J Mol Evol (1991) 32:328–332

Journal of Molecular Evolution

High Genetic Stability in Natural Populations of the Plant RNA Virus Tobacco Mild Green Mosaic Virus

Emilio Rodríguez-Cerezo,[1] Santiago F. Elena,[2] Andrés Moya,[2] and Fernando García-Arenal[1]

[1] Departamento de Patología Vegetal, E.T.S.I. Agrónomos, 28040 Madrid, Spain
[2] Departamento de Genética, Facultad de Biología, Dr. Moliner 50, 46100 Burjassot, Valencia, Spain

Summary. Quantitative studies on the genetic variation of plant viruses are very scarce, in spite of their theoretical and applied importance. We report here on the genetic variability of field isolates of the plant RNA virus tobacco mild green mosaic virus (TMGMV) naturally infecting the wild plant *Nicotiana glauca* Grah. The populations studied were composed of a high number of haplotypes. Two main features are found regarding TMGMV variation: First, there is no correlation between genetic proximity of isolates and geographic proximity of the sites from which they were obtained; and second, the estimated divergence among haplotypes is low, and values are maintained regardless of the scale of the distance between the sites from which the isolates come. No comparable studies have been done with a plant RNA virus, and these two features seem to be unique for this system as compared with other RNA viruses.

Key words: Genetic stability — Plant RNA viruses — Tobacco mild green mosaic virus

Introduction

In the past 15 years molecular methods of genome comparison have been extensively applied to bacterial and animal RNA viruses, yielding data on which models that describe the structure and evolution of RNA genomes have been built, giving a general picture of high genetic variation in RNA viruses (Domingo and Holland 1988). In spite of the theoretical and applied importance of genetic studies of the populations of plant pathogens (Wolfe and Caten 1987), quantitative studies on the variation of plant viruses are very scarce (Van Vloten-Doting and Bol 1988). We present here for the first time an analysis of the genetic variation of randomly sampled populations of an RNA virus naturally infecting a wild plant. The chosen system was tobacco mild green mosaic tobamovirus (TMGMV), a plus sense RNA virus with a 6.4-kb monopartite genome (Wetter 1986), that in nature infects *Nicotiana glauca* Grah., a wild Solanaceae found in areas of Mediterranean climate throughout the world. The data obtained show that this plant RNA virus is exceedingly stable genetically.

Materials and Methods

Virus Isolates and Sampling. Sixty-nine isolates of TMGMV naturally infecting *N. glauca* in SE Spain were collected in 1985 and 1986 through two types of sampling: a large-scale sample (LS) over a 300-km transect covering the optimal area of the distribution in Spain of the host plant (Fig. 1A), and five small-scale samples (SS) in five of the discrete stands in which the plant grows in this area (stands E, C, S, R, and P in Fig. 1A and B). These stands appear as well-defined groups isolated from other *N. glauca* plants. Plants were always randomly sampled by sampling one plant each x meters (x varying with the scale of the sample) on previously fixed itineraries. Four TMGMV isolates from Australia, the gift of Dr. J.W. Randles, named as in Randles et al. (1981), were also included in the study.

Genomic Comparisons. Isolates from individual plants were biologically cloned through single lesion passage in *Nicotiana tabacum* L. Xanthi-nc and multiplied in *N. tabacum* Samsum. We had shown previously that clones represent the uncloned field isolate from which they were derived (Rodriguez-Cerezo and García-Arenal 1989), and for a number of isolates it was shown

Offprint requests to: F. García-Arenal

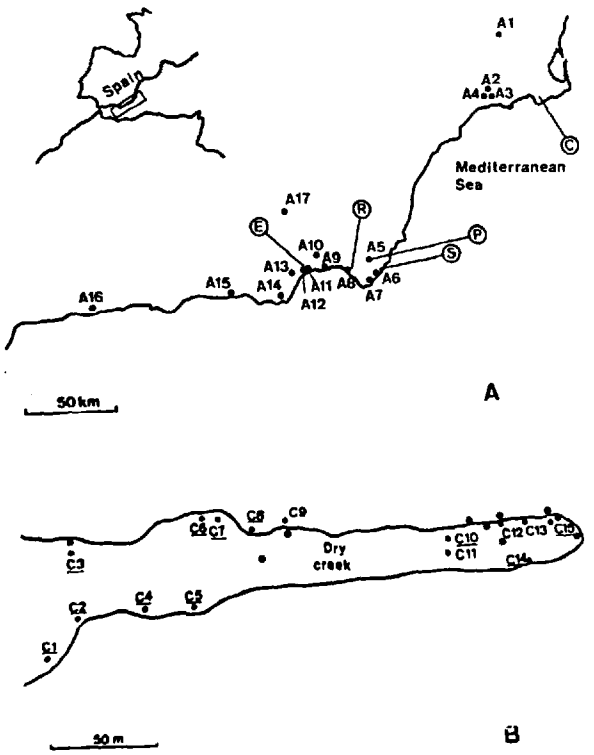

Fig. 1. Area of sampling of TMGMV from *N. glauca* in SE Spain. **A** Large-scale sample; A1–17 indicates sites where the 17 LS isolates were obtained; E, S, P, R, and C indicate the position of five *N. glauca* stands from where the 52 (10 for E, 11 for S, 10 for P, 10 for R, 11 for C) SS isolates were obtained. Isolates were named by a letter (A, E, S, P, R, C) and a number: consecutive numbers indicate plants sampled consecutively and thus, nearest sites of isolate obtainment. **B** Map of stand C indicating the position of all *N. glauca* plants (*, ●), of those tested for viral infection (*, Ci), and of those infected by TMGMV (●, Ci).

that the clone and the parent uncloned isolate had the same fingerprint. Thus, it is legitimate to compare clones derived from field isolates, a procedure otherwise unavoidable, as mixed infections with other viruses were found to be very frequent in *N. glauca*. The genomic RNAs were compared by two-dimensional electrophoretic maps of oligonucleotides generated by complete digestion with RNase T1 (T1 fingerprints) as in García-Arenal et al. (1984), except that 25% polyacrylamide gels were used for the second dimensions. RNase T1 cleaves internucleotide linkages on the 3′ side of guanylate residues (Gp), so that from a large heteropolymeric RNA a series of oligonucleotides are produced each having a 3′ Gp residue. Only the large oligonucleotides (≥10 bases), representing unique sequences, have been considered for the analysis of the relatedness of the different genomic RNAs, according to Aaronson et al.'s (1982) evaluation of the method. For TMGMV RNA a typical fingerprint (Fig. 2) resolves about 55 oligonucleotides in the size range of 10–25 bases. The size distribution of the oligonucleotides has been determined by sequencing (not shown), and they were found to represent 12% of the 6400-nucleotide-long TMGMV RNA.

The intra- and interpopulational divergences were estimated as follows. Let m_i and m_j be the number of oligonucleotides for RNA sequences i and j, and m_{ij} the number of oligonucleotides shared by the two sequences. The fraction of shared oligonucleotides, F_{ij}, is estimated by $\hat{F}_{ij} = 2m_{ij}/(m_i + m_j)$. According to Nei (1987) the mean number of nucleotide substitutions, d_{ij}, per site can be estimated by $\hat{d}_{ij} = -(2/r)\log_e \hat{G}$, where r is the length of the recognition sequence (in our case r = 1) and G is estimated by the iteration formula $\hat{G} = [\hat{F}(3 - 2G_1)]^{1/4}$ where G_1 is a trial value of G, for example, $G = F^{1/4}$. The iterative computation is done until $G = G_1$. Usually a few cycles of iteration are sufficient. Once d_{ij} is estimated, we can estimate the intra- and interpopulational divergences. Following Nei (1987), if there are n different haplotypes in populations k and l, and x_i and y_j are the sample frequencies of the ith and jth haplotypes for populations k and l, respectively, then the average number of nucleotide substitutions for a randomly chosen pair of haplotypes, D_k, in population k can be estimated by

$$\hat{D}_k = n_k/(n_k - 1) \sum_{ij} x_i x_j \hat{d}_{ij}$$

where n_k is the number of sequences sampled. The net average number of nucleotide substitutions, D_{kl}, between RNA haplotypes from k and l is estimated by

$$\hat{D}_{kl} = \sum_{ij} x_i y_j \hat{d}_{ij} - \tfrac{1}{2}(\hat{D}_k + \hat{D}_l)$$

where $\hat{d}_{ij}$ is the nucleotide substitution between the ith haplotype from k and jth haplotype from l. The net nucleotide substitution, or the average minimum genetic distance, between s populations is given by

$$\bar{D}_m = \sum_{k \neq l} \hat{D}_{kl}/[s(s - 1)]$$

Similarities (shared oligonucleotides) among fingerprint types (Nei 1987) were used in UPGMA cluster analyses. Fitch and Margoliash's method using $\hat{d}_{ij}$ estimates between haplotypes was also used to construct a phylogenetic tree.

Results

Genetic Variation in Populations of TMGMV

Sixty-nine out of 90 *N. glauca* plants, gathered in both LS and SS sampling, were found to be infected with TMGMV regardless of their showing symptoms or not, which suggests a long virus/host association typical of endemic situations.

The viral populations studied were found to be composed of a high number of different haplotypes: the four AUS isolates correspond to four different haplotypes and the 17 LS isolates to 15. All these haplotypes appear to be closely related as indicated by the fraction of shared oligonucleotides between two given haplotypes, which averages 0.930 ± 0.031 for the LS group of isolates and 0.919 ± 0.042 for the LS and AUS isolates pooled together. The 52 SS isolates correspond to 32 haplotypes in which only five appear in two different stands, and none in more than two stands. The average fraction of shared oligonucleotides between two haplotypes of this group is 0.950 ± 0.020. Two thirds of the spots defining a haplotype were found to be invariant for all the 51 haplotypes.

The estimated values for intrapopulational divergence are small in all cases: 0.0142 on the average for the five populations belonging to the SS isolate group (Table 1A), 0.0216 for the LS isolates, and 0.0342 for the pooled LS + AUS isolates. The net interpopulational divergence values are also small, stressing again the limited divergence among haplotypes: they average 0.0026 when comparing the

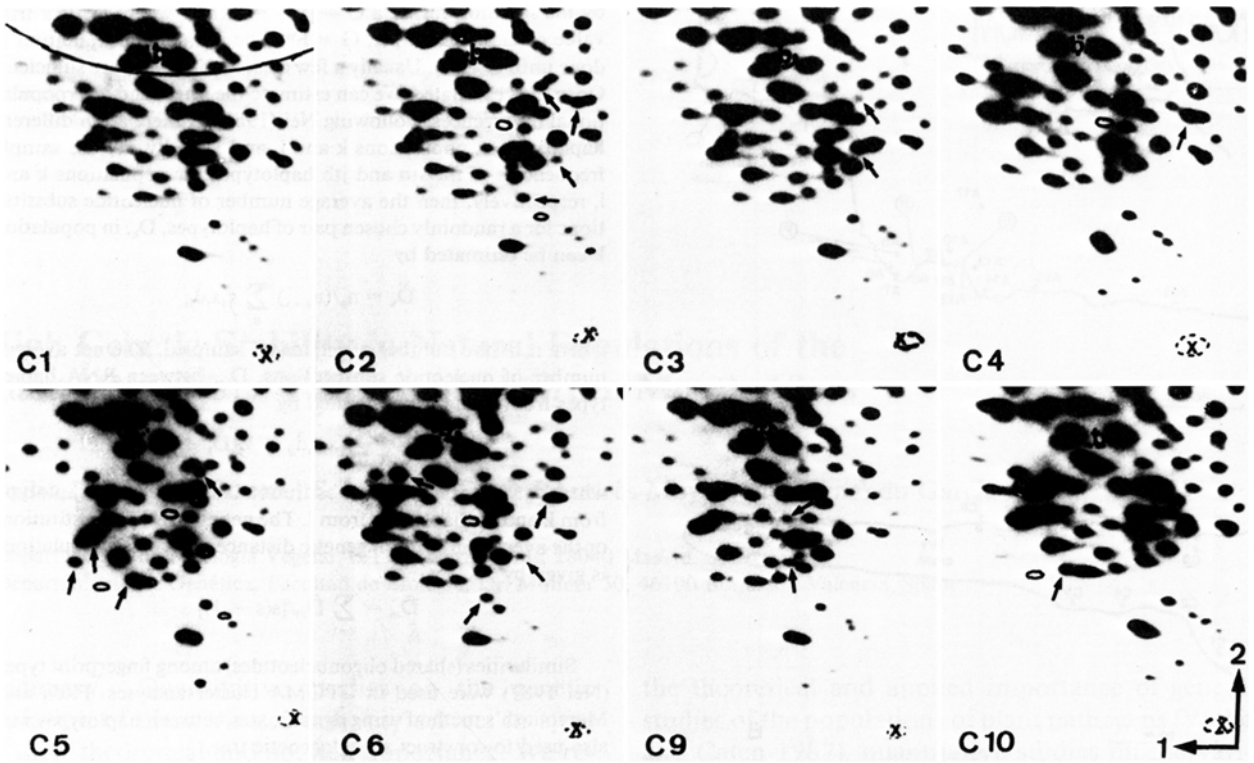

Fig. 2. Fingerprints of TMGMV corresponding to isolates in SS-C (C1 = C7 = C8, C2, C3, C4, C5 = C11, C6, C9, and C10 as in Fig. 1B). Additional (→) and missing (O) oligonucleotides with respect to C1 are indicated. (x) and (b), positions of xylencyanol and bromophenol blue, respectively. Oligonucleotides used for comparisons (≥10 n long) are positioned below the line in C1. Arrows in C10 indicate direction of first and second electrophoresis.

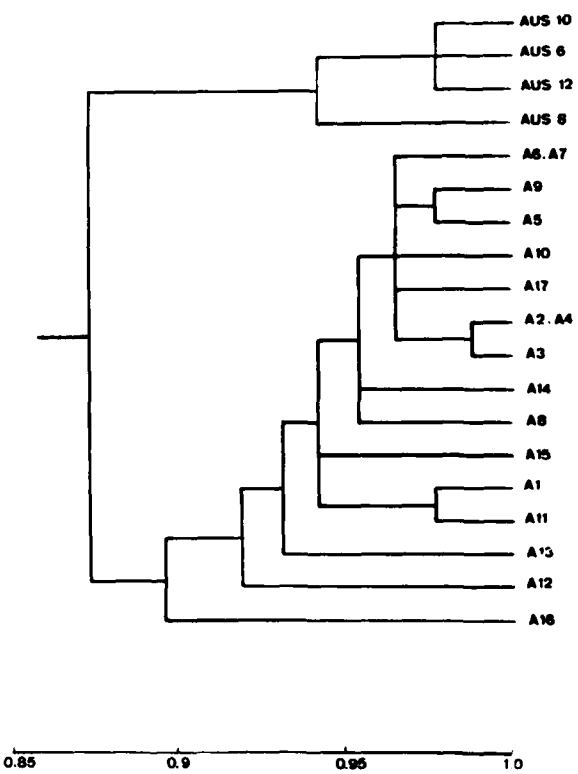

Fig. 3. UPGMA cluster analysis of AUS and LS haplotypes of TMGMV. Scale represents values of the fraction of shared oligonucleotides (F_{ij}) between two given haplotypes.

five populations of SS isolates (Table 1B) and reach only 0.0380 when the LS and AUS isolates are compared.

Relationships among Haplotypes

UPGMA clustering of LS and AUS isolates on the basis of the values of shared oligonucleotides among pairs of haplotypes (Fig. 3) shows that the Australian isolates group together separately from the Spanish ones. Within the subgroup of Spanish isolates, a lack of correlation between genetic and geographic proximity of their sites of origin is observed. To further investigate this point, a matrix of geographic distances was built on data of the longitude, latitude, and altitude of each stand (data not shown) and compared with the matrix of the net interpopulational genetic distances of the five groups of SS isolates (Table 1B) by means of Mantel's nonparametric test (Manly 1985). No significant correlation between both matrices was found.

Evolutionary relationships among haplotypes were established by Fitch and Margoliash's method. The same type of unrooted tree was obtained with haplotypes belonging to either the LS or SS isolates, or to a mixture of both (Fig. 4): haplotypes diverge

Table 1. Values of intrapopulational (A) and interpopulational (B) divergences among SS isolates coming from different stands of *N. glauca*.

A		B					
Stand	Divergence	Stand	E	S	P	R	C
E	0.0121	E	—	0.0018	0.0023	0.0012	0.0024
S	0.0146	S		—	0.0013	0.0025	0.0040
P	0.0191	P			—	0.0024	0.0047
R	0.0154	R				—	0.0035
C	0.0100	Average minimum genetic distance					
Mean value	0.0142	$D_m = 0.0026$					

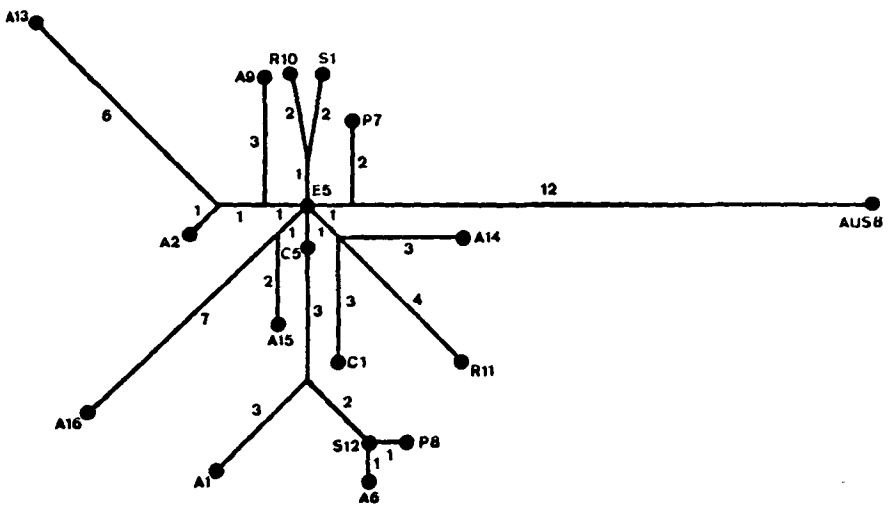

Fig. 4. Unrooted phylogenetic tree following Fitch and Margoliash's method of 17 randomly chosen TMGMV haplotypes corresponding to LS isolates (Ai) or to SS isolates (Ei, Si, Ri, Pi, Ci). The AUS8 isolate was introduced as outgroup.

from a central (possible ancestral) haplotype, but intermediate haplotypes arising by mutation accumulation have not been detected in our screening. Again, the AUS isolate used as outgroup is only partially differentiated from the Spanish ones, and again there is no correlation between the geographic origin of the Spanish isolates and their position in the tree.

Discussion

The pattern of variability here described for TMGMV is similar to the one reported for isolates sampled during epidemics in pepper caused by the closely related pepper mild mottle tobamovirus (PMMV) (Rodríguez-Cerezo et al. 1989). These viruses are similar in the limited divergence among haplotypes that do not conform to clear evolutionary lines. Nevertheless there is a significant ($P < 0.05$) difference in the distribution of haplotype frequencies: for PMMV 4 haplotypes (out of 9) account for 75% of the isolates, whereas for TMGMV the 4 more frequent haplotypes (of 32) represent 35% of the isolates, and 19 haplotypes represent 75% of the isolates. This may be compared with what has been frequently reported for phytopathogenic fungi (Burdon 1987; Wolfe and Caten 1987): in crop epidemics fungal populations are dominated by a few genotypes, being more heterogeneous in endemics on wild hosts. This difference has been related to host factors, both genetic (i.e., higher genetic heterogeneity in wild hosts) and ecological (i.e., density and distribution as it affects host availability), which could also account for what is observed for PMMV and TMGMV.

The variability model for TMGMV is also similar to models reported for animal RNA viruses such as influenza C (Buonagurio et al. 1986a) or foot-and-mouth disease (Dopazo et al. 1988) in which isolates from different evolutionary lines cocirculate, and is clearly different from those that evolved rapidly by mutation accumulation as reported for influenza A (Buonagurio et al. 1986b), enterovirus EV70 (Takeda et al. 1984), or vesicular stomatitis virus (Nichol et al. 1989).

Two outstanding features of TMGMV variation need to be stressed: (1) The lack of correlation between the genetic proximity of the isolates and the geographical proximity of the sites from which they were obtained. This is not explained in epidemio-

logical terms: the only known way of natural spread for TMGMV is direct contact between plants, as there are no known vectors and the virus is not seed transmitted (Randles et al. 1981). (2) The narrow limit of divergence among haplotypes, most dramatically reflected in the low values for divergences, that are maintained regardless of the scale of the distance between the sites from which the isolates come (a few meters in an *N. glauca* stand for SS isolates, 300 km for LS isolates, and thousands of km when AUS isolates are considered). Although high genetic stability has been reported for plant viruses (Block et al. 1987; Van Vloten-Doting and Bol 1988), a quantitative analysis such as the one presented here has not been reported previously, and it is not easy to explain the high genetic stability detected. It does not seem likely that the error frequency for TMGMV RNA replicase would be lower than for other RNA viruses, for which similar values have been reported (Smith and Inglis 1987; Domingo and Holland 1988). An alternative hypothesis would be to consider the functional constraints (negative selection) to variation that would limit the sites in the genome that can vary, as is suggested by the high proportion of sites invariant among the haplotypes (Kimura 1985). This would in turn place an upper limit on the number of possible variants, and it could explain the two above-stressed features. Selective factors, related to the nature of plant infection by viruses and to the heterogeneous nature of RNA genome populations, cannot be discarded and may contribute to the observed variability limits: a single virus particle could initiate infection into a plant cell, and infected cells could thus become resistant to later infections by related strains (Matthews 1981). Under these conditions, each new infection event would select the most frequent variant from the heterogeneous population of related genome sequences replicating into the plant and acting as inoculum source. It is relevant to note that under equilibrium conditions in tissues where viral accumulation has reached its maximum the heterogeneous genome population for TMGMV has been shown (Rodriguez-Cerezo and García-Arenal 1989) to be largely dominated by a prevailing (master) sequence.

Acknowledgments. We thank J.W. Randles (Adelaide, South Australia) for sending isolates and J.M. Malpica (Madrid, Spain) and F. González-Candelas (Valencia, Spain) for helpful discussions during the progress of this work. Financial support from Comisión Asesora para la Investigación Científica y Técnica (Grant PR83-2032) (Spain) and from Dirección General de Investigación Científica y Técnica (Grant PB86-0517) (Spain) is acknowledged.

References

Aaronson RP, Young JF, Palese P (1982) Oligonucleotide mapping: evaluation of its sensitivity by computer simulation. Nucleic Acids Res 10:237–246

Block J, Gibbs A, Mackenzie A (1987) The classification of tymoviruses by cDNA–RNA hybridization and other measures of relatedness. Arch Virol 96:225–240

Buonagurio DA, Nakada S, Fitch WM, Palese P (1986a) Epidemiology of influenza C virus in man: multiple evolutionary lineages and low rate of change. Virology 153:12–22

Buonagurio DA, Nakada S, Parvin JD, Krystal M, Palese P, Fitch WM (1986b) Evolution of human influenza A viruses over 50 years: rapid, uniform rate of change in NS gene. Science 232:980–982

Burdon JJ (1987) Diseases and plant population biology. Cambridge University Press, Cambridge, England

Domingo E, Holland JJ (1988) High error rates, population equilibrium and evolution of RNA replication systems, vol III. In: Domingo E, Holland JJ, Ahlquist P (eds) RNA genetics. CRC, Boca Raton FL, pp 3–35

Dopazo J, Sobrino F, Palma EL, Domingo E, Moya A (1988) Gene encoding capsid protein VP 1 of foot-and-mouth disease virus: a quasispecies model of molecular evolution. Proc Natl Acad Sci USA 85:6811–6815

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

Kimura M (1985) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England

Manly BFJ (1985) The statistic of natural selection. Chapman & Hill, London

Matthews REF (1981) Plant virology. Academic Press, New York

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York, pp 101, 106, and 276

Nichol ST, Rowe JE, Fitch WM (1989) Glycoprotein evolution of vesicular stomatitis virus New Jersey. Virology 168:281–291

Randles JW, Palukaitis P, Davies C (1981) Natural distribution, spread, and variation in the tobacco mosaic virus infecting *Nicotiana glauca* in Australia. Ann Appl Biol 98:109–119

Rodríguez-Cerezo E, García-Arenal F (1989) Genetic heterogeneity of the RNA genome population of the plant virus U5-TMV. Virology 170:418–423

Rodríguez-Cerezo E, Moya A, García-Arenal F (1989) Variability and evolution of the plant RNA virus pepper mild mottle virus. J Virol 63:2198–2203

Smith DB, Inglis SCJ (1987) The mutation rate and variability of eukaryotic viruses: an analytical review. J Gen Virol 68: 2729–2740

Takeda N, Miyamura K, Ogino T, Natori S, Yamazaki S, Sakurai S, Nakazono N, Ishii K, Kono P (1984) Evolution of enterovirus type 70: oligonucleotide mapping analysis of RNA genome. Virology 134:375–389

Van Vloten-Doting L, Bol J (1988) Variability, mutant selection, and mutant stability in plant RNA viruses, vol. III. In: Domingo E, Holland JJ, Ahlquist P (eds) RNA genetics. CRC, Boca Raton FL, pp 37–51

Wetter C (1986) Tobacco mild green mosaic virus, vol 2. In: Van Regenmortel MHV, Fraenkel-Conrat H (eds) The plant viruses. Plenum, New York, pp 205–220

Wolfe MS, Caten CE (1987) Populations of plant pathogens: their dynamics and genetics. Blackwell, Oxford

Received July 25, 1990/Revised October 12, 1990