Arch Virol (1999) 144: 1513–1526

Archives of Virology

Genome evolution of tobacco mosaic virus populations during long-term passaging in a diverse range of hosts

C. M. Kearney, M. J. Thomson*, and **K. E. Roland**

Department of Biology, Baylor University, Waco, Texas, U.S.A.

Accepted March 29, 1999

Summary. The effects of host changes on plant virus genome evolution was studied by nucleotide sequencing. A single tobacco (*Nicotiana tabacum* cv. Xanthi) plant was inoculated with in vitro transcripts from a plasmid clone of tobacco mosaic tobamovirus (TMV). This initial viral population was then transferred 11–12 times in parallel populations in 7 plant host species (1–4 replicates each) over a period of 413–515 days. Virion RNA was then isolated, reverse transcribed, amplified, cloned in bacteria, and sequenced. Portions of the coat protein, movement protein, and replicase genes were sequenced. Fourteen unique mutations were detected from a total of 188 clones (35,607 bases) sequenced, indicating a relatively small overall mutation rate of 3.1×10^{-4} nucleotide substitutions/base-year. A small Ka/Ks value of 0.09 was also found, indicating selection against amino acid changes. Eighty-five percent of the substitutions were transitions. A G'_{ST} value of 0.7 for the coat protein gene suggested that host type affected sequence changes in this region of the genome, but χ^2 analysis did not support this conclusion. This is the first study using sequencing to compare representative sample sections of a plant viral genome following a major selective disturbance such as extended passaging in an alternate host.

Introduction

Tobacco mosaic virus (TMV), a 6395 base (+) RNA virus, is the type member of the genus Tobamovirus. Different studies on tobamovirus populations have noted either RNA sequence stability or rapid evolution, depending on the evolutionary conditions placed on the virus and methods of detection of mutants and haplotypes. Rodríguez-Cerezo et al. [31] examined epiphytotics of pepper

*Present address: Department of Plant Breeding, Cornell University, Ithaca, New York, U.S.A.

mild mottle tobamovirus in commercial greenhouses by RNase fingerprinting over a 4-year period and noted genetic stability in one sequence haplotype that dominated over the entire period of the study. An examination of 53 isolates of tobacco mild green mosaic tobamovirus (TMGMV) from different parts of the world were noted for their relatively small sequence diversity values [7]. A similar study [20] also found wild TMGMV isolates which had a small populational diversity. In a greenhouse experiment [32], a field population of TMGMV from *Nicotiana glauca* was serially passaged through *Nicotiana tabacum* cv. Samsun and found to have maintained the original sequence (as determined by RNase T1 fingerprinting of 12% of the genome) even after 23 passages. However, other greenhouse experiments have detected that tobamoviruses may evolve rapidly. TMV populations directly derived from local lesions have been shown to harbor a great variety of variants, some of which are thought have arisen from mutation [8]. Studies on a necrotic lesion mutant of TMV found that certain physiological changes in the host induced the rapid domination of this mutant in the TMV population [15, 16]. The success of this mutant was also found to be affected by host species in a similarly rapid manner in another study [1].

These experimental studies indicate that TMV is capable of rapid evolution if sufficient selective pressure were provided, a trait of viruses of both plants [26, 35] and animals [5, 24]. We sought to determine the effect of an experimental host shift on the nucleotide sequences of portions of the TMV genome after passaging in the new host. The founder population was from a single *Nicotiana tabacum* cv. Xanthi plant inoculated with in vitro transcripts from a TMV genome cloned as DNA in bacteria [6]. We chose a range of plant hosts for extended passaging. We selected three regions to sequence, comprising 9.2% of the TMV genome, and used a common amplification-cloning-sequencing method to assay viral sequence microheterogeneity [28, 29, 34]. We found that the host shift resulted in evolutionary rates less than those seen in field studies and, furthermore, that the frequencies of transitions and silent mutations indicated stabilizing selection during the host passages.

Materials and methods

Virus and host plants

An initial population of TMV was made from in vitro transcripts produced from *Escherichia coli* RNA polymerase and *Pst*-I-linearized plasmid pTMV204 [6], which contains the full length wild type (U1) sequence of TMV driven by the lambda phage pPM1 promoter. The in vitro transcribed RNA was inoculated onto a single *N. tabacum* cv. Xanthi (tobacco) plant by rubbing the leaves with the RNA in buffer with abrasive as previously described [6]. After the virus had replicated in this single plant for four weeks, the resulting population was inoculated (as described below) onto 4 replicates of each of the 7 host species shown in Table 1.

Passaging and virion harvest

The viral populations were transferred to virus-free plants of the same species roughly once per month for a total of 11–12 passages (Table 1). Each transmission began by collecting leaf

Table 1. Host plants in which the TMV populations were grown

Host species	Family	Host response	Number of passages	Total days passaged
Nicotiana tabacum cv. Xanthi (tobacco)	Solanaceae	Systemic mosaic	12	440
Solanum nigrum (nightshade)	Solanaceae	Systemic mosaic or no symptoms	12	440
Collinsia heterophylla	Scrophulariaceae	Systemic mosaic	12	440
Phacelia campanularia	Hydrophyllaceae	Systemic mosaic	12	482
Fagopyrum esculentum (buckwheat)	Polygonaceae	Localized to inoculated leaf; no symptoms	11	514
Tagetes erecta (marigold)	Asteraceae	Localized to inoculated leaf, no symptoms	12	515
Plantago sp. (meadow plantain)	Plantaginaceae	Localized necrotic or chlorotic ringspots	12	413

pieces from all portions of infected plants, including both inoculated and systemically infected leaves and both symptomatic and nonsymptomatic leaves. Roughly 200 mg of leaf pieces were ground using an autoclaved mortar and pestle into 2 ml of glycine-phosphate buffer (50 mM glycine, 30 mM potassium phosphate, pH 9.2) containing 1% diatomaceous earth (Celite 503) as an abrasive. Approximately 1 ml of this slurry was then manually inoculated onto 1–3 healthy plants in one pot. Our designation of a "replicate viral population" thus consists of 1–3 infected plants in one pot and all plants were sampled when making the inoculum for the next passage. The viral titer was monitored for every passage by inoculating a portion of each of the passaging inocula onto *N. tabacum* cv. Xanthi nc, a necrotic local lesion host. After 11–12 passages, all parts of the final group of plants were harvested for virion extraction. Virions were prepared by chloroform/butanol extraction and polyethylene glycol (PEG) precipitation [11, 14]. Virions were then stored frozen at −20 °C until viral RNA extraction.

Virion RNA extraction and RT-PCR

Virions were suspended in 10 mM Tris (pH 8.0), 2 mM EDTA, 2% sodium dodecyl sulfate, incubated at 37 °C for 15 min. and then extracted twice with phenol/chloroform (1:1 v/v) and once with chloroform. The RNA was precipitated by adding NaCl to 0.2 M and 2.5 volumes of 100% ethanol. The pellet was resuspended in 15 ul of DEPC-treated water and stored at −80 °C. Alternatively, RNA was isolated from virions with the RNeasy minikit from Qiagen, Inc. To produce cDNA, 300–500 ng of viral RNA was heated for 5 min at 65 °C and then cooled on ice, followed by addition of MMLV reverse transcriptase (Promega), commercial reaction mixtures, and primer T-43, which is complementary to TMV-U1 sequence 6 312–6 292 (numbering after Goelet et al. [9]), which is 85 bases from the 3′ terminus of the TMV genome. Reverse transcriptase was destroyed by incubating at 75 °C for 5 min, and 1/5 of the cDNA reaction was added to the polymerase chain reaction, which was catalyzed by Vent DNA polymerase (New England Biolabs). PCR was done in a Stratagene Robocycler and consisted of an initial denaturation at 94 °C for 2–5 min, followed by 30 cycles of 94 °C/45 sec,

55 °C/30 sec, 72–74 °C/80 sec, followed by a single cycle of 3–10 min at 72–74 °C. Two separate PCR products were produced in two different reactions from each replicate (Fig. 1). One product comprised 1.4 kb of the TMV replicase gene. This PCR used primers M-1 (TMV nts. 1 295–1 316) and M-2 (nts. 2 678–2 657). The other product comprised 1.3 kb of the movement protein and coat protein genes and was primed with T-43 (the RT primer) and T-51 (nts. 5 010–5 036). Alternatively, T-12 (nts. 5 456–5 475) and WD-16 (nts. 6 245–6 225) were used to prime the coat/movement gene PCR. All PCR products were examined by gel electrophoresis in 1% agarose. The PCR DNA routinely gave a single, strong band of the expected length (data not shown).

Cloning and plasmid DNA sequencing

PCR products were purified and cloned, and the clones were then sequenced to reveal populational hetereogeneity of the TMV cultures. PCR products were purified by two chloroform extractions followed by ethanol precipitation with ammonium acetate and stored at −20 °C. The products were then digested with appropriate restriction enzymes and ligated into a pBluescript SK- cloning vector (Stratagene), and *E. coli* (JM109) competent cells [12] (frozen storage protocol 2) were transformed with the ligation mix. The replicase PCR product was digested with *Hin*d III and *Sac*II, which are native TMV sites interior to the priming sites. The coat/movement PCR product was digested with *Sal* I and *Bam* HI, the former being incorporated into the primer T-51 sequence and the latter into T-43. Alternatively, coat/movement PCR products produced by primers T-12 and WD-16 were ligated into the PCRscript vector (Stratagene), without using the optional *Pfu* DNA polymerase polishing step (Stratagene protocols), and XL1-Blue MRF' Kan cells were transformed (Stratagene). Transformants were selected on LB/lactose plates containing 300 nM 2,3,5-triphenyl-2H-tetrazolium chloride and 100 mg/liter ampicillin; colonies containing inserts were chosen by color reaction. Plasmid minipreparations utilized modified alkaline lysis procedures, either from reference [3] followed by a Sepharose CL-6B (Pharmacia) spun column [21], or using a PerfectPrep kit (5′–3′, Inc.). Plasmid DNA was sequenced using the dideoxynucleotide chain termination method, essentially following Sequenase protocols (USB/Amersham). Alternatively, cycle sequencing was performed using Thermosequenase (USB/Amersham). A 200 base region in the coat protein gene (Fig. 1) was sequenced using primer T-58 (nts. 5 686–5 706); a 190

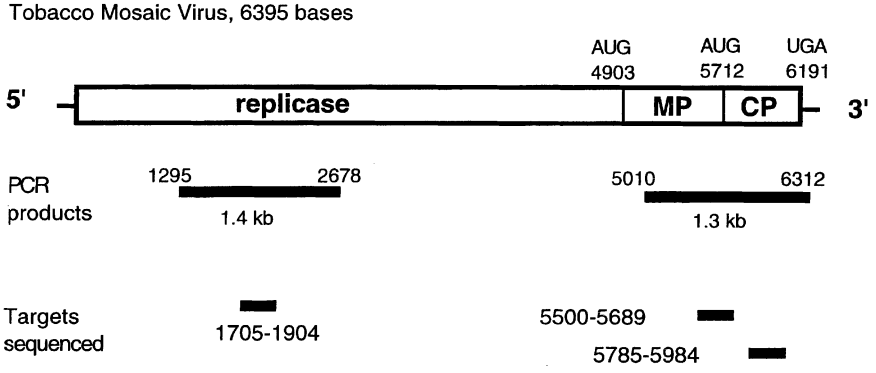

Fig. 1. Genome of tobacco mosaic virus showing the regions amplified by PCR and the target regions sequenced after subsequent cloning. The first and last nucleotide for each region is shown. *MP* Movement protein gene; *CP* coat protein gene. The start sites and one stop site are shown for MP and CP. The open boxes represent ORFs while the terminal lines represent the 5′ and 3′ untranslated regions

base region in the movement protein gene was sequenced with D-33 (5 780–5 763); and a 200 base region in the replicase region was sequenced with primer J-1 (1 603–1 625). Reads of 180–190 bases, rather than 200 bases, were taken if the upper bands were too compressed to be accurately resolved.

Analysis of sequence data

G'_{ST} values [22] indicate the proportion of variation due to interpopulational variation versus intrapopulational variation. For the purpose of the G'_{ST} calculation only, all the replicate populations that came from the same host species were considered as a single "population." In this way, the extent of the contribution of host shift to populational diversity could be estimated by comparing the differences between the seven host populations with the variation within each population. Each unique viral sequence variant derived from one plant replicate was considered a haplotype; sequences not differing from the initial sequence were considered as a single haplotype. Separate calculations were made for each target region. It was first necessary to determine the genetic distance between each pair of haplotypes, using the method of Jukes and Cantor [13]. The diversity within each population and between each pair of populations was then determined as described by Nei, p. 276 [22]. The proportion of variation due to populational (host type) differences versus variation within populations was calculated from this data as a G'_{ST} value of each target region, as in Nei, pp. 198–191 [22].

To estimate the proportion of nonsynonymous mutations versus synonymous mutations that occurred in the target sequences during passaging, the method of Pamilo and Bianchi [25] and Li [18] was employed. For the purposes of calculating this proportion, a composite sequence containing all substitutions detected in the study was compared to the original sequence. This was done for the coat protein region data alone and for all three target regions grouped together. The Ka/Ks value estimates the proportion of nonsilent mutations in the population after host shift.

Results

The initial viral population and passaging

The serial passages of TMV (Table 1) were initiated with a population derived from a single *N. tabacum* cv. Xanthi plant inoculated with in vitro transcripts of the bacterial plasmid , pTMV204 [6]. pTMV204 contains a sequence of TMV-U1 (wild type) which was cloned from a viral culture that had been maintained in *N. tabacum* cv. Xanthi continuously for 40 years in a greenhouse (W. O. Dawson, pers. comm.). The host plants used in extended passaging (Table 1) represented six different plant families and a range of host responses to TMV, including necrotic resistance, restriction of infection to inoculated leaves, systemic spread, and the presence or absence of symptoms. Four replicate viral populations grown in each of the seven host species were passaged initially, but, at various points in the experiment, six populations failed to infect the next set of healthy plants, leaving a total of 22 populations at the end of the experiment.

Mutations accumulating in three target areas after passaging

Three regions of the TMV genome were targetted for a detailed examination of sequence heterogeneity by sequencing clones derived from the TMV populations.

We chose three target areas (Fig. 1) after aligning published protein sequences of tobamovirus genes and identifying phylogenically conserved and variable regions (data not shown). These regions were: (1) a conserved region of the coat protein gene; (2) the very variable 3′ region of the movement protein gene; and (3) a moderately variable region of the replicase gene. The replicase region corresponded to the region between the putative methylase and helicase domains of the replicase [10].

As seen in Tables 2–4, the frequency of mutations which accrued during passaging differed with each region but not as might be expected from viral phylogenetic comparisons. The coat protein gene target acquired the most nucleotide substitutions, the movement protein gene target the least, and replicase gene target an intermediate amount (Tables 2–4). No one host species appeared to induce an especially large number of mutations in the viral sequences examined. To examine the effect of host type quantitatively, an analysis similar to that used by Fraile et al. [7] was employed, as outlined in Nei [22]. G'_{ST} values were calculated for each set of mutations which occurred in different hosts. These values indicate the proportion of variation (ranging from 0 to 1) which can be attributed to host type compared to the variation found within each host type. A G'_{ST} value of 0.7 was calculated for the coat protein gene data (Table 2), with values of 0.5 and 0.5 for the movement protein and replicase gene data (Tables 3 and 4)

Table 2. Unique mutations found in the coat protein gene region (TMV nts. 5785–5984) of viral populations after long term passaging

Host	Number of replicate populations	Number of clones sequenced	Number of mutations	Number of bases sequenced	Combined no. mutations per base per year
Tobacco	4	13	1	2416	3×10^{-4}
Nightshade	4	12	1	2377	4×10^{-4}
Collinsia	3	11	1	2108	4×10^{-4}
Phacelia	1	3	1	595	1×10^{-3}
Buckwheat	3	11	1[a]	2196	3×10^{-4}
Marigold	4	13	2	2453	6×10^{-4}
Plantain	2	6	1[b]	1176	8×10^{-4}
Control[c]	1	30	0	6,000	0
Totals[d]	21	69	8	13,321	5×10^{-4}

[a]Two identical mutations were detected in two out of a total of four clones sampled from the same replicate population in buckwheat and were counted as one unique mutation

[b]Three identical mutations were detected out of a total of three clones sampled from the same replicate population in plantain and were counted as one unique mutation

[c]Control was a single tobacco plant inoculated with the same extract from an in vitro transcript-inoculated tobacco plant that was used to initiate all test passages

[d]Excluding control data

Table 3. Mutations found in the movement protein gene region (TMV nts. 5500–5689) of viral populations after long term passaging

Host	Number of replicate populations	Number of clones sequenced	Number of mutations	Number of bases sequenced	Combined no. mutations per base per year
Tobacco	3	7	0	1176	0
Nightshade	4	8	1	1458	6×10^{-4}
Collinsia	3	9	0	1578	0
Phacelia	1	5	0	950	0
Buckwheat	3	16	1	2660	3×10^{-4}
Marigold	4	9	0	1823	0
Plantain	2	7	0	1330	0
Totals	21	54	2	10,975	1×10^{-4}

Table 4. Unique mutations accumulating in the replicase gene region (TMV nts. 1705–1904) of viral populations after long term passaging

Host	Number of replicate populations	Number of clones sequenced	Number of mutations	Number of bases sequenced	Combined no. mutations per base per year
Tobacco	4	17	1	3400	2×10^{-4}
Nightshade	4	13	3[a]	2499	1×10^{-3}
Collinsia	3	5	0	942	0
Phacelia	1	3	0	600	0
Buckwheat	3	10	0	1870	0
Marigold	4	6	0	1200	0
Plantain	2	4	0	800	0
Totals	21	58	4	11,311	3×10^{-4}

[a]Five identical mutations were detected in a total of five clones sampled from the same replicate population in nightshade and were counted as one unique mutation; this, added to two other unique mutations, gave a total of three unique mutations (see Table 6)

respectively. This suggested that, for the coat protein target at least, much of the observed variance is linked with the passage host. To test for significance, a χ^2 was calculated from the data of Table 5 with the null hypothesis that the mutation frequencies were the same for the seven host species. The observed p-value of 0.69 demonstrated that there was no evidence for the effect of host type in this data set.

The evolutionary rate after host shift was calculated using the summary data (Table 5). The overall mutation frequency was 3.1×10^{-4} substitutions per base

Table 5. Total numbers of unique mutations accumulating in all target regions in each host species

Host	Number of clones sequenced	Number of mutations	Number of bases sequenced	Number of mutations per base per year
Tobacco	37	2	6,992	2×10^{-4}
Nightshade	33	5	6,334	6×10^{-4}
Collinsia	25	1	4,628	2×10^{-4}
Phacelia	11	1	2,145	4×10^{-4}
Buckwheat	37	2	6,726	2×10^{-4}
Marigold	28	2	5,476	3×10^{-4}
Plantain	17	1	3,306	3×10^{-4}
Totals	188	14	35,607	3.1×10^{-4}

per year for all hosts and target regions, or 2.0 substitutions per TMV genome per year. Dividing the hosts into two symptom groups, viral populations passaged on systemic hosts (see Table 1) had an average of 4×10^{-4} unique substitutions per base per year, while those on localizing hosts had a frequency of 2×10^{-4}. These figures are based on unique mutations; e.g., the A $\rightarrow$ G transition at base 1721 was counted only once in these calculations, though it was recorded in five clones from the same population (Table 6). In fact, sets of identical mutations were detected in three different replicate populations representing three different host species. To ascertain that the mutations observed in the test populations were not artifactual, a negative control was analyzed. This consisted of a tobacco plant (*N. tabacum* cv. Xanthi) inoculated with the stored frozen inoculum derived from the original transcript-infected tobacco plant used to initiate the study. Clones derived from the coat protein target gene of TMV RNA from this plant had no mutations in 6 000 bases sequenced (Table 2), compared to 8 unique mutations in 13,321 bases of the coat protein gene target of all host lineages.

Patterns were also evident in the type of mutations in all the populations (Table 6). Of the 13 unique substitutions, 11 were synonymous (85%). These data were used to calculate a Ka/Ks proportion with respect to the original sequence. This value estimates the proportion of nonsynonymous vs. synonymous substitutions that occurred. A Ka/Ks value of 0.09 was calculated for all three targets combined and a value of 0.07 was calculated for the coat protein target alone. A strong bias towards transitions was also apparent, as 11 out of 13 unique substitutions (85%) were transitions. The nucleotide 5948 was mutated in three separate replicate populations and two different host species, suggesting adaptive evolution. Finally, a single deletion in the movement protein gene of population 3 in nightshade was found which eliminated the final 16 amino acids from the C terminus of the movement protein.

Table 6. Types and locations of all mutations detected in viral populations after long term passaging

Host	Region sequenced[a]	Replicate population[b] code	Location on TMV sequence[c]	Codon changed	Mutation type
Tobacco	CP	2	5978	GAC⇒GAU	silent
	REP	4	1739	UAC⇒UAU	silent
Nightshade	CP	3	5948	CCG⇒CCA	silent
	MP	3	5647	AAA⇒ AA	deletion[d]
	REP	4	1721	GAA⇒GAG	silent
	REP	4	1721	GAA⇒GAG	silent
	REP	4	1721	GAA⇒GAG	silent
	REP	4	1721	GAA⇒GAG	silent
		(same clone)[e]	1760	UCG⇒UCA	silent
	REP	4	1721	GAA⇒GAG	silent
		(same clone)[f]	1844	GCG⇒GCA	silent
Collinsia	CP	3	5810	GGA⇒GGU	silent
Phacelia	CP	3	5849	CAA⇒CAG	silent
Buckwheat	CP	1	5900	UUC⇒UUU	silent
	CP	1	5900	UUC⇒UUU	silent
	MP	1	5597	GUG⇒GCG	val⇒ala
Marigold	CP	1	5948	CCG⇒CCA	silent
	CP	4	5948	CCG⇒CCC	silent
Plantain	CP	3	5959	GCA⇒GUA	ala⇒pro
	CP	3	5959	GCA⇒GUA	ala⇒pro
	CP	3	5959	GCA⇒GUA	ala⇒pro

[a] *CP* Coat protein gene; *MP* movement protein gene; *REP* replicase gene

[b] A total of four replicate viral populations were initiated per host and were labeled 1 through 4

[c] Sequence numbering as in [9]

[d] A single base deletion of nt. 5647 which results in a frameshift that would truncate the movement protein by 16 amino acids at the C-terminus

[e] Two mutations in same sequence of one clone (positions 1721 and 1760)

[f] Two mutations in same sequence of one clone (positions 1721 and 1844)

Discussion

There are previously published indications that tobamovirus populations evolve slowly. As mentioned earlier, studies on field populations of tobamoviruses have found the population to be stable [7, 20, 31, 33]. Similarly, a long-term passaging experiment in *N. tabacum* starting with a field isolate from *N. glauca* found little sequence change as detected by RNase protection assays [32].

However, the shifting of a viral population from one host to another can have dramatic effects on viral populations, including TMV populations. Bald [2] found different variants of TMV in different host species in a natural environment, and Pelham et al. [26] found resistance-breaking TMV strains in TMV-resistant tomatoes. Khan and Jones [16] studied the effect of the host plant development on the

proportion of TMV variants which produce necrotic lesions in *Nicotiana sylvestris* (NL variants). These variants dominated the floral stalk of *N. sylvestris*, but were often not detected in the rosette of leaves at the base of the plant. In another study, selection of these mutants also occurred during culture of TMV-U1-infected *N. sylvestris* tissues [15]. Aldaoud et al. [1] found host effects on the evolution of these populations, including a decrease in the proportion of NL variants to levels that could not be detected in a TMV-U1 background after three passages in *Physalis floridana* but not in other hosts. There are numerous examples of animal virus adaptations to host species, such as a 2-million-fold fitness adaptation change, demonstrated when vesicular stomatitis virus was shifted from BHK cells to sandfly cells [24]. The host-specific pathogenicity of simian immunodeficiency virus shifted dramatically in just one passage in an alternate host [5], as did the composition of the populations [34].

In contrast to these studies, our viral populations, despite being grown in different hosts, appeared to experience an evolutionary rate lower than most described for field isolates for other viruses. We calculated 3.1×10^{-4} substitutions per base per year (0.03%) to be the evolutionary rate for our populations. A study of a pepper mild mottle tobamovirus epiphytotic occuring in commercial greenhouses estimated a maximum of 4% sequence divergence between haplotypes occurring at the beginning and end of the 4-year study period [31]. Both experiments examined approximately 9% of the genome. A rate of 5×10^{-4} substitutions per year was found for the measles H gene by sequencing 75 natural isolates [30], which is close to our estimated rate after host shifting. Other estimated annual evolution rates from animal RNA virus field isolates have been larger [4, 17, 23]. It should be noted that all the works cited above (except [23]) have used direct sequencing or RNase protection assays to find the consensus sequence of each field isolate and then compared the sequences of all the isolates. Our work examined multiple clones per sample and thus included sample heterogeneity in the evolutionary rate estimate. Only two of the mutations (nt. 1721 and nt. 5959 in Table 6) in 22 populations might have been prevalent enough in our populations to have been detected by direct sequencing or RNase protection assay. Thus our rate determinations might need to be adjusted to an even smaller value to be comparable to the work cited above. However, with all studies examining the frequency of mutation accumulation in virus populations, a far greater number of mutants may have appeared and disappeared during the time course of the observation period. If purifying selection is operative, the low mutation frequency may be due more to selection than to an intrinsically low mutation rate. Such selective pressure is suggested by the frequencies of silent, transitional, and repeated mutations as discussed below.

The combined data (Table 6) suggested a tendency toward amino acid sequence conservation, despite nucleotide mutation, through the host shift. First, there was a preponderance of silent (85%) as compared to nonsilent substitutions, as indicated more precisely by the Ka/Ks values of 0.09 for all three regions examined. These values are less (fewer nonsilent substitutions) than most recorded for other RNA viruses (see, e.g., Table 7.9 in Li [19]). This is generally thought

to be an indicator of "negative" or purifying selection which favors the prevalent protein sequence and selects against variants [27]. A similar purifying selection was seen in analyses of field isolates of another tobamovirus [7, 20]. We found a strong bias towards transitions (85%) over transversions; this bias is greater than most of a wide range of mammalian genes (Table 7.2 in Li [19]). Transitions exchange similarly sized nucleotides (either purines or pyrimidines) so perhaps this bias was due to a need to maintain a certain structure in the RNA.

Seven of the 21 total mutations observed were found more than once; though only a small number of mutations were recorded, this may indicate a predominance of certain mutants occuring in some populations. From one virus replicate population in buckwheat, two out of four clones had the same silent substitution; in one plantain population, three out of three clones encoded the same ala $\rightarrow$ pro substitution; and in one nightshade population, five out of five clones had the same silent mutation (Tables 2, 4 and 6). Two of the identical mutations in the plantain set co-occurred with different second site mutations (Table 6). The presence of these sets of identical mutations could be due to adaptive evolution or simply to stochastic effects such as bottlenecks.

The stability of the genomic sequence and the small rate of sequence change observed in our populations corroborates results obtained in another study using another tobamovirus [32]. In that study, the virus was obtained from naturally infected *Nicotina glauca* plants from the field. The virus was then cloned via local lesion passage and then transferred to *Nicotiana tabacum* cv. Samsun for systemic infection. In four different passage series of 20 passages each in Samsun tobacco, no new variants appeared, as determined by T1 fingerprinting. A single variant was detected after further passaging. A field study with another tobamovirus [33] found one haplotype to be prevalent over a period of eight years, though a replacement of other haplotypes occurred during this period as well. It may be that some greenhouse cultures, especially those derived from in vitro transcripts, are qualitatively different from field cultures in their evolution. For example, Khan and Jones [16] noted the absence of genetic variants in a TMV culture derived from an in vitro transcripts compared to one obtained from a natural source and passaged many years in the greenhouse. Perhaps the evolutionary rates for tobamoviruses in the field are driven by a replacement of fairly stable sequences while those in our viral populations were dependent on the *de novo* development of new variants. Though the error rates of RNA polymerases are known to be large, there may be other factors which may limit the development of new variants, such as bottlenecks in cell-to-cell movement and phloem loading or in plant-to-plant transmission. These were not examined in these studies, but such comparisons may be interesting for future research.

The small number of mutations detected in this study did not allow us to determine the effect of particular host types on the evolution of the virus population. From the mutation frequency data, we did calculate a G'_{ST} value for the coat protein sequence which was indicative of greater sequence diversity between host types compared to diversity within the host types. However, χ^2 analysis using the combined data set found no significant connection between host type and the

sequence changes observed. It is interesting that none of the seven hosts induced a burst of evolutionary change in the viral sequences examined, in spite of the differing types and levels of host resistance encountered by the virus (ringspot lesions, systemic infection, nonsymptomatic restriction to the inoculated leaf). These data demonstrate that much longer target regions or many more replicates need to be examined to significantly define host-directed evolution over this time scale. As we examined only 9.2% of the genome, we easily may not have detected hypervariable sites affected by host type. Alternatively, evolutionary genetic markers, such as the NL trait discussed above, could be utilized to follow host-induced evolutionary patterns.

Acknowledgements

Our thanks to Vitaly Boyko for assistance with the computer alignment of tobamovirus sequences, to Jack Ewing for help with χ^2 analysis, and to Fernando García-Arenal for advice on sequence data analysis. This grant was funded by USDA grant No. 94-39210-0371.

References

1. Aldaoud R, Dawson WO, Jones GE (1989) Rapid, random evolution of the genetic structure of replicating tobacco mosaic virus populations. Intervirology 30: 227–233
2. Bald JG (1960) Forms of tobacco mosaic virus. Nature 188: 645–647
3. Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 7: 1 513–1 520
4. Chao YC, Tang HS, Hsu CT (1994) Evolution rate of hepatitis delta virus RNA isolated in Taiwan. J Med Virol 43: 397–403
5. Courgnaud V, Laure F, Fultz PN, Montagnier L, Brechot C, Sonigo P (1992) Genetic differences accounting for evolution and pathogenicity of simian immunodeficiency virus from a sooty mangabey monkey after cross-species transmission to a pig-tailed macaque. J Virol 66: 414–419
6. Dawson WO, Beck DL, Knorr DA, Grantham GL (1986) cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. Proc Natl Acad Sci USA 83: 1 832–1 836
7. Fraile A, Malpica JM, Aranda MA, Rodriguez CE, Garcia AF (1996) Genetic diversity in tobacco mild green mosaic tobamovirus infecting the wild plant *Nicotiana glauca*. Virology 223: 148–155
8. García-Arenal F, Palukaitis P, Zaitlin M (1984) Strains and mutants of tobacco mosaic virus are both found in virus derived from single-lesion-passaged inoculum. Virology 132: 131–137
9. Goelet P, Lomonossoff GP, Butler PJG, Akam ME, Gait MJ, Karn J (1982) Nucleotide sequence of tobacco mosaic virus RNA. Proc Natl Acad Sci USA 79: 5 818–5 822
10. Goldbach R (1990) Genome similarities between positive-strand RNA viruses from plants and animals. In: Brinton MA, Heinz FX (eds) New aspects of positive-strand RNA viruses. American Society for Microbiology, Washington, pp 3–11
11. Gooding GV Jr, Hebert TT (1967) A simple technique for purification of tobacco mosaic virus in large quantities. Phytopathology 57: 1 285
12. Hanahan D (1985) Techniques for transformation of *E. coli*. In: Glover DM (ed) DNA cloning, vol 1. IRL Press, Oxford, pp 109–135
13. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

14. Kearney CM, Chapman SN, Turpen TH, Dawson WO (1995) High level gene expression in plants using RNA viruses as transient expression vectors. In: Gelvin S (ed) Plant Molecular biology manual, vol L1. Kluwer Academic, Dordrecht, pp 1–16
15. Khan IA, Jones GE (1989) Accumulation of necrotic lesion inducing variants in TMV-infected plantlets derived from leaf disks of *Nicotiana sylvestris*. Can J Bot 67: 984–989
16. Khan IA, Jones GE (1989) Selection for a specific tobacco mosaic virus variant during bolting of *Nicotiana sylvestris*. Can J Bot 67: 88–94
17. Lee CM, Bih FY, Chao YC, Govindarajan S, Lai MM (1992) Evolution of hepatitis delta virus RNA during chronic infection. Virology 188: 265–73
18. Li W-H (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36: 96–99
19. Li W-H (1997) Molecular evolution. Sinauer Associates, Sunderland, MA
20. Moya A, Rodríguez-Cerezo E, García-Arenal F (1993) Genetic structure of natural populations of the plant RNA virus tobacco mild green mosaic virus. Mol Biol Evol 10: 449–456
21. Murphy G, Kavanagh T (1988) Speeding up the sequencing of double-stranded DNA. Nucleic Acids Res 16: 5 198
22. Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
23. Ni YH, Chang MH, Chen PJ, Lin HH, Hsu HY (1997) Evolution of hepatitis C virus quasispecies in mothers and infants infected through mother-to-infant transmission. J Hepatol 26: 967–974
24. Novella IS, Clarke DK, Quer J, Duarte EA, Lee CH, Weaver SC, Elena SF, Moya A, Domingo E, Holland JJ (1995) Extreme fitness differences in mammalian and insect hosts after continuous replication of vesicular stomatitis virus in sandfly cells. J Virol 69: 6 805–6 809
25. Pamilo P, Bianchi NO (1993) Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol Biol Evol 10: 271–281
26. Pelham J, Fletcher JT, Hawkins JH (1970) The establishment of a new strain of tobacco mosaic virus resulting from the use of resistant varieties of tomato. Ann Appl Biol 65: 293–297
27. Plikat U, Nieselt-Struwe K, Meyerhans A (1997) Genetic drift can dominate short-term human immunodeficiency virus type 1 nef quasispecies evolution in vivo. J Virol 71: 4 233–4 240
28. Plyusnin A, Cheng Y, Lehvaslaiho H, Vaheri A (1996) Quasispecies in wild-type tula hantavirus populations. J Virol 70: 9 060–9 063
29. Polyak SJ, McArdle S, Liu SL, Sullivan DG, Chung M, Hofgartner WT, Carithers RL, Jr., McMahon BJ, Mullins JI, Corey L, Gretch DR (1998) Evolution of hepatitis C virus quasispecies in hypervariable region 1 and the putative interferon sensitivity-determining region during interferon therapy and natural infection. J Virol 72: 4 288–4 296
30. Rima BK, Earle JA, Baczko K, ter Meulen V, Liebert UG, Carstens C, Carabana J, Caballero M, Celma ML, Fernandez-Munoz R (1997) Sequence divergence of measles virus haemagglutinin during natural evolution and adaptation to cell culture. J Gen Virol 78: 97–106
31. Rodríguez-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: 328–332
32. Rodríguez-Cerezo E, García-Arenal F (1989) Genetic heterogeneity of the RNA genome population of the plant virus U5-TMV. Virology 170: 418–423
33. Rodríguez-Cerezo E, Moya A, García-Arenal F (1989) Variability and evolution of the plant RNA virus pepper mild mottle virus. J Virol 63: 2 198–2 203

34. Valli PJ, Goudsmit J (1998) Structured-tree topology and adaptive evolution of the simian immunodeficiency virus SIVsm envelope during serial passage in rhesus macaques according to likelihood mapping and quartet puzzling. J Virol 72: 3 673–3 683
35. Zitter TA, Murakishi HH (1969) Nature of increased virulence in tobacco mosaic virus after passage in resistant tomato plants. Phytopathology 59: 1 736–1 739

Authors' address: Dr. C. M. Kearney, Department of Biology, Baylor University, P.O. Box 97388, Waco, TX, 76798, U.S.A.

Received February 9, 1999