SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
VIROLOGY
APPLIED MATHEMATICS
EVOLUTIONARY THEORY

Received
3 January 2014

Accepted
21 March 2014

Published
9 April 2014

Correspondence and requests for materials should be addressed to J.S. (josep.sardanes@upf.edu)

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

Josep Sardanyés[1,2], Carles Simó[3], Regina Martínez[4], Ricard V. Solé[1,2,5] & Santiago F. Elena[5,6]

[1]ICREA-Complex Systems Lab, Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain, [2]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Maritim de la Barceloneta 37, 08003 Barcelona, Spain, [3]Departament de Matemàtica Aplicada i Anàlisi (Universitat de Barcelona), Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain, [4]Departament de Matemàtiques, Edifici C. Facultat de Ciències (Universitat Autònoma de Barcelona) 08193, Bellaterra, Spain, [5]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA, [6]Instituto 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 quasispecies[1–3], initially conceived to explore prebiotic evolution, has played a key role in understanding the population dynamics and evolution of RNA viruses[4–7]. Due to the complexity of these pathogens, some theoretical assumptions and predictions fail to catch important properties of viral dynamics and adaptation[8]. For instance, assumptions made considering simple fitness landscapes generally predict the presence of transitions towards error threshold or extinction regimes as mutation rates increase[9–11]. However, biological evolution proceeds on more complex, rugged fitness landscapes[3,12,13] and, outstandingly, not all mutations exert the same effect on viral fitness[14,15].

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 peaks[16], neutrality and robustness[17,18], spatial effects[10,19–21], complementation during coinfection[22,23], and different modes of replication and epistasis[21,24], 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 evolution[25–27]. Several experiments quantifying the fraction of spontaneous mutations as either selectively beneficial, neutral, deleterious or lethal have been performed in, e.g., *Arabidopsis thaliana*[28], *Drosophila melanogaster*[29,30], bacteria[31], and viruses[14,15,32,33]. 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 replicators[34,35], 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 system[36]. 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

implicitly depends on the fitness landscape[37]; e.g., for the single-peak fitness landscape, the error threshold occurs when the homogeneous cloud of mutants with lower fitness outcompetes the wild-type sequence[1,2,10,11]. In order to introduce a more realistic fitness landscape into the quasispecies model, as well as to consider realistic DMFE that determine how mutation moves the quasispecies within this landscape, we analyze a phenotypic quasispecies mathematical model incorporating variable mutational fitness effects (see Figure 1).

Here, we will first investigate the impact of increasing mutation rate on the phenotypic distributions and on the viability of the quasispecies integrating available experimental data on the DMFE for two RNA viruses: the animal pathogen *Vesicular stomatitis virus*[14] and the plant pathogen *Tobacco etch virus*[15]. Second, we will investigate the dynamics for the general model with and without phenotypic reversion, since we are especially interested on a detailed analysis of the dependence of the results with respect to the model parameters. Finally, we perform an extensive search in the parameter space to evaluate the likeliness of finding parameter combinations that impair the success of viral phenotypes, driving the quasispecies towards lethal and deleterious vertices of the phenotypic sequence space.

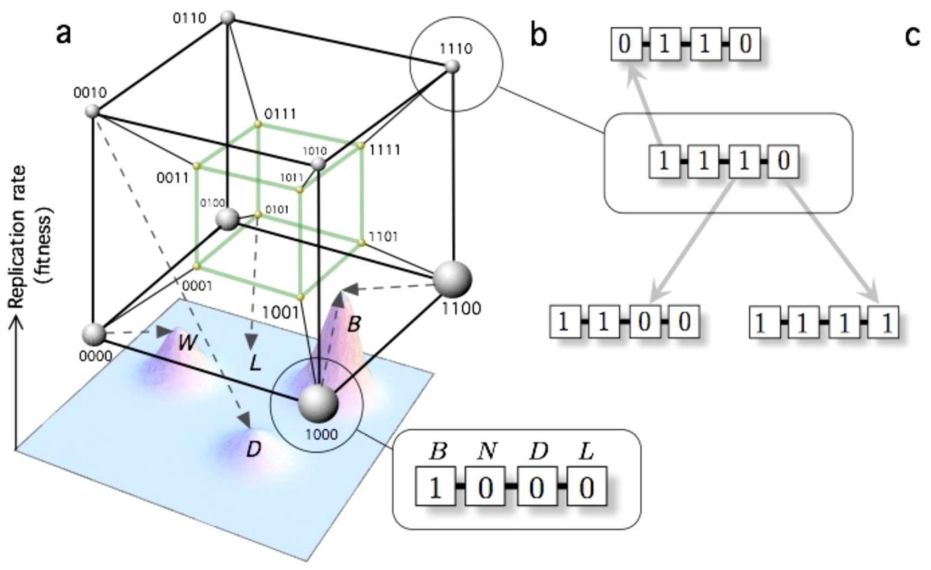
Results

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[10,19,22], while other models have incorporated backward mutations in the dynamics of quasispecies[16,38]. The case with no backward mutations (no phenotypic reversion) can be studied in our model by setting $\delta = 0$ (being δ 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: $(\mathbf{x}_k^* = \vec{0},\ x_{1100}^*,\ x_{1110}^*,\ x_{1101}^*,\ x_{1111}^*)$, where $k \in \mathcal{H}^v \setminus \{1100, 1110, 1101, 1111\}$, and

$$x_{1100}^* = s_D(1-\mu)^2/\Psi,$$
$$x_{1110}^* = (1+s_B)\lambda_D\mu(1-\mu)/\Psi,$$
$$x_{1101}^* = \lambda_L\mu s_D(1-\mu)/\Psi,$$
$$x_{1111}^* = \lambda_L\mu(1-s_D+s_B)\lambda_D\mu/\Psi,$$

with $\Psi = s_D(1-\mu)^2 + [(1+s_B)\lambda_D\mu + s_D\lambda_L\mu](1-\mu) + (1-s_D+s_B)\lambda_D\lambda_L\mu^2$. Here s_D and s_B are the selection coefficients tied to deleterious and beneficial mutations, μ is mutation rate, and λ_D and λ_L 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_{1100}^*, x_{1110}^*, and the lethals x_{1101}^* and x_{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_{1100}^*) and the sequence with neutral, beneficial, and deleterious mutations (x_{1110}^*). The other two sequences have the lethal phenotype (i.e., x_{1101}^* and x_{1111}^*). These results perfectly match with the time series shown in Figure S1. Moreover, projections of the dynamics in the simplex suggest that



Sequence	Phenotype	Fitness (A_i)
0000	wild–type, W	1
0001	lethal, L	0
0010	deleterious, D	$1-s_D$
1000	beneficial, B	$1+s_B$
0100	W	1
0101	L	0
0110	D	$1-s_D$
0111	L	0
0011	L	0
1001	L	0
1010	BD	$1+s_B-s_D$
1011	L	0
1100	B	$1+s_B$
1101	L	0
1110	BD	$1+s_B-s_D$
1111	L	0

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_{1010} and x_{1110} will occupy peaks W, B, or D, depending on the values of the selection coefficients s_B and s_D. (b) Example of replication with mutation for string x_{1110}, with replication rate $A_{1110} = 1 + s_B - s_D$. 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 $\lambda_B = 0$ 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 $\mu = 1 - [w_{unfit}/w_{fit}]$, where w_{unfit} and w_{fit} are, respectively, the lowest and the highest fitnesses of the sequences[39]. 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_{unfit} should be weighted in the previous calculation to include these lethal members, which might prevent the error catastrophe.

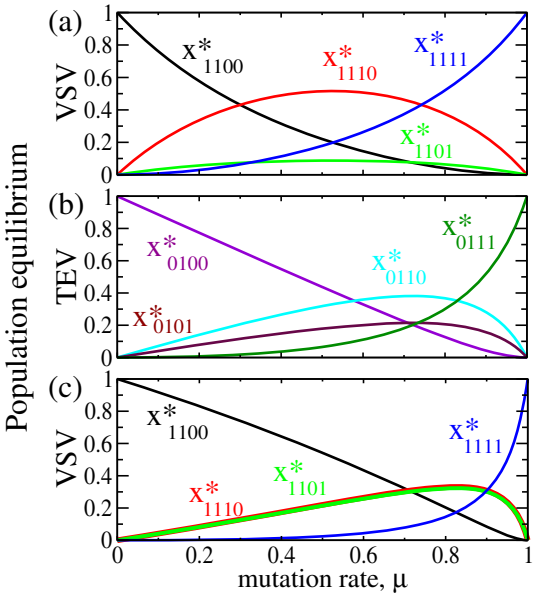


Figure 2 | Impact of increasing mutation rate in the population equilibria without phenotypic reversion. Here we use $\delta = 0$ (i.e., no phenotypic reversion) and DMFE data for *Vesicular stomatitis virus* (VSV)[14] and *Tobacco etch virus* (TEV)[15]. In (a) and (c), we display the equilibrium populations for VSV: x_{1100} (black); x_{1110} (red); x_{1101} (green); and x_{1111} (blue). In (b) we show equilibrium populations for TEV: x_{0100} (violet); x_{0110} (cyan); x_{0101} (maroon); and x_{0111} (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) $s_B = 0.25 < s_D = 0.9$.

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., $x^*_{0000} \lesssim 10^{-7}$, 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 $x^*_{0000} \lesssim 10^{-3}$ (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 $x_{1000,1100}$) 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., $\delta \gtrsim 0$). 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 every 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 λ_k. These analyses can be found in Supplementary Information Sections S4, S5, S6, and S7 (see also next sections).

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[40], 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, $\sum_{i\in L} x_i = N^*$, i.e., sum of the populations of lethal sequences.
- *Scenario (B)*: N^* lethal plus deleterious mutants, $\sum_{i\in L\cup D} x_i = N^*$, i.e., sum of scenarios (A) and (C).

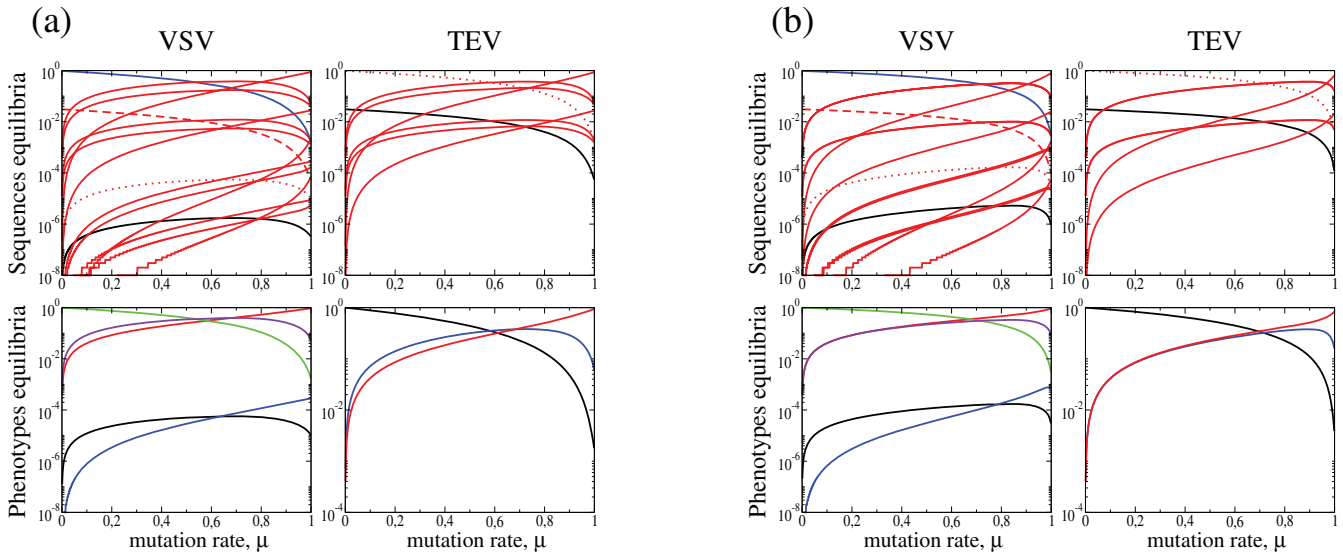


Figure 3 | Population equilibria for sequences and phenotypes at increasing mutation considering phenotypic reversion. We here consider phenotypic reversion ($\delta = 10^{-3}$) and DMFE data for VSV and TEV (in linear-log scale) with: (a) $s_B = s_D = 0.5$ and (b) $s_B = 0.25 < s_D = 0.9$. In the first row we display sequences: x_{0000} (black line), and the pool of mutants in red lines (except for strings x_{0100} (dotted red line), x_{1000} (dashed red line); and x_{1100} (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_{1000} plus x_{1100} (beneficial, *B*, in green); x_{0010} 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_{1010} plus x_{1110} (sequences with beneficial and deleterious mutations, in violet). Sequences x_{1010} and x_{1110} can be included in phenotypic classes *W*, *B*, or *D*, depending on the values of the selection coefficients [i.e., $s_D = s_B$ (*W*), $s_D < s_B$ (*B*), and $s_D > s_B$ (*D*)].

- *Scenario (C)*: N^* deleterious mutants, $\sum_{i\in D} x_i = N^*$, 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 MonteCarlo (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, $\langle p^* \rangle \cdot 100$, that push the quasispecies towards lethal vertices of the phenotypic space is extremely small (see parameter spaces (λ_L, μ) and (μ, λ_B) for scenario (A) in Figure 4). For instance, such a percentage is about 1.88% using $N^* \geq 0.9$. 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 λ_L. 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_B) that is displayed in Figure 4. Concerning to scenario (B), where we put together lethal and deleterious genomes, the percentages of parameter combinations pushing the whole quasispecies to this scenario is larger than for scenario (A). For instance, the probability of finding parameter combinations with a quasispecies composed by $N^* \geq 0.9$ lethal plus deleterious sequences is about 6.8%, which is still a low value. Here, as we identified in the previous analyses, selection coefficients do not play an important role in this scenario. Finally, the percentage of parameter combinations pushing the quasispecies to deleterious nodes in the phenotypic sequence space is also extremely low. We may notice also that having a population dominated by deleterious mutants does not imply that viral sequences will not be able to replicate, since some of the deleterious sequences will be able to replicate whenever s_D is small. For instance, a sequence with the deleterious phenotype will be able to replicate at a rate $A = 0.8$ when $s_D = 0.2$.

The exploration of the parameter space considering phenotypic reversion reveals that the percentage of parameter combinations driving to lethality decreases even more (see Supplementary Information Section S7 for further details). All together, the previous results suggest that the quasispecies is very unlikely to be driven toward full lethality by manipulating the parameters of the model, including mutation rates (see Figures 2–4, Supplementary Information Section S7, and Figures S5–S7). We performed an ANOVA test with the data displayed in Table I, taking "scenario" and "phenotypic reversion" as fixed factors, and "N^*" as a variable and covariable. Factor "scenario" had a significant effect ($P < 0.0001$) with the following rank: (B) > (A) > (C). In other words, $\langle p^* \rangle$ was larger in scenario (B) than in (C), with (A) occupying an intermediate position. "Phenotypic reversion" had a significant effect ($P = 0.0165$), with lower values of $\langle p^* \rangle$ when phenotypic reversion was allowed and higher when it was not allowed, meaning that phenotypic reversion actually makes lethality much more difficult. The covariable "N^*" also had a significant effect ($P < 0.0001$), with the largest values of $\langle p^* \rangle$ when N^* is lower, indicating that the probability of having the majority of sequences in lethal vertices was small in this case. Of all interactions between the factors and the

Table I | Statistics obtained from the MonteCarlo (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)

		No phenotypic reversion ($\delta = 0$)		Phenotypic reversion ($\delta > 0$)	
Scenario	$\geq N^*$	$\langle p^* \rangle \cdot 100$	±SD	$\langle p^* \rangle \cdot 100$	±SD
(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

covariable, the only significant one was "scenario" × "N^*" ($P < 0.0001$), meaning that differences between scenarios are bigger for small N^* than for large values of N^*.

The previous estimations of the percentages of parameter combinations pushing 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 $t = 10^2$, $t = 10^3$, and $t = 10^4$). This strategy allows us to explore the likelihood of these scenarios taking place also during transients. The results, displayed in Supplementary Information Table S1, revealed 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 $\langle p(t) \rangle \approx 0.0066$, 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 ($P < 0.0001$): time has a different behavior for each scenario ($P < 0.0001$). For scenario (A), $\langle p(t) \rangle$ 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 $\langle p(t) \rangle$ 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 $N(t)$.

Transient regimes as a function of δ and other parameters. We are particularly interested in the effect of the probability of phenotypic reversion, δ, on transient times (see Model and Methods section). Our results indicate that the smaller δ the longer is the time needed to approach the fixed point at the given distance η. This time depends on two things: the component of the initial data in the eigenvector of the matrix A in the direction of $\mathbf{x}_*$ and on the eigenvalue of $DF(\mathbf{x}_*)$ closest to zero. As discussed in Supplementary Information Section S5, the smaller is δ 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 η to $\mathbf{x}_*$ is not produced at the final integration time $t_{\max}$ is extremely small: around 0.00002 for $\delta = 10^{-3}$ and increases to 0.00022 for $\delta = 0$.

Figure S8(left) shows the results using $M^* = 10^8$ 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 τ is represented in $\log_{10}$ scale. The value of $t_{\max}$ is set equal to 10^8.

What happens if η is reduced? Of course, the transient time, until the distance to $\mathbf{x}_*$ is less than η, will increase. To illustrate this fact we reproduce in Figure S8(right) the same curves in the left plot for $\delta = 10^{-3}, 10^{-5}$ and 0 for $\eta = 10^{-3}$ (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 $t_{\max} = 10^8$ but M^* is reduced to 10^6.

Now, while for $\delta = 10^{-3}$ still only a small fraction, around 0.00005, of cases do not approach $\mathbf{x}_*$ at the distance η 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 $\mathbf{x}_*$ is of the form $\exp(-\gamma\delta^{1/2})$ where γ is the coefficient of $\delta^{1/2}$ in Eq. (33) of the Supplementary Information. On the other hand, for $\delta = 0$ the approach to $\mathbf{x}_*$ 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 τ for different values of $\varphi(\tau)$. We have selected values of $\varphi(\tau)$ of the form 0.1 k, $k = 1, \ldots, 9$. Let τ_1 be the value of τ when $\eta = \eta_1 = 10^{-3}$ and τ_2 the one for $\eta = \eta_2 = 10^{-6}$. For $\delta > 0$ the decrease of η should satisfy, approximately,

$$\eta_2 = \eta_1 \exp\Big(-\gamma\delta^{1/2}(\tau_2 - \tau_1)\Big).$$

The value of γ depends on the average value of the random parameters and also on the value of $\varphi(\tau)$.

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 $\log_{10}(\tau_2 - \tau_1)$ for different values of $\varphi(\tau)$. The differences are very close to 1, as expected. On the opposite case, $\delta = 0$, the green curve shows $\log_{10}(\tau_2/\tau_1)$. It is almost exactly equal to 3, in perfect agreement with the fact that $\log_{10}(\eta_2/\eta_1) = -3$.

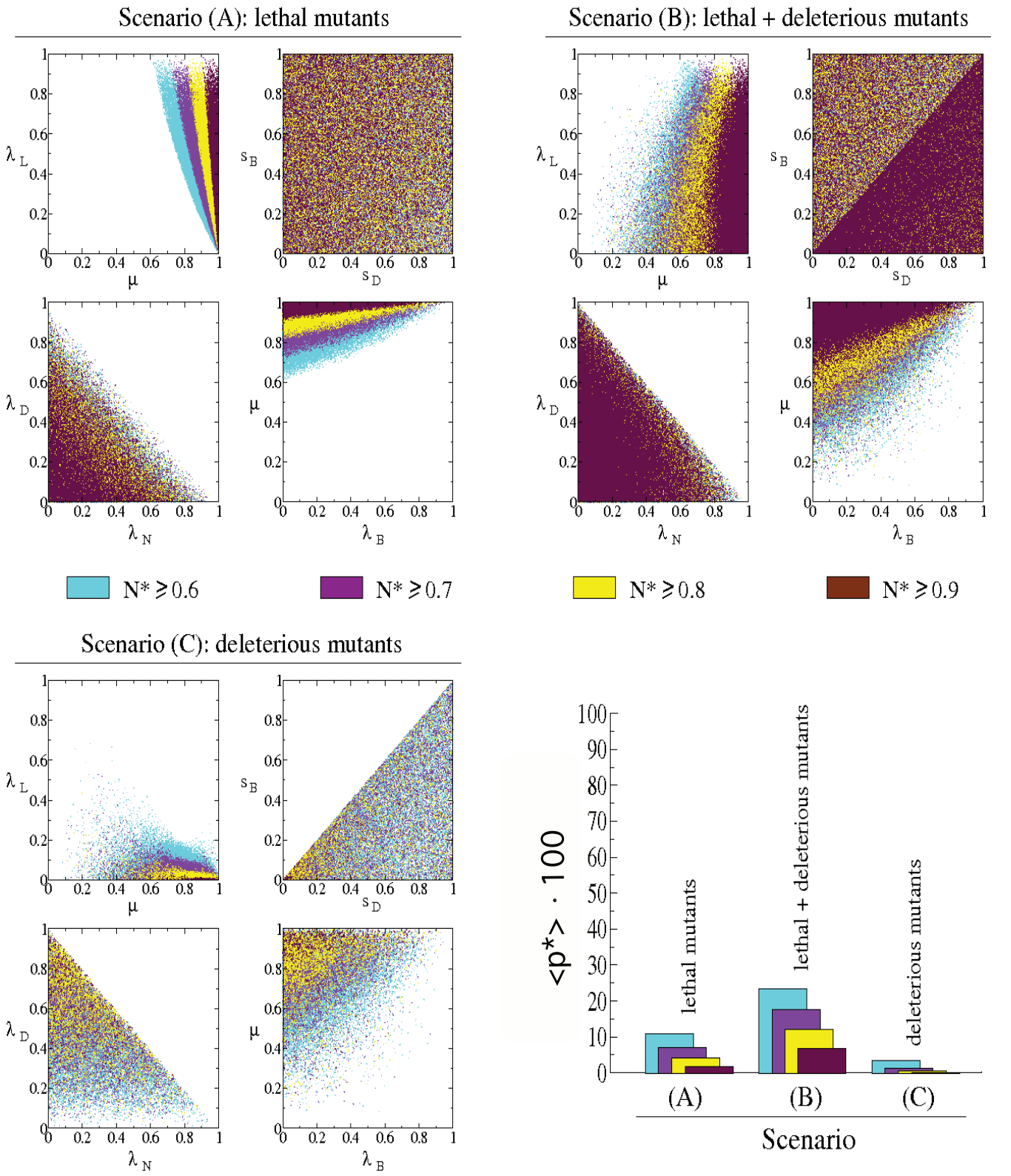


Figure 4 | Exploration of the parameter space of Eqs. (1) without phenotypic reversion (i.e., $\delta = 0$). We plot two-dimensional parameter space projections using $M = 10^6$ iterations of the MonteCarlo (MC) algorithm displaying parameter values pushing the quasispecies towards scenarios (A), (B), 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 $y^* > x^*$ (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 $\langle p^* \rangle = \langle c(N^*)/M \rangle$, 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[25,26]. Although the DMFE has been experimentally quantified for several viruses[14,15,32,33], previous theoretical models[34,35] 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[1,2]: error-prone replicators have a limited tolerance to mutations, beyond which an error catastrophes 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[39,40,42,43]. 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[44,45]. Further experimental evidence confirming the effect of mutagens in decreasing infectivity and replication was later described for HIV-1 in tissue cultures[46]. Similar results were also found for *Lymphocytic choriomeningitis virus*, hantaviruses, and *Hepatitis C virus*, among others (see review[47]). 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 (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[41] (e.g., lethal defection[48]), or bottlenecks during transmission events that turn on Muller's ratchet (shown to operate in phages $\phi 6$[49] and *MS2*[50] VSV[51], TEV[52], *Foot-and-mouth disease virus*[53], and *human immunodeficiency virus type 1*[54]). 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[8,36,41]. 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[55,56]. Genomic instability[57,58] is a hallmark of most malignant tumors, which could be ablated by increased mutagenesis[59–61]. 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[62]. The potential success of suppression of tumor growth through mutation-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[1]. The model describes the time dynamics of i classes of binary sequences with population numbers x_i able to carry different mutational types that can replicate and mutate according to the differential equations:

$$\frac{dx_i}{dt} = (1-\mu)A_i x_i + \mu \sum_{\langle j \rangle_i} \lambda_k^{[j\to i]} \beta^{[j\to i]} A_j x_j - \Phi x_i, \qquad (1)$$

with $i,j \in \mathcal{H}^v$, being $\mathcal{H}^v$ the configuration or sequence space (i.e., boolean v-dimensional hypercube), with $\mathcal{H}^v = \{1, \ldots, 2^v\}$, 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, 1^{st} bit), neutral (N, 2^{nd} bit), deleterious (D, 3^{rd} bit), and lethal (L, 4^{th} bit), as we display in Figure 1. Our model becomes a generalization of the classic Eigen's quasispecies model[2], 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 μ. The second term corresponds to the influx of mutant strings j from orthant neighbors (denoted as $<j>_i$) of $\mathcal{H}^v$ producing string i by mutation. Parameters $\lambda_k^{[j\to i]} \geq 0$, with $\sum_k \lambda_k^{[j\to i]} = 1$, 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 $\beta^{[j\to i]}$ is the probability of occurrence of transitions $0 \to 1$ and $1 \to 0$ (i.e., phenotypic reversion) from sequence j to sequence i during replication, with:

$$\beta^{[j\to i]} = \begin{cases} 1 & \text{if} \quad \text{transition is } 0\to 1, \\ 0 \leq \delta \ll 1 & \text{if} \quad \text{transition is } 1\to 0. \end{cases}$$

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). The total population is held constant (i.e., $\sum_{i=1}^{n} x_i = 1$) by a compensating dilution flux with rate coefficient $\mathbf{\Phi}$. The outflow term, $\mathbf{\Phi}$, is obtained under the constant population condition, i. e., $\sum_{i=1}^{n} x_i = 1$, giving:

$$\Phi = \sum_{i=1}^{n} \left[(1-\mu) A_i x_i + \mu \sum_{\langle j \rangle_i} \lambda_k^{[j\to i]} \beta^{[j\to i]} A_j x_j \right]. \quad (2)$$

The state space of system (1) is a high-dimensional simplex (i.e., $2^\nu - 1$ simplex) represented by the nonnegative vectors $\mathbb{R}_+^n = \left\{ (x_{0000}, \ldots, x_{1111}) \in \mathbb{R}^n | x_i \geq 0, i \in \mathcal{H}^\nu, \sum_i x_i = 1, i = 1, \ldots, n \right\}$. As initial conditions we will use (if not otherwise specified) $x_{0000}(0) = 1$, setting all other variables to zero.

All 2^4 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 $x_{0000,0100}$ (wild-type phenotype and the neutral mutants), with replicative fitness $A_{0000,0100} = 1$.
- Peak (B) populated by sequences $x_{1100,1000}$ (beneficial mutants plus the neutral mutants), with replication rate $A_{1000,1100} = 1 + s_B$.
- Peak (D) populated by sequences $x_{0010,0110}$ (deleterious mutant and the neutral mutants), with replicative fitness $A_{0010,0110} = 1 - s_D$.
- The lowest replication value of the fitness landscape is given by the lethal phenotypes, (L), given by sequences $x_{abc1} = 0$, $a, b, c \in \{0, 1\}$, 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 $x_{0010,0110}$ and with replication rates $A_{1010,1110} = 1 + s_B - s_D$) will populate fitness 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 μ, s_B, s_D, λ_B, λ_D, λ_L, λ_N, and δ. We set $0 < s_B \leq 1$, to ensure that beneficial mutations always increase replication rates. We also set $0 < s_D \lesssim 1$ (if not otherwise specified) to ensure that mutations are always deleterious, also avoiding a 100% decrease in fitness that would result in a lethal mutant (lethal mutants are explicitly introduced as variables in our dynamical system). Note that for $s_B > s_D$, mutants $x_{1010,1110}$ are fitter than the wild-type string and will be considered as beneficial, while for $s_B < s_D$ their fitnesses are lower than the wild-type fitness and thus will be considered as deleterious. For combined mutations (i.e., strings carrying two or more types of mutations) we assume additive fitness effects. Although epistatic interactions have been described for RNA viral genomes[63], as a first approach we will here use additive fitness effects. Hence, the ranges for these parameters are: $\mu \in (0, 1)$, $s_B \in (0, 1]$, $s_D \in [0, 1)$, $\lambda_B \geq 0$, $\lambda_D \geq 0$, $\lambda_L \geq 0$, $\lambda_N \geq 0$, $\lambda_B + \lambda_D + \lambda_L + \lambda_N = 1$, $\delta \geq 0$.

It is important to notice that Eigen's model describes the dynamics of explicit genome sequences[2,3,6]. We here simplify the system to sequences where mutations are not nucleotide substitutions but changes involving possible phenotypic transitions. Hence, our sequences are not viral genomes but an abstract representation of the genomes in terms of phenotypic traits (see Ref. 64 for other phenotypic approaches to viral quasispecies). In this sense, the relationship between our modeling approach and true viral genomes becomes obvious from the identification of groups of viral genomes that belong to the same fitness phenotypic classes (i.e, all have the same fitness regardless the specific mutations they carry in their genomes). These phenotypic classes are the entities represented in our model. Furthermore, keeping in mind our goal of exploring the role of the DMFE in quasispecies dynamics and transitions, we necessarily made several simplifying assumptions that allow for an analytical treatment. These assumptions and their limitations are discussed in Supplementary Information Section S1.

Experimental data. In this article we use experimental DMFE data for VSV and TEV as a case study for viral quasispecies. Sanjuán *et al.*[14] generated a collection of single-nucleotide substitution mutants of VSV and evaluated their fitnesses *in vitro*. They found that the frequencies of the four different fitness classes were (using our notation) $\lambda_B^{[j\to i]} = 0.042$; $\lambda_N^{[j\to i]} = 0.2707$; $\lambda_D^{[j\to i]} = 0.2917$; and $\lambda_L^{[j\to i]} = 0.3956$. Using a similar experimental approach but in this case measuring fitness effects *in vivo*, Carrasco *et al.*[15] found that such frequencies for TEV were $\lambda_B^{[j\to i]} = 0$; $\lambda_N^{[j\to i]} = 0.227$; $\lambda_D^{[j\to i]} = 0.364$; and $\lambda_L^{[j\to i]} = 0.409$. To simplify notation, in our model we will use λ_k instead of $\lambda_k^{[j\to i]}$. We refer the reader to references 14, 15 for further details.

Numerical algorithms. We are especially interested in performing an extensive exploration of the model parameters to address several issues. Our interests here are mainly two:

a) Given arbitrary parameters of the model, we desire to compute the population values of beneficial, deleterious, and lethal sequences at the equilibrium. Furthermore, as typically the parameters are not well known or, simply, non available, we want to establish the probability that the fraction of sequences of each one of the previous types exceeds some fraction N^* (being N^* a given fixed population equilibrium value). For this later case, we are mainly interested in quantifying the likelihood of finding parametric scenarios impairing viral sequences success, i.e., parameter combinations generating deleterious and/or lethal phenotypes. To do so, we will use the population scenarios *(A)*, *(B)*, and *(C)* (defined in the Results Section: Exploration of the parameter space). As mentioned, in all these scenarios we want to have a measure of the fraction of parameters for which these scenarios reach values of population size exceeding a given threshold, N^*. The domain $\mathcal{P}$ where we locate the parameters is a cube $[0, 1]^4$ corresponding to the values of μ, s_B, s_D, δ times a three-dimensional simplex Σ_3 in the parameters λ_B, λ_D, λ_L, λ_N with the constrain $\lambda_B + \lambda_D + \lambda_L + \lambda_N = 1$. Different assumptions can be made on the probability density in $\mathcal{P} = [0,1]^4 \times \Sigma_3$.

b) How fast go the initial conditions (typically the vector $x(0) = (1, 0, \ldots, 0)^T$) to the attractor to which they tend? In other words, what can be said about transients? This is irrelevant if we assume that the evolution time is as large as desired. However, for finite time evolution, T, the distance from $x(T)$ to the attractor can be large.

Let us describe the algorithms to carry out the computations in both cases.

In principle, if we consider a uniform probability density in $\mathcal{P}$ to answer question a) amounts to compute the volume of a semi algebraic set in the parameter space. Indeed, let us denote a value of the parameters simply as $\Pi \in \mathcal{P}$. The amount of viral sequences of a given class corresponding to the equilibrium associated with $\mathbf{\Pi}$ can be denoted as $g(\mathbf{\Pi})$. It is clear, due to the dependence of the matrix A (see Supplementary Inforamtion Section S4) and, hence, of the characteristic polynomial and the eigenvectors, that g is an algebraic function. Therefore the volume of the region of interest in Σ is bounded by the hypersurface $g(\mathbf{\Pi}) = N^*$ and boundaries of Σ.

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 Σ_3 using a standard method.
- When all the parameters are available we look for the dominant zero of the polynomial p_B in Eq. (24) of the Supplementary Information. Let us denote it as σ_3 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 $c(N^*)$ in which $g(\mathbf{\Pi}) \geq N^*$. The desired probability is $p^* = c(N^*)/M$. 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 (A), (B), and (C) at a given time, $N(t)$. Then, the previous probability will be represented as $p(t) = c(N(t))/M$.

In a large part of the computations the parameter δ is not taken with uniform probability but uniform in log scale, say in the range $[\delta_1, \delta_2]$. Then the value of $\log(\delta)$ is generated with a uniform distribution in $[\log(\delta_1), \log(\delta_2)]$ and proceed as before.

It can be also interesting to fix δ and proceed in a similar way for a set of values of δ. 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

$$\left(\int_{\delta_1}^{\delta_2} f(\delta) d\delta \right)^{-1} \int_{\delta_1}^{\delta_2} p(\delta) f(\delta) d\delta. \quad (3)$$

We pass now to the point b). Given parameters $\mathbf{\Pi}$ we can compute the equilibrium x_*. On the other hand, we can start the integration of system (3) of the SI from the initial point $x(0) = (1, 0, \ldots, 0)^T$ until the current point $x(t)$ reaches a given distance η from x_*. Typically we have used the value $\eta = 10^{-3}$ and the distance has been computed using the $|\,|_\infty$ norm. The integrations are carried out using the Runge-Kutta-Fehlberg RKF78 method, with automatic step size control and local relative tolerance 10^{-15}. After every step the distance to x_* 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 δ as is illustrated below and in the Results section.

Computation of transient times. In order to see how parameter δ, i.e., phenotypic reversion, affects the transients, we carry out the following experiment. For a fixed value of δ 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 $x(0) = (1, 0, \ldots, 0)^T$. The integration is stopped when the distance to the fixed point is a small fixed value η. The distance is measured as $\eta = \max_{i=1,\ldots,16} |x_i(t) - (x_*)_i|$. For a given value τ of the time we record the fraction $\varphi(\tau)$ of M^* such that the distance η is reached for a value of $t < \tau$. Then we display $\varphi(\tau)$ as a function of τ (see Figure S8). Of course, for practical reasons, the integration time is limited to a fixed large $t_{\max}$.

1. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwiss.* **58**, 465–523 (1971).

2. Eigen, M. & Schuster, P. *The hypercycle. A principle of natural self-organization* [Eigen, M. & Schuster, P. (eds.)] (Springer-Verlag, Berlin, 1979).
3. Schuster, P. Evolution on "*realistic*" fitness landscapes. Phase transitions, strong quasispecies, and neutrality. Santa Fe Institute Working Paper #12-06-006, 1–94 (2012).
4. Domingo, E., Webster, E. & Holland, J. J. *Origin and evolution of viruses* [Domingo, E., Parrish, C. R. & Holland, J. J. (eds.)] (Academic, San Diego, USA, 1999).
5. Wilke, C. O. Quasispecies theory in the context of population genetics. *BMC Evol. Biol.* **5**, 44–51 (2005).
6. Eigen, M., McCaskill, J. & Schuster, P. The Molecular quasi-species. *Adv. Chem. Phys.* **75**, 149–263 (1989).
7. Domingo, E., Biebricher, C., Eigen, M. & Holland, J. J. *Quasispecies and RNA virus evolution: Principles and consequences* (Landes Bioscience, Austin, TX, 2001).
8. Manrubia, S. C., Domingo, E. & Lázaro, E. Pathways to extinction: beyond the error threshold. *Phil. Trans. R. Soc. B* **365**, 1943–1952 (2010).
9. Franz, S. & Peliti, L. Error threshold in simple landscapes. *J. Phys. A: Math. Gen.* **30**, 4481–4487 (1997).
10. Satorras, R. P. & Solé, R. V. Field theory for a reaction-diffusion model of quasispecies dynamics. *Phys. Rev. E* **64**, 051909278 (2001).
11. Saakian, D. B. & Chin-Kun, H. Exact solution of the Eigen model with general fitness functions and degradation rates. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4935–4939 (2005).
12. Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
13. Kauffman, S. & Levin, S. Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).
14. Sanjuán, 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.* **101**, 8396–8401 (2004).
15. 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.* **81**, 12979–12984 (2007).
16. Saakian, D. B., Muñoz, E., Chin-Kun, H. & Deem, M. W. Quasispecies theory for multiple-peak fitness landscapes. *Phys. Rev. E* **73**, 041913 (2006).
17. Sardanyés, J., Solé, R. V. & Elena, S. F. Simple quasispecies models for the survival-of-the-flattest effect: the role of space. *J. theor. Biol.* **250**, 560–568 (2006).
18. Wilke, C. O. Selection for fitness versus selection for robustness in RNA secondary structure folding. *Evolution* **55**, 2412–2420 (2001).
19. McCaskill, J. S. & Altemeyer, S. Error threshold for spatially resolved evolution in the quasispecies model. *Phys. Rev. Lett.* **86**, 5819–5822 (2001).
20. Toyabe, S. & Sano, M. Spatial suppression of error catastrophe in a growing pattern. *Physica D* **230**, 1–8 (2005).
21. Sardanyés, J. & Elena, S. F. Quasispecies spatial models for RNA viruses with different replication modes and infection strategies. *PLOS ONE* **6**, e24884 (2011).
22. Sardanyés, J. & Elena, S. F. Error threshold in RNA quasispecies models with complementation. *J. Theor. Biol.* **265**, 278–286 (2010).
23. Iranzo, J. & Manrubia, S. C. Evolutionary dynamics of genome segmentation in multipartite viruses. *Proc. R. Soc. B* **279**, 3812–3819 (2012).
24. Sardanyés, J., Solé, R. V. & Elena, S. F. Replication mode and landscape topology differentially affect RNA virus mutational load and robustness. *J. Virol.* **83**, 12579–12589 (2009).
25. Sanjuán, R. Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies. *Phil. Trans. R. Soc. B* **365**, 1975 (2010).
26. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
27. Manrubia, S. C. Modelling viral evolution and adaptation: challenges and rewards. *Curr. Op. Virol.* **2**, 531–537 (2012).
28. Keightley, P. D. & Lynch, M. Toward a realistic model of mutations affecting fitness. *Evolution* **57**, 683–685 (2003).
29. Keightley, P. D. The distribution of mutation effects on viability in *Drosophila melanogaster*. *Genetics* **138**, 1315–1322 (1994).
30. Loewe, L. & Charlesworth, B. Inferring the distribution of mutational effects on fitness in *Drosophila*. *Biol. Lett.* **2**, 426–430 (2006).
31. 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).
32. Elena, S. F. & Moya, A. Rate of deleterious mutation and the distribution of its effects on fitness in *Vesicular stomatitis virus*. *J. Evol. Biol.* **12**, 1078–1088 (1999).
33. van Opiijnen, 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).
34. Orr, H. A. The distribution of fitness effects among beneficial mutations. *Genetics* **163**, 1519–1526 (2003).
35. 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).
36. 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).
37. Schuster, P. Mathematical modeling of evolution. Solved and open problems. *Theory Biosci.* **130**, 71–89 (2011).
38. Lorenz, D. M., Park, J.-M. & Deem, M. W. Evolutionary processes in finite populations. *Phys. Rev. E* **87**, 022704 (2013).
39. Bull, J. J., Meyers, L. A. & Lachmann, M. Quasispecies made simple. *PLOS Comput. Biol.* **1**, e61 (2005).
40. Bull, J. J., Sanjuán, R. & Wilke, C. O. Theory of lethal mutagenesis for viruses. *J. Virol.* **81**, 2930–2939 (2007).
41. Wylie, C. S. & Shakhnovic, E. L. Mutation induced extinction in finite populations: lethal mutagenesis and lethal isolation. *PLOS Comput. Biol.* **8**, e1002609 (2012).
42. Aguirre, J., Buldú, J. M. & Manrubia, S. C. A trade-off between neutrality and adaptability limits the optimization of viral quasispecies. *J. Theor. Biol.* **261**, 148–155 (2001).
43. 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).
44. Holland, J. *et al*. Rapid evolution of RNA genomes. *Science* **215**, 1577–1585 (1982).
45. Holland, J. J., Domingo, E., de la Torre, J. C. & Steinhauer, D. A. Mutation frequencies at defined single codon sites in vesicular stomatitis virus and poliovirus can be increased only slightly by chemical mutagenesis. *J. Virol.* **64**, 3960–3962 (1990).
46. Loeb, L. A. *et al*. Lethal mutagenesis of HIV with mutagenic nucleoside analogs. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1492–1497 (1999).
47. Anderson, J. P., Daifuku, R. & Loeb, L. A. Viral error catastrophe by mutagenic nucleosides. *Annu. Rev. Microbiol.* **58**, 183–205 (2004).
48. Grande-Pérez, A. *et al*. Suppression of viral infectivity through lethal defection. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4448–4452 (2005).
49. Chao, L. Fitness of RNA virus decreased by Muller's ratchet. *Nature* **348**, 454–455 (1990).
50. de la Peña, M., Elena, S. F. & Moya, A. Effect of deleterious mutation-accumulation on the fitness of RNA bacteriophage MS2. *Evolution* **54**, 686–691 (2000).
51. Duarte, E. *et al*. Rapid fitness losses in mammalian RNA virus clones due to Muller's ratchet. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6015–6019 (1992).
52. de la Iglesia, F. & Elena, S. F. Fitness declines in *Tobacco etch virus* upon serial bottleneck transfers. *J. Virol.* **81**, 4941–4947 (2007).
53. Escarmís, C., Lázaro, E. & Manrubia, S. C. Population bottlenecks in quasispecies dynamics. *Curr. Top. Microbiol. Immunol.* **299**, 141–170 (2006).
54. Yuste, E. *et al*. (1999) Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. *J. Virol.* **73**, 2745–2751 (1999).
55. Solé, R. V. & Deisboeck, T. S. An error catastrophe in cancer? *J. Theor. Biol.* **228**, 47–54 (2004).
56. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
57. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
58. Cahill, D. P., Kinzler, K. W., Vogelstein, B. & Lengauer, C. Genetic instability and darwinian selection in tumours. *Trends Cell. Biol.* **9**, M57–60 (1999).
59. Fox, M. In vitro mutagenesis by anti-cancer drugs. *Chemotherapy* 315–322 (1976).
60. Fox, E. J. & Loeb, L. A. Lethal mutagenesis: Targeting the mutator phenotype in cancer. *Seminars in Cancer Biology* **20**, 353–359 (2010).
61. Loeb, L. A. Human cancers express the mutator phenotypes: origin, consequences and targeting. *Nature* **11**, 450–457 (2011).
62. Vogelstein, B. & Kinzler, K. W. Cancer genes and the pathways they control. *Nature Medicine* **10**, 789–99 (2004).
63. Elena, S. F., Solé, R. V. & Sardanyés, J. Simple genomes, complex interactions: Epistasis in RNA virus. *Chaos* **20**, 026106 (2010).
64. Reidys, C., Forst, C. V. & Schuster, P. Replication and mutation on neutral networks. *Bull. Math. Biol.* **63**, 57–94 (2001).

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, 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).

Author contributions

J.S. and S.F.E. built the mathematical model. C.S. and R.M. carried out the analysis of the mathematical model and the related theoretical calculations. J.S. and C.S. produced the numerical data. J.S., C.S., R.M. and S.F.E. analyzed the numerical data. J.S., C.S., R.M., R.V.S. and S.F.E. wrote the manuscript. All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Sardanyés, J., Simó, C., Martínez, R., Solé, R.V. & Elena, S.F. Variability in mutational fitness effects prevents full lethal transitions in large quasispecies populations. *Sci. Rep.* **4**, 4625; DOI:10.1038/srep04625 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. The images in this article are included in the article's Creative Commons license, unless indicated otherwise in the image credit; if the image is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the image. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/

SCIENTIFIC REPORTS

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)."