Shrinkage of Genome Size in a Plant RNA Virus upon Transfer of an Essential Viral Gene into the Host Genome

Nicolas Tromas1, Mark P. Zwart1, Javier Forment1, and Santiago F. Elena1,2,*

1Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, Valencia, Spain
2The Santa Fe Institute, Santa Fe, New Mexico
*Corresponding author: E-mail: santiago.elena@csic.es.
Accepted: February 17, 2014

Abstract
Nonretroviral integrated RNA viruses (NIRVs) are genes of nonretroviral RNA viruses found in the genomes of many eukaryotic organisms. NIRVs are thought to sometimes confer virus resistance, meaning that they could impact spread of the virus in the host population. However, a NIRV that is expressed may also impact the evolution of virus populations within host organisms. Here, we experimentally addressed the evolution of a virus in a host expressing a NIRV using Tobacco etch virus (TEV), a plant RNA virus, and transgenic tobacco plants expressing its replicase, Nib. We found that a virus missing the Nib gene, TEV-C1 Nib, which is incapable of autonomous replication in wild-type plants, had a higher fitness than the full-length TEV in the transgenic plants. Moreover, when the full-length TEV was evolved by serial passages in transgenic plants, we observed genomic deletions within Nib—and in some cases the adjacent cistrons—starting from the first passage. When we passed TEV and TEV-C1 Nib in transgenic plants, we found mutations in proteolytic sites, but these only occurred in TEV-C1 Nib lineages, suggesting the adaptation of polyprotein processing to altered Nib expression. These results raise the possibility that NIRV expression can indeed induce the deletion of the corresponding genes in the viral genome, resulting in the formation of viruses that are replication defective in hosts that do not express the same NIRV. Moreover, virus genome evolution was contingent upon the deletion of the viral replicase, suggesting NIRV expression could also alter patterns of virus evolution.

Key words: experimental evolution, genome evolution, nonretroviral integrated RNA viruses, plant virus, RNA virus, virus evolution.

Introduction
The genome size of RNA viruses is highly variable, ranging from less than 4 kb for members of the Narnaviridae to more than 30 kb for some Coronaviridae. However, most RNA viruses have genome sizes in the 5–15-kb range (Holmes 2009). What could explain this wide range of genome sizes in RNA viruses? Small genomes probably replicate faster than large ones, while genome size limits enzymatic complexity. This limited complexity results in lower-fidelity replication enzymes, which in turn leads to higher variability (Smith and Szathmáry 1995). On the other hand, larger genomes will replicate more slowly, while being less constrained in the complexity of the proteins they code for. Therefore, larger genomes allow for evolutionary innovations such as polymerases with greater replication fidelity (Denison et al. 2011) or the modulation of host immune responses (Shackelton and Holmes 2004). However, the evolution of larger genomes and greater complexity may be impeded by Eigen’s paradox: to obtain proofreading activity, an RNA virus must already have a large genome (Eigen 1971; Smith and Szathmáry 1995). It has therefore been suggested that RNA virus genome size evolves to the upper limit imposed by mutation rate (Swetina and Schuster 1982).

If genome size of RNA viruses is limited by a tradeoff between RNA-dependent RNA polymerases (RdRp) fidelity and decreased replication rates, then examples of enlargement in genome size should be scarce for RNA viruses. There are, however, some examples of the integration of host genetic material into viral RNA genomes, leading to the largest known RNA viruses. For instance, the genome of coronaviruses contains homologous sequences of at least five cellular enzymes that are associated with RNA processing and that may be involved in modulating mutation rate (Snijder et al. 2003). Acquisition of a host’s ubiquitin gene by certain strains of bovine viral diarrhea virus resulted in increased pathogenicity (Meyers et al. 1989). Plant closterovirus genomes...
encode for a homologue of cellular molecular chaperone HSP70 (Dolja et al. 2006) that is required for virion assembly (Satyanarayana et al. 2000). Finally, a particularly interesting case is the ubiquitous presence of DNA/RNA repair AlkB-like proteins in several groups of plant viruses, including closterovirus, flexiviruses, tymoviruses, and even one potyvirus (Blackberry virus Y; Susaimuthu et al. 2008), whose function has been postulated to be both repair of nucleic acid alkylation and suppression of posttranscriptional gene silencing (Aravind and Koonin 2001; Bratlie and Drabløs 2005). Similarly, the insertion of sequences into a viral genome generally leads to their rapid removal, as has been seen in the case of heterologous genes (Chapman et al. 1992; Guo et al. 1998; Chung et al. 2007; Paar et al. 2007; Zwart et al. 2014) and repetitions of endogenous noncoding sequences (Nagai et al. 2003; Gritsun and Gould 2006). Finally, most viruses have a tendency to evolve defective interfering viruses under conditions promoting cellular coinfection (Huang 1973; Roux et al. 1991), again suggesting a link between genome size and replication. Although demographic conditions may allow for the maintenance of extraneous sequences (Dolja et al. 1993; Zwart et al. 2014), overall it appears as if genome size is limited and under purifying selection, and that the incorporation or maintenance of foreign sequences comes at a cost to viruses (Sakai et al. 1999; Marks et al. 2005; Holmes 2009; Zwart et al. 2014).

On the other hand, genome size does not appear to be under the same selective pressure in eukaryotes, probably due to the smaller effective population sizes in higher organisms (Lynch et al. 2006). Although the existence of many retroviral sequences in eukaryotic genomes was already well documented (Gifford and Tristem 2003), it has been recently found that there are also many nonretroviral RNA virus genes in eukaryotic genomes. These nonretroviral integrated RNA viruses (NIRVs) have been found in mammoths (Geuking et al. 2009; Belyi et al. 2010; Horie et al. 2010; Katzourakis and Gifford 2010), plants (Hong et al. 1998; Chiba et al. 2011), and fungi (Taylor and Bruenn 2009). In fungi, NIRVs are very common and have in some cases become functional genes (Koonin 2010). In one case, the entire viral RdRp was integrated and conserved, probably conferring viral resistance to the host (Taylor and Bruenn 2009). It has also been suggested that NIRVs can confer resistance (Bertsch et al. 2009) or tolerance (Flegel 2009) against the original virus to plant or animal hosts. Indeed, it has been a common practice to generate commercial transgenic plants resistant to viruses by inserting full or partial viral genes into the genome of the plants of agronomic interest (reviewed in Kundu and Mandal [2001]); these transgenic plants can be considered as artificial cases of NIRVs.

Because the reading frame is conserved in many NIRVs (Taylor and Bruenn 2009) they can in principle be expressed. In practice, this will depend on their location in the genome and the state of the host cell. Some NIRVs are, however, clearly expressed by the host (Taylor and Bruenn 2009). Host NIRV expression has been proposed to mediate host immunity against the cognate virus (Bertsch et al. 2009; Flegel 2009; Taylor and Bruenn 2009). However, host NIRV expression may also have important implications for the evolution of the cognate virus. The absence of an effective NIRV-induced host immune response and sufficiently high NIRV expression levels could lead to complementation of host proteins, making the NIRV-expressed genes in the cognate virus functionally redundant. What would be the impact of NIRV expression on viral replication? Is there any selective advantage to using in trans the protein produced from the NIRV as a functional viral protein? How will a virus population evolve in such a situation of functional redundancy? Finally, how do such large-scale genomic perturbations affect the regulation of expression and accumulation of all other viral components? To our knowledge these questions have not been addressed in any study, while given the prevalence of NIRVs in natural host populations and biosafety issues linked to the use of transgenic plants expressing viral genes for resistance, it is a highly relevant topic.

Here, we evaluated the patterns of genomic evolution of Tobacco etch virus (TEV, genus Potyvirus, family Potyviridae) in transgenic Nicotiana tabacum plants expressing Nlβ. The TEV RdRp (Li and Carrington 1995). This system is an experimental model for exploring what the effects of high levels of NIRV expression are on virus infection and evolution. During infection of these transgenic plants, plant-derived Nlβ expression makes virus-derived Nlβ-expression redundant (Li and Carrington 1995). We therefore expect that there may be relaxed selection for maintenance of the Nlβ cistron in the viral genome, which might lead to selection of variants with genomic deletions at the Nlβ locus, due to the replicative advantage associated with a smaller genome size. Ultimately, the expression of Nlβ might therefore induce genome shrinkage and result in a virus population that can no longer autonomously infect wild-type plants. We first determined whether the deletion of viral Nlβ led to an increase in viral fitness in Nlβ-expressing plants. Then, we passaged a full-length wild-type virus in Nlβ-expressing plants and used next-generation sequencing (NGS) to check for deletions in TEV Nlβ loci. Finally, we preempted the complete deletion of TEV Nlβ in Nlβ-expressing plants and studied the effects of Nlβ complementation in trans on genomic changes.

Materials and Methods

Virus Genotypes

The pTEV7DA infectious clone (Dolja et al. 1992) was used as a source for TEV. A TEV genotype was produced that lacked the full replicase gene (ΔNlβ) by inverse polymerase chain reaction (PCR) using Pfu turbo DNA polymerase (Stratagene) and primers conserving the proteolytic Nla/Nlβ.
Concentrated saps of TEV and TEV-ΔNIb were obtained by grinding 500 mg of 7 days postinoculation (dpi) infected tissue in a mortar with 500 μl of grinding buffer. Nicotiana tabacum 35S::NIb plants were inoculated with a 1:1 mixture of homogenates of both viruses. Twenty plants were inoculated by abrasion of the third true leaf with 30 μl mixed homogenate, and 20 plants were taken as a mock-inoculated control. After 7, 21, and 63 dpi, total RNA was extracted from the youngest leaves of five plants. The region flanking the NIb cistron was reverse transcribed using Moloney Murine leukemia virus (M-MLV) reverse transcriptase (RT) (Fermentas) and primer 97-101-R (supplementary table S1, Supplementary Material online). PCR amplification was performed using the high-fidelity Phusion DNA polymerase (Finnzymes) and the 73-80-F/92-96-R primers (supplementary table S1, Supplementary Material online). PCR products were resolved on a 1% agarose gel to verify the presence of both genotypes. RT-qPCR was performed to compare viral accumulation in case of direct competition between the two genotypes. RT-qPCR for the CP cistron was used to determine the total level of viral accumulation (i.e., accumulation of both genotypes), using the specific primers TEV-CP-qPCR-F and TEV-CP-qPCR-R (supplementary table S1, Supplementary Material online). RT-qPCRs were performed using One-Step SYBR PrimeScript RT-PCR kit II (Takara) following the instructions provided by the manufacturer. CP was chosen because it locates in the 3' end of TEV genome and hence would only quantify complete genomes but not partial incomplete amplicons. Each RNA sample was quantified three times in independent experiments. Amplifications were done using the ABI PRISM Sequence Analyzer 7000 (Applied Biosystems). The thermal profile was as follows: RT phase consisted of 5 min at 42 °C followed by 10 s at 95 °C, and PCR phase of 40 cycles of 5 s at 95 °C and 31 s at 60 °C. Quantification results were examined using SDS7000 software v. 1.2.3 (Applied Biosystems). RT-qPCR for NlaProNIb cistrons was then performed to determine viral accumulation of the TEV genotype only, using specific primers TEV-NlaProNIb-F and TEV-NlaProNIb-R (supplementary table S1, Supplementary Material online) and otherwise identical conditions when compared with the quantification of the CP. The ratio of TEV-ΔNIb to TEV (R) is then 

\[ R = \frac{n_{CP} - n_{NlaProNIb}}{n_{NlaProNIb}} \]

where \( n_{CP} \) and \( n_{NlaProNIb} \) are the copy numbers of the CP and NlaProNIb, respectively, as measured by RT-qPCR. To estimate the replicative advantage \( W \) (Carrasco et al. 2007) of TEV-ΔNIb, we performed linear regression on the infection time in days and the log-transformed ratio data. The antilog-transformed slope is \( W \).

Experimental Evolution Protocol

Infectious plasmids pTEV7DA and pTEV7DA-ΔNIb were linearized with BglII (Takara) and transcribed into 5'-capped RNAs using the SP6 mMESSAGE mMACHINE kit (Ambion Inc). Transcripts were prepared as described previously (Carrasco et al. 2007). Two batches of five 4-week-old N. tabacum 35S::NIb plants were inoculated mechanically on the third true leaf either with TEV transcripts (4–7 μg) in inoculation buffer (3% polyethyleneglycol 8000, pH = 7.0) or TEV-ΔNIb transcripts (20–25 μg) in the same inoculation buffer. In all cases, TEV symptoms appeared 6–7 dpi.

Two evolution experiments were performed with these viruses. In the first experiment, conceived to study the appearance of genomic deletions de novo, four 3-week passages of TEV and TEV-ΔNIb were performed in N. tabacum 35S::NIb plants. At 21 dpi, all symptomatic leaves of each of the five lineages per virus were collected. Five hundred milligrams of symptomatic tissue was ground in 500 μl of grinding buffer (50 mM potassium phosphate pH = 7.0, 3% polyethyleneglycol 6000). Ten microliters of the resulting sap was then mixed with inoculation buffer and used to infect a batch of five healthy transgenic plants by abrasion for each TEV genotype. In the second experiment, conceived to evaluate the subsequent evolution and fitness optimization of genotypes missing the NIb cistron, 30 one-week passages of TEV and TEV-ΔNIb were performed in N. tabacum 35S::NIb plants. At 7 dpi, symptomatic leaves were collected for each of the five lineages per virus. Infectious sap was prepared as described earlier.

TEV Genomic RNA Extraction and Purification for Sequence Analysis

To analyze the lineages of TEV-ΔNIb and TEV evolved in N. tabacum 35S::NIb (second experiment above), total RNA was extracted using RNAeasy Plant Mini Kit (Qiagen) from symptomatic leaves of five N. tabacum 35S::NIb infected by the evolved lineages of TEV and five N. tabacum 35S::NIb infected by the evolved lineages of TEV-ΔNIb. To verify the presence of deletions, the full genome of each genotype was reverse transcribed using M-MLV RT (Fermentas) and primer 97-101-R (supplementary table S1, Supplementary Material online), then PCR amplified using the Ultra High-Fidelity DNA polymerase Phusion (Finnzymes) and primers 73-80-F and 97-101-R (supplementary table S1, Supplementary Material online). Deletions were confirmed by resolution of PCR products on 1% agarose gels.

Subsequently, the full genome was reverse transcribed using M-MLV RT (Fermentas) and primers 97-101-R, 67-77-R, and 40-45-R and PCR amplified using the high-fidelity DNA polymerase Phusion (Finnzymes) and three pairs of primers 2-10-F/40-45-R; 45-48-F/67-77-R; and 73-80-F/97-101-R (supplementary table S1, Supplementary Material online). By using these pairs, we ensured that the mRNA from the
transgene was not amplified and that we obtained only TEV sequences from viral genomes. PCR products were resolved on 1% agarose gels to verify the presence of deletions. For both genotypes, TEV and TEV-ΔNlb, the three amplicons were purified and sequenced by GenoScreen (Lille, France) using BIGDYE 3.1 and a 96-capillars AB3730XL sequencing system (Applied Biosystems) with overlapping readouts using the same set of primers as in Agudelo-Romero et al. (2008). Contigs were assembled using GENEIOUS version 4.8.

High-Throughput Sequencing and Data Analysis
To analyze the diversity of deletions that could occur in Nlb cistron during the evolution of TEV in N. tabacum 35S::Nlb (first experiment described above), the full genome of each genotype was reverse transcribed using M-MLV RT (Fermentas) and primer R97-101, then PCR amplified using the Ultra High-Fidelity DNA polymerase Phusion (Finnzymes) and primers F73-80 and R97-101 (supplementary table S1, Supplementary Material online). Seven independent RT-PCR replicates were performed for each reaction and then pooled, to limit possible PCR artifacts. These PCR amplicons were sequenced by GenoScreen (Lille, France) with an Illumina HiSeq2000 equipment (Illumina Inc.) using a paired-end (2 x 100 b) protocol. The reads were already demultiplexed and adapters removed. Data were then analyzed with GSNAP (Wu and Nacu 2010). GSNAP aligned both single-end and paired-end reads. GSNAP is designed to detect short and long-distance splicing, but in our case, we used it to detect deletions instead of introns. GSNAP was used with its default set of parameters, and results were parsed using custom Perl scripts: 1) we excluded reads that are mapped on its default set of parameters, and results were parsed using custom Perl scripts: 1) we excluded reads that are mapped on 

Analysis of TEV-ΔNlb Mutants
Each mutation was introduced in the pMTEV infectious clone (Bedoya and Daro’s 2010) by site-directed mutagenesis using the Quickchange II XL kit (Stratagene) and following the indications given by the manufacturer. To minimize unwanted errors during the mutagenesis process, the kit incorporates Pfu Ultra High-Fidelity Turbo DNA polymerase (Stratagene). The amplification conditions were 1 min at 95°C (initial denaturation), followed by 18 cycles consisting of 30 s at 95°C, 45 s at 65°C and 18 min at 68°C, and a final extension step of 28 min at 68°C.

Infectious plasmids of each mutant were linearized with BgIII (Takara) and transcribed into 5’-capped RNAs. Transcripts were prepared as described previously. Two batches of five 4-week-old N. tabacum 35S::Nlb plants were, respectively, inoculated mechanically on the third true leaf with TEV transcripts (20–30 μg) in inoculation buffer. In all cases, the first symptoms appeared 6–7 dpi. Seven dpi, infectious saps were obtained by grinding 500 mg of TEV mutant-infected tissue in a mortar with 500 μl of grinding buffer; then, three 4-week-old N. tabacum 35S::Nlb plants for every TEV mutants were, respectively, inoculated mechanically on the third true leaf with 15 μl of the corresponding sap.

Total RNA was extracted from the youngest leaves of each of the three plants per TEV mutant 2.67 dpi and then RNAs were purified. RT-qPCR with the CP primers was performed to compare viral accumulation of each mutant to the TEV-ΔNlb ancestor as described previously. For each genotype, the Malthusian growth rate per day was computed as m = (1/t)log Qt, where Qt is amount of TEV RNA (in pg) per 100 ng of total plant RNA quantified at t = 2.67 dpi. Absolute fitness was then defined as W = e^m (Crow and Kimura 1970).

Results and Discussion
TEV-ΔNlb Has Greater Within-Host Fitness than TEV in N. tabacum 35S::Nlb
Li and Carrington (1995) generated transgenic N. tabacum cv. Xanthi plants expressing TEV Nlb under control of the Cauliflower mosaic virus (CaMV) 35S promoter, referred to here as N. tabacum 35S::Nlb. TEV variants with disruption of the polymerase GDD amino acid motif, the nuclear localization signal, or the Nlb N-terminal cleavage site could replicate in these transgenic plants, whereas replication was not supported in wild-type plants. Moreover, a TEV variant missing Nlb altogether could also infect these transgenic plants and induce symptoms after 7 dpi, similar to the wild-type virus. Tromas and Elena (2010) also generated a TEV variant missing the Nlb cistron, referred to here as TEV-ΔNlb. TEV-ΔNlb can infect N. tabacum 35S::Nlb, but not the wild-type N. tabacum, in agreement with previous results (Li and Carrington 1995). TEV is the full-length wild-type virus, capable of infecting both wild-type N. tabacum and 35S::Nlb transgenic plants. When wild-type N. tabacum plants were coinoculated with different mixtures of TEV:TEV-ΔNlb (1:100 to 100:1), we detected only infection by TEV. Coinfected wild-type plants were probably not observed because the low cellular multiplicity of infection of TEV in N. tabacum (Tromas et al. 2014) precludes cellular coinfection, especially early in infection.

TEV-ΔNlb has a 16% smaller genome than TEV. If 1) a relationship between genome size and virus replication exists (Sakai et al. 1999; Zwart et al. 2014) and 2) dependence
on in trans NlB expression for replication does not impose a large fitness burden to the virus, then NlB deletion in the virus genome will cause a faster replication and a selective advantage over a full-length wild-type virus. We therefore inoculated 15 N. tabacum 35S::NlB plants with a 1:1 mixture of ground infectious tissue with TEV and TEV-ΔNlB and measured the ratio of the two variants at 7, 21, and 63 dpi using quantitative reverse transcription PCR (RT-qPCR). Five plants were analyzed for each time point, meaning that each data point represents an independent measurement. On the basis of these data, we estimated the selective advantage for TEV as the per day selection rate constant W (Materials and Methods). A significant, upward-sloping linear regression was observed between time and the ratio of the TEV-ΔNlB:TEV (fig. 1; \( r^2 = 0.632, F_{1.14} = 22.330, p < 0.001 \)), meaning that the ratio of the viruses shifts significantly toward TEV-ΔNlB, and this virus therefore has a significantly higher competitive fitness. We estimated that \( W = 1.028 \pm 0.006 \) d⁻¹ (±1 SD), suggesting that TEV-ΔNlB will slowly displace the wild-type virus in transgenic plants. If genome size was inversely proportional to fitness and the 16% reduction in genome size were the only determinant of fitness for TEV-ΔNlB, we would expect \( W = 1/(1-0.16) = 1.19 \) d⁻¹. The 95% confidence interval for our estimate of W is 1.016–1.040 d⁻¹, suggesting that 1) there is a cost to depend on in trans NlB expression, but 2) this cost is partially compensated for by faster replication of TEV-ΔNlB, resulting in a higher W than the wild-type virus. Our data therefore support the idea of a relationship between genome size and fitness, and suggest that viruses harboring deletions in NlB will outcompete the wild-type virus in NlB-expressing plants. Therefore, NIRV expression might also lead to conditions favoring the replication of virus variants, with deletions in the gene(s) encoded for by the NIRV. However, the accumulation of viruses with large genomic deletions due to in trans expression of virus’ genes will also depend on mutational supply, specifically how often recombination generates viruses with deletions in NlB that conserve the reading frame in downstream genes.

Full-Length Genomes Were Not Recovered by Recombination between TEV-ΔNlB and the Transcripts from the 35S::NlB Transgene

Borja et al. (1999) reported that inoculation of transgenic N. benthamiana plants expressing the CP protein of tomato bushy stunt tomsusvirus (TBSV) with TBSV mutants carrying defective CP resulted in the restoration of wild-type virus throughout a double-recombination event between the viral genome and the mRNA transcribed from the transgene. In sharp contrast, several authors have failed to detect recombination events between transgenes and potyviruses (Chung et al. 2007; Dietrich et al. 2007). To test for possible recombination events in our system, we performed an evolution experiment consisting of five independent lineages of TEV-ΔNlB serially passaged four times in N. tabacum 35S::NlB transgenic plants, with each passage being 3 weeks long. At the end of the evolution experiment, we analyzed the evolved lineages by RT-PCR using a combination of primers that will specifically amplify full-genome viruses (supplementary table S1, Supplementary Material online). We found no evidence for recombination between TEV-ΔNlB and host-derived NlB transcripts. Therefore, we conclude that, in our experimental set up, regeneration of the wild-type genome by recombination is an unlikely phenomenon. This outcome is congruent with the estimates of within-host fitness for TEV and TEV-ΔNlB; recombinants analogous to the wild-type virus would be displaced by TEV-ΔNlB, and the process could occur rapidly as de novo variation will initially be present at low frequencies in the virus population. On the other hand, this result may reflect the possibility that recombination between TEV-ΔNlB and the transgene transcripts is very rare, as shown for other potyviruses (Chung et al. 2007; Dietrich et al. 2007).

Rapid Occurrence of Large Genomic Mutations in TEV upon Passaging in N. tabacum 35S::NlB Plants

To test whether there will indeed be rapid accumulation of virus variants with large deletions as suggested by estimates of W, we performed serial passage experiments. Here, we chose to perform four serial passages, with each passage having 3 weeks of duration, because genome size appears to be under stronger selection in long passages (Zwart et al. 2014). Five independent lineages of TEV were evolved. After four passages, deletions were detected by RT-PCR in two lineages: Lineage 2 and Lineage 3 (fig. 2A). These two deletions (nucleotide positions 7621–8484 and 7046–8500) conserved the
reading frame and were situated within the NiB cistron (positions 6982–8517 of the wild-type genome).

To detect virus deletion variants with much greater sensitivity, Illumina NGS was performed on these five TEV lineages evolved for four 3-week passages. A TEV genome region encompassing the NiAPro, NiB, and CP cistrons (positions 6377–9473 of the wild-type genome) was amplified and sequenced (Materials and Methods). We developed a new approach for identifying deletions, adapting GSNAP (Wu and Nacu 2010) to detect deletions instead of introns (Materials and Methods).

As a negative control, the same region was analyzed using NGS data for five lineages of TEV-eGFP evolved for three 9-week passages in wild-type N. tabacum plants (Zwart et al. 2014). No deletions were detected in the entire region, suggesting that our NGS analysis does not generate artifacts. The frequencies of deletion variants (see supplementary table S1, Supplementary Material online) in systemically infected leaves of transgenic plants were significantly higher than the frequencies found in our negative controls (Fisher’s exact test, P < 0.001). The NGS results were consistent with RT-PCR data; the deletion variant detected by RT-PCR in Lineage 2 (positions 7621–8484) and Lineage 3 (positions 7046–8500) was the variant found at the highest frequency by NGS, 0.249 and 0.246, respectively. Sixty-four different deletion variants were observed in total, ranging in frequency from 8.44/C2 10/C0 5 (positions 6538–8478; Lineage 4) to 0.249 (positions 7621–8484; Lineage 2). Interestingly, 6.25% of deletion variants disrupted the reading frame of the coding sequence. Moreover, 43.75% of deletion variants extended into genes flanking NiB, meaning that probably only the remaining 56.25% variants could be viable when transcomplemented by the NiB expressed by the plant.

In 17 deletion variants, the 5′ end of the deletion extended into NiAPro; the frequencies of these variants in the virus populations range from 8.44 × 10−5 (positions 6538–8478; Lineage 4) to 1.73 × 10−3 (positions 6460–8478; Lineage 4). Furthermore, in eight other variants the deletion started within NiAPro (upmost 5′ position being 6516) and ended within CP (downmost 3′ position being 9163), varying in frequency from 1.02 × 10−4 (positions 6923–9013; Lineage 4) to 1.88 × 10−3 (positions 6569–8746; Lineage 1). The 43.75% of genotypes carrying deletions larger than the NiB cistron will produce
viruses with truncated cistrons NiPro, CP, or both. Moreover, deletions beyond the NiPro cistron will lead to the loss of proteolytic cleavage sites and result in fusion of the truncated proteins. These genomes will probably be defective particles and some may additionally interfere with replication of the full-length virus. These defective particles can only replicate if they are complemented by a full-length TEV or possibly TEV-ΔNiPro, and some may additionally interfere with replication of the viral genome. These genomes will probably be defective particles and be removed by the cellular defense mechanisms.

The ends of the genomic deletions found by NGS (fig. 2C and D) were not uniformly distributed (Kolmogorov–Smirnov tests: 5' end n = 42, P < 0.001; 3' end n = 23, P = 0.001). Surprisingly, the 5' start of the genomic deletion was clustered in NiPro, suggesting a possible recombination hotspot outside of the NiPro cistron. For the 3' end of the deletion, ending sites clustered toward the 3' end of NiPro cistron. This clustering is consistent with the idea that we are more likely to detect deletion variants that only affect part of NiPro and are, therefore, likely viable by complementation in trans.

Our results demonstrate that rapid deletions of NiPro occurred upon passage in N. tabacum 35S::NiPro. In two lineages, these variants reached high frequencies, comprising a fourth of the population. These results are therefore a proof of principle that NiPro expression could have a similar effect on a virus population: allowing deletion mutants to be maintained in the virus populations, while also providing conditions in which these mutants could reach high frequencies. The immediate impact of these deletions will be that a part of the virus population is replication defective in host individuals that do not express the NiPro. This could have important ramifications for viral spread, and virus–host coevolution. On the other hand, the further evolution of a virus population in a host expressing virus genes may also be contingent upon the presence of genomic deletions in the viral genome.

**TEV Evolution in N. tabacum 35S::NiPro Plants Is Contingent upon the Presence of NiPro in the Viral Genome**

As we just described, serial passages of TEV in N. tabacum 35S::NiPro led to the rapid occurrence of a whole spectrum of deletion variants. How would these deletion mutants further evolve in transgenic plants, and would their evolution be different from that of a full-length TEV? To address this question and gain insight into the long-term evolution of viruses harboring NiPro-induced deletions, we performed a second evolution experiment for both TEV and TEV-ΔNiPro passages in N. tabacum 35S::NiPro, with five independent lineages for each virus. Li and Carrington (1995) showed that beta-glucuronidase (GUS) activity in leaves systemically infected by the TEV-ΔNiPro-GUS mutant was comparable to that of the parental TEV-GUS. We found a similar result for the accumulation of TEV and TEV-ΔNiPro in N. tabacum 35S::NiPro using RT-qPCR (fig. 3A): there was not a significant difference in virus accumulation at 7 dpi (t = 1.394, P = 0.201). The infection dynamics of the two viruses should therefore be similar, allowing us to compare the evolutionary outcomes. However, to compare the evolution of the full-length TEV and TEV-ΔNiPro, we had to devise a setup in which the integrity of the NiPro gene is ensured for TEV. Two previous studies have found that marker genes were highly stable in the viral genome when many 1-week passages were performed, although the markers were quickly lost in longer passages (Dolja et al. 1993; Zwart et al. 2014). By performing 30-one week passages with TEV and TEV-ΔNiPro in N. tabacum 35S::NiPro, we therefore hoped to avoid the occurrence of large genomic deletions in TEV. On the other hand, we realize that selection may not operate as effectively as during longer passages, due to short passage duration and frequent bottleneck associated to the mechanical transmissions among plants.

After 30 weeks of evolution, we measured virus accumulation of the evolved lineages (fig. 3B). The data were analyzed by ANOVA (table 1), and we found no effect of the virus type (TEV or TEV-ΔNiPro) on accumulation, although there was a strong effect of lineage. The significant lineage-dependent effects suggest that random factors have played an important role during evolution, implicating genetic drift effects due to short passage duration. When pairwise comparisons were made between the evolved lineages and their respective ancestral viruses, Lineage 1 of TEV-ΔNiPro had a significantly lower accumulation (t-test with Holm–Bonferroni correction: t = 4.001, P = 0.005). Accumulation of all other lineages was not significantly different from the ancestral virus, although the data suggest Lineage 3 and Lineage 5 of TEV-ΔNiPro also have low accumulation (fig. 3B).

All evolved lineages of TEV and TEV-ΔNiPro were fully sequenced to determine whether there were differences in their patterns of molecular evolution. We were particularly interested in understanding whether the use of NiPro in trans may have affected the regulation of expression and accumulation of all other viral components. Mutational spectra for both viral genotypes were dominated by point mutations as not a single insertion or deletion was observed (table 2). As anticipated, the passaging setup used leads to the retention of NiPro cistron by all TEV lineages, and by contrasting the evolution of TEV and TEV-ΔNiPro, we are therefore considering the evolutionary
effects of having a full-length or truncated genome. There were no differences between ratios of transitions to transversions; the ratio was 8.5 for TEV and 6.0 for TEV-DNiB (Fisher’s exact test: $P = 1.000$). Given that it is biochemically easier to produce transitions than transversions, it is not surprising to observe these biased ratios. Likewise, there was not a significant difference in the ratio of synonymous to nonsynonymous substitutions between the two viruses (Fisher’s exact test: $P = 0.330$); for TEV, this ratio was 1.375 whereas for TEV-DNiB it was 0.636. For both genotypes, mutations were evenly distributed along the genome (fig. 4A; Kolgomorov–Smirnov test: TEV $n = 19$, $P = 0.330$; TEV-DNiB $n = 21$, $P = 0.688$). An interesting example of convergent evolution was mutation A6805G (K2269E) that appeared in 80% of TEV lineages and in 40% of TEV-DNiB lineages. This nonsynonymous mutation is localized in a superficial domain of NiAPro (Nunn et al. 2005), and it could affect the proteinase function or modulate the interactions of NiAPro with other proteins. This same mutation has also been previously reported in serial-passage evolution experiments of TEV in N. tabacum: in 10% of lineages after 15 one-week passages (Bedhomme et al. 2012), and in 30% of lineages after three 9-week passages (Zwart et al. 2014).

In three lineages of TEV-DNiB, we observed three nonsynonymous mutations localized in the 6K1/CI junction (fig. 4B): G3618U (E1206D) (Lineage 1), A3622G (I1208V) (Lineage 5), and C3629U (T1210M) (Lineage 3), affecting residues P6, P4, and P2 of the proteolytic site, respectively (Adams et al. 2005). A conserved heptapeptide sequence has been identified at this site [E-X-X-Y-X-Q/S (or G)] at the 6K1/CI junction (Carrington and Dougherty 1987, 1988; Carrington et al. 1988; Adams et al. 2005). Each mutation was observed in a single lineage, and combinations thereof were not observed.
The E1206D mutation at the P6 site may only weakly affect the proteolytic processing as amino acids E and D are widely represented in Potyviridae family at this site (Adams et al. 2005). On the other hand, I1208V at position P4 should have a stronger impact on processing activity; in only 7% of all Potyviridae is I present at this site, whereas V is present in 80% (Adams et al. 2005). Similarly, T1210M in P2 may also have a strong impact on processivity; T is found at the P2 site in 40% and F (15%) are the most common in position P2 among the Potyviridae, M is represented with a frequency of only 2%. Generally, mutations at positions P2/P4 may involve regulation of the cleavage rate (Dougherty et al. 1989).

**Fig. 4.**—(A) Mutations found in TEV-ΔNlb and TEV consensus genome after 30 weekly passages in N. tabacum 35S::Nlb transgenic plants. Black circles represent synonymous mutation, gray circles nonsynonymous mutation, and white circle mutations localized in UTRs. (B) TEV-ΔNlb mutations on the NlaPro proteolytic 6K1/CI site. P6 to P2’ (E-X-X-Y-X-Q-G/S) are key amino acids constituting the NlaPro binding site. Cleavage site is between P1 and P1’ (vertical arrow). (E) (26.5%) and (D) (12.8%) are the most frequent amino acids at P6 position in the family Potyviridae; V (80%) is the most frequent in position P4; H (40%) and F (15%) are the most common in position P2 among the Potyviridae, M is represented with a frequency of only 2%. Generally, mutations at positions P2/P4 may involve regulation of the cleavage rate (Dougherty et al. 1989). (C) Accumulation of mutants E1206D, I1208V, T1210M, K2269E, and the combinations K2269E/E1206D, K2269E/I1208V, and K2269E/T1210M. Mutant viruses belong to four groups (A–D) with homogeneous accumulation (Holm-Bonferro post hoc test, P < 0.05).
in only 2% of the Potyviridae (2%). Generally, mutation at sites P2/P4 may involve regulation of the cleavage rate (Dougherty et al. 1989).

We conclude that the evolution of TEV and TEV-ΔNIb in *N. tabacum* 35S::NIb was similar when transmission bottlenecks are close in time. Although there was a significant lineage-dependent effect on accumulation, only one evolved lineage of TEV-ΔNIb had a significantly lower accumulation than its ancestral virus. Moreover, patterns of molecular evolution were similar. There were no significant differences in the ratios of transitions to transversions, synonymous to nonsynonymous mutations, and the overall distribution of mutations. The one mutation that occurred in multiple lineages, A6805G, occurred in both lineages of TEV and TEV-ΔNIb. Therefore, the only conspicuous difference between the two viruses is mutations localized in the 6K1/CI junction.

### Analysis of Mutations in the 6K1/CI Junction of TEV-ΔNIb

The K2269E, E1206D, I1208V, and T1210M mutations were introduced in the pMTEV infectious clone by site-directed mutagenesis. As K2269E is common among the different lineages, we also generated clones containing this mutation in combination with the other three. As a proxy for within-host fitness, we measured viral accumulation of each simple and double mutant at 2.67 dpi by RT-qPCR in three biological replicates. Data were fitted to a GLM with a log-link function and gamma distributed error structure, using replicate as factor and group as nested within virus genotype. A significant effect of virus genotype was observed between the two viruses is mutations localized in the ancestral virus. The mutations we found in evolved lineages, A6805G, occurred in both lineages of TEV and TEV-ΔNIb. Therefore, the only conspicuous difference between the two viruses is mutations localized in the 6K1/CI junction.

Concluding Remarks

NIRVs are thought to impact viruses by—in some cases—conferring virus resistance to host organisms (Koonin, 2010). However, these virus-derived sequences could also impact the evolution of virus populations. Here, we used an experimental model system to study the effects of NIRV expression on virus populations, using TEV and transgenic tobacco plants expressing its replicate, NIb. We found that 1) the loss of NIb cistron leads to higher within-host fitness in transgenic plants, 2) that genomic deletions quickly accumulate in a full-length virus passaged in transgenic plants, and 3) that the evolution of a virus population can be contingent upon the deletion of virus genes, although many aspects of phenotypic and molecular evolution were unchanged. Overall, our results therefore suggest that the expression of NIRVs could have far-reaching consequences for virus evolution. Moreover, these effects can occur on relatively short timescales (i.e., deletion variants were generated de novo and reached frequencies of ~0.25 within four 3-week passages). These changes on short timescales support the notion that plausible epidemiological models of virus spread in NIRV-expressing host populations need to take virus evolution into consideration. For example, if large genomic deletions occur rapidly in NIRV-expressing hosts, then eventually many virus-infected hosts will only be infectious to those hosts expressing the NIRV. The generation of various classes of infectious hosts could in this case decrease viral transmission between hosts, an effect that could be further augmented by the clustering of NIRV-expressing hosts.

Our results provide further support for the notion of a relationship between genome size and replication/within-host competitive fitness (Sakai et al. 1999; Marks et al. 2005; Zwart et al. 2010, 2014), although there are clearly limitations to this idea (Bull et al. 2004). Not only could TEV-ΔNIb outcompete TEV in transgenic plants, also all independent TEV lineages accumulate various NIb deletions upon passing. We identified 64 different deletion variants, most of which conserved the reading frame. In 17 deletion variants the deletion extended far into the upstream cistron, NLaPro, while for eight deletion variants, the deletion also extended into the downstream cistron, CP. Because both NLaPro and CP proteins are essential for virus replication and these deletions removed one or more proteolytic

---

cleavage sites, the virus variants probably cannot replicate autonomously, even in the transgenic plants. Therefore, the virus population in the simulated NIRV-expressing host appears to have generated two types of defective viruses: 1) those with deletions in Nib only and can replicate autonomously in the transgenic plants but not in the wild-type plants and 2) those with deletions extending beyond the Nib cistron, which cannot replicate autonomously in either host. On the basis of these observations, we speculate that the NIRV-expressing hosts may be a source of defective genotypes in natural virus populations. If a deletion genotype arises de novo in a NIRV-expressing host, it will not require coinfection with a full-length virus to be maintained in the population. Consequently, it may more easily reach intermediate to high frequencies in a NIRV-expressing plant, allowing the defective virus to be sustained even in non-NIRV hosts by coinfection with full-length viruses (Montville et al. 2005). Complementation can slow down the rate at which deleterious alleles are eliminated from the virus population (Froissart et al. 2004; Sardanye and Elena 2010) and promote selection for defective-interfering particles as cheater genotypes that reduce mean population fitness (Turner and Chao 1999).

Our study has important implications for the development and agronomic use of transgenic plants resistant to viruses generated by inserting complete viral genes into the plant genome, in particular by inserting viral replicase genes (reviewed in Kundu and Mandal [2001]). From a biosafety standpoint, it has been known for several years that recombination between viral sequences transgenically expressed by plants and virus genomes may occur (Turturo et al. 2008; Morroni et al. 2009), raising concerns about whether this transgenic strategy would favor the emergence of new viruses (Tepfer 2002; Thompson and Tepfer 2010). Here, we have illustrated a different evolutionary scenario: the expression by the plant of functional viral proteins creates the conditions in which viruses with long or even full deletions of the complemented protein may dominate the population, thus precluding their spread in nontransgenic plants.

Finally, it is fair to mention that our experimental system remains artificial to some extent. Although many examples of DNA viruses using host enzymes for replication exist (e.g., geminiviruses and polydnaviruses), no RNA virus has been described so far that does not encode for its own functional RdRp, even though certain plants encode functional RdRps of viral origin in their genomes, for example, the narnavirus-like RdRp found in the mitochondria of Arabidopsis thaliana (Hong et al. 1998). In this respect, a critical reader may argue that RdRp was not the best choice for studying the pseudogenization of viral genes in the presence of a homologous NIRV. However, our results should be strong enough to convince such a reader that even a highly constrained gene, such as RdRp can be lost from viral genomes if provided in trans by the host.

Supplementary Material
Supplementary table S1 is available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

Acknowledgments
The authors thank Francisca de la Iglesia, Paula Agudo, and Ángels Prósper for excellent technical support, Dr José-Antonio Darós for gifting pMTEV, and the Bioinformatics Core Service at the IBMCP for the support provided. This project was supported by grant 22371 from the John Templeton Foundation to S.F.E. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. Additional support was provided by the Spanish Ministerio de Economía y Competitividad (MINECO) grant BFU2009-06993 and European Commission FP7-ICT program grant 610427-EvoEvo to S.F.E., by a predoctoral fellowship from MINECO to N.T., and by MINECO grant JCI2011-10379 and Rubicon grant from the Netherlands Organization for Scientific Research (www.nwo.nl) to M.P.Z.

Literature Cited


Associate editor: Eugene Koonin