Population structure within lineages of *Wheat streak mosaic virus* derived from a common founding event exhibits stochastic variation inconsistent with the deterministic quasi-species model

Roy French, Drake C. Stenger *

Agricultural Research Service, United States Department of Agriculture and Department of Plant Pathology, University of Nebraska, Lincoln, NE 68583, USA

Abstract

Structure of *Wheat streak mosaic virus* (WSMV) populations derived from a common founding event and subjected to serial passage at high multiplicity of infection (MOI) was evaluated. The founding population was generated by limiting dilution inoculation. Lineages of known pedigree were sampled at passage 9 (two populations) and at passage 15, with (three populations) or without mixing (four populations) of lineages at passage 10. Polymorphism within each population was assessed by sequencing $17-21$ clones containing a 1371 nt region (WSMV-Sidney 81 nts $8001-9371$) encompassing the entire coat protein cistron and flanking regions. Mutation frequency averaged $\sim 5.0 \times 10^{-4}$/nt across all populations and ranged from 2.4 to 11.6×10^{-4}/nt within populations, but did not consistently increase or decrease with the number of passages removed from the founding population. Shared substitutions (19 nonsynonymous, 10 synonymous, and 3 noncoding) occurred at 32 sites among 44 haplotypes. Only four substitutions became fixed (frequency = 100%) within a population and nearly one third (10/32) never achieved a frequency of 10% or greater in any sampled population. Shared substitutions were randomly distributed with respect to genome position, with transitions outnumbering transversions 5.4:1 and a clear bias for A to G and U to C substitutions. Haplotype composition of each population was unique with complexity of each population varying unpredictably, in that the number and frequency of haplotypes within a lineage were not correlated with number of passages removed from the founding population or whether the population was derived from a single or mixed lineage. The simplest explanation is that plant virus lineages, even those propagated at high MOI, are subject to frequent, narrow genetic bottlenecks during systemic movement that result in low effective population size and stochastic changes in population structure upon serial passage. Published by Elsevier Inc.

Keywords: Potyvirus; Tritimovirus; Population genetics; Genetic drift; Genetic bottlenecks

Introduction

RNA viruses replicate utilizing an RNA dependent RNA polymerase (RdRp) with no known proofreading functions (Steinhauser et al., 1992). As a consequence, error due to misincorporation of bases during replication is much higher for RNA viruses (Drake, 1993; Drake and Holland, 1999) than that of double-stranded DNA viruses for which polymerase proofreading and subsequent mismatch repair reduce the mutation rate by several log-units. Thus, RNA virus populations typically are composed of a collection of sequence variants. Often, this variation within an RNA virus population has been described as synonymous with a 'quasi-species'. However, the two terms,

'variation' and 'quasi-species', are not necessarily interchangeable. Although a quasi-species population certainly has variation, not all populations that vary are quasi-species.

The quasi-species model (Domingo, 2002; Domingo et al., 1985, 2001; Eigen, 1996) differs from other models of population genetics in several key attributes. The quasi-species model requires that the population in question have a high mutation rate (which seems to be the case for all RNA viruses), and be extremely large with all members of the population being in direct competition with one another. As a consequence, quasi-species population structure is entirely driven by selection. Being deterministic, the quasi-species model does not allow for stochastic changes in population structure, such as those due to genetic drift. Furthermore, the unit of selection is the group rather than the individual, such that for a quasi-species population it is the frequency distribution of variants

* Corresponding author. Fax: +1 402 472 4020.
  E-mail address: dstenger@unlnotes.unl.edu (D.C. Stenger).

that is selected for rather than individual genotypes. A commonly cited advantage is that a quasi-species population may adapt quickly to changes in selective conditions by having a very large pool of pre-existing variants from which to draw (Domingo and Holland, 1994, 1997).

The quasi-species model does describe certain attributes of bacterial or animal RNA viruses grown in tissue culture (Domingo et al., 1978; Novella, 2003). However, whether RNA viruses behave as quasi-species under natural conditions has been questioned (Choi et al., 2001; Hall et al., 2001b; Holmes and Moya, 2002; Jenkins et al., 2001; Moya et al., 2004). The conditions under which RNA viral populations develop within a plant (Hall et al., 2001a, 2001b; French and Stenger, 2003) are very different from the experimental systems upon which the quasi-species model are based. Plant tissues are highly organized and cells are separated from one another by cell walls. Thus, plant virus genomes (encapsidated or otherwise) are not simply released into the extracellular environment and free to attach to and enter any cell of the host population. Instead, plant virus genomes may only move among adjacent cells sharing plasmodesmata connections or long distance via phloem (for reviews see Ghoshroy et al., 1997; Carrington, 1999; Waigmann et al., 2004). These spatial restrictions on movement of virus genomes between cells are potential bottlenecks that plant viruses encounter within an infected plant.

Recently, three independent studies (French and Stenger, 2003; Sacristán et al., 2003; Li and Roossinck, 2004) estimated the severity of genetic bottlenecks imposed upon plant viruses during systemic movement. Although the three virus systems studied (*Wheat streak mosaic virus* [WSMV], *Tobacco mosaic virus* [TMV]), and *Cucumber mosaic virus* [CMV]) are unrelated and infect dicotyledonous (TMV and CMV) or monocotyledonous (WSMV) hosts, the estimates of movement bottleneck severity were remarkably similar: the number of plant virus genomes giving rise to all progeny in a plant organ may be

measured in units! The implication of these somewhat startling results is that although plant viruses have extremely large census populations, the effective population size (e.g., the proportion of genomes that serve as replication templates giving rise to progeny) is very low. If indeed plant virus populations have high mutation rates but low effective population sizes, then genetic drift is possible and stochastic variation should result such that the quasi-species model is not applicable. In this report, we examine the population structure of WSMV lineages derived from a common founding event to assess whether temporal changes in population structure were deterministic or stochastic.

Results

Generation of populations

The founding population S10 was established previously by limiting dilution inoculation of the WSMV-Sidney 81 isolate (Hall et al., 2001b). The pedigree of populations derived from S10 after multiple passages at high multiplicity of infection (MOI) is presented in Fig. 1. Populations WB9 and BA9 have been described previously (Hall et al., 2001b) and represent two separate lineages generated by nine high MOI serial passages removed from S10. The WB9 and BA9 populations served as inocula to generate four additional lineages (two from each passage 9 inoculum) that were passaged at high MOI an additional six times to generate four populations (WB15a, WB15b, BA15a, and BA15b) removed from S10 by 15 passages. WB9 and BA9 inocula were also mixed in equal proportions to establish three mixed lineages that were serially passaged at high MOI six times to generate three populations (Mix15A, Mix15B, and Mix15C) that also were removed from S10 by 15 passages. These 10 populations share a single common ancestor that established the S10 population upon the limiting dilution inoculation founding event, such that diversity
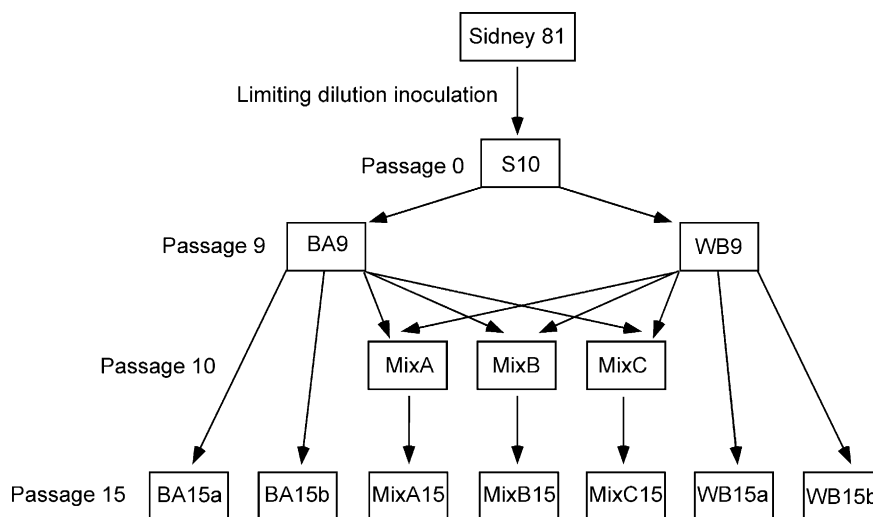

Fig. 1. Pedigree of passaged populations derived from *Wheat streak mosaic virus* isolate Sidney 81. The initial population (S10) was established by limiting dilution inoculation and passaged at high multiplicity of infection (MOI) nine times in wheat (WB9) or barley (BA9). Each passage 9 population was divided into two lineages and passaged an additional six times in wheat to produce four passage 15 populations (BA15a, BA15b, WB15a, and WB15b). Equal amounts of inocula from WB9 and BA9 were mixed to establish three mixed lineages (MixA, MixB, and MixC) at passage 10 that were passaged an additional five times in wheat at high MOI to generate three mixed populations at passage 15 (MixA15, MixB15, and MixC15).

of the population at that point in time was extremely low, if not completely eliminated. Since the initial founding event, each of these 10 populations has distinct histories due to separation and/or mixing of lineages in space and time.

Polymorphic sites

Each of the 10 populations was sampled for genetic diversity by nucleotide (nt) sequencing of a 1371 nt region corresponding to Sidney 81 nts 8001–9371 (GenBank Accession AF057533; Stenger et al., 1998). This genomic region contains the entire coat protein (CP) cistron and flanking regions, including the 3′-proximal portion of the upstream NIb cistron and most of the 3′-untranslated region (3′-UTR) positioned downstream of the CP cistron. Templates used for sequencing were cloned from reverse transcription-polymerase chain reaction (RT-PCR) products. A total of 196 RT-PCR clones (17–21 per population) were sequenced such that the entire data set was comprised of 268,716 nts.

Mutations resident in the 10 populations are summarized in Table 1. The entire data set contained 192 different mutations present at 189 polymorphic sites. Of these mutations, 160 were singletons, in which a specific mutation appeared only once in a single clone throughout the entire data set. The 160 singletons (154 substitutions, 4 single-base deletions, and 1 two-base deletion) were distributed among 158 polymorphic sites, as two polymorphic sites each had two different singleton substitution mutations. For simplicity, each position of the two-base deletion was treated as an independent mutation. Only shared mutations (e.g., those in which the identical substitution appeared at the same site in more than one clone) were present at the remaining 31 polymorphic sites. Only one polymorphic site (nt 8248) yielded both singleton and shared mutations, such that a total of 32 distinct shared mutations were present at 32 polymorphic sites in the entire data set (Fig. 2). As the frequency of shared mutations at any single site varied from 2/196 to 69/196, the total number of nts classified as shared

mutations was 280, corresponding to ~0.1% of the 268,716 nts sequenced. Thus, total number of nts classified as polymorphic was 440 (160 singletons plus 280 shared).

Estimation of mutation frequency/nt attributed to RT-PCR error

Mutations observed in the data set could have occurred in vivo (RdRp error) or in vitro (RT-PCR error). The former class of mutations is of interest, the latter class is artifactual. Although it is not possible to specify in vivo or in vitro origin for a given mutation present in a cloned sequence, the data may be used to calculate the mutation frequency/nt due to RdRp, if a correction factor (e. g., the mutation frequency/nt due to RT-PCR) is known. The overall error rate due to RT-PCR was estimated to be 2.6×10^{-4}/nt, based on control experiments using the same RT-PCR conditions (Hall et al., 2001b). Therefore, in vitro error due to RT-PCR accounted for only 70 of 440 polymorphic nts (268,716 total nts sequenced multiplied by the RT-PCR error rate of 2.6×10^{-4}/nt) present in the entire data set. As most RT-PCR induced errors would be expected to occur as singletons (Smith et al., 1997), most shared substitutions in the data set may be considered as having arisen in vivo.

Estimation of expected homoplasy

Shared mutations can be either identical by common ancestry or identical by independent recurrent substitutions at the same site (homoplasy). The potential for the latter depends on both the number of samples sequenced and the amount of sequence diversity. Thus, the probability of recurrent substitution grows as the observed number of polymorphic sites in the sample increases. The expected amount of homoplasy due to sampling cannot be calculated, but it can be assessed by Monte Carlo simulations (Maynard Smith and Smith, 1998). The particular simulation algorithm employed is conservative, in that only two substitution states (i.e., transitions) are allowed

Table 1
Nucleotide sequence polymorphism present in populations derived from WSMV-Sidney 81 limiting dilution subisolate S10 after nine passages (WB9, BA9), 15 passages (WB15a, WB15b, BA15a, BA15b), or six passages after mixing WB9 and BA9 inocula (MixA15, MixB15, and MixC15)

Population	No. of clones sequenced	No. nt[a]	Observed no. of mutations			Mutation frequency/nt	
			Total	Singletons	Shared[b]	Uncorrected total	Corrected total[c]
S10	17	23,307	16	15	1	6.9×10^{-4}	4.2×10^{-4}
WB9	20	27,420	23	16	7	8.4×10^{-4}	5.8×10^{-4}
BA9	21	28,791	27	21	6	9.4×10^{-4}	6.8×10^{-4}
MixA15	20	27,420	17	10	7	6.2×10^{-4}	3.6×10^{-4}
MixB15	20	27,420	28	18	10	10.2×10^{-4}	7.6×10^{-4}
MixC15	20	27,420	16	11	5	5.8×10^{-4}	3.2×10^{-4}
WB15a	19	26,049	19	14	5	7.3×10^{-4}	4.7×10^{-4}
WB15b	20	27,420	25	17	8	9.1×10^{-4}	6.2×10^{-4}
BA15a	19	26,049	13	10	3	5.0×10^{-4}	2.4×10^{-4}
BA15b	20	27,420	39	28	11	14.2×10^{-4}	11.6×10^{-4}
Entire data set	196	268,716	192[d]	160	32[d]	7.1×10^{-4}	4.5×10^{-4} (5×10^{-4})[e]

[a] For each clone sequenced, 1371 nt positions were determined.
[b] Number of sites having a mutation that occurred more than once in the entire data set.
[c] All changes attributed to RT-PCR error (2.6×10^{-4}/nt) were subtracted from uncorrected total mutations to yield corrected total mutation frequencies/nt.
[d] Each site exhibiting shared polymorphism counted once in total regardless of frequency in the entire data set.
[e] Corrected for expected homoplasy.

Fig. 2. Shared substitutions and haplotypes present in 10 *Wheat streak mosaic virus* (WSMV) populations of known pedigree. Haplotypes designated by letters (A to qq) present in each population are shown in which nucleotide (nt) substitutions relative to WSMV-Sidney 81 are denoted in red. Shared nucleotide substitutions denoted in lower case occurred as a singleton in at least one population; shared substitutions present only as singletons within two populations are denoted as lower case and boxed. Underline denotes unique substitution occurring at site 8248 also bearing a shared, but different substitution. WSMV-Sidney 81 nt coordinates for each polymorphic site bearing a shared substitution are indicated at the top. The frequency of each haplotype within a population is indicated at the right.

per site. This simplification is reasonable given the actual data (see below) where transitions outnumbered transversions. A series of 5000 simulation runs for 5 to 250 polymorphic sites sampled from a 1371 nt sequence was performed. As mentioned above, there were 189 polymorphic sites in the total data set. Sampling 189 mutated sites requires a median of 203 "draws", meaning that about 14 sites with homoplasy are

expected across all populations (Fig. 3). An obvious sign of recurrent mutation is when more than one base substitution occurs at the same polymorphic site. There were three such instances in the WSMV data set: nts 8081, 8248, and 9128. At nts 8366, 8495, 8583, 8635, 8774, 8899, 9186, and 9243, the same singleton substitutions occurred in each of two populations, and nts 8403 and 8932 were singleton substitutions
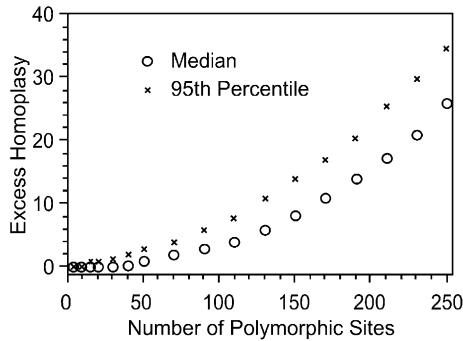
Fig. 3. Graphical depiction of expected number of recurrent mutations in generating samples with a given number of total polymorphic sites. Results presented are from 5000 independent Monte Carlo simulations by sequentially assigning mutations to positions at random on a sequence 1371 nt in length until the total number of polymorphic site reached the values given on the x axis. Median values of excess mutations required to generate samples containing the specified number of polymorphic sites are shown on the y axis. Mean values were very close to the medians, as were the modal values. The 95 percentile confidence values are also presented.

in one of the populations (Fig. 2), making them plausible examples of homoplasy as well. During 5000 simulated sampling runs of 189 polymorphic sites, 2287 runs (46%) had 13 or fewer excess mutations, suggesting that most of the expected instances of homoplasy in the data were credibly identified.

Within each of the 10 populations, however, there were far fewer polymorphic sites, ranging from 16 to 39 (Table 1), with each requiring a median of zero extra draws. In fact, homoplasy does not become apparent until 50 or more polymorphic sites are sampled (Fig. 3) for sequences of this length. A reasonable scenario, given the expected amount of homoplasy within and between populations, is that shared sites within populations are identical by descent, whereas singleton substitutions within a population that were shared between populations most likely are identical by state.

Viral mutation rate

Table 1 lists mutation frequency/nt calculated for each population and the entire data set, with the RT-PCR mutation frequency/nt subtracted from the uncorrected total to yield total corrected mutation frequency/nt. The corrected mutation frequency/nt for the entire data set was 4.5×10^{-4}/nt, a value similar to those determined for a more limited WSMV data set (Hall et al., 2001b) or observed with other plant viruses (Schneider and Roossinck, 2000, 2001). Note that if excess homoplasy (e.g., shared substitutions identical by state rather than descent) is accounted for, the viral mutation rate becomes

approximately 10% greater ($\sim 5.0 \times 10^{-4}$/nt). Although the corrected mutation frequency/nt ranged between 2.4 and 11.6×10^{-4}/nt among the 10 populations (without accounting for excess homoplasy, which is minimal within each population), there was no obvious trend, upwards or downwards, in the magnitude of the corrected mutation frequency/nt with respect to passage or lineage, such that populations sampled at passages 9 or 15 (with or without mixing) were not any more likely to have a mutation frequency greater than or less than that of the founding S10 population (Table 1).

Analysis of shared substitution mutations

The specific substitutions present at each polymorphic site bearing shared mutations are presented in Fig. 2. Among shared substitutions, transitions (27/32) vastly outnumbered transversions (5/32) to yield a transition/transversion (ts/tv) ratio of 5.4. This ts/tv ratio is similar to that (4.8) calculated for a data set of closely related consensus sequences of WSMV isolates resident in temperate North America (Stenger et al., 2002). Nearly half of the transitions appearing among the shared substitutions were A to G despite the fact that the A:G ratio is only 1.1 (Table 2). This same bias for A to G transitions was observed for CMV (Schneider and Roossinck, 2000) but not with the *Satellite tobacco mosaic virus*, for which the most common transition was G to A (Kurath et al., 1992). It is interesting to note that the second most common transition observed in passaged populations of WSMV was U to C (Table 2), such that this complementary set of substitutions (A to G and U to C) accounted for over three-fourths of all transitions. This observation suggests that bias for one of these transitions may be greater than indicated in the data as, for example, A to G substitutions occurring during minus-strand synthesis would have been scored in Table 2 as U to C transitions in plus-strand.

The genomic distribution of shared mutations identified within the entire data set is depicted in Fig. 4. The distribution of substitutions as a function of genome position was evaluated using the nonparametric one-sample runs test (Zar, 1974). The genomic position of all shared mutations was random ($P > 0.24$), as was the genomic distribution of shared nonsynonymous substitutions ($P > 0.24$). There were too few synonymous or noncoding shared substitutions to evaluate genomic distribution by the one-sample runs test. Shared mutations were predominately nonsynonymous substitutions (19/32), with the remaining shared mutations resulting in synonymous substitutions (10/32) or noncoding substitutions (3/32). These results differ from those among consensus sequences of ~50 WSMV field isolates, in which synonymous substitutions were by far more frequent and the genomic

Table 2

Transition and transversion mutations[a] among shared substitutions present within WSMV populations derived from a single common founding event

Transition mutations (27 total)				Transversion mutations (5 total)							
A to G	G to A	C to U	U to C	A to C	C to A	A to U	U to A	C to G	G to C	G to U	U to G
13	3	3	8	3	0	2	0	0	0	0	0

[a] Empirical base frequencies for the entire data set were A: 0.301, C: 0.212, G: 0.273, and T: 0.214.
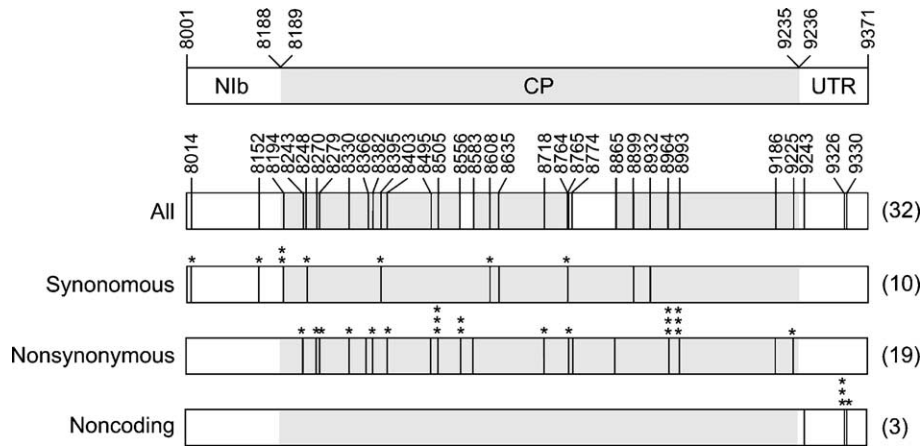
Fig. 4. Genomic distribution of shared substitutions present in 10 populations derived from *Wheat streak mosaic virus* (WSMV) isolate Sidney 81. At the top is a partial genome map of WSMV-Sidney 81 indicating the region sequenced and analyzed for polymorphism. Below are partial genome maps that denote the location and nucleotide coordinates of all shared substitutions, or only synonymous, nonsynonymous, or noncoding substitutions. The number in parentheses at right indicates the total number of each type of substitution. Asterisks refer to the maximum frequency of a substitution within any single population: 100% = three asterisks; ≥50%, <100% = two asterisks; ≥10%, <50% = one asterisk; <10% = no asterisk.

distribution of nonsynonymous substitutions was nonrandom (Stenger et al., 2002). These differences in relative frequency and genomic distribution of nonsynonymous substitutions among the two WSMV data sets may be related to the observation that all substitutions scored among isolates of WSMV in the previous study represented mutations that had achieved a frequency of >50% (e.g., they constituted the consensus sequence), whereas in the passage study described here, only 6/32 shared substitutions occurred within a population at a frequency of >50% and only 4/32 became fixed (frequency = 100%) within a population (Fig. 4). At the same time, it may be that some of the nucleotide substitutions observed in this study are deleterious and, therefore, never become established in field populations.

Analysis of haplotypes within and between populations

The 32 shared substitutions were distributed among 44 haplotypes (Fig. 2). Composition (occurrence) and complexity (number and relative frequencies) of haplotypes present in S10 and each passaged population are presented in Fig. 5. The founding S10 population contained two haplotypes, designated A and A′. Haplotype A was present in 16/17 clones recovered from the S10 population and is identical in sequence to GenBank Accession AF057533 and that of an infectious clone of the Sidney 81 isolate (Choi et al., 1999). Haplotype A′ occurred once in the S10 sample and contained a single A to G substitution at nt 8899 relative to haplotype A. In Hall et al. (2001b), haplotype A′ was not differentiated from the singleton population but in this analysis, this same uncommon shared substitution appeared once in haplotype qq present in population BA15b (Figs. 2, 5).

While many pairs of substitutions were linked, haplotypes representing all possible four allelic combinations were found for several substitution pairs (Fig. 2). The relaxed linkage disequilibrium between these latter pairs may represent recombination events during virus growth. Applying the four

gamete test of Hudson and Kaplan (1985), at least one recombination event each occurred in the histories of populations WB9 (between nts 8403 and 8764) and BA9 (between nts 8014 and 8718), and at least two events occurred in the history of MixB (between nts 8270 and 8505, and between nts 8765 and 9326). The other populations did not have haplotypes with site configurations exhibiting all four allelic types.

Haplotype composition varied among the passaged populations (Fig. 5). Most of the 44 haplotypes were observed only within a single population, with relatively few haplotypes appearing in more than one passage of the same lineage. Furthermore, even for those haplotypes that did appear in more than one passage (haplotypes A, A′, B, F, L, and aa), there was no obvious selection for a particular haplotype. For example, although haplotype B (which varies from haplotype A by a single C to U transition at nt 8556) represented the majority in population WB9, in lineages derived from WB9, haplotype B increased in frequency in one population (WB15a) but was not recovered from the second (WB15b). Furthermore, in the three lineages in which WB9 served as inoculum in a mixture with BA9, haplotype B was detected in only one of three mixed lineages at passage 15. Other haplotypes appearing in more than one population shared similar fates (Fig. 5).

The populations also varied in haplotype complexity. Sampling of populations S10, WB15a, BA15a, and MixC15 yielded three or fewer haplotypes, with each population containing a single predominant haplotype constituting ~89–95% of the sample. However, a different haplotype dominated each of these four populations. In contrast, other populations (BA9, BA15b, WB15b, and MixB15) exhibited increased complexity in that more haplotypes were sampled (6–10 per population), but none of these haplotypes achieved majority status in a population. The remaining two populations (WB9 and MixA15) each had a single haplotype accounting for 60% of the sampled population but also had a significant fraction of the population (40%) composed of 5–6 minority haplotypes. Collectively, the haplotype data indicate that population
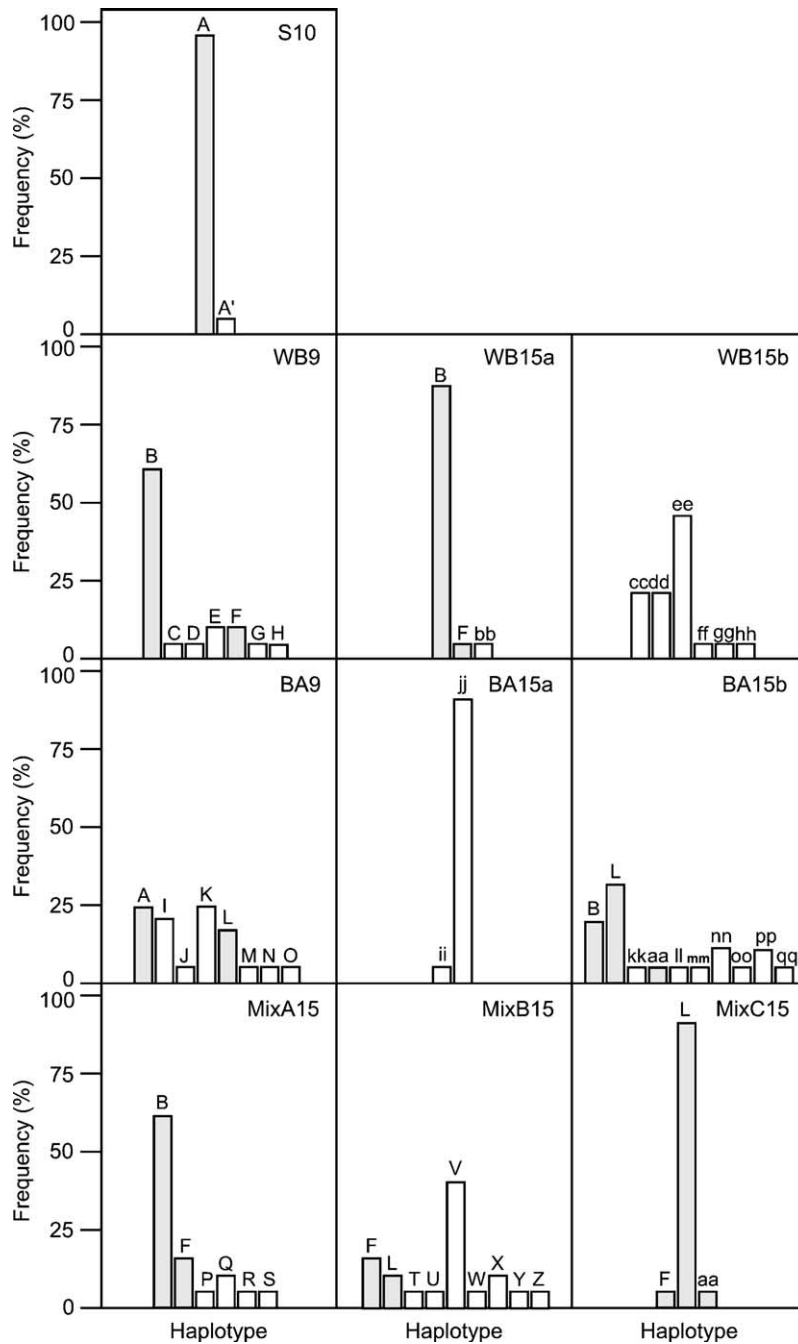
Fig. 5. Histograms of haplotype composition and complexity for 10 *Wheat streak mosaic virus* (WSMV) populations of known pedigree and derived from WSMV-Sidney 81. Letters denote specific haplotypes present in each population. Frequency of haplotypes within a population is denoted by height of the bars. Shaded bars represent haplotypes that occurred in more than one population.

structure varied unpredictably in both space and time, and that attempts to increase the complexity of a population through mixed inocula were not any more or less likely to produce a lineage of increased complexity relative to an unmixed lineage.

F statistics (Wright, 1951) were calculated to quantify between or among population divergence using haplotype frequencies (Gst; Nei, 1987) or pairwise differences between sequences (Fst; Hudson et al., 1992). These statistics estimate the proportion of total diversity due to between population differences. Gst between WB9 and BA9 was 0.1445 and Fst was 0.337. By passage 15, Gst and Fst had increased to 0.4692

and 0.7069, respectively, among BA15b, BA15b, WB15a WB15b, to 0.3298 and 0.4445 among MixA15, MixB15, and MixC15, and to 0.4370 and 0.5955 among all passage 15 populations. Thus, populations from different lineages became more divergent from each other over time.

Analysis of change in haplotype frequencies after passage

There were five haplotypes each in populations WB9 and BA9 which were, respectively, the source populations for WB15a and WB15b, or BA15a and B15b. Both WB9 and BA9

Table 3
Estimation of effective population size (N_e) from analysis of variance in *Wheat streak mosaic virus* haplotype frequencies over time

Population	Fc[a]	N_e (95% CI)[b]
MixA15	0.1935	50 (18, 155)
MixB15	0.1814	56 (20, 189)
MixC15	0.3273	24 (16, 123)
WB15a	0.1036	163 (49, ∞)
WB15b	0.6667	11 (5, 20)
BA15a	0.5714	13 (5, 26)
BA15b	0.5467	13 (6, 27)
Arithmetic mean		47
Harmonic mean		21

[a] Wright's *F* statistic calculated using a modification of Eq. (8) in Waples (1989) to account for haploid data.
[b] Per Eqs. (11) and (16) in Waples (1989), modified for haploid data. CI = confidence interval.

were source populations for MixA15, MixB15, and MixC15. Thus, it is possible to estimate effective population size (N_e) from observed temporal changes in haplotype frequencies. For this analysis, passage 15 haplotypes were grouped by inferred passage 9 ancestries. A variation of Wright's F-(Fc) statistic was calculated for MixA15, MixB15, MixC15, WB15a, WB15b, BA15a, and B15b (Table 3). From this, estimates of N_e per passage, and their 95% confidence intervals (CI) were obtained (Table 3). Most populations were consistent, with surprisingly low values for N_e. The 95% CI upper bound for these estimates was less than 200 for all except WB15a. Maximum likelihood analysis of haplotype data, pooled across lineages, was done using the MCLEEPS computer program estimated N_e to be 14 (95% CI = 9, 24). The average point estimate of N_e across lineages was 47 (Table 3), but a better way to pool several independent measures of N_e (Lande and Barrowclough, 1987; Luikart et al., 1999) is to use the harmonic mean, which was 21 (Table 3). This latter mean also was more consistent with the maximum likelihood result. It should be noted, however, that because each passage consisted of inoculation to (and subsequent inoculum pooled from) five plants, N_e for virus populations resident within an individual host plant must be less than that calculated here for groups of five host plants.

Discussion

Populations derived from a common founding event vary stochastically

Following serial passage, all populations derived from the S10 founding population differed from one another with respect to haplotype composition and complexity. This result is consistent with previous passage experiments with WSMV (Hall et al., 2001b), in which changes in consensus sequences were tracked in multiple lineages derived from a common source. In the present series of experiments, we examined the derived populations in greater detail such that information regarding the complexity and composition of nonconsensus variation also was measured. Given that the derived populations were subjected to identical selection forces (with the possible

exception of BA9 in passages 1–9 in which the host was barley rather than wheat), a deterministic outcome should have resulted in the same distribution of haplotypes in each lineage. This clearly was not the case with respect to nonsingleton variation, in which the probability of mutants having (i) served as replication templates, (ii) engaged in competition with other haplotypes, and (iii) survived the sieve of selection was extremely likely. Instead, haplotype composition and complexity of the derived populations were unpredictable, in that there was no correlation between population structure and the number of passages removed from the founding population or between population structure and the complexity of inocula, which was artificially increased to generate the mixed lineages. Collectively, these results are consistent with repeated and severe bottlenecks imposed upon population structure during the course of the experiment, such that stochastic changes in population structure (e.g., genetic drift) occurred.

That MOI of the inoculum determines that the severity of transmission bottlenecks is intuitive, and has been experimentally demonstrated with WSMV by mechanical inoculation at low MOI (Hall et al., 2001b) or by single mite vector transmission assays (Hall et al., 2001a). In both cases, severe transmission bottlenecks resulted in rapid changes in consensus sequence which were apparent after a single passage. Because all serial passage inoculations in this experiment used high MOI inocula, these experiments were conducted under conditions in which transmission bottlenecks were as large as the experimental system permits. Nonetheless, evidence of bottlenecks was apparent in haplotype composition and complexity of the derived populations. This was substantiated by analysis of haplotype frequency changes over time with N_e averaging between 21 and 47 per passage. This suggests that at high MOI, bottlenecks during systemic movement in planta were more severe than those occurring at transmission. However, it should be noted that natural transmission bottlenecks, especially under conditions of low vector populations, also are likely to be severe (Hall et al., 2001a), such that bottlenecks during propagation of a viral lineage in the field may be even more severe than those measured in the passage experiments described here. The combined effects of transmission and movement bottlenecks, in conjunction with genetic isolation of lineages imposed by cross-protection (Hall et al., 2001a) should facilitate divergence among closely related lineages in relatively brief periods of time. This is exactly the pattern observed for consensus sequences of isolates recovered from the U.S. population of WSMV, which appears to be derived from a single founding event approximately 100 years ago (Stenger et al., 2002).

It is instructive to compare results presented here for plant to plant passage of WSMV to those for serial passage of vesicular stomatitis virus (VSV) in tissue culture (Novella et al., 2004). Consensus sequences were determined for 10 lineages after 80 passages of 2×10^5 plaque forming units of VSV. A total of 107 point mutations were found at 77/11,161 sites. Of these, 26 sites underwent parallel mutation in two or more lineages (Novella et al., 2004). Thus, the rate of mutations rising to prominence in populations of VSV in liquid culture is substantially lower than that of WSMV in plants, and there

are signs of convergent evolution among VSV lineages but none were found among WSMV lineages. While other factors are no doubt involved, large differences in N_e can explain in part why the outcome with VSV tended to be deterministic while the outcome with WSMV was essentially random. Clearly, all RNA viruses do not follow the same trajectories of mutation, selection, and drift, as has been sometimes either assumed or implied.

New mutations are constantly added to and removed from the population

The bulk of viral RNA replication within a plant is a linear process in which many progeny are made one at a time by serial transcription of a template (French and Stenger, 2003 and references therein). Therefore, mutations introduced into progeny molecules are largely independent events, and will be incorporated into progeny molecules at a rate determined solely by viral RdRp fidelity. As the majority of progeny molecules do not serve as replication templates but are instead sequestered as virions, nearly all of these mutations will be rare as individuals, but collectively common as a class if the error rate is sufficiently high. Most of these new mutations will appear in sampled populations as singletons, due to the large number of possible single substitution events (for WSMV 28,152 [3 × 9384 nts] different single substitution mutants are possible; 37,536 if deletions are included). Given an RdRp error rate of 5×10^{-4}/nt (~5 mutations per genome copy), it is likely that a substantial fraction of all possible substitutions would be generated de novo in a single cell in which $\sim 10^5$ progeny were produced. However, since the systemic movement bottleneck is measured in units (French and Stenger, 2003), it is impossible for most mutants generated in a single cell to move to the next cell. Thus, while new mutations are constantly being generated, movement bottlenecks are constantly eliminating nearly all of these mutations from the effective population. Therefore, the frequency of singleton mutations observed in a population should remain constant in time so long as the RdRp error rate remains constant. This appears to be the case for WSMV, in which most (8/10) populations yielded a corrected mutation frequency/nt within one standard deviation of the mean for all passaged populations, and all but one population had a corrected mutation frequency/nt within a 95% confidence interval of the mean (±1.96 standard deviations). These results suggest that the RdRp error rate remained constant among the different lineages sampled at different times (passages). However, this does not mean that RdRp error rates are not subject to change under certain circumstances, as appears to be the case for CMV passaged in different hosts (Schneider and Roossinck, 2001), suggesting that host factors contributing to the viral replicase complex may influence polymerase fidelity.

If most mutants do not serve as replication templates and are constantly being removed from the effective population, what is the advantage of producing such mutants? We know for certain that some mutations do in fact serve as templates and survive both movement and transmission bottlenecks, other-wise viruses would not evolve. There is an additional benefit to constantly producing and encapsidating new mutations: given enough time, each mutation eventually will be tested for fitness upon transmission. Therefore, having a high error rate allows for constant regeneration of variation, such that even with severe movement and transmission bottlenecks, a plant virus population contains a reservoir of alternative genotypes that may be more fit than the consensus sequence as selective conditions change. However, this is not group selection, as each mutant genome is tested essentially individually when transmitted under low MOI circumstances, which is the most likely scenario in nature.

Structure and dynamics of plant virus populations are inconsistent with the quasi-species model

Although the mutation rate of WSMV appears similar to that of other RNA viruses (e.g., it is high), and satisfies one criterion of a quasi-species, the derived populations of WSMV examined here varied stochastically and have the signature of severe bottlenecks and correspondingly low N_e. Thus, WSMV (and most likely all plant virus) populations lack essential features required of a quasi-species. As has been suggested earlier (Holmes and Moya, 2002; Moya et al., 2000, 2004), such populations are more appropriately studied using classical population genetics theory. In other words, we propose (i) that WSMV population structure is dependent upon the combined action of both selection and stochastic processes, (ii) that the individual genome is the unit of selection, and (iii) that WSMV populations are shaped by the same evolutionary processes that apply to cellular organisms. Consequently, there is no need to invoke a special model (e.g., quasi-species) to describe population attributes of WSMV or, perhaps, all viruses replicating in plants.

Materials and methods

Population establishment and passages

The founding population S10 (established from WSMV-Sidney 81 by limiting dilution inoculation) and two populations (BA9 and WB9) derived from S10 and sampled at passage 9 have been described previously (Hall et al., 2001a). In this study, the BA9 and WB9 populations were each separated into two lineages and passaged an additional six times in wheat, with the resulting populations (BA15a, BA15b, WB15a, and WB15b) sampled at passage 15. Three mixed populations also were established by combining equal amounts of WB9 and BA9 inocula. After six passages in wheat, the mixed populations (Mix15A, Mix15B, and Mix15C) were sampled. A pedigree of the 10 populations sampled is presented in Fig. 1. Each passage consisted of five plants grown in a common pot. All populations were passaged in wheat cultivar 'Arapahoe', with the exception of BA9, which was generated by nine serial passages in barley, as described previously (Hall et al., 2001b). Systemically infected leaf tissue from the five plants of a common pot was harvested 21 days

post-inoculation (dpi) and pooled for each lineage at each passage and used as inocula (diluted 1/10 w/v) to establish the next passage. For each population sampled, systemically infected leaf tissue (pooled from the five plants grown in a common pot) was harvested 21 dpi and stored at −70 °C. To minimize environmental differences, all passages were conducted in a growth chamber under controlled conditions (20 °C, 16 h daily photoperiod).

Cloning and sequencing

Total nucleic acids were extracted (McNeil et al., 1996) at passage 0 (S10), at passage 9 (WB9 and BA9), and at passage 15 from unmixed (WB15a, WB15b, BA15a, BA15b) or mixed (Mix15A, Mix15B, and Mix15C) populations. RT-PCR was conducted as described previously (Hall et al., 2001a) using the Expand High Fidelity System (Roche) to reduce the frequency of in vitro substitutions incorporated during amplification. PCR amplification using the primer set CGATTTTTTTTTTT-TTTTGCCGTCGCCCT (downstream primer) and AGCTTAT-GAGCGCTATTATTGCAGCAT (upstream primer) resulted in a product of 1429 bp that was cloned into pGEM-T (Promega). For each population, a total of 17–21 clones were completely sequenced on both strands using a combination of universal and custom primers. The sequences of each clone were assembled and trimmed to remove primer annealing regions using the program Sequencher 4.1 (Gene Codes, Ann Arbor), such that the sequence used in subsequent analyses consisted of 1371 nts, corresponding to Sidney 81 nts 8001–9371. All sequences for all clones of S10 and WB9, and 13 clones of BA9 were determined by Hall et al. (2001b); sequences of all other clones were determined in this study.

Analysis of population structure

Sequences of 196 clones derived from the 10 populations sampled were aligned using the program CLUSTAL X (Thompson et al., 1997). The entire data set was comprised of 268,716 nts. Polymorphic sites were classified as singletons if a substitution (or single base deletion) occurred at a nucleotide position once in the entire data set. Polymorphic sites at which a specific nucleotide substitution occurred in 2 or more clones in the entire data set were classified as shared substitutions. The sequence of each clone was assigned to a haplotype based on the occurrence of shared substitutions. Singletons were not considered in haplotype designation as this class of polymorphism includes most errors generated by RT-PCR (Smith et al., 1997) and is uninformative with respect to relationships within and between populations. For each population sampled, structure was assessed by scoring for haplotype composition (which haplotypes were present), and haplotype complexity (the number and frequency of each haplotype).

Estimates of mutation frequency (substitutions/nt) for each population sampled and for the entire data set was partitioned into that produced in vitro by RT-PCR, and that produced in vivo by viral polymerase. Corrected total mutation frequency was calculated by subtracting the mutation frequency/nt attributed to RT-PCR error (2.6×10^{-4}/nt), as determined previously (Hall et al., 2001b) for the specific RT-PCR conditions employed in this study, from the total uncorrected mutation frequency/nt. Shared substitutions at the same polymorphic site within populations were considered to be due to a single mutation event, such that each shared substitution was counted only once in the calculation of mutation frequency/nt. Nucleotide compositions were determined with the DAMBE program (Xia and Xie, 2001). Minimum number of recombination events (Hudson and Kaplan, 1985) were determined using DnaSP 4.0 (Rozas et al., 2003), which also was used to calculate population F statistics. The expected amount of homoplasy (Maynard Smith and Smith, 1998) was determined by Monte Carlo random sampling of mutations on a sequence length of 1371 until 189 polymorphic sites were generated. Results were reported as the median of 5000 independent simulations.

For each polymorphic site, shared substitutions were plotted as a function of genome position, and scored as nonsynonymous or synonymous, depending upon whether the mutation results in a change in the deduced protein product, or as noncoding if present within the 3′-UTR. The frequency of each shared substitution within a population was assessed independent of haplotype frequency and classified into frequency categories of uncommon (<10%), minority (≥10%, <50%), consensus (≥50%, <100%), and fixed (100%). Distribution of shared mutations as a function of genome position was assessed for deviations from randomness using the nonparametric one-way runs test (Zar, 1974). Point estimates of effective population size, N_e, were obtained through analysis of temporal changes in haplotype frequency using Eqs. (8), (11), and (16) in Waples (1989), except these were adjusted to reflect that sampled individuals are haploid rather than diploid. In addition, a maximum likelihood method to estimate N_e was done using MCLEEPS version 1.1 (Anderson et al., 2000).

References

Anderson, E.C., Williamson, E.G., Thompson, E.A., 2000. Monte Carlo evaluation of the likelihood for Ne from temporally-spaced samples. Genetics 156, 2109–2118.

Carrington, J.C., 1999. Reinventing plant virus movement. Trends Microbiol. 7, 312–313.

Choi, I.-R., French, R., Hein, G.L., Stenger, D.C., 1999. Fully biologically active in vitro transcripts of the eriophyid mite-transmitted wheat streak mosaic tritimovirus. Phytopathology 89, 11182–11185.

Choi, I.-R., Hall, J.S., Henry, M., Zhang, L., Hein, G.L., French, R., Stenger, D.C., 2001. Contributions of genetic drift and negative selection on the

evolution of three strains of wheat streak mosaic tritimovirus. Arch. Virol. 146, 619–628.

Domingo, E., 2002. Quasispecies theory in virology. J. Virol. 76, 463–465.

Domingo, E., Holland, J.J., 1994. Mutation rates and rapid evolution of RNA viruses. In: Morse, S.S. (Ed.), The Evolutionary Biology of Viruses. Raven Press, New York, NY, pp. 161–184.

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

Domingo, E., Sabo, D., Taniguchi, T., Weissman, C., 1978. Nucleotide sequence heterogeneity of an RNA phage population. Cell 13, 735–744.

Domingo, E., Martínez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortín, J., López-Galindez, C., Pérez-Bréna, P., Villanueva, N., Nájera, R., VandePol, S., Steinhauser, D., DePolo, N., Holland, J., 1985. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance—A review. Gene 40, 1–8.

Domingo, E., Biebricher, C.K., Eigen, M., Holland, J.J., 2001. Quasispecies and RNA Virus Evolution: Principles and Consequences. Landes Bioscience, Georgetown, TX.

Drake, J.W., 1993. Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. U. S. A. 90, 4171–4175.

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

Eigen, M., 1996. On the nature of virus quasispecies. Trends Microbiol. 4, 216–218.

French, R., Stenger, D.C., 2003. Evolution of wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. Annu. Rev. Phytopathol. 41, 199–214.

Ghoshroy, S., Lartey, R., Sheng, J., Citovsky, V., 1997. Transport of proteins and nucleic acids through plasmodesmata. Annu. Rev. Plant Physiol. Plant Mol. Biol. 48, 27–50.

Hall, J.S., French, R., Hein, G.L., Morris, T.J., Stenger, D.C., 2001a. Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. Virology 282, 230–236.

Hall, J.S., French, R., Morris, T.J., Stenger, D.C., 2001b. Structure and temporal dynamics of populations within wheat streak mosaic virus isolates. J. Virol. 75, 10231–10243.

Holmes, E.C., Moya, A., 2002. Is the quasispecies concept relevant to RNA viruses? J. Virol. 76, 460–462.

Hudson, R., Kaplan, N., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147–164.

Hudson, R.R., Slatkin, M., Maddison, W.P., 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589.

Jenkins, G.M., Worobey, M., Woelk, C.H., Holmes, E.C., 2001. Evidence for the non-quasispecies evolution of RNA viruses. Mol. Biol. Evol. 18, 987–994.

Kurath, G., Rey, C.M.E., Dodds, J.A., 1992. Analysis of 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, 233–244.

Lande, R., Barrowclough, G.F., 1987. Effective population size, genetic variation, and their use in population management. In: Soulé, M.E. (Ed.), Viable Populations for Conservation. Cambridge Univ. Press, Cambridge, UK, pp. 87–123.

Li, H., Roossinck, M.J., 2004. Genetic bottlenecks reduce population variation in an experimental RNA virus population. J. Virol., 10582–10587.

Luikart, G., Cornuet, J.M., Allendorf, F.W., 1999. Temporal changes in allele frequencies provide estimates of population bottleneck size. Conserv. Biol. 13, 523–530.

Maynard Smith, J., Smith, N.H., 1998. Detecting recombination from gene trees. Mol. Biol. Evol. 15, 590–599.

McNeil, J.E., French, R., Hein, G.L., Baenziger, P.S., Eskridge, K.M., 1996. Characterization of genetic variability among natural populations of wheat streak mosaic virus. Phytopathology 86, 1222–1227.

Moya, A., Elena, S.F., Bracho, A., Miralles, R., Barr, E., 2000. The evolution of RNA viruses: a population genetics view. Proc. Natl. Acad. Sci. U. S. A. 97, 6967–6973.

Moya, A., Holmes, E.C., González-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev., Microbiol. 2, 279–288.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia Univ. Press, New York.

Novella, I.S., 2003. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. Curr. Opin. Microbiol. 6, 399–405.

Novella, I.S., Zárate, S., Metzgar, D., Ebendick-Corpus, B.E., 2004. Positive selection of synonymous mutations in vesicular stomatitis virus. J. Mol. Biol. 342, 1415–1421.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Sacristán, S., Malpica, J.M., Fraile, A., García-Arenal, F., 2003. Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants. J. Virol. 77, 9906–9911.

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

Schneider, W.L., Roossinck, M.J., 2001. Genetic diversity in RNA virus quasispecies is controlled by virus–host interactions. J. Virol. 75, 6566–6571.

Smith, D.B., McAllister, J., Casino, C., Simmonds, P., 1997. Virus 'quasispecies': making a mountain out of a molehill? J. Gen. Virol. 78, 1511–1519.

Steninhauser, D.A., Domingo, E., Holland, J.J., 1992. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene 122, 288–2811.

Stenger, D.C., Hall, J.S., Choi, I.-R., 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, 782–787.

Stenger, D.C., Seifers, D.L., 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.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24, 4876–4882.

Waigmann, E., Ueki, S., Trutnyeva, K., Citovsky, V., 2004. The ins and outs of nondestructive cell-to-cell and systemic movement of plant viruses. Crit. Rev. Plant Sci. 23, 195–250.

Waples, R., 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. Genetics 121, 379–391.

Wright, S., 1951. The genetical structure of populations. Ann. Eugenics 15, 323–354.

Xia, X., Xie, Z., 2001. DAMBE: data analysis in molecular biology and evolution. J. Hered. 92, 371–373.

Zar, J.H., 1974. Biostatistical Analysis. Prentice Hall Englewood Cliffs, New Jersey, pp. 307–308.