

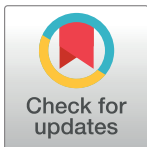
## RESEARCH ARTICLE

# Why are viral genomes so fragile? The bottleneck hypothesis

Nono S. C. Merleau<sup>1</sup>, Sophie Pénisson<sup>2,3</sup>, Philip J. Gerrish<sup>4</sup>, Santiago F. Elena<sup>5,6</sup>, Matteo Smerlak<sup>1\*</sup>

**1** Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany, **2** Université Paris Est Créteil, CNRS, LAMA, Creteil, France, **3** Université Gustave Eiffel, LAMA, Marne-la-Vallée, France, **4** University of New Mexico, Albuquerque, New Mexico, United States of America, **5** Instituto de Biología Integrativa de Sistemas (I<sup>2</sup>SysBio), CSIC-Universitat de València, València, Spain, **6** Santa Fe Institute, Santa Fe, New Mexico, United States of America

\* [smerlak@mis.mpg.de](mailto:smerlak@mis.mpg.de)



## OPEN ACCESS

**Citation:** Merleau NSC, Pénisson S, Gerrish PJ, Elena SF, Smerlak M (2021) Why are viral genomes so fragile? The bottleneck hypothesis. PLoS Comput Biol 17(7): e1009128. <https://doi.org/10.1371/journal.pcbi.1009128>

**Editor:** Katia Koelle, Emory University, UNITED STATES

**Received:** January 28, 2021

**Accepted:** May 28, 2021

**Published:** July 8, 2021

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1009128>

**Copyright:** © 2021 Merleau et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Code for replication is available at <https://github.com/strevol-mpi-mis/EvoEpi>.

**Funding:** Funding for this work was provided by the Alexander von Humboldt Foundation in the

## Abstract

If they undergo new mutations at each replication cycle, why are RNA viral genomes so fragile, with most mutations being either strongly deleterious or lethal? Here we provide theoretical and numerical evidence for the hypothesis that genetic fragility is partly an evolutionary response to the multiple population bottlenecks experienced by viral populations at various stages of their life cycles. Modelling within-host viral populations as multi-type branching processes, we show that mutational fragility lowers the rate at which Muller's ratchet clicks and increases the survival probability through multiple bottlenecks. In the context of a susceptible-exposed-infectious-recovered epidemiological model, we find that the attack rate of fragile viral strains can exceed that of more robust strains, particularly at low infectivities and high mutation rates. Our findings highlight the importance of demographic events such as transmission bottlenecks in shaping the genetic architecture of viral pathogens.

## Author summary

Given that most mutations are deleterious, high mutation rates carry a significant evolutionary cost. To reduce this burden, an obvious evolutionary solution would be to reduce the fitness cost of mutations by becoming more robust; this solution is indeed selected in populations of constantly large size. Here, we show that when populations regularly experience bottlenecks, as viruses do upon transmission to a new host, a less obvious solution becomes more viable: namely, to *increase* the fitness cost of mutations so that unfit mutants are less likely to fix at each passage. This could explain why viruses—especially RNA viruses—do in fact have very fragile genomes.

## Introduction

From tobacco mosaic virus to poliovirus and SARS-CoV-2, some of the most consequential plant, animal and human pathogens are RNA viruses. In spite of their tiny genomes, these

framework of the Sofja Kovalevskaja Award endowed by the German Federal Ministry of Education and Research to M.S. Work in València was supported by Spain Agencia Estatal de Investigación - FEDER grant PID2019-103998GB-I00 and Generalitat Valenciana grant PROMETEO2019/012 to S.F.E. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

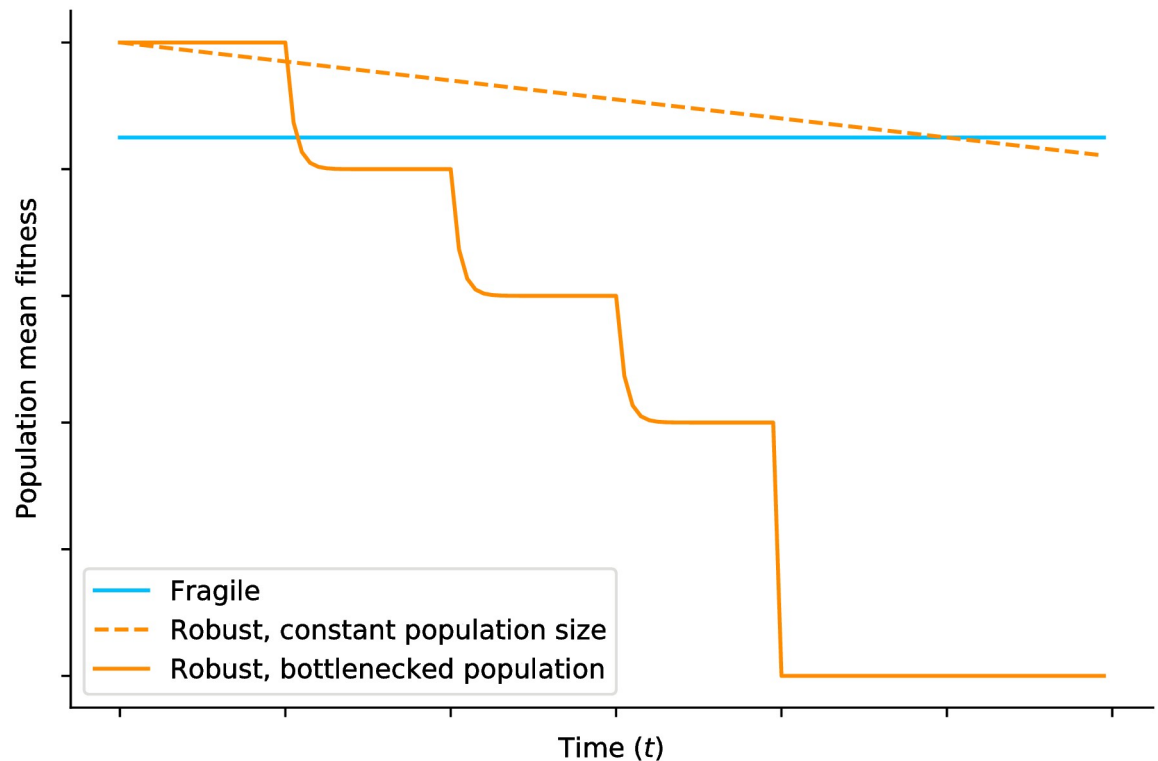
organisms find adaptive solutions to environmental challenges such as hosts' immune response, pervasive differences in susceptible cell types, switches in host and vector species, and antiviral drugs [1–3]. Such remarkable evolvability has been linked to the error-prone replication and short generation times of RNA viruses [4]. But high mutation rates are a double-edged sword: while replication errors provide the fuel necessary for rapid adaptation, they also increase the genetic load on viral populations, which in turn imposes a limit to genome size [5, 6]. Moreover, evidence gathered from diverse viral systems [7–11] shows that high mutation rates coupled with strong population bottlenecks (e.g. associated with airborne or fomite transmission events) turn on Muller's ratchet [12], resulting in the loss of fit genotypes [13, 14]. As fitness declines, populations risk experiencing a mutational meltdown, with low fitness genotypes unable to restore large population sizes and deleterious mutations accumulating at an ever increasing rate [15, 16]. How do RNA viruses manage to persist in spite of these challenges?

RNA viruses may have evolved specific mechanisms to maintain genome integrity in the face of high mutation rates [17–20]. Proposed mechanisms include complementation at high multiplicity of infection (MOI) during transmission e.g., by physically aggregating viral particles [21, 22]; the use of stamping machine, rather than geometric, replication mechanisms [23]; the segmentation of viral genomes with reassortments during mixed infections or increased recombination rates, two simple forms of sex that reduce mutational load [24, 25]; or the co-opting of cellular chaperones (e.g., heat-shock proteins) to assist at different stages of the replication cycle [26]. It is possible that all of these mechanisms (and others yet to be discovered) play a role in mitigating the damage done by mutations.

An extreme of such mitigation is one in which mutations cause no damage, i.e., they are neutral. And yet it is unrealistic to suppose that an organism can continue to accumulate mutations forever with no cost. This begs the question: how many mutations can an organism accumulate before a fitness cost is incurred? And how does that fitness cost then increase as further mutations accumulate? One common model of viral evolution, the neutral network model, supposes that a virus has a complex network of mutational neighbors (and neighbors of neighbors, etc) which incur zero fitness cost. Any mutant outside of this network, however, is non-viable. In this model, a viral population can mitigate the cost of mutation by evolving towards the center (away from the extremities) of such a network, where most mutations are neutral [27–30].

Counter-intuitively, the opposite strategy of *maximizing* mutational damage may also be a key component of the evolutionary response to low-fidelity replication. Both empirical and theoretical arguments support this hypothesis. First, viral RNA genomes are highly compacted, contain overlapping reading frames, encode for polyproteins that need to be precisely processed post-translationally, and express multi-functional proteins involved in different processes along the infection cycle; these structural properties all predict large deleterious effects of most mutations [5, 6], as is indeed observed [31]. Second, analyses of Muller's ratchet show that the risk of mutational meltdown is highest when deleterious effects are moderate [16]. This is because weak deleterious mutations negatively impact population fitness without being strongly selected against (Fig 1). This observation has led several authors to the conclusion that genetic fragility is in fact selected for in the high mutation rate regime of evolution [32–34].

In this paper we show that the population bottlenecks experienced by RNA viruses contribute to promoting fragile genetic architectures. Our contribution is twofold. First, we model population bottlenecks (and the genetic drift they induce, including Muller's ratchet) explicitly using multi-type branching processes [35, 36]; such bottlenecks are not easily modeled within more common approaches based on weak-mutation strong-selection limit [33] or quasi-species theory [32, 34]. Using general results in branching process theory, we derive expressions for the survival probability of a population through repeated bottlenecks as a function of the



**Fig 1. Schematic of fitness dynamics.** Robust genomes (orange) initially have a lower mutational load, and therefore higher population mean fitness, than fragile genomes (blue). But they also fix deleterious mutations more frequently, leading to a gradual decline in population fitness (dashed line); this effect is more pronounced when bottlenecks weaken the strength of selection and accelerate the clicking of Muller's ratchet (continuous line).

<https://doi.org/10.1371/journal.pcbi.1009128.g001>

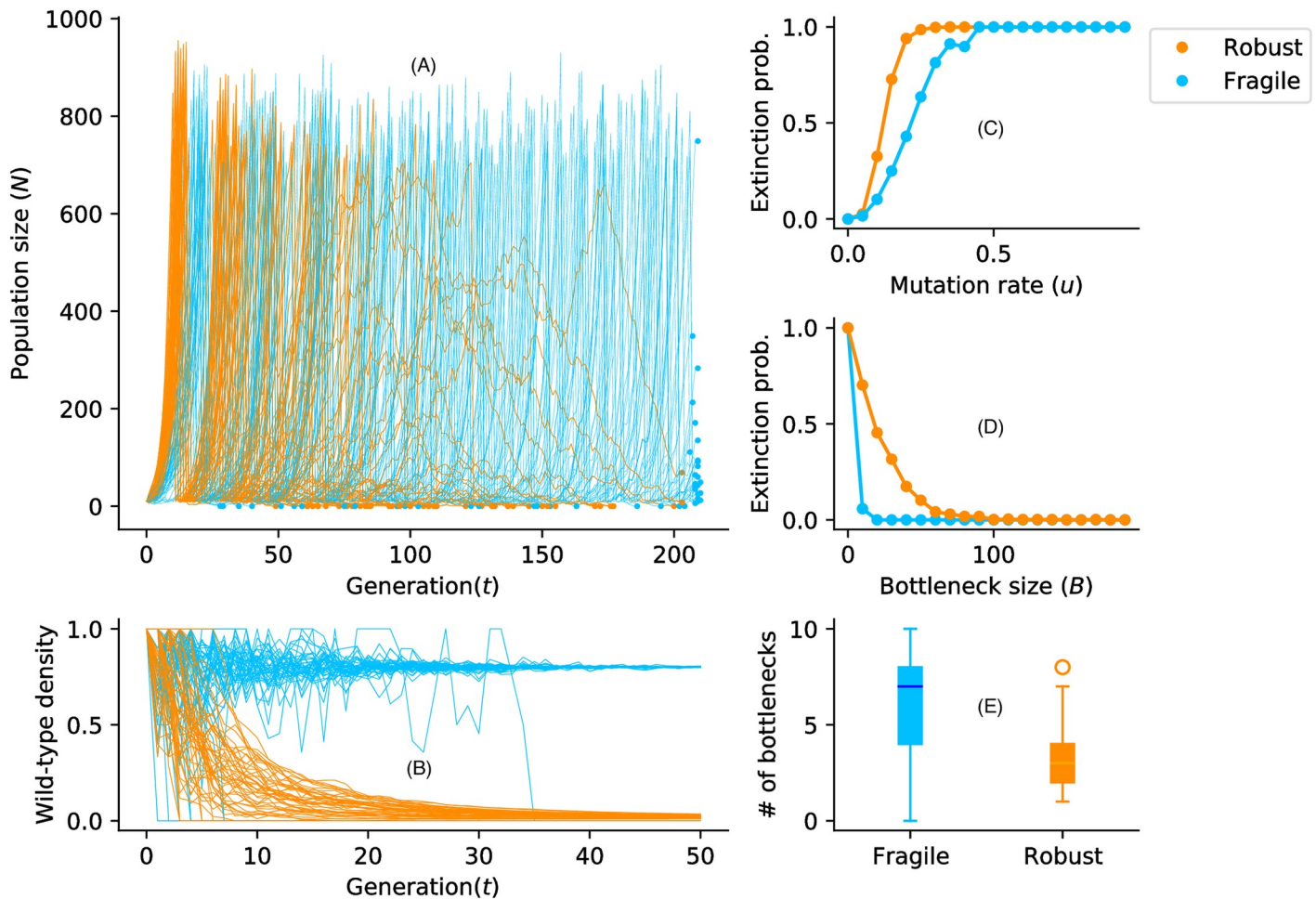
deleterious effect of mutations, confirming that fragility can be advantageous in the long run. Second, we consider the epidemiology of genetic fragility using a simple agent-based compartmental model. Evolutionary epidemiology [37] is an emerging field focusing on the interactions between evolutionary and epidemiological dynamics that seeks to explain the evolution of virulence and other properties of pathogens. Here we ask under what epidemiological conditions a fragile strain can have a higher attack rate than a robust one. We find that two parameters determine whether or not this is possible: the mutation rate  $u$  and the infection rate  $\beta$  (both relative to the recovery rate), with high  $u$  and low  $\beta$  both favouring fragile viruses.

Our model builds upon the standard model of Muller's ratchet [13], under which all mutations have the same deleterious effect (in particular, none is lethal), and do not interact epistatically. However, positive epistasis among deleterious mutations has in fact been shown to be a pervasive phenomenon in compacted RNA genomes [38]. Likewise, the fraction of mutations that are lethal is usually large for RNA viruses [39]. We rationalize these simplifications by noting that relaxing them would further increase the long-term advantage of fragile genomes, reinforcing the argument for the bottleneck hypothesis.

## Results

### Muller's ratchet in expanding viral populations

Consider a small viral population with absolute fitness  $w_0 > 1$  and genomic mutation rate  $u$ . Assume that all mutations are deleterious with the same effect  $s_d$ , such that an individual



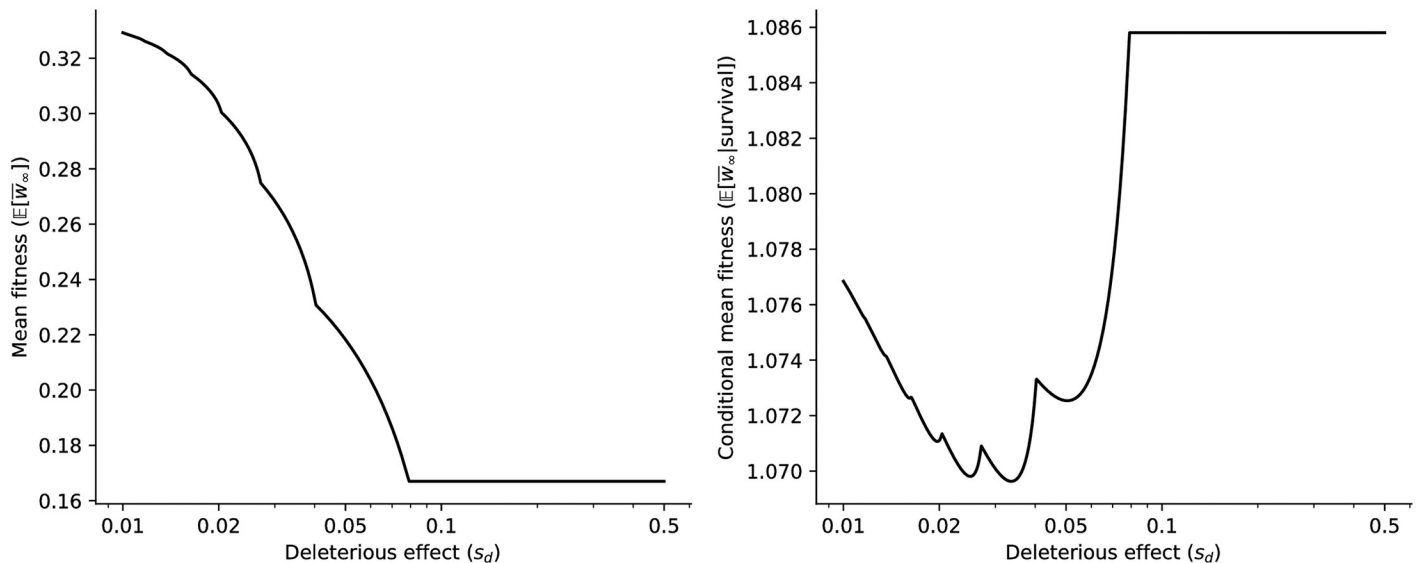
**Fig 2. Viral populations through multiple bottlenecks.** For an initial population size  $N = 10$  with fitness  $w_0 = 1.5$  and a mutation rate  $u = 0.2$ , viral particles reproduce, mutate, and die out. Once the populations reach the carrying capacity  $C = 800$ , a sub-population of size  $B = 10$  is sampled and the branching process is restarted. (A) Populations of robust viruses grow faster but go extinct more often after multiple bottlenecks. (B) The wild-type (individual with fitness  $w_0$ ) density in the populations over time. (C) Effect of mutation rate on the extinction probability for a fixed bottleneck size  $B = 10$ . (D) Effect of  $B$  for a fixed mutation rate ( $u = 0.2$ ) on the extinction probability. (E) Number of bottlenecks before extinction.

<https://doi.org/10.1371/journal.pcbi.1009128.g002>

carrying  $i$  mutations—an “ $i$ -mutant”—has fitness  $w_i \equiv w_0(1 - s_d)^i$ . The dynamics of such a population falls under two broad alternatives: either it goes extinct through demographic fluctuations, or it grows to unbounded sizes with an asymptotic mean fitness  $1 < \bar{w}_\infty \leq w_0$ ; the latter outcome is only possible if  $w_0 e^{-u} > 1$ . This is illustrated in Fig 2 for a low value of  $s_d$  (a “robust” type) and a high value of  $s_d$  (a “fragile” type).

Key to the fate of the population is the onset of Muller’s ratchet, *i.e.* the extinction of fit genotypes through genetic drift in small populations. The ratchet mostly clicks during the early states of the expansion, when the population size is smallest and extinction is likely. But it may also start clicking later through some rare fluctuation; if this click is followed by another click, and then another, the population can start shrinking again into mutational meltdown. If the ratchet clicks too many times, namely more than  $K = \max\{k: w_k e^{-u} > 1\}$  times, extinction is unavoidable: any mutant carrying more than  $K$  mutations has absolute fitness smaller than one, and therefore generates a subcritical branching process.

Robust and fragile populations experience Muller’s ratchet differently. For a robust genotype, mutations have a small deleterious effect, and are therefore under weak negative



**Fig 3. Analytical results for expected asymptotic population mean fitness.** Both are unconditional (left) and conditional on survival (right) as a function of the deleterious effect of mutations  $s_d$ , for a population with  $w_0 = 1.2$  and  $u = 0.1$ . While a more robust strain has a lower mutational load and therefore a higher population mean fitness (left), it is also more vulnerable to Muller's ratchet, implying that surviving populations tend to have lower mean fitness (right).

<https://doi.org/10.1371/journal.pcbi.1009128.g003>

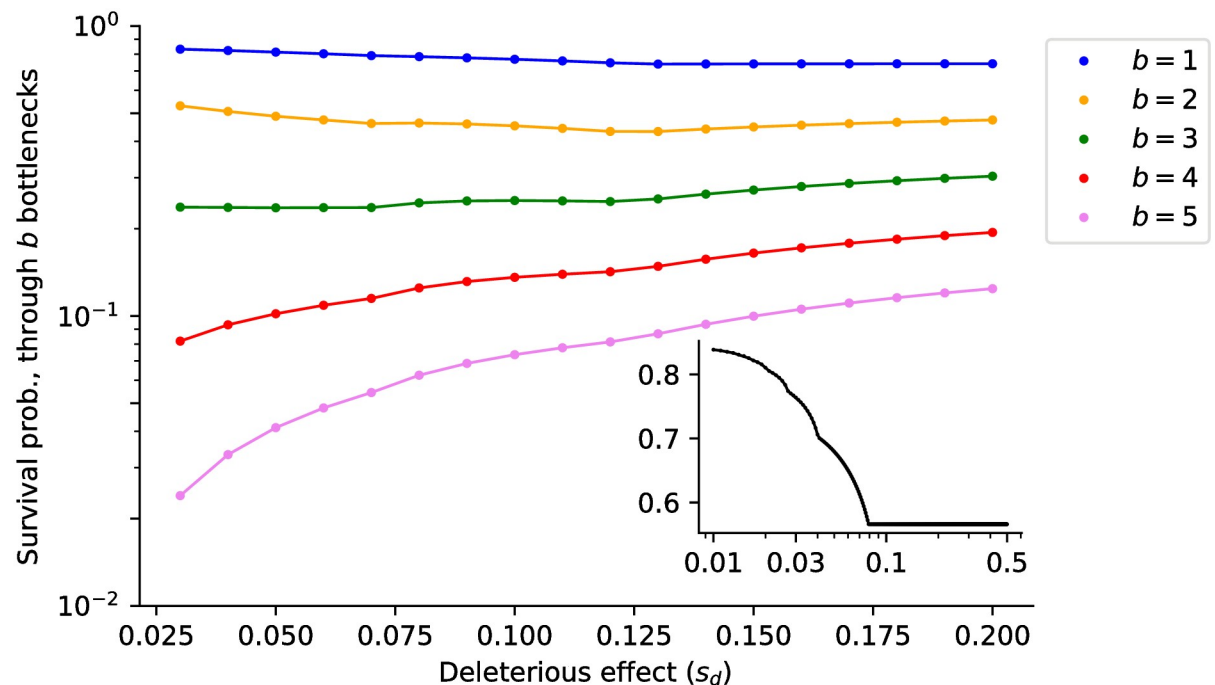
selection. This makes a click of the ratchet quite likely. On the other hand, each click comes at a low fitness cost, and so the population can withstand a relatively large number of clicks. Fragile populations, by contrast, are under strong negative selection against deleterious mutations, hence clicks are rarer; when they do occur, though, extinction becomes almost certain. Which fares better?

Using branching process theory we compute analytically the probability  $p_k$  that the fittest surviving individual carries exactly  $k$  mutations, *i.e.* that Muller's ratchet click  $k$  times (Eq (1) in Methods). From this, we find that both the survival probability  $p_{\text{surv}} = \sum_{k=0}^K p_k$  and the expected asymptotic population mean fitness  $E(\bar{w}_{\infty})$  (Eq (2)) are decreasing functions of  $s_d$ , consistent with the idea that, when all mutations are deleterious, mutational robustness is an evolutionary advantage. But there is a caveat: because Muller's ratchet clicks less frequently when genomes are fragile, the populations that do emerge from the expansion tend to be mutations-free. As a result, the asymptotic mean population fitness *conditional on survival*  $E(\bar{w}_{\infty} | \text{survival})$  turns out to be non-monotonic in  $s_d$ , and in fact to be maximized for large deleterious effects (Fig 3).

### Survival of viral population through multiple bottlenecks

The adaptive value of virus' mutational fragility becomes apparent when we consider a succession of bottleneck-expansion cycles (Fig 2), corresponding to viral transmission followed by within-host replication. This process can be modeled as a Markov chain on the space of post-bottleneck populations. Let  $B$  denote the size of the population after the transmission bottleneck. (In some cases,  $B$  can be as small as 1 [40, 41]).

We model virus transmission as sampling without replacement from a surviving population as described above, resulting in a new founding population with composition  $\mathbf{n} = (n_0, \dots, n_K)$ , where  $n_i$  is the number of  $i$ -mutants and we omit the subcritical mutants whose lineage will go extinct with probability one (hence  $\sum_{k=0}^K n_k \leq B$ ). After within-host expansion and transmission sampling, this population will give rise to a new founding population with composition



**Fig 4. Analytical results for the survival probability vs. mutational fragility.** For a population with  $w_0 = 1.2$  and  $u = 0.05$  undergoing  $b$  bottlenecks of size  $B = 5$ , it is computed using the bottleneck-to-bottleneck Markov chain  $P(\mathbf{n} \rightarrow \mathbf{m})$ . The higher the number of bottlenecks a population has to survive, the more fragility is selected for. The inset figure shows the survival probability for  $b = 0$ ,  $w_0 = 1.2$  and  $u = 0.05$ .

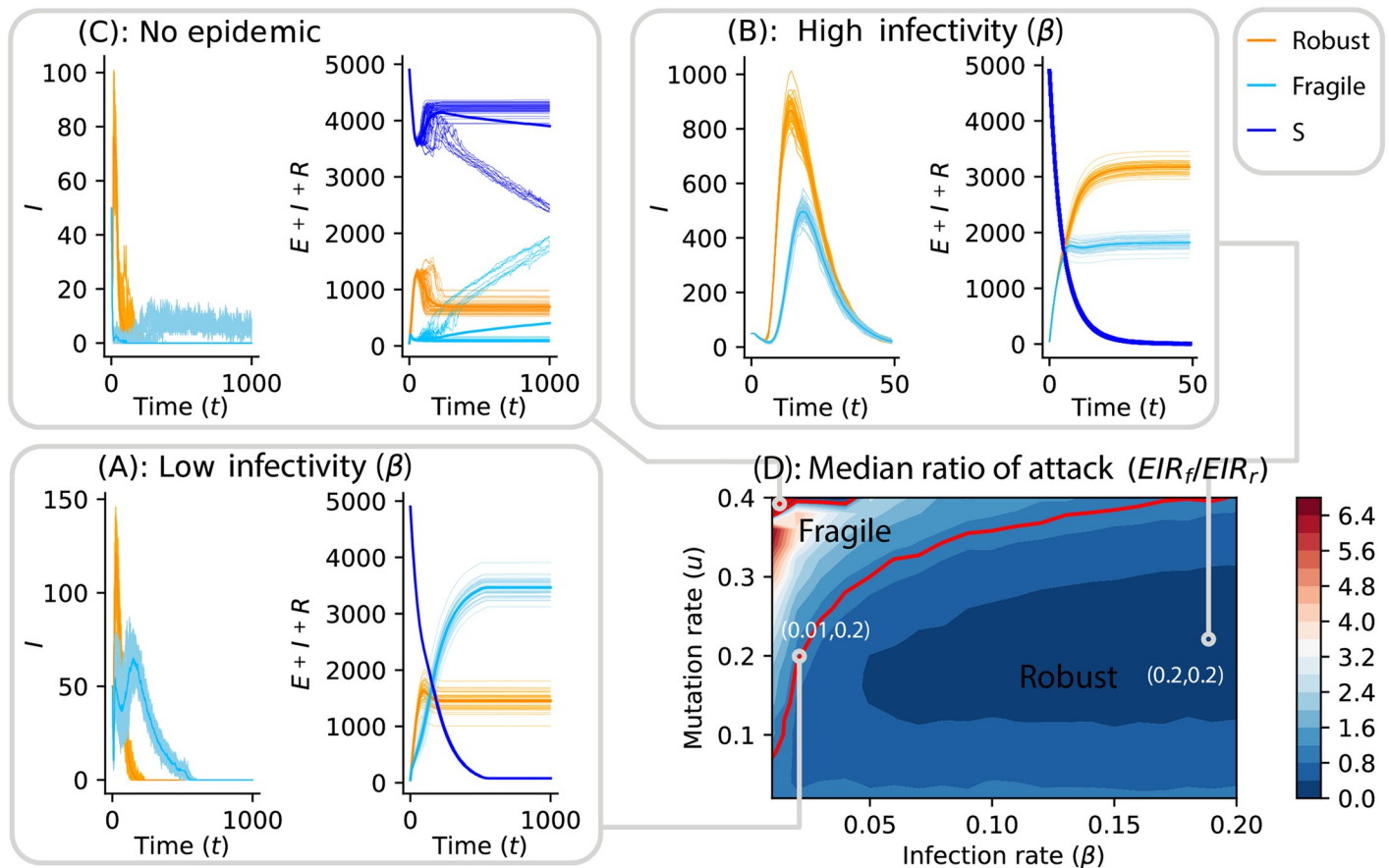
<https://doi.org/10.1371/journal.pcbi.1009128.g004>

$\mathbf{m} = (m_0, \dots, m_K)$  with a probability  $P(\mathbf{n} \rightarrow \mathbf{m})$  given explicitly in (Eq 10 in S1 Text). From this Markov chain, we can compute the probability that a population can survive any given number of bottlenecks (Eq (4)).

Fig 4 shows the survival probability after up to five bottlenecks of size  $B = 5$  as a function of the deleterious effect  $s_d$ . Although non-monotonic, this probability is maximized at large  $s_d$  when the number of bottlenecks increases, *i.e.* extinction becomes less likely for more fragile genomes. The results in the previous paragraph explain why: fragile populations are protected against Muller's ratchet, hence each new infection starts from fit founders. By contrast, robust genomes accumulate deleterious mutations; after several transmission bottlenecks, the founder particles tend to have low fitness and become increasingly unlikely to give rise to surviving lineages. It is as if robustness promoted mutational meltdown on a longer time scale—a meta-population meltdown.

### Epidemiology of fragility

How would these effects play out in an epidemic outbreak? Would the lower propensity of fragile genomes to suffer meltdown make up for their lower initial population fitness? To investigate this question we consider a Susceptible-Exposed-Infectious-Recovered (SEIR) model defined as follows. When a susceptible  $S$  meets an infectious individual  $I$ , a sample of the latter's viral population with size  $B$  is transmitted and  $S$  becomes exposed  $E$ . After this event, the within-host viral population carried by  $E$  can either (i) go extinct, in which case  $E$  returns to the susceptible compartment ( $E \rightarrow S$ ) or (ii) grow exponentially during an incubation period  $\tau$  until it reaches a critical threshold  $C$  which makes the host infectious ( $E \rightarrow I$ ). Which is more likely depends on the viral genetic parameters ( $w_0$ ,  $s_d$ ,  $u$ ) and, as shown in the previous section, on the numbers of supercritical viral particles transmitted to the new host,

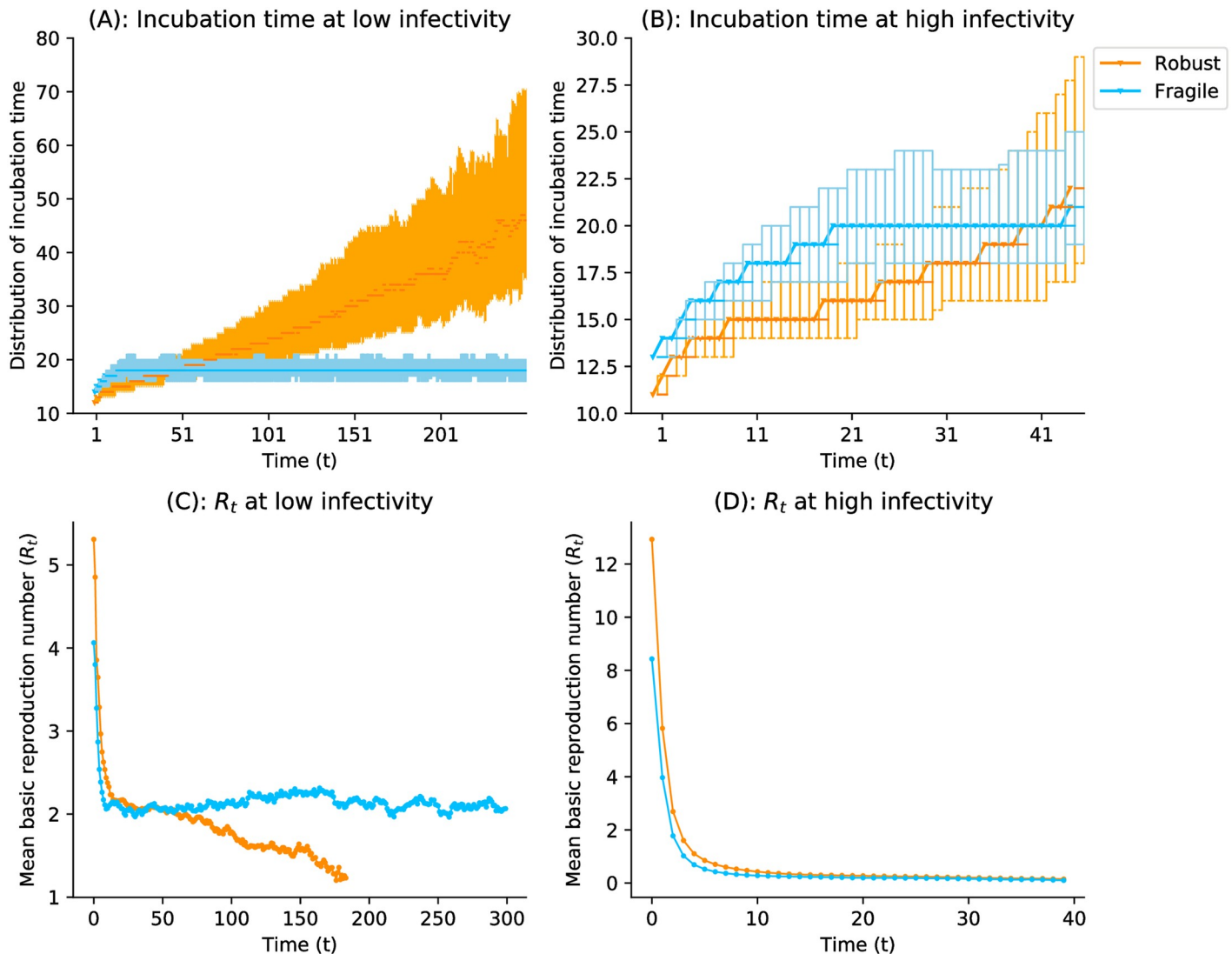


**Fig 5. Agent-based simulation of the SEIR model.** (A) At low infection rate ( $\beta = 0.01$ ) and high mutation rate  $u = 0.2$ , the robust genomes are more infectious in the beginning of the epidemic than the fragile ones, but as time goes on, it turns out that the fragile genomes become more virulent. (B) Using the same mutation rate and the highest infectivity gives the robust genome a chance to become the most infectious because it reproduces faster, and its population undergoes fewer bottlenecks. (C) At low infection rate and very high mutation rate ( $u = 0.4$ ), the disease does not spread in the population, due to the high extinction probability of the viral populations. (D) Median ratio of attack rate (which is the total number of fragile infections divided by the robust ones) across 50 runs for each  $(\beta, u)$  parameters. The red line shows the region of parameter  $(\beta, u)$  where both fragile and robust strains have equal attack rate.

<https://doi.org/10.1371/journal.pcbi.1009128.g005>

described by some vector  $\mathbf{n}$ . The incubation time is random as well, with a probability distribution depending on  $\mathbf{n}$  and on the number of clicks of Muller's ratchet during the expansion, see Eq (5). (We show in S1 Text that this distribution can be approximated by a Gamma distribution depending on these parameters.) The host population is assumed to be well-mixed (no spatial structure): at each time step and for each pair  $(S, I)$ , transmission occurs with a probability  $\beta$ . Recovery in turn takes place at a rate  $\gamma$ .

Fig 5 presents the results of simulations where a susceptible population of size  $N = 5000$  is seeded with 100 infectious individuals, half of which carry a robust viral strain ( $s_d = 0.05$ ) and the other half a fragile one ( $s_d = 0.9$ ). Three regimes emerge depending on the transmission rate  $\beta$  and the mutation rate  $u$  (at fixed bottleneck size  $B = 5$ ). When infectivity is high and the population quickly becomes completely infected, the robust strain fares better due to its higher initial fitness and faster growth rate. When infectivity is low, however, fragile strains prove to have a much higher attack rate (We define the attack rate of a strain as the number of hosts infected by that strain.). As noted in Fig 5A, the robust strains are more virulent in the early stages of the epidemic. But as time goes on, the robust genomes lose fitness and increase their post-bottleneck extinction probabilities. That is also the reason why there is a small decrease in



**Fig 6. Incubation time distributions and mean basic reproduction number at low ( $\beta = 0.01$ ) and high ( $\beta = 0.2$ ) infectivity, with viral parameters  $w_0 = 1.5$ ,  $u = 0.1$  and  $B = 5$ .** (A) The epidemic lasts longer and the robust viral populations lose fitness. The later transmitted viral particles take a longer time to grow up to C. In contrast, the incubation times of the fragile strains stay on average constant. (B) On short-time epidemic, the mean incubation time of the robust strain stays smaller than the fragile ones. (C) The epidemic lasts longer, wherein the beginning, the reproduction number of the robust strain is higher than the fragile one but later on, when the robust strain's  $R_t$  decreases, the fragile one stays constant. (D) The virus takes over the population in the early stage of the epidemic when the robust strains are the most virulent.

<https://doi.org/10.1371/journal.pcbi.1009128.g006>

the total number of robust genome infections in ( $E + R + I$ )-populations, as exposed hosts return to the susceptible compartment (Fig 5A). At a late stage ( $t > 200$ ), mutationally robust viruses are too degraded to make their host infectious. Finally, very high mutation rates and low infectivities do not allow for the epidemic to spread, as all viral populations (robust and fragile) quickly go extinct. These findings are summarized in Fig 5D in terms of the median ratio of attack rates across 50 runs for each parameter set.

Underlying this contrast is the distribution of incubation times, which differs for the robust and fragile strains (Fig 6A and 6B). The same observation can be drawn by looking at the mean basic reproduction number of both robust and fragile strains at low infectivity (and

respectively at high infectivity) (see Fig 6C and 6D). Over time, the basic reproduction number  $R_t$  of fragile strains stays constant while the robust one decreases.

## Discussion

There is a growing appreciation for the fact that population bottlenecks are not just a fundamental aspect of the life cycle of viruses, but that they also play a key role in their evolution [42, 43]. The effect of bottlenecks is not always detrimental: bottlenecks can effectively remove cheaters, *e.g.* defective interfering viruses [44], or enhance the effectiveness of selection if beneficial alleles act in *trans* [45]. In addition, genetic bottlenecks can facilitate traveling across the characteristically rugged fitness landscapes of RNA viruses [38], where it is easy for viruses to become trapped at suboptimal fitness peaks [46, 47]; by relaxing the intensity of selection, bottlenecks enable the exploration of new regions of the landscape.

In this paper we have explored another aspect of highly mutable populations subjected to periodic bottlenecks: they experience a strong evolutionary pressure towards genetic fragility. Earlier work has established that intermediate deleterious effects  $s_d$  maximize the strength and speed of Muller's ratchet [16, 32]; similar results have been reported more recently in terms of "ratchet robustness" [48] or "drift robustness" [33]. We find this U-shaped pattern in the context of expanding populations as well, *e.g.* for the asymptotic population mean fitness given survival (Fig 3); the unconditional population mean fitness, by contrast, always decreases with fragility, which is consistent with another analysis of the advantage of mutational robustness [34]. By modelling changes in population sizes with branching processes, we allowed a more complete picture to emerge. In this picture, the evolution of high neutrality [27, 49] is not incompatible with selection for maximally deleterious mutations [32] and anti-redundancy [34]. An agent-based SEIR epidemiological model further reveals the determinants of genetic fragility, highlighting the importance of epidemic transmission parameters.

The evolution of viruses is often described in terms of fitness landscapes and their topographies. Our findings highlight the limitation of this picture: the motion of evolving population in genotype space depends on the structure of the genotype-phenotype-fitness mapping, but also on mutation rates and life cycle parameters such as the frequency and stringency of population bottlenecks. As a result, the evolution of mutational robustness—or mutational fragility—cannot be construed solely as the search for an optimal region in the fitness landscape, be it the highest peak or the flattest plateau. Our multi-type branching process approach is suggestive that a composite definition of fitness might be more predictive of evolutionary success in the present context, namely, a definition that takes account of both offspring number (Malthusian fitness) and long-term survival probability. The augmented evolutionary relevance of this definition to the present context is manifest in the comparison between the left and right panels of Fig 3.

To be sure, our model relies on simplifying assumptions, mainly pertaining to the nature of underlying mutational landscapes and to epidemiological details. The assumption that all mutations having the same deleterious effect is a common one that simplifies the mathematics. Fitness effects of deleterious mutations are more realistically modeled as a random variable with a continuous, heavy-tailed distribution (*e.g.* the Gamma and Weibull distributions have been previously used to satisfactorily fit experimental data [50, 51]) or a U-shaped distribution to incorporate lethal mutations. As already noted, we make another common assumption that does not necessarily hold for real viral genomes, namely, the independence of mutational effects.

Evidence pervasively suggests that positive epistasis is the norm for compacted viral RNA genomes [38], including many instances of compensatory mutations. Indeed, it was shown long ago that if deleterious alleles interact synergistically, they are more efficiently removed from the population and thereby slow the advance of Muller's ratchet [52]. Finally, our evolutionary epidemiological approach ignores co- or super-infections, wherein multiple viral strains infect the same host. This last point is subtle. On the one hand, multiple infections allow competition among strains to occur within hosts as well as across hosts, which may reinforce the advantage of fragile strains in the low  $\beta$  regime. On the other hand, multiple infections allow the sharing of gene products among different genotypes within a cell, thus compensating for deleterious effects and eventually contributing to the accumulation of more mildly deleterious mutations in the population [53].

If position on the robustness-fragility spectrum is a heritable and variable trait, then that trait is subject to adjustment by indirect selection [54]. How this trait evolves will be determined by a tradeoff between the individual vs population cost of mutation. Under low mutation rates, selection on this trait at the individual and population levels can coincide. Under higher mutation rates and low effective population sizes (as studied here), however, they can diverge, favoring increased but unsustainable robustness at the individual level, but favoring increased and sustainable fragility at the population level.

This tradeoff, however, becomes more complex in changing environments: if a change in environment is lethal to the fragile genome, and if "evolutionary rescue" [55–57]—through, for example, the viability of an immediate mutational neighbor—is not feasible, this may tip the balance toward robustness. This interplay between environmental change and the evolution of genetic architecture has been dubbed *plastogenetic congruence* [58] in the context of RNA viruses which can face dramatic, unpredictable and fast fluctuations in their environments. This interplay between environment and genetic architecture has been observed in several experimental studies [59, 60].

A study of particular relevance to our work [61] found that poliovirus at high multiplicity of infection (MOI) contained more genetic variability, including many neutral mutations, and were more mutationally robust. By contrast, populations at low MOI evolved greater fragility. These findings are entirely consistent with our theoretical findings here.

This study, however, introduced a further complexity that added some realism: the poliovirus populations were then used to infect hosts that differed in their susceptibility to infection (a heterogeneous environment). Intriguingly, and counter to arguments put forth by [58], it was the more fragile viruses that were more successful adapting to the more restrictive novel hosts. This finding supports the notion that evolutionary rescue of fragiles may be effective enough in RNA viruses that fragiles can even be superior in heterogeneous or changing environments.

Clearly, there is more work to be done to determine how environmental heterogeneity in space and time tip the balance in favor of robustness vs fragility. And incorporating beneficial mutations into our analyses would add further realism. Beneficial mutations are typically rare and, when mutation rates are high, can arise on genetic backgrounds contaminated by deleterious mutations (a "ruby in the rubbish" [62]). It would be interesting to determine which has the higher probability of fixation: 1) a beneficial arising on a robust, only slightly eroded, background, or 2) a beneficial arising on a fragile background, where it arises on equal footing with the rest of the population but at an effectively reduced rate. These considerations warrant a follow-up study.

Overall, our findings can be summarized with a classic metaphor. Robust genomes are hares: they grow fast but accumulate mutational damage through Muller's ratchet, which jeopardizes their potential for long-term survival. Fragile genomes are turtles: they grow

more slowly but weather bottlenecks more reliably and have higher long-term survival probabilities. Turtles may seem weak individually, but as a group they have survived for hundreds of millions of years. Similarly, fragile viral genomes are individually vulnerable to deleterious mutations; at the meta-population level, however, they may hold the key to evolutionary resilience.

## Materials and methods

The evolution of viral populations experiencing Muller's ratchet and going through bottlenecks is modeled by a multitype branching process, in which a "type" corresponds to the number of accumulated deleterious mutations. We refer to [S1 Text](#) for a detailed description of the model and proofs of the following results.

Starting with  $\mathbf{n}$  supercritical mutants, the extinction probability  $p_{\text{ext},k}(\mathbf{n})$  of all types up to type  $k$  in the population is given by the element-wise product  $\mathbf{q}_k^n$ , where  $\mathbf{q}_k$  is the smallest solution to a fixed-point equation involving the generating function of the branching process. Muller's ratchet click probability  $p_k(\mathbf{n})$  that the fittest surviving individuals carry  $k$  mutations is then

$$p_k(\mathbf{n}) = p_{\text{ext},k-1}(\mathbf{n}) - p_{\text{ext},k}(\mathbf{n}) \quad (1)$$

and the population survival probability  $p_{\text{surv}}(\mathbf{n}) = \sum_{k=0}^K p_k(\mathbf{n})$ .

On the event that the fittest surviving individuals carry  $k$  mutations, the asymptotic proportion of  $k + i$ -mutants in the population is  $e^{-u/s_d} (u/s_d)^i / i!$ . The asymptotic mean population fitness is thus a random variable  $\bar{w}_\infty$  equal to  $w_k e^{-u}$  with probability  $p_k(\mathbf{n})$  for  $0 \leq k \leq K$ , and to 0 otherwise. Therefore

$$\begin{cases} \mathbb{E}(\bar{w}_\infty) = e^{-u} \sum_{k=0}^K w_k p_k(\mathbf{n}), \\ \mathbb{E}(\bar{w}_\infty \mid \text{survival}) = e^{-u} \sum_{k=0}^K w_k p_k(\mathbf{n}) / p_{\text{surv}}(\mathbf{n}). \end{cases} \quad (2)$$

When going through a bottleneck, this population gives rise to a new founding population with composition  $\mathbf{m}$ , according to transition probability

$$P(\mathbf{n} \rightarrow \mathbf{m}) = \sum_{k=0}^K Q_k(\mathbf{m}) p_k(\mathbf{n}) + \mathbf{1}_{\mathbf{m}=0} (1 - p_{\text{surv}}(\mathbf{n})), \quad (3)$$

where  $Q_k(\mathbf{m})$  is the probability of getting  $\mathbf{m}$  supercritical mutants in the sample of size  $B$ . The resulting transition matrix  $P$  then provides an explicit expression of the probability for a viral population with initial state  $\mathbf{n}$  to become extinct after going through at most  $b$  bottlenecks:

$$p_{\text{bottlenecks},b}(\mathbf{n}) = \mathbf{P}^b(\mathbf{n} \rightarrow \mathbf{0}). \quad (4)$$

In our SEIR epidemiological model, the time spent in the exposed state depends on  $\mathbf{n}$  and on the number of Muller's ratchet clicks during the expansion, or equivalently on the type  $k$  of the fittest surviving individuals. We show that this incubation period  $\tau_{\mathbf{n},k}$  is

$$\tau_{\mathbf{n},k} = \left\lceil \frac{\ln(C/W_{\mathbf{n},k}) - u/s_d}{\ln w_k - u} \right\rceil, \quad (5)$$

where  $W_{\mathbf{n},k}$  is a positive random variable and  $\lceil \cdot \rceil$  stands for the ceiling function. As exposed in [S1 Text](#),  $W_{\mathbf{n},k}$  can be expressed as the sum of a random number of independent exponential

random variables with common parameter

$$1 + \frac{1}{w_k e^{-u}} W[-w_k e^{-u} e^{-w_k e^{-u}}], \quad (6)$$

where  $W$  stands for the principal solution of the Lambert function.

Simulations were carried out using an agent-based approach in Python. We have created a repository at [github.com/strevol-mpi-mis/EvoEpi](https://github.com/strevol-mpi-mis/EvoEpi) containing the code and minimum documentation.

Parameter values used in all the simulations reported in this study were chosen to be representative of RNA virus. Genomic mutation rate values,  $u$ , were taken in the range 0.05 – 0.4, as described in [31]. [31] compared the DMFE for five different viruses (three with RNA and two with DNA genomes) and found that the mean  $s_d$  value was remarkably constant (around 0.11) but with a large variance and a bimodal shape characterized by many mutations of small effect and 20 – 40% of lethal mutations. To match with these values, in our simulations we took  $s_d$  in the range 0.05 – 0.9. Finally, the transmission bottleneck size,  $B$ , varies greatly across virus and routes of transmission [44], though a consensus exists that very few particles shall be enough to start a new infection [41]. Therefore, we enhance the strength of bottlenecks, we chose  $B$  values in the low range 1 – 10.

## Supporting information

**S1 Text. Mathematical appendices: Muller's ratchet in expanding populations (A); Survival through multiple bottlenecks (B); Epidemiology of genetic fragility (C); Explicit formulas (D).**

(PDF)

## Acknowledgments

We thank the members of the Structure of Evolution group at MPI MiS for useful discussions.

## Author Contributions

**Conceptualization:** Philip J. Gerrish, Santiago F. Elena, Matteo Smerlak.

**Formal analysis:** Sophie Pénisson.

**Funding acquisition:** Matteo Smerlak.

**Investigation:** Nono S. C. Merleau, Sophie Pénisson, Philip J. Gerrish, Matteo Smerlak.

**Methodology:** Sophie Pénisson, Matteo Smerlak.

**Software:** Nono S. C. Merleau.

**Supervision:** Matteo Smerlak.

**Validation:** Sophie Pénisson, Matteo Smerlak.

**Visualization:** Nono S. C. Merleau.

**Writing – original draft:** Sophie Pénisson, Philip J. Gerrish, Matteo Smerlak.

**Writing – review & editing:** Nono S. C. Merleau, Sophie Pénisson, Philip J. Gerrish, Santiago F. Elena, Matteo Smerlak.

## References

1. Elena SF, Sanjuán R. Virus Evolution: Insights from an Experimental Approach. *Annual Review of Ecology Evolution, and Systematics*. 2007; 38(1):27–52. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095637>
2. Holmes EC. The Evolutionary Genetics of Emerging Viruses. *Annual Review of Ecology Evolution, and Systematics*. 2009; 40(1):353–372. <https://doi.org/10.1146/annurev.ecolsys.110308.120248>
3. Andino R, Domingo E. Viral quasispecies. *Virology*. 2015; 479–480:46–51. <https://doi.org/10.1016/j.virol.2015.03.022> PMID: 25824477
4. Domingo E, Perales C. Viral quasispecies. *PLOS Genetics*. 2019; 15(10):e1008271. <https://doi.org/10.1371/journal.pgen.1008271> PMID: 31622336
5. Elena SF, Sanjuán R. Adaptive Value of High Mutation Rates of RNA Viruses: Separating Causes from Consequences. *Journal of Virology*. 2005; 79(18):11555–11558. <https://doi.org/10.1128/JVI.79.18.11555-11558.2005> PMID: 16140732
6. Belshaw R, Gardner A, Rambaut A, Pybus OG. Pacing a small cage: mutation and RNA viruses. *Trends in Ecology & Evolution*. 2008; 23(4):188–193. <https://doi.org/10.1016/j.tree.2007.11.010> PMID: 18295930
7. Chao L. Fitness of RNA virus decreased by Muller's ratchet. *Nature*. 1990; 348(6300):454–455. <https://doi.org/10.1038/348454a0> PMID: 2247152
8. Donis RO. Muller's ratchet and flu virus. *Nature*. 1991; 353(6342):308–309. <https://doi.org/10.1038/353308b0> PMID: 1922335
9. Duarte E, Clarke D, Moya A, Domingo E, Holland J. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proceedings of the National Academy of Sciences*. 1992; 89(13):6015–6019. <https://doi.org/10.1073/pnas.89.13.6015> PMID: 1321432
10. Yuste E, Sánchez-Palomino S, Casado C, Domingo E, López-Galíndez C. Drastic Fitness Loss in Human Immunodeficiency Virus Type 1 upon Serial Bottleneck Events. *Journal of Virology*. 1999; 73(4):2745–2751. <https://doi.org/10.1128/JVI.73.4.2745-2751.1999> PMID: 10074121
11. de la Iglesia F, Elena SF. Fitness Declines in Tobacco Etch Virus upon Serial Bottleneck Transfers. *Journal of Virology*. 2007; 81(10):4941–4947. <https://doi.org/10.1128/JVI.02528-06> PMID: 17344305
12. Muller HJ. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 1964; 1(1):2–9. [https://doi.org/10.1016/0027-5107\(64\)90047-8](https://doi.org/10.1016/0027-5107(64)90047-8) PMID: 14195748
13. Haigh J. The accumulation of deleterious genes in a population—Muller's Ratchet. *Theoretical Population Biology*. 1978; 14(2):251–267. [https://doi.org/10.1016/0040-5809\(78\)90027-8](https://doi.org/10.1016/0040-5809(78)90027-8) PMID: 746491
14. Stephan W, Chao L, Smale JG. The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genetical Research*. 1993; 61(3):225–231. <https://doi.org/10.1017/S0016672300031384> PMID: 8365659
15. Lynch M, Bürger R, Butcher D, Gabriel W. The Mutational Meltdown in Asexual Populations. *Journal of Heredity*. 1993; 84(5):339–344. <https://doi.org/10.1093/oxfordjournals.jhered.a111354> PMID: 8409355
16. Gabriel W, Lynch M, Bürger R. Muller's ratchet and mutational meltdowns. *Evolution*. 1993; 47(6):1744–1757. <https://doi.org/10.2307/2410218> PMID: 28567994
17. Wagner A, Stadler PF. Viral RNA and evolved mutational robustness. *Journal of Experimental Zoology*. 1999; 285(2):119–127. [https://doi.org/10.1002/\(SICI\)1097-010X\(19990815\)285:2%3C119::AID-JEZ4%3E3.0.CO;2-D](https://doi.org/10.1002/(SICI)1097-010X(19990815)285:2%3C119::AID-JEZ4%3E3.0.CO;2-D) PMID: 10440723
18. Elena SF, Carrasco P, Daròs JA, Sanjuán R. Mechanisms of genetic robustness in RNA viruses. *EMBO Reports*. 2006; 7(2):168–173. <https://doi.org/10.1038/sj.embor.7400636> PMID: 16452927
19. Elena SF. RNA virus genetic robustness: possible causes and some consequences. *Current Opinion in Virology*. 2012; 2(5):525–530. <https://doi.org/10.1016/j.coviro.2012.06.008> PMID: 22818515
20. Llaure AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology*. 2013; 11(5):327–336. <https://doi.org/10.1038/nrmicro3003> PMID: 23524517
21. Cuevas JM, Durán-Moreno M, Sanjuán R. Multi-virion infectious units arise from free viral particles in an enveloped virus. *Nature Microbiology*. 2017; 2(7). <https://doi.org/10.1038/nrmicrobiol.2017.78> PMID: 28530650
22. Sardanyés J, Elena S. Error threshold in RNA quasispecies models with complementation. *J Theor Biol*. 2010; 265:278–86. <https://doi.org/10.1016/j.jtbi.2010.05.018> PMID: 20493884
23. Martínez F, Sardanyés J, Elena SF, Daròs JA. Dynamics of a Plant RNA Virus Intracellular Accumulation: Stamping Machine vs. Geometric Replication. *Genetics*. 2011; 188(3):637–646. <https://doi.org/10.1534/genetics.111.129114> PMID: 21515574

24. Montville R, Froissart R, Remold SK, Tenaillon O, Turner PE. Evolution of Mutational Robustness in an RNA Virus. *PLoS Biology*. 2005; 3(11):e381. <https://doi.org/10.1371/journal.pbio.0030381> PMID: 16248678
25. Ojosnegros S, García-Arriaza J, Escarmís C, Manrubia S, Perales C, Arias A, et al. Viral genome segmentation can result from a trade-off between genetic content and particle stability. *PLoS Genet*. 2011; 7:e1001344. <https://doi.org/10.1371/journal.pgen.1001344> PMID: 21437265
26. Geller R, Taguwa S, Frydman J. Broad action of Hsp90 as a host chaperone required for viral replication. *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research*. 2012; 1823(3):698–706. <https://doi.org/10.1016/j.bbamcr.2011.11.007> PMID: 22154817
27. van Nimwegen E, Crutchfield JP, Huynen M. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences*. 1999; 96(17):9716–9720. <https://doi.org/10.1073/pnas.96.17.9716> PMID: 10449760
28. Wilke CO. Selection for fitness versus selection for robustness in RNA secondary structure folding. *Evolution*. 2001; 55(12):2412–2420. <https://doi.org/10.1111/j.0014-3820.2001.tb00756.x> PMID: 11831657
29. Codoñer FM, Darós JA, Solé RV, Elena SF. The Fittest versus the Flattest: Experimental Confirmation of the Quasispecies Effect with Subviral Pathogens. *PLoS Pathogens*. 2006; 2(12):e136. <https://doi.org/10.1371/journal.ppat.0020136> PMID: 17196038
30. Smerlak M. Neutral quasispecies evolution and the maximal entropy random walk. *Science Advances*. 2021; 7(16):eabb2376. <https://doi.org/10.1126/sciadv.abb2376> PMID: 33853768
31. Sanjuán R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010; 365(1548):1975–1982. <https://doi.org/10.1098/rstb.2010.0063> PMID: 20478892
32. O'Fallon BD, Adler FR, Proulx SR. Quasi-species evolution in subdivided populations favours maximally deleterious mutations. *Proceedings of the Royal Society B: Biological Sciences*. 2007; 274(1629):3159–3164. <https://doi.org/10.1098/rspb.2007.1228> PMID: 17939983
33. LaBar T, Adami C. Evolution of drift robustness in small populations. *Nature Communications*. 2017; 8(1). <https://doi.org/10.1038/s41467-017-01003-7> PMID: 29044114
34. Krakauer DC, Plotkin JB. Redundancy antiredundancy, and the robustness of genomes. *Proceedings of the National Academy of Sciences*. 2002; 99(3):1405–1409. <https://doi.org/10.1073/pnas.032668599>
35. Kimmel M, Axelrod DE. *Branching Processes in Biology*. Springer New York; 2002. Available from: <https://doi.org/10.1007%2Fb97371>.
36. Haccou P, Jagers P, Vatutin VA. *Branching Processes*. Cambridge University Press; 2005. Available from: <https://doi.org/10.1017%2Fcb09780511629136>.
37. Restif O. Evolutionary epidemiology 20 years on: Challenges and prospects. *Infection Genetics and Evolution*. 2009; 9(1):108–123. <https://doi.org/10.1016/j.meegid.2008.09.007> PMID: 18977460
38. Elena S, Solé R, Sardanyés J. Simple genomes, complex interactions: epistasis in RNA virus. *Chaos*. 2010; 20:026106. <https://doi.org/10.1063/1.3449300> PMID: 20590335
39. Sanjuán R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Philos Trans R Soc Lond B Biol Sci*. 2010; 365:1975–82. <https://doi.org/10.1098/rstb.2010.0063>
40. Zwart MP, Hemerik L, Cory JS, de Visser JAGM, Bianchi FJJA, Oers MMV, et al. An experimental test of the independent action hypothesis in virus–insect pathosystems. *Proceedings of the Royal Society B: Biological Sciences*. 2009; 276(1665):2233–2242. <https://doi.org/10.1098/rspb.2009.0064> PMID: 19324752
41. Zwart MP, Darós JA, Elena SF. One Is Enough: In Vivo Effective Population Size Is Dose-Dependent for a Plant RNA Virus. *PLoS Pathogens*. 2011; 7(7):e1002122. <https://doi.org/10.1371/journal.ppat.1002122> PMID: 21750676
42. Bergstrom CT, McElhany P, Real LA. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proceedings of the National Academy of Sciences*. 1999; 96(9):5095–5100. <https://doi.org/10.1073/pnas.96.9.5095> PMID: 10220424
43. Elena SF, Sanjuán R, Bordería AV, Turner PE. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. *Infection Genetics and Evolution*. 2001; 1(1):41–48. [https://doi.org/10.1016/S1567-1348\(01\)00006-5](https://doi.org/10.1016/S1567-1348(01)00006-5) PMID: 12798049
44. Zwart MP, Elena SF. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annual Review of Virology*. 2015; 2(1):161–179. <https://doi.org/10.1146/annurev-virology-100114-055135> PMID: 26958911

45. Miyashita S, Ishibashi K, Kishino H, Ishikawa M. Viruses roll the dice: the stochastic behavior of viral genome molecules accelerates viral adaptation at the cell and tissue levels. *PLoS Biol.* 2015; 13: e1002094. <https://doi.org/10.1371/journal.pbio.1002094> PMID: 25781391
46. Burch C, Chao L. Evolvability of an RNA virus is determined by its mutational neighbourhood. *Nature.* 2000; 406:625–8. <https://doi.org/10.1038/35020564> PMID: 10949302
47. Cervera H, Lalić J, Elena S. Efficient escape from local optima in a highly rugged fitness landscape by evolving RNA virus populations. *Proc Biol Sci.* 2016; 283. <https://doi.org/10.1098/rspb.2016.0984> PMID: 27534955
48. Lan Y, Trout A, Weinreich DM, Wylie CS. Natural selection can favor ratchet robustness over mutational robustness. *BioRxiv.* 2017.
49. Wilke C, Wang J, Ofria C, Lenski R, Adami C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature.* 2001; 412:331–3. <https://doi.org/10.1038/35085569> PMID: 11460163
50. Sanjuan R, Moya A, Elena SF. The contribution of epistasis to the architecture of fitness in an RNA virus. *Proceedings of the National Academy of Sciences.* 2004; 101(43):15376–15379. <https://doi.org/10.1073/pnas.0404125101>
51. Carrasco P, de la Iglesia F, Elena SF. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in Tobacco etch virus. *Journal of virology.* 2007; 81(23):12979–12984. <https://doi.org/10.1128/JVI.00524-07> PMID: 17898073
52. Kondrashov AS. Muller's ratchet under epistatic selection. *Genetics.* 1994; 136(4):1469–1473. <https://doi.org/10.1093/genetics/136.4.1469> PMID: 8013919
53. Tromas N, Zwart MP, Lafforgue G, Elena SF. Within-host spatiotemporal dynamics of plant virus infection at the cellular level. *PLoS Genet.* 2014; 10(2):e1004186. <https://doi.org/10.1371/journal.pgen.1004186> PMID: 24586207
54. Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. Evolution and detection of genetic robustness. *Evolution.* 2003; 57(9):1959–1972. <https://doi.org/10.1111/j.0014-3820.2003.tb00377.x> PMID: 14575319
55. Gomulkiewicz R, Holt RD. When does evolution by natural selection prevent extinction? *Evolution.* 1995; p. 201–207. <https://doi.org/10.2307/2410305> PMID: 28593677
56. Gonzalez A, Ronce O, Ferriere R, Hochberg ME. Evolutionary rescue: an emerging focus at the intersection between ecology and evolution; 2013.
57. Anciaux Y, Lambert A, Ronce O, Roques L, Martin G. Population persistence under high mutation rate: from evolutionary rescue to lethal mutagenesis. *Evolution.* 2019; 73(8):1517–1532. <https://doi.org/10.1111/evo.13771> PMID: 31134614
58. Ancel LW, Fontana W. Plasticity, evolvability, and modularity in RNA. *Journal of Experimental Zoology.* 2000; 288(3):242–283. [https://doi.org/10.1002/1097-010X\(20001015\)288:3%3C242::AID-JEZ5%3E3.0.CO;2-O](https://doi.org/10.1002/1097-010X(20001015)288:3%3C242::AID-JEZ5%3E3.0.CO;2-O) PMID: 11069142
59. Domingo-Calap P, Pereira-Gómez M, Sanjuán R. Selection for thermostability can lead to the emergence of mutational robustness in an RNA virus. *Journal of Evolutionary Biology.* 2010; 23(11):2453–2460. <https://doi.org/10.1111/j.1420-9101.2010.02107.x> PMID: 20831733
60. Butković A, González R, Cobo I, Elena SF. Adaptation of turnip mosaic potyvirus to a specific niche reduces its genetic and environmental robustness. *Virus Evolution.* 2020; 6(2):veaa041. <https://doi.org/10.1093/ve/veaa041> PMID: 32782826
61. Stern A, Bianco S, Te Yeh M, Wright C, Butcher K, Tang C, et al. Costs and Benefits of Mutational Robustness in RNA Viruses. *Cell Reports.* 2014; 8(4):1026–1036. <https://doi.org/10.1016/j.celrep.2014.07.011> PMID: 25127138
62. Peck JR. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics.* 1994; 137(2):597–606. <https://doi.org/10.1093/genetics/137.2.597> PMID: 8070669