Replication Mode and Landscape Topology Affect Differentially RNA Virus Mutational Load and Robustness

Josep Sardanyés[1*] Ricard V. Solé[1,3] Santiago F. Elena[2,3]

Complex Systems Lab (ICREA-UPF), Barcelona Biomedical Research Park (PRBB-GRIB), Dr. Aiguader 88, 08003 Barcelona, Spain[1], Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, Ingeniero Fausto Elio s/n, 46022 València, Spain[2], and The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA[3]

[*]Corresponding. Mailing address: Complex Systems Lab (ICREA-UPF), Barcelona Biomedical Research Park, Dr. Aiguader 88, 08003 Barcelona, Spain. E-mail: josep.sardanes@upf.edu.

Running title: replication mode and mutational load

Word counts: 244 words in the abstract, 7403 words in the main text, excluding references, table and figure legends.

Regardless of genome polarity, intermediaries of complementary sense must be synthesized and used as templates for the production of new genomic strands. Depending on whether these new genomic strands become themselves templates for producing extra antigenomic ones, thus giving rise to a geometric growth, or only the firstly synthetized antigenomic strands can be used to this end, thus following Luria's stamping machine model, the abundance and distribution of mutant genomes will be different. Here we propose mathematical and bit string models that allow distinguishing between stamping machine and geometric replication. We have observed that, regardless the topology of the fitness landscape, the critical mutation rate at which the master sequence disappears increases as the mechanism of replication switches from purely geometric to stamping machine. We also found that for a wide range of mutation rates, large effect mutations do not accumulate regardless the scheme of replication. However, mild mutations accumulate more in the geometric model. Furthermore, at high mutation rates, geometric growth leads to a population collapse for intermediate values of mutational effects at which the stamping machine still produces master genomes. We observed that the critical mutation rate was weakly dependent on the strength of antagonistic epistasis but strongly dependent on the synergistic epistasis. In conclusion, we have shown that RNA viruses may increase their robustness against the accumulation of deleterious mutations by replicating as stamping machines, and that the magnitude of this benefit depends on the topology of the fitness landscape assumed.

The mode of RNA virus replication has important consequences for understanding the rates at which deleterious mutations accumulate and the statistical properties of the cloud of mutants around the master sequence (4, 17). For the sake of illustration, let's assume that the infecting virus has an mRNA sense genome (positive strand), such as for example picorna-like viruses. The different steps of the infectious cycle are schematically illustrated in Figure 1a. The first step of infection is the uncoating of the RNA molecule, followed by its translation into structural and nonstructural proteins, the latter including the replicase. The replicase then copies the genomic strand to make antigenomic (negative sense) strands. These are used as templates to produce the genomic progeny that will accumulate in the cell, serve as templates for translation and, following encapsidation by coat proteins, form new virions. If the antigenomic strands produced at the first round are the only templates for producing the entire progeny of genomic strands, the distribution of mutations per genome within an infected cell is expected to be Poisson because mutants do not replicate (25). Consequently, the fraction of mutation-free genomes produced is given by the Poisson null class $e^{-\mu L}$, where μ is the per site mutation rate and L the genome length. This scheme of replication corresponds to the linear stamping machine replication model first proposed by Luria (24); hereafter SMR. However, if all genomic strand progeny can also immediately serve as templates for additional rounds of antigenomic strands synthesis, the replication model is effectively geometric (GR) and the distribution of mutant genomes per cell increases in variance because mutant

progeny is producing itself more mutant viruses. In this case, the distribution of mutant genomes conforms to the Luria-Delbrück distribution (9). The fraction of mutation-free genomes produced would depend on the number of replication rounds experienced, τ, according to $e^{-\mu L \tau}$. Therefore, it is straightforward to see that GR will produce $e^{-\mu L}/e^{-\mu L \tau}$ more mutant genomes than SMR. If only a fraction of the genomic strand progeny replicates, then the replication model will be a mixture of GR and SMR that deviates from the Poisson expectation as much as the GR contribution. The effect of replication mode in virus mutational load can be better understood with the following example. The genomic mutation rate of *Tobacco mosaic virus* was estimated in the range $0.043 \leq \mu L \leq 0.063$ per replication round and about 40 viral particles produced per infected tobacco cell, which is equivalent to $\tau = 5.322$ generations of GR (26). Therefore, the excess production of mutant genomes for GR over SMR will lie in the range $16.38 \leq f \leq 23.76$.

Experimental data support different models of replication for different viruses. For example, bacteriophage *T*2 is thought to replicate mostly following GR because the number of mutants per infected cell fails to fit a Poisson distribution (25). However, phage $\phi X174$ data fit well the Poisson distribution and, hence, is thought to replicate according to a SMR model (7). Lying within these two extremes, phage $\phi 6$ slightly deviated from the Poisson expectation, an observation interpreted as a result of a mixed model in which some progeny was also able of replicating (4). Plant positive strand RNA viruses are also thought to reproduce mostly according to a SMR model (15, 18). Despite the apparent importance of the

model of RNA virus replication on the accumulation of mutant genomes, most mathematical models proposed to study the dynamics of RNA virus populations rely on the assumption of GR. For example, the most commonly used theoretical paradigm for the study of virus evolution, Eigen's quasispecies picture (10, 13), assumes geometric growth and the Swetina-Schuster fitness landscape (41). This approach assumes that all mutations have the same deleterious fitness effect and that fitness does not depend on the number of mutations carried by a genome (i.e., the fitness of genomes having 1, 10 or 100 mutations would be identical) (Figure 1d).

How to model viral replication is a timely research topic in virology. The growth of a virus in its host cell is a complex process. In seeking to understand this process and the effect of the interactions between the macromolecules involved in viral growth, the crossing of disciplines as biochemistry, molecular biology, population genetics, and nonlinear dynamical systems might provide a powerful way to study the overall behavior of virus dynamics. In this sense, models play a crucial role for a qualitative and a quantitative study of virus replication, being also useful to predict the system's behavior time evolution as well as to analyze its sensitivity with respect to parameter changes. Furthermore, insights into interactions of viruses with host cells might help us to improve our understanding of virus-mediated diseases and to develop antiviral strategies (37). Several models of intracellular viral growth kinetics can be found in the literature, ranging from simple (unstructured) models capturing the basic replication processes (12, 19, 39, 46) to the so-called structured models that consider replication in different cellular

compartments such as membranes, endosomes, cytoplasm or nucleus. Some examples of structured models have been developed for bacteriophage *T*7 (17, 45), *Human immunodeficiency virus* type 1 (29), subgenomic *Hepatitis C virus* (6), *Influenza A virus* (37), or *Vesicular stomatitis virus* (VSV) (24). However, to the extent of our knowledge and despite its relevance, none has considered the effect that non-geometric modes of replication in combination with different fitness landscapes may have on the accumulation of mutant genomes.

In the first part of this work we analyze a mathematical model describing the single-cell reproductive cycle of positive-sense RNA viruses that make no subgenomic mRNAs and encode a single polyprotein that is self-processed into structural and nonstructural proteins, *sensu* picorna-like viruses (see Figure 1a). This model describes the main features of cytoplasmatic intracellular amplification of a viral genome (see (1) and references therein) without getting deep into the peculiarities of any particular system. By doing so, the model remain as general as possible and would be applicable to a wide variety of RNA viruses. As far as focusing into intracellular replicative dynamics, we are not explicitly considering cell lyses and among-cells transmission. The main goal of this first part is to analyze the effect of SMR and GR under the simplistic but mathematically convenient Swetina-Schuster fitness landscape (41). For each replication scheme, we will compare the dynamics of non-mutated and mutated strains of both polarities, the value of the critical mutation rate for which the non-mutated strain disappears from the population and the sensitivity to mutations. The model considers explicitly both genomic and antigenomic master and mutant strands, the

viral polyprotein, the translational complexes and the mature virions. Whenever possible, model parameter values were chosen to be in the same range than experimental determinations taken from the literature (Table 1). For those parameters for which no experimental value was found, we ran a sensitivity analysis with over a million parameter combinations and selected values providing the most consistent outcome. In the second part of our study, we confirm the validity of the results by using a stochastic model involving digital genomes and incorporating more complex and realistic fitness landscapes to be cast representative for RNA viruses.

MATERIALS AND METHODS

Mathematical model. Our quasispecies mathematical model of intracellular viral replication is based on the scheme shown in Figure 1a (see Table 1 for details on the notation used hereafter). The model is used to analyze the dynamics of replication of positive-sense RNA viruses that make no subgenomic mRNAs. We explicitly define the polarity of the strands constituting the quasispecies, studying its dynamics using the Swetina-Schuster fitness landscape (41) (Figure 1d). In this landscape, the wildtype sequence has fitness 1, while all mutant genotypes (regardless the number and nature of mutations) have equal fitness < 1 (41) and, thus, represents a single sharp peak emerging from a flat surface. In such scenario we may divide the population in either master or mutant genomic and antigenomic viral strands. The pool of mutant strands of each polarity is thus grouped in an average sequence different from the master one. Therefore, the variables of this

dynamical system are given by the genomic (+) and antigenomic (–) viral strands, $\psi_{0,1\pm}$, being the master ones indicated with subindex 0 and the mutants with subindex 1. Moreover we also consider as variables the viral polyprotein, p, the translation complex formed by genomic RNAs and ribosomes, T_c, and the mature viral particles, V.

Our model assumes that all the interacting macromolecules are homogeneously mixed and that mutant genomic RNAs (ψ_{1+}) fail to be translated into the polyprotein and thus they do not compete for the available ribosomes. This assumption is reasonable because, as confirmed below, for moderately large mutational effects ($\Lambda^+ << \Gamma^+$), the concentration of mutant genomic RNAs should be low compared to the wildtype RNAs, thus making the amount of ribosome-RNA complexes for mutant strains quickly drop to zero. For instance, one can envision these mutations as introducing stop codons or inducing conformational change in the ribosomal entry sites hindering its interaction with the ribosomes. More importantly, with this assumption, we eliminate the mathematical complications of complementation, keeping the model more simple and focused. Next we proceed to give a detailed explanation of the processes described by the model. Four steps are considered here, (a) to (d) below, which correspond to the main phases of viral replication inside the host cell (Figure 1a).

(a) Translation complex kinetics. Upon entry and uncoating, the genomic strand binds with the cellular ribosomes forming the translational complexes. Following Dahari *et al*. (6), the amount of free available ribosomes, R^{av}, is used as an upper bound to the formation of the translation complexes, T_c, and is given by $R^{av} = R^{tot} –$

T_c. Note that here we assume that the total number of ribosomes, R^{tot}, is constant and the number of available ribosomes decreases due to the formation of the translational complexes. The dynamics of the translational complex is then defined by

$$\frac{dC}{dt} = k_R R_0 + \varepsilon_T C - k_1 R C - k_1 C, \quad (1)$$

where the parameter k_R is the effective interaction rate between the genomic RNA and the available ribosomes. The second term denotes the dissociation of the translation complex after which the ribosomes and the genomic RNA become again available. We consider that the translation complex is degraded at a rate $\varepsilon_T << k_1$.

(b) Viral polyprotein dynamics. The dynamics of the viral polyprotein, p, depends on the presence of the translation complexes, the formation of mature virions, its intrinsic degradation and on the replication of viral strands. An appropriate description reads

$$\frac{dp}{dt} = k_1 C - \gamma p = k_1 C + \varepsilon_T C - \gamma p - \mathcal{R} p. \quad (2)$$

Here the function $\mathcal{R}p = \frac{dp}{dt} - \gamma p + p_0 + \varepsilon_T C - \gamma p - p_0 - \varepsilon_T C + k_1 C - \gamma p + p_1 + \varepsilon_T C - \gamma p - p_1 -$ determines the amount of nonstructural protein involved in replication (see below for the definition of the terms involved in function $\mathcal{R}p$). The function $\varphi(pp+)$, which corresponds to the formation of virions by encapsidation of the genomic strands, is defined as $\frac{dV}{dt} = k_2 C - \gamma V$. Here k_2 is the encapsidation rate, and m the number of monomers of structural protein required for building up a mature virion. This equation assumes that virion assembly follows an m-order mass action kinetics. The polyprotein is proteolitically self-processed giving place to the formation of

both structural (capsid) and nonstructural (replicase) proteins, which represent, respectively, $1 - \beta$ and β of the entire polyprotein (19). The polyprotein degrades at rate ε_p.

(c) RNA synthesis and degradation. Four different classes of RNA sequences, $x_j \in \{\square 0+, \square 0-, \square 1+, \square 1-\} = \mathbf{x}$, are considered and their growth is limited by a logistic-like function that imposes a limit due to finite cellular resources (i.e., carrying capacity $\square = 1$)

$\mathcal{L}\square = 1 - 1\square\square = 01\square\square + + \square\square -$.

The replicase uses a given strand as template to synthesize its perfectly complementary sequence at a rate $\Gamma(1 - \mu)$, being μ the average mutation rate and Γ the replication rate of the master strands. Hence, master sequences will generate mutant antigenomic sequences at a rate $\Gamma\mu$. Mutant strands replicate at rates $\Lambda < < \Gamma$ because we have assumed that mutations are deleterious. Note that backward mutations are assumed to be so infrequent that their effect may be neglected. The concentration of free strands will grow following

$$\square\square 0+\square\square = \square + \mathcal{L}\square - \square\square 0+ + \square 1\square\square - \square\square 0+ - \square\square\square\square\square\square\square 0+, \qquad (3)$$

where $\square\square 0+$, represents the decay of RNA molecules as a consequence of inherent degradation or the action of cellular RNases which is assumed to be the same for all strands. The three last terms in the right-hand side correspond to the dissociation from the translation machinery, the sequestration rate due to the formation of new viral particles, and the capture by free ribosomes to produce new translational complexes, respectively. The growth rate r_+ incorporates the presence of $\square 0-$

templates, the fraction of the polyprotein used as replicase (βp), and the fidelity of replication, $1 - \mu$, as $\Phi+=\Phi-1-\Phi\Phi0-\Phi\Phi$. For the $\Phi0-$ strands, the dynamics is now given by

$$\Phi\Phi0-\Phi\Phi=\Phi-\mathcal{L}\Phi-\Phi\Phi0-, \quad (4)$$

where now $\Phi-=\Phi\Phi-\Phi+1-\Phi\Phi0+\Phi\Phi$, (see below for the definition of the function $\Phi\Phi-$). Similarly, we can build the equations for the mutant populations as follows

$$\Phi\Phi1+\Phi\Phi=\Phi+'\mathcal{L}\Phi-\Phi\Phi1+-\Phi\Phi1+ \quad (5)$$

and

$$\Phi\Phi1-\Phi\Phi=\Phi-'\mathcal{L}\Phi-\Phi\Phi1- \quad (6)$$

with their replication rates now given by $\Phi+'=\Phi\Phi\Phi-\Phi\Phi0-+\Phi-\Phi1-$ and $\Phi-'=\Phi\Phi-\Phi\Phi\Phi+\Phi\Phi0++\Phi+\Phi1+$, consistently with the reactions outlined in Figure 1.

Note that the differences in the replication rates of both genomic and antigenomic strands allow us to analyze both SMR and GR kinetics. To model GR we set $\Gamma^+ = \Gamma^-$, i.e., all the synthesized strands are allowed to replicate. For SMR, however, we use $\Gamma^+ << \Gamma^-$, that is, the infectious genomic RNA entering into the host cell synthesizes one or very few negative copies which are then used as the only templates for the synthesis of new genomic strands in a Luria's stamping machine strategy. Indeed, to further stress the assumption that at the beginning of the infectious cycle only the antigenomic strands need to be produced but, as infectious progresses, this production has to be shut off to favor production of genomic strands, we assume a negative feedback of antigenomic strands concentration on its own rate of production. In mathematical terms, this constraint

can be incorporated by setting $\gamma_{-}=1+\gamma=01\gamma_{-}$ in the production of antigenomic strands from genomic ones. For GR, $\gamma_{-}=1$.

(d) Formation of viral particles. The new virions are produced both from master and mutant genomic strands combined with the structural proteins. From the previous steps, we can see that the formation of mature virions, V, will follow

$$\gamma\gamma\gamma\gamma=\gamma=01\gamma\gamma\gamma+-\gamma\gamma. \quad (7)$$

Where the first left-hand side term represents the encapsidation of positive strands and the second is introduced to control the amount of virions (e.g., degradation of viral particles and elimination of mature particles that may leak out the cell).

For the sake of simplicity we hereafter will use (by default): $\Gamma_{-} = 1$, $\Lambda_{\pm} = \Gamma_{-}/10$ (i.e., assuming that mutants are largely deleterious and have reduced replication rate in a factor of 1/10). The 10% average fitness effect was chosen to be in the same order of magnitude that experimental data for the effect of single point mutations gathered for VSV (33) and *Tobacco etch virus* (3). The other parameters will be either explored in this work or remain fixed at the biologically meaningful values provided in Table 1.

Stochastic dynamics of digital genomes. During the first stages of the infection and due to the stochastic nature of transmission events, cells are usually invaded by one or few viral particles. Therefore, a stochastic description of the replication process would better capture the fluctuations due to small population sizes (see (39) and references therein). We use an unstructured discrete model of *in silico* genome evolution considering a bit string description of the population structure (35, 38) which allows us to explicitly simulate the complex and

heterogeneous structure of populations of replicators. Although a real RNA is composed by a four-letter alphabet, we use Leuthäusser's approach by considering that each bit would represents either purines or pyrimidines (22, 23). Digital strands will thus be represented as chains of bits. Each chain will have L = 32 bits and a maximum population size of N = 1000 chains will be allowed.

We define a population of digital genomes. We indicate as S_i^+ and S_i^- positive and negative strands, respectively. A given string will be defined as $S_i^\pm$ = ($S_{i1}^\pm$, $S_{i2}^\pm$, ..., $S_{iL}^\pm$), with $S_{ij}^\pm \in \{0, 1\}$. The genomic and antigenomic master sequences in our model (indicated as S_m) are chosen to be S_m^+ = (11...1) and S_m^- = (00...0), respectively. We initially "inoculate" our system with $N(0)$ replicating strings. These strings can now replicate (generating complementary strands) and mutate. For instance, each bit in S_i^+ can mutate, i.e., $S_{ij}^+ S_{ij1}^- S_{ij}^+ = S_{ij}^-$, with a given mutation probability μ_b per bit and replication cycle. They also degrade with probability ε (Figure 1c). We first investigate the Swetina-Schuster fitness landscape (41) used in the previous mathematical model (Figure 1d). In this case, the master sequences have the highest fitness: their replication probabilities are set to Γ_m = 1 whereas all other strings replicate with an arbitrary probability Λ_m = 0.1. Then, we move to more complex and realistic fitness landscapes that are described by the following general form

$$\Gamma_i^\pm = 1 - \xi d_{im}$$

where d_{im} is the Hamming distance between string $S_i^\pm$ and the master sequence $S_m^\pm$ and ξ measures the sign and strength of epistasis. If ξ = 1, then we have an

additive fitness landscape in which all mutations have the same effect ($1/L$) and thus fitness declines linearly with the number of mutations (Figure 1d). Values of $\xi > 1$ or $0 < \xi < 1$ correspond to synergistic and antagonistic epistasis, respectively (Figure 1d). This latter case is in good agreement with the results of recent experiments showing that, on average, antagonistic epistasis predominate in RNA viruses (2, 30, 34, 42).

The simulation algorithm repeats, at every generation τ, 1000 times the replication and degradation rules. This is a standard Metropolis Importance Sampling Monte Carlo updating scheme which ensures that, on average, the rules are applied to all the population of strings and defines the time scale (21). To differentiate between both types of replication we implemented the following strategy: when a genomic strand replicates producing an antigenomic one, the later will always keep replicating (unless degraded). On the contrary, when an antigenomic strand replicates, the synthesized genomic strand will become a replicator with probability ρ. Note that whenever $\rho = 1$ all the progeny strands copied from the negative templates will replicate in the following generations and replication will be purely geometric. However, if $\rho << 1$, the negative strands will be mainly used as templates while the positive ones will not replicate. With this second strategy the kinetics will be closer to the SMR. Indeed, to potentiate the effect of SMR, the non-replicating genomic strands are not degraded, and the degradation probability for the replicating sequences is kept very low. We also consider differential replication rates for each strategy of replication, by using $\delta\Gamma^+$ and $\delta\Lambda^+$. For SMR we set $\delta = 0.1$ and the genomic strands will synthesize few

antigenomic ones. For GR we use $\delta = 1$ and all the synthesized strands will be used as templates for further replication.

All numerical analyses of the deterministic mathematical model were done using a C program implemented to solve the differential equations with the standard fourth order Runge-Kutta method (40) using a constant time step size of $\Delta t = 0.1$. The stochastic bit string model was also implemented in a C program (available upon request).

RESULTS

Mathematical model

Quantitative differences in the accumulation of master and mutant genomes of both polarities. The first important analysis for any replication model involving antigenomic intermediates of replication should be to explore the kinetics of accumulation of both genomic and antigenomic strands and to understand how mutations accumulate in both strands. The effects of mutation rate in each replicating strategy are first illustrated in Figure 2. We show the temporal dynamics of all the viral strands for each replication strategy using three different mutation rates. The genomic strands (solid line) achieve higher equilibrium concentrations than the antigenomic ones (dashed line) for the SMR. For GR, however, the genomic and antigenomic strands always asymptotically achieve identical equilibria. With $\mu = 0.1$ (Figure 2a) the master strands (black trajectories) achieve population equilibria higher than the mutant strands regardless the replication kinetics. If mutation rate is increased ($\mu = 0.3$; Figure 2b), then the

concentration of the mutant strands grows while the master strands concentration decreases, although for SMR the concentration of mutant genomic strands achieves a higher value than the antigenomic master strands. For either replication mechanisms both master strands achieve higher population densities than the mutant strands if mutation rates are low. However, if mutation rate is increased (μ = 0.65) mutant strands dominate the population (Figure 2c). However, two important differences exist between replication modes. While SMR still produces non mutated genomes of both polarities and the population reaches a non trivial equilibrium, GR fails to sustain replication and the population extincts at the long term.

Figure 2d illustrates the initial growth kinetics for each replication mode for the genomic strands (in linear-log scale). For GR the initial growth phase is exponential. For the SMR strategy we obtain subexponential growth kinetics. The model also illustrates another basic difference between both replication strategies, namely the temporal dynamics of the ratio of positive to negative strands. For SMR, and with the parameter values used in this study, such ratio reaches an equilibrium at which the concentration of positive strands is approximately 3.5 times larger than that of negative strands (data not shown). However, as one may expect from the GR, this ratio flattens off at one, indicating an equal production of both strands (data not shown).

In conclusion, SMR is a more efficient replication mechanism: for low mutation rates, it produces more genomic strands than GR without the necessity of generating equivalent amounts of antigenomic strands. At high mutation rates, GR

drives to the extinction of the population while SMR still generates low amounts of mutation-free genomes. In the next section, we will explore under which mutation rates the master non-mutated strand disappears from the population.

SMR is compatible with a higher critical mutation rate. One of the more important predictions of the quasispecies theory is the existence of a critical mutation rate, μ_c, beyond which the wildtype sequences does not exist anymore in the population. As already mentioned in the Introduction, Eigen's model implicitly assumes GR. The question we are addressing in this section is how the mode of replication may impact μ_c. As expected, for SMR the equilibrium concentration for both master genomic ($[0+]$) and antigenomic ($[0-]$) strands as well as for both mutant strands ($[1+]$ and $[1-]$) is asymmetric (Figure 3a). This actually means that we have a higher production of positive strands from the negative ones. On the contrary, for GR the equilibrium concentrations for the master and the mutant strands are the same for both genomic and antigenomic strands. The critical mutation rate μ_c, involving the extinction of the master sequences, is shown to be 10,07% larger for SMR (μ_c = 0.765, Figure 3a) than for GR (μ_c = 0.695, Figure 3b), suggesting that the extinction of master genomes under SMR replication takes place at mutation rates larger than if genomes replicate according to GR.

From these results we can conclude that a virus replicating as a SMR would benefit from having a higher μ_c. In other words, it will be able of maintaining a well defined population structure at mutation rates for which a virus replicating as a GR will have undergone a collapse of the non-mutated strands.

SMR is more robust to the accumulation of slightly deleterious mutations.

So far, we have shown that a virus replicating as a SMR produces less mutant genomes and can sustain the population at mutation rates which are prohibitive for a virus replicating as a GR. Next, we sought to explore the effect of the severity of mutational effects on the accumulation of master and mutant strands of both polarities. The severity of mutations was computed as the ratio between the average replication rates of mutant Λ^+ and of master Γ^+ genomic strands. This ratio will be one for neutral mutations ($\Lambda^+ = \Gamma^+$) and zero for lethal ones. Γ^+ was fixed to 0.1 for SMR and to 1 for GR while $\Lambda^+ \in [0, \Gamma^+]$. Figure 4 shows the equilibrium population densities for each genomic class and their dependence on mutational severity. At relatively low mutation rates (μ = 0.15) the positive master sequence remains dominant regardless the replication mode. As expected, strong effect mutations accumulate less than those causing mild effects, irrespective of the mode of replication. However, GR is more sensitive to the accumulation of mild mutations than SMR (Figure 4a), as indicated by the steeper slope for the positive master strands. A similar situation occurs at intermediate mutation rates (μ = 0.25; Figure 4b): both replication modes accumulate more mild- than strong-effect mutations, but the GR accumulates proportionally more mild mutations. At higher mutation rates (μ = 0.6; Figure 4c) results change for SMR in an important way, that is, positive master genomes are not dominant anymore for the entire range of mutation severities and, instead, the mutant ones become the most abundant class, although it is still possible to recover the master genome along all the range of mutational severities at ~10% population frequency. Mild mutations are still the most commonly fixed ones. However, GR collapses at intermediate mutational

severities ($\Lambda^+ \approx 0.45$) and all genotypes get extinguished due to the accumulation of small-effect mutations.

Another difference between SMR and GR schemes is that at low mutation rates the second-most abundant genotype for SMR is the antigenomic master strand irrespective of the severity of mutations, whereas antigenomic mutants are the second most abundant class for GR. At intermediate mutation rates, the genomic mutants become the second most abundant class when replication occurs via SMR, and their frequency rises up as mutation rate increases. These results are in agreement to those presented in the previous section and support the notion that SMR model of virus replication is not only compatible with higher mutation rates but also it is more robust to the severity of mutations.

In general, strong effect mutations will have very low impact in the fitness of populations, with the extreme case being lethal mutations, which do not contribute to the next generation. By contrast, mutations of mild effect will accumulate in the populations since selection is poorly efficient removing them. We have shown here that viral populations with SMR accumulate less mild effects mutations than if replication was GR and, therefore, selection will be more efficient in keeping the population free of deleterious alleles (19).

Digital genomes

To incorporate the stochasticity inherent to viral infection and replication as well as to be able of analyzing different fitness landscape, in the following three sections we will move from deterministic models based on differential equations to stochastic models of simulation with digital genomes.

The equilibrium genotypic distributions differ among SMR and GR and show different dynamics towards the loss of the master sequence. We begin our study of the stochastic model by analyzing the effect of mutations across four different fitness landscapes for both replication modes. The variable we are measuring in this section is the equilibrium concentration for the master genomic sequences and its mutant spectrum using the per-bit mutation probability μ_b, and an initial population size of $N(0) = 50$ genomes. For the epistatic landscapes (Figure 1d), the epistasis coefficients were set to $\xi = 1.4$ and $\xi = 0.6$ for the synergistic and antagonistic cases, respectively. The critical mutation probability per bit, □□□, is defined as the lowest mutation value for which the master genomic strands concentration □□+ < 10^{-4} and disappears from the population. The results of these simulations are shown in Figure 5. In the case of the Swetina-Schuster single sharp peak landscape (Figure 1d), the extinction of the master sequence occurs at □□□ ≈ 0.156 for the SMR model and at □□□ ≈ 0.070 for the GR. Moreover, the composition of the mutant spectrum is shown to differ according to the replication strategy. It is well known that for the combination of a Swetina-Schuster landscape and GR the mutant spectrum suffers a sharp phase transition at □□□ and each mutant genome reaches a steady-state concentration that only depends on its mutational coupling (13). However, we show that this is not the case for the SMR, since such sharp phase transition is never observed and different genomes rise and decrease in frequency depending on mutation rate (Figure 5).

For more realistic landscapes, □□□ is also larger for the SMR than for the GR.

The difference between both replication strategies is maximized for the antagonistic landscape (7.43-fold difference) and minimized for the synergistic landscape (4.72-fold difference), while the additive landscape produces a 6.41-fold difference in □□□. No sharp phase transition is observed for any of these three landscapes.

Therefore, a very important conclusion can be drawn from this section: the sharp phase transition usually known as error threshold is characteristic of populations of GR replicators and if the fitness landscapes conforms the Swetina-Schuster model. Failure to meet any of these requirements results in a lack of sharp phase transition at the critical mutation rate.

Statistical analysis of the distribution of the number of mutations per genome. Next we sought to explore the effect of the two extreme models of RNA virus replication and of landscape topology on the properties of the cloud of mutants generated around the master sequence. Figure 6a shows the population frequency of master genomes for both replication modes and the four different fitness landscapes. A clear effect of the replication mode is observed. While a master genome replicating according to SMR will remain in the population at frequencies over the extinction threshold (dotted line; 10^{-4}) for $\mu_b \leq 0.148$, irrespective of the landscape employed, a master genome using a GR strategy always crosses the extinction threshold at much lower mutation rates ($\mu_b \leq 0.031$; Figure 6a). Indeed, irrespective of the fitness landscape and the mutation rate employed in the simulations, the distributions of the number of mutations per genome were significantly different both in location (Figure 6b; Mann-Whitney test, $P < 0.001$ in all comparisons) and shape (Figure 6c; Kolmogorov-Smirnov test,

$P < 0.001$ in all comparisons); with GR always generating a much larger average mutational load and a larger variance in the number of mutations per genome. Figure 6b reflects once again the sharp phase transition characteristic of the GR in a Swetina-Schuster landscape, with genomes increasing their average number of mutations up to 16 (Figure 6b). For the other landscapes, the average number of mutations per genome increases smoothly. For the SMR, the landscape topology does not affect in a large extent the average number of mutations per genome. However, for the GR, the antagonistic landscape generates the largest mutational load for all mutation rates tested. By contrast, the synergistic landscape generates the lowest mutational loads. Therefore, the magnitude of the benefit of SMR over GR (i.e., the difference between the surfaces under both curves) varies across landscapes, being the largest for the Swetina-Schuster, then for the antagonistic, then the additive and finally the smallest difference corresponding to the synergistic landscape.

Similarly, the standard deviation of the number of mutations per genome is also affected by the interplay between the mode of replication and the landscape topology (Figure 6c). Again, the sharp phase transition characteristic for the combination of Swetina-Schuster landscape and GR is reflected in the observed sudden increase in variability in the number of mutations per genome as mutation rate approaches the critical value. However, this fast increase was only transient until the average mutational load reached its maximum value. Afterwards, the variance in the number of mutations sharply decreased and asymptotically approached the standard deviation value observed for the SMR (Figure 6c). For the

other fitness landscapes, the behavior is monotonous. At low mutation rates GR produces larger variances than SMR regardless the landscape topology, with the following rank order: antagonistic, additive and synergistic. As mutation rate increases, however, the differences among GR and SMR are reduced and the curves intersect at some value of μ_b (Figure 6c) beyond which GR produces less variable populations than SMR. The precise value at which SMR losses the advantage over GR depends on the landscape topology; being the synergistic the first one, followed by the antagonistic and finally the additive.

As described in the Introduction, under a purely SMR mechanism, the number of mutations per genome should conform to a Poisson distribution, whereas for a GR mechanism the distribution departs from the Poisson model and fits the more complex Luria-Delbrück distribution. To confirm this expectation, we run Kolmogorov-Smirnov tests for deviations of the Poisson null hypothesis. As expected, under the SMR, the numbers of mutations per genome were Poisson distributed for all combinations of μ_b and landscapes ($P \geq 0.050$ in all tests). Furthermore, the distributions of mutations per genome generated under the GR scheme departed from null Poisson model for values of $\mu_b \geq 0.035$ for the Swetina-Schuster, additive and antagonistic landscapes ($P \leq 0.019$ in all tests) and of $\mu_b = 0.015$ for the synergistic one ($P = 0.011$).

The take home message from this section is that if one looks for differences between SMR and GR in terms of where in sequence space is placed the cloud of mutants and how big it is, the results strongly depends on the fitness landscape. Whereas SMR is quite insensitive to the topology, GR is greatly affected.

Interplay between epistasis and critical mutation rate. In the previous section, we observed that for a given replication mode the most divergent results were always observed for the two extremes of the epistatic landscape. To further characterize the effect of the epistasis in the replication dynamics, next we sought to explore the effect of the fitness landscape curvature (ξ) on σ_{max}. To do so, we run simulation experiments similar to those described in the above sections but varying ξ in 0.1 intervals (Figure 7). Irrespective of the mode of replication, the shape of the relationship was similar: antagonistic epistasis ($\xi < 1$) barely affects σ_{max}, whereas increasing the strength of synergistic epistasis ($\xi > 1$) allows for almost linear increases in σ_{max} of up to 14% for $\xi = 2$. In good agreement with the results shown in previous sections, what most strongly determines the magnitude of σ_{max} is the mechanism of replication, being, around one order of magnitude larger for SMR than for GR across the entire range of ξ.

From these results we conclude that the critical mutation rate is insensitive to antagonistic epistasis but it dramatically increases as epistasis become more and more synergistic. The biological meaning of this difference is clear. Antagonistic epistasis means that genomes with multiple mutations are fitter than expected under a multiplicative model and, therefore, may persist in the population for long periods of time before being removed by selection. Therefore, the extinction of the master genotype is only weakly affected by changes in σ_{max}. By contrast, increasing synergistic epistases means that the pernicious effect of multiple mutations goes beyond the multiplicative expectation and thus, mutant genomes

will be removed from the population in a very efficient way by natural selection; henceforth, only the master sequence survives and produces progeny. This "population robustness" effect allows □□□ increasing up to the point in which it is not possible anymore to produce progeny of master sequences from a master template.

DISCUSSION

It is known that the mode of virus replication can change the rates at which deleterious mutations occur. Depending on whether genomes replicate according to a stamping machine model or geometrically one may observe different fractions of mutation-free genomes in the offspring. Luria's stamping machine replication (25) implies that the copies produced during the first round of replication will be the only templates for the generation of the entire viral population. Experimental data suggest that $\phi X174$ replicates in such a way (7). However, in geometric replication, as for example in the case of phage *T*2 (25), all the synthesized copies during infection will serve as templates in the following generations. Intermediate situations have also been described for other viruses like the phage $\phi 6$ (4). As far as we know, previous attempts to model virus replication have not integrated the many features that we have explored here: mode of replication involving genomic and antigenomic strands, topology of fitness landscape and mutation rates.

In this work we analyze a model describing the within-cell replication of positive-sense RNA viruses that make no subgenomic mRNAs and encode a single polyprotein that is self-processed into structural and nonstructural components.

We develop a mathematical model using as key variables the genomic and antigenomic RNA strands (considering either mutation-free and mutant genomes), the viral polyprotein, the translational complexes formed by viral RNA and cellular ribosomes, and the mature virions. However, the model is defined in quite general terms, so it can be easily modified to account for any other genomic organization. Genetic heterogeneity is introduced with a simplified quasispecies structure of the strands considering master genomes (which have the higher fitness) and the pool of deleterious mutants for each strands polarity, which are grouped into a single variable denoting an average mutant sequence different from the master one. We analyze a Swetina-Schuster single-peak fitness landscape (41) which assumes that mutations have a large deleterious effect. We explore the critical mutation rate at which master sequences disappear and the sensitivity to mutations for viral populations replicating as SMR or GR. We have shown that the extinction of the master genome takes place at a higher mutation rate for the SMR, indicating that this strategy is less sensitive to the effect of deleterious mutations. On the contrary, GR displays a lower critical mutation rate, being more sensitive to mutation. Consistently, we have also shown that mild mutational effects tend to accumulate more in the GR, and that under such a growth kinetics the population collapses at intermediate mutation rates. However, the SMR strategy is less sensitive to the accumulation of mild mutations and viral strands continue existing for mutation rates that produce a collapse of viral populations replicating geometrically.

A critical assumption of our model is that mutations have two immediate effects. First, they preclude the right interaction of mutant RNAs with ribosomes

thus fully avoiding their translation and, second, they reduce their efficiency interacting with the wildtype replicase thus negatively affecting their transcription. Certainly, with the exception of mutations generating stop codons, severely modifying ribosomal entry sites or altering signals for replicase binding, mutations will not necessarily have such assumed severe effect. By removing these assumptions and moving to the proteins the effect of mutations, we will clearly increase the realism of our models. However, this increase in realism will come with a cost in terms of model complexity due to the necessity of incorporating interactions between mutant proteins and the four RNA species, each interaction governed by a different set of parameters. At the one side, complementation of defective genomes by wildtype proteins will increase the chances of mutant genomes to persist in the viral population, effectively reducing the strength of purifying selection (44). However, this possibility does not represent a problem in our formulation because we assume that mutant genotypes are only replicated and encapsidated by wildtype proteins. At the other side, the presence of proteins with impaired functions will reduce the replication/encapsidation efficiency of both wildtype and mutant genomic RNAs. This second option has not been taken into consideration in our current formulation and is being the subject of ongoing research. Finally, readers should keep in mind that our model is an extension of the basic quasispecies model (12, 13) that so much has impacted the virology community. As in Eigen's original formulation, the fitness of mutant genotypes does not depend on mutations affecting translated proteins but is an inherent property of the RNA molecules.

To gain further insights into the replication process and the genesis and fate of variability, we complement the analysis of the mathematical model with a model of digital genomes which considers the intrinsic noise due to small population sizes and incorporates different fitness landscapes, including the above mentioned Swetina-Schuster single-peaked plus the more realistic and rugged additive and epistatic ones. This model consistently shows that the critical per-bit mutation rate is higher for the SMR case and the mutational load lower regardless the fitness landscape used in the simulations. However, the topography of the fitness landscapes does, indeed, affect the results. The error threshold and the transition to error catastrophe predicted by the quasispecies theory (13) only holds for the Swetina-Schuster landscape and for GR. However, by assuming a SMR model, which should be the norm for many viruses, even with this unrealistic fitness landscape the transition disappears. Furthermore, neither the additive nor the epistatic (on its two configurations) landscapes show such transition. Therefore, all the results here shown suggest that the existence of an error threshold is very sensitive to the election of the topology of the fitness landscape and of the mode of replication and, thus, caution needs to be put against its application to real RNA virus populations for which nothing is known about their fitness landscapes nor their mode of replication.

Recent results suggest that antagonistic epistasis should dominate in viral genomes (2, 30, 34, 42). This being the case, our results suggest that natural selection would have favored viruses using a SMR mechanism, as a way of reducing mutational load. With the combination of antagonistic epistasis and SMR,

not only the average number of mutations per genome will be minimized and the critical mutation rate increased, but also the range of mutation rates in which SMR is still a better choice than GR is larger than for a synergistic landscape. It is generally assumed that RNA viruses have mutation rates in the order of magnitude of the inverse of genome length (11). For the digital genomes used here, this would be equivalent to a mutation rate of 1/32 = 0.031. Even at this value, the combination of an antagonistic landscape and of a SMR model provides the best solution. Antagonistic epistasis usually appears in small genomes as a consequence of the lack of genetic redundancy that also makes genomes extremely sensitive to the effect of mutations (20, 32). Due to their parasitic lifestyle, RNA viruses have been selected for fast replication at the cost of an increased sensitivity to mutations. However, a pure SMR strategy may minimize the impact of mutation accumulation and preserve in the population mutation-free genomes at frequencies that would not be possible by adopting a pure GR strategy. In this sense, SMR may represent a mechanism of genetic robustness.

Whether RNA viruses may have evolved some mechanisms to buffer the deleterious effects of mutations has recently attracted the attention of researchers (5, 27, 28, 31, 43). Robustness is defined as a reduced sensitivity to perturbations affecting phenotypic expression (8). RNA virus populations may owe their robustness to several of the following mechanisms (reviewed in (16)). First, the above mentioned hypersensitivity to mutational effects of individual genomes translates into robustness at the population level as a consequence of a more efficient purifying selection that maintains average fitness high (20). Second, high

mutation rates characteristic of RNA viruses may impose a strong selective pressure that pushes virus populations towards regions of the sequence space where the density of neutral mutations is higher (5, 31). Third, the variable and random ploidy of viruses and the frequent coinfection events enhance the possibility of genetic complementation. Fourth, segregation of segments during mixed infections and homologous recombination are forms of sex that may recreate mutation-free genomes (28). Fifth, cellular buffering mechanisms (e.g., heat-shock proteins) can be utilized by the viruses in their own benefit as an extrinsic source of robustness. The results reported in this study suggest that, in addition to these five potential mechanisms, compacted genomes characterized by an excess of antagonistic epistasis may benefit from a SMR model because they will accumulate less deleterious mutations, have a higher critical mutation rate and suffer in a lesser extent from the effect of deleterious mutations, that is, they will have increased their robustness.

ACKNOWLEDGEMENTS

We thank the constructive comments and suggestions from the three reviewers of a previous version of this work. This work has been supported by the Human Frontier Science Program Organization grant RGP12/2008. Work in València was also funded by the Spanish Ministerio de Ciencia e Innovación grant BFU2006-14819-CO2-01/BMC. We also acknowledge support from The Santa Fe Institute.

REFERENCES

1. **Ball, L. A.** 1995. *Fields Virology*, chapter 5. Replication Strategies of RNA Viruses. Lippincott Williams & Wilkins. Fourth Edition.

2. **Bonhoeffer, S., C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos.** 2004. Evidence for positive epistasis in HIV-1. Science **306**:1547-1550.

3. **Carrasco, P., F. de la Iglesia, and S. F. Elena.** 2007. Distribution of fitness and virulence effects caused by single-nucleotide substitutions in *Tobacco etch virus*. J. Virol. **81**:12979-12984.

4. **Chao, L., C. U. Rang, and L. E. Wong.** 2002. Distribution of spontaneous mutants and inferences about the replication mode of the RNA bacteriophage ϕ6. J. Virol. **76**:3276–3281.

5. **Codoñer, F. M., J. A. Daròs, R. V. Solé, and S. F. Elena.** 2006. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. PLoS Pathog. **2**:e136

6. **Dahari, H. R. M. Ribeiro, C. M. Rice, and A. S. Perelson**. 2007. Mathematical modeling of subgenomic *Hepatitis C virus* replication in Huh-7 cells. J. Virol. **81**:750–760.

7. **Denhart, D. and R. B. Silver**. 1966. An analysis of the clone size distribution of φX174 mutants and recombinants. Virology **30**:10–19.

8. **de Visser, J. A. G. M., J. Hermisson, G. P. Wagner, L. Ancel Meyers, H. Bagheri-Chaichian, J. L. Blanchard, L. Chao, J. M. Cheverud, S. F. Elena, W. Fontana, G. Gibson, T. F. Hansen, D. C. Krakauer, R. C. Lewontin, C. Ofria,**

S. H. Rice, G. von Dassow, A. Wagner, and M. C. Whitlock. 2003. Evolution and detection of genetic robustness. Evolution **57**:1959-1972.

9. **Dewanji, A., E. G. Luebeck, and S. H. Moolgavkar**. 2005. A generalized Luria-Delbrück model. Math. Biosci. **197**:140–152.

10. **Domingo, E., E. Baranowski, C. M. Ruiz-Jarabo, A. M. Martín-Hernández, J. C. Sáiz, and C. Escarmís**. 1998. Quasispecies structure and persistence of RNA viruses. Emerging Infect. Dis. **4**:521–527.

11. **Drake, J. W. and J. J. Holland.** 1999. Mutation rate among RNA viruses. Proc. Natl. Acad. Sci. USA **96**:13910-13913.

12. **Eigen, M., C. K. Biebricher, M. Gebinoga, and W. C. Gardiner**. 1991. The hypercycle. Coupling of RNA and protein synthesis in the infection cycle of an RNA bacteriophage. Biochemistry **30**:11005–11018.

13. **Eigen, M., J. McCaskill, and P. Schuster**. 1989. The molecular quasispecies. Adv. Chem. Phys. **75**:149–263.

14. **Elena, S. F.** 2001. Evolutionary history conditions the timing of transmission in *Vesicular stomatitis virus*. Infect. Genet. Evol. **1**:151-159.

15. **Elena, S. F., P. Agudelo-Romero, P. Carrasco, F. M. Codoñer, S. Martín, C. Torres-Barceló, and R. Sanjuán**. 2008. Experimental evolution of plant RNA viruses. Heredity **100**:478–483.

16. **Elena, S. F., P. Carrasco, J. A. Daròs, and R. Sanjuán**. 2006. Mechanisms of genetic robustness in RNA viruses. EMBO Rep. **7**:168–173.

17. **Endy, D., D. Kong, and J. Yin**. 1997. Intracellular kinetics of a growing virus: A genetically structured simulation for the bacteriphage T7. Biotechnol. Bioeng.

55:375–389.

18. **French, R. and D. C. Stenger.** 2003. Evolution of *Wheat streak mosaic virus*: dynamics of population growth within plants may explain limited variation. Annu. Rev. Phytopathol. **41**:199–214.

19. **Krakauer, D. C. and N. L. Komarova**. 2003. Levels of selection in positive-strand virus dynamics. J. Evol. Biol. **16**:64–73.

20. **Krakauer, D. C. and J. B. Plotkin**. 2002. Redundancy, antiredundancy, and the robustness of genomes. Proc. Natl. Acad. Sci. USA **99**:1405-1409.

21. **Landau, D. P. and K. Binder.** 2000. *A Guide to Monte Carlo Simulations in Statistical Physics*, pp. 70-73. Cambridge University Press, Cambridge, UK.

22. **Leuthäusser, I.** 1986. An exact correspondence between Eigen's evolution model and a two-dimensional Ising system. J. Chem. Phys. **84**:1884–1885.

23. **Leuthäusser,** I. 1987. Statistical mechanics of Eigen's evolution model. J. Stat. Phys. **48**:343–360.

24. **Lim, K., V. Lang, T. Lam, and J. Yin**. 2006. Model-based design of growth-attenuated viruses. PLoS Comput. Biol. **2**(9):e116.

25. **Luria. S. E.** 1951. The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage production. Cold Spring Harbor Symp. Quant. Biol. **16**:463–470.

26. **Malpica, L. M., A. Fraile, I. Moreno, C. I. Obies, J. W. Drake, and F. García-Arenal**. 2002. The rate and character of spontaneous mutation in an RNA virus. Genetics **162**:1505–1511.

27. **Martín, V., A. Grande-Pérez, and E. Domingo.** 2008. No evidence of selection

for mutational robustness during lethal mutagenesis of *Lymphocytic choriomeningitis virus*. Virology **378**:185-192.

28. **Montville, R., R. Froissart, S. K. Remold, O. Tenaillon, and P. E. Turner.** 2005. Evolution of mutational robustness in an RNA virus. PLoS Biol. **3**:e381.

29. **Reddy, B. and J. Yin**. 1999. Quantitative intracellular kinetics of HIV type 1. AIDS Res. Hum. Retrovir. **15**:273–283.

30. **Sanjuán, R**. 2006. Quantifying antagonistic epistasis in a multifunctional RNA secondary structure of the *Rous sarcoma virus*. J. Gen. Virol. **87**:1595-1602.

31. **Sanjuán, R., J. M. Cuevas, V. Furió, E. C. Holmes, and A. Moya.** 2007. Selection for robustness in mutagenized RNA viruses. PLoS Genet. **3**:93.

32. **Sanjuán, R. and S. F. Elena**. 2006. Epistasis correlates to genomic complexity. Proc. Natl. Acad. Sci. USA **103**:14402-14405.

33. **Sanjuán, R., A. Moya, and S. F. Elena.** 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. USA **101**: 8396-8401.

34. **Sanjuán, R., A. Moya, and S. F. Elena.** 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. Proc. Natl. Acad. Sci. USA **101**:15376-15379.

35. **Sardanyés, J., S. F. Elena, and R. V. Solé**. 2008. Simple quasispecies models for the survival-of-the-flattest effect: the role of space. J. Theor. Biol. **250**:560–568.

36. **Shukla, D.D., C. W. Ward, and A. A. Brunt.** 1994. The *Potyviridae*. Wallingford: CAB International.

37. **Sidorenko, Y. and U. Reichl.** 2004. Structured model of Influenza virus

replication in MDCK cells. Biotech. Bioeng. **88**:1–14.

38. **Solé, R. V., J. Sardanyés, J. Díez, and A. Mas**. 2006. Information catastrophe in RNA viruses through replication thresholds. J. Theor. Biol. **240**:353–359.

39. **Srivastava, R., L. You, J. Summers, and J. Yin**. 2002. Stochastic vs. deterministic modeling of intracellular viral kinetics. J. Theor. Biol. **218**:309–321.

40. **Strogatz, S. H.** 2000. *Nonlinear Dynamics and Chaos with applications to Physics, Biology, Chemistry, and Engineering*. Westview Press.

41. **Swetina, J. and P. Schuster**. 1982. Self-replication with errors. A model for polynucleotide replication. Biophys. Chem. **16**:329.

42. **van Opijnen, T., M. C. Boerlijst, and B. Berkhout.** 2006. Effects of random mutation in the *Human immunodeficiency virus* type 1 transcriptional promoter on viral fitness in different host cell environments. J. Virol. **80**:6678-6685.

43. **Wagner, A. and P. F. Schuster.** 1999. Viral RNA and evolved mutational robustness. J. Exp. Zool. **285**:119-127.

44. **Wilke, C. O. and I. S. Novella**. 2003. Phenotypic mixing and hiding may contribute to memory of viral quasispecies. BMC Microbiol. **3**:11.

45. **You, L., P. F. Suthers, and J. Yin**. 2002. Effects of *Escherichia coli* physiology on growth of phage T7 in vivo and *in silico*. J. Bacteriol. **184**:1888–1894.

46. **Zhdanov, V. P.** 2004. Bifurcation in a generic model of intracellular viral kinetics. J. Phys. A Math. Gen. **37**:L63–L66.

FIG. 1. (a) Schematic representation of the virus infectious cycle. During infection, the viral particle enters into the host cell and after uncoating, the genomic RNA (acting as mRNA) forms the translational complex, T_c, by binding with ribosomes, that directs the synthesis of the viral polyprotein, p, which is converted to both structural and nonstructural proteins, at rates $1 - \beta$ and β, respectively. The nonstructural proteins are involved in the synthesis of genomic and antigenomic RNAs. We distinguish between master and mutant genomic (+) and antigenomic (–) strands. We simulate SMR with $\Gamma^{+} << \Gamma^{-}$, where the initial genomic RNA directs the synthesis of one or very few negative copies which are used as templates for the synthesis of new genomic strands. To model GR we use $\Gamma^{+} = \Gamma^{-}$, where all the synthesized strands replicate at the same rate (see Table 1 for a description of the variables and the parameters used in the model). (b) Sequence spaces for a population of genomic and antigenomic binary strands of length $L = 3$. The consideration of genomic and antigenomic senses can be interpreted as two coupled hypercubes $\mathcal{H}$□±. Each node of the hypercube generates by replication of the complementary strand, □□± (blue line), or a mutant one □□± (yellow line), in the coupled sequence space. (c) Diagram of the replication, mutation and degradation rules implemented for the simulations. (d) Fitness landscapes analyzed in this study. The plot shows the effect of increasing numbers of mutations on the fitness of the genotype: the Swetina-Schuster (black), additive (blue), synergistic (green), antagonistic (red) fitness landscapes. The parameters defining the landscapes shown in panel (d) are arbitrary.

1

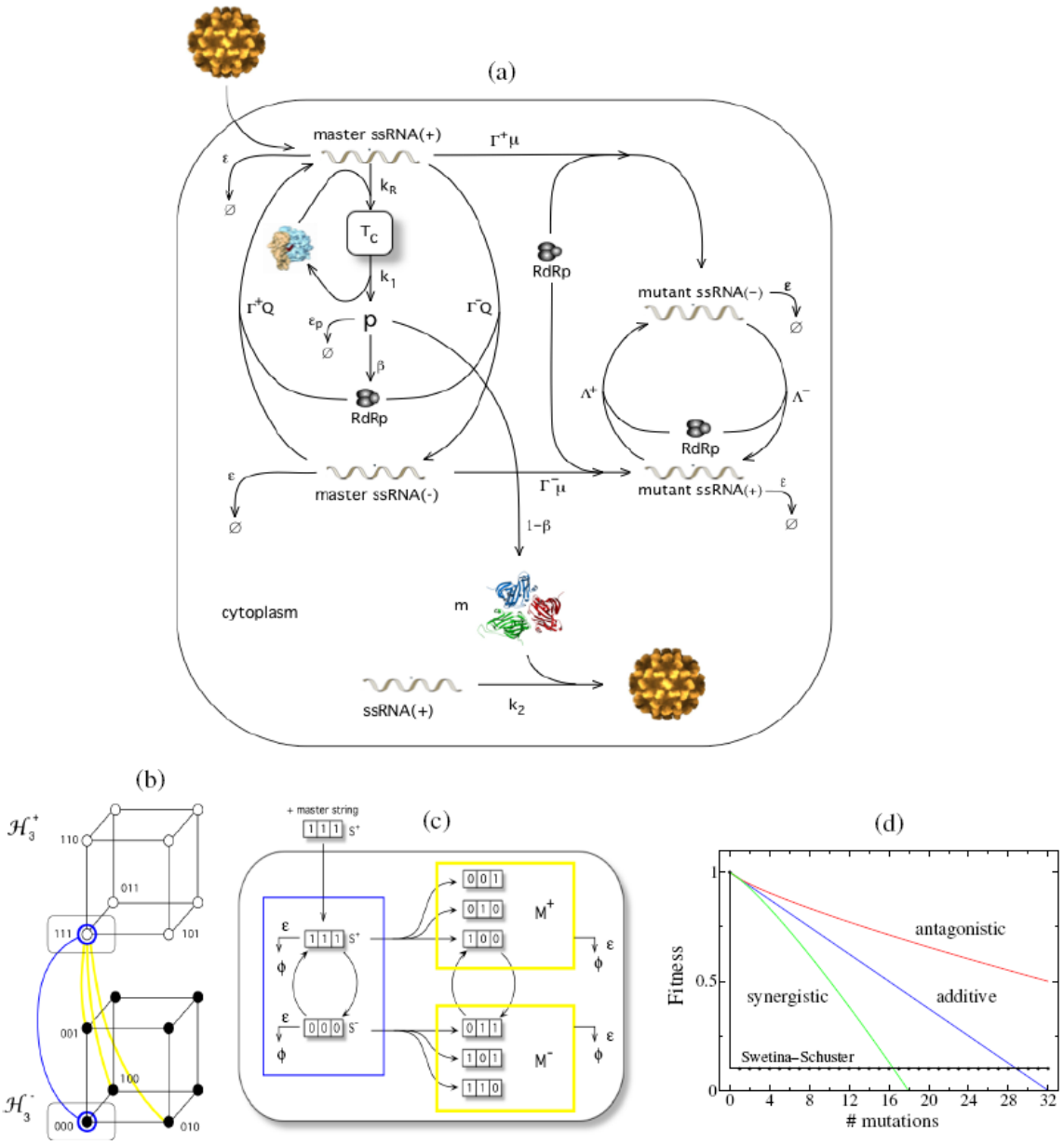
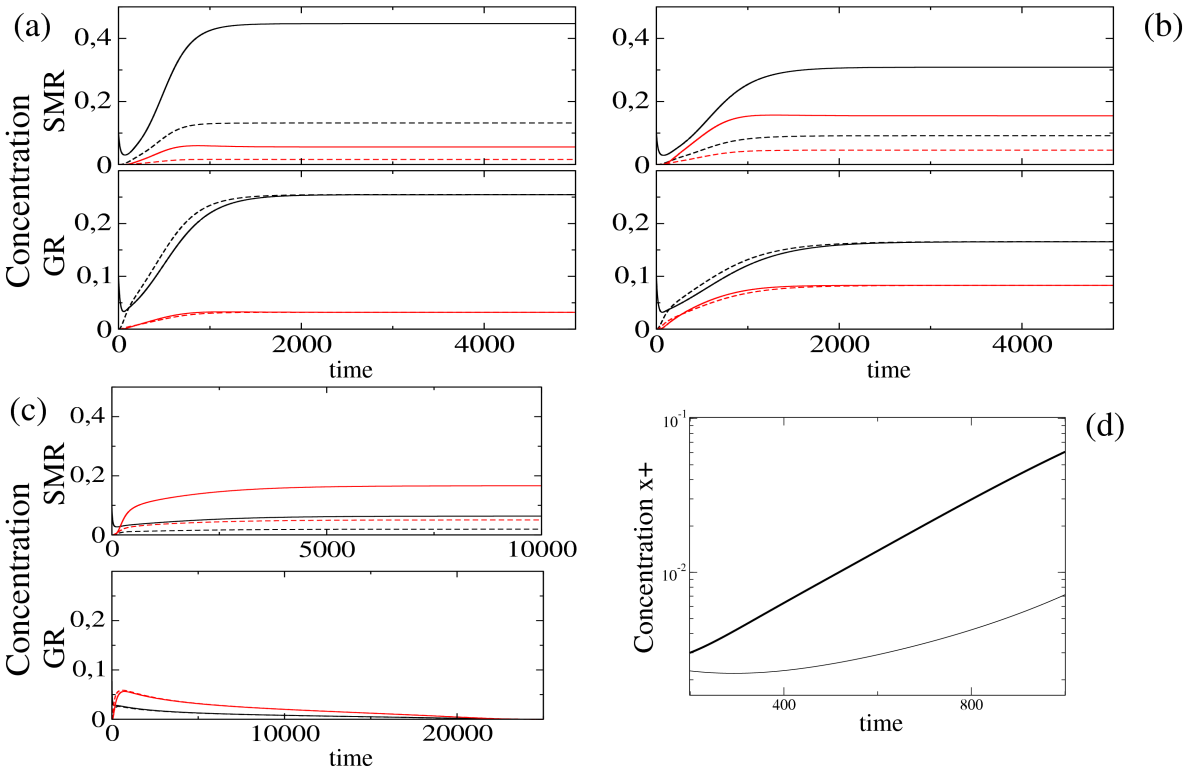

(a)
master ssRNA(+)
$\Gamma^+\mu$
k_R
T_C
k_1
Γ^+Q
ε_p
p
Γ^-Q
β
RdRp
RdRp
mutant ssRNA(−)
ε
Λ^+
Λ^-
RdRp
master ssRNA(-)
$\Gamma^-\mu$
mutant ssRNA(+)
$1-\beta$
cytoplasm
m
ssRNA(+)
k_2
(b)
$\mathcal{H}_3^+$
110
011
111
101
001
100
$\mathcal{H}_3^-$
000
010
(c)
+ master string
1 1 1 S+
M^+
M^-
0 0 1
0 1 0
1 0 0
0 1 1
1 0 1
1 1 0
0 0 0 S−
(d)
Fitness
mutations
antagonistic
synergistic
additive
Swetina–Schuster
0
0.5
1
0
4
8
12
16
20
24
28
32

FIG. 2. Solutions of the mathematical model. Time series for master (black) and mutant (red) strands at the following mutation rates: (a) μ = 0.1, (b) μ = 0.3 and (c) μ = 0.65. Genomic and antigenomic strands are indicated, respectively, with solid and dashed lines. In all the plots we show the time evolution for SMR (upper panel, with Γ^{+} = 0.1) and GR (lower panel, with Γ^{+} = 1). Panel (d) shows the initial amplification phase for the genomic strands (in linear-log scale) for GR (thick line) and SMR (thin line) using μ = 0.3. Initial condition x_{0+0} = 0.1 and all other parameters as in Table 1.



1

2 FIG. 3. Equilibrium concentrations for master (⍰*0±*) and mutant (⍰*1±*) strands
3 against mutation rate, μ, for SMR (a) and GR (b) modes. All other parameters are
4 as in Table 1. To facilitate the comparison, the arrow indicates the critical mutation
5 rate obtained for the SMR
6 model.

7

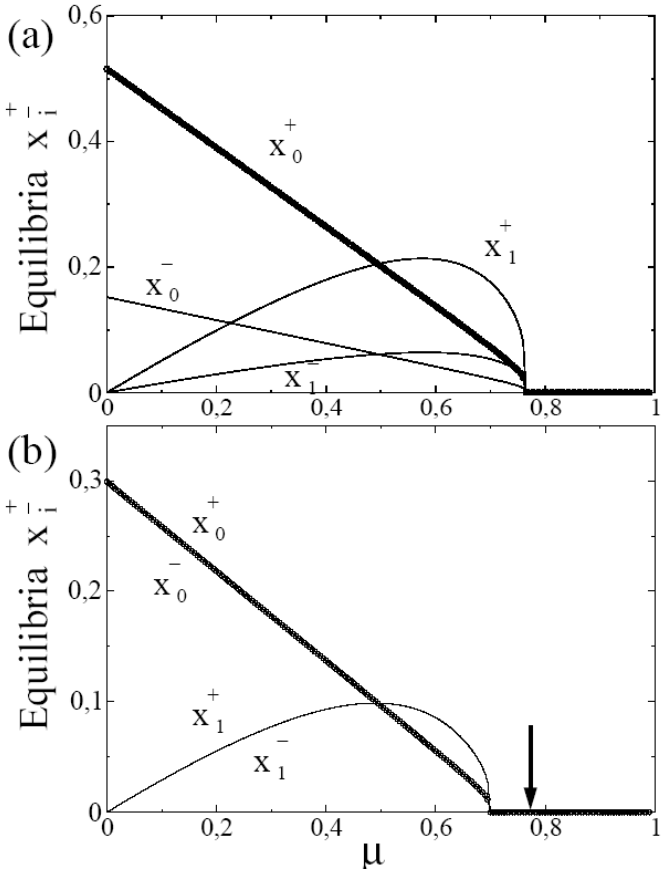
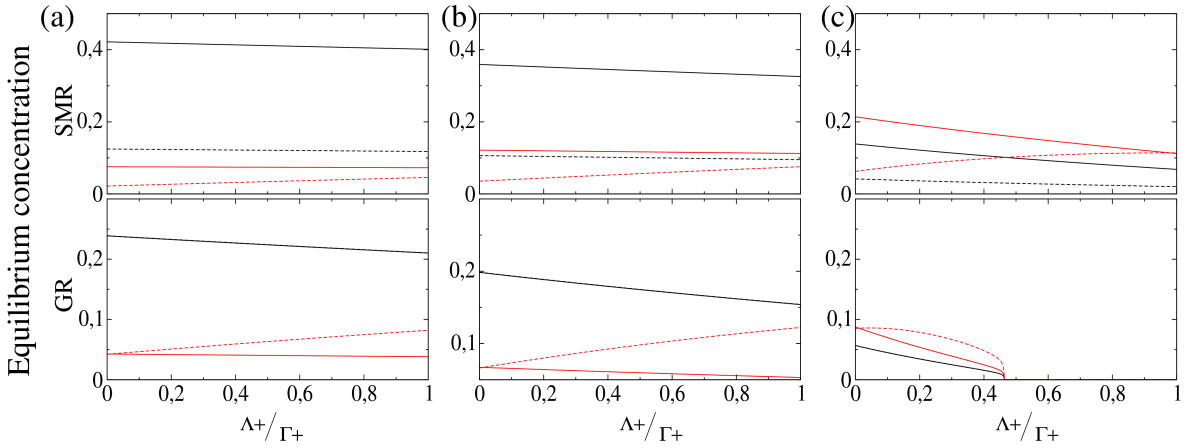


FIG. 4. Severity of mutations and accumulation of the different strands. Relative effect of mutations on the equilibrium concentration of the strands measured as the ratio Λ^+/Γ^+ for different values of the mutation rate: (a) μ = 0.15, (b) μ = 0.25 and (c) μ = 0.35. In all the plots we represent (upper) SMR using Γ^+ = 0.1 and (lower) GR with Γ^+ = 1. Initial condition $0+0$ = 0.1 and all other parameters as in Table 1.



1

FIG. 5. Average equilibrium concentration of genomic strands in the bit string model for the four fitness landscapes studied as a function of per bit mutation rate. From left to right: Swetina-Schuster, antagonistic epistasis (ξ = 0.6), additive (ξ = 1), and synergistic epistasis (ξ = 1.4). The upper row represents the SMR and the lower row the GR models. For the Swetina-Schuster ε = 0.01 and ε = 0.001 for all other landscapes. In all the diagrams we show the normalized population numbers (averaging over 200 independent replicates when τ = 20000) for genomic master strands (thick line) and their mutant spectrum (thin lines). $N(0)$ = 50 in all simulations.

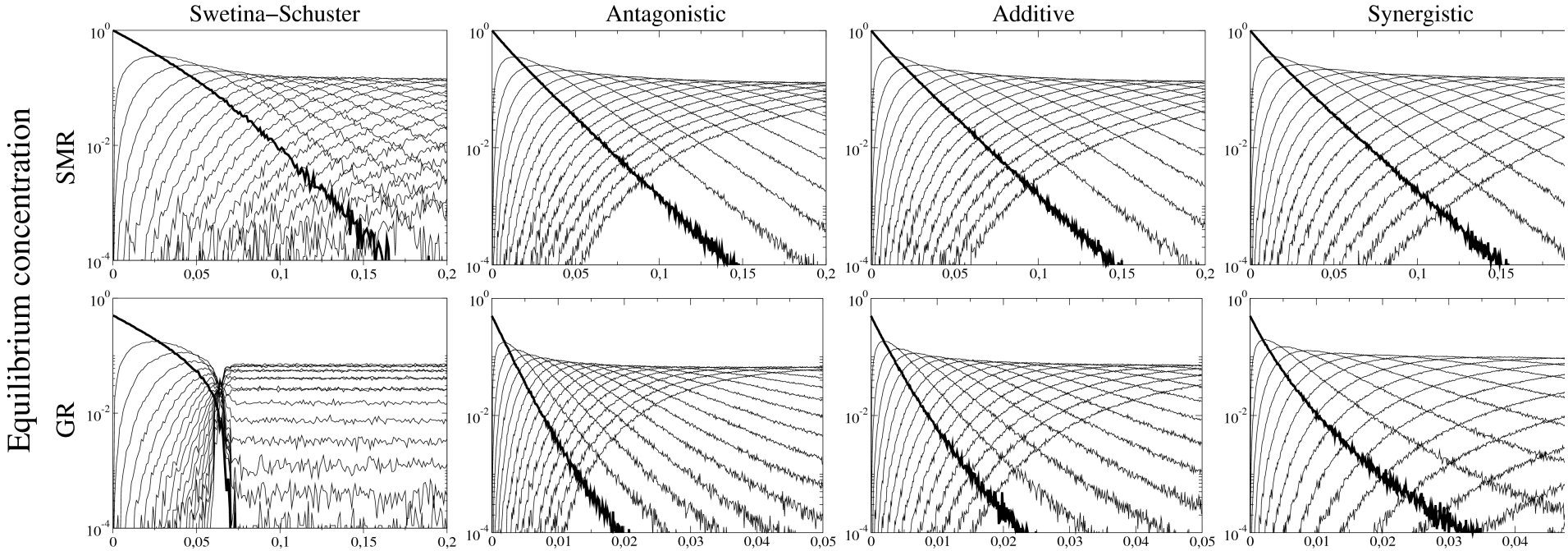


FIG. 6. Interplay between mutation rate, mode of replication and the topology of the fitness landscape. SMR and GR are represented with solid and dashed lines, respectively. Landscapes are represented with different colors (see legend for information). (a) Frequency of the mutation-free genomic strands. The dotted horizontal line represents the extinction threshold. (b) Average number of mutations per genome as a function of mutation rate. (c) Standard deviation of the number of mutations per genome.

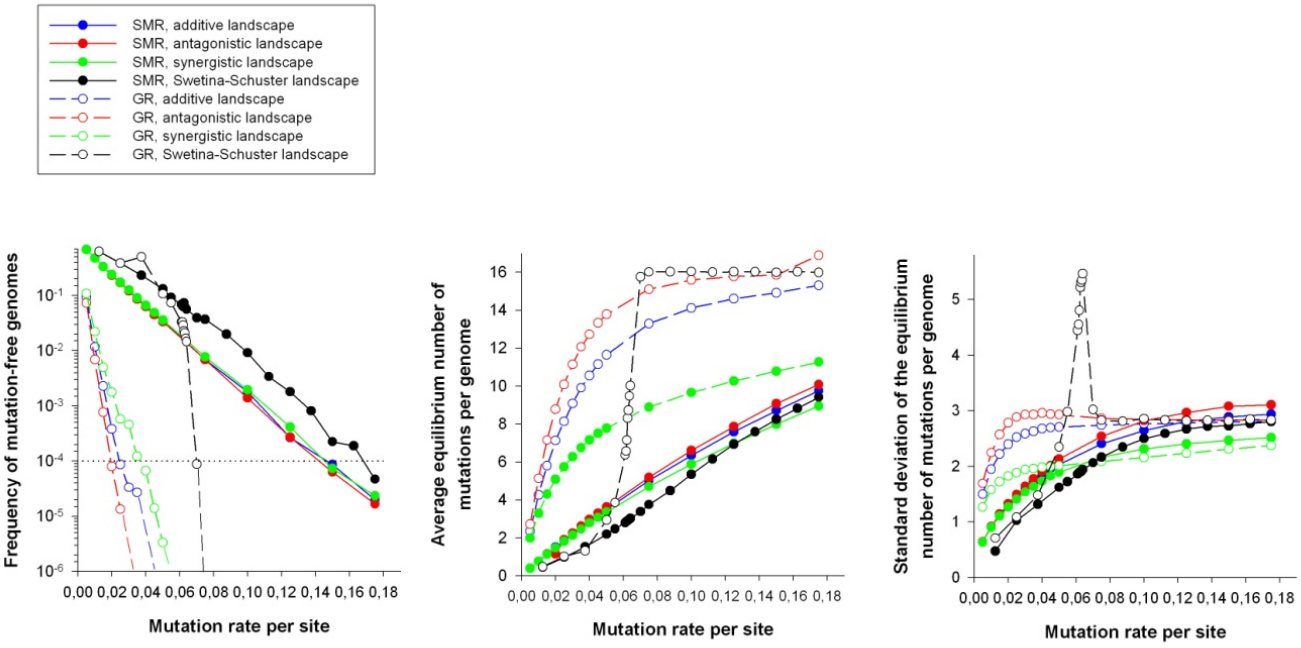


FIG. 7. Mean critical mutation rate, $\bar{\mu}_{cb}$, as a function of the sign and strength of epistasis, ξ. Each data point is the result of averaging 20 independent runs at equilibrium. Error bars represent SD. Such population equilibria values were obtained averaging the entire population at generation 20000 over 200 independent runs. The vertical dotted line at $\xi = 1$ corresponds to the additive fitness landscape. (a) SMR, (b) GR.

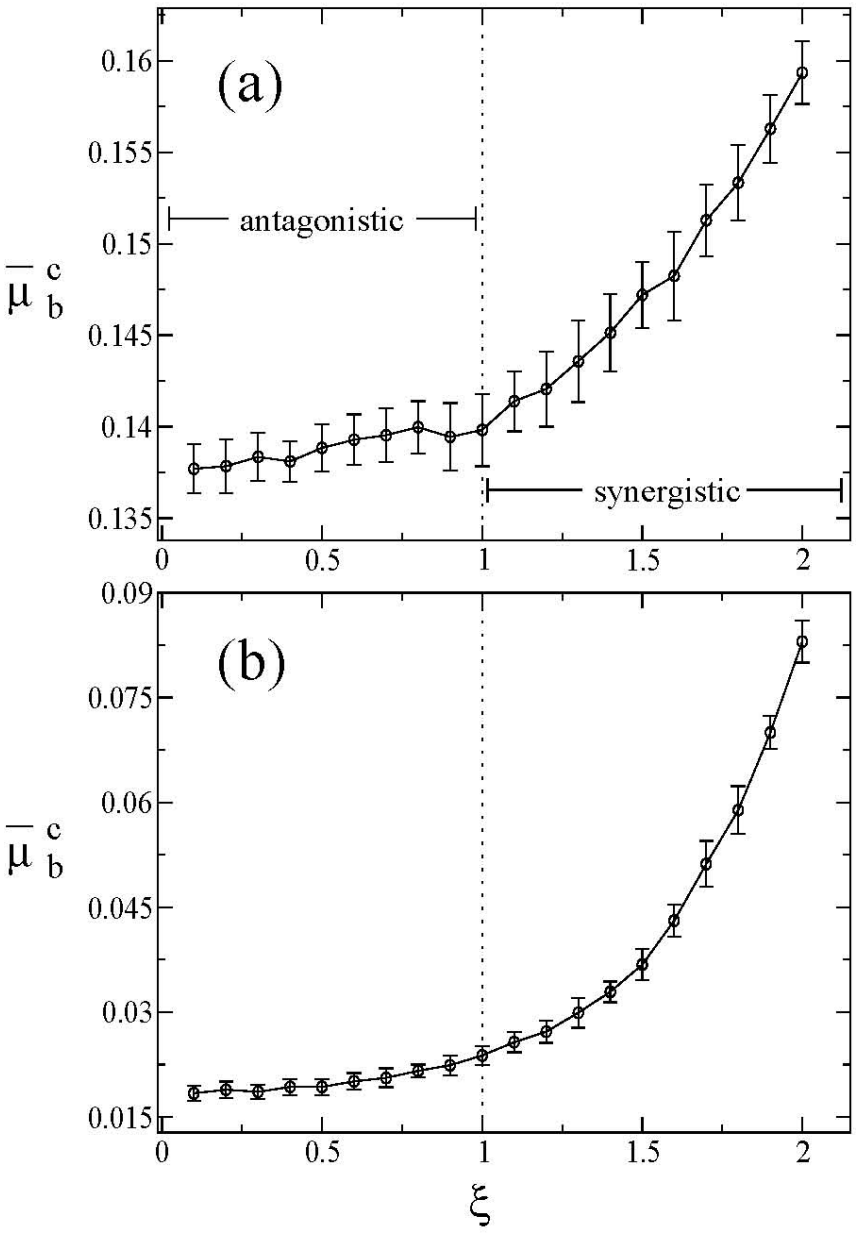


TABLE 1. Notations used in the differential equations model and the values assigned to each parameter.

Notation	Description	Value
β	Fraction of the viral polyprotein used as replicase, being 1 – β, the fraction used as structural proteins	0.9*
ε	Strands degradation rate	0.005†
ε_T	Degradation rate of the translational complex	10^{-5}#
ε_p	Degradation rate of the viral polyprotein	1.5×10^{-3}§
$\Gamma^{\pm}$	Probabilities of replication for the genomic (+) and antigenomic (–) master strands	0 - 1
Λ^{*}	Probability of replication for the genomic and antigenomic mutant strands	0 – 0.1
k_R	Effective interaction rate between the master genomic strands and the available ribosomes	0.04#
k_1	Dissociation rate of the master genomic strand from the translational complex	0.02#
k_2	Encapsidation probability of genomic strands	0.75
m	Number of monomers of structural protein necessary for building up a virion	200$
μ	Average mutation rate	0 - 1
R^{tot}	Constant number of ribosomes inside the cell	1
σ	Rate of elimination of mature virions (either by degradation or by leaking out of the cell)	3.5×10^{-3}‡

*The nonstructural genomic components represent ~90% of the genome in picorna-like viruses (1).

†RNA stability depends on intrinsic properties of the molecule such as sequence structure, presence of CAP or VPg or poly(A) tail in as much as from the action of cellular RNases and thus it must be independent on the replication model.

We assume that mutations have negligible effects on RNA stability. Since we found no useful values in the literature, we

performed a sensitivity analysis for ε in the range 0.002 – 0.01 exploring more than a million parameter combinations. We obtained a good qualitative convergence of equilibrium concentrations inside this range.

[#]These values were chosen in such a way that the ratio of formation and degradation plus dissociation rates lies in the same range than in (6).

[§]Once the polyprotein has been produced, we assume that either it can be degraded at rate ε_p or it is cleaved with absolute efficiency into the mature peptides, which is equivalent to a cleavage rate = 1, as in the range reported in (6). The exact value employed in our simulations was obtained from the same sensitivity analyses described for ε.

[$]For picorna-like viruses, the number of monomers of structural proteins required for encapsidation is in the range of hundreds to thousands. The value has been fixed to make it similar to the potyviruses (36).

[‡]Virions degradation rate has been fixed two orders of magnitude smaller than the formation rate (k_2) as reported for VSV (14).

Initial conditions: $[0+0] = 0.01$, $[0-0] = [1+0] = [1-0] = T_c(0) = p(0) = V(0) = 0$ if not otherwise specified.