The rate and spectrum of spontaneous mutations in a plant RNA virus

Nicolas Tromas* and Santiago F. Elena*§

*Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, 46022 València, Spain, §The Santa Fe Institute, Santa Fe, NM 87501, USA

Running head: Mutation in TEV

Keywords: experimental evolution; mutant spectrum; mutation rate; Tobacco etch potyvirus; virus evolution

Corresponding author: Instituto de Biología molecular y Celular de Plantas, CSIC-UPV, Campus UPV CPI 8E, Ingeniero Fausto Elio s/n, 46022 València, Spain. Phone: +34 963 877 895; Fax: +34 963 877 859; E-mail: sfelena@ibmcp.upv.es.
Knowing mutation rates and the molecular spectrum of spontaneous mutations is important to understand how the genetic composition of viral populations evolves. Previous studies have shown that the rate of spontaneous mutations for RNA viruses widely varies between 0.01 and 2 mutations per genome and generation, with plant RNA viruses always occupying the lower side of this range. However, this peculiarity of plant RNA viruses is based in a very limited number of studies. Here we analyze the spontaneous mutational spectrum and the mutation rate of *Tobacco etch potyvirus*, a model system of positive sense RNA viruses. Our experimental setup minimizes the action of purifying selection on the mutational spectrum thus giving a picture of what types of mutations are produced by the viral replicase. As expected for a neutral mutational target, we found that transitions and nonsynonymous (including a few stop codons and small deletions) mutations were the most abundant type. This spectrum was notably different from the one previously described for another plant virus. Next, using two different methods, we have estimated that the spontaneous mutation rate for this virus was in the range $10^{-6}$ - $10^{-5}$ mutations per site and generation. Our estimates are in the same biological ballpark than previous values reported for plant RNA viruses. This finding gives further support to the idea that plant RNA viruses may have lower mutation rates than their animal counterparts.
The rate of spontaneous mutation is a key parameter to understand the genetic structure of populations over time. Mutation represents the primary source of genetic variation on which natural selection and genetic drift operate. Although the exact value of mutation rate is important for several evolutionary theories, yet accurate estimates are only available for a handful of organisms. RNA viruses show mutation rates that are orders of magnitude higher than those of their DNA-based hosts and in the range of 0.03 to 2 per genome and replication round (Drake et al. 1998; Drake and Holland 1999; Chao et al. 2002). This difference results from the lack of proofreading activity of the virus-encoded RNA-dependent RNA-polymerases (Steinhauer et al. 1992). The evolutionary causes of such elevated mutation rates remain unknown and it is commonly accepted that they may be beneficial as a mechanism to escape from the strong selective pressures imposed by the host’s defense mechanisms, thought not necessarily evolved in response to natural selection (Elena and Sanjuán 2005; Clune et al. 2008). Indeed, in the short term, a too high mutation rate has pernicious effects on viral fitness since most of the mutations produced are deleterious (Bonhoeffer et al. 2004; Sanjuán et al. 2004).

In the case of plant RNA viruses, it has been repeatedly reported that their populations are highly genetically stable (Rodríguez-Cerezo et al. 1991; Fraile et al. 1997; Marco and Aranda 2005; Herránz et al. 2008) in comparison with their animal counterparts, although reports of higher substitution rates also exist (Fargette et al. 2008; Gibbs et al. 2008). This peculiar behavior might be due in part to stronger stabilizing selection, weaker immune-mediated positive selection (García-Arenal et al. 2001), the existence of strong bottlenecks during
cell-to-cell movement and systemic colonization of distal tissues (Hall et al. 2001; Sacristán et al. 2003; Li and Roossinck 2004), severe bottlenecks during vector-mediated transmission (Ali et al. 2006; Mouri et al. 2007; Betancourt et al. 2008), or differences in the replication mode compared to lytic animal viruses (French and Stenger 2003; Sardanyés et al. 2009). Another more obvious possibility is that, indeed, plant viruses have lower mutation rates than other RNA viruses. Indeed the only two available direct estimates of mutation rates for plant viruses are both in the lower side of the range usually accepted for animal riboviruses: 0.10 - 0.13 per genome and generation for Tobacco mosaic virus (TMV) (Malpica et al. 2002) and 0.28 for Tobacco etch virus (TEV) (Sanjuán et al. 2009). However, none of these estimates is perfect. Although in the TMV experiments particular care was taken to measure mutation rate in a long target protected from the action of purifying selection (hence deleterious mutations remain in the population), uncertainties exist related to the number of infection cycles elapsed during the mutation-accumulation phase and the fraction of mutations that produced a selectable phenotype. In the case of TEV, the estimate should be taken as an upper limit because selection was operating during the mutation-accumulation phase. Furthermore, the estimate is in the same order of magnitude than the methodological error.

To further evaluate whether plant RNA viruses show unusually low mutation rates, we have developed a new empirical method that allows estimating the mutation rate and the spectrum of spontaneous mutations produced during an in vivo infectious process. The viral model system chosen for this experiment has been TEV (family Potyviridae, genus Potyvirus), a
prototypical example of positive sense RNA virus that has also become a model for virus experimental evolution. The method is based in the analysis of the temporal accumulation of mutations in a 1536 nt long neutral viral target. TEV genome size is 9539 nt long (GeneBank DQ986288) and encodes a large polyprotein of 346 kDa that self-processes into at least nine mature proteins. One of these proteins, the nuclear inclusion protein b (NIb) has RNA-dependent RNA-polymerase activity (URCUQUI-INCHIMA et al. 2001). This protein forms inclusions in the nucleus of infected plants and is required in the cytoplasm for replication complexes during viral RNA synthesis. NIb is the only protein that can be provided functionally in trans (LI and CARRINGTON 1995). Taking advantage of this property, we infected Nicotiana tabacum transgenic plants expressing TEV NIb and followed the accumulation of mutations in the viral copy of NIb. This experimental system minimizes the effect of purifying selection on the virus-encoded NIb due to complementation by the transgene.

MATERIALS AND METHODS

Virus and plants: The pTEV7DA infectious clone (DOLJA et al. 1992) was used as source for TEV. A TEV genotype was produced that lacked the full replicase gene (ΔNIb) by inverse PCR using Pfu turbo DNA polymerase (Stratagene) and primers conserving the proteolytic NIa-NIb and NIb-CP sites (5’-TTGCGAGTACACCCAATTCACTCATGAGTTGAGTGCCTCCTT-3’ and 5’-AGTGGCACTGTGGGTGGCTGGTAAGAAGAAA-3’, respectively). The resulting clone was named pTEV7DA-ΔNIb.
Two different genotypes of *N. tabacum* L. were used in these experiments, the wild-type tobacco var. Xanthi and the transgenic *Nt::NIb* line derived from var. Samsun by Li and CARRINGTON (1995). These transgenic plants express TEV NIb protein in a stable and functional manner. Prior to starting our experiments, the presence of the transgene was confirmed by PCR using Taq polymerase (Roche) and the primers F90-95 (5′-GCTGTATTGAAAGTGCGAC-3′ identical to bases 7767 - 7786 of TEV NIb) and R86-91 (5′-AGGCCCAACTCTCCGAAAG-3′ complementary to bases 8084 - 8102 of TEV NIb). The expression of the gene also was confirmed by RT-PCR. *Moloney murine leukemia virus* reverse transcriptase (MMLV RT) (Fermentas) was used to obtain cDNA from plants RNA extracts using primer R92-96 (5′-GCAAACTGCTCATGTGTGG-3′ complementary to bases 8761 - 8779 of CP gene). Then this cDNA was amplified using Taq and primers F90-95 and R86-91. Finally, the biological activity of the NIb protein encoded by the transgene was confirmed by inoculating batches of *Nt::NIb* plants with infectious RNAs from both viruses. All *Nt::NIb* plants inoculated with either TEV (n = 20) or TEV-ΔNIb (n = 10) developed a systemic infection after 6 - 7 days post inoculation (dpi). By contrast, none of the wild-type plants inoculated with TEV-ΔNIb (n = 5) became infected, while all plants inoculated with TEV (n = 5) were so. Furthermore, these results confirm that the presence of any putative RNA secondary folding structure within the NIb coding sequence is necessary for completing the infectious cycle of the virus.

**Experimental procedure:** Infectious plasmid pTEV7DA was linearized with BglII (Takara) and transcribed into 5’-capped RNAs using the SP6
mMESSAGE mMACHINE kit (Ambion Inc). Transcripts were precipitated (1.5 volumes of DEPC-treated water, 1.5 volumes of 7.5 M LiCl, 50 mM EDTA), collected, and resuspended in DEPC-treated water (CARRASCO et al. 2007). RNA integrity was assessed by gel electrophoresis and its concentration spectrophotometrically determined using a Biophotometer (Eppendorf). Twenty 4-weeks old Nt::NIb plants were inoculated mechanically on the third true leaf with TEV transcripts (4 - 7 µg) and 10% of inoculation buffer (100 mg/mL carborundum, 0.5 M K$_2$HPO$_4$, 3% PEG8000, pH = 7). In all cases, first symptoms appeared 6-7 days dpi.

Total RNA was extracted using RNeasy Plant Mini Kit (Quiagen) from symptomatic leaves of 3 Nt::NIb plants at 5, 10, 15, 20, 25, and 60 dpi. One of the plants at 20 dpi was not sampled because it dried out. The full NIb gene was reverse-transcribed using MMLV RT and primer R92-96 and PCR-amplified using the high fidelity PrimeSTAR HS DNA polymerase (Takara Bio Inc) and primers F73-80 (5’-TCATTACAAACAAGCACTTG-3’ identical to bases 6377 - 6396 of TEV Nla gene) and R92-96. By using this pair of primers we ensure that the mRNA from the transgene is not amplified and only NIb sequences from viral genomes will be so. PCR products of 2403 pb were gel purified with Zymoclean (Zymo Research), cloned into the plasmid pUC19/Smal (Fermentas) and used to transform Escherichia coli DH5α. At least 25 clones per plant were purified and sent out for sequencing by GenoScreen (www.genoscreen.fr) using BIGDYE 3.1 and a 96-capillars ABI3730XL sequencing system (Applied Biosystems). The following five internal primers were used for fully sequencing NIb with overlapping readouts: F1 5’-
GCAAACCTGAAGAGCCTTTTCAG-3'; F2 5’-
GCATGCTCATCACAAGCTCAAG-3'; F3 5’-
GTGGATGATTCAACAATCAATTTTATGAT-3'; F4 5’-
ACCAGCGTCAACACCAGCAC-3'; F5 5’-
GATCTGTCCCATTCCAAATAGAAAC-3’.

Contigs were assembled using GENEIOUS version 4.7 (www.geneious.com). The number of clones that rendered useful sequences was 472 (instead of the 500 submitted for sequencing). The number of sequenced clones per plant ranged between 12 and 34, with a median value of 24.

**Mutation rate estimations:** Two different approaches have been used to estimate TEV mutation rate. In the first approach, we proceeded as follows.

For a given plant the number of clones sequenced that contained zero, one, two, ..., $k$ mutations was fitted to a Poisson distribution with parameter $\lambda = \mu l T$, where $\lambda$ is the expected number of mutations per clone, $T$ the number of generations of viral replication, $l = 1536$ the length of the amplicon, and $\mu$ the mutation rate per base and per generation (m/b/g). Defining generations *in vivo* in plant viral populations is troublesome, given that a viral population colonizing a plant is not replicating synchronously but with overlapping generations. A good approximation is to define viral generations as the number of cycles of cell infections (Malpica *et al.* 2002). For this definition to be operative, it is necessary first to have an estimate of the average number of viruses produced per infected cell. By performing one-step accumulation curves in tobacco protoplasts, F. Martínez, S. F. Elena and J. A. Daròs (unpublished results) have estimated that, on average, an infected cell yields

---

8
1555 genomes (quantified by real-time quantitative RT-PCR). To estimate the number of generations experienced by TEV at the time points where the samples were taken, we revisited previously published data on the kinetics of TEV accumulation (CARRASCO et al. 2007). Reanalyzing these data, we found that the model that better describes TEV accumulation within an infected plant was a 4-parameters Gompertz growth equation \((R^2 = 0.975)\) (CAMPBELL and MADDEN 1990). From the parameters of the model and using the above estimate of virus yield per cell, it is possible to calculate that during the exponential growth phase, the viral population experienced 3.156 generations per day, but this number reduces as growth rate flats off and the carrying capacity of the system is reached. After estimating the number of generations corresponding to each sampling day, it is then possible to transform the above per clone mutation rate values into the biologically meaningful scale of mutations per base and per generation using the simple expression \(\mu = \lambda/IT\). Each plant has been treated as an independent replicate, rendering 19 estimates of \(\mu\).

For the second approach, we focused only on putatively lethal mutations, that is, mutations generating frameships or stop codons. Readers need to recall that the only ORF encoded by TEV genome is translated into a single polyprotein. Our method is based on the fact that amino acid substitutions affecting NIb would in turn be neutral because the \textit{trans} complementation provided by the host (and the best evidence of such active \textit{trans} complementation is the ability of TEV-\DeltaNIb to infect \textit{Nt::NIb} plants). However, frameship mutations and stop codons affecting the NIb sequence would be
lethal because they will produce a virus deficient not only in N1b but also in CP, the gene downstream from N1b, which is not complemented by the host. In haploid populations at the mutation-selection balance, the frequency of deleterious mutations, $p$, is given by $p = \mu / s$, where $s$ is the selection coefficient. For lethal mutations, however, $s = 1$, then $\mu_L = p$ and the equilibrium is reached instantaneously because all lethal mutations have been generated in the previous generation (Crow and Kimura 1970). In other words, this method provides an estimate of mutation rate per replication event ($m/b/r$) rather than by generation, as in the first method. Following Cuevas et al. (2009), it is possible to calculate a mutation rate for the $i$th amplicon using the expression

$$\mu_{\text{NSTMT},i} = \frac{1}{n} \sum_{i=1}^{K} W_i,$$

(1)

where $n$ is the total number of nonsense mutational targets (NSMT: sites that can generate a stop codon after a single nucleotide substitution) in an amplicon, $W_i$ a weighting factor for the two types of nonsense mutations ($W_i = 3$ if only one of the three possible mutations in a NSMT produces a stop codon and $W_i = 1.5$ if two out of three possible produce a stop codon), and $K$ is the total number of observed nonsense mutations in the amplicon. According to the standard genetic code, there are 18 NSMT-containing codons and 19 different NSMTs (the UGG codon contains two). In our experiments, we have 472 independent estimates of $\mu_{\text{NSTMT}}$. If the frequency of insertions and deletions is $\mu_{\text{indel}}$ (it can be computed using the Poisson distribution, as described above), then $\mu_L$ can be estimated as
\[ \mu_L = \mu_{\text{indel}} + \frac{1}{472} \sum_{j=1}^{472} \mu_{\text{NSMT},j} \]  

(2)

where \( \mu_{\text{NSMT},j} \) is estimated using Equation 1.

Hereafter, we will use the notation \( \mu_L \) when referring to the estimated based on the frequency of lethals (units of m/b/r) and reserve the notation \( \mu \) for the Poisson estimate (units of m/b/g).

**Statistical analyses:** All statistical tests have been performed using SPSS version 16. All molecular evolutionary analyses were done using MEGA4 (Tamura et al. 2007).

**RESULTS AND DISCUSSION**

**Characterization of the mutant spectrum:** Table 1 summarizes the spectrum of mutations characterized for the 472 clones sequenced, which added up to 724992 nucleotides sequenced. A file containing the full sequence alignment is available upon request. Fifty-two mutations have been identified, 46 of which were nucleotide substitutions and six deletions. Not a single insertion has been observed. The number of mutations per amplicon distributed as follows: 427 amplicons had no mutation, 15 carried a single mutation and five had two mutations. This distribution does not depart from the expectation from the Poisson null model (Kolmogorov-Smirnov test, \( P = 1 \)). Among base substitutions, 33 were transitions and 13 transversions. Consistent with the principle that transitions are biochemically more likely to occur than
transversions, the maximum composite likelihood estimate of the overall transitions to transversions rates ratio was 2.161. This excess of transitions also occurs when only purines (4.262) or only pyrimidines (6.681) are considered. Indeed, the observed frequencies of transitions among purines and among pyrimidines are equally likely (Figure 1) and are by far the most frequent type of mutation (Table 1, Figure 1). Therefore, we can conclude that TEV N1b polymerase spontaneously produces, on average, two-thirds transitions and one-third transversions. Cuevas et al. (2009) have reported a similar result for hepatitis C virus. If purifying selection would not be canceled out by N1b trans complementation, this ratio would be surely more biased towards transitions, since they are more often silent than transversions.

Under the observed mutational spectrum, the equilibrium base-composition achieved only by mutation (in the absence of purifying selection) would be 31.1% A, 25.4% U, 17.8% C, and 25.7% G. This distribution significantly deviates from what is expected just by shear chance (\( \chi^2 = 55.505, 3 \) d.f., \( P < 0.001 \)). The deviation is mainly driven by the unbalanced composition in purines, with a large excess of A (24.5%) that compensates for the large defect in G (−28.9%).

We have observed that 16 mutations were synonymous and 30 were nonsynonymous. Two of the nonsynonymous mutations generated a stop codon. At least eight of the nonsynonymous substitutions could induce a major deformation on N1b folding by replacing polar or charged side chains by apolar ones (E20G, Q462P, H355L, and E507A) or apolar side chains by polar ones (F106S, G200S and W417R). Three substitutions (L143P, D146H and D276Y)
lead to a strong change in the length of the size chain. Additionally nine substitutions (D248N, A270V, D276Y, R283Q, I302L, D348N, H355L, T381I, and Q387Stop) may be affecting the putative active site of Nlb (PFAM00680). Among the deletions, three cases involved single nucleotide eliminations and in only one case three contiguous nucleotides were deleted. To evaluate whether this pattern of synonymous and nonsynonymous changes in Nlb is compatible with a model of neutral evolution, we have estimated the difference between substitution rates per nonsynonymous ($d_N$) and synonymous ($d_S$) sites (using Nei-Gojobori’s modified method and bootstrap SEM). If a gene accumulates changes in a neutral manner, then $E(d_N - d_S) = 0$. The observed value of $d_N - d_S = (5.537 \pm 4.133) \times 10^{-4}$ is not significantly different from zero ($z = 1.340$, $P = 0.090$), failing to reject the null hypothesis of neutral evolution.

In conclusion, the pattern of nucleotide substitutions observed in the Nlb sequence is consistent with the expectation for a neutral target, validating our methodology for protecting a viral sequence from purifying selection.

**Comparison of TEV mutant spectrum with that observed for other plant viruses:** TEV spontaneous mutational spectrum differs in several aspects from the other only one reported for plant viruses, TMV (MALPICA et al. 2002). First, TMV mutational spectrum is dominated by insertions and deletions (69% of all mutations belong to these categories). Deletions were both short (five cases with 1 – 3 nt deleted) and long (seven cases with up to 100 nt deleted). Insertions were also short (1 nt) and large (4 cases with poly(A) insertions). In sharp contrast, only 9.8% of mutations in TEV mutational spectrum were short deletions, and not a single insertion has been observed. This difference is
largely significant (Fisher’s exact test, \( P < 0.001 \)) and it suggests either that TEV N1b replicase is more processive \textit{in vivo} than TMV replicase or that the difference is due to the experimental setup. In this regard, \textsc{Malpica et al.} (2002) used the MP protein expressed in \textit{trans} as target for measuring mutation rate on the viral copy of MP. However, MP has a positive regulatory effect on the formation of TMV replication complex (\textsc{Beachy and Heinlein} 2000) and, hence, may favor template switching and a higher rate of deletions and insertions. By contrast, in our experiments N1b is expressed in excess concentration from the transgene and, thus, N1b molecules may remain attached to the RNA molecules reducing the likelihood of template switching.

The second noticeable difference between both mutational spectra refers to the ratio of synonymous to nonsynonymous substitutions. For TMV the ratio is 1:10, whereas for TEV it is about five times larger (16:30). This difference may simply reflect that the method employed by \textsc{Malpica et al.} (2002) was less efficient than our method to protect deleterious point mutations from purifying selection, although this explanation is unsatisfactory given the large amount of deletions maintained in TMV populations. However, to honor the truth, it is worth saying that this 5-fold difference was not statistically significant (Fisher’s exact test, \( P = 0.146 \)).

A third difference is that the ratio of transitions to transversions was roughly 1.0 for TMV whereas it was > 2.0 for TEV. Given that it is biochemically easier to produce transitions than transversions, the deficit of the former type observed for TMV may reflect a preference of its replicase for
transversions or, as MALPICA et al. (2002) honestly recognized, to a problem associated to their small sample size.

Finally, MALPICA et al. (2002) found striking the high frequency of mutant genotypes carrying multiple mutations. The distribution of mutations per mutant TMV amplicon had a median of one and a range of 1 – 3. In our case, the distribution had also a median of one and a range of 1 – 2. From a statistical point of view, both distributions are undistinguishable in shape (Kolmogorov-Smirnov test, $P = 0.199$) and location (Mann-Whitney test, $P = 0.929$). Therefore, we would not consider striking at all finding a minor proportion of amplicons carrying more than one mutation: it is just what is expected for the mutation rates estimated below.

**Estimates of the mutation rate:** Applying the first method described in the Material and Methods section, that is, counting the number of mutations per amplicon per plant, we have obtained 19 independent estimates of the spontaneous mutation rate. The estimates ranged from $0 \leq \mu \leq 1.340 \times 10^{-5}$ m/b/g. The distribution of estimates was Gaussian (Kolmogorov-Smirnov test, $P = 0.944$) with mean $\mu = 4.754 \times 10^{-6}$ m/b/g and standard deviation $\sigma = 3.540 \times 10^{-6}$ m/b/g. This average value was significantly greater than zero (one sample $t$-test: $t_{18} = 5.854$, 1-tailed $P < 0.001$). As a way to evaluate the statistical power associated with this test, we constructed the 95% confidence interval around the mean as $3.048 \times 10^{-6} \leq \mu \leq 6.460 \times 10^{-6}$ m/b/g, which excludes the zero. Therefore, according to these values, we conclude that the genomic mutation rate of TEV is $0.045 \pm 0.008$ (SEM) per generation.
Next, we sought for applying the lethal alleles method to obtain a second estimate of the mutation rate that is not affected by our assumptions about generation time. To compute the first term in Equation 2, we proceeded as above and fitted the observed number of deletions per amplicon per plant to a Poisson distribution, obtaining 19 independent estimates of $\mu_{\text{indel}}$. The average rate of deletion mutations was $\mu_{\text{indel}} = (3.787 \pm 1.558) \times 10^{-7}$ deletions/b/r, a value that was significantly greater than zero (one sample $t$-test: $t_{18} = 2.431$, 1-tailed $P = 0.013$). Next we focused in the computation of the second term in Equation 2, the number of nonsense mutations observed. Only two out of the ∼725 Kb sequenced were stop codons (hence $K = 2$ in Equation 2). As a consequence of codon usage bias, the actual number of NSMT in our sample is 7.46% instead of the expected 10.34%. Taking this source of bias into consideration and after correcting for the three possible nucleotide substitutions per site (see Material and Methods), the second term in Equation 2 results in $(6.295 \pm 0.556) \times 10^{-5}$ m/NSMT/r. Therefore, the estimate of the spontaneous lethal mutation rate is $\mu_L = (6.299 \pm 0.558) \times 10^{-5}$ m/b/r or, expressed into the per genome scale, $0.601 \pm 0.053$ per replication event.

This $\mu_L$ value is 13.356 times higher than the $\mu$ estimate obtained using the first method, being the difference highly significant (2-samples $t$-test, $t_{36} = 10.328$, $P < 0.001$). Therefore, we conclude that both methods produced very dissimilar estimates of the in vivo genomic mutation rate for TEV. What may produce this discrepancy between both estimates? The lethality method has the advantage of being independent from generation time. However, it is strongly dependent on whether the mutations considered are truly lethal. Deviations
from this assumption imply that the estimate immediately becomes an upper-limit of the true value. In infected cells wherein multiple genomes may coexist, genomes carrying deletions or stop codons (e.g., putative lethals) can still be replicated by the pool of polymerases, encapsidated into wild-type capsides and moved cell-to-cell and even systemically. In other words, complementation with functional proteins makes lethal mutations behave as effectively neutral ones and thus they can increase frequency in the population; they will be effectively lethal only when infecting a cell alone. An alternative consideration is that, as defined above, one generation involves many replication rounds. Assuming that $\mu_L$ has not been biased by complementation, the 13.356 fold difference between estimates can be interpreted as the number of replication events within an infected cell. Nonetheless, we can conservatively conclude that the above $\mu_L$ estimate must be taken as an upper-limit estimate of the true mutation rate: $\mu \leq \mu_L$.

**Comparison of TEV mutation rate with those obtained for other RNA viruses:** The only previous direct estimate of mutation rate for another plant virus, TMV, was in the range $1.452 \times 10^{-5} - 2.060 \times 10^{-5}$ m/b/g (MALPICA et al. 2002), values lying well within our two estimates. In a recent study, SANJUÁN et al. (2009) estimated TEV upper-limit mutation rate as $(2.96 \pm 0.32) \times 10^{-5}$ m/b/g, a value also within our both estimates and that, as expected, it was an overestimation of the actual mutation rate. In the same study, these authors performed a literature survey for upper-limit estimates of per site mutation rates for four plant viruses. All the compiled studies were methodologically similar and relied on the characterization of the mutant spectrum from
individual plants inoculated with a viral clone (i.e., close to zero starting genetic diversity). In neither of these studies was genetic variation protected from purifying selection (Sanjuán et al. 2009). The median upper-limit mutation rate estimated was $7.74 \times 10^{-4} \text{ m/b/g}$, which was in the range of values estimated for animal RNA viruses and some RNA bacteriophages (Drake and Holland 1999) but still 12.29-fold larger than our upper-limit estimate.

Our data allow us to conclude that the mutation rate of TEV is slightly lower than previously estimated by Sanjuán et al. (2009) and very similar to the only other direct estimation available for another RNA plant virus, TMV (Malpica et al. 2002). All these estimates are within a narrow range of values in the lower side of estimates reported for RNA animal viruses and bacteriophages. This agreement may suggest that plant RNA viruses have lower mutation rates than their animal counterparts. Indeed, this difference in mutation rates may help to partially explain why the rates of molecular evolution of most RNA plant viruses are apparently lower than those observed for RNA animal viruses (Rodríguez-Cerezo et al. 1991; Fraile et al. 1997; Marco and Aranda 2005; Herránz et al. 2008). This difference between animal and plant RNA viruses raises an intriguing question: given that plant and animal RNA viruses do not form separated phylogenetic groups and that they are basically replicated by similar polymerases, why plant RNA viruses show significantly lower mutation rates? We can imagine several scenarios to explain this difference. First, obviously, this may not be the rule and just by chance the two plant viruses for which mutation has been estimated turn out to have polymerases of particularly good fidelity. A second obvious possibility is
that most values for animal RNA viruses are, actually, upper-limit estimations. In this sense, it has been reported that Yellow fever virus polymerase has an error rate as low as $1.9 \times 10^{-7}$ m/b/g (Pugachev et al. 2002). Third, the difference is real and results from differences in the selective pressures that modulated the evolution of mutation rates in both types of hosts. This implies that mutation rate has been tuned by natural selection higher in animal RNA viruses than in their plant relatives because the former represent a more stressful environment, perhaps in the form of more diverse cell types or stronger antiviral responses (e.g., the adaptive immune system; Kamp et al. 2003). However, whether virus’ mutation rates have been optimized by natural selection or are byproducts of a parasitic fast lifestyle still needs to be confirmed (Elena and Sanjuán 2005).

Pita et al. (2007) have suggested a dependency of mutation rate on the host where the virus is replicating. We have measured TEV mutation rate in its natural host, where we expect that selection may have optimized it to minimize the impact of deleterious mutations. An open an interesting question is to confirm that the estimate of mutation rate for TEV would be higher in a different host.

Potential pitfalls and considerations: In this study we have used a high-fidelity DNA polymerase to minimize the probability that observed mutations may be due to PCR errors. According to the information provided by the manufacturer, the PrimeSTAR HS DNA polymerase is about 2 times more accurate than Pfu due to its improved and robust 3’ → 5’ exonuclease activity and its error rate has been estimated to be $1.60 \times 10^{-6}$ m/b/PCR cycle (catalog.takara-bio.co.jp). Since we run PCRs for 30 cycles, we expect an error
rate per amplicon of $30 \times 1.6 \times 10^6 = 4.8 \times 10^5$ m/b. Henceforth, we may expect $724992 \times 4.8 \times 10^5 = 34$ mutations in our sample to be due to errors during PCR. Unfortunately, this is not the only source of error; the error rate of MMLV RT is around $3.3 \times 10^5$ m/b/r (AREZI and HOGREFE 2007), which means that we may expect as well $724992 \times 3.3 \times 10^5 = 24$ mutations to be produced during retrotranscription. Since we have obtained 52 mutations, someone may argue that all of them must result from errors during either retrotranscription or PCR amplification (actually, we are observing less mutations than expected by methodological errors). This being the case, the mutation rate of TEV would be $< 10^{-9}$ m/b/g, a value that is, by all means, absurdly low and, thus rejectable. Furthermore, the estimate of the error rate of PrimeSTAR HS polymerase should be taken with strong precaution. It is surprising the manufacturer claims that the enzyme has improved fidelity compared with Pfu but the estimate they provide is indistinguishable from values reported for Pfu, $1.3 \times 10^6$ m/b/PCR cycle (CLINE et al. 1996; BRACHO et al. 1998). Therefore, we can conclude that even if (some) unwanted mutations are produced during the RT-PCR amplification, the estimated mutation rates are still on the low side of previous reports.

We thank Prof. J. C. Carrington (Oregon State University) for kindly providing the *N. tabacum* NLb transgenic plants and the pTEV7DA clon, Dr. J. A. Darós, Dr. R. Sanjuán and our labmates for advice and discussion and F. de la Iglesia for excellent technical assistance. This study was supported by grants BFU2006-14819-C02-01/BMC and BFU2009-06993 from the Spanish Ministerio
de Ciencia e Innovación (MICINN). N. T. is supported by a FPI fellowship from MICINN.

LITERATURE CITED


# TABLE 1

**Numbers of mutations by type and observed substitution matrix**

<table>
<thead>
<tr>
<th>Type of mutation</th>
<th>Number</th>
<th>Substitution matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>51</td>
<td>A U G C</td>
</tr>
<tr>
<td>Base substitutions</td>
<td>46</td>
<td>A - 3 4 7</td>
</tr>
<tr>
<td>Transitions</td>
<td>33</td>
<td>U 3 - 7 0</td>
</tr>
<tr>
<td>Transversions</td>
<td>13</td>
<td>G 0 9 - 0</td>
</tr>
<tr>
<td>Synonymous</td>
<td>16</td>
<td>C 10 2 1 -</td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>30 (2 stops)</td>
<td></td>
</tr>
<tr>
<td>Deletions</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>1-nt</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>3-nt</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1.— Observed frequencies for the different types of nucleotides substitutions. Each column groups mutations rendering complementary pairs and, thus, can occur during the synthesis of the genomic or antigenomic strains. The LaPlace estimator of the frequency has been used to minimize the bias due to small sample size (Agresti and Coull 1998). Error bars represent the 95% confidence interval for the estimator.