Upper-limit mutation rate estimation for a plant RNA virus

Rafael Sanjuán1,†, Patricia Agudelo-Romero1,† and Santiago F. Elena1,2,*

1Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV; Campus UPV CPI SE, Ingeniero Fausto Elío s/n, 46022 Valencia, Spain
2The Santa Fe Institute, Santa Fe, NM 87501, USA
*Author for correspondence (sfelena@sbmep.upv.es).
†These authors contributed equally to this work.

It is generally accepted that mutation rates of RNA viruses are inherently high due to the lack of proofreading mechanisms. However, direct estimates of mutation rate are surprisingly scarce, in particular for plant viruses. Here, based on the analysis of in vivo mutation frequencies in tobacco etch virus, we calculate an upper-bound mutation rate estimation of $3 \times 10^{-7}$ per site and per round of replication; a value which turns out to be undistinguishable from the methodological error. Nonetheless, the value is barely on the lower side of the range accepted for RNA viruses, although in good agreement with the only direct estimate obtained for other plant viruses. These observations suggest that, perhaps, differences in the selective pressures operating during plant virus evolution may have driven their mutation rates towards values lower than those characteristic of other RNA viruses infecting bacteria or animals.

Keywords: experimental evolution; mutation rate; plant virus; virus evolution

1. INTRODUCTION

The rate of spontaneous mutation is a key evolutionary parameter, yet accurate estimations are available only for a handful of organisms. One of the most widely accepted differences in mutation rates is the one between RNA viruses and dsDNA organisms, the former mutating orders of magnitude more frequently than the latter. Such elevated mutation rates, though not necessarily evolved in response to natural selection (Elena & Sanjuán 2005; Clune et al. 2008), can explain why RNA virus show rapid evolution at the molecular level, although other factors such as replication speed, transmission mode or genonomic architecture have to be considered as well (Duffy et al. 2008). However, our current knowledge of RNA virus mutability is mainly based on work with animal viruses and bacteriophages (Elena et al. 2008).

High genetic stability has frequently been reported for plant RNA viruses (Rodriguez-Cerezo et al. 1991; Fraile et al. 1997; Marco & Aranda 2005; Herránz et al. 2008), although substitution rates within the range of animal RNA viruses have also been reported (Fargette et al. 2008; Gibbs et al. 2008). This peculiar behaviour might be due to stronger selective constraints operating on plant viruses, weaker immune-mediated positive selection (Garcia-Arenal et al. 2001), the existence of strong bottlenecks during cell-to-cell movement and systemic colonization of distal tissues (Hall et al. 2001a; Sacristán et al. 2003; Li & Roossink 2004), or differences in replication mode compared with lytic viruses (French & Stenger 2003). Another, more obvious, possibility is that plant viruses show reduced rates of spontaneous mutation. However, the only available direct mutation rate estimation for a plant virus is for the tobacco mosaic virus (TMV; Málpica et al. 2002). In this pioneer study, mutations at the viral movement protein (MP) gene were sampled from TMV populations infecting transgenic Nicotiana tabacum plants that expressed the MP protein, hence minimizing selection against TMV MP mutants. A mutation rate of $2.4 \times 10^{-5}$ mutations/base/replication round (m/b/r) was obtained, which lies on the lower side of the range usually accepted for animal riboviruses (Drake & Holland 1999), although this estimate was subject to uncertainties related to the number of infection cycles elapsed during the mutation-accumulation phase and the fraction of mutations that produced a selectable phenotype.

To further investigate whether plant viruses show unusually low rates of mutation, we followed a strategy that allowed us to obtain an upper-bound estimation for the mutation rate, based on the quantification of genetic variability in tobacco etch virus (TEV), a positive-stranded RNA virus belonging to the Potyviridae family. Then, we compared this estimate with those available in the literature.

2. MATERIAL AND METHODS

(a) Viral populations

Chenopodium quinoa leaves were inoculated with a TEV extract and six necrotic local lesions were isolated. This procedure (the equivalent of plaque-picking in lytic viruses) introduces a severe bottleneck in the viral population, effectively eliminating genetic variation (de la Iglesia & Elena 2007). Each of these isolates was used to infect a different four-week old N. tabacum plant and, following the appearance of systemic symptoms, virions were extracted from whole plants as previously described (Carrasco et al. 2007). Any mutation detected would necessarily be generated after the bottleneck.

(b) Molecular methods

The region encompassing nucleotides 7808–9437, which corresponds to the 5’ end of the viral polyomerase gene N1b, the entire capsid protein gene and the proximal portion of the 3’ untranslated region, was sequenced. Moloney murine leukaemia virus (MMLV) reverse transcriptase (RT) (Fermentas) was used to obtain cDNA from the viral extract and the high-fidelity Phusion DNA polymerase (Finnzymes) was used to obtain a PCR fragment, which was subsequently gel-purified, ligated to a pUC19/SmaI (Fermentas) and used for transformation of Escherichia coli DH5α. Plasmid DNA was purified and sequenced using the ABI prism BigDye terminator cycle sequencing kit v. 3.1 (Applied Biosystems), vector-based primers and internal primers. Labelled products were resolved in an ABI 3100 genetic analyser (Applied Biosystems).

(c) Estimating mutation rate from mutant frequencies

Following Drake & Holland (1999), the frequency of mutations per base after c cycles of cell infection is given by $f = 2^mc$, where $m$ is the fraction of mutations observed in a randomly sampled population of size $c$. The fraction of mutations $f$ that fall within $c$ was calculated as $f = 2^mc$, where $m$ is the fraction of mutations observed in a randomly sampled population of size $c$.
the mutation rate per base and copying event and ‘2’ reflects the fact that offspring genomic strands are produced via the synthesis of intermediary antigenomic strains. For lethal mutations, \( c = 1 \) and thus a direct estimation of \( \mu \) can be obtained. However, since \( c \) is unknown in our experiment, \( f \) can only be used to obtain an upper-limit estimator of the real mutation rate (\( \mu (f) = \mu_{\text{max}} \)). We estimate \( \mu_{\text{max}} \) in samples from TEV-infected plants by high-fidelity RPF-PCR, molecular cloning and sequencing.

### 3. RESULTS AND DISCUSSION

Out of 83 clones and a total of 135 207 nucleotides sequenced, only four mutations were observed, yielding a mutation rate (\( \mu_{\text{max}} \)) of \( \mu_{\text{max}} = (2.96 \pm 0.32) \times 10^{-5} \) m/b/r. Two mutations were found in the \( Nlb \) gene, one synonymous (C8029U) and the other (C8248U) with a probable major effect on the protein folding, since it replaces P2750 by S, an amino acid with a small polar radical. Two other non-synonymous mutations were found in the CP gene, C8980A (Q2994K) and C9008A (A3003D). The first mutation replaced a positively charged side chain by another one with similar characteristics. The second one may have a stronger effect on fitness since it replaced a small apolar radical by a larger side chain with net negative charge. Assuming that only P2750S and A3003D may be lethal for the virus, then the estimated mutation rate would be \( \mu = 1.48 \times 10^{-5} \) m/b/r. This value is in good agreement with the indirect estimate reported by Carrasco et al. (2007) for the deleterious genomic mutation rate.

The use of a high-fidelity DNA polymerase minimizes the probability that the observed mutations were due to PCR errors, but reverse transcription artefacts are more likely since the error rate of MMLV RT is probably higher than \( 3.3 \times 10^{-3} \) m/b/r (Arczi & Hogrefe 2007). Therefore, we can only safely conclude that the obtained \( \mu_{\text{max}} \) at least for the sequenced region, in the particular host employed and experimental conditions, is as low as that for a common RT enzyme.

How do our data compare with the levels of genetic variability reported for other plant RNA viruses? Table 1 compiles \( \mu_{\text{max}} \) for four different viruses on different hosts. These values were in all cases estimated after characterizing the mutant spectrum from individual plants. The median \( \mu_{\text{max}} \) across viruses and experiments is \( 4.74 \times 10^{-4} \) m/b/r, which is in the range of estimated rates for animal RNA viruses and some RNA bacteriophages (Drake & Holland 1999). It is worth noting that estimates are homogeneous for the five different viruses (two-way non-parametric ANOVA: \( H = 1.140, d.f. = 2, p = 0.566 \)), although they significantly vary across host plants (\( H = 22.424, d.f. = 9, p = 0.008 \)), suggesting that virus mutation rates depend on the environment in which they replicate, as suggested by Pita et al. (2007).

The approximately 16-fold lower \( \mu_{\text{max}} \) obtained for TEV compared with the above median value can be attributed to several factors. It is possible that TEV shows an unusually low rate of spontaneous mutation or that tobacco plants are hosts in which mutation rate is particularly low. It must also be noted that the region we have sequenced contains \( cis \)-acting secondary structure motifs necessary for genome replication (Haldeman-Cahill et al. 1998), and mutations affecting these motifs render replication-incompetent viruses. Therefore, our \( \mu_{\text{max}} \) value might be closer to the actual mutation rate than \( \mu_{\text{max}} \) values obtained in studies where a less-constrained region was sequenced. Hence, most of the genetic variation sampled in the studies are listed in table 1, and in previous work with animal RNA viruses and bacteriophages might correspond to neutral or nearly neutral mutations that had been segregating in the population for long periods of time, thus reaching relatively high frequencies, or that some of these mutations were beneficial. In our experiments, by contrast, long-term variation could not be present because infections were initiated from a single lesion-forming unit. Furthermore, the virus used in our experiments had been replicating in tobacco plants for 30 passages prior to the assays without showing significant changes in several fitness traits (data not shown), suggesting that the mutations sampled were probably not beneficial.

### Table 1. Upper-limit estimate for the mutation rate (\( \mu_{\text{max}} \)) for several RNA plant viruses.

<table>
<thead>
<tr>
<th>Virus</th>
<th>Host</th>
<th>( \mu_{\text{max}} ) (( \pm ) s.e.m.) \times 10^{-4}</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMV</td>
<td>Capsicum annuum</td>
<td>15.34±0.71</td>
<td>Schneider &amp; Roossinck (2001)</td>
</tr>
<tr>
<td></td>
<td>C. annuum</td>
<td>1.39±0.07</td>
<td>Pita et al. (2007)</td>
</tr>
<tr>
<td></td>
<td>Nicotiana benthamiana</td>
<td>6.44±0.95</td>
<td>Schneider &amp; Roossinck (2000)</td>
</tr>
<tr>
<td></td>
<td>Nicotiana tabacum</td>
<td>0.20±0.09</td>
<td>Pita et al. (2007)</td>
</tr>
<tr>
<td>CCMV</td>
<td>N. benthamiana</td>
<td>5.29±4.93</td>
<td>Schneider &amp; Roossinck (2000)</td>
</tr>
<tr>
<td>TMV</td>
<td>C. annuum</td>
<td>11.02±0.12</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Collinsia heterophylla</td>
<td>4.74</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Fragopyrum esculentum</td>
<td>4.55</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Lycopersicum esculentum</td>
<td>1.45±0.51</td>
<td>Schneider &amp; Roossinck (2001)</td>
</tr>
<tr>
<td></td>
<td>N. benthamiana</td>
<td>4.21±0.69</td>
<td>Schneider &amp; Roossinck (2000)</td>
</tr>
<tr>
<td></td>
<td>N. tabacum</td>
<td>4.14</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>N. tabacum</td>
<td>0.24±0.00</td>
<td>Malpica et al. (2002)</td>
</tr>
<tr>
<td></td>
<td>Phacelia campanularia</td>
<td>16.81</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Plantago sp.</td>
<td>8.50</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Solanum nigrum</td>
<td>4.21</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td></td>
<td>Tagetes erecta</td>
<td>8.15</td>
<td>Kearney et al. (1999)</td>
</tr>
<tr>
<td>WSMV</td>
<td>Zea mays</td>
<td>9.01±0.90</td>
<td>Hall et al. (2001b)</td>
</tr>
</tbody>
</table>

\*CMV, cucumber mosaic virus; CCMV, cowpea chlorotic mottle virus; TMV, tobacco mosaic virus; WSMV, wheat streak mosaic virus.

\*This study reported a \( \mu \) estimation instead of the \( \mu_{\text{max}} \) upper limit.
Our data allow us to conclude that the mutation rate of TEV is in the lower range of those reported for other RNA viruses, but, obviously, we cannot conclude that plant RNA viruses show low mutation rates in general. However, the fact that our estimate is very close to the only previous estimate available for a plant RNA virus (Malpica et al. 2002) at least suggests so. Finally, slow rates of molecular evolution in plant viruses could be partly explained by low mutation rates, but the complex relationships between mutation and evolution need to be considered in detail to explain why some plant viruses show unusually high levels of genetic stability while others evolve fast.

We thank Francisca de la Iglesia for technical assistance. This work was supported by grant BFU2008-03978/BMC from the Spanish MICINN. R.S. also acknowledges current support from MICINN grant BFU2008-03978/BMC and the Ramón y Cajal programme.


