Evidence for a role of phenotypic mutations in virus adaptation

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

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

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

[†]These authors contributed equally to this work

Abstract

Viruses repurpose the host molecular machinery for their own proliferation, block host antiviral factors and recruit host proteins for processes essential for virus propagation. Cross-species transmission requires that the virus can establish crucial interactions in the two different environments of the new and the old hosts. To explore the molecular mechanisms behind host promiscuity, we challenged a lytic phage to propagate in a host in which a protein essential for the assembly of a functional viral replisome had been modified to hinder its recruitment. The virus adapted to the engineered host without losing the capability to propagate in the original host, but no mutations that could directly explain the recruitment of the modified protein were fixed in the viral DNA genome. Adaptation, however, correlated with mutations in the gene for the viral RNA polymerase, supporting that transcription errors led to phenotypic mutations that contributed to promiscuous recruitment. Some key molecular interactions need only occur a few times per host cell to allow virus replication. Our results then support that such virus-host interactions may be mediated by mutant proteins present at very low concentrations. The possibility arises that phenotypic mutations facilitate cross-species transmission and contribute to evasion of antiviral strategies.

Mistakes in protein synthesis, due to translation and transcription errors, are common and lead to the so-called phenotypic mutations[1,2]. Translation error rates are generally higher than transcription error rates. Yet, transcription errors may have a stronger impact at the protein level, since an mRNA molecule is typically translated many times[3]. 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[1,2]. On the other hand, some phenotypic mutations could provide crucial functional advantages under certain conditions. It has been theorized that phenotypic mutations might play an evolutionary role by allowing organism survival until functionally useful mutations appear at the genetic level[4,5]. The laboratory evolution experiments reported here support that phenotypic mutations may contribute to virus adaptation, a possibility with important potential implications for the mechanisms of cross-species transmission and evasion of antiviral strategies.

Bacteriophage T7 recruits the small (~110 residues) protein thioredoxin from the *E. coli* host to be a part of its four-protein replisome[6]. Upon recruitment, thioredoxin becomes an essential processivity factor for the viral DNA polymerase[7] (Fig. 1a). Binding of *E. 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[6,8,9]. This interaction is extremely tight (dissociation constant ~5 nanomolar) showing that the TBD has evolved to specifically bind the host thioredoxin[6]. It follows that replacing *E. coli* thioredoxin with an alternative thioredoxin could hinder recruitment and potentially prevent phage propagation in *E. coli*[10]. A simple way to perform such replacement involves complementing a knockout *E. coli* Trx^- strain with a plasmid bearing the alternative thioredoxin gene[10]. 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[10]. 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 wt 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[10] and has been used as a representation the original host throughout this work.

As alternative thioredoxin we used a protein, LPBCA thioredoxin, previously obtained and characterized in detail as part of an ancestral reconstruction study[11-14]. 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 (Fig. 1b) and efficient interaction with the thioredoxin binding domain of the viral DNA polymerase is not likely to occur[10]. 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.

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, they are observed only occasionally[10]. It appears then that only a tiny fraction of the virions in the phage sample (~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 (Fig. S1). Here, we have evolved such "anomalous" virions for efficient propagation in the engineered host. Schematically, our evolution experiments can be represented as:

$$V_0 \Longrightarrow V_1 \overset{R_1}{\rightarrow} V_2 \overset{R_2}{\rightarrow} V_3 \overset{R_3}{\rightarrow} V_4 \overset{R_4}{\rightarrow} \cdots\cdots$$

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 ten-fold dilutions and plaque selection (see Methods for details). These assays immediately lead to numbers of virus particles (plaque forming units of 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.

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 3x3 mm^2 were used to start each next round (except for the V_1 sample for which the whole plaque was used, since it was smaller than 3x3 mm^2). Both experiments (Fig. 2) 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 Fig. 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 one month from plates stored at 4 ºC. In one case (experiment in lower panel in Fig. 2), the protocol was changed after re-starting by eliminating an amplification step (see legend of Fig. 2 for details). As expected, these one-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.

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 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. Between 1 and 3 genetic mutations appeared in the DNA polymerase in most experiments (see Fig. 1a 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 (Fig. 1a), clustering in the region of the polymerase domain close to the bound DNA and in the exonuclease domain involved in proofreading. Therefore, although none of the mutations found can reasonably explain thioredoxin recruitment in the engineered host, it is plausible that they increase replication errors, thus promoting mutations in other viral proteins that trigger processes that eventually lead to recruitment.

We next performed 3 additional evolution experiments using a modified protocol that allows next generation sequencing of the phage genome. 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 above, 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. Also, 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.

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 above. Adaptation was thus immediately clear from the general trend towards 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 (Fig. 3a). 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 propagate in the original and the engineered hosts, they do not necessarily do so with the same efficiency. As shown in Fig. 3b, 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 (Fig. 3b). Upon virus adaptation, however, lysis times for the engineered host approach those for the original host, although to an extent that it is variable. Lysis times for the engineered host always remain larger in one experiment (labelled A), while they eventually become somewhat smaller than the lysis times for the original host in another experiment (labelled 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 adaption, the experiments in Fig. 3b can be ranked C>B>A.

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 or readings at the position >0.95) were determined for all viral genes. These results are shown in Fig. 4. Secondly, for genes of particular interest, we considered all SNVs fulfilling these 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. The results are given in Fig. 5 for the viral DNA polymerase and in Fig. 6 for the viral RNA polymerase. In both cases, 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. Several inferences can be made from these sequence analyses:

1) The results obtained for the DNA polymerase are consistent with those derived using Sanger sequencing discussed above. That is, the few mutations observed at high frequency do not appear in the TBD (Fig. 1a) 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), while any genetic mutation that directly enabled recruitment in the engineered host would have been enriched by selection.

2) The mutations in the viral DNA polymerase increase replication errors. This is evident from the comparison with a control experiment using the original virus (V_0) in the original host (Figs. 4, 5 and 6). Note that the smaller number of SNVs in the control experiment at both high and low frequencies is not an artefact related to a lower number of readings in the control, as it is visually apparent in Fig. S3.

3) Virus adaption 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, while very few SNVs are found in the control experiment using the original virus in the original host (Figs. 1c, 4 and 6). Again, the low number of SNVs in the control is not an artefact due to an insufficient number of readings (Fig. S3). Secondly, the C>B>A engineered vs. original host adaptation pattern (Fig. 3b) correlates with the number of high frequency mutations is observed in the gene for viral RNA polymerase. In particular, the largest number of high-frequency mutations in the viral RNA polymerase occur in experiment C (Figs. 4 and 5), where a very efficient adaptation to the engineered host is observed (Fig. 3b). Certainly, experiment C also shows increased number of mutations in class III genes, 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 below, mutations in the RNA polymerase provide a straightforward explanation for the observed virus adaption.

Once genetic mutations 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 to transcription errors. Since viruses, however, do not encode a translation machinery, the obvious suspect is the viral RNA polymerase. This enzyme[15] 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[16] indicate that, under the appropriate pressure, the T7 RNA polymerase can evolve towards increased error rates by accepting mutations in different regions of the molecule. 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.

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 molecules with different sets of phenotypic mutations. Some of these molecules will still recruit the original thioredoxin, while some other will have the capability to recruit the alternative LPBCA thioredoxin. The putative ancestral protein displays higher stability[11] and better folding kinetics properties[14,17] 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.

It is important to note at this point that genetic mutations at the level of DNA could hardly explain the recruitment promiscuity over many rounds of evolution, as we find in this work. Since the interaction between thioredoxin and the TBD domain of the DNA polymerase is highly specific, mutations that enable the recruitment of LPBCA thioredoxin will likely impair the recruitment of *E. coli* thioredoxin, causing a strong trade-off. Certainly, it is conceivable that some mutations at the genetic level could lead to an ensemble of different protein conformations, thus generating a diversity of interaction capabilities[18]. However, it is difficult to see how this mechanism could apply in our system, since no mutations at the level of DNA are fixed in the TBD domain of the viral DNA polymerase. On the other hand, diversity at the phenotypic level should easily bypass the trade-off inherent to a very tight intermolecular interaction.

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, the impressive capability of the phage to adapt to the engineered host quickly and efficiently, supports 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. Several potential examples are briefly discussed below.

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. Certainly, many viruses are quasispecies[19] and diversity at the genetic level may contribute to the required interaction promiscuity. Still, replication error rates generate inheritable genetic diversity which cannot increase above the threshold that leads to the error catastrophe. On the other hand, transcription errors may lead to an additional layer of sequence diversity that, while promoting adaptation, it is renewed, not amplified, by successive replication cycles. This 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[20,21] potentially leading to “useful” (for the virus) phenotypic diversity in the spike and other structural proteins. 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[22], 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 towards antivirals[23,24] by complementing the effect of the few mutations that appear at the genetic level.

Obviously, the possibilities suggested in the preceding paragraph 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[25].

ACKNOWLEDGEMENTS

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 BarrosoDelJesus is gratefully acknowledged. We also thank Dr. John Beckwith and Dr. Dana Boyd (Harvard University) for kindly providing knockout strains used in this work.

REFERENCES

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

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

3. Traverse, C.C. & Ochman, H. Conserved rates and patterns of transcription errors across bacterial growth states and lifestyles. *Proc. Natl. Acad. Sci. USA* **113,** 3311-3316 (2016).

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

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

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

7. Etson, C.M., Hamdan, S.M., Richardson, C.C. & van Oijen, A.M. Thioredoxin supresses microscoping hoping of T7 DNA polymerase on duplex DNA. *Proc. Natl. Acad. Sci. USA* **107,** 1900-1905 (2010).

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

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

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

11. Perez-Jimenez R, Ingles-Prieto, A., Zao, Z.-M., Sanchez-Romero, I., Alegre-Cebolalda J., Kosuri, P., Garcia-Manyes S., Kappock TJ, Tanokura, M., Holmgren A., Sanchez-Ruiz, J.M., Gaucher, E.A. & Fernandez, J.M. *Nat. Struct. Mol. Biol.* **18,** 592-596 (2011).

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

13. Risso, V.A., Manssour-Triedo, F., Delgado-Delgado, A., Arco, R., Barroso-delJesus, A., Ingles-Prieto, A., Godoy-Ruiz, R., Gavira, J.A., Gaucher, E.A., Ibarra-Molero, B. & Sanchez-Ruiz, J.M. Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary hystory. *Mol. Biol. Evol.* **119,** 1323-1333 (2015).

14. 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. & Sanchez-Ruiz, J.M.

Non-conservation of folding rates in the thioredoxin family reveals degradation of ancestral unassisted-folding. *Biochem. J.* **476,** 3631-3647 (2019).

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

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

17. Naganathan, A.N. Molecular origins of folding rate differences in the thioredoxin family. *Biochem. J.* **477,** 1083-1087 (2020).

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

19. Eigen, M., McCaskill, J. & Schuster, P. Molecular quasi-species. *J. Phys. Chem.* **92,** 6881-6891 (1998).

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

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

22. Ellebedy, A.H. & Ahmed, R. Antiviral vaccines: challenges and advances. Chapter 15, pp 283-310, part VI of "The Vaccine Book" (second edition), Barry R. Bloom & Paul-Henri Lambert, eds. Academic Press (2016).

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

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

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

METHODS

Strains used in this work

The *E. coli* strains used in this work have been previously described in detail[10]. Briefly, we use as the original host the *E. coli* strain DHB4. 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). As previously described (Delgado et al., 2017), 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. 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 as previously described (Delgado et al.). 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 described previously in detail (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[10]. 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[10] was included in the rounds before the one-month break in the experiment shown in the lower panel of Fig. 2.

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[26]. The quality of the reads, the number of repeated sequences and the adapter content were above the recommended threshold values. We called and annotated all variants on the basis of the following steps performed by Snippy[27]. Burrow Wheeler Alignment (Li and Durbin, 2009) was used to align the reads against the reference genome (GCF_000844825 from GeneBank). Samtools[28] was used to sort, mark and remove duplicate sequences. Variant calling was performed by Freebayes[29] using the following parameters: "-P 0 – C 10 –min-repeat-entropy 1.5 –strict-vcf –q 13 –m 60". The obtained variants were filtered by bcftools[30] using a quality filter over 100 and a minimum number of 10 reads per position. The remaining variants were processed by snpEff[31] 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 using a fraction of at least 0.01 where at least two reads should support the alternative allele. Raw sequencing data are available in the Sequence Read Archive (SRA) under the PRJNA656432 BioProject accession number.

ADDITIONAL REFERENCES FOR METHODS

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

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

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

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

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

31. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & Ruden, D.M. 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 (2012).

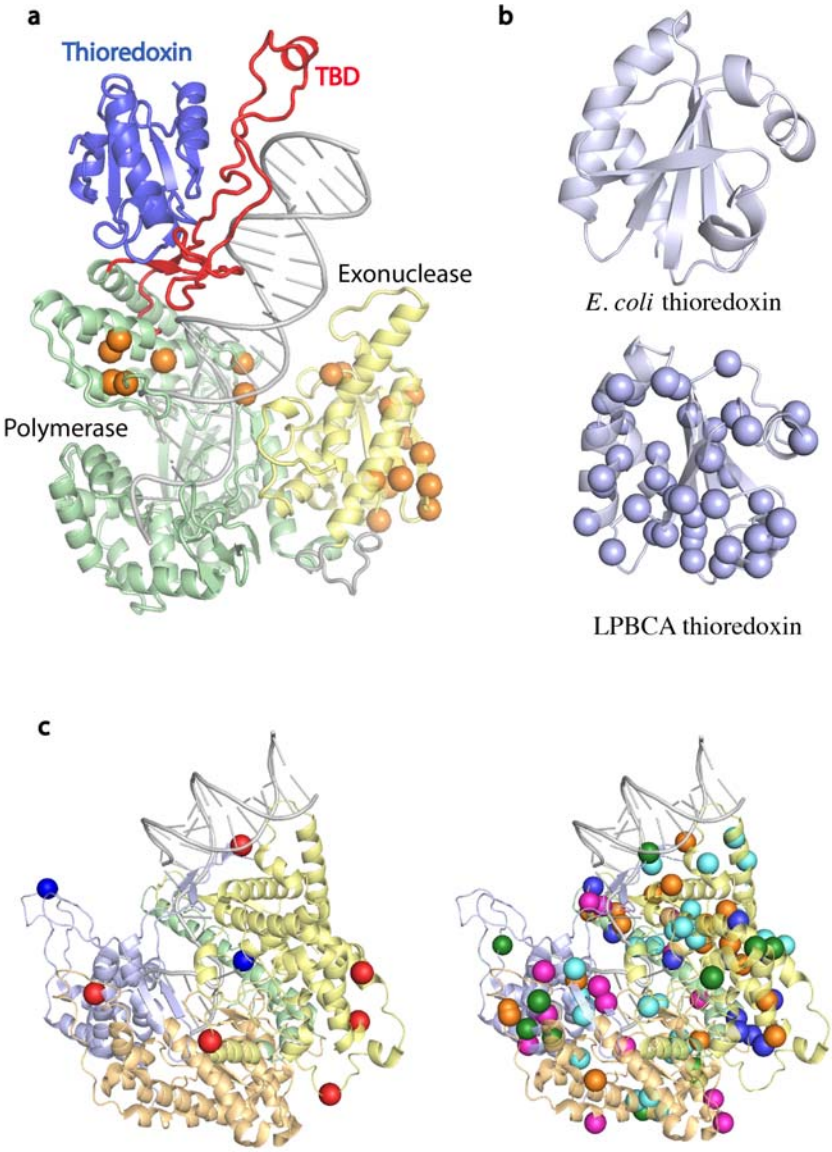

Fig. 1|3D structures of the main molecular players in this work. a, Structure of the viral DNA polymerase interacting with its processivity factor, *E. coli* thioredoxin (PDB ID 1T8E). The thioredoxin binding domain of the viral polymerase is labelled with TBD. Orange spheres are used to highlight the positions at which mutations are fixed during viral adaptation to the engineered host. We include the results from 14 evolution experiments in which Sanger sequencing was performed (see text for details) and {A,B,C} experiments in which next generation sequencing was carried out (see text and Fig. 3 for details). **b,** Structures of *E. coli* thioredoxin (PDB ID 2TRX)) and the alternative LPBCA thioredoxin (PDB ID 2YJ7). Spheres in the latter structure indicate the positions at which the amino acid residues differ between the two proteins. **c,** Structure of the viral RNA polymerase (PDB ID 1QLN). In the structure in the left, spheres are used to highlight the positions at which mutations are observed at high frequency in evolution experiments B (blue) and C (red). In the structure at the right, spheres indicate the positions at which low frequency mutations are observed in experiment A: V_1(cyan), V_2(blue), V_3(orange), V_4(pink), V_5(green). See text and Fig. 3 for the relevant details on the {A,B,C} set of experiments.

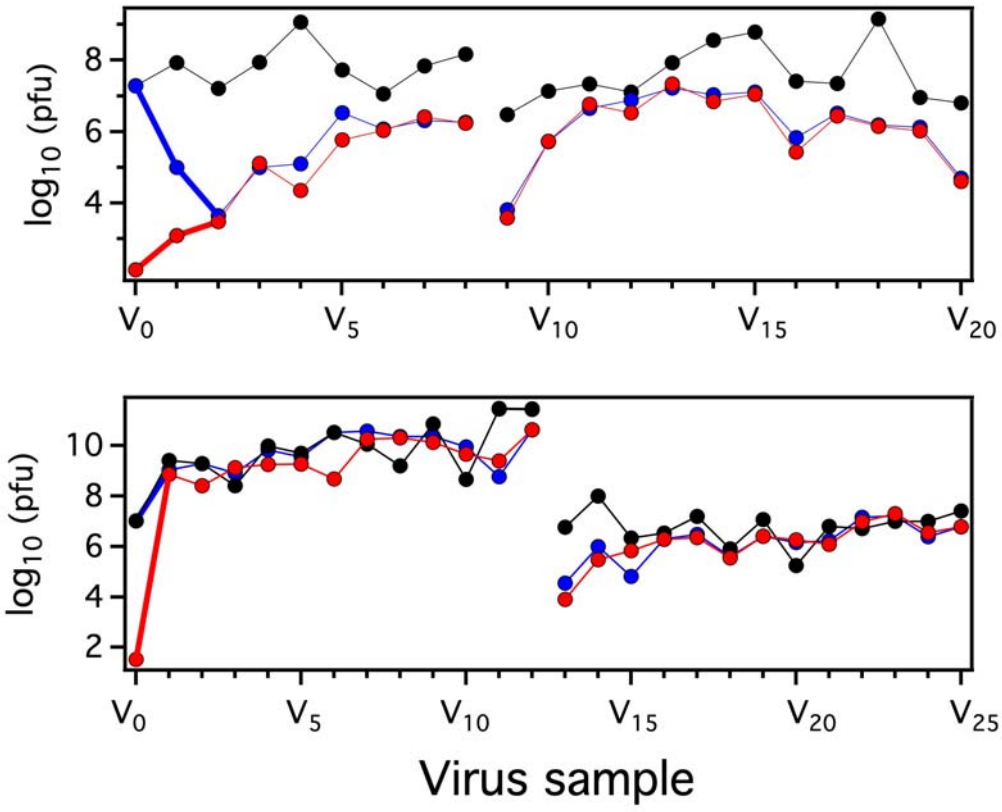

Fig. 2| Propagation of bacteriophage T7 in engineered *E. coli* cells with a modified thioredoxin gene. Number of plaque forming units (pfu) for virus samples from two laboratory evolution experiments performed as described in the text. Pfu values determined using the engineered host (red) and the original 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.

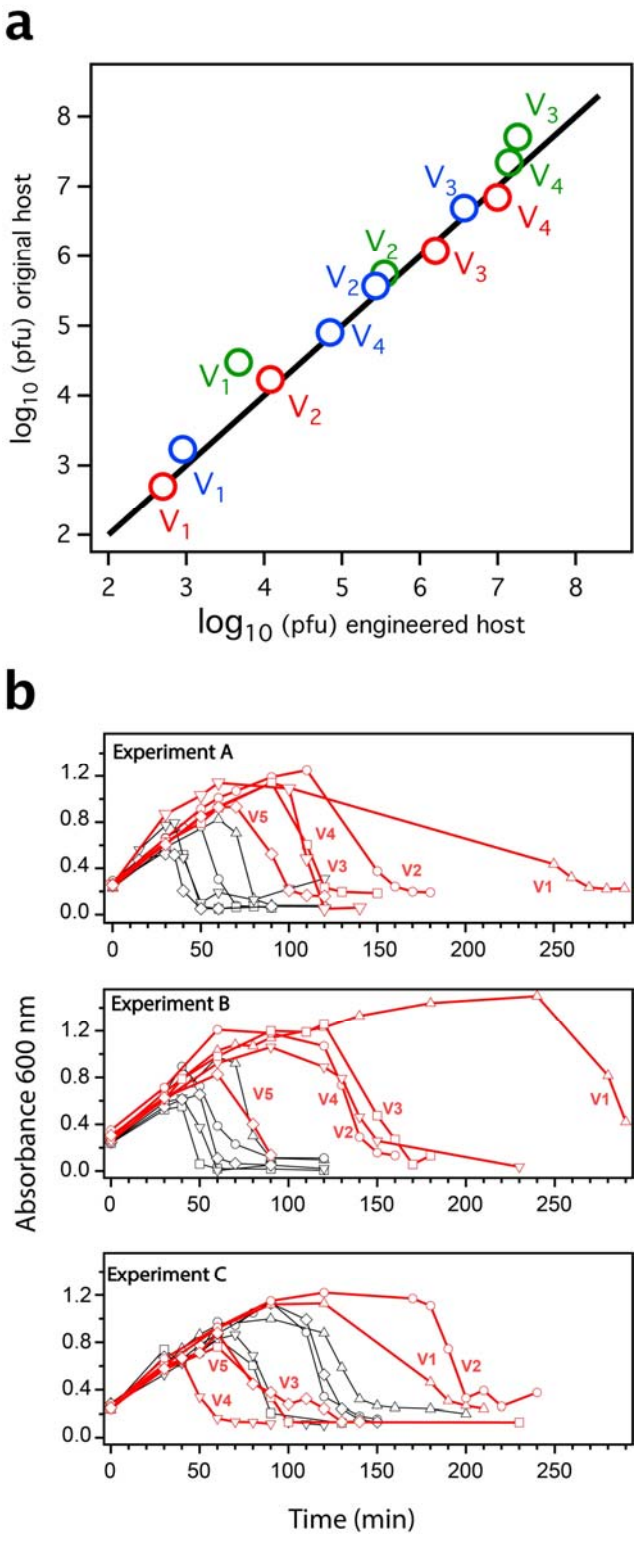

Fig. 3|Engineered host versus original host adaptation during laboratory virus evolution. Data from the {A,B,C} set of laboratory evolution experiments (see text for details). The three experiments involve selection for propagation in the engineered host. **a,** Pfu values for virus samples from the 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(green), B (blue), C (red). **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.

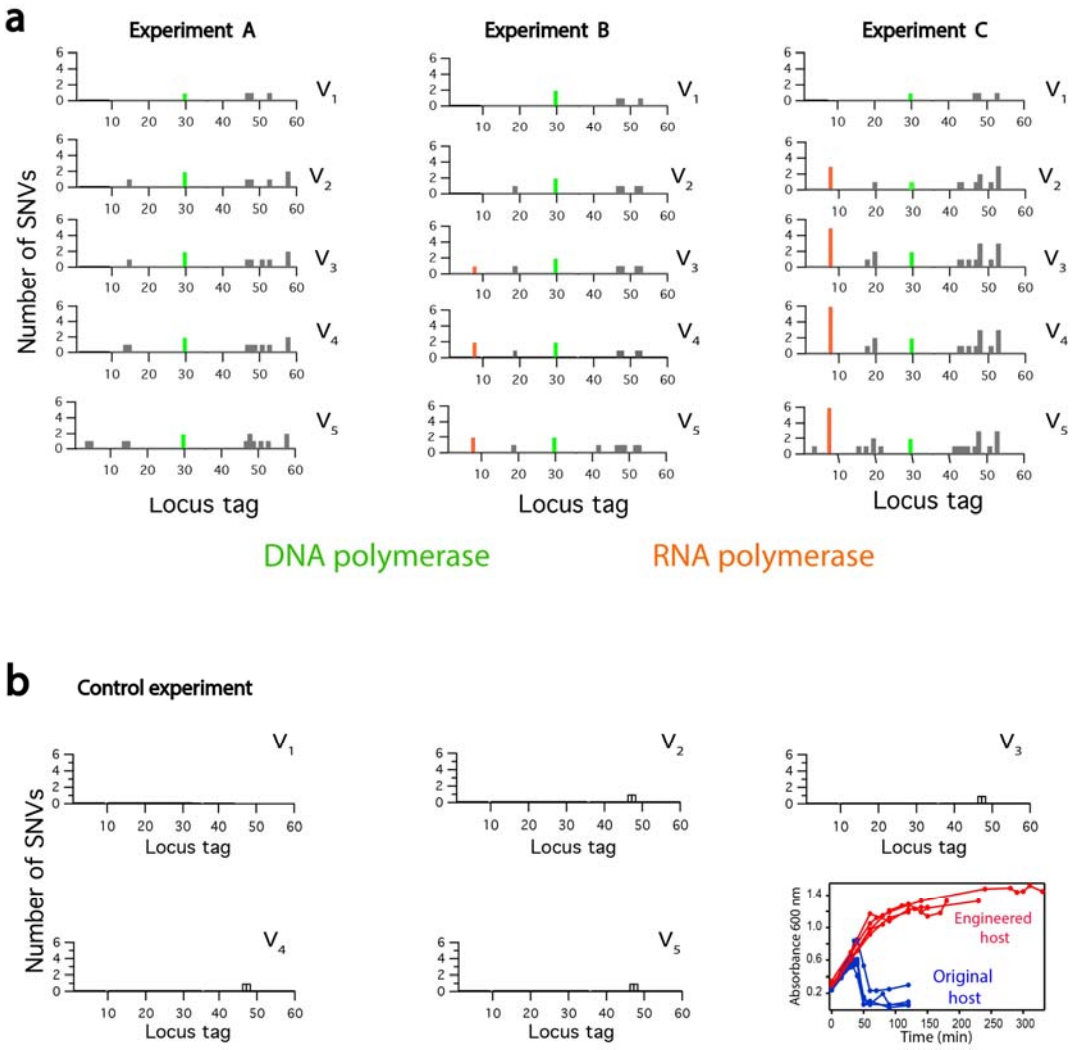

Fig. 4|Sequence changes in the viral genome during laboratory evolution. a, Number of high frequency (fraction>0.95) mutations for the different viral genes during the evolution experiments of the{A,B,C} set. The viral DNA and RNA polymerases have been highlighted. Data were obtained by Illumina sequencing 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 color 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.

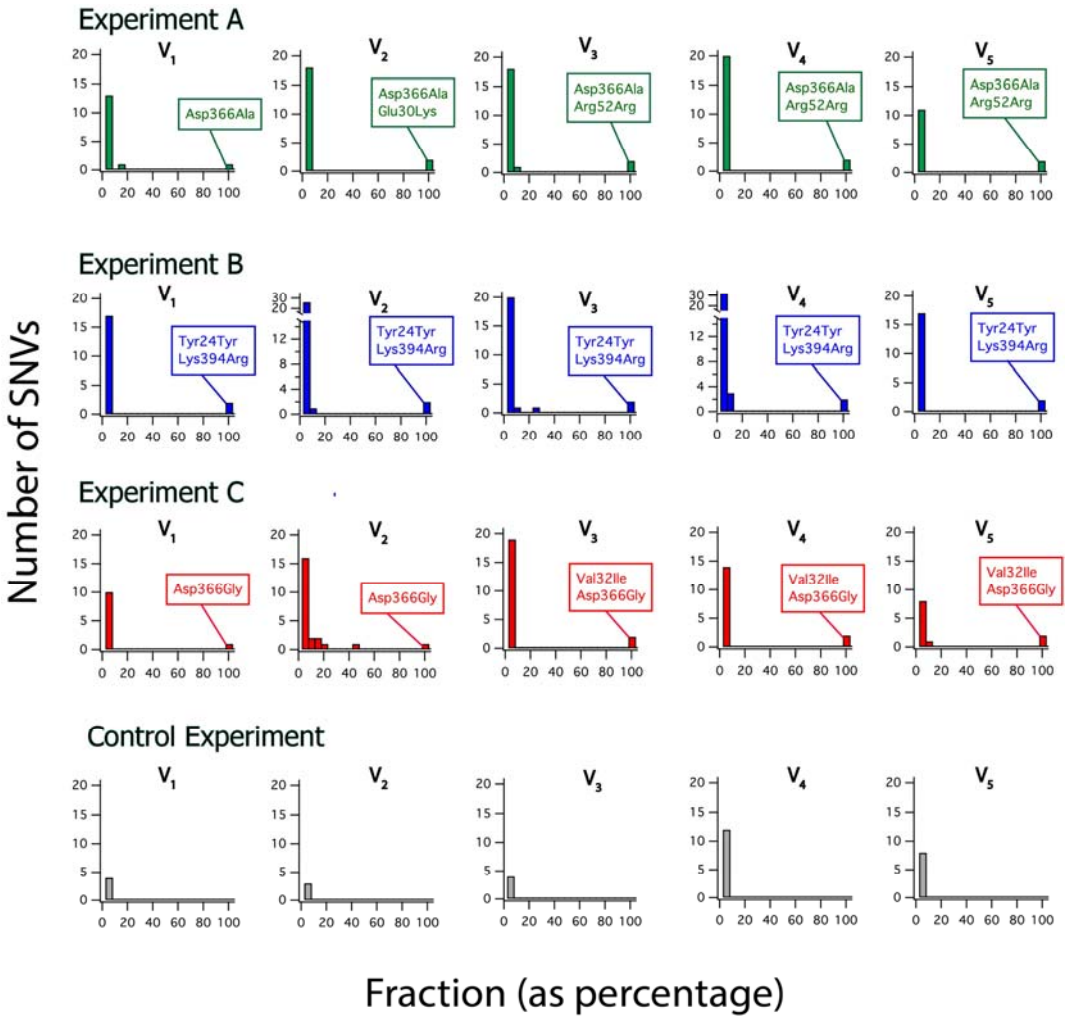

Fig. 5|Sequence changes in the viral DNA polymerase during laboratory virus evolution. Single nucleotide variants (SNVs) for the gene of the viral DNA polymerase for virus samples from the three evolution experiments of Fig. 3. Data were obtained by Illumina sequencing of the DNA extracted after lysis 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 cotraslational folding. Data for a control experiment corresponding to the propagation of the original virus in the original *E. coli* host are also included.

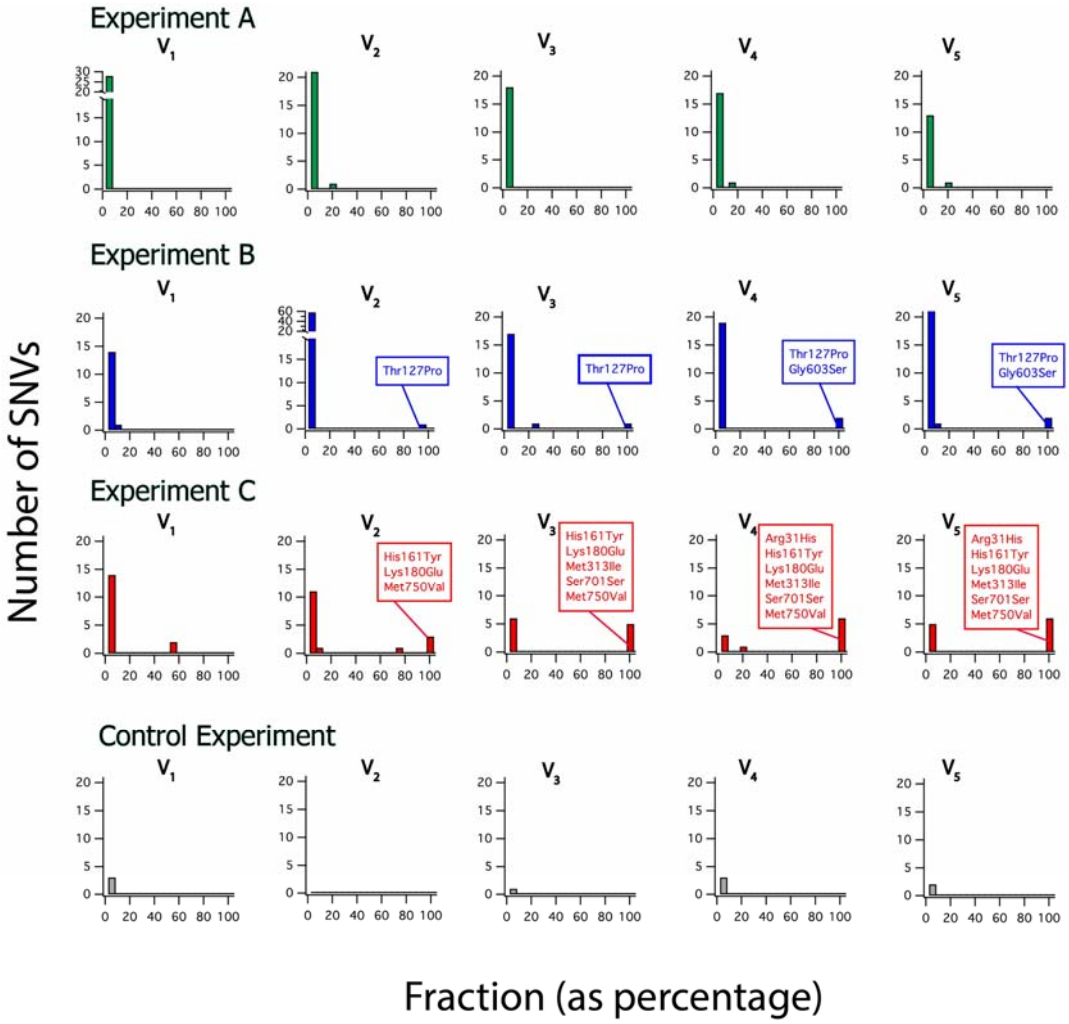

Fig. 6|Sequence changes in the viral RNA polymerase during laboratory virus evolution. Single nucleotide variants (SNVs) for the gene of the viral RNA polymerase for virus samples from the three evolution experiments of Fig. 3. Data were obtained by Illumina sequencing of the DNA extracted after lysis 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 are also included.