

1 **Contents Category:** Plant Viruses – RNA

2

3 **Estimation of the *in vivo* recombination rate for a plant RNA**

4 **virus**

5

6 **Nicolas Tromas,<sup>1</sup> Mark P. Zwart,<sup>1</sup> Maité Poulain,<sup>2</sup> and Santiago F. Elena<sup>1,3</sup>**

7

8 <sup>1</sup>Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de  
9 Investigaciones Científicas-UPV, 46022 Valencia, Spain

10 <sup>2</sup>Genoscreen, 1 Rue du Professeur Calmette, 59000 Lille, France

11 <sup>3</sup>The Santa Fe Institute, Santa Fe, New Mexico 87501, USA

12

13 **Running title:** TEV recombination rate

14 **Keywords:** experimental evolution; plant virus; *Tobacco etch potyvirus*; virus  
15 evolution; virus replication

16 **Word count:** Summary: 214. Full manuscript: 5099. Number of figures: 2. Number  
17 of tables: 3.

18

19 **Correspondence:**

20 Prof. Santiago F. Elena, E-mail: [santiago.elena@csic.es](mailto:santiago.elena@csic.es)

21 Dr. Nicolas Tromas, E-mail: [ntromas@ibmcp.upv.es](mailto:ntromas@ibmcp.upv.es)

22 IBMCP (CSIC-UPV), Campus UPV CPI 8E, Ingeniero Fausto Elio s/n, 46022  
23 Valencia, Spain; Phone: +34 963 877 895; Fax: +34 963 877 859.

24

25 **ABSTRACT**

26 Phylogenomic evidence suggests that recombination is an important evolutionary force  
27 for potyviruses, one of the larger families of plant RNA viruses. However, mixed-  
28 genotype potyvirus infections are marked by low levels of cellular coinfection,  
29 precluding template switching and recombination events between virus genotypes  
30 during genomic RNA replication. To reconcile these conflicting observations, we have  
31 evaluated the *in vivo* recombination rate ( $r_g$ ) of *Tobacco etch potyvirus* by coinfecting  
32 plants with pairs of genotypes marked with engineered restriction sites as neutral  
33 markers. The recombination rate was then estimated using two different approaches: (i)  
34 a classical approach that assumes recombination between marked genotypes can occur  
35 in the whole virus population, rendering an estimate of  $r_g = 7.762 \times 10^{-8}$  events per  
36 nucleotide site per generation; (ii) an alternative method that assumes recombination  
37 between marked genotypes can occur only in coinfecting cells, rendering a much higher  
38 estimate of  $r_g = 3.427 \times 10^{-5}$  events per nucleotide site per generation. This last estimate  
39 is similar to the TEV mutation rate, suggesting that recombination should be at least as  
40 important as point mutation in creating variability. Finally, we compared our mutation  
41 and recombination rate estimates to those reported for animal RNA viruses. Our  
42 analysis suggests that high recombination rates may be an unavoidable consequence of  
43 selection for fast replication at the cost of low fidelity.

44

## 45 INTRODUCTION

46 Three different mechanisms are ultimately responsible for the observed high genetic  
47 variability of plant RNA viruses: mutation, recombination and segment reassortment.  
48 Both the roles of mutation (Aranda *et al.*, 1997; Malpica *et al.*, 2002; Sanjuán *et al.*,  
49 2009; Tromas & Elena, 2010) and recombination (Aranda *et al.*, 1997; Bonnet *et al.*,  
50 2005; Chen *et al.*, 2002; Fernández-Cuartero *et al.*, 1994; Froissart *et al.*, 2005; Martín  
51 *et al.*, 2009; van der Walt *et al.*, 2009) in the evolution of plant RNA viruses have been  
52 extensively studied. Although it has been widely suggested that both high mutation and  
53 recombination rates are beneficial *per se*, they could also be byproducts of the parasitic  
54 lifestyle of viruses that favors fast replication over high fidelity (Belshaw *et al.*, 2007;  
55 Elena & Sanjuán 2005) and of the modularity of viral RNA genomes (Martin *et al.*,  
56 2005; Simon-Loriere & Holmes, 2011).

57 Recombination in RNA viruses can be defined mechanistically as an exchange of  
58 genetic material between at least two different viral genomes caused by replicase-driven  
59 template switching. Three classes of recombination have been proposed, and are  
60 distinguished by the precise mechanism of template switching (Nagy & Simon, 1997;  
61 Simon-Loriere & Holmes, 2011; Sztuba-Solińska *et al.*, 2011). The first class is base-  
62 pairing dependent and therefore requires a perfect alignment of the donor and acceptor  
63 RNA molecules. The second recombination class does not require similarity between  
64 sequences, but instead requires similarity between RNA secondary structures or *cis*-  
65 acting replication elements. The last recombination class combines characteristics of  
66 the two first classes. RNA virus recombination frequencies and rates have been  
67 estimated with a wide battery of techniques, including assays based on combinations of  
68 viruses with genetic markers in cell culture (Kirkegaard & Baltimore, 1986; Levy *et al.*,  
69 2004; Reiter *et al.*, 2011) and *in vivo* (Bruyere *et al.*, 2000; Froissart *et al.*, 2005; Pita &

70 Roossinck, 2013; Urbanowicz *et al.*, 2005), and by phylogenetic methods (Chare &  
71 Holmes, 2005; Codoñer & Elena, 2008; Martín *et al.*, 2009; Ohshima *et al.*, 2007;  
72 Revers *et al.*, 1996). Overall, quantitative estimates are highly variable, ranging from  
73  $1.4 \times 10^{-5}$  recombination events per site per generation for *Human immunodeficiency*  
74 *virus* type 1 (HIV-1) within a host (Neher & Leitner, 2010) to  $4 \times 10^{-8}$  in the case of  
75 *Hepatitis C virus* (HCV) (Reiter *et al.*, 2011). Froissart *et al.* (2005) reported the first *in*  
76 *planta* recombination rate for a plant virus, *Cauliflower mosaic virus* (CaMV). They  
77 found that recombination for this pararetrovirus was frequent and estimated its rate to be  
78  $4 \times 10^{-5}$  events per nucleotide site and per replication cycle. In a recent report, Pita &  
79 Roossinck (2013) described frequent recombination events for *Cucumber mosaic virus*  
80 (CMV), although their experimental design did not allow for estimation of  
81 recombination rates. Unfortunately, *in planta* estimates for recombination rate of  
82 single-stranded positive-sense RNA viruses, the most common amongst known plant  
83 viruses, are still missing.

84 Potyviruses represent a particularly interesting model system for studying  
85 recombination, since two apparently conflicting observations have been made. On the  
86 one hand, phylogenetic evidence suggests that unusually high frequencies of  
87 recombination occur (Chare & Holmes, 2005; Revers *et al.*, 1996). On the other hand,  
88 during mixed-genotype potyvirus infections low levels of cellular coinfection have been  
89 observed (Dietrich & Maiss, 2003; N.T. Tromas, M.P. Zwart, G. Lafforgue, S.F. Elena,  
90 unpublished manuscript). Template switching between virus genotypes can only occur  
91 if the genomic RNA of two virus genotypes is being replicated in the same cell, and  
92 hence these low levels of cellular coinfection form an impediment to recombination  
93 between virus genotypes. These conflicting observations therefore call for an  
94 experimental evaluation of the potyvirus recombination rate. Moreover, they raise the

95 question of what the effects of cellular coinfection, or paucity thereof, may have on  
96 estimates of the recombination rate.

97 In this study, we provide an estimate of the recombination rate of *Tobacco etch virus*  
98 (TEV; genus *Potyvirus*, family *Potyviridae*) during a single infection cycle in its  
99 primary host *Nicotiana tabacum*. TEV is a prototypical single-strand positive-sense  
100 RNA virus that encodes a 346-kDa polypeptide that self-processes into ten mature  
101 proteins (Riechmann *et al.*, 1992) plus an additional peptide resulting from a +2 frame  
102 shift within the third cistron during translation (Chung *et al.*, 2008). Our strategy  
103 consisted of inoculating equimolar mixtures of pairs of engineered genotypes carrying  
104 different neutral markers, characterizing the virus populations resulting from systemic  
105 infection and subsequently using different approaches to estimate the recombination  
106 rate. We have hereby provided the first estimate of a potyvirus recombination rate, and  
107 considered in detail the effects of cellular coinfection on recombination between virus  
108 genotypes.

109

## 110 **RESULTS AND DISCUSSION**

111 We used the pTEV-7DA (GenBank DQ986288) infectious clone (Dolja *et al.*, 1992) as  
112 a source for TEV. To analyze the TEV recombination rate, we introduced four neutral  
113 genetic markers, in the form of artificial restriction sites, along the TEV genome. New  
114 *AscI* and *PmeI* restriction sites were created at positions 402 and 3735, respectively,  
115 whereas natural restrictions sites *Eco47III* and *SalI* were removed from positions 4969  
116 and 7166, respectively (Fig. 1A). *N. tabacum* plants were then infected with *in vitro*  
117 synthesized RNA of each virus variant. We observed no differences in the time until  
118 the onset of TEV symptoms among the marked viruses; in all cases symptoms appeared  
119 6 - 7 days post inoculation (dpi). Furthermore, the accumulation of TEV genomes was

120 measured by RT-qPCR 7 dpi for each engineered genotype (Fig. S1, supplementary  
121 material). No differences were observed between the marked viruses (Model II nested  
122 ANOVA,  $F_{3,10} = 0.216$ ,  $P = 0.883$ ), thus confirming the neutrality of the markers.  
123 For the actual experiments to measure the recombination frequency, four-week-old *N.*  
124 *tabacum* plants were mechanically inoculated on the third true leaf with 7  $\mu$ g of an  
125 equimolecular mixture of RNA transcripts. The three combinations assayed were  
126 *AscI/PmeI*, *PmeI/Eco47III*, and *Eco47III/SalI* (Fig. 1). Each combination was  
127 inoculated on 5 plants, and after 15 days we harvested all systemically infected leaves.  
128 RT-PCR was used to amplify a region containing the two markers (see Materials and  
129 Methods). We adapted our PCR protocol to use relatively small amounts of template  
130 cDNA and a low number of cycles to avoid the formation of recombinants during this  
131 step. In 140 control reactions, we did not find any false positives (see Materials and  
132 Methods).

133 We sequenced individual clones originating from each combination in order to check  
134 for recombinants (Table 1). For the *PmeI/Eco47I* mixture, one out of the five plants  
135 was not infected. The average observed frequency of recombinant genotypes per  
136 marker combination varied from zero for the *Eco47III/SalI* combination, where no  
137 recombinant genotypes were detected, to 1.85% for the *AscI/PmeI* combination (Table  
138 1). Contrary to our expectations, we did not find a significant relationship between the  
139 distance between markers and the frequency of recombinant genotypes (Spearman  
140 correlation coefficient,  $r = 0.243$ , 12 d.f.,  $P = 0.402$ ). This result is probably due to low  
141 statistical power, given that we sampled few recombinants per marker combination.  
142 When frequencies of recombinant genotypes were transformed into recombination rates  
143 per site (Kosambi, 1944), the grand mean recombination rate ( $\pm 1$  SEM) for all three  
144 combinations was estimated to be  $r = (3.388 \pm 2.973) \times 10^{-6}$  recombination events per

145 nucleotide site (r/s). Marker combination had no effect on recombination rate (Kruskal-  
146 Wallis test,  $\chi^2 = 4.735$ , 3 d.f.,  $P = 0.094$ ), suggesting that the frequency of recombinant  
147 genotypes did not vary along the TEV genome. Similarly, distance between markers  
148 did not have a significant effect on recombination rate (Spearman correlation  
149 coefficient,  $r = 0.264$ , 12 d.f.,  $P = 0.362$ ), as has been observed previously (Anderson *et*  
150 *al.*, 1998; Froissart *et al.*, 2005). Finally, we rescaled the estimate of recombination rate  
151 to the more biologically meaningful units of recombination events per nucleotide site  
152 and per generation (r/s/g). Given a generation time of  $2.91 \pm 0.58$  ( $\pm 1$  SEM) generations  
153 per day (Martínez *et al.*, 2011), 15 dpi are equivalent to  $43.65 \pm 8.70$  generations and  
154 therefore  $r_g = (7.762 \pm 6.985) \times 10^{-8}$  r/s/g.

155

#### 156 **Factors biasing recombination rate estimation for TEV: Unbalanced mixtures and** 157 **cellular coinfection**

158 The classic method we used to estimate recombination rate assumes that the two marked  
159 virus genotypes are present at the same frequency during infection (i.e., a balanced  
160 mixture). For an unbalanced mixture of parentals, it is likely that there will be fewer  
161 opportunities for recombination between the two parental genotypes, as has been  
162 observed for *Murine leukemia virus* (MuLV) (Anderson *et al.*, 1998). As we can only  
163 detect recombination events between the two parental genomes, this may result in a  
164 lower observed recombination frequency and therefore considerably lower the estimated  
165 recombination rate. Although we carefully quantified and mixed both genotypes to  
166 ensure the 1:1 initial ratio, in most individual plants the frequency of the genotypes  
167 changed considerably after 15 dpi. However, for the three marker combinations  
168 included in the analysis the mean frequency of the genotypes across plants did not  
169 deviate significantly from the inoculum mixture (one-sample *t*-test against a value of

170 0.5: *AscI/PmeI*:  $t = 0.638$ , 4 d.f.,  $P = 0.558$ ; *PmeI/Eco47III*:  $t = 1.155$ , 3 d.f.,  $P = 0.332$ ;  
171 *Eco47III/SaII*:  $t = 0.408$ , 4 d.f.,  $P = 0.704$ ), which suggests variation may be due to  
172 genetic drift rather than to differences in fitness.

173 To further evaluate the effect of the marker composition we explored the dependence of  
174 the estimates of recombination rate on the observed ratio of the less abundant to the  
175 most abundant parental genotype. A linear regression showed a significant relationship  
176 between this ratio and the recombination rate ( $R^2 = 0.395$ ;  $F_{1,12} = 7.839$ ,  $P = 0.016$ ),  
177 confirming that the estimated recombination rate decreased as the ratio of parental  
178 genotypes departed from the hypothetical 1:1. Indeed, the regression equation  $r = a +$   
179  $b \times \text{RATIO}$  can be used to correct for the effect of the unbiased mixtures. In the worse  
180 scenario, that is when *RATIO* has the largest possible departure from the 1:1  
181 expectation, the (underestimated) expected recombination rate is simply the intercept of  
182 the equation  $a = (0.852 \pm 1.578) \times 10^{-6}$  r/s. By contrast, in the optimal situation, namely  
183 when the mixture is well balanced (e.g., *RATIO* = 1), then the expected recombination  
184 rate is  $a + b$ . Hence,  $r = (2.101 \pm 0.878) \times 10^{-5}$  r/s, or  $r_g = (4.813 \pm 2.228) \times 10^{-7}$  r/s/g, a  
185 value that is 6.2 times larger than the value reported in the previous section.

186 A second factor that could bias recombination rate estimates for TEV is the low  
187 frequency of cellular coinfection. Recombination between the marked parental viruses  
188 can only occur in those cells that are coinfecting by both viruses. Therefore detectable  
189 recombination will only occur in a fraction of infected cells, even if the parental ratio is  
190 perfectly balanced. If the parental ratio is not balanced, there will be fewer coinfecting  
191 cells and consequently even less detectable recombination events. However, only one  
192 virus genotype will be present in most infected cells (Dietrich & Maiss, 2003). In  
193 another study, the rate of cellular coinfection was measured by flow cytometry on  
194 protoplasts extract from *N. tabacum* plants infected with two TEV variants carrying



195 different fluorescent markers (N.T. Tromas, M.P. Zwart, G. Lafforgue, S.F. Elena,  
196 unpublished manuscript). Cellular coinfection was highest at 10 dpi, being  $0.138 \pm 0.029$   
197 % of infected cells. Low levels of cellular coinfection are therefore probably a general  
198 characteristic of potyvirus infection, and are probably a common impediment to  
199 recombination between different potyvirus variants. In the next section, we develop an  
200 approach to incorporate the level of coinfection into the estimation of  $r_g$ .

201

### 202 **Maximum-likelihood estimate of the recombination rate**

203 We developed a more sophisticated maximum-likelihood-based method to estimate  
204 recombination rates that takes into account details of the infection process. First, we  
205 incorporate a time-varying cellular multiplicity of infection (*MOI*). The *MOI* is the  
206 number of virions infecting a cell, which changes over the course of plant virus  
207 infection (González-Jara *et al.*, 2008; Gutiérrez *et al.*, 2010; Zwart *et al.*, 2013). A  
208 time-dependent mathematical function for *MOI* was determined based on empirical  
209 estimates of *MOI* over time (N.T. Tromas, M.P. Zwart, G. Lafforgue, S.F. Elena,  
210 unpublished manuscript) (see Materials and Methods). If the *MOI* and the frequency of  
211 parental genotypes are known, these two variables can be used to predict the expected  
212 frequency of cellular coinfection for each time point in each plant.

213 Second, we considered the effects of virus expansion in the host plant on the expected  
214 frequency of recombinants, for two reasons: (i) recombinants that are generated early in  
215 an expanding population can have a large number of descendants and therefore reach  
216 high frequencies, (ii) recombinants will probably only reach appreciable frequencies if  
217 they occur in cells where replication occurs, since they can then be replicated within the  
218 cell and possibly be transmitted to other cells. Plant viruses move locally by cell-to-cell  
219 movement, and each cell can only infect those cells adjacent to it (Dietrich & Maiss,

220 2003; González-Jara *et al.*, 2008), and hence a virus variant can be “trapped” if it is  
221 surrounded by other variants (Zwart *et al.*, 2011). Hence, observable recombination  
222 events in this setup can only occur a short time after initial infection of coinfecting cells.  
223 One approach for capturing the dynamics of virus colonization of the host is to estimate  
224 the time-varying cellular contagion rate ( $C$ ): the number of cells infected per infected  
225 cell per day. Time-varying estimates of  $C$  were made for TEV infection of *N. tabacum*  
226 based on flow cytometry data on protoplasts (N.T. Tromas, M.P. Zwart, G. Lafforgue,  
227 S.F. Elena, unpublished manuscript), and we used these estimates here to obtain a  
228 mathematical function predicting changes in  $C$  over time (see Materials and Methods).  
229 We then developed an infection model incorporating both changes in cellular  
230 coinfection (as predicted by the  $MOI$ ) and virus expansion (as embodied by  $C$ ), which  
231 predicts the frequency of recombinants. We then used a maximum-likelihood-based  
232 method to fit the model to the data by comparing the predicted and observed frequency  
233 of recombinants, and hereby estimate the recombination rate (see Materials and  
234 Methods). Bootstrapping was used to estimate its 95% confidence interval (CI). Using  
235 this approach we estimated the recombination rate to be  $r_g = 3.427 \times 10^{-5}$  r/s/g (95% CI:  
236  $1.346 \times 10^{-5}$  -  $5.998 \times 10^{-5}$ ).

237 Note that the cellular contagion rate is highest early in infection when the  $MOI$  is still  
238 low, further limiting cellular coinfection to a small time window and hereby severely  
239 limiting opportunities for recombination between virus genotypes. The effects of such  
240 details of the infection process can only be captured by the fitted model, and not by  
241 simple corrections for unbalanced mixtures of the rate of coinfection. This estimate of  
242  $r_g$  is ca. 71-fold higher than the estimate obtained above ignoring the details of TEV  
243 colonization of plant tissues. It should be noted, however, that even the maximum-  
244 likelihood-based estimate that considers details of the infection process is probably best

245 seen as a lower limit of the recombination rate. The model corrects for the effects of the  
246 segregation at the cellular level, but any further segregation of the genotypes at lower  
247 levels of organization (i.e, replication complexes within the cell) is not considered.  
248 Moreover, we have tailored this experimental system for considering homologous  
249 recombination, whereas non-homologous recombination events are also known to  
250 occur.

251

### 252 **Comparison with estimates of recombination rate for other plant RNA viruses**

253 The only previous estimate of recombination rate for a plant virus during real infection  
254 conditions was reported by Froissart *et al.* (2005) for CaMV, with  $r_g$  have an estimated  
255 range of  $(2 - 4) \times 10^{-5}$  r/s/g. This estimate is very similar to our maximum-likelihood-  
256 based estimate using an infection model  $(1 - 6) \times 10^{-5}$  r/s/g, although Froissart *et al.*  
257 (2005) incorporated no corrections for levels of cellular coinfection. However, CaMV  
258 reaches high cellular coinfection and *MOI* values (Gutiérrez *et al.*, 2010), unlike TEV  
259 and other potyviruses (Dietrich & Maiss, 2003; N.T. Tromas, M.P. Zwart, G. Lafforgue,  
260 S.F. Elena, unpublished manuscript). Such a correction is therefore unlikely to  
261 appreciably alter estimates of the CaMV recombination rate, and we propose that a  
262 comparison of these two estimates is meaningful and suggests recombination rates are  
263 similar.

264 The recombination rate for the tripartite *Brome mosaic virus* (BMV) has been evaluated  
265 in several studies (Bruyere *et al.*, 2000; Olsthoorn *et al.*, 2002; Urbanowicz *et al.*,  
266 2005). Unfortunately, comparison of these studies and our own one is not  
267 straightforward for several reasons: (i) none of these studies made a rigorous statistical  
268 data analyses and just reported counts of recombinant and parental genomes, (ii) each  
269 study focused on a particular genomic region, which may or may not be representative

270 for the whole genome, and (iii) there are no data from which to estimate the number of  
271 generations per day for BMV. Nonetheless, it is still possible to compute the frequency  
272 of recombinants and  $r$  from the numbers provided in different tables and figures of these  
273 studies. Bruyere *et al.* (2000) introduced several restriction site markers in BMV  
274 RNA3, and inoculated the non-natural host *Chenopodium quinoa* with balanced  
275 mixtures. Variable numbers of local lesions were analyzed for the presence of parental  
276 and recombinant genomes. Averaging across the four experimental replicates described  
277 in their Table 1, we estimated  $r = (3.388 \pm 2.973) \times 10^{-6}$  r/s. In a follow-up study  
278 Urbanowicz *et al.* (2005) used a highly similar method to estimate the recombination  
279 frequencies for RNA1 and RNA2. Using the data contained in their Figs. 2 and 3, we  
280 estimated  $r = (1.739 \pm 0.433) \times 10^{-4}$  r/s for RNA1 and  $r = (2.490 \pm 0.400) \times 10^{-4}$  r/s for  
281 RNA2, averaging across experiments for RNA2. The results for the three segments are  
282 homogeneous (Kruskal-Wallis test,  $\chi^2 = 2.444$ , 2 d.f.,  $P = 0.295$ ) and thus we can  
283 estimate an average genome-wide recombination rate per site for BMV of  $r =$   
284  $(2.104 \pm 0.268) \times 10^{-4}$  r/s. This value is far higher than our estimate of  $r$  for TEV, and  
285 suggests that TEV recombines less than BMV, a multipartite positive-stranded RNA  
286 virus. Unfortunately, to our knowledge, the frequency of cellular coinfection for BMV  
287 is still unknown as it is the number of generations per day.

288

### 289 **Relationship between mutation and recombination rates for RNA viruses**

290 Tromas & Elena (2010) estimated the TEV point mutation rate to be in the range  $(0.475$   
291  $- 6.299) \times 10^{-5}$  mutations per site per generation. The mutation rate range is therefore,  
292 given the uncertainties associated to both estimates, similar to the TEV recombination  
293 rate estimated with the infection model. This similarity suggests that mutation and

294 recombination may have a similar impact on the generation of genetic diversity for  
295 TEV.

296 If mutation and recombination rates are indeed similar, this has important evolutionary  
297 implications. For example, high mutation rates in combination with small population  
298 sizes during viral transmission turn on Muller's ratchet, a phenomenon already shown  
299 to operate in experimental TEV populations transmitted throughout very dramatic  
300 bottlenecks (De la Iglesia & Elena, 2007). A high recombination rate could counteract  
301 the effect of Muller's ratchet by recreating mutation-free genomes. The main  
302 evolutionary advantage of recombination may be as a mechanism of sex that helps  
303 purge deleterious mutations (Muller, 1964; Simon-Loriere & Holmes, 2011). In a  
304 similar vein, high recombination rates may have also be advantageous because they  
305 speed up the rate of adaptation by bringing together beneficial mutations that otherwise  
306 would exist in different genomes, thus minimizing the effect of clonal interference  
307 (Fisher, 1930; Muller, 1932; Simon-Loriere & Holmes, 2011). In addition to these  
308 "fitness advantage" theories, other models have been brought forward to explain viral  
309 recombination. One intriguing possibility is that RNA virus recombination may have  
310 evolved as a byproduct of the high nucleotide incorporation rate of viral RNA  
311 polymerases (Simon-Loriere & Holmes, 2011): the faster the action of the replicase, the  
312 more slippery it becomes. RNA virus high mutation rates are also a likely consequence  
313 of the tradeoff between fast replication and accuracy (Belshaw *et al.*, 2007; Elena &  
314 Sanjuán, 2005), and one would then expect a positive correlation between mutation and  
315 recombination rates across RNA viruses.

316 To test this prediction, we searched the literature on RNA viruses for cases in which  
317 estimates of both the mutation and recombination rates are available. Unfortunately, the  
318 set is just limited to the following seven cases: HIV-1 (Batorsky *et al.*, 2011; Jezt *et al.*,

319 2000), HCV (Reiter *et al.*, 2011), *Mouse hepatitis virus* (MHV; Baric *et al.*, 1990),  
320 MuLV (Anderson *et al.*, 1997; Zhuang *et al.*, 2006), *Poliovirus* (PV; Duggal *et al.*,  
321 1997; Jarvis & Kirkegaard, 1992; King, 1988), *Spleen necrosis virus* (SNV; Hu &  
322 Temin, 1989), and TEV (this study). When more than one estimate existed for one  
323 virus, the average was taken. Mutation rates were taken from Sanjuán *et al.* (2010).  
324 Note that the low recombination rate of HCV measured by Reiter *et al.* (2011) has been  
325 questioned by others (González-Candelas *et al.*, 2011). Fig. 2 illustrates the relationship  
326 between mutation and recombination rates. A positive correlation exists between both  
327 traits, which becomes highly significant if the discordant data point for HCV is removed  
328 from the computation (Pearson's  $r = 0.963$ , 4 d.f.,  $P = 0.002$ ). Therefore, these data  
329 support the hypothesis that TEV recombination rate may be a side effect of selection for  
330 fast but error prone replication, rather than being selected for the fitness advantages it  
331 may provide in the long run.

332

## 333 **METHODS**

334 **Generation of restriction sites as genetic markers.** All mutations necessary to create  
335 or remove restriction sites were introduced by PCR-directed mutagenesis using the  
336 Quickchange<sup>®</sup> II XL kit (Stratagene) and following the indications given by the  
337 manufacturer. Primer pairs (Table 2) for mutagenesis were designed following  
338 Stratagene's recommendations. To minimize unwanted errors during the mutagenesis  
339 process, the kit incorporates the *PfuUltra*<sup>™</sup> high fidelity DNA polymerase (Stratagene).  
340 The amplification conditions were 1 min at 95°C (initial denaturation), followed by 18  
341 cycles consisting of 30 s at 95 °C, 45 s at 65 °C and 18 min at 68 °C, and a final  
342 extension step of 28 min at 68°C. PCR products were digested with *DpnI* (New  
343 England Biolabs) to remove the parental methylated strands and transformed into

344 electrocompetent *Escherichia coli* DH5 $\alpha$ . At least 15 clones were sequenced to  
345 confirm the successful incorporation of desired mutations.

346

347 **Neutrality of restriction markers.** Sequence-validated plasmids containing the  
348 corresponding restriction site maker were linearized with *Bgl*III (Takara) and transcribed  
349 into 5'-capped RNAs using the SP6 mMMESSAGE mMACHINE kit (Ambion Inc).  
350 Transcripts were precipitated (1.5 volumes of DEPC-treated water, 1.5 volumes of 7.5  
351 M LiCl, 50 mM EDTA), collected and resuspended in DEPC-treated water (Carrasco *et*  
352 *al.* 2007). RNA integrity was assessed by gel electrophoresis and concentration was  
353 spectrophotometrically determined using a Biophotometer (Eppendorf).

354 To evaluate the neutrality of the four markers, we proceeded as follows. Three four-  
355 week-old *N. tabacum* plants were inoculated by abrasion on the third true leaf with 7  $\mu$ g  
356 of transcribed RNA from each individual marker as described elsewhere (Carrasco *et*  
357 *al.*, 2007). Inoculated plants were placed in a BSL-2 greenhouse at 25 °C and 16 h  
358 light/8 h dark period. RT-qPCR was performed as described elsewhere (Lalić *et al.*,  
359 2011).

360

361 **Coinoculation experiments and restriction analysis.** Total RNA was extracted using  
362 InviTrap<sup>®</sup> Spin Plant RNA Mini Kit (Invitex) from the symptomatic leaflets for each of  
363 five plants 15 dpi. Each plant was analyzed separately, thus providing independent  
364 replicates of the recombination rate among pairs of markers. For each combination of  
365 restriction makers, the regions of interest were reverse transcribed using the following  
366 reaction mixture: 1 $\times$  RT reaction buffer (Fermentas), 0.2 mM each dNTP, 0.25  $\mu$ M  
367 forward primer (Table 3), 0.2  $\mu$ L RNase inhibitor, 40 units of Moloney MuLV reverse  
368 transcriptase (Fermentas), 10 ng total RNA, and DEPC-treated water to complete 20  $\mu$ L

369 reaction volume. Five PCR reactions for each RT reaction were then done using the  
370 following reaction mixture: 1× HF buffer (Finzymes), 0.2 mM each dNTP, 0.25 μM  
371 each primer (Table 3), 0.5 units of ultra high-fidelity Phusion DNA polymerase  
372 (Finzymes), 1 μL DMSO, 3 μL from the reverse-transcription reaction, and DEPC-  
373 treated water to complete a reaction volume of 25 μL. The cycling conditions were  
374 optimized to limit false-positive recombination events: 1 min at 98 °C; followed by 25  
375 cycles consisting of 8 s at 98 °C, 25 s at 57 °C and 25 s/kb at 72 °C; and then a final  
376 extension of 5 min at 72 °C. PCR products were gel purified with GeneJET™ Gel  
377 Extraction Kit (Fermentas), cloned into the plasmid pUC19/*Sma*I (Fermentas), and used  
378 to transform *E. coli*. The analyses of a large number of clones were performed by  
379 amplifying the region of interest using colonies-PCRs: 1× Taq buffer (Roche), 0.2 mM  
380 each dNTP, 0.25 μM each primer (Table 3), 2 units of Taq DNA polymerase (Roche)  
381 and sequencing by Genoscreen (<http://www.genoscreen.fr>) using BIGDYE 3.1 and a  
382 96-capillars ABI3730XL sequencing system (Applied Biosystems). Sequences were  
383 analyzed using GENEIOUS version 4.8 (<http://www.geneious.com>). The number of  
384 clones that rendered useful sequences was 525 (instead of the 672 submitted for  
385 sequencing). The number of sequenced clones per plant ranged between 10 and 47,  
386 with a median value of 43.

387 For each pair of markers, four progeny genotypes are expected, the two parentals (Fig.  
388 1B, left column) and the two recombinants (Figure 1B, right column).

389

390 **Minimizing the formation of recombinant molecules during RT-PCR.** A  
391 worrisome aspect of PCR based studies of virus variability is the phenomenon of PCR-  
392 mediated recombination, or chimera formation (Meyerhans *et al.*, 1990). In a recent  
393 study, Lahr & Katz (2009) have shown that using the Phusion DNA polymerase, no



394 more than 30 amplification cycles and a low initial template concentration minimized  
395 chimera formation. To determine whether the RT-PCR conditions used in our  
396 experiments may have favored the formation of chimeras, we first inoculated two four-  
397 week-old *N. tabacum* plants with respectively TEV RNAs containing *PmeI* and  
398 *Eco47III* markers. After 15 dpi, total RNA from each plant was extracted and mixed in  
399 a 1:1 (w:w) ratio. Three serial dilutions of this mixture were made: 50 ng/μL, 5 ng/μL  
400 and 0.5 ng/μL. Each of these dilutions was then used as template for an RT-PCR  
401 experiment (25 cycles, Phusion DNA polymerase). We failed to get an amplification  
402 product for the highest dilution (0.5 ng/μL). For the two others dilutions, PCR products  
403 were purified, cloned and transformed into *E. coli*. Forty-eight clones were genotyped  
404 by restriction analysis with *PmeI* and *Eco47III* for the 50 ng/μL dilution, finding two  
405 false positives. One hundred and seven clones were likewise genotyped for the 5 ng/μL  
406 dilution without observing false positives; 33 additional clones were genotyped by  
407 sequencing for this dilution, with no false positives observed. No homologous or non-  
408 homologous recombination events were observed, confirming these conditions avoid  
409 false positive results. These results also confirm that the concentration of template  
410 RNA molecules and the number of PCR cycles determine the chances of generating  
411 chimera molecules during RT-PCR.

412

413 **Estimation of  $r$  and  $r_g$ .** The frequency of recombinant genotypes ( $f$ ) on each analyzed  
414 plant was estimated as  $f = \text{Recombinants}/(\text{Recombinants} + \text{Parentals})$ . The  
415 recombination rate between restriction markers was computed according to Kosambi  
416 (1944) equation:

417 
$$r = \frac{1}{4L} \ln \frac{1+2f}{1-2f},$$

418 where  $L$  is the physical distance separating the two markers (in nucleotides). The units  
 419 of the resulting estimate are recombination events per site (r/s). The Kosambi method  
 420 was chosen to minimize the potential effect that multiple crossovers may have in the  
 421 inference of  $r$ . The independent estimates for each pairs of markers were averaged and  
 422 the corresponding SEM computed. A genome-wide  $r$  was computed averaging all  
 423 estimates obtained from different combinations of restriction markers and experimental  
 424 replicates. Note that  $r$  was divided by the number of generations (15 days  $\times$  2.91  
 425 generations per day; Martínez *et al.*, 2011) to obtain  $r_g$ .

426

427 **Maximum-likelihood estimate of  $r_g$ .** To obtain a function for  $MOI$  over time ( $t$  in  
 428 days),  $m_t$ , we fitted a logistic model to  $MOI$  estimates for TEV infection of *N. tabacum*  
 429 leaves 3, 5, 6, and 7 on days 3, 5, 7, and 10 dpi (N.T. Tromas, M.P. Zwart, G.

430 Lafforgue, S.F. Elena, unpublished manuscript), such that  $m_t = \kappa / \left[ 1 - \left( 1 - \frac{\kappa}{m_0} \right) e^{-\nu t} \right]$ .

431 Here  $m_t = m_0$  is the  $MOI$  at  $t = 0$ ,  $\kappa$  is the maximum attainable value for  $MOI$  and  $\nu$  is the  
 432 initial rate of increase in  $MOI$ . A grid search was used to estimate  $m_0 = 1.12 \times 10^{-3}$ ,  $\kappa =$   
 433 0.47 and  $\nu = 1.02$ , by minimizing the residual sum of squares. From  $m_t$  the frequency of

434 coinfecting cells can be calculated as  $c_t = \left( 1 - e^{-m_t \alpha} \right) \left( 1 - e^{-m_t (1-\alpha)} \right) / \left( 1 - e^{-m_t} \right)$ , where  $\alpha$  is the

435 LaPlace binomial point estimator of the frequency of a parental genotype. Note that  $m_t$   
 436 is the  $MOI$  in all cells, including those that are not infected, and we must therefore  
 437 divide the fraction of coinfecting cells by the fraction of infected cells  $\left( 1 - e^{-m_t} \right)$ .

438 To obtain a function for the change in  $C$  (the cellular contagion rate), we fitted an  
 439 exponential function to estimates of  $C$  for TEV infection of pooled data of different  
 440 leaves of *N. tabacum* on days 3, 5, 7, and 10 dpi (N.T. Tromas, M.P. Zwart, G.

441 Lafforgue, S.F. Elena, unpublished manuscript), such that  $C_t = C_0 e^{-\gamma t}$ . Here,  $C = C_0$  if  $t$

442 = 0 and  $\gamma$  is the decay rate. A grid search was used to estimate  $C_0 = 45.0$  and  $\gamma = 0.84$ ,  
 443 by minimizing the residual sum of squares. The difference equation for the number of  
 444 infected cells is  $n_{t+1} = (1 + C_t)n_t$ .

445 To estimate the expected final frequency of recombinants, we first estimate the  
 446 frequency of *de novo* recombinants each day ( $y$ ) as determined by Kosambi (1944):

447  $y = \frac{\theta}{2} \tanh(2Lr_g)$ , where  $\theta$  is the number of virus generations per day of infection

448 (2.91). The frequency of recombinants on a given day ( $f_t$ ) is then  $f_t = f_{t-1} + yC_t$ . The  
 449 frequency of recombinants from the previous day is included, since we are considering  
 450 only those recombinants that occur in the newly infected cells, but assume that newly  
 451 occurring recombinants will be maintained in the virus population. The window of one  
 452 day for allowing recombination to occur in newly infected cells is probably  
 453 conservative given the rapid advance of TEV infection (Dolja *et al.*, 1993) and our  
 454 estimate of  $\theta = 2.91$  generations per day (Martinez *et al.*, 2011). We can then estimate

455 the expected frequency of recombinants in the final population  $f_{15} = \frac{1}{n_{15}} \sum_{t=1}^{15} n_t C_t f_t$ , where

456  $n_t$  is the number of infected cells after  $t$  days. This equation in essence estimates the  
 457 mean frequency of recombinants occurring over days, weighted by the amount of  
 458 expansion occurring on a particular day ( $n_t C_t$ ). In order to estimate  $r_g$ , we minimized  
 459 the negative log-likelihood by a grid search. The likelihood of the corresponding  $f_{15}$

460 value is given by:  $L(f_{15} | g, h) = \binom{g}{h} f^h (1-f)^{g-h}$ , where  $g$  is the total number of clones

461 sequenced, and  $h$  is the number of sequenced clones that were recombinant.

462

463 **ACKNOWLEDGMENTS**

464 We thank Francisca de la Iglesia and Àngels Pròsper for excellent technical assistance,  
465 José A. Daròs for methodological advice, José M. Cuevas for critical reading of the  
466 manuscript, and other lab members for helpful discussions. This work was supported  
467 by grants BFU2009-06993 and BFU2012-30805 from the Spanish Secretaría de Estado  
468 de Investigación, Desarrollo e Innovación. N.T. was supported by a pre-doctoral  
469 fellowship from the former Spanish Ministerio de Ciencia e Innovación.

470

## 471 REFERENCES

472 **Anderson, J. A., Bowman, E. H. & Hu, W. S. (1998).** Retroviral recombination rates  
473 do not increase linearly with marker distance and are limited by the size of the  
474 recombining subpopulation. *J Virol* **72**, 1195-1202.

475 **Aranda, M. A., Fraile, A., Dopazo, J., Malpica, J. M. & García-Arenal, F. (1997).**  
476 Contribution of mutation and RNA recombination to the evolution of a plant  
477 pathogenic RNA. *J Mol Evol* **44**, 81-88.

478 **Baric, R. S., Fu, K., Schaad, M. C. & Stohlman, S. A. (1990).** Establishing a genetic  
479 recombination map for murine coronavirus strain A59 complementation groups.  
480 *Virology* **177**, 646-656.

481 **Batorsky, R., Kearney, M. F., Palmer, S. E., Maldarelli, F., Rouzine, I. M. &**  
482 **Coffin, J. M. (2011).** Estimate of effective recombination rate and average  
483 selection coefficient for HIV in chronic infection. *Proc Natl Acad Sci USA* **108**,  
484 5661-5666.

485 **Belshaw, R., Gardner, A., Rambaut, A. & Pybus, O. G. (2007).** Pacing a small cage:  
486 mutation and RNA viruses. *Trends Ecol Evol* **23**, 188-193.

- 487 **Bonnet, J., Fraile, A., Sacristán, S., Malpica, J. M. & García-Arenal, F. (2005).**  
488 Role of recombination in the evolution of natural populations of *Cucumber mosaic*  
489 *virus*, a tripartite RNA plant virus. *Virology* **332**, 359-368.
- 490 **Bruyere, A., Wantroba, M., Flasiński, S., Dzianott, A. & Bujarski, J. J. (2000).**  
491 Frequent homologous recombination events between molecules of one RNA  
492 component in a multipartite RNA virus. *J Virol* **74**, 4214-4219.
- 493 **Carrasco, P., Daròs, J. A., Agudelo-Romero, P. & Elena, S. F. (2007).** A real-time  
494 RT-PCR assay for quantifying the fitness of *Tobacco etch virus* in competition  
495 experiments. *J Virol Meth* **139**, 181-188.
- 496 **Chare, E. R. & Holmes, E. C. (2005).** A phylogenetic survey of recombination  
497 frequency in plant RNA viruses. *Arch Virol* **151**, 933-946.
- 498 **Chen, Y. K., Goldbach, R. & Prins, M. (2002).** Inter- and intramolecular  
499 recombinations in the *Cucumber mosaic virus* genome related to adaptation to  
500 *Alstroemeria*. *J Virol* **76**, 4119-4124.
- 501 **Chung, B. Y. W., Miller, W. A., Atkins, J. F. & Firth, A. E. (2008).** An overlapping  
502 essential gene in the *Potyviridae*. *Proc Natl Acad Sci USA* **15**, 5897-5902.
- 503 **Codoñer, F. M. & Elena, S. F. (2008).** The promiscuous evolutionary history of the  
504 family *Bromoviridae*. *J Gen Virol* **89**, 1739-1747.
- 505 **De la Iglesia, F. & Elena, S. F. (2007).** Fitness declines in *Tobacco etch virus* upon  
506 serial bottleneck transfers. *J Virol* **10**, 4941-4947.
- 507 **Dietrich, C. & Maiss, E. (2003).** Fluorescent labelling reveals spatial separation of  
508 potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J Gen Virol*  
509 **84**, 2871-2876.

510 **Dolja, V. V., Herndon, K. L., Pirone, T. P. & Carrington, J. C. (1993).** Spontaneous  
511 mutagenesis of a plant potyvirus genome after insertion of a foreign gene. *J Virol*  
512 **67**, 5967-5975.

513 **Dolja, V. V., McBride, H. J. & Carrington, J. C. (1992).** Tagging of plant potyvirus  
514 replication and movement by insertion of  $\beta$ -glucuronidase into the viral  
515 polyprotein. *Proc Natl Acad Sci USA* **89**, 10208-10212.

516 **Duggal, R., Cuconati, A., Gromeier, M. & Wimmer, E. (1997).** Genetic  
517 recombination of poliovirus in a cell-free system. *Proc Natl Acad Sci USA* **94**,  
518 13786-13791.

519 **Elena, S. F. & Sanjuán, R. (2005).** Adaptive value of high mutation rates of RNA  
520 viruses: separating causes from consequences. *J Virol* **79**, 11555-11558.

521 **Fernández-Cuartero, B., Burgyán, J., Aranda, M. A., Salánki, K., Moriones, E. &**  
522 **García-Arenal, F. (1994).** Increase in relative fitness of a plant virus RNA  
523 associated with its recombinant nature. *Virology* **203**, 373-377.

524 **Fisher, R. A. (1930).** *The Genetical Theory of Natural Selection*. University Press,  
525 Oxford, UK.

526 **Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S. & Michalakis, Y. (2005).**  
527 Recombination every day: abundant recombination in a virus during a single multi-  
528 cellular host infection. *PLoS Biol* **3**, e89.

529 **González-Candelas, F., López-Labrador, F. X. & Bracho, M. A. (2011).**  
530 Recombination in *Hepatitis C virus*. *Viruses* **3**, 2006-2024

531 **González-Jara, P., Fraile, A., Cantó, T. & García-Arenal, F. (2009).** The  
532 multiplicity of infection of a plant virus varies during colonization of its eukaryotic  
533 host. *J Virol* **83**, 7487-7494.

534 **Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y. & Blanc, S.**  
535 **(2010).** Dynamics of the multiplicity of cellular infection in a plant virus. *PLoS*  
536 *Pathog* **6**, e1001113.

537 **Hu, W. S. & Temin, H. M. (1990).** Genetic consequences of packaging two RNA  
538 genomes in one retroviral particles: pseudodiploidy and high rate of genetic  
539 recombination. *Proc Natl Acad Sci USA* **87**, 1556-1560.

540 **Jarvis, T. C. & Kirkegaard, K. (1992).** Poliovirus RNA recombination: mechanistic  
541 studies in the absence of selection. *EMBO J* **11**, 135-145.

542 **Jetzt, A. E., Yu, H., Klarmann, G. J., Ron, Y., Preston, B. D. & Dougherty, J. P.**  
543 **(2000).** High rate of recombination throughout the *Human immunodeficiency virus*  
544 type 1 genome. *J Virol* **74**, 1234-1240.

545 **King, A. M. (1988).** Preferred sites of recombination in poliovirus RNA: an analysis of  
546 40 intertypic cross-over sequences. *Nucl Acids Res* **16**, 11705-11723.

547 **Kirkegaard, K. & Baltimore, D. (1986).** The mechanism of RNA recombination in  
548 poliovirus. *Cell* **47**, 433-443.

549 **Kosambi, D. D. (1944).** The estimation of map distance from recombination values.  
550 *Ann Eugen* **12**, 172-175.

551 **Lahr, D. J. G. & Katz, L. A. (2009).** Reducing the impact of PCR-mediated  
552 recombination in molecular evolution and environmental studies using a new-  
553 generation high-fidelity DNA polymerase. *BioTechniques* **47**, 857-866.

554 **Lalić, J., Cuevas, J. M. & Elena, S. F. (2011).** Effect of host species on the  
555 distribution of mutational fitness effects for an RNA virus. *PLoS Genet* **7**,  
556 e1002378.

557 **Levy, D. N., Aldrovandi, G. M., Kutsch, O. & Shaw, G. M. (2004).** Dynamics of  
558 HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci USA* **101**, 4204-  
559 4209.

560 **Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W. & García-Arenal,**  
561 **F. (2002).** The rate and character of spontaneous mutation in an RNA virus.  
562 *Genetics* **162**, 1505-1511.

563 **Martin, D. P., van der Walt, E., Posada, D. & Rybicki, E. P. (2005).** The  
564 evolutionary value of recombination is constrained by genome modularity. *PLoS*  
565 *Genet* **1**, e51.

566 **Martín, S., Sambade, A. Rubio, L., Vives, M. C., Moya, P., Guerri, J., Elena, S. F.**  
567 **& Moreno, P. (2009).** Contribution of recombination and selection to molecular  
568 evolution of *Citrus tristeza virus*. *J Gen Virol* **90**, 1527-1538.

569 **Martínez, F., Sardanyés, J., Elena, S. F. & Daròs, J. A. (2011).** Dynamics of a plant  
570 RNA virus intracellular accumulation: stamping machine vs. geometric replication.  
571 *Genetics* **188**, 637-464.

572 **Meyerhans, A., Vartanian, J. P. & Wain-Hobson, S. (1990).** DNA recombination  
573 during PCR. *Nucl Acids Res* **18**, 1687-1691.

574 **Muller, H. J. (1932).** Some genetic aspects of sex. *Am Nat* **66**, 118-138.

575 **Muller, H. J. (1964).** The relation of recombination to mutational advance. *Mut Res* **1**,  
576 2-9.

577 **Nagy, P. D. & Simon, A. E. (1997).** New insights into the mechanisms of RNA  
578 recombination. *Virology* **235**, 1-9.

579 **Neher, R. A. & Leitner, T. (2010).** Recombination rate and selection strength in HIV  
580 intra-patient evolution. *PLoS Comput Biol* **6**, e1000660.



581 **Ohshima, K., Tomitaka, Y., Wood, J. T., Minematsu, Y., Kajiyama, H.,**  
582 **Tomimura, K. & Gibbs, A. J. (2007).** Pattern of recombination in *Turnip mosaic*  
583 *virus* genomic sequences indicate hotspots of recombination. *J. Gen. Virol.* **88**, 298-  
584 315.

585 **Olsthoorn, R. C. L., Bruyere, A., Dzianott, A. & Bujarski, J. J. (2002).** RNA  
586 recombination in *Brome mosaic virus*: effects of strand-specific stem-loop inserts. *J*  
587 *Virol* **76**, 12654-12662.

588 **Pita, J. S. & Roossinck, M. J. (2013).** Fixation of emerging interviral recombinants in  
589 *Cucumber mosaic virus* populations. *J Virol* **87**, 1264-1269.

590 **Reiter, J., Pérez-Vilaró, G., Scheller, N., Mina, L. B., Díez, J. & Meyerhans, A.**  
591 **(2011).** *Hepatitis C virus* RNA recombination in cell culture. *J Hepatol* **55**, 777-  
592 783.

593 **Revers, F., Le Gall, O., Candresse, T., Le Romancer, M. & Dunez, J. (1996).**  
594 Frequent occurrence of recombinant potyvirus isolates. *J Gen Virol* **77**, 1953-1965.

595 **Riechmann, J. L., Laín, S. & García, J. A. (1992).** Highlights and prospects of  
596 potyvirus molecular biology. *J Gen Virol* **73**, 1-16.

597 **Sanjuán, R., Agudelo-Romero, P. & Elena, S. F. (2009).** Upper-limit mutation rate  
598 for a plant RNA virus. *Biol Lett* **5**, 394-396.

599 **Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. (2010).** Viral  
600 mutation rates. *J Virol* **84**, 9733-9748.

601 **Simon-Loriere, E. & Holmes, E. C. (2011).** Why do RNA viruses recombine? *Nat Rev*  
602 *Microbiol* **9**, 617-626.

603 **Sztuba-Solińska, J., Urbanowicz, A., Figlerowicz, M. & Bujarski, J. J. (2011).** RNA-  
604 RNA recombination in plant virus replication and evolution. *Annu Rev Phytopathol*  
605 **49**, 415-443.

606 **Tromas, N. & Elena, S. F. (2010).** The rate and spectrum of spontaneous mutations in  
607 a plant RNA virus. *Genetics* **185**, 983-989.

608 **Urbanowicz, A., Alejska, M., Formanowicz, P., Błażewicz, J., Figlerowicz, M. &**  
609 **Bujarski, J. J. (2005).** Homologous crossover among molecules of *Brome mosaic*  
610 *bromovirus* RNA1 or RNA2 segments *in vivo*. *J Virol* **79**, 5732-5742.

611 **Van der Walt, E., Rybicki, E. P., Varsani, A., Polston, J. E., Billharz, R.,**  
612 **Donaldson, L., Monjane, A. L. & Martin, D. P. (2009).** Rapid host adaptation by  
613 extensive recombination. *J Gen Virol* **90**, 734-746.

614 **Zhuang, J., Mukherjee, S., Ron, Y. & Dougherty, J. P. (2006).** High rate of genetic  
615 recombination in *Murine leukemia virus*: implications for influencing proviral  
616 ploidy. *J Virol* **80**, 6706-6711.

617 **Zwart, M. P., Daròs, J. A. & Elena, S. F. (2011).** One is enough: *in vivo* effective  
618 population size is dose-dependent for a plant RNA virus. *PLoS Pathog* **7**,  
619 e1002122.

620 **Zwart, M. P., Tromas, N. & Elena, S. F. (2013).** Model-selection-based approach for  
621 calculating cellular multiplicity of infection during virus colonization of multi-  
622 cellular hosts. *PLoS ONE* **8**, e64657.

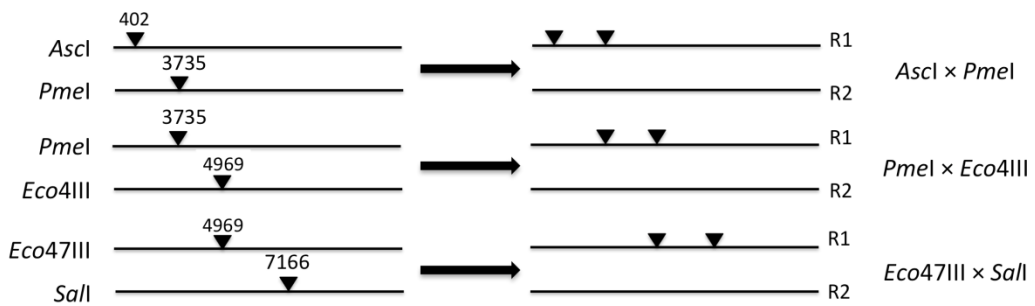
623

624 **Fig. 1.** (A) Location of the different restriction site markers in TEV genome. (B)  
 625 Expected restriction profile for each pair of markers. The left column shows the two  
 626 parental genotypes and the right column the two recombinant ones.

A



B



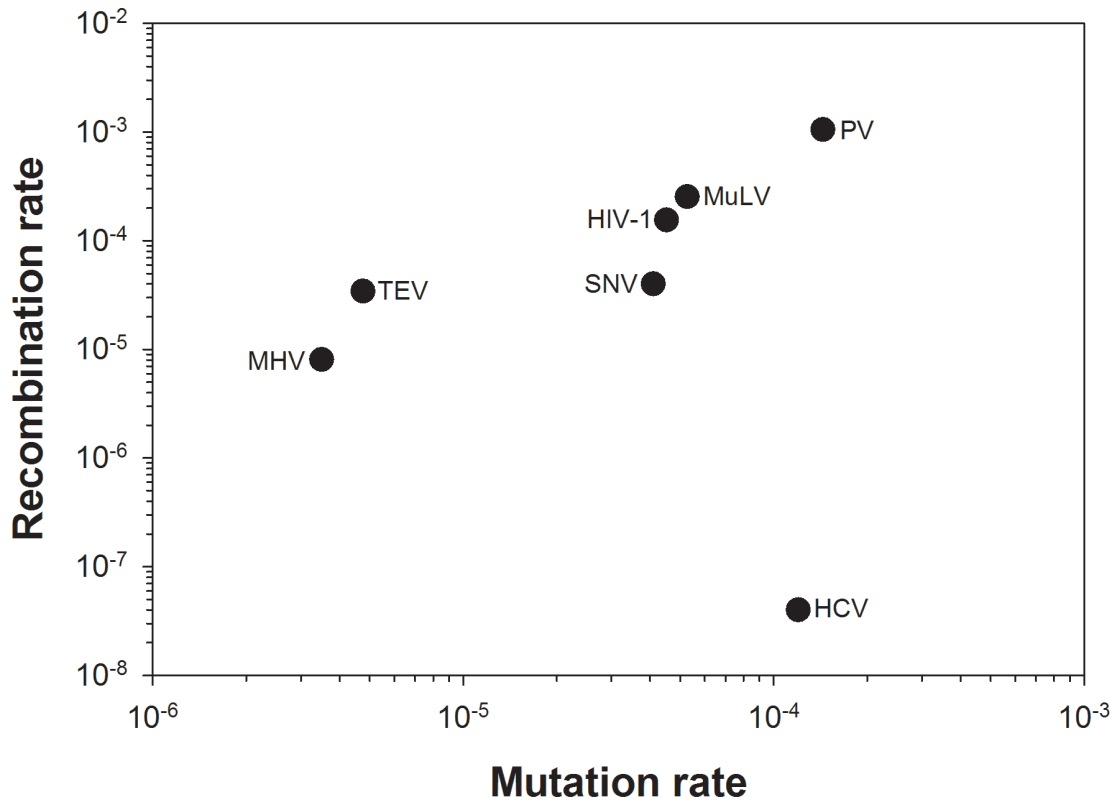
627

628

629

630 **Fig. 2.** Relationship between mutation and recombination rates for seven RNA viruses

631 and retroviruses.



632  
633

**Table 1.** Observed frequency of homologous recombinants ( $f$ ) and estimated recombination rate per site ( $r$ ).

<b>Combination</b>	<b>Fragment</b>	<b>Plants</b>	$f (\pm 1 \text{ SEM}) (\%)$	$r (\pm 1 \text{ SEM}) (\times 10^{-6})$
	<b>size (nt)</b>	<b>analyzed</b>		
<i>AscI/PmeI</i>	3334	5	1.854±0.125	5.561±0.376
<i>PmeI/Eco47III</i>	1234	4	0.568±0.084	4.604±0.684
<i>Eco47III/SalI</i>	2197	5	0	0

**Table 2.** Artificial restriction sites engineered as genetic markers for this study

Restriction		Genomic position		Mutagenesis
enzyme	Cistron	for the cut		primer (5' to 3')*
<i>AscI</i>	P1	402		TTATCTTGGTC <u>GCGCGCC</u> CTCACCCATGGC
<i>PmeI</i>	CI	3735		AGCCTTCCTGGAGTCAC <u>GTTTAA</u> CAATGGTGGAACAACCA
<i>Eco47III</i>	CI	4969		AGTCATACATGACAAGCTGAA <u>ACGTT</u> TTAAGCTACACACTTGTGAG
<i>SalI</i>	NIb	7166		GATGGGAGCATATAAGCCA <u>CCCGACT</u> TAATAGAGAGGCG

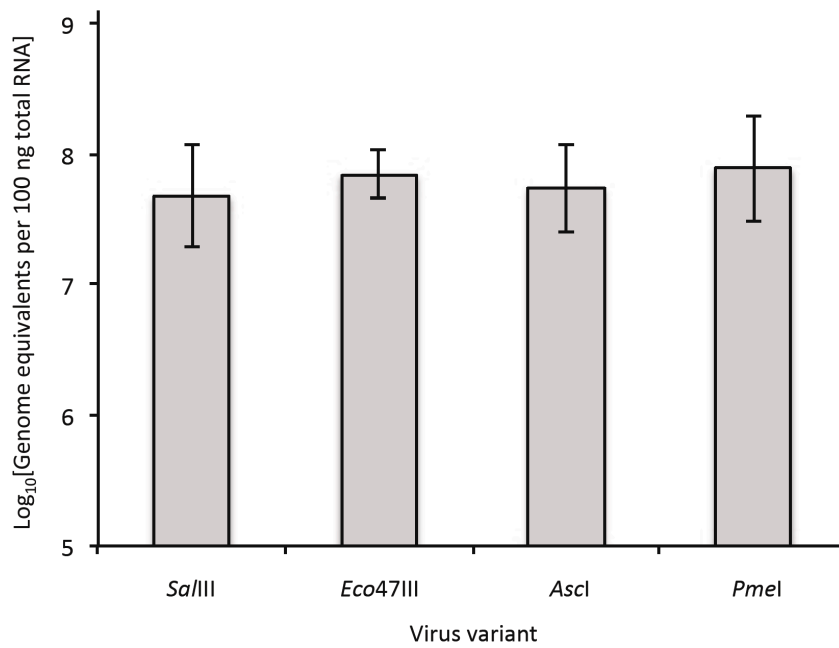
\*Restriction sites are underlined. Mutagenized sites are indicated with bold cursive type.

**Table 3.** Primers used to amplify the region containing the pair of restriction sites

<b>Combination</b>	<b>TEV genome</b>	
	<b>position (5')</b>	<b>Sequence (5' to 3')</b>
<i>AscI/PmeI</i>	46	GCAATCAAGCATTCTACTTC
	3894	ATCCAACAGCACCTCTCAC
<i>PmeI/Eco47III</i>	3541	TTGACGCTGAGCGGAGTGATGG
	5275	CTATTGATGCATGCTAGAGTC
<i>Eco47III/SalI</i>	4972	TTAAGCTACACACTTGTGAGAC
	7394	TTCTTTCTTCTTGCCTTTG

**Supplementary Material for “Estimation of the *in vivo* recombination rate for a plant RNA virus”**

Nicolas Tromas, Mark P. Zwart, Maïté Poulain, and Santiago F. Elena



**Fig. S1.** Mean log<sub>10</sub>-transformed accumulation levels of the four marked virus variants, as measured by RT-qPCR. Error bars represent 1 standard deviation.