Variability in mutational fitness effects prevents full lethal transitions in large quasispecies populations

Josep Sardanyés1,2, Carles Simó3, Regina Martínez4, Ricard V. Solé1,2,5 & Santiago F. Elena6

1ICREA-Complex Systems Lab, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain, 2Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain, 3Departament de Matemàtica Aplicada i Anàlisi (Universitat de Barcelona), Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain, 4Departament de Matemàtiques, Edifici C. Facultat de Ciències (Universitat Autònoma de Barcelona) 08193, Bellaterra, Spain, 5Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, 6Instituto de Biología Molecular y Celular de Plantas (CSIC-UPV), Ingeniero Fausto Elio s/n, 46022 València, Spain.

The distribution of mutational fitness effects (DMFE) is crucial to the evolutionary fate of quasispecies. In this article we analyze the effect of the DMFE on the dynamics of a large quasispecies by means of a phenotypic version of the classic Eigen's model that incorporates beneficial, neutral, deleterious, and lethal mutations. By parameterizing the model with available experimental data on the DMFE of Vesicular stomatitis virus (VSV) and Tobacco etch virus (TEV), we found that increasing mutation does not totally push the entire viral quasispecies towards deleterious or lethal regions of the phenotypic sequence space. The probability of finding regions in the parameter space of the general model that results in a quasispecies only composed by lethal phenotypes is extremely small at equilibrium and in transient times. The implications of our findings can be extended to other scenarios, such as lethal mutagenesis or genomically unstable cancer, where increased mutagenesis has been suggested as a potential therapy.

Eigen’s model of molecular quasispecies1–3, initially conceived to explore prebiotic evolution, has played a key role in understanding the population dynamics and evolution of RNA viruses4–7. Due to the complexity of these pathogens, some theoretical assumptions and predictions fail to catch important properties of viral dynamics and adaptation8. For instance, assumptions made considering simple fitness landscapes generally predict the presence of transitions towards error threshold or extinction regimes as mutation rates increase9–11. However, biological evolution proceeds on more complex, rugged fitness landscapes3,12,13 and, outstandingly, not all mutations exert the same effect on viral fitness14–15.

Efforts have been made to expand the basic quasispecies theory by relaxing its original assumptions and incorporating more virus-realistic features: fitness landscapes with multiple peaks16, neutrality and robustness17,18, spatial effects19–21, complementation during coinfection22,23, and different modes of replication and epistasis24–26, among others. Despite all these modeling approaches, there is still a major factor that remains poorly explored by the standard quasispecies model: realistic distributions of mutational fitness effects (DMFE). DMFE may exert a strong impact in the evolutionary dynamics of a large quasispecies population. Therefore, our aim in this study is to provide a broader framework by incorporating experimentally available DMFE within quasispecies evolutionary dynamics.

The fitness effects of mutations are central to evolution25–27. Several experiments quantifying the fraction of spontaneous mutations as either selectively beneficial, neutral, deleterious or lethal have been performed in, e.g., Arabidopsis thaliana28, Drosophila melanogaster29–30, bacteria31, and viruses32–33. The efforts made by virologists in describing DMFE have not been mirrored by theoretical research in quasispecies populations. As yet, few models have explicitly considered the DMFE in asexual replicators34,35, and none of them have parameterized the standard quasispecies model with experimentally available data on DMFE. Furthermore, the impact of the DMFE on the error threshold and how they may promote or constrain the generation of deleterious or lethal phenotypes in a large quasispecies population (e.g., in persistently infected hosts) remains unknown.

The error threshold is a theoretical average mutation rate that sets a maximum limit for maintenance of genetic information encoded by a replicating system36. Usually, the error threshold takes place when the pool of mutants displaces the wild-type sequence due to increased mutation. It is important to mention that the error threshold...
mutations (no phenotypic reversion) can be studied in our model by setting \( \delta = 0 \) (being \( \delta \) the probability of phenotypic reversion in the quasispecies, see Model and Methods section). Under our model assumptions, together with \( \delta = 0 \) and considering no production of beneficial mutations i.e., \( \lambda_B \neq 0 \) (see Model section), Eqs. (1) have two equilibrium points (given as \( Q_3 \) and \( Q_4 \) in Section S4 of Supplementary Information). When no phenotypic reversion is allowed (i.e., \( \delta = 0 \)), the relevant fixed point for VSV data is given by \( Q_3 \) (see SI Section S4). This fixed point, which is an attractor under the parameter values for VSV, can be represented for the sake of clarity as: \( x_i^* = 0 \), \( x_i^{1100} \), \( x_i^{1110} \), \( x_i^{1101} \), \( x_i^{1111} \), where \( k \in \mathbb{H}^+ \{1100, 1110, 1101, 1111\} \), and

\[
\begin{align*}
  x_i^{1100} &= s_D(1 - \mu)/\Psi, \\
  x_i^{1110} &= (1 + s_B)\lambda_B(1 - \mu)/\Psi, \\
  x_i^{1101} &= \lambda_D\mu s_B(1 - \mu)/\Psi, \\
  x_i^{1111} &= \lambda_D\mu(1 - s_D + s_B)\lambda_B/\Psi,
\end{align*}
\]

with \( \Psi = s_D(1 - \mu)^2 + [(1 + s_B)\lambda_B(1 - \mu) + (1 - s_D + s_B)\lambda_D\mu s_B]/\Psi \). Here \( s_D \) and \( s_B \) are the selection coefficients tied to deleterious and beneficial mutations, \( \mu \) is mutation rate, and \( \lambda_D \) and \( \lambda_B \) are the frequencies of production of deleterious and lethal phenotypes during the process of replication and mutation (see Model and Methods section for further details on the parameters of the model). The previous equilibrium point involves the persistence of four different sequences: \( x_i^{1100} \), \( x_i^{1110} \), \( x_i^{1101} \), \( x_i^{1111} \). That is, when \( \delta = 0 \) and \( \lambda_B > 0 \) (i.e., no phenotypic reversion for VSV), 12 of the 16 different sequences of the quasispecies will asymptotically achieve extinction, and four types of sequences will only compose the quasispecies: a neutral mutant with the beneficial phenotype \( x_i^{1100} \) and the sequence with neutral, beneficial, and deleterious mutations \( x_i^{1111} \). The other two sequences have the lethal phenotype (i.e., \( x_i^{1101} \) and \( x_i^{1111} \)). These results perfectly match with the time series shown in Figure S1. Moreover, projections of the dynamics in the simplex suggest that

**Results**

We refer the reader to the Model and Methods section and to Section 1 in the Supplementary Information for a detailed description of the mathematical model and, in particular, for the notation and meaning of variables and parameters that we use in what follows.

**Dynamics with realistic DMFE.** First, we present results on the model Eqs. (1) incorporating DMFE data for VSV and TEV and considering that phenotypic reversions (via backward or compensatory mutations) are not possible. Previous theoretical work has assumed that the likelihood of backward mutations was extremely small due to the length of RNA viral genomes\(^{14,19,22}\), while other models have incorporated backward mutations in the dynamics of quasispecies\(^{16,34}\). The case with no backward

---

**Figure 1** | Population structure of the phenotypic quasispecies. Sequences can carry beneficial (B), neutral (N), deleterious (D), and lethal (L) mutations (bit 0 means non-mutated, and bit 1 means mutated). (a) Phenotypic sequence space where ball size indicates strings’ replicative fitness. The inner green cube contains the lethal mutants. The fitness of some strings is shown on a fitness landscape with dashed arrows. The landscape has three different peaks available through mutation: wild-type (W), beneficial (B), and deleterious (D) peaks, while the ground (light blue) indicates the lethal (L) phenotype with zero replication. Strings \( x_i^{1100} \) and \( x_i^{1110} \) will occupy peaks W, B, or D, depending on the values of the selection coefficients \( s_D \) and \( s_B \). (b) Example of replication with mutation for string \( x_i^{1110} \) with replication rate \( A_i^{1100} = 1 + s_D - s_B \). Gray arrows denote mutation transitions. (c) Sequences of the quasispecies (1st column), their phenotypes (2nd column), and their fitnesses determined by the replication rates, \( A_i \) (3rd column).
such a fixed point is globally stable (Figure S2), which has been confirmed analytically in SI Section S4.

The effect of increasing mutation involves the extinction of the non-mutated genomes at equilibrium (Figure 2). Such extinction is mainly due to outcompetition by the beneficial phenotype $x_{1100}$, which is present in the quasispecies in a wide range of mutation rates, independently of the values of the selection coefficients $s_B$ and $s_D$ [Figure 2(a) and (c)]. Two of the sequences found at equilibrium are lethal mutants, but the quasispecies for VSV is mainly dominated by the neutral mutants carrying a beneficial mutation and by sequence $x_{1110}$ in a wide range of mutation rates. The case with $s_B = 0.25$ and $s_D = 0.9$ actually makes the sequence $x_{1100}$ much more dominant in a large range of mutational values [Figure 2(c)]. For more details on the dynamics using VSV DMFE data see SI Section S4.

A qualitatively different dynamics is found for the DMFE data for TEV without phenotypic reversions. Here $\lambda_E = 0$ gives rise to a different fixed point given by $Q_1$ (see SI Section S4). For this case, the equilibrium also consists of four sequences: $x_{0100}, x_{0110}$, and the lethals $x_{0101}$ and $x_{0111}$. Similarly to VSV, the dominant sequences at increasing mutation are the sequences with only neutral and with both neutral and deleterious mutations. Lethal sequences become abundant only when mutation rate is extremely high [Figure 2(b)]. Some examples of the time dynamics and phase portraits for TEV DMFE data can be found in Figs. S3 and S4.

Interestingly, our results reveal that the fittest phenotype does not go extinct as we might expect. The usual error threshold sets in when $\mu = 1 - [w_{\text{unfit}}/w_{\text{fit}}]$, where $w_{\text{unfit}}$ and $w_{\text{fit}}$ are, respectively, the lowest and the highest fitness of the sequences$^{30}$. For example, according to this prediction, we would expect the extinction of the sequence 1100 for VSV when $\mu > \mu_c = 1 - [(1 - 0.25)/1] = 0.25$. However, the panels (a) and (c) of Figure 2 show persistence of 1100 at this mutation rate and higher. Hence, the simple expectation previously defined fails. The reason for this phenomenon may be due to the fact that the unfit subpopulation (1110) is constantly loosing members to the lethal class (1111), and then $w_{\text{unfit}}$ should be weighted in the previous calculation to include these lethal members, which might prevent the error catastrophe.

When phenotypic reversions are allowed, i.e., $\delta \neq 0$, the non-mutated strings (i.e., wild-type string $x_{0000}$) also decrease their population density dramatically as mutation is slightly increased. This finding is common for both VSV and TEV DMFE data. However, such a population is maintained by keeping its equilibrium in small populations for VSV DMFE data, i.e., $x_{0000} \lesssim 10^{-3}$, for the whole range of mutation rates (see Figure 3 and panels (d–f) in Figure S1). This effect also takes place for different values of the selection coefficients but is less dramatic for TEV data, where $x_{0000} \approx 10^{-3}$ (Figure 3 and panels (d–f) in Figure S3). A plausible explanation for this effect is that as TEV does not produce beneficial mutants, competition is weaker and thus the wild-type phenotype can persist even for high mutation rates. For VSV, however, the equilibrium populations of the strings with the beneficial phenotype (i.e., sequences $x_{1000}, x_{1100}$) undergo a dramatic increase for very low mutation rates in all the scenarios analyzed in Figure 3. Typically, such sequences decrease their population numbers at increasing mutation rates (see also Figure S5). Hence, for both VSV and TEV examples, when phenotypic reversions are allowed, no error thresholds exist at which the wild-type sequence (and its neutral mutants) disappear from the population. For the particular case considering phenotypic reversion, explicit analytical approximations can be derived assuming that the probability of phenotypic reversion is small (i.e., $\delta \approx 0$). We refer the reader to Supplementary Information Sections S5 and S6 for detailed calculations.

We notice that for VSV there is an available beneficial mutation, and since reversions are rare, both mutational and selective pressures favor it, and all sequences at equilibrium have 1 in the first position. Similarly, the neutral bit is favored by mutation and neutral selection; thus even sequence has it as well. For TEV the situation is similar, but since there are no beneficial mutations, the first bit is always 0. In either case, this is essentially a 2-locus problem.

In the previous analyses we have considered experimental DMFE to characterize the dynamics of Eq. (1). Since the proportion of mutations that are effectively neutral, beneficial, and deleterious may vary for other species, we provide analytical and numerical results of the model as a function of all parameters, including arbitrary values of the DMFE, given by $\lambda_E$. These analyses can be found in Supplementary Information Sections S4, S5, S6, and S7 (see also next sections).

**Figure 2 | Impact of increasing mutation rate in the population equilibria without phenotypic reversion.** Here we use $\delta = 0$ (i.e., no phenotypic reversion) and DMFE data for *Vesicular stomatitis virus* (VSV)$^{34}$ and *Tobacco etch virus* (TEV)$^{35}$. In (a) and (c), we display the equilibrium populations for VSV: $x_{1100}$ (black); $x_{1110}$ (red); $x_{1010}$ (green); and $x_{1111}$ (blue). In (b) we show equilibrium populations for TEV: $x_{1000}$ (purple); $x_{1100}$ (cyan); $x_{1010}$ (maroon); and $x_{1111}$ (dark green). For both VSV and TEV, all other variables have zero population numbers at equilibrium. We used: (a) $s_B = 1 > s_D = 0.25$; (b) $s_B = s_D = 0.5$; and (c) $s_B = 0.25 < s_D = 0.9$.

**Exploration of the parameter space.** The previous results without phenotypic reversion reveal the existence of error thresholds shifting the quasispecies towards beneficial, neutral, deleterious, and lethal regions of the phenotypic sequence space in an asymmetric manner. The error threshold has been discussed as a phenomenon causing the loss of viral genetic information due to increased mutation$^{19}$. Actually, the error threshold is an evolutionary transition in sequence space that can delay or even prevent extinction by moving the population towards genotypes that are robust to extinction$^{19}$, as our results indicate, i.e., we characterized persistence of beneficial and neutral phenotypes at equilibrium for a wide range of mutation rates (see Figures 2 and 3).

In this context, our theoretical approach allows us to address the following question: what is the likelihood of finding scenarios in which genomes fail to replicate and thus the viral population is only composed of deleterious and/or lethal phenotypes? To explore this question, let us define three different scenarios given by a quasispecies with different equilibrium population values, $N^*$, as follows:

- **Scenario (A):** $N^*$ lethal mutants: $\sum_{i \in T} x_i = N^*$, i.e., sum of the populations of lethal sequences.
- **Scenario (B):** $N^*$ lethal plus deleterious mutants, $\sum_{i \in T \cup J} x_i = N^*$, i.e., sum of scenarios (A) and (C).
Figure 3 | Population equilibria for sequences and phenotypes at increasing mutation considering phenotypic reversion. We here consider phenotypic reversion ($\delta = 10^{-4}$) and DMFE data for VSV and TEV (in linear-log scale) with: (a) $s_D = s_L = 0.5$ and (b) $s_D = 0.25 < s_L = 0.9$. In the first row we display sequences $x_{0000}^*$ (black line), and the pool of mutants in red lines (except for strings $x_{1100}$ (dotted red line), $x_{1000}$ (dashed red line); and $x_{1110}$ (blue line)). For clarity, phenotypes in the second row are grouped as follows: sequences $x_{0000}^*$ plus $x_{0100}$ (wild-type, W, in black); $x_{0100}$ plus $x_{1100}^*$ (beneficial, B, in green); $x_{0100}$ plus $x_{0110}$ (deleterious, D, in blue); and lethal ($L$ in red) with the sum of all sequences with bit 1 in the last position; and $x_{0100}$ plus $x_{1110}$ (sequences with beneficial and deleterious mutations, in violet). Sequences $x_{0100}$ and $x_{1110}$ can be included in phenotypic classes $W$, $B$, or $D$, depending on the values of the selection coefficients [i.e., $s_D = s_L (W)$, $s_D < s_L (B)$, and $s_D > s_L (D)$].

- **Scenario (C):** $N^*$ deleterious mutants, $\sum_{B \in D} x_i = N^*$, i.e., sum of the sequences with the deleterious phenotypes, and, if $s_D > s_B$, adding also the sequences with both deleterious and beneficial mutations with fitness $A = 1 + s_B - s_D$.

All our previous analyses using VSV and TEV data suggest that the possibility to push the entire quasispecies towards lethality by increasing mutation (scenario (A) with $N^* = 1$) may be very unlikely due to variable DMFE. To test the generality of this result, we perform an extensive search in the parameter space of Eqs. (1), which allows us to identify those parameter combinations fulfilling scenarios (A), (B), and (C) for different values of $N^*$. To do so, we use a Monte-Carlo (MC) algorithm to randomly sample the parameter space of Eqs. (1) (see Model and Methods section). The results obtained without phenotypic reversion are shown in Table I and in Figure 4. We found that the mean percentage of parameter combinations, $(p^*) \cdot 100$, that push the quasispecies towards lethal vertices of the phenotypic space is extremely small (see parameter spaces $(\lambda_1, \mu)$ and $(\mu, \lambda_B)$ for scenario (A) in Figure 4). For instance, such a percentage is about 1.88% using $N^* \approx 0.9$. As expected, the parameter values responsible of increasing populations of lethal sequences are mainly those combinations involving extremely high mutation rates, although such mutation can diminish at increasing values of $\lambda_1$. However, as we display in the histogram of Figure 4, the likelihood to push the whole quasispecies towards lethality is very small. Specifically, the histogram displays the mean percentage of parameter combinations for each value of $N^*$ and scenario, averaged over 100 independent replicas, where each replica corresponds to $M = 10^6$ iterations of the MC algorithm.

Our results also reveal that when beneficial mutations are less common, mutation rate does not need to be so high to produce some fraction of the population composed by lethal sequences, although it is also very unlikely to have a quasispecies fully composed by lethal sequences (it simply will not replicate). We also found that the selection coefficients have a little effect in scenario (A), since no clear pattern is shown for the different values of $N^*$ in the parameter space projection given by $(s_D, s_L)$ that is displayed in Figure 4.
The previous estimates of the percents of the quasispecies towards population values \(N^*\) for scenarios (A), (B), and (C) were computed at equilibrium. But, how do these probabilities behave in transient time? To answer this question, we performed similar analyses for \(N(t)\) instead of \(N^*\). That is, we used the MC algorithm to quantify the fraction of parameter combinations involving different population values for scenarios (A), (B), and (C) at a given time, \(t\) (we used \(t = 10^2, t = 10^3, t = 10^4\)). This strategy allows us to explore the likelihood of these scenarios taking place also during transients. The results, displayed in Supplementary Information Table S1, showed that such percentages are indeed smaller during transients, especially when \(\delta = 0\). For instance, the probability of finding parameter combinations that push the quasispecies towards scenario (A) considering \(N(t = 100) \geq 0.9\), is \(p(t) = 0.0066\), which is approximately 3-fold lower than the same probability computed at equilibrium. As expected, such probabilities get close to the probabilities computed at equilibrium for larger values of time. Similar results were found for scenarios (B) and (C) (see Supplementary Information Table S1).

The ANOVA of data from Table S1, found a net effect of time (\(P < 0.0001\)) on the percentage for each scenario (\(P < 0.0001\)). For scenario (A), \(p(t)\) for the first evaluated time (i.e., \(t = 10^2\)) is significantly lower than for the other two scenarios, which remain the same. In scenario (B), this difference was even smaller, and in (C) this difference was not found. These data indicate that for \(t = 10^2\) the convergence to equilibrium in scenario (A) has not been reached and thus the system is in the transient state, but for scenarios (B) and (C) the system is almost at equilibrium. We notice that the values of \(p(t)\) displayed in Table S1, are typically lower in transient time, meaning that the probability of pushing the quasispecies toward lethal vertices of the phenotypic sequence space is smaller during the transient time, although such values closer to equilibrium also remain very small, specially for large values of \(N(t)\).

Table 1: Statistics obtained from the Monte Carlo (MC) exploration of the parameter space of Eqs. (1). For \(\delta = 0\) (no phenotypic reversion) and for \(\delta > 0\) (phenotypic reversion), we computed the mean percentage (±SD) of parameter combinations, \(\langle p^* \rangle \cdot 100\), pushing the quasispecies towards scenarios (A), (B), and (C) for different population equilibria, \(N^*\). Each data value was averaged over 100 independent replicas, and, for each replica, we ran \(M = 10^6\) MC iterations. The same data represented in histograms for \(\delta = 0\) and \(\delta > 0\) are displayed, respectively, in Figure 4 and Figure S6(b).

| Scenario | \(\geq N^*\) | No phenotypic reversion \((\delta = 0)\) | Phenotypic reversion \((\delta > 0)\) |
|----------|-------------|---------------------------------|---------------------------------|
| (A)      | 0.6         | 10.7471 ± 0.0031                 | 7.3394 ± 0.0272                 |
|          | 0.7         | 7.1483 ± 0.0026                  | 4.0553 ± 0.0212                 |
|          | 0.8         | 4.2387 ± 0.0019                  | 1.7490 ± 0.0124                 |
|          | 0.9         | 1.8815 ± 0.0012                  | 0.3808 ± 0.0064                 |
| (B)      | 0.6         | 23.7100 ± 0.0039                 | 20.27 ± 0.0422                  |
|          | 0.7         | 17.5503 ± 0.0036                 | 14.161 ± 0.0352                 |
|          | 0.8         | 12.1097 ± 0.0031                 | 8.544 ± 0.0264                  |
|          | 0.9         | 6.8064 ± 0.0024                  | 3.478 ± 0.0183                  |
| (C)      | 0.6         | 3.3089 ± 0.0018                  | 3.0518 ± 0.0154                 |
|          | 0.7         | 1.4216 ± 0.0012                  | 1.148 ± 0.0103                  |
|          | 0.8         | 0.4308 ± 0.0006                  | 0.2547 ± 0.0048                 |
|          | 0.9         | 0.0572 ± 0.0002                  | 0.0145 ± 0.0011                 |

Transient regimes as a function of \(\delta\) and other parameters. We are particularly interested in the effect of the probability of phenotypic reversion, \(\delta\), on transient times (see Model and Methods section). Our results indicate that the smaller \(\delta\) the longer is the time needed to approach the fixed point at the given distance \(\eta\). This time depends on two things: the component of the initial data in the eigenvector of the matrix \(A\) in the direction of \(x^*\) and on the eigenvalue of \(DP(x^*)\) closest to zero. As discussed in Supplementary Information Section S5, the smaller \(\delta\) the closest is the eigenvalue to zero and, hence, the required time is larger.

It is worth to remark that the fraction of \(M^*\) for which the approximation to distance \(x_0\) to \(x^*\) is not produced at the final integration time \(t_{\text{max}}\) is extremely small: around 0.00002 for \(\delta = 10^{-2}\) and increases to 0.00022 for \(\delta = 0\).

Figure S8(left) shows the results using \(M^* = 10^4\) and \(\eta = 10^{-3}\) for the values \(\delta = 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}\), and \(\delta = 0\), from left to right. On the horizontal axis \(\tau\) is represented in log10 scale. The value of \(t_{\text{max}}\) is set equal to 106.

What happens if \(\eta\) is reduced? Of course, the transient time, until the distance to \(x_0\) is less than \(\eta\), will increase. To illustrate this fact we reproduce in Figure S8(right) the same curves in the left plot for \(\delta = 10^{-5}, 10^{-6}\) and 0 for \(\eta = 10^{-5}\) (in red) and we show the corresponding curves for \(\eta = 10^{-6}\) in blue. We see that in the first two cases the curves shift a little to the right, while in the \(\delta = 0\) case they shift by a large amount. The computations for \(\eta = 10^{-6}\) use again \(t_{\text{max}} = 10^6\) but \(M^*\) is reduced to 106.

Now, while for \(\delta = 10^{-5}\) still only a small fraction, around 0.00005, of cases do not approach \(x^*\) at the distance \(\eta\) at the maximum time, this fraction increases dramatically to 0.076 for \(\delta = 0\).

According to the theory developed in Supplementary Information Sections S4 and S5, if \(\delta > 0\) is small the final approach to \(x^*\) is of the form \(\exp(-\gamma \delta^{1/2})\) where \(\gamma\) is the coefficient of \(\delta^{1/2}\) in Eq. (33) of the Supplementary Information. On the other hand, for \(\delta = 0\) the approach to \(x^*\) is like a constant divided by \(t\).

From the data which allow to produce Figure S8(right), one can compute, by Lagrange interpolation, the values of \(\tau\) for different values of \(\phi(t)\). We have selected values of \(\phi(t)\) of the form 0.1, \(k = 1, \ldots, 9\). Let \(\tau_1\) be the value of \(\tau\) when \(\eta = \eta_1 = 10^{-3}\) and \(\tau_2\) the one for \(\eta = \eta_2 = 10^{-6}\). For \(\delta > 0\) the decrease of \(\eta\) should satisfy, approximately,

\[
\eta_2 = \eta_1 \exp\left(-\gamma \delta^{1/2}(\tau_2 - \tau_1)\right).
\]

The value of \(\gamma\) depends on the average value of the random parameters and also on the value of \(\phi(t)\).

Hence, the decrease of \(\delta^{1/2}\) by a factor of 10 should be compensated by an increase of \(\tau_2 - \tau_1\) also by a factor of 10. This is shown in Figure S9. The red and blue curves show \(\log_{10}\left(\tau_2 - \tau_1\right)\) for different values of \(\phi(t)\). The differences are very close to 1, as expected. On the opposite case, \(\delta = 0\), the green curve shows \(\log_{10}\left(\tau_2 / \tau_1\right)\). It is almost exactly equal to 3, in perfect agreement with the fact that \(\log_{10}(\eta_2 / \eta_1) = -3\).
Figure 4 | Exploration of the parameter space of Eqs. (1) without phenotypic reversion (i.e., $\delta = 0$). We plot two-dimensional parameter space projections using $M = 10^6$ iterations of the MonteCarlo (MC) algorithm displaying parameter values pushing the quasispecies towards scenarios (A), (B), and (C) for different population equilibria, $N^*$: $N^* \geq 0.6$ (cyan), $N^* \geq 0.7$ (violet), $N^* \geq 0.8$ (yellow), and $N^* \geq 0.9$ (maroon). Note that $N^* \geq x^*$ includes all values of $y^*$ (e.g., $N^* \geq 0.6$ includes all other $N^*$ values analyzed). The histogram displays the mean percentage of parameter combinations, $\langle p^* \rangle \cdot 100$, with $\langle p^* \rangle = (c(N^*))/M$, fulfilling scenarios (A), (B), and (C) for the investigated values of $N^*$ (using the same colors for each value of $N^*$ previously described). Each bar was computed over 100 independent replicas of the MC algorithm (each replica run for $M = 10^6$ iterations), standard deviation (SD) not shown (see Table I below).
Summarizing, this analysis and numerical checks allow obtaining good indications on the rate of approach to the fixed point.

Discussion
In this article we have extended Eigen’s quasispecies model using a phenotypic description of the quasispecies that considers variable distributions of mutational fitness effects (DMFE), which are known to be crucial to RNA virus evolution. Although the DMFE has been experimentally quantified for several viruses, previous theoretical models have not deeply analyzed the effect of the DMFE in persistently infected hosts. Moreover, none of them included these experimental data in the standard quasispecies model. Here we intend to cover this gap, paying special attention to the interplay between the DMFE and mutation in possible deleterious or lethal scenarios in large viral quasispecies populations. The interest in characterizing the effects of artificially-increased mutation (mutagenesis) in viral replicators comes from the fundamental prediction of the classic quasispecies theory: error-prone replicators have a limited tolerance to mutations, beyond which an error catastrophe takes place and the genetic information is lost. Several models have shown that when mutation rate is increased, the population moves towards lower fitness classes, and the average fitness of the quasispecies or the average growth rate per generation diminishes. In such a case, for discrete systems, when the average growth rate becomes smaller than one, the population enters into an extinction regime that causes the disappearance of the quasispecies.

Previous works on the induction of mutagenesis in poliovirus and VSV revealed a 100-fold decrease in viral titers. Further experimental evidence confirming the effect of mutagens in decreasing infectivity and replication was later described for HIV-1 in tissue cultures. Similar results were also found for Lymphocytic choriomeningitis virus, hantaviruses, and Hepatitis C virus, among others (see review). Despite these previous investigations, the feasibility of entirely pushing a quasispecies towards lethality by manipulating external parameters such as mutation rates still remains a matter of debate, especially in persistently infected patients, i.e., in quasispecies with large population numbers. Our model, aimed to answer this question, predicted that an external forcing towards lethality of sequences with large population numbers. Our model, aimed to answer this question, predicted that an external forcing towards lethality (e.g., increased mutagenesis) would be unlikely to be found in large quasispecies populations. The generation of different mutational types seems enough to ensure the persistence of viable viral genomes, even if highly deleterious or lethal mutations are constantly produced.

Our theoretical approach does not include other possible sources that could enhance viral extinction or involve significant viral fitness decrease. For instance, we consider soft selection, i.e., no degradation of sequences, because we are mainly interested in the net effects of mutation rate, DMFE, and phenotypic reversions in the viability and lethality of the viral phenotypes. Also, we are not considering discreteness (individual based model) or finite populations with stochasticity (e.g., lethal deletion), or bottlenecks during transmission events that turn on Muller’s ratchet (shown to operate in phages and human immunodeficiency virus type I). As mentioned, the standard quasispecies model (and thus our model as well) does not consider sources of stochasticity, suggested to be important in realistic experimental systems as well as during initial infections, when the number of viral particles can be extremely low. However, as we previously discussed, we considered a deterministic quasispecies model as a proxy to characterize the effect of the DMFE during a persistent infection, for which intrinsic or demographic noise of viral populations might be negligible. Further research should include these factors in quasispecies dynamics together with DMFE.

Our model predicts that increasing mutation rates may not be enough to drive viral quasispecies towards highly deleterious or lethal phenotypes in large populations, as may be those typical of persistent infections. It would be interesting to perform experiments like those described in Refs. 14 and 15, in which the properties of the DMFE were evaluated, but at artificially increased mutation rate. Such experiments would facilitate the comprehension of the role played by the DMFE at increasing mutation in possible lethal scenarios for viral quasispecies in vivo, also allowing to test our prediction. Moreover, these experiments would also give clues about the interplay between mutation and the DMFE, an important issue, that, to the extent of our knowledge, remains unknown.

We notice that our results may also be relevant in the context of lethal mutagenesis. Lethal mutagenesis refers to a viral population’s extinction through an excess of mutations, often promoted by mutagenic nucleotide analogs administered during viral replication. Lethal mutagenesis is a demographic phenomenon that operates in finite populations, not considered in our model. Nonetheless, our model indicates that the chances to enter into a regime of declining viral populations may be very unlikely, since the variability in the DMFE will constantly keep available phenotypes able to successfully replicate in the population.

As a final point, the theoretical framework and the results presented here, developed for viral quasispecies, may also have implications in carcinogenesis. In cancer, the maximum amount of genetic instability that can be tolerated by tumor cells, suggested to also behave as quasispecies, has been proposed. Genomic instability is a hallmark of most malignant tumors, which could be ablated by increased mutagenesis. Our model suggests that the success of such strategies should be weighted by the corresponding distribution of mutational fitness effects, which can be defined by taking into account the distinct roles played by genes affecting proliferation, DNA repair or apoptosis, along with house-keeping. The potential success of suppression of tumor growth through mutagen-induced lethality as well as the resistance of some clones to drugs should be considered as two faces of the unstable tumor dynamics, which may require considering the repertoire of mutational fitness effects.

Methods
Model. To analyze the impact of the DMFE on viral quasispecies we build a phenotypic mathematical model using Eigen’s formulation. The model describes the dynamics of i classes of binary sequences with population numbers xi, able to carry different mutational types that can replicate and mutate according to the differential equations:

\[
\frac{d x_i}{d t} = (1 - \mu)A_i x_i + \mu \sum_j \frac{g_j}{\sum_k g_k} \beta^{-i-j} A_i x_j - \Phi x_i,
\]

with \(i,j \in \mathcal{H}^+\), being \(\mathcal{H}^+\) the configuration or sequence space (i.e., boolean v-dimensional hypercube), with \(\mathcal{H}^+ = \{1, \ldots, 2^v\}\), and \(v = 4\), resulting in a quasispecies composed by \(n = 16\) different sequences. Each sequence has 4 bits, each of them corresponding to each particular type of mutation: beneficial (B, 1st bit), neutral (N, 2nd bit), deleterious (D, 3rd bit), and lethal (L, 4th bit), as we display in Figure 1. Our model becomes a generalization of the classic Eigen’s quasispecies model, which incorporates new parameters to introduce variance in fitness effects due to mutation and, eventually, phenotypic reversion. We refer the reader to Ref. 38 for a stochastic model similar to Eqs. (1).

The first term of Eqs. (1) is the error-free replication of string \(i\) with replication rate \(A_i\), and mutation rate \(\mu\). The second term corresponds to the influx of mutant strings \(j\) from onthant neighbors (denoted as \(\prec \prec j \succ \succ \)) of \(i\) producing string \(j\) by mutation. Parameters \(\beta^{-i-j} \geq 0\), with \(\sum_j \beta^{-i-j} = 1\), denote the fraction of mutations with a fitness effect \(k \in \{B, N, D, L\}\) from sequence \(j\) to sequence \(i\) during error-prone replication. These parameters are used to incorporate the DMFE in our model (see Section Experimental data below). Parameter \(\beta^{-i-j}\) is the probability of occurrence of transitions \(0 \rightarrow 1\) and \(1 \rightarrow 0\) (i.e., phenotypic reversion) from sequence \(j\) to sequence \(i\) during replication, with:

\[
\beta^{-i-j} = \begin{cases} 
1 & \text{if transition is } 0 \rightarrow 1, \\
0 & \text{if transition is } 1 \rightarrow 0.
\end{cases}
\]

We notice that our approach allows us to model the process of mutation considering no backward mutations (i.e., no phenotypic reversion), setting \(\delta = 0\); as well as to model phenotypic reversions due to backward or compensatory mutations (with \(\delta > 0\)).
0. The total population is held constant (i.e., $\sum_{i=1}^{n} x_i = 1$) by a compensating dilution flux with rate coefficient $\Phi$. The outflow term, $\Phi$, is obtained under the constant population condition, i.e., $\sum_{i=1}^{n} x_i = 1$:

$$\Phi = \sum_{i=1}^{n} \left(1 - \mu A_i x_i + \sum_{j=1}^{q} \frac{A_i}{j!} \frac{\partial^j}{\partial x^j} A_i x_i \right).$$

The state space of system (1) is a high-dimensional simplex (i.e., $2^n - 1$ simplex) represented by the nonnegative vectors

$$P^n = \{ \mathbf{x} \geq 0, \sum_{i=1}^{n} x_i = 1, 1, \ldots, n \}.$$  

As initial conditions we will use (if not otherwise specified) $x_{\text{mut}}(0) = 1$, setting all other variables to zero.

All 2^n different strings of the quasispecies can be mapped into a phenotypic space (fitness landscape) with four different fitness classes and three different peaks (Figure 1(a)):

- Peak (W) populated by sequences x0000,1000 (wild-type phenotype and the neutral mutants), with replicative fitness $A_{00001000} = 1$.
- Peak (B) populated by sequences x1100,1010 (beneficial mutants plus the neutral mutants), with replication rate $A_{11001010} = 1 + \gamma$.
- Peak (D) populated by sequences x1000,0010 (deleterious mutant and the neutral mutants), with replicative fitness $A_{10000010} = 1 - \delta p$.
- The lowest replication value of the fitness landscape is given by the lethal phenotypes, (L), given by sequences $x_{\text{mut}} = 0, a, b, c, e \in \{0, 1\}$, with replication rate equal to zero.

We refer the reader to Supplementary Information Section S1 which contains all i equations obtained from (1). Sequences with both beneficial and deleterious mutations (given by $x_{\text{mut}} = 1100, 1010$ with four different fitness classes and three different peaks (W), (B), or (D) depending on the values of the selection coefficients, $s_B$ and $s_D$. Note all fitness classes include the neutral mutants. Equations (1) depend on the parameters $\mu, s_B, s_D, s_L, \lambda_D, \lambda_D, \lambda_L$ and $\delta$. We set $0 < \delta \leq 1, to ensure that beneficial mutations always increase replication rates. We also set $0 < \delta \leq 1, to ensure that beneficial mutations always increase replication rates.

Using a (pseudo-)random number generator with uniform distribution in $[0, 1]$ 4. Then, we generate points with uniform distribution in $[0, 1]$ 4 corresponding to the values of $x_{\text{mut}}$, $x_{\text{mut}}$, $x_{\text{mut}}$, $x_{\text{mut}}$.

It is important to notice that Eigen's model describes the dynamics of explicit genome macromolecules. Inforamtion Section S4) and, hence, of the characteristic polynomial and the eigenvalues, that is an algebraic function. Therefore the volume of the region of interest in $\Sigma$ is bounded by the hypersurface $g(\Omega) = N^* and boundaries of $\Sigma$.

However, even in the case that $g$ is explicitly available, this is of little use because of the dimensionality. Moreover, in the generic case $g$ is not explicitly known. This suggests to use a MonteCarlo (MC) method to compute the volume of interest. The steps are as follows:

- Using a pseudo-random number generator with uniform distribution in $[0, 1]$, we generate four values to obtain a point in $[0, 1]^4$. Then, we generate points with uniform distribution in $\Sigma$, using a standard method.
- When all the parameters are available we look for the dominant zero of the polynomial $p_2$ in (Eq. (24) of the Supplementary Information). Let us denote it as $s_1$, as in Supplementary Information Section S3. Then, we proceed to the computation of the components of the fixed point using the formulas (25) and (26) of the Supplementary Information.

The process is repeated M times (M large, a typical value being $M = 10^8$). Then, for the different values of $N^*$ used in the study we count the number of cases $n(N^*)$ in which $g(\Omega) > N^*$. The desired probability is $p = n(N^*)/M$. In many of our analyses we will represent such a probability as a percentage (i.e., meaning the percentage of parameter combinations driving the quasispecies towards a given scenario for a given value of $N^*$, if not otherwise specified). Alternatively, we can also be interested in transient times (see Results Section), and perform a similar analysis considering a given population value for scenarios (A1, B1, and C1) at a given time, $t$. Then, the previous probability will be represented as $p = n(N(\Omega))/M$.

In a large part of the computations the parameter $\delta$ is not taken with uniform probability but uniform in log scale, say in the range $[\delta_1, \delta_2]$. Then, the value of $\log(d)$ is generated with a uniform distribution in $[\log(d_1), \log(d_2)]$ and proceed as before. It can be also interesting to fix $\delta$ and proceed in a similar way for a set of values of $\delta$. Then we can recover a probability function $p(\delta)$ by interpolation. The desired global probability with respect to a density $f(\delta)$ is obtained by integration and normalization

$$\int_{\delta_1}^{\delta_2} f(\delta) d\delta = \int_{\delta_1}^{\delta_2} \frac{1}{\delta_2 - \delta_1} \times p(\delta) f(\delta) d\delta = \delta.$$ (3)

We pass now to the point b). Given parameters $\mathbf{I}$ we can compute the equilibrium $x_0$. On the other hand, we can start the integration of system (3) of the SI from the initial point $x(0) = (1, 0, \ldots, 0)$ until the current point $x(t)$ reaches a given distance $\eta$ from $x_0$. Typically we have used the value $\eta = 10^{-3}$ and the distance has been computed using the $\| \|$ norm. The integrations are carried out using the Runge-Kutta-Fehlberg RKF78 method, with automatic step size control and local relative tolerance $10^{-12}$. After every step the distance to $x_0$ is checked. This allows obtaining the distribution of transient times and the effect of the different parameters on it, with special emphasis on the role of $\delta$ as is illustrated below and in the Results section.

**Computation of transient times.** In order to see how parameter $\delta$, i.e., phenotypic reversion, affects the transients, we carry out the following experiment. For a fixed value of $\delta$ we generate a random sample of $M^*$ sets of the remaining parameters, as described in Numerical algorithms subsection above. Then, for each set of parameters, the integration of Eq. (2) in the Supplementary Information is started with $x(0) = (1, 0, \ldots, 0)$. The integration is stopped when the distance to the fixed point is small enough (a fixed value $\epsilon$). The distance is measured as $\| x - x_0 \|$. For a given value $\epsilon$ of the time we record the fraction $\phi(\epsilon)$ of $M^*$ such that the distance $\| x - x_0 \|$ reached for a value of $\epsilon < \epsilon$. Then we display $\phi(\epsilon)$ as a function of $\epsilon$ (see Figure S8). Of course, for practical reasons, the integration time is limited to a fixed long $t_{\text{max}}$.

1. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Nature*, **58**, 485–523 (1971).
31. Imhof, M. & Schlotterer, C. Fitness effects of advantageous mutations in evolving populations. *Nature*, 123–126 (2012).
30. Loewe, L. & Charlesworth, B. Inferring the distribution of mutational effects on fitness landscapes. *J. Theor. Biol.*, 123–126 (2012).
29. Manrubia, S. C. Modelling viral evolution and adaptation: challenges and rewards. *Phys. Rev. E.*, 123–126 (2012).
28. Wilke, C. O. Quasispecies theory in the context of population genetics. *BMC Evol. Biol.*, 123–126 (2012).
27. Eigen, M., McCaskill, J. S. & Altman, S. Error threshold in RNA quasispecies models with complex replication modes and infection strategies. *Phys. Rev. E.*, 123–126 (2012).
26. Sanjuan, R., Moya, A. & Elena, S. F. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc. Natl. Acad. Sci. U.S.A.*, 123–126 (2012).
25. Carrasco, P., de la Iglesia, F. & Elena, S. F. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. *J. Virol.*, 123–126 (2012).
24. Sardanyés, J., Sole, R. V. & Elena, S. F. Simple quasispecies models for the survival-of-the-fittest effect: the role of space. *J. Theor. Biol.*, 123–126 (2012).
23. Iranzo, J. & Manrubia, S. C. Evolutionary dynamics of genome segmentation in *Drosophila melanogaster*. *Genetics*, 123–126 (2012).
22. Loewe, L. A. & Charlesworth, B. Inferring the distribution of mutational effects on fitness landscapes. *J. Theor. Biol.*, 123–126 (2012).
21. Sanjuan, R. Mutual fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Phil. Trans. R. Soc. B.*, 123–126 (2012).
20. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.*, 123–126 (2012).
19. Manrubia, S. C. Modelling viral evolution and adaptation: challenges and rewards. *Curr. Op. Virol.*, 123–126 (2012).
18. Keightley, P. D. & Lynch, M. Toward a realistic model of mutations affecting fitness. *Evolution*, 57, 683–685 (2003).
17. Keightley, P. D. The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics*, 138, 1315–1322 (1994).
16. Loewe, L. & Charlesworth, B. Inferring the distribution of mutational effects on fitness in *Drosophila*. *Biol. Lett.*, 2, 426–430 (2006).
15. Imhof, M. & Schlötterer, C. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl. Acad. Sci. U.S.A.*, 98, 1113–1117 (2001).
14. Elena, S. F. & Moya, A. Rate of deleterious mutations and the distribution of its effects on fitness in *Vesicular stomatitis virus*. *J. Virol.*, 12, 1078–1088 (1999).
13. van Opijnen, T., Boerlijst, M. C. & Berkhout, B. Effects of random mutations in the human immunodeficiency virus type 1 transcriptional promoter on viral fitness in different host cell environments. *J. Virol.*, 80, 6678–6685 (2006).
12. Orr, H. A. The distribution of fitness effects among beneficial mutations. *Genetics*, 163, 1519–1526 (2003).
11. Antoneli, F., Bosco, F., Castro, D. & Janini, L. M. Virus replication as a phenotypic version of polynucleotide evolution. *Bull. Math. Biol.*, 75, 602–628 (2013).
10. Perales, C., Iranzo, J., Manrubia, S. & Domingo, E. The impact of quasispecies dynamics on the use of therapeutics. *Trends in Microbiol.*, 20, 595–603 (2012).
9. Schuster, P. Mathematical modeling of evolution. *Solved and open problems. Theory Biosci.*, 130, 71–89 (2011).
8. Lorenz, D. M., Park, J.-M. & Deem, M. W. Evolutionary processes in finite populations. *Phys. Rev. E.*, 87, 022704 (2013).
7. Ball, J. J., Meyers, L. A. & Lachmann, M. Quasispecies made simple. *PLOS Comput. Biol.*, 1, e61 (2005).
6. Bull, J. J., Sanjuán, R. & Wilke, C. O. Theory of lethal mutagenesis for viruses. *J. Virol.*, 81, 2930–2939 (2007).
5. Wylie, C. S. & Shakhnovich, E. L. Mutation induced extinction in finite populations: lethal mutagenesis and lethal isolation. *PLOS Comput. Biol.*, 8, e1002609 (2012).
4. Aguirre, J., Buldá, J. M. & Manrubia, S. C. A trade-off between neutrality and adaptability limits the optimization of viral quasispecies. *J. Theor. Biol.*, 261, 145–151 (2005).
3. Cuesta, J. A., Aguirre, J., Capitán, J. A. & Manrubia, S. C. Struggle for space: viral extinction through competition for cells. *Phys. Rev. Lett.*, 106, 028104 (2011).
2. Domingo, E., Webster, E. & Holland, J. J. Origin and evolution of viruses. *Proc. Natl. Acad. Sci. U.S.A.*, 90, 1487–1497 (1993).
1. Anderson, J. P., Daifuku, R. & Loeb, L. A. Viral error catastrophe by mutagenic nucleosides. *Annu. Rev. Microbiol.*, 58, 183–205 (2004).

Acknowledgments

We thank the members of the Complex Systems Lab as well as Phillip Gerrish, Susanna C. Manrubia, and Ernest Fontich for their helpful comments. The authors acknowledge the computing facilities of the Dynamical Systems Group (Universitat de Barcelona). This work was partially funded by the Botín Foundation (JS, SFE), by the Spanish Secretaria de Estado de Investigación, Desarrollo e Innovación grants MTM2010-16425 (CS, RM) and BFU2012-30805 (SFE), by grant 2009-SGR-67 from the Catalan government (CS, RM), by grant NSF PHY05-51164 (JS, SFE), and by the Santa Fe Institute (RVS, SFE). All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at www.nature.com/scientificreports
Corrigendum: Variability in mutational fitness effects prevents full lethal transitions in large quasispecies populations

Josep Sardanyés, Carles Simó, Regina Martínez, Ricard V. Solé & Santiago F. Elena

Scientific Reports 4:4625; doi: 10.1038/srep04625; published online 09 April 2014; updated 30 June 2015

This Article contains an error in the Acknowledgements section:

“This work was partially funded by the Botín Foundation (JS, RVS), by the Spanish Secretaria de Estado de Investigación, Desarrollo e Innovación grants MTM2010-16425 (CS, RM) and BFU2012-30805 (SFE), by grant 2009-SGR-67 from the Catalan government (CS, RM), by grant NSF PHY05-51164 (JS, SFE), and by the Santa Fe Institute (RVS, SFE).”

should read:

“This work was partially funded by the Spanish grant FIS2012-39288 (RVS), by the Botín Foundation (JS, RVS), by the Spanish Secretaria de Estado de Investigación, Desarrollo e Innovación grants MTM2010-16425 (CS, RM) and BFU2012-30805 (SFE), by grant 2009-SGR-67 from the Catalan government (CS, RM), by grant NSF PHY05-51164 (JS, SFE), and by the Santa Fe Institute (RVS, SFE).”