

Multiple Barriers to the Evolution of Alternative Gene Orders in a Positive-Strand RNA Virus

Anouk Willemsen,^{*1} Mark P. Zwart,^{*2} Nicolas Tromas,^{*3} Eszter Majer,^{*} José-Antonio Daròs,^{*}
and Santiago F. Elena^{*†}

^{*}Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas–Universidad Politécnica de Valencia, 46022 Valencia, Spain, and [†]Santa Fe Institute, New Mexico 87501

ORCID ID: 0000-0002-8511-3244 (A.W.)

ABSTRACT The order in which genes are organized within a genome is generally not conserved between distantly related species. However, within virus orders and families, strong conservation of gene order is observed. The factors that constrain or promote gene-order diversity are largely unknown, although the regulation of gene expression is one important constraint for viruses. Here we investigate why gene order is conserved for a positive-strand RNA virus encoding a single polyprotein in the context of its authentic multicellular host. Initially, we identified the most plausible trajectory by which alternative gene orders could evolve. Subsequently, we studied the accessibility of key steps along this evolutionary trajectory by constructing two virus intermediates: (1) duplication of a gene followed by (2) loss of the ancestral gene. We identified five barriers to the evolution of alternative gene orders. First, the number of viable positions for reordering is limited. Second, the within-host fitness of viruses with gene duplications is low compared to the wild-type virus. Third, after duplication, the ancestral gene copy is always maintained and never the duplicated one. Fourth, viruses with an alternative gene order have even lower fitness than viruses with gene duplications. Fifth, after more than half a year of evolution in isolation, viruses with an alternative gene order are still vastly inferior to the wild-type virus. Our results show that all steps along plausible evolutionary trajectories to alternative gene orders are highly unlikely. Hence, the inaccessibility of these trajectories probably contributes to the conservation of gene order in present-day viruses.

KEYWORDS gene order; genome evolution; virus evolution; experimental evolution

THE organization of genes within a genome can vary greatly between phylogenetically distant species. Several comparative studies of bacterial, archaeal, and eukaryotic genomes have concluded that, in general, gene order is not conserved (Himmelreich *et al.* 1997; Kolstø 1997; Koonin and Galperin 1997; Siefert *et al.* 1997; Watanabe *et al.* 1997; Dandekar *et al.* 1998; Rocha 2008). In stark contrast, gene order within virus orders and families is often conserved.

Viral genomes tend to be smaller, with minimal intergenic sequences and in some cases overlapping genes (Lynch 2006; Belshaw *et al.* 2007; Koonin 2009). The reasons why a particular gene order supports the required patterns of virus gene expression and virus replication have in many cases also been elucidated. For example, different expression levels for viral gene products can arise through the generation of sub-genomic RNAs (de Haan *et al.* 2003), frameshifts (Chung *et al.* 2008), stuttering of the RNA polymerase in intergenic regions (Wertz *et al.* 1998), having multiple genome segments with different regulatory elements (Sullivan and Ahlquist 1997), or varying the frequency of different genome segments (Sicard *et al.* 2013). Altering gene order in viral genomes therefore can be associated with great fitness costs (Novella *et al.* 2004; Springman *et al.* 2005), and rearrangement of essential genes is not always reversible (Wertz *et al.* 1998). Nevertheless, it is not always obvious why gene order has been so well conserved in viruses.

Phylogenetic approaches have helped to unveil interesting patterns in gene-order evolution. An intriguing example

Copyright © 2016 by the Genetics Society of America

doi: 10.1534/genetics.115.185017

Manuscript received November 19, 2015; accepted for publication February 7, 2016; published Early Online February 10, 2016.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1.

¹Corresponding author: Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas–Universidad Politécnica de Valencia, Campus UPV, CPI 8E, Laboratory 2.0.4, Ingeniero Fausto Elio s/n, 46022 Valencia, Spain.
E-mail: anwil1@ibmcp.upv.es

²Present address: Institute of Theoretical Physics, University of Cologne, Zùlpicher Strasse 77, D-50937 Köln, Germany.

³Present address: Department of Biological Sciences, University of Montreal, Pavillon Marie-Victorin, 90 avenue Vincent-d'Indy, Montreal, QC H2V 2S9, Canada.

is the endornaviruses found in plants, fungi, and protists (Valverde *et al.* 1990; Wakarchuk and Hamilton 1990; Fukuhara *et al.* 2006), which have acquired domains with similar functions from these different hosts (Song *et al.* 2013). Despite their distinct origins, these domains have a strict functional order within the *Endornavirus* genus (Roossinck *et al.* 2011; Song *et al.* 2013), even though they highly vary as to presence or absence. Whereas phylogenetic approaches can identify patterns in gene-order evolution, experimental evolution potentially could shed light on the short-term dynamics and underlying mechanisms. The evolution of gene order has been explored experimentally for phage T7 (Springman *et al.* 2005) and *Vesicular stomatitis virus* (VSV) (Pesko *et al.* 2015).

T7 has a double-stranded DNA genome of ~40 kb. The T7 genome contains three promoters for the *Escherichia coli* RNA polymerase, and host-mediated transcription draws the first part of the T7 genome into the cell. Once the T7 RNA polymerase protein in this early region is expressed, it initiates transcription for the rest of the genome from its associated promoters, internalizing the remaining part of the T7 genome and achieving a high level of transcription of the late genes. The artificial repositioning of the T7 RNA polymerase downstream of its normal location resulted in a delay of the phage life cycle and had severe impacts on viral fitness (Endy *et al.* 2000; Springman *et al.* 2005). Subsequent experimental evolution led to only a modest recovery in fitness (Springman *et al.* 2005). In one evolved line, the RNA polymerase was restored to the wild-type position, but at the same time, other genes in T7 genome were relocated, and a full regain of fitness was not observed. This study demonstrates that gene order is important for fitness and that the wild-type levels of fitness are not rapidly re-evolved after reorganizing the genome.

VSV is a nonsegmented negative-strand RNA virus with a genome size of ~11.2 kb that encodes five proteins. Transcription of VSV is regulated by a single promoter located at the 3' end of the genome. Stuttering of the VSV RNA polymerase causes greater messenger RNA (mRNA) production in upstream genes, which is a strategy to regulate gene expression. Gene order in VSV was altered by moving the nucleocapsid (*N*) gene, located at the 3' end, sequentially downstream in the genome (Wertz *et al.* 1998). This led to a stepwise decrease in *N* mRNA production and protein expression (Wertz *et al.* 1998). The initial fitness of the VSV variants was low (Novella *et al.* 2004), but fitness gains were observed in evolutionary time, and fitness improved the most for the variant with the lowest initial fitness (Pesko *et al.* 2015). Nevertheless, the variant with the wild-type gene order still grew better than the other variants.

T7 and VSV are different in nature, gene content, and structure, and both use different replication strategies. Despite these differences, gene order is important for the regulation of gene expression in both viruses. Moreover, the different constraints on gene-order evolution observed in these two studies raise the question of their general applicability. Do most viruses and viral genome architectures suffer

from the major constraints, as has been observed for T7? What about viruses that do not use promoters for the transcription of mRNA, such as the positive-strand RNA viruses? Many emerging viruses with large societal impacts, as well as viral model systems with great relevance to fundamental research, are positive-strand RNA viruses, making it relevant to address these questions.

The positive-strand RNA viruses represent the largest group of viruses (Francki *et al.* 1991) and are classified into three tribes: picorna-, alpha-, and flavi-like (Koonin and Dolja 1993; Fauquet *et al.* 2005). These viruses are characterized by conserved gene clusters and especially the helicase-polymerase arrangement, where the helicase gene is typically located upstream of the polymerase gene (Koonin and Dolja 1993). In particular, the picorna-like tribe is identified by the partial conservation of core genes that consist of the RNA-dependent RNA polymerase (RdRp), a chymotrypsin-like protease (3CPro), a superfamily 3 helicase (S3H), and a genome-linked protein (VPg) (Koonin and Dolja 1993; Fauquet *et al.* 2005; Koonin *et al.* 2008). Moreover, core genes tend to form ordered arrays, whereas noncore genes are responsible for genome reorganization and recombination between distant groups of viruses from all three tribes (Koonin and Dolja 1993; Fauquet *et al.* 2005).

To study the evolution of alternative gene orders in positive-strand RNA viruses in the context of a real multicellular-host infection, we used the picorna-like *Tobacco etch virus* (TEV; genus *Potyvirus*, family Potyviridae). TEV has a 9.5 kb genome that codes for a single polyprotein that is further processed into 11 mature peptides (Figure 1A). Because it is composed of positive-strand RNA, the TEV genome can be immediately translated on entering a cell. Unlike the bacteriophage T7 and VSV, replication of TEV is not regulated by a promoter but by the VPg protein linked to the 5' end of the genome, which helps to initiate RNA replication (Puustinen and Mäkinen 2004). Then the viral NIa-Pro protease is responsible for processing the polyprotein at most of its proteolytic cleavage sites (Revers and García 2015), except for the processing of the first two proteins, P1 serine protease and HC-Pro cysteine protease, which are self-cleaving. Therefore, the rate of synthesis of the mature proteins depends on three factors: the amount of positive-strand RNA accessible to ribosomes, the rate and effectiveness of translation into the polyprotein, and the efficiency of its proteolysis by the viral proteases. Within the Potyviridae family, gene order has been strictly conserved, including the *Bymovirus* genus, which has evolved a segmented bipartite genome (Revers and García 2015).

Given the need for correct polyprotein processing, a polyprotein-mediated gene-expression strategy is likely to impose constraints on gene-order evolution. Rearranged viral genomes must conserve proteolytic cleavage sites, and as a consequence, most recombination events are likely to disrupt polyprotein processing. However, even if they do not, would the resulting viruses be viable? The RdRp of TEV is coded by the *NIb* gene, and deletion thereof leads to virus variants that

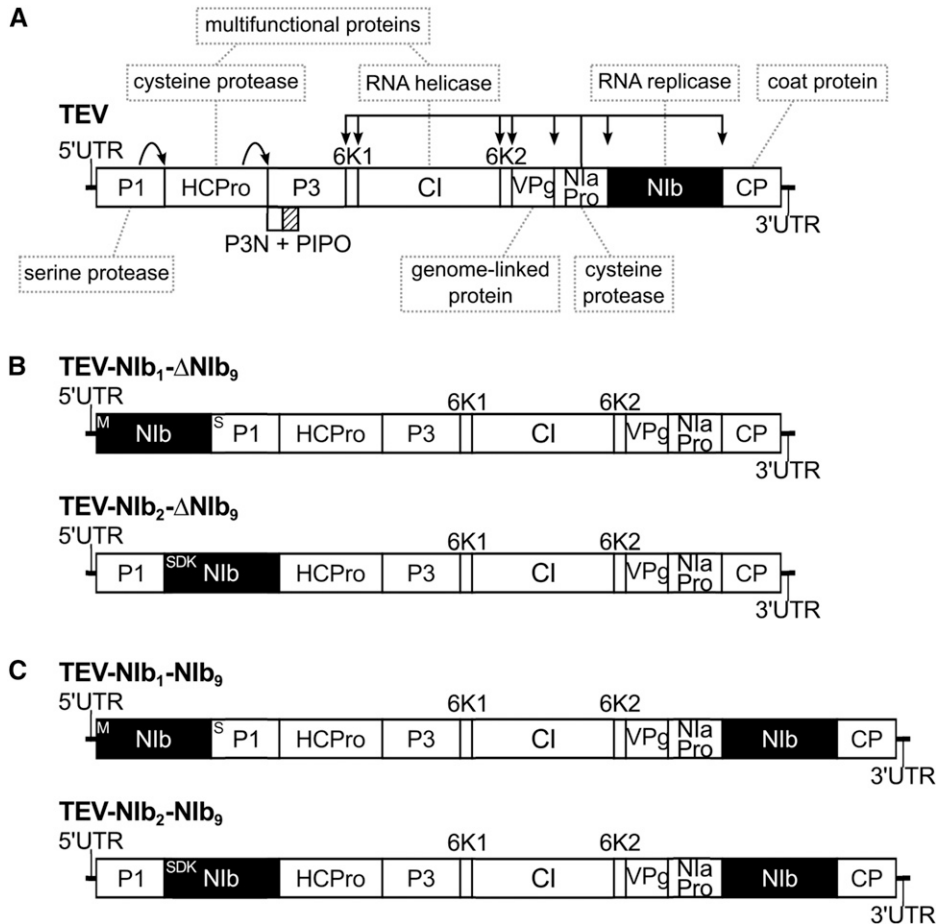


Figure 1 Schematic representations of the different *Tobacco etch virus* genotypes used in this study. The wild-type TEV (A) codes for 11 mature peptides, including P3N-PIPO embedded within the P3 protein at a +1 frameshift. Relevant names of the proteins products are indicated in the gray dashed boxes. For more information on the features and proposed functions of the proteins, consult Revers and García (2015). For constructing the viral genotypes with the Nib replicase in an alternative position, a copy of *Nib* was moved to the first and second positions within the TEV genome, removing *Nib* from its original position (B). For constructing the viral genotypes with a duplication of *Nib*, a second copy of *Nib* was introduced into the first and second positions within the TEV genome, preserving the original *Nib* and thereby generating a duplication event (C). For simplification, P3N-PIPO is drawn only at the wild-type TEV.

cannot replicate on their own (Li and Carrington 1995). In a previous study, we considered whether virus infectivity was maintained when the *Nib* gene was relocated to all possible intergenic positions in the TEV genome without maintaining the original *Nib* copy (Majer *et al.* 2014). Only two of nine viruses with reordered genomes were viable: the genotypes with *Nib* relocated to the first two intergenic positions (Figure 1B). A variant with *Nib* relocated to the third intergenic position was not infectious in wild-type plants, while it could cause infection in transgenic plants expressing NIB *in trans*, albeit at a lower frequency (Majer *et al.* 2014). Moreover, in these cases, we always found an exact deletion of the *Nib* gene, and therefore, we do not consider this a viable virus for reordering.

In this study, we used experimental evolution to better understand the dynamics of genome architecture evolution and gene-order conservation. Why has gene order been conserved within positive-strand RNA virus orders and families? Are there accessible evolutionary trajectories to alternative orders, or is a lack of accessible trajectories an important impediment in present-day viruses? Here we use the two viable reordered TEV variants to address these issues. We first explore whether there are accessible evolutionary trajectories that result in these two viable reordered viruses. To consider whether such natural trajectories exist, genomes containing duplications

of the *Nib* gene were generated, tested for viability, and evolved in plants. We consistently found that the *Nib* gene at the alternative position was rapidly lost owing to the occurrence of large genomic deletions. Finally, we explored the evolutionary potential of viruses with reordered genomes by evolving viruses with a single *Nib* gene in an alternative position. We then measured virus accumulation and fitness and used next-generation sequencing to identify genomic changes. Although we found evidence for adaptation of these reordered viruses in terms of increasing virus accumulation, they were still less fit than the wild-type virus. This study therefore revealed multiple barriers to the evolution of alternative gene orders.

Materials and Methods

Viral constructs, virus stocks, and plant infections

TEV-Nib₁-ΔNib₉, TEV-Nib₂-ΔNib₉, TEV-Nib₁-Nib₉, and TEV-Nib₂-Nib₉ were generated from complementary DNA (cDNA) clones constructed using plasmid pGTEV_a, which consists of a TEV infectious cDNA (accession no. DQ986288, including two silent mutations, G273A and A1119G) flanked by *Cauliflower mosaic virus* (CaMV) 35S promoter and terminator in a binary vector derived from pCLEAN-G181 (Thole *et al.* 2007).

Clones were constructed using standard molecular biology techniques, including PCR amplification of cDNAs with the high-fidelity Phusion DNA Polymerase (Thermo Scientific), DNA digestion with *Eco31I* (Thermo Scientific) for assembly of DNA fragments (Engler *et al.* 2009), DNA ligation with T4 DNA Ligase (Thermo Scientific), and transformation of *E. coli* DH5 α by electroporation. Sanger sequencing confirmed the sequences of the resulting plasmids. The ancestral and resulting binary plasmids were transformed in *Agrobacterium tumefaciens* C58C1 harboring helper plasmid pCLEAN-S48 (Thole *et al.* 2007). *Nicotiana tabacum* L cv. Xanthi (NN) plants were agroinoculated with *A. tumefaciens* cultures (Bedoya and Daròs 2010), and symptomatic tissue was collected 7 days postinoculation (dpi). To generate large virus stocks, the collected tissue was homogenized, ground into fine powder using liquid nitrogen and a mortar, and resuspended 1:1 in phosphate buffer (50 mM KH₂PO₄, pH 7.0, and 3% polyethylene glycol 6000). The third true leaf of 4-week-old *N. tabacum* plants was mechanically inoculated with 50 μ l of the TEV genotypes. *N. tabacum* plants were kept in a BSL-2 greenhouse at 24° with 16 hr of light. All systemically infected tissues were harvested 7 dpi and stored at –80°.

For the serial passage experiments, 500 mg of homogenized stock tissue was ground into fine powder and diluted in 500 μ l of phosphate buffer. From this mixture, 20 μ l was then mechanically inoculated on the third true leaf using carborundum. Depending on the virus variant and the passage duration, seven or five independent replicates were used. At the end of the designated passage duration (3 or 9 weeks), all leaves above the inoculated leaf were collected and stored at –80°. For subsequent passages, the frozen tissue was homogenized, and a sample of the homogenized tissue was ground and resuspended with an equal amount of phosphate buffer (Zwart *et al.* 2014). Then 20 μ l was rub-inoculated on the third true leaf. All the experiments involving plant infections have been performed following the Spanish National Guidelines for Plant Research.

Reverse transcription polymerase chain reaction (RT-PCR)

To determine whether deletions had occurred at the *Nib* locus, RNA was extracted from 100 mg of homogenized infected tissue using the InviTrap Spin Plant RNA Mini Kit (Strattec Molecular). Reverse transcription (RT) was performed using M-MuLV Reverse Transcriptase (Thermo Scientific) and the reverse primer 5'-CGCACTACATAGGAGAATTAG-3' located in the 3' UTR of the TEV genome. PCR was then performed with Taq DNA Polymerase (Roche) and primers flanking the *Nib* gene inserted at the first and second positions—forward 5'-GCAATCAAGCATTCTACTTC-3' and reverse 5'-CCTGATATGTTTCCTGATAAC-3'—as well as primers flanking the *Nib* gene at the original position—forward 5'-TCATTACAAA CAAGCACTTG-3' and reverse 5'-GCAAAGTCTCATGTG TGG-3'. PCR products were electrophoresed using 1% agarose gel. Based on the amplicon size and the genome size of the ancestral viruses, genome size was estimated for the

corresponding evolved viruses, assuming that deletions occurred only within the amplified region.

Accumulation and within-host competitive fitness assays

Prior to performing assays, the viral copy numbers per 100 mg of tissue of the ancestral virus stocks and all evolved lineages were determined for subsequent assays. The InviTrap Spin Plant RNA Mini Kit (Strattec Molecular) was used to isolate the total RNA of 100 mg of homogenized infected tissue. Real-time quantitative RT-PCR (RT-qPCR) was then performed using the One-Step SYBR PrimeScript RT-PCR Kit II (Takara) in accordance with the manufacturer's instructions in a StepOnePlus Real-Time PCR System (Applied Biosystems). Specific primers for the *CP* (*coat protein*) gene were used—forward 5'-TTGGTCTTGATGGCAACGTG-3' and reverse 5'-TGTGCCGTTCACTGTCTTCCT-3'. StepOne Software v.2.2.2 (Applied Biosystems) was used to analyze the data. The concentration of genome equivalents per 100 mg of tissue was then normalized to that of the sample with the lowest concentration using phosphate buffer.

For the accumulation assays, 4-week-old *N. tabacum* plants were inoculated with 50 μ l of these dilutions of ground tissue. For each ancestral and evolved lineage, three independent plant replicates were used. Plant height was measured (data not shown) and leaf tissue was harvested 7 dpi. Virus accumulation then was determined by means of RT-qPCR for the *CP* gene for TEV, TEV-Nib₁- Δ Nib₉, TEV-Nib₂- Δ Nib₉, TEV-Nib₁-Nib₉, TEV-Nib₂-Nib₉, and the evolved lineages of these viruses.

To measure within-host competitive fitness, we used TEV carrying an enhanced green fluorescent protein (TEV-eGFP) (Bedoya and Daròs 2010) as a common competitor. TEV-eGFP has proven to be stable for up to 6 weeks (using 1- and 3-week serial passages) in *N. tabacum* (Zwart *et al.* 2014) and is therefore not subjected to *eGFP* loss in our 1-week-long competition experiments. All ancestral and evolved viral lineages were again normalized to the sample with the lowest concentration, and 1:1 mixtures of TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ genome equivalents were made with TEV-eGFP. For both TEV-Nib₁- Δ Nib₉ and TEV-Nib₂- Δ Nib₉, the ratio was 10 genome equivalents set to 1 of TEV-eGFP. The mixture was mechanically inoculated on *N. tabacum* plants using three independent plant replicates per viral lineage. The plant leaves were collected at 7 dpi and stored at –80°. RNA was extracted from 100 mg of homogenized tissue as described earlier. RT-qPCR for the *CP* gene was used to determine total viral accumulation, and independent RT-qPCR reactions also were performed for the *eGFP* sequence using specific primers—forward 5'-CGACAACCAC TACCTGAGCA-3' and reverse 5'-GAACTCCAGCAGGAC CATGT-3'. The ratio *R* of the evolved and ancestral lineages to TEV-eGFP is then $R = (n_{CP} - n_{eGFP})/n_{eGFP}$, where n_{CP} and n_{eGFP} are the RT-qPCR measured copy numbers of *CP* and *eGFP*, respectively. Then we can estimate the within-host competitive fitness as $W = \sqrt[4]{R_t/R_0}$, where R_0 is the ratio at

the start of the experiment, and R_t is the ratio after t days of competition (Carrasco *et al.* 2007).

Sanger sequencing

To determine the exact positions of the deletions detected by RT-PCR in the evolved TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ lineages, the genomes were partly sequenced by Sanger's method. RT was performed using AccuScript Hi-Fi (Agilent Technologies) reverse transcriptase and the reverse primer 5'-TTGCACCTTGTGTGACCAC-3' located in the *P3* gene of the TEV genome. PCR was then performed with Phusion DNA Polymerase (Thermo Scientific) and primers flanking the deletions of *Nib* at the first and second positions—forward 5'-GCAATCAAGCATTCTACTTC-3' and reverse 5'-CCTGATATGTTTCCTGATAAC-3'. Sanger sequencing was performed at GenoScreen (Lille, France, www.genoscreen.com) with an ABI3730XL DNA analyzer. Six sequencing reactions were done per lineage using the same two outer primers as used for PCR amplification plus four inner primers—forward 5'-CAATTGTTGCAAGTGTGC-3' and 5'-ACACGTACTGGCTGCAGCG-3' and reverse 5'-GCTCTTCTTGCTAATGAT-3' and 5'-ATGGTATGAAGAATGCCTC-3'. Sequences were assembled using Geneious v.8.0.3 (www.geneious.com), and the start and end positions of the deletions were determined. Based on the original TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ reference sequences, new reference sequences were constructed for each of the evolved lineages.

Illumina next-generation sequencing (NGS), variants, and SNP calling

For Illumina NGS of the evolved and ancestral lineages, the viral genomes were amplified by RT-PCR using AccuScript Hi-Fi Reverse Transcriptase (Agilent Technologies) and Phusion DNA Polymerase (Thermo Scientific), with six independent replicates that were pooled. TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ were amplified using three primer sets (set 1: 5'-GCAATCAAGCATTCTACTTCTATTGCAGC-3' and 5'-CCTGATATGTTTCCTGATAAC-3'; set 2: 5'-ACACGTACTGGCTGTCAGCG-3' and 5'-CATCAATGTCAATGGTACAC-3'; set 3: 5'-CCCGTGAAACTCAAGATAG-3' and 5'-CGCACTACATAGGAGAATTAG-3'). TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ were amplified using three different primer sets (set 1: 5'-GCAATCAAGCATTCTACTTCTATTGCAGC-3' and 5'-TATGGAAGTCTGTGGATTTCCAGATCC-3'; set 2: 5'-TTGACGCTGAGCGGAGTGATGG-3' and 5'-AATGCTTCCAGAATATGCC-3'; set 3: 5'-TCATTACAAACAAGCACTTG-3' and 5'-CGCACTACATAGGAGAATTAG-3'). Equimolar mixtures of the three PCR products were made. We did not have any Illumina data for the TEV-Nib₁-Nib₉ 3WL1 (3-week passages, lineage 1) because this sample was lost during the preprocessing steps for sequencing. Sequencing was performed at GenoScreen. Illumina HiSeq2500 2 × 100-bp paired-end libraries with dual-index adaptors were prepared along with an internal PhiX control. Libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina). Sequencing quality control was performed by GenoScreen

based on PhiX error rate and Q30 values. Read artifact filtering and quality trimming (3' minimum Q28 and minimum read length of 50 bp) were done using FASTX-Toolkit v.0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Dereplication of the reads and 5' quality trimming requiring a minimum of Q28 were done using PRINSEQ-lite v.0.20.4 (Schmieder and Edwards 2011). Reads containing undefined nucleotides (N) were discarded. As an initial mapping step, the evolved sequences were mapped using Bowtie v.2.2.4 (Langmead and Salzberg 2012) against their corresponding ancestral sequence: TEV (GenBank accession number DQ986288, including two silent mutations: G273A and A1119G), TEV-Nib₁-ΔNib₉ (GenBank accession number KT203714), TEV-Nib₂-ΔNib₉ (GenBank accession number KT203715), TEV-Nib₁-Nib₉ ancestral (GenBank accession number KT203712), TEV-Nib₂-Nib₉ ancestral (GenBank accession number KT203713), and against the evolved TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ references including the corresponding deletions. Subsequently, mutations were detected using SAMtools' mpileup (Li *et al.* 2009) in the evolved lineages compared with their ancestral lineages. We were only interested in mutations at a frequency of >10%. Therefore, we present frequencies as reported by SAMtools, which has a low sensitivity for detecting low-frequency variants (Spencer *et al.* 2014).

After the initial premapping step, error correction was done using Polisher v.2.0.8 (available for academic use from the Joint Genome Institute), and consensus sequences were defined for every lineage. Subsequently, the cleaned reads were remapped using Bowtie v.2.2.4 against the corresponding consensus sequence for every lineage. The remapping was efficient, with about 84–89% and 89–92% of the paired reads mapping exactly one time for the TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ lineages, respectively. For the TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ lineages, this was 86–89% and 85–89%, respectively. For each new consensus, SNPs within each virus population were identified using SAMtools' mpileup and VarScan v.2.3.7 (Koboldt *et al.* 2012). For SNP calling, maximum coverage was set to 40,000, and SNPs with a frequency of <1% were discarded.

Data availability

The sequences of the ancestral viral stocks were submitted to GenBank (GenBank accession numbers KT203711—KT203715). Supplemental Material, Figure S1 and Figure S2 show the distribution of the SNP frequencies in the evolved virus lineages. Table S1 shows the results of a non-linear regression analysis of the viral genome sizes over time. Table S2 shows the high-frequency mutations of the evolved TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ lineages compared with their ancestral sequence. Table S3 shows the within-population sequence variation of the evolved and ancestral TEV-Nib₁-Nib₉ lineages after the remapping step. Table S4 shows the within-population sequence variation of the evolved and ancestral TEV-Nib₂-Nib₉ lineages after the remapping step. Table S5 shows the high-frequency mutations of the evolved TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ lineages compared

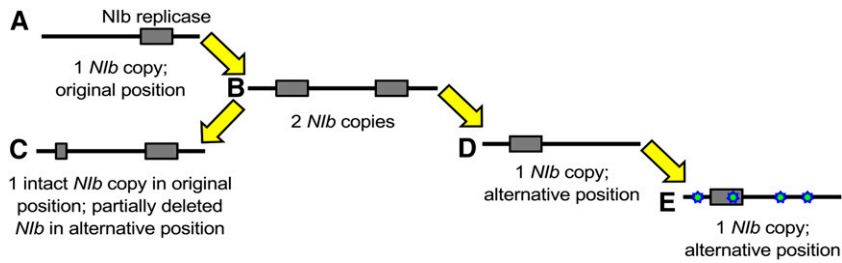


Figure 2 Plausible evolutionary trajectory for reordering of an essential gene in a potyvirus genome. First, the essential gene must be duplicated (variant a → variant b) to ensure that the virus can replicate autonomously at all times. Subsequently, we anticipate that one copy of the duplicate gene will be lost: either the second new copy is deleted (variant c) or the alternative gene copy is deleted (variant d). In the latter event, further evolution will take place to accommodate this reordered genome (variant e).

with their ancestral sequence. Table S6 shows the within-population sequence variation of the evolved and ancestral TEV-*Nib*₁- Δ *Nib*₉ lineages after the remapping step. Table S7 shows the within-population sequence variation of the evolved and ancestral TEV-*Nib*₂- Δ *Nib*₉ lineages after the remapping step.

Results

Study framework: plausible evolutionary trajectories to alternative gene orders

For the rearrangement of genes in a viral genome, many potential evolutionary trajectories can be envisioned. Here we use “plausible” to denote trajectories that could be traversed by the virus in terms of consistently maintaining replication (*i.e.*, virological perspective), whereas we use the term “accessible” to denote trajectories that could be traversed if the fitness and stability of intermediate steps are considered (*i.e.*, evolutionary perspective). In the potyvirus model we are considering, a number of constraints conspire to effectively limit the number of plausible trajectories to one. First, we are considering repositioning of the essential *Nib* replicase, which must be present in every cell that will contribute to infection and between-host transmission. When plants are inoculated with multiple potyvirus genotypes, the observed rate of cellular co-infection is typically very low, with the main exception being early infection prior to systemic movement (Dietrich and Maiss 2003; Zwart *et al.* 2011; Tromas *et al.* 2014a; Gutiérrez *et al.* 2015). It is therefore not surprising that while TEV missing the *Nib* gene (TEV- Δ *Nib*) can autonomously infect plants expressing *Nib* (Li and Carrington 1995), it cannot co-infect wild-type tobacco plants when co-inoculated with a wild-type virus (Tromas *et al.* 2014b). Since each intermediate along a reordering trajectory must be capable of autonomous replication, a plausible trajectory for an essential gene—such as *Nib*—will necessarily involve a gene-duplication event. Second, although higher gene expression as a consequence of gene duplication may have benefits, the increase in genome size and the possible disruption of the expression of other genes could significantly reduce viral fitness. Therefore, the most plausible trajectory to a new gene order is gene duplication, followed by deletion of the ancestral copy (Figure 2). A complication in this process is that variants with a deletion of the new gene copy may very well be favored owing to the fine-tuning of polyprotein processing and expression levels at this position.

A deletion of the new gene copy would, however, be a cul-de-sac along this evolutionary trajectory (Figure 2, variant c). Alternatively, viruses with only the new copy may be fixed (Figure 2, variant d), successively followed by the evolutionary fine-tuning (*e.g.*, adaptive mutations) of the relocated gene (Figure 2, variant e).

Given that biological constraints strongly suggest that an evolutionary trajectory through gene duplication is the most plausible route to gene reordering, we decided to study this route in detail for the repositioning of *Nib* in the TEV genome. To get a complete picture of the likelihood that a new gene order can evolve by means of this route, we decided to consider the following five key steps: (1) the fitness of viruses with gene duplications, rendering an indication of how long such variants can persist, (2) the evolutionary potential of viruses with gene duplications, focusing on the stability of the new gene copy because this will show whether the viruses can act as a bridge to the evolution of a new gene order, (3) the fitness of viruses with a single *Nib* copy in an alternative position to determine the likely fate of such viruses in the background of their direct ancestor, the corresponding double *Nib* genotype, and (4) the evolutionary potential of viruses with a single *Nib* copy in an alternative position because should such a virus occur and have low fitness, we can infer whether—following a period of reproductive isolation—it could eventually be competitive with wild-type viruses.

From the outset, however, we are faced with a barrier to the evolution of alternative gene orders for TEV: there are only two alternative positions to place the *Nib* replicase for which viruses appear to be viable (Majer *et al.* 2014): (1) before the *P1* serine protease gene and (2) between *P1* and the *HC-Pro* cysteine protease genes (Figure 1B). A first barrier to the evolution of alternative gene order is therefore the number of potentially viable intergenic sites, which is limited to only two of nine for TEV. Recombination events leading to the movement of a gene, as well as conservation of the reading frame and polyprotein processing, will be rare, and in addition, all other things being equal, seven of nine of these events will lead down trajectories that are ultimately cul-de-sacs. The number of viable alternative positions, and therefore the effective supply of first-step recombinants leading to alternative gene order, is almost 10-fold smaller than that suggested by the mutational supply alone.

We therefore focus on those evolutionary pathways to reach the two viable alternative gene orders, and consequently, four different TEV genotypes were constructed.

The *Nib* gene was inserted at the first and second positions in the TEV genome while preserving *Nib* at the original position (Figure 1C). Henceforth, we refer to these viruses with a duplication of *Nib* as TEV-NIB₁-NIB₉ and TEV-NIB₂-NIB₉ (see Table 1), with subscripts denoting the intergenic positions of *Nib*. Note that NIB₉ is referring to *Nib* at its original position. We also generated viruses in which *Nib* was moved to alternative positions and the original gene was deleted (Figure 1B). We refer to these viruses with a single *Nib* copy at alternative positions as TEV-NIB₁-ΔNIB₉ and TEV-NIB₂-ΔNIB₉ (Table 1). For all viruses generated, the termini of the relocated *Nib* were adjusted such that this protein is properly translated and sites for cleavage from the viral polyprotein are provided (Figure 1), similar to the original proteolytic cleavage sites at the corresponding positions. For these four key intermediate viruses along an evolutionary trajectory to altered gene order, we could then measure virulence, viral accumulation, and competitive fitness and study their evolutionary potential.

Viruses with a duplication of *Nib* have reduced fitness and accumulation

TEV-NIB₁-NIB₉ and TEV-NIB₂-NIB₉ were reconstituted from infectious clones and inoculated in *N. tabacum* L. cv. Xanthinc (NN) plants. The collected tissue from these plants served as the starting material for all succeeding experiments. Subsequently, we measured within-host competitive fitness (*W*) by individually competing the two viruses with *Nib* duplications against the wild-type virus carrying a GFP marker (TEV-eGFP) and viral accumulation by measuring the number of virions (genome equivalents) present in the host plant in the absence of a competing virus. After normalizing the number of virions for each viral genotype to the same concentration, both competition and accumulation experiments were performed for a total duration of 1 week. For both ancestral viruses TEV-NIB₁-NIB₉ and TEV-NIB₂-NIB₉, we observed statistically significant decreases—compared to the ancestral wild-type virus—in competitive fitness (Figure 3A; TEV-NIB₁-NIB₉: $t_4 = -6.379$, $P = 0.003$; TEV-NIB₂-NIB₉: $t_4 = -8.348$, $P = 0.001$) and accumulation (Figure 4A; TEV-NIB₁-NIB₉: $t_4 = -45.097$, $P < 0.001$; TEV-NIB₂-NIB₉: $t_4 = -8.650$, $P < 0.001$). The magnitude of change in competitive fitness for both genotypes is similar (~0.6), but the magnitude of change in accumulation is much larger for TEV-NIB₂-NIB₉ (0.054) than for TEV-NIB₁-NIB₉ (0.001). Compare the light gray bars labeled “ancestral” in panel A of both Figure 3 and Figure 4 to see how the duplicated viruses performed compared to the wild-type virus. Because of the increase in genome size, we were not surprised that duplication of *Nib* leads to a virus with reductions in these fitness components. Nevertheless, this observation has an important implication; the first step that must be taken along the plausible evolutionary route we have suggested is already unlikely. Viruses with *Nib* duplications cannot be maintained in virus populations for long periods of time. These viruses therefore must establish a bridgehead

Table 1 TEV genotypes constructed for this study

Viral genotype	Original <i>Nib</i> gene present?	Alternative position of <i>Nib</i> gene
TEV-NIB ₁ -ΔNIB ₉	No	1
TEV-NIB ₂ -ΔNIB ₉	No	2
TEV-NIB ₁ -NIB ₉	Yes	1
TEV-NIB ₂ -NIB ₉	Yes	2

in viral populations by means of genetic drift, from whence they can continue along the evolutionary trajectory to alternative gene order. Consequently, the low fitness of viruses with *Nib* duplications constitutes a second barrier to reordering.

An evolutionary cul-de-sac: after duplication, *Nib* is pervasively deleted from an alternative position

The next key stage in the evolution of alternative gene order is to determine the evolutionary potential of viruses with gene duplications TEV-NIB₁-NIB₉ and TEV-NIB₂-NIB₉. The genotypes containing two *Nib* copies (Figure 1C) were therefore evolved in *N. tabacum* plants for a total of 27 weeks using both nine 3-week passages and three 9-week passages. Our choice for passage duration was based on a previous study in which we showed that selection appears to act more strongly for longer-duration passages, while nonfunctional sequences are more stable during shorter passages (Zwart *et al.* 2014). Hence, these conditions may fulfill requirements for further evolution of duplicated viruses by (1) retaining the duplicated gene copy while (2) still allowing for selection—and not mainly genetic drift—to act on these virus populations. At the start of the evolution experiment, the TEV-NIB₁-NIB₉ and TEV-NIB₂-NIB₉ viruses had reduced infectivity and showed little or no symptoms, consisting of fewer than seven sparse small chlorotic spots per leaf. These symptoms are much weaker than typical symptoms observed for wild-type TEV, which consist of vein clearing, mosaic mottling, chlorosis, and stunting of the leaves together with reduced plant growth (Velasquez *et al.* 2014; Revers and García 2015). However, during the evolution experiment, these symptoms changed to wild-type-like symptoms, indicated by the green lines in Figure 5. In the first 3- and 9-week passages, the reduced symptoms turned into mild wild-type-like symptoms for the TEV-NIB₁-NIB₉ lineages (Figure 5). For the TEV-NIB₂-NIB₉ lineages, mild wild-type-like symptoms appeared later: for the 3-week lineages in passage 3 (9 weeks on the *x*-axis in Figure 5) and for the 9-week lineages in passage 2 (18 weeks on the *x*-axis in Figure 5). At the end of the evolution experiment (27 weeks), all lineages of both viruses showed wild-type-like symptoms in tobacco plants.

Through RT-PCR, deletions were detected in the second alternative copy of *Nib* but never in the original one. Genome size for all the evolved lineages was estimated at every passage (black symbols and continuous lines in Figure 5). As deletions occur, the viral genome reduces in size to one that is similar to that of wild-type TEV. An exponential decay model [$S(t) = S(0) + a \exp(-bt)$] was fitted to the estimated genome size *S* over time *t* for every lineage (red line in Figure 5

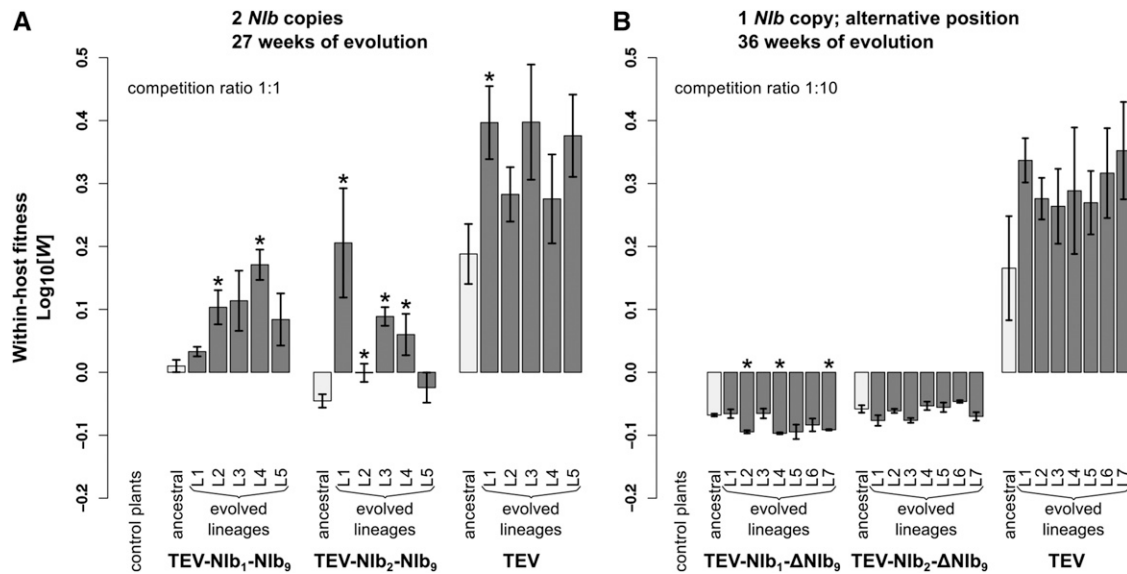


Figure 3 Within-host competitive fitness of the evolved and ancestral lineages. Fitness W , as determined by competition experiments and RT-qPCR, of the different viral genotypes with a duplication of *Nib* (A) and with *Nib* moved to an alternative position (B) with wild-type TEV and healthy control plants with respect to a common competitor, TEV-eGFP. In A, the competition experiment was started with an inoculum containing a 1:1 mixture of both competitors, while a 1:10 mixture (common competitor and virus of interest) was used in B. The ancestral lineages are indicated with light-gray bars and the evolved lineages with dark-gray bars. The viruses in A were evolved using five replicate lineages each (L1–L5) for a total of 27 weeks, and viruses in B were evolved using seven replicate lineages each (L1–L7) for a total of 36 weeks. Evolved lineages that tested significantly different compared to their ancestral lineage are indicated with an asterisk (*t*-test with Holm-Bonferroni correction for multiple tests).

and Table S1), and the rates of size change b were compared by means of a generalized linear model (GLM) using a gamma probability distribution. Whereas *PASSAGE DURATION* did not have a significant effect on genome size, both virus *GENOTYPE* and the interaction *GENOTYPE* \times *PASSAGE DURATION* were significant (Table 2). The alternative *Nib* copy was deleted more quickly in the TEV-*Nib*₁-*Nib*₉ lineages than in the TEV-*Nib*₂-*Nib*₉ lineages (Figure 5; compare A and B to C and D), whereas the interaction term suggests that the effect of genotype is particularly strong for the 3-week passages (Table 2). Additionally, in the TEV-*Nib*₂-*Nib*₉ lineages, there appears to be more variation in the time points at which the second *Nib* copy was deleted in the evolution experiment (Figure 5, C and D). The alternative *Nib* copy from TEV-*Nib*₂-*Nib*₉ is therefore more stable, suggesting that the second position in the TEV genome is a more accessible trajectory for duplication of a gene and subsequent reorganization within the genome. The decrease in genome size also appears to correlate with the appearance of stronger symptoms (Figure 5). At the end of the evolution experiment, all lineages had a genome size that was very similar to that of the wild-type virus. All TEV-*Nib*₁-*Nib*₉ 3-week lineages and four of five 9-week lineages even evolved to a genome size smaller than the wild-type virus by deleting a part of the 5' UTR. For TEV-*Nib*₂-*Nib*₉, four of five 3-week lineages and three of five 9-week lineages evolved to a smaller genome size by deleting part of the *HC-Pro* cysteine protease.

For all the evolved lineages, we then measured within-host competitive fitness and accumulation (Figure 3A and Figure 4A).

Compare bars labeled “ancestral” and “evolved lineages” in Figure 3A. When pairwise comparisons were made between the ancestral virus and the evolved lineages (*t*-test with Holm-Bonferroni correction), significant increases in within-host competitive fitness were found for two of five TEV-*Nib*₁-*Nib*₉ lineages and for four of five TEV-*Nib*₂-*Nib*₉ lineages (Figure 3A and Table 3). However, fitness never reached the fitness of the wild-type virus, and therefore, we did find a significant effect of treatment [ANOVA with *post hoc* Tukey honestly significant difference (HSD)] comparing the evolved TEV-*Nib*₁-*Nib*₉ and TEV-*Nib*₂-*Nib*₉ lineages to the evolved TEV lineages (Table 4). Now compare bars labeled “ancestral” and “evolved lineages” in Figure 4A. Whereas no significant increases in accumulation were found comparing the ancestral virus and the evolved lineages of the wild-type TEV, we did find significant increases in viral accumulation for all the evolved lineages of both TEV-*Nib*₁-*Nib*₉ and TEV-*Nib*₂-*Nib*₉. And these lineages reached similar accumulation levels as the evolved wild-type lineages. However, we did find a significant effect of treatment on accumulation comparing the evolved lineages of the viral genotypes owing to the differences between the evolved TEV-*Nib*₁-*Nib*₉ and the wild-type TEV lineages (Table 4).

The evolution of viruses with two *Nib* copies results in an increase in fitness related to the reduction in genome size. However, we consistently observed the deletion of *Nib* at an alternative position, leading back to the ancestral wild-type virus, and we never observed the deletion of *Nib* at its ancestral position. This evolutionary cul-de-sac therefore represents a third barrier to reordering.

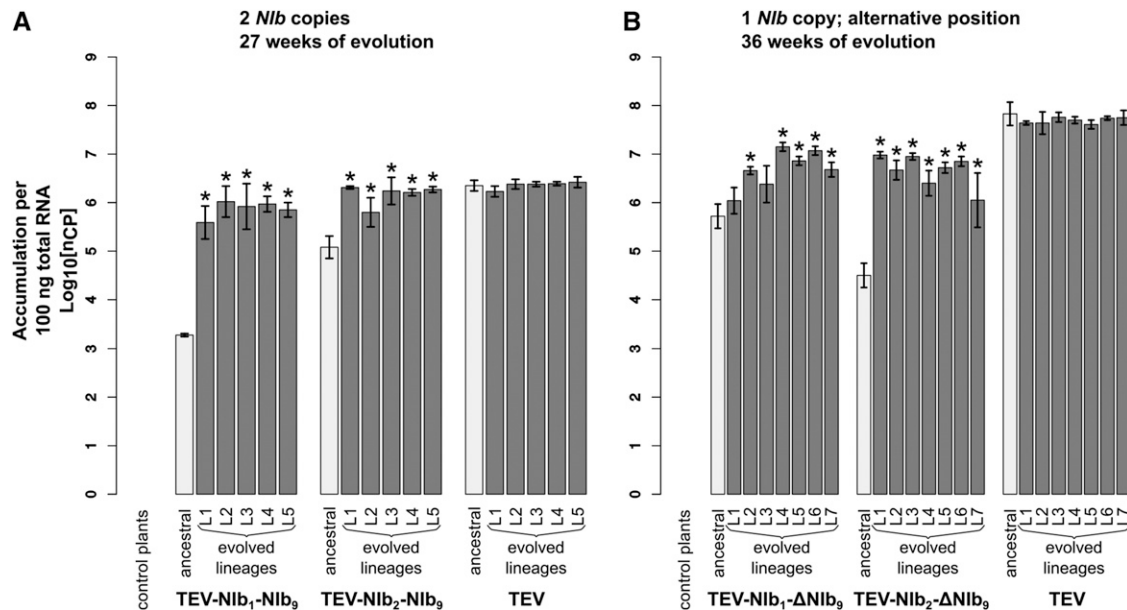


Figure 4 Virus accumulation of the evolved and ancestral lineages. Virus accumulation, as measured by RT-qPCR, of the different viral genotypes with a duplication of *Nib* (A) and with *Nib* moved to an alternative position (B) with the wild-type TEV and healthy control plants. The ancestral lineages are indicated with light-gray bars and the evolved lineages with dark-gray bars. The viruses in A were evolved using five replicate lineages each (L1–L5) for a total of 27 weeks, and the viruses in B were evolved using seven replicate lineages each (L1–L7) for a total of 36 weeks. Evolved lineages that tested significantly different compared to their ancestral lineage are indicated with an asterisk (*t*-test with Holm-Bonferroni correction for multiple tests).

Whole-genome sequences of evolved lineages of viruses with a *Nib* duplication

All evolved and ancestral lineages described in this study have been fully sequenced using the Illumina technology. The sequences of the ancestral lineages were used as an initial reference for the evolved lineages. Furthermore, for the genotypes that originally had two *Nib* copies, parts of the genome were sequenced by Sanger to determine the exact deletion sites, which have been detected previously by RT-PCR (Figure 5). Most deletion variants were used to construct new reference sequences for each of the evolved TEV-Nib₁-NIB₉ and TEV-Nib₂-NIB₉ lineages. After the initial mapping step, mutations were detected in the evolved lineages compared to their corresponding ancestor.

At the sequence level, the main changes in TEV-Nib₁-NIB₉ and TEV-Nib₂-NIB₉ were large genomic deletions in the second *Nib* copy. In other words, for both viruses, we consistently observed pseudogenization of the duplicated *Nib* copy, in accordance with RT-PCR results. In TEV-Nib₁-NIB₉, none of these deletions included the N-terminus of P1, while for nine of ten lineages these deletions included the 3' end of the 5' UTR (Figure 6). Only in half of the TEV-Nib₁-NIB₉ lineages is the reading frame maintained after pseudogenization. Because the deletion occurs at the start of the genome, it is not necessary to maintain the reading frame as long as the original second methionine in P1 is preserved. In TEV-Nib₂-NIB₉, none of the deletions included the C-terminus of P1, but for seven of ten lineages the deletions included the N-terminal region of the HC-Pro cysteine protease (Figure 6), similar to results obtained by previous studies (Dolja *et al.* 1993;

Zwart *et al.* 2014). HC-Pro cysteine protease is a multifunctional protein, and the N-terminal region of HC-Pro cysteine protease is implicated in transmission by aphids (Thornbury *et al.* 1990; Atreya *et al.* 1992) and is not essential for replication and movement (Dolja *et al.* 1993; Cronin *et al.* 1995). In the other genotype, TEV-Nib₂-NIB₉, the reading frame was maintained in the sequence of all lineages after pseudogenization. This could be explained simply by the fact that these lineages depend on only one methionine codon at the beginning of the polyprotein, at the start of the coding region in P1.

However, for these lineages, little evidence for adaptive evolution was found at the level of single-nucleotide mutations (Figure 6, Table S2, Table S3, and Table S4). We determined mutations present in the evolved lineages with respect to their corresponding ancestral sequences. In the TEV-Nib₁-NIB₉ 3-week lineages, we detected three high-frequency (>10%) convergent nonsynonymous mutations occurring in two of five lineages located in the pseudogenized alternative *Nib* copy (A1643U) at the 3' end of *CI* (U7066C) and in *VPg* (U7703A). For the TEV-Nib₁-NIB₉ 9-week lineages, we observed one convergent nonsynonymous mutation that occurred in two of five lineages in *NiA-Pro* (A8347G). The same mutation also was found in TEV wild-type 3- and 9-week lineages. For the TEV-Nib₂-NIB₉ 3-week lineages, no repeated mutations were found, and in the 9-week lineages, one convergent synonymous mutation was found in *CI* (C6351U) for two of five lineages. For more information of the mutations found in both genotypes, see Table S2. After remapping the cleaned reads against a new defined consensus sequence for each lineage, we looked at the variation

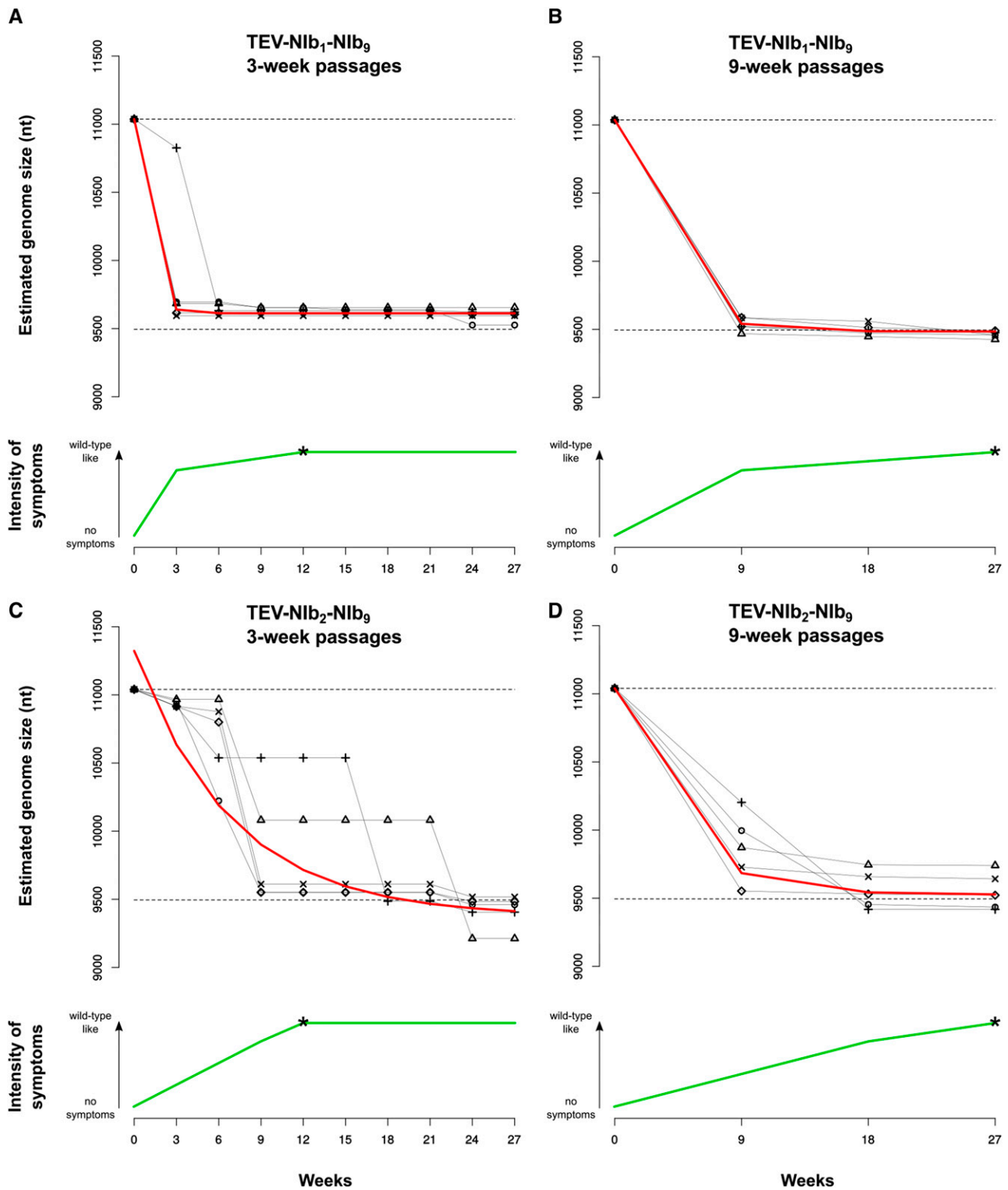


Figure 5 Symptomatology and genome size over time of genotypes with a duplication of *Nib*. The different panels display how the genome size and symptomatology of the TEV-*Nib*₁-*Nib*₉ (A and B) and TEV-*Nib*₂-*Nib*₉ (C and D) genotypes change along the evolution experiments performing 3-week (A and C) and 9-week (B and D) serial passages. The green lines in each panel plot the observed severity of symptoms. The arrows on the left sides of the plots indicate the gradient from no symptoms to wild-type-like symptoms. The asterisk on each green line indicates the passage where we first observed wild-type-like symptoms (as described in the text). Note that for all genotypes we observed few or no symptoms at the beginning and wild-type-like symptoms at the end of the evolution experiment. The black symbols connected by continuous black lines represent the estimated genome size of the different viral lineages along the evolution experiment. The upper dashed lines represent the genome size of the ancestral viruses, and the lower dashed lines represent the genome size of the wild-type virus. The red line plots the median estimates of a nonlinear regression model. Note that the reduction in genome size appears to be negatively correlated with the severity of symptoms developed.

Table 2 GLM analysis of the genome size data

Term	LRT	df	P
Intercept	28.420	1	<0.001
GENOTYPE	10.280	1	0.001
PASSAGE DURATION	0.005	1	0.943
GENOTYPE × PASSAGE DURATION	4.707	1	0.030

within each lineage. SNPs were detected at a frequency as low as 1%. In the evolved TEV-Nib₁-Nib₉ lineages, a total of 301 SNPs were detected, with a median of 36 (27–45) per lineage. In the evolved TEV-Nib₂-Nib₉ lineages, a total of 220 SNPs were detected, with a median of 23.5 (4–44) per lineage. In both virus genotypes, most of the SNPs were present at low frequency, with a higher percentage of synonymous (TEV-Nib₁-Nib₉, 66.4%; TEV-Nib₂-Nib₉, 64.5%) vs. nonsynonymous changes (Figure S1). However, the difference in the distribution of synonymous vs. nonsynonymous SNP frequency was not significant (Kolmogorov-Smirnov test; TEV-Nib₁-Nib₉: $D = 0.146$, $P = 0.073$; TEV-Nib₂-Nib₉: $D = 0.144$, $P = 0.217$). For more details on the frequency of the SNPs within every lineage, see Table S3 and Table S4.

The results from the whole-genome data are congruent with the RT-PCR results and phenotypic assays: the main change in the evolved lineages of the viruses with *Nib* duplications is the deletion of the second copy, which turns into a virus that is in all respects similar to the ancestral wild-type virus. Although there are some convergent single-nucleotide mutations, these occur only in a small fraction of lineages and, moreover, are often shared between the TEV, TEV-Nib₁-Nib₉, and TEV-Nib₂-Nib₉ lineages. These mutations therefore appear to represent general adaptations, without a strong link to the transient presence of the second *Nib* copy.

Viruses with *Nib* moved to an alternative position have further reductions in fitness and viral accumulation

The viruses for which *Nib* was moved to an alternative position without conserving the original *Nib* copy, TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉, were reconstituted from infectious clones. Subsequently, we measured their within-host competitive fitness and viral accumulation. Compare bars labeled “ancestral” in both Figure 3B and Figure 4B. For both TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉, we observed significant decreases—compared to the wild-type virus—in competitive fitness (Figure 3B; TEV-Nib₁-ΔNib₉: $t_4 = -4.897$, $P = 0.008$; TEV-Nib₂-ΔNib₉: $t_4 = -4.692$, $P = 0.009$) and accumulation (Figure 4B; TEV-Nib₁-ΔNib₉: $t_4 = -10.463$, $P < 0.001$; TEV-Nib₂-ΔNib₉: $t_4 = -16.453$, $P < 0.001$). Deletion of the original *Nib* copy therefore leads to further reductions in viral fitness, suggesting that TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ will be rapidly outcompeted by both their direct ancestors (TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉, respectively) and viruses with deletions in the new copy of *Nib* (fitness similar to wild-type virus). Therefore, we are confronted with a fourth evolutionary barrier: viruses with a single copy of *Nib* in an alternative position cannot outcompete the duplicated

virus, meaning that they must be maintained, or probably fixed, by genetic drift to have the opportunity to undergo further evolution.

Limited short-term evolutionary potential of viruses with *Nib* moved to an alternative position

In the evolutionary trajectory we have postulated, the final step is the evolution of a virus with *Nib* only in an alternative position. If these viruses managed to occur through a series of chance events and could exist in isolation from their ancestral viruses for a period of time, what would their evolutionary fate be? Could these viral populations readily converge on a fitness peak that allowed them to be comparable or superior to the ancestral TEV?

To address these questions, TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ were evolved in tobacco plants for a total of 36 weeks using four 9-week serial passages. We did not perform 3-week passages because we expected these genomes to be stable and therefore considered only a condition with maximal selection and minimal drift by intermittent bottlenecks. Both reordered viruses have very low infectivity, and serial passage was completed successfully for only seven of ten lineages. For both TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉, tobacco plants had very weak or no symptoms of infection, and these symptoms did not become more severe over time. Therefore, infection had to be confirmed by RT-PCR. Two different regions of the reordered genomes were amplified using primers flanking the region of the new *Nib* position as well as primers flanking the original position of the removed *Nib*. No evidence was found of any insertions or deletions within these amplified sites, indicating that reordered viruses were stable over time and that there was no presence of wild-type-like viruses.

For all the evolved lineages, we measured within-host competitive fitness and accumulation (Figure 3B and Figure 4B). Compare bars labeled “ancestral” and “evolved lineages” in Figure 3B. Pairwise comparisons between the ancestral and evolved lineages showed there were no significant increases in within-host competitive fitness (Figure 3B and Table 3), while in three of five TEV-Nib₁-ΔNib₉ lineages we even found significant decreases. Now compare bars labeled “ancestral” and “evolved lineages” in Figure 4B. Accumulation levels of the wild-type virus did not change significantly compared to the ancestral TEV, while for five of seven of the evolved TEV-Nib₁-ΔNib₉ lineages and all the TEV-Nib₂-ΔNib₉ lineages, accumulation increased significantly (Figure 4B and Table 3). However, these accumulation levels never reached the same levels as the wild-type virus. Comparing only the evolved lineages, there was a significant effect of treatment (ANOVA with *post hoc* Tukey HSD; Table 4) on viral accumulation and within-host competitive fitness, indicating that the wild-type TEV outperforms the two reordered viruses for both fitness components.

Whole-genome sequences of evolved lineages of viruses with a single *Nib* copy at an alternative position

We found evidence of adaptive convergent evolution comparing the evolved and ancestral lineages containing one

Table 3 Two-sample t-test (4 df) with Holm-Bonferroni correction comparing ancestral against evolved lineages

	Genotype	Lineage	Accumulation		Within-host fitness		
			<i>t</i>	<i>P</i>	<i>t</i>	<i>P</i>	
2 <i>Nib</i> copies	TEV- <i>Nib</i> ₁ - <i>Nib</i> ₉	9WL1	-11.744	<0.001 ^a	-3.101	0.036	
		9WL2	-14.697	<0.001 ^a	-5.610	0.005 ^a	
		9WL3	-9.690	0.001 ^a	-3.663	0.022	
		9WL4	-27.962	<0.001 ^a	-10.713	<0.001 ^a	
		9WL5	-29.966	<0.001 ^a	-2.991	0.040	
	TEV- <i>Nib</i> ₂ - <i>Nib</i> ₉	9WL1	-9.305	0.001 ^a	-4.849	0.003 ^a	
		9WL2	-3.284	0.030 ^a	-4.358	0.012 ^a	
		9WL3	-5.600	0.005 ^a	-12.814	<0.001 ^a	
		9WL4	-8.222	0.001 ^a	-5.348	0.006 ^a	
		9WL5	-8.832	0.001 ^a	-1.405	0.233	
	TEV	9WL1	1.3475	0.249	-4.836	0.008 ^a	
		9WL2	-0.2692	0.801	-2.549	0.063	
		9WL3	-0.4693	0.663	-3.542	0.024	
		9WL4	-0.5268	0.626	-1.782	0.149	
		9WL5	-0.7859	0.476	-4.028	0.016	
	1 <i>Nib</i> copy, alternative position	TEV- <i>Nib</i> ₁ -Δ <i>Nib</i> ₉	9WL1	-1.521	0.203	-0.395	0.713
			9WL2	-6.133	0.004 ^a	15.041	<0.001 ^a
			9WL3	-2.534	0.064	-0.515	0.634
			9WL4	-9.322	0.001 ^a	19.454	<0.001 ^a
9WL5			-7.363	0.002 ^a	3.986	0.016	
9WL6			-8.738	0.001 ^a	2.652	0.057	
9WL7			-5.670	0.005 ^a	17.220	<0.001 ^a	
TEV- <i>Nib</i> ₂ -Δ <i>Nib</i> ₉		9WL1	-16.224	<0.001 ^a	3.139	0.035	
		9WL2	-11.531	<0.001 ^a	0.805	0.466	
		9WL3	-16.134	<0.001 ^a	4.564	0.010	
		9WL4	-9.068	0.001 ^a	-0.989	0.379	
		9WL5	-13.961	<0.001 ^a	-0.502	0.642	
		9WL6	-14.943	<0.001 ^a	-3.544	0.024	
		9WL7	-4.320	0.012 ^a	2.374	0.077	
TEV		9WL1	1.377	0.241	-3.311	0.030	
		9WL2	0.986	0.380	-2.152	0.098	
		9WL3	0.476	0.659	-1.674	0.170	
		9WL4	0.908	0.415	-1.642	0.176	
		9WL5	1.464	0.217	-1.867	0.135	
		9WL6	0.621	0.569	-2.388	0.075	
		9WL7	0.497	0.646	-2.865	0.046	

^a Significant after Holm-Bonferroni correction.

reordered *Nib* copy (Figure 7, Table S5, Table S6, and Table S7). Mutations in the TEV-*Nib*₁-Δ*Nib*₉ lineages were found in (1) the reordered *Nib* gene at the first position (U428C), (2) in *P1* around the proteolytic cleavage site of *Nib* and *P1* (U1688C, U1697C, and U1697A), and (3) in *Nia-Pro* (U8210C and A8347G). Mutation U1688C modifies the start codon of *P1* (M563T). This is explained by the introduction of an additional start codon at the first position of the reordered genome that makes the original methionine redundant. Mutations in the TEV-*Nib*₂-Δ*Nib*₉ lineages were found (1) in the reordered *Nib* gene at the second position (G1066A, G1090A, G1264A, and U1346C), (2) in *HC-Pro* (G3213U, A3632G, and U3803C), (3) in *P3* (A4016G), and (4) in *Nia-Pro* (U8285C and A8350G). The former mutation in *Nia-Pro* (U8285C) also was found in one lineage of TEV-*Nib*₁-Δ*Nib*₉, and the latter mutation (A8350G) also was found in the evolved lineages of TEV-*Nib*₁-Δ*Nib*₉ and the wild-type TEV. Not a single mutation was detected in *VPg*, which is putatively involved in translation and replication.

For more information of the mutations found in both genotypes, see Table S5.

As for the within-population variation, in the evolved TEV-*Nib*₁-Δ*Nib*₉ lineages, we detected 137 SNPs, with a median of 22 (7–37) per lineage. In the evolved TEV-*Nib*₂-Δ*Nib*₉ lineages, 155 SNPs were detected, with a median of 22 (18–48) per lineage. In both virus genotypes, most of the SNPs were low-frequency SNPs, with a higher percentage of synonymous changes (58.4%) in the TEV-*Nib*₁-Δ*Nib*₉ lineages, while in the TEV-*Nib*₂-Δ*Nib*₉ lineages, the percentage of nonsynonymous changes was higher (63.9%) (see Figure S2). However, no significant difference was found in the distribution of synonymous vs. nonsynonymous SNP frequencies (Kolmogorov-Smirnov test; TEV-*Nib*₁-Δ*Nib*₉: *D* = 0.203, *P* = 0.078; TEV-*Nib*₂-Δ*Nib*₉: *D* = 0.152, *P* = 0.318). For more information on the frequency of the SNPs within each lineage, see Table S6 and Table S7.

The mutations appeared to be contingent on the ancestral genotype; most of the convergent mutations that were found

Table 4 Nested ANOVAs on accumulation and within-host fitness of evolved lineages with a *post hoc* Tukey HSD test

	Trait	Source of variation	df	SS	F	P	Treatments compared	P
2 <i>Nib</i> copies	Accumulation	Treatment	2	1.818	19.476	<0.001	TEV-Nib ₁ -Nib ₉ TEV-Nib ₂ -Nib ₉	0.002
		Lineage within treatment	12	0.922	1.646	0.131	TEV-Nib ₁ -Nib ₉ TEV	<0.001
		Error	30	1.400			TEV-Nib ₂ -Nib ₉ TEV	0.053
	Within-host fitness	Treatment	2	0.665	117.296	<0.001	TEV-Nib ₁ -Nib ₉ TEV-Nib ₂ -Nib ₉	0.593
		Lineage within treatment	12	0.208	6.124	<0.001	TEV-Nib ₁ -Nib ₉ TEV	<0.001
		Error	32	0.091			TEV-Nib ₂ -Nib ₉ TEV	<0.001
1 <i>Nib</i> copy, alternative position	Accumulation	Treatment	2	14.451	18.236	<0.001	TEV-Nib ₁ -ΔNib ₉ TEV-Nib ₂ -ΔNib ₉	0.832
		Lineage within treatment	18	4.778	6.699	<0.001	TEV-Nib ₁ -ΔNib ₉ TEV	<0.001
		Error	42	1.664			TEV-Nib ₂ -ΔNib ₉ TEV	<0.001
	Within-host fitness	Treatment	2	1.966	686.952	<0.001	TEV-Nib ₁ -ΔNib ₉ TEV-Nib ₂ -ΔNib ₉	0.161
		Lineage within treatment	18	0.028	1.078	0.405	TEV-Nib ₁ -ΔNib ₉ TEV	<0.001
		Error	42	0.060			TEV-Nib ₂ -ΔNib ₉ TEV	<0.001

in the lineages of one reordered genotype were not found in the other genotype. All convergent mutations found were nonsynonymous. Note that the convergent mutations are present at different frequencies and that none of these convergent mutations were fixed in all the replicate lineages (Table S5).

For TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉, we therefore see markedly different patterns of genome evolution than for TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉. In agreement with RT-PCR results, TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ show no signs of major genomic rearrangements, such as the large deletions seen in TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉. Given the strong selection, which undoubtedly would occur for a variant with wild-type gene order, these results suggest that the mutation supply and nonviability of viruses without *Nib* are major limiting factors. This result provides support for our conjecture that gene-order evolution involving essential genes must occur through gene duplications in potyviruses. However, the observed convergent single-nucleotide mutations are congruent with the observed improvement in virus accumulation, suggesting that selection is acting on these virus populations.

Studying the evolutionary potential of the reordered viruses TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉, although demonstrating clear signatures of adaptive evolution, therefore also shows further barriers to the evolution of alternative gene orders. First, increased accumulation and sequence-level convergent evolution illustrate that adaptive evolution occurred, meaning that the reordered viruses, despite initially having low levels of accumulation, are still somewhat evolvable and show marked improvement in key fitness components. Conversely, accumulation was still significantly lower than that of the wild-type virus. Furthermore, within-host competitive fitness remained similar to that of the ancestral TEV-Nib₁-ΔNib₉ or TEV-Nib₂-ΔNib₉ or, strikingly, even decreased in some lineages. These observations suggest that even after 36 weeks of evolution under conditions that optimize selection, these reordered viruses remain grossly inferior competitors to the wild-type virus. Therefore, even if a virus population should overcome all four of these barriers, a

final evolutionary barrier to the reordering of potyvirus genomes remains.

Discussion

In this study, we have explored whether the most plausible evolutionary trajectory for the rearrangement of gene order in a positive-strand RNA potyvirus is accessible. Overall, we have identified five barriers to the evolution of viruses with the essential *Nib* replicase gene moved to an alternative position. First, only two of nine viruses with *Nib* moved to an alternative position are viable. Second, the fitness of viruses with *Nib* duplications was low, meaning that such viruses would be quickly displaced from populations if they arose (Figure 8 shows a summary of fitness data). Third, for viruses with gene duplications, the new *Nib* copy was lost in all lineages, whereas the original copy was maintained. This propensity represents a second cul-de-sac because loss of the new *Nib* copy entails a trajectory that leads back to the ancestral gene order. Fourth, viruses with only a single *Nib* copy at an alternative position had low fitness and accumulation, notably lower than viruses with duplications (Figure 8). Therefore, to reach a virus with a single *Nib* copy in an alternative position, two rare recombination events must occur within a small time window because the intermediate step is unstable. Moreover, the low-fitness recombinants would need to be maintained or fixed by genetic drift because they would be outcompeted otherwise. Fifth, even if this unlikely sequence of events occurs and the resulting virus becomes reproductively isolated, after more than half a year of evolution under optimal conditions, this virus still would not stand a chance in head-to-head competition with its ancestor (Figure 8). We therefore conclude that—under the conditions we have considered—the evolution of alternative gene orders for TEV is highly unlikely because the evolutionary trajectory to alternative gene order we have studied is not accessible. These results suggest that one reason gene order has been conserved in potyviruses is therefore the lack of accessible trajectories to alternative gene orders.

The observation that the evolved virus lineages with an alternative gene order have improved accumulation while

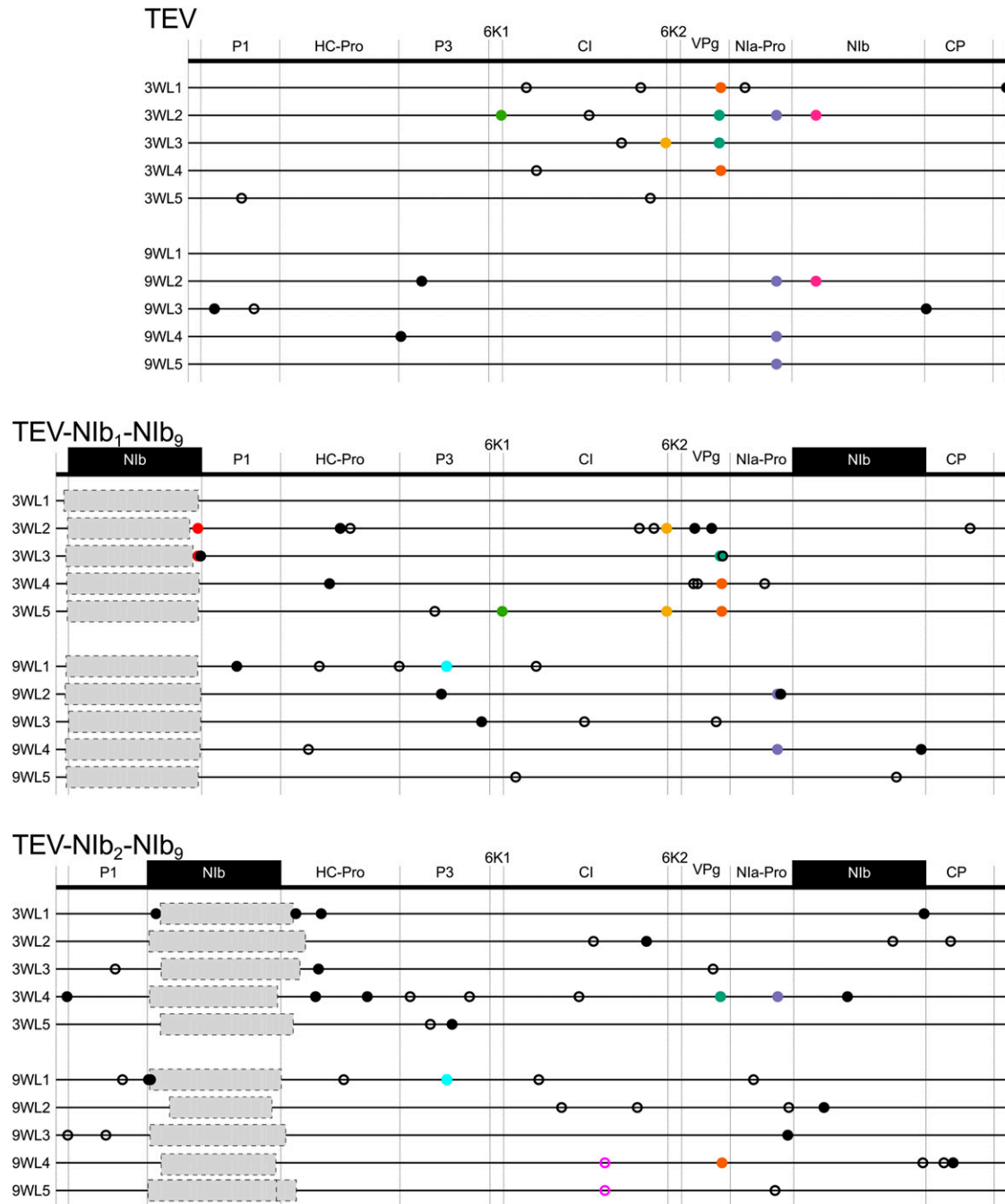


Figure 6 Genomes of the evolved lineages with a duplication of *Nib*. Mutations were detected using NGS data of the evolved TEV-*Nib*₁-*Nib*₉ and TEV-*Nib*₂-*Nib*₉ lineages compared to their ancestral lineages. The wild-type TEV is given for comparative purposes. The names on the left identify lineages (e.g., 3WL1 is the final population of 3-week passages, lineage 1). No NGS data were available for TEV-*Nib*₁-*Nib*₉ 3WL1. Filled circles and open circles represent nonsynonymous and synonymous substitutions, respectively. Black substitutions occur in only one lineage, whereas color-coded substitutions are repeated in two or more lineages. Note that the mutations are present at different frequencies, as reported by SAMtools (>10%). Gray boxes indicate genomic deletions in the majority variant, as detected by Sanger sequencing technology.

having unchanged or deteriorated within-host fitness is noteworthy. Our serial passage experiment was conducted in single-host plants, and we would therefore expect within-host competitive fitness to improve. In other words, the virus variant that is present at highest frequency in the final population has the highest probability of carrying over to the next round of infection, irrespective of the level of accumulation.

We hypothesize that unchanged or lowered competitive fitness is probably due to low infection levels during serial passaging of TEV-*Nib*₁- Δ *Nib*₉ and TEV-*Nib*₂- Δ *Nib*₉. These low infection levels would result in limited direct competition for space within the host, with adaptation occurring by means of more prudent use of cellular resources and improved repression of host immune responses, for example. Evolution during

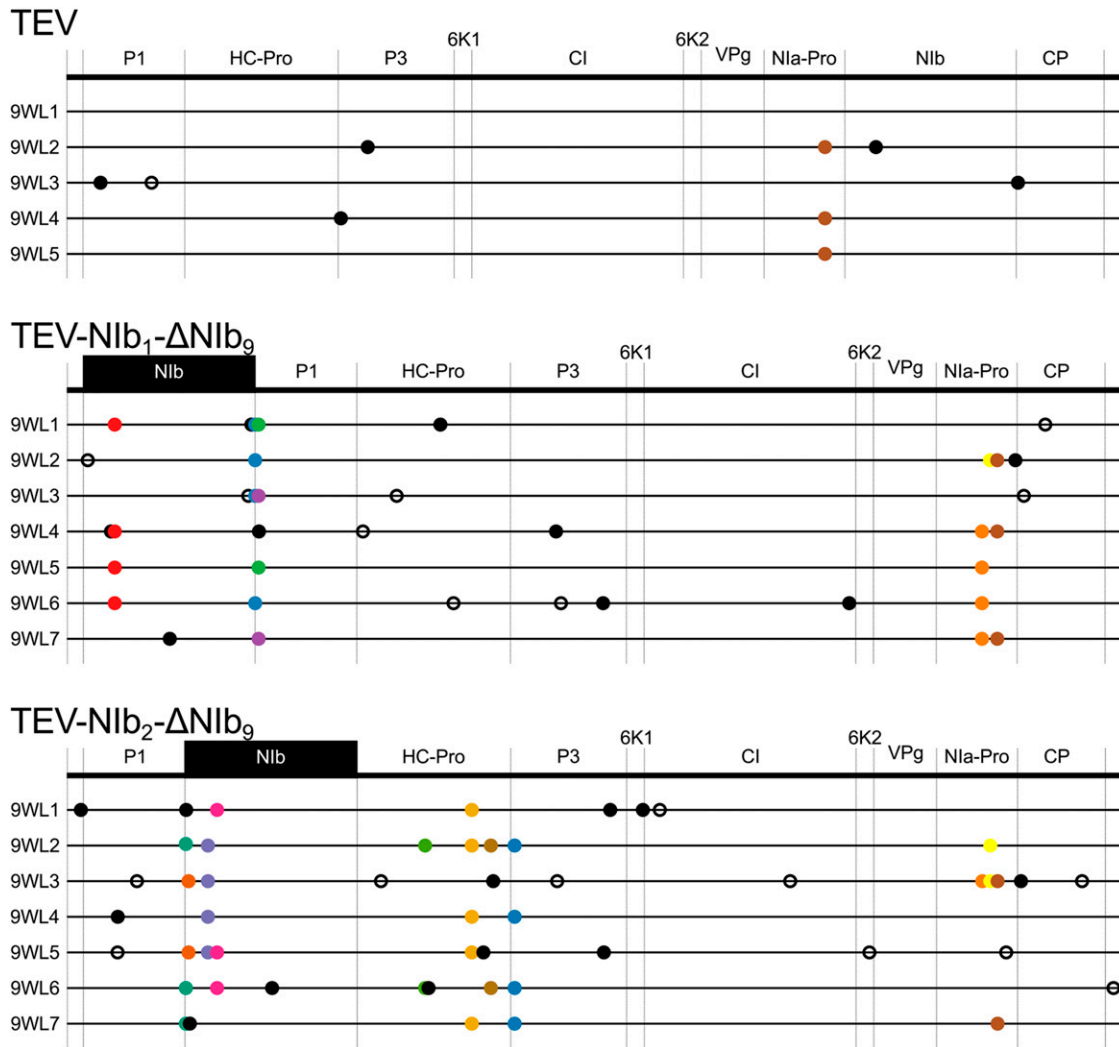


Figure 7 Genomes of the evolved lineages with *Nib* moved to an alternative position. Mutations were detected using NGS data of the evolved TEV-*Nib*₁- Δ *Nib*₉ and TEV-*Nib*₂- Δ *Nib*₉ lineages compared to their ancestral lineages. The wild-type TEV is given for comparative purposes. The names on the left identify lineages (e.g., 9WL1 is the final population of 9-week passages, lineage 1). Filled circles and open circles represent nonsynonymous and synonymous substitutions, respectively. Black substitutions occur in only one lineage, whereas color-coded substitutions are repeated in two or more lineages. Note that the mutations are present at different frequencies, as reported by SAMtools (>10%).

low-level infections therefore may not improve performance in direct competition and may even lower it as a result of antagonistic pleiotropy, given that accumulation and competitive fitness do not correlate for TEV (Zwart *et al.* 2014). Since the common competitor used in the fitness assays is a strain derived from TEV, high-level infections will occur in these assays. For these reasons, fitness improvements in TEV-*Nib*₁- Δ *Nib*₉ and TEV-*Nib*₂- Δ *Nib*₉ lineages may very well not be conspicuous in direct competitions with the ancestral TEV and can only be detected systemically by measuring virus accumulation.

The first step in our proposed model, gene duplication, is a rare event in the recent history of RNA viruses (Simon-Loriere and Holmes 2013). The few cases describing gene duplication in RNA viruses seem to occur through either homologous or nonhomologous recombination (Forss and Schaller 1982; Tristem *et al.* 1990; Boyko *et al.* 1992; Walker *et al.* 1992; Wang and Walker 1993; Karasev *et al.* 1995; LaPierre *et al.*

1999; Peng *et al.* 2001; Valli *et al.* 2007; Simon-Loriere and Holmes 2013). In our experiments, the gene duplications are engineered artificially, while loss of the duplicated gene generates shorter genomes that are efficiently selected for, probably by means of slippage of the RdRp. We also observed deletions in the N-terminal region of the HC-Pro cysteine protease because there is no selection to maintain the intact HC-Pro. However, these specific deletions probably would not perform well in nature because the resulting truncated protein would preclude aphid-mediated transmission.

We speculate that the rearrangement of genes within the TEV genome could lead to differences in efficiency of cleavage at the proteolytic sites. Given the lower conservation of gene order and sequence homology at the 5' end in potyviruses, repositioning of genes to this side of the genome would least disturb polyprotein processing. Therefore, it is not surprising that the two positions available for relocation of the *Nib*

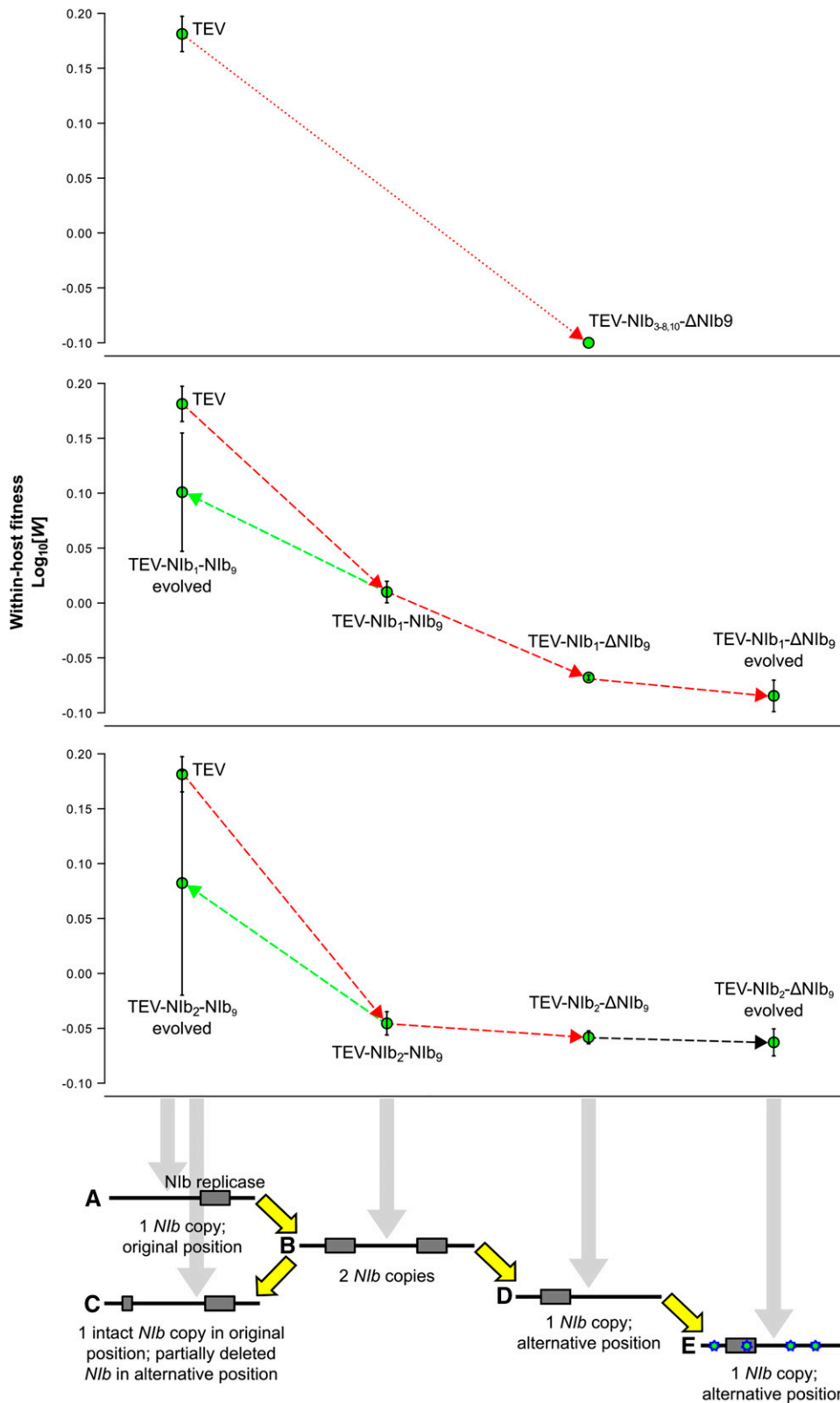


Figure 8 Within-host competitive fitness of all trajectories explored. For all the trajectories proposed (x-axis), the fitness (W at the y-axis) was plotted. The viruses with the *Nib* gene moved to intergenic positions 3–8 and 10 are non-viable genotypes (upper panel; $W < 0$). The trajectories of viruses with *Nib* moved to intergenic positions 1 and 2 are indicated in the middle and lower panels, respectively. Red lines indicate a decrease in fitness, black lines indicate no significant change in fitness, and green lines indicate an increase in fitness. Note how the ancestral and evolved TEV-Nib₁- Δ Nib₉ and TEV-Nib₂- Δ Nib₉ lineages have an inferior within-host competitive fitness compared to their ancestral TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ viruses (respectively) and the wild-type TEV.

replicase gene are located at the N-terminus, before and between the *P1* and *HC-Pro* genes. Within the Potyviridae family, both *P1* and the multifunctional *HC-Pro* have the lowest sequence conservation (Adams *et al.* 2005; Revers and García

2015). *P1* is not an essential gene (Verchot and Carrington 1995a), but the cleavage that separates *P1* and *HC-Pro* is required (Verchot and Carrington 1995b). The introduction of *Nib* before or between *P1* and *HC-Pro* could result in a delay in

this separation or even make *P1* inactive, which, in turn, would result in a low accumulation of the virus (Verchot and Carrington 1995a). We do observe that the viral genomes with an alternative gene order have low accumulation levels, but the accumulation levels are able to improve over time with the same gene order. This suggests that the adaptive mutations alone are responsible for the increase in accumulation levels. Nevertheless, these adaptive mutations could not compensate for the rearrangements done because within-host competitive fitness remains low throughout evolutionary time.

It can be postulated that the early expression of the *Nib* gene would have advantages for viral replication. However, the position of *Nib* in potyviruses' polyprotein seems to be optimized for its interaction with VPg for initiation of replication. Furthermore, it is thought that the interaction with both VPg and *Nia-Pro* targets *Nib* to the membranous structures where viral RNA replication takes place (Duffresne *et al.* 2008). Thus, the movement of *Nib* away from these interacting proteins might result in a delay in the interaction, replication, and gene expression. This is a difficult, maybe even impossible, burden to overcome for a virus.

The relocation of *Nib* in our model system did lead to the evolution of adaptive or compensatory mutations that allowed for an increase in viral accumulation. Interestingly, convergent mutations occur in both the relocated *Nib* and two regions that are responsible for proteolytic activity: the C-terminal region of *HC-Pro* and the main viral protease *Nia-Pro*. When relocating the *Nib* replicase to the first position in the genome, and thereby introducing a second methionine codon before *Nib* in the viral genome, we observe a mutation that changes the original methionine (in P1) to a threonine. Therefore, a reversion of *Nib* to its original position is not a subsequent step we would expect to see in this viral genotype.

Comparing our study to the experimental evolution done on the repositioning of the T7 RNA polymerase (Springman *et al.* 2005), similar results were obtained despite T7 having a different genome composition, architecture, and replication and gene expression strategies. In T7 studies, fitness is measured as the rate of population growth, which is comparable to our accumulation measurements. As in our study, the fitness of the rearranged phages improved but never reached the wild-type level. Additionally, as in our study, the reordered T7 genomes are stable over a long period of time. Concerning these similarities, the lack of accessible evolutionary trajectories to alternative gene orders cannot be entirely explained by the most obvious impediment: that potyviruses with reordered genomes most likely would be nonviable owing to improper autocatalytic processing of the polyprotein. In our study, experimental evolution of the reordered viruses illustrates other important barriers to gene order evolution.

In VSV, the evolution of alternative gene orders is more plausible than for TEV because the movement of the *N* gene does not affect viability of the virus (Wertz *et al.* 1998). The evolvability of variants of VSV with the *N* in alternative positions was higher in a cell line from an alternative host (Pesko *et al.* 2015). Our results for the evolution of TEV-*Nib*₁- Δ *Nib*₉

and TEV-*Nib*₂- Δ *Nib*₉ are comparable, although we considered evolution in a permissive host and, as such, expected to see little adaptation of the ancestral TEV. Pesko *et al.* (2015) did not consider the evolutionary trajectory leading up to the formation of a virus with a rearranged genome. However, at least in the cell-culture environment where MOI can be high for VSV, this may not be as big an impediment as for TEV. A computational study suggests that point mutations in VSV intergenic regions preceded or coevolved with fixation of the wild-type gene order, resulting in a sub-optimal genome organization (Lim and Yin 2009).

For our experiments, we have considered only one present-day potyvirus as well as variants with rearranged gene orders derived from it. Whether our results will extend to other present-day potyviruses is a valid question, but it is equally important to consider that these results may have only a limited bearing on the potential for gene-order evolution in ancestral potyviruses. As a result of epistasis, adaptive evolution can limit accessible evolutionary trajectories (Salverda *et al.* 2011), while purifying selection can result in entrenchment (Shah *et al.* 2015). It is therefore plausible that evolutionary trajectories to alternative gene orders may have been more accessible to ancestral viruses. We therefore do not rule out that other factors may have been important in conserving potyvirus gene order, in particular, at early time points in their evolution.

What conditions could make the evolutionary trajectory we have studied accessible for present-day viruses or could open up alternative evolutionary trajectories? We have explored evolvability of a single potyvirus genotype at different points along the trajectory to alternative gene orders. It is possible that certain genotypes could be less constrained, although given the low fitness of the intermediates considered and the instability of the new *Nib* copy, this potentiating variation probably would have to preexist prior to the first recombination step leading to gene duplication. Prime candidates for mutations that may mitigate such constraints are the convergent mutations found in the evolved lineages of TEV-*Nib*₁- Δ *Nib*₉ and TEV-*Nib*₂- Δ *Nib*₉. An alternative host species in which high MOI occurs also could open up new evolutionary trajectories by allowing complementation between viral genomes, thereby widening the set of plausible trajectories beyond only those involving gene duplication. We think that it is unlikely that such hosts exist, however, given that (1) in general, MOI estimates for plant RNA viruses tend to be low (Zwart *et al.* 2013), and (2) alternative hosts would tend to be semipermissive, and we therefore would intuitively expect lower infection levels and MOI in such a host. We think that the most promising avenue for further research on alternative gene orders is to consider the impact of potentiating mutations.

By showing these different barriers to alternative gene orders in viruses, we expect to drive further research on the diversity of gene order over different organisms. Our results serve as a roadmap for testing which factors constrain or promote gene-order conservation across different viruses and could be compared to the great diversity of gene order in other taxa.

Acknowledgments

We thank Francisca de la Iglesia and Paula Agudo for excellent technical assistance. This work was supported by the John Templeton Foundation (grant 22371), by the European Commission 7th Framework Program EvoEvo Project (grant ICT610427), and by the Spanish Ministerio de Economía y Competitividad (MINECO) (grants BFU2012-30805 to S.F.E. and BIO2011-26741 and BIO2014-54269-R to J.A.D.). A.W. was supported by a John Templeton Foundation grant and the EvoEvo Project. M.P.Z. was supported by a Juan de la Cierva postdoctoral contract from MINECO. E.M. was the recipient of a predoctoral fellowship (AP2012-3751) from the Spanish Ministerio de Educación, Cultura y Deporte. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. The authors declare that they have no competing interests.

Author contributions: A.W., M.P.Z., and S.F.E. designed the study. A.W., M.P.Z., and N.T. performed the experiments. E.M. and J.A.D. prepared the viral constructs. A.W., M.P.Z., and S.F.E. analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Literature Cited

- Adams, M. J., J. F. Antoniw, and C. M. Fauquet, 2005 Molecular criteria for genus and species discrimination within the family Potyviridae. *Arch. Virol.* 150: 459–479.
- Atreya, C. D., P. L. Atreya, D. W. Thornbury, and T. P. Pirone, 1992 Site-directed mutations in the potyvirus HC-Pro gene affect helper component activity, virus accumulation, and symptom expression in infected tobacco plants. *Virology* 191: 106–111.
- Bedoya, L. C., and J.-A. Daròs, 2010 Stability of *Tobacco etch virus* infectious clones in plasmid vectors. *Virus Res.* 149: 234–240.
- Belshaw, R., O. G. Pybus, and A. Rambaut, 2007 The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res.* 17: 1496–1504.
- Boyko, V. P., A. V. Karasev, A. A. Agranovsky, E. V. Koonin, and V. V. Dolja, 1992 Coat protein gene duplication in a filamentous RNA virus of plants. *Proc. Natl. Acad. Sci. USA* 89: 9156–9160.
- Carrasco, P., J.-A. Daròs, P. Agudelo-Romero, and S. F. Elena, 2007 A real-time RT-PCR assay for quantifying the fitness of tobacco etch virus in competition experiments. *J. Virol. Methods* 139: 181–188.
- Chung, B. Y., W. A. Miller, J. F. Atkins, and A. E. Firth, 2008 An overlapping essential gene in the Potyviridae. *Proc. Natl. Acad. Sci. USA* 105: 5897–5902.
- Cronin, S., J. Verchot, R. Haldeman-Cahill, M. C. Schaad, and J. C. Carrington, 1995 Long-distance movement factor: a transport function of the potyvirus helper component proteinase. *Plant Cell* 7: 549–559.
- Dandekar, T., B. Snel, M. Huynen, and P. Bork, 1998 Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.* 23: 324–328.
- de Haan, C. A. M., L. van Genne, J. N. Stoop, H. Volders, and P. J. M. Rottier, 2003 Coronaviruses as vectors: position dependence of foreign gene expression. *J. Virol.* 77: 11312–11323.
- Dietrich, C., and E. Maiss, 2003 Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected *Nicotiana benthamiana* plants. *J. Gen. Virol.* 84: 2871–2876.
- Dolja, V. V., K. L. Herndon, T. P. Pirone, and J. C. Carrington, 1993 Spontaneous mutagenesis of a plant potyvirus genome after insertion of a foreign gene. *J. Virol.* 67: 5968–5975.
- Dufresne, P. J., K. Thivierge, S. Cotton, C. Beauchemin, C. Ide *et al.*, 2008 Heat shock 70 protein interaction with *Turnip mosaic virus* RNA-dependent RNA polymerase within virus-induced membrane vesicles. *Virology* 374: 217–227.
- Endy, D., L. You, J. Yin, and I. J. Molineux, 2000 Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. USA* 97: 5375–5380.
- Engler, C., R. Gruetzner, R. Kandzia, and S. Marillonnet, 2009 Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One* 4: e5553.
- Fauquet, C. M., M. A. Mayo, J. Maniloff, U. Desselberger, and L. A. Ball, 2005 *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego.
- Forss, S., and H. Schaller, 1982 A tandem repeat gene in a picornavirus. *Nucleic Acids Res.* 10: 6441–6450.
- Francki, R. I. B., C. M. Fauquet, D. L. Knudson, and F. Brown, 1991 Classification and nomenclature of viruses: fifth report of the International Committee on Taxonomy of Viruses. *Arch. Virol.* 17: 1–405.
- Fukuhara, T., R. Koga, N. Aoki, C. Yuki, N. Yamamoto *et al.*, 2006 The wide distribution of endornaviruses, large double-stranded RNA replicons with plasmid-like properties. *Arch. Virol.* 151: 995–1002.
- Gutiérrez, S., E. Pirolles, M. Yvon, V. Baecker, Y. Michalakis *et al.*, 2015 The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. *J. Virol.* 89: 9665–9675.
- Himmelreich, R., H. Plagens, H. Hilbert, B. Reiner, and R. Herrmann, 1997 Comparative analysis of the genomes of the bacteria *Mycoplasma pneumoniae* and *Mycoplasma genitalium*. *Nucleic Acids Res.* 25: 701–712.
- Karasev, A. V., V. P. Boyko, S. Gowda, O. V. Nikolaeva, M. E. Hilf *et al.*, 1995 Complete sequence of the citrus tristeza virus RNA genome. *Virology* 208: 511–520.
- Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan *et al.*, 2012 VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22: 568–576.
- Kolstø, A. B., 1997 Dynamic bacterial genome organization. *Mol. Microbiol.* 24: 241–248.
- Koonin, E. V., 2009 Evolution of genome architecture. *Int. J. Biochem. Cell Biol.* 41: 298–306.
- Koonin, E. V., and V. V. Dolja, 1993 Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* 28: 375–430.
- Koonin, E. V., and M. Y. Galperin, 1997 Prokaryotic genomes: the emerging paradigm of genome-based microbiology. *Curr. Opin. Genet. Dev.* 7: 757–763.
- Koonin, E. V., Y. I. Wolf, K. Nagasaki, and V. V. Dolja, 2008 The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. *Nat. Rev. Microbiol.* 6: 925–939.
- Langmead, B., and S. L. Salzberg, 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9: 357–359.
- LaPierre, L. A., D. L. Holzschu, P. R. Bowser, and J. W. Casey, 1999 Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: evidence for a gene duplication. *J. Virol.* 73: 9393–9403.

- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Li, X. H., and J. C. Carrington, 1995 Complementation of tobacco etch potyvirus mutants by active RNA polymerase expressed in transgenic cells. *Proc. Natl. Acad. Sci. USA* 92: 457–461.
- Lim, K., and J. Yin, 2009 Computational fitness landscape for all gene-order permutations of an RNA virus. *PLOS Comput. Biol.* 5: e1000283.
- Lynch, M., 2006 Streamlining and simplification of microbial genome architecture. *Annu. Rev. Microbiol.* 60: 327–349.
- Majer, E., Z. Salvador, M. P. Zwart, A. Willemsen, S. F. Elena *et al.*, 2014 Relocation of the N1b gene in the tobacco etch potyvirus genome. *J. Virol.* 88: 4586–4590.
- Novella, I. S., L. A. Ball, and G. W. Wertz, 2004 Fitness analyses of vesicular stomatitis strains with rearranged genomes reveal replicative disadvantages. *J. Virol.* 78: 9837–9841.
- Peng, C. W., V. V. Peremyslov, A. R. Mushegian, W. O. Dawson, and V. V. Dolja, 2001 Functional specialization and evolution of leader proteinases in the family *Closteroviridae*. *J. Virol.* 75: 12153–12160.
- Pesko, K., E. A. Voigt, A. Swick, V. J. Morley, C. Timm *et al.*, 2015 Genome rearrangement affects RNA virus adaptability on prostate cancer cells. *Front. Genet.* 6: 121.
- Puustinen, P., and K. Mäkinen, 2004 Uridylylation of the potyvirus VPg by viral replicase N1b correlates with the nucleotide binding capacity of VPg. *J. Biol. Chem.* 279: 38103–38110.
- Revers, F., and J. A. García, 2015 Molecular biology of potyviruses. *Adv. Virus Res.* 92: 101–199.
- Rocha, E. P. C., 2008 The organization of the bacterial genome. *Annu. Rev. Genet.* 42: 211–233.
- Roossinck, M. J., S. Sabanadzovic, R. Okada, and R. A. Valverde, 2011 The remarkable evolutionary history of endornaviruses. *J. Gen. Virol.* 92: 2674–2678.
- Salverda, M. L. M., E. Dellus, F. A. Gorter, A. J. M. Debets, J. van der Oost *et al.*, 2011 Initial mutations direct alternative pathways of protein evolution. *PLoS Genet.* 7: e1001321.
- Schmieder, R., and R. Edwards, 2011 Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27: 863–864.
- Shah, P., D. M. McCandlish, and J. B. Plotkin, 2015 Contingency and entrenchment in protein evolution under purifying selection. *Proc. Natl. Acad. Sci. USA* 112: E3226–E3235.
- Sicard, A., M. Yvon, T. Timchenko, B. Gronenborn, Y. Michalakos *et al.*, 2013 Gene copy number is differentially regulated in a multipartite virus. *Nat. Commun.* 4: 2248.
- Siefert, J. L., K. A. Martin, F. Abdi, W. R. Widger, and G. E. Fox, 1997 Conserved gene clusters in bacterial genomes provide further support for the primacy of RNA. *J. Mol. Evol.* 45: 467–472.
- Simon-Loriere, E., and E. C. Holmes, 2013 Gene duplication is infrequent in the recent evolutionary history of RNA viruses. *Mol. Biol. Evol.* 30: 1263–1269.
- Song, D., W. K. Cho, S.-H. Park, Y. Jo, and K.-H. Kim, 2013 Evolution of and horizontal gene transfer in the *Endornavirus* genus. *PLoS One* 8: e64270.
- Spencer, D. H., M. Tyagi, F. Vallania, A. J. Bredemeyer, J. D. Pfeifer *et al.*, 2014 Performance of common analysis methods for detecting low-frequency single nucleotide variants in targeted next-generation sequence data. *J. Mol. Diagn.* 16: 75–88.
- Springman, R., M. R. Badgett, I. J. Molineux, and J. J. Bull, 2005 Gene order constrains adaptation in bacteriophage T7. *Virology* 341: 141–152.
- Sullivan, M. L., and P. Ahlquist, 1997 cis-Acting signals in bromovirus RNA replication and gene expression: networking with viral proteins and host factors. *Semin. Virol.* 8: 221–230.
- Thole, V., B. Worland, J. W. Snape, and P. Vain, 2007 The pCLEAN dual binary vector system for *Agrobacterium*-mediated plant transformation. *Plant Physiol.* 145: 1211–1219.
- Thornbury, D. W., C. A. Patterson, J. T. Dessens, and T. P. Pirone, 1990 Comparative sequence of the helper component (HC) region of *Potato virus Y* and a HC-defective strain, *Potato virus C*. *Virology* 178: 573–578.
- Tristem, M., C. Marshall, A. Karpas, J. Petrik, and F. Hill, 1990 Origin of vpx in lentiviruses. *Nature* 347: 341–342.
- Tromas, N., M. P. Zwart, G. Laffogues, and S. F. Elena, 2014a Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS Genet.* 10: e1004186.
- Tromas, N., M. P. Zwart, J. Forment, and S. F. Elena, 2014b Shrinkage of genome size in a plant RNA virus upon transfer of an essential viral gene into the host genome. *Genome Biol. Evol.* 6: 538–550.
- Valli, A., J. J. López-Moya, and J. A. García, 2007 Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family Potyviridae. *J. Gen. Virol.* 88: 1016–1028.
- Valverde, R. A., S. Nameth, O. Abdallah, O. Al-Musa, P. Desjardins *et al.*, 1990 Indigenous double-stranded RNA from pepper (*Capsicum annum*). *Plant Sci.* 67: 195–201.
- Velasquez, N., M. J. Hossain, and J. F. Murphy, 2014 Differential disease symptoms and full-length genome sequence analysis for three strains of *Tobacco etch virus*. *Virus Genes* 50: 442–449.
- Verchot, J., and J. C. Carrington, 1995a Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification. *J. Virol.* 69: 3668–3674.
- Verchot, J., and J. C. Carrington, 1995b Debilitation of plant potyvirus infectivity by P1 proteinase-inactivating mutations and restoration by second-site modifications. *J. Virol.* 69: 1582–1590.
- Wakarchuk, D. A., and R. I. Hamilton, 1990 Partial nucleotide sequence from enigmatic dsRNAs in *Phaseolus vulgaris*. *Plant Mol. Biol.* 14: 637–639.
- Walker, P. J., K. A. Byrne, G. A. Riding, J. A. Cowley, Y. Wang *et al.*, 1992 The genome of bovine ephemeral fever rhabdovirus contains two related glycoprotein genes. *Virology* 191: 49–61.
- Wang, Y., and P. J. Walker, 1993 Adelaide river rhabdovirus expresses consecutive glycoprotein genes as polycistronic mRNAs: new evidence of gene duplication as an evolutionary process. *Virology* 195: 719–731.
- Watanabe, H., H. Mori, T. Itoh, and T. Gojobori, 1997 Genome plasticity as a paradigm of eubacteria evolution. *J. Mol. Evol.* 44 (Suppl. 1): S57–S64.
- Wertz, G. W., V. P. Perepelitsa, and L. A. Ball, 1998 Gene rearrangement attenuates expression and lethality of a nonsegmented negative strand RNA virus. *Proc. Natl. Acad. Sci. USA* 95: 3501–3506.
- Zwart, M. P., J.-A. Daròs, and S. F. Elena, 2011 One is enough: *in vivo* effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog.* 7: e1002122.
- Zwart, M. P., N. Tromas, and S. F. Elena, 2013 Model-selection-based approach for calculating cellular multiplicity of infection during virus colonization of multi-cellular hosts. *PLoS One* 8: e64657.
- Zwart, M. P., A. Willemsen, J. A. Daròs, and S. F. Elena, 2014 Experimental evolution of pseudogenization and gene loss in a plant RNA virus. *Mol. Biol. Evol.* 31: 121–134.

Communicating editor: J. J. Bull

GENETICS

Supporting Information

www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1

Multiple Barriers to the Evolution of Alternative Gene Orders in a Positive-Strand RNA Virus

**Anouk Willemsen, Mark P. Zwart, Nicolas Tromas, Eszter Majer, José-Antonio Daròs,
and Santiago F. Elena**

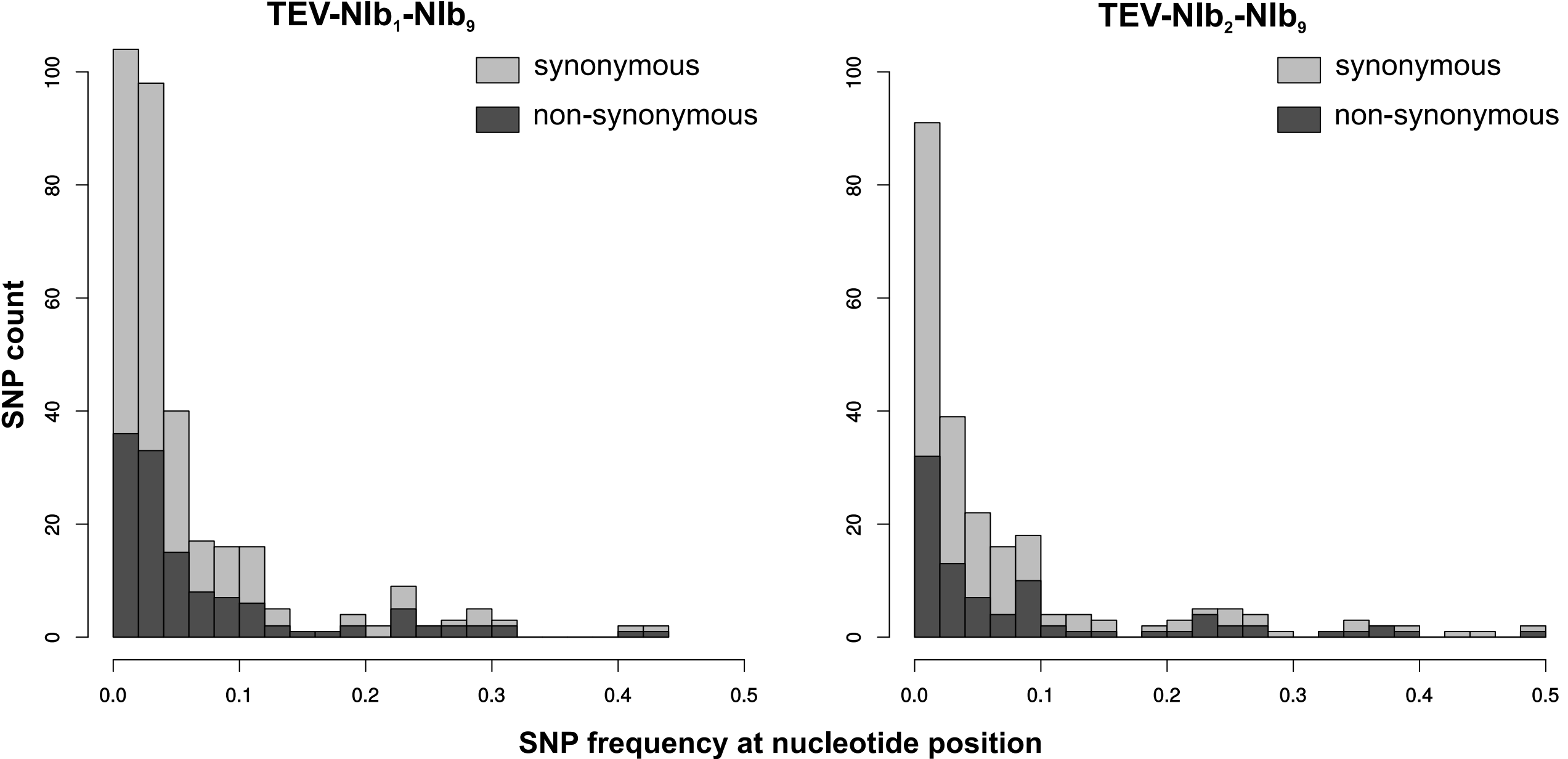


Figure S1. The distribution of SNP frequencies in the evolved TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ lineages

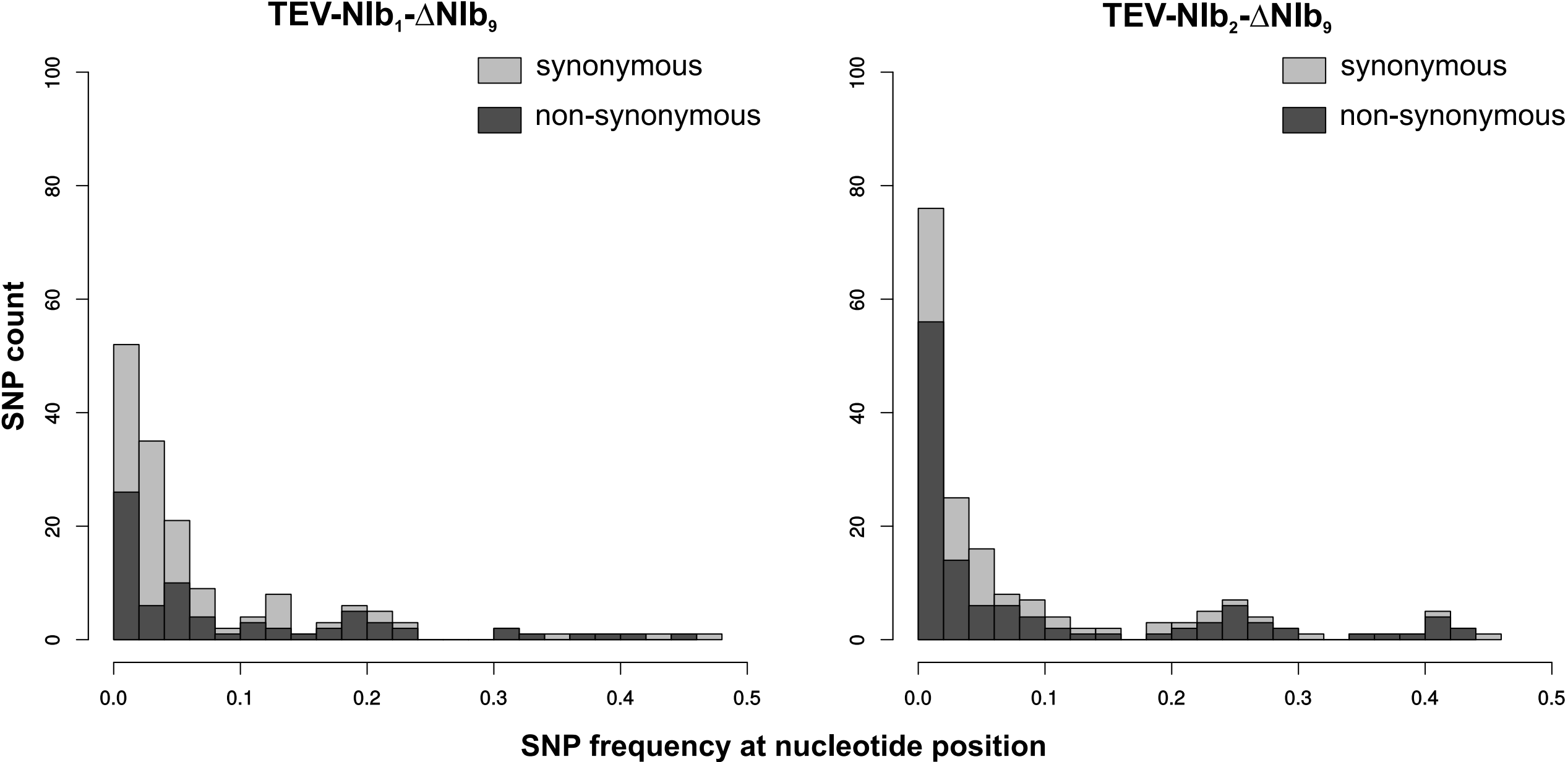


Figure S2. The distribution of SNP frequencies in the evolved TEV-Nib₁-ΔNib₉ and TEV-Nib₂-ΔNib₉ lineages

Table S1. Nonlinear regression analysis of genome sizes over time. (.xlsx, 15 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS1.xlsx

Table S2. High frequency mutations of evolved TEV-Nib₁-Nib₉ and TEV-Nib₂-Nib₉ lineages as compared to their ancestral sequence. (.xlsx, 18 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS2.xlsx

**Table S3. Within population sequence variation of the evolved and ancestral TEV-Nib₁-
Nib₉ lineages (.xlsx, 42 KB)**

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS3.xlsx

**Table S4. Within population sequence variation of the evolved and ancestral TEV-Nib₁-
Nib₉ lineages (.xlsx, 34 KB)**

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS4.xlsx

Table S5. High frequency mutations of evolved TEV-NIb₁-ΔNIb₉ and TEV-NIb₂-ΔNIb₉ lineages as compared to their ancestral sequence. (.xlsx, 17 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS5.xlsx

Table S6. Within population sequence variation of the evolved and ancestral TEV-Nib₁- Δ Nib₉ lineages. (.xlsx, 26 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS6.xlsx

Table S7. Within population sequence variation of the evolved and ancestral TEV-Nib₂- Δ Nib₉ lineages (.xlsx, 36 KB)

Available for download as a .xlsx file at
www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.185017/-/DC1/TableS7.xlsx