The Rate and Character of Spontaneous Mutation in an RNA Virus

José M. Malpica,* Aurora Fraile,† Ignacio Moreno,† Clara I. Obies,†
John W. Drake‡ and Fernando García-Arenal†,1

*Departamento de Protección Vegetal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Carretera de La Coruña Km. 7.5, 28040-Madrid, Spain, †Departamento de Biotecnología, Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Politécnica de Madrid, 28040-Madrid, Spain and ‡Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709-2233

ABSTRACT

Estimates of spontaneous mutation rates for RNA viruses are few and uncertain, most notably due to their dependence on tiny mutation reporter sequences that may not well represent the whole genome. We report here an estimate of the spontaneous mutation rate of tobacco mosaic virus using an 804-base cognate mutational target, the viral *MP* gene that encodes the movement protein (MP). Selection against newly arising mutants was countered by providing MP function from a transgene. The estimated genomic mutation rate was on the lower side of the range previously estimated for lytic animal riboviruses. We also present the first unbiased riboviral mutational spectrum. The proportion of base substitutions is the same as that in a retrovirus but is lower than that in most DNA-based organisms. Although the *MP* mutant frequency was 0.02–0.05, 35% of the sequenced mutants contained two or more mutations. Therefore, the mutation process in populations of TMV and perhaps of riboviruses generally differs profoundly from that in populations of DNA-based microbes and may be strongly influenced by a subpopulation of mutator polymerases.

M ANY viruses have RNA genomes (riboviruses) and cause important diseases of humans and domestic animals and plants. Countering ribovirus diseases is often hampered by the rapid evolution of ribovirus populations, thus providing a major incentive to study ribovirus evolution (Nathanson *et al.* 1995). In addition, riboviruses provide good models for investigating evolutionary processes (*e.g.*, Burch and Chao 2000; Moya *et al.* 2000). Ribovirus mutant frequencies are generally much higher than those of DNA-based microbes (Drake *et al.* 1998), and the notion that riboviruses have high rates of spontaneous mutation is now widely accepted (Domingo and Holland 1997). However, this conclusion rests on few and uncertain estimates of spontaneous mutation rates. Data exist only for lytic animal viruses, and the available estimates are less reliable than the estimates for viruses or cellular microbes with DNA genomes (Drake *et al.* 1998; Drake and Holland 1999).

A major problem in quantifying the rates of spontaneous mutation of riboviruses is that most values are based on results obtained using very small, potentially unrepresentative mutational targets (Drake 1993). It would be far better to use a gene-sized target, which should preferably be a cognate viral gene rather than a transgene so that the viral replication complex is replicating

a coevolved sequence. A second source of uncertainty derives from many data having been obtained from serial-passage experiments, during which bottlenecks, population replacements, or approaches to selection-mutation equilibrium may occur (Drake *et al.* 1998; Drake and Holland 1999). Also, selection and phenotypic masking need to be estimated to convert mutation frequencies into mutation rates (Drake *et al.* 1998). The final major source of uncertainty derives from a nearly total lack of information about riboviral mutational spectra (Drake and Holland 1999).

Here we present a new estimate of the spontaneous mutation rate of tobacco mosaic virus (TMV), which circumvents several of these problems. The estimate is based on a large mutational target, the 804-base TMV *MP* gene that encodes the viral movement protein, which is a cognate sequence for the viral replicase. The experimental system minimizes selection by providing the *MP* function *in trans* and reduces uncertainties about population history by avoiding serial passaging. We also present the first mutational spectrum of a ribovirus. This spectrum is very different from those of most DNA-based organisms.

MATERIALS AND METHODS

Virus and plants: TMV was derived from a biologically active cDNA clone (Dawson *et al.* 1986), the gift of W. O. Dawson (University of Florida). A mutant lethal for cell-to-cell movement (*MP562∇*) was derived from this clone by inserting four

[1] *Corresponding author:* Departamento de Biotecnología, Escuela Técnica Superior de Ingenieros Agrónomos, Ciudad Universitaria, 28040-Madrid, Spain. E-mail: fga@bit.etsia.upm.es

nucleotides (5′-CATG-3′) after *MP* nucleotide 562, thus creating a frameshift mutation.

Tobacco plants were mechanically inoculated with RNA transcribed from these clones by T7 RNA polymerase (Dawson *et al.* 1986). Four different tobacco genotypes were used. Plants of cv. Xanthi-nn (*nn* plants) are systemically infected by TMV. Plants of the nearly isogenic line Xanthi-NN (*NN* plants) carry the resistance gene *N* that limits TMV infection to necrotic local lesions (nll) on the inoculated leaf. Line 277 (Deom *et al.* 1987) of Xanthi-nn plants (*nn-MP* plants) constitutively produces movement protein (MP) from a transgene. Line 2005 (Deom *et al.* 1991) of Xanthi-NN plants (*NN-MP* plants) also constitutively produces MP from a transgene. These plants were the gift of R. N. Beachy (International Laboratory for Tropical Agricultural Biotechnology, St. Louis). In *nn-MP* and *NN-MP* plants, MP produced by the transgene complements the cell-to-cell movement of TMV mutants impaired in this function (Deom *et al.* 1987, 1991). Note that nll permit ready cloning and quantification of infectious virus particles in a manner similar to plaque formation by lytic animal viruses and bacteriophages.

Experimental procedure: Five microliters of a 0.2-μg/ml solution of RNA in 0.1 M Na_2HPO_4 transcribed from the wild-type TMV cDNA clone was inoculated by gently rubbing the upper epidermis of the interveinal spaces of fully expanded leaves of *nn-MP* plants. The infectivity of the inoculum was simultaneously quantified by nll assay (Matthews 1991, pp. 12–16) on *NN* plants. Inoculated plants were kept in a growth chamber at 25° with a 16-hr photoperiod. Three days after inoculation, 23.5 mg of fresh tissue from the inoculated leaf was ground in 6 ml of 0.1 M Na_2HPO_4. Part of this extract was used to determine the virus yield ($N - N_0$, where N is the size of virus population after multiplication in the nn-MP leaf and N_0 is the initial infective input) by nll assay in *NN* plants. The rest was used to inoculate fully expanded leaves of *NN-MP* plants. After 4 days in the greenhouse, nll appeared in these plants. Each lesion was ground in a drop of 0.1 M Na_2HPO_4 and used to inoculate leaves of both *NN* and *NN-MP* plants. If the TMV clone that caused the initial nll on the *NN-MP* plant was competent for cell-to-cell movement, new nll formed in leaves of both *NN* and *NN-MP* plants. If the TMV clone had a mutation that impaired cell-to-cell movement, nll formed in the leaves of *NN-MP* plants but not *NN* plants. Thus, this procedure detects mutations that impair cell-to-cell movement and are thus lethal for the virus. Because the only TMV factor involved in cell-to-cell movement is the MP (Citovsky 1999), the mutational target is the 804-base *MP* gene. This experimental procedure was repeated nine times, and in each case, a parallel control experiment was done in which the initial multiplication of the cDNA-derived clone was in leaves of *nn* plants. In four of these experiments, the percentage of infected cells in leaves of the systemic host in which the virus initially multiplied was estimated by counting immunofluorescent protoplasts isolated from the inoculated leaf after incubation with an antiserum directed against TMV and a fluorescein isothiocyanate-conjugated secondary antibody, as described in Moreno *et al.* (1997).

Total RNA was extracted from leaves of *MP*-transgenic plants infected with the mutants. The *MP* gene was amplified by reverse transcription and polymerase chain reaction (PCR) for 30 cycles with Taq polymerase, using the primers 5′-AGT TACTTGTGGTGAAGG-3′, complementary to bases 5877–5894 in the *CP* gene (*i.e.*, 165 bases downstream of the *MP* stop codon), and 5′-CTCCAGGTTCGTTTGTTT-3′, identical to bases 4831–4848 in the gene (58 bases upstream of the *MP* initiation codon), which encodes the 183-kD protein. PCR products were cloned into the plasmid pGEM-T (Sambrook *et al.* 1989; Promega, Madison, WI) and multiplied in *Escherichia coli* DH 5α. The *MP* sequence of some randomly chosen mutants was determined. Two or three independently obtained cDNA clones were sequenced for each of these mutants to avoid errors introduced during reverse transcription and PCR amplification. To estimate these errors, we also determined the *MP* sequence of 25 clones independently obtained from transcripts of the original TMV cDNA clone. Sequencing was done in an automatic DNA sequencer (ABI Prism; Perkin Elmer, Branchburg, NJ) using two sets of primers. The first set included primers at the forward and reverse sequencing sites of pUC/M13. The second set included primers 5′-GCTA TAACCACCCAGGAC-3′, identical to TMV bases 5281–5298, and 5′-GATCGAAACTTTGCAAGCC-3′ complementary to bases 5522–5540, both within *MP*.

RESULTS AND DISCUSSION

Frequency of mutants defective in cell-to-cell movement: Nll assays of the TMV population resulting from multiplication of the initial RNA inoculum transcribed from the TMV cDNA clone gave similar values for *nn* and *nn-MP* plants, an average of 1.8×10^7 infectious units per inoculated leaf interveinal area.

Next, nll derived from the initial multiplication in *nn-MP* plants were transferred to *NN* and to *NN-MP* plants. Of 1820 nll, 1777 induced nll in both plant lines, while 43 induced nll only in *NN-MP* plants. Thus, 43 out of 1820 clones carried mutations lethal for cell-to-cell movement and the frequency of such mutants was $f_{+MP} = 0.0236$. The error rate *in vitro* of the T7 RNA polymerase used to transcribe the RNA inoculum from the cDNA clone is 10^{-6}–10^{-5} per base (Boyer *et al.* 1992), so that mutants arising during transcription of the 804-base target might contribute 0.008–0.0008 to f_{+MP}. Therefore, the contribution to f_{+MP} from transcription errors is likely to be either small or negligible.

The mutational spectrum: Because no mutational spectrum was available for a ribovirus, recent estimates of their spontaneous mutation rates were derived, using a correction factor derived from DNA-based microbes (Drake 1991; Drake and Holland 1999). To ascertain the characteristics of spontaneous TMV mutations, we determined the *MP* sequence in 23 mutants randomly sampled from the total set. We also determined the *MP* sequence of 25 independent clones derived from transcription products of the original cDNA clone, finding only one transition [A → G at position 295 (A295G), producing arg → gly, phenotype unknown] in one clone; thus, the sum of errors of transcription, reverse transcription, and PCR is $1/20,100 = 5 \times 10^{-5}$. Because at least two independent clones were sequenced for each mutant, errors introduced during reverse transcription or PCR did not affect our results. Six of the sequenced mutants were obtained in *nn* plants (A mutants) and 17 in *nn-MP* plants (B mutants), but there was no significant difference between the A and B spectra. The similarity between the mutations in the A and B mutants shows that the B mutations were not conditional on the transgenic plant, for instance, as a result

```
                             ▽  A                                        Δ,Δ
AUGGCUCUAGUUGUUAAAGGAAAAGUGAAUAUCAAUGAGUUUAUCGACCUGUCAAAAAUG 060

      U  Δ                            Δ                 ▽
GAGAAGAUCUUACCGUCGAUGUUUACCCCUGUAAAGAGUGUUAUGUGUUCCAAAGUUGAU 120
                                                U
AAAAUAAUGGUUCAUGAGAAUGAGUCAUUGUCAGAGGUGAACCUUCUUAAAGGAGUUAAG 180

CUUAUUGAUAGUGGAUACGUCUGUUUAGCCGGUUUGGUCGUCACGGGCGAGUGGAACUUG 240
                                      U
CCUGACAAUUGCAGAGGAGGUGUGAGCGUGUGUCUGGUGGACAAAAGGAUGGAAAGAGCC 300

           A                        ▽        ▽
GACGAGGCCACUCUCGGAUCUUACUACACAGCAGCUGCAAAGAAAAGAUUUCAGUUCAAG 360

                          ▽,▽84A  C                        ▽
GUCGUUCCCAAUUAUGCUAUAACCACCCAGGACGCGAUGAAAAACGUCUGGCAAGUUUUA 420

GUUAAUAUUAGAAAAUGUGAAGAUGUCAGCGGGUUUCUGUCCGCUUUCUCUCGGAGUUUGUG 480
                  (Δ3)                U  ▽
UCGGUGUGUAUUGUUUAUAGAAAUAAUAUAAAAUUAGGUUUGAGAGAGAAGAUUACAAAC 540

GUGAGAGACGGAGGGCCCAUGGAACUUACAGAAGAAGUCGUUGAUGAGUUCAUGGAAGAU 600
                                  G644U•▽10A
                                  ▽4A,▽84A
GUCCCUAUGUCGAUCAGGCUUGCAAAGUUUCGAUCUCGAACCGGAAAAAAGAGUGAUGUC 660

                    C              GG
CGCAAAGGGAAAAAUAGUAGUAGUGAUCGGUCAGUGCCGAACAAGAACUAUAGAAAUGUU 720
          ▽
AAGGAUUUUGGAGGAAUGAGUUUUAAAAAGAAUAAUUUAAUCGAUGAUGAUUCGGAGGCU 780

ACUGUCGCCGAAUCGGAUUCGUUUUAA
```
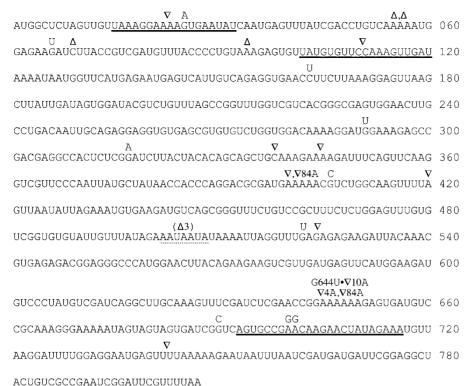
FIGURE 1.—The *MP* mutational spectrum. Single letters above the *MP* sequence indicate a base substitution at that postion (G, guanine; C, cytosine; A, adenine; U, uracil). ▽ indicates an insertion of a base immediately before, within, or immediately after the target base(s); Δ indicates a deletion of one base (either the unrepeated base beneath the symbol or a base from within the sequence repeat); (Δ3) indicates a deletion of either AAU or AUA from the dot-underlined sequence. Thick underlines indicate longer deletions, except that mutants A23 = Δ412(30–441), B8 = B216 = Δ68(84–151), and B1192 = Δ101(55–155) are not shown. The changes in mutant B74 converted the sequence 643-GGA$_6$GAGU-654 to GUA$_5$CA$_4$UGA$_4$GAGU, and the changes in mutant A400 converted the same sequence to GGA$_{78}$GA$_{13}$GU, perhaps by replicating the A$_6$ run once, recopying it 11 times, and then once more but including the following GA, and then twice more simply as A$_6$ again.

of recombination between viral RNA and transgenic mRNA. The pooled mutational spectrum appears in Figure 1. The mutations are listed by type in Table 1 and by distribution among mutants in Table 2.

Of the 11 base substitutions, 5 were transitions and 6 were transversions. Most of the mutations (8/11) arose at G sites. The misincorporated base was often (8/11) the same as a base adjacent to the mutated base; this ratio differs from the random expectation of $\frac{1}{4} + \frac{1}{4} - \frac{1}{16} = \frac{7}{16}$ (the probability that the upstream base is the same, or that the downstream base is the same, but that both the upstream and downstream bases are not the same, as the base adjacent to the mutated base) with $P = 0.0536$ in a two-tailed association test. One substitution (G690C in mutant A292) produced a synonymous codon and 10 produced amino acid substitutions; no chain-termination mutations were observed. The synonymous mutation was accompanied by a single-base deletion that would suffice to produce the mutant phenotype. Because 6 of the 10 missense mutations arose in mutants that contained one or more non-base-substitution mutations (NBSs; *i.e.*, an addition or deletion of one or more bases), some of these 6 may lack a mutant phenotype in the absence of the other mutation(s) in the same mutant.

Five of the 6 mutants from *nn* plants and 14 of the 17 mutants from *nn-MP* plants contained at least one NBS. The NBSs comprised 12 deletions and 12 insertions. NBSs appear to be favored by sequence context because most (8 out of 12) single-base insertions and deletions, and all 4 large insertions, were at runs of 2

or more of the same base (10 out of 12 at A runs), and 1 deletion was of either AAU or AUA from the fugue-like sequence AAUAAUA. All 3 single-base insertions that were not in runs were nevertheless duplications of a base in the target sequence, strongly suggesting iterative copying of a single template base; 2 of the 3 were adjacent to short mononucleotide runs, suggesting a role for mutagenesis by transient misalignment (KUNKEL and SONI 1988). None of the 4 large insertions were tandem duplications, which are often seen in other organisms, but were instead insertions of from 4 to 84 A's into A runs; 2 of these insertions were complex (see Figure 1 legend). The 7 large deletions all arose at or near direct

TABLE 1

Numbers of mutations by type

Kind of mutation	No.
Total	35
Base substitutions	11
Transitions	5
Transversions	6
At G	8
Additions and deletions	24
−1 base in runs	3
−1 base not in runs	1
−3 from AAUAAUA	1
+1 base in runs	5
+1 (XBY → XBBY)	3
Large deletions	7
Poly(A) insertions	4

TABLE 2

Nucleotide changes in specific mutants

Mutant	Change(s)
A2	∇C(110–111)
A23	Δ412(30–441)
A292	ΔA(54–58) + G690C
A394	G317A
A400	∇84A(645–650) + A701G
A939	ΔA(93–95)
B3	G66U
B8	Δ68(84–151)
B59	∇4A(645–650)
B74	G644U•∇10A(645–650)
B101	G522U
B211	∇A(21–24) + ∇C(338)
B216	Δ68(84–151)
B303	ΔC(69) + G406C + Δ24(693–716)
B420	G291U + Δ(AAU or AUA)(502–508)
B553	G25A
B895	Δ19(102–120) + ∇A(400–404) + ∇U(741–744)
B1084	Δ18(15–32) + ΔA(54–58) + A700G
B1098	∇A(343–346)
B1106	C163U + ∇84A(400–404)
B1184	∇A(420)
B1192	Δ101(55–155)
B1600	∇G(524)

A-mutants arose in *nn* plants, B-mutants in *nn-MP* plants. Entries of the form "G690C" describe a base substitution at the indicated position; *i.e.*, the G at position 690 is replaced by a C. Entries of the form 4A indicate that four A residues were added or deleted as indicated. (G, guanine; C, cytosine; A, adenine; U, uracil). Δ indicates a deletion of the indicated base(s), either of the single base or within a run of identical bases at the positions indicated in parentheses. ∇ indicates an insertion of the indicated base(s) immediately before, within, or immediately after the indicated target base(s); note that there are $n + 1$ possible interbase insertion sites for a target of n bases. • in mutant B74 indicates that the complex set of changes (see Figure 1 legend) probably arose in a single event.

sequence repeats. The large deletions in both B895 and B1084 arose between UU $\cdots$ U, beginning before either the first or the second U of the doublet and leaving one U undeleted. The large deletion in B303 arose at A $\cdots$ AAA, removing all four A residues. The A23 deletion of bases 30–441 is associated with repeats in two ways, first as a deletion at AA $\cdots$ A that removes all three A residues and second as GAAAAGUGAA U $\cdots$ GAAAUGUGAAGA, where the underlined sequence is deleted. (This repeat is an imperfect 9 out of 10 bases, and the deletion includes one G that is absent from the repeat.) The B1192 deletion of bases 54–154 or 55–155 is also associated with repeats in two ways, first as a deletion at AAAAA $\cdots$ A that removes five A residues and second as a deletion at UGUCA $\cdots$ UGUCAGA, where the underlined sequence is deleted. (Thus, the deletion includes a GA doublet that is absent from the repeat.) This deletion is in the same region as the deletion of bases 94–151 that appeared indepen-

dently in two clones obtained in different *nn-MP* plants (B8 and B216). The present spectrum is too sparsely populated for the analysis of hot spots, but three clones (B8, B216, and B1192) had large deletions within the same region (55–155), three clones (A400, B59, and B74) had insertions in the same region (645–650), and two (B895 and B1106) had insertions in a third region (399–404). The structures of most of the NBS mutants are compatible with polymerase slippage (KUNKEL 1992), as has been described for other RNA-dependent RNA polymerases (RAMÍREZ *et al.* 1995).

The high proportion of NBSs, and the fact that 50% of them involved from three to many bases, suggest that, in addition to being error prone, RNA-dependent RNA polymerases may have low processivity *in vivo*. The large deletions could have arisen by template switching in which the replication complex pauses, allowing several terminal bases of the progeny strand to dissociate from the template strand, and then reassociates with the same or another template at a different position. This mechanism has been invoked to explain the high frequency of recombination in riboviruses (BRUYERE *et al.* 2000) as well as the formation of defective genomes (WHITE and MORRIS 1999). Template switching during reverse transcription has been proposed to produce the same effects in retroviruses (PATHAK and HU 1997).

The TMV spontaneous mutational spectrum differs from those reported for most DNA-based organisms (GROGAN *et al.* 2001) in the much greater frequency of NBS mutations. Only 11/35 or 31% of these TMV mutations were base substitutions compared to a mean of 69% for most DNA-based organisms. Such a low percentage of base-substitution mutations has been reported only for spleen necrosis retrovirus (PATHAK and TEMIN 1990) and for the acidothermophilic archeaon *Sulfolobus acidocaldarius* (GROGAN *et al.* 2001). Until now, mutations in riboviruses have been considered to be mostly base substitutions with a high proportion of transitions, the resulting mutants forming a swarm of sequence variants around a consensus sequence (that is, a quasi-species; DOMINGO and HOLLAND 1997). Our results suggest that transitions and transversions arise at roughly the same frequency, although the sample size is too small to make the point strongly. It is also sometimes suggested that the high mutation rates of riboviruses are of adaptive value (*e.g.*, DOMINGO and HOLLAND 1997; DOMINGO 2000). In TMV, however, NBS mutations predominate, and most of these will be almost immediately lethal. Thus, the results presented here challenge widely accepted views of ribovirus mutation and evolution.

The most striking aspect of the mutation process in TMV is the high frequency of mutants carrying multiple mutations. Although the mutant frequency was $\sim$2.3% (or $\sim$5%, taking into account the incomplete rescue of the MP function by the transgene as described below), 8 of 23 mutants contained multiple mutations (5 mutants with two mutations each and 3 mutants with three each;

Table 2). There was no significant difference in the kinds of mutations in single and multiple mutants, and the *MP* multiple mutations tend to be widely separated from each other. This aspect of the mutational spectrum of TMV is discussed later.

The spontaneous mutation rate: The course of TMV infection in the inoculated tobacco leaf is very different from the course of infection by lytic viruses in cell suspensions. Progeny TMV is not released from infected cells and only a fraction of the particles move through the plasmodesmal connections to neighboring cells. After inoculation in the adaxial (upper) epidermis, infection reaches the underlying mesophyll and then spreads mostly radially in the plane of the leaf blade. From each infected cell, only a few neighboring cells, r on average, are infected. If the number of initially infected cells (first infection cycle) is n_0, then at the ith infection cycle, the number of newly infected cells n_i will be $n_0 r^{(i-1)}$. At the time of virus harvest, the total number of infected cells N will be $N = n_0(r^c - 1)/(r - 1)$, where c is the average number of consecutive infectious cycles that accumulate between the initial infection and the final harvest as in DRAKE and HOLLAND (1999). Rearranging,

$$c = \log[(N/n_0)(r - 1) + 1]/\log(r). \quad (1)$$

Let v be the average number of viruses accumulated per cell and let μ be the mutation rate to lethals per replication. Assume that there is negligible back mutation from lethal to functional. Then, at the ith infection cycle, $n_0 v r^{(i-1)}$ and $n_0 v r^{(i-1)} 2i\mu$ will be the numbers of total viruses and of mutants, respectively, produced in this infection cycle. The frequency of mutants, f, will be the ratio of the sums to the cth term of the two progressions with the general terms above:

$$f = 2\mu[(cr^c)/(r^c - 1) - 1/(r - 1)]. \quad (2)$$

Note that as r becomes large, this expression approaches that proposed by DRAKE and HOLLAND (1999), $f = 2\mu c$.

Equation 1 and, consequently, Equation 2 apply to *nn-MP* plants where *MP* mutants can move without restriction. In *nn* plants, cells infected only by a mutant defective in movement will not produce adjacent infections. Thus, at the ith infection cycle, the number of newly infected cells will be $n_0 r^{(i-1)}[1 - (2\mu)^\pi][1 - (4\mu)^\pi] \cdots [1 - \{2(i - 2)\mu\}^\pi]$, where π is the average number of viruses infecting a new cell. The ratio between n_i in *nn* and *nn-MP* plants will be largest at the last cycle of cell infection, c, and for $\pi = 1$. Under these conditions, in *nn* plants,

$$N \approx n_0 r^{(c-1)}[1 - \mu\{2 + 4 + \ldots + 2(c - 2)\}]$$
$$= n_0 r^{(c-1)}[1 - \mu(c - 1)(c - 2)]. \quad (3)$$

The difference in the value of c between *nn* and *nn-MP* plants is negligible if $c \ll \mu^{-0.5}$, which holds for our

experiments (see below). Thus, we use the same value of c for both types of plants.

Inoculating *nn-MP* leaves with TMV RNA transcripts produced ~30 foci of infection per inoculated leaf. Thus, $n_0 = 30$. By 3 days after inoculation, our counts of leaf cells that reacted with a TMV antiserum indicated that $1.57 \pm 0.35\%$ of epidermis and mesophyll cells in the inoculated leaf were infected by TMV. The total number of cells in a tobacco leaf is $\sim 3 \times 10^7$ (our estimates with inoculated leaves and MATTHEWS 1970, p. 166), and the inoculated interveinal space represents one-tenth of the leaf; thus, the number of infected cells after 3 days, N, is $\sim 4.5 \times 10^4$. Because virus multiplication in the systemic host yielded an average of 1.8×10^7 infectious units (see first paragraph of this section), these values give ~40 infectious units per cell. The efficiency of the nll assay is very low, $\sim 10^4$–10^5 particles per nll (GARCÍA-ARENAL *et al.* 1984). Thus, our data agree well with other reports of TMV accumulation in tobacco cells (MATTHEWS 1970, pp. 201–204). From the anatomy of the tobacco leaf (our histological observations with the inoculated plants and ESAU 1965, pp. 427–431), it can be concluded that, from each infected cell, infection will reach 3 to 6 abutting cells; thus, r will be 3–6 or, on average, 4.5, and viral foci will be small with only ~1.5% of leaf cells infected. For $r = 3$–6, the data above and Equation 1 give $c = 5.0$–7.3 with $c = 5.7$ for the average $r = 4.5$.

A traditional route from mutant frequency to mutation rate requires correcting the mutant frequency for those base substitutions that fail to produce a mutant phenotype, while assuming that virtually all base addition and deletion mutations do produce a mutant phenotype. When the mutant collection includes base substitutions generating chain-termination codons, total base substitutions can be estimated by applying an appropriate multiplier of the chain-termination mutations, such as $^{64}/_3$ (DRAKE 1991). In the absence of chain-termination base substitutions, as here, we can fall back only on data for DNA-based microbes, in which an average correction factor of $B = 4.78$ was obtained for the ratio of all substitutions to detected substitutions (GROGAN *et al.* 2001). Then the corrected *MP* mutation frequency $f_c = (\text{mutations/mutant}) f_{+MP}[\text{NBS} + B(\text{base substitutions})]/(\text{all mutations}) = (27/17)0.0236[19 + (4.78 \times 8)]/27 = 0.079$. The corrected frequency per average base for the 804-base *MP* gene is 0.99×10^{-4}. The TMV genome contains 6395 bases (GOELET *et al.* 1982), so the average mutation frequency per genome is 0.63. Using Equation 2 with $r = 3$–6 and $c = 5.0$–7.3, the average rate per base pair $\mu_b = 7.26$–10.3×10^{-6} (9.12×10^{-6} for $r = 4.5$) and the genomic rate $\mu_g = 0.049$–0.066 (0.058 for $r = 4.5$). Similar values are obtained using the simpler formulation $\mu = f/2c$ (DRAKE and HOLLAND 1999), in which case $\mu_b = 6.8$–9.9×10^{-6} and the genomic rate $\mu_g = 0.043$–0.063.

If the rescue of lethal mutants in *nn-MP* plants was incomplete, then the mutation rates would be underestimated. To analyze this possibility, RNA transcripts from wild-type and *MP562∇* cDNA clones were inoculated at similar concentrations in leaves of *nn-MP* plants. Extracts from inoculated leaves 3 days after inoculation were then used to inoculate *NN-MP* plants. When nll appeared, individual nll were transferred to leaves of *NN-MP* and *NN* plants and the ratio of nll that caused new nll only in *NN-MP* plants or in both plant genotypes was used to estimate the relative success of the two viral genotypes. This ratio was 1.8 ± 0.4 in three replicate experiments. A second experiment was done using virion RNA (rather than RNA transcripts) as the initial inoculum with similar results. Hence, the TMV mutant impaired for cell-to-cell movement accumulated in the leaves of *nn-MP* plants at a rate of ~ 0.5 of that of the wild type. These results are in good agreement with a previous report (DEOM *et al.* 1987). Thus, if this selection factor applies to all the mutations, about one-half of the lethal mutants would not be detected, and the mutation rate would be twice as large as estimated above, that is, $\mu_g = 0.10$–0.13.

A survey of mutation rates based on very small targets among lytic animal viruses yielded a median μ_g of 0.76 with a lower 96% confidence value of 0.18 (DRAKE and HOLLAND 1999). The latter value is a little higher than our value of 0.11. Our value might be slightly underestimated if heterogeneity in the architecture of cell-to-cell connections in the leaf mesophyll produced a lower value of *c* than the value we used. However, the TMV mutation rate may simply be lower than that for lytic riboviruses infecting mammalian cells. In support of this possibility, populations of lytic riboviruses were extinguished when their mutation frequencies were increased ~ 2.5-fold (HOLLAND *et al.* 1990), whereas the TMV mutation frequency could be increased ~ 80-fold by nitrous acid with little loss of infectivity and starting off with single-hit kinetics (GIERER and MUNDRY 1958).

Because *MP* mutants accumulate to a frequency of $\leq 5\%$ while about one-third of them contain multiple mutations, our values of μ_g are likely to be averages over a mixture of mostly lower but sometimes higher rates. An excess of multiple mutations has also been recorded in a mouse bearing a bacterial reporter gene (BUETTNER *et al.* 2000) and in an ordinary bacteriophage T4 gene (BEBENEK *et al.* 2001). The origin of the mutations in Table 2 can be approximated as a bimodal set containing a 98% majority with a mutation frequency of $\sim 10^{-2}$ and a 2% minority with a mutation frequency >1.

Two hypotheses might explain the large excess of multiple *MP* mutations. In one, a small fraction of TMV genomes encode mutator replicases, perhaps because of the combination of an intrinsically high mutation rate and the large fraction of the genome (76%) that encodes the replicase (GOELET *et al.* 1982) within which

many mutations might be expected to generate mutator mutations. This hypothesis is strongly disfavored because riboviruses are unlikely to survive such strong mutator mutations, whereas our MP mutants exhibit nll of normal size and time of appearance. (This hypothesis cannot be easily restored by proposing that the MP mutations are separated from mutator replicase genes by recombination.)

The alternative hypothesis is transient hypermutable replication similar to the "transient mutators" postulated by NINIO (1991). For instance, an error of transcription or translation might produce a nontransmissible mutator replicase, or a conventionally encoded mutator replicase might operate distributively within a pool of mostly nonmutator replicases, or the replication accessory proteins might be abnormal in amount or in their quaternary interactions, or an abnormal interaction might be established between the replicase and the primer template. Such transient hypermutation would probably have to be delimited to only a fraction of the genome to avoid producing lethal mutations in regions outside of the *MP* gene and thus preventing us from detecting the multiple *MP* mutations in the first place.

LITERATURE CITED

BEBENEK, A., H. K. DRESSMAN, G. T. CARVER, S. NG, V. PETROV *et al.*, 2001 Interacting fidelity defects in the replicative DNA polymerase of bacteriophage RB69. J. Biol. Chem. **276:** 10387–10397.

BOYER, J. C., K. BEBENEK and T. A. KUNKEL, 1992 Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. Proc. Natl. Acad. Sci. USA **89:** 6919–6923.

BRUYERE, A., M. WANTROBA, S. FLASINSKI, A. DZIANOTT and J. J. BUJARSKI, 2000 Frequent homologous recombination events between molecules of one RNA component in a multipartite RNA virus. J. Virol. **74:** 4214–4219.

BUETTNER, V. L., K. A. HILL, W. A. SCARINGE and S. S. SOMMER, 2000 Evidence that proximal multiple mutations in Big Blue transgenic mice are dependent events. Mutat. Res. **452:** 219–229.

BURCH, C. L., and L. CHAO, 2000 Evolvability of an RNA virus is determined by its mutational neighborhood. Nature **406:** 625–628.

CITOVSKY, V., 1999 Tobacco mosaic virus: a pioneer of cell-to-cell movement. Philos. Trans. R. Soc. Lond. B **354:** 637–643.

DAWSON, W. O., D. L. BECK, D. A. KNORR and G. L. GRANTHAM, 1986 cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. Proc. Natl. Acad. Sci. USA **83:** 1832–1836.

DEOM, C. M., M. J. OLIVER and R. M. BEACHY, 1987 The 30-kilodalton gene product of tobacco mosaic virus potentiates virus movement. Science **237:** 389–394.

DEOM, C. M., S. WOLF, C. A. HOLT, E. J. LUCAS and R. N. BEACHY, 1991 Altered function of the tobacco mosaic virus movement protein in a hypersensitive host. Virology **180:** 251–256.

DOMINGO, E., 2000 Viruses at the edge of adaptation. Virology **270:** 251–253.

DOMINGO, E., and J. J. HOLLAND, 1997 RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. **51:** 151–178.

DRAKE, J. W., 1991 A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88:** 7160–7164.

DRAKE, J. W., 1993 Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. USA **90:** 4171–4175.

DRAKE, J. W., and J. J. HOLLAND, 1999 Mutation rates among lytic RNA viruses. Proc. Natl. Acad. Sci. USA **96:** 13910–13913.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH and J. F. CROW, 1998 Rates of spontaneous mutation. Genetics **148:** 1667–1686.

ESAU, K., 1965 *Plant Anatomy.* John Wiley & Sons, New York.

GARCÍA-ARENAL, F., P. PALUKAITIS and M. ZAITLIN, 1984 Strains and mutants of tobacco mosaic virus are both found in virus derived from single-lesion passaged inoculum. Virology **167:** 201–206.

GIERER, A., and K. W. MUNDRY, 1958 Production of mutants of tobacco mosaic virus by chemical alteration of its ribonucleic acid *in vitro.* Nature **182:** 1457–1458.

GOELET, P., G. P. LOMONOSSOFF, P. J. BUTTLER, M. E. AKAM, M. J. GAIT *et al.,* 1982 Nucleotide sequence of tobacco mosaic virus RNA. Proc. Natl. Acad. Sci. USA **79:** 5818–5822.

GROGAN, D. W., G. T. CARVER and J. W. DRAKE, 2001 Genetic fidelity under harsh conditions: analysis of spontaneous mutation in the thermoacidophilic archaeon *Sulfolobus acidocaldarius.* Proc. Natl. Acad. Sci. USA **98:** 7928–7933.

HOLLAND, J. J., E. DOMINGO, J. C. DE LA TORRE and D. A. STEINHAUER, 1990 Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. J. Virol. **64:** 3960–3962.

KUNKEL, T. A., 1992 DNA replication fidelity. J. Biol. Chem. **267:** 18251–18254.

KUNKEL, T. A., and A. SONI, 1988 Mutagenesis by transient misalignment. J. Biol. Chem. **263:** 14784–14789.

MATTHEWS, R. E. F., 1970 *Plant Virology,* Ed. 1. Academic Press, San Diego.

MATTHEWS, R. E. F., 1991 *Plant Virology,* Ed. 3. Academic Press, San Diego.

MORENO, I. M., J. M. MALPICA, E. RODRÍGUEZ-CEREZO and F. GARCÍA-ARENAL, 1997 A mutation in tomato aspermy cucumovirus that abolishes cell-to-cell movement is maintained to high levels in the viral RNA population by complementation. J. Virol. **71:** 9157–9162.

MOYA, A., S. F. ELENA, A. BRACHO, R. MIRALLES and E. BARRIO, 2000 The evolution of RNA viruses: a population genetics view. Proc. Natl. Acad. Sci. USA **97:** 6967–6973.

NATHANSON, N., K. A. McGANN and J. WILESMITH, 1995 The evolution of virus diseases: their emergence, epidemicity and control, pp. 31–46 in *Molecular Basis of Virus Evolution,* edited by A. GIBBS, C. H. CALISHER and F. GARCÍA-ARENAL. Cambridge University Press, Cambridge, UK.

NINIO, J., 1991 Transient mutators: a semiquantitative analysis of the influence of translation and transcription errors on mutation rates. Genetics **129:** 957–962.

PATHAK, V. K., and W.-S. HU, 1997 "Might as well jump!": template switching by retroviral reverse transcriptase, defective genome formation, and recombination. Semin. Virol. **8:** 141–150.

PATHAK, V. K., and H. M. TEMIN, 1990 Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. Proc. Natl. Acad. Sci. USA **87:** 6019–6023.

RAMÍREZ, B. C., P. BARBIER, K. SÉRON, A.-L. HAENNI and F. BERNARDI, 1995 Molecular mechanisms of point mutation in RNA viruses, pp. 105–118 in *Molecular Basis of Virus Evolution,* edited by A. GIBBS, C. H. CALISHER and F. GARCÍA-ARENAL. Cambridge University Press, Cambridge, UK.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

WHITE, K. A., and T. J. MORRIS, 1999 Defective and defective interfering RNAs of monopartite plus-strand RNA plant viruses, pp. 1–17 in *Satellites and Defective Viral RNAs,* edited by P. K. VOGT and A. O. JACKSON. Springer, Berlin.

Communicating editor: S. LOVETT