THE EFFECT OF WITHIN-HOST VIRUS POPULATION GROWTH AND INTERSPECIFIC COMPETITION ON APHID TRANSMISSION AND POPULATION STRUCTURE OF BARLEY YELLOW DWARF VIRUS

A Dissertation
Presented to the Faculty of the Graduate School of Cornell University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

by
Gerod Sharper Hall
January 2007
THE EFFECT OF WITHIN-HOST VIRUS POPULATION GROWTH AND
INTERSPECIFIC COMPETITION ON APHID TRANSMISSION AND
POPULATION STRUCTURE OF BARLEY YELLOW DWARF VIRUS

Gerod Sharper Hall, Ph.D.
Cornell University 2007

Barley yellow dwarf virus (BYDV) (Luteoviridae) species PAV and PAS are ecologically similar in that they share aphid vectors and host species, but in agricultural fields in New York State we found that the prevalence of PAV was three times greater than that of PAS. To determine if differences in within-host population growth rate affect vector transmission efficiency, disease spread and the distribution of virus types in the host community, this study evaluated the biological characteristics of PAV and PAS species isolates in common agricultural hosts and the outcome of competitive interactions between species in mixed infections. In singly infected plants, PAS population size was 20% greater than that of PAV at 8 days post inoculation (DPI), but by 33 DPI the population size of PAV was 10% greater than that of PAS. In doubly infected plants, by 33 DPI the population size of PAV was 40% greater than that of PAS.

There was no difference in the transmission efficiency of PAV and PAS by Rhopalosiphum padi from singly infected plants at 30 DPI. But when transmission assays were performed 60 DPI, the transmission success of PAV was significantly greater than that of PAS. The greater transmission efficiency of PAV at late stages of infection did not translate to greater spread of PAV isolates in barley, oat or wheat plots in the field experiment. In the field, the susceptibility of plants significantly declined 18 days post plant emergence, suggesting that the development of resistance as plants matured may have arrested virus spread before asymmetry in the distribution of PAV and PAS could occur. In general, disease spread further in wheat than oat or
barley plots. Taken together these results suggest that the identity of host species and vector population dynamics in relation to the availability of susceptible hosts are key determinants of the disease prevalence in the host community. Virus multiplication within hosts may influence the relative abundance of PAV and PAS in natural populations if there is a greater likelihood of virus transmission at different stages of host infection.
BIOGRAPHICAL SKETCH

Gerod Sharper Hall was born to Florence Grayer-Walker and Eugene Gerod Hall in Detroit, Mich. Gerod’s life experiences as well as the morals and values instilled by his parents have lead him to become an extraordinarily compassionate, sincere and contemplative person. Gerod is well known for his giving spirit, his vivaciousness, his positive outlook on life, his all around good cheer and of course his humility.
ACKNOWLEDGMENTS

I thank my committee chair Alison (Sunny) Power for her support, generosity and faith in my ability to succeed. Sunny has been an amazing example of how I might use my training as scientist to aid people and natural communities. I also thank my other committee members John Losey and Michael Milgroom who have been instrumental in my intellectual development. I thank all members of the Power Laboratory for their completely indispensable help with the day-to-day work one must do to execute a research project and I also thank them for their continued friendship. I thank my collaborator Damon Little without whom chapters two-four would never have materialized and without whom my time as a graduate student and my future on the planet are wholly unimaginable.
# TABLE OF CONTENTS

Biographical sketch

Acknowledgments

List of Tables

List of Figures

CHAPTER ONE: Selective Constraint and Genetic Differentiation in Geographically Distant Barley yellow dwarf virus Populations

Abstract

Introduction

Materials and Methods

Results

Discussion

References

CHAPTER TWO: Quantitative sequencing: A new method for analyzing the within-host dynamics of closely related virus genotypes

Abstract

Introduction

Materials and Methods

Results

Discussion

References
CHAPTER THREE: *Vector dynamics in relation to host phenology will influence the disease prevalence and population structure of Barley yellow dwarf virus.*

Abstract.................................................................46
Introduction............................................................47
Materials and Methods.............................................49
Results.................................................................55
Discussion.............................................................59
References..............................................................65

CHAPTER FOUR: *Interference competition between Barley yellow dwarf virus species affects vector transmission.*

Abstract.................................................................69
Introduction............................................................70
Materials and Methods.............................................72
Results.................................................................76
Discussion.............................................................80
References..............................................................87

Appendix..................................................................90
LIST OF TABLES

Table 1.1. Percent nucleotide identity between the coat protein and polymerase gene of BYDV PAV and PAS

Table 1.2. Nucleotide diversity per site in the polymerase, coat protein and movement protein gene of BYDV-PAV and PAS in New York and Morocco

Table 1.3. Maximum likelihood analysis of codon substitution for detection of positive selection in the polymerase, coat protein and movement protein gene of BYDV-PAV and PAS

Table 1.4. Population statistics and neutrality tests based on variation in the polymerase, coat protein and movement protein of BYDV-PAV and PAS

Table 1.5. Relative selective constraints (dN/dS+constant) calculated for polymerase, coat protein and movement protein of BYDV-PAV and PAS

Table 2.1. Line of best fit when the frequency of BYDV-PAV template in the standard sample was regressed onto PAV as a proportion of the total peak area or height

Table 2.2. Sum of the squared differences between expected and calculated PAV concentration in a group of ten BYDV-PAV/PAS template mixtures

Table 3.1. Number of inoculated virus source plants that became infected with BYDV-PAV and PAS isolates in field plots
Table 3.2. Parameter estimates for logistic regression model of the probability that the inoculated virus source plants became infected…………………………………………………60

Table 4.1. Number of plants that became infected with BYDV-PAV and PAS after inoculation by aphids previously fed on plants in a given inoculation treatment……83
LIST OF FIGURES

Figure 1.1. Maximum likelihood trees of Luteoviridae species isolates calculated using partial polymerase, complete coat protein, complete movement protein or complete genome nucleotide sequences..............................................................12

Figure 2.1. Standard curves for peak measurements at polymorphic sites 2074 and 2088..................................................................................................................................................37

Figure 2.2. Proportion of PAV template in double inoculated plants harvested 33 days post inoculation with PAV.........................................................................................................................39

Figure 2.3. Proportion of PAS template in doubly inoculated plants harvested 33 days post inoculation with PAS..........................................................................................................................41

Figure 3.1. Percent of barley, oat and wheat plants that became infected with BYDV-PAV or PAS when inoculated using two or eight aphids carrying the appropriate strain........................................................................................................................................56

Figure 3.2. Relative concentration of BYDV-PAV template in wheat at 8, 20, 33 and 45 days post inoculation........................................................................................................................................57

Figure 3.3. Relative concentration of BYDV-PAV template in barley, oat and wheat plants harvested 14 days post inoculation........................................................................................................58
Figure 3.4. Mean distance BYDV-PAV and PAS isolates spread from inoculated source plants in barley, oat and wheat plants in field plots........................................63

Figure 4.1. Mean proportion of BYDV-PAV template when PAV was the protecting or challenging virus.................................................................77

Figure 4.2. Mean proportion of BYDV-PAS template when PAS was the protecting or challenging virus.................................................................79
CHAPTER ONE

Selective Constraint and Genetic Differentiation in Geographically Distant Barley yellow dwarf virus Populations

Abstract

Numerous studies have documented molecular variability in plant virus populations but few have assessed the relative contribution of natural selection and genetic drift in generating the observed pattern of diversity. To this end we examine how gene function, environment and phylogenetic history affect genetic diversity and population structure of the PAV and PAS species of barley yellow dwarf virus (Luteoviridae).

Three functional classes of genes were analyzed: transcription related (RdRp), structural (CP) and movement related (MP). The results indicate that there were no inherent differences, in terms of total diversity or diversity at synonymous or nonsynonymous nucleotide sites, between functional classes of genes or populations. Rather, selective constraints on a gene may be more or less relaxed depending on its function and the phylogenetic history of the population sampled. The CP of the PAS species, but not the PAV species, was genetically differentiated between regions. This is most likely due to genetic drift since there was no evidence that any gene deviated from a neutral model of evolution or is under positive selection. In general, the MP was under considerably less functional constraint than structural or replication related proteins and four positively selected codon sites were identified. Mutations at these sites differentiate species and geographical subpopulations, so presumably they have aided the virus in adaptation to its host environment and contributed to intra and interspecies diversification.
INTRODUCTION

Plant viruses may affect the fitness of their hosts by reducing host survivorship, fecundity or competitive ability relative to uninfected individuals in the population. Virus strains may differ markedly in the severity of symptoms they induce (Anderson et al., 1991, Bencharki et al., 1999) and their ability to infect a given host (Moury et al., 2001, Sacristan et al., 2005). Variation among strains in virulence, infectivity and transmission may affect patterns of disease spread and, thereby, host population dynamics in natural systems (Raybould et al., 1999) or crop yield in agricultural systems. The behavior and constraints of a virus upon host infection surely have a genetic underpinning. Thus, it is necessary to acquire knowledge of genetic diversity in pathogen populations to better understand the role they play in ecological processes and as impediments to agricultural production. This study investigates qualitative (genetic differentiation) and quantitative (genetic variation) differences in genomic content among geographically distant populations of two barley yellow dwarf virus (BYDV) species. This is a first step in understanding what evolutionary forces are acting on the viral genome.

Barley yellow dwarf disease is caused by Luteoviridae species in the genera Luteovirus and Polerovirus. Each virus species has a distinct aphid transmission phenotype and the acronym for the species is derived from this specificity. Luteoviridae species commonly isolated from grain crops include: GAV (Luteovirus) transmitted most efficiently by Schizaphis graminum and Sitobion avenae (Wang et al., 2001), MAV (Luteovirus) transmitted most efficiently by S. avenae (formerly Macrosiphum avenae), PAV (Luteovirus) transmitted most efficiently by Rhopalosiphum padi and S. avenae, SGV (unassigned to a genus within the family) transmitted most efficiently by S. graminum and RPV (Polerovirus) transmitted most efficiently by R. padi (Rochow, 1969, Rochow & Muller, 1971). BYD disease has
significant impacts in agricultural and natural plant communities. It is the most economically damaging viral disease of grain crops worldwide (Lister & Ranieri, 1995) and in grasslands it may contribute to shifts in community composition due to asymmetrical fitness effects on exotic and native grass species (Malmstrom et al., 2005a, Malmstrom et al., 2005b). Among the species listed above PAV is the most widely distributed and economically important. To reflect significant variation in coat protein (CP) sequence among isolates PAV has recently been divided into two species, PAV and PAS (Mayo, 2002). Sympatric populations of both species have been identified in Morocco (Bencharki et al., 1999), New York State (Chay et al., 1996a) and France (Mastari et al., 1998). BYDV is composed of a single-stranded, positive sense RNA with six open reading frames (ORF) in total (Miller et al., 2002). Three viral genes were analyzed in this study, ORFs 2, 3 and 4. ORF2 encodes the viral RNA-dependent RNA-polymerase (RdRp) and is responsible for the replication of all viral RNAs (Koev et al., 2002). ORF3 encodes the major component of the CP which is required for virion assembly (Mohan et al., 1995) and is thereby a prerequisite for aphid transmission (Gildow, 1987, Gildow, 1993) and systemic plant infection (Filichkin et al., 1994). ORF 4 encodes the movement protein (MP) which is required for the virus to spread systemically in the host (Chay et al., 1996b). ORF 4 is completely embedded within ORF 3 but is translated + 1 base pair out of the CP reading frame (Dinesh-Kumar & Miller, 1993).

There is currently a limited understanding of what factors influence genetic diversification and genome evolution in BYDV. This is due in part to a lack of studies which investigate variation in nonstructural genes and explore how the genetic variation found in natural virus populations effects gene function. Through an analysis of diversity and selective constraint in multiple types of genes and across virus subpopulations, the present study was able to separate out the effects of gene
function, environment and phylogenetic history in creating the pattern of diversity observed in the virus population. Thus, this study evaluates the role of selection in promoting or limiting genetic variability and provides a bridge between descriptive studies of population diversity and studies of protein structure and function (Moury, 2004).

Materials and Methods

Virus isolates

The New York State virus isolates analyzed in this study were obtained from field caught *R. padi* or from field collected corn (*Zea mays*), wheat (*Triticum aestivum*), and reed-canary grass (*Phalaris arundinacea*) during the time period 1998-2003. *Luteoviridae* nucleotide sequences from other geographic regions were obtained using the Genbank database.

Immunocapture, reverse transcription-polymerase chain reaction (RT-PCR) and DNA sequencing

Virions were isolated from plant extracts by first grinding tissue in a mixture of dry ice and phosphate-buffered saline solution followed by immunocapture in anti-BYDV-PAV antibody coated microfuge tubes. In the same tubes RT-PCR was used to amplify the CP nucleotide sequence (603 bp) or a portion of the RdRp (654 bp). RT-PCR of the CP was performed using oligonucleotide primers designed by Bencharki *et al.* (1999). The Primer Select program (DNASTAR software package) along with full-length PAV sequences of isolates PAS-129 and PAV-Aus available in Genbank were used to design primers used to amplify the RdRp. Reverse transcription was performed using oligonucleotide 5′-TTAGGGTCAACTCCGAATGATTC-3′ and PCR using the former and oligonucleotide 5′-GCGCCTAAGTGGAACACG-3′. Single-step RT-PCR was
carried out in a 50 µl reaction volume containing 25 mM Tris pH 8.8, 10 mM KCl, 10 mM dTT, 10 mM (NH₄)₂SO₄, 2.5 mM MgSO₄, 1 µM each primer, 0.4 ng ml⁻¹ BSA, 200 µM dNTPs, 30 units RNase inhibitor, 50 units SuperScript II RNase H- RT and 5 units Taq polymerase (all enzymes Invitrogen). Thermocycling conditions were: 1 cycle of 45 min at 42 °C for reverse transcription, 1 cycle of 2 min at 95° C for inactivation of reverse transcriptase, 36 cycles of 30 s at 94° C, 30 s 53° C for the coat protein or 56° C for the RdRp, 1 min at 72° C, and a final extension of 10 min at 72° C. PCR products were purified using the Qiagen PCR clean-up kit then submitted to the Cornell University Bioresource Center for direct sequencing. Sequencing was carried out in both the forward and reverse directions with the primers listed above.

**Phylogenetic and nucleotide diversity analyses**

Sequences were aligned with the ClustalW algorithm of the Megalign program (DNASTAR software package). Alignments of the RdRp, CP and MP were unambiguous but manual adjustments were made to the alignment of the complete genome sequences. In PAUP version 4.0b10 (Swofford, 2000) maximum-likelihood trees for the RdRp, CP, and MP (PAV and PAS isolates only) were constructed using the 2 ST model of nucleotide substitution with base frequencies and the shape of the gamma distribution estimated from the data. The resultant tree topologies were used for maximum-likelihood (ML) analysis of codon substitution. To explicitly examine phylogenetic relationships among BYDV species a second set of trees was constructed using the same procedure outlined above expect MAV, GAV, SGV, SbDV (*Soybean dwarf virus, Luteoviridae*) and RPV (used as outgroup) sequences were included in the analysis. The addition of these sequences did not alter the topology of the PAV/PAS region of the tree. Thus, excluding the additional sequences, the trees depicted in Fig. 1 A, B and C are identical to those used for the PAML analysis. For each gene
robustness of the nodes of the phylogenetic tree was assessed by bootstrap percentages computed after 100 resamplings.

DnaSP version 4.0 (Rozas et al., 2003) was used to estimate total nucleotide diversity per nucleotide site ($\pi$) (Nei, 1987). To compare diversity values Friedman’s nonparametric test of population central values and a series of pre-planned comparisons were implemented in SAS version 9.1. The first set of tests addressed if there is a difference in diversity between the RdRp and CP or MP genes within species or homologous genes in different species (New York state population only). A second set of tests addressed if diversity in the CP or MP differs within species in different geographic regions or between species in the same geographic region.

**Tests of selective neutrality and population differentiation**

Estimation of population parameters ($s$, $k$, $\theta$) and Tajima’s D (Tajima, 1989), Fu and Li’s F*, and Fu and Li’s D* (Fu & Li, 1993) tests of selective neutrality were performed with DnaSP (Rozas et al., 2003). Tajima’s D compares nucleotide diversity with the number of segregating sites, which are expected to be equal if mutations are selectively neutral. Fu and Li’s D* statistic is based on differences between the number of singletons (mutations appearing only once among the sequences) and the total number of mutations. Fu and Li’s F* statistic is based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences. SNAP Workbench (Price & Carbone, 2003) was used to implement the programs Seqtomatrix (Hudson et al., 1992), which facilitated the conversion of aligned CP sequences to a distance matrix, and Permtest (Hudson et al., 1992), which tested for geographic subdivision in the virus populations. Permtest calculates Hudson’s $K_{ST}$ statistic of genetic differentiation. $K_{ST}$ is equal to $1 - K_{S}/K_{T}$, where $K_{S}$ is a weighted average of $K_1$ and $K_2$ (average number of differences between sequences in subpopulations 1 and 2, respectively) and $K_{T}$
represents the average number of differences between two sequences regardless of their subpopulation. The null hypothesis of no genetic differentiation will be rejected \((P < 0.05)\) when \(K_S\) is small and \(K_{ST}\) is close to one.

**Tests of positive selection and selective constraint**

The ratio of nonsynonymous (\(dN\)) to synonymous (\(dS\)) nucleotide substitutions \((\omega = dN/dS)\) can be used to measure the degree of functional constraint for the maintenance of the encoded protein (Li, 1993). An omega ratio greater than one indicates that nonsynonymous substitutions have a higher probability of fixation than synonymous substitutions and, presumably, offer a fitness advantage to the protein (positive selection). An omega ratio close to zero indicates that nonsynonymous substitutions are less likely than synonymous substitutions and the gene will be conserved at the amino acid level (negative selection). An omega ratio equal to one indicates neutral evolution. Two approaches were taken to determine the mode and strength of selection acting on the viral genome. To test for positive selection in the RdRp, CP and MP, ML models of codon substitution were implemented in PAML version 3.14 (Yang, 1997). I then compared selective constraint on these genes among PAV and PAS isolates collected in New York and Morocco.

The ML models employed allow omega to vary among codon sites but remain constant across lineages in the phylogeny. The models implemented were: \(M_0\) which assumes one \(\omega\) for all codon sites, \(M_1\) fixes sites as either invariant \((\omega_0 = 0)\) or neutral \((\omega_1 = 1)\), \(M_2\) adds a third site class to the neutral model to allow for positive selection \((\omega_2 > 1)\), \(M_7\) assumes a beta distribution for \(\omega\) \((0 < \omega < 1)\) and \(M_8\) adds another site class to the beta model to allow for positively selected sites (Yang et al., 2000). A likelihood ratio test (LRT) was used to determine whether the positive selection model \((M_2, M_8)\) fit the data significantly better than the neutral model \((M_1, M_7)\) (Yang et al., 2000). To construct the LRT statistic twice the log-likelihood difference between
the general model and the null model was compared with a chi-square with degrees of freedom equal to the difference in the number of parameters between the two models. Bayes empirical Bayes (BEB) analysis was used to infer to what class, conserved, neutral or positively selected, a codon site belongs (Yang *et al.*, 2005). Codon sites with omega values greater than one and where posterior probabilities summed to be greater than 95% were identified as potentially under positive selection.

To assess the magnitude of the selective constraint acting on the RdRp, CP and MP genes dS and dN values were generated from pairwise sequence comparisons using the method of Nei and Gojobori (Nei & Gojobori, 1986) implemented in PAML version 3.14. To avoid dividing by zero two approaches were used: a constant was added to the dS for one synonymous substitution dN/(dS+constant) and the ratio dN/(dN+dS) was calculated (Mishmar *et al.*, 2003). Friedman’s test and a series of pre-planned comparisons were used to compare dN/(dS+constant) or dN/(dN+dS) between species and geographic regions. I first tested the null hypotheses that within a given species selective constraint does not differ between the RdRp and CP or MP genes and in different species selective constraint does not differ between homologous genes. I then tested the null hypothesis that selective constraint on the CP or MP gene does not vary across geographic regions or species.

**Results**

*Phylogenetic relationships among isolates*

Maximum-likelihood methods were used to determine the phylogenetic relationship among PAV and PAS isolates collected from a number of geographic regions. CP sequence identity between New York state isolates was similar to that described for populations sampled in other geographic regions (Bencharki *et al.*, 1999, Bisnieks *et al.*, 2004) (Table 1.1). For the RdRp there is low within species diversity,
Table 1.1. Percent nucleotide identity within the coat protein (600 nt) (above diagonal) and RNA-dependent RNA polymerase (654 nt) (below diagonal) coding regions for PAV and PAS isolates collected in New York State.

<table>
<thead>
<tr>
<th>No. of isolates</th>
<th>Viruses</th>
<th>PAV</th>
<th>PAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>PAV</td>
<td>100-97</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>PAS</td>
<td>100-99</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PAV</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>PAS</td>
<td>78-80</td>
<td>97-98</td>
</tr>
</tbody>
</table>

but between species diversity is greater than that found in the CP (Table 1.1). The RdRp (Fig. 1A) tree places PAV and MAV in a monophyletic clade that excludes PAS, but the CP (Fig. 1B) and MP (Fig. 1C) phylogenies place PAV and PAS in a monophyletic clade with MAV sister to both species. The placement of CN-PAV also differs between gene trees. It clusters on a branch sister to all PAV and PAS isolates in the MP tree but its RdRp is clearly derived from the same lineage as PAS. Analysis of the complete genomes of a number of different Luteoviridae species supports the topology of the RdRp tree, i.e. MAV and PAV form a monophyletic clade and CN-PAV and PAS form a monophyletic clade (Fig. 1D). It also shows that SGV, which is unassigned to a genus in the Luteoviridae, forms a monophyletic clade with Luteovirus isolates and is only distantly related to the Polerovirus RPV. Disagreement between RdRp and CP phylogenies could be the result of selection and/or genetic drift that has lead to divergence of the CN-PAV and MAV CP sequences. It is also possible that ancestral Luteoviruses have experienced recombination in the CP. The position of isolate ASL-1 in the CP and RdRp trees also warrants further investigation. ASL-1 shares greater nucleotide sequence identity with PAS isolates in its CP but greater identity with PAV isolates in its RdRp sequence. ASL-1 may be a product of recombination between PAV and PAS or, alternatively, CP similarity could be the result of convergent evolution.
Table 1.2. Nucleotide diversity per site ($\pi$) in the RNA-dependent RNA polymerase (RdRp), coat protein (CP) and movement protein (MP) of BYDV-PAV and PAS isolates collected in New York and Morocco. Numbers in parentheses are standard deviations of estimates.

<table>
<thead>
<tr>
<th>Region</th>
<th>Species</th>
<th>RdRp</th>
<th>CP</th>
<th>MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>PAV</td>
<td>0.017 ($0.003)^{ab}$</td>
<td>0.025 ($0.007)^{ac}$</td>
<td>0.024 ($0.006)^{bd}$</td>
</tr>
<tr>
<td></td>
<td>PAS</td>
<td>0.02 ($0.002)^{ab}$</td>
<td>0.01 ($0.001)^{Ac}$</td>
<td>0.009 ($0.002)^{Bd}$</td>
</tr>
<tr>
<td>Morocco</td>
<td>PAV</td>
<td>*</td>
<td>0.028 ($0.006)^{C}$</td>
<td>0.026 ($0.006)^{D}$</td>
</tr>
<tr>
<td></td>
<td>PAS</td>
<td>*</td>
<td>0.006 ($0.002)^{C}$</td>
<td>0.003 ($0.001)^{D}$</td>
</tr>
</tbody>
</table>

DnaSP was used to estimate $\pi$ as the average of pairwise comparisons among sequences in a population (Nei 1987). Freidman’s nonparametric test of population central values was used to compare $\pi$ values. Values with the same letter were included in the same analysis and capitalization denotes values found to be significantly different ($P < 0.05$). * No sequences available for analysis.

**Qualitative and quantitative differences in genomic content among species and geographical subpopulations**

To better understand patterns of genetic diversity in the virus population, Hudson’s $K_{ST}$ test of population subdivision (Hudson et al., 1992) was performed. Hudson’s test showed evidence of genetic differentiation in the CP of the New York and Moroccan PAS populations ($P < 0.01$, $K_S = 4.1$, $K_T = 16.0$, $K_{ST} = 0.74$). It did not, however, detect differentiation in the PAV population ($P = 0.16$, $K_S = 15.5$, $K_T = 16.3$, $K_{ST} = 0.051$) despite clear segregation of Moroccan and New York isolates within cluster groups (Fig. 1.1). This may be due to the clustering of the PAV CP into two groups (previously described in (Bisnieks et al., 2004)), each containing New York and Moroccan isolates, which could act to inflate $K_S$ and in turn lower the $K_{ST}$ value. I then compared total nucleotide diversity across genes and geographic regions between and within virus species (Table 1.2). It was found that for the RdRp and CP comparison there is a significant interaction between gene and species (df = 1, $F = 8.92$, $P < 0.01$), thereby indicating that the probability that two randomly chosen isolates differ at a given nucleotide site is dependent upon the identity and phylogenetic history of the gene. This is reflected by the pre-planned comparisons.
found to yield significant $P$ values. The CP of PAS is less diverse than its RdRp ($P < 0.01$) and less diverse than the CP of PAV ($P < 0.01$). Similar results were obtained in comparison of the RdRp to the MP, in that, there was no singular effect of gene or species and the MP of PAS was less diverse than the PAS RdRp ($P < 0.01$) and less diverse than the MP of PAV ($P < 0.01$). I next asked if diversity in the CP or MP differs between species in the same geographic region and within species between geographic regions? For the CP gene the effect of geographic region on nucleotide diversity was dependent upon the phylogenetic history of the isolate (df = 1, $F = 8.28$, $P < 0.01$). It was also found that the PAS population in Morocco is significantly less diverse than the PAS population in New York ($P < 0.01$) and the PAV population in Morocco ($P < 0.01$). Diversity did not differ between the PAV population in New York state and Morocco. An identical pattern was found in analysis of the MP, where Moroccan PAS isolates had significantly lower nucleotide diversity than all other populations ($P < 0.01$) but diversity did not differ between species in New York or between geographic regions for PAV.

Tests of positive selection and selective neutrality and comparisons of selective constraint

To determine if natural selection played a role in generating the genetic diversity observed in the virus population I used ML models of codon substitution to test for positive selection. Models that allow for positively selected codon sites, M2 and M8, did not fit the data significantly better than models which assume that all sites are either conserved or neutral, M1 or M7 (Table 1.3). The BEB analysis did, however, identify four codon sites in the MP that are potentially undergoing positive selection (posterior probability of site belonging to class $\omega_2 > 95\%$ with M8 and $> 90\%$ with M2). No positively selected sites were detected in the RdRp or the CP.
Figure 1.1. Maximum likelihood trees of *Luteoviridae* species isolates calculated using (A) partial RNA-dependent RNA polymerase nucleotide sequences (ML score = 3126), (B) complete coat protein nucleotide sequences (ML score = 3101), (C) complete movement protein nucleotide sequences (ML score = 2099) and (D) complete genome sequences (except SGV, open reading frames 2, 3 and 4 only) (ML score = 44053). PAV isolates in italics and PAS isolates in bold. All nodes with less than 70% bootstrap support out of 100 replicates were collapsed to polytomies.
Mutations at codon site 9 (PAV-Aus used as reference) were region specific but not species specific. All virus isolates collected in Morocco have lysine at this site; some New York PAV isolates, some New York PAS isolates and CN-PAV have glutamic acid; and some New York PAS isolates have aspartic acid. The alanine mutation at codon 68 was unique to New York PAS isolates. All other PAS and PAV isolates have valine at this site. At two sites MP substitutions were differentiated between species rather than between geographic regions. At codon site 133 all PAS isolates have asparagine, PAV isolates in one cluster group have glycine and PAV isolates in the other cluster group have arginine. At codon site 145 all PAV isolates have threonine, all PAS isolates have methionine and CN-PAV has glutamic acid. These selection events may have occurred before populations were separated on different continents. None of the test statistics generated from Tajima’s D and Fu and Li’s F* and D* were significant \( (P > 0.10; \text{Table 1.4}) \). This indicates that the hypothesis of selective neutrality for the CP, RdRp and MP in the New York and Moroccan PAV and PAS populations cannot be rejected.

To examine selective constraints experienced by the RdRp, CP and MP I compared \( dS/(dS + \text{constant}) \) and \( dS/(dN + dS) \) for each gene across geographic regions and lineages. Both methods underestimate \( \omega \) but the \( dN/(dN+dS) \) has the further shortcoming that it does not allow values greater than one. Similar distributions and \( P \) values were obtained using either method, thus only values calculated by \( dS/(dS + \text{constant}) \) are presented in Table 1.5. To evaluate if selective constraint is correlated with gene function we compared the RdRp to the CP or MP. Results show that selective constraint is dependent upon the phylogenetic history of a gene and its function \( (df = 1, F = 12.7, P < 0.01) \). This is due to significantly lower selective constraint on the PAS CP than the PAS RdRp \( (P < 0.01) \) or the PAV CP \( (P \)
Table 1.3. Maximum-likelihood analysis of codon substitution (Yang, 1997) for detection of positive selection in the RNA-dependent RNA polymerase (RdRp), coat protein (CP) and movement protein (MP) of PAV and PAS isolates collected in New York and Morocco.

<table>
<thead>
<tr>
<th>Gene</th>
<th>mean $\omega$</th>
<th>$2\Delta L$ M2 vs M1 $P$ value</th>
<th>$2\Delta L$ M8 vs M7 $P$ value</th>
<th>positively selected sites</th>
<th>A A changes</th>
</tr>
</thead>
<tbody>
<tr>
<td>RdRp</td>
<td>0.05</td>
<td>17.4</td>
<td>0.74</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>CP</td>
<td>0.37</td>
<td>1.2</td>
<td>0.55</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>MP</td>
<td>0.67</td>
<td>4.2</td>
<td>0.12</td>
<td>4</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Mean $\omega$ ($d_N/d_S$) for best fit model among M0, M1 and M2, model M0 assumes one ratio for all codon sites, M1 and M7 assume $\omega = 0$ or $\omega = 1$ for all sites, M2 and M8 allow $\omega > 1$ (Yang et al. 2000). †Twice the difference in -log likelihood between the general and reduced models. ‡Probability that $2\Delta L$ is smaller than a chi-squared with 2 degrees of freedom. §Codon belongs to site class $\omega > 1$ with 90% confidence for M2 and 95% confidence for M8, number indicates first nucleotide position of codon site with PAV-Aus used as reference sequence.
Table 1.4. Population statistics and neutrality tests based on variation in the RNA-dependent RNA polymerase (RdRp), coat protein (CP) and movement protein (MP) of PAV and PAS isolates collected in New York and Morocco.

<table>
<thead>
<tr>
<th>Geographic Region</th>
<th>gene</th>
<th>species</th>
<th>l</th>
<th>n</th>
<th>s</th>
<th>k</th>
<th>θ</th>
<th>Tajima's D statistic</th>
<th>Fu and Li's D* statistic</th>
<th>Fu and Li's F* statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>RdRp</td>
<td>PAV</td>
<td>654</td>
<td>5</td>
<td>29</td>
<td>12</td>
<td>0.022</td>
<td>-1.03</td>
<td>-1.03</td>
<td>-1.11</td>
</tr>
<tr>
<td></td>
<td>RdRp</td>
<td>PAS</td>
<td>654</td>
<td>7</td>
<td>33</td>
<td>12.95</td>
<td>0.021</td>
<td>-0.381</td>
<td>-0.288</td>
<td>-0.342</td>
</tr>
<tr>
<td>New York</td>
<td>CP</td>
<td>PAV</td>
<td>603</td>
<td>5</td>
<td>25</td>
<td>14.2</td>
<td>0.021</td>
<td>1.36</td>
<td>1.36</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>PAS</td>
<td>603</td>
<td>7</td>
<td>14</td>
<td>6.19</td>
<td>0.0096</td>
<td>0.459</td>
<td>0.636</td>
<td>0.656</td>
</tr>
<tr>
<td>Morocco</td>
<td>CP</td>
<td>PAV</td>
<td>603</td>
<td>5</td>
<td>32</td>
<td>16.9</td>
<td>0.0257</td>
<td>0.751</td>
<td>0.799</td>
<td>0.852</td>
</tr>
<tr>
<td></td>
<td>CP</td>
<td>PAS</td>
<td>603</td>
<td>7</td>
<td>10</td>
<td>3.33</td>
<td>0.007</td>
<td>-0.984</td>
<td>-1.05</td>
<td>-1.14</td>
</tr>
<tr>
<td>New York</td>
<td>MP</td>
<td>PAV</td>
<td>462</td>
<td>5</td>
<td>19</td>
<td>11.2</td>
<td>0.02</td>
<td>1.23</td>
<td>1.39</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>MP</td>
<td>PAS</td>
<td>462</td>
<td>7</td>
<td>9</td>
<td>4.38</td>
<td>0.008</td>
<td>1.02</td>
<td>1.18</td>
<td>1.25</td>
</tr>
<tr>
<td>Morocco</td>
<td>MP</td>
<td>PAV</td>
<td>462</td>
<td>5</td>
<td>22</td>
<td>12</td>
<td>0.024</td>
<td>1.011</td>
<td>1.011</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>MP</td>
<td>PAS</td>
<td>462</td>
<td>7</td>
<td>4</td>
<td>1.14</td>
<td>0.0036</td>
<td>-1.434</td>
<td>-1.51</td>
<td>-1.62</td>
</tr>
</tbody>
</table>

* l, sequence length; n, sample size; s, number of segregating sites; k, average number of nucleotide differences in pairwise sequence comparisons (Tajima 1983); θ, population mean mutation rate per site (Watterson's θ estimator) is a function of the number of segregating sites and the sample size. † Tajima's D, Fu and Li's D* and F* tests measure the departure from neutrality for all mutations in a genomic region (Tajima 1989, Fu and Li 1993). Values for neutrality tests were all not significant (P > 0.1).
Thus, neither the RdRp nor the CP is necessarily more functionally constrained than the other. For the RdRp/MP comparison there was a significant effect of function (df = 1, F = 2.43, P < 0.01) due a relaxation of selective constraints for the MP of PAV (P < 0.01) and PAS (P < 0.01). I next evaluated the effect of geographic region on the selective constraint experienced by the CP and MP genes. For the CP selective constraints were more relaxed on the PAS population in New York and Morocco, thus phylogenetic history (df = 1, F = 23.18, P < 0.01) but not geographic location (df = 1, F = 0.09, P = 0.78) has a significant impact on selective constraint. For the MP the effect of geographic location on selective constraint is dependent upon the phylogenetic history of the isolate (df = 1, F = 11.08, P < 0.01). This is reflected by the pre-planned comparisons, where the PAS population in Morocco was found to be more constrained than the PAS population in New York and (P < 0.01) and PAV population Morocco (P < 0.01).

Table 1.5. Relative selective constraints [dN/(dS+constant)] calculated for the RNA-dependent RNA polymerase (RdRp), coat protein (CP) and movement protein (MP) of PAV and PAS isolates collected in New York and Morocco. Numbers in parentheses are the standard errors of estimates.

<table>
<thead>
<tr>
<th>Region</th>
<th>Species</th>
<th>RdRp</th>
<th>CP</th>
<th>MP</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>PAV</td>
<td>0.052 (0.017) &lt;sup&gt;ab&lt;/sup&gt;</td>
<td>0.055 (0.025) &lt;sup&gt;ac&lt;/sup&gt;</td>
<td>1.0 (0.32) &lt;sup&gt;Bd&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>PAS</td>
<td>0.035 (0.026) &lt;sup&gt;ab&lt;/sup&gt;</td>
<td>0.44 (0.07) &lt;sup&gt;AC&lt;/sup&gt;</td>
<td>0.71 (0.14) &lt;sup&gt;Bd&lt;/sup&gt;</td>
</tr>
<tr>
<td>Morocco</td>
<td>PAV</td>
<td>*</td>
<td>0.11 (0.03) &lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.05 (0.29) &lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>PAS</td>
<td>*</td>
<td>0.30 (0.05) &lt;sup&gt;c&lt;/sup&gt;</td>
<td>(0.008) &lt;sup&gt;D&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

PAML (Yang 1997) was used to estimate dN and dS values from pairwise comparisons of sequences in a population using the method of Nei and Gojobori (1986). Freidman’s nonparametric test of population central values was used to compare relative selective constraint. Values with the same letter were included in the same analysis and capitalization denotes values found to be significantly different (P < 0.05). *No sequences available for analysis.
DISCUSSION

Phylogenetic relationships and genetic structure

Previous studies of genetic variation in BYDV populations have focused exclusively on genes in the 3’ half of the viral genome (Bencharki et al., 1999, Chay et al., 1996a, Mastari et al., 1998). The findings of the present study demonstrate that in order to understand the phylogenetic history of luteoviruses it is necessary to analyze genes in the 5’ and 3’ halves of the genome. For instance, analysis of the RdRp, which is near the 5’ terminus, has unmasked potential recombination between the CP of PAV isolate ASL-1 and PAS. Recent findings that Luteoviridae species Sugarcane yellow leaf virus has undergone recombination with PAV in this same genomic region (Smith et al., 2000) lends some support to this hypothesis, but it is also possible that there has been convergence of the ASL-1 CP to a PAS like sequence. This study has also found inconsistent evolutionary relationships among MAV, PAV and PAS isolates. In the RdRp phylogeny MAV and PAV form a monophyletic clade independent of CN-PAV and PAS which also form a monophyletic clade. In the MP phylogeny, however, PAV and PAS are sister taxa, CN-PAV is sister to both species, and MAV is sister to all three virus types. Analysis of the complete genome sequences from a number of Luteoviridae species supported the topology of the RdRp tree. It could be that the position of MAV in the CP is the result of the loss or change in genetic motifs that govern transmission by R. padi. CP sequences that regulate aphid species specific virus transmission have not yet been uncovered but studies have demonstrated that vector selection does exert evolutionary pressure on luteovirus populations (Power & Gray, 1995, Zavaleta et al., 2001).

Geographic subdivision has been reported for the global populations of Tomato spotted wilt virus (TSWV) (Tsompana et al., 2005), Cucurbit yellow stunting disorder virus (Rubio et al., 2001) and Turnip mosaic virus (Tomimura et al., 2003), as well as,
for regional populations of *Rice yellow mottle virus* (RYMV) (Pinel et al., 2000), *Sweet potato chlorotic stunt virus* (Alicai et al., 1999) and *Kennedyda yellow mosaic virus* (Skotnicki et al., 1996). Several of these studies invoke virus adaptation in response to new hosts or transmission modalities to explain the observed pattern of variation but few present evidence that geographical variants differ in their biology or that there is a correlation between positively selected codon sites and the geographic region from where the population was sampled. In the absence of such evidence one could also attribute geographic subpopulation structure to transmission bottlenecks followed by genetic drift. As selection and genetic drift can both lead to a decrease in diversity within populations and an increase in diversity between populations, it is often times difficult to distinguish the effects of these processes. But selection and drift need not be mutually exclusive forces. A recent study by Choi and colleagues (Choi et al., 2001) found considerable divergence between the American and Mexican populations of *Wheat streak mosaic virus* (WSMV), mostly at synonymous nucleotide sites. They conclude from these data that negative selection and drift acting simultaneously have contributed to the evolution of WSMV strains. The present study found significant genetic differentiation between the CP of the Moroccan and New York PAS populations but no support for the hypothesis that either population deviates from a neutral model of evolution. Thus, it appears that for PAS, like WSMV, genetic drift rather than adaptation is responsible for divergence between the two populations.

**Genetic diversity and selective constraint**

Total nucleotide diversity was estimated to be 0.028 and 0.029 in the CP and 0.017 and 0.020 in the RdRp for the New York PAV and PAS populations respectively. These values are similar to or lower than total nucleotide diversity values reported for other insect-transmitted plant RNA viruses, such as 0.035 for TSWV (Tsirpana et al., 2005), 0.068 for *Citrus tristeza virus*, 0.07 for *Groundnut rosette virus* and 0.194 for
RYMV (Garcia-Arenal et al., 2001). Diversity did not significantly differ across genes, species, or geographic regions. The only exception was the Moroccan PAS population which has much lower nucleotide diversity in its CP (0.006) when compared to the Moroccan PAV population (.028) or the New York PAS population (.025). This study did not find a significant difference in selective constraint on the New York or Moroccan PAS CP. Thus, it is likely that the Moroccan population has been through a genetic bottleneck more recently than the New York population, such that there has been less time for population expansion and re-establishment of diversity. In New York selective constraint did not differ between the PAV CP (0.055), the PAV RdRp (0.052) or the PAS RdRp (0.035) but was significantly greater for the PAS CP (0.44). This indicates that genetic variation in either gene class is not necessarily more constrained due to the function of the encoded protein. I also found that selective constraint does not differ between the PAS CP in New York and Morocco and it is less constrained than the PAV CP in both environmental contexts. Because PAV and PAS are sympatric in these locations and share the same vector specificity it might be that relaxed selective constraints on the PAS CP are due to differences in plasticity of the protein rather than differences in the source or magnitude of the selection pressure acting on the virus population.

Utilizing statistical models that allow heterogeneous $\omega$ ratios among codon sites researchers have found evidence of diversifying selection in genes of CMV (Moury, 2004), Potato virus Y (Moury et al., 2002) and TSWV (Tsompana et al., 2005). Positively selected amino acid sites were detected in structural proteins of each of these viruses and in transcription related proteins of CMV and TSWV. For BYDV-PAV and PAS the MP was under the least selective constraint when compared to genes in other functional classes, except in the Moroccan PAS population where it seemed to be more constrained than any gene sampled from any other population. But as discussed above
this could be the result of a recent population bottleneck. Four codon sites in the MP are putatively affected by positive selection. Three of these sites are located in the C-proximal part of the protein which is known to be the nucleic acid binding domain of the Potato leaf roll virus (PLRV) (Luteoviridae, Polerovirus) MP (Tacke et al., 1991). The other site is located in the N-terminal part of the protein. Experimental manipulation of the PLRV nucleotide sequence suggest that protein/protein interactions occur in this domain (Tacke et al., 1993). In interpreting the results of the PAML analysis it is necessary to keep in mind that synonymous sites in the MP correspond to nonsynonymous sites in the CP. This may influence estimates of $\omega$ for the MP by reducing the rate of synonymous substitutions. Relatively high $\omega$ estimates have been shown for overlapping reading frames in CMV (Moury et al., 2002) and PLRV (Guyader & Ducray, 2002).

In summary, the present study found that PAV and PAS are more distantly related with respect to their RdRp than CP genes. It is not well understood how genetic variation in the BYDV RdRp relates to virus accumulation in host tissues, host range and disease severity. Future research must address this issue as the RdRp is a significant source of genetic variation at the intra and interspecies level. The results of this study indicate that genetic variation in the RdRp is not correlated with the aphid transmission phenotype of the virus because analysis of the RdRp gene alone cannot differentiate between BYDV species. This reflects the different selection pressures acting on the RdRp and CP. Despite the different roles they play in the virus infection cycle, selective constraint experienced by the RdRp or CP did not differ based solely upon their function. Rather, functional constraint was determined by gene function, phylogenetic history and selection pressures in a particular environment. Given our findings that New York PAS populations are genetically distinct from the Moroccan populations, it should not be assumed that isolates of the same species in different
geographic regions will have similar biological characteristics. It may be necessary to independently evaluate the infection properties of isolates present in each area where BYDV hinders grain production. This information will aid growers in the development of grain varieties with the appropriate resistance, and will link features in the viral genome to virus transmission dynamics and effects on host fitness.
References


CHAPTER TWO

Quantitative sequencing: A new method for analyzing the within-host dynamics of closely related virus genotypes

Abstract

Empirical studies of within-host competitive interactions between virus genotypes are quite rare because there has not been a good method for quantifying the population size of closely related genotypes in a single host. This study utilized a new method called quantitative sequencing to measure the template concentration of two *Barley yellow dwarf virus* (BYDV; *Luteoviridae*) species in doubly infected wheat plants. The competition assays we performed varied the order of inoculation and the time interval between the first and second inoculation. Reverse transcription-PCR and direct sequencing was performed on nucleic acid extracts from experimental plants. We developed a PERL script (polySNP) that uses PHRED to automatically extract relative peak areas and heights from sequencing chromatograms at polymorphic sites. Peak measurements from experimental samples were compared to a standard curve generated by mixing RNA transcribed *in vitro* from BYDV-PAV and PAS templates in several ratios (ranging from 1:9 to 9:1 PAV:PAS) prior to RT-PCR amplification and sequencing. The frequency of RNA template added to a sample was regressed onto the proportion of the chromatogram peak height or area corresponding to one virus species. The function of the best fit line was used to calculate template frequency in the experimental samples. In singly infected plants, virus species had equal population sizes. In mixed infections there was reciprocal cross-protection between virus species, but regardless of the order of inoculation or the length of the inoculation delay PAV eventually dominated the virus population.
Introduction

The virus population in a host community (Kong et al., 2000, Kurath et al., 1993, Tsompana et al., 2005) and within an individual host (Schneider & Roossinck, 2000) can exhibit a high degree of genetic diversity. Consequently, for a given genotype other genetic variants will be part of the selective environment within a host. Theoretical models have demonstrated that competitive interactions between pathogen genotypes have important consequences for within-host pathogen population dynamics, pathogen transmission between hosts, and the evolution of pathogen virulence (Levin & Pimentel, 1981, Nowak & May, 1994). Despite their evolutionary importance, empirical studies of virus genetic variants in mixed infections are still quite rare due, in part, to the lack of an effective method for quantifying the population size of closely related virus genotypes in a single host. This study utilizes a new method called quantitative sequencing to measure the population size of two species of Barley yellow dwarf virus (BYDV; Luteoviridae) in mixed infections. BYDV has a positive sense ssRNA genome of approximately 5.6 kb (Miller et al., 2002). It infects many wild and cultivated grass species and is an economically important pathogen of grain crops worldwide (D’arcy, 1995). The PAV and PAS species analyzed in this study are ecologically similar, sharing the same aphid vectors (Rhopalosiphum padi and Sitobion avenae) (Chay et al., 1996) and host species (many members of the family Poacea). Also, PAV and PAS have approximately 90% and 78% nucleotide sequence identity in their capsid protein (Bisnieks et al., 2004) and polymerase genes (Hall, 2006), respectively. The effect of one virus species on the population growth of the other during mixed infections has not yet been reported, but plants infected with both species have been identified in natural plant populations in France (Mastari et al., 1998) and New York State (Chapter 3).
For genetic variants that differ at few nucleotide sites it may be difficult or impossible to develop genotype specific primers, probes or antibodies capable of differentiating between them in a mixed infection. RNase protection and restriction fragment analyses have been employed to discriminate between virus types that have small sequence differences (Kurath et al., 1993, Tenllado et al., 1997), but these techniques are generally not implemented in a quantitative manner. As such, they can only determine if the end result of competition is coexistence or complete exclusion. Measurement of the relative area or height of sequencing chromatogram peaks at polymorphic sites can result in quantitative data if a standard curve is constructed by mixing known amounts of viral template and performing PCR amplification followed by direct sequencing (Wilkening et al., 2005). Direct nucleotide sequencing has two major advantages over the previously listed methods. First, direct sequencing can differentiate templates that differ by as little as one nucleotide. Secondly, direct sequencing is less costly and requires less time to develop and optimize than antibody or other quantitative PCR based techniques.

Quantitative sequencing has previously been used to estimate allele frequency and mutation frequency in pooled DNA samples (Amos et al., 2000, Wilkening et al., 2005). The data are quantitative because the height or area of the peaks in a sequencing chromatogram at polymorphic nucleotide positions is proportional to the corresponding template concentration at the start of the sequencing reaction. Quantitative sequencing has never before been applied to RNA templates nor has it been used to measure the relative concentration of virus template in a host as is done in the present study. To facilitate the analysis of experimental samples we have developed a PERL script called polySNP that uses PHRED to automatically extract peak area and height data from multiple single nucleotide polymorphisms (SNPs) in
sequence chromatograms. The objective of the present study is to evaluate the accuracy of quantitative sequencing when applied to mixtures of RNA templates. As an example, quantitative sequencing is used to determine if the presence of one BYDV species in the host inhibits the establishment of another species and if varying the time interval between inoculations affects the stability of the competitive hierarchy.

**Materials and methods**

*Inoculation of experimental plants*

Wheat (*Triticum aestium*) plants were grown from seed in six inch pots and randomly assigned to one of seven inoculation treatments: single inoculation with PAV, single inoculation with PAS, simultaneous inoculation with PAV and PAS, PAV challenged with PAS after a three day delay, PAV challenged with PAS after a fifteen day delay, PAS challenged with PAV after a three day delay, and PAS challenged with PAV after a fifteen day delay. To inoculate plants *R. padi* from laboratory maintained disease-free colonies were fed for 48 hours on detached leaves infected with the appropriate virus isolate. Aphids were then transferred to healthy plants and allowed an inoculation access period of five days. Inoculation with the protecting strain was carried out when plants were at the two-leaf stage. PAV and PAS isolates used in this study were collected from agricultural fields in central New York State. Isolate PAS-129 was obtained in 1992 from migrating alate *R. padi* alighting on winter wheat (Chay *et al.*, 1996). PAV isolate Fa2k298 was obtained in 1998 from apterous *R. padi* collected in oat fields. Virus isolates were maintained in oat (*Avena sativa*) plants in the greenhouse since the initial isolation.

*Nucleic acid extraction, reverse transcription-PCR and sequencing*
At 33 days post-inoculation (DPI) with the protecting strain leaves were harvested from at least three plants per inoculation treatment. Equal amounts of leaf tissue from plants singly inoculated with PAV or PAS (100 mg each) were combined into one sample prior to the nucleic acid extraction. For all other treatments the total nucleic acid extraction was performed on 200 mg of leaf tissue from a single plant. Leaf samples were homogenized with a glass rod in 1.5 mL microfuge tubes in 900 µL extraction buffer (7 M urea, 0.3 M NaCl, 50 mM tris pH 8.0, 20 mM EDTA and 1% SDS). The homogenate was incubated at 65° C for 10 minutes then extracted with 650 µl phenol:chloroform:isoamyl alcohol (25:24:1). Nucleic acids were precipitated by adding 700 µL of supernatant to a 700 µL solution containing isopropanol and ammonium acetate such that the final concentration of the solution was 0.7 M pH 5.2. Single-step RT-PCR was carried out in a 50 µL reaction volume containing 25 mM tris pH 8.8, 10 mM KCl, 10 mM dTT, 10 mM (NH₄)₂SO₄, 2.5 mM MgSO₄, 1 µM each forward (5′-AGAGGCCACAGAATGTCCGG-3′) and reverse (5′-GTTCAGCTTCAACACCCAGC-3′) oligonucleotide primers, 200 µM dNTPs, 30 units RNase inhibitor, 50 units SuperScript II RNase H-reverse transcriptase and 5 units Taq polymerase (all enzymes Invitrogen, Carlsbad, Calif.). This primer combination amplifies a 106-nucleotide region of the viral RNA-dependent RNA-polymerase. Both virus species have exactly the same sequence in the primer binding region. Genbank accession numbers for Fa2k298 and PAS-129 sequences containing this gene region are DQ286379 and DQ286383, respectively. Thermocycling conditions were: reverse transcription 1 cycle of 40 min at 42° C, inactivation of reverse transcriptase and pre-denaturation 1 cycle of 2 min at 95° C, DNA amplification 35 cycles of 30 s at 94° C, 30 s at 60° C, 1 min at 72° C, and a final extension of 10 min at 72° C. RT-PCR products were purified using the Invitrogen
PCR clean-up kit following the protocol prescribed by the manufacturer and then submitted to the Cornell University Bioresource Center for direct sequencing using an ABI 3730 sequencer with Big Dye Terminator chemistry and AmpliTaQ-FS DNA polymerase. Sequencing was carried out with the forward amplification primer.

In vitro transcription of PAV and PAS polymerase gene template

PAS and PAV isolates were RT-PCR amplified with primers and under conditions described above. The reverse primer was modified by the addition of a T7 site at the 5' end, (5'-TAATACGACTCACTATAGGGTTCAGC-3'). Unincorporated primers and dNTPs were separated from the amplicon using the QIAquick PCR Purification kit (Qiagen Inc., Valencia, Calif.) following the manufacturer's instructions. Approximately 250 ng of purified DNA amplicon was used as a template for RNA synthesis using the MAXIscript in vitro RNA transcription kit (Ambion Inc., Austin, Tex.) and the supplied T7 polymerase. After 1 hour of transcription at 37°C, the DNA template was destroyed by the addition of DNase I (Roche Applied Science, Mannheim, Germany) followed by a 15 minute incubation at 37°C. The DNase was deactivated by a 10 minute incubation at 75°C (with 2.2 mM EDTA to protect the RNA). The transcribed RNA was further purified by ethanol precipitation in the presence of 45 mM ammonium acetate. The Nanodrop (Wilmington, Del.) spectrophotometer was used to quantify the yield of RNA.

Construction of standard curves and analysis of experimental plants

In order to determine the relative concentration of PAV and PAS in experimental plants a standard curve was developed using in vitro transcribed RNA. PAV and PAS templates were mixed in eight ratios (from 1:9 to 9:1 PAV:PAS). Mixing was replicated three times for each ratio, after which samples were RT-PCR amplified (according to the protocol above) and submitted for direct sequencing.
Within the 106-nucleotide region sequenced nine SNPs differentiate virus species, five sites that were consistently resolved in the chromatograms were chosen for further analysis. The area and height of the peaks for each of these SNPs was extracted from the sequence chromatograms by the PERL script polySNP (available at http://staging.nybg.org/polySNP.html). To extract data from the chromatograms, polySNP first produced .seq, .phd, and .poly files by running PHRED version 0.020425.c (Ewing et al., 1998) on each chromatogram file with a “-trim cutoff” value of 0.10. The pre-aligned reference sequences (DQ286379 and DQ286383) were aligned to the .seq file (a FASTA format file containing the called bases) using the “-profile” option of MUSCLE version 3.6. (Edgar, 2004). The alignment was used to locate dimorphic SNP positions in the .phd file (a text file containing the called bases, a quality assessment for each base, and location of the peak in the chromatogram) which was in turn used to find the corresponding line in the .poly file (a text file containing peak areas and heights for each chromatogram position). After confirming that the peaks correspond to the known variation in the SNP, the peak areas and heights were extracted and the relative peak areas or heights were calculated for each virus species. The relative proportions (PAV / PAV + PAS) were used instead of the ratio of the relative peak heights (PAV/PAS), as has been done in previous studies (Wilkening et al., 2005), because we were interested in relative viral transcript concentration rather than the relative magnitude of separation between concentrations.

For each SNP the amount of virus template added to the RT-PCR reaction was regressed onto the height or area of the PAV peak as a proportion of the total peak height or area. This procedure minimized the error in our estimate of virus concentration (Hilsel and Hirsch, 2005). Given the shape of the plotted data, Minitab14 (Minitab Inc., State College, Pa.) was used to fit a linear, exponential,
quadratic or cubic model. Following the same procedures described above, standard curves were also generated from mixtures of RT-PCR products amplified from plants singly infected with PAV and PAS.

To test their accuracy, standard curves were used to estimate relative virus concentration in a group of ten samples prepared by mixing *in vitro* transcribed PAS and PAV RNA. The proportion of PAV added to the sample was read from the first column of a random number table. Chi-square goodness-of-fit tests were used to test the null hypothesis: the quantity of PAV added to the reaction did not significantly differ from the quantity calculated from the standard curve for a given polymorphic site across the group of random samples. Paired t-tests were used to test the null hypothesis that for a given SNP there was no significant difference in virus concentration as calculated from peak area or height standard curves. A single factor analysis of variance implemented was used to compare relative PAV or PAS concentration between inoculation treatments. Each of the five SNPs were treated as independent measures of virus concentration. Individual comparisons between treatments were carried out with Tukey’s procedure ($\alpha = 0.05$). All statistical analyses performed with Minitab 14.

**Results**

*Standard curves*

In total, ten standard curve plots were generated and depending on the position of the SNP and peak measurement (height or area) linear, exponential, quadratic or cubic models were found to provide the best fit regression line (Table 2.1). Relative virus concentration calculated from peak area and height using RNA and DNA
standard curves did not significantly differ from the expected concentrations of PAV and PAS in a group of random template mixtures (chi-square, $\alpha = 0.05$; Table 2.2).

Table 2.1. Best fit line when the frequency of PAV template in standard sample was regressed onto PAV as a proportion of the total peak area or height (PAV / PAV + PAS) at five nucleotide sites that differentiate PAV and PAS.

<table>
<thead>
<tr>
<th>Site*</th>
<th>template†</th>
<th>peak measurement</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>2074</td>
<td>RNA</td>
<td>area</td>
<td>$- 0.097 + 15.95 \text{ area} - 6.73 \text{ area}^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$0.13 + 25.53 \text{ height} - 31.9 \text{ height}^2 + 15.5 \text{ height}^3$</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>area</td>
<td>$- 0.006 + 15.24 \text{ area} - 5.75 \text{ area}^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.27 + 23.6 \text{ height} - 29.3 \text{ height}^2 + 16.22 \text{ height}^3$</td>
</tr>
<tr>
<td>2080</td>
<td>RNA</td>
<td>area</td>
<td>$- 0.33 + 9.63 \text{ area}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.27 + 23.6 \text{ height} - 29.3 \text{ height}^2 + 16.22 \text{ height}^3$</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>area</td>
<td>$- 0.82 + 10.83 \text{ area}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.19 + 9.049 \text{ area}$</td>
</tr>
<tr>
<td>2083</td>
<td>RNA</td>
<td>area</td>
<td>$- 0.33 + 9.63 \text{ area}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.33 + 9.63 \text{ area}$</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>area</td>
<td>$- 0.33 + 9.63 \text{ area}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.33 + 9.63 \text{ area}$</td>
</tr>
<tr>
<td>2086</td>
<td>RNA</td>
<td>area</td>
<td>$- 0.39 + 33.93 \text{ area} - 47.38 \text{ area}^2 + 22.88 \text{ area}^3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$3.96 - 17.24 \text{ height} + 22.06 \text{ height}^2$</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>area</td>
<td>$1.6 + 8.05 \text{ area}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.82 + 10.83 \text{ area}$</td>
</tr>
<tr>
<td>2088</td>
<td>RNA</td>
<td>area</td>
<td>$5.92 - 20.92 \text{ area} + 23.96 \text{ area}^2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$9.34 \text{ height}^{2.136}$</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>area</td>
<td>$- 1.45 + 19.45 \text{ height} - 46.86 \text{ area}^2 + 37.7 \text{ area}^3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$3.96 - 17.24 \text{ height} + 22.06 \text{ height}^2$</td>
</tr>
<tr>
<td></td>
<td>RNA</td>
<td>area</td>
<td>$9.34 \text{ height}^{2.136}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 1.45 + 19.45 \text{ height} - 46.86 \text{ area}^2 + 37.7 \text{ area}^3$</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>area</td>
<td>$0.91 + 33.97 \text{ area} - 57.21 \text{ area}^2 + 31.60 \text{ area}^3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>height</td>
<td>$- 0.12 + 41.8 \text{ height} - 77.32 \text{ height}^2 + 47 \text{ height}^3$</td>
</tr>
</tbody>
</table>

*, complete genome of PAV-Aus (M21347) used as the reference sequence; †, standard curves were created by mixing in vitro transcribed viral RNA or RT-PCR products in ratios from (1:9 to 9:1 PAV:PAS).

Goodness-of-fit test rejected the null hypothesis only when virus concentration in the random template samples was estimated from the RNA standard curve for peak heights at site 2088 (full-length genome sequence of PAV-Aus used as reference; Genbank accession number M21347). Examples of standard curves generated from peak area measurements at sites 2074 and 2088 are shown in Fig. 2.1. For the standard curves that were found to be good predictors of virus concentration, there was no significant difference in concentration calculated from peak areas or heights ($P$
Figure 2.1. Standard curves for peak area measurements at polymorphic sites 2074 (A, B) and 2088 (C, D) (complete genome of PAV-Aus used as reference sequence). Standard curves were created by mixing in vitro transcribed viral RNA (A, C) or (B, D) RT-PCR products in ratios from 1:9 to 9:1 (PAV : PAS).
> 0.05 for all comparisons). Peak area standard curves yielded the lowest sum of the squared differences across all SNPs (Table 2.2); therefore, these curves were used to calculate virus concentration in the experimental samples.

Table 2.2. Sum of the squared differences between the expected and calculated PAV concentration in a group of ten random PAV/PAS template mixtures. The quantity of virus added to the sample was chosen randomly. Relative PAV concentration was calculated from standard curves created by mixing in vitro transcribed viral RNA or RT-PCR products in ratios from 1:9 to 9:1 (PAV:PAS).

<table>
<thead>
<tr>
<th>Site*</th>
<th>area</th>
<th>height</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site</td>
<td>RNA</td>
<td>DNA</td>
</tr>
<tr>
<td>2074</td>
<td>1.5</td>
<td>1.4</td>
</tr>
<tr>
<td>2080</td>
<td>6.1</td>
<td>3.2</td>
</tr>
<tr>
<td>2083</td>
<td>8.7</td>
<td>10.6</td>
</tr>
<tr>
<td>2086</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>2088</td>
<td>20.4</td>
<td>7</td>
</tr>
<tr>
<td>total</td>
<td>45.7</td>
<td>25.3</td>
</tr>
</tbody>
</table>

*, complete genome of PAV-Aus used as the reference sequence; a, The null hypothesis of no difference between expected and calculated PAV concentration was rejected in a Chi-square goodness-of-fit test (α = 0.05). The chi-square failed to reject the null for all other calculated values.

Competitive interactions between PAV and PAS in mixed infections

The relative proportion of PAV significantly differed between inoculation treatments (F = 15.73, P < 0.01) (Fig. 2.2). The proportion of PAV to PAS in samples where leaf tissue from singly infected plants were combined prior to the nucleic acid extraction (‘one:one’) was less than that of plants simultaneously infected (P = 0.05). If there was no competitive interaction between species, one would expect to see the same relative concentration of PAV in both treatments. An inoculation delay of 15 days before challenge with PAS allowed significantly greater population growth of PAV when compared to other treatments. The relative proportion of PAS significantly differed between inoculation treatments (F = 11.66, P = 0.01) when PAS was inoculated to host first (Fig. 2.3). PAS concentration was greater in singly infected
plants, than plants simultaneously inoculated with both virus species. A three-day delay before challenge with PAV did not allow for greater population growth of PAS. If, however, there was a fifteen-day delay before challenge with PAV, PAS concentration was significantly greater than in other treatments.

Discussion

Standard curves generated from in vitro transcribed RNA and RT-PCR products were able to accurately estimate template concentration using peak height or area from sequence chromatograms. Since standard curves derived from RT-PCR
products perform as well as the RNA standard curves, in vitro transcription of the RNA template appears to be unnecessary for conducting a quantitative sequencing analysis. This suggests that the reverse-transcription step is completely quantitative. For SNP 2088 there is a strong incorporation bias for the base in the PAS template (T versus A). The same incorporation bias was observed in the standard curves constructed from RT-PCR products. Thus, the sequencing reaction or the normalizing function of PHRED, not the reverse transcription process preferentially selected the T in this case.

In applying quantitative sequencing to plants singly and doubly infected with PAS and PAV, it was found that the ratio of PAV to PAS template was 55:45 in singly infected plants at 33 DPI. However, when both species are inoculated to plants at the same time PAV dominates the virus population (70:30 PAV to PAS). When either virus was the primary infection a three-day delay before challenge with the other species did not change the outcome of competition, i.e. the relative template ratio was 70:30 PAV to PAS in both treatments. It appears that three days was not enough establishment time to significantly enhance or ameliorate the negative effects of competition on PAS. A similar result was found when BYDV-MAV and PAV were inoculated to the same host (Wen et al., 1991). A three day interval between the first (MAV) and second (PAV) inoculation inhibits population growth of the challenge virus at early infection stages, but by 30 DPI PAV dominates the virus population. If closely related variants of MAV are inoculated to the same host, a three day interval before challenge inoculation lead to almost complete exclusion of the challenging strain. In the present study, a fifteen-day inoculation interval lead to a stronger inhibition in accumulation of the challenge virus. This effect was more pronounced for PAV which is clearly the stronger competitor.
A limitation of the quantitative sequencing method in viral competition studies is that total virus population size in singly and doubly infected plants must be independently measured in order to understand the mechanism of competition. For example, the dominance of PAV regardless of the order of inoculation or the starting concentrations of each of the competitors does not necessarily indicate that there was direct resource competition between PAV and PAS. It may be that PAV exploits some part of the host cell machinery that is unused by PAS or the presence of PAS complements the growth of PAV but PAV does not have a reciprocal effect on PAS. Thus, the population size of PAV may be greater than that of PAS in relative terms but
it has no direct negative impact on the accumulation of PAS. It is also possible that the population size of one or both viruses is reduced by a host defense mechanism, such as virus-induced gene silencing, which would spuriously appear as competition between viruses. Measuring total virus population size (PAV + PAS) in singly and doubly infected plants would allow one to determine if there is synergy, mutual inhibition or exploitation competition in mixed infections.

A previous study found that genetic variants of satellite tobacco mosaic virus that differ at only five nucleotide sites across their entire genome can restrict host colonization and replication of the other variant depending on which is inoculated to the host first (Kurath & Dodds, 1994). One can speculate that in every replicating virus population there are competitive interactions between the progeny genomes produced early and those produced later in infection. As a result competition might enhance the effect of stochastic processes (such as transmission bottlenecks and founder effects) on genetic diversity and structure in the virus population. For the reasons described above none of the methods most commonly used to measure virus concentration in hosts could be applied to test such a hypothesis. As shown in the present study, quantitative sequencing can accurately and efficiently address questions related to the within-host dynamics of closely related pathogen genotypes.
References


CHAPTER THREE

Vector dynamics in relation to host phenology will influence the disease prevalence and population structure of Barley yellow dwarf virus

Abstract

*Barley yellow dwarf virus* (BYDV) (*Luteoviridae*) species PAV and PAS are ecologically similar in that they share aphid vectors and host species, but in agricultural fields in New York State, prevalence of PAV was found to be three times greater than that of PAS. To determine if differences in within-host population growth rate affect vector transmission efficiency, disease spread and ultimately the distribution of virus types in the host community, this study evaluated the biological characteristics of PAV and PAS species isolates in common agricultural hosts. Within infected hosts population size of PAS was greater than that of PAV at early stages of infection, but at late stages of infection population size of PAV was greater than that of PAS. At late stages of infection PAV isolates were transmitted more successfully than PAS by *Rhopalosiphum padi* when two but not eight aphids were used to carry out inoculations. The greater transmission efficiency of PAV in the laboratory study did not translate to greater spread of PAV isolates in barley, oat or wheat plots in a field experiment. In the field, the susceptibility of plants significantly declined 18 days post plant emergence, suggesting that the development of resistance as plants matured may have arrested virus spread before asymmetry in the distribution of PAV and PAS could occur. Regardless of plant age, wheat plants were more easily infected than oat or barley and disease spread farther in these hosts. Taken together these results suggest that the identity of host species and vector population dynamics in relation to the availability of susceptible hosts are key determinants of the disease prevalence in the host community.
Introduction

In order to be efficiently transmitted between hosts, pathogens must multiply to a certain density within an infected host. If a pathogen population is genetically diverse, differences among variants in their ability to exploit hosts may lead to differences in the distribution and abundance of pathogen types in the host community. Theoretical models predict that the pathogen type with the greatest rate of population growth within the host will also have the greatest prevalence in the host community (Bremermann & Thieme, 1989, Nowak & May, 1994). This body of work has mainly focused on pathogens transmitted by wind or direct contact, but in order for the model and its predictions to be valid for insect-transmitted pathogens there must be a positive relationship between within-host pathogen multiplication and the vector acquisition rate. The concentration of virus in the host impacts mosquito acquisition of West Nile virus (Komar et al., 2003) and Dengue virus (both Flaviviridae) (Armstrong & Rico-Hesse, 2001). Also, virus availability in plants that serve as virus sources for aphids has a significant impact on the transmission of Cucumoviruses (Bromoviridae)(Banik & Zitter, 1990, Pirone & Megahed, 1966) and Potato virus Y (Potyviridae) (Marte et al., 1991). Many studies that document genetic diversity in plant virus populations do not experimentally assess phenotypic differences among pathogen types. Thus, it is often difficult to determine what factors (differential transmission rates, competitive interactions between genotypes or stochastic demographic processes) contribute to the structure of the virus population. The present study explores the effect of within-host population growth on vector transmission and pathogen dispersal for two closely related species of Barley yellow dwarf virus (BYDV) (Luteoviridae).
BYDV is the name generally applied to a complex of viruses that cause yellowing and stunting in a wide array of grass species, are confined to cells in the phloem compartment of the host, and are obligately aphid transmitted. Several of these viruses have been assigned to one of two genera, *Luteovirus* (GAV, MAV, PAV and PAS) or *Polerovirus* (RPV), while others have remained unassigned (RMV, SGV). The acronyms for the virus species are derived from the names of their principal aphid vectors. PAV and PAS have the same vectors, *Rhopalosiphum padi* and *Sitobion avenae*. PAV and PAS have a single-stranded, positive-sense RNA genome approximately 5.6 kb in size. PAS and PAV are distinguished from each other by approximately 10% and 22% nucleotide divergence in their structural and replication related genes, respectively (Bencharki et al., 1999, Hall, 2006). PAS has also been found to produce greater symptom severity than PAV on susceptible and resistant varieties of oat (Chay et al., 1996) and barley (Bencharki et al., 1999). PAV and PAS have sympatric distributions in many grain producing regions worldwide (Bencharki et al., 1999, Chay et al., 1996, Mastari et al., 1998), but the distribution of each in a particular location is unknown because the majority of published field surveys conflate the prevalence of the two species. This is because PAV and PAS are indistinguishable by the most commonly employed antibody based survey methods (Chay et al., 1996). Thus, while variation in PAV/PAS genes has received intense scrutiny (Bencharki et al., 1999, Bisnieks et al., 2004, Mastari & Lapierre, 1999), the genetic structure of the virus population has been little studied (but see Mastari et al., 1998).

Aphid transmission efficiency of BYDV is sensitive to virus concentration in the source leaf (Gray et al., 1991, Gray et al., 1993). Thus, it is possible that the relative distributions of PAV and PAS in the field could be the result of differential
accumulation rates in hosts. There are only a handful of studies that examine how the host-pathogen interaction affects virus disease epidemics (Gray et al., 1994, Padgett et al., 1990, Steinlage et al., 2002). In each of these studies the within-host virus dynamics were manipulated through the use of susceptible and resistant host genotypes, and in every case plots with resistant plants (in which virus replication would be restricted) had significantly less spatial or temporal spread of disease.

BYDV is commonly isolated from multiple species of wild and cultivated grasses and there is evidence that PAV and PAS respond differently to some host species. There may then be a link between host identity and virus population structure via the effect of hosts on the probability that the aphid acquires a given genotype. No study has yet examined the epidemic spread of disease while varying both the host and virus genotype. This study compares PAV and PAS replication in a number of different host species in greenhouse grown plants in addition to initiating and monitoring epidemics in field plots. The context for the experiments is provided by a survey over one growing season of PAV and PAS populations in local agricultural fields. The combination of approaches undertaken in this study allows an evaluation of the effects of within-host virus population dynamics and host community diversity on the structure of a virus population in a managed plant community.

Materials and Methods

Field survey of PAV and PAS populations

From May-July 2002 wheat plants were collected from agricultural fields in central New York State. Only field edges were sampled but within this area plants were selected randomly. Double antibody sandwich-enzyme linked immunosorbent assay (DAS-ELISA) with PAV polyclonal antibodies (Agdia, Inc) was performed (as
described in Gray et al. 1991) to detect virus infected plants. PAV polyclonal antibodies also react with PAS isolates (Chay et al., 1996). Total nucleic acids were extracted from plants that showed an antibody reaction according to the following protocol: A glass rod was used to homogenize 200 mg of leaf tissue in 900 µl extraction buffer (7 M urea, 0.3 M NaCl, 50 mM tris pH 8.0, 20 mM EDTA and 1% SDS). The homogenate was incubated at 65°C for 10 minutes and clarified with phenol:chloroform:isoamyl alcohol (25:24:1). Nucleic acids were precipitated in a solution containing one volume 0.7 M ammonium acetate pH 5.2 and isopropanol. Reverse transcription-PCR (RT-PCR) was used to amplify the major coat protein gene (Hall, 2006), followed by digestion of the products with the restriction enzyme BstNI (Bencharki et al., 1999) to differentiate PAV and PAS infected plants.

**Virus isolates**

All virus isolates used in the greenhouse experiments were collected from agricultural fields in central New York State. PAS-129 was isolated in 1992 from migrating alate *R. padi* alighting on winter wheat and infected leaf tissue was provided to our laboratory by S. Gray (USDA, Cornell University). PAV-FA2k298 was isolated in 1998 from apterous *R. padi* collected from oat fields. PAV-WS32 and PAS-WS179 were isolated in the winter of 2004 from wheat plants. Isolates were established in greenhouse culture plants by placing field caught aphids on individually caged, healthy oats seedlings. Alternatively, adult *R. padi* (from laboratory maintained, disease-free colonies) were fed for 48 hours on field collected leaves then transferred to oat seedlings. Aphids were allowed a five day inoculation access period, after which plants were fumigated and placed in the greenhouse. Hereafter PAV-FA2k298, PAV-WS32, PAS-129 and PAS-WS179 will be referred to as PAV 1, PAV 2, PAS 1 and PAS 2, respectively.
**Virus population size and virus transmission efficiency**

Three separate experiments were conducted in order to evaluate virus population size in infected plants and the transmission efficiency of PAV and PAS isolates by *R. padi*. Experiment 1 examined the impact of virus concentration in the host and vector density on the transmission of PAV and PAS to three species of host plants. Experiment 2 examined the relative concentration of PAV and PAS template in one host species over time and virus transmission efficiency by single aphids. Experiment 3 compared the relative concentration of PAV and PAS template at 14 days post inoculation (DPI) between three species of host plants. In Exp. 1 *Hordeum vulgare* (barley cv. Romulus), *Avena sativa* (oat cv. Astro) and *Triticum aestium* (wheat cv. Huras) were individually grown from seed in four-inch pots. At the two- or three-leaf stage plants were challenged with two or eight *R. padi* previously fed on detached oat leaves infected with one of the four virus isolates. Oat plants were used as virus source approximately two months after their initial inoculation. A minimum of ten plants were inoculated per treatment. After an inoculation access period of five days plants were sprayed with insecticidal oil and placed in the greenhouse. At 21 DPI plants were harvested and analyzed for virus content by ELISA. A multi-factor ANOVA was used to analyze the proportion of plants that became infected with PAV versus PAS isolates. The main effects of the analysis were virus isolate, host species and the number of inoculating aphids. Pair-wise comparisons were made with Tukey’s procedure (α = 0.05). All statistical analyses were carried out in SAS version 9.1.

In Exp. 2, 24 wheat plants were grown from seed in four inch pots and twelve plants were inoculated with PAS 1 or PAV 1. At 8, 20, 33 and 45 DPI leaves from three plants in each virus treatment were harvested and stored at -20°C. Leaves were
also collected at 30 DPI to use as source tissue for *R. padi* feeding. Aphids were fed on leaves for 48 hours then singly transferred to individual, healthy oat seedlings. At 14 DPI ELISA was used to determine the number of indicator plants that became infected. In Exp. 3 barley, oat and wheat were grown from seed in four inch pots and three plants of each species were infected with PAS 1 or PAV 2. All plants were harvested at 14 DPI.

In Exp. 2 and Exp. 3 quantitative sequencing was used to compare virus content in PAV and PAS infected plants. We followed the quantitative sequencing protocol provided in Hall (Chapter 2). At each harvest, date equal amounts of leaf tissue from plants infected with different virus species were combined prior to nucleic acid extraction. Extracts were used as template for RT-PCR amplification of a 106-nucleotide region of the viral RNA-dependent RNA polymerase gene. Amplified products were submitted to the Cornell University Bioresource Center for direct sequencing using an ABI 3730 sequencer with Big Dye Terminator chemistry and AmpliTaq-FS DNA polymerase. The amplified region contains nine sites that differentiate PAV and PAS. The five sites that were retained after quality trimming were chosen for further analysis. PHRED as invoked by polySNP was used to calculate the height of peaks at these sites in the sequence chromatograms. It is possible to determine the relative concentration of PAV and PAS template in the experimental sample by comparing the peak heights to a standard curve of peak heights versus template concentration. Standard curves were generated by mixing known quantities (as measured by spectrophotometry, NanoDrop, Wilmington, Del.) of *in vitro* transcribed viral RNA template in several ratios (ranging from 1:9 to 9:1 PAV to PAS) prior to RT-PCR amplification and sequencing. The mixing of viral template and the subsequent sequencing analysis was replicated three times. For each
dimorphic site the height of peaks was plotted against the frequency of PAV template added to the reaction. The function of the regression line was used to determine the relative concentration of PAV template in the experimental samples when given the corresponding value for peak height.

*Design of field experiment*

In the summer of 2004 the epidemic spread of PAV and PAS isolates in barley, oat and wheat was evaluated in a field experiment. Plant species and virus isolates were randomly assigned to 4 x 1m plots arranged in four blocks. There was 1m of bare earth between plots in the same block and 2 m of bare earth between blocks. On June 7 and 8 seeds were hand sown in five rows per plot at a density of 2.5 (oat and wheat) or 2 (barley) bushels per acre. Hand weeding of the experimental field was done as needed over the course of the growing season. Plots were uncaged; therefore, one plot of each species per block was uninoculated (control plots) so that one could assess whether plots became contaminated with inoculum from outside the plot. Plants that would serve as the source of virus infection within each plot were inoculated on one of three dates post plant emergence (DPE), 18 DPE, 26 DPE and 36 DPE. On the first inoculation date, four plots of each plant species were infected with each virus isolate. Plots were inoculated by placing twenty viruliferous adult *R. padi* on two adjacent plants in the center (two meters from the end) in the center row. At 26 DPE three to six previously uninoculated plots per block of each plant species were infected with a given virus isolate. Inoculations were carried out as on the first date. On the third date plots inoculated at 26 DPE were inoculated a second time with ten viruliferous aphids placed on one plant located one meter and three meters from the end in the center row. After all species had reached growth stage 60 (anthesis) (Zadoks et al., 1974) DAS-ELISA was used to identify infected plants. A roller leaf
press was used to extract sap from collected plants. Sampling began with the inoculated source plants. Then two contiguous plants on both sides of the source plant were collected. If at least one of the two plants was infected two additional plants along the center row were analyzed. If neither of the plants was infected, infection was no longer monitored along that length of row. Sampling was repeated along the row until neither plant in a two plant sample was found to be infected.

Two types of statistical analyses were performed. First, logistic regression was used to model the probability that the inoculated source plant became infected. Predictor variables in the model were plant species inoculated, virus species and plant age at the time of inoculation. The logistic procedure estimates regression coefficients and evaluates their significance via maximum likelihood methods. If a predictor variable was found to be non-significant, the parameter was dropped from the model and the model was refit. The coefficient of the predictor variable can be interpreted as the change in odds of disease for plants that are in the membership category versus those in the reference category. One can determine the odds of disease (probability of disease / 1 - probability of disease) given the predictors in the model by exponentiation of the regression coefficients. To evaluate if the two-way interaction terms significantly improved our ability to predict disease, a likelihood ratio test was applied to the data with and without each set of terms (plant species * isolate, inoculation date * plant species, and isolate * inoculation date). Likelihood ratio tests were performed by comparing the change in -2 ln(likelihood) between the full and reduced models against a chi-square with degrees of freedom equal to the number of parameters dropped from the model (df = 1, \( \alpha = 0.05 \)). The second analysis performed was a multi-factor ANOVA of the number of plants that were found to be infected.
Results

PAV and PAS incidence in agricultural fields

Of the 262 total plants sampled, forty were found to be infected with PAV or PAS. Restriction fragment analysis of the virus content in these plants revealed that 24 were infected with PAV, 7 were infected with PAS, 2 were infected with both PAV and PAS and we were unable to recover RNA from seven plants.

Virus population size and virus transmission efficiency

In Exp. 1, barley, oat and wheat plants were inoculated with PAV or PAS isolates and harvested 21 DPI. The infection status of plants was dependent upon the number of aphids used to carry out the inoculation (F = 9.5, df = 1, P = 0.02) and the challenging isolate (F = 13.4, df = 3, P = 0.005) but not the plant species inoculated (F = 1.4, df = 3, P = 0.33). All two-way interaction terms, aphid number * plant species (F = 1.35, df = 2, P = 0.33), aphid number * isolate (F = 1.47, df = 3, P = 0.32) and isolate*plant species (F = 1.47, df = 6, P = 0.33), were non-significant. When inoculations were carried out with two aphids there was a significant difference in the number of plants that became infected with PAS as opposed to PAV isolates (P = 0.007) (Fig. 3.1). When eight aphids were used for the inoculation infection success of PAS 2 was not different from PAV 1 (P = 0.32) or PAV 2 (P = 0.32). Infection success of PAS 1 did not, however, become equivalent to that of PAV 1 (P = 0.03) or PAV 2 (P = 0.03).

In Exp. 2, virus concentration in singly infected plants was estimated by combining equal amounts of leaf tissue prior to nucleic acid extraction, RT-PCR and quantitative sequencing analysis. Virus concentration varied significantly over time (F = 43.01, P < 0.01). Multiple comparisons between assay dates were performed with Tukey’s test (α = 0.05). Relative PAV concentration significantly increased between 8
(42%) and 20 DPI (86%) (Fig 3.2). Between 20 and 33 DPI (53%) relative PAV concentration decreased. This is most likely due to a decrease in PAV population size at late stages of infection (as has been noticed in other studies (Boovaraghand et al., 2003, Ranieri et al., 1993), rather than an increase in the population size of PAS. There was no significant change in PAV template concentration between 33 and 45 DPI. Mean transmission efficiency (± standard deviation) of PAV 1 and PAS 1 did not differ significantly (67.2% ± 21.3% and 61.7% ±28.6% respectively) from infected leaf tissue collected at 30 DPI.

![Figure 3.1. Percent of barley, oat and wheat plants that became infected with barley yellow dwarf virus-PAV (cross-hatched bars) or PAS (dotted bars) isolates when inoculated with two or eight aphids carrying the appropriate strain. Experimental treatments are ranked according to the absorbance value of the infected leaves on which aphids were fed (source leaf) prior to inoculation. Source leaves are oat in all cases.](image)

In Exp. 3, barley, oat and wheat singly infected with PAV and PAS were collected at 14 DPI. The relative proportion (± standard deviation) of PAV template was 41% (± 5.6), 44% (± 2.8) and 43% (± 4.4) for barley, oat and wheat respectively.
(Fig 3.3). Thus, at 14 DPI the population size of PAV and PAS does not vary across the host species assayed in this study.

A logistic regression was used to model the probability that the source plants would become infected after inoculation. The total number of infected source plants is given in Table 3.1. Plant age at the time of inoculation ($x^2 = 41.1, P < 0.01$) and plant species ($x^2 = 22.5, P < 0.05$), but not virus species ($x^2 = 1.47, P = 0.23$) were significant predictors of the odds of disease. The odds of disease significantly decreased when plants were inoculated 26 ($P < 0.01$) or 36 DPE ($P < 0.01$) instead of 18 DPE (Table 3.2). There was no significant change in the odds of disease when plants were inoculated 26 versus 36 DPE ($P = 0.16$). Thus, a significant decrease in plant susceptibility to virus infection occurred after 18 DPE. There was a significant
decrease in the odds of disease when inoculating oat versus wheat \( (P < 0.01) \) and barley versus wheat \( (P = 0.045) \). The odds of disease significantly increased if barley was inoculated as opposed to oat \( (P < 0.01) \). The two-way interaction terms were not significant predictors of the probability of disease (data not shown).

There was too little disease spread in plots inoculated 26 or 36 DPE to perform a meaningful statistical analysis; therefore, only data for plots inoculated 18 DPE are presented here (Fig 3.4). In these plots the identity of the host species \( (df = 2, F = 4.4, P = 0.02) \) but not virus isolate \( (df = 3, F = 0.89, P = 0.46) \) had a significant effect on the distance of disease spread. The distance of disease spread in wheat plots was significantly greater than barley \( (P = 0.03) \) or oat plots \( (P = 0.009) \), but there was no

![Figure 3.3](image-url)

Figure 3.3. Relative concentration of PAV template in barley, oat and wheat plants harvested 14 days post inoculation. Equal amounts of leaf tissue from plants singly infected with PAV and PAS were combined prior to quantitative sequencing analysis. Box plots show the median, 10th, 25th, 75th and 95th percentiles.
difference between barley and oat ($P = 0.47$) plots. To evaluate whether experimental plots may have experienced contamination with outside inoculum, five plants randomly collected between 1 m and 3 m of row length in the center row of the control plots were assayed for virus. There were three wheat plants and one barley plant infected in block 1, two wheat plants infected in block 2, one wheat plant infected in block 3 and no plants infected in block 4. Based upon these data it seems that there might have been preferential colonization of wheat plots by aphid immigrants. Block was found to have no effect on our measure of disease spread ($df = 3, F = 0.34, P = 0.8$), indicating that if there was contamination of plots with virus from outside of the plot it occurred randomly across blocks.

**Discussion**

In this study the survey of natural BYDV populations conducted over one growing season found that in wheat fields the incidence of PAV was at least three times greater than that of PAS. One cannot draw definitive conclusions about the population structure of PAV and PAS from this limited data set but it does suggest testable hypotheses. Specifically, this study focused on the multiplication rate of PAV and PAS isolates in different agricultural host species and the effect of virus concentration on vector acquisition and disease spread.

Two different virus transmission assays were performed in this study. In the first experiment, antigen concentration was much higher in source leaves infected with PAV versus PAS isolates and transmission success (with two but not eight aphids) was greater for PAV isolates. Plants at an earlier stage of infection were used as virus source in the second experiment. Concentration of PAV and PAS was nearly equal in these plants (54:46 PAV to PAS) and there was no significant difference in the number of single *R. padi* transmitting either virus species. Chay *et al.* (1996) also
reported no difference in transmission efficiency of PAV and PAS isolates by *R. padi* or *S. avenae*. The greater transmission efficiency of PAV observed in Exp. 1 is probably due to its higher concentration in the host rather than differences in how virus species interact with factors inside the aphid.

Table 3.1. Number of inoculated virus source plants that became infected with barley yellow dwarf virus-PAV and PAS isolates in field plots. Plants were inoculated at 18, 26 and 36 days post emergence (DPE).

<table>
<thead>
<tr>
<th>Host species</th>
<th>Isolate</th>
<th>Inoculation 18 DPE</th>
<th>Inoculation 26 DPE</th>
<th>Inoculation 36 DPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>PAV 1</td>
<td>8/8*</td>
<td>1†</td>
<td>5/5</td>
</tr>
<tr>
<td></td>
<td>PAV 2</td>
<td>6/7</td>
<td>85</td>
<td>8/11</td>
</tr>
<tr>
<td></td>
<td>PAS 1</td>
<td>7/8</td>
<td>88</td>
<td>4/7</td>
</tr>
<tr>
<td></td>
<td>PAS 2</td>
<td>4/7</td>
<td>57</td>
<td>3/6</td>
</tr>
<tr>
<td></td>
<td>PAV 1</td>
<td>7/8</td>
<td>88</td>
<td>3/12</td>
</tr>
<tr>
<td></td>
<td>PAV 2</td>
<td>2/7</td>
<td>29</td>
<td>3/8</td>
</tr>
<tr>
<td></td>
<td>PAS 1</td>
<td>7/8</td>
<td>86</td>
<td>0/6</td>
</tr>
<tr>
<td></td>
<td>PAS 2</td>
<td>4/8</td>
<td>50</td>
<td>2/6</td>
</tr>
<tr>
<td></td>
<td>PAV 1</td>
<td>6/7</td>
<td>86</td>
<td>1/6</td>
</tr>
<tr>
<td></td>
<td>PAV 1</td>
<td>6/8</td>
<td>75</td>
<td>3/6</td>
</tr>
<tr>
<td></td>
<td>PAS 1</td>
<td>6/8</td>
<td>75</td>
<td>4/10</td>
</tr>
<tr>
<td></td>
<td>PAS 2</td>
<td>7/8</td>
<td>88</td>
<td>4/10</td>
</tr>
</tbody>
</table>

*, number of plants infected / number of plants assayed (some inoculated plants died before they could be assayed for virus infection); †, percent of infected plants

Virus population size in barley, oat and wheat was analyzed at 14 DPI and in all cases the mean proportion of PAS template was 60%. A survey of BYDV populations in France reported that 98% of viruses isolated from *Lolium multiflorum* belonged to the PAV species and 83% isolated from barley belonged to the PAS species. The authors conclude from this that hosts play a role in the selection and maintenance of diversity in BYDV. The findings of the present study may conflict with the authors’ interpretation of the survey data because we did not find any difference in the infectivity or population growth of a given virus species in any of the
host species tested. The response of French PAV and PAS isolates to infection in barley and *L. multiflorum* warrants testing experimentally, because asymmetry in the distribution of virus species could be due to stochastic demographic processes, such as transmission bottlenecks or founder effects. The Mastari *et al.* (1998) study also raises questions about the host preferences and potential population subdivision of field populations of BYDV vectors. If vectors move randomly among wild and cultivated grass species it would be selectively disadvantageous for the virus to become adapted to one host.

Table 3.2. Parameter estimates for logistic regression model. The probability that the inoculated source plant became infected is modeled as a function host age at the time of inoculation, the host species inoculated and the virus isolate. Plants were inoculated at 18, 26 and 36 days post emergence (DPE).

<table>
<thead>
<tr>
<th>Reference category</th>
<th>Membership category</th>
<th>Coefficient *</th>
<th>Change in odds (P value) †</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inoc. 18 DPE</td>
<td>Inoc. 26 DPE</td>
<td>-1.6</td>
<td>0.2 (0.0001)</td>
</tr>
<tr>
<td>Inoc. 18 DPE</td>
<td>Inoc. 36 DPE</td>
<td>-2.06</td>
<td>0.13 (0.0001)</td>
</tr>
<tr>
<td>Inoc. 26 DPE</td>
<td>Inoc. 36 DPE</td>
<td>-0.46</td>
<td>0.63 (0.16)</td>
</tr>
<tr>
<td>wheat</td>
<td>barley</td>
<td>-0.67</td>
<td>0.51 (0.045)</td>
</tr>
<tr>
<td>wheat</td>
<td>oat</td>
<td>-1.6</td>
<td>0.2 (0.0001)</td>
</tr>
<tr>
<td>oat</td>
<td>barley</td>
<td>0.93</td>
<td>2.5 (0.006)</td>
</tr>
</tbody>
</table>

*, coefficient of predictor variable in the membership category; †, change in the odds of disease for plants in the membership category versus those in the reference category.

Although we observed greater transmission success of PAV isolates in the greenhouse, this did not translate to a higher incidence of PAV in the field experiment. There are three explanations for the lack of difference in the incidence of the two strains in the experimental plots. First, in the greenhouse component of the study the success of PAS infection increased to that of PAV infection when eight aphids instead of two were used to carry out inoculations. This implies that any effect of within-host
multiplication on virus population structure will be dependent on the number of vectors present. It is possible that the number of aphids in the field plots overwhelmed any effect differences in virus multiplication might have otherwise had on the spread of PAV and PAS, i.e. at high aphid densities PAV and PAS will have a similar prevalence in the population. Conversely, it is possible that there were not enough aphids seeded into field plots and this could have the restricted the spread of both strains. Since the odds of becoming diseased significantly decreased for plants inoculated after 18 DPE, it is also plausible that an increase in resistance as plants matured arrested disease spread and prevented the development of asymmetry in the distribution of PAV and PAS. Mature plant resistance and its depressive effects on disease spread have been documented for *Wheat dwarf virus* (*Geminiviridae*) (Lindblad & Sigvald, 2004). The pattern of virus prevalence observed in natural populations may be due to differences in transmission rate between virus species, but this study may not have been able to demonstrate this effect because conditions in the field experiment may not have adequately mirrored those of natural populations. The most notable difference was that the predominant agricultural practice in the region is to sow grain crops in the fall and let them over winter as seedlings. As a result, aphid migrants leaving summer crops have an opportunity to initiate and enlarge disease foci while plants are still very young.

In the field study, wheat was more easily infected by all virus isolates and disease spread further in these plots than oat or barley plots. Thus, the main factor affecting the progress of the BYDV epidemic in the field experiment was the identity of the host. A study by Power (1991) monitored the behavior of BYDV vectors in diverse host communities and found that vector movement rates and disease spread were significantly impacted by the genotype of the host. It could be that in wheat
plots aphids fed more efficiently, had greater population growth or moved in such a way that leads to more disease spread. Taken as a whole the results of the present study indicate that many components of the vector’s biology, including population density, activity relative to the availability of susceptible hosts and behavior in response to the host genotype, will be important determinants of disease outcomes in the host community in general. There is no evidence of vector-associated selection of PAV and PAS with the *R. padi* genotypes used in this study. It appears that the transmission efficiency of PAV and PAS is dependent on their concentration in the host and not differential interactions between the virus species and the aphid vector. In order to confirm this hypothesis it is necessary to conduct transmission assays at early stages of infection. If this hypothesis is correct then epidemiological factors,
such as vector activity in relation to virus multiplication in hosts, will have important consequences for the relative abundance of virus species.
References


CHAPTER FOUR

Interference Competition Between Barley Yellow dwarf virus species affects vector transmission

Abstract

The outcome of within-host interactions between pathogen strains will affect the genetic diversity and evolutionary dynamics of the pathogen population. The PAV and PAS species of Barley yellow dwarf virus (Luteoviridae) are vectored by the same aphid species, have sympatric distributions in many geographic regions and both have been isolated from single hosts. The goal of the present study was to assess how competitive interactions between PAV and PAS within hosts may affect the structure of the virus population. There was no evidence that PAS had a negative impact on PAV population growth if there was a three-day delay before PAV was challenged with PAS or if both species were inoculated to a host simultaneously. PAS population size decreased over time in all doubly inoculated plants, but there was less impact of superinfection if there was a fifteen-day interval before challenge inoculation with PAV. There was no difference in the transmission efficiency of PAV and PAS by Rhopalosiphum padi from singly infected plants, but transmission success from doubly infected plants was associated with relative virus concentration in the host. PAV is the stronger competitor within the host, but it does not completely cross-protect plants when it is inoculated to a host first or completely take over a host when it is inoculated second. Thus, we found no evidence of competitive exclusion in this system, and diversity could be maintained in the pathogen population even though the transmission success of PAS decreases over time in doubly infected plants.
Introduction

The presence of different virus genotypes in the same host may influence within-host population dynamics (Levin & Bull, 1994), pathogen transmission and disease spread (Mansky et al., 1995, Moreno et al., 1997). As the selective environment of a given virus genotype is made up of a diverse group of genotypes, mixed infections may also influence the evolution of virus biological characteristics such as virulence (Levin & Pimentel, 1981, Nowak & May, 1994, Van Baalen & Sabelis, 1995). In agricultural and natural plant communities, mixed genotype infections are common. Thus, in many plant communities there is the potential for within host interactions between viruses to affect the structure of the virus community and the structure of the host community. Mixed genotype pathogen infections may result in negative host impacts in two ways. First, two coinfecting viruses may act synergistically to produce more severe symptoms and higher virus titers than either virus alone. Secondly, within-host competition between genotypes can select for genotypes with greater replication rates and, thereby, increase the virulence of the pathogen population. This study sought to determine how competition between two species of Barley yellow dwarf virus (BYDV) (Luteoviridae) affects the pattern of virus population growth and the consequences for vector transmission.

Competitive interactions among pathogen genotypes within hosts may influence pathogen multiplication and, as a result, the dispersal of genotypes between hosts. Theoretical models have been developed to describe the evolutionary and epidemiological dynamics that result from different types of competitive interactions. Many of these models assume a positive relationship between pathogen multiplication and virulence (defined as pathogen induced host mortality). Thus, pathogen evolution toward higher rates of multiplication will be constrained by the decreased transmission
of virulent genotypes, due to mortality shortening the duration of host infectiousness. In superinfection models one pathogen genotype can infect and ‘take over’ a host that is already infected by another genotype. Superinfection does not necessarily lead to competitive exclusion. If the weaker intra-host competitor has a greater inter-host advantage by allowing the host to live longer, both genotypes can coexist in the population (Nowak & May, 1994). In cross-protection models where host infection by one genotype (‘protector’) prevents the establishment of infection by another genotype (‘challenger’), the genotype with the highest transmission rate will dominate the population and exclude all others (Zhang & Holt, 2001). If genotypes can infect and be transmitted from the same host (co-infection), each will proliferate according to their individual reproduction rate (Frank, 1992, Van Baalen & Sabelis, 1995).

Very few studies have identified within-host competition as a factor affecting the distribution and abundance of virus genotypes in a host community. Fraile et al. (1997) found that Tobacco mosaic virus (TMV) attained only one tenth the concentration in plants doubly infected with Tobacco mild green mosaic virus (TMGMV) (Tobamovirus) as singly infected plants. They propose that a decrease in the transmissibility of TMV, as the result of superinfection by TMGMV, is the most likely cause for the disappearance of TMV from wild Nicotiana glauca populations in Australia. Also, BYDV species PAV and MAV exhibit reciprocal cross-protection (Wen et al., 1991). Zhang and Holt (2001) were able to fit a cross-protection competition model to field data from New York State showing a shift in dominance from MAV to PAV over a thirty year period. A field survey conducted by Hall (chapter 3) found that in fields in central New York State the prevalence of PAV is three times greater than that of BYDV-PAS. This pattern may have been influenced by within-host interactions between virus species, and as such may represent
coexistence between the two viruses at different population densities or an intermediate stage in the complete exclusion of one virus.

The grass infecting members of the Luteoviridae have been grouped into two genera, luteovirus (BYDV) containing GAV, MAV, PAV and PAS and polerovirus (Cereal yellow dwarf virus, CYDV) containing RPV. Two species, SGV and RMV, have yet to be assigned to a genus within the family. The virus species acronyms are derived from the name of the aphid species that most efficiently facilitate virus transmission. PAV and PAS are transmitted by Rhopalosiphum padi and Sitobion avenae. PAV and PAS have sympatric distributions in many parts of their range, share host species and have been identified in mixed infections. PAV and PAS are distinguished by 10% nucleotide divergence in their coat protein gene (Bencharki et al., 1999) and 22% nucleotide divergence in their RNA-dependent RNA polymerase gene (RDRP) (Hall, 2006). Also, some PAS isolates have been shown to induce more severe symptoms than PAV isolates on barley and oat (Bencharki et al., 1999, Chay et al., 1996). In the present study we carried out a series of competition assays to better understand how within-host interactions between virus species affect the structure of the virus population. To determine if there is a stable competitive hierarchy among species we varied the order of inoculation and the time interval between the first and second inoculation. The potential effects of competition on virus population structure were evaluated by aphid transmission assays of virus from the experimental plants.

Materials and Methods

Inoculation of experimental plants

Wheat (Triticum aestium) plants were individually grown from seed in six inch pots and randomly assigned to one of seven inoculation treatments: single inoculation
with PAV, single inoculation with PAS, simultaneous inoculation with PAV and PAS, PAV challenged with PAS after a three day delay, PAV challenged with PAS after a fifteen day delay, PAS challenged with PAV after a three day delay, and PAS challenged with PAV after a fifteen day delay. There were twelve plants per inoculation treatment. Protection inoculations and single inoculations were carried out when plants were ten days old. To inoculate plants *R. padi* from laboratory maintained disease-free colonies were fed for 48 hours on detached oat leaves infected with PAV or PAS isolates then transferred to individual wheat seedlings (eight aphids for single inoculations and four aphids for protection inoculations). Three days after the initial inoculation, plants in the short delay treatment were challenged with a second virus species transmitted by an additional four aphids. After a total inoculation access period of five days, all pots were sprayed with insecticidal oil and placed in the greenhouse. Fifteen days after the protection, inoculation plants in the long delay treatment were challenged with a second virus species transmitted by an additional four aphids. PAV and PAS isolates used in this study were collected from agricultural fields in central New York State. Isolate PAS-129 was obtained in 1992 from migrating alate *R. padi* alighting on winter wheat (Chay *et al.* 1996). PAV isolate Fa2k298 was obtained in 1998 from apterous *R. padi* collected from oat fields. Viruses were maintained in continuous culture in greenhouse grown oat plants since their initial isolation.

*Nucleic acid extraction, reverse transcription-PCR and sequencing*

At 8, 20, 33, and 45 DPI (days post inoculation with the protecting strain) leaves were harvested from at least three plants per treatment and stored at -80 °C. Quantitative sequencing was used to measure the relative concentration of PAV of PAS in infected plants. A full description of the protocol has been previously
described (Chapter 2). Briefly, equal amounts of leaf tissue from plants singly inoculated with PAV and PAS isolates were combined into one sample prior to the nucleic acid extraction (‘one:one’). For other treatments a total nucleic acid extraction was performed on 200 mg of tissue from a single leaf. Leaf tissue was homogenized with a glass rod in 900 μL extraction buffer (7 M urea, 0.3 M NaCl, 50 mM tris pH 8.0, 20 mM EDTA and 1% SDS). The homogenate was incubated for ten minutes at 60°C and extracted with phenol:chloroform:isoamyl alcohol (25:24:1). The supernatant was precipitated in 3 M ammonium acetate and one volume of isopropanol. Reverse transcription-PCR (RT-PCR) was used to amplify a 106-nucleotide region of the viral RNA-dependent RNA polymerase. Single-step RT-PCR was carried out in a 50 μL reaction volume containing 25 mM tris pH 8.8, 10 mM KCl, 10 mM dTT, 10 mM (NH₄)₂SO₄, 2.5 mM MgSO₄, 1 μM forward (5′-AGAGGC CACAGAATGTCCGG-3′) and reverse (5′-GTTCAGCTTCAACACCCAGC-3′) oligonucleotide primers, 200 μM dNTPs, 30 units RNase inhibitor, 50 units RNase H reverse transcriptase and 5 units Taq polymerase (all enzymes Invitrogen, Carlsbad, Calif.). Thermocycling conditions were: 1 cycle of 45 min at 42°C for reverse transcription, 1 cycle of 2 min at 95°C for inactivation of reverse transcriptase, 35 cycles of 30 s at 94°C, 30 s 60°C, 1 min at 72°C, and a final extension of 10 min at 72°C. RT-PCR products were purified using the Invitrogen PCR clean-up kit following the protocol prescribed by the manufacturer and then submitted to the Cornell University Bioresource Center for direct sequencing using an ABI 3730 sequencer with Big Dye Terminator chemistry and AmpliTaq-FS DNA polymerase. Sequencing was carried out in the forward direction with the primer indicated above. 

Quantitative sequencing
The gene region sequenced has nine nucleotide sites that differentiate PAV and PAS. The five sites that were consistently retained across sequencing runs after quality trimming \((P = 0.10)\) with PHRED version 0.020425.c (Ewing et al., 1998) were used for analysis. PHRED (with the assistance of MUSCLE (Edgar, 2004)) as invoked by polySNP was used to estimate peak height at polymorphic sites from sequence chromatograms. As previously described, a standard curve for each of the five sites was generated by mixing known quantities \textit{in vitro} transcribed viral RNA template in ratios from 1:9 to 9:1 (PAV:PAS) prior to RT-PCR and sequencing. PAV as a proportion of the total peak height was plotted against the frequency of template added to the sequencing reaction. The equation of the best-fit line was used to calculate the relative concentration of virus in the experimental sample when given the corresponding peak height (regression functions for each standard curve can be found in chapter 2). For a given sample, each polymorphic site was treated as an independent measure of virus concentration. In order to determine if the addition of another virus affected the population growth of PAV or PAS, a two-factor analysis of variance (with main effects inoculation treatment and day) was implemented in SAS 9.1. To make pair-wise comparisons, t-tests were used test the null hypothesis that the difference between the least squares mean of each category is equal to zero. Specifically, these test assessed the effect of double infection (one:one versus simultaneous inoculation) and the effect of inoculation delay (simultaneous inoculation versus inoculation delay) on virus population growth. To determine if the concentration of PAV and PAS differs within a treatment, a t-test was used to test the null hypothesis that the mean proportion of PAV does not significantly differ from 50\% of the total virus concentration (i.e. chromatogram peak area). To formally assess the strength of cross-protection a group of control plants must be inoculated
with PAV and PAS (single and simultaneous infection) at the same time as each challenge inoculation. This ensures that plants in the inoculation delay treatments will be compared to control plants that are infected when they are the same age. This study did not control for the effect of plant age on the pattern of virus accumulation, nonetheless, concentration of the challenge virus was plotted against the number of days the virus was in the plant. This allows one to address in a general way if the population growth of one virus species is affected by prior infection with the other species.

_Aphid transmission from experimental plants_

At 30 DPI leaves were collected from at least two plants in all treatments. Single non-viruliferous aphids were fed on detached leaves for 48 hours then individually transferred to single, healthy oat seedlings. At 14 DPI enzyme linked immunosorbent assay with PAV polyclonal antibodies was used to identify virus infected plants. PAV antibodies also react with PAS. To determine which virus species was the cause of the infection, RT-PCR was performed on total plant nucleic acid extracts (protocol and primers as described above). Amplified products were digested with restriction enzymes _Ase_ I which specifically digests PAV and _Nci_ I which specifically digests PAS according the supplier’s recommendations (New England Biolabs, Ipswich, Mass.).

**Results**

_The effect of competition on PAV population growth_

At 8 DPI there was no difference in the relative proportion of PAV in singly infected plants (47%, _P_ = 0.17) versus those inoculated with both virus species simultaneously (33%) (Fig 4.1). When PAV was challenged with PAS after a three-
day delay, PAV represented 85% of the virus population. By 20 DPI relative concentration of PAV increased to 58% in simultaneously inoculated plants (58%), but this is significantly less than its relative concentration in singly infected plants (90%, $P < 0.01$). At 33 DPI relative PAV concentration in simultaneously inoculated plants (66%) is greater than that observed in singly infected plants (55%, $P < 0.01$), but there is no difference in virus concentration if PAS follows PAV after a three-day inoculation delay (64%, $P = 0.4$). In the latter treatment, PAV concentration decreased between 20 and 33 DPI ($P = 0.02$), but as this decrease was also noticed in singly infected plants it is most likely not due to a competitive effect of PAS on PAV. If there was a fifteen-day interval before PAS inoculation, PAV represented 92% of the virus template population at 20 and 33 DPI, but decreased to approximately 65% by 45 DPI. In fact at 45 DPI, PAV was approximately 70% of the virus template in all double inoculation treatments. This is in contrast to singly infected plants, where the relative concentration of PAV (53%) was significantly less than other treatments ($P < 0.05$).

The effect of competition on PAS population growth

At 8 DPI PAS represented a greater proportion of the virus population in simultaneously infected plants than singly infected plants ($P = 0.02$) (Fig. 4.2A). A three-day delay before challenge inoculation with PAV did not enhance the population growth of PAS ($P = 0.16$). In simultaneously infected plants PAS was the dominant member of the virus population (63%, $t = 3.08$, $P < 0.01$), but as in all doubly infected plants, there was until 33 DPI. This is probably due to an increase in PAV concentration and a concomitant decrease in PAS concentration after its peak near 8 DPI. When PAV follows PAS after a fifteen-day interval, the decline in PAS population size begins soon after the challenge inoculation. However, at 33 ($P < 0.01$)
and 45 DPI ($P < 0.01$) the population size of PAS is greater when there is fifteen-day interval before challenge inoculation when compared to plants simultaneously inoculated.

A) 

![Graph A](image1)

B) 

![Graph B](image2)

Figure 4.1. Mean proportion of PAV template when PAV was the protecting virus (A) or the challenging virus (B). Inoculation treatments were: equal amounts of leaf tissue from plants singly infected with PAV and PAS paired prior to nucleic acid extraction (one:one), simultaneous inoculation with PAV and PAS, PAV challenged with PAS three days later (PAV 3 day), PAV challenged with PAS fifteen days later (PAV 15 day), PAS challenged with PAV three days later (PAS 3 day) and PAS challenged with PAV fifteen days later (PAS 15 day).
**Interference competition (cross-protection) between PAV and PAS**

When compared to plants that were simultaneously inoculated, a three-day interval between PAS (protector) and PAV (challenger) inoculation had very little impact on the rate of PAV population growth (Fig 4.1B). If, however, there was a fifteen day interval before challenge inoculation with PAV, its concentration was greatly reduced. For example, at 30 days post challenge inoculation with PAV, PAV was approximately 70% of the total viral population in the three-day delay treatment but was only 50% after a fifteen-day delay. A three-day or fifteen-day delay before challenge inoculation with PAS greatly reduced relative concentration of PAS when compared to plants simultaneously inoculated with both species (Fig 4.2B). Regardless of the length of time before PAS inoculation, by 30 days post challenge PAS represented approximately 35% of the virus in all double inoculation treatments.

**The effect of virus competitive interactions on aphid transmission**

Mean transmission efficiency (± standard deviation) by *R. padi* from singly infected plants was 67.2% (± 21.3) for PAV and 61.7% (± 28.6%) for PAS. Thus, there is no evidence that PAV and PAS differ in their ability to be transmitted from single infections at 30 DPI. RT-PCR was able to recover both PAV and PAS from all doubly inoculated plants used for aphid feeding. Aphids, however, recovered PAS less frequently than PAV from plants in four of the five doubly inoculation treatments (Table 4.1). When aphids were used to transmit virus from plants that received PAV fifteen days after PAS 100% of infected indicator plants carried PAS while 71% carried PAV. This is also the only treatment where indicator plants were found to be infected with PAS only. In the other treatments 100% of the infected plants assayed
carried PAV while a smaller fraction carried PAS. No PAS was transmitted by aphids from plants where PAS followed PAV after a fifteen day interval.

Figure 4.2. Mean proportion of PAS template when PAS was the protecting virus (A) or the challenging virus (B). Inoculation treatments were: equal amounts of leaf tissue from plants singly infected with PAV and PAS paired prior to nucleic acid extraction (one:one), simultaneous inoculation with PAV and PAS, PAV challenged with PAS three days later (PAV 3 day), PAV challenged with PAS fifteen days later (PAV 15 day), PAS challenged with PAV three days later (PAS 3 day) and PAS challenged with PAV fifteen days later (PAS 15 day).
Table. 4.1. Number of plants that became infected with the PAV and PAS after inoculation by aphids previously fed on leaf tissue from plants in the indicated inoculation treatment. Aphids were fed on source tissue thirty days after protection inoculation.

<table>
<thead>
<tr>
<th>Protecting virus</th>
<th>Challenging virus</th>
<th>Inoculation interval</th>
<th>Number of plants assayed*</th>
<th>Plants carrying PAV</th>
<th>Plants carrying PAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAV</td>
<td>none</td>
<td>none</td>
<td>57</td>
<td>38</td>
<td>-</td>
</tr>
<tr>
<td>PAS</td>
<td>none</td>
<td>none</td>
<td>60</td>
<td>-</td>
<td>36</td>
</tr>
<tr>
<td>PAV</td>
<td>PAS</td>
<td>none</td>
<td>27</td>
<td>27</td>
<td>12</td>
</tr>
<tr>
<td>PAV</td>
<td>PAS</td>
<td>3</td>
<td>18</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>PAV</td>
<td>PAS</td>
<td>15</td>
<td>13</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>PAS</td>
<td>PAV</td>
<td>3</td>
<td>18</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>PAS</td>
<td>PAV</td>
<td>15</td>
<td>7</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

*, Restriction fragment analysis with enzymes that specifically digest PAV or PAS was used to determine the infection status of inoculated plants.

Discussion

Irrespective of the order of inoculation or the time interval between protection and challenge PAV eventually came to dominate the virus population. In four of the five double inoculation treatments a 70:30 PAV to PAS template ratio was established by 30 days post challenge and maintained until the end of the experiment at 45 DPI. The movement of virus concentration towards one ratio despite differences in the initial virus concentrations may indicate that 70:30 is the stable equilibrium population density. Lotka-Volterra models that describe competition between two species competing for the same resources allow long-term coexistence provided the carrying capacity and competitive effect of each species takes on a certain set of values. In order to draw clear conclusions about the long-term coexistence of PAV and PAS the period of competition should be extended to the flowering stage, but Lotka-Volterra models may provide a useful framework for understanding what parameters contribute to the coexistence of competing virus genotypes. This is of considerable importance because coexistence between virus genetic variants in a single host provides a
mechanism for the maintenance of virus genetic diversity in the larger host community.

The observed 70:30 PAV to PAS template ratio could also be the result of apparent competition between virus species mediated by the host defense response. The term apparent competition refers to indirect interactions between species that negatively impact the population growth of one of the competitors. Population growth of PAV and RPV RNA in singly infected wheat plants was found to decrease after 6 DPI (Boovaraghan et al., 2003). The authors hypothesize that virus-induced gene silencing may have lead to the degradation of the viral RNA. In the current study, the 70:30 PAV to PAS template ratio could be explained by an asymmetrical effect of plant defenses on the population growth of the two species. Apparent competition mediated by the host immune system has been demonstrated for coinfecting genotypes of the rodent malaria parasite Plasmodium chabaudi (Raberg et al., 2006). However, in mixed infections competitive suppression of one clone was alleviated but not absent in immunodeficient mice, indicating that there was also competition between clones for host resources. The virus genes under selection will differ depending on the source of the competitive effect. Exploitation competition will select for PAS genotypes that have higher replication rates. If the host mediates the dominance of PAV, this will select for PAS isolates that can evade or disable host defenses. Thus, exploitation competition and apparent competition may lead to different evolutionary outcomes for the pathogen population.

PAV is clearly the stronger competitor, but both species can interfere with the population growth of the other if transmitted to the host first. Consequently, the persistence of cross-protection (and conversely the rate of superinfection), but not the competitive hierarchy among species is dependent on the inoculation order and the
time interval between protection and challenge. There were two instances where virus concentration was very different between replicates within the same treatment. At 8 DPI when PAV followed PAS after a three-day inoculation delay, the proportion of PAS template was 90%, 75% and 23%. In the same treatment but in plants sampled at 20 DPI (different plants than those sampled earlier), PAS concentration was measured to be 86%, 41%, and 33%. These results suggest that there is variation in the ability of PAS to protect against PAV when there is a short time interval between protection and challenge. It is possible that the magnitude of protection is dependent on the initial dose of the protecting virus or the proximity between the sites of protection and challenge inoculation. Neither of these factors was controlled between replicates because live aphids were used to transmit virus from source tissue and whole plants, not single leaves, were infested at the time of inoculation.

In order for cross-protection to occur, the protecting virus must be in the cell at the time of challenge inoculation. A short inoculation interval may limit the ability of PAS to occupy the cells that will become infected upon challenge inoculation with PAV. Longer inoculation intervals enhanced the persistence of cross-protection between genetic variants of Tobacco mosaic virus (Tobamovirus) (Tenllado et al., 1997), genetic variants of Satellite tobacco mosaic virus (STMV) (Kurath & Dodds, 1994), Beet soilborne mosaic virus and Beet necrotic yellow vein virus (Furovirus) (Mahmood & Rush, 1999) and several luteovirus species, including PAV and MAV, PAV and SGV, and genetic variants of MAV (Wen et al., 1991).

The mechanism responsible for cross-protection is not understood, and experimental studies suggest that several mechanisms are involved in the specific interference between two coinfecting viruses. A preexisting virus infection could inhibit the uncoating of the challenge virus, thereby preventing the initiation of the
replicative cycle (Sherwood & Fulton, 1982), or antisense binding could inhibit the translation or transcription of the challenge virus nucleic acids (Palukaitis & Zaitlin, 1984). If these direct interactions between virus variants are responsible for cross-protection, then it can be considered a mechanism of interference competition (reviewed in Sherwood, 1987). However, neither of these mechanisms can explain mutual inhibition between MAV, PAV and SGV in mixed infections (Jedlinski & Brown, 1965, Wen et al., 1991). Jedlinski and Brown (1965) found complete recovery of oat plants when simultaneously infected with PAV and MAV isolates at the one-leaf stage. This suggests a host defense response against virus infection. It is possible that gene silencing activated by the initial virus infection slows the accumulation of the challenge virus RNA. A number of studies have documented a decrease in virus antigen (Gray et al., 1993, Ranieri et al., 1993) or RNA concentration (Boovaraghan et al., 2003) over time in BYDV infected plants, which allows one to speculate that single virus infections can illicit a defense response by the host.

Typically, the degree of relatedness among pathogen types is also associated with the incidence and persistence of cross-protection. Wen et al. (1991) performed cross-protection assays with a number of Luteoviridae species. There was a greater degree of cross-protection among more closely related pairs of competing viruses. For example, there was no cross-protection when RPV was challenged with PAV or RMV, but there was strong cross-protection when MAV was challenged with another MAV isolate or PAV. Apparent competition and interference competition models are compatible with these observations. There is, however, strong evidence that the virus coat protein itself can significantly inhibit secondary infection by closely related virus types (Sherwood and Fulton, 1982, Sherwood, 1987).
The genotype of the virus, the genotype of the aphid, virus concentration in the source tissue and the length of time the aphid feeds are key determinants of whether the aphid will acquire the virus from an infected plant and inoculate a healthy plant during the feeding process (reviewed in Gray and Gildow, 2003). For the aphids and virus isolates used in this study the probability of successful transmission from singly infected plants was 67% and 61% for PAV and PAS, respectively. The frequency of transmission from doubly infected plants was associated with the relative concentration of virus template in the source tissue. For example, if there was fifteen-day interval before challenge with PAV, the PAV to PAS template ratio was 50:50 and 71% of infected indicator plants carried PAV and 100% carried PAS. If there was a fifteen-day interval before challenge with PAS, the template ratio was 92:8 and 100% of indicator plants became infected with PAV and no plants became infected with PAS. In simultaneously inoculated plants the proportion of PAV to PAS template was 66:34 and 100% of infected indicator plant carried PAV and 24% carried PAS. This study found no evidence that aphids selectively transmit PAV or PAS isolates, rather the observed transmission bottleneck is the result of interference competition, exploitation competition and/or apparent competition between the two virus species. Recent studies with *Citrus tristeza virus* (*Closteroviridae*) and *Cucumber mosaic virus* (*Bromoviridae*) (Ali et al., 2006, d'Urso et al., 2000) have reported that aphid transmission acts as a bottleneck that affects the genetic structure of the virus population. Neither of these studies addresses the possibility that within-host competition between virus variants affected their availability for aphid transmission. A one year field survey of PAV and PAS isolates in agricultural fields in upstate New York found that the total disease prevalence was 15.2 %, and at least 60% of these plants were infected with PAV isolates and 17.5% were infected PAS isolates. Of the
infected plants 2% contained both PAV and PAS isolates, which is what one would expect by chance given the total number of infected plants. It is most likely that when disease incidence is low, as was observed in the field survey, PAV and PAS epidemics proceed independently. But in years where there is relatively high disease incidence, one would expect competition between PAV and PAS to influence the structure of the virus population.
References


APPENDIX

Geographic region of origin and accession numbers of virus isolates used in the phylogenetic analysis.

<table>
<thead>
<tr>
<th>Isolate name *</th>
<th>Region of origin</th>
<th>Genomic regions analyzed §</th>
<th>Database accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>MA952 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007923</td>
</tr>
<tr>
<td>MA9501 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007929</td>
</tr>
<tr>
<td>MA9502 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007918</td>
</tr>
<tr>
<td>MA9504 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007919</td>
</tr>
<tr>
<td>MA9513 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007924</td>
</tr>
<tr>
<td>MA9508 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007921</td>
</tr>
<tr>
<td>MA9505 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007920</td>
</tr>
<tr>
<td>MA9511 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007922</td>
</tr>
<tr>
<td>MA9514 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007925</td>
</tr>
<tr>
<td>MA9516 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007926</td>
</tr>
<tr>
<td>MA9517 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007927</td>
</tr>
<tr>
<td>MA9415 †</td>
<td>Morocco</td>
<td>ORF 3/4</td>
<td>AJ007928</td>
</tr>
<tr>
<td>FA2K298 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285674, DQ286379</td>
</tr>
<tr>
<td>W195 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285673, DQ286375</td>
</tr>
<tr>
<td>WG2 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285672, DQ286376</td>
</tr>
<tr>
<td>WS3603 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285671, DQ286378</td>
</tr>
<tr>
<td>WG13 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285675, DQ286377</td>
</tr>
<tr>
<td>CC99 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285677, DQ286385</td>
</tr>
<tr>
<td>WD2 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285676, DQ286384</td>
</tr>
<tr>
<td>WS5003 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285681, DQ286380</td>
</tr>
<tr>
<td>WS6603 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285680, DQ286381</td>
</tr>
<tr>
<td>PAS-64 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285679, DQ286382</td>
</tr>
<tr>
<td>PAS-129 ‡</td>
<td>New York</td>
<td>ORF 2/3/4, CG</td>
<td>DQ285682, DQ286383</td>
</tr>
<tr>
<td>WF12 ‡</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>DQ285678, DQ286386</td>
</tr>
<tr>
<td>PAV-Aus †</td>
<td>Australia</td>
<td>ORF 2/3/4, CG</td>
<td>M21347</td>
</tr>
<tr>
<td>CAN-1 †</td>
<td>Canada</td>
<td>ORF 2</td>
<td>AJ810419</td>
</tr>
<tr>
<td>PH2 ‡</td>
<td>California</td>
<td>ORF 2/3/4</td>
<td>AY819231</td>
</tr>
<tr>
<td>CN-PAV †</td>
<td>China</td>
<td>ORF 2/3/4, CG</td>
<td>AF1912967</td>
</tr>
<tr>
<td>GAV-30, GP1 †</td>
<td>China</td>
<td>ORF 2/3/4, CG</td>
<td>AF338909, NC004666</td>
</tr>
<tr>
<td>ASL-1 †</td>
<td>Germany</td>
<td>ORF 2/3/4, CG</td>
<td>AJ810418</td>
</tr>
<tr>
<td>Fin-PAV †</td>
<td>Finland</td>
<td>ORF 2</td>
<td>AY926496</td>
</tr>
<tr>
<td>SGV-AY, NY †</td>
<td>New York</td>
<td>ORF 2/3/4</td>
<td>AY541039, U06865</td>
</tr>
<tr>
<td>MAV †</td>
<td>New York</td>
<td>ORF 2/3/4, CG</td>
<td>D11028</td>
</tr>
<tr>
<td>P-PAV †</td>
<td>Pennsylvania</td>
<td>ORF 2/3/4, CG</td>
<td>D11032</td>
</tr>
<tr>
<td>RPV †</td>
<td>New York</td>
<td>CG</td>
<td>NC_004751</td>
</tr>
<tr>
<td>SbDV-DS,TAS1 †</td>
<td>Japan</td>
<td>CG</td>
<td>AB038149, L24049</td>
</tr>
</tbody>
</table>

PAV isolates in italics and PAS isolates in bold, † Virus isolate obtained from Genbank, ‡ Virus isolates sequenced in this study, § ORF2 encodes the RNA-dependent RNA polymerase, ORF3 encodes the coat protein, ORF4 encodes the movement protein and CG is the complete genome sequence.