop and non-crop plants as potential reservoir hosts of *Alfalfa mosaic virus* and *Cucumber mosaic virus* for spread to commercial snap bean. *Plant Dis* **96**, 506-514.

94. Mu¨ller, H. L. (1964). The relation of recombination to mutational advance. Mutat Res **1**, 2–9.

95. Nault, B. A., Shah, D. A., Dillard, H. R. & McFaul, A. R. (2004). Seasonal and spatial dynamics of alate aphid dispersal in snap bean fields in proximity to alfalfa and implication for virus management. Environ Entomol **33**, 1593- 1601.

96. Nault, B. A., Shah, D. A. & Taylor, A. G. (2006). Viruses and aphids everywhere in New York snap bean fields in 2005, pp. 74-76. *In* Proceedings of the 2006 Empire State Fruit and Vegetable EXPO. February 13-16, 2006. Cornell Coop. Exten., Syracuse, NY.

97. Neegaard, P. (1977). Survey of seed-borne viruses. In: Seed Pathology. John Wiley & Sons, New York, pp. 80–116.

98. O'Keefe, D. C., Berryman, D. I., Coutts, B. A. & Jones, R. A. C. (2007). Lack of seed coat contamination with *Cucumber mosaic virus* in lupin permits reliable, large-scale detection of seed transmission in seed samples. *Plant Dis* **91**, 504–508.

99. Owens, R. A. & Kaper, J. M. (1977). *Cucumber mosaic virus* – associated RNA 5. II. In vitro translation in a wheat germ protein- synthesis system. *Virology* **80**, 196- 203.

100. Owen, J. & Palukaitis, P. (1988). Characterization of *cucumber mosaic virus*. I. Molecular heterogeneity mapping of RNA 3 in eight CMV strains. *Virology* **166**, 495–502.

101. Palukaitis, P. (1988). Pathogenicity regulation by satellite RNAs of *cucumber mosaic virus*: minor nucleotide sequence changes alter host responses. *Mol Plant Microb Interact* **1**, 175- 181.

102. Palukaitis, P., Roossinck, M.J., Dietzgen, R.G. & Francki, R.I.B. (1992). *Cucumber mosaic virus*. *Adv Virus Res* **41**, 281–348.

103. Palukaitis, P. & Roossinck, M. J. (1996). Spontaneous change of a benign satellite RNA of *cucumber mosaic virus* to a pathogenic variant. *Nature Biotech* **14**, 1264–1268.

104. Palukaitis, P. & Garcia-Arenal, F. (2003). *Cucumoviruses*. *Adv Virus Res* **62**, 241-323.

105. Peden, K. W. C. & Symons, R. H. (1973). *Cucumber mosaic virus* contains a functionally divided genome. *Virology* **53**, 487–492.

106. Perry, K.L., Zhang, L., Shintaku, M.H. & Palukaitis, P. (1994). Mapping determinants in *cucumber mosaic virus* for transmission by *Aphis gossypii*. *Virology* **205**, 591–595.

107. Perry, K.L., Zhang, L. & Palukaitis, P. (1995). Differential transmission of *cucumber mosaic virus* by two aphids: mutation in the coat protein restore transmission by *Aphis gossypii* but not by *Myzus persicae*. *Phytopathology* **85**, 1143.

108. Perry, K. L., L. Zhang & Palukaitis, P. (1998). Amino acid changes in the coat protein of *cucumber mosaic virus* differentially affect transmission by the aphids *Myzus persicae* and *Aphis gossypii*. *Virology* **242**, 204–210.

109. Pierrugues, O., Guilbaud, L., Fernandes-Delmond, I., Faber, F., Tepfer, M. & Jacquemond, M. (2007). Biological properties and relative fitness of inter-subgroup *cucumber mosaic virus* RNA 3 recombinants produced in Vitro. *J Gen Virol* **88**, 2852-2861.

110. Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* **8**, 3867–3874.

111. Raccah, B. & Fereres, A. (2009). Plant Virus Transmission by Insects. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.A0021525.a0000760.pub2

112. Richards, K. E., Jonard, G., Jacquemond, M. & Lot, H. (1978). Nucleotide sequence of *cucumber mosaic virus* associated RNA 5. *Virology* **89**, 395–408.

113. Rizos, H., Gunn, L. V., Pares, R. D. & Gillings, M. R. (1992). Differentiation of *cucumber mosaic viru*s isolates using the polymerase chain reaction. *J Gen Virol* **73**, 2099–2103.

114. Rodrı´guez-Cerezo, E., Elena, S. F., Moya, A. & Garcı´a-Arenal, F. (1991). High genetic stability in natural populations of the plant RNA virus *Tobacco mild green mosaic virus*. *J Mol Evol* **32**, 328–332.

115. Rodriguez- Alvarado, G., Kurath, G. & Dodds, J. A. (1995). Heterogenity in pepper isolates of cucumber mosaic virus. *Plant Dis* **79**, 450- 455.

116. Roossinck, M. J., Sleat, D. & Palukaitis, P. (1992). Satellite RNAs of plant viruses: Structures and biological effects. *Microbiol Rev* **56**, 265–279

117. Roossinck, M. J., Zhang, L. & Hellward, K. H. (1999). Rearrangements in the 5` nontranslated region and phylogenetic analyses of *cucumber mosaic virus* RNA 3 indicate radial evolution of three subgroups. *J Virol* **73**, 6752-6758.

118. Roossinck, M. J. (2002). Evolutionary history of *cucumber mosaic virus* deduced by phylogenetic analyses. *J Virol* **76**, 3382–3387.

119. Rybicki, E. P. (1995). The *Bromoviridae*. In *Virus Taxonomy*. *Sixth Report of the International Committee on Taxonomy of Viruses*, pp. 450-457. Edited by Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., and Summers, M. D.Vienna & New York: Springer-Verlag. divided genome. *Virology* **53**, 487–492.

120. Sacrista´n, S., Malpica, J. M., Fraile, A. & Garcı´a-Arenal, F. (2003). Estimation of population bottleneck during systemic movement of *Tobacco mosaic virus* in tobacco plants. *J Virol* **77**, 9906–9911.

121. Sanjua´n, R., Agudelo-Romero, P. & Elena, S. F. (2009). Upper limit mutation rate estimation for a plant RNA virus. *Biol Lett* **5**, 394–396.

122. Schneider, W. L. & Roossinck, M. J. (2000). Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. *J Virol* **74**, 3130–3134.

123. Schwinghamer, M. W. & Symons, R. H. (1977). Translation of the four major RNA species of *cucumber mosaic virus* in plant and animal cell-free systems and in toad oocytes. *Virology* **79**, 88–108.

124. Sclavounos, A.P., Voloudakis, A. E., Arabatzis, C. & Kyriakopoulou, P. E. A. (2006). Severe Hellenic CMV Tomato Isolate: Symptom Variability in Tobacco, Characterization and Discrimination of Variants. *Eur J Plant Path* **115**, 163-172.

125. Shah, D. A., Dillard, H. R., Mazumdar-Leighton, S., Gonsalves, D. & Nault, B. (2006). Incidence, spatial patterns, and associations among viruses in snap bean and alfalfa in New York. *Plant Dis* 90, 203-210.

126. Shimura, H., Pantaleo, V., Ishihara, T., Myojo, N. & Inaba, J. I. (2011). A Viral Satellite RNA Induces Yellow Symptoms on Tobacco by Targeting a Gene Involved in Chlorophyll Biosynthesis using the RNA Silencing Machinery. *PLoS Pathog* **7**: e1002021. doi:10.1371/journal.ppat.1002021

127. Simon, A. E., Roossinck, M. J. & Havelda, Z. (2004). Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu Rev Phytopathol* **42**, 415– 437

128. Sleat, D. E. & Palukaitis, P. (1992). A single nucleotide change within a plant virus satellite RNA alters the host specificity of disease induction. *Plant J* **2**, 43–49.

129. Sleat, D. E., Zhang, L. & Palukaitis, P. (1994). Mapping determinants within *cucumber mosaic virus* and its satellite RNA for the induction of necrosis in tomato plants. *Mol Plant-Microbe Interact* **7**, 189–195.

130. Smith, N. A, Eamens, A. L. & Wang, M. B. (2011). Viral Small Interfering RNAs Target Host Genes to Mediate Disease Symptoms in Plants. PLoS Pathog **7**, e1002022. doi:10.1371/journal.ppat.1002022

131. Soards, A. J., Murphy, A. M., Palukaitis, P. & Carr, J. P. (2002). Virulence and differential local and systemic spread of *cucumber mosaic virus* in tobacco are affected by the CMV 2b protein. *Mol Plant-Microbe Interact* **15**, 647–653.

132. Srivastava, A., Chandra, G. & Raj, S. K. (2004). Coat protein and movement protein gene based molecular characterization of *Amaranthus* strain of *Cucumber mosaic virus*. *Acta Virol* **48**, 229‑239

133. Strange, R.N. & Scott, P.R. (2005). PLANT DISEASE: A Threat to Global Food Security. *Annu Rev Phytopathol* **43**, 83–116

134. Sullistyowati, E., Mitter, N., Bastiaan- Net, S., Roossinck, M. J. & Dietzgen, R. G. (2004). Host range, symptom expression and RNA 3 sequence analyses of six Australian strains of *Cucumber mosaic virus*. *Aust Plant Pathol* **33**, 505-512.

135. Suzuki, M., Kuwata, S., Kataoka, J., Masuta, C., Nitta, N. & Takanami, Y. (1991). Functional analysis of deletion mutants of *cucumber mosaic virus* RNA3 using an in vitro transcription system. *Virology* **183**, 106–113.

136. Szilassy, D., Sala´nki, K. & Bala´zs, E. (1999). Stunting induced by *cucumber mosaic cucumovirus*-infected *Nicotiana glutinosa* is determined by a single amino acid residue in the coat protein. *Mol Plant-Microbe Interact* **12**, 1105–1113.

137. Takanami, Y. A. (1981). A striking change in symptoms on *cucumber mosaic virus*-infected tobacco plants induced by a satellite RNA. Virology **109**, 120-126.

138. Tomlinson, J. A. (1987). Epidemiology and control of virus diseases of vegetables. *Ann Appl Biol* 110, 661– 681.

139. Tousignant, M. E. & Kaper, J. M. (1998). Control of *cucumber mosaic virus* using viral satellites. In ‘‘Plant-Microbe Interactions and Biological Control’’ (G. J. Boland and L. D. Kuykendall, eds.), pp. 283–296. Dekker, New York.

140. Tromas, N. & Elena, S. F. (2010). The Rate and Spectrum of Spontaneous Mutations in a Plant RNA Virus. *Genetics* **185**, 983-989.

141. Vaquero, C., Sanz, A. I., Serra, M. T. & Garcı´a-Luque, I. (1996). Accumulation kinetics of CMV RNA 3-encoded proteins and subcellular localization of the 3a protein in infected and transgenic tobacco plants. *Arch Virol* **141**, 987–999.

142. Wahyuni, W. S., Dietzgen, R. G., Hanada, K. & Francki, R. I. B. (1992). Serological and biological variation between and within subgroup I and II strains of *cucumber mosaic virus*. *Plant Pathol* **41**, 282-297.

143. Walkey, D. (1991). Applied plant virology, 2nd ed. Chapman & Hall, New York.

144. Wang, M., Bian, X., Wu, L., Liu, L., Smith, N., Isenegger, D., Wu, R., Masuta, C., Vance, V., Watson, J., Rezaian, A., Dennis, E. & Waterhouse, P. (2004). On the role of RNA silencing in the pathogenicity and evolution of viroids and viral satellites. *Proc Natl Acad Sci U S A* **101**, 3275-3280.

145. Whipple, O. C. & Walker, J. C. (1941). Strains of *cucumber mosaic virus* pathogenic on bean and pea. *J Agric Res* **62**, 0027-0060.

146. Yang, Y., Kim, K. S. & Anderson, E. J. (1997). Seed transmission of *Cucumber mosaic virus* in spinach. *Phytopathology* **87**, 924–931.

147. Zhang, L., Kim, C. H. & Palukaitis, P. (1994b). The chlorosis-induction domain of the satellite RNA of *cucumber mosaic virus*: Identifying sequences that affect accumulation and the degree of chlorosis. *Mol Plant-Microbe Interact* **7**, 208–213.

Chapter 2

PHYLOGEOGRAPHY AND GENETIC DIVERSITY OF *CUCUMBER MOSAIC CUCUMOVIRUS* (CMV) ISOLATES IN THE UNITED STATES*

*A slightly modified version of this chapter will be submitted for publication to *PLoS ONE* as:

Nouri S, Falk B-W, Groves R-L. (2012). Phylogeography and Genetic diversity of *Cucumber mosaic cucumovirus* (CMV) in the United States.

Abstract

Cucumber mosaic cucumovirus (CMV) has a worldwide distribution and the widest host range of any known plant virus. During the 2000s, epidemics of CMV severely affected the production of important processing crops such as snap bean and pepper in Midwest and Northeastern United States. Virus diversity leading to emergence of new strains is often considered a significant factor in the appearance of virus epidemics. In addition to epidemics, new disease phenotypes arising from genetic exchanges or mutation can compromise the effectiveness of plant disease management strategies. Here, we investigated the phylogeographic structure and genetic variation of CMV isolates collected from different regions of US. Nucleotide diversity (π) was low for US CMV isolates. Phylogenetic analysis based on neighbor joining (NJ) trees did not show a significant association between phylogenetic groups and environmental factors such as geographical origin, collection year, or plant host species. Sequence and phylogenetic analyses revealed that subgroup I is predominate in the US and further suggests that the CMV population is a mixture of subgroups IA and IB. Furthermore, phylogenetic analysis suggests the likely reassortment between subgroups IA and IB. Based on phylogenetic and computational analysis, recombination between subgroups I and II was detected in RNA3. Neutrality tests illustrated that negative selection was the major force operating upon the CMV genome, although some positively- selected sites were detected in the 2a, 2b and CP proteins. Together, this suggests that different regions of the CMV genome are under different evolutionary constraints. Taken together, these results delineate the CMV strain composition in the US, suggest that recombination and reassortment among strain groups occur but at a low frequency, and point towards CMV genomic regions that differ in types of selection pressure.

Introduction

Cucumber mosaic cucomovirus (CMV) is one of the most common plant viruses causing disease in various crop and non-crop plants in the United States and worldwide. *Cucumber mosaic virus*, the type species of the genus *Cucumovirus* in the family *Bromoviridae*, is one of the most widespread plant viruses. The host range of CMV includes over 1,000 plant species comprised of both crop and non-crop weed species (68, 69) and is transmitted by over 75 species of aphids in a non-persistent manner (69). The genome of CMV contains three, positive-sense, single-stranded RNAs packaged in separated particles and two subgenomic RNAs (15, 69, 70). RNA 1 encodes the 1a protein which contains methyl transferase and helicase motifs (28, 31, 34, 39). RNA 2 encodes two proteins, 2a and 2b. The 2a protein encodes RNA-dependent-RNA polymerase (RdRp) (6, 31, 37, 75) which, together with the 1a protein, forms the viral components of the replicase complex (32). The 2b protein is a multifunctional protein and functions in host-specific, long-distance movement, symptom induction, and as a virulence determinant by suppressing gene silencing (7, 14, 15, 16). Moreover, a recent study has demonstrated that the 2b gene determines the selection of inter-viral recombination (89). The CMV RNA 3 encodes two proteins, 3a, a cell-to-cell movement protein (MP) (16) and 3b or capsid protein (CP); this latter protein being translated from a sub-genomic RNA 4 (87) which is also involved in cell-to-cell movement, virus assembly and aphid-mediated transmission (3, 8, 61, 72, 73, 94).

CMV strains have been classified into two main subgroups designated as subgroups I and II based on serology (69, 100), nucleic acid hybridization (66), RT-PCR followed by RFLP (78) and nucleotide sequence identity (69, 81). These two subgroups have showed 75% nucleotide identity (81). Subgroup I is more heterogeneous than subgroup II (68). Further analysis of the CP gene and 5` non-translated region (NTR) of RNA3 has led to division of subgroup I into IA and

IB with 92-95% nucleotide identity between these two subgroups (81, 82). Phylogenetic analysis of some CMV strains showed that the estimated trees for various Open Reading Frames (ORFs) located on the different RNAs are not congruent and do not completely support the subgrouping from CP ORF analysis. This indicates that different RNAs may have independent evolutionary histories (81). The subgroups are not evenly distributed across agricultural regions. Subgroups IA and II have a worldwide distribution (59), while subgroup IB is principally restricted to Asia (81). Subgroup IB was recently reported in both Hawaii and California (51, 52, 82). CMV is a helper virus for a linear, single-stranded satellite RNA (satRNA) with 332-405 nucleotide residues that can modulate, attenuate, or exacerbate the symptoms induced by CMV depending on the particular CMV strain and host (68, 69, 82). RNA viruses undergo rapid genetic variation and random mutation, recombination and reassortment are the most common sources of RNA virus evolution and variability (17, 23, 25, 35, 58). Reassortment among RNAs within all three subgroups of CMV has been reported (5, 10, 51, 81). Moreover, recombination in the 5` and 3` non-translated regions (NTR), between ORFs 3a and 3b (5, 11, 20, 68, 74, 82) and in natural populations containing satRNA (2), have been shown to be additional sources of variability in the virus population. Hence, CMV is a heterogenic group with significant variation among isolates documented in different crops providing the ability to rapidly evolve in unique environments with shifting selection pressures. (22, 44, 51, 52, 80, 81, 93). Studies focusing on this genetic diversity and sources of variation in the viral populations are important to better understand the evolutionary mechanisms that generate variation.

CMV has been endemic in many parts of the United States for decades. However, during the 2000s, a series of the virus epidemics emerged that affected several processing crops including succulent snap bean (*Phaseolus vulgaris* L.) and pepper (*Capsicum annuum* L.) in agriculturally

important regions of the upper Midwest and Northeast US (24, 29, 47, 62). A significant influence on the appearance of virus epidemics is variation in virus population. However, there is currently insufficient molecular-based phylogenetic information to describe bio-geographical differences among CMV isolates in the United States. In the current study, we have attempted to characterize the geographic variation of CMV isolates collected among locations on a national scale and make comparisons among these US isolates with reference or historical isolates. We determined sources of genetic variation acting upon these isolates and discuss potential evolutionary mechanisms acting upon domestic CMV isolates included in the investigation. In a practical context, this information can be very useful towards the development of more comprehensive control strategies for virus diseases that attempt to integrate plant resistance.

Materials and Methods

Virus isolates and propagation

CMV isolates were collected from a range of naturally infected host plants as either dried or fresh leaf tissue, and from a range of geographic locations throughout the US representing several collection years. Each of the geographic locations was considered as a representative of a unique agriculture region of US. Five CMV isolates (PV544NJU04, PV30MDH85, PV59AZSU84, PV243AZM77 and PV29WIC76) were originally obtained from ATCC (Manassa, VA, USA) as infected plant tissue. Prior to characterization, all infected leaf tissues were first mechanically inoculated into both *Nicotiana benthamiana* and small sugar pumpkin (*Cucurbita pepo* L. cv. 'Small Sugar'). The leaves were ground in 0.1 M potassium phosphate buffer (0.03 M potassium phosphate, pH 7.4, containing 0.05 % Na2SO3) and leaves were mechanically sap-inoculated. Inoculated plants were maintained in an insect-proof greenhouse at

21–25 °C and 16:8 (L:D) photoperiod for 4 weeks post-inoculation. Leaves from symptomatic plants were then vacuum-dried and stored at -20 °C prior to analysis.

RNA extraction and RT- PCR amplification

Total RNA extractions were performed on all samples using an RNeasy Mini kit (Qiagen, CA, USA) according to the manufacturer's instructions. The first strand cDNAs were synthesized in a 20 µl volume of 1X Superscript III reaction buffer (Invitrogen, Carlsbad, USA), containing 0.5 mM dNTP mix, 5 mM DTT, 40 U RNaseOut, 200 U of SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, USA), and 20 pmol of a specific reverse primer for each region (51). The first strand cDNAs were synthesized at 53˚C for 45 minutes. Subsequent PCR reactions were conducted in 25 µl volume of 1X Go Taq Flexi DNA Polymerase reaction buffer (Promega, WI, USA), containing 0.2 mM dNTP mix, 1.1 mM MgCl2, 0.75 U of Go Taq Flexi DNA Polymerase (Promega, WI, USA) and 12.5 mM of each forward and reverse primer (51). The thermal cycles were as follows: 5 min at 94˚C followed by 22 cycles at 94˚C for 30s, 54˚C (CP, MP, 2a and 3` NTR) and 51˚C (1a and 2b) for 30s, 72˚C for 1 minute (80s for 1a) and finished by 72˚C for 7 minutes.

Sequencing

Gel purifications were carried out using a QIA quick gel extraction (QIAGEN, CA, USA). Purified products for each gene / isolate combination were bi-directionally sequenced using a model 377 ABI PRISM DNA sequencer (Perkin-Elmer, Fremont, CA) in the Automated DNA Sequencing Facility of the University of California-Davis. Consensus sequences of each gene / isolate combination obtained using the NTI Advance 11 program (Invitrogen, Carlsbad, USA) and later used for phylogenetic analysis.

Sequence Alignment

Multiple nucleotide sequence alignments were performed using CLUSTALW in MEGA version 5 (97). Alignments were manually edited to maximize nucleotide homology, to ensure proper reading frame in the coding regions, and to limit the number of gaps present in the sequence record.

Phylogenetic analysis and estimation of population genetic parameters

Aligned CMV sequences were assessed using DnaSP version 5.1 (83) software, to estimate genetic diversity and other population genetic parameters. Phylogenetic trees were constructed for six genomic segments by the neighbor-joining (NJ) method (84) using the MEGA program with 1,000 bootstrap replicates under the Kimura 2-parameter model (41). Bootstrap consensus trees were rooted using *Peanut stunt virus* (PSV) and *Tomato aspermy virus* (TAV) as the out-groups based on the current knowledge of the *Bromoviridae* (12). All branches with <70% bootstrap support were judged inconclusive and were collapsed (18, 33). For further analyses, reference isolate sequences were included as representatives of CMV subgroups IA, IB and II and these were downloaded from the NCBI Nucleotide Database (www.ncbi.nlm.nih.gov/nuccore).

Neutrality tests (Tests for selection)

To assess the potential for selection pressure to be imposed upon CMV coding regions, non-synonymous (dN, amino-acid altering) and synonymous (dS, silent) substitution rates and their associated ratios (dN/dS= ω) were estimated for each segment by using the bootstrap method with 500 replicates under the Kumar method (46) in MEGA version 5 (97). To determine the site specific selection pressure in each region, three complementary maximum-likelihood methods

including single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL) (44, 45) were implemented in the Hyphy package available at the Datamonkey web server (http://www. Datamonkey.org). To classify a site as positively or negatively selected, the cutoff P-value was selected to be 0.1 for SLAC and FEL. For REL, a Bayes factor of 50 was selected as the cut-off value. The most appropriate nucleotide substitution models were selected for each gene by the software and only selections determined to be significant by at least two methods were considered as a significant positive.

Recombination Analysis

To detect any possible recombination between different CMV isolates, the automatic recombination scans of the sequence alignments were carried out using the RDP3 program (56). In total, 7 recombination detection methods were implemented and included RDP (56), Bootscan (85), GENECONV (67), Maximum Chi Square (77, 92), Chimaera (77), Sister Scanning (25) and 3SQE (4). These methods were used to detect recombinant sequences, possible recombination break points, and identification of potential parental sequences among the CMV isolates. The program was run using the default settings plus the Bonferroni corrected *p*-value cut-off (α=0.05). Recombination events were considered as significant if four or more methods had a consensus P- value $\leq$0.01, in addition to phylogenetic evidence of recombination. The results obtained in the recombination analysis by RDP were confirmed using a boot scanning method (85) in the SimPlot program (54). The window width and the step size were set to 200 and 20 bp, respectively.

Results

Genetic diversity of selected CMV isolates

A total of thirty-five CMV isolates (Table 1) were included in this study and cDNA fragments representing 6 genomic regions of CMV (1a, 2a, 2b, MP, CP and 3`NTR of RNA3) were amplified from CMV isolates with RT-PCR. The six regions generated amplicons of 1000, 653, 378, 848, 678 and 315 bp in length, respectively for each of the viral genomic segments. The number of isolates which were amplified with specific primer sets for each genomic segment is illustrated in Table 2. Associated population genetic parameters were estimated (Table 2) including π, the average pairwise nucleotide difference per site and Θ_w, the mutation rate from segregating number, and these estimators were used as two indicators of genetic diversity for each gene. Overall, the genetic diversity for US CMV isolates was low with a mean genetic diversity of 0.034. Specifically, the 2b region showed the highest genetic variation among CMV coding regions followed by MP, 2a, CP and 1a considering both genetic variation estimations.

Phylogenetic relationship of CMV isolates

Phylogenetic trees were constructed for US CMV isolates based on partial nucleotide sequences of 1a and 2a and full nucleotide sequences of 2b, MP, CP and the 3`NTR of RNA 3 (Appendix 1). Fifteen reference sequences obtained from GenBank were included in phylogenetic analysis and phylogenetic trees were re-constructed (Fig. 1). Divergence of subgroups I and II, with high supporting values, was observed in all phylogenetic trees. Based on these analyses, subgroup I was the predominant subgroup among the US isolates included in this investigation (Fig. 1). Two isolates belonging to subgroup II were recognized according to the MP tree, but further analysis suggested that these may have been recombinants (details in the recombination analysis section) and were removed from the data set for the remainder of the analyses. The presence of both

subgroups IA and IB in US was confirmed by phylogenetic analyses and the divergence of these two subgroups was clear in most trees (Fig. 1). According to the CP phylogenetic tree, historical isolates included in this study represented a mixture of both subgroup IA and IB, while all new isolates collected in the last decade (e.g. 2002 – present) belonged to subgroup IA (Fig. 1e). These newly collected isolates were sampled from snap bean, with the exception of "HWH10" and "PV544NJU04" isolates, which were collected from other hosts. Furthermore, all subgroup IB isolates examined in this study were clustered together in the CP gene tree, separately from two other subgroup IB reference isolates (OHW and 2A1IL) previously collected from various locations in the US. These newly identified subgroup IB isolates in the US illustrated a close phylogenetic relationship with the 'Nt9' and 'Tfn' reference isolates, also determined to be subgroup IB strains and isolated in Taiwan and Italy, respectively, with a high supporting values (Fig. 1e). Upon closer examination of the MP tree, US subgroup IB isolates were placed in two separate clades (Appendix 1d and Fig. 1d). Two additional subgroup IB isolates ('PV29WIC76' and 'PV243AZM77') were separated from the remaining isolates and formed a unique clade which illustrated a close phylogenetic relationship with a reference subgroup IB strain ('SD') submitted from China also with a high supporting value (Fig. 1d). Similar to the CP gene tree, the remainder of subgroup IB isolates formed a cluster with 'Nt9' and 'Tfn' strains (Fig. 1d). In both CP and MP gene trees, subgroup IA US isolates formed a large clade with reference isolates from the same subgroups (Figs. 1d and 1e). Based on the 2a gene tree, subgroup IB clustering followed a similar trend as the CP gene tree (Fig. 1b). In contrast, two branches were seen inside of the subgroup IA clade forming a group with higher phylogenetic relationship to the 'Leg' isolate obtained from Japan (Fig. 1b). The topology of the 2b gene tree was different from 2a (Figs. 1b and 1c) which may be related to independent evolutionary histories of these two genes

originating from the same RNA. The similar pattern of the 2a gene subgroup IB clustering was observed for four identified subgroup IB isolates in the 2b gene tree (Fig. 1c). No specific clustering was observed inside of subgroup IA based on this gene (Fig. 1c and Appendix 1c).

More branching was observed in subgroup IA based on the 1a tree (Fig. 1a). It was interesting that all US isolates, with the exception of the 'HWH10' and 'KNYS09' isolates formed a unique clad inside of subgroup IA (Fig. 1a). Although these two isolates formed a separate clade inside the 1a tree, they were still recognized as subgroup IA (Fig. 1a and Appendix 1a) isolates. This result suggested that this gene may be conserved among the US CMV isolates included in the current study. Overall, no clear, significant associations between phylogenetic groups with location (e.g. state or origin of collection), host plant species, or collection year was observed and this result was consistent for all gene trees examined (Fig. 1 and Appendix 1).

Reassortment trace

To assess the potential for genetic exchange resulting from reassortment among the representatives of the US CMV population, phylogenetic trees of full nucleotide sequences of 2b, MP, CP, 3`NTR of RNA3, as well as partial sequences of 1a and 2a together with 15 additional reference CMV isolates from GenBank were compared. These phylogenetic comparisons suggested at least three obvious reassortants that had occurred in US isolates included in the current study. Specifically, analysis of the RNA3 (MP, CP and 3`NTR) and RNA2 (2a and 2b) placed the isolates 'PV243', 'PV29WI' and 'CaNY' into subgroup IB, whereas they belonged to subgroup IA based on analysis of the RNA1 (1a) segment (Fig. 1). These data suggested the potential for these three isolates to be the result of natural reassortment between subgroup IA and IB isolates at some point in their history.

Recombination analysis

Phylogenetic analysis further illustrated evidence for recombination between two CMV isolates included in this research. Two isolates (‘BORU93’ and ‘HORPR94’) were assigned to subgroup I based on the phylogenetic trees of the CP and 3`NTR genomic regions of the RNA3 (Fig. 2b and 2c), but were assigned to subgroup II according to the MP tree (Fig. 2a). Since these three regions occur on the RNA3, recombination may be a possible explanation for this outcome. To further investigate the accuracy of this phylogenetic signature of recombination, we concatenated the nucleotide sequences of MP, CP and the 3`NTR and subsequently evaluated them in the RDP3 package. Here, the RDP3 package scanned the aligned sequences using multiple methods including RDP, BootScan, GENECOV, Maxichi, Chimaera, Siscan and 3SEQ. Six out of seven methods implemented in this package suggested recombination events for these two isolates with highly significant p-values (Table 3). Moreover, the RDP3 program could detect the ‘NNY’ and ‘Trk7’ isolates as the possible major and minor parents, respectively with a high level of confidence for both recombinants. Positions 1 and 840 in each sequence were detected as the beginning and ending breakpoints for both isolates, respectively. Phylogenetic trees were constructed separately according to the non-recombination and detected recombination regions (Fig. 3). Both the ‘BORU93’ and ‘HORPR94’ isolates were assigned as subgroup IA based on the non-recombination region (Fig. 3a), while inclusion of the recombination region clustered these two isolates within subgroup II isolates (Fig. 3b).

To confirm the result obtained from the RDP3, phylogenetic and bootscan analyses using ‘BORU93’ and ‘HORPR94’ isolates as the query sequences were performed with Simplot package. We used the assumed parents estimated by RDP3 (NNY and Trk7) and representatives from both subgroups I and II (NNY, PNY, Fny and IA for subgroup I and Trk7, LS, and Q for

subgroup II) as the references. When the phylogenetic analysis was carried out for ‘BORU93’, a recombination point was detected at position 825 (Fig. 4a) of the sequence alignment, while for ‘HORPr94’, this position was nucleotide 841 (Fig.4b). Both of these positions were close to the position 844 estimated by RDP3 program. In order to confirm the results obtained by Simplot, similar sequences were used for a bootscanning analysis. The basic principle of bootscanning is that ‘mosaicism’ is suggested when one observes high levels of phylogenetic relatedness between a query sequence and more than one reference sequences in different genomic regions (85). Evidence of recombination is typically considered to be supported when 70% of permuted trees support a particular grouping of sequences. Bootscan analyses demonstrated that the ‘BORU93’ and ‘HORPr94’ isolates were built from a movement protein region related to the isolate from subgroup II, and a CP and 3`NTR region related to the isolate from subgroup IA (Fig. 5).

We also examined sequence alignments of other genomic regions to detect any possible recombination event by RDP3. Several isolates in each region were predicated to experience recombination events, but less than four out of seven methods with non-expected, significant p-values supported them. Also, Bootscan analysis with the Simplot program did not confirm these isolates as candidate recombinants. Therefore, they were not considered as candidate recombinants in our analyses.

Selection pressure on different coding regions

To determine the direction of the selective constraints imposed on different genomic regions of CMV and compare the degree of this selection among the coding regions, the patterns of selection in these genes was further analyzed. For this, the ratio of non-synonymous and synonymous substitutions was estimated for each segment (Table 2). There was no evidence for

positive selection in any of the CMV coding regions among the isolates included in this investigation. The mean ω values estimated from pairwise comparisons between sequences were less than 1 for all regions with the highest value for the 2b gene (Table 2). These ratios of ω indicated that CMV coding regions were subjected to negative or purifying selection. The 1a, MP and CP genes showed low ω ratios, suggesting high selective pressure, whereas this ratio was 5 to 9 times greater for 2a and 2b (Table 2), suggesting a greater tolerance for amino acid substitutions among these latter two genes. To identify selection at individual codons in each coding region, we applied three complementary maximum- likelihood methods (SLAC, FEL and REL). Overall, the estimated ω ratios using these methods further confirmed the result obtained from pairwise comparisons. However, sites 72 (V → A, V → M, A → V or V → T) in the 2b, 25 (P → L or P → S) in the CP and 76 (D → V or D → Q) in the 2a proteins were accepted as positively selected codons by two methods (Table 4). No positively selected codons were detected in the 1a and MP genes.

Discussion

Changes in the genetic composition of a virus population and new phenotypes which can arise as a result of genetic exchanges (e.g. reassortment and recombination), can compromise the effectiveness of disease control strategies. Therefore, an improved understanding of genetic structure of CMV and the associated factors or selective forces driving virus evolution can help to design improved disease management strategies. In the current study, our goal was to analyze the phylogeography and genetic diversity of CMV. Furthermore, we investigated sources of variation in the genetic diversity of CMV. The US CMV isolates included in this investigation exhibited a low genetic diversity. The observed low genetic diversity among isolates in this investigation is not surprising, however, as similar results have previously been reported for

populations of this virus in large geographic regions including California (51) and Spain (5, 20). A founder effect has been suggested as a partial explanation potentially shaping the observed genetic structure of CMV (20, 51). Genetic bottleneck(s) may also have contributed to this observed low genetic diversity and function to minimize the extent of genetic variation. Genetic bottlenecks during CMV systemic movement in host plants and CMV transmission through the aphid vector(s) have previously been reported (1, 50). This low genetic variation is true for most plant virus populations (23) and our finding is consistent with the concept that genetic stability is the rule in natural plant virus populations (23). Among US CMV coding regions in the current study, the 2b gene demonstrated more genetic variability (π=0.051) when comparing among all other regions. This result was consistent with previous investigations by Liu *et al.* (2004) and Lin *et al.* (2009), demonstrating that the 2b gene possessed the greatest diversity in the CMV genome. In our study, there was no clear association between phylogenetic groups and other biotic (host plant species) or abiotic factors (geographic location or year). Here again, our results are in agreement with previous findings obtained from California (51), Spain (22), and China (53) illustrating that the genetic structure of CMV varied randomly, without association with location, year, or host plant species. However, our sampling procedure may have had some limitations in this study. For example, we may have had insufficient replicates from each geographic region and host plant to conclude that genetic structure of CMV varied randomly. All CMV isolates included in this investigation belonged to subgroup I. If we consider these isolates as a reasonable approximation of the US CMV population, it suggests that subgroup I is of greater prevalence than subgroup II in this region, and this observation is consistent with previous results (22, 63). Within subgroup I, IA has a worldwide distribution (59), while most of the previously described subgroup IB strains were from Asia (82), although the presence of this

subgroup in other regions such as Italy (81) and Greece (88) has been reported. In the United States, the presence of two CMV subgroup IB strains isolated from pepper in 1990 in CA and one strain isolated from banana in 1995 in Hawaii was reported for the first time (51, 82). Sequence analysis in the current study detected several subgroup IB isolates among historic CMV isolates and phylogenetic analysis further revealed the presence of this specific subgroup in the US prior to these references. However, no CMV subgroup IB was detected among recently collected isolates sampled from a single host, snap bean in affected regions by the recent virus outbreaks.

Reassortment in RNA viruses with segmented genome has been shown for animal and plant viruses (9, 30, 51, 71, 79, 81, 87, 101, 102). Here, we report phylogenetic evidence for natural, inter-subgroup reassortment between subgroups IA and IB for three CMV isolates included in this study. This type of reassortment has been previously reported for CMV isolates (5, 51, 81). Although, we were only able to partially assess the RNA1 (~1Kb), we believe that this sequence is sufficient to distinguish subgroup IA and IB from each other.

This study also detected two isolates as the natural recombinants from a subgroup II pattern for the MP gene and a subgroup IA pattern for CP gene. Recombination was confirmed by both phylogenetic and computational analyses designed to detect recombination events. In similar studies investigating natural CMV populations in Spain, approximately 17% of the sequenced isolates possessed evidence for recombination derived principally from the RNA 3 of CMV subgroup IA and IB, with MP CMV (IA)/CP CMV (IB) as the most prevalent type of recombinant (5). To our knowledge, this is the first report of recombination between subgroups I and II at the RNA 3 in natural populations of CMV. To construct phylogenetic trees, however,

we removed the sequences of these two recombinant isolates from our data because recombination can mislead the phylogenetic estimation procedures (76).

A low frequency of genetic exchange (reassortment and recombination) among CMV isolates assessed in this research agrees with the previous results obtained from analysis of the genetic structure of field population of CMV in Spain and California (5, 10, 20, 51) and illustrates that these events are counter- selected in CMV natural population (19).

An analysis of natural selection showed that negative (or purifying) selection was the predominant evolutionary force operating upon all CMV coding regions. This type of selection imposed on CMV encoded proteins has been shown with CMV population of California and China (51, 59). On the average, the evolutionary constraints exerted on proteins 1a, MP and CP were larger than 2b and 2a. This observation is in agreement with the previous observation from California CMV populations (51) showing that the 2b and 2a proteins are more flexible with regard to amino acid substitutions and is also consistent with the idea that different coding regions of CMV are under different constraints. However, the ω ratio was higher for 2a and 2b genes in our research when compared to the CA ω values, likely because we had a diverse collection of CMV isolates from both subgroups IA and IB and different geographic locations throughout the US. Selection can be associated with various factors such as structural features of the virus, host plant and a vector. Garcia-Arenal *et al.* (2001) illustrated that negative selection predominates during evolution of plant viruses when the entire genome is assayed and that this purifying selection is principally due to the internal and external constraints (23). In the case of CMV, almost all encoded proteins have direct or indirect interactions either with the host plant (1a, 2a, 2b, MP and CP) or the insect vector (CP). Moreover, CMV possesses a very broad host range and can be transmitted by a large number of aphid vectors, but there remains some degree

of specificity for both transmission and infectivity by this virus (42, 72, 90, 95, 96). In order for this virus to effectively adapt to this level of host and vector variation, we expected to see some degree of positive (diversifying) selection in portions of the CMV coding regions similar to those presented for other plant viruses in the past surveys (27, 59, 60, 86, 98). In the CP gene, we noted that site 25 was accepted as the positively selected codon in subgroup IA. This result corroborated previous reports illustrating diversifying selection for this codon (59). This amino acid is located in the folded portion of CP (92) and it affects virus transmission by aphid (72). Hence, a positive selection pattern in this CP region could be related to the role of different aphid species in selection of different virus variants (59). Furthermore, Site 76 in the 2a protein was shown to be under positive selection in the current study. These changes were detected in only a portion of subgroup IB isolates. Referencing the *brome mosaic virus* (BMV), another member of the family *Bromoviridae*, the N-terminal, 115 amino acids of the 2a protein are necessary to interact with the helicase domain of 1a protein in the yeast two- hybrid system (40, 65). Hence, the detected positively selected site in the CMV 2a protein is probably involved in this interaction and is necessary for replication. Positively selected site 72 in the 2b protein located in the overlapping region with ORF2a was strongly accepted by all three maximum-likelihood methods, demonstrating a strong diversifying selection on this particular codon. Here again, 2b is a multifunctional protein involved in virus, long- distance movement, symptom induction, silencing suppressor and as a pathogenicity determinant (7, 14, 15, 16). Therefore, there is likely a high level of interaction between this protein and host components and, at least in the case of long- distance movement, this interaction may regulate host specificity (16). This host-specific function and the extremely wide host range of CMV allow this protein to be considerably more tolerant to nucleotide and amino acid changes (The highest estimated ω value among all coding

regions). Multiple amino acid substitutions at this particular site suggest a greater selection potential for adaptation. However, no specific amino acid / host relationship was recognized for this particular site in this study. Compared to the known domains of the 2b protein with known functions, this specific codon does not belong to any (48, 49). Hence, we have insufficient information at the current time to propose reasons of this selection. However, we hypothesize that amino acids located in this region of the protein are involved in an important function, because CMV isolates from California showed a greater average number of non-synonymous mutations than synonymous in a short region of the 2b gene (codons 81-93) (51). Also, we have demonstrated in our current investigations (Chapter 3 of this thesis) that codon 80 of the 2b protein of CMV field isolates in Wisconsin snap bean is under positive selection. Therefore, an improved understanding of the function of this region remains as an interesting aspect that warrants further investigation.

Although, the selection determination methods applied in the current study detected two positively selected sites 249 and 258 in the 1a and MP proteins, respectively, but we did not accept them because only one out of three methods confirmed these selections. Both of these sites have been reported as the positively selected codons in the previous studies (13, 59).

It is also noteworthy here to mention that positively selected codons reported for the 2a and 2b proteins in this research, have not been reported in the previous investigations (59) and may be unique. However, the number of positively selected sites for the CP and 1a genes among the range of CMV isolates included in this investigation was much less. There are possible explanations for these differences. First of all, different methods for detecting positive selection, may well account for the observed differences between these two studies. Secondly, we selected isolates from natural populations of CMV in our research composed of subgroups IA and IB in

US, while the previous research included only available CMV sequences in GenBank belonging to subgroups IA, IB and II obtained from both experimental and natural conditions with undefined constraints imposed upon the coding regions. Finally, we sequenced ORF1a only partially. Therefore, if we had assessed the whole gene, perhaps we would have detected more positive codons with a higher level of statistical confidence.

Acknowledgments

We thank our colleagues from different Universities and Research Centers in US who provided us CMV infected plant tissues, in specific Drs. Said Gabriel, Bob Martin, Keith Perry, Mark Fuchs, Wayne Borth and Bryce Falk.

Table 1. CMV isolates collected in the United States and included in the phylogenetic analysis.

Isolate	Geographic origin	Host	Date collected	Isolate	Geographic origin	Host	Date collected
AORU93	Oregon	Unknown	1993	NNYS09	New York	Snap Bean	2009
BORU93	Oregon	Unknown	1993	ONYS09	New York	Snap Bean	2009
FORU93	Oregon	Unknown	1993	HWH10	Hawaii	*Commelina diffusa*	2010
GORU93	Oregon	Unknown	1993	NS3WIS09	Wisconsin	Snap bean	2009
HORPr94	Oregon	*Primula* sp.	1994	TWIS08	Wisconsin	Snap bean	2008
KYKTTo08	Kentucky	Tobacco	2008	WWIS07	Wisconsin	Snap bean	2007
CaNYU90	New York	Unknown	Before 1994	MirrorWIS07	Wisconsin	Snap bean	2007
CENYP90	New York	Pepper	Before 1991	3ARS50	Arkansas	Spinach	1950
PNYU90	New York	Unknown	Before 1991	113CAT90	California	Tomato	1990
NKNYC90	New York	Cucumber	Before 1991	116CAP90	California	Pepper	1990
V154NYT85	New York	Tomato	Before 1989	160ECAP90	California	Pepper	1990
V85NYT80	New York	Tomato	Before 1982	MDCAP93	California	Pepper	1993
PV243AZM77	Arizona	Mungbean	1977	144ICAP90	California	Pepper	1990
PV29WIC76	Wisconsin	Cucumber	1976	INYS09	New York	Snap Bean	2009
PV30MDH85	Maryland	*Commelina nudiflura*	1985	JNYS09	New York	Snap Bean	2009
PV544NJU04	New Jersey	Unknown	2004	KNYS09	New York	Snap Bean	2009
PV59AZSU84	Arizona	Sugar beet	1984	MNYS09	New York	Snap Bean	2009
LNYS09	New York	Snap Bean	2009	-	-	-	-

Table 2. Population genetic parameters estimated for coding regions of US CMV isolates using DnaSP and MEGA program.

Genomic region	Number of isolates	[1]S	[2]η	[3]π	[4]Θ_w	[5]dS	[6]dN	[7]ω (dN/dS)
1a	29	68	70	0.013	0.017	0.033±0.005	0.005±0.002	0.151
2a	31	67	69	0.034	0.030	0.068±0.012	0.022±0.005	0.323
2b	30	74	76	0.051	0.058	0.072±0.017	0.041±0.009	0.570
MP	30	116	123	0.039	0.034	0.111±0.013	0.011±0.003	0.100
CP	31	72	76	0.033	0.028	0.098±0.013	0.006±0.002	0.061

[1]S: Total number of segregating sites

[2]η: Total number of Mutations

[3]π: Nucleotide Diversity, average pairwise nucleotide difference per site

[4]Θ_w: Mutation rate estimated from S

[5]dS: The average number of pairwise differences per synonymous site

[6]dN: The average number of pairwise differences per non-synonymous site

[7]dS and dN were estimated by Kumar method

Table 3. Probability estimates for the 'BORU93' and 'HORPr94' isolates to be considered as recombinants.

Concatenated nucleotide sequences of MP, CP and 3`NTR regions was used as a query. Probability values *(p*-value) were set at level of 0.05 for all methods and the program was implemented using the default settings.

Algorithm	***p*-value**
RDP	-
Geneconve	3.361×10^{-50}
BootScan	6.209×10^{-55}
Maxichi	2.649×10^{-29}
Chimaera	1.676×10^{-29}
SisCan	1.773×10^{-30}
3SEQ	1.172×10^{-100}

Table 4. Codon positions of the coding regions in the US CMV populations affected by positive selection.

Positively selected codons were determined using three maximum-likelihood methods. Critical *p*- values for SLAC and FEL were 0.1 and the minimum Bayes factor value was set at 50. The best nucleotide substitution model for each coding region was selected as follows: 1a and CP, HKY85; 2b and MP, TrN93; 2a, REV. **Sites detected as statistically significant by two methods. ***Site detected as statistically significant by all three methods.

Coding region	Site	SLAC dN-dS	SLAC p-value	FEL dN-dS	FEL p-value	REL dN-dS	Bayes Factor
2a	76**	5.453	0.401	46.346	0.070	1.369	88.763
2b	72***	7.664	0.099	34.739	0.032	7.787	693.8643
CP	25**	8.110	0.296	23.710	0.077	0.662	334.124

Figure 1. CMV phylogenetic analysis based on six genomic regions including US and 15 reference isolates.

The rooted condensed tree (cutoff value of 70%) was constructed for US CMV isolates based on the a) 1a (Partial sequence); b) 2a (Partial sequence); c) 2b (Complete sequence); d) MP (Complete sequence); e) CP (Complete sequence); and f) 3`NTR of RNA3 complete sequence by the neighbor-joining method using the Kimura-2 parameters model of nucleotide substitution implemented in MEGA 5. The robustness of branching patterns was tested by 1,000 bootstrap replications and bootstrap values are indicated at nodes. The trees were rooted by using PSTV and TAV as out-groups. CMV subgroups are indicated by vertical brackets. Reference isolates from GenBank are indicated by red stars and the accession numbers were listed in Appendix 3.

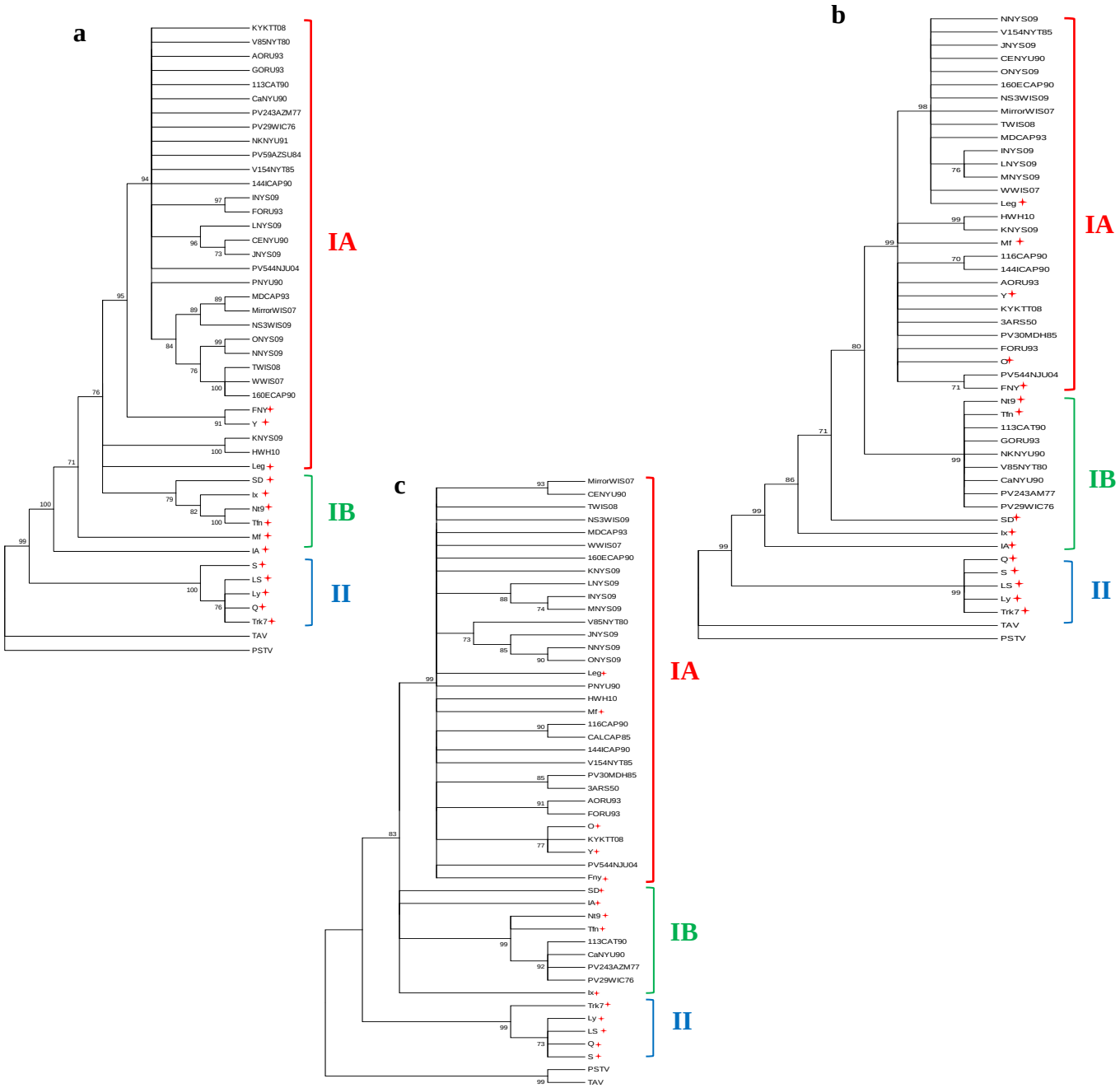


Figure 1- Continued

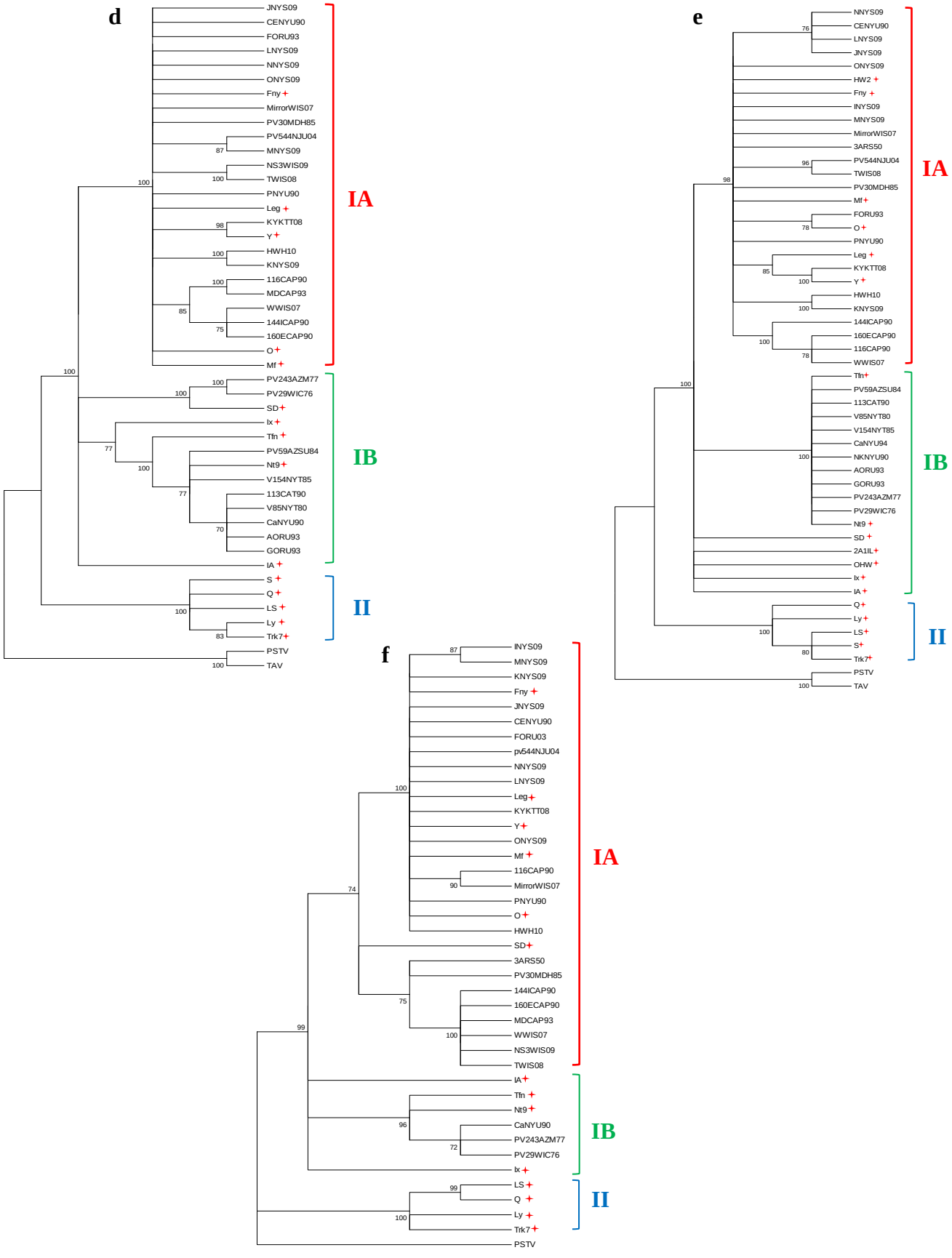


Figure 2. Phylogenetic position of two US CMV isolates, 'BORU93' and 'HORPr94' based on a) MP, b) CP, c) 3`NTR of RNA3.

CMV subgroups are indicated by vertical brackets and the position of these two isolates is marked by orange arrows.

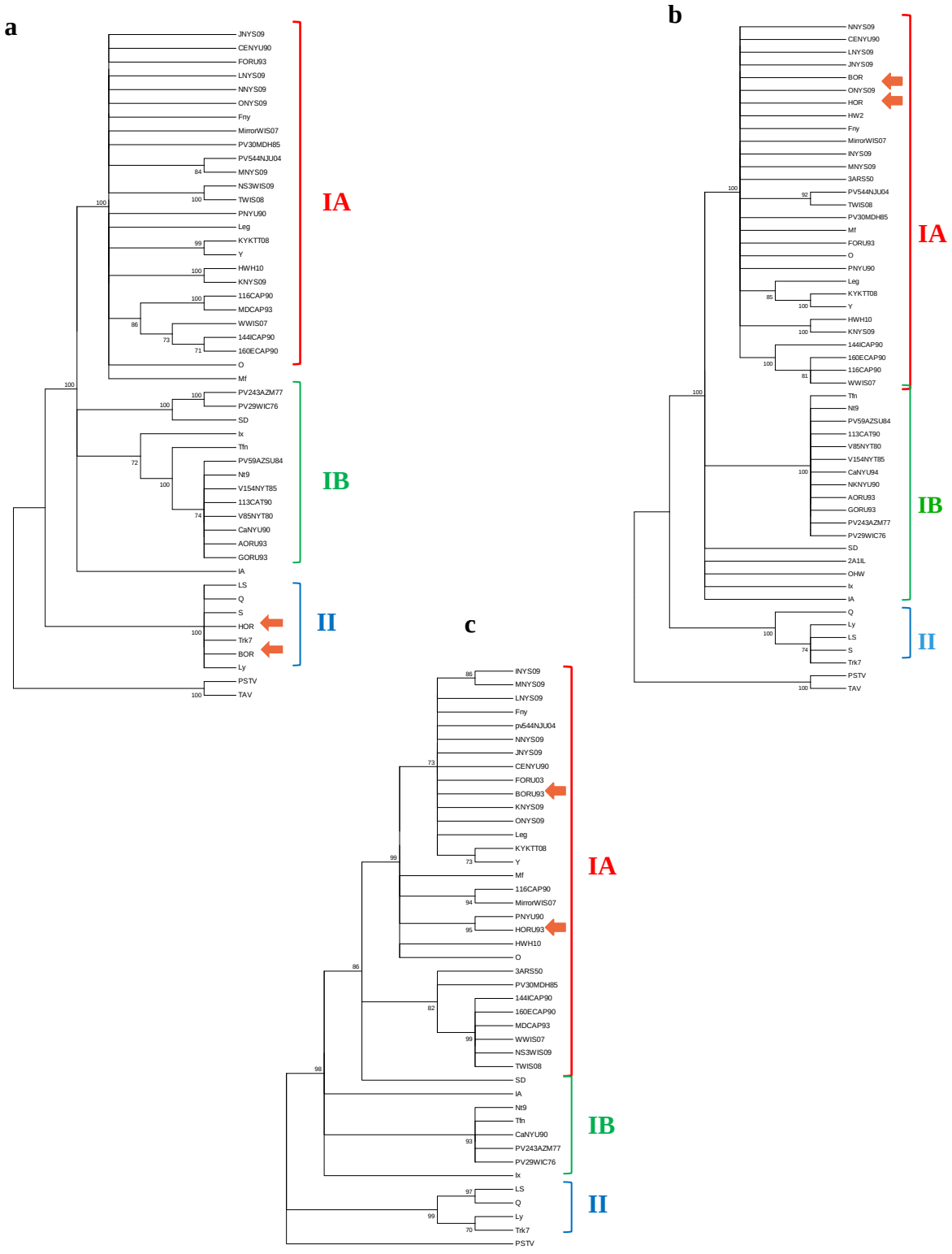
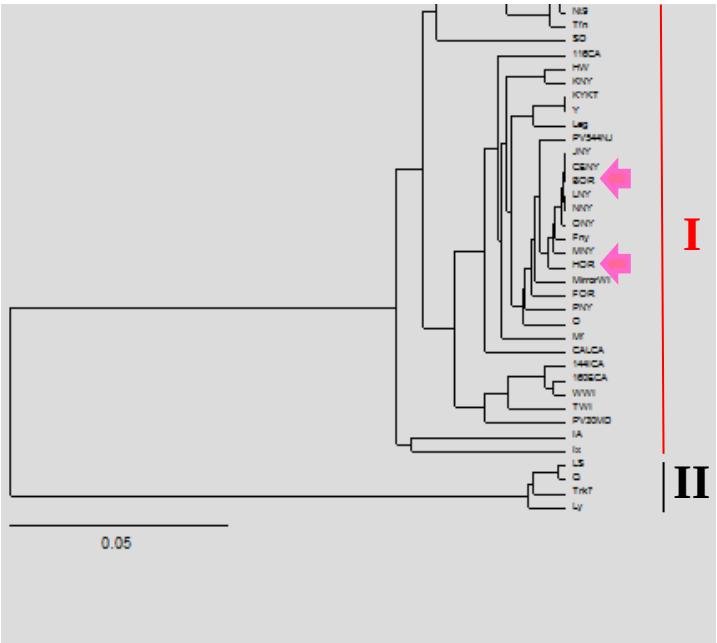


Figure 3. Concatenated trees constructed based on the non-recombination (a) and recombination (b) regions detected for 'BORU93' and 'HORPr94' isolates.

'BORU93' and 'HORPr94' were grouped with other subgroup I isolates based on the non-recombination while these two particular isolates were clustered with the reference isolates belonging to subgroup II according to the recombination region. CMV subgroups I and II are indicated by vertical brackets and the position of the two isolates is marked by pink arrows.

a



b

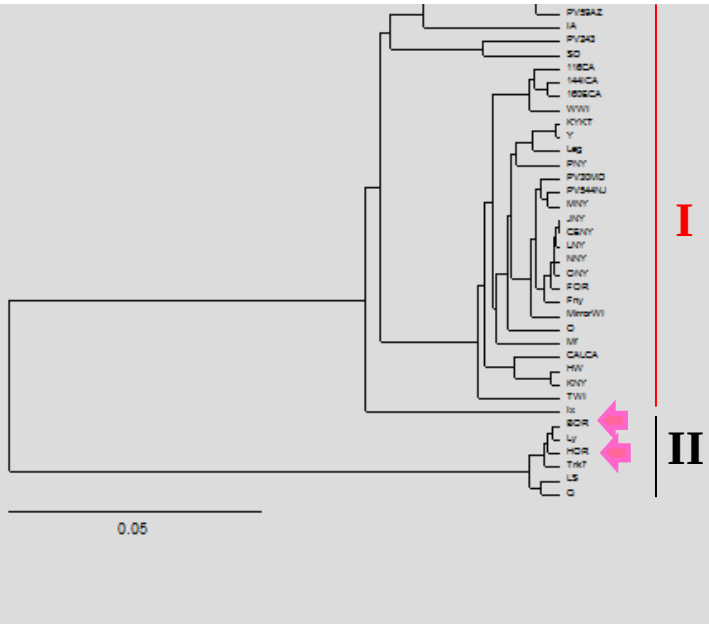


Figure 4. Phylogenetic relationship of CMV isolates on the basis of concatenation of nucleotide sequences of MP, CP and 3`NTR using Simplot.

Two US CMV isolates, 'BORU93' (a) and 'HORPr94' (b) were used as the query sequence and six CMV isolates as reference sequences. Y-axis varies in identity percentage within a sliding window of 200 bp and a step size of 20 bp. Black vertical dashed line shows the proposed recombination break point. Sequences compared with the query sequence are indicated in the legend.

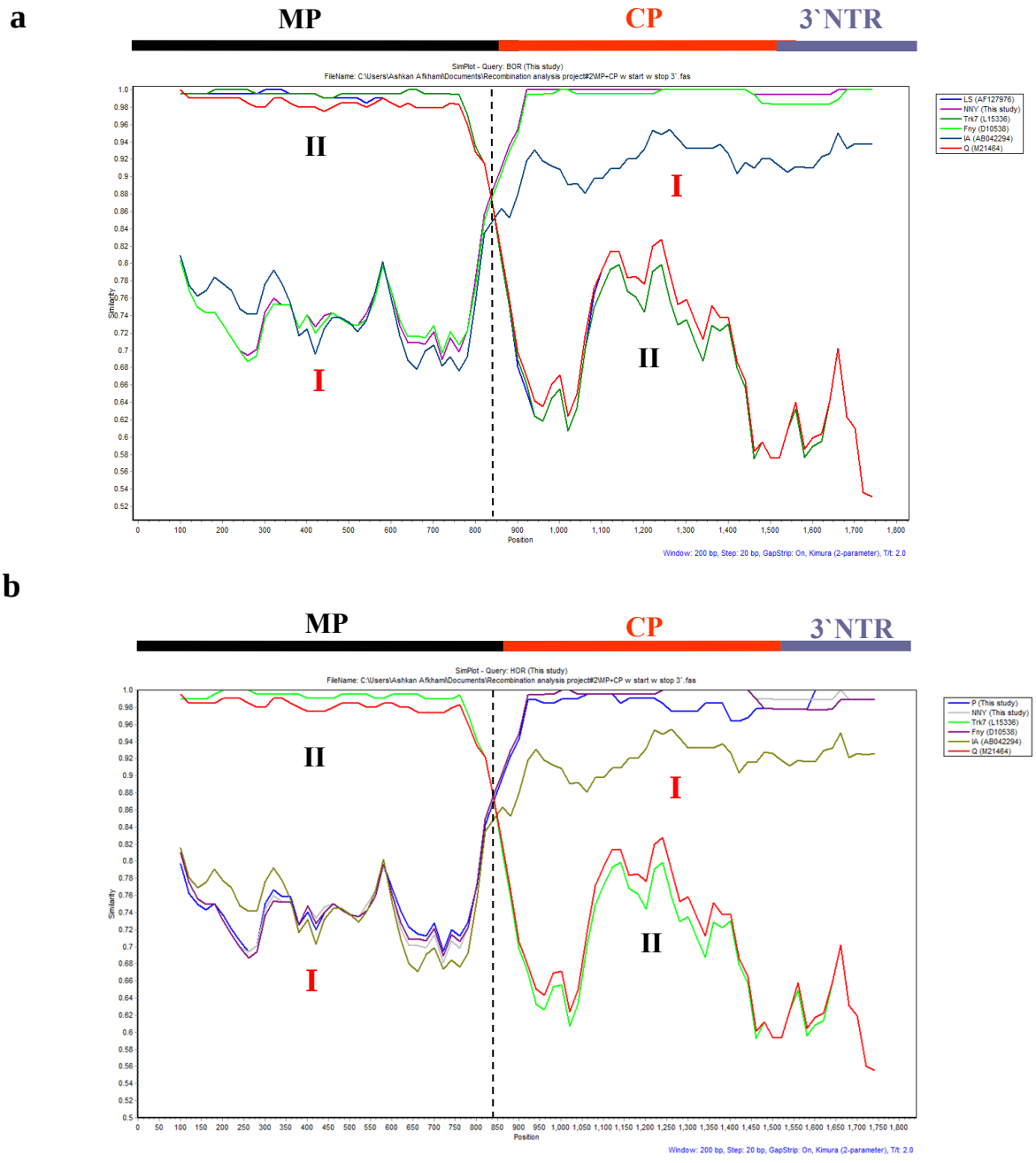
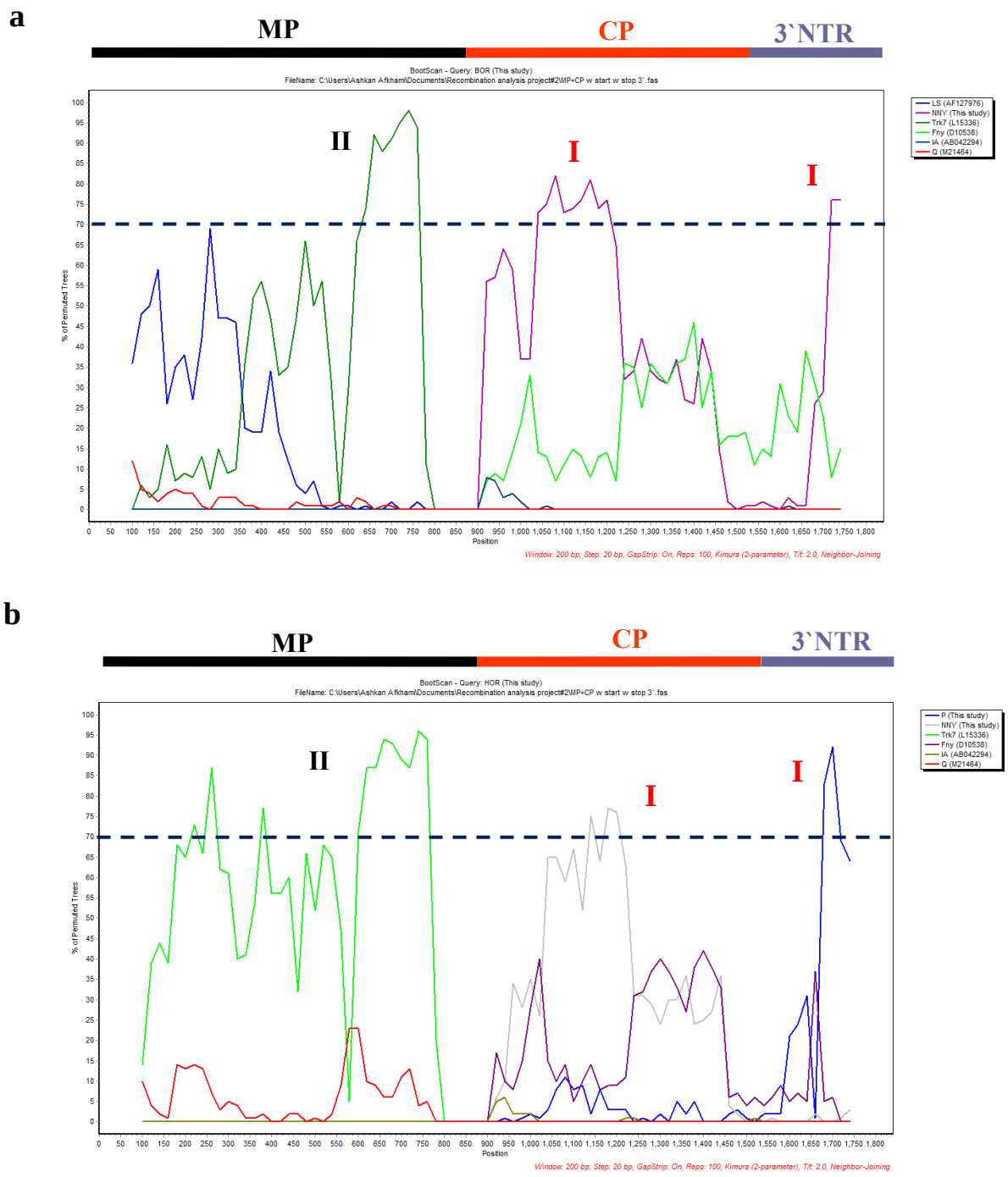


Figure 5. Bootscan analysis with 'BORU93' (a) and 'HORPr94' (b) isolates as the query sequence.

Query sequences are illustrated on the upper portion of the figure. Sequences compared with the query sequence are indicated in the legend. Analysis was performed with a sliding window of 200 bp and a step size of 20 bp. The y-axis illustrates the percentage of permuted trees in which each selected isolate cluster with the query sequence. The 70% cutoff level, representing possible recombination, is indicated by the dashed line.



References

1- Ali, A., Li, H., Schneider, W. L., Sherman, D. J., Gray, S., Smith, D. & Roossinck, M. J. (2006). Analysis of genetic bottlenecks during Horizontal transmission of *Cucumber mosaic virus*. *J Gen Virol* **80**, 8345- 8350.

2- Aranda, M. A., Fraile, A., Dopazo, J., Malpica, J. M. & Garcı´a-Arenal, F. (1997). Contribution of mutation and RNA recombination to the evolution of plant pathogenic RNA. *J Mol Evol* **44**, 81–88.

3- Boccard, F. & Baulcombe, D. (1993). Mutational analysis of cis-acting sequences and gene function in RNA3 of *cucumber mosaic virus*. *Virology* **193**, 563–578.

4- Boni, M. F., Posada, D. & Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **176**, 1035-1047.

5- Bonnet, J., Fraile, A., Sacrista´ n, S., Malpica, J. M. & Garcı´a-Arenal, F. (2005). Role of recombination in the evolution of natural populations of *Cucumber mosaic virus*, a tripartite RNA plant virus. *Virology* **332**, 359–368.

6- Bruenn, J. A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res* 19, 217–226.

7- Brigneti, G., Voinnet, O., Li, W.X., Ji, L. H., Ding, S. W. & Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J* **17**, 6739–6746.

8- Canto, T., Prior, D. A. M., Hellwald, K. H., Oparka, K. J. & Palukaitis, P. (1997). Characterization of *cucumber mosaic virus*. IV. Movement protein and coat protein are both essential for cell-to-cell movement of *cucumber mosaic virus*. *Virology* **237**, 237–248.

9- Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455.

10- Chen, Y., Chen, J., Zhang, H., Tang, X. & Du, Z. (2007). Molecular evidence and sequence analysis of a natural reassortant between *Cucumber mosaic virus* subgroup IA and II strains. *Virus Genes* **35**, 405- 413.

11- Chen, Y. K., Goldbach, R. & Prins, M. (2002). Inter- and intramolecular recombinations in the *cucumber mosaic virus* genome related to adaptation to alstromeria. *J Virol* **76**, 4119- 4124.

12- Codoner, F. M. & Elena, S. F. (2008). The promiscuous evolutionary history of the family *Bromoviridae*. *J Gen Virol* **89**, 1739-1747.

13- Davino, S., Panno, S., Rangel, E. A., Davino, M., Bellardi, M. G. & Rubio, L. (2012). Population genetics of *cucumber mosaic virus* infecting medicinal, aromatic and ornamental plants from northern Italy. *Arch Virol* **157**, 739- 745.

14- Diaz-Pendon, A., Li, F., Li, W.X. & Ding, S.W. (2007). Suppression of Antiviral Silencing by *Cucumber Mosaic Virus* 2b Protein in Arabidopsis Is Associated with Drastically Reduced Accumulation of Three Classes of Viral Small Interfering RNAs. *Plant Cell* **19**, 2053–2063.

15- Ding, S.W., Anderson, B. J., Haase, H. R. & Symons, R. H. (1994). New overlapping gene encoded by the *cucumber mosaic virus* genome. *Virology* **198**, 593–601.

16- Ding, S. W., Li, W. X. & Symons, R. H. (1995). A novel naturally occurring hybrid gene encoded by a plant RNA virus facilitates long distance virus movement. *EMBO J* **14**, 5762–577.

17- Domingo, E., Holland, J. J. & Ahlquist, P. (1988). RNA Genetics CRC Press, Boca Raton, FL

18- Efron, B., Halloran, E. & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci USA* **93**, 7085–7090.

19- Escriu, F., Fraile, A. & Garcı´a-Arenal, F. (2007). Constraints to genetic exchange support gene coadaptation in a tripartite RNA virus. *PLoS Pathog* **3**, e8. doi:10.1371/

20- Fraile, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M. & Garcı´a-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *J Virol* **71**, 934–940.

21- Fujiwara T., Giesmann-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.

22- Garcia-Arenal, F., Escriu, F., Aranda, M. A., Alonso-Prados, J. L., Malpica, J. M. & Fraile, A. (2000). Molecular epidemiology of *Cucumber mosaic virus* and its satellite RNA. *Virus Res* **71**, 1-8.

23- Garcia-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* **39**, 157-186.

24- German, T. L., Thompson, A. & Wills, D. K. (2004). Statewide distribution of virus problems on processing beans. In. Proc 2004 Midwest Food Processors Association Processing crops Conference. Feb 23-24, La Cross, WI Department of HORPr94ticulture, University of Wisconsin.

25- Gibbs, M. J., Armstrong, J. S. & Gibbs, A. J. (2000). Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573-582.

26- Gibbs, A., Calisher, C. H. & Garcia-Arenal, F. (1995). Molecular Basis of Viral Evolution. Cambridge University Press, New York, NY

27- Glasa, M., Marie-Jeanne, V., Moury, B., Ku´ dela, O. & Quiot, J. B. (2002). Molecular variability of the P3-6K1 genomic region among geographically and biologically distinct isolates of *Plum pox virus*. *Arch Virol* **147**, 563–575.

28- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1988). A novel super family of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett* 235, 16–24.

29- Grau, C. R., Stevenson, W. R. & Mondjana, A. M. (2002). Viruses causing losses on processing beans in the Midwest, In. Proc 2002 Wisconsin fertilizer, Aglime and Pest Management Conference, **41**, 248-256.

30- Guan, Y., Poon, L. L. M., Cheung, C. Y., Ellis, T. M. & Lim, W. (2004). H5N1 influenza: A protean pandemic threat. *Proc Natl Acad Sci U S A* **101**, 8156– 8161.

31- Habili, N. & Symons, R. H. (1989). Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res* **17**, 9543–9555.

32- Hayes, R. J. & Buck, K. W. (1990). Complete replication of a eukaryotic virus RNA in vitro by a purified RNA-dependent RNA polymerase. *Cell* **63**, 363–368.

33- Hillis, D. M. & Bull, J. J. (19930. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**, 182– 192.

34- Hodgman, T. C. (1988). A new superfamily of replicative proteins. *Nature* **333**, 22–23.

35- Holland, J. J. (1992). Genetic Diversity of RNA Viruses. *Curr Top Microbiol Immunol* **176**, 1-20.

36- Hong, J. S., Ohnishi, S., Masuta, C., Choi, J. K. & Ryu, K. H. (2007). Infection of soybean by *cucumber mosaic virus* as determined by viral movement protein. *Arch Virol* **152**, 321-328.

37- Ishihama, A. & Barbier, P. (1994). Molecular anatomy of viral RNA-directed RNA polymerase. *Arch Virol* **134**, 235-258.

38- Itaya, A., Hickman, H., Bao, Y., Nelson, R. & Ding, B. (1997). Cell to- cell trafficking of *Cucumber mosaic virus* movement protein: Green fluorescent protein fusion produced by biolistic gene bombardment in tobacco. *Plant J* **12**, 1223–1230.

39- Kadare´, G. & Haenni, A. L. (1997). Virus-encoded RNA helicases. *J Virol* **71**, 2583–2590.

40- Kao, C. C. & Ahlquist, P. (1992). Identification of the domains required for direct interaction of the helicase-like and polymerase-like RNA replication proteins of *brome mosaic virus*. *J Virol* **66**, 7293–7302.

41- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120.

42- Kobori, T., Miyagawa, M., Nishioka. K., Ohki, S. T. & Osaki, T. (2002). Amino acid 129 of *cucumber mosaic virus* coat protein determines local symptom expression and systemic movement in *Tetragonia expansa*, *Momordica charantia* and *Physalis floridana*. *J Gen Plant Pathol* **68**, 81–88.

43- Koundal, V., Hag, Q. M. & Praveen, S. (2011). Characterization, genetic diversity, and evolutionary link of *Cucumber mosaic virus* strain New Delhi from India. *Biochem Genet* **49**, 25-38.

44- Kosakovsky-Pond., S .L. & Frost, S .D .W. (2005a). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531- 2533.

45- Kosakovsky-Pond, S.L., Frost, S. D. W., Grossman, Z., Gravenor, M .B., Richman, D. D. & Brown, A. J. (2006). Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLOS Comput Biol* **2**, e62, doi: 10.1371/journal. Pcbi.0020062.

46- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in *Bioinformatics* **5**, 150-163.

47- Larsen, R. C., Miklas, P. N., Eastwell, K. C., Grau, C. R. & Mondjana, A. (20020. A virus disease complex devastating late season snap bean production in the Midwest. *Annu Rep Bean Improv Coop* **45**, 36- 37.

48- Lewsey, M., Surette, M., Robertson, F. C., Ziebell, H., Choi, S. H. & Ryu, K. H. (2009). The role of the *cucumber mosaic virus* 2b protein in viral movement and symptom induction. *Mol Plant-Microbe Interact* **22**, 642-54.

49- Lewsey, M., González, G., Kalinina, N. O., Palukaitis, P., Canto, T. & Carr1, J. P. (2010). Symptom induction and RNA silencing suppression by the *cucumber mosaic virus* 2b protein. *Plant Signal Behav* 5, 705- 708.

50- Li, H. & Roossinck, M. J. (2004). Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J Virol* **78**, 10582– 10587.

51- Lin, H. X, Rubio, L., Smythe, A. & Falk, B. W. (2004). Molecular population genetics of cucumber mosaic virus in California: evidence for founder effects and reassortments. *J Virol* **78**, 6666–6675

52- Lin, H. X., Rubio, L., Smythe, A., Jiminez, M. & Falk, B. W. (2003). Genetic diversity and biological variation among California isolates of *Cucumber mosaic virus*. *J Gen Virol* **84**, 249–258.

53- Liu, Y.Y., Yu, S. L., Lan, Y. F., Zhang, C. L. & Hou, S. S. (2009). Molecular variability of five *Cucumber mosaic virus* isolates from China. *Acta Virologica* **53**, 89-97.

54- Lole, S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiencey virus Type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**, 152–160.

55- Martin, D. & Rybicki, E. (2000). RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563.

56- Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: Recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.

57- Maynard Smith, J. (1992). Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126-129.

58- Morse, S. S. (1994). The Evolutionary Biology of Viruses. Raven Press, New York, NY.

59- Moury, B. (2004). Differential Selection of Genes of *Cucumber Mosaic Virus* Subgroups. *Mol Biol Evol* **21**, 1602–1611.

60- Moury, B., Morel, C., Johansen, E. & Jacquemond, M. (2002). Evidence for diversifying selection in potato virus Y and in the coat protein of other potyviruses. *J Gen Virol* **83**, 2563–2573.

61- Mossop, D. W. & Francki, R. I. B. (1977). Association of RNA 3 with aphid transmission of *cucumber mosaic virus*. *Virology* **81**,177–181.

62- Nault, B. A., Shah, D. A. & Taylor, A. G. (2006). Viruses and aphids everywhere in New York snap bean fields in 2005, pp. 74-76. *In* Proceedings of the 2006 Empire State Fruit and Vegetable EXPO. February 13-16, 2006. Cornell Coop. Exten., Syracuse, NY.

63- Ng, J. & Perry, K. L. (1999). Stability of the aphid transmission phenotype in *cucumber mosaic virus*. *Plant Pathol* **48**, 288- 394.

64- Oparka, K.J., Boevink, P. & Santa Cruz, S. (1996). Studying the movement of plant viruses using green fluorescent protein. *Trends Plant Sci* **1**, 412–418.

65- O'Reilly, E. K., Paul, J. D. & Kao, C. C. (1997). Analysis of the interaction of viral replication proteins by using the yeast two-hybrid assay. *J Virol* **71**, 7526-7532.

66- Owen, J. & Palukaitis, P. (1988). Characterization of *cucumber mosaic virus*. I. Molecular heterogeneity mapping of RNA 3 in eight CMV strains. *Virology* **166**, 495-502.

67- Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225.

68- Palukaitis, P. & Garcia-Arenal, F. (2003). *Cucumoviruses*. *Adv Virus Res* **62**, 241– 323.

69- Palukaitis, P., M. J. Roossinck, R. G. Dietzgen & Francki, R. I. B. (1992). *Cucumber mosaic virus*. *Adv Virus Res* **41**, 281-348.

70- Peden, K. W. C. & Symons, R. H. (1973). *Cucumber mosaic virus* contains a functionally divided genome. *Virology* **53**, 487–492.

71- Perry, K. L. & Francki, R. I. B. (1992). Insect-mediated transmission of mixed and reassorted *cucumovirus* genomic RNAs. *J Gen Virol* **73**, 2105– 2114.

72- Perry, K. L., Zhang, L. & Palukaitis, P. (1998). Amino acid changes in the coat protein of *cucumber mosaic virus* differentially affect transmission by the aphids *Myzus persicae* and *Aphis gossypii*. *Virology* **242**, 204–210.

73- Perry, K. L., Zhang, L., Shintaku, M. H. & Palukaitis, P. (1994). Mapping determinants in *cucumber mosaic virus* for transmission by *Aphis gossypii*. *Virology* **205**, 591–595.

74- Pierrugues, O., Guilbaud, L., Fernandes-Delmond, I., Faber, F., Tepfer, M. & Jacquemond, M. (2007). Biological properties and relative fitness of inter-subgroup *cucumber mosaic virus* RNA 3 recombinants produced in Vitro. *J Gen Virol* **88**, 2852-2861.

75- Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* **8**, 3867–3874.

76- Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **54**, 396-402.

77- Posada, D. & Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci U S A* **98**, 13757-13762.

78- Rizos, H., Gunn, L. V., Pares, R. D. & Gillings, M. R. (1992). Differentiation of *cucumber mosaic virus* isolates using the polymerase chain reaction. *J Gen Virol* **73**, 2099–2103.

79- Robinson, D. J., Hamilton, W. D. O., Harrison, B. D. & Baulcombe, D. C. (1987). Two anomalous tobravirus isolates: evidence for RNA recombination in nature. *J Gen Virol* **68**, 2551–2561.

80- Rodriguez- Alvarado, G., Kurath, G. & Dodds, J. A. (1995). Heterogenity in pepper isolates of *cucumber mosaic virus*. *Plant Dis* **79**, 450- 455.

81- Roossinck, M. J. (2002). Evolutionary history of *Cucumber mosaic virus* deduced by phylogenetic analyses. *J Virol* **76**, 3382–3387.

82- Roossinck, M. J., Zhang, L. & Hellwald, K. H. (1999). Rearrangements in the 59 nontranslated region and phylogenetic analysis of *cucumber mosaic virus* RNA 3 indicate radial evolution of three subgroups. *J Virol* **73**, 6752–6758.

83- Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496.

84- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

85- Salminen, M. O., Carr, J. K., Burke, D. S. & McCutchan, F. E. (1995). Identification of breakpoints in intergenotypic recombinants of HIV-1 by bootscanning. *AIDS Res Hum Retrov* 11, 1423-1425.

86- Schirmer, A., Link, D., Cognat, V., Moury, B., Beuve, M., Meunier, A., Bragard, C., Gilmer, D. & Lemaire, O. (2005). Phylogenetic analysis of isolates of Beet necrotic yellow vein virus collected worldwide. *J Gen Virol* **86**, 2897-2911.

87- Schwinghamer, M. W. & Symons, R. H. (1977). Translation of the four major RNA species of *cucumber mosaic virus* in plant and animal cell-free systems and in toad oocytes. *Virology* **79**, 88–108.

88- Sclavounos, A. P., Voloudakis, A. E., Arabatzis, C. & Kyriakopoulou, P. E. (2006). A severe hellenic CMV tomato isolate: symptom variability in tobacco, characterization and discrimination of variants. *Eur J Plant Pathol* **115**, 163–172.

89- Shi, B. J, Symons, R. H, Palukaitis, P. (2008). The *cucumovirus* 2b gene drives selection of inter-viral recombinants affecting the crossover site, the acceptor RNA and the rate of selection. *Nucleic Acids Res* **36**, 1057–1071.

90- Shintaku, M. H., Zhang, L. & Palukaitis, P. (1992). A single amino acid substitution in the coat protein of *cucumber mosaic virus* induces chlorosis in tobacco. *Plant Cell* **4**, 751–757.

91- Silander, O. K., Weinreich, D. M., Wright, K. M., O'Keefe, K. J. & Rang, C. U. (2005). Widespread genetic exchange among terrestrial bacteriophages. *Proc Natl Acad Sci U S A* **102**, 19009- 19014.

92- Smith, T. J., Chase, E., Schmidt. T. & Perry, K. L. (2000). The structure of *cucumber mosaic virus* and comparison to *cowpea chlorotic mottle virus*. *J Virol* **74**, 7 578–7586.

93- Sullistyowati, E., Mitter, N., Bastiaan- Net, S., Roossinck, M. J. & Dietzgen, R. G. (2004). Host range, symptom expression and RNA 3 sequence analyses of six Australian strains of *Cucumber mosaic virus*. *Aust Plant Pathol* **33**, 505-512.

94- Suzuki, M., Kuwata, S., Kataoka, J., Masuta, C., Nitta, N. & Takanami, Y. (1991). Functional analysis of deletion mutants of *cucumber mosaic virus* RNA3 using an in vitro transcription system. *Virology* **183**,106–113.

95- Suzuki, M., Kuwata, S., Masuta, C. & Takanami, Y. (1995). Point mutations in the coat protein of *cucumber mosaic virus* affect symptom expression and virion accumulation in tobacco. *J Virol* **76**, 1791–1799.

96- Takeshita, M., Suzuki, M. & Takanami, Y. (2001). Combination of amino acids in the 3a protein and the coat protein of *cucumber mosaic virus* determines symptom expression and virus spread in bottle gourd. *Arch Virol* **146**, 697–711.

97- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* **28**, 2731-2739.

98- Tsompana, M., Abad, J., Purugganan, M. & Moyer, J. W. (2005). The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. *Mol Ecol* **14**, 53-66.

99- Vaquero, C., Turner, A. P., Demangeat, G., Sanz, A., Serra, M. T., Roberts, K. & Garcia-Luque, I. (1994). The 3a movement protein from *Cucumber mosaic virus* increases the gating capacity of plasmodesmata in transgenic tobacco plants. *J Gen Virol* **75**, 3193– 3197.

100- Wahyuni, W. S., Dietzgen, R. G., Hanada, K. & Francki, R. I. B. (1992). Serological and biological variation between and within subgroup I and II strains of *cucumber mosaic virus*. *Plant Pathol* **41**, 282-297.

101- Webster, R. G., Bean, W. J. & Gorman, O. T. (1995). Evolution of influenza viruses, p. 531–544. In A. J. Gibbs, and C. H. Calishe, and F. Garcia-Arenal (ed.), Molecular basis of viral evolution. Cambridge University Press, Cambridge, England.

102- White, P. S., Morales, F. J. & Roosinck, M. J. (1995). Interspecific reassortment in the evolution of *cucumoviruses*. *Virology* **207**, 334–337.

Chapter 3

POPULATION GENETICS AND PHYLOGENETIC ANALYSIS OF *CUCUMBER MOSAIC CUCUMOVIRUS* (CMV) FIELD ISOLATES COLLECTED IN SNAP BEAN AND PEPPER IN WISCONSIN*

*A slightly modified version of this chapter will be submitted for publication to *Phytopathology* as:

Nouri, S., Groves, C. L., and Groves, R. L. (2012). Population genetics and phylogenetic analysis of natural populations of *Cucumber mosaic cucumovirus* (CMV): Implications for the novel epidemics in Wisconsin.

Abstract

Cucumber mosaic cucomovirus (CMV) is one of the most common and economically important plant viruses with a worldwide distribution and the widest host range among all known plant viruses. Recent outbreaks of CMV, which started in 2000 in the Midwest and Northeast US, have caused significant economic losses to processing snap bean (*Phaseolus vulgaris* L.) and pepper (*Capsicum annuum* L.). Our analysis provides a genotyping profile of CMV 2b and CP genes from field isolates in Wisconsin from 2007, 2008, and 2010, and describes the phylogenetic relationships of these isolates within and among affected fields and two crop hosts in the state. The nucleotide diversity (π) was estimated at 0.055 and 0.036 for the 2b and CP genes, respectively. Phylogenetic analysis placed all selected isolates in the CMV subgroup I clade. Both subgroups IA and IB were detected among Wisconsin CMV isolates in both processing snap bean and pepper crops. The distribution of these subgroups was not correlated with their geographical location or collection year, but it was associated with the host plant. Subgroups IA and IB were detected in pepper, whereas only subgroup 1A was found in snap bean. The subgroup IA isolates collected in Wisconsin possessed a high sequence identity with other U.S. subgroup IA reference isolates including historic CMV isolates from New York, another region recently challenged with CMV outbreaks. In contrast, subgroup IB isolates collected from pepper were different from other reference subgroup IB isolates. The CP and 2b genes of the newly detected subgroup IB isolates in Wisconsin were more closely related with two other isolates originating from Illinois and China, respectively. When comparing the CMV subpopulations within and among fields, limited genetic variation was detected among isolates using the 2b and CP genomic regions. The frequency of gene flow was limited for the 2b while it was frequent for the CP gene. Intersubpopulation genetic distance of CMV isolates between two hosts was higher

than intrasubpopulation suggesting genetic differentiation between two CMV subpopulations. Neutrality tests further illustrated that negative selection was the major evolutionary force on the 2b and CP coding regions although some positively selected codons were detected inside of each gene (often evidence of interaction with a host or vector).

Introduction

Cucumber mosaic virus (CMV), the type species of the genus *Cucumovirus* in the family *Bromoviridae,* is one of the most common and economically important plant viruses with more than 1000 host species (57). CMV is transmitted by over 75 species of aphids in a non-persistent manner (58). The genome of CMV contains three positive-sense, single-stranded RNAs packaged in separate particles and two subgenomic RNAs (12, 58, 60). RNA 1 encodes the 1a protein which is a replicase subunit and which contains methyl transferase and helicase motifs (24, 26, 29, 33). RNA 2 encodes two proteins, 2a and 2b. The 2a protein encodes RNA-dependent-RNA polymerase (RdRp) (6, 32, 26, 63) which together with the 1a protein forms the viral components of the replicase complex (27). The 2b protein encoded via a subgenomic RNA derived from RNA2, is a multifunctional protein and functions in host-specific, long-distance movement, symptom induction, and as a virulence determinant by suppressing gene silencing (5, 11, 12, 13). The CMV RNA 3 encodes two proteins, 3a, a cell-to-cell movement protein (MP) (13) and 3b or capsid protein (CP); this latter protein being translated from a sub-genomic RNA 4 (71). CP is involved in cell-to-cell movement, virus assembly and aphid- mediated transmission (3, 7, 51, 61, 62, 75). Moreover, some strains of CMV carry an extra RNA molecule known as satellite RNA. This satRNA is a linear, single-stranded RNA molecule, ranging in size from 332 to 405 nucleotides (21), which can modify symptom expression induced by the helper virus, ranging from attenuation to increased severity (57, 58, 68). CMV isolates were previously placed into two subgroups, designated as subgroups I and II based on serology (58, 80), nucleic acid hybridization (56), RT-PCR followed by RFLP (65), and nucleotide sequence identity (58, 67). The two groups have an average of 75% nucleotide identity overall (67). Subgroup I is more heterogeneous than subgroup II (57). Analysis of the CP gene and 5` non-translated region (NTR) of RNA3 led to further division of subgroup I into IA and IB with

92-95% nucleotide identity among strains across these groups (67, 68). It has been shown that estimated trees for different Open Reading Frames (ORFs) located on the different RNAs are not congruent, suggesting that different RNAs may have independent evolutionary histories (67). Subgroups IA and II are found worldwide in distribution (52), while the majority of subgroup IB isolates are principally restricted to East Asia (67). Subgroup IB was recently confirmed to be in Hawaii and California (45, 68). Hence, CMV is a heterogenic group with significant variation among isolates documented in different crops and geographic locations and as an RNA virus, it able to rapidly evolve in unique environments with shifting selection pressures. (19, 39, 45, 46, 66, 67, 74).

CMV has been endemic in many parts of the United States for decades. Beginning in 2000, however, a significant increase in virus outbreaks emerged in the processing snap bean (*Phaseolus vulgaris* L.) industry in Wisconsin and later became prevalent in this crop in other parts of Midwest and Northeastern U.S. (e.g. New York and Pennsylvania) (22, 25, 41, 53). Succulent snap bean is grown as a canning crop in the upper Midwest and is considered one of the most important commercially grown vegetables in the U.S. with an estimated $161 M dollar in revenue (Natural Agricultural Statistics Service, 2011) and with Wisconsin as a domestic leader on total production. Moreover, in 2007, the first among a series of CMV outbreaks was observed in processing pepper (*Capsicum annuum* L.) fields in Wisconsin (Thomas L. German, University of Wisconsin, personal observation). This has become a serious concern because observed disease incidence approaching 100% had been observed in some snap bean fields (54). Explanations for the emergence of these novel virus outbreaks are not well known.

Changes in the population structure of viruses, or virus diversity, has been considered as a significant factor in the appearance of novel virus epidemics. However, the spatial structure of

CMV genetic diversity, especially as it relates to plant host type, has not been fully investigated. Hence, we implemented a molecular, phylogenetic analysis of CMV isolates based on two important genomic regions, the 2b and CP genes, collected from selected field sites over the course of a three year study. We chose the 2b gene because it is a multifunctional gene and is involved in host-specific symptom induction (14, 43). Moreover, the presence of different subgroups of CMV in the selected field locations can be determined by comparing CP sequences among isolates. Two susceptible crops, succulent snap bean and pepper, were included in the study as these were the two major virus-affected crops observed in outbreaks early in the 2000's.

Materials and Methods

Virus isolates

In total, nine succulent snap bean fields and five pepper fields were sampled in 3 years. Three snap bean and two pepper fields were sampled in both 2007 and 2008. In 2010, our collection was limited to three snap bean and one pepper field. All sample fields were located in central Wisconsin and specifically in the irrigated processing vegetable growing region known as the Central Sands. All symptomatic plant samples in each year were collected between August and October in 2007, 2008 and 2010. Specifically, a total of 25 symptomatic plants were collected from each field. From among these field / year combinations, four plants, each showing a very consistent set of foliar symptoms, were further subsampled from each field. Individual fields, host plants and collection year represented a potentially different population in this study and were subsequently termed sub-populations during analysis. From each plant, single trifoliate leaf possessing disease symptoms including mottling, blistering and mosaic were sampled. Collected leaves were immediately processed after sampling and the remaining foliage was kept in -80 °C ultra-cold freezer.

RNA extraction and RT- PCR amplification

Total RNAs were extracted from leaf samples using an RNeasy Mini kit (Qiagen, CA, USA) according to manufacturer's instructions. Two sets of primers (45) were used in one-step RT-PCR to amplify the 2b and CP genes. Amplification was carried out in a 25 µl volume of 1X Go Taq Flexi DNA polymerase reaction buffer (Promega, WI, USA), containing 0.2 mM dNTP mix, 5 mM DTT, 1.1 mM MgCl2, 10 U of SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad ,USA), 0.75 U of Go Taq Flexi DNA Polymerase (Promega, WI, USA) and 12.5 mM of each primer. The PCR program was initiated at 53˚C for 45 minutes, followed by 25 cycles at 94˚C for 30s, 54˚C (CP) and 51 ˚C (2b) for 30s, 72˚C for 1 minute and finished by 72˚C for 7 minutes. Products were separated on a 1.2% agarose gel to initially screen for the presence of product.

Sequencing

Observed bands of the appropriate size for the 2b (370 bp) and CP (675 bp) genes, respectively, were excised and purified using a QIA quick gel extraction (QIAGEN, CA, USA). Purified products for each gene / isolate combination were bi-directionally sequenced using a BigDye termination sequencing kit and gene specific primers in an automated DNA sequencer (ABI Prism 310 Genetic Analyzer, Applied Biosystem) in the DNA Sequencing Facility at Biotechnology Center of University of Wisconsin-Madison.

Sequence alignment, phylogenetic analysis and estimation of nucleotide diversity

Consensus sequences of each gene / isolate combination obtained using the DNASTAR (Madison, WI, USA) were later used for phylogenetic analysis. Multiple nucleotide sequence alignments were performed using CLUSTALW in MEGA 5 (78). The sequence alignments were

cleaned manually. Phylogenetic trees were constructed for each genomic segment by the neighbor-joining (NJ) method (70) using the MEGA 5 program with 1,000 bootstrap replicates under the Kimura 2- parameter model (34). Bootstrap consensus trees were rooted using *Peanut stunt virus* (PSV) and *Tomato aspermy virus* (TAV) as the out-groups based on the current knowledge of the *Bromoviridae* (8). All branches with <70% bootstrap support were judged inconclusive and were collapsed (15, 28). For further analyses, reference isolate sequences as representatives of CMV subgroups IA, IB and II, were downloaded from the NCBI Nucleotide Database. We also used some previously recognized CMV isolates from the US (Chapter 2) to compare with the current WI isolates. A portion of these extra selected isolates were representatives of CMV from other geographic regions with the recent history of virus outbreak. The list of the reference CMV sequences obtained from GenBank has been illustrated in Appendix 3. The overall nucleotide diversity (π) among isolates and other genetic parameters for each gene were determined using the DnaSP version 5.1 (69).

Measuring genetic differentiation

To measure the genetic variation of CMV subpopulations within and among fields, estimates of nucleotide diversity were calculated within and among fields using DnaSP 5. To test if CMV subpopulations between fields were genetically different, three independent permutation tests, K_S*, Z and Snn, which represent the most powerful sequence-based, statistical tests for genetic differentiation, with 1,000 permutation replicates implemented in DnaSP for the 2b and CP genes (30, 31). K_S* is the weighted average of the nucleotide differences between two sequences from two subpopulations and it is close to zero under the null hypothesis (no genetic differentiation). The Z or rank statistic is the weighted sum of two assigned ranks obtained from a matrix of pairwise number of differences for all pairs of sequences. Snn is the frequency of

finding the nearest neighbor of sequences in the same locality. A Snn value is in the range of 0 to 1 for undifferentiated to high subpopulation differentiation, respectively (30). When taken together, the null hypothesis is rejected if estimates for K*, Z and Snn resulted in corresponding P- values < 0.05. Also, DnaSP was used to estimate another parameter, F_{ST}, fixation index, a measure of the proportion of the total genetic variance contained in a subpopulation relative to the total genetic variance (83). F_{ST} values range between 0 and 1, with respect to undifferentiated to fully differentiated subpopulations. Furthermore, the level of gene flow can be estimated by F_{ST}, with an absolute value > 0.33 normally used as a standard for infrequent gene flow.

In order to identify genetic diversity of subpopulations on different host species and year of collection, mean genetic distance was estimated within and between hosts and years using MEGA 5 with 500 replicates. Pairwise F_{ST} values between hosts and also years were used to estimate genetic differentiation between subpopulations by the host species and collection year.

Natural selection analysis

To assess any type of natural selection pressure on either the 2b or CP genes of the representative CMV isolates, non-synonymous (dN) and synonymous (dS) substitution rates and their associated ratios (dN/dS= ω) were estimated for each segment by using the bootstrap method with 500 replicates in MEGA version 5. To determine the site specific selection pressure in each region, three complementary maximum-likelihood methods including single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL) (37, 38) were implemented in the Hyphy package available at the Datamonkey web server (http://www. Datamonkey.org). To classify a site as positively or negatively selected, the cutoff P-value was established as 0.1 for SLAC and FEL. For REL, a Bayes factor of 50 was selected as the cut-off value. The most appropriate nucleotide substitution models were selected for each gene by the

software and only selections determined to be significant by at least two methods were considered as a significant positive.

Results

Genotype profile and phylogenetic relationship of CMV field isolates

The molecular signature of selected CMV field isolates in WI (Table 1) was determined based on upon two genomic elements. To provide a genotype profile for CMV isolates in this area, we performed phylogenetic analyses of fifty-three and forty-six isolates focusing on the 2b and CP genes collected from different fields in the Central Sands vegetable production region. Nucleotide sequences of both genes were aligned and used to construct NJ trees with 1,000 bootstrap replications and a Kimura 2-parameter model using MEGA (Fig. 1). As illustrated in Figure 1, selected CMV isolates formed two main clades, with high bootstrap values. When including both genes, the majority of isolates fell inside of clade CMV-1. Sequences of selected isolates in the current study were compared with those of other available CMV isolates. This analysis placed all Wisconsin collected CMV isolates in the subgroup I designation and further illustrated that isolates in clade CMV-1 (Fig. 1) were a combination of host associations of snap bean and pepper belonging to subgroup IA (Fig. 2). All Wisconsin collected isolates in clade CMV-2 (Fig. 1) were from pepper and belonged to subgroup IB of CMV (Fig. 2). No association was documented between phylogenetic groups and geographic locations or collection years. Wisconsin subgroup IA isolates showed 97% and 96% sequence identity when compared to six, contemporary (2009) and five historical New York snap bean isolates, respectively, based on the 2b gene. This sequence identity was 97% and 95% for the CP gene, respectively.

The 2b and CP phylogenetic trees constructed with the Wisconsin sequences and reference CMV isolates, grouped subgroup IB isolates, in the current study, separately from the PV29 isolate

belonging to the same subgroup isolated from cucumber in 1975 from Wisconsin (Fig. 2). According to these trees, the newly detected subgroup IB Wisconsin sequences demonstrated a close phylogenetic relationship with 2A1IL isolate (Illinois) collected in 1999 from pepper (Fig. 2b) and an SD isolate submitted from China (Fig. 2a), when comparing the CP and 2b genes, respectively.

Genetic diversity of selected CMV isolates and genetic differentiation between subpopulations

To investigate the genetic diversity among selected CMV isolates in Wisconsin, population genetic parameters for the 2b and CP genes were estimated using DnaSP 5 (Table 2). Overall, π values (nucleotide diversity, average number of nucleotide differences per site between two sequences) for our CMV isolates ranged between 0.055 and 0.036 for the 2b and CP genes (Table 2). The mean nucleotide diversity was 0.017 and 0.014 when comparing isolates within fields based on the 2b and CP genes, respectively (Tables 3 and 4). However, three pepper fields (JP08, CAP07 and E3P10) demonstrated higher within field nucleotide diversity based on both genes (Tables 3 and 4). This result could be related to the unique composition of CMV in these three specific fields, which was composed of both subgroups IA and IB (Fig. 1 and 2), whereas CMV isolates in other fields exclusively belonged to either subgroup IA or IB. To measure the genetic variation among fields, pairwise nucleotide diversity (π) was estimated using DnaSP and these estimates ranged between 0.000 and 0.067, plus 0.000 and 0.052 for the 2b and CP genes, respectively (Tables 3 and 4). Statistical tests, K_S*, Z, Snn and F_{ST} were applied to test the hypothesis that subpopulations between fields were genetically differentiated. The null hypothesis (no genetic differentiation) was rejected for less than half of the field comparisons (33 out of 91 comparisons ~ 36.2% for 2b and 18 out of 91 ~ 19.7% for CP) for both genes

supported by P- values less than 0.05 (Tables 3 and 4). F_{ST} values ranged between 0 and 0.75, but most comparisons (55 out of 91 ~ 60%) exceeded 0.33 for the 2b gene (Appendix 2). In contrast, more than half of the comparisons (51 out of 91 ~ 56%) resulted in F_{ST} values < 0.33 for the CP gene. Mean values of F_{ST} ranged between 0.36 and 0.28 for the 2b and CP genes, respectively. These values suggested a limited and frequent gene flow for the 2b and CP genes, respectively. In order to measure the genetic diversity of the CMV subpopulation between the two different host species (snap bean and pepper), within and between mean distances were calculated using MEGA with 500 bootstrap iterations under the Kimura 2-parameter model (Table 5a). Estimated within host mean distances was low in snap bean and higher in pepper for both genomic elements (Table 5a). Between subpopulation diversity values were greater in their order of magnitude to those corresponding to within subpopulation diversity values (Table 5a). This indicates that there is differentiation of the population according to the host species. Pairwise F_{ST} values between these two host species was also applied to test if CMV subpopulations were genetically differentiated. F_{ST} values between these two hosts ranged between 0.53 and 0.46 for the 2b and CP genes, respectively, which is sufficiently high to differentiate CMV subpopulations between the two hosts. Phylogenetic analysis confirmed this result in which subpopulations of CMV in pepper and snap bean were not fully differentiated based on the 2b and CP phylogenetic trees. A total of 12 out of 19 isolates collected from pepper formed a cluster in the 2b tree; however, the remaining seven isolates were grouped with other snap bean isolates (Fig. 1a). This clustering remained consistent regarding the CP phylogenetic tree (Fig. 2a). CMV isolates in this study were collected during 2007, 2008 and 2010. To evaluate genetic differentiation of CMV subpopulations between these three years, we considered subpopulation as those isolates that were sampled during a given growing season and

genetic distance values were estimated within and between years (Tables 5b and 5c) as previously described for host species. Intrapopulation diversity was similar for 2007 and 2008, whereas 2010 subpopulations illustrated the lowest genetic diversity. Interpopulation diversity values were similar to estimated intra-subpopulation values for 2007 and 2008, suggesting no genetic differentiation between these two years (Tables 5a and 5b). In contrast, computed genetic diversity between 2010 and the two other included years was greater than within subpopulation diversity (Tables 5a and 5b). This indicates that CMV subpopulations in 2010 were genetically differentiated more so than 2007 and 2008. Estimated Fst values were 0 when comparing sample years 2007 and 2008, suggesting that CMV subpopulations in these two years were similar (Tables 5a and 5b). Calculated Fst values between subpopulations of 2010 and the other years confirmed the result obtained from genetic diversity comparisons and further illustrated that there was some degree of differentiation between 2010 and CMV subpopulations in both remaining years for both genomic regions (Tables 5a and 5b).

Overall, F_{ST} and test statistic estimates illustrated that CMV subpopulations among Wisconsin fields, based on the 2b and CP genes, were not genetically differentiated. Moreover, phylogenetic analysis and F_{ST} values suggested genetically differentiated CMV subpopulations between the two host species examined in a common geographic area (e.g. Central Sands region). The genetically differentiated CMV subpopulation in 2010 was additionally confirmed by within and between-subpopulation mean genetic distance comparisons and also Fst values.

Selection pressure on the two coding regions

The genetic distances at non-synonymous (dN) and synonymous (dS) sites and the dN/ dS ratios were estimated for the 2b and CP genes to determine the direction of selective constraints imposed on the coding regions (Table 2). The genetic distance at non-synonymous sites between

isolates was higher in the 2b gene (0.047) than CP (0.009). The mean ω values estimated from pairwise comparisons between sequences were less than 1 for both 2b and CP genes suggesting overall negative (purifying) selection imposed on these coding regions. However, complementary maximum- likelihood methods (SLAC, FEL and REL) implemented using the Hyphy package under HKY85 model could recognize single sites (codons) inside of coding regions potentially undergoing positive selection. Data in Table 6 illustrates the position of these positively selected areas within 2b and CP genes. Site 24 (Q → L) in the 2b gene was accepted as the positively selected codon most strongly recognized by all three methods, while Sites 3 (L → I) and 80 (P → S) in the 2b and codon 25 (S → P) in the CP were detected by two methods with high supporting values. The positively selected amino acid change at sites 24 and 3 of the 2b was present in 14 and 6 isolates, respectively, collected from snap bean and phylogenetic group 1 (CMV IA) in various years and fields. A change at the site 80 of the 2b protein was detected in 17 isolates from snap bean and pepper (both subgroups) collected in 2007 and 2008. This latter change in pepper was present in only subgroup IB. In the CP coding region, site 25 was positively selected in 31 isolates from among different hosts, years and both phylogenetic groups.

Discussion

We surveyed the genetic diversity and phylogenetic relationship of CMV isolates in Wisconsin collected from different fields, years and two host plants. A distance-based NJ method was applied to construct the phylogenetic trees based on the nucleotide sequences of the 2b and CP genes. Overall, CMV field isolates in Wisconsin demonstrated low genetic diversity, with π estimates of 0.055 and 0.036 for the 2b and CP genes, respectively. Our results are consistent with previous studies investigations of natural CMV populations in Spain and California

illustrating a low genetic diversity (17, 45, 46). Additionally, Wisconsin CMV isolates included in this investigation fell into subgroup I which was in agreement with the results obtained from the previous studies which showed a higher incidence of subgroup I more so than subgroup II (10, 17, 19, 46, 66, chapter 2 of this thesis). One possible reason for not detecting any subgroup II isolates among our field collections could be related to the different adaptive responses of these two subgroups to temperature. Specifically, it has been shown that CMV subgroup II has a lower thermal optima than subgroup I and this latter subgroup is predominate in collections obtained during the summer (9, 17), the season in which isolates were collected in the current investigation. Taken together, the structure of field CMV isolates in Wisconsin was comprised of a combination of subgroup IA and IB. However, the distribution of these two subgroups was different between two hosts. Our data showed that all isolates collected from snap bean belonged to subgroup IA, whereas isolates from pepper were more diverse, composed of both subgroups IA and IB with the latter being predominant (Figs. 1a and 2a). It has been shown previously that mixed infections of subgroups IA and IB were common in affected tomato fields in Spain (4, 17). No significant association was seen between phylogenetic groups and other factors such as geographic region or collection date, which was again consistent with previous investigations (19, 45, 47). In contrast, an observed association between phylogenetic subgroups and host plant was noted, and this observation is inconsistent with previous studies in Spain and California, USA (19, 45). A question that emerges here centers on the variation in genetic diversity and specifically subgroups, explained by host plant. Specifically, why is the population of CMV in snap bean composed of only the specific subgroup IA, with low intra-host genetic variation, whereas populations in pepper were more diverse and containing both subgroups. One possible explanation is related to host-virus strain specificity. There is evidence suggesting that specific

strains of CMV, termed “legume strains”, can cause systemic infection in legumes such as bean (16, 34, 40, 50, 64, 82). Although no investigation has been performed to determine the genetic profile(s) of CMV legume strains, and furthermore there is no evidence to suggest that subgroup IB is unable to cause infection in succulent snap bean.

Another possible explanation for the extremely low genetic diversity among CMV isolates in snap bean, and furthermore the existence of only a single subgroup (IA) in this crop, could result from the ‘founder effect’ which has previously been suggested to play a role in shaping the genetic structure of CMV in California and Spain (17, 45). Founder effect is often invoked to explain the low genetic diversity of certain populations of various plant viruses, including the *rice yellow mottle virus* population in western Africa, *rice strip virus* in China and *cucurbit yellow stunting disorder virus* in Spain (20, 48, 81). Introduction or selection of a relatively few CMV variants in a discrete field, area or region could result in a lack of significant genetic variation. Genetic bottlenecks may contribute to the founder effect and reduced genetic variation ultimately leading to a lack of intra-population differentiation. Genetic bottlenecks during CMV systemic movement in the host as well as during transmission by aphid vectors have previously been reported (1, 44). However, we cannot rule out the possibility of the presence of both subgroups IA and IB in snap bean fields in the past, with only the emergence of subgroup IA recently. Furthermore, we included only affected plants showing typical CMV symptoms in this study. Hence, it is possible that subgroup IB in snap bean was among asymptomatic, non-included plants.

We did observe a high level of sequence identity (98%) between the legume subgroup IA CMV isolates in snap bean and the same subgroup in pepper and phylogenetic trees clustered them together (Fig. 1a and 2a). Although, the legume strains of CMV can cause systemic infection in

legumes, these strains have also been isolated from pepper (59), spinach (18) and lettuce (79). Therefore, our observation of a high level of similarity between subgroup IA isolates in these two hosts is not surprising.

Moreover, it was interesting to observe that the CP phylogenetic tree classified subgroup IB isolates, recognized in the current survey in pepper, separately from other available CMV subgroup IB isolates in other parts of the world and as well as the 2A1IL isolate from pepper in Illinois (55). In the case of the 2b gene, the sequence of the 2A1IL isolate was not available in GenBank for comparison, but a close phylogenetic relationship was determined between Wisconsin subgroup IB and the SD-CMV isolate from China, suggesting the potential for a long-distance migration (gene flow) between these two geographic regions for this specific gene. Furthermore, a high level of sequence identity was found between six CMV subgroup IA isolates collected from snap bean fields in 2009 in New York, another region with a recent history of virus outbreak and recently collected, Wisconsin subgroup IA isolates. Although, we did not have historic subgroup IA CMV isolates from Wisconsin, we were able to compare these to the historic, New York isolates. Sequence comparisons demonstrated a range of 95-97% identity between the CMV isolates recently collected from snap bean in Wisconsin and New York, to the historic New York isolates, respectively. Hence, these results suggest that the recent virus outbreaks in selected regions of U.S. may have not been due to invasion of a new subgroup IA serotype.

It is worthwhile to note here that coincident with the increasing incidence of the virus epidemic in the selected regions, the soybean aphid (*Aphis glycines*), a novel and highly efficient vector for CMV (23, 53), was introduced to the United States (and specifically the State of Wisconsin) from Asia in 2000. Although there is insufficient evidence to demonstrate that there may be

vector selection pressure for specific subgroups of CMV, it has been shown that there is some degree of specificity for transmission of this virus by other aphid species (61). Similar biases in transmissibility have been reported for other plant virus–vector systems. For example, whitefly *Bemisia tabaci* populations from various geographical locations had different transmission efficiencies with different isolates of *tomato leaf curl geminivirus* or other geminiviruses (2, 49). Hence, one proposed scenario for the recent virus epidemics in select regions of US is that when the first CMV outbreak occurred in Wisconsin, one or a few genetic types established an initial population and this genetic type (subgroup) may have been selected by the soybean aphid and was subsequently distributed rapidly and widely due to the tremendous number of aphids during dispersal. Another plausible explanation might be a constriction, or genetic bottleneck, which may happen at the end of the growing season. Specifically, the virus may successfully overwinter in only a very discrete set of biennial or perennial host plants, and in the following growing season, a very wide range of crop hosts are again recolonized from this reservoir. Marco and Arenda (2005) proposed a similar scenario for the *Cucurbit yellow stunting disorder virus* epidemics in Spain. However, in this instance, genetic bottlenecks resulting from host plant constraints during overwintering may be unlikely because of the year-round maintenance of the vector and plant populations (48). Further research would be required to link this aphid with the observed subgroup distributions in the current study.

The genetic variation (nucleotide diversity) within fields was very low for both genes suggesting that the CMV population in individual sample fields could have been derived from a single origin and a predominate haplotype. However, isolates from three pepper fields (JP08, CAP07 and E3P10) showed genetic variation among isolates in the same sample field location (Tables 3 and 4). This result was not very surprising, because isolates collected from these two fields

belonged to both subgroups IA and IB (Figs. 1a and 2a) and these two subgroups of CMV show 5-8% nucleotide sequence difference (67, 68). Overall, genetic differentiation of CMV subpopulations among fields for both genes was not supported in most comparisons. However, this differentiation for 2b was seen approximately twice as often as the CP gene. This result was consistent with the estimated F_{ST} values illustrating that gene flow among fields for the CP gene was more frequent than the 2b.

Although the majority of CMV isolates sampled from pepper were grouped together and formed a cluster, phylogenetic trees did not place these CMV isolates sampled from the two host species into completely separate clades. A few CMV isolates from pepper were grouped with isolates in snap bean. It shows that the CMV subpopulations in these two hosts cannot by fully differentiate among the host types. This degree of genetic differentiation is also consistent with the limited gene flow (F_{ST} value=0.53 and 0.46) between these two subpopulations. Estimated genetic diversity for CMV subpopulations by year illustrated that the 2010 subpopulation had the lowest genetic diversity among all three years. One possible reason for this low amount of variation could be related to the number of CMV subgroup IB isolates collected from pepper fields in this particular year. We had two subgroup IB versus 11 subgroup IA isolates in the 2010 CMV subpopulation and it was previously shown that subgroup IA isolates included in this study were genetically uniform.

Comparison of synonymous to non-synonymous substitutions and calculated ω values showed that the CMV population in Wisconsin fields is experiencing purifying selection upon both genes, especially CP as has been shown in other studies for this virus (45, 52). Although, negative selection has been demonstrated as the predominate constraint imposed on coding regions during the evolution of plant viruses and purifying selection is due to both internal and

external constraints (20), there is some degree of specificity for both transmission and infectivity by some plant viruses such as CMV (36, 61, 72, 76, 77). CMV shows a broad host range and can be transmitted by a large number of aphid vectors. Hence, in order to adapt to this level of host and vector variation, we expected to see some degree of positive (diversifying) selection in some codons in the CMV coding regions. In a previous study to identify positively selected amino acid changes in different coding regions of CMV, no evidence of positive selection was found in the 2b gene (52), whereas this was detected in three positively selected sites in the 2b gene in the current study (Table 6). The 2b protein is a multifunctional protein involved in long-distance movement, virulence and is a suppressor of RNA silencing (5, 11, 12, 13). Hence, this protein interacts directly with the host and there is evidence that this interaction is host specific (14). This host-specific function, coupled with the extremely wide host range of CMV, allow this protein to be considerably more tolerant to nucleotide and amino acid changes. Positively selected amino acids 24 and 3 in the 2b gene have been located in NLS1 (nuclear localization signal 1) and the N-terminal region of the protein which are required for symptom induction, respectively (42, 43). RNA silencing suppression is another assumed function for the NLS region of the CMV 2b protein (43). It suggests that host- virus interaction plays an important role in selection and adaptation. Although, we could not find any information about the specific function of codon 80 in the 2b coding region of CMV in the current literature, it seems that amino acids located in this region are involved in an important function because CMV isolates from California showed a greater average number of non-synonymous mutations than synonymous in a short region of 2b (codons 81-93) and close to our detected site (45). It is noteworthy that the majority of isolates exhibiting positive selection at sites 3 and 24 were collected from snap bean, while site 80 was detected among subgroup IA isolates collected from

both snap bean and pepper. It may suggest some degree of host adaptation selection pressure for CMV selected isolates. For the CP gene, our approach found site 25 under positive selection. This result confirmed the previous report showing diversifying selection for this codon (52). This amino acid is located in the folded region of the CP (73) and it affects virus transmission by the aphid vector (61). Hence, a positive selection pattern in this codon of the CP gene could be related to the role of different aphid species in selection of different virus variants (52).

In conclusion, CMV displays low genetic variation in Wisconsin processing vegetable crop fields for the 2b and CP genes. Our analyses provided a genotyping profile of CMV field isolates in the state, and this allowed us to investigate the genetic composition of CMV populations in this agriculturally important region. Subgroups IA and IB were prevalent in the subset of CMV isolates sampled, but the distribution of subgroup IB appears to be host-specific. Therefore, further investigation of more CMV isolates in other host plants in this geographic region can generate more information about the composition and genetic structure of the virus. Purifying selection was determined as the predominant evolutionary force acting on the 2b and CP genes, although certain amino acids under positive selection could be detected. Our data also confirmed mixed infections of subgroups in pepper fields which can lead to the possibility for reassortment. Hence, assessments of the phylogenetic relationships among CMV isolates in other genomic segments of CMV for the selected isolates in this study will provide valuable information about the most important genetic variation and evolutionary source in this segmented virus.

Table 1. CMV isolates collected in Wisconsin through the experimental interval 2007, 2008 and 2010.

Field ID	Number of isolates for 2b/CP	Host Plant	Collection Date	Farm Identity
CA	4/3	Pepper	2007	Heath
BD	4/4	Snap bean	2007	Hartung Bros.
DB	4/4	Snap bean	2007	Hartung Bros.
DA	4/3	Snap bean	2007	Affeldt
HP	3/3	Pepper	2007	Heath
Y	3/3	Snap bean	2008	Hartung Bros.
C	4/3	Pepper	2008	Heath
B	3/3	Snap bean	2008	Seneca
J	4/4	Pepper	2008	Heath
H	4/4	Snap bean	2008	Hartung Bros.
K24	4/3	Snap bean	2010	Hancock AES
SC2	4/3	Snap bean	2010	Seneca
S1C	4/3	Snap bean	2010	Seneca
E3	4/3	Pepper	2010	Heath

Table 2. Population genetic parameter estimates for Wisconsin CMV isolates.

Genomic region	Number of isolates	S[1]	η [2]	π[3]	Θ_w[4]	dS[5]	dN[6]	ω (dN/dS)[7]
2b	56	49	49	0.055	0.035	0.093±0.019	0.047±0.010	0.505
CP	46	78	80	0.036	0.027	0.108±0.014	0.009±0.003	0.083

[1]S: Total number of segregating sites
[2]η: Total number of Mutations
[3] π: Nucleotide Diversity, average pairwise nucleotide difference per site
[4]Θ_w: Mutation rate estimated from S
[5]dS: The average number of pairwise differences per synonymous site
[6]dN: The average number of pairwise differences per non-synonymous site
[7]ω (dS/dN) were estimated by Kumar method

Table 3. Nucleotide diversity (π) within (in bold) and between CMV subpopulations for the 2b gene by field.

An asterisk indicates sample pairs for which a null hypothesis of no genetic differentiation was rejected (P<0.05) using genetic differentiation estimates by Ks*, Z and Snn implemented in DnaSP program.

Field	J	Y	H	DA	BD	C	DB	S1C	B	K24	CA	HP	SC2	E3
J	**0.067**													
Y	0.043	**0.000**												
H	0.046*	0.013	**0.000**											
DA	0.041	0.010	0.017	**0.014**										
BD	0.043*	0.006	0.015	0.015*	**0.008**									
C	0.067	0.068	0.076*	0.070*	0.074*	**0.001**								
DB	0.042*	0.005	0.018	0.013	0.005	0.070*	**0.002**							
S1C	0.040	0.013	0.013*	0.017	0.017	0.077*	0.016	**0.012**						
B	0.044	0.006	0.019*	0.015	0.009	0.072*	0.006	0.014	**0.007**					
K24	0.037	0.006*	0.018*	0.013	0.008	0.074*	0.007	0.014	0.006	**0.008**				
CA	0.071	0.065	0.068*	0.063	0.064*	0.031*	0.061	0.068	0.069	0.061*	**0.053**			
HP	0.072	0.073*	0.078*	0.073*	0.076*	0.000	0.072	0.080*	0.077	0.075*	0.034	**0.000**		
SC2	0.035	0.009	0.014*	0.013*	0.013*	0.076*	0.014	0.010	0.010	0.011	0.067	0.078*	**0.000**	
E3	0.059	0.040	0.046*	0.041	0.041	0.072	0.039*	0.042	0.042	0.038	0.065	0.076	0.038*	**0.067**

Table 4. Nucleotide diversity (π) within (in bold) and between CMV subpopulations for the CP gene by field.

An asterisk indicates sample pairs for which a null hypothesis of no genetic differentiation was rejected (P<0.05) using genetic differentiation estimates by Ks*, Z and Snn implemented in DnaSP program.

Field	J	Y	H	DA	BD	C	DB	S1C	B	K24	CA	HP	SC2	E3
J	**0.037**													
Y	0.030	**0.011**												
H	0.035*	0.022*	**0.001**											
DA	0.038	0.023	0.015*	**0.028**										
BD	0.020*	0.012*	0.020*	0.020*	**0.001**									
C	0.042	0.046	0.045	0.051	0.041	**0.000**								
DB	0.024*	0.005*	0.020	0.019*	0.008	0.040	**0.000**							
S1C	0.030	0.015	0.027*	0.029	0.013	0.047	0.010*	**0.020**						
B	0.025	0.012	0.023*	0.024	0.006	0.044	0.007*	0.014	**0.010**					
K24	0.022	0.012	0.022	0.024	0.003	0.043	0.008	0.013	0.005	**0.001**				
CA	0.043	0.043	0.043*	0.050	0.037*	0.022	0.034	0.045	0.041	0.041	**0.044**			
HP	0.042	0.048	0.046	0.053	0.042	0.001	0.041	0.048	0.045	0.044	0.023	**0.002**		
SC2	0.022	0.011	0.021	0.023	0.002	0.042	0.007	0.012	0.005	0.000	0.040	0.044	**0.000**	
E3	0.038	0.032	0.037*	0.042	0.025	0.038	0.024*	0.034	0.029	0.028	0.044	0.039	0.027	**0.052**

Table 5. Within- and between-subpopulations genetic distance (D) for the 2b and CP genes of CMV by host species and year.

A subpopulation here is considered as a set of isolates from a given host species (Snap bean and Pepper) (a) or collected during a growing season (years 2007, 2008 and 2010) (b is for 2b gene and c for CP gene). Genetic distance refers to the average number of nucleotide substitutions between two randomly selected sequences in a population estimated by 1,000 bootstrap replicates under the Kimura 2-parameter model in MEGA5 program. F_{ST} values were calculated using DnaSP 5. F_{ST} values range between 0 and 1 for undifferentiated to fully differentiated populations. Also, the level of gene flow can be estimated by F_{ST}. The absolute value > 0.33 is normally used as a standard for infrequent gene flow (S = Snap bean, P = Pepper).*Estimated F_{ST} values.

a

Gene	D			F_{ST}
	Within S	Within P	Between S and P	Between S and P
2b	0.017±0.004	0.075±0.011	0.097±0.015	0.53
CP	0.020±0.004	0.039±0.006	0.055±0.007	0.46

b

Year	2007	2008	2010
2007	0.068±0.011	0.069±0.010/0.000*	0.061±0.010/0.200*
2008		0.074±0.011	0.058±0.009/0.112*
2010			0.029±0.005

c

Year	2007	2008	2010
2007	0.043±0.006	0.042±0.005/0.000*	0.034±0.004/0.116*
2008		0.044±0.006	0.035±0.004/0.084*
2010			0.021±0.03

Table 6. Codon positions of the 2b and CP coding regions of Wisconsin CMV isolates affected by positive selection.

Positively selected codons were determined using three maximum-likelihood methods. *p*- value for SLAC and FEL was 0.1 and the minimum Bayes factor value was set at 50. *Sites were detected as statistically significant by two methods. ** Sites were detected as statistically significant by all three methods.

Coding region	Site	SLAC dN-dS	SLAC p-value	FEL dN-dS	FEL p-value	REL dN-dS	Bayes Factor
2b	3*	43.915	0.126	37.813	0.024	2.741	1527.270
	24**	60.396	0.085	77.468	0.028	2.709	667.389
	80*	28.244	0.296	18.551	0.098	2.698	234.077
CP	25*	15.258	0.296	24.209	0.077	2.074	4399.340

Figure 1. Bootstrap majority rule (70%) consensus trees for the 2b and CP genes of Wisconsin CMV isolates in snap bean and pepper fields.

Constructed phylogenetic trees for the 2b (a) and CP (b) genes using the neighbor-joining (NJ) method with 1000 bootstrap replicates under the Kimura 2-parameter model implemented in MEGA 5. Phylogenetic trees were rooted using *Peanut stunt virus* (PSTV) and *Tomato aspermy virus* (TAV) as out-groups. CMV subgroups have been indicated by vertical brackets (P = pepper; S = snap bean; Numbers at the end of taxon represent collection year).

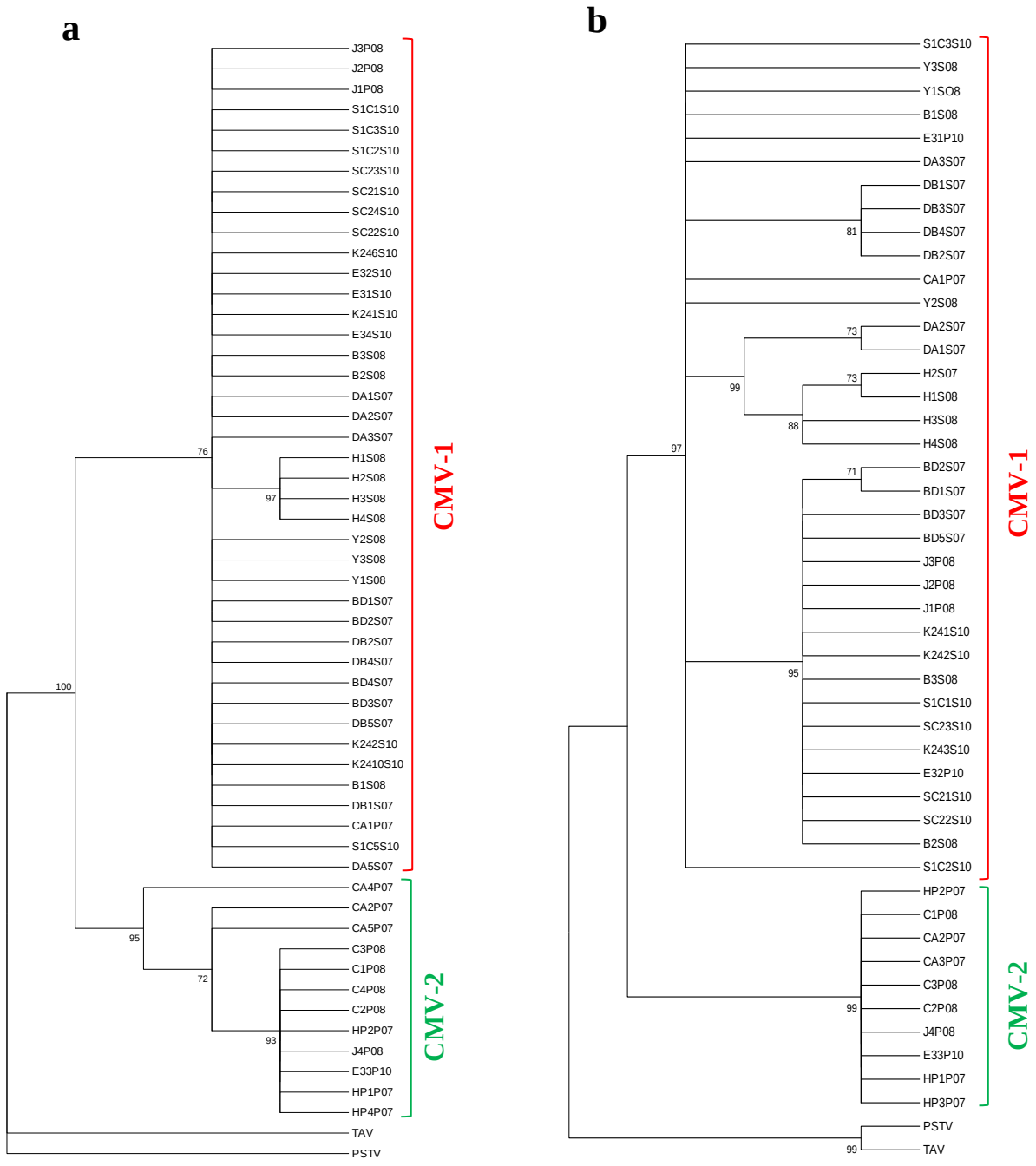
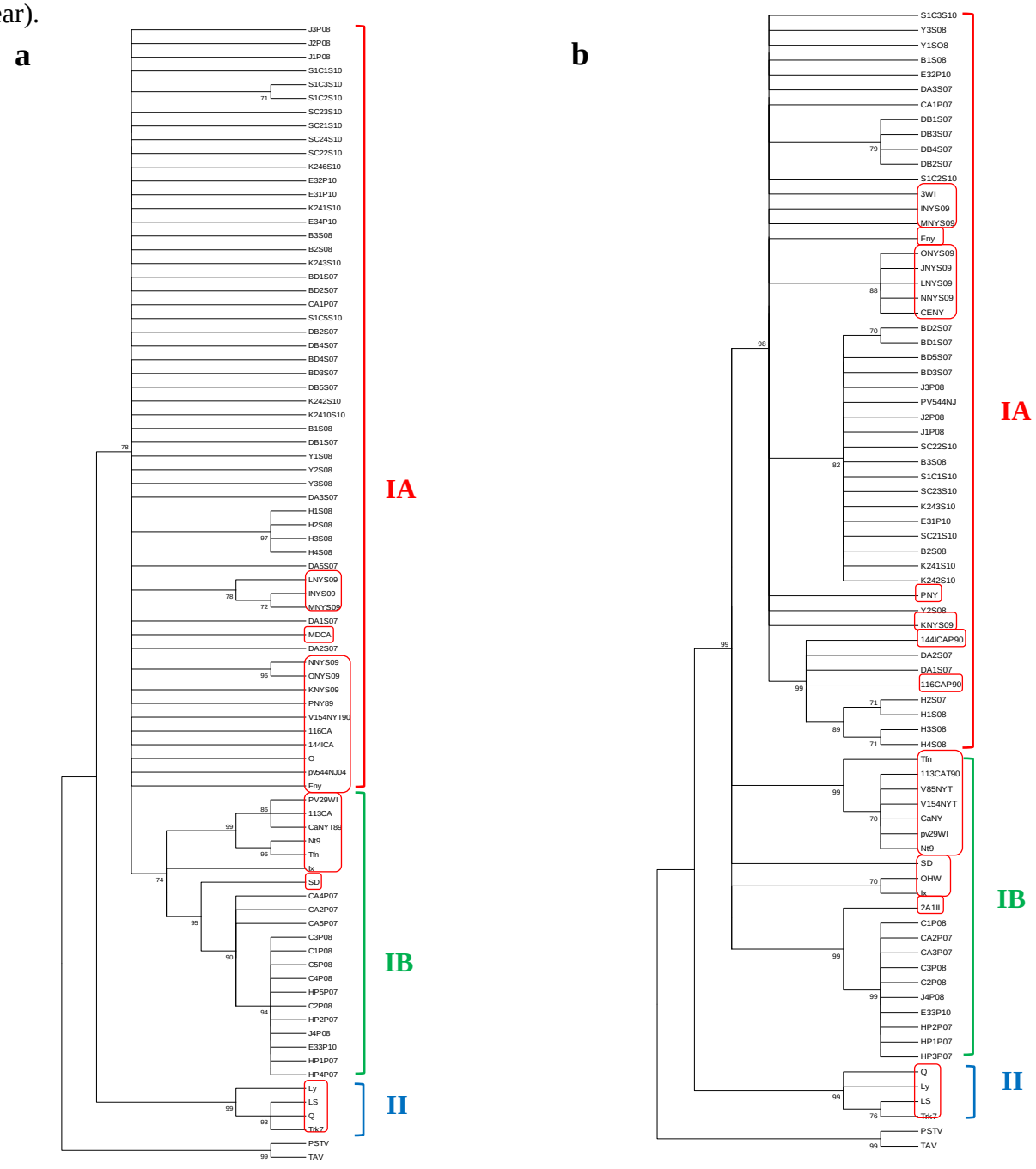


Figure 2. Bootstrap majority rule (70%) consensus trees for the 2b and CP genes of Wisconsin and reference CMV isolates.

Constructed phylogenetic trees for the 2b (a) and CP (b) genes using the neighbor-joining (NJ) method with 1000 bootstrap replicates under the Kimura 2-parameter model implemented in MEGA 5. Selected CMV isolates from GenBank and sixteen additional CMV isolates previously collected in the U.S. were included as reference isolates (Appendix 3). Reference isolates have been marked with red boxes. Phylogenetic trees were rooted using *Peanut stunt virus* (PSTV) and *Tomato aspermy virus* (TAV) as out-groups. CMV subgroups have been indicated by vertical brackets (P = pepper; S = snap bean; Numbers at the end of taxon represent collection year).



References

1- Ali, A., Li, H., Schneider, W. L., Sherman, D. J., Gray, S., Smith, D. & Roossinck, M. J. (2006). Analysis of genetic bottlenecks during horizontal transmission of *Cucumber mosaic virus*. *J Gen Virol* **80**, 8345- 8350.

2- Bedford, I. D., Briddon, R. W., Rosell, R. & Markham, P. G. (1994). Geminivirus transmission and biological characterization of *Bemisia tabaci* (Gennadius) biotypes from different geographic regions. *Ann Appl Biol* **125**, 311–325.

3- Boccard, F. & Baulcombe, D. (1993). Mutational analysis of cis-acting sequences and gene function in RNA3 of *cucumber mosaic virus*. *Virology* **193**, 563–578.

4- Bonnet, J., Fraile, A., Sacrista´ n, S., Malpica, J. M. & Garcı´a-Arenal, F. (2005). Role of recombination in the evolution of natural populations of *Cucumber mosaic virus*, a tripartite RNA plant virus. *Virology* **332**, 359–368.

5- Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W. & Baulcombe, D. C. (1998). Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J* **17**, 6739–6746.

6- Bruenn, J. A. (1991). Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases. *Nucleic Acids Res* **19**, 217–226.

7- Canto, T., Prior, D. A. M., Hellwald, K. H., Oparka, K. J. & Palukaitis, P. (1997). Characterization of *cucumber mosaic virus*. IV. Movement protein and coat protein are both essential for cell-to-cell movement of *cucumber mosaic virus*. *Virology* **237**, 237–248.

8- Codoner, F. M. & Elena, S. F. (2008). The promiscuous evolutionary history of the family *Bromoviridae*. *J Gen Virol* **89**, 1739-1747.

9- Daniel, J. & Campbell, R. N. (1992). Characterization of *Cucumber Mosaic Virus* Isolates from California. *Plant Dis* **76**, 1245- 1250.

10- Davino, S., Bellardi, M. G., Dibella, M., Davino, D. & Bertaccini, A. (2005). Characterization of a *Cucumber mosaic virus* isolate infecting *Mandevilla sanderi* (Hemsl.) Woodson. *Phytopathol Mediterr* **44**, 220- 225

11- Diaz-Pendon, J. A., Li, F., Li, W. X. & Ding, S. W. (2007). Suppression of antiviral silencing by *cucumber mosaic virus* 2b protein in *Arabidopsis* is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**, 2053-2063.

12- Ding, S.W., Anderson, B. J., Haase, H. R. & Symons, R. H. (1994). New overlapping gene encoded by the *cucumber mosaic virus* genome. *Virology* **198**, 593–601.

13- 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.

14- Ding, S. W., Li, W. X. & Symons, R. H. (1995). A novel naturally occurring hybrid gene encoded by a plant RNA virus facilitates long distance virus movement. *EMBO J* **14**, 5762–577.

15- Efron, B., Halloran, E. & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc Natl Acad Sci USA* **93**, 7085–7090.

16- Fischer, H. U. & Lockhart, B. E. L. (1976). A strain of *cucumber mosaic virus* isolated from cowpeas in Morocco. *Phytopathol Z* **85**, 132-138.

17- Fraile, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M. & Garcı´a-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *J Virol* **71**, 934–940.

18- Fulton, J. P. (1950). Studies on strains of *cucumber virus 1* from spinach. *Phytopathology* **40**, 729 736.

19- Garcia-Arenal, F., Escriu, F., Aranda, M. A., Alonso-Prados, J. L., Malpica, J. M. & Fraile, A. (2000). Molecular epidemiology of *Cucumber mosaic virus* and its satellite RNA. *Virus Res* **71**, 1-8

20- Garcia-Arenal, F., Fraile, A. & Malpica, J. M. (2001). Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol* **39**, 157-186.

21- Garcı´a-Arenal, F. & Palukaitis, P. (1999). Structure and functional relationships of satellite RNAs of *cucumber mosaic virus*. *Curr Top Microbiol Immunol* **239**, 37–63.

22- German, T. L., Thompson, A. & Wills, D. K. (2004). Statewide distribution of virus problems on processing beans. In. Proc 2004 Midwest Food Processors Association Processing crops Conference. Feb 23-24, La Cross, WI Department of Horticulture, University of Wisconsin.

23- Gildow, F. E., Shah, D. A., Sackett, W. M., Butzel, T., Nault, B. A. & Fleischer, S. J. (2008). Transmission efficiency of *Cucumber mosaic virus* by aphids associated with virus epidemics in snap bean. *Phytopathology* **98**, 1233-1241.

24- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1988). A novel super family of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. *FEBS Lett* **235**, 16–24.

25- Grau, C. R., Stevenson, W. R. & Mondjana, A. M. (2002). Viruses causing losses on processing beans in the Midwest, In. Proc 2002 Wisconsin fertilizer, Aglime and Pest Management Conference, **41**, 248-256.

26- Habili, N., and Symons, R. H. 1989. Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases. *Nucleic Acids Res* **17**, 9543–9555.

27- Hayes, R. J. & Buck, K. W. (1990). Complete replication of a eukaryotic virus RNA in vitro by a purified RNA-dependent RNA polymerase. *Cell* **63**, 363–368.

28- Hillis, D. M. & Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst Biol* **42**, 182– 192.

29- Hodgman, T. C. (1988). A new superfamily of replicative proteins. *Nature* 333, 22–23.

30- Hudson, R. R. (2000). A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–2014.

31- Hudson, R. R., Boos, D. D. & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Mol Biol Evol* **9**, 138–151.

32- Ishihama, A. & Barbier, P. (1994). Molecular anatomy of viral RNA-directed RNA polymerase. *Arch Virol* **134**, 235-258.

33- Kadare´, G. & Haenni, A. L. (1997). Virus-encoded RNA helicases. *J Virol* **71**, 2583–2590.

34- Kim, C. H. & Palukaitis, P. (1997). The plant defense response to *cucumber mosaic virus* in cowpea is elicited by the viral polymerase gene and affects virus accumulation in single cells. *EMBO J* **16**, 4060–4068.

35- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120.

36- Kobori, T., Miyagawa, M., Nishioka. K., Ohki, S. T. & Osaki, T. (2002). Amino acid 129 of *cucumber mosaic virus* coat protein determines local symptom expression and systemic

movement in *Tetragonia expansa*, *Momordica charantia* and *Physalis floridana*. *J Gen Plant Pathol* **68**, 81–88.

37- Kosakovsky-Pond., S. L. & Frost, S .D .W. (2005a). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531- 2533.

38- Kosakovsky-Pond, S. L., Frost, S. D. W., Grossman, Z., Gravenor, M. B., Richman, D. D. & Brown, A. J. (2006). Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLOS Comput Biol* **2**, e62, doi: 10.1371/journal. Pcbi.0020062.

39- Koundal, V., Hag, Q. M. & Praveen, S. (2011). Characterization, genetic diversity, and evolutionary link of *Cucumber mosaic virus* strain New Delhi from India. *Biochem Genet* **49**, 25-38.

40- Kuwite, C. A. & Purcifull, D. E. (1982). Some properties of a *cucumber mosaic virus* strain isolated from winged bean in Florida. *Plant Dis* **66**, 1071-1073.

41- Larsen, R. C., Miklas, P. N., Eastwell, K. C., Grau, C. R. & Mondjana, A. (2002). A virus disease complex devastating late season snap bean production in the Midwest. *Annu Rep Bean Improv Coop* **45**, 36- 37.

42- Lewsey, M., Surette, M., Robertson, F. C., Ziebell, H., Choi, S. H. & Ryu, K. H. (2009). The role of the *cucumber mosaic virus* 2b protein in viral movement and symptom induction. *Mol Plant-Microbe Interact* **22**, 642-54.

43- Lewsey, M., González, G., Kalinina, N. O., Palukaitis, P., Canto, T. & Carr1, J. P. (2010). Symptom induction and RNA silencing suppression by the *cucumber mosaic virus* 2b protein. *Plant Signal Behav* **5**, 705- 708.

44- Li, H. & Roossinck, M. J. (2004). Genetic bottlenecks reduce population variation in an experimental RNA virus population. *J Virol* **78**, 10582– 10587.

45- Lin, H. X, Rubio, L., Smythe, A. & Falk, B. W. (2004). Molecular population genetics of *cucumber mosaic virus* in California: evidence for founder effects and reassortments. *J Virol* **78**, 6666–6675

46- Lin, H. X., Rubio, L., Smythe, A., Jiminez, M. & Falk, B. W. (2003). Genetic diversity and biological variation among California isolates of *Cucumber mosaic virus*. *J Gen Virol* **84**,249–258.

47- Liu, Y.Y., Yu, S. L., Lan, Y. F., Zhang, C. L. & Hou, S. S. (2009). Molecular variability of five *Cucumber mosaic virus* isolates from China. *Acta Virologica* **53**, 89-97.

48- Marco, C. F. & Aranda M. A. (2005). Genetic diversity of a natural population of *Cucurbit yellow stunting disorder virus*. *J Gen Virol* **86**, 815- 822.

49- McGrath, P. F., Harrison, B. D. 1995. Transmission of *tomato leaf curl geminiviruses* by *Bemisia tabaci*: effects of virus isolate and vector biotype. *Ann Appl Biol* **126**, 307–316.

50- Meiners, J. P., Smith, F. F., Lawson, R. H., Waterworth, H. E. & Alconero, R. (1974). A strain of *cucumber mosaic virus* seed- borne in bean (*Phaseolus vulgaris* L.). (Abstr.). *Proc Phytopathol Soc* **1**, 137.

51- Mossop, D. W. & Francki, R. I. B. (1977). Association of RNA 3 with aphid transmission of *cucumber mosaic virus*. *Virology* **81**, 177–181.

52- Moury, B. (2004). Differential selection of genes of *cucumber mosaic virus* subgroups. *Mol Viol Evol* **21**, 1602-1611.

53- Nault, B. A., Shah, D. A., Dillard, H. R. & McFaul, A. R. (2004). Seasonal and spatial dynamics of alate aphid dispersal in snap bean fields in proximity to alfalfa and implication for virus management. *Environ Entomol* **33**, 1593- 1601.

54- Nault, B. A., Shah, D. A. & Taylor, A. G. (2006). Viruses and aphids everywhere in New York snap bean fields in 2005, pp. 74-76. *In* Proceedings of the 2006 Empire State Fruit and Vegetable EXPO. February 13-16, 2006. Cornell Coop. Exten., Syracuse, NY.

55- Ng, J. & Perry, K. L. (1999). Stability of the aphid transmission phenotype in *cucumber mosaic virus*. *Plant Pathol* **48**, 388-394.

56- Owen, J., and Palukaitis, P. 1988. Characterization of *cucumber mosaic virus*. I. Molecular heterogeneity mapping of RNA 3 in eight CMV strains. Virology 166:495–502.

57- Palukaitis, P. & Garcia-Arenal, F. (2003). *Cucumoviruses*. *Adv Virus Res* **62**, 241– 323.

58- Palukaitis, P., Roossinck, M. J. Dietzgen, R. G. & Francki, R. I. B. (1992). *Cucumber mosaic virus*. *Adv Virus Res* **41**, 281-348.

59- Paulus, G. C., Thomason, J. & Weathers, L. G. (1963). A soilborne virus disease of pepper in California. *Phytopathology* **53**, 885-886.

60- Peden, K. W. C. & Symons, R. H. 1973. *Cucumber mosaic virus* contains a functionally divided genome. *Virology* **53**, 487–492.

61- Perry, K. L., Zhang, L. & Palukaitis, P. (1998). Amino acid changes in the coat protein of *cucumber mosaic virus* differentially affect transmission by the aphids *Myzus persicae* and *Aphis gossypii*. *Virology* **242**, 204–210.

62- Perry, K. L., Zhang, L., Shintaku, M. H. & Palukaitis, P. (1994). Mapping determinants in *cucumber mosaic virus* for transmission by *Aphis gossypii*. *Virology* **205**, 591–595.

63- Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* **8**, 3867–3874.

64- Provvidenti, R. (1976). Reaction of *Phaseolus* and *Macroptilium* species to a strain of *cucumber mosaic virus*. *Plant Dis Rep* **60**, 289-293.

65- Rizos, H., Gunn, L. V., Pares, R. D. & Gillings, M. R. (1992). Differentiation of *cucumber mosaic virus* isolates using the polymerase chain reaction. *J Gen Virol* **73**, 2099–2103.

66- Rodriguez- Alvarado, G., Kurath, G. & Dodds, J. A. (1995). Heterogenity in pepper isolates of *cucumber mosaic virus*. *Plant Dis* **79**, 450- 455.

67- Roossinck, M. J. (2002). Evolutionary history of *cucumber mosaic virus* deduced by phylogenetic analyses. *J Virol* **76**, 3382–3387.

68- Roossinck, M. J., Zhang, L. & Hellwald, K. H. (1999). Rearrangements in the 59 nontranslated region and phylogenetic analysis of *cucumber mosaic virus* RNA 3 indicate radial evolution of three subgroups. *J Virol* **73**, 6752–6758.

69- Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496.

70- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

71- Schwinghamer, M. W. & Symons, R. H. (1977). Translation of the four major RNA species of *cucumber mosaic virus* in plant and animal cell-free systems and in toad oocytes. *Virology* **79**, 88–108.

72- Shintaku, M. H., Zhang, L. & Palukaitis, P. (1992). A single amino acid substitution in the coat protein of *cucumber mosaic virus* induces chlorosis in tobacco. *Plant Cell* **4**, 751–757.

73- Smith, T. J., Chase, E., Schmidt. T. & Perry, K. L. (2000). The structure of *cucumber mosaic virus* and comparison to *cowpea chlorotic mottle virus*. *J Virol* **74**, 7578–7586.

74- Sullistyowati, E., Mitter, N., Bastiaan- Net, S., Roossinck, M. J. & Dietzgen, R. G. (2004). Host range, symptom expression and RNA 3 sequence analyses of six Australian strains of *Cucumber mosaic virus*. *Aust Plant Pathol* **33**, 505-512.

75- Suzuki, M., Kuwata, S., Kataoka, J., Masuta, C., Nitta, N. & Takanami, Y. (1991). Functional analysis of deletion mutants of *cucumber mosaic virus* RNA3 using an in vitro transcription system. *Virology* **183**, 106–113.

76- Suzuki, M., Kuwata, S., Masuta, C. & Takanami, Y. (1995). Point mutations in the coat protein of *cucumber mosaic virus* affect symptom expression and virion accumulation in tobacco. *J Virol* **76**, 1791–1799.

77- Takeshita, M., Suzuki, M. & Takanami,Y. (2001). Combination of amino acids in the 3a protein and the coat protein of *cucumber mosaic virus* determines symptom expression and virus spread in bottle gourd. *Arch Virol* **146**, 697–711.

78- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* **28**, 2731-2739.

79- Tomlinson, J. A., Carter, A. L., Dale, W. T. & Simpson, C. J. (1970). Weed plants as source of *cucumber mosaic virus*. *Ann Appl Biol* **66**, 11-16.

80- Wahyuni, W. S., Dietzgen, R. G., Hanada, K., and Francki, R. I. B. 1992. Serological and biological variation between and within subgroup I and II strains of *cucumber mosaic virus*. Plant Pathol. 41: 282-297.

81- Wei, T. Y., Yang, J. G., Liao, F. R., Gao, F. L., Lu, L. M., Zhang, X. T., Li, F., Wu, Z. J., Lin, Q. Y,, Xie, L. H. & Lin, H. X. (2009). Genetic diversity and population structure of rice stripe virus in China. *J Gen Virol* **90**, 1025-1034

82- Whipple, O. C. & Walker, J. C. (1941). Strains of *cucumber mosaic virus* pathogenic on bean and pea. *J Agric Res* **62**, 0027-0060.

83- Wright, S. (1951). The genetical structure of populations. *Ann Eugen* **15**, 323-354.

Chapter 4

ASSOCIATION OF NEW SATELLITE RNAs WITH NATURAL INFECTIONS OF *CUCUMBER MOSAIC CUCUMOVIRUS* (CMV) IN SUCCULENT SNAP BEAN AND THE ROLE OF EXTRA RNAs IN SYMPTOM DEVELOPMENT*

* The first part of this chapter (satellite RNA discovery) is a slightly modified version of the following publication:

Nouri S, Falk BW, Groves RL. 2012. A new satellite RNA is associated with natural infections of *cucumber mosaic virus* in succulent snap bean. *Arch Virol* **157**: 375- 378.

*A slightly modified version of the second part of this chapter (symptom development bioassay) will be submitted for publication as:

Nouri, S., Palukaitis, P., and Groves, R. L. 2012. Symptom enhancement by two newly characterized satellite RNAs of *Cucumber mosaic cucumovirus* (CMV) in snap bean.

Abstract

From ~2000 to the present, snap bean (*Phaseolus vulgaris* L.) production in the Midwest and Northeast United States has been affected by a complex of virus disease epidemics. *Cucumber mosaic cucomovirus* (CMV) was consistently recovered from symptomatic and asymptomatic plants infected with CMV during surveys conducted in 2007, 2008 and 2010 in Wisconsin. Satellite RNAs (satRNA) were detected in 71% of selected samples by RT-PCR. Comparison of satRNA sequences with currently available CMV satRNA sequences suggested these to be novel satRNAs. A greenhouse bioassay used to measure symptom development illustrated that the co-association of the two satRNAs with CMV infections exacerbated symptoms in snap bean and hastened the onset of symptom expression in this host. The effects on symptom development were host specific since the new satRNAs attenuated CMV symptoms in tobacco.

Introduction

Snap bean (*Phaseolus vulgaris* L.) is one of the most important commercially grown vegetables in the United States with 161 M dollars in revenue (Natural Agricultural Statistics Service, 2011) with Wisconsin as a domestic production leader. In 2000, a widespread *Cucumber mosaic cucomovirus* (CMV) outbreak significantly affected the snap bean processing industry in Wisconsin and then progressed to the other portions of Midwest and Northeast United States (New York and Pennsylvania States) (23, 27, 46, 53), plus Ontario, Canada (46). In some cases, nearly 100% of plants in snap bean fields were infected with CMV (54). In the US, CMV has been regarded as endemic for many years, yet it had only intermittently affected snap bean production prior to 2000 (15, 78).

CMV is one of the most economically important plant viruses with a worldwide distribution and a potentially severe impact on cultivated crops (59). *Cucumber mosaic cucomovirus*, the type species of the genus *Cucumovirus* in the family *Bromoviridae*, is also one of the most widespread plant viruses. The host range of CMV includes over 1,000 plant species (56, 59) and the virus is transmitted by over 75 species of aphids in a non-persistent manner (59). The genome of CMV contains three positive-sense, single-stranded RNAs packaged in separated particles and two subgenomic RNAs (13, 59, 60). RNA 1 and 2 encode the 1a and 2a proteins; respectively constitute two subunits of the virus replicase complex (31). The 2b protein, encoded via a subgenomic RNA derived from RNA2, is a multifunctional protein and functions in host-specific, long-distance movement, symptom induction, and as a virulence determinant by suppressing gene silencing (7, 12, 13, 14). The CMV RNA 3 encodes two proteins, 3a, a cell-to-cell movement protein (MP) (14) and 3b or capsid protein (CP); this latter protein being translated from a sub-genomic RNA 4 (71) which is also involved in cell-to-cell movement,

virus assembly and aphid-mediated transmission (6, 9, 51, 61, 62, 75). Furthermore, CMV is a heterogenic group divided into two subgroups designated as subgroups I and II based on serological (59, 77) and nucleotide sequence identities (59, 66), with each subgroup sharing approximately 75% nucleotide identity (66). Analysis of the CP gene and 5` non-translated region (NTR) of the RNA3 has led to further division of subgroup I into designations of IA and IB, sharing approximately 92-95% nucleotide identity (66, 69). There is also evidence showing a relationship between some strains of CMV and specific host plants such as legumes (19, 42, 45, 49, 78).

Some strains of CMV carry an extra RNA molecule known as satellite RNA (satRNA). This satRNA is a linear, single-stranded RNA molecule, ranging in size from 332 to 405 nucleotides (21, 73). These small RNAs do not encode any protein and are dependent on CMV for replication, encapsidation and transmission (26, 68). A unique feature of satRNAs is their ability to modify symptom expression induced by the helper virus, ranging from attenuation to increased severity depending on the CMV genotype (57, 67), host plant (22, 39, 44, 50) and satRNA (57). Most CMV satRNAs reduce the replication and accumulation of their helper virus, which results in symptom attenuation (5, 29, 33, 36, 47). However, the presence of hypervirulent satRNA in the CMV population has been described to cause severe systemic chlorosis in tobacco, pepper and tomato (25, 55, 76). The co-occurrence of the satRNA with natural CMV populations has further been linked to four important historical epidemics in tomato in different parts of the world (11, 35, 40, 43).

Consistent with these previous reports, one hypothesis that we propose as a partial explanation of the recent CMV epidemics, is the co-occurrence, or presence of satRNA in affected snap bean

fields. Specifically, our objective in the current study was to investigate the relationship of this extra RNA and disease symptom development within susceptible snap bean.

Towards this end, we collected CMV infected samples from symptomatic snap bean fields in Wisconsin to document the possible association of satRNA with these field-collected isolates. Following characterization, we investigated the role of field-collected satRNA in disease symptom development using greenhouse inoculation bioassays including a satellite free, legume-specific CMV infectious clone that was in the same subgroup as the CMV in the snap bean fields and two satRNA variants isolated from Wisconsin field samples.

Materials and Methods

Sampling and virus isolates

Five leaf samples from snap bean (*Phaseolus vulgaris* L.) plants showing disease symptoms, including mottling, blistering and stunting, were collected from 12 fields in Wisconsin during the 2007, 2008 and 2010 production seasons between July and September in each year. Samples were initially returned to the laboratory and assayed for the presence of CMV. Three sets of CMV-infected trifoliate leaves were then flash frozen in liquid nitrogen and held in a freezer at -80 °C until RNA extraction were completed.

RNA extraction and RT- PCR amplification

Total RNA was extracted from leaf samples using an RNeasy Mini kit (Qiagen, CA, USA) according to the manufacturer's instructions. To select CMV isolates, the CMV coat protein (CP) gene was targeted for amplification with a specific primer pair (CP F: 5`-TTGAGTCGAGTCATGGACAAATC-3' and CP R:5'-AACACGGAATCAGACTGGGAG-3') (48). One-step RT-PCR amplification was carried out in 25 µl volume of 1X Go Taq Flexi DNA

Polymerase reaction buffer (Promega, WI, USA), containing 0.2 mM dNTP mix, 5 mM DTT, 1.1 mM MgCl2, 10 U of SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad ,USA), 0.75 U of Go Taq Flexi DNA Polymerase (Promega, WI, USA) and 12.5 mM of each primer. The first strand cDNAs were synthesized at 53˚C for 45 minutes and then PCR reactions were conducted by 25 cycles in 94˚C for 30s, 54˚C for 30s, 72˚C for 1 minute and finished by 72˚C for 7 minutes. RT-PCR with specific satRNA primer pairs (CMV sat F: 5`GGGAATTCATTTAGGTGACACTATAGTTTTGTTTG3` and CMV sat R: 5`GGGGTCTAGACCCGGGTCCTG3`) was conducted for a subset of CMV isolates (30). The RT-PCR reaction and PCR program was the same as described above for the CP gene, except that the annealing temperature was 56 ˚C for satRNA amplification. Products were separated on a 2% TAE agarose gel to initially screen for the presence or absence of satRNA. Mock, buffer-inoculated plant and a CMV Ky-strain with associated satRNA, were used as negative and positive controls (30), respectively. Gel purification was performed using a QIAquick Gel Purification Kit (QIAGEN, CA, USA).

Cloning, sequencing and sequence comparison

The purified PCR products of satRNA were directly sequenced. Ten satRNA cDNAs were cloned using the pGEMT Easy Vector kit (Promega, WI, USA) and transformed in chemically competent cells of *Escherichia coli* strain DH5α according to the manufacturer's instructions. Three colonies from each isolate were selected from the plates and screened by PCR for the presence of appropriately sized inserts using the M13F and M13R primers. Minipreps were performed using the Qiaprep Spin Miniprep kit (Qiagen, CA, USA) and the satRNA cDNA inserts were sequenced using a BigDye termination sequencing kit and universal M13F and M13R primers in an automated DNA sequencer (ABI Prism 310 Genetic Analyzer, Applied

Biosystem) at the DNA Sequencing Facility at Biotechnology Center of the University of Wisconsin-Madison. Consensus sequences were obtained with the DNASTAR (Madison, WI, USA) and products were blasted against GenBank sequences using the local alignment tool, BLASTn (2).

Purification of CMV satellite RNA

The total RNAs of two satRNA positive CMV isolates confirmed by nucleotide sequencing were separated by electrophoresis on a 1% agarose denaturing gel containing glyoxal/DMSO (8). These isolates were obtained from a single snap bean field in Wisconsin. The total RNAs of reference controls, including a healthy plant and the CMV Ky-strain carrying satRNA, were used as negative and positive controls, respectively to determine satRNA bands on the gel. A 0.1-2 Kb RNA ladder (Invitrogen, Carlsbad, USA) was also run next to the RNA samples to estimate the approximate size of satRNA bands. The region of the gel containing the satellite RNA was excised and gel slices were ground into a paste with nuclease free plastic pestles, and eluted overnight with 0.5M ammonium acetate (Fisher Scientific, Waltham, MA), 1mM EDTA and 0.2% SDS (Fisher Scientific, Waltham, MA) (70). The eluted RNAs were transferred to new tubes and precipitated with 3V of cold ethanol. RNA concentrations were measured with the Nanodrop apparatus (Thermo Scientific, USA) and RNA quality was analyzed by a denaturing gel electrophoresis. Total RNAs were also examined by RT-PCR with satRNA specific primers, as described previously, to check the size of amplicons. These isolated satRNAs were referred to as IR-WI and Sb-WI- satRNAs. They were stored at -80 °C for future experiments.

Plasmid construction and in vitro transcription

Viral RNAs were transcribed from full-length infectious cDNA clones (kindly donated by Dr. Peter Palukaitis) and obtained from CMV-B RNA2 (a legume-specific, systemically-infecting strain of CMV), CMV-Fny RNA1 (pFny-109) and CMV-Fny RNA3 (pFny-309) (65). Obtained constructs were transformed into chemically competent cells of *Escherichia coli* strain JM101 (Agilent, CA, USA) according to the manufacturer's instructions and DNA was extracted with a Qiaprep Spin Miniprep kit (Qiagen, CA, USA). The plasmids were linearized with Pst1 and blunt-ended by treatment with T4 DNA polymerase prior to application as a template for in vitro transcription using mMESSAGE mMACHINE T7 kit (Ambion, Forster, CA, USA), according to the manufacture's instruction. Transcripts were precipitated with ethanol and then dissolved in nuclease free water.

Inoculation source preparation

RNA transcripts obtained from the above experiment were mechanically, sap-inoculated into *Nicotiana benthamiana* plants as described previously (65, 74). Equal amounts of each RNA (~ 200 µg) were mixed with EFS buffer (1.6 µg / µL bentonite, 0.05M phosphate buffer, pH 7.2 and carborundum powder) and inoculated onto 6-7 fully expanded leaves of 5-8 week old *N.benthamiana* plants (20 microliter per plant). Inoculated plants were maintained in an insect-proof growth chamber at 22 °C and a 16:8 (L:D) photoperiod. Inoculated plants were examined at 14 days post inoculation (dpi) by RT-PCR with a CMV CP specific primer, as described previously. CMV positive leaves were later used as a satRNA free, CMV inoculation source for the remainder of the study. To obtain isolates carrying satRNA, each purified satRNA (~ 25 µg) was mechanically inoculated into the fully expanded leaves of *N. benthamiana* plants after being mixed with aforementioned, CMV RNA transcripts (74). We also inoculated *N. benthamiana*

plants with satRNA alone to confirm that no genomic RNA was transferred with satRNA from gel. Inoculated plants were held in insect-free cages in a separate growth chamber under the same conditions, as described above. The plants were assayed at 14 dpi by RT-PCR with CMV satRNA specific primer. Infected leaves were used as a CMV+ satRNA inoculation sources for the remaining experiments described in the next section. Plants inoculated with phosphate buffer (pH 7.0) were grown under similar conditions as all groups to compare symptoms and to also serve as a negative control in RT-PCR reactions.

Symptom development assay in snap bean and tobacco

To test the effect of each satRNA on disease symptom development, replicate sets of even-aged snap bean plants (*Phaseolus vulgaris,* cv. 'Hystyle') were mechanically sap-inoculated onto fully expanded cotyledons and the first true trifoliate leaves with plant sap extracted from leaves from among the 3 groups of *N. benthamiana* each containing CMV, CMV+ IR-WI- satRNA, and CMV+ Sb-WI-satRNA, in 0.1M potassium phosphate buffer. A total of 10 plants, plus a mock inoculated buffer control, were considered as a replicate set with 3 experimental replications arranged as a randomized complete block design using time as a blocking factor. Replicate sets of plants were maintained separately in insect-proof cages within greenhouse rooms with a 14:10 (L:D) photoperiod and temperatures ranging between 22-26 and 18-24°C during the daylight and nighttime hours, respectively. The same experimental procedure was performed for tobacco (*Nicotiana tabacum* cv. 'Xanthi'). In these experiments, plants were inoculated at the four- to six true leaf stage and again maintained in insect-proof cages under similar temperature and photoperiod conditions for 3 weeks post-inoculation.

Influence of satRNA constructs on symptom expression and pod yield and quality in succulent snap bean

Following the 7 dpi period, symptoms of disease were recorded daily for the next 20. Foliar plant disease symptoms recorded included mosaic, mottling and rugosity. We also measured the plant heights from different experimental groups. The presence of both satRNA and CMV genomic RNAs were confirmed at 21 dpi in the inoculated plants by RT-PCR using CP and satRNA-specific primers (30, 48). To investigate the potential influence of experimental constructs on pod yield and quality, inoculated snap bean plants from among all constructs were allowed to grow for a total of 53 days to produce marketable pods. Fresh above-ground biomass, pod number and shape were recorded for all pods and subsequently compared among experimental treatments.

Results

Identification of satellite RNA in natural infections of *Cucumber mosaic cucomovirus* (CMV) in succulent snap bean

Fifty-three isolates of CMV were obtained from symptomatic snap bean plants in 12 separate fields in Wisconsin (Table 1). Fragments of approximately 380 bp were amplified with RT-PCR using satRNA primer from 38 of 53 isolates (71.6%) and from all 12 fields (Table 1; Fig. 1). The identity of the satRNAs was confirmed by direct sequencing the amplified cDNA fragments.

Characterizations of satRNAs and relationship to other known CMV satRNAs

The isolated CMV satRNAs obtained from Wisconsin fields varied in nucleotide sequence. In some cases, multiple variants were isolated from a single field. Since our goal was to examine effects of satRNA on CMV symptom expression and not to assess satRNA diversity, we

randomly selected two satRNA variants obtained from a single field. As previously mentioned, these isolates are referred to as IR-WI (GenBank accession number JF834526) and Sb-WI-satRNA (GenBank accession number to be determined) with 338 and 339 nucleotide length, respectively. These satRNA were 96% identical, with 12 substitutions and one deletion (Fig. 2). Comparison of these novel CMV satRNAs with previously described CMV satRNAs available in GenBank showed that IR-WI and Sb-WI- satRNA sequences shared the highest nucleotide identity with a Spanish satRNA (ca. 94%) (4) and an E-satRNA from Japan (ca. 93%) (32), respectively. The IR-WI and Spanish satRNAs were differentiated by 17 nucleotide substitutions, two deletions and one insertion (Fig. 3a) and the Sb-WI and E-satRNAs included 22 nucleotide substitutions and one deletion (Fig. 3b). The two satRNAs isolated from Wisconsin snap bean fields, IR-WI and Sb-WI-satRNA, shared 94% and 90% identity, respectively, with the 1CARNA5 satRNA, an attenuating satRNA, originally isolated from cucumber (*Cucumis sativus* L.) in Wisconsin and described in 1983 (10). Also, the range of sequence identity between these newly detected satRNAs in Wisconsin and two previously reported satRNAs, CMV-WL (25) and B- satRNA (16) in New York, another U. S. snap bean production region with a recent history of CMV outbreaks, was 89 to 91%. B- satRNA is an attenuating variant from a strain of CMV isolated from red kidney bean (63), whereas CMV-WL is a chlorogenic satRNA on tomato.

Determining the potential role of newly characterized satRNAs in disease symptom development, pod quality, and pod yield in succulent snap bean

To characterize the role of the newly emerged satRNAs in the development of disease symptoms induced by the helper virus (CMV) in snap bean, the two recently discovered satRNA variants in this study, IR-WI and Sb-WI-satRNAs were tested in snap bean, in combination with a satRNA-

free legume- specific CMV. The satRNA-free CMV was used as a reference control for comparison of symptom production and yield impacts. We did not inoculate constructs directly into snap bean plants, because establishing infection through the use of transcripts is inconsistent in *P. vulgaris* (P. Palukaitis, Personal communication). Therefore, constructs were inoculated into *N.benthamiana* plants, prior to inoculation into *P. vulgaris*. The sap of *N. benthamiana* plants that were PCR positive for CMV or CMV and satRNA were used as an inoculum source for snap bean. Plants inoculated with satRNA alone did not show any symptom and these plants did not support satRNA replication based on RT-PCR assays. The satRNA-free CMV inoculations into snap bean produced only a mild mosaic and green mottling pattern, which appeared in as few as 14 to 16 dpi (Table 2; Fig. 4a). Disease symptoms progressed and by 25 dpi, there was more obvious mottling and leaf malformation in infected plants (Fig. 4b). In contrast, plants inoculated with the CMV/IR-WI or Sb-WI- satRNA had severe initial symptoms including leaf rugosity and blistering (Fig. 5a) which appeared in as few as 7 to 10 dpi (Table 2). And shortly thereafter, symptoms continued to develop into a more severe blistering and leaf deformation at 14 dpi (Fig. 5b). No significant differences were observed with respect to the plant height (Table 3). To determine the effects of newly characterized satRNAs on the quantity and quality of pods, the above-ground biomass of infected plants and the total number of pods were measured at the completion of the experiment (Table 3). No significant differences were observed between experimental treatments including CMV alone and CMV plus satRNAs with respect to the measured parameters; however, they were different from healthy controls (Table 3). The majority of pods generated from infected plants with CMV/satRNA combinations produced a characteristic “C-shaped” pod symptom when compared to pods of infected plants by

CMV alone and the healthy control plants (Fig. 6). The presence of satRNA in all infected plants was confirmed by RT-PCR with satRNA specific primer (Fig. 7).

The influence of IR-WI-satRNA and Sb-WI-satRNA on symptom development in tobacco

To further test the role of the newly discovered satRNAs in symptom modification in other plant species, we conducted an additional inoculation bioassay on tobacco plants. Table 4 illustrates the results obtained from this bioassay. Plants infected with CMV alone possessed chlorotic spots mainly along the margins of leaves and veins which later became translucent (Fig.8a). The mixed infection of CMV with either IR-WI or Sb-WI- satRNA induced an oak leaf and ring spot symptom at 10 dpi initially in the inoculated leaf and in the newest leaf above the mechanical inoculation point. The progression of these symptoms to the upper leaves, however was abated at approximately 20 dpi, and leaves above this point remained asymptomatic (Fig. 8b and 9).

Discussion

We found that satRNAs were prevalent in snap bean with CMV symptoms in Wisconsin fields. We identified two novel CMV satRNAs and showed that they increase CMV symptom severity on snap bean, but alter symptoms and decrease CMV severity on tobacco. Our results support the hypothesis that novel satRNAs play an important role in the recent CMV epidemic on snap bean in the United States.

CMV is endemic in most agricultural areas of the United States. CMV was first reported in 1916 as the causal agent of a plant disease in cucumber and muskmelon in Wisconsin and Michigan (15) and also in cucumber in New York (34). In 2000, snap bean growers in Wisconsin experienced a significant increase in several plant-infecting viruses, with CMV contributing significantly to these epidemics. Since 2000, the incidence of CMV in snap beans has steadily

increased in different parts of the country including the Midwest and portions of the Northeast including Pennsylvania and New York (23, 24, 27, 46, 53). The presence of an extra RNA molecule (satRNA) in the genome of some CMV isolates can intensify symptom expression induced by the helper virus (11, 25, 35, 40, 43, 55, 76) and four devastating CMV epidemics around the world in tomato were the result of this association (11, 35, 40, 43). To our knowledge, this is the first report describing an association of CMV and satRNA in snap bean fields. Although examination of the genetic diversity among CMV satRNAs was not the purpose of this study, we observed variation among the two satRNA sequenced. Heterogeneity in nucleotide sequence has been documented among CMV satRNAs in both experimental (e.g. greenhouse) and field conditions in various studies (1, 3, 20, 28, 44, 64) and it seems plausible that the populations of satRNA in the affected areas in U.S. is not exceptional.

The origin of these new satRNAs is not clear. A report of the isolation of an attenuating variant (1-CARNA5) from a CMV isolate previously reported on cucumber (*Cucumis sativus*) in Wisconsin was published in 1983 (10), thus CMV satRNA may have been present in Wisconsin for decades. However, sequence comparisons showed that the newly determined satRNA sequences are different from the previous report and are more closely related to satRNA from Spain and Japan. Our data provide additional evidence demonstrating that CMV satRNAs consist of a population of closely related sequence variants.

To determine the role of the newly characterized satRNAs in symptom expression, we inoculated two variants, IR-WI and Sb-WI-satRNAs that are 96% identical and that were obtained from single snap bean field, onto two hosts, snap bean and tobacco. Our rationale arose from previous research suggesting that even single nucleotide changes can modify the behavior of satRNA and its effects on symptom expression (58). Furthermore, it has been shown that solanaceous plants

(e.g. *N. tabacum*) are good hosts for satRNA, supporting high levels of satRNA accumulation (21, 38, 50). Results from the greenhouse-based bioassay show that these novel satRNAs impact CMV symptom severity and time to development in snap bean. Although this association did not lead to particular disease phenotypes, including necrogenic or chlorogenic symptoms previously reported for pathogenic CMV satRNAs in tomato, tobacco and pepper (11, 25, 35, 40, 43, 55, 76), the increased severity of symptoms and shorter expression time in the presence of both satRNA variants was obvious.

The satRNA impacted pod quality by causing pod distortion when co-inoculated with CMV, but did not appear to impact pod number or the weight of the above-ground plant biomass. This result was unexpected based on our observations in snap bean fields, where we did observe a reduction in the number of pods apparently due to viral disease. The outcome of CMV-satRNA interactions depend on several factors, including the host, CMV genotype, satRNA variant and environmental conditions, and field conditions are difficult to replicate in greenhouse assays (21, 37). We also cannot rule out the possibility that the associated molecular and biological properties of the CMV infectious clone used in this experiment may have interacted differently with the satRNA and the plant when compared to the wild-type virus.

In contrast, both satRNA variants used in this research attenuated the symptoms in tobacco in combination with CMV. These results both illustrate that the host plays an important role in satRNA effects on CMV symptom development.

Our results suggest that satRNA epidemiology differs in snap bean compared to crops in which CMV satRNAs have previously been examined. According to a field survey for the satRNA of CMV in New York State, the frequency of satRNA incidence in field populations of CMV in tomato, pepper, lettuce and cucurbits is low (41), and in most natural population instances

satRNAs are selected against because of their effect on virus fitness and potential vector transmission limitations (5, 17, 18, 40, 47). Hence, the high incidence of satRNA in snap bean infected with CMV may reflect a more favorable set of conditions for satRNA maintenance or transmission. To partially explain this, and coincident with the emergence of CMV epidemics, the soybean aphid (*Aphis glycines* Matsumura), was also first observed in Wisconsin in July of 2000 and the populations of this insect rapidly dispersed across several North Central states. This insect is a competent vector of several nonpersistent viruses, including CMV (24, 53). With rapid increases in the aphid populations observed annually, the likelihood of these insects visiting and probing a myriad different host plants during migration flights, the virus and its associated satRNA have the potential to spread very rapidly. Therefore, one explanation for the recent epidemic of satRNA in snap bean CMV isolates may result from endemic satRNA in CMV populations present in the North Central states for many years combined with the introduction of a novel competent vector that is present in high numbers and that prefers legume crops. The satRNA incidence was probably not of concern previously because it attenuates symptoms in reservoir plants and also due to the potential disadvantage of satRNA for aphid transmission (5, 17, 18, 47). The introduction of this novel vector, the wide host range of the vector, the virus, and the satRNA, and the prevalence of satRNA in CMV-infected snap beans may drive further differentiation of CMV satRNA that cause severe symptoms on other CMV hosts.

It has been shown that both crop and non-crop reservoir plants containing CMV and satRNA play an important role as the source of primary inoculum (3, 35, 71). The results obtained from a recent survey in Wisconsin showed that *Asclepias syriaca* (common milkweed), cultivated *Cucurbita* spp., or perhaps *Capsicum annuum* (pepper), are significant inoculum sources of CMV in snap bean (52). The incidence of satRNA in the two first plant species has not been

investigated as yet, but we have detected satRNA in CMV isolates from pepper at a low frequency (data not shown). More detailed examination of CMV reservoir plants and an exploration for sources of satRNA may provide important new information about the possible source of these parasite RNAs in the affected areas.

Acknowledgments

We thank Dr. Peter Palukaitis (Scottish Crop Research Institute) for providing us B-CMV infectious clone.

Table 1. CMV isolates from Wisconsin and associated frequency of occurrence of satRNA.

Field	No. CMV isolates	No. satRNA with CMV	Year
SB	4	4	2007
BD	5	3	2007
MR	5	2	2007
K24	4	4	2010
S1C	5	4	2010
SC2	5	5	2010
B	3	1	2008
S	5	4	2007
DA	4	3	2007
H	5	3	2008
WK	4	2	2007
FCS	4	3	2007
Total	53	38	-

Table 2. Snap bean disease symptom development bioassay results.

This assay was conducted with three constructs over three experimental replicates. * Number of snap bean plants with symptoms in each replication. [§] Observed symptoms included mosaic and mottling. [Ψ] Mottling, blistering and leaf rogusity symptoms were observed.

Treatment (Constructs)	No. inoculated plant	Rep. 1*	Rep. 2*	Rep. 3*	First symptom appearance (dpi)
CMV	10	9	7	8	14[§]
CMV+IR-WI satRNA	10	7	8	8	7 [Ψ]
CMV+Sb-WI satRNA	10	9	9	9	9 [Ψ]

Table 3. Effect of infection by CMV and CMV+ satRNA on biomass, plant height and pods number in snap bean.

Data presented illustrate a mean ± SE among 10 plants from each experimental replicate.

Treatment	Fresh above- ground biomass (g)	Plant height (cm)	Number of pod per plant
Mock	411.95±27.68	38.75±0.74	18
CMV	360.11±29.26	34.70±0.38	12
CMV+IR-WI satRNA	352.20±32.04	33.06±0.35	10
CMV+Sb-WI satRNA	354.07±31.57	33.41±0.40	10

Table 4. *N. tabacum* disease symptom development bioassay results.

This assay was conducted with three constructs and three experimental replicates. *Number of tobacco plants illustrating symptoms in each replication. § Observed symptoms included chlorotic spots and vein clearing. Ψ oak leaf/ asymptomatic plants.

Treatment (Constructs)	No. inoculated plant	Rep. 1*	Rep. 2*	Rep. 3*	First symptom appearance (dpi)
CMV	10	8	9	10	14§
CMV+IR-WI satRNA	10	8	9	8	10/ asymptomatic Ψ
CMV+Sb-WI satRNA	10	9	8	9	10/ asymptomatic Ψ

Figure 1. Detection of satRNA in field isolates of CMV by RT-PCR.

M: 1Kb DNA ladder (Promega); 1- 12) CMV isolates collected from snap bean fields in Wisconsin and New York; 13) CMV Ky-strain carrying satRNA (positive control); 14) mock inoculated plant with phosphate buffer (pH 7.0) (negative control).

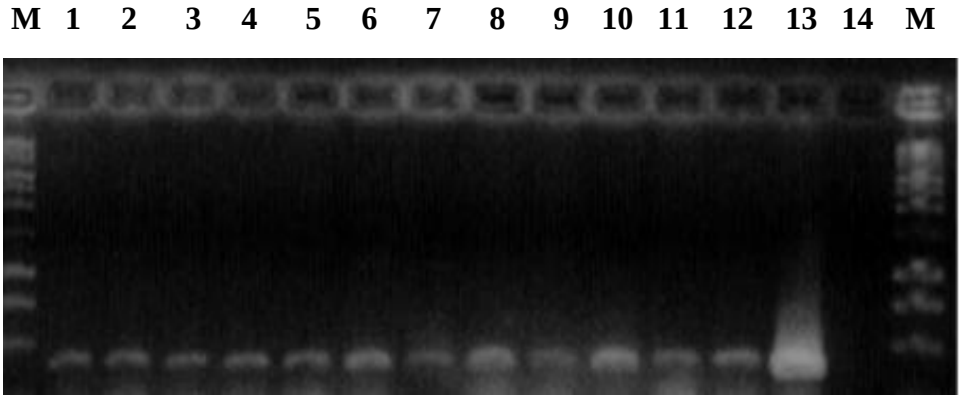


Figure 2. Alignment of IR-WI and Sb-WI-satRNA sequences.

These two variants were differentiated by 12 nucleotide substitutions and one deletion. GenBank accession numbers: IR-WI-satRNA, JF834526; Sb-WI-satRNA, the accession number is not available at the time of thesis submission. Variations in the nucleotide sequence are illustrated by boxes.

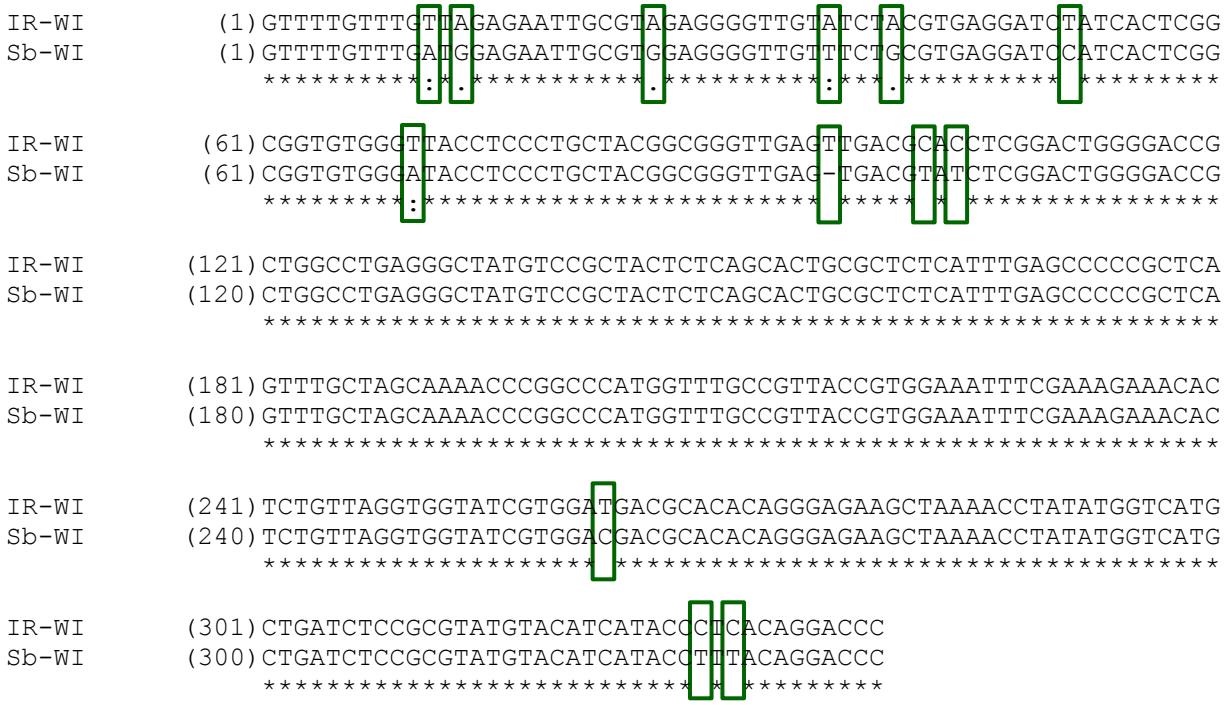


Figure 3. Alignment of IR-WI-satRNA and Sb-WI-satRNA with Spanish satellite RNA and E-satRNA sequences, respectively.

a) IR-WI-satRNA demonstrated the greatest sequence similarity with a Spanish satRNA (isolate To/1989/16.1) with 17 nucleotide substitutions, two deletions and one insertion. GenBank accession numbers: IR-WI, JF834526; Spanish-satRNA, Z75870. b) Sb-WI-satRNA showed the highest sequence similarity with E-satRNA with 22 nucleotide substitution and one deletion. GenBank accession numbers: Sb-WI-satRNA, the accession number is not available at the time of thesis submission; E-satRNA, M20844. Nucleotide variation is illustrated in box.

a

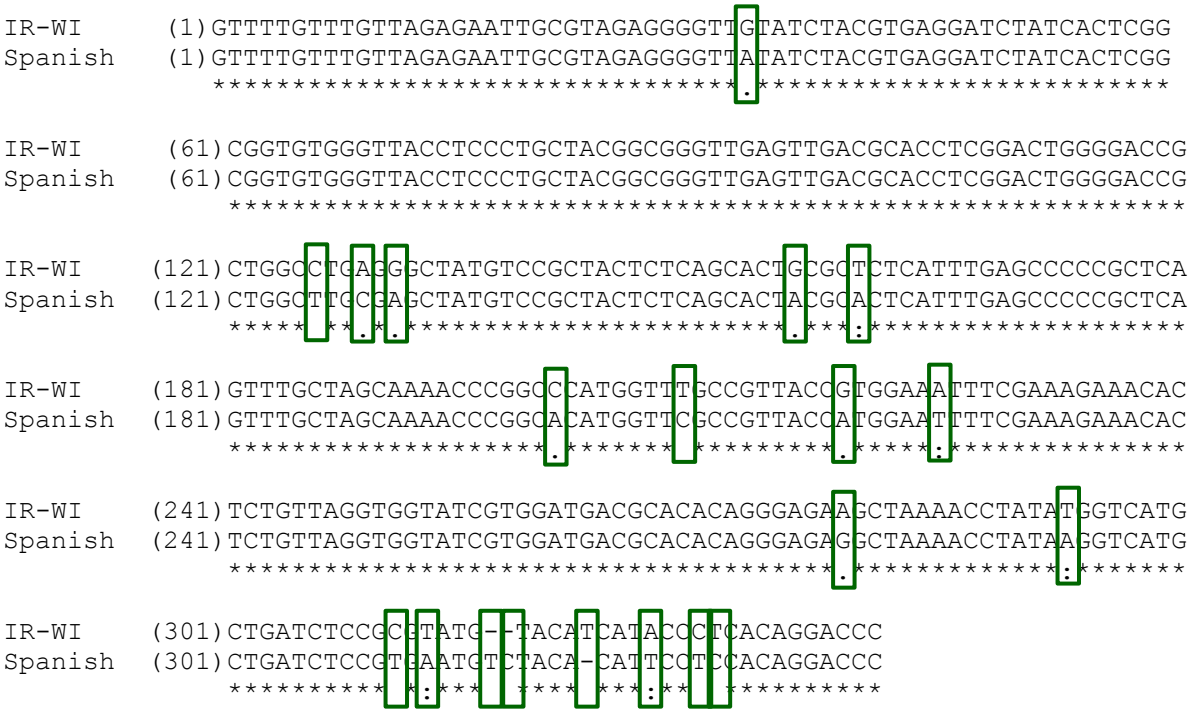


Figure 3- continued.

b

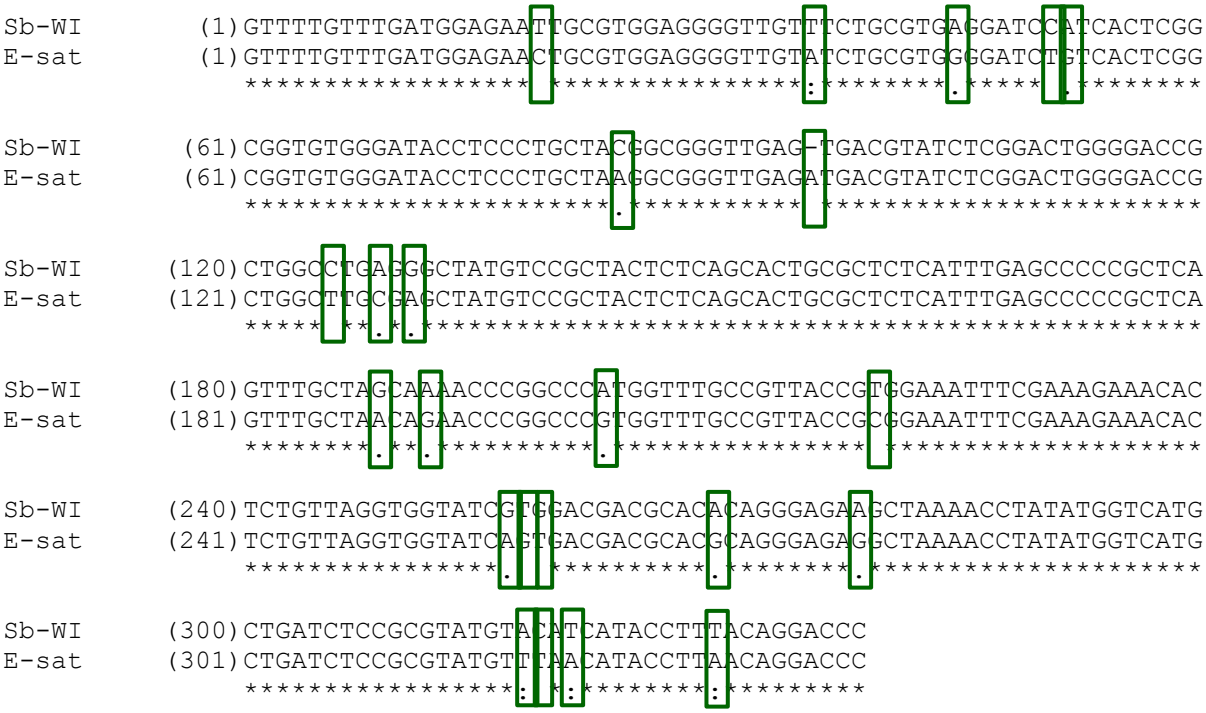
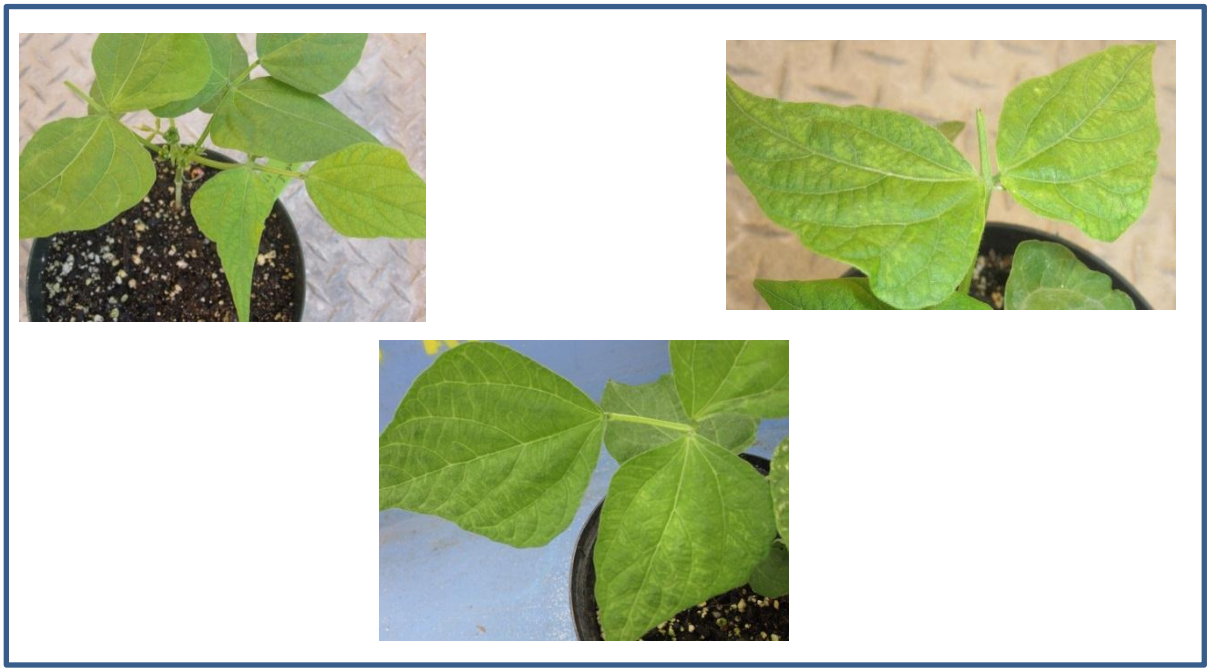


Figure 4. Symptoms caused by satRNA- free CMV on snap bean (*Phaseolus vulgaris* cv. 'Hystyle').

a) Mild mosaic and green mottling at 14 dpi; b) Severe mottling and leaf malformation at 25 dpi.

a



b

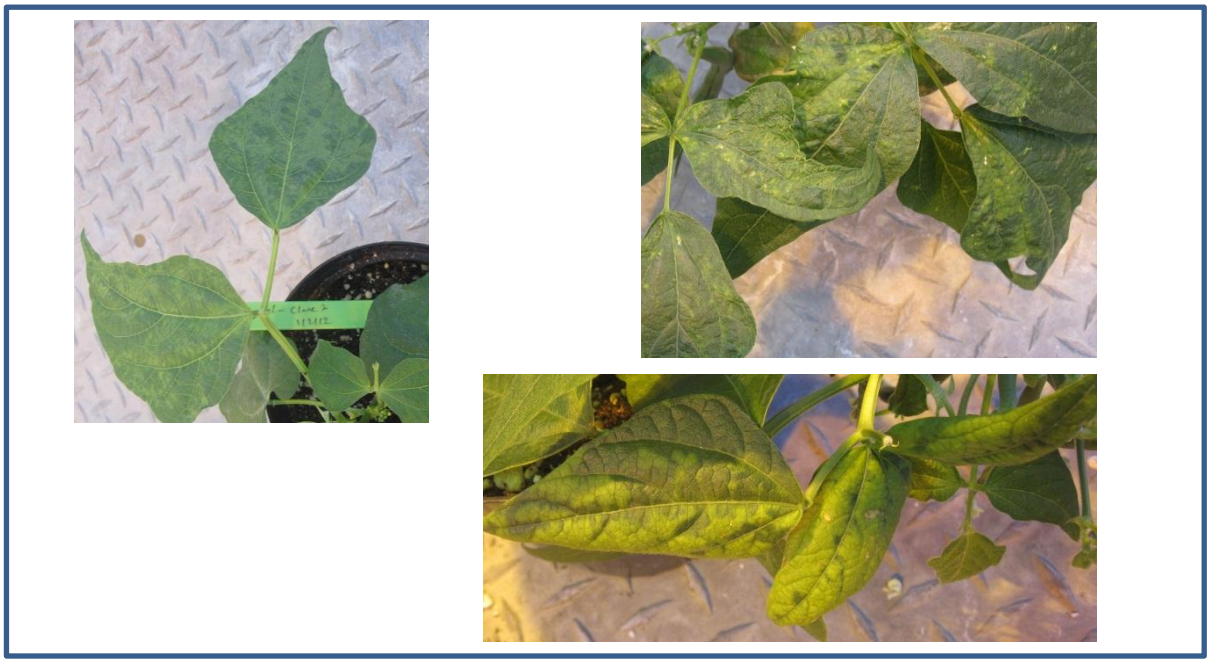


Figure 5. Symptoms caused by two, newly characterized satRNAs in combination with a legume CMV strain on snap bean (*Phaseolus vulgaris* cv. 'Hystyle')

Induced symptoms are shown at 7 dpi (a) and 14 dpi (b).

a

CMV IR-WI-satRNA

CMV Sb-WI-satRNA

b

CMV IR-WI-satRNA

CMV Sb-WI-satRNA

Figure 6. Effects of novel satRNAs on snap bean pod quality.

Pod distortion and a characteristic “C-shaped” pod symptom were consistently observed among CMV/satRNA combinations with either IR-WI or Sb-WI.

CMV+satRNA **CMV**

Figure 7. The presence of satRNA was confirmed in inoculated snap bean plants by RT-PCR.

M: 1Kb DNA ladder (Promega); 1) Wild-type CMV isolate from Wisconsin snap bean field; 2) Mock inoculated plants with phosphate buffer (pH 7.0); 3) CMV Ky strain containing satRNA (positive control); 4-6) Snap been plants inoculated by CMV alone, CMV plus Sb-WI-satRNA and CMV plus IR-WI-satRNA, respectively.

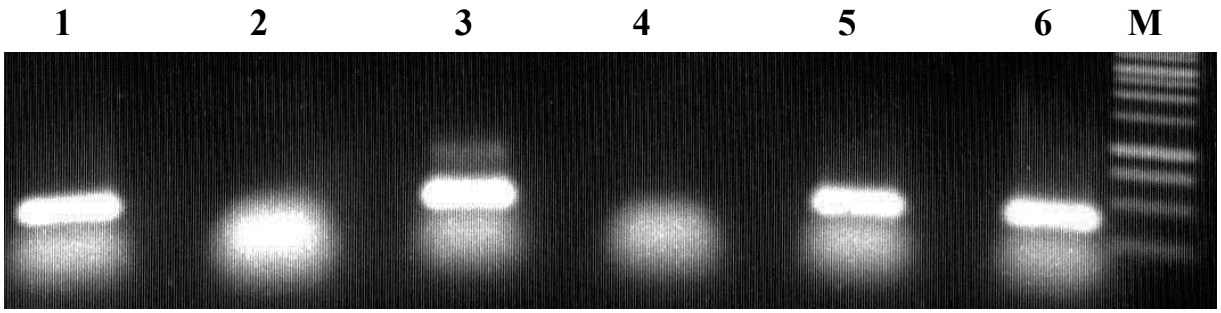


Figure 8. Symptoms induced by CMV alone (CMV) and in combination with the new satRNAs (CMV+IR-WI-satRNA and CMV+Sb-WI-satRNA) on tobacco (*N.tabacum* cv. 'Xanthi').

a) Mild symptoms were often observed over the whole inoculated plant, with only localized oak-leaf symptoms on or immediately above the inoculated leaf; b) Symptoms observed on the single inoculated leaf; b-a; Mock, b-b; CMV, b-c; CMV+ satRNA (IR-WI and Sb-WI) at 10 dpi, B-d; CMV+ satRNA (IR-WI and Sb-WI) at 20 dpi.

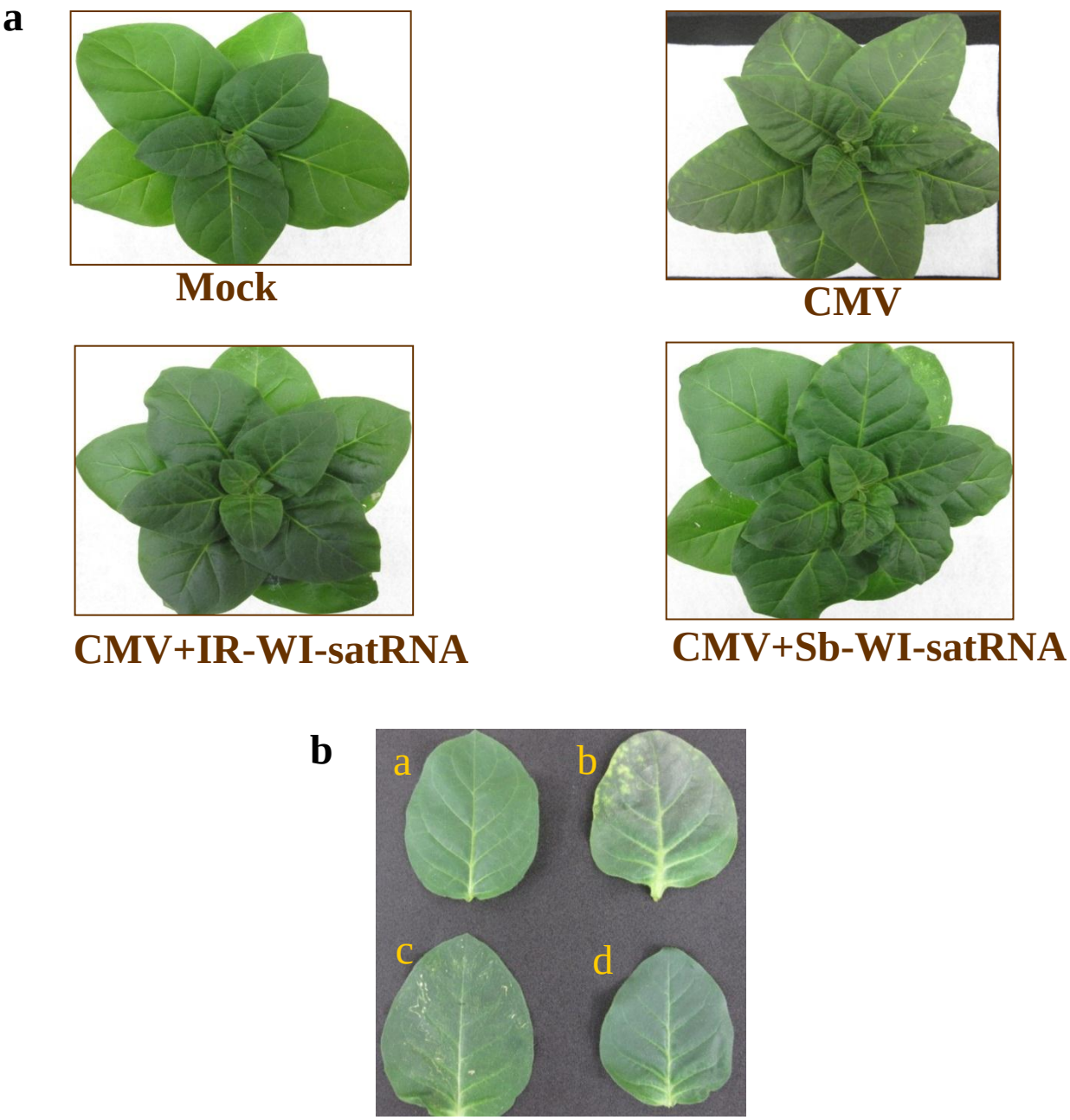
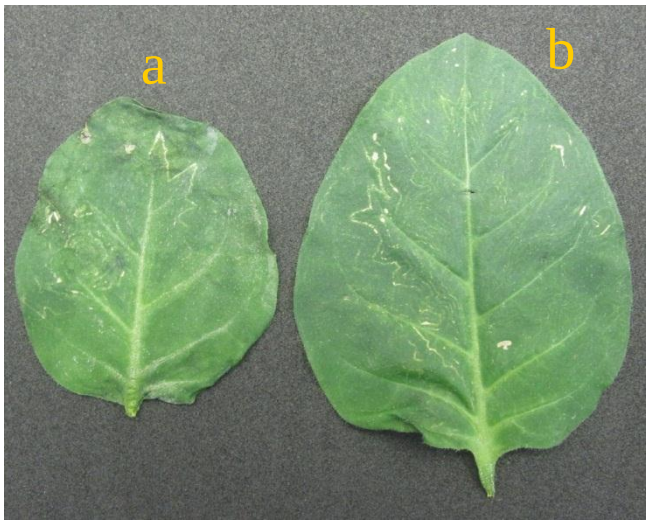


Figure 9. Symptoms induced by CMV in combination with IR-WI and Sb-WI-satRNA at 10 dpi in tobacco (*N.tabacum* cv. 'Xanthi').

Symptoms are illustrated on the leaves above mechanical inoculation points. a) Ring spot and oak leaf symptoms; b) oak leaf symptom.



References

1- Alonso-Prados, J. L., Aranda, M. A., Malpica, J. M., Garcı´a-Arenal, F. & Fraile, A. (1998). Satellite RNA of *cucumber mosaic cucumovirus* spreads epidemically in natural populations of its helper virus. *Phytopathology* **88**, 520–524.

2- Altschul, S. F., Gish, W., Miller, W., Mayers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403- 410.

3- Aranda, M. A., Fraile, A. & Garcı´a-Arenal, F. (1993). Genetic variability and evolution of the satellite RNA of *cucumber mosaic virus* during natural epidemics. *J Virol* **67**, 5896–5901.

4- Aranda, M. A, Fraile, A., Dopazo, J., Malpica, J. M. & Garcia-Arenal, F. (1997). Contribution of mutation and RNA recombination to the evolution of a plant pathogenic RNA. *J Mol Evol* **44**, 81-88

5- Betancourt, M., Fraile, A. & Garcı´a-Arenal, F. (2011). *Cucumber mosaic virus* satellite RNAs that induce similar symptoms in melon plants show large differences in fitness. *J Gen Virol* **92**, 1930–1938

6- Boccard, F. & Baulcombe, D. (1993). Mutational analysis of cis-acting sequences and gene function in RNA3 of *cucumber mosaic virus*. *Virology* **193**, 563–578.

7- Brigneti, G., Voinnet, O., Li, W. X., Ji, L. H., Ding, S. W. & Baulcombe, D. C. 1998. Viral pathogenicity determinants are suppressors of transgene silencing in *Nicotiana benthamiana*. *EMBO J* 17, 6739–6746.

8- Brown, T., Mackey, K. & Du, T. (2004). Analysis of RNA by Northern and Slot Blot Hybridization. *Curr Protoc Mol Biol* **67**, 4.9.1–4.9.19.

9- Canto, T., Prior, D. A. M., Hellwald, K. H., Oparka, K. J. & Palukaitis, P. (1997). Characterization of *cucumber mosaic virus*. IV. Movement protein and coat protein are both essential for cell-to-cell movement of *cucumber mosaic virus*. *Virology* **237**,237–248.

10- Collmer, C. W, Tousignant, M. E. & Kaper, J. M. (1983). *Cucumber mosaic virus*-associated RNA 5. X. The complete nucleotide sequence of a CARNA5 incapable of inducing tomato necrosis. *Virology* **127**, 230- 234

11- Crescenzi, A., Barbarossa, L., Cillo, F., Di Franco, A., Vovlas, N. & Gallitelli, D. (1993). Role of *cucumber mosaic virus* and its satellite RNA in the etiology of tomato fruit necrosis in Italy. *Arch Virol* **131**, 321–333.

12- Diaz-Pendon, J. A., Li, F., Li, W. X. & Ding, S. W. (2007). Suppression of antiviral silencing by *cucumber mosaic virus* 2b protein in *Arabidopsis* is associated with drastically reduced accumulation of three classes of viral small interfering RNAs. *Plant Cell* **19**, 2053-2063.

13- Ding, S. W., Anderson, B. J., Haase, H. R. & Symon, R. H. (1994). New overlapping gene encoded by the *cucumber mosaic virus* genome. *Virology* **198**, 593-601.

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

15- Doolittle, S. P. (1916). A new infectious mosaic disease of cucumber. *Phytopathology* **6**, 145–147

16- Edwards, M. C., Gonsalves, D. & Provvidenti, R. (1983). Genetic analysis of *cucumber mosaic virus* in relation to host resistance: location of determinants for pathogenicity to certain legumes and *lectuca saligna*. *Phytopathology* **73**, 269- 273

17- Escriu, F., Fraile, A. & García-Arenal, F. (2000). Evolution of virulence in natural populations of the satellite RNA of *Cucumber mosaic virus*. *Phytopathology* **90**, 480-485.

18- Escriu, F., Perry, K. L. & García-Arenal, F. (2000). Transmissibility of *Cucumber mosaic virus* by *Aphis gossypii* correlates with viral accumulation and is affected by the presence of its satellite RNA. *Phytopathology* **90**, 1068-1072.

19- Fischer, H. U. , Lockhart, B. E. L. 1976. A strain of *cucumber mosaic virus* isolated from cowpeas in Morocco. *Phytopathol Z* 85, 132-138.

20- Garcia-Arenal, F., Escriu, F., Aranda, M. A., Alonso-Prados, J. L., Malpica, J. M. & Fraile A. (2000). Molecular epidemiology of *cucmber mosaic virus* and its satellite RNA. *Virus Res* **71**, 1-8.

21- Garcı´a-Arenal, F. & Palukaitis, P. (1999). Structure and functional relationships of satellite RNAs of *cucumber mosaic virus*. *Curr Top Microbiol Immunol* **239**, 37–63.

22- Garcı´a-Luque, I., Kaper, J. M., Dı´az-Ruiz, J. R. & Rubio-Huertos, M. (1984). Emergence and characterisation of satellite RNAs associated with Spanish *cucumber mosaic virus* isolates. *J Gen Virol* **6**5, 539–547.

23- German, T. L., Thompson, A. & Wills, D. K. (2004). Statewide distribution of virus problems on processing beans. In. Proc 2004 Midwest Food Processors Association Processing crops Conference. Feb 23-24, La Cross, WI Department of Horticulture, University of Wisconsin.

24- Gildow, F. E., Shah, D. A., Sackett, W. M., Butzler, T., Nault, B. A. & Fleischer, S. J. (2008). Transmission efficiency of *Cucumber mosaic virus* by aphids associated with virus epidemics in snap bean. *Phytopathology* **98**, 1233-1241.

25- Gonsalves, D., Provvidenti, R. & Edwards, M. C. (1982). Tomato white leaf: the relation of an apparent satellite RNA and *cucumber mosaic virus*. *Phytopathology* **72**, 1533–38.

26- Gorden, K. H. J. & Symons, R. H. (1983). Satellite RNA of *cucumber mosaic virus* forms a secondary structure with partial 3`- terminal homologyto genomal RNAs. *Nucleic Acids Res* **11**, 947-960.

27- Grau, C. R., Stevenson, W. R. & Mondjana, A. M. (2002). Viruses causing losses on processing beans in the Midwest, In. Proc 2002 Wisconsin fertilizer, Aglime and Pest Management Conference, **41**, 248-256.

28- Grieco, F., Lanave, C. & Gallitelli, D. (1997). Evolutionary dynamics of *cucumber mosaic virus* satellite RNA during natural epidemics in Italy. *Virology* **229**, 166–174.

29- Habili, N. & Kaper, J. M. (1981). *Cucumber mosaic virus*-associated RNA 5. VII. Double-stranded form accumulation and disease attenuation in tobacco. *Virology* **112**, 250–261.

30- Hajimorad, M. R., Ghabrial, S. A. & Roossinck, M. J. (2009). De novo emergence of a novel satellite RNA of *cucumber mosaic virus* following serial passages of the virus derived from RNA transcripts. *Arch Viral* 154, 137-140

31- Hayes, R. J. & Buck, K. W. (1990). Complete replication of a eukaryotic virus RNAin vitro by a purified RNA-dependent RNA polymerase. *Cell* **63**, 363–368.

32- Hidaka, S., Hanada, K., Ishikawa, K. & Miura, K. (1988). Complete nucleotide sequence of two new satellite RNAs associated with *cucumber mosaic virus*. *Virology* **164**, 326–333.

33- Hou, W., Duan, C., Fang, R., Zhou, X. & Guo, H. (2011). Satellite RNA reduces expression of the 2b suppressor protein resulting in the attenuation of symptoms caused by *Cucumber mosaic virus* infection. *Mol Plant Path* DOI: 10.1111/J.1364-3703.2010.00696.X

34- Jagger, I. C. (1916). Experiments with the cucumber mosaic disease. *Phytopathology* **6**, 148.

35- Jord´a, C., Alfaro, A., Aranda, M. A., Moriones, E. & Garc´ıa-Arenal, F. (1992). Epidemic of *cucumber mosaic virus* plus satelliteRNA in tomatoes in eastern Spain. *Plant Dis* **76**, 363–66.

36- Kaper, J. M. 1982. Rapid synthesis of double-stranded *cucumber mosaic virus*-associated RNA 5: Mechanism controlling viral pathogenesis? *Biochem Biophys Res Commun* **105**, 1014–1022.

37- Kaper, J. M., Geletka, L. M., Wu, G. S. & Tousignant, M. E. (1995). Effect of temperature on *cucumber mosaic virus* satellite-induced lethal tomato necrosis is helper virus strain dependent. *Arch Virol* **140**, 65-74.

38- Kaper, J. M. & Tousignant, M. E. (1977). *Cucumber Mosaic Virus*-Associated RNA 5 I. Role of Host Plant and Helper Strain in Determining Amount of Associated RNA 5 with Virions. *Virology* **80**, 186-195.

39- Kaper, J. M., Tousignant, M. E. & Steen, M. T. (1988). *Cucumber mosaic virus* associated RNA 5. XI. Comparison of 14 CARNA 5 variants relates ability to induce tomato necrosis to a conserved nucleotide sequence. *Virology* **163**, 284–292.

40- Kaper, J. M. & Waterworth, H. E. (1977). *Cucumber mosaic virus*- associated RNA 5: Causal agent for tomato necrosis. *Science* **196**, 429- 431.

41- Kearney, C. M., Zitter, T. A. & Gonsalves, D. (1990). A field survey for serogroups and the satellite RNA of *cucumber mosaic virus*. *Phytopathology* **80**, 1238–1243.

42- Kim, C. H. & Palukaitis, P. (1997). The plant defense response to *cucumber mosaic virus* in cowpea is elicited by the viral polymerase gene and affects virus accumulation in single cells. *EMBO J* 16, 4060–4068.

43- Kosaka, Y., Hanada, K., Fukunishi, T. & Tochihara, H. (1989). *Cucumber mosaic virus* tomato necrotic disease in Kyoto prefecture. *Ann Phytopathol Soc Jpn* **55**, 229- 232.

44- Kurath, G. & Palukaitis, P. (1990). Serial passage of infectious transcripts of a *cucumber mosaic virus* satellite RNA clone results in sequence heterogeneity. *Virology* **176**, 8–15.

45- Kuwite, C. A. & Purcifull, D. E. (1982). Some properties of a *cucumber mosaic virus* strain isolated from winged bean in Florida. *Plant Dis* **66**, 1071-1073.

46- Larsen, R. C., Miklas, P. N., Eastwell, K. C., Grau, C. R. & Mondjana, A. (2002). A virus disease complex devastating late season snap bean production in the Midwest. *Annu Rep Bean Improv Coop* **45**, 36- 37.

47- Liao, Q., Zhu, L., Du, Z., Zeng, R., Feng, J. & Chen, J. (2007). Satellite RNA-mediated Reduction of *Cucumber Mosaic Virus* Genomic RNAs Accumulation in *Nicotiana tabacum.* *Acta Biochim Biophys Sin* **39**, 217–223

48- Lin, H. X., Rubio, L., Smythe A. & Falk, B. W. (2004). Molecular population genetics of *cucumber mosaic virus* in California: Evidence for founder effects and reassortments. *J Virol* **78**, 6666- 6675.

49- Meiners, J. P., Smith, F. F., Lawson, R. H., Waterworth, H. E. & Alconero, R. (1974). A strain of *cucumber mosaic virus* seed- borne in bean (*Phaseolus vulgaris* L.). (Abstr.). *Proc Phytopathol Soc* **1**, 137.

50- Moriones, E., Fraile, A. & Garcı´a-Arenal, F. (1991). Host-associated selection of sequence variants from a satellite RNA of *cucumber mosaic virus*. *Virology* **184**, 465–468.

51- Mossop, D. W. & Francki, R. I. B. (1977). Association of RNA 3 with aphid transmission of *cucumber mosaic virus*. *Virology* **81**, 177–181.

52- Mueller, E. E., Groves, R. L. & Gratton, C. (2012). Crop and non-crop plants as potential reservoir hosts of *Alfalfa mosaic virus* and *Cucumber mosaic virus* for spread to commercial snap bean. *Plant Dis* **96**, 506-514.

53- Nault, B. A., Shah, D. A., Dillard, H. R. & McFaul, A. R. (2004). Seasonal and spatial dynamics of alate aphid dispersal in snap bean fields in proximity to alfalfa and implication for virus management. *Environ Entomol* **33**, 1593- 1601.

54- Nault, B. A., Shah, D. A. & Taylor, A. G. (2006). Viruses and aphids everywhere in New York snap bean fields in 2005, pp. 74-76. *In* Proceedings of the 2006 Empire State Fruit and Vegetable EXPO. February 13-16, 2006. Cornell Coop. Exten., Syracuse, NY.

55- Palukaitis, P. (1988). Pathogenicity regulation by satellite RNAs of *cucumber mosaic virus*: minor nucleotide sequence changes alter host responses. *Mol Plant Microb Interact* **1**, 175- 181.

56- Palukaitis, P. & Garcia-Arenal, F. (2003). *Cucumoviruses*. *Adv Virus Res* **62**, 241-323

57- Palukaitis, P. & Roossinck, M. J. (1995). Variation in the hypervariable region of *cucumber mosaic virus* satellite RNAs is affected by the helper virus and the initial sequence context. *Virology* **206**, 765–768.

58- Palukaitis, P. & Roossinck, M. J. (1996). Spontaneous change of a benign satellite RNA of *cucumber mosaic virus* to a pathogenic variant. *Nat Biotechnol* **14**, 1264–1268.

59- Palukaitis, P., Roossinck, M.J., Dietzgen, R.G. & Francki, R.I.B. (1992). *Cucumber mosaic virus*. *Adv Virus Res* **41**, 281–348.

60- Peden, K. W. C. & Symons, R. H. (1973). *Cucumber mosaic virus* contains a functionally divided genome. *Virology* **53**, 487–492.

61- Perry, K. L., Zhang, L. & Palukaitis, P. (1998). Amino acid changes in the coat protein of *cucumber mosaic virus* differentially affect transmission by the aphids *Myzus persicae* and *Aphis gossypii*. *Virology* **242**, 204–210.

62- Perry, K. L., Zhang, L., Shintaku, M. H &, Palukaitis, P. (1994). Mapping determinants in *cucumber mosaic virus* for transmission by *Aphis gossypii*. *Virology* **205**, 591–595.

63- Provvidenti, R. (1976). Reaction of *Phaseolus* and *Macroptilium* species to a strain of *cucumber mosaic virus*. *Plant Dis Rep* **60**, 289-293

64- Richards, K. E., Jonard, G., Jacquemond, M. & Lot, H. (1978). Nucleotide Sequence of *Cucumber Mosaic Virus*-Associated RNA 5. *Virology* **89**, 395- 408.

65- Rizzo, T. M. & Palukaitis, P. (1990). Construction of full-length cDNA clones of *cucumber mosaic virus* RNAs 1, 2 and 3: generation of infectious RNA transcripts. *Mol Gen Genet* **222**, 249-256.

66- Roossinck, M. J. (2002). Evolutionary history of *cucumber mosaic virus* deduced by phylogenetic analyses. *J Virol* **76**, 3382–3387.

67- Roossinck, M. J. & Palukaitis, P. (1995). Genetic analysis of helper virus-specific selective amplification of *cucumber mosaic virus* satellite RNAs. *J Mol Evol* **40**, 25–29.

68- Roossinck, M. J., Sleat, D. & Palukaitis, P. (1999). Satellite RNAs of plant viruses: structures and biological effects. *Microbiol Rev* **56**, 265- 279.

69- Roossinck, M. J., Zhang, L. & Hellward, K. H. (1999). Rearrangements in the 5` nontranslated region and phylogenetic analyses of *cucumber mosaic virus* RNA 3 indicate radial evolution of three subgroups. *J Virol* **73**, 6752-6758.

70- Sambrook, J. & Russell, D. W. (2001). 'Molecular cloning. A laboratory manual', Cold Spring Harbor Laboratory Press.

71- Schwinghamer, M. W. & Symons, R. H. (1977). Translation of the four major RNA species of *cucumber mosaic virus* in plant and animal cell-free systems and in toad oocytes. *Virology* **79**, 88–108.

72- Simon, A. E., Roossinck, M. J. & Havelda, Z. (2004). Plant virus satellite and defective interfering RNAs: new paradigms for a new century. *Annu Rev Phytopathol* **42**, 415– 437

73- Skoric, D., Krajacis, M., Perica, M. C., Halupecki, E., Topic, J. & Lgrc-Barcia, J. (2000). *Cucumber mosaic virus* (*Cucumovirus*) and associated satRNA in weed species under the natural epidemic conditions of tomato lethal necrosis in Croatia. *Z Planzenk* 107, 304- 309.

74- Sleat, D. E. & Palukaitis, P. (1990). Site-directed mutagenesis of a plant viral satellite RNA changes its phenotype from ameliorative to necrogenic. *Proc Nati Acad Sci USA* **87**, 2946-2950.

75- Suzuki, M., Kuwata, S., Kataoka, J., Masuta, C., Nitta, N. & Takanami, Y. (1991). Functional analysis of deletion mutants of *cucumber mosaic virus* RNA3 using an in vitro transcription system. *Virology* **183**, 106–113.

76- Takanami, Y. A. (1981). A striking change in symptoms on *cucumber mosaic virus*-infected tobacco plants induced by a satellite RNA. *Virology* **109**, 120-126.

77- Wahyuni, W. S., Dietzgen, R. G., Hanada, K. & Francki, R. I. B. (1992). Serological and biological variation between and within subgroup I and II strains of *cucumber mosaic virus*. *Plant Pathol* **41**, 282-297.

78- Whipple, O. C. & Walker, J. C. (1941). Strains of *cucumber mosaic virus* pathogenic on bean and pea. *J Agric Res* **62**, 0027-0060.

Chapter 5

Conclusions and Future Directions

The world of science is full of mystery and uncertainty; and the answer to one question may reveal linkages to several other questions. This dissertation research is not an exception! We hope that the results obtained from this research have increased our knowledge about CMV genetic structure and components of the virus epidemiology, but it has also raised a number of questions which can help to formulate new hypotheses and ultimately new research projects. Here, I will highlight a portion of these questions and propose experiments that might be helpful towards their resolution.

The aim of this thesis was to better understand the molecular epidemiology and genetic variation among selected *Cucumber mosaic cucumovirus* (CMV) isolates. In the last decade, CMV epidemics occurred in discrete regions of the US; however, the reason(s) behind these virus outbreaks is not well known and several hypotheses have emerged to explain their occurrence. Candidate explanations for the emergence of this virus complex includes, but is not limited to, (1) the arrival and establishment of a novel aphid vector (*A. glycines*), (2) a shift in production practices in the commercial crop including planting dates, plant protection inputs, or selection of new varieties, (3) changes in the genetic structure of the virus populations or the emergence of novel virus strains resulting from introductions, (4) the co-association of pathogenic satRNA which could exacerbate symptom expression in the susceptible crop hosts. This dissertation research attempts to investigate two of these candidate hypotheses and further suggests explanations for the likelihood of their occurrence. Specifically, phylogenetic analyses describing the current and historical genetic diversity of CMV isolates were performed at the local, state and national levels in the US. Also, following the discovery of satRNA in CMV field isolates, disease symptom development assays were conducted in a greenhouse environment to determine the potential role of this RNA in symptom induction.

One significant conclusion that emerged from our investigation of CMV isolates suggests that there was a low likelihood of emergence of novel subgroup or strains of CMV that explains recent epidemics in snap bean. Here again, our observations suggest low genetic variation at both the local and national levels, and limited development of spatially explicit, structured populations. Even within the state of Wisconsin, the genetic diversity of CMV isolates in snap bean, a major host affected by the recent virus epidemics was low and the genetic structure of CMV was composed of only subgroup IA in this specific host. This observation further suggests the potential role of a founder effect, or the contribution of only a few discrete genotypes that largely describe the genetic composition of CMV in the state.

One additional explanation for these virus epidemics is the adaptation of a new quasispecies variant or subgroup (subgroup IA in this case) of CMV, specific to the new vector for this virus, the soybean aphid. In turn, this quasispecies variant or subgroup may have become more widely distributed due to the magnitude of flights often observed with this aphid vector. Sequence comparisons and phylogenetic analyses, however, illustrated that the recent virus outbreaks were not likely due to the invasion of a new subgroup or genotype of CMV in snap bean. A low frequency of genetic exchange illustrates that these events are counter-selected in CMV field isolates.

Furthermore, purifying selection was determined as the predominant natural force on the CMV genome among isolates included in these investigations and at both geographic levels. Together with a lack of observed shifts in the genetic structure of the CMV populations, these results may lessen concerns the use of transgenic plants as a very effective CMV control strategy. However, some positively selected codons were detected in the genomic regions of CMV which are

involved in plant host defense. This observation may, however, increase the potential for the emergence of resistance-breaking isolates of CMV.

Outlined in Chapter 2 of this thesis, three natural reassortants were recognized based on phylogenetic analyses. Additionally, there were at least three other isolates (113CAT90, AORU93 and GORU93) whose 1a gene belonged to the subgroup IA, whereas their associated 2a, 2b, MP and CP genes belonged to subgroup IB. We could not, however, strongly suggest that these were reassortants, because of our inability to amplify their associated 3`NTR regions using PCR. Hence, it would be worthwhile to amplify this region of the aforementioned isolates in future investigations to more fully characterize their phylogenetic position in the 3`NTR trees.

Phylogenetic analysis of CMV field isolates in Wisconsin suggests that subgroups IA and IB together constitute the genetic structure of this virus in this geographic region. However, the distribution of these subgroups was generally observed to be host specific. The only host plant in which we could detect subgroup IA was succulent snap bean (*Phaseolus vulgaris* L.), whereas the subgroup composition of CMV in pepper was a mixture of both IA and IB. One question that emerges from this observation is *"what, if any, biological significance or relevance may be behind this host-subgroup interaction?", "is subgroup IB incapable of systemic infection in snap bean?", "could there be functional differences in the binding affinities of these two subgroups to the principal aphid vector species (e.g. Aphis glycines Matsumura)?".* A relatively simple and replicated inoculation bioassay of succulent snap bean with extracted sap of subgroup IB (e.g. subgroup IB isolate from pepper) could provide an answer to this first question. To test the transmission efficiency question, we would propose to compare a few different aphid species (e.g. *A. glycines, Rhopalosiphum padi, R. maidis, Therioaphis trifolii*), using appropriate acquisition and inoculation access periods, with the two subgroups of CMV.

Another unique outcome of this research was the discovery of satellite RNA (satRNA) in association with natural populations of CMV in snap bean fields. Through sequence analysis, we were able to demonstrate that these CMV satRNAs in snap bean are novel and supplementary greenhouse experiments, using a satRNA- free CMV infectious clone, further demonstrated that this extra RNA can enhance CMV symptom induction in snap bean. However, the specific mechanism(s) by which satRNA influences symptoms in this particular crop is unclear. The results of a recent study have demonstrated that satRNA can reduce the expression of the 2b suppressor protein resulting in attenuation of induced symptoms by the helper virus (2), but how this interaction results in disease symptom enhancement is unknown. *Does satRNA increase expression of the 2b protein lead to the further enhancement, or intensification of symptoms?* To answer this question, it may be appropriate to measure the relative proportions of the 2b gene using qPCR during infection of the plant with the CMV+ satRNA construct and to compare this with the satRNA-free CMV infected host plant. Furthermore, it may be intriguing to learn about the up or down-regulation of host genes during infection using microarray analyses with extracted RNA from both wild-type and CMV+ satRNA constructs. Although, a common bean whole genome DNA microarray is not currently available, it has been shown that transcript profiling in common bean can be performed in an analogous system, specifically using the currently available soybean Gene Chip (3, 5). Presently, the whole genome sequence of *P. vulgaris* (common bean) is being pursued (1) and upon completion of this project, a common bean microarray can be synthesized. Additionally, after more than 40 years of research investigating the interaction of CMV satRNAs, a recent study has clearly demonstrated that satRNA uses the RNA silencing machinery in tobacco to target a host gene involved in chlorophyll synthesis and symptom induction (4). Although our investigations of snap bean,

CMV and CMV satRNAs, did not lead to predictable necrogenic or chlorogenic phenotypes, the severity of symptoms and shorter expression time in the presence of both satRNA variants is intriguing in this context. Therefore, one emerging hypothesis would suggest that satRNA may target the host defense genes. Hence, it would be worthwhile to search for the presence of short interfering RNAs (siRNAs) in infected plants containing CMV+ satRNA. Taken together, results obtained from both microarray and siRNA traces would provide sufficient, preliminary information to assist in designing future experiments.

References

1- Common Bean Coordinated Agricultural Project. [http://www.reeis.usda.gov/web/crisprojectpages/219849.html].

2- Hou, W., Duan, C., Fang, R., Zhou, X. & Guo, H. (2011). Satellite RNA reduces expression of the 2b suppressor protein resulting in the attenuation of symptoms caused by *Cucumber mosaic virus* infection. *Mol Plant Path* DOI: 10.1111/J.1364-3703.2010.00696.X

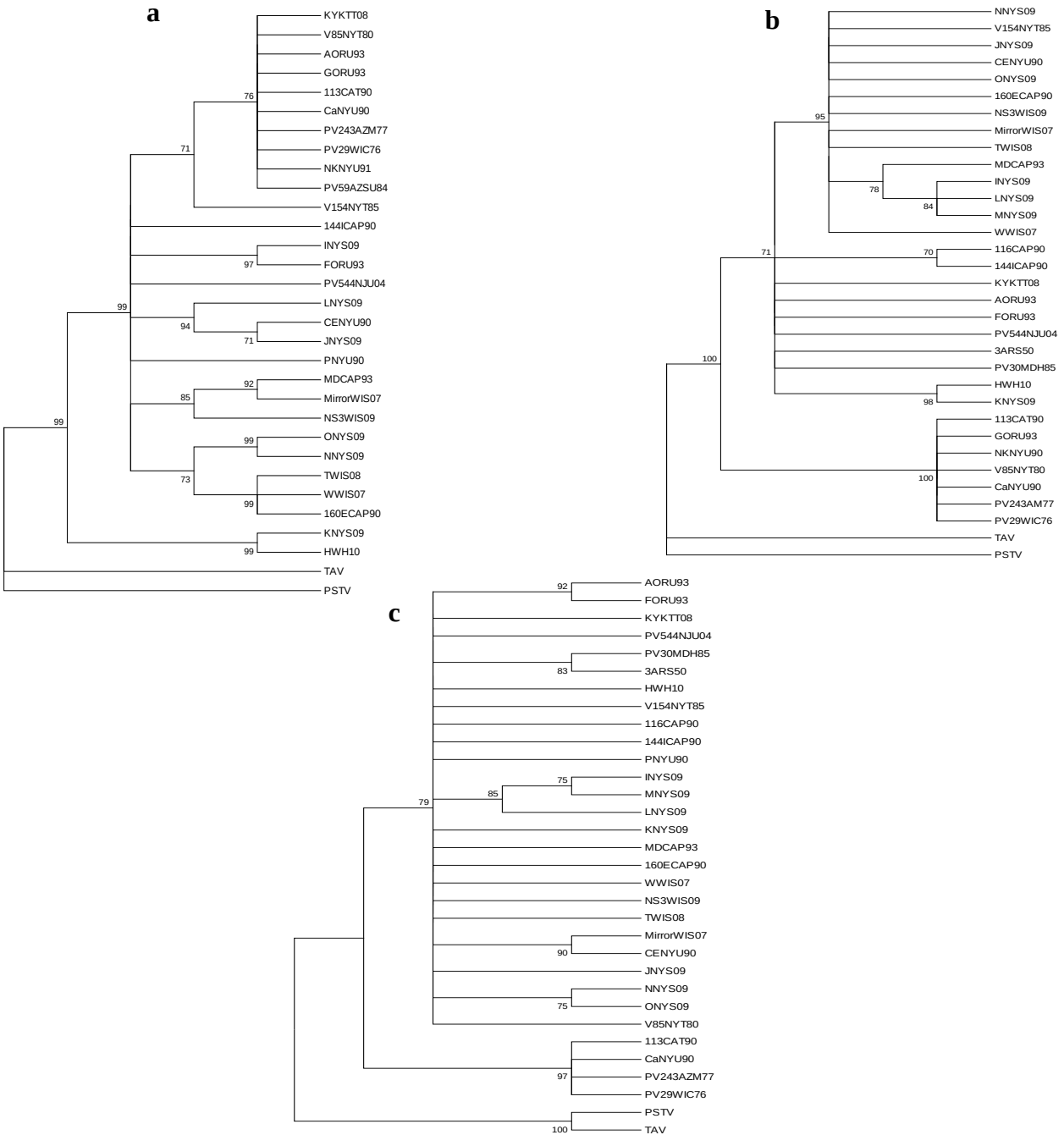
3- Mensack, M. M., Fitzgerald, V. K., Ryan, E. P., Lewis, M. R., Thompson, H. J. & Brick, M. A. (2010). Evaluation of diversity among common beans (*Phaseolus vulgaris L.*) from two centers of domestication using 'omics' technologies. *BMC Genomics* **11**, 686- 696.

4- Shimura, H., Pantaleo, V., Ishihara, T., Myojo, N. & Inaba, J. I. 2011. A Viral Satellite RNA Induces Yellow Symptoms on Tobacco by Targeting a Gene Involved in Chlorophyll Biosynthesis using the RNA Silencing Machinery. *PLoS Pathog* **7**, e1002021. doi:10.1371/journal.ppat.1002021

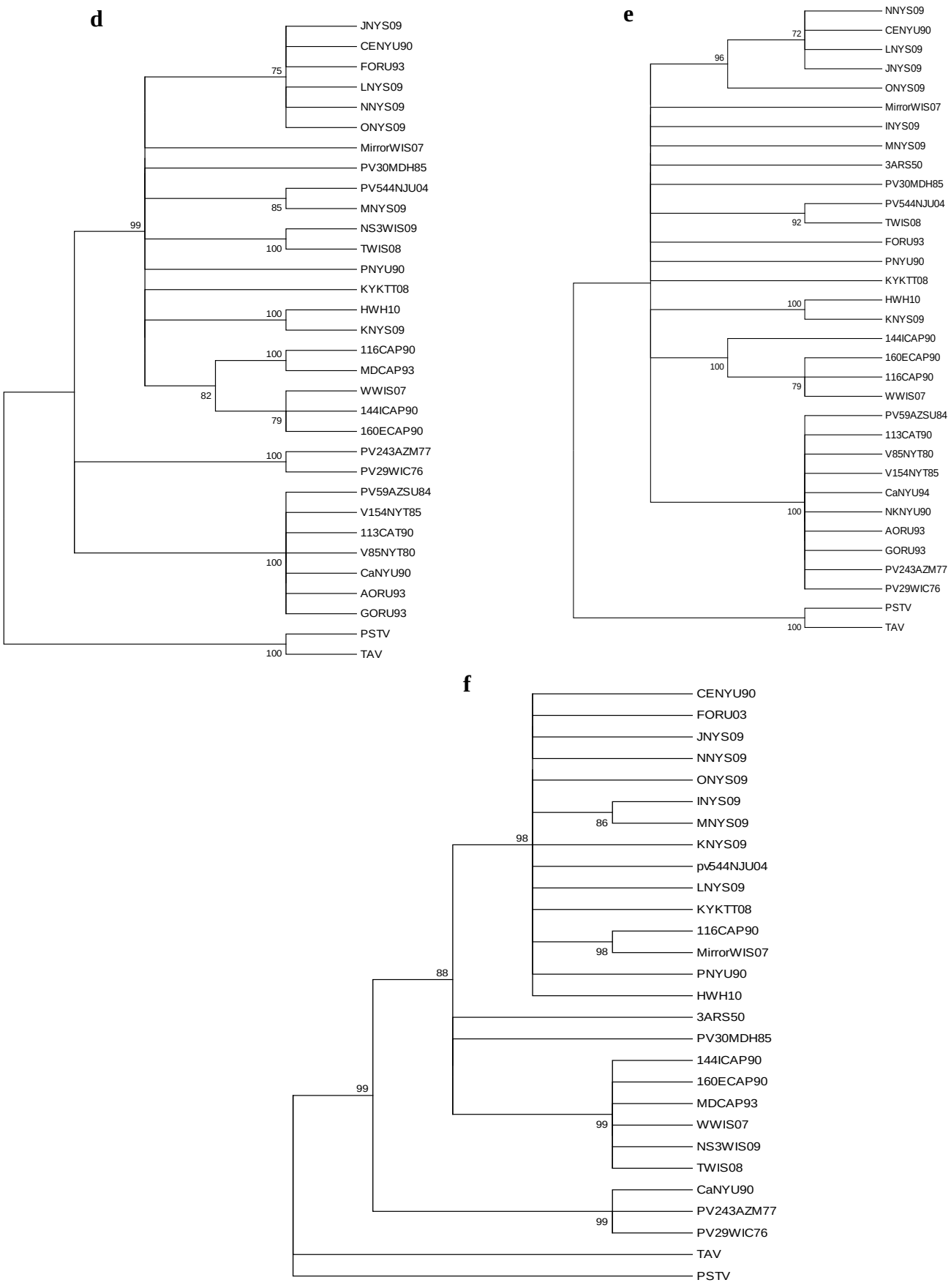
5- Yang, S. S., Valdés-López, O., Xu, W. W., Bucciarellil, B., Gronwald, J. W., Hernández, G. & Vance, C. P. (2010). RTerseaanrchs acrtricilept profiling of common bean (*Phaseolus vulgaris* L.) using the GeneChip® Soybean Genome Array: optimizing analysis by masking biased probes. *BMC Biol* **10**, 85- 100.

Appendix 1. Phylogenetic trees constructed for US CMV isolates.

The rooted condensed tree was constructed by the NJ method using the Kimura 2-parameter model of nucleotide substitution in MEGA version 5 based on a) 1a (Partial sequence); b) 2a (Partial sequence); c) 2b (Complete sequence); d) MP (Complete sequence); e) CP (Complete sequence); f) the complete 3`NTR of RNA3 sequence. The robustness of branching patterns was tested by 1,000 bootstrap replicates. Bootstrap values are indicated at nodes. All branches supported by < 70% were collapsed. Trees were rooted using PSTV and TAV as the out-group.



Appendix 1- continued



Appendix 2. Test statistics and parameter estimates assessed for genetic differentiation of CMV subpopulations between fields.

K_S*, Z and Snn are three permutation statistical tests of genetic differentiation. P-values <0.05 resulted in rejection of the null hypothesis (no genetic differentiation) between two subpopulations and these have been marked by asterisks. F_{ST} provides an estimate of the extent of genetic differentiation and gene flow. $F_{ST} > 0.33$ indicates infrequent gene flow.

Gene	Compared fields	K_S*	P-value	Z	P-value	Snn	P-value	F_{ST}
2b	J vs. Y	1.245	0.051	8.00	0.137	0.857	0.051	0.196
	J vs. H	0.934	0.024*	8.50	0.021*	1.000	0.024*	0.178
	J vs. DA	1.750	0.125	12.50	0.125	0.875	0.043*	0.026
	J vs. BD	1.470	0.029*	9.25	0.029*	0.875	0.025*	0.230
	J vs. C	1.107	0.139	12.56	0.131	0.750	0.139	0.650
	J vs. DB	1.165	0.030*	9.50	0.030*	0.875	0.030*	0.270
	J vs. S1C	1.558	0.054	9.00	0.054	0.750	0.059	0.013
	J vs. B	1.533	0.143	8.00	0.080	0.850	0.061	0.150
	J vs. K24	1.397	0.156	13.74	0.077	0.880	0.156	0.034
	J vs. CA	1.921	0.056	18.33	0.145	0.750	0.088	0.250
	J vs. HP	0.935	0.156	10.25	0.156	0.750	0.156	0.660
	J vs. SC2	0.934	1.000	8.50	1.000	1.000	1.000	0.088
	J vs. E3	1.903	0.480	12.25	0.226	0.750	0.220	-0.248
	Y vs. H	0.000	1.000	4.00	1.000	1.000	1.000	0.520
	Y vs. DA	1.090	0.141	9.66	0.391	0.875	0.058	0.133
	Y vs. BD	0.723	0.054	7.83	0.111	0.890	0.054	0.530
	Y vs. C	0.231	0.159	5.95	0.159	1.000	0.094	0.555
	Y vs. DB	0.308	0.408	4.00	0.408	1.000	0.408	0.333
	Y vs. S1C	0.832	0.052	7.66	0.164	0.850	0.054	0.476
	Y vs. B	0.462	0.110	4.00	0.092	0.833	0.110	0.400
	Y vs. K24	0.812	0.043*	12.65	0.221	0.940	0.014*	0.400
	Y vs. CA	1.316	0.069	5.00	0.069	0.850	0.069	0.700
	Y vs. HP	0.000	0.031*	4.00	0.031*	1.000	0.031*	0.680
	Y vs. SC2	0.000	1.000	4.00	1.000	1.000	1.000	0.456
	Y vs. E3	1.268	0.229	8.000	0.143	0.877	0.053	0.083
	H vs. DA	0.818	0.133	6.66	0.133	0.910	0.158	0.388
	H vs. BD	0.542	0.142	5.50	0.158	1.000	0.142	0.533
	H vs. C	0.173	0.024*	7.42	0.024*	1.000	0.024*	0.750

2b	H vs. DB	0.231	0.459	5.50	0.459	1.000	0.459	0.333
	H vs. S1C	0.624	0.020*	5.50	0.020*	1.000	0.020*	0.496
	H vs. B	0.308	0.037*	4.00	0.037*	1.000	0.037*	0.530
	H vs. K24	0.649	0.006*	7.34	0.006*	1.000	0.006*	0.433
	H vs. CA	0.987	0.035*	6.50	0.035*	0.870	0.035*	0.730
	H vs. HP	0.000	0.029*	5.50	0.029*	1.000	0.029*	0.722
	H vs. SC2	0.000	0.024*	5.30	0.024*	1.000	0.024*	0.540
	H vs. E3	0.951	0.029*	9.50	0.029*	1.000	0.029*	0.153
	DA vs. BD	1.361	0.028*	7.00	0.028*	1.000	0.028*	0.440
	DA vs. C	0.991	0.023*	7.70	0.023*	1.000	0.023*	0.537
	DA vs. DB	1.041	0.059	7.58	0.199	1.000	0.059	0.224
	DA vs. S1C	1.442	0.096	9.58	0.096	0.900	0.096	0.238
	DA vs. CA	1.805	0.063	8.20	0.063	0.610	0.025*	0.602
	DA vs. B	1.399	0.362	6.44	0.699	1.000	0.054	0.095
	DA vs. K24	1.304	0.149	12.26	0.144	1.000	0.149	0.130
	DA vs. HP	0.818	0.038*	5.50	0.038*	1.000	0.038*	0.588
	DA vs. SC2	0.818	0.028*	8.16	0.028*	1.000	0.028*	0.140
	DA vs. E3	1.769	0.104	12.50	0.104	0.875	0.062	0.024
	BD vs. C	0.716	0.023*	7.65	0.023*	1.000	0.023*	0.523
	BD vs. DB	0.773	0.196	12.41	0.196	0.610	0.196	0.250
	BD vs. S1C	1.167	0.056	6.91	0.011	0.870	0.056	0.428
	BD vs. B	1.031	0.232	7.83	0.122	0.670	0.176	0.310
	BD vs. K24	1.080	0.254	15.36	0.176	0.620	0.151	0.166
	BD vs. CA	1.529	0.031*	6.83	0.031*	0.870	0.031*	0.660
	BD vs. HP	0.542	0.027*	5.50	0.027*	1.000	0.027*	0.499
	BD vs. SC2	0.542	0.029*	5.50	0.029*	1.000	0.029*	0.533
	BD vs. E3	1.494	0.024*	10.00	0.510	0.875	0.024*	0.138
	C vs. DB	0.404	0.036*	7.55	0.036*	1.000	0.036*	0.588
	C vs. S1C	0.797	0.028*	7.62	0.032*	1.000	0.028*	0.686
	C vs. B	0.539	0.023*	6.06	0.230	1.000	0.023*	0.581
	C vs. K24	0.788	0.005*	9.50	0.005*	1.000	0.005*	0.600
	C vs. CA	1.160	0.036*	10.62	0.036*	1.000	0.036*	0.078
	C vs. HP	0.173	1.000	17.80	1.000	0.430	1.000	0.000
	C vs. SC2	0.173	0.021*	7.40	0.021*	1.000	0.021*	0.600
	C vs. E3	1.124	0.132	10.00	0.132	0.718	0.135	0.553
	DB vs. S1C	0.855	0.025*	6.66	0.081	0.870	0.051	0.566
	DB vs. B	0.616	0.149	7.00	0.087	0.710	0.332	0.466
	DB vs. K24	0.834	0.116	14.99	0.109	0.660	0.176	0.384
	DB vs. CA	1.218	0.135	6.50	0.148	0.870	0.170	0.469
	DB vs. HP	0.231	0.421	5.50	0.421	1.000	0.421	0.333
	DB vs. SC2	0.231	0.436	5.30	0.436	1.000	0.436	0.333

2b	DB vs. E3	1.182	0.025*	9.30	0.025*	0.875	0.028*	0.333
	K24 vs. CA	1.439	0.010*	9.01	0.010*	0.880	0.025*	0.667
	K24 vs. B	1.043	0.584	14.80	0.699	0.620	0.290	-0.171
	K24 vs. S1C	1.149	0.151	11.57	0.151	0.880	0.151	0.124
	K24 vs. HP	0.649	0.011*	7.34	0.011*	1.000	0.011*	0.481
	K24 vs. SC2	0.649	0.442	9.26	0.442	1.000	0.258	0.333
	K24 vs. E3	1.494	0.722	15.60	0.722	0.380	0.729	-0.013
	CA vs. B	1.624	0.069	5.11	0.069	0.800	0.050	0.681
	CA vs. HP	0.987	0.372	8.50	0.412	1.000	0.256	0.082
	CA vs. S1C	1.611	0.059	7.10	0.059	0.750	0.127	0.666
	CA vs. SC2	0.987	0.133	6.50	0.147	0.870	0.133	0.660
	CA vs. E3	1.987	0.128	11.30	0.141	0.870	0.132	0.251
	S1C vs. SC2	0.624	0.147	8.50	0.147	1.000	0.147	0.083
	S1C vs. E3	1.575	0.087	11.80	0.087	0.875	0.087	0.049
	S1C vs. B	1.140	0.056	7.11	0.103	0.800	0.049*	0.200
	S1C vs. HP	0.624	0.019*	5.50	0.019*	1.000	0.019*	0.677
	HP vs. B	0.308	0.150	4.00	0.164	1.000	0.095	0.500
	HP vs. SC2	0.000	0.029*	5.50	0.029*	1.000	0.029*	0.489
	HP vs. E3	0.951	0.023*	9.50	0.023*	1.000	0.023*	0.088
	SC2 vs. E3	0.000	1.000	5.50	1.000	1.000	1.000	0.431
	B vs. SC2	0.308	0.432	5.33	0.432	1.000	0.124	0.000
	B vs. E3	1.576	1.000	10.33	1.000	0.857	0.121	0.000
CP	J vs. Y	2.093	0.057	8.555	0.158	0.857	0.780	0.119
	J vs. H	1.387	0.033*	8.500	0.033*	1.000	0.033*	0.324
	J vs. DA	2.431	0.039*	9.333	0.255	0.714	0.130	0.214
	J vs. BD	1.387	0.027*	10.083	0.027*	0.750	0.096	0.095
	J vs. C	1.541	0.152	7.833	0.152	0.714	0.152	0.657
	J vs. DB	1.156	0.021*	9.500	0.021*	0.875	0.021*	0.241
	J vs. S1C	2.422*	0.311	9.777	0.353	0.714	0.142	0.025
	J vs. B	2.074	0.375	10.222	0.432	0.857	0.053	-0.016
	J vs. K24	1.695	0.354	8.777	0.196	0.785	0.123	0.073
	J vs. CA	2.387	0.225	8.888	0.225	0.571	0.225	0.112
	J vs. HP	1.817	0.140	7.000	0.088	0.571	0.265	0.645
	J vs. SC2	1.541	0.468	9.333	0.298	0.857	0.052	0.051
	J vs. E3	2.66	0.527	10.444	0.527	0.571	0.399	-0.277
	Y vs. H	0.860	0.025*	4.000	0.025*	1.000	0.025*	0.441
	Y vs. DA	2.164	0.169	6.000	0.169	0.777	0.103	0.250

CP	Y vs. BD	0.860	0.024*	4.444	0.024*	1.000	0.024*	0.314
	Y vs. C	0.828	0.094	2.500	0.094	1.000	0.094	0.266
	Y vs. DB	0.552	0.032*	6.666	0.032*	0.857	0.032*	0.153
	Y vs. S1C	2.149	0.290	7.000	0.462	0.500	0.389	-0.021
	Y vs. B	1.627	0.198	6.500	0.422	0.830	0.198	0.192
	Y vs. K24	1.059	0.101	4.500	0.101	1.000	0.101	0.250
	Y vs. CA	2.097	0.207	4.166	0.105	0.833	0.207	0.477
	Y vs. HP	1.242	0.087	2.500	0.087	1.000	0.087	0.235
	Y vs. SC2	0.828	0.087	4.500	0.087	1.000	0.087	0.214
	Y vs. E3	2.514	0.312	6.666	0.312	0.666	0.177	0.028
	H vs. DA	1.198	0.028*	6.555	0.028*	0.857	0.073	0.134
	H vs. BD	0.462	0.036*	5.500	0.036*	1.000	0.094	0.332
	H vs. C	0.308	0.410	4.000	0.410	1.000	0.410	0.332
	H vs. DB	0.231	0.432	5.500	0.432	1.000	0.282	0.323
	H vs. S1C	1.180	0.020*	4.000	0.020*	1.000	0.020*	0.400
	H vs. B	0.840	0.027*	4.000	0.027*	1.000	0.027*	0.543
	H vs. K24	0.462	0.198	4.000	0.198	1.000	0.198	0.428
	H vs. CA	1.153	0.020*	4.888	0.020*	0.857	0.075	0.540
	H vs. HP	0.584	0.242	4.000	0.146	1.000	0.099	0.350
	H vs. SC2	0.038	0.452	4.000	0.452	1.000	0.452	0.333
	H vs. E3	1.432*	0.037*	6.888	0.037*	1.000	0.037*	0.117
	DA vs. BD	1.198	0.029*	6.666	0.029*	0.857	0.084	0.477
	DA vs. C	1.335	0.101	2.500	0.101	1.000	0.101	0.404
	DA vs. DB	0.890	0.032*	6.666	0.032*	0.857	0.032*	0.417
	DA vs. S1C	2.656	0.309	5.333	0.214	0.666	0.309	0.270
	DA vs. CA	2.604	0.194	4.500	0.194	0.666	0.589	0.393
	DA vs. B	2.135	0.102	5.833	0.322	0.666	0.618	0.329
	DA vs. K24	1.566	0.103	5.500	0.204	0.833	0.103	0.516
	DA vs. HP	1.749	0.100	2.500	0.100	1.000	0.100	0.387
	DA vs. SC2	1.335	0.119	5.500	0.409	0.833	0.119	0.517
	DA vs. E3	3.022	0.315	6.750	0.409	0.666	0.315	0.010
	BD vs. C	0.308	0.447	4.000	0.447	1.000	0.235	0.333
	BD vs. DB	0.231	0.401	5.500	0.401	1.000	0.401	0.333
	BD vs. S1C	1.180	0.162	6.111	0.162	0.857	0.097	0.382
	BD vs. B	0.840	0.224	5.777	0.322	1.000	0.224	0.043
	BD vs. K24	0.462	0.190	4.000	0.190	1.000	0.190	0.428
	BD vs. CA	1.153	0.023*	4.888	0.023*	0.857	0.023*	0.500
	BD vs. HP	0.584	0.391	4.000	0.391	1.000	0.391	0.000
	BD vs. SC2	0.308	0.432	4.000	0.432	1.000	0.412	0.333
	BD vs. E3	1.432	0.036*	7.111	0.036*	0.857	0.092	0.117
	C vs. DB	0.000	0.084	4.000	0.095	1.000	0.084	0.487

CP	C vs. S1C	1.321	0.095	2.500	0.095	1.000	0.905	0.333
	C vs. B	0.799	0.121	2.500	0.121	1.000	0.121	0.501
	C vs. K24	0.231	0.097	2.500	0.097	1.000	0.097	0.500
	C vs. CA	1.268	1.000	7.000	1.000	0.400	1.000	0.000
	C vs. HP	0.414	0.401	6.750	0.401	0.416	0.401	0.000
	C vs. SC2	0.000	0.1000	2.500	0.1000	1.000	0.1000	0.532
	C vs. E3	1.686	0.400	5.750	0.400	0.666	0.400	0.446
	DB vs. S1C	0.880	0.034*	7.111	0.034*	0.714	0.034*	0.130
	DB vs. B	0.532	0.024*	6.666	0.024*	0.857	0.024*	0.523
	DB vs. K24	0.154	0.131	4.000	0.131	1.000	0.131	0.620
	DB vs. CA	0.845	0.146	4.888	0.146	0.857	0.124	0.546
	DB vs. HP	0.276	0.131	4.000	0.131	1.000	0.131	0.252
	DB vs. SC2	0.000	0.099	4.000	0.099	1.000	0.099	0.147
	DB vs. E3	1.124	0.027*	7.333	0.027*	0.714	0.027*	0.100
	K24 vs. CA	1.499	0.105	3.500	0.105	0.833	0.105	0.575
	K24 vs. B	1.030	0.370	6.333	0.370	0.611	0.189	0.083
	K24 vs. S1C	1.552	0.179	5.916	0.179	0.583	0.378	0.275
	K24 vs. HP	0.645	0.103	2.500	0.103	1.000	0.103	0.250
	K24 vs. SC2	0.231	0.383	5.500	0.383	0.666	0.383	0.500
	K24 vs. E3	1.917	0.213	5.916	0.213	0.583	0.395	0.110
	CA vs. B	2.068	0.088	4.166	0.088	0.666	0.622	0.467
	CA vs. HP	1.683	1.000	7.166	1.000	0.388	1.000	0.000
	CA vs. S1C	2.590	0.319	4.666	0.104	0.500	0.406	0.394
	CA vs. SC2	1.268	0.079	3.500	0.079	0.833	0.079	0.576
	CA vs. E3	2.955	0.707	7.083	0.0407	0.250	1.000	-0.174
	S1C vs. SC2	1.321	0.381	6.250	0.381	0.541	0.381	0.259
	S1C vs. B	2.120	1.000	7.333	1.000	0.250	1.000	-0.139
	S1C vs. HP	1.735	0.106	2.500	0.106	1.000	0.106	0.340
	S1C vs. E3	3.007	0.808	7.250	0.808	0.250	0.808	-0.094
	HP vs. B	1.213	0.091	2.500	0.091	1.000	0.091	0.286
	HP vs. SC2	0.414	0.100	2.500	0.100	1.000	0.100	0.477
	HP vs. E3	2.100	0.394	5.666	0.202	0.500	0.394	0.436
	B vs. SC2	0.799	1.000	7.000	1.000	1.000	1.000	0.000
	B vs. E3	2.485	1.000	7.416	1.000	0.250	1.000	-0.085
	SC2 vs. E3	1.686	0.387	6.250	0.387	0.541	0.411	0.096

Appendix 3. Accession numbers and country of origin of reference isolates and out-groups.

Strain	Accession numbers			Origin	Subgroup
	RNA1	RNA2	RNA3		
Fny	D00356	D00355	D10538	United States	IA
Y	D12537	D12538	D12499	Japan	IA
O	-	D10209	D00385	Japan	IA
Mf	AJ276479	AJ276480	AJ276481	South Korea	IA
Leg	D16403	D16406	D16405	Japan	IA
Nt9	D28778	D28779	D28780	Taiwan	IB
Tfn	Y16924	Y16925	Y16926	Italy	IB
IA	AB042292	AB042293	AB042294	Indonesia	IB
Ix	U20220	U20218	U20219	Philippines	IB
SD	AF071551	D86330	AB008777	China	IB
Q	X02733	X00985	M21464	Australia	II
LS	AF416899	AF416900	AF127976	United States	II
LY	AF198101	AF198102	AF198103	Australia	II
Trk7	AJ007933	AJ007934	L15336	Hungary	II
S	Y10884	Y10885	U37227	South Africa	II
2AIIL	-	-	AJ271416	United States	IB
OHW	-	-	U31220	United States	IB
HW2	-	-	U31219	United States	IB
ER-PSV	U15728	U15729	U15730	United States	-
TAV	D10044	D10663	AJ277268	Australia	-