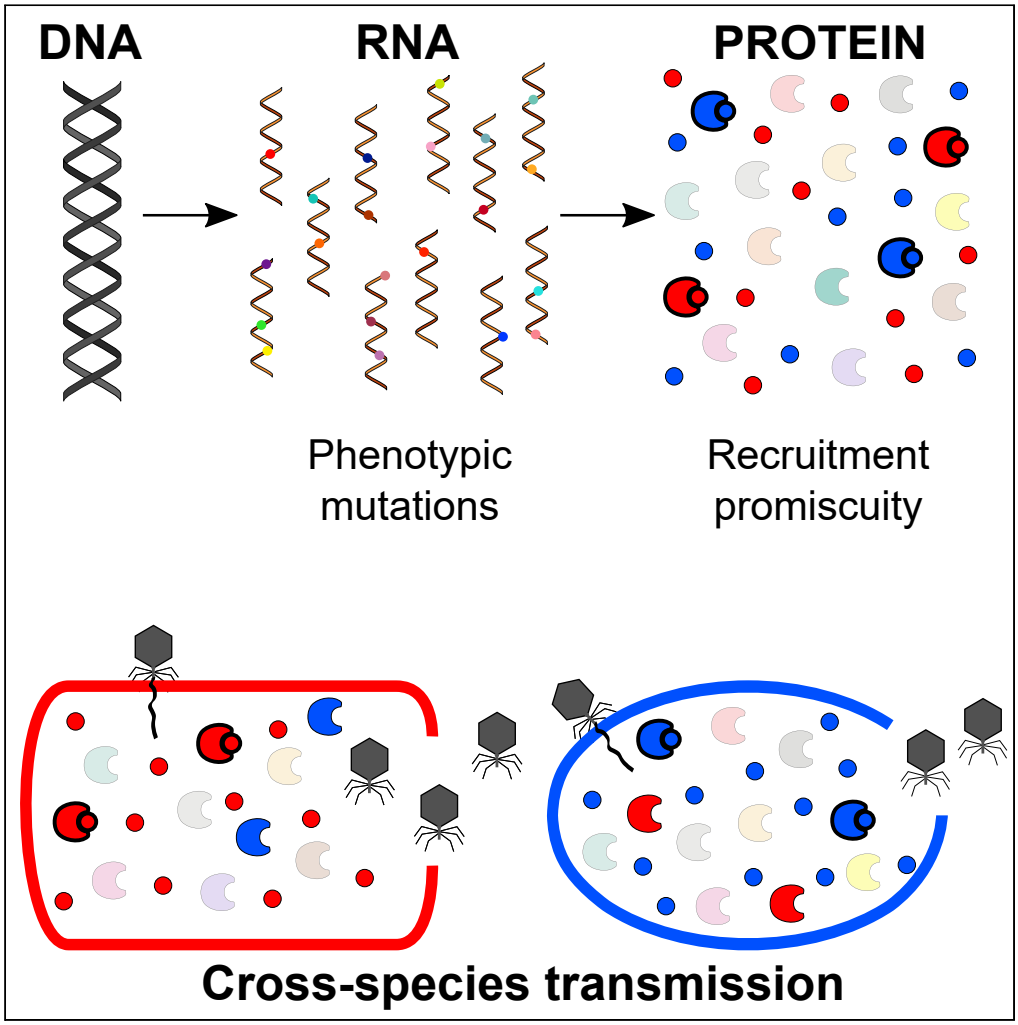
iScience

CellPress
OPEN ACCESS

Article

Evidence for a role of phenotypic mutations in virus adaptation



Raquel Luzon-Hidalgo, Valeria A. Risso, Asuncion Delgado, Eduardo Andrés-León, Beatriz Ibarra-Molero, Jose M. Sanchez-Ruiz

sanchezr@ugr.es

Highlights

Phage adapts to a host modified to hinder the essential recruitment of a host protein

No genetic mutations are fixed at the engineered virus-host interaction interface

Adaptation is likely linked to phenotypic mutations caused by transcription errors

Sub-genomic RNAs may enable this kind of adaptation mechanism in coronaviruses

Luzon-Hidalgo et al., iScience 24, 102257
April 23, 2021
https://doi.org/10.1016/j.isci.2021.102257

iScience

CellPress
OPEN ACCESS

Article

Evidence for a role of phenotypic mutations in virus adaptation

Raquel Luzon-Hidalgo,[1,3] Valeria A. Risso,[1,3] Asuncion Delgado,[1] Eduardo Andrés-León,[2] Beatriz Ibarra-Molero,[1] and Jose M. Sanchez-Ruiz[1,4,*]

SUMMARY

Viruses interact extensively with the host molecular machinery, but the underlying mechanisms are poorly understood. Bacteriophage T7 recruits the small protein thioredoxin of the *Escherichia coli* host as an essential processivity factor for the viral DNA polymerase. We challenged the phage to propagate in a host in which thioredoxin had been extensively modified to hamper its recruitment. The virus adapted to the engineered host without losing the capability to propagate in the original host, but no genetic mutations were fixed in the thioredoxin binding domain of the viral DNA polymerase. Virus adaptation correlated with mutations in the viral RNA polymerase, supporting that promiscuous thioredoxin recruitment was enabled by phenotypic mutations caused by transcription errors. These results point to a mechanism of virus adaptation that may play a role in cross-species transmission. We propose that phenotypic mutations may generally contribute to the capability of viruses to evade antiviral strategies.

INTRODUCTION

Viruses repurpose the host molecular machinery for their own proliferation, block or evade antiviral factors, and recruit host proteins (proviral factors) for processes critical for virus propagation. Viruses occasionally jump between species, sometimes with fatal consequences for the new host. Homolog proteins that carry out the same tasks in two different organisms likely share functionality and 3D-structure but differ in amino acid sequence, and, consequently, they also differ in the chemical composition of their exposed surfaces available for intermolecular interaction. In view of this, cross-species transmission is quite remarkable, as it implies that the virus can simultaneously establish extensive and effective interactions in the two quite different molecular environments of the new and the old hosts. To explore fundamental mechanisms behind this molecular-level promiscuity, we have carried out laboratory evolution experiments in which a lytic phage is challenged to propagate in an engineered host that had been modified to hinder the recruitment of a known proviral factor. Our experiments target a specific virus-host interaction for which extensive functional and structural information is available, thus considerably facilitating the interpretation of the experimental results. Indeed, although we observe quick and efficient evolutionary adaptation of the virus to the engineered host, DNA sequencing reveals that no inheritable, genetic mutations occur at the targeted virus-host interaction. This essentially leaves phenotypic mutations caused by transcription errors as the only possible mechanism of adaptation in this case, a scenario that is actually consistent with the mutational pattern we find in the viral RNA polymerase.

Mistakes in protein synthesis, due to translation and transcription errors, are common and lead to the so-called phenotypic mutations (Drummond and Wilke, 2009; Goldsmith and Tawfik, 2009). Translation error rates are generally higher than transcription error rates. Yet, transcription errors may have a stronger impact at the protein level, because an mRNA molecule is typically translated many times (Traverse and Ochman, 2016). Regardless of their origin, phenotypic mutations are not inherited and are often regarded as a burden to the organism because they may result in a substantial amount of misfolded protein molecules that are non-functional and that may actually be harmful (Drummond and Wilke, 2009; Goldsmith and Tawfik, 2009). On the other hand, some phenotypic mutations could provide crucial functional advantages under certain conditions. It has been theorized that phenotypic mutations may play an evolutionary role by allowing organism survival until functionally useful mutations appear at the genetic level (Whitehead et al., 2008; Petrovic et al., 2018). The laboratory evolution experiments reported here support this hypothesis while providing evidence for a hitherto unknown mechanism of virus adaptation.

[1]Departamento de Quimica Fisica. Facultad de Ciencias, Unidad de Excelencia de Quimica Aplicada a Biomedicina y Medioambiente (UEQ), Universidad de Granada, Granada 18071, Spain

[2]Unidad de Bioinformática. Instituto de Parasitología y Biomedicina "López Neyra", CSIC, Armilla, Granada 18016, Spain

[3]These authors contributed equally

[4]Lead contact

*Correspondence: sanchezr@ugr.es

https://doi.org/10.1016/j.isci.2021.102257

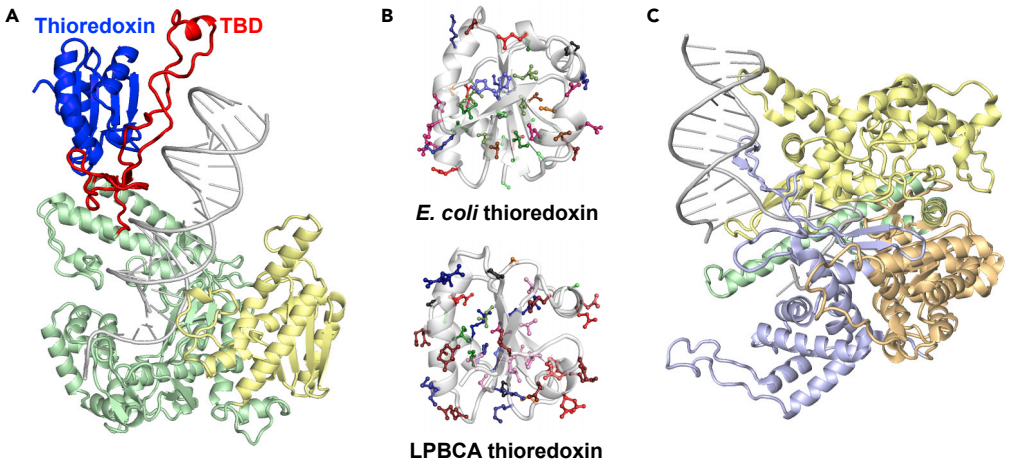


Figure 1. 3D structures of the main molecular players in this work

(A) Structure of the viral DNA polymerase interacting with its processivity factor, *E. coli* thioredoxin (PDB ID 1T8E). The thioredoxin-binding domain of the viral polymerase is labeled with TBD.

(B) Structures of *E. coli* thioredoxin (PDB ID 2TRX) and the alternative LPBCA thioredoxin (PDB ID 2YJ7) used in this work. The amino acid residues that differ between the two structures are shown in stick representation.

(C) Structure of the viral RNA polymerase (PDB ID 1QLN).

More generally, we argue here that an adaptation mechanism based on phenotypic mutations has important implications beyond the specific model system used here to demonstrate it. Many key intermolecular interactions may need to occur only a few times per host cell to allow virus propagation and can, therefore, be realized by mutant proteins present at very low levels. It follows that diversity at the phenotypic level caused by transcription errors may generate an extensive capability to establish/avoid interactions in different molecular environments, thus facilitating cross-species transmission and potentially contributing to the capability of viruses to evade antiviral strategies.

RESULTS

Design of the engineered host

Bacteriophage T7 recruits the small (~110 residues) protein thioredoxin from the *Escherichia coli* host to be a part of its four-protein replisome (Hamdan and Richardson, 2009). Upon recruitment, thioredoxin becomes an essential processivity factor for the viral DNA polymerase (Etson et al., 2010) (Figure 1A). In the absence of thioredoxin, the processivity of the DNA polymerase is only 1–15 nucleotides per binding event (versus several hundred upon thioredoxin binding), a very low level that does not lead to virus replication and propagation in the knockout strains used in this work (see below). Binding of *Escherichia coli* thioredoxin to the viral DNA polymerase is mediated by the interaction with a unique 76-residue fragment often referred to as the thioredoxin-binding domain or TBD (Hamdan and Richardson, 2009; Akabayov et al., 2010; Lee and Richardson, 2011). This interaction is extremely tight (dissociation constant ~5 nanomolar), showing that the TBD has evolved to specifically bind the host thioredoxin (Hamdan and Richardson, 2009). It follows that replacing *E. coli* thioredoxin with an alternative thioredoxin could hinder recruitment and potentially prevent phage propagation in *E. coli* (Delgado et al., 2017). A simple way to perform such replacement involves complementing a knockout *E. coli* Trx^- strain with a plasmid bearing the alternative thioredoxin gene (Delgado et al., 2017). The engineered host used in the experiments reported here is further modified to allow phenotypic mutations linked to viral transcription errors to occur in both partners of the targeted DNA polymerase-thioredoxin interaction. That is, the DNA polymerase is transcribed by the viral T7 RNA polymerase, and, in addition, we placed the alternative thioredoxin gene in the plasmid under a promoter of the viral RNA polymerase. The gene for the wild-type T7 RNA polymerase was inserted in the host chromosome, which ensures the presence of a significant amount of thioredoxin even before phage infection. In order to have available a suitable control host, the same procedure was performed with an *E. coli* Trx^- complemented with a plasmid bearing the gene of *E. coli* thioredoxin. The resulting strain is similar to *E. coli* in terms of growth and susceptibility to bacteriophage T7 (Delgado et al., 2017) and has been used as a representation of the original host throughout this work.

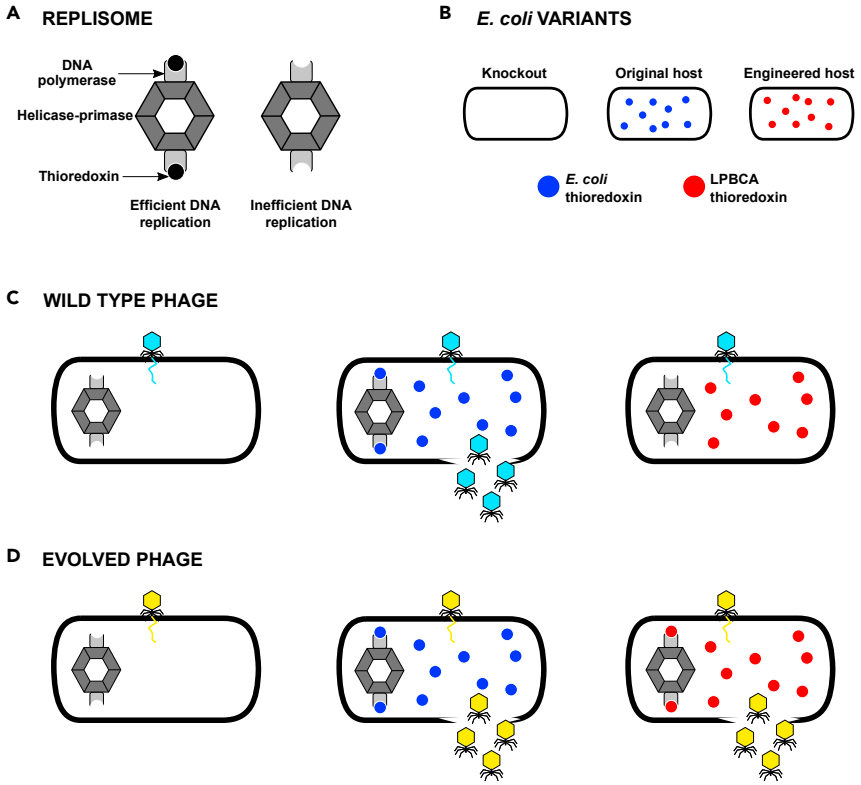


Figure 2. *E. coli* variants used in this work and their behavior as hosts for phage replication

(A) *E. coli* thioredoxin is recruited for the T7 bacteriophage replisome as an essential processivity factor for the viral DNA polymerase. A cartoon representation of the replisome is shown here. See Figure 1 in Hamdan and Richardson (2009) for a more realistic model.

(B) Differences between the three *E. coli* variants used in this work are related to thioredoxin. An *E. coli* trx− knockout strain lacks thioredoxin. A control strain expresses the "normal" *E. coli* thioredoxin, and it is used as our representation of the original host. The engineered host expresses the alternative LPBCA thioredoxin, which displays only 57% sequence identity with *E. coli* thioredoxin (see Figure 1B).

(C) DNA replication for wild-type T7 bacteriophage requires recruitment of *E. coli* thioredoxin, and efficient virus propagation is only observed with the original host.

(D) Results reported in this work show, however, that the virus can evolve to propagate in the engineered host. The evolved virus does not propagate in the knockout host, but still propagates efficiently in the original host, pointing to a promiscuous capability to recruit both *E. coli* thioredoxin and LPBCA thioredoxin, as symbolically depicted by the replisomes shown.

As alternative thioredoxin we used a protein, LPBCA thioredoxin, previously obtained and characterized in detail as part of an ancestral reconstruction study (Perez-Jimenez et al., 2011; Ingles-Prieto et al., 2013). Actually, LPBCA stands for "last common ancestor of cyanobacterial, deinococcus and thermos groups." LPBCA thioredoxin shares function and 3D-structure with *E. coli* thioredoxin but displays only 57% sequence identity with its modern counterpart. As a result, the amino acid composition of its exposed protein surfaces is substantially altered with respect to *E. coli* thioredoxin (Figure 1B), and efficient interaction with the thioredoxin binding domain of the viral DNA polymerase is not likely to occur (Delgado et al., 2017). The putative ancestral nature of LPBCA thioredoxin is, of course, immaterial for this work. The key features of LPBCA thioredoxin in the context of this work are that (1) it poses an *a priori* tough challenge to the virus and (2) we know beforehand that mutations enabling recruitment of the alternative thioredoxin must be at the TBD-thioredoxin interaction surface. This fact facilitates considerably the interpretation of the experimental data. Figure 2 shows schematically the *E. coli* variants used in this work as well as their main features as hosts for phage replication.

General design of the laboratory evolution experiments

In plaque assays with the *E. coli* Trx^- strain complemented with LPBCA thioredoxin (i.e., our engineered host), plaques are only observed at the highest virus concentrations used and, even under those conditions,

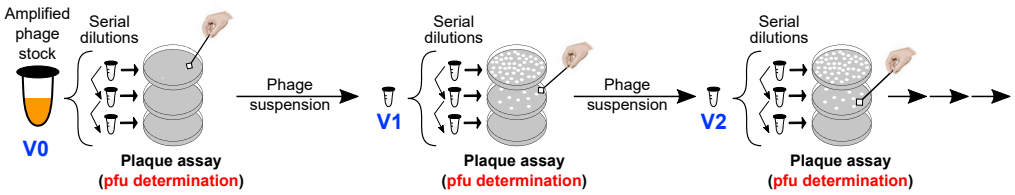


Figure 3. Initial phage evolution experiments performed in this work

In each evolutionary round, plaque assays with serial dilutions of a phage suspension are performed, and, therefore, number of plaque-forming units in the suspension can be calculated. At each round, a well-defined plaque from the highest dilution at which plaques are observed is picked and suspended in buffer to yield the phage suspension used to start the next round (see main text and transparent methods for further details). This protocol was used in the two long evolution experiments that led to pfu profiles of Figure 4 below.

they are observed only occasionally (Delgado et al., 2017). It appears then that only a tiny fraction of the virions in the phage sample ($\sim 10^{-7}$ or less) can actually propagate in the engineered host. Furthermore, the propagation is initially inefficient, as judged by the very small size of the plaques they generate (Figure S1). Here, we have evolved such "anomalous" virions for efficient propagation in the engineered host. Schematically, our evolution experiments can be represented as

$$V_0 \Rightarrow V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} V_3 \xrightarrow{R_3} V_4 \xrightarrow{R_4} \cdot\cdot\cdot\cdot\cdot\cdot$$

where V_0 is the original virus sample, V_1 is a virus sample obtained from the initial propagation in the engineered host (double-line arrow), and the V_2, V_3, etc. samples result from the R_1, R_2, etc. evolutionary rounds, each involving plaque assays for propagation in the engineered host with 10-fold dilutions and plaque selection (see transparent methods for details and Figure 3 for a pictorial representation of the evolution experiments). These assays immediately lead to numbers of virus particles (plaque-forming units or pfu) that could infect the engineered host. For each plaque selected from the rounds of adaptation to the engineered host, we also performed plaque assays to determine the number of particles that could infect the original host.

Evolution experiments reveal promiscuous viral adaptation

We first performed two long evolution experiments consisting of an evolutionary round per day during several weeks. In these experiments, samples corresponding to a plaque surface of 3 × 3 mm^2 were used to start each next round (except for the V_1 sample for which the whole plaque was used, because it was smaller than 3 × 3 mm^2). Both experiments (Figure 4) revealed a fast adaptation of the virus to engineered host, as shown by increases in plaque size and in the numbers of plaque-forming units (see Figure S2 for a representative example of the plaque size increases observed in this work). In fact, the numbers of plaque-forming units determined using the engineered and the original host became similar to each other within the first 1–2 rounds and, surprisingly, the two numbers remained similar over many rounds of evolution. The two experiments were stopped at some point and restarted after about 1 month from plates stored at 4°C. In one case (experiment in lower panel of Figure 4), the protocol was changed after re-starting by eliminating an amplification step (see legend of Figure 4 for details). As expected, these 1-month breaks are reflected in sudden drops in pfu numbers. Nevertheless, the pfu numbers determined using the original and the engineered hosts still remained very similar to each other. The congruence between the two pfu numbers is quite remarkable because we challenged the virus to propagate in the engineered host, and we did not impose any selective pressure for propagation in the original host. Therefore, the capability of the evolved virus to propagate in the two hosts cannot be explained by assuming two genetically differentiated virus sub-species in our samples, because the subspecies competent to propagate in the engineered host would have quickly dominated the virus population in the evolution experiments. Obviously, whatever the virus adaptation mechanism is, it is intrinsically promiscuous, meaning that each infective virion can propagate in both hosts.

Virus adaptation and host promiscuity are not explained by genetic mutations

It is important to note that none of the many evolved virus samples studied in this work was found to propagate in the knock-out *E. coli* Trx^- strain that lacks thioredoxins. Therefore, the observed host promiscuity is not due to the virus evolving a capability to assemble a functional replisome without the assistance of

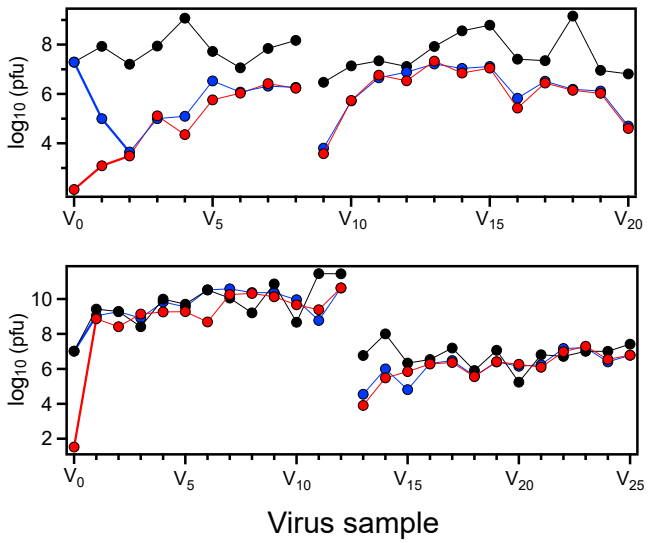


Figure 4. Propagation of bacteriophage T7 in engineered *E. coli* cells with a modified thioredoxin gene

Number of plaque-forming units (pfu) for virus samples from two laboratory evolution experiments performed as described in the text. Pfu values determined using the engineered host (red) and the original host (blue) are shown. It is important to note, however, that only selection for propagation in the engineered host has been applied in these experiments. Data from control experiments are also shown (black). These control experiments used the original (non-evolved) phage and the original *E. coli* host and were carried out concurrently with the evolution experiments. Experiments were interrupted after a substantial number of rounds and restarted from stored plaques after about a month and continued for an additional number of rounds. The interruption is apparent from the break in the lines that connect the experimental data points.

thioredoxin. Rather, it must be linked to a capability of the viral DNA polymerase to recruit the thioredoxins in both the original and the engineered hosts. In order to determine the mutations in the viral DNA polymerase gene that could potentially be responsible for the promiscuous recruitment, we carried out 14 "short" evolution experiments and we performed PCR followed by Sanger sequencing on samples from the early rounds (see Figure 5 for a pictorial representation of these evolution experiments). Between 1 and 3 genetic mutations appeared in the DNA polymerase in most experiments (see Figure 6 and also Table S1 for details), but none of them occurred in the thioredoxin-binding domain. Interestingly, however, the 21 positions at which mutations are found (including data from the {A,B,C} experiments discussed below) define a clear structural pattern (Figure 6), clustering in the region of the polymerase domain close to the bound DNA and in the exonuclease domain involved in proofreading. Therefore, none of the

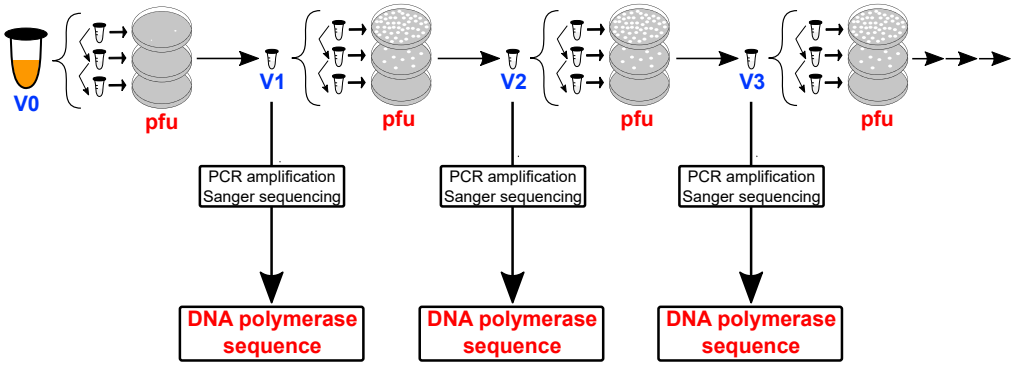


Figure 5. Evolution experiments modified to include DNA polymerase sequencing

The protocol described in Figure 3 can be easily modified to include PCR amplification and Sanger sequencing of the DNA polymerase for each phage suspension. This modified protocol was used in the 14 short evolution experiments that led to the sequence information on the viral DNA polymerase shown in Figure 6 and Table S1.

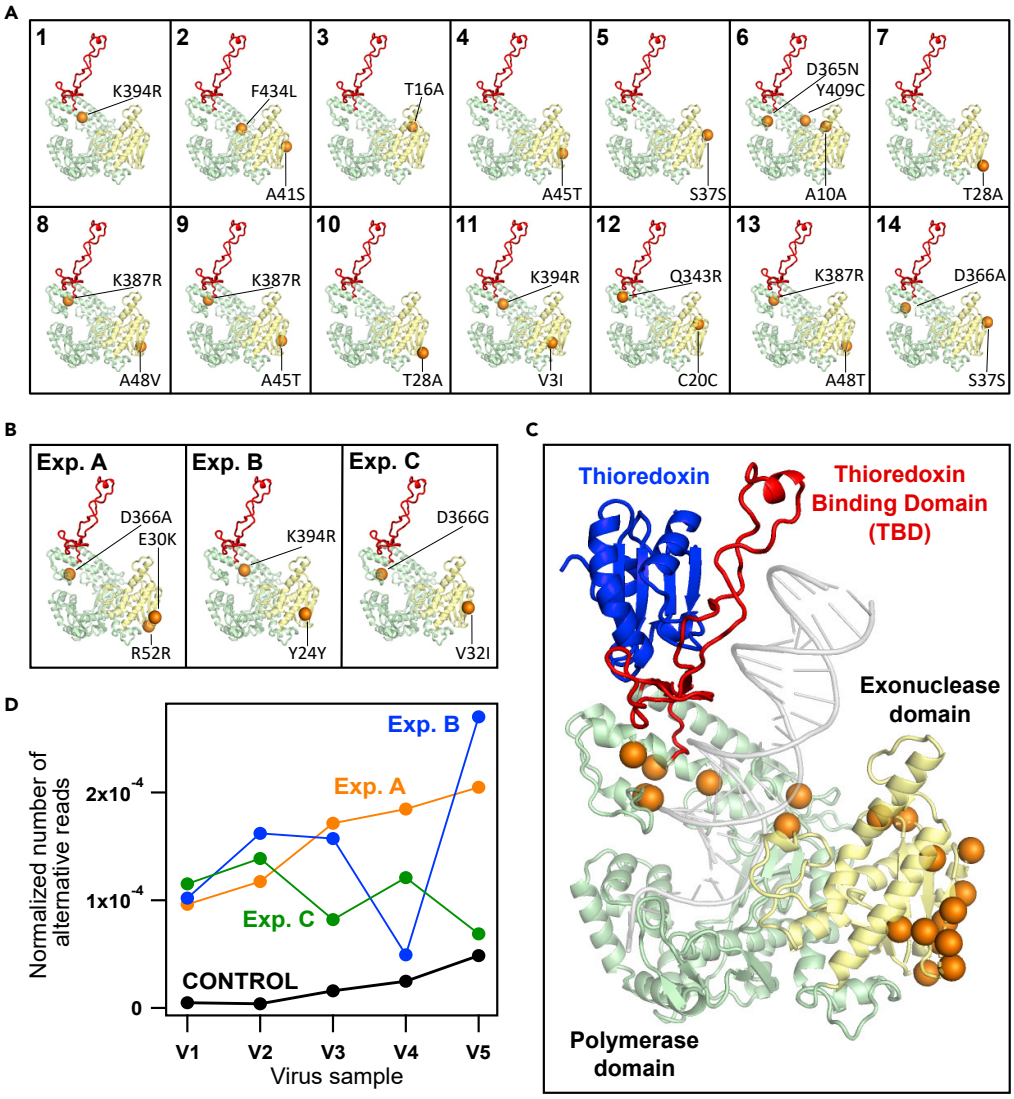


Figure 6. Structure of the viral DNA polymerase showing mutations appearing during phage adaptation to the engineered host

(A) Mutations found in 14 "short" evolution experiments consisting of 2–3 rounds after PCR and Sanger sequencing of the DNA polymerase (see main text, Figure 5 and Table S1 for details).

(B) Mutations observed at high frequency (fraction >0.95) in the {A,B,C} set of evolution experiments, as determined from next-generation sequencing of the viral genome (see Figure 7 and main text for details). The mutations shown are those determined for the virus samples obtained after the fourth round (V5 samples). See Table S2, Figure 9 below and Figure S3 for the mutations in all viral samples during the evolution experiment.

(C) All sites at which mutations were found are highlighted with spheres in a single DNA polymerase structure. It is visually apparent that no mutations were found in the thioredoxin-binding domain (TBD).

(D) Plot of normalized number of alternative reads in the genome of virus samples from the several rounds of the {A,B,C} experiments compared with a control experiment in which the non-evolved phage was allowed to propagate in the original host. These values support that the mutations fixed in the DNA polymerase during the {A,B,C} experiments (panel B) increase the replication errors (see text for details).

mutations found can reasonably explain thioredoxin recruitment in the engineered host and, certainly, they do not allow efficient replication in the absence of thioredoxin, because the evolved virus does not propagate in the knock-out *E. coli* Trx^{-} strain. However, it is plausible that these mutations increase replication errors, thus promoting mutations in other viral proteins that trigger processes that eventually lead to recruitment. This interpretation is supported by experiments and analyses given further below.

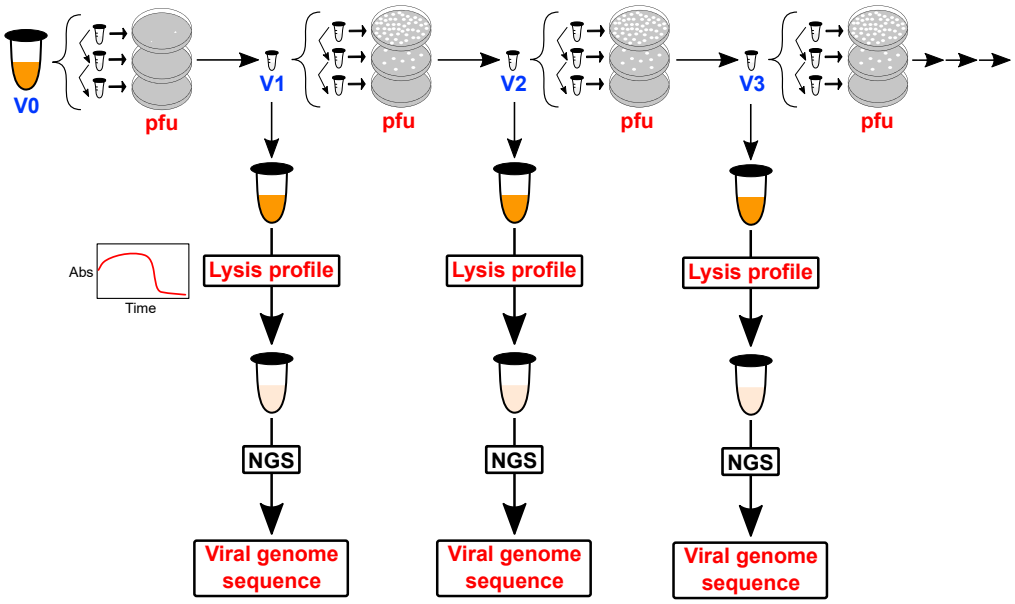


Figure 7. Evolution experiments modified to include the determination of lysis curves and NGS of the viral genome

The evolution experiments described in Figure 3 can be easily modified to determine *E. coli* lysis profiles (turbidity *versus* time) for the phage from the suspensions at each evolutionary round (see transparent methods for details). Furthermore, DNA extraction after lysis allows next-generation sequencing of the viral genome at each evolutionary round (see transparent methods for details). This modified protocol was used in the {A,B,C} set of evolution experiments that led to the lysis profiles of Figure 8 and the sequence information provided in Figures 9, 10, S3–S7 and Table S2.

Virus adaptation probed by lysis profiles

Host promiscuity, as described in the preceding sections, was inferred from the congruence between the numbers of plaque-forming units determined using the engineered host and the control, "original" host. The implication is, therefore, that each infective virion particle from the evolved virus sample is capable of replicating in both hosts. This, however, does not necessarily mean that the efficiency of virus replication is the same in the two hosts. In order to investigate replication efficiency, we performed three additional evolution experiments using a modified protocol that allows lysis profiles to be determined (see Figure 7 for a pictorial representation of these experiments). We label these experiments as A, B, and C and collectively refer to them as to the {A,B,C} set. Briefly, experiments were performed as described earlier, but, in addition, aliquots of the virus suspension at each round were used to assess lysis in solution with the original host, the engineered host, and, as a control, with the knockout *E. coli* Trx^{-} strain.

In the three experiments of the {A,B,C} set, the phage adapted to the engineered host similarly to what was observed in the previous evolution experiments discussed earlier. Adaptation was thus immediately clear from the general trend toward increased plaque sizes in the first rounds. It is to be noted, however, that, in order to maximize the amount of sample for DNA extraction, the largest plaque was *fully* removed from the agar plate with the smaller number of plaques to start each next round in the {A,B,C} experiments. This leads to large increases in the determined pfu numbers over the evolution experiment. Actually, the combined pfu data for the three experiments span 5 orders of magnitude (Figure 8A). Still, there is clear congruence between the pfu values determined using the engineered and the original hosts over this very wide range, showing again that the mechanism of virus adaptation is intrinsically promiscuous.

Lysis profiles for experiments of the {A,B,C} set reveal that, although infective virions can replicate in the original and the engineered hosts, they certainly do not do so with the same efficiency, at least in the first evolutionary rounds. As shown in Figure 8B, lysis times decrease over the evolution rounds reflecting, at least in part, the increases in viral load associated with the use of full plaques to start rounds. More relevant is the fact that, initially, lysis times are clearly larger for the engineered host as compared with the original host (Figure 8B). Upon virus adaptation, however, lysis times for the engineered host approach those for

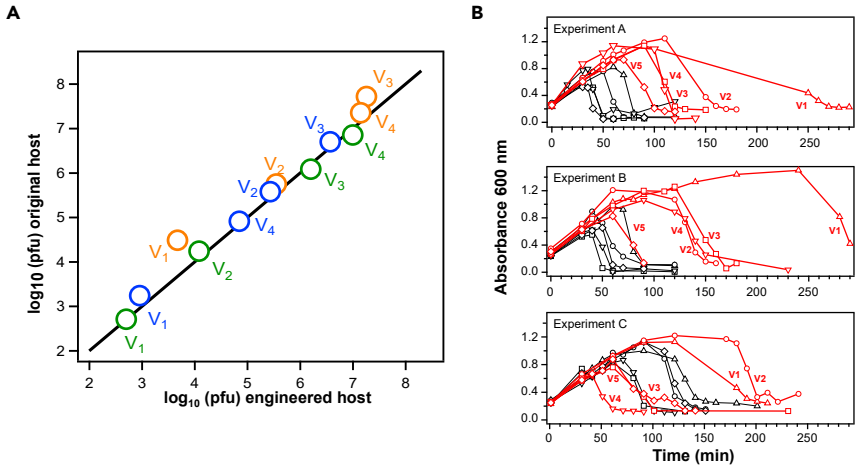


Figure 8. Engineered host versus original host adaptation during laboratory virus evolution

Data from the {A,B,C} set of laboratory evolution experiments (see text for details). The three experiments involve selection for propagation in the engineered host.

(A) Pfu values for virus samples from the evolution experiments determined with the original host versus the values determined with the engineered host. Note that the values span about 5 orders of magnitude. The line is not a fit but represents the equality of the two pfu values. Color code identifies the experiment: A (orange), B (blue), C (green).

(B) Lysis plots of absorbance at 600 nm versus time for the virus samples from the three evolution experiments. Absorbance at 600 nm reflects turbidity and lysis is revealed by an absorbance drop. Profiles determined using the engineered host (red) and the original host (black) are shown. For both the original host and the engineered host, symbols identify the virus sample: triangles (V_1), circles (V_2), squares (V_3), down-pointing triangles (V_4), diamonds (V_5). Lysis was not observed in control experiments with the knockout *E. coli* Trx$^-$ strain that lacks thioredoxin.

the original host, although to an extent that it is variable. Lysis times for the engineered host always remain larger in one experiment (labeled A), whereas they eventually become somewhat smaller than the lysis times for the original host in another experiment (labeled C). As expected, no lysis was observed with the knockout *E. coli* Trx$^-$ strain that lacks thioredoxin, not even after overnight incubation. Overall, from the point of view of engineered versus original host replication efficiency, the experiments in Figure 8B can be ranked C > B > A, with the later virus samples from experiment C being the most efficient at replicating in the engineered host.

Next-generation sequencing studies on virus adaptation to the engineered host

The protocol used for the {A,B,C} experiments allows sequencing of the viral genome at each evolutionary round (see Figure 7). Thus, DNA was extracted from the lysed engineered host samples and used for Illumina next generation sequencing. Certainly, lysis experiments involve an additional evolution step in solution that can potentially result in additional mutations. Still, as it will be apparent from the discussions further below, the DNA sequencing information from the lysed samples provides a clear picture of the molecular mechanism behind viral adaptation.

The Illumina sequencing data for the {A,B,C} set were processed in two ways. First, single nucleotide variants (SNVs) that appeared at high frequency (fraction over the total of readings at the position >0.95) were determined for all viral genes. These results are shown in Figures 9 and S7 and highlighted in yellow in Table S2. Secondly, for genes of particular interest, we considered all SNVs fulfilling the following two criteria: (1) the fraction of the nucleotide variant at a given position over the total number of readings was 0.01 or higher; (2) the SNV is observed at least two times. Overall, for both, the DNA polymerase and the RNA polymerase, the selected SNVs define clearly bimodal distributions, including low-frequency SNVs (fraction over total number of readings between 0.01 and 0.05) and high-frequency SNVs (fraction between 0.95 and 1) with no SNVs at intermediate fractions in most cases (Figures S3 and S4). The numbers of low-frequency SNVs are shown in Figure 10, whereas the mutations themselves are given in Table S2. Note that most of these low-frequency SNVs appear only transiently, are not enriched, and do not become fixed. On the other hand, for most of the mutations that do become fixed (see files highlighted in yellow in Table S2), the mutation is often detected at a lower level in an early round.

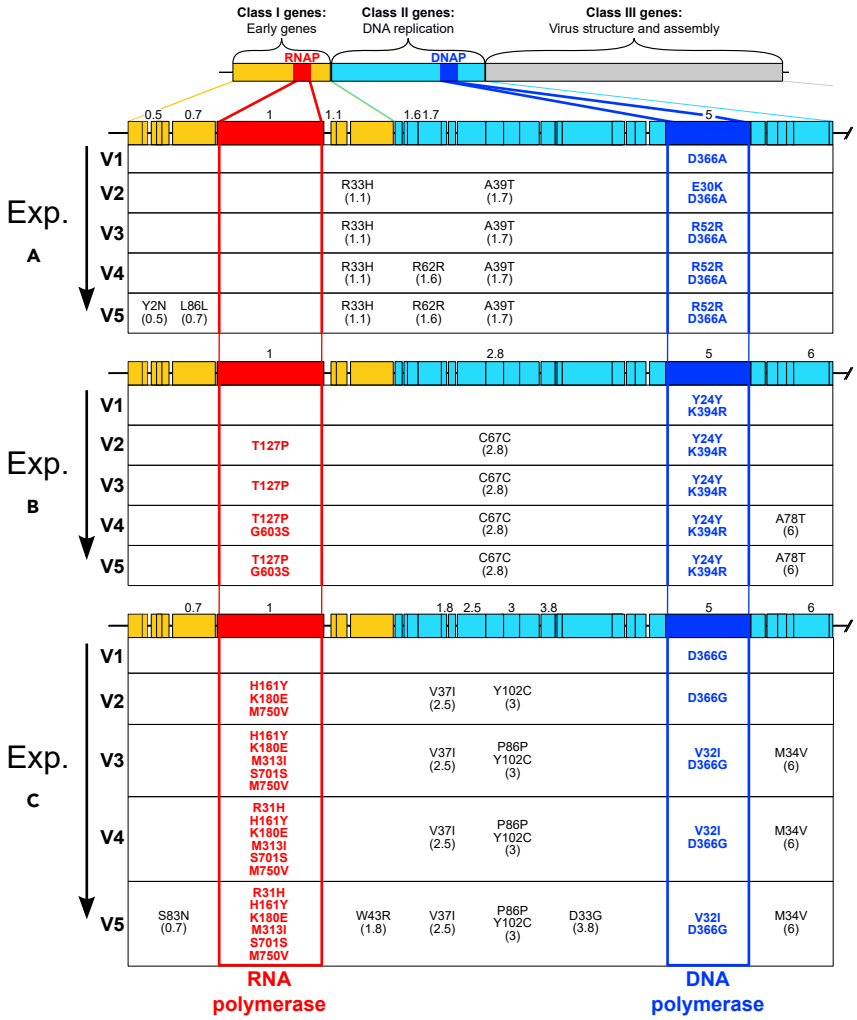


Figure 9. Sequence changes in the viral genome during laboratory evolution

Number of high-frequency (fraction>0.95) mutations for the different viral genes during the evolution experiments of the {A,B,C} set. For the sake of clarity, mutations in class III genes are not shown here, but can be found in Figure S7. The viral DNA and RNA polymerases have been highlighted. For all other mutations, the gene is indicated within brackets. Data were obtained by Illumina sequencing of the DNA extracted after lysing of the engineered host samples (see main text and Figure 7 for details). A control experiment in which the non-evolved virus propagates in the original host yielded no mutations in the class I and class II genes and very few mutations in class III genes (see Figure S6).

Several inferences can be made from these sequencind studies and anlyses: (1) The results obtained for the DNA polymerase are consistent with those derived using Sanger sequencing discussed earlier. That is, the few mutations observed at high frequency do not appear in the TBD (Figure 6B) and cannot directly explain promiscuous thioredoxin recruitment. Certainly, SNVs at low frequencies occasionally occur in the TBD. However, they do so transitorily (see Table S2 for details), whereas any genetic mutation that directly enabled recruitment in the engineered host would have been enriched by selection. Also, no mutations are fixed in the helicase-primase (product of gene gp4), which interacts with the DNA polymerase in the replisome. Certainly, one mutation is fixed in gp2.5 during experiment C (Figure 9), but the role of this protein in the replisome is to coat the ssDNA in the lagging strand (Hamdan and Richardson, 2009). Therefore, mutations in gp2.5 cannot reasonably explain promiscuous thioredoxin recruitment by the DNA polymerase.

(2) The mutations in the viral DNA polymerase increase replication errors. This is easily inferred from the numbers of low-frequency single-nucleotide variants in the virus samples from the {A,B,C} experiments

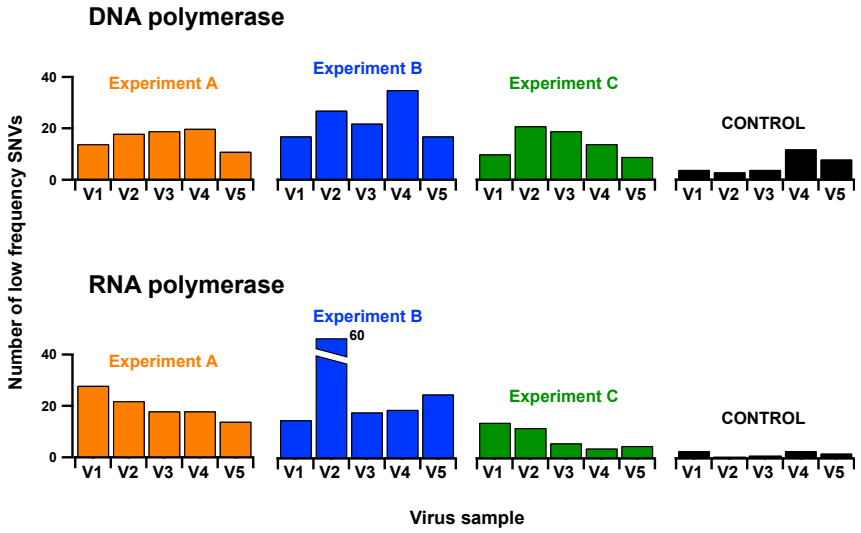


Figure 10. Number of low-frequency SNVs during virus adaptation to the engineered host

The numbers shown for the viral DNA and RNA polymerases were obtained from the next-generation sequencing of the viral genome from the virus samples (V1–V5) at the several evolutionary rounds of the {A,B,C} set of experiments. Only total numbers of low-frequency SNV's are shown here; see Table S2 for more detailed information. For comparison, we also include values in a control experiment in which the non-evolved phage was allowed to propagate in the original host. For the sake of clarity, the bar for the RNA polymerase from the V2 sample in experiment B has been truncated (the actual value in this case is about 60 SNVs).

as compared with a control experiment in which the non-evolved virus propagates in the control host (note that we exclude high-frequency SNVs from this analysis, because their accumulation could potentially be the result of natural selection fixing mutations with a selective advantage). In fact, the pattern of much lower numbers of SNVs in the control experiments is already apparent for the DNA polymerase and RNA polymerase in Figure 10 (see also Figures S3 and S4), and it is not an artifact related to a lower number of readings in the control, as it is visually apparent in Figure S5. This result is confirmed by a more thorough analysis of the whole viral genome, as described in the following section.

We selected SNVs in the whole viral genome with frequency less than 0.2 as given by the ratio of the number of nucleotides detected as an alternative to the number of nucleotides identical to those in the reference genome in the same genomic position. To exclude variants with very low frequencies that may be due to sequencing errors, only those with a frequency greater than 0.01 were taken into account. Furthermore, to eliminate the possible noise of mutation hotspots outside the coding regions, as intergenic regions, only variants coded within the 60 genes of the virus were included. Before comparing the numbers of SNVs from different virus samples, they were normalized through division by the total number of positions read in each sample, that is, by a value that corresponds to the sample coverage (average number of reads per nucleotide) times the size of the genome. The resulting numbers of normalized alternative reads are shown in Figure 6D for the several virus samples of the {A,B,C} experiments, together with the values for a control experiment in which the non-evolved virus propagates in the original host. Lower numbers are obtained for the control as compared with any of the {A,B,C} experiments, a result which is statistically significant as supported by T Test p values of 8×10^{-4} (samples from experiment A), 10^{-2} (samples from experiment B), and 1.5×10^{-3} (samples form experiment C).

(3) Virus adaptation to the engineered host correlates with mutations in the gene for the viral RNA polymerase. First, large numbers of SNVs, at both high and low frequencies, are determined for the RNA polymerase in samples from the {A,B,C} experiments, whereas very few SNVs are found in the control experiment using the original virus in the original host (Figures 9, 10 and S4). Again, the low number of SNVs in the control is not an artifact due to an insufficient number of readings (Figure S5). Secondly, the C > B > A engineered versus original host adaptation pattern (Figure 8B) correlates with the number of high-frequency mutations observed in the gene for viral RNA polymerase. In particular, the largest number of high-frequency mutations in the viral RNA polymerase occurs in experiment C (Figure 9), where a very efficient

adaptation to the engineered host is observed (Figure 8B). Certainly, experiment C also shows increased number of mutations in class III genes (Figures S6 and S7), but these are involved in virion assembly and host lysis and cannot be connected with replisome assembly in any reasonable way. On the other hand, as we elaborate later (see Discussion), mutations in the RNA polymerase provide a straightforward explanation for the observed virus adaptation, as they may have an effect on replisome assembly linked to the possibility of transcription errors (see Discussion for details).

Assessing the error levels of viral RNA polymerases

We hypothesize (see Discussion for details) that transcription errors play a key role in the virus adaptation reported in this work. We deemed convenient, therefore, to qualitatively test the error level of the RNA polymerase from the last round of experiment C using a known procedure (Goldsmith and Tawfik, 2009) based on the determination of antibiotic resistance. This experiment is briefly described in the following section.

We complemented a JM109 *E. coli* strain with two plasmids, one bearing the viral RNA polymerase gene under the promotor of the *E. coli* RNA polymerase and another one bearing the gene of TEM-1 β-lactamase under the promotor of the viral RNA polymerase gene. Expression of the viral RNA polymerase is induced by arabinose, and that of β-lactamase is induced by IPTG. β-lactamase degrades β-lactam antibiotics, and, in principle, its expression would make *E. coli* resistant to ampicillin. However, in the cells we are using for these experiments, the β-lactamase gene is transcribed by the viral RNA polymerase, and, therefore, antibiotic resistance will be compromised to some extent by the error-prone nature of the RNA polymerase, as a substantial number of the phenotypic mutations caused by transcription errors will be disruptive and compromise the proper folding of β-lactamase to yield an active enzyme (Goldsmith and Tawfik, 2009). We performed experiments in which cell growth was monitored in solution upon 2-fold serial dilutions of ampicillin and determined values of inhibitory ampicillin concentration (IC: maximum ampicillin concentration at which growth is observed) for both wild-type T7 RNA polymerase and the variant with the 5 non-silent mutations from the last round of experiment C (Figure 11). In the absence of arabinose induction, the viral RNA polymerase is not synthesized, the β-lactamase gene is not transcribed, and, therefore, the IC value is very low and essentially identical to the value obtained using *E. coli* cells not-complemented with the plasmid bearing the β-lactamase gene. Upon arabinose (and IPTG) induction, RNA polymerase is synthesized, the β-lactamase gene is transcribed, and substantial resistance to ampicillin emerges. However, the IC values are about 16-fold higher with the wild-type viral RNA polymerase as compared with the variant from evolution experiment C. This supports that the mutations fixed in the viral RNA polymerase during evolution experiment C do increase its transcription error rate. Interestingly, western blotting analysis (Figures 11B) suggest a somewhat higher expression level for β-lactamase when it is transcribed by RNA polymerase from the evolved virus. One possibility then is that the mutations accumulated in the viral RNA polymerase during experiment C bring about a transcription process that is both faster and more error-prone.

DISCUSSION

Key results from our phage evolution experiments

We challenged bacteriophage T7 to propagate in a host that had been modified to hinder the recruitment of a protein that is essential for the assembly of a functional viral replisome. The modification targets a very specific virus-host interaction involving thioredoxin (a proviral factor) and the thioredoxin binding domain (TBD) of the viral DNA polymerase. Our laboratory evolution experiments yield some surprising results. First, the virus adapted to the engineered host (i.e., it evolved to exploit the modified proviral factor) without losing its competence to propagate in the original host (i.e., it retained the capability to recruit the unmodified proviral factor). Secondly, genetic mutations cannot explain the observed adaptation. This is, of course, shown by the fact that no genetic mutations were found in the TBD domain, but it is also apparent from the observed pattern of promiscuity. Because the interaction between thioredoxin and the TBD domain of the DNA polymerase is highly specific, genetic mutations that could enable the recruitment of the modified proviral factor would likely impair the recruitment of *E. coli* thioredoxin, causing a strong trade-off, which is not observed in our evolution experiments.

Comparison with previous reports of promiscuity in virus adaptation

It is interesting to compare our results with those recently obtained by Petrie et al. (2018) in a laboratory evolution study on the adaptation of bacteriophage λ to a new receptor. Promiscuity (here the capability

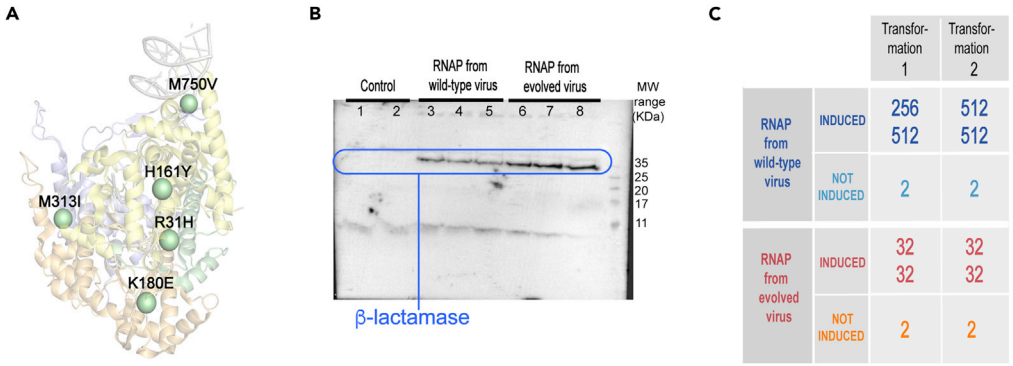


Figure 11. Assessment of the error level of viral RNA polymerases

(A) 3D-structure of the viral RNA polymerase showing the positions at which non-silent mutations were fixed during evolution experiment C (see text and Figure 9 for details).

(B) Western blotting analysis of the expression of the antibiotic resistance enzyme β-lactamase when its gene is transcribed by viral RNA polymerases, both from the wild-type virus and from the evolved virus from experiment C (see text and Figure 9 for details). Three independent experiments were run with each RNA polymerase. "Control" refers to two independent experiments in which the RNA polymerase used (wild type) was not induced and, therefore, the β-lactamase gene was not transcribed.

(C) Resistance to the antibiotic ampicillin of *E. coli* strains in which the gene of the antibiotic resistance enzyme β-lactamase is transcribed by viral RNA polymerases. Values of the inhibitory ampicillin concentration (IC: maximum ampicillin concentration at which growth is observed) are given for independent experiments starting with different clones from two transformations. IC values (μg/mL) are very low when the RNA polymerases are not induced (no arabinose added), and the gene of the β-lactamase is not transcribed. Upon induction of the polymerases, antibiotic resistance is observed, but lower IC values observed with the RNA polymerase from the evolved virus (from experiment C) support an enhanced error level (see text for details). Two-fold serial dilutions of ampicillin were used in these experiments.

to exploit both the new and the old receptors) was also found in this study and attributed to genetic mutations that induced an ensemble of different protein conformations and generated a diversity of interaction capabilities (Petrie et al., 2018). However, it is difficult to see how this mechanism could apply to our system, because no mutations at the level of DNA are fixed in the TBD domain of the viral DNA polymerase. Certainly, conformational dynamics is known to be subject to long-distance effects (Petrovic et al., 2018), and we could consider the possibility that distant mutations in the DNA polymerase caused conformational diversity in the TBD. Still, we have performed sequence determinations for a substantial number (17) of evolution experiments and have found typically 1–3 mutations per experiment, spreading overall over 21 positions in the structure of the DNA polymerase (Figure 6D). It would be unrealistic to assume that mutations at so many different positions show the same capability to trigger, through long-distance effects, the conformational heterogeneity in the TBD that it is specifically required to enable promiscuous recruitment.

The role of transcription errors on phage adaptation to the engineered host

Once genetic mutations in the DNA polymerase are ruled out as responsible for the observed virus adaptation, the only remaining possibility is that mutations at the phenotypic level caused by protein synthesis errors enable promiscuous thioredoxin recruitment. This possibility is reasonable, because viral replisome assembly likely needs only occur a few times per host cell to allow virus propagation and can therefore be mediated by protein variants present at very low level. Phenotypic mutations can be due to translation errors or transcription errors. Because viruses, however, do not encode a translation machinery, the obvious suspect is the viral RNA polymerase. This enzyme (Cheetham and Steitz, 2000) transcribes the viral genes involved in virion assembly and host cell lysis (class III genes), as well as those involved in DNA replication (class II genes), including the gene for the DNA polymerase. Directed evolution studies (Brakmann and Grzeszik, 2001) indicate that, under the appropriate pressure, the T7 RNA polymerase can evolve toward increased error rates by accepting mutations in different regions of the molecule. Indeed, our experiments shown in Figure 11, in which the gene for the antibiotic resistance enzyme β-lactamase gene is transcribed by the RNA polymerase, support that the mutations fixed in the viral RNA polymerase during evolution experiment C do increase its error rate. Overall, therefore, mutations in the gene for the viral RNA polymerase could increase transcription error rates and lead to phenotypic mutations, some of which would occur at the TBD/thioredoxin interaction region and enable recruitment. Initially, the adaptation would rely on

variants of the RNA polymerase present at low level (likely the situation of experiment A), although eventually selection will lead to fixation of mutations, as it is already observed in experiments B and C (Figure 9).

A mechanism of phage adaptation to the engineered host based on phenotypic mutations

To summarize, our results are consistent with a mechanism of adaptation to the engineered host that involves the following steps: (1) increased replication rates are brought about by mutations in the viral DNA polymerase; (2) replication errors promote genetic mutations throughout the viral genome including the gene of the viral RNA polymerase; (3) increased transcription errors cause phenotypic mutations that enable the interaction between the DNA polymerase and thioredoxin to be established in the engineered host. This mechanism provides a simple and convincing explanation for the capability of the evolved phage to propagate in both the original and the engineered hosts. Focusing for illustration on the DNA polymerase, transcription errors will generate a population of molecules with different sets of phenotypic mutations. Some of these molecules will still recruit the original thioredoxin, whereas some other will have the capability to recruit the alternative LPBCA thioredoxin. Note that the putative ancestral protein we have used displays higher stability (Perez-Jimenez et al., 2011) and better folding kinetics properties (Gamiz-Arco et al., 2019) than its modern *E. coli* counterpart, which should contribute to somewhat higher levels of folded protein *in vivo*. This explains that lysis times for the engineered host eventually become somewhat smaller than the lysis times for the original host in experiment C (Figure 8B).

Of course, it may seem at first surprising that, according to our proposed mechanism, phenotypic mutations enable virus replication in the engineered host while they compromise antibiotic resistance in the experiment described in Figures 11C in which the gene for the antibiotic resistance enzyme β-lactamase gene is transcribed by the RNA polymerase from the evolved virus. However, it is important to note that the required numbers of functional enzymes greatly differ between the two scenarios. Antibiotic resistance requires a substantial amount of active β-lactamase and will therefore be compromised by phenotypic mutations which, in most cases, impair proper folding to yield an active enzyme. On the other hand, virus infection of a cell typically results in the generation of a comparatively small number of virions, and replication, therefore, could be achieved on the basis of just a few efficient replisomes. In this scenario, phenotypic mutations could certainly lead to misfolded proteins but also to a small, but sufficient, number of variants capable to form efficient replisomes and enable replication.

General implications of this work

The experimental studies reported here make use of a host that has been engineered to display a very specific difference with the original host and also to promote phenotypic mutations at the targeted host-virus interaction. Still, we argue that phenotypic mutations caused by transcription errors may provide a general mechanism for virus adaptation. Virus infection of a cell typically results in the generation of a not too large number of new virions. It follows that many of the key intermolecular interactions involved in virus infection and replication need only occur a few times per host cell and could be mediated by very low levels of protein variants with enabling phenotypic mutations. To illustrate the plausible generality of the proposed adaptation mechanism, several potential examples are briefly discussed later.

Virus cross-species transmission requires that the virus has the remarkable capacity to establish key interactions in the two different molecular environments of the old and the new host. Diversity at the phenotypic level caused by transcription errors may generate an extensive capability to establish/avoid interactions in different molecular environments, thus facilitating cross-species transmission. The adaptation mechanism based on phenotypic mutations may be immediately implemented in RNA viruses in which both replication and transcription are performed by error-prone RNA-dependent RNA polymerases. For instance, RNA synthesis in coronaviruses produces, not only genomic RNA but also subgenomic RNAs that are not encapsulated in the assembled virions but that function as mRNAs for downstream genes (Fehr and Perlman, 2015; de Wilde et al., 2018), potentially leading to "useful" (for the virus) phenotypic diversity in other structural proteins including the spike (of which there are many copies in the virion surface). Furthermore, transcription errors may conceivably lead to viral protein variants capable of evading antiviral strategies. Phenotypic mutations might, for instance, allow the evasion of antibody neutralization, a phenomenon that plausibly contributes, for instance, to the so-called influenza puzzle (Ellebedy and Ahmed, 2016), i.e., the fact that influenza remains a health problem despite repeated exposure of the population worldwide to natural infection and to influenza viral proteins through vaccination. Another intriguing possibility is that phenotypic mutations contribute to the very low genetic barriers sometimes observed for resistance toward

antivirals (Pennings, 2012; Irwin et al., 2016) by complementing the effect of the few mutations that appear at the genetic level. Obviously, these suggested possibilities rely on the notion that protein variants with suitable phenotypic mutations and present at very low concentrations may enable key processes for virus infection and replication. The interest in determining low-level mutations in viral mRNAs thus emerges. This task is compromised by normal errors of next-generation sequencing and the errors introduced in reverse transcription steps. Still, methods to circumvent these problems have been developed in recent years (Lu et al., 2020).

Limitations of the study

We propose that phenotypic mutations play a role in phage adaptation to the engineered host mainly because we find the obvious alternative explanations to be highly unlikely or impossible in the light of the reported experimental results. The proposal is certainly supported by the mutational pattern we find for the viral RNA polymerase. Yet, the molecular mechanisms of virus replication are not fully understood. Therefore, we cannot be absolutely certain that our analyses have ruled out all possible explanations that rely upon a role for genetic mutations. Consequently, it is conceivable (although we do not think it is likely) that our proposal is an instance of the so-called Holmesian fallacy. Work currently in progress in our laboratory aims at deriving more direct evidence for the role of phenotypic mutations in virus adaptation.

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jose M. Sanchez-Ruiz (sanchezr@ugr.es).

Materials availability

All strains generated in this study are available upon request. This study did not generate new unique reagents.

Data and code availability

Additional data that support the findings of this study are available from the lead contact upon reasonable request. Raw sequencing data are available in the Sequence Read Archive (SRA) under the PRJNA656432 BioProject accession number (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA656432).

METHODS

All methods can be found in the accompanying transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102257.

ACKNOWLEDGMENTS

This work was supported by Spanish Ministry of Economy and Competitiveness/FEDER Funds Grant RTI2018-097142-B-100 and by Human Frontier Science Program Grant RGP0041/2017. Viral genome library preparation and Illumina sequencing were carried out at the IPBLN Genomics Facility (CSIC, Granada, Spain), and the assistance of Dr. Alicia Barroso del Jesus is gratefully acknowledged. We also thank Dr. Jon Beckwith and Dr. Dana Boyd (Harvard University) for kindly providing knockout strains used in this work.

AUTHORS CONTRIBUTION

Conceptualization, B.I.-M. V.A.R., and J.M.S.-R.; Methodology, A.D., R.L.-H., V.A.R., and E.A.-L.; Experiments, R.L.-H. and V.A.R.; Formal Analyses, R.L.-H., B.I.-M., and V.A.R; Writing—Original Draft, J.M.S.-R; Writing—Review & Editing, all authors; Supervision and Funding Acquisition, J.M.S.-R.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 9, 2020
Revised: January 22, 2021
Accepted: February 26, 2021
Published: April 23, 2021

REFERENCES

Akabayov, B., Akabayov, S.R., Lee, S.J., Tabor, S., Kulczyk, A.W., and Richardson, C.C. (2010). Conformational dynamics of bacteriophage T7 DNA polymerase and its processivity factor, *Escherichia coli* thioredoxin. Proc. Natl. Acad. Sci. U S A *107*, 15033–15038.

Brakmann, S., and Grzeszik, S. (2001). An error-prone T7 RNA polymerase mutant generated by directed evolution. ChemBioChem *2*, 212–219.

Cheetham, G.M., and Steitz, T.A. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. *10*, 117–123.

de Wilde, A.H., Snijder, E.J., Kikkkert, M., and van Hemert, M.J. (2018). Host factors in coronavirus replication. Curr. Top. Microbiol. *419*, 1–42.

Delgado, A., Arco, R., Ibarra-Molero, B., and Sanchez-Ruiz, J.M. (2017). Using resurrected ancestral proviral proteins to engineer virus resistance. Cell Rep. *19*, 1247–1256.

Drummond, A.D., and Wilke, C.O. (2009). The evolutionary consequences of erroneous protein synthesis. Nat. Rev. Genet. *10*, 715–724.

Ellebedy, A.H., and Ahmed, R. (2016). Antiviral vaccines: challenges and advances. In Part VI of "The Vaccine Book", second edition, *15*, B.R. Bloom and P.H. Lambert, eds (Academic Press), pp. 283–310.

Etson, C.M., Hamdan, S.M., Richardson, C.C., and van Oijen, A.M. (2010). Thioredoxin suppresses microscopic hopping of T7 DNA polymerase on duplex DNA. Proc. Natl. Acad. Sci. U S A *107*, 1900–1905.

Fehr, A.R., and Perlman, S. (2015). Coronaviruses: an overview of their replication and pathogenesis. Coronaviruses *1282*, 1–23.

Gamiz-Arco, G., Risso, V.A., Candel, A.M., Ingles-Prieto, A., Romero-Romero, M.L., Gaucher, E.A., Gavira, J.A., Ibarra-Molero, B., and Sanchez-Ruiz, J.M. (2019). Non-conservation of folding rates in the thioredoxin family reveals degradation of ancestral unassisted-folding. Biochem. J. *476*, 3631–3647.

Goldsmith, M., and Tawfik, D.S. (2009). Potential role of phenotypic mutations in the evolution of protein expression and stability. Proc. Natl. Acad. Sci. U S A *106*, 6197–6202.

Hamdan, S.M., and Richardson, C.C. (2009). Motors, switches and contacts in the replisome. Annu. Rev. Biochem. *78*, 205–243.

Ingles-Prieto, A., Ibarra-Molero, B., Delgado-Delgado, A., Perez-Jimenez, R., Fernandez, J.M., Gaucher, E.A., Sanchez-Ruiz, E.A., and Gavira, J.A. (2013). Conservation of protein structure over four billion years. Structure *21*, 1690–1697.

Irwin, K.K., Renzette, N., Kowalik, T.F., and Jensen, J.D. (2016). Antiviral drug resistance as an adaptive process. Virus Evol. *2*, vew014.

Lee, S.J., and Richardson, C.C. (2011). Choreography of bacteriophage T7 DNA replication. Curr. Opin. Chem. Biol. *15*, 580–586.

Lu, I.-N., Muller, C.P., and He, F.Q. (2020). Applying next-generation sequencing to unravel the mutational landscape in viral quasispecies. Virus Res. *283*, 197963.

Pennings, P.S. (2012). Standing genetic variation and the evolution of drug resistance in HIV. PLoS Comput. Biol. *8*, e1002527.

Perez-Jimenez, R., Ingles-Prieto, A., Zao, Z.-M., Sanchez-Romero, I., Alegre-Cebollada, J., Kosuri, P., Garcia-Manyes, S., Kappock, T.J., Tanokura, M., Holmgren, A., et al. (2011). Single-molecule paleoenzymology probes the chemistry of resurrected enzymes. Nat. Struct. Mol. Biol. *18*, 592–596.

Petrie, K.L., Palmer, N.D., Johnson, D.T., Medina, S.J., Yan, S.J., Li, V., Burmeister, A.R., and Meyer, J.R. (2018). Destabilizing mutations encode nongenetic variation that drives evolutionary innovation. Science *359*, 1542–1545.

Petrovic, D., Risso, V.A., Kamerlin, S.C.L., and Sanchez-Ruiz, J.M. (2018). Conformational dynamics and enzyme evolution. J. R. Soc. Interface *15*, 20180330.

Traverse, C.C., and Ochman, H. (2016). Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. Proc. Natl. Acad. Sci. U S A *113*, 3311–3316.

Whitehead, D.J., Wilke, C.O., Vernazobres, D., and Bornberg-Bauer, E. (2008). The look-ahead effect of phenotypic mutations. Biol. Direct *3*, 18.

iScience, Volume *24*

Supplemental information

Evidence for a role of phenotypic mutations in virus adaptation

Raquel Luzon-Hidalgo, Valeria A. Risso, Asuncion Delgado, Eduardo Andrés-León, Beatriz Ibarra-Molero, and Jose M. Sanchez-Ruiz

SUPPLEMENTAL FIGURES:

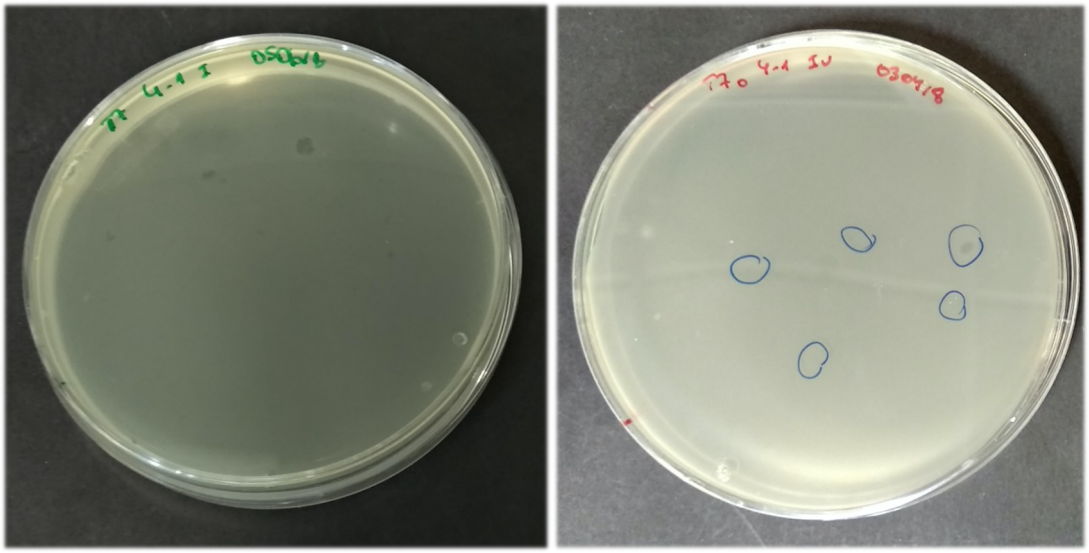
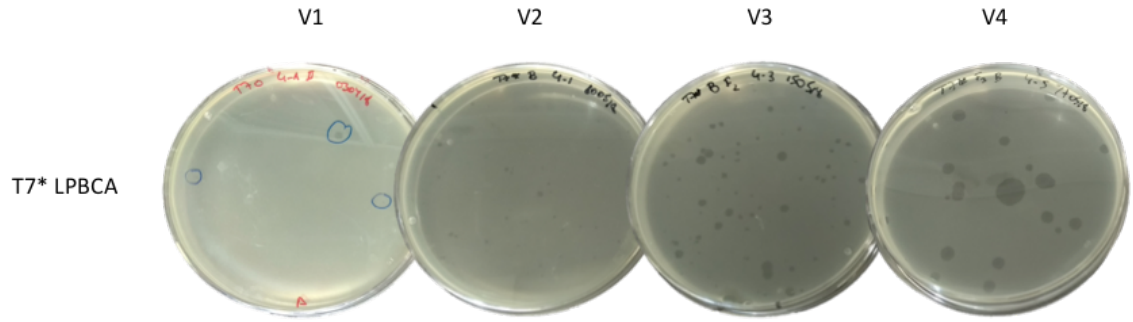


Fig S1. Pictures of representative plaques corresponding to the initial infection of the engineered *E. coli* strain that contains the alternative LPBCA thioredoxin (the engineered host). Related to Figure 3.



EXPERIMENT 2 (SANGER)

Fig. S2. Pictures showing a representative increase in the size of plaques along the evolution experiment in which the virus adapted to an engineered host containing the alternative LPBCA thioredoxin (engineered host). Related to Figures 4 and 5.

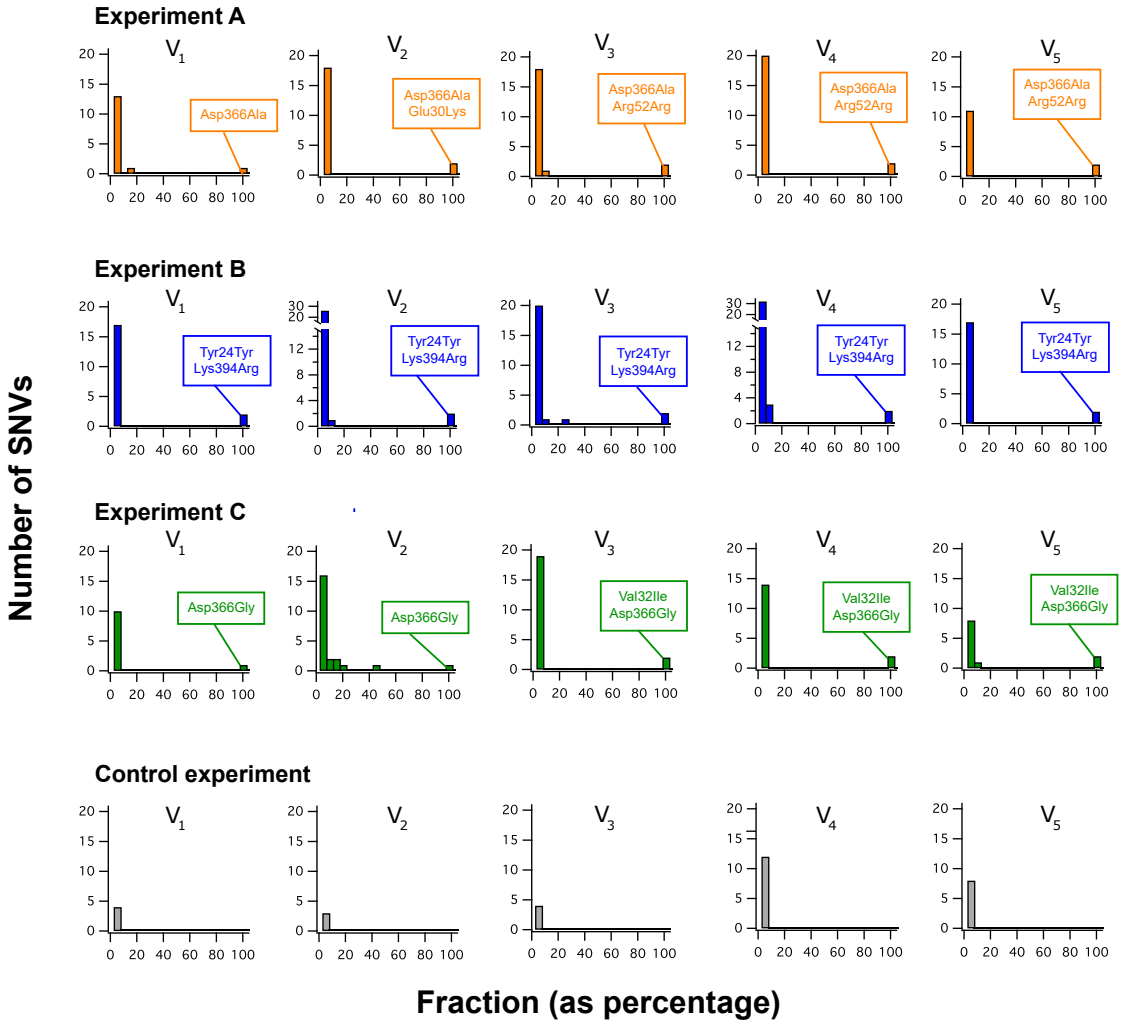


Fig. S3. Sequence changes in the viral DNA polymerase during laboratory virus evolution. Related to Figures 7, 8, 9 and 10. Single nucleotide variants (SNVs) for the gene of the viral DNA polymerase for virus samples from the three evolution experiments of Figure 8. Data were obtained by Illumina sequencing of the DNA extracted after lysis of the engineered host samples. Mutations are binned according to its frequency and the plots show the number of mutations for each 0.05 bin. The mutations that are "fixed" (fraction of occurrence between 0.95 and 1) have been identified. One silent mutation is included because, perhaps, it cannot be ruled out that it could have some effect at the level of RNA structure that translates into some effects on the structure resulting from cotranslational folding. Data for a control experiment corresponding to the propagation of the original virus in the original *E. coli* host is also included.

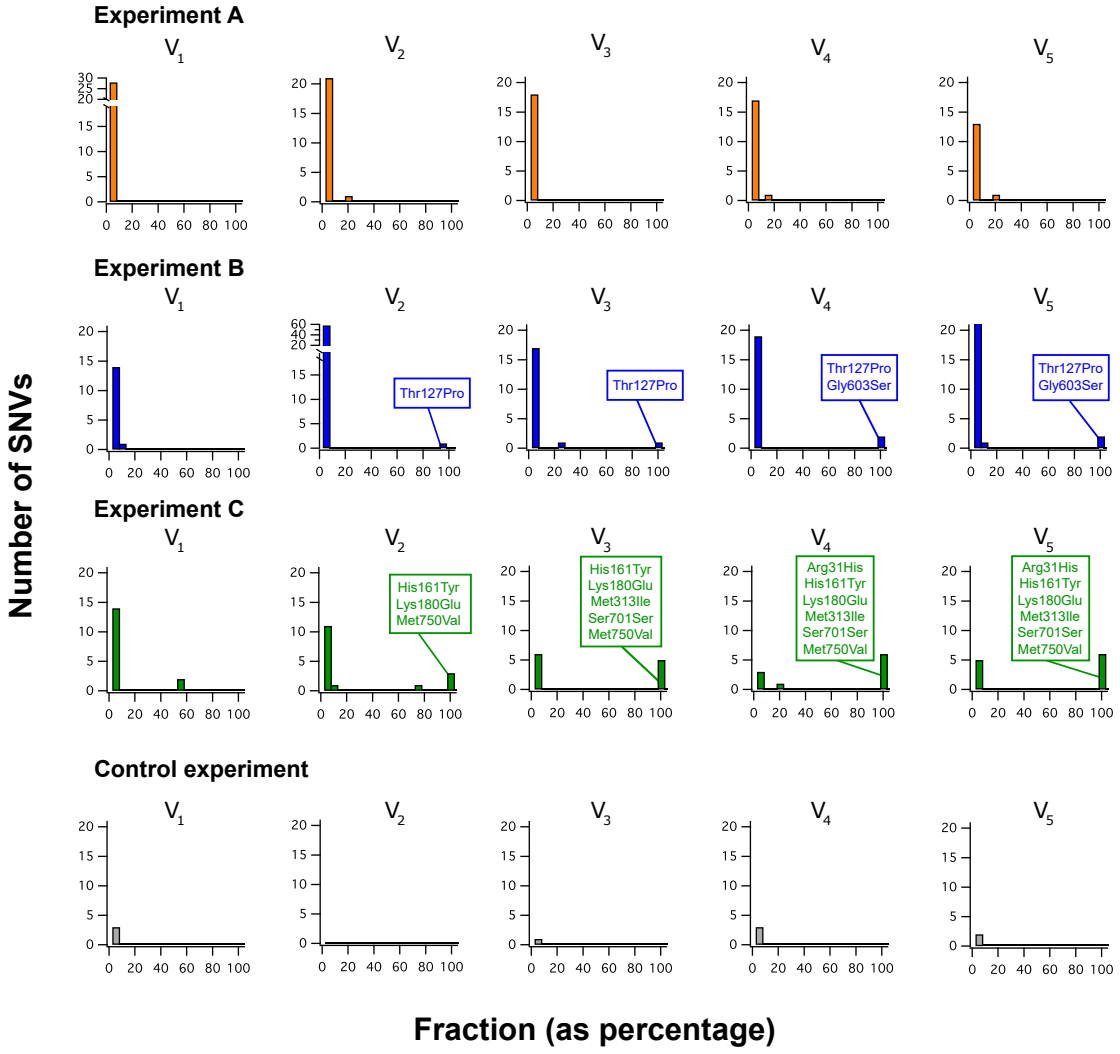


Fig. S4. Sequence changes in the viral RNA polymerase during laboratory virus evolution. Related to Figures 7, 8, 9 and 10. Single nucleotide variants (SNVs) for the gene of the viral RNA polymerase for virus samples from the three evolution experiments of Figure 8. Data were obtained by Illumina sequencing of the DNA extracted after lysis of the engineered host samples. Mutations are binned according to its frequency and the plots show the number of mutations for each 0.05 bin. The mutations that are "fixed" (fraction of occurrence between 0.95 and 1) have been identified. Data for a control experiment corresponding to the propagation of the original virus in the original *E. coli* host is also included.

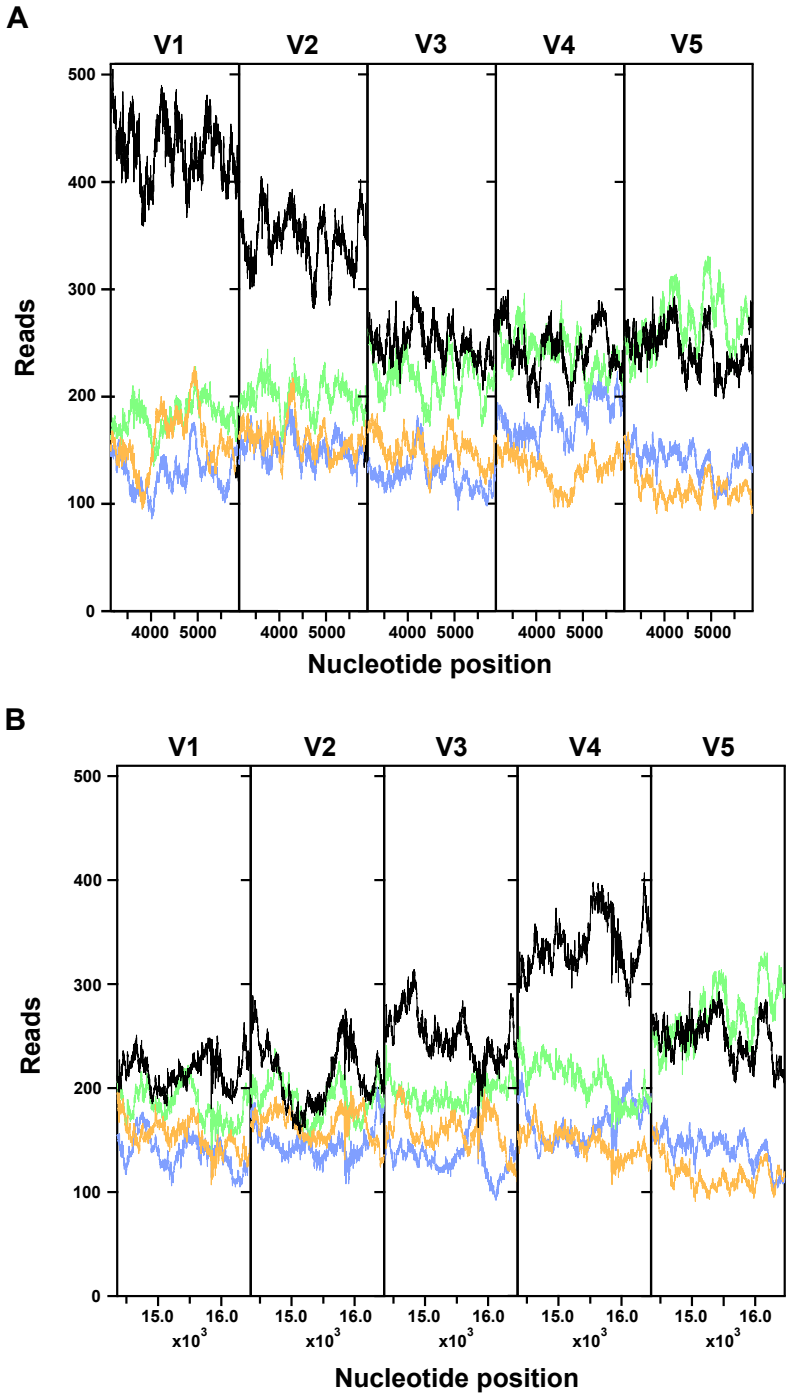


Fig. S5 Total number of readings for the genes of the viral RNA (A) and DNA (B) polymerases determined from Illumina sequencing of samples from the different rounds of the evolution experiment of the {A,B,C} set. Related to Figures 8, 9 and 10. Data for control experiments mimicking the evolution experiment but using the original virus (V_0) in the original host are also shown.

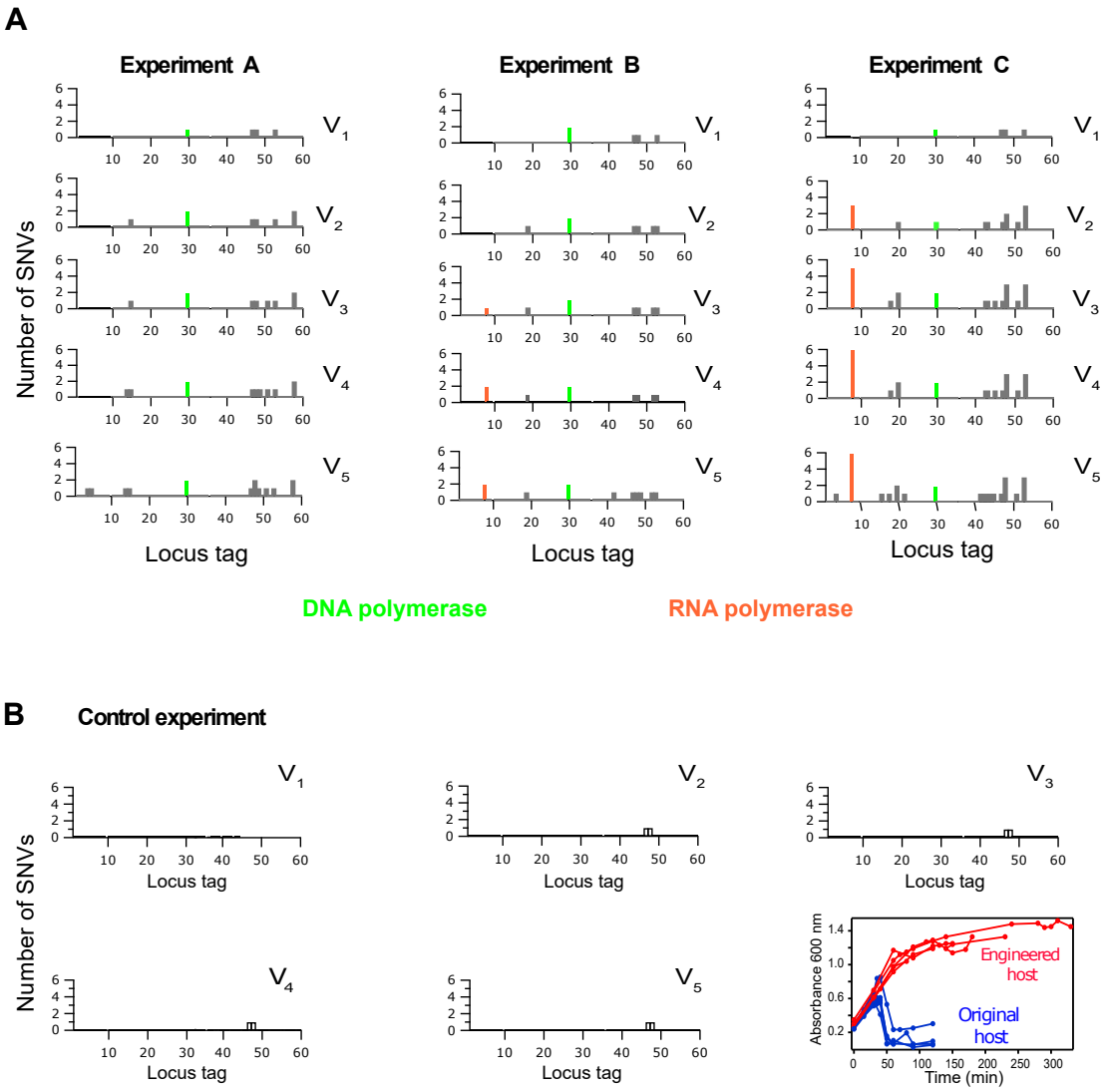


Fig. S6. Sequence changes in the viral genome during laboratory evolution. Related to Figures 8 and 9. A, Number of high frequency (fraction>0.95) mutations for the different viral genes during the evolution experiments of the {A,B,C} set. The viral DNA and RNA polymerases have been highlighted. Data were obtained by Illumina sequencing of the DNA extracted after lysing of the engineered host samples. B, Same as in A, but for a control experiment in which the non-evolved phage propagates in the original host. Black is used here to highlight that this is a control experiment. A panel with lysis curves for this control is included. No selection for propagation in the engineered host is applied in this control experiment and the phage does not evolve the capability to lyse the engineered host.

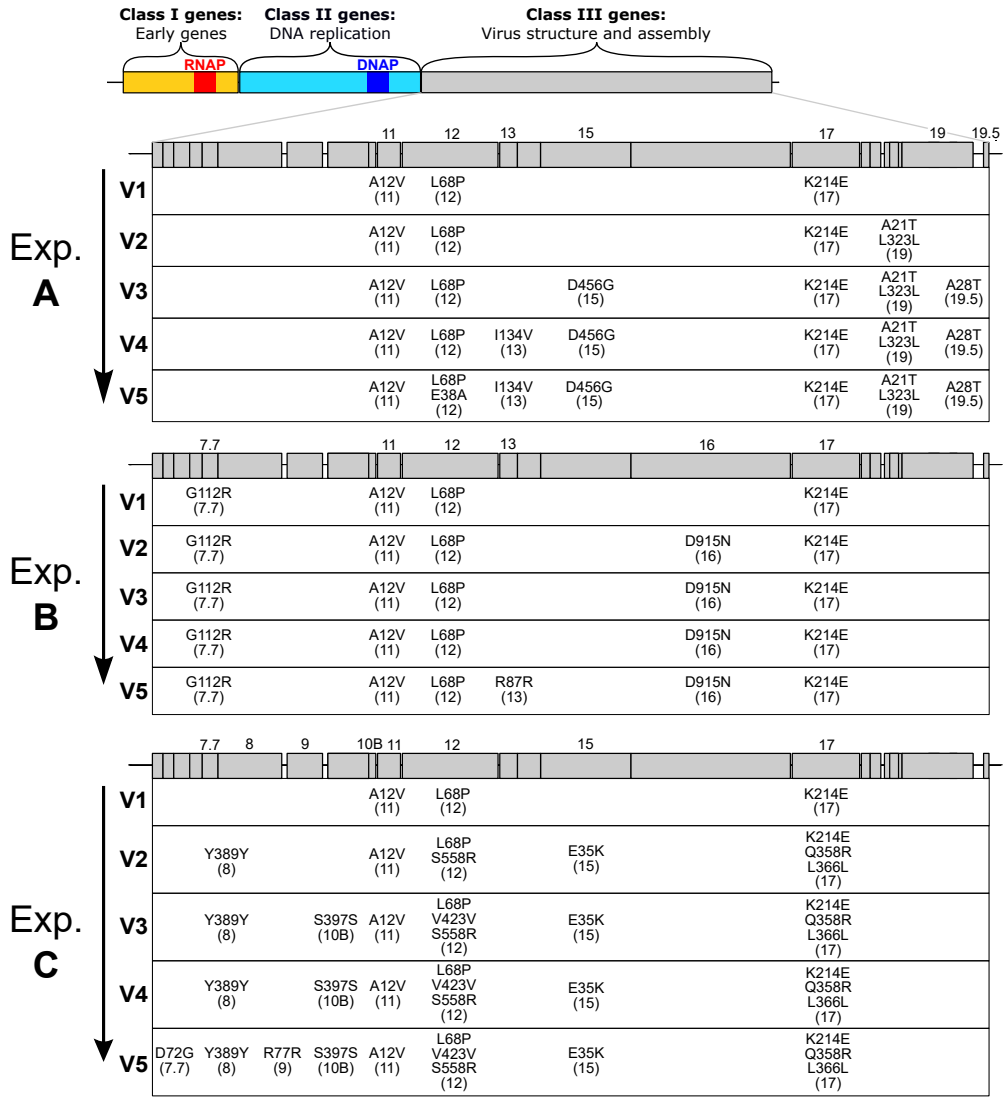


Fig. S7. Sequence changes in the viral genome during laboratory evolution. Related to Figure 7 and 9. Number of high frequency (fraction>0.95) mutations for the different viral genes during the evolution experiments of the {A,B,C} set. Only mutations in class III genes are shown here. See Figure 9 in the main text for mutations in class I and class II genes. Data were obtained by Illumina sequencing of the DNA extracted after lysing of the engineered host samples (see main text and Figure 7 for details). A control experiment in which the non-evolved virus propagates in the original host yielded very few mutations in class III genes (see Figure S6).

SUPPLEMENTAL TABLES:

Table S1. Mutations in the viral DNA polymerase determined from PCR and Sanger sequencing on samples from the early rounds of 14 evolution experiments. Related to Figures 5 and 6.

Experiment	POS	NT_CHANGE	AA_CHANGE	V1	V2	V3
1	15533	1181A>G	Lys394Arg	X	X	ND
2	14473	121G>T	Ala41Ser	X	ND	X
2	15652	1300T>C	Phe434Leu	-	ND	X
3	14398	46A>G	Thr16Ala	X	ND	X
4	14485	133G>A	Ala45Thr	X	ND	X
5	14463	111T>C	Ser37Ser	X	ND	X
6	14382	30C>T	Ala10Ala	X	ND	X
6	15445	1093G>A	Asp365Asn	-	ND	X
6	15578	1226A>G	Tyr409Cys	-	ND	X
7	14434	82A>G	Thr28Ala	X	ND	X
8	14495	143C>T	Ala48Val	X	X	ND
8	15512	1160A>G	Lys387Arg	X	X	ND
9	14485	133G>A	Ala45Thr	X	X	ND
9	15512	1160A>G	Lys387Arg	X	X	ND
10	14434	82A>G	Thr28Ala	X	-	ND
11	14359	7G>A	Val3Ile	X	X	ND
11	15533	1181A>G	Lys394Arg	X	X	ND
12	14412	60C>T	Cys20Cys	X	X	ND
12	15380	1028A>G	Gln343Arg	X	X	ND
13	14494	142G>A	Ala48Thr	X	X	ND
13	15512	1160A>G	Lys387Arg	X	X	ND
14	14463	111T>C	Ser37Ser	X	X	ND
14	15449	1097A>C	Asp366Ala	X	X	ND

TRANSPARENT METHODS

Strains used in this work

We use as the original host the *E. coli* strain DHB4 (Delgado et al., 2017). The engineered host is based upon a receptor strain FA41 which is a DHB4 strain deficient in the two thioredoxins identified in the cytoplasm of *E. coli*: thioredoxin 1, which we refer throughout the text simply as *E. coli* thioredoxin and which is the thioredoxin recruited by the non-evolved virus for its replisome, and thioredoxin 2, which is induced under some stress conditions and has an additional zinc-binding domain. This DHB4 *trxA trxC* strain was a gift from Jon Beckwith (Harvard Medical School). We further performed the curation of the F' factor (which would prevent phage infection) and lysogenization (using the λDE3 Lysogenization kit from Novagen) to introduce in the bacterial chromosome the gene of the T7 RNA polymerase (Delgado et al., 2017). This results in the *E. coli* Trx^- strain that has been used as a control throughout this work. The gene for LPBCA thioredoxin was cloned in pET30a(+) under a T7 RNA polymerase promoter (Delgado et al. 2017). This system is leaky and leads to a basal expression level, even under non-inducing conditions. Complementation of the *E. coli* Trx^- strain with this plasmid produced the engineered host extensively studied in this work. The gene for *E. coli* thioredoxin was also cloned in pET30a(+) and complementation of the *E. coli* Trx^- strain with this plasmid produced the version of the original host used in this work.

Plaque assays

Plaque assays were performed as follows (Delgado et al., 2017). Briefly, phage samples were serially diluted in Tris buffer 20 mM pH 7.4 (including 100 mM NaCl, 10 mM $MgSO_4$). Strains were grown to an absorbance of ~0.5 in LB medium at 37 ºC and 100 μL of the cell suspension were mixed with 100 μL of a phage suspension and the resulting mix was combined with 4 mL of molten agar. Plates were incubated at 37 ºC overnight or 15 hours for the experiments A, B and C (Fig. 3). Numbers of plaques were determined from visual inspection and counting. Phage titer was calculated upon serial dilution experiments from the number of plaques at the two highest phage dilutions at which plaques are observed.

Lysis experiments and DNA extraction

Lysis experiments were carried out using a straightforward modification of the protocol we previously used for the determination of generation times (Delgado et al., 2017). Briefly, preinocula in LB were incubated overnight at 37 ºC. Cultures were then diluted 1/200 in fresh medium and the absorbance at 600 nm was determined as function of time for about 4 hours typically while keeping the temperature at 37 ºC. A volume of virus sample was added when the absorbance was 0.25-0.3. After lysis, cultures were collected and kept at -20 ºC until DNA extraction. For this, samples were thawed on ice, mixed with 1/60 chloroform, vortexed and centrifuged at 4000g for 10 minutes. DNA was isolated from the supernatant using the phage DNA isolation kit from Norgen (CAT#46800).

Evolution experiments

The evolution experiments were initiated by a standard plaque assay in which the host engineered strain was infected with an appropriate dilution of wild type phage suspension. Initially, after overnight incubation a clearly well-isolated large plaque was

picked up from the agar plate and transferred to a 1 mL of buffer (Tris buffer 20 mM pH 7.4, 100 mM NaCl, 10 mM $MgSO_4$). Virus particles were allowed to diffuse for 2 hours and titration of the phage suspension was performed. Subsequently, one 100 µl aliquot was used to start the next round of evolution by infection of a fresh culture of the engineered host. An additional amplification step (Delgado et al., 2017) was included in the rounds before the one-month break in the experiment shown in the lower panel of Figure 4.

Genomic DNA sequencing

gDNA samples from ~50 isolates of bacteriophage T7 were extracted with the phage isolation kit from Norgen (CAT#46800) and quantified using a Nanodrop One (ThermoFisher). Subsequently, gDNA was purified using MucleoMag Beads (Omega) without size selection. The quality of the gDNA was evaluated by Qubit dsDNA HS Assay kit (ThermoFisher) and 0.8% agarose gel electrophoresis. 37 libraries were constructed with an input of 100-300 ng of DNA using the NexteraTM DNA Flex Library Preparation kit with only 5 PCR cycles in the indexing step. The quality of the libraries was validated by the Qubit dsDNA HS Assay kit (ThermoFisher) with a 2100 Bioanalyzer (Agilent Technologies). Libraries were sequenced on an Illumina MiSeq producing 1170569 of 2x250 bp read, i.e., 61762 average reads per sample, reaching an average coverage of 369x.

Initially, the sequences for each isolate were subjected to quality evaluation using FASTQC software, which provides an extensive report of the quality of the reads (Andrews, 2010). On the basis of the generated report, we determined a very high drip in the quality value of the last sequenced nucleotide, which was, therefore, eliminated. We also verified that the amount of Ns and adapters included in the sequences was almost negligible. The average quality of all reads was above 30. After this quality filter, we called and annotated all variants on the basis of the following steps performed by Snippy (Seemann, 2015). Burrow Wheeler Alignment (Li and Durbin, 2009) was used to align the reads against the reference genome (GCF_000844825 from GeneBank). Samtools (Li and Durbin, 2009) was used to sort, mark and remove duplicate sequences. Variant calling was performed by Freebayes (Garrison and Marth, 2012) using the following parameters: "-P 0 – C 10 –min-repeat-entropy 1.5 –strict-vcf –q 15 –m 30" which involves the exclusion of reads having a mapping quality of less than 30. Furthermore, alleles were excluded from the analysis if their supporting base quality was less than 15. The remaining variants were processed by snpEff (Cingolani et al., 2012) to annotate and predict their effects on genes and proteins. The single nucleotide variants discussed in the text are the outcome of an additional filtering process by bcftools (Danecek and McCarthy, 2017) to eliminate further possible errors. Variant determination at a given position required at least 10 sequences including the position and a mapping quality of at least 100. Finally, we eliminated all those variants that did not have at least 2 reads supporting the alternative nucleotide or that had a frequency lower than 0.01 (the frequency is calculated by dividing the number of reads that support the alternative by the number of reads whole nucleotide is equal to the reference). Raw sequencing data are available in the Sequence Read Archive (SRA) under the PRJNA656432 BioProject accession number.

Antibiotic-resistance assessment of RNA polymerase error level

We used *E. coli* JM109 competent cells (Agilent), the following plasmids:

pACYC -T7 RNAP-wt, encoding the wild type RNA polymerase under lacI repression and induced by addition of L-arabinose, pACYC -T7 RNAP-variant, encoding the RNA polymerase from the viral V5 sample of experiment C, under lacI repression and induced by addition of L-arabinose, pET30a(+)-TEM-1, encoding the TEM-1 β-lactamase gene under the control of the T7 operon and induced by addition of IPTG, and the following media for bacterial growth, LB medium. 10 g/L Bacto -Tryptone, 5 g/L Bacto -Yeast Extract, 5 g/L NaCl, autoclave to sterilize. Allow the auto-claved medium to cool to 55ºC and add chloramphenicol (35 mg/mL), kanamycine (50 mg/mL) and glucose (1%). For LB plates, 1.5% Bacto-agar (15 g/L) was added prior to autoclave. SOC medium. 2% Tryptone (bacto), 0.5% Yeast Extract, 10 mM NaCl, 2.5 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose.

Transformation of plasmid DNA into *E. coli* using the heat shock method was performed using well-established protocols. Briefly, after a short incubation in ice (30 min), a mixture of chemically competent bacteria and DNA is placed at 42°C for 45 seconds (heat shock) and then placed back in ice. SOC media is added and the transformed cells are incubated at 37°C for 1 h with agitation. Cells were plated on LB agar plates with chloramphenicol (35 μg/mL), kanamycine (50 μg/mL) and glucose (1%) and grown without selection (O/N, 37 °C). Colonies carrying a TEM-1 and T7 RNAP-wt or TEM-1 and T7 RNAP-variant were picked and individually grown in 5 mL of "Liquid growth media" [LB medium with chloramphenicol (35 μg/mL), kanamycine (50 μg/mL) and glucose (0.2%)] for 16 h at 37 °C with shaking.

To test ampicillin resistance, cultures were diluted 1/100 in fresh ''Liquid growth media" and strains were grown to an absorbance of ~0.2 add L-arabinose to be concentration final 0.2%, and culture at 37ºC for 5 h. Then, β-lactamase expression was induced by addition of 0.5 mM IPTG for 2 h (37 °C, shaking). Finally, after de induction the cultures were diluted (1:100) in fresh "Liquid growth media" with IPTG 0.5 mM and varying concentrations of ampicillin (0, 2, 4, 8, 16, 32, 64,128, 256, 512, 1024, 2048 μg/mL). Culture were grown overnight at 37 °C and the inhibitory concentration for TEM-1 were determined from visual inspection. The inhibitory concentration of each transformant, were assayed in duplicate with two independent transformations.

To assess the expression levels of TEM-1 β-lactamase when its gene is transcribed by the viral RNA polymerases, cultures (prepared as described above) were diluted 1/100 in fresh ''Liquid growth media", the strains were grown to an absorbance of ~0.2, L-arabinose to be concentration final 0.2% was added and the cultures were grown at 37ºC for 5 h. β-lactamase expression was then induced by addition of 0.5 mM IPTG. After 2 hours at 37 °C with shaking, the cultures were diluted (1:100) in fresh "Liquid growth media" with IPTG 0.5 mM and grown overnight at 37 °C. 5 ml of bacterial cultures were then centrifuged for 10 min at 4000 rpm at 4ºC. Pellets were resuspended in 2 ml buffer [10 mM potassium phosphate buffer, pH 6.5, 1 mM EDTA, 20% (w/v) sucrose, 1 mg ml^{-1} lysozyme]. Cell lysis was performed by sonication using a Vibra cell ultrasonics processor (1 seg. pulse, 60% amplitude, eight cycles) with cooling. Equal volumes of cell lysis preps were loaded onto a 15% PAGE-gel. The cell lysis preparations were analyzed by Western blotting. Western blots were revealed with monoclonal Anti-TEM-1 β-lactamase antibody produced in mouse (Invitrogen MA1-

10712) and Horseradish Peroxidase (HRP)-conjugated goat anti-mouse antibodies (Santa Cruz Biotechnology). After chemiluminescence (ECL) detection, the membrane was photographed.

SUPPLEMENTAL REFERENCES

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly *6,* 80–92.

Danecek, P., and McCarthy, S.A. (2017). BCFtools/csq: haplotype-aware variant consequences. Bioinformatic*s 33,* 2037–2039.

Delgado, A., Arco, R., Ibarra-Molero, B., and Sanchez-Ruiz, J.M. (2017). Using resurrected ancestral proviral proteins to engineer virus resistance. Cell Reports *19,* 1247-1256.

Garrison, E., and Marth, G. (2012). Haplotype-based detection from short-read sequencing. arXiv:1207.3907.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–60.

Seemann, T. (2015). Snippy-Rapid haploid variant calling and core SNP phylogeny. GitHub. Available at: github.com/tseemann/snippy/.