April 2003

EVOLUTION OF WHEAT STREAK MOSAIC VIRUS: Dynamics of Population Growth Within Plants May Explain Limited Variation

Roy C. French
University of Nebraska-Lincoln, rfrench2@unl.edu

Drake C. Stenger
University of Nebraska-Lincoln

Annu. Rev. Phytopathol. 2003. 41:199–214
doi: 10.1146/annurev.phyto.41.052002.095559
First published online as a Review in Advance on April 29, 2003

EVOLUTION OF WHEAT STREAK MOSAIC VIRUS: Dynamics of Population Growth Within Plants May Explain Limited Variation*

Roy French and Drake C. Stenger
USDA-ARS, 344 Keim Hall, University of Nebraska, Lincoln, Nebraska 68583; email: rfrench@unlnotes.unl.edu; dstenger@unlnotes.unl.edu

Key Words population genetics, plant virus, Tritimovirus, Potyviridae

■ **Abstract** Like many other plant RNA viruses, *Wheat streak mosaic virus* (WSMV) sequence diversity within and among infected plants is low given the large number of virions produced. This may be explained by considering aspects of plant virus life history. Intracellular replication of RNA viruses is predominately linear, not exponential, which means that the rate at which mutations accumulate also is linear. Bottlenecks during systemic movement further limit diversity. Analysis of mixed infections with two WSMV isolates suggests that about four viral genomes participate in systemic invasion of each tiller. Low effective population size increases the role of stochastic processes on dynamics of plant virus population genetics and evolution. Despite low pair-wise diversity among isolates, the number of polymorphic sites within the U.S. population is about the same as between divergent strains or a sister species. Characteristics of polymorphism in the WSMV coat protein gene suggest that most variation appears neutral.

INTRODUCTION

Wheat streak mosaic virus (WSMV) is the type species of the genus *Tritimovirus* within the family Potyviridae (67) and is found in most major wheat-growing regions of the world (4). WSMV has a relatively broad host range encompassing many plants in the grass family. The virus infects all varieties of wheat (*Triticum aestivum* L), and most isolates can infect barley (*Hordeum vulgare* L.) and oats (*Avena sativa* L.). Some varieties of maize (*Zea mays* L.) and millet (*Panicum*, *Setaria*, and *Echinocholoa* L. spp.) also are susceptible to WSMV (4). Although transmitted by an eriophyid mite (64) rather than aphids, the genome organization and cytopathology of WSMV is essentially the same as viruses in the genus

Potyvirus. The ~9384-nucleotide (nt) WSMV RNA is translated as a single polyprotein that is processed by three viral proteinases (9–11).

WSMV is an especially serious pathogen of wheat in the Great Plains region of the United States. WSMV is estimated to reduce annual wheat yields by about 5% per year in the region (12) but local disease outbreaks can be very destructive, with yield losses approaching 100%. Some factors involved in the epidemiology of WSMV have been identified (4, 12, 64, 65, 72) but many unresolved questions remain. For example, the disease is often associated with the presence of volunteer wheat, which serves as an over-summering reservoir for both WSMV and its eriophyid mite vector, the wheat curl mite (*Aceria tosichella* Keifer [Amrine]), but widespread disease epidemics also occur unpredictably.

Polymorphism may be discerned at all levels of population structure. While a WSMV isolate may be defined by a consensus sequence, closely related genotypic variants are found within single isolates (24). Numerous WSMV genotypes are present within field populations and may be distinguishable by restriction fragment length polymorphism (RFLP) analysis (42), or in some instances by serology (43, 69). Divergent strains of WSMV occur in the United States, Mexico, and Eurasia (3–5, 7, 8, 40–42, 48, 55, 69). Given the observed levels of genetic diversity, population genetic analyses may generate insights into WSMV epidemiology and evolutionary history while providing an interesting case study for understanding dynamics of plant virus populations in general.

Population genetics is primarily the study of processes that affect changes in allele frequencies in populations over time and space (25). Such processes include mutation, natural selection, genetic drift, migration, population growth or decline, and recombination. Quantitative population genetic study of plant viruses is in its infancy, with few suitable data sets available. A recent review of plant virus population genetics provides a cogent and thoughtful discussion of the issues awaiting resolution and outlines future areas of research (18). Variation among isolates of plant viruses is well documented (1, 2, 8, 30, 32–34, 42, 45, 48–51, 54, 61–63, 66, 69, 71). However, diversity within a plant virus population is generally less than might be expected given their population sizes, so that most seem to be genetically stable (18).

A generally held view is that while plant RNA viruses have the potential to vary because of error-prone polymerases that lack proofreading capabilities, the resulting population does not exhibit extreme variability (20). This implies that selection plays an important role to restrict diversity and experimental results are often interpreted in this light (58, 59). For the most part, little direct evidence for stringent selection is available and other explanations are possible (18, 44). Through the use and manipulation of infectious cDNA clones of plant viruses, we know that many viral genomes tolerate change. Restriction endonuclease cleavage sites and fairly large insertions are easily introduced into WSMV (11), and mutagenesis studies utilizing methods such alanine scanning (21) or codon insertions (31) would not be very informative unless many of the introduced changes were tolerated. Thus, there seems to be a paradox in that plant RNA viruses have the potential to vary and

have the capacity to tolerate introduced changes, yet natural populations generally exhibit limited variation (1, 18, 50). Here we suggest that inherent dynamics of virus population growth within plants is one explanation.

Diversity can be measured as the number of nt positions that are variable in the population sample (i.e., the number of segregating polymorphic sites and its summary statistic *theta*) (46) or as the mean pairwise nt sequence diversity (summary statistic *pi*) (46) among all individuals in the sample. Under the model of selective neutrality both *theta* and *pi* are equal to $2N \times mu$, where N is population size and *mu* is the population mutation rate. Genetic stability in this context means that either N or *mu*, or both, are smaller than expected. The neutral theory of molecular evolution was first developed by Kimura (28) and later modified by Ohta (47). It assumes that most mutations are either selectively neutral or slightly deleterious. Because genetic processes are random and population sizes are finite, the spread of such mutations also is stochastic. Under the neutral model allele frequencies vary from generation to generation owing to a random sampling process, called random genetic drift. Population size controls the rate of drift where a new neutral mutation takes an average of N generations to become fixed. Note that although an average time to fixation may be calculated, the process is stochastic, such that any single mutant may vary with respect to time of fixation, or even be lost (become extinct) from the population.

Selection also is affected by drift. If the population size is small enough, genetic drift can bring even disadvantageous mutations to fixation. On the other hand, selectively advantageous mutations usually become fixed, with time to fixation much more rapid than the fixation of a neutral variant. However, even advantageous mutations undergo genetic drift and are occasionally lost from the population (13, 15, 36, 52). The applicability of the neutral model has been much debated but it is generally agreed that it serves as a useful null hypothesis for testing evolutionary models and theories. Often *Ne*, the effective population size, is substituted for actual population size. *Ne* is often much less than N due to population bottlenecks and the variance among individuals in the population with respect to reproduction. In the next section we argue that relatively few viral genomes produce many copies, whereas most produce no copies at all, such that very low values of *Ne* are not unrealistic for plant virus populations.

RNA VIRUSES ARE MADE ONE (NOT TWO) AT A TIME

The mode of replication of positive-strand RNA plant viruses has important consequences for understanding the rates at which mutations accumulate. We go into some detail here because RNA virus replication is quite different from that of organisms such as bacteria or double-stranded (ds) DNA viruses that likely are more familiar to population geneticists. A generalization of plus strand plant virus replication is as follows: A viral genome enters an uninfected cell (through a wound or from an adjacent cell). If the genome is contained in a virion, the first step

is unencapsidation, followed by translation of the RNA to produce viral proteins including one or more required to generate the viral replicase. The replicase then copies the positive strand to make negative strands. These in turn are used as templates to produce multiple progeny plus strands that accumulate in the cell and, following encapsidation by coat protein, form new virions. The key feature here is that multiple plus strands are copied from each negative strand template, i.e., the bulk of plus strands are produced by a linear process of independent, successive rounds of transcription. Some progeny plus strands may be copied to provide more negative strand templates, such that a certain amount of exponential growth may occur as well. Nevertheless, the majority of plus strand production should resemble the linear "stamping machine" model envisioned by Luria (37).

During replication of plus strand RNA viruses many more plus strands are made than negative strands, but quantitative data on the dynamics of intracellular replication of plant viruses are not available. Two quantitative studies of intracellular replication of RNA bacteriophages may serve at least as qualitative guides. First is a study of Q-beta phage growth dynamics (16). The intracellular growth cycle lasts for about 40 min. During the first 10 min post inoculation little viral protein or RNA synthesis is seen. Over the next 5 to 7 min synthesis rates of both increase explosively at an exponential, or perhaps at an even faster hyperbolic, rate (16). As all proteins required for RNA replication and packaging are provided in *trans*, there is little selection pressure on progeny plus strands other than the ability to be encapsidated. For the next 20 to 25 min plus strand RNA and virions are produced at a flat, linear rate of about 1000 copies/cell/min until cell lysis occurs. About half of the plus strand RNA progeny are packaged so that the ~10,000 virions produced contain RNAs made almost exclusively during the linear replication phase. Thus, intracellular growth of Q-beta is largely stamping machine-like with kinetics far different from the exponential growth kinetics that predominate during successive rounds of phage release by cell lysis and infection of new bacterial cells (16).

A second example is intracellular replication of phage phi6 (6). Like positive strand RNA viruses, replication of dsRNA-containing phi6 is semiconservative, with positive strands being displaced by successive and independent copying of the negative strand template. Here the evidence of linear growth is not from measuring intracellular growth kinetics but arises from a genetic analysis approach first used by Luria (37). First, we define some terms. An RNA polymerase copying a template may introduce an erroneous nucleotide substitution, or mutation, into a progeny strand. The strand bearing the mutation, and any of its subsequent copies, are considered mutants. As mutations occur at random because of polymerase error (substitution bias of the polymerase notwithstanding), the expected distribution of mutations across infected cells (whether they burst or not) is a Poisson distribution. However, the expected distribution of mutants across infected cells may or may not be expected to be Poisson, depending on the mode of intracellular replication. The logic of this is as follows: If intracellular growth is exponential, cells in which a mutation arose early in the replication cycle will contain more mutant progeny than cells in which the mutation occurred later in the replication

cycle. This is because progeny strands also serve as templates in the exponential growth model. At one extreme are cells where a mutation occurred during the first round of duplication; half the virions in such cells would bear mutant genomes. A cell where the mutation occurred during the last duplication event would harbor only a single mutant progeny. Thus, the number of mutants per cell is expected to fluctuate widely from one cell to another and follow a Luria-Delbruck distribution (35, 38). On the other hand, if multiple progeny were produced by copying the same template, i.e., a linear, stamping machine model, mutations would arise at random independently in virus progeny, so that the distribution of mutants would be Poisson. The two scenarios can be readily distinguished because a Poisson distribution has a variance/mean ratio equal to 1 whereas the variance/mean ratio for a Luria-Delbruck distribution is much greater than 1. Note that if intracellular growth is a hybrid of exponential and linear modes of growth, the experimental variance/mean ratio of the number of mutants/cell will lie between the two distributions. For phi6, the experimentally observed variance/mean value was 1.37, i.e., nearly Poisson (6). By computer simulations, Chao et al. (6) determined that the excess variance, $1.37 - 1 = 0.37$, would arise if 1% of intracellular RNA progeny left progeny of their own. Thus, the mode of phi6 growth is, therefore, 99% linear.

Luria (37) found that intracellular replication of the dsDNA phage T2 was exponential. This is because the DNA of T2 reproduces by a binary process as each chromosome produces 2, 4, 8, etc., copies. This result leant credence to the notion that phage reproduction was essentially the same as that of bacteria and other cellular organisms. Since then, most biologists have been unaware of Luria's hypothetical counterexample of a stamping machine mode of growth.

The implications of exponential versus linear growth of viruses are several fold. Double stranded DNA viruses have a genome mutation rate (average number of mutations per genome produced, mu g) of about 0.002, whereas for RNA viruses mu g is about 1 (15b). Under the stamping machine growth model, the fraction of genomes with no mutations, $f(0)$, following one round of intracellular growth is the Poisson null class, i.e., $f(0) = \exp(-\text{mu g})$. For exponential growth $f(0)$ depends on the number of doublings: after one doubling $f(0) = \exp(-\text{mu g})$, after two doublings $f(0) = \exp(-2\ \text{mu g})$, after three doublings $f(0) = \exp(-3\ \text{mu g})$, etc., so $f(0) = \exp(-\text{n} \times \text{mu g})$ for n doublings. Consider a dsDNA virus and an RNA virus each yielding 1000 copies per infected cell. Under the stamping machine (linear) model this requires 999 copying steps. For binary replication, just under 10 doublings are required to produce 999 progeny because $2^{10} = 1024$. For the dsDNA virus $f(0)$ is $\exp(-0.002) = 0.998$ for linear growth, and $\exp(-10 \times 0.002) = 0.980$ for exponential growth. The mutational "savings" of linear versus exponential growth is about 1.8%. However, for the RNA virus with a high per genome mutation rate, the cost of exponential growth is evolutionarily unacceptable and leads to mutational meltdown. For the RNA virus $f(0)$ is $\exp(-1) = 0.37$ and $\exp(-10 \times 1) = 0.000045$ for linear and exponential growth, respectively. The mutational "savings" of linear growth in this case is more than 800,000%! Thus, there is much to be gained by RNA viruses replicating under a linear mode versus

an exponential mode. The trade-off in speed of replication (999 replication rounds versus 10 to produce 1000 progeny from a single progenitor) is minor compared to the potentially catastrophic loss of progeny genomes bearing zero mutations. Said another way, exponential growth only becomes a viable strategy when the per genome mutation rate is low. Double-stranded DNA viruses have solved this problem by utilizing proof-reading functions of DNA polymerases. As far as we know RNA viral polymerases do not proofread, such that genome repair is limited to RNA recombination, a mechanism that apparently has limits with respect to the rate of genome repair.

There are, of course, differences between the RNA phage examples above and plant viruses. Plant viruses are more prolific than other viruses, generating several orders of magnitude more intracellular virions. Cells infected with tobacco mosaic virus often contain hexagonal crystals comprised of 10^7 to 10^8 virions (19). If these were produced by purely exponential growth, essentially no genomes free of mutations would remain [$f(0) \approx 3 \times 10^{-12}$], suggesting intracellular growth must be largely linear to avoid mutational meltdown. Second, in infected plant cells virus replication eventually stops but cells do not lyse, such that the vast majority of progeny genomes are encapsidated and remain in the cell in which they were produced. These sequestered genomes are not in competition among themselves or among encapsidated genomes in other parts of the plant, and have yet to be tested for fitness (and never will be unless transmitted to another plant). With lytic viruses, progeny virions are released and all have the opportunity to initiate new rounds of infection until host cells are no longer available. Lytic viruses, therefore, are expected to have extracellular growth rates many times higher than plant viruses. Consequently, selection for rapid intracellular replication and release rates is especially relaxed for plant viruses and adoption of a slow, mainly linear replication mode seems to be an almost inevitable evolutionary outcome. In this light, genetic stability of plant viruses compared to lytic viruses is more apparent than real; plant viruses simply vary less than would be expected if their replication was mainly exponential.

The above discussion focused on RNA viruses. However, the distribution of mutants of the single-stranded DNA phage PhiX174 across infected cells also is Poisson (14). PhiX174 replicates by a rolling circle mechanism so it is likely that all viruses, including the plant-infecting geminiviruses (68), replicating this way also should be examples of stamping machine growth.

For WSMV there are other factors contributing to reproductive variance. Its mite vector only feeds on epidermal tissue so the probability that virions in other tissues are transmitted to a new host is zero. Moreover, after wheat plants mature in midsummer, an enormous quantity of virus (almost, but not quite all) is lost from the population. In Kansas, about 4×10^6 hectares of wheat are planted annually. If 5% of about 2.5×10^6 plants/hectare are infected by WSMV and each infected plant contains 10^{12} virus particles, the yearly loss from the WSMV gene pool is 5×10^{23} genomes in this state alone! This must have an effect on effective population size and corresponding diversity levels.

EVIDENCE FOR INTRAPLANT BOTTLENECKS IN THE LIFE HISTORY OF WSMV

Having hypothesized that *Ne* should be low for plant viruses, is there any experimental evidence to support this? To be sure, limited genotypic variation is seen within individual WSMV isolates (24) or among a collection of 54 isolates from temperate North America (69). Both observations are consistent with a low value of *Ne*; however, this interpretation must be in reference to an evolutionary model. What is required is an independent approach to estimate *Ne*. We examined the fate of two WSMV strains, Type and Sidney 81, in mixed infections (23). When inoculated sequentially onto wheat these two strains exhibit the well-known phenomenon of cross-protection, but when inoculated simultaneously they can coexist in the same plant. As wheat plants grow they develop multiple side shoots, or tillers. Developing tillers on plants that are systemically infected with virus represent new tissue to be invaded once vascular connections are made between the primary shoot and the emerging secondary shoot. Tillers of dually infected plants were assayed for the presence of Type and Sidney 81 by RT-PCR and RFLP analysis and the majority of these were found to contain both strains, but a number of tillers contained only one or the other strain. Thus, occasional spatial segregation of genotypes within an infected plant was observed (23). It is then possible to estimate the "ploidy" of virus colonization of new tillers by fitting observed Type and Sidney 81 frequencies to a binomial distribution.

Let the proportions of Sidney 81 and Type be s and t, respectively, with $(s + t) = 1$ across all observations. If the number of colonizing viral genomes were 2, we would have the familiar binomial expansion for diploids, namely:

$$(t + t)^2 = s^2 + 2st + t^2 = 1,$$

with s^2 being the proportion of tillers infected with Sidney 81 only and t^2 the proportion of tillers with Type only (the two homozygous frequencies) and $2st$ being the proportion of heterozygous (both Type and Sidney 81 present) tillers. This can be rearranged to give:

$$s^2 + t^2 = 1 - 2st,$$

such that there are three frequency classes, i.e., Sidney 81 only, Type only, and both Sidney 81 and Type.
For the general case, $(s + t)^n = 1$, we have

$$s^n + t^n = 1 - (\text{all terms containing } st^{n-1}, s^2t^{n-2}, \text{etc.}).$$

Thus, a Chi Square goodness-of-fit test to estimate the best value of n may be done using the same three frequency classes, with 1 degree of freedom (d.f. = 3–1 for 3 data classes, less 1 d.f. for determining proportions s and t from the same data). Of 108 tillers, 8 had only Sidney 81, 5 had only Type, and 95 had both

TABLE 1 Fit to a binomial distribution of tiller infection frequencies (data from Reference 24) for *Wheat streak mosaic virus* (WSMV) strains Sidney 81 and Type in mixed infections

Observed[a]	***s*-only**	***s* + *t***	***t*-only**	**Chi-Square**	***P*-value**[c]
Tiller data:	8	95	5		
Expected[b]:					
$(s + t)^2$	28	54	25	61.4	$P < 10^{-11}$
$(s + t)^3$	13.5	81	13.5	10.0	$P < 0.002$
$(s + t)^4$	6.2	94.6	7.3	1.2	$P = 0.273$
$(s + t)^5$	3	101.2	3.7	9.2	$P < 0.003$
$(s + t)^6$	1.5	104.6	1.9	34.1	$P < 10^{-8}$

[a]Observed *s*-only, only WSMV-Sidney 81 detected; $s + t$, both WSMV-Sidney 81 and WSMV-Type present; *t*-only, only WSMV-Type detected.

[b]Expected values for the three frequency classes for each given power of *n*; not all sum to 108 due to rounding.

[c]Probabilty of obtaining a larger value of Chi-Square if the given binomial expression were true (1 d.f.).

strains (23). The proportion *s* is $(8 + 95)/((8 + 95) + (5 + 95)) = 0.51$, whereas $t = 0.49$ indicating the two strains are equally able to systemically spread within infected plants. The fit of observed to expected values of binomial distributions for several powers of *n* is shown in Table 1. Clearly, the best fit is for $(s + t)^4$, with the fit for $(s + t)^3$ and $(s + t)^5$ well outside a 99% confidence interval. This suggests that, on average, only 4 WSMV genomes are competing successfully to establish an infection as a tiller develops. Following establishment, many other genomes enter these tillers but are prevented from colonizing the tissue, likely because of cross-protection. Consistent with this was a control experiment where plants were coinfected with WSMV-Sidney 81 and *Brome mosaic virus* (BMV). No cross-protection occurs between these two unrelated viruses. All 68 tillers examined contained both viruses (23), i.e., no segregation was seen. The best fit binomial distribution for data without any observed segregation is for *n* to approach infinity. Thus, WSMV-Sidney 81 and BMV had essentially unfettered opportunity to invade each tiller.

A similar RT-PCR and RFLP analysis also was done from 1-mm disks of tissue from dually infected leaves (23). Of 55 samples, 3 had only Sidney 81 and 16 had only Type. The proportion of Sidney 81 in the sample, *s*, was 0.43, and that of Type, *t*, was 0.57. In this case the best-fitting binomial distribution was for $n = 2.6$ with a 95% confidence interval between 2.2 and 3.1, suggesting that the distribution of virus genomes is patchy and that between two and three WSMV genomes compete to colonize "territories" as small as 1 mm in diameter during systemic infection. The surprising conclusion for both the leaf disk and tiller experiments is that the numbers of successful competing genomes, 2 to 3, and 4, respectively, are very low given that there are an estimated 10^5–10^6 virions in a single infected cell. The bottleneck during systemic movement is severe indeed.

The RT-PCR and RFLP method used to distinguish Type and Sidney 81 was fairly crude. One may argue that samples determined to be "Sidney 81 only" were, say, 90% Sidney 81 and 10% Type. To account for this, the three frequency classes used to calculate expected values can be adjusted. For example, one can define "Sidney 81 only" to be the range of 100% Sidney 81 to 90% Sidney 81 and 10% Type, "both" to be 11%–89% Sidney 81 and 11%–89% Type, and "Type only" to be 0%–10% Sidney 81 and 100%–90% Type. A conservative supposition is that a 20:80 mixture of the two isolates can be discerned and accurately scored as a mixture. We bin together the range (Type:Sidney 81) from 0:100 to 20:80 as Sidney 81 only, 100:0 to 80:20 as Type only, and mixed infections being in the range 21:79 to 79:21. Using the observed Sidney 81 (0.51) and Type (0.49) proportions from the tiller experiment, we can test how well larger values of n fit the observed data using the cumulative binomial distribution. For $n = 10$, the probability of observing from $0/10$ to $2/10$ Type:Sidney 81 is 0.062; for $n = 100$, the probability of observing from $0/100$ to $20/100$ Type:Sidney 81 is 1.8×10^{-9}; for $n = 1000$, the probability of observing from $0/1000$ to $200/1000$ Type:Sidney 81 is 1×10^{-80}. Thus, given a 20% lower detection limit for identifying single infections, $n = 10$ competing WSMV genomes is within a 90% (two-tailed) confidence interval, and $n = 12$ is just outside a 95% confidence interval. The likelihood that the potential number of genomes competing to invade a tiller is 100, much less 1000, is vanishingly low. Whether the systemic movement bottleneck is 4, 10, or 12 individual genomes, it is a surprisingly small number of founding genomes that are the progenitors of all progeny generated in a tiller. Note that as systemic movement within a plant may be defined as the equivalent of extracellular growth for a lytic virus, the number of plant virus genomes involved in this phase of exponential growth is miniscule compared to that of a lytic virus.

It is instructive to view these results in terms of predicted diversity under selective neutrality (28). Assuming a mutation rate of 2.3×10^{-5} per nucleotide (18) and the effective population size determined above in the range of 4 to 10, *pi* should be 1.8×10^{-4} to 4.6×10^{-4}. Within-isolate average value of *pi* for four genes of Type was 0.4×10^{-4}, whereas that of Sidney 81 was 3×10^{-4} (calculated from data in Reference 24). These values are remarkably close to the predicted range for genetic segregation within infected plants.

There is a subtle difference between taking several WSMV genomes at random and the intraplant bottlenecks described above. The number of WSMV genomes entering a tiller was not directly measured; rather the average number of winning genomes per tiller was determined. To be a winner means that other genomes were excluded through cross-protection. It is doubtful that infection with a debilitated virus could prevent subsequent invasion by a more fit genotype. The small number of genotypes that become successfully established, therefore, are not chosen from a pool of all genotypes present, but rather are chosen from a pool of more or less equally competitive genotypes. The bottleneck is random but it is not blind to selection.

EVOLUTIONARY TRENDS IN WSMV AND A CLOSE RELATIVE

Frequently, studies of virus diversity summarize the data as phylogenetic trees. Sequences are aligned and trees reconstructed by parsimony, maximum likelihood, or distance measures. However, the sequence alignment itself provides a rich data source for other types of population genetic and evolutionary inferences. One can think of each nucleotide position in an alignment as an evolutionary scorekeeper. Some positions may be inherently tolerant to variation and others may be less changeable. Within a sample of 54 consensus sequences derived from 53 WSMV isolates, polymorphic sites within the coat protein cistron and 3′ noncoding region had a complex distribution (69). Out of 378 codons and a total of 1267 nt sites, there were 272 variable nt sites, 150 of which occurred only once in the sample. There were almost 5 times more substitutions at the third codon position (183 of 378 codons) than at the first (34 of 378 codons) or second codon positions (38 of 378 codons), and just over 13% (17 of 130 nts) of sites in the noncoding region were polymorphic. Selection would tend to remove deleterious genotypes from a field population such that less fit lineages are infrequently sampled. While a few mutations may result in increased fitness, positive selection should result in fixation with a concomitant loss of diversity in the field population. That the bulk of variation among isolate consensus sequences are at synonymous sites suggests that much of the diversity within WSMV is neutral with respect to fitness (69).

At the same time positions tolerant to variation might not be variable in a sample due to chance alone. Mutations leading to polymorphisms present in higher frequency in the sample may predate mutations generating less frequent polymorphisms or may have attained higher frequencies through random drift. A given set of sampled individuals represents an unique outcome of the past. That is, the sample represents the results of a single, unreplicated, evolutionary experiment. Consider five progeny sampled from a population. If we could see into the past, all are connected to each other by a single, specific genealogy. The problem is that while that genealogy exists, we do not know what it is. Keeping in mind the random nature of evolutionary processes such as mutation and recombination, a single phylogenetic tree oversimplifies the situation by discounting how uncertain the history of our sample actually is. Imagine we have protein-coding sequence alignment from the five individuals. Say at position 6 three sequences have an A and two have a G. The assumption is that all with A are related to each other and to a most common recent ancestor (MCRA), and the same for the two sequences with G. Further, the MCRAs of the A and G sequences also have a single MCRA in the more distant past. Other sites will accumulate mutations at random along the genealogy and, if recombination has occurred, different sites may have different genealogies. Processes such as a recent selective sweep may cause one lineage to have fewer accumulated mutations than others. A widely used approach to incorporate such uncertainty, termed the coalescent (17, 27, 29), is to treat both mutations and genealogies as random variables. This allows one to statistically

test whether patterns of variation are inconsistent with a given model of evolution. Several computer programs incorporating coalescent-based statistical tests such as DNASP (53) and SITES (26) are available.

It is possible to calculate expected distributions of allele frequencies for strictly neutral populations (70), for models allowing population growth (60), selection (either positive or negative) (56), and pseudohitchhiking (22). Growth, negative selection, and pseudohitchhiking each may perturb the neutral allele size class spectrum, resulting in overrepresentation of rare sites. However, in each of these models the number of sites occurring at intermediate frequencies are concomitantly reduced. Given the observed data (69), estimates of best-fitting parameter values were obtained for the three models by maximum likelihood. For each model, the proportion of predicted singletons was much lower than the observed value (69). A similar skewed distribution was observed for WSMV variants within single isolates after serial passage in the laboratory (24). A possible explanation is that error-prone viral replication continually adds new mutations unique to individual plants. This would result in a hybrid distribution: a Poisson distribution resulting from mutation, and a second distribution resulting from a combination of genealogy and population history of the sample.

Until recently, the potyvirus *Oat necrotic mottle virus* (ONMV) was officially listed as a definitive member of the genus *Rymovirus*. However, this taxonomic placement was not due to hard evidence but was, more or less, a default position. The 3′-terminal 1729 nucleotides (nts) of two ONMV isolates were sequenced to clarify the relationship of ONMV among other potyviruses (48). Phylogenetic analysis provides unambiguous evidence that ONMV should be considered a member of the genus *Tritimovirus* and that it is closely related to WSMV. Specifically, ONMV shares 73%–74% (nt) and 79%–81% [amino acid (aa)] identity with nine WSMV isolates from the Czech Republic, Hungary, Iran, Mexico, Russia, Turkey, and the United States.

The coat protein (CP) cistron of ONMV was 12 codons shorter than the corresponding WSMV consensus sequence. However, sequence identities were sufficiently high such that the 11 sequences could be aligned with little ambiguity. The missing codons in ONMV were near the CP amino-terminus, a region that also is variable among WSMV isolates. Given the codon-based alignment, several nonuniform patterns of divergence between and within the two virus species became apparent. The two species differ by an average of 106 aa replacements out of a total of 527 codons. The vast majority (86%) of replacement codons have nt substitutions at multiple positions within each codon, and even 14 of 217 silent codons involved multiple nt substitutions. Further, only 54 aa replacements between ONMV and WSMV were at sites monomorphic in the WSMV data set. Thus, it seems that the rate of evolution varies widely among sites for these two viruses and that not all substitution events are independent of each other.

Polymorphism and diversity may be statistically analyzed in a number of ways. The McDonald-Kreitman test (39) examines the ratio of fixed differences and polymorphic sites partitioned into silent and replacement substitution categories.

TABLE 2 Variation within the coat protein cistron among *Wheat streak mosaic virus* and *Oat necrotic mottle virus* partitioned into between (fixed) or within (polymorphic) species categories that are further partitioned based upon whether the variation results from silent (synonymous) or replacement (nonsynonymous) nucleotide substitutions

	Fixed	**Polymorphic**
Synonymous	134	397
Nonsynonymous	83	62

For the WSMV-ONMV data the results presented in Table 2. The ratio of polymorphic to fixed silent sites is not the same as the same ratio for replacement sites ($P < 10^{-5}$). The relatively fewer replacement polymorphic sites suggest these are under negative selection. The Poisson random field model (PRF) provides a theoretical framework for estimating parameters proportional to mutation and selection rates from allele frequency data (56, 57). The 539 aligned codons of nine diverse WSMV isolates were analyzed for sequence differences occurring among one, two, three, or four isolates. The observed allele frequency distribution for silent substitutions (exclusive of multiple substitutions in the same codon) was class 1:130; class 2:25; class 3:39, and class 4:94. PRF maximum likelihood estimates for *mu*, the mutation parameter, was 41.4 ± 4.8 and for *S*, the selection parameter, was 1.00 ± 1.71. The latter parameter was not significantly different from $S = 0$ ($P = 0.103$) by the log ratio test. The estimate for *mu* (replacement) was 51.2 ± 10.5 and the *S* (replacement) estimate was -3.81 ± 1.48, which is significantly different from null hypothesis of $S = 0$ ($P < 10^{-8}$). Nevertheless, the selection rate against replacement substitutions is rather weak, and the ratio of synonymous and nonsynonymous substitution rates remain highly correlated (48).

Within species transitions (*ts*) are about fourfold more frequent than transversions (*tv*), while between WSMV and ONMV the *ts*/*tv* ratio is 1.1. A possible reason for this is that some sites have undergone sequential substitutions over time. Sites experiencing rarer transversion events may be less likely to revert so that transversions "catch up" with transitions as sequences continue to diverge. This indicates that one or both branches leading to their common ancestor may be longer than lengths based on the number of nt differences alone.

CONCLUSIONS

Much work remains to understand and explain population genetic dynamics of plant viruses. As more studies accumulate it should be possible to distinguish patterns of variability due to a particular virus's life history from those due to more

general features of virus biology. We have presented some, no doubt, controversial hypotheses as to why plant viruses may be less variable than their bacterial and animal counterparts and why RNA viruses in general may exhibit less variation than might be expected from their very large populations sizes and high mutation rates. Specifically, we propose that virus growth within plants is predominately linear, rather than exponential. As a consequence, the effective population size is very low, with systemic movement bottlenecks quite severe. In the case of WSMV, in which the primary host (wheat) is uniformly susceptible, it is perhaps not surprising that the bulk of variation observed among isolates in a field population, having survived the sieve of selection, is neutral with respect to fitness. Although selection remains an important force, we point out that plant virus genomes are actually quite plastic and that stochastic processes also significantly affect the evolution of plant viruses such that the outcome in any single virus lineage is unpredictable. It is our hope that this discussion will be thought-provoking and promote future experimental and theoretical studies adding new insights into plant virus population genetic and evolutionary biology.

The *Annual Review of Phytopathology* is online at http://phyto.annualreviews.org

LITERATURE CITED

1. Albiach-Marti MR, Mawassi M, Gowda S, Satyanarayana T, Hilf ME, et al. 2000. Sequences of citrus tristeza virus separated in time and space are essentially identical. *J. Virol.* 74:6856–65
2. Aleman-Verdaguer ME, Goudou-Urbino C, Dubern J, Beachy RN, Fauquet C. 1997. Analysis of the sequence diversity of the P1, HC, P3, NIb, and CP genomic regions of several yam mosaic potyvirus isolates: implications for the intraspecies molecular diversity of potyviruses. *J. Gen. Virol.* 78:1253–64
3. Brakke MK. 1958. Properties, assay, and purification of wheat streak mosaic virus. *Phytopathology* 48:439–45
4. Brakke MK. 1971. Wheat streak mosaic virus. *CMI/AAB Descriptions of Plant Viruses*, No. 48. Wellesbourne, UK: Assoc. Appl. Biol.
5. Brakke MK, Skopp RN, Lane LC. 1990. Degradation of wheat streak mosaic virus capsid protein during leaf senescence. *Phytopathology* 80:1401–5
6. Chao L, Rang CU, Wong LE. 2002. Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage phi 6. *J. Virol.* 76:3276–328
7. Chenault KD, Hunger RM, Sherwood JL. 1996. Comparison of the nucleotide sequence of the coat protein open reading frame of nine isolates of wheat streak mosaic Rymovirus. *Virus Genes* 13:187–88
8. Choi IR, Hall JS, Henry M, Zhang L, Hein GL, et al. 2001. Contributions of genetic drift and negative selection on the evolution of three strains of wheat streak mosaic tritimovirus. *Arch. Virol.* 146:619–28
9. Choi IR, Horken KM, Stenger DC, French R. 2002. Mapping of the P1 proteinase cleavage site in the polyprotein of *Wheat streak mosaic virus* (genus *Tritimovirus*). *J. Gen. Virol.* 83:443–50
10. Choi IR, Stenger DC, French R. 2000. Multiple interactions among proteins encoded by the mite-transmitted wheat streak mosaic tritimovirus. *Virology* 267: 185–98

11. Choi IR, Stenger DC, Morris TJ, French R. 2000. A plant virus vector for systemic expression of foreign genes in cereals. *Plant J.* 23:547–55
12. Christian ML, Willis WG. 1993. Survival of wheat streak mosaic virus in grass hosts Kansas from wheat harvest to fall emergence. *Plant Dis.* 77:239–42
13. de la Torre JC, Holland JJ. 1990. RNA virus quasispecies populations can suppress vastly superior mutant progeny. *J. Virol.* 64:6278–81
14. Denhardt D, Silver RB. 1966. An analysis of the clone size distribution of PhiX174 mutants and recombinants. *Virology* 30:10–19
15. Domingo E, Escarmis C, Menendez-Arias L, Holland JJ. 1999. Viral quasispecies and fitness variations. See Ref. 15a, pp. 141–62

15a. Domingo E, Webster R, Holland JJ, eds. 1999. *Origin and Evolution of Viruses.* San Diego: Academic

15b. Drake J, Holland J. 1999. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* 96:13910–13

16. Eigen M, Biebricher CK, Gebinoga M, Gardiner WC. 1991. The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry* 30:11005–18
17. Fu YX, Li WH. 1999. Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theor. Popul. Biol.* 56:1–10
18. García-Arenal F, Fraile A, Malpica JM. 2001. Variability and genetic structure of plant virus populations. *Annu. Rev. Phytopathol.* 39:157–86
19. Gibbs A, Harrison B. 1976. *Plant Virology, The Principles.* London: Edward Arnold
20. Gibbs AJ, Keese PL, Gibbs MJ, García-Arenal F. 1999. Plant virus evolution: past, present and future. See Ref. 15a, pp. 263–85
21. Giesman-Cookmeyer D, Lommel SA. 1993. Alanine scanning mutagenesis of a plant virus movement protein identifies three functional domains. *Plant Cell* 5:973–82
22. Gillespie JH. 2000. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* 155:909–19
23. Hall JS, French R, Hein GL, Morris TJ, Stenger DC. 2001. Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. *Virology* 282:230–36
24. Hall JS, French R, Morris TJ, Stenger DC. 2001. Structure and temporal dynamics of populations within wheat streak mosaic virus isolates. *J. Virol.* 75:10231–43
25. Hartl DL, Clark AG. 1998. *Principles of Population Genetics*. Sunderland: Sinauer. 3rd ed.
26. Hey J, Wakeley J. 1997. A coalescent estimator of the population recombination rate. *Genetics* 145:833–46
27. Hudson RR. 1991. Gene genealogies and the coalescent process. In *Oxford Surveys in Evolutionary Biology*, ed. D Futuyama, J Antonovics, 7:1–44. Oxford: Oxford Univ. Press
28. Kimura M. 1968. Evolutionary rate at the molecular level. *Nature* 217:624–26
29. Kingman JC. 1982. On the genealogy of large populations. *J. Appl. Prob.* 19:27–34
30. Kong P, Rubio L, Polek M, Falk BW. 2000. Population structure and genetic diversity within California *Citrus tristeza virus* (CTV) isolates. *Virus Genes* 21:139–45
31. Kroner PA, Young BM, Ahlquist P. 1990. Analysis of the role of brome mosaic virus 1a protein domains in RNA replication, using linker insertion mutagenesis. *J. Virol.* 64:6110–20
32. Kurath G, Heick JA. Dodds JA. 1993. RNase protection analyses show high genetic diversity among field isolates of satellite tobacco mosaic virus. *Virology* 194:414–18
33. Kurath G, Palukaitis P. 1990. Serial

passage of infectious transcripts of a cucumber mosaic virus satellite RNA clone results in sequence heterogeneity. *Virology* 176:8–15
34. Kurath G, Rey CME, Dodds JA. 1992. Analysis of genetic heterogeneity within the type strain of satellite tobacco mosaic virus reveals several variants and a strong bias for G to A substitution mutations. *Virology* 189:233–44
35. Lea DE, Coulson CA. 1949. The distribution of the numbers of mutants in bacterial populations. *Genetics* 49:264–85
36. Li WH. 1997. *Molecular Evolution*. Sunderland: Sinauer
37. Luria SE. 1951. The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. *Cold Spring Harbor Symp. Quant. Biol.* 16:463–70
38. Luria SE, Delbruch M. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511
39. McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–54
40. McKinney HH. 1937. Mosaic diseases of wheat and related cereals. *US Dep. Agric. Circ.* 442:1–23
41. McKinney HH. 1956. Interference and synergy—their possible use in identifying certain mosaic viruses of cereals and indicating degrees of relationship. *Plant Dis. Report.* 40:898–903
42. McNeil JE, French R, Hein GL, Baenziger PS, Eskridge KM. 1996. Characterization of genetic variability among natural populations of wheat streak mosaic virus. *Phytopathology* 86:1222–27
43. Montana JR, Hunger RM, Sherwood JL. 1996. Serological characterization of wheat streak mosaic virus isolates. *Plant Dis.* 80:1239–44
44. Moya A, Elena SF, Bracho A, Miralles R, Barrio E. 2000. The evolution of RNA viruses: a population genetics view. *Proc. Natl. Acad. Sci. USA* 97:6967–73
45. Naraghi-Arani P, Daubert S, Rowhani A. 2001. Quasispecies nature of the genome of *Grapevine fanleaf virus*. *J. Gen. Virol.* 82:1791–95
46. Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press
47. Ohta T. 1973. Slightly deleterious substitutions in evolution. *Nature* 246:96–98
48. Rabenstein F, Seifers DL, Schubert J, French R, Stenger DC. 2002. Phylogenetic relationships, strain diversity, and biogeography of tritimoviruses. *J. Gen. Virol.* 83:895–906
49. Rodríquez-Cerezo E, Elena SF, Moya A, García-Arenal F. 1991. High genetic stability in natural populations of the plant RNA virus tobacco mild green mosaic virus. *J. Mol. Evol.* 32:328–32
50. Rodríquez-Cerezo E. García-Arenal F. 1989. Genetic heterogeneity of the RNA genome population of the plant virus U5-TMV. *Virology* 170:418–23
51. Roossinck MJ, Zhang L, Hellwald KH. 1999. Rearrangements in the 5′ nontranslated region and phylogenetic analyses of cucumber mosaic virus RNA 3 indicate radial evolution of three subgroups. *J. Virol.* 73:6752–58
52. Rosenberg NA, Nordborg M. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3:380–90
53. Rozas J, Rozas R. 1999. DNASP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–75
54. Rubio L, Abou-Jawdah Y, Lin HX, Falk BW. 2001. Geographically distinct isolates of the crinivirus *Cucurbit yellow stunting disorder virus* show very low genetic diversity in the coat protein gene. *J. Gen. Virol.* 82:929–33
55. Sanchez-Sanchez H, Henry M, Cardenas-Soriano E. Alviso-Villasana HF. 2001. Identification of *Wheat streak mosaic virus* and its vector *Aceria tosichella* in Mexico. *Plant Dis.* 85:13–7
56. Sawyer SA. 1997. Estimating selection

and mutation rates from a random field model for polymorphic sites. In *Progress in Population Genetics and Human Evolution; IMA Volumes in Mathematics and its Applications*, ed. P Donnelly, S Tavaré pp. 193–206. New York: Springer-Verlag
57. Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics* 132:1161–76
58. Schneider WL, Roossinck MJ. 2000. Evolutionarily related Sindbis-like plant viruses maintain different levels of population diversity in a common host. *J. Virol.* 74:3130–34
59. Schneider WL, Roossinck MJ. 2001. Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. *J. Virol.* 75:6566–71
60. Shen PD, Wang F, Underhill PA, Franco C, Yang WH, et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* 97:7354–59
61. Skotnicki ML, Ding SW, Mackenzie AM, Mo JQ, Gibbs AJ. 1993. RNA hybrid mismatch polymorphisms in Australian populations of turnip yellow mosaic tymovirus. *Arch. Virol.* 132:83–99
62. Skotnicki ML, Mackenzie AM, Gibbs AJ. 1992. Turnip yellow mosaic virus variants produced from DNA clones encoding their genomes. *Arch. Virol.* 127:25–35
63. Skotnicki ML, Mackenzie AM, Gibbs AJ. 1996. Genetic variation in populations of *Kennedya yellow mosaic tymovirus*. *Arch. Virol.* 141:99–110
64. Slykhuis JT. 1955. *Aceria tulipae* Keifer (Acarina Eriophyidae) in relation to the spread of wheat streak mosaic. *Phytopathology* 45:116–28
65. Slykhuis JT, Bell W. 1963. New evidence on the distribution of wheat streak mosaic virus and the relation of isolates from Rumania, Jordan, and Canada. *Phytopathology* 53:236–37
66. Stavalone L, Alioto D, Ragozzino A, Laliberte' JF 1998. Variability among turnip mosaic potyvirus isolates. *Phytopathology* 88:1200–4
67. Stenger DC, Hall JS, Choi IR, French R. 1998. Phylogenetic relationships within the family Potyviridae: *Wheat streak mosaic virus* and *Brome streak mosaic virus* are not members of the genus *Rymovirus*. *Phytopathology* 88:782–87
68. Stenger DC, Revington GN, Stevenson MC, Bisaro DM. 1991. Replicational release of geminivirus genomes from tandemly repeated copies: evidence for rolling circle replication of a plant viral DNA. *Proc. Natl. Acad. Sci. USA* 88: 8029–33
69. Stenger DC, Seifers DL, French R. 2002. Patterns of polymorphism in *Wheat streak mosaic virus*: sequence space explored by a clade of closely related viral genotypes rivals that between the most divergent strains. *Virology* 302:58–70
70. Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:597–601
71. van Vloten-Doting L, Bol JF. 1988. Variability, mutant selection, and mutant stability in plant RNA viruses. In *RNA Genetics*, ed. E Domingo, JJ Holland, P Ahlquist, 3:37–45. Boca Raton, FL: CRC Press
72. Wiese MV. 1977. Wheat streak mosaic. In *The Compendium of Wheat Diseases*, pp. 75–76. St. Paul: Am. Phytopathol. Soc.