# Tempo and Mode of Plant RNA Virus Escape from RNAi-Mediated Resistance

Guillaume Lafforgue,[1†] Fernando Martínez,[1†] Josep Sardanyés,[1†] Francisca de la Iglesia,[1] Qi-Wen Niu,[2] Shih-Shun Lin,[2§] Ricard V. Solé,[3,4,5] Nam-Hai Chua,[2] José-Antonio Daròs,[1] and Santiago F. Elena[1,5*]

*Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universitat Politècnica de València, València, Spain,*[1] *Laboratory of Plant Biology, Rockefeller University, New York NY, USA,*[2] *Complex Systems Laboratory, ICREA-Universitat Pompeu Fabra, Barcelona, Spain,*[3] *Instituto de Biología Evolutiva, Universitat Pompeu Fabra-Consejo Superior de Investigaciones Científicas, Barcelona, Spain,*[4] *and The Santa Fe Institute, Santa Fe NM, USA*[5]

[*]Corresponding author. Mailing address: IBMCP (CSIC-UPV), Campus UPV CPI 8E, Ingeniero Fausto Elio s/n, 46022 València, Spain. Phone: 34 963 877 895. Fax: 34 963 877 859. E-mail: sfelena@ibmcp.upv.es.

[†]G.L., F.M. and J.S. contributed equally to this study.

[¶]Current address: Institute of Biotechnology, National Taiwan University, Taipei, Taiwan.

**A biotechnological application of artificial microRNAs (amiR) is the generation of plants resistant to virus infection. This resistance has proven to be highly effective and sequence-specific. However, before these transgenic plants can be deployed in the fields, it is important to evaluate the likelihood of emergence of resistance-breaking mutants. Two issues are of particular interest: (i) whether such mutants can arise in non-transgenic plants that may act as reservoirs; and (ii) whether suboptimal expression of the transgene, resulting in sub-inhibitory concentrations of the amiR would favor the emergence of escape mutants. To address the first issue, we experimentally evolved independent lineages of *Turnip mosaic virus* (TuMV, family *Potyviridae*) in fully susceptible wild-type *Arabidopsis thaliana* plants and then simulated the spill over of the evolving virus to the fully resistant *A. thaliana* transgenic plants. To address the second issue, the evolution phase took place in transgenic plants that expressed the amiR at sub-inhibitory concentrations. Our results show that TuMV populations replicating in susceptible hosts accumulated resistance-breaking alleles that resulted in overcoming the resistance of fully resistant plants. The rate at which resistance was broken was 7 times faster for TuMV populations that experienced sub-inhibitory concentrations of the antiviral amiR. Molecular characterization of escape alleles showed that all contained at least one nucleotide substitution in the target sequence, generally a transition of the G-to-A and C-to-U types, with many instances of convergent molecular evolution. To better understand the viral population dynamics taking place within each host, as well as to evaluate relevant population genetic parameters, we performed *in silico* simulations of the experiments. Together, our results contribute to the rational management of amiR-based antiviral resistance in plants.**

The natural function of plant microRNAs (miRNA) is to regulate the abundance of target mRNAs by guiding the RNA-induced silencing complex (RISC) to cleave the corresponding complementary sequence. It has also been shown that changes within the miRNA 21-nt sequence do not affect its biogenesis and maturation (24,46) and this finding opened the possibility for redesigning the miRNA sequence to target different transcripts using different pre-miRNAs as backbones (38,39,50). One of the many applications of this technology is to produce artificial miRNAs (amiR) targeting viral genomes thus generating transgenic plants that are resistant to viral infection (38,39). Niu et al. (38) used the pre-miRNA159a precursor to express two amiR159 with sequences complementary to the RNA genome of *Turnip yellow mosaic virus* (TYMV) and of *Turnip mosaic virus* (TuMV), respectively. The amiR159-P69 was designed to target the sequence of the P69 silencing suppressor protein of TYMV. Similarly, the amiR159-HCPro was designed to target the sequence of TuMV silencing suppressor HC-Pro. Transgenic expression of the amiRs conferred high levels of specific resistance against the corresponding virus.

Similar to the case of virus-resistant transgenic plants, a gene-silencing mechanism (RNAi) has been used in *in vitro* assays as antiviral therapeutics to inhibit the replication of several human viruses, including *Human immunodeficiency virus* type 1 (12), *Hepatitis C virus* (30) and *Influenza A virus* (22). A major issue confronting these *in vitro* assays with mammalian viruses, however, has been the emergence of resistance variants (4,17,23,52). These variants differ from the wild-type (WT) virus by one or more point mutations in the 21-nt target sequence leading to imperfect matching and hence not being properly or efficiently processed by RISC (40,48,52). The RNAi machinery tolerates changes in certain positions of the 21-nt target but is particularly sensitive to changes in the central positions (particularly positions 9 and 11) (19,51). Lin et al. (31) extended these observations to the case of plant viruses. Using transgenic *Arabidopsis thaliana* plants expressing the amiR159-P69 described above and an engineered version of TuMV that carried an insert corresponding to the 21-nt TYMV P69 target, it was shown that conservation of positions 3 – 6, 9 and 12 were absolutely essential for RISC to cleave the viral genome; changes in positions 2, 10, 11, 13, 15, and 18 had a moderate effect on the cleavage efficiency whereas changes in the remaining nine positions had very minor effect in the processing efficiency (31). Furthermore, when viruses mutated at every one of the 21-nt target were allowed to replicate in the amiR159-P69 transgenic plants, deletions of variable length or additional changes at alternative sites arose and increased in frequency in the viral population, further jeopardizing the resistance of the transgenic plants (31). It is important to note that in this experiment the 21-nt target was non-coding thus selection only operated at the RNA level and was not constrained by protein-coding requirements.

All together, these results demonstrate that changes in certain sites within the 21-nt target may generate virus escape variants. Yet, the relevance, if any, of these escape variants in

natural viral populations remains to be established. In other words, to evaluate the viability of antiviral therapies based on the transgenic expression of amiRs in crops, it is essential to evaluate the likelihood of viral populations infecting susceptible reservoir host species to contain escape variants that may be subsequently transmitted to the transgenic crops by vectors. Moreover, it is also pivotal to evaluate whether variations in the expression of the amiRs transgenes, especially at sub-inhibitory concentrations, would affect the accumulation and evolution of escape viral mutants. More specifically, we are interested to address the following issues: What is the likelihood of escape mutations arising and accumulating in a WT host population? Does partial resistance favor the accumulation of escape mutants? What sites in the 21-nt target are more critical for escaping from the RNAi surveillance? What are the basic population genetic parameters governing the escape process? To address these issues we performed two sets of evolution experiments, together with their corresponding *in silico* simulations. In the first set, 25 independent TuMV populations evolving in fully susceptible WT *A. thaliana* plants were periodically tested for the presence of escape mutants by challenging fully resistant *A. thaliana* plants (Fig. 1A). We observed a steady increase in the number of evolving lineages able to breaking the resistance. The second set of experiments was similar to the first, except that 25 independent TuMV populations were evolved in partially susceptible *A. thaliana* plants expressing sub-inhibitory concentrations of amiR159-HCPro. We found that resistance breaking occurred faster in this second experiment. In all cases, changes in the 21-nt target sequence were observed. These results show that escape variants maintained at low frequency in sensitive and partially resistant transgenic plants were quickly filtered out upon transmission to fully resistant transgenic plants. The *in silico* simulation algorithm was used to evaluate population genetic parameters governing the evolutionary dynamics of escape mutants.

## MATERIALS AND METHODS

**Plant material and growth conditions.** Two transgenic *A. thaliana* Col-0 lines expressing amiR159-HCPro were used in this study: 10-4 and 12-4 (38). Seeds corresponded to homozygous T4 generations. Plants were maintained in a growth chamber under 16 h light 25 ºC/8 h darkness 22 ºC.

**Quantification of amiR159-HCPro expression.** Total RNA was extracted and purified from *A. thaliana* tissue using the Trizol reagent (Invitrogen). RNA was precipitated with isopropanol, resuspended in $H_2O$ and quantified by spectrophotometry. Quantification of amiR159-HCPro in RNA preparations was performed by RT-qPCR in triplicate (45). Standards were prepared by adding known amounts of the synthetic oligoribonucleotide [5'-r(ACUUGCUCACGCACUCGACUG)-3', corresponding in sequence to amiR159-HCPro] to a non-transgenic *A. thaliana* total RNA preparation. RT reactions were done in 10 µl with 100 ng

total RNA, 1 pmol primer I (5'-GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGAC**CAGTCG**-3' (in bold sequence complementary to amiR159-HCPro) and 30 U M-MuLV RT (Fermentas) incubating 10 min at 25 ºC, 45 min at 42 ºC, 10 min at 50 ºC, 5 min at 60 ºC and finally 15 min at 70 ºC. qPCR was done in 20 µl with 2 µl of RT reaction and 10 pmol each primer PII (5'-CGGCGG**ACTTGCTCACGCACT**-3', in bold sequence homologous to amiR159-HCPro) and PIII (5'-GTGCAGGGTCCGAGGT-3', homologous to sequence underlined in PI) using the Maxima SYBR Green Master Mix (Fermentas) and incubating 10 min at 95 ºC followed of 40 cycles of 15 s at 95 ºC, 30 s at 60 ºC and 30 s at 72 ºC.

**Population passages and evaluation of pathogenicity in *A. thaliana* 12-4 plants.** As a source of TuMV inoculum for all our experiments we used a large stock of infectious sap obtained from TuMV-infected *Nicotiana benthamiana* inoculated with a plasmid containing a TuMV cDNA (Genbank accession no. AF530055.2) under the control of *Cauliflower mosaic virus* 35S promoter. This TuMV sequence variant corresponds to isolate YC5 from calla lily (*Zantedeschia* sp.) (10). Plant infectious saps were obtained grinding infected tissue in a mortar with 20 volumes of grinding buffer 50 mM potassium phosphate pH 7.0, 3% PEG6000.

Fig. 1A summarizes the experimental design for the evolution experiments. Aliquots of 5 µL of 10% Carborundum in grinding buffer were applied on three different *A. thaliana* leaves and inoculation was done mechanically by gently rubbing with a cotton swab soaked with infectious sap. Twenty-five WT *A. thaliana* and twenty-five 10-4 transgenic plants were initially inoculated. Each plant represented the starting point for an independent evolution lineage. Fourteen dpi, symptomatic tissue was collected for each lineage and homogenized in grinding buffer. A portion of the resulting saps was used to inoculate the next set of plants. The remaining portion of the homogenized sap was frozen at −80 ºC for further characterization. A third portion was used in the challenging experiments designed to estimate pathogenicity in 12-4 plants. This procedure was repeated until all 50 evolutionary lineages overcame the resistance in 12-4 line. Once a lineage was capable of breaking resistance, it was removed from the passaging experiment.

For the pathogenicity test experiments, 20 plants of the 12-4 line were inoculated as described above. Plants were visually checked for the presence of symptoms 14 dpi and the frequency of infected plants, that is pathogenicity, recorded. These challenging experiments were performed after every evolutionary passage for each one of the 50 evolving lineages. A pilot experiment showed that infection always concurred with symptoms development. A lineage was considered as able of breaking resistance if at least one 12-4 plant showed symptoms.

**Sequence analysis of the 21-nt target region.** The region around the 21-nt target of the amiR159-HCPro was sequenced in virus populations breaking resistance. Total RNA from

infected *A. thaliana* 12-4 transgenic plants was purified using silica columns (Zymo Research) and a viral cDNA amplified by RT-PCR. RT reactions were carried out in 10 µL with 50 µl M-MuLV RT and 5 pmol primer IV (5'-CCTGGTGACAGTAAAGCATATAATGG-3') for 45 min at 42 ºC, 5 min at 50 ºC and 5 min at 60 ºC. One µL of the RT reaction was used for PCR amplification in 20 µL with 0.4 U Phusion DNA polymerase (Finnzymes) and 10 pmol each primer PV (5'-GACAATGAGTCACAAGATTGTGCACTTT-3') and PVI (5'-CATGAGTGTCCTCCCATTCTGTCCC-3') incubating 30 s at 98 ºC, 30 cycles of 10 s at 98 ºC, 30 s at 55 ºC and 30 s at 72 ºC, and a final extension of 10 min at 72 ºC. Amplification products were separated by electrophoresis in a 1% agarose gel and the TuMV cDNAs matching the expected 1427 bp eluted and sequenced with primer PVII (5'-AAACGATTCTTCAGCAACTACTTTG-3').

**Simulation algorithm.** The experiments were simulated using a bit string Monte Carlo model (20) in which digital genomes were represented by binary strings, $\boldsymbol{S}$, of length $L$ = 31 bits. The digital genomes explicitly considered the 21-nt of the amiR159-HCPro target and added 10 more bits, each corresponding to one of the 10 viral cistrons (Fig. 1B). We made this distinction to disentangle the effects due to mutation in the target (evaluated at the challenging step of the experiment) from those associated to changes in other viral genes and that determine the overall fitness of the virus. Maximum string population size was set to $N_{max}$ = 5000 genomes. As in the experiments, the simulation model considered 25 independent lineages (Fig. 1B). Each lineage started with a sample of size $N < N_{max}$ of WT genomes. For each lineage we let the population to experience $\tau$ replication events. At each event, two locations in the population are randomly chosen. If location $i$ already contains a string, it is copied to site $j$ with probability $P_{ij} = \frac{1}{1+\exp(-\Delta f_{ij}/T)}$ that depends on the fitness difference $\Delta f_{ij} = f_i - f_j$ between strings $\boldsymbol{S}_i$ and $\boldsymbol{S}_j$ (if site $j$ is empty $f_j = 0$). $T$ is the Boltzman temperature, which is a measure of the noise tied to replication events and it was fixed to $T = 0.2$. The fitness of a given string, $\boldsymbol{S}_k$, is obtained from the binary composition of the 10 loci. We consider four types of deleterious fitness landscapes: the standard additive, antagonistic and synergistic ones, plus one in which mutations in the bits representing the 10 viral cistrons were considered as lethal. For the three deleterious landscapes we compute the fitness as $f_k = 1 - d_H^{\xi}/10$, where $d_H$ is the Hamming distance (i.e., how many different bits we have) between sequence $k$ and the corresponding loci of the WT genome. $\xi$ measures the sign and strength of epistasis: $\xi = 1$ if additive, $\xi < 1$ if antagonistic and $\xi > 1$ if synergistic (42). During replication, each bit of the amiR159-HCPro target can mutate with probability $\mu$. The other 10 loci of the strings mutate with probability $\mu_{l_i} = 3\mu\nu/2l_i$, where $l_i$ is the length of locus $i$ and the 2/3 is introduced to consider, as a first approximation, that mutations at third codon positions are neutral. This

correction was done to ensure that all loci mutate proportionally to their length. In order to differentiate between the experiments carried out in WT and 10-4 *A. thaliana* plants, we consider that if the string chosen for replication is the WT genome it will be degraded with probability $\varepsilon = 0$ for simulations in WT and $\varepsilon > 0$ for simulations in 10-4.

As previously mentioned, for each lineage we let the population to evolve over $\tau$ replication events according to the previous rules. Then, we take two random samples of size $N$ (Fig. 1B). The first sample is used to initiate the next population (simulating the next passage in the experimental evolutionary lineages) until resistance is broken. The second sample is used to evaluate the likelihood of resistance-breaking as follows. For each string $\boldsymbol{S}_i$ in the second sample we evaluated its pathogenicity as $\theta(\mathbf{S}_i) = 1 - \prod_{k=1}^{21}\left[1 - \lambda(S_{ik})\right]$, being $\lambda(S_{ik})$ the empirical probability that a change in position $k$ of the 21-nt target will be an escape mutation (frequency data from Fig. 4 corrected using the Laplace estimator). Next we evaluated the likelihood of resistance breaking for this second sample, after 20 trials (the number of plants inoculated during the challenging experiments), as $P_b = 1 - \left(1 - \langle p \rangle\right)^{20}$, where $\langle p \rangle = \frac{1}{N}\sum_{i=1}^{N}\theta(\mathbf{S}_i)$ is the average pathogenicity of all the strains contained in the sample. If $P_b \geq 0.05$ we assumed that resistance was broken. For a sample of 20 plants this threshold means at least one plant becoming symptomatic.

**Data fitting and parameter inference.** To fit the experimental data to the simulation model and to infer relevant population parameters we used an optimization algorithm (OA) (33) that systematically searched the parameter space defined by $C = \{\tau, \mu, N, \xi, \varepsilon\}$ as follows. First, we defined a starting population of 150 parameter sets $C_1(0)$, $C_2(0)$, …, $C_{150}(0)$. The parameter values for each one of these $C_h(0)$ parameter sets was randomly assigned within the following ranges: $1 \leq \tau \leq 10^5$, $1 \leq N \leq N_{max}$, $10^{-7} \leq \mu \leq 10^{-3}$, $0.2 \leq \xi \leq 1.8$, and $0.1 \leq \varepsilon \leq 0.5$. For each one of these parameter sets, we run the simulation algorithm described in the two previous paragraphs. At the end of each simulation we compared the observed cumulative frequencies at passage $j$ shown in Fig. 3, $\rho_{obs}(j)$, with those obtained in the simulation, $\rho_i(j)$, using the equation $d_h = \sum_{j=1}^{28}\left|\rho_h(j) - \rho_{obs}(j)\right|$, that represents a distance value between the empirical and the simulated data. This procedure generates a vector of 150 $d_h$ values between observed and simulated data. Then we computed the average distance $D = \frac{1}{150}\sum_{h=1}^{150} d_h$ from all the 150 parameter sets and chose those sets with a distance smaller than $D$ as starting point for the next iteration of the OA, $C_i(1)$. Since less than 150 parameter sets are left for the next iteration, the rest of sets are generated by adding small perturbations to the retained parameter sets. The

whole process is repeated until no change is observed in $\langle d(t) \rangle$ after $t$ iterations of the OA. Notice that for WT plants $C = \{\tau, \mu, N, \xi\}$ since amiR-mediated degradation was fixed to $\varepsilon = 0$.

## RESULTS

***A. thaliana* lines 10-4 and 12-4 differ in amiR159-HCPro expression and susceptibility to TuMV infection.** First, we evaluated whether TuMV had the same level of pathogenicity, $p$, in both *A. thaliana* transgenic lines 10-4 and 12-4. All plants were inoculated at Boyes' stage 1.03 (i.e., three rosette leaves are greater than 1 mm in length) (6) and with TuMV infectious sap applied by gentle abrasion on leaves of the same position on the plant. None of the 30 inoculated 12-4 plants developed symptoms of infection 14 days post inoculation (dpi) ($p$ = 0.000±0.048; ±95% CI computed using Wald adjusted method). In sharp contrast 152 out of 218 inoculated 10-4 plants developed obvious symptoms after the same period of time ($p$ = 0.697±0.063). The difference between results obtained from 10-4 and from 12-4 was highly significant (Fisher's exact test $P < 0.001$). Significantly, TuMV pathogenicity in 10-4 plants was only 11.80% smaller than in the fully susceptible WT plants (166 out of 210, $p$ = 0.791±0.057), although this small difference was still statistically significant (Fisher's exact test $P = 0.035$).

To elucidate the difference in pathogenicity between the two transgenic lines, we first evaluated whether there was any difference in the overall accumulation of amiR159-HCPro. To this end, we analyzed by RT-qPCR the concentration of the amiR accumulated in sets of each transgenic line at Boyes' stage 1.03 (the developmental stage at which the above pathogenicity tests were performed). Twelve 12-4 plants and eleven 10-4 plants were analyzed; three independent quantifications were obtained for each plant. The data were analyzed using a general linear model (GLM) using "plant genotype" as main random factor and "plant replicate" nested within plant genotype. This analysis showed that significant heterogeneity exists among plants of the same genotype ($\chi^2$ = 264.698, 21 d.f., $P < 0.001$). Despite this heterogeneity, the differences among genotypes were highly significant ($\chi^2$ = 389.442, 2 d.f., $P < 0.001$). On average, 12-4 plants accumulated 111.367±6.998 pg amiR159-HCPro per mg of plant tissue (hereafter errors will represent ±1 SEM), whereas 10-4 plants accumulated 4.961±1.370 pg/mg (i.e, 22.45-fold less compared to 12-4 plants).

Second, we characterized the temporal pattern of accumulation of amiR159-HCPro in 10-4 leaves whose developmental stage was equivalent to those inoculated in the pathogenicity tests (e.g., the zero in the ordinate corresponds to Boyes' stage 1.03). Four independent 10-4 plants were analyzed at each time point and the estimates averaged among plants. Fig. 2A shows that the amount of amiR159-HCPro accumulated per ng of plant total RNA increased in a non-linear fashion as a leaf developed. Indeed, during the first days of the experiment the increase in

amiR159-HCPro concentration was minor but accumulation significantly accelerated 10 days after the beginning of the experiment (i.e., accumulation was not linear but exponential; Fig. 2A).

Third, we sought for differences in the amount of amiR159-HCPro at different leaves of the same plants (at Boyes' stage 1.06, i.e., six rosette leaves are greater than1 mm) to see whether this accumulation pattern was consistent among plants. To do so, we estimated the concentration of amiR in each of six leaves from four different plants. Fig. 2B shows the observed pattern of amiR159-HCPro accumulation. A GLM model in which "plant" was treated as a random factor and "leaf" as a covariable, highlighted several interesting results. First, the amount of amiR159-HCPro significantly varied among leaves at different developmental stages, significantly increasing as leaves become older ($\chi^2$ = 88.713, 1 d.f., $P$ < 0.001). Second, in agreement with our first test, plants were heterogeneous in their average amount of accumulated amiR159-HCPro ($\chi^2$ = 497.603, 4 d.f., $P$ < 0.001). Third, differences existed among plants in the rate at which amiR159-HCPro increased in concentration (test for homogeneity of slopes: $\chi^2$ = 96.531, 3 d.f., $P$ < 0.001). In other words, early stochastic events during development determined the initial amount of amiR159-HCPro that would characterize a leaf, and these differences further amplify as leaves expand and develop.

Therefore, all these analyses lead to the conclusion that the transgenic line 10-4 shows incomplete genetic penetrance (i.e., not all individual transgenic plants are resistant) and variable gene expressivity for resistance (i.e., not all resistant individuals express the amiR159-HCPro at the same level). By contrast, line 12-4 shows complete genetic penetrance of the resistance trait. These phenotypic differences are due to differences in the amount and timing of expression of the amiR159-HCPro. Rather than an issue, we will take full advantage of 10-4 peculiarity to evaluate the effect of evolving TuMV populations under sub-inhibitory and variable expression of the amiR159-HCPro.

**Resistance breaking in TuMV populations evolving in WT *A. thaliana* plants at the mutation-drift balance.** We aimed to evaluate the likelihood that TuMV populations replicating and evolving in fully susceptible WT *A. thaliana* hosts contained escape mutants able to overcoming the resistance mediated by the amiR159-HCPro. To this end, 25 independent evolution lineages were founded by inoculating WT *A. thaliana* plants with sap obtained from a pool of *N. benthamiana* plants previously inoculated with an infectious TuMV cDNA genome. Therefore, the amount of genetic variability in the inoculum will not be zero but the lowest technically possible. All plants were inoculated with the same amount of this infectious sap. All plants became infected as confirmed by the presence of symptoms. Every 14 dpi, infected plants were sampled; one portion of the sample was used to inoculate the following set of plants, another portion was stored for future analyses and a third portion used to challenge twenty 12-4 transgenic plants per evolving lineage (total 20 × 25 = 500 plants per

challenge experiment; see Fig. 1). The pathogenicity of each evolving lineage at each passage was evaluated by inspection of symptoms; a lineage was considered as able of breaking the resistance if it was able of infecting at least one 12-4 plant in the challenging experiments (i.e., pathogenicity ≥ 0.05). We hypothesized that mutants in the amiR159-HCPro target will arise and stay in the population at the mutation-drift balance and they will be transferred to the 12-4 plants during challenging in a rather stochastic manner. The black line in Fig. 3 shows the cumulative frequency of lineages that overcame resistance at each passage. The first break out occurred at passage 6, and all 25 lineages were capable of breaking resistance after 28 passages. A Kaplan-Meier regression shows that the median time for resistance-breaking was of 14.000±0.480 passages in WT *A. thaliana*.

**Resistance breaking in TuMV populations evolving in partially susceptible *A. thaliana* 10-4 plants.** Next, we sought to evaluate the effect that TuMV replication under sub-inhibitory concentrations of the amiR159-HCPro had on resistance durability. To this end, we repeated the evolution experiment by performing serial passages in the partially resistant *A. thaliana* 10-4 plants; all other operations were kept identical. We reasoned that in this case the TuMV populations infecting plants would be under the selective pressure imposed by the presence of the amiR in the cells but at concentrations that may still allow viral replication. We predicted that under such situation escape mutations would have a selective advantage and accumulate in the population at the mutation-selection-drift balance, at frequencies higher than in the previous experiment. This would allow for a faster resistance-breaking after challenging the 12-4 plants. The red line in Fig. 3 illustrates the time course accumulation of lineages able to breaking the resistance. As we predicted, lineages broke resistance faster than in the previous experiment, with many of them already containing escape mutants after the first passage and all 25 being able to do so after only eight passages. A Kaplan-Meier regression shows that, in this case, the median time for resistance-breaking was 2.000±0.343 passages in 10-4, a value that is significantly smaller than that obtained for the 12-4-evolved lineages (Mantel-Cox test: $\chi^2$ = 54.971, 1 d.f., $P < 0.001$).

**Changes in the amiR159-HCPro target.** After determining that a TuMV lineage was capable of escaping from amiR159-HCPro-mediated resistance, we sought to characterize the genetic changes associated to its new phenotype. Based on results by Lin et al. with TuMV (31), supported by previous accumulated knowledge from HIV-1 (4,17,52-48) and poliovirus (23) cell culture experiments, we hypothesized that in all cases, the dominant TuMV genotype in the infected 12-4 plants after challenging would carry at least one mutation in the target sequence. To test this expectation, we obtained the 21-nt target consensus sequence for the viral population replicating in each 12-4 plants. Tables 1 and 2 show the different escape alleles found in TuMV populations evolving in WT *A. thaliana* and 10-4, respectively. Regarding Table 1, a total of 10 different alleles have been characterized, although four of them (alleles 1,

2, 3, and 4) have been pervasively seen in more than one lineage, in a clear example of convergent evolution. The two most common nucleotide substitutions were a synonymous one at target site 11 (in 10 cases) and a nonsynonymous one at position 12 (in 7 instances) that gave rise to a conservative amino acid replacement V to M in the HC-Pro protein. Half of the alleles contained a single substitution (1, 2, 4, 8, and 10), whereas the other half contained two mutations. Four of these substitutions were synonymous and eight were associated to amino acid replacements. Interestingly, lineages 6, 11 and 23 all showed a polymorphism at position 20 of the target. In all three cases one of the coexisting alleles was a synonymous substitution, whereas the other one involved a conservative amino acid replacement K to N in the HC-Pro.

Regarding Table 2, seven escape alleles were identified in the TuMV populations evolved in the partially resistant 10-4 plants. Four of them were not observed in the populations evolving in WT *A. thaliana* (alleles 11, 12, 13 and 14), although only one of the mutations involved in these alleles was not previously observed (the A to C nonsynonymous change a position 19 of allele 11). The two most common alleles in this experiment were also those observed in the first experiment (1 and 2): the synonymous substitution at position 11 of the target (in 12 cases) and the second most abundant one the nonsynonymous replacement at site 12 (in 8 cases).

Pooling data from both experiments, 52 of the 55 observed mutations were transitions, with G to A and C to U changes dominating the mutational spectrum. Consistent with the principle that transitions are biochemically more likely than transversions, the maximum composite likelihood estimate of the overall transitions to transversions rates ratio is 14.176. This excess also occurs when purines (20.599) and pyrimidines (40.639) are considered separately. It is well known that viral coding regions show an excess of transitions over transversions (9,26,31,43). Three reasons can account for this bias: (i) the underlying mechanisms of mutation render transitions easier than transversions, (ii) the redundancy of the genetic code is expected to make the average effect of transitions smaller than of transversions and (iii) RNA editing by deaminase-like enzymes have been shown to induce transition mutations in single-stranded regions of certain viral genomes (3).

Convergent evolution would imply that the frequency distribution of changes along the 21-nt target should be similar in both experiments. Fig. 4 shows these distributions for both types of TuMV populations. A homogeneity test detected no differences among both pattern distributions ($\chi^2$ = 8.388, 11 d.f., $P$ = 0.678), thus supporting the notion of widespread convergent evolution, likely driven by the selective advantage of mutations at sites 11 and 12 of the target.

**Estimates of population genetic parameters by *in silico* simulations.** To provide new insights into the above results as well as to evaluate the range of population parameters compatible with our observations, we simulated the two evolution experiments using digital viral genomes replicating, mutating and subjected to transmission bottlenecks as in the

experiments (see Methods and Fig. 1B). We performed a search of parameter space using an optimization algorithm (OA) to find a set of parameters that minimized the distance between the data shown in Fig. 3 and those simulated. For the simulations of the evolution experiments carried out in WT *A. thaliana* plants (i.e., without sequence-specific degradation), we analyzed a total of 393 runs of the EA: 129 assuming that mutations had additive effects, 210 runs assuming they interacted epistatically and 54 assuming mutations outside the target and affecting other genes being lethal. Each run of the OA consisted of 400 generations, with a population of parameter sets of 150, resulting in more than 25 million simulations. The parameter set that generated the lowest and more robust distance ($d$ = 0.56, $R^2$ = 0.976, $F_{1,27}$ = 1114.571, $P < 0.001$) between the experiments and the simulation model was obtained with the additive fitness landscape with parameters: $\langle\tau\rangle$ = 13918.23±75.64 viral replications between passages, $\langle\mu\rangle = (4.11\pm0.33)\times10^{-5}$ mutations per site and generation and $\langle N_e\rangle$ = 956.89±23.76 digital viruses transmitted per bottleneck event (i.e., ~19% of the total population). Fig. 5A shows the results of the simulation obtained with this set of parameters. The simulated values of the frequency of lineages escaping from the amiR159-HCPro are shown with red dots on top of the black line that represents the experimental data.

For the evolution experiments in the 10-4 partially resistant plants (i.e., with sequence-specific degradation), we followed the same procedure, although we restricted the study only to the additive fitness landscape (that gave the best fit for the WT plants) and added a degradation rate $\varepsilon > 0$ to the parameter set. This degradation rate simulated the assumption that 10-4 plants expressed the amiR159-HCPro and hence the silencing machinery may still be capable of degrading a fraction of the viral population (i.e., the strings containing a WT target sequence are degraded with probability $\varepsilon$; see Methods). For this case, we run 150 replicas of the OA; thus exploring a total of six million simulations. Among all these simulations, the parameter combination providing the smallest distance between experimental and simulated data ($d$ = 0.16, $R^2$ = 0.995, $F_{1,7}$ = 733.253, $P < 0.001$) was: $\langle\tau\rangle$ = 5629.51±63.79, $\langle\mu\rangle = (7.69\pm1.12)\times10^{-5}$, $\langle N_e\rangle$ = 68.61±11.78 (i.e., ~1.4% of the potential maximum population size), and $\langle\varepsilon\rangle$ = 0.223±0.098 per genome. The best fitting to the experimental data is shown in Fig. 5B. As before, the red dots represent the simulated values for this parameter set. Not surprisingly, the mutation rates estimated for both experiments are on the same order of magnitude and close to the only experimental value reported for potyviruses (43).

The degradation of genomes containing non-mutated amiR159-HCPro targets in 10-4 plants has two interlinked effects. First, a reduction of 92.83% in $\langle N_e\rangle$: not all genomes contained in the inoculum were able of replicating in the partially susceptible plants and a certain fraction is degraded. Second, we expect an apparent reduction in the number of viral replication events supported by the two plant genotypes. In Col-0 plants all the viral progeny produced may

eventually contribute to future replications. At the other hand, in 10-4 plants, we expect part of the progeny to be degraded by the amiRs and hence not contributing to future replications. The model catches this expectation, and shows that 10-4 plants supported ~2.5 times fewer replication events than the WT plants. Consistently, the viral populations replicating in 10-4 plants did not reach carrying capacity and, therefore, the number of transmitted genomes to the next infection cycle was 13.95-fold smaller. This reduction in the size of the transmitted population enhances the effect of genetic drift in the 10-4 lineages. This being said, it is important to recall that two different evolutionary regimes are at play on each plant genotype. In the fully susceptible WT plants, purifying selection and drift should be the only factor affecting allele frequencies, since mutations in the target would be either deleterious or neutral; deleterious alleles will not reach high frequencies. The time to fixation of a neutral allele whose initial frequency is negligible is $4\langle N_e \rangle$ = 3827.56 generations (27), which is smaller than the estimated number of viral replications $\langle \tau \rangle$, thus making likely that some neutral alleles in the target would drift to high frequencies in the population. By contrast, in partially resistant 10-4 plants positive selection also enters in the picture, since escape alleles will clearly be beneficial in the presence of the amiR159-HCPro. Indeed, we can estimate that the average selection coefficient of such a beneficial allele to survive drift should be $\langle s \rangle > 1/\langle N_e \rangle = 0.015$ (27), a low value that ensures that many mutations conferring resistance will survive drift.

Finally, the mutation rates estimated for both experiments are in the range $4 - 8 \times 10^{-5}$ per site, values that are very close to recent estimates obtained for other potyvirus, *Tobacco etch virus* (41,43), and, more generally, for other plant viruses (33,41). This excellent agreement gives support to the validity of our modeling approach as well as to the conclusions derived from it.

## DISCUSSION

The long-term effectiveness of genetic resistances against plant viruses is constantly being challenged by the evolutionary potential of RNA viruses (21), creating the necessity to develop new resistance strategies. In the early nineties it was recognized that transgenic expression of virus-derived sequences resulted in a highly efficient defense against plant viruses (32), being this defense mediated by the post-transcriptional degradation of RNA genomes guided by virus-derived small interfering RNAs (siRNAs) (25). In recent years, plants have been engineered using this approach that are resistant to virus infection (11,18,29,36,49). However, transgenic expression of long viral sequences raises biosafety concerns regarding the possibility of recombination and generation of new potentially virulent strains (44). Taking advantage of the functional similarities between siRNA and miRNAs, Niu et al. (38) modified the backbone of *A. thaliana* pre-miRNA159 replacing it by short 21-nt viral sequences, resulting in highly specific

resistant plants. This approach has at least two advantages compared with the expression of long viral sequences. First, it should have fewer off-target effects, as the amiR sequences are shorter than those required for homology-dependent gene silencing. Second, recombination is not a concern anymore given the shortness of the amiRs. However, this approach may still raise a major concern: the high mutability of RNA viruses makes it likely that resistant virus variants will emerge, as already observed in *in vitro* experiments with mammalian viruses (4,12,17,22,23,30,40,48,52). The objective of the present work was to evaluate the likelihood of emergence of such escape variants in viral populations replicating in fully susceptible reservoir plants as well as in plants expressing the resistance at sub-inhibitory levels. Toward accomplishing this objective we have performed two different evolution experiments using the pathosystem TuMV/*A. thaliana*, together with the *in silico* computational models simulating both evolution experiments. The first experiment was designed to mimic the situation in which crops of resistant transgenic plants coexisted with crops of fully susceptible ones that acted as virus reservoirs. In this case, we observed an increase in the number of evolving lineages that were capable of successfully infecting the fully resistant host. Such escape mutants should most likely be neutral, or perhaps even slightly deleterious maintained by complementation, in the evolving population. Our second experiment was aimed to mimic the situation in which the expression level of the antiviral amiR was variable among plants, with some of them having suboptimal levels that allow for virus replication and selection of escape variants. In the second case we found that these populations accumulated escape mutations at a much higher frequency and, therefore, were able to successfully infecting the fully resistant hosts at earlier times in virus evolution. This second result was highly predictable since it recapitulates the evolution of bacteria at antibiotic concentrations below the minimum inhibitory concentration (14) and which has been solidly established. At sub-inhibitory concentrations of the antiviral amiR, mutant genotypes gain a fitness advantage given their ability to replicate despite the presence of the antiviral amiR whereas wild-type genomes may still suffer from the inhibitory effects. This fitness advantage results in accumulation of escape alleles above what is expected for the first experiment.

In all fifty cases, molecular characterization of the escape mutants confirmed the presence of mutations in the amiR159-HCPro target. In agreement with the mutant spectra described for other viruses, including TuMV, we have observed an excess of transition mutations (9,26,31,43). Particularly interesting is the fact that G to A and C to U transitions represented 95% of all mutations observed. These transitions are from the particular type induced by cellular cytidine deaminases involved in innate immune responses to viral infection (13), a phenomenon particularly well described in HIV-1 and other retroviruses (16) but hitherto not described for RNA viruses. This observation is in good agreement to that of Lin et al. (31), thus giving additional support to the hypothesis that as an antiviral strategy plants may have an

RNA-editing system that induces hypermutagenesis in viral genomes. We note that *A. thaliana* contains a family of nine paralogous genes that have been annotated as cytidine deaminases owing to their homology to the *CDA1* locus (47).

Indeed, mutations were unevenly distributed along the 21-nt target and mainly concentrated in positions 11 and 12, in a clear case of convergent evolution at the molecular level. Convergent evolution is a widespread phenomenon in RNA viruses both in experimental (8,15,53) as well as in natural (5,35) populations. Although these convergences could in principle be explained from a neutralist point of view as resulting from mutational bias, it is more likely that parallel and convergent substitutions are adaptive. This pattern would result from viruses facing identical selective pressures, with few alternative adaptive pathways, as expected for their simple and compacted genomes. In agreement with our observation, Lin et al. (31) classified position 11 as moderately crucial and position 12 as critical for resistance-breaking, although other sites qualified as crucial do not show high frequency of variation in our experiments. In contrast to the study of Lin et al. (31) in which the targeted sequence was neutral to the virus, here the amiR159-HCPro targets a coding region of TuMV HC-Pro cistron and consequently mutations in escape variants must result form the balance between avoiding recognition by the amiR159-HCPro and retaining biological function. Indeed, this coding effect may explain why Lin et al. observed an excess of critical positions at the 5' end of the amiR. Additionally, a potential explanation for convergence in these two central sites relies on the fact that imperfect pairing with central mismatches in small RNA-target hybrids promotes translational repression as it excludes slicing (7). This observation suggests the possibility that imperfect pairing between he amiR and mutated targets might lead to translational repression rather than viral RNA cleavage. In contrast to the catalytic effects of amiR-mediated viral RNA cleavage, translational repression requires stoichiometric amounts of amiRs and therefore is not as efficient. Inefficient translation inhibition might allow for residual viral replication and progeny virus can still escape the repression by fixing changes in the target sequence.

All in all, our results suggest that the durability of amiR-based resistance may be too short in time as to make it a profitable approach. However, this assertion has to be carefully considered in the context that we designed our experiments in such a way that they represent the most favorable possible situation for resistance-breaking. For instance, our challenge experiments were done with inocula that represent 1-20% of the whole viral population, according to our simulations. In a natural situation in the field, transmission would be mediated by vectors, which impose more dramatic bottlenecks, in the order of units per vector and transmission event (1,2,37), thus minimizing the likelihood of transmitting very low frequency escape alleles; although large vector populations will contribute to increasing the chances of transmission. Furthermore, the way we sample viral populations, homogenizing the whole plant, provided transmission probability to all genomes present in the plant. The spatial structure imposed by

plant architecture limits gene flow among distal parts of the plant, up to the point that each part may be dominated by different viral genotypes (28). This means that variants may not reach high frequency within the whole metapopulation despite having some local fitness advantage. Therefore, by feeding on particular leaves, vectors would miss loading escape mutants that may be abundant in other parts of the plant. All these factors, plus surely some additional ones, increase the stochasticity of escape alleles spilling over from their reservoirs to the amiR transgenic crops, thus perhaps increasing the resistance durability. Another factor that may affect durability, as suggested by our results, is the amount at which the amiR is expressed. We have shown that sub-inhibitory expression levels would indeed select for resistance alleles, facilitating their spread in transgenic populations. This adds a cautionary note for biotechnologists when selecting their new transgenic plants. Another way of increasing the resistance durability could be to express more than one amiR in a transgenic crop, targeting different highly conserved RNA sequences in the viral genome, or combining amiR-mediated resistance with other genetic resistances. By combining multiple amiRs into a single plant, the likelihood of resistance-breaking will drop down exponentially. Currently, we are exploring this possibility in the laboratory.

## ACKNOWLEDGEMENTS

We thank Dr. S.D. Yeh for kindly providing p35STuMV and J. Forment for technical computational assistance.

This work was supported by the Human Frontiers Science Program Organization grant RGP12/2008, the Generalitat Valenciana grant PROMETEO/2010/019, and CSIC grant 2010TW0015. We also acknowledge support from the Santa Fe Institute.

## REFERENCES

1. **Ali, A., et al**. 2006. Analysis of genetic bottlenecks during horizontal transmission of *Cucumber mosaic virus*. J. Virol. **80**:8345-8350.
2. **Betancourt, M., A. Fereres, A. Fraile, and F. García-Arenal.** 2008. Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. J. Virol. **82**:12416-12421.
3. **Bishop, K.N., R.K. Holmes, A.M. Sheehy, and M.H. Malim.** 2004. APOBEC-mediated editing of viral RNA. Science **305**:645.
4. **Boden, D., O. Pusch, F. Lee, L. Tucker, and B. Ramratnam.** 2003. Human immunodeficiency virus type 1 escape from RNA interference. J. Virol. **77**:11531-11535.
5. **Boucher, C.A., et al.** 1992. Ordered appearance of zidovudine resistance mutations during treatment of 18 human immunodeficiency virus-positive subjects. J. Infect. Dis. **165**:105-110.

6. **Boyes, D.C., et al.** 2001. Growth stage-based phenotypic analysis of *Arabidopsis*: a model for high throughput functional genomics in plants. Plant Cell **13**:1499-1510.
7. **Brodersen, P., et al**. 2008. Widespread translational inhibition by plant miRNAs and siRNAs. Science **320**:1185-1190.
8. **Bull, J.J., et al.** 1997. Exceptional convergent evolution in a virus. Genetics **147**:1497-1507.
9. **Burch, C.L., S. Gudayer, D. Samarov, and H. Shen.** 2007. Experimental estimates of the abundance and effects of nearly neutral mutations in the RNA virus ϕ6. Genetics **176**:467-476.
10. **Chen, C.C., et al.** 2003. Identification of *Turnip mosaic virus* isolates causing yellow stripe and spot on calla lily. Plant Dis. **87**:901-905.
11. **Chen, Y.K., D. Lohuis, R. Goldbach, M. Prins.** 2004. High frequency induction of RNA-mediated resistance against *Cucumber mosaic virus* using inverted repeat constructs. Mol. Breeding **14**:215-226.
12. **Coburn, G.A., and B.R. Cullen. 2002.** Potent and specific inhibition of human immunodeficiency virus type 1 replication by RNA interference. J. Virol. **76**:9225-9231.
13. **Conticello, S.G., C.J. Thomas, S.K. Petersen-Mahrt, and M.S. Neuberger.** 2005. Evolution of the AID/APOBEC family of polynucleotide (deoxy)cytidine deaminases. Mol. Biol. Evol. **22**:367-377.
14. **Couce, A, and Blázquez, J.** 2009. Side effects of antibiotics on genetic variability. FEMS Microbiol. Rev. 33:531-538.
15. **Cuevas, J.M., S.F. Elena, and A. Moya.** 2002. Molecular basis of adaptive convergence in experimental populations of RNA viruses. Genetics **162**:533-542.
16. **Cullen, B.R.** 2006. Role and mechanism of action of the APOBEC3 family of antiretroviral resistance factors. J. Virol. **80**:1067-1076.
17. **Das, A.T., et al.** 2004. Human immunodeficiency virus type 1 escapes from RNA interference-mediated inhibition. J. Virol. **78**:2601-2605.
18. **Di Nicola-Negri, E., A. Brunetti, M. Tavazza, and V. Ilardi.** 2005. Hairpin RNA-mediated silencing of *Plum pox virus* P1 and HC-Pro genes for efficient and predictable resistance to the virus. Transgenic Res. **14**:989-994.
19. **Elbashir, S.M., J. Martínez, A. Patkaniowska, W. Lendeckel, and T. Tuschl.** 2001. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. EMBO J. **20**:6877-6888.
20. **Elena, S.F., R.V. Solé, and J. Sardanyés.** 2010. Simple genomes, complex interactions: epistasis in RNA virus. Chaos **20**:e26106.
21. **García-Arenal, F., and B.A. McDonald.** 2003. An analysis of the durability of resistance to plant viruses. Phytopathology **93**:941-952.

22. **Ge, Q., et al.** 2003. RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. Proc. Natl. Acad. Sci. USA **100**:2718-2723.
23. **Gitlin, L., J.K. Stone, and R. Andino.** 2005. Poliovirus escape from RNA interference: short interfering RNA-target recognition and implications for therapeutic approaches. J. Virol. **79**:1027-1035.
24. **Guo, H.S., Q. Xie, J.F. Fei, and N.H. Chua.** 2005 MicroRNA directs mRNA cleavage of the transcription factor NAC1 to downregulate auxin signals for *Arabidopsis* lateral root development. Plant Cell **17**:1376-1386.
25. **Hamilton, A.J., and D.C. Baulcombe**. 1999. A species of small antisense RNA in posttranscriptional gene silencing in plants. Science **286**:950-952.
26. **Haydon, D., N. Knowles, and J. McCauley.** 1998. Methods for the detection of non-random base substitution in virus genes: models of synonymous nucleotide substitution in picornavirus genes. Virus Genes **16**:253-266.
27. **Hedrick, P.W.** 2004. Genetics of Populations. Sudbury: Jones and Bartlett Publishers. pp. 376-379.
28. **Jridi, C., J.F. Martin, V. Mareie-Jeanne, G. Labonne, and S. Blanc.** 2006. Distinct viral populations differentiate and evolve independently in a single perennial host plant. J. Virol. **80**:2349-2357.
29. **Kalantidis, K., S. Psaradakis, M. Tabler, and M. Tsagris.** 2002. The occurrence of CMV-specific short RNAs in transgenic tobacco expressing virus-derived double-stranded RNA is indicative of resistance to the virus. Mol. Plant-Microb. Interact. **15**:826-833
30. **Krönke, J., et al.** 2004. Alternative approaches for efficient inhibition of hepatitis C virus RNA replication by small interfering RNAs. J. Virol. **78**:3436-3446.
31. **Lin, S.S., et al.** 2009. Molecular evolution of a viral non-coding sequence under the selective pressure of amiRNA-mediated silencing. PLoS Pathog. **5**:e1000312.
32. **Lindbo, J.A., and W.G. Dougherty.** 2005. Plant pathology and RNAi: a brief history. Annu. Rev. Phytopathol. **43**:191-204.
33. **Malpica, J.M., A. Fraile, I. Moreno, C.I. Obies, J.W. Drake, and F. García-Arenal.** 2002. The rate and character of spontaneous mutation in an RNA virus. Genetics **162**:1505-1511.
34. **Marín, J., and R.V. Solé**. 1999. Macroevolutionary algorithms: a new optimization method on fitness landscapes. IEEE Trans. Evol. Comput. **3**:272-286.
35. **Martínez-Picado, J., et al.** 2000. Antiretroviral resistance during successful therapy of HIV type 1 infection. Proc. Natl. Acad. Sci. USA **97**:10948-10953.
36. **Missiou, A., et al.** 2004. Generation of transgenic potato plants highly resistant to *Potato virus Y* (PVY) through RNA silencing. Mol. Breeding **14**:185-197.

37. **Moury, B., F. Fabre, and R. Senoussi.** 2007. Estimation of the number of viral particles transmitted by an insect vector. Proc. Natl. Acad. Sci. USA **45**:17891-17896.
38. **Niu, Q.W., et al.** 2006. Expression of artificial microRNAs in transgenic *Arabidopsis thaliana* confers virus resistance. Nat. Biotechnol. **24**:1420-1428.
39. **Qu J, J. Ye, and R. Fang.** 2007. Artificial microRNA-mediated virus resistance in plants. J. Virol. **81**:6690-6699.
40. **Sabariegos, R., M. Giménez-Barcons, N. Tàpia, B. Clotet, and M.A. Martínez.** 2006. Sequence homology required by human immunodeficiency virus type 1 to escape from short interfering RNAs. J. Virol. **80**:571-577.
41. **Sanjuán, R., P. Agudelo-Romero, and S.F. Elena.** 2009. Upper-limit mutation rate estimation for a plant RNA virus. Biol. Lett. **5**:394-396.
42. **Sardanyés, J., R.V. Solé, and S.F. Elena**. 2009. Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. J. Virol. **83**, 12579-12589.
43. **Tromas, N., and S.F. Elena**. 2010. The rate and spectrum of spontaneous mutations in a plant RNA virus. Genetics **185**:983-989.
44. **Turturro, C., et al.** 2008. Evaluation of potential risks associated with recombination in transgenic plants expressing viral sequences. J. Gen. Virol. **89**:327-335.
45. **Varkonyi-Gasic, E., R. Wu, M. Wood, E.F. Walton, and R.P. Hellens.** 2007. Protocol: a highly sensitive RT-PCR method for detection and quantification of microRNAs. Plant Meth. **3**:12.
46. **Vaucheret, H., F. Vázquez, P. Crete, and D.P. Bartel.** 2004. The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. Genes Dev. **18**:1187-1197.
47. **Vicenzetti, S., et al.** 1999. Cloning, expression, and purification of cytidine deaminase from *Arabidopsis thaliana*. Protein Expr. Purif. **15**:8-15.
48. **Von Eije, K.J., O. ter Brake, and B. Berkhout.** 2008. Human immunodeficiency virus type 1 escape is restricted when conserved genome sequences are targeted by RNA interference. J. Virol. **82**:2895-2903.
49. **Wang, M.B., D.C. Abbot, and P.M. Waterhouse.** 2000. A single copy of a virus-derived transgene encoding hairpin RNA gives immunity to *Barley yellow dwarf virus*. Mol. Plant. Pathol. **1**:347-356.
50. **Warthmann, N., H. Chen, S. Ossowski, D. Weigel, and P. Hervé**. 2008. Highly specific gene silencing by artificial microRNAs in rice. PLoS ONE **3**:e1829
51. **Westerhout, E.M., and B. Berkhout.** 2007. A systematic analysis of the effect of target RNA structure on RNA interference. Nucl. Acids Res. **35**:4322-4330.

52. **Westerhout, E.M., M. Ooms, M. Vink, A.T. Das, and B. Berkhout.** 2005. HIV-1 can escape from RNA interference by evolving an alternative structure in its RNA genome. Nucl. Acids Res. **33**:796-804.
53. **Wichman, H.A., M.R. Badgett, L.A. Scott, C.M. Boulianne, and J.J. Bull.** 1999. Different trajectories of parallel evolution during viral adaptation. Science **285**:411-424.

## Figure Legends

FIG. 1. (A) Schematic representation of the experimental design. For illustrative purposes we only show one of the WT *A. thaliana* evolved lineages. The same protocol was repeated for the 10-4 lineages with the exception that serial passages were performed on *A. thaliana* 10-4 transgenic plants. In the illustrated example, resistance breaking occurred at passage 2 (note symptoms in the corresponding 12-4 plants). (B) Schematic diagram of the *in silico* simulation model. (a) Each lineage was considered as a population of bit-strings containing the 21-bit of the target region plus 10 loci each corresponding to the different cistrons in TuMV genome. The model simulated within-host viral replication with mutation and bottleneck transmission between passages. For the simulations of virus evolution in WT *A. thaliana* plants we did not consider target-specific degradation of strings, while for simulating the evolution in 10-4 plants we included a degradation probability $\varepsilon$ for strings with a WT target sequence. (b) Digital genome of TuMV, where the target sequence has been explicitly considered.

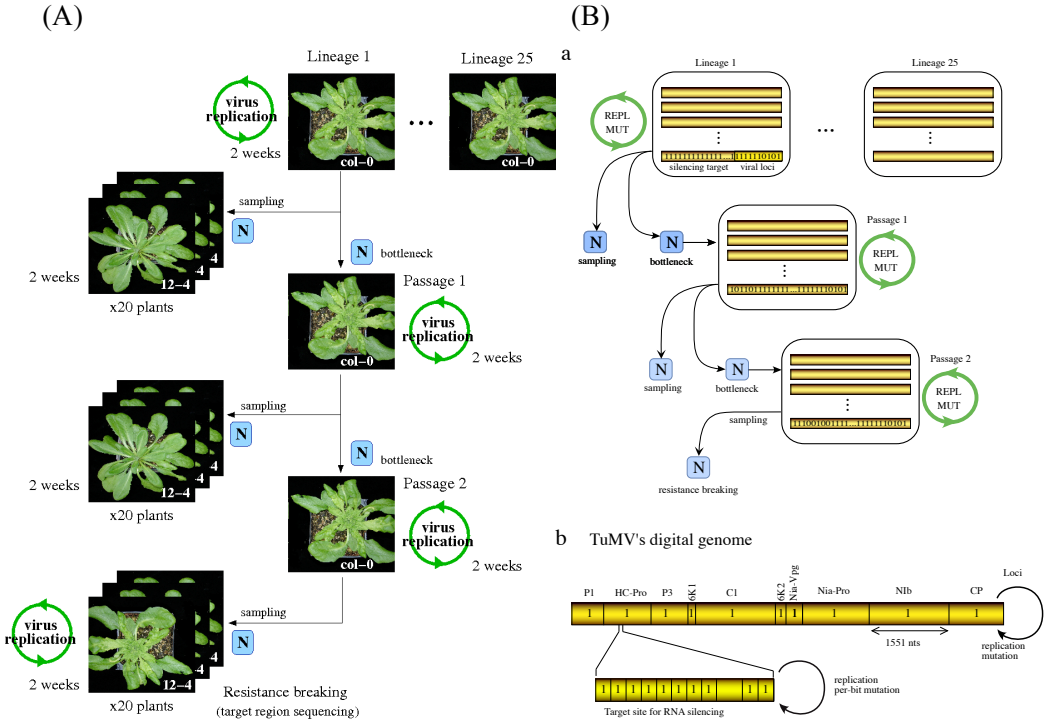

FIG. 2. Pattern of amiR159-HCPro accumulation in the partially resistant transgenic lineage 10-4. A. Curve of amiR159-HCPro accumulation in the leaf inoculated in the pathogenicity tests (units of pg of amiR159-HCPro per ng of total plant RNA). The dashed line represents the fit to a 2-parameters exponential growth model ($R^2$ = 0.990, $F_{1,4}$ = 405.167, $P$ < 0.001). B. Pattern of amiR159-HCPro accumulation in six leaves that differ in their developmental stage from four different plants (units of amiR159-HCPro molecules per ng of total plant RNA). Each plant is represented by a different color. In all cases, error bars represent ±1 SEM.

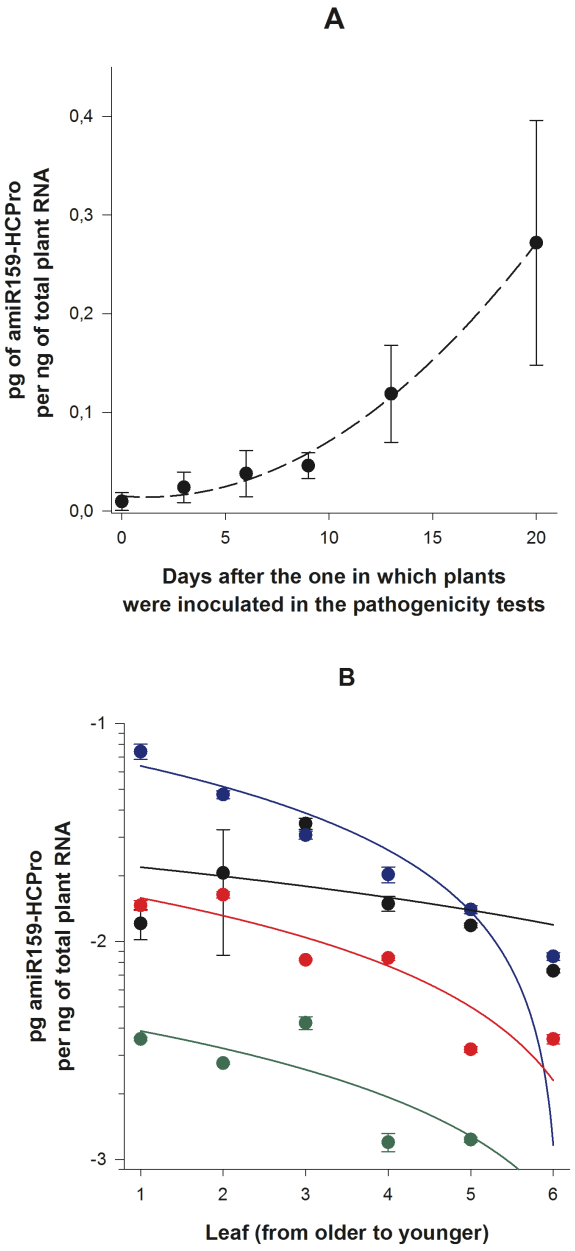

FIG. 3. Cumulative frequency of lineages capable of escaping from the amiR159-HCPro resistance. The black line corresponds to the lineages evolved in WT *A. thaliana* plants. The red line corresponds to the lineages evolved in partially resistant *A. thaliana* 10-4 plants. The ability of TuMV evolving populations to escape from the amiR159-HCPro was evaluated in 12-4 plants that were fully resistant to the ancestral TuMV genotype. A population was considered as able of escaping from the resistance when at least one 12-4 plant was infected.

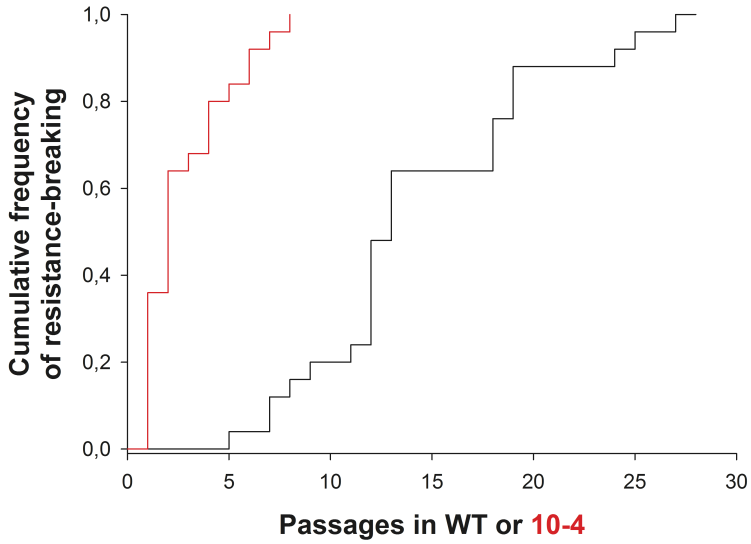

FIG. 4. Distribution of mutations in escape viruses along the amiR159-HCPro target sequence. Black bars correspond to the frequency of mutations that arose in WT *A. thaliana* plants; white bars to those observed in the 10-4 transgenic line.

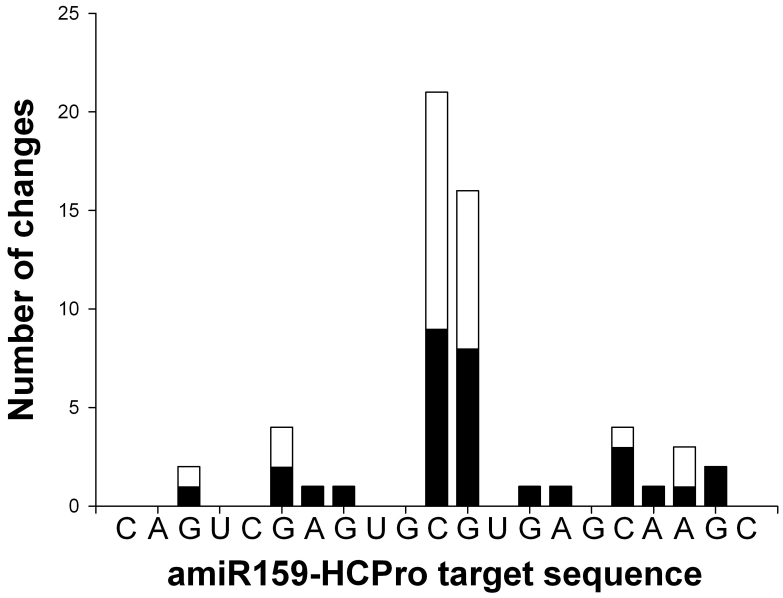

FIG. 5. Results of the simulation studies for the set of parameters that showed the best fit to data shown in Figure 3. A. Simulation results for the WT *A. thaliana*-evolved TuMV lineages. B. Simulation results for the TuMV lineages evolved in partially resistant 10-4 plants. The red dots correspond to the best-fitting trajectory obtained from the most optimized parameter set. Red bars denote the SD among $10^3$ runs of the simulation model using the best-fitting parameters.

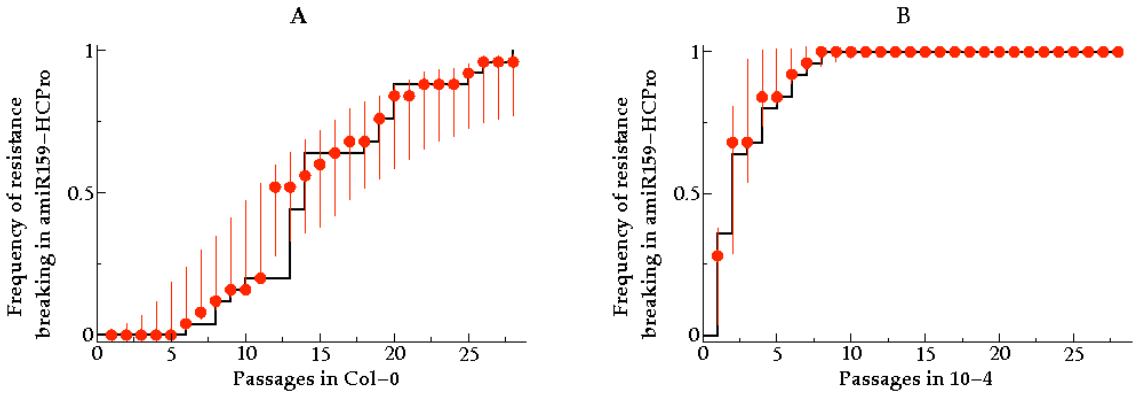

TABLE 1. Escape alleles found in the TuMV populations evolved in fully susceptible WT *A. thaliana* plants.

| **Allele**[a] | **Lineage (passage)**[b] | **Type of mutation**[c] |
|---|---|---|
| ACA GUC GAG UGC GUG AGC AAG UUA | Ancestral WT | |
| 1 ACA GUC GAG UGU GUG AGC AAG UUA | 12(6), 18(9), 9(13), 7(14), 19(14), 21(14), 20(19) | Synonymous |
| 2 ACA GUC GAG UGC AUG AGC AAG UUA | 10(10), 3(14), 8(14), 22(14), 24(14), 4(19), 16(20) | V→M |
| 3 ACA GUC GAG UGU GUG AGC AAN UUA | 6(14), 11(27), 23(27) | synonymous/synonymous or K→N |
| 4 ACA GUC GAG UGC GUG AGU AAG UUA | 14(13), 1(27) | Synonymous |
| 5 ACA GUC AAG UGC GUA AGC AAG UUA | 15(8) | E→K/synonymous |
| 6 ACA GUC GAG UGC GUG GGU AAG UUA | 25(8) | S→G |
| 7 ACA GUC GUA UGC GUG AGC AAG UUA | 13(14) | E→V |
| 8 ACA GUC GAG UGC GUG AGC AGG UUA | 5(18) | K→R |
| 9 ACA AUC AAG UGC GUG AGC AAG UUA | 2(20) | V→I/E→K |
| 10 ACA GUC GAG UGC GUG AGC GAG UUA | 17(20) | K→E |

[a] Underlined are the 21-nt of the target. The mutated sites are marked in red.

[b] The lineage (and passage) in which each allele was observed are indicated.

[c] The last column indicates whether the mutations was synonymous or involved an amino acid replacement.

TABLE 2. Escape alleles found in TuMV populations evolved in partially resistant 10-4 plants.

| Allele | Lineage (passage) | Type of mutation |
|---|---|---|
| ACA GUC GAG UGC GUG AGC AAG UUA | Ancestral WT | |
| 1 ACA GUC GAG UGU GUG AGC AAG UUA | 3(1), 6(1), 8(1), 13(1), 15(1), 14(2), 21(2), 25(2), 7(4), 11(4), 22(5) | synonymous |
| 2 ACA GUC GAG UGC AUG AGC AAG UUA | 2(1), 9(1), 17(2), 10(3), 4(4), 18(6), 24(7), 20(8) | V→M |
| 11 ACA GUC GAG UGC GUG AGC ACG UUA | 1(2), 19(6) | K→T |
| 12 ACA GUC AAG UGU GUG AGC AAG UUA | 5(1) | E→K/synonymous |
| 13 ACA AUC GAG UGC GUG AGC AAG UUA | 16(1) | V→I |
| 14 ACA GUC AAG UGC GUG AGC AAG UUA | 12(2) | E→K |
| 4 ACA GUC GAG UGC GUG AGU AAG UUA | 23(2) | synonymous |