

Simple genomes, complex interactions: Epistasis in RNA virus

Santiago F. Elena, Ricard V. Solé, and Josep Sardanyés

Citation: *Chaos* **20**, 026106 (2010); doi: 10.1063/1.3449300

View online: <http://dx.doi.org/10.1063/1.3449300>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/chaos/20/2?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Noise analysis of genome-scale protein synthesis using a discrete computational model of translation](#)

J. Chem. Phys. **143**, 044109 (2015); 10.1063/1.4926536

[Bidirectional transport of motor-driven cargoes in cell: A random walk with memory](#)

AIP Conf. Proc. **1512**, 140 (2013); 10.1063/1.4790950

[Kinetic models of the interference of gene transcription to ncRNA and mRNA](#)

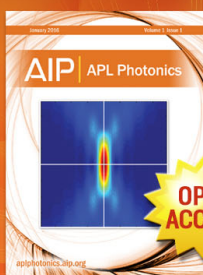
Chaos **21**, 023135 (2011); 10.1063/1.3605464

[Computing the conformational entropy for RNA folds](#)

J. Chem. Phys. **132**, 235104 (2010); 10.1063/1.3447385

[Building Three-Dimensional Ribonucleic Acid Structures](#)

Comput. Sci. Eng. **5**, 44 (2003); 10.1109/MCISE.2003.1225860



Launching in 2016!

The future of applied photonics research is here

OPEN
ACCESS

AIP | **APL
Photonics**

Simple genomes, complex interactions: Epistasis in RNA virus

Santiago F. Elena,^{1,2,a)} Ricard V. Solé,^{2,3,4} and Josep Sardanyés¹

¹Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, Ingeniero Fausto Elio s/n, 46022 València, Spain

²Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA

³ICREA-Complex Systems Lab, Parc de Recerca Biomèdica de Barcelona (PRBB), Universitat Pompeu Fabra, Dr. Aiguader 88, 08003 Barcelona, Spain

⁴Institut de Biologia Evolutiva, Consejo Superior de Investigaciones Científicas-UPF, Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain

(Received 12 February 2010; accepted 19 May 2010; published online 30 June 2010)

Owed to their reduced size and low number of proteins encoded, RNA viruses and other subviral pathogens are often considered as being genetically too simple. However, this structural simplicity also creates the necessity for viral RNA sequences to encode for more than one protein and for proteins to carry out multiple functions, all together resulting in complex patterns of genetic interactions. In this work we will first review the experimental studies revealing that the architecture of viral genomes is dominated by antagonistic interactions among loci. Second, we will also review mathematical models and provide a description of computational tools for the study of RNA virus dynamics and evolution. As an application of these tools, we will finish this review article by analyzing a stochastic bit-string model of *in silico* virus replication. This model analyzes the interplay between epistasis and the mode of replication on determining the population load of deleterious mutations. The model suggests that, for a given mutation rate, the deleterious mutational load is always larger when epistasis is predominantly antagonistic than when synergism is the rule. However, the magnitude of this effect is larger if replication occurs geometrically than if it proceeds linearly. © 2010 American Institute of Physics. [doi:10.1063/1.3449300]

Epistasis occurs when the phenotypic effect of a mutation is conditional to the presence of other mutations in the genome. Despite their structural simplicity and reduced size, the genomes of RNA viruses show complex patterns of epistatic interactions between and within genes. The existence of such complex patterns has profound implications in the evolutionary dynamics of these pathogens. Here we review evidences for epistasis on RNA viruses gathered from experiments and from computational studies. We also provide a description of some theoretical tools that can be used to investigate the dynamics and evolution of RNA viruses. Finally, we analyze an *in silico* bit-string stochastic model of virus replication exploring the effect of replication mode and epistasis on the load of deleterious mutations.

I. INTRODUCTION

Back in the late 1970s, Domingo and co-workers^{1,2} pioneered the studies of the genetic composition and dynamics of genetic variability in RNA virus populations. The theoretical framework they used for understanding their results was Eigen's quasispecies theory.³ Since then, quasispecies theory has become the dominant theoretical framework for explaining virus evolution. Concepts such as master sequence, mutant cloud, sequence space, error threshold and catastrophe, lethal mutagenesis, quasispecies effect, or quasispecies memory have been drawn from Eigen's theory and are now

of common use among virologists. Despite this extensive application, the quasispecies theory was initially developed in the framework of prebiotic evolution, providing models for the earlier evolution of biological information.³ The standard quasispecies models consider large, perfectly mixed, noninteracting populations of error-prone replicator molecules. Roughly, quasispecies are clouds of genotypes that appear in a population at mutation-selection balance. It has also served as starting point for new mathematical developments.⁴ For developing his theory, Eigen made a number of biologically unrealistic assumptions that have been pointed and criticized by evolutionary geneticists,^{5,6} namely, constant environments, large and constant population sizes, all mutations having the same fitness effects, and no interactions between mutations. Therefore, it is important to distinguish between theoretical quasispecies and real viral populations.

During the recent years, one of Eigen's assumptions has received special attention from experimentalists: the independence among sites in viral RNA genomes. Highly compacted genomes, with many examples of multifunctional proteins or overlapping genes, are expected to show strong epistases. Epistatic interactions are important in evolutionary genetics almost whenever multilocus genetics matters and plays a central role in the evolution of genetic systems such as sex and recombination, ploidy, genomic segmentation and modularity, genetic incompatibility and speciation, mechanisms of mutational robustness, the accumulation of deleterious mutations through genetic drift, and the rate of adaptive

^{a)}Electronic mail: sfelena@ibmcp.upv.es.

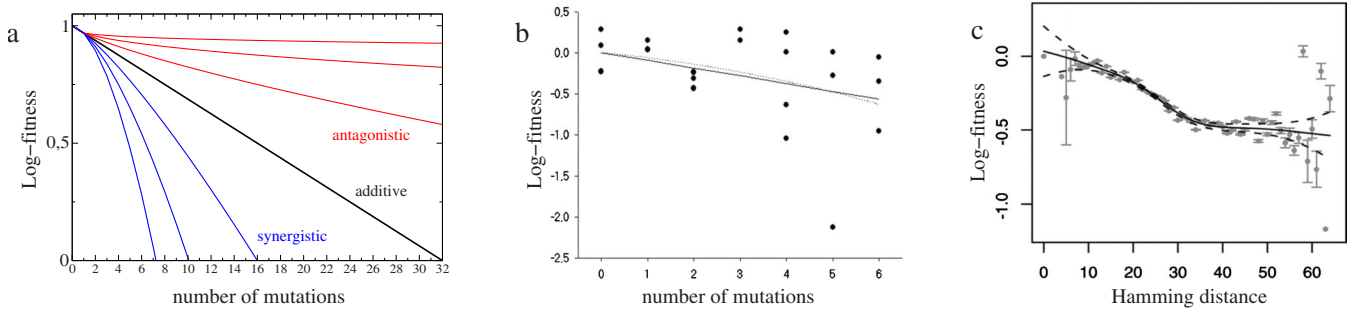


FIG. 1. (Color online) (a) Simple fitness landscapes of the form $f_k = 1 - d_H^{\xi} \nu^{-1}$, ν being the length of the genome, d_H the Hamming distance between the k th sequence and the master one (i.e., the number of mutations), and ξ measuring the sign and strength of epistasis. The landscape assumes that all mutations are deleterious. Red (upper three) lines represent examples of antagonistic epistasis ($\xi < 1$), the black (central) line represents the null model of no epistasis $\xi = 1$, and the blue curves (lower three lines) represent cases of synergistic epistasis ($\xi > 1$). In (b) and (c) we display experimental results showing the decline in viral fitness at increasing number of mutations for FMDV and HIV-1, respectively [see Elena (Ref. 8) and Bonhoeffer *et al.* (Ref. 9) for more details].

evolution (reviewed in Ref. 7). Epistasis occurs when the phenotypic effects of a mutation change depending on the presence or absence of other mutations in the genome. Hereafter, we will focus on the role of epistasis in determining the most important phenotypic trait: fitness. Under the multiplicative model, mutations act independently and the total fitness of a genotype equals the product of the fitness effects of mutations on different loci. For deleterious mutations, independent effects can be expressed as a linear decline of log fitness as the number of mutations increases [Fig. 1(a)]. When mutations do not act independently, the fitness effect of a mutation changes depending on the presence of other mutations in the genome. Epistasis changes the shape of the log-fitness function. Synergistic (negative) epistasis results if deleterious mutations are more harmful together than would be expected from their separate effects and produces accelerated fitness losses with increasing mutational load. However, if log fitness declines slower than predicted by the multiplicative model, then we talk about antagonistic (positive) epistasis (see Fig. 1).

A theoretical approach to the problem of how populations evolve under epistatic interactions requires a number of assumptions. Epistasis represents one of the most relevant problems in understanding the nature and consequences of complex interactions within biological complex systems. The impact of different genes in affecting the fitness of others, but in general in modifying their behavior, has been at the center of early research in genetic networks^{10,11} as well as in recent *in silico* studies seeking to understand the evolution of genomic complexity and mutational robustness.^{12,13} Epistasis represents a problem of high dimensionality. For a genome ν nucleotides long, the total number of possible interactions is $2^{\nu} - \nu - 1$, which for viral genomes of several thousands nucleotides tends to infinity. A good first approximation, empirically convenient, is to focus only on the $\nu(\nu-1)/2$ possible pairwise interactions. Obviously, variance must exist in the sign and strength of epistasis within a genome (i.e., each particular pair of mutations interacts on a different way), but what we are interested in this review is whether the architecture of viral genomes favors a particular type of epistasis over the others.

In this article, we will review current experimental evidences exploring the existence of epistasis in RNA genomes.

We will also present some useful tools for the theoretical and computational analyses of RNA virus dynamics and evolution. Then we will move to the analysis of an *in silico* model which simulates RNA virus replication. With this model we will analyze different modes of replication, paying special attention to the interplay between epistasis and replication strategy on the accumulation of mutant genomes and population fitness and robustness.

II. THE EXISTENCE OF EPISTASIS IN RNA VIRUS GENOMES

A. Preliminary indirect observations

A pervasive observation has been that after treatment with antiviral drugs, pre-existing mutants able of successfully replicating in the presence of the drug rise in frequency and get fixed in the population. When these escape mutants are characterized, they usually contain several mutations. Indeed, studies with *human immunodeficiency virus* type 1 (HIV-1) showed that mutations conferring resistance to protease inhibitors appeared sequentially and in a defined order.¹⁴⁻¹⁶ Moreover, the analyses of such combinations showed that the first mutation(s) fixed confer(s) the resistance phenotype whereas mutations fixed later are usually compensating for possible side effects of the resistance mutations.¹⁷ When recreated in the absence of the resistance mutation, the fitness effect of the compensatory ones was usually deleterious. This context-dependent effect is a fundamental characteristic of epistasis. However, since compensatory mutations imply a particular type of epistatic interaction whose effect is observed after the action of selection, they provide limited information to our question of whether the architecture of RNA genomes favors a particular type of epistasis.

Another source of evidence supporting the existence of epistasis among loci in RNA genomes comes from the widespread evidence of genetic convergence in experimental evolution studies. Whenever independent lineages were evolved in the same constant environment, the resulting populations tended to fix similar sets of mutations.¹⁸⁻²¹ These observations suggest that selection may not act upon single nucleotide changes but on covariation groups. The effect of each mutation in the group is beneficial only in the genetic context

of the other mutations. By themselves, their effects may not necessarily be beneficial. As it happened with compensatory mutations, the existence of covariation groups selected during adaptation is not informing about overall tendencies in the sign and strength of epistasis.

Phylogenomics also provides a valuable source of information about epistasis in viral genomes. For example, compensatory changes could occur in viral multifunctional proteins. If a mutation alters favorably one function but also compromises the second function, then a compensatory mutation may be required to restore overall fitness. Subsequently, that pair of mutations may increase in frequency in the population and be detected in a large sample of sequences. By contrast, synergistic epistasis among mutations generates unfit individuals and will therefore be undetectable through the analysis of gene sequence data. Shapiro *et al.*²² evaluated the likelihood of co-occurring pairs of mutations along the phylogenies of 177 RNA virus genes. These analyses revealed widespread evidence for antagonistic epistatic interactions both at synonymous and nonsynonymous sites and for both clonal and recombining viruses.

B. First direct tests: Mutation-accumulation experiments

The first systematic exploration of epistasis in RNA viruses was performed by Elena.⁸ This author analyzed the changes in log fitness associated with the accumulation of deleterious mutations during bottleneck transmissions of *foot-and-mouth disease virus* (FMDV) and the associated molecular changes.²³ The analyses concluded that the decline in log fitness was compatible with a multiplicative model [Fig. 1(b)]. Similar type of data can be found in the extensive literature on the design of live attenuated viruses for vaccination. Rational vaccine design depends upon the measurement of individual and combined effects of deleterious mutations and, therefore, these data are well-suited for seeking epistasis. Burch *et al.*²⁴ undertook the task of analyzing this literature and gathered information from 29 studies using 14 different RNA viruses. They found no consistent tendency toward a particular form of epistasis. However, the statistical power associated with the small sample sizes of these two studies does not allow detecting weak epistasis.

A mutation-accumulation (MA) approach was also used by Burch and Chao²⁵ to study epistasis in the bacteriophage $\phi 6$. Assessing the average sign of epistasis from MA data requires determining the average fitness effect for a set of starting genotypes that already differ in fitness. Burch and Chao²⁵ started their study using a collection of genotypes that differed in an unknown number of loci resulting in measurable differences in fitness. These lineages were used in a MA experiment consisting of a single bottleneck transfer. After this additional transfer, the fitness of a large number of clones was determined per each genotype. Given the high mutation rates of RNA viruses, clones within each genotype are expected to differ in a certain number of mutations and thus in fitness. If mutations fixed during the bottleneck passage act additively on their genetic background, then the regression between the estimated average mutational effects (across clones) and the fitness of the starting genotype would

have slope one. By contrast, epistasis would produce a deviation from this expectation. Burch and Chao²⁵ determined that the slope of the regression was significantly greater than 1.0 and concluded that, on average, epistasis in $\phi 6$ was antagonistic.

In a recent article, de la Iglesia and Elena²⁶ reported results from a MA experiment with *tobacco etch virus* (TEV). Twenty independent MA lineages were transferred through consecutive bottlenecks of size one performed by isolating lesions (equivalent to plaques in an agar plate) in the host *Chenopodium quinoa*. The pattern of log fitness decline with the number of bottleneck transfers was consistent with antagonistic epistasis [Fig. 1(a)]. Furthermore, a negative correlation was observed between the strength of epistasis and the severity of mutational effects: strongly deleterious mutations were involved in more antagonistic interaction whereas milder mutations were involved in weaker antagonistic or even synergistic interactions.

C. Systematic genome-wide screenings

All attempts to quantify epistasis on viral RNA genomes described in Secs. II A and II B relied on statistical inferences from indirect phenotypic estimates. These indirect approaches are all imperfect. For example, MA experiments suffer from at least three limitations: (i) a lack of knowledge of the number of mutations accumulated, (ii) the fact that a lack of deviations from a linear fitness decline may indicate the absence of epistasis or widespread epistasis in which antagonistic and synergistic interactions are equally common, and (iii) selection may be operating within growing plaques thus biasing the sampling against very low fitness genotypes (not to say that lethal mutations will be completely missed, although lethal mutations are irrelevant for epistasis). Phylogenomic approaches also suffer from at least a major limitation: the set of mutations observed is not random but those that have been fixed by selection. A straightforward approach to avoid these problems relies on constructing genotypes containing exactly the same mutations alone and in combination, then measuring their individual and combined fitness effects and, finally, comparing the results with predictions generated under the null hypothesis of multiplicative fitness effects.

Sanjuán *et al.*²⁷ took for the first time this direct approach to explore the distribution of epistasis in the genome of *vesicular stomatitis virus* (VSV). These authors created a collection of pairs of single-nucleotide substitution mutants and quantified their fitness effects. Then, they randomly drew pairs of mutations from the collection and created the corresponding double mutants. The first significant observation was that variance exists in the sign and strength of epistasis, with many pairs being synergistic, others acting multiplicatively and many acting antagonistically. Despite this heterogeneity, the average epistasis value was significantly antagonistic for pairs of deleterious mutations but also for pairs of beneficial mutations. A second interesting finding was that epistases were equally like within and between genes. A third interesting result was that in a few instances, combining a pair of viable mutations created nonviable genotype, an extreme case of synergism known as *synthetic lethality*. Finally,

when two beneficial mutations were combined, the fitness of the double mutant was generally lower than for each single mutant, a situation known as *decompensatory epistasis*.

Using both MA data and estimates from pairs of known-effect mutations, Bonhoeffer *et al.*⁹ looked for epistasis in HIV-1 genome [Fig. 1(c)]. In good agreement with the VSV data, these authors reported significant variance in epistasis but with the average value being significantly antagonistic. However, in the recent years two studies have reported contradicting results for HIV-1. First, van Opijnen *et al.*²⁸ found that pairs of mutations affecting the transcriptional promoter of HIV-1 interacted in a multiplicative way. The results of Parera *et al.*²⁹ were closer to the latter. In this study, 114 pairs of mutations were created within the HIV-1 protease gene. The average value was multiplicative, although significant variance in the sign of epistasis was found and synergistic pairs were more frequent than antagonistic ones. It is worth mentioning that the set of 29-point mutations included in the study was not randomly chosen but all represent deleterious but viable cases. Interestingly, a large fraction (~40%) of pairs created synthetic lethals. Three possible explanations may account for the difference between these two latter studies and the first one. First, mutations in a regulatory region are expected to affect the levels of expression of multiple genes at the same time and thus they may not be a representative case of random mutations but very pleiotropic ones. Second, the sample size in the latter experiment was also rather limited (eight pairs). Third, as pointed out by Wang *et al.*,³⁰ the results of Bonhoeffer *et al.* may be biased because of the inability of their method to recover lethal mutations.

Sanjuán³¹ has explored the existence of epistasis among mutations in an RNA structure essential for initiation of the reverse transcription, RNA packaging, and virion formation of *Rous sarcoma virus* (RSV). Therefore, mutations affecting the folding may exert pleiotropic effects on several viral functions. Polymerase chain reaction (PCR)-based mutagenesis followed by artificial selection was used to create collections of genotypes that contained from one to nine mutations in the stem-loop structure. In good agreement with VSV and first HIV-1 results, average epistasis was significantly antagonistic.

D. Epistasis in the simplest replicons: The viroids

Viroids represent a class of subviral plant pathogens whose genome is constituted by a small (246–401 nucleotides long) single-stranded circular and covalently closed RNA molecule with a high degree of self-complementation, resulting in a compact folding. Indeed, RNA folding represents, together with the symptoms induced in susceptible hosts, the only identifiable phenotype of most viroids. Based on phylogenetic studies and shared structural and phenomenological properties, viroids are classified into two families. The *Pospiviroidae* replicate in the nucleus and fold into a rodlike structure of minimal free energy with five structural domains. The *Avsunviroidae* replicate in the chloroplast, fold into a less organized structure, and contain autocatalytic ribozymes. Phylogenetic covariations as well as *in vivo* cross-

linking studies support the validity of the predicted secondary structures. Sanjuán *et al.*³² undertook the task of estimating computationally the average sign and strength of epistasis for the 29 known viroid species. To do so, they tested the effect of a large number of pairs of random mutations in the RNA folding. Fitness was evaluated as the difference between the minimum free energy folding obtained for the mutant viroid and that obtained for the wild-type sequence. Three major findings were reported. First, in all but one case, epistasis was predominantly antagonistic. Second, the strength of antagonistic epistasis decreased with genome complexity. Third, as it was observed in the TEV experiments described above, the intensity of epistasis was negatively correlated with the severity of mutational effects: viroid species with large average mutational effects showed stronger antagonistic epistasis whereas viroids with mild average mutational effects showed weaker antagonisms.

III. EVOLUTIONARY CONSEQUENCES OF EPISTASIS

The above review supports the view that the type of epistasis in RNA viral genomes is predominantly antagonistic. This finding has some important evolutionary consequences. The first one has to do with the rate of adaptation. In a recent study, Sanjuán *et al.*³³ explored the effect of different types of epistasis in the rate of compensatory evolution in VSV. As expected, the largest fitness improvements were associated to pairs of mutations interacting synergistically. By contrast, the slowest rates were associated to antagonistic pairs. The reason for this is obvious: if mutations interact synergistically, by compensating the first mutation the gain in fitness obtained is larger than expected if mutations were additive. By contrast, if mutations interact antagonistically, by compensating the first mutation the magnitude of the fitness gain is lower than expected because the pernicious effect of the second mutation is fully expressed.

The second important evolutionary implication of antagonistic epistasis has to do with the existence of sex and recombination in some RNA virus species. According to the Fisher–Muller theory, the advantage of sex and recombination relies in combining beneficial mutations into the same genome and thus speeding up the rate of adaptation. However, antagonistic epistasis among beneficial mutations (in particular in the form of decompensatory epistasis) would not necessarily imply a benefit in terms of adaptive evolution. Therefore, antagonistic epistases impose a constraint to the rate of virus evolution. Sex may still be beneficial for RNA viruses as a mechanism to purge deleterious mutations from populations. However, according to the mutational deterministic hypothesis,³⁴ an excess of synergistic interactions is required to compensate for the twofold cost of sexual reproduction compared with clonality. In this context, Azevedo *et al.*³⁵ explored an artificial gene network model, showing that synergistic epistasis evolved as a by-product of selection for genetic robustness because of recombination, whereas antagonistic epistasis evolved under asexual reproduction. Moreover, other works combining recombination and gene-regulatory networks showed that when epistasis and reproductive strategy are allowed to coevolve, asexual reproduction outcompetes sexual reproduction.³⁶ Furthermore, the

existence of variability among loci in the sign and strength of epistasis, and specially the abundance of antagonistic epistasis, decreases the parameter space over which sex may evolve.³⁷ All together, this may explain why VSV has evolved as a strictly asexual organism. However, these considerations pose a serious problem for highly recombining viruses such as HIV-1, RSV, TEV, and $\phi 6$. Therefore, why RNA viruses have different forms of sex remains unexplained.

Another quite common observation is that the value of average epistasis negatively correlates with the average effect of deleterious mutations. This correlation seems to be a ubiquitous phenomenon that has also been observed in other systems of simple replicons: digital organisms,^{38,39} *in silico* RNA folding,^{39,40} and *in silico* models of bacteriophage T7 life cycle,⁴¹ suggesting that it may represent a general property of genetic systems. As pointed out by Wilke *et al.*,⁴⁰ this correlation is consistent with the hypothesis that epistasis depends on the location of the reference genotypes relative to other high-fitness genotypes. However, the existence of this correlation does not explain the overall preponderance of antagonistic epistasis although it has an important implication: it means that both parameters cannot be evolutionary optimized independently. If selection has to favor a reduction in the effect of mutations, i.e., an increase in individual robustness, then the strength of antagonistic epistasis needs to be relaxed. Therefore, dominance of antagonistic epistasis on RNA viruses thus points to a lack of mechanisms of individual robustness, an obvious conclusion given their lack of genetic and functional redundancy.

The parasitic runaway lifestyle of RNA viruses, by favoring fast replication and genomic compactation, creates antagonistic epistasis and strong sensitivity to deleterious mutations (reviewed in Ref. 42). However, individual hypersensitivity to mutations creates an unexpected outcome: mutant genomes of very low fitness are removed from the population in a very efficient way. The efficiency of natural selection to purge deleterious mutations from a population depends on the product between population size and the average effect of mutations on fitness. RNA viruses usually reach very large population sizes even within a single infected host. Therefore, the combination of large populations with strong deleterious mutational effects makes selection extremely efficient. If deleterious alleles are efficiently removed, the frequency of mutation-free genomes is high and hence the average fitness remains close to the optimal value.

Another consequence of genome compactation and parasitic lifestyle is that viruses have opted for low fidelity replication strategies. A large mutation rate creates a different selective pressure that may push populations toward regions of genotypic space where the density of neutral mutations is higher.^{43,44} As a consequence, both individual genomes and the whole population may gain an enhanced robustness against deleterious mutations. Notice that robustness in this case arises from a completely different mechanism (and somehow opposed) than the one described in the previous paragraph, based on individual hypersensitivity to mutations. The existence of different robustness mechanisms in RNA virus populations has been recently postulated.⁴² In addition

to the population robustness just mentioned, high multiplicity of infection, sex and recombination, the use of cellular chaperons, or the election of a stamping-machine replication mode⁴⁸ may also contribute to increase robustness.

IV. USING DIGITAL GENOMES TO MODEL VIRUS DYNAMICS

A useful mathematical and computational modeling of RNA viruses requires considering a number of simplifications involving the nature of genomes, their internal organization, and the nature of RNA viruses themselves. Typically, all these approaches sacrifice some realism to the benefit of a clear understanding of key mechanisms. In particular, the complex structure of the virus environment (its host organism) and the fact that viral particles are not just strings of nucleotides are simply neglected. Moreover, the details of the virus-host interactions and the molecular peculiarities of the viral life cycle are not taken into account. And yet, even under these simplistic assumptions, relevant properties of virus dynamics can be captured. In this section, we summarize some standard models of RNA viruses that make use of a digital genome description, their predictions, and how epistatic interactions can be taken into account.

Mathematical and computational models of RNA virus dynamics have considered several levels of dynamical complexity, mainly at the population level.⁴⁵ These approaches make use of a diverse array of approximations, from ordinary differential equations^{46–48} to stochastic spatial dynamics.^{49,50} One of the most successful theoretical approaches to the modeling of RNA viruses is given by digital genomes. In this approach, a mapping between RNA sequence, defined as a chain of nucleotides involving a four-letter alphabet Ω , and a binary sequence is obtained,

$$\mathcal{F}: \Omega = \{U, G, A, C\} \rightarrow \Sigma = \{0, 1\},$$

or, alternatively, to a Boolean representation where we use spins instead of bits,

$$\mathcal{F}_s: \Omega = \{U, G, A, C\} \rightarrow \Sigma = \{+1, -1\}.$$

Both approaches can be shown to be equivalent^{51,52} (since the mapping has the same nature) but the second exploits some advantages of considering “up” and “down” spins in order to describe the microscopic dynamics.

The string $S_i = (S_{i1}, \dots, S_{i\nu})$ of length ν describes a digital genome, i.e., a sequence of purines and pyrimidines that only incorporate the linear information encoded by the string. In order to introduce further information concerning the functional relevance of this sequence, we need to introduce an additional mapping, namely, the sequence-fitness measure (see Fig. 2 left) defined as

$$f: \Sigma^\nu \rightarrow U \subset \mathcal{R},$$

which maps the string into a scalar value, i.e., $S_i \in \Sigma^\nu \rightarrow f(S_i)$. The nature of this mapping is the central problem considered here. If this functional relation is such that bits are essentially independent from each other, no epistatic components will be at play. Instead, if the function f is such

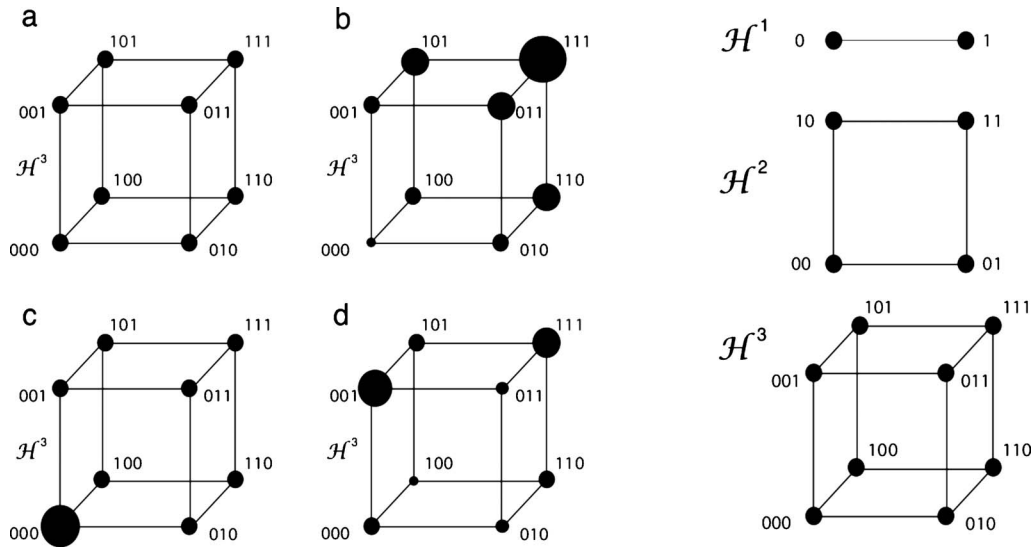


FIG. 2. (Left) The fitness $f(S)$ associated to each vertex of the sequence space for digital genomes allows defining the trajectories followed through string evolution. Here, using the simple 3-dimensional case, \mathcal{H}^3 , four standard cases are presented. In this plot, the size of the nodes is proportional to their fitness value. Here we have (a) flat, (b) Fujiyama, (c) Sweetina–Schuster (sharp peak), and (d) rugged landscapes, respectively. In the flat landscape all the strings share the same fitness. In the Fujiyama landscape the fitness decreases linearly and proportionally to the number of different bits from the master sequence (string 111 in this example). In the Sweetina–Schuster landscape the population is divided in two fitness classes, the master sequence [represented with the 000 string in (c)], which has the highest fitness, and the pool of mutants that have the same lower fitness. Finally, in the rugged landscape, each string has a different fitness. (Right) The first three (from top to bottom) small landscapes associated to one, two, and three bits, respectively. As the number of bits grows, the size of the hypercube increases exponentially. Each landscape \mathcal{H}^ν involves 2^ν sequences.

that different parts of the system influence others in some nontrivial ways, then epistasis will be at play.

The digital genome metaphor allows us using an abstract, multidimensional representation of the potential set of states accessible to a ν -bit digital genome. This hypercube $\mathcal{H}^\nu = \Sigma^\nu$ provides, at low dimensionality (see Fig. 2 right), intuitions concerning the behavior of strings under the action of mutation and selection. If only small mutation rates $\mu \ll 1$ are considered, transitions between sequences will take place involving only nearest neighbors in sequence space, thus differing in only one bit. In general, for a given mutation rate, two sequences \mathbf{S} and \mathbf{S}' will be obtained from each other with a given probability,

$$W_\mu(\mathbf{S} \rightarrow \mathbf{S}') = \mu^{d_H(\mathbf{S}, \mathbf{S}')} (1 - \mu)^{\nu - d_H(\mathbf{S}, \mathbf{S}')},$$

where $d_H(\mathbf{S}, \mathbf{S}')$ is the Hamming distance between the two sequences, i.e., the number of different bits, given by

$$d_H(\mathbf{S}, \mathbf{S}') = \sum_{i=1}^{\nu} (1 - \delta_{S_i, S'_i}),$$

where $\delta_{i,j}$ is the Kronecker delta with $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$. Alternatively, we can also use

$$d_H(\mathbf{S}, \mathbf{S}') = \frac{1}{2} \left(\nu - \sum_{i=1}^{\nu} S_i S'_i \right).$$

The meaning of W_μ is very easy to interpret in probabilistic terms: it is just the probability of having exactly d_H differences between the two digital genomes. This function allows introducing the dynamics associated to mutations as transition probabilities. For the spin mapping, the transition probabilities can be rewritten as follows:

$$W_\mu(\mathbf{S} \rightarrow \mathbf{S}') = \exp[d_H \log \mu + (\nu - d_H) \log(1 - \mu)],$$

which gives, after some algebra,

$$W_\mu(\mathbf{S} \rightarrow \mathbf{S}') = e^{\nu \log(1 - \mu)} \times \exp \left[\frac{1}{2} \left(\nu - \sum_{i=1}^{\nu} S_i S'_i \right) \log \left(\frac{\mu}{1 - \mu} \right) \right],$$

or, in a compact form,

$$W_\mu(\mathbf{S} \rightarrow \mathbf{S}') = \mathcal{N} \exp \left(-\beta \sum_{i=1}^{\nu} S_i S'_i \right),$$

where the β term, defined as $\beta = \log(\mu/(1 - \mu))/2$, can be interpreted in terms of a temperature⁵⁵ and \mathcal{N} is a normalization constant. Actually, the product of binary units on the sum of the last expression is very close to the energy functions associated to spin interaction in magnetic systems. As it occurs with them, phase transitions can arise in RNA virus dynamics.

Once the hypercube and the fitness function associated to each potential vertex are defined, we need to describe the dynamics. If $N(\mathbf{S}, t)$ indicates the fraction of strings having a given sequence $\mathbf{S} \in \Sigma^\nu$, Eigen’s formulation of the population dynamics would be described as a set of nonlinear differential equations,

$$\frac{dN(\mathbf{S}, t)}{dt} = \sum_{\mathbf{S}'} W_\mu(\mathbf{S}' \rightarrow \mathbf{S}) f(\mathbf{S}') N(\mathbf{S}', t) - \left(\sum_{\mathbf{S}'} f(\mathbf{S}') N(\mathbf{S}', t) \right) N(\mathbf{S}, t). \quad (1)$$

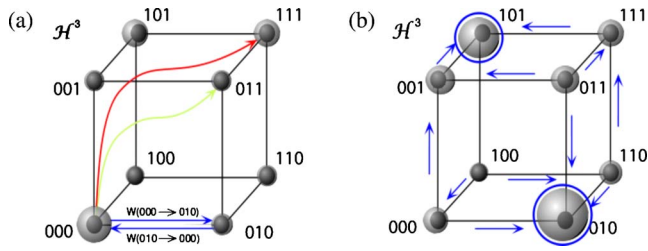


FIG. 3. (Color online) (a) Transitions between different sequences in a 3D digital genome model. Once strings start replicating and mutating, the populations will spread through sequence space. At small mutation rates, the typical transition (lower blue arrows) will involve nearest neighbors in the hypercube, although transitions involving changes in more than one bit are also possible (upper arrows). (b) In the NK model, epistatic interactions among loci within the sequence introduce ruggedness in the landscape structure. Such ruggedness, controlled by the value of K , is visible in the presence of multiple local maxima in the fitness landscape. Depending on where a population starts, it will end up in different local peaks. Here two peaks exist, and mutations involving one-bit changes create a set of flows that reach each local optimum after a given number of adaptive walks.

The first part in the right-hand side of Eq. (1) corresponds to positive contributions to the abundance of \mathbf{S} due to transitions from other strings to the one considered here. The second term includes all the reverse events leaving the node occupied by \mathbf{S} . As we can see we consider both the explicit fitness function and the transition probabilities from each pair of sequences. In Fig. 3(a) we illustrate the information described in the previous equation using a $\nu=3$ system. The spheres located at each node indicate the population size and the three potential transitions from 000 to other strings differing in one, two, or three bits.

Alternatively, a discrete-time formulation is sometimes more convenient, and in this case the set of equations would be given by

$$N(\mathbf{S}, t + 1) = \frac{\sum_{\mathbf{S}'} W_{\mu}(\mathbf{S}' \rightarrow \mathbf{S}) f(\mathbf{S}') N(\mathbf{S}', t)}{\sum_{\mathbf{S}'} f(\mathbf{S}') N(\mathbf{S}', t)}. \quad (2)$$

Simple fitness landscapes can be defined from our previous definitions. One of the simplest cases involves a single-peak organization. Specifically, the fitness function is described by

$$f(\mathbf{S}) = f_0 \delta_{S, S_m} + (1 - \delta_{S, S_m}),$$

with $f_0 > 0$ and S_m being a given string, hereafter the master sequence, usually taken as $S_{mi}=1$ for all $i=1, \dots, \nu$. This landscape, known as the Swetina-Schuster fitness landscape,⁵³ describes a system where one sequence has a high fitness and all others share the same lower fitness [see Fig. 2(c) left]. Despite its obvious simple nature, this model displays a very interesting phenomenon, the so-called error catastrophe: there is a critical mutation rate beyond which information and Darwinian selection cannot be preserved and random drift arises. While using this model we will consider that the population size $N = \sum_{\mathbf{S}} N(\mathbf{S})$ is very large and that mutation rates are small. Using the single peak landscape as defined above, we can use the discrete dynamical equation for the quasispecies dynamics in order to obtain its equilibrium distribution. From Eq. (2) we have

$$\begin{aligned} N(\mathbf{S}_m, t + 1) &= \frac{\sum_{\mathbf{S}'} W_{\mu}(\mathbf{S}' \rightarrow \mathbf{S}_m) [f_0 \delta_{S', S_m} + (1 - \delta_{S', S_m})] N(\mathbf{S}', t)}{\sum_{\mathbf{S}'} [f_0 \delta_{S', S_m} + (1 - \delta_{S', S_m})] N(\mathbf{S}', t)} \\ &= \frac{(1 - \mu)^{\nu} f_0 N(\mathbf{S}_m, t) + \sum_{\mathbf{S}' \neq \mathbf{S}_m} W_{\mu}(\mathbf{S}' \rightarrow \mathbf{S}_m) N(\mathbf{S}', t)}{f_0 N(\mathbf{S}_m, t) + \sum_{\mathbf{S}' \neq \mathbf{S}_m} N(\mathbf{S}', t)} \\ &\approx \frac{f_0 N(\mathbf{S}_m, t)}{f_0 N(\mathbf{S}_m, t) + (1 - N(\mathbf{S}_m, t))}. \end{aligned}$$

The last equation has been obtained by assuming that backward mutations can be neglected. After some algebra, and using $(1 - \mu)^{\nu} \approx \exp(-\mu\nu)$, it can be shown that the stationary distribution of the master sequence [obtained when $N(\mathbf{S}_m, t + 1) = N(\mathbf{S}_m, t) \equiv N^*(S_m)$] is given by

$$N^*(S_m) \approx \frac{f_0 e^{-\mu\nu} - 1}{f_0 - 1}.$$

This condition allows us to calculate the critical mutation rate μ_c defining the error threshold beyond which the master sequence is no longer conserved. From the critical condition $N^*(S_m) = 0$, we can see that master sequences will be observed provided that $\mu\nu \leq \ln f_0$, which gives the threshold value,

$$\mu_c = \frac{\ln f_0}{\nu},$$

thus predicting that the maximum allowed mutation rate for a replicator displaying mutations scales inversely with genome length. If we increase mutation rates beyond this value, the population will no longer be able to preserve itself at the peak in this simple landscape.

The sharp peak landscape defines an extreme in a hierarchy of models introducing different levels of dependencies among genes. A different approximation deals with landscapes in a much more general way by allowing them to display a number of local maxima. The best known of these models is Kauffman's NK model defined on a hypercube (as above). The two main parameters of the NK model are the number of loci N and the average number of other loci K which epistatically influence the fitness contribution of each locus.⁶⁶ It was originally proposed as a representation of haploid genomes involving two alleles per locus with additive contributions to fitness from different loci. Once again a fitness function is introduced, $f = f(S_{i1}, \dots, S_{i\nu})$, and changes in the traits are assumed to occur by means of single, one-bit steps. These single-chain events are consistent with our assumption of small mutation rates. In this way, a given string obtained by inaccurate replication allows to perform a random *adaptive walk* from a given node toward one of its ν nearest neighbors if this leads to an increase in fitness. A direct consequence of this process is that once a local peak is reached, no further changes are allowed to occur. This is clearly different from the assumptions made above, which assume the presence of a preferred sequence around which other sequences have a lower fitness value. In the context of NK landscapes, a local peak is very simply defined: if all nearest neighbors in the hypercube are less fit, we have a

local maximum in fitness. The basic idea is illustrated in Fig. 3(b). How can we construct a system displaying a NK landscape? Kauffman suggested a simple approach using fitness tables: for each element S_{ij} if it is influenced by K other elements. Each element contributes in an additive way to the overall fitness. In other words, if we consider the two-locus model and assume that a given locus i contributes to the global fitness associated to \mathbf{S} by an amount $f_i(\mathbf{S}) \in [0, 1]$, the global fitness is given by the average value,

$$f(\mathbf{S}) = \frac{1}{\nu} \sum_{i=1}^{\nu} f_i(\mathbf{S}).$$

As K grows, the ruggedness of the landscape increases, since more complex interactions are allowed to occur.

An interesting feature of the NK model is that, because of its simplicity, it allows predicting some expected patterns of evolutionary dynamics. As an example, let us consider that fitness values are random and uncorrelated, i.e., if $f(S_{i1}, \dots, S_{i\nu}) = \eta$ where $\eta \in [0, 1]$ is a random number with uniform distribution. This random fitness landscape has many local fitness peaks. This number M_L , which is very large, is given by

$$M_L(\nu) = \frac{2^\nu}{\nu + 1}, \quad (3)$$

and implies that our digital viruses can get trapped into a very large number of optima. Note that Eq. (3) arises within the particular and simplistic context in which genotypic fitness values are independent and identical distributed. To see this, let us consider the number of neighbors of a given node and compute the probability that this node is a local maximum. The chance that it is the fittest among its ν neighbors and itself, given the random choice of values, is simply $P_1 = 1/(\nu + 1)$. Since there are 2^ν possible strings, the fraction of those who are local maxima is $M_L(\nu) = 2^\nu P_1$.

The presence of epistatic interactions can be introduced in a general form by the following functional form for the sequence-dependent fitness:

$$f(\mathbf{S}) = q(1 - s)^{d_H(\mathbf{S}, \mathbf{S}_m)^\xi} + 1 - q.$$

Here the Hamming distance introduces a measure of how far we are from the master sequence having the highest replication rate. In order to illustrate the effect of varying ξ , let us consider the special case $s = q = 1/2$. For these parameters, the fitness function reads

$$f(\mathbf{S}) = \frac{1}{2} \left[\left(\frac{1}{2} \right)^{d_H^\xi} + 1 \right].$$

If we plot this for different values of ξ , we can appreciate the impact of epistasis on the shape of the fitness curve. The three scenarios are present in Fig. 1(a). As we can see, the fitness decay that takes place as we move away from the master sequence is different by depending on the degree of epistasis. These involve antagonistic, null, and synergistic epistasis.

The impact of ξ can be easily determined using a Taylor expansion of $f(\mathbf{S})$. It is easy to show that three curves having

the same q and s values intersect at $d_H = 1$. At this point, they decay differently. Close to this point we have

$$f(d_H, \xi) = f(1, \xi) + \left(\frac{\partial f}{\partial d_H} \right)_{d_H=1} (1 - d_H).$$

This gives a linear relation $f(d_H, \xi) = G(\xi)(1 - d_H)$ with $G(\xi) = 3(1 + \xi \log(1/2))/4$. The higher the epistasis parameter ξ the larger the local slope and the faster the fitness decay, whereas for $\xi < 1/2$ the opposite effect will be at work.

Following the previous formulation, it can be shown that, for large population size and small mutation rates, for $\xi = 1$ we obtain (as expected) the previous result of the error threshold under sharp peak. However, as we tune the degree of epistasis, substantial changes occur. For $q = 1$ and $\xi < 1$ the critical mutation rate is given by

$$\mu_c = \lambda \nu^{\xi-1}.$$

These results are only examples of the potential formalism applied to digital genomes.⁵⁴ They have been useful mainly in the theoretical arena, although relevant results have been obtained when compared to experimental data.^{55,67}

V. A CASE STUDY: REPLICATION MODE, EPISTASIS, AND DELETERIOUS MUTATIONS *IN SILICO*

In this section we aim to illustrate the use of bit-string models for exploring the effect of different parameters on virus population dynamics. We extend recent results obtained by Sardanyés *et al.*⁴⁸ analyzing the effect of the mode of replication together with different fitness landscapes on the mutational load and robustness of positive-sense RNA viruses. Different viruses may use alternative replication strategies. For instance the so-called linear or stamping machine replication (SMR) implies that the progeny of strands during viral replication is mainly produced using as template the antigenomic strand produced from the genomic one first entering the cell. According to this model of replication, the number of mutations in the viral genomes per infected cell follows a Poisson distribution. Such a mutant distribution was found for bacteriophage $\phi X174$.⁵⁶ On the opposite side, for the geometric mode of replication (GR) every synthesized strand can be used as template for new rounds of replication, which generate mutants that continue further replication. Henceforth, the number of mutations per genome and infected cell should follow a more complex distribution known as Luria–Delbrück. For instance, the distribution of mutants for bacteriophage T2 failed to fit a Poisson distribution and thus is thought to mostly replicate according to the GR mode.⁵⁷ We note that modes of replication between the SMR and GR extremes may also exist, as was shown for bacteriophage $\phi 6$, whose mutant spectrum slightly differed from the Poisson distribution.⁵⁸

Here we will specifically use the stochastic bit-string model developed in Ref. 48. We define a population of N strings of length ν . Such strings can be genomic (\mathbf{S}_k^+) or antigenomic (\mathbf{S}_k^-) templates with $\mathbf{S}_k^\pm = (S_{k1}^\pm, S_{k2}^\pm, \dots, S_{k\nu}^\pm)$ and $S_{ki}^\pm \in \{0, 1\}$. The strings of each polarity represent a vertex S_k^\pm which belongs to the sequence space, \mathcal{H}^ν (there will exist two sequence spaces, the genomic and the antigenomic ones,

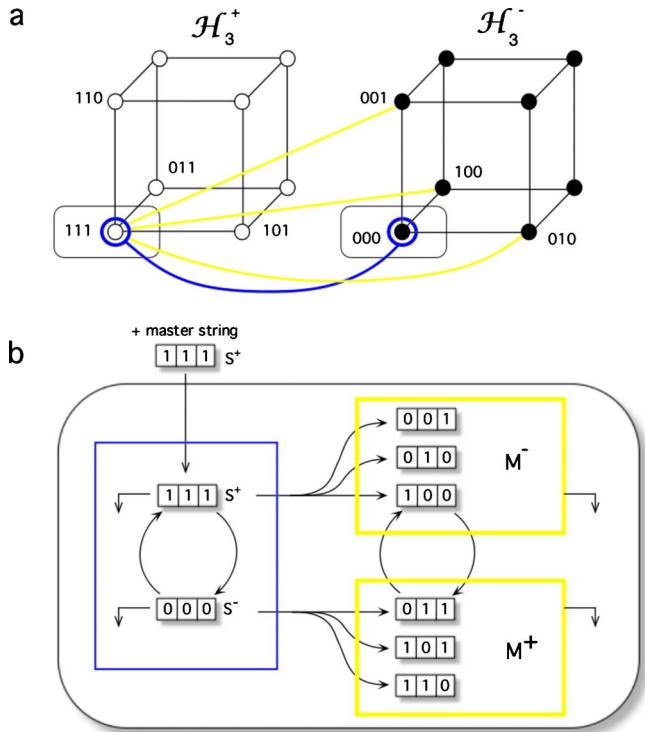


FIG. 4. (Color online) (a) Sequence spaces for a population of genomic (left) and antigenomic (right) binary strands exemplified for strings of length ($L=3$) bits. The consideration of genomic and antigenomic senses can be interpreted as two coupled hypercubes. The population located at each node of the hypercube is generated by error-free replication of the complementary strand, S^\pm (single lower blue line), or by a mutant one, M^\pm (three yellow lines), in the coupled sequence spaces. (b) Schematic diagram of replication, mutation, and degradation rules implemented for the simulation model studied in Sec. V [see Sardanyés *et al.* (Ref. 48) for further details].

coupled by mutation) [see Fig. 4(a)]. In our model, the all-1 and all-0 strings define the genomic and antigenomic master sequences, respectively. The algorithm starts inoculating the system with $N(0)$ genomic master sequences, which can replicate and mutate. It is worth mentioning that back mutations are allowed in the bit-string approach, which can result in fitness recoveries. The fitness for a given k -string of (+) or (-) polarity is represented by its replication rate Γ^\pm with

$$\Gamma^\pm = 1 - d_m^\xi \nu^{-1},$$

$d_m = \sum_{i=1}^L |S_{ki}^\pm - S_{mi}^\pm|$ being the Hamming distance between S_k^\pm and the master sequence S_m^\pm . The parameter ξ introduces the sign and strength of epistasis [see Fig. 1(a)]. The simulation algorithm is a standard Monte Carlo method. At each time generation we apply N times the rules of replication and degradation to ensure that, on average, all strings in the population are updated once per generation. For SMR, the replication state-transition rule is as follows: when a positive-sense string replicates, it generates a negative one that undergoes further replication. However, when a negative-sense string replicates, the synthesized positive-sense string will become a replicator with probability σ . By keeping $\sigma \ll 1$, the negative strings will be mainly used as templates while genomic templates will not participate in the production of complementary strings. On the contrary, for GR all strings will replicate proportionally to its fitness in-

dependent of their polarity [see Material and Methods in Ref. 48 and Fig. 4(b)]. In our simulations we have used $N=1000$, $N(0)=50$, and $\nu=32$. The null model is represented by the additive landscape ($\xi=1$). To explore the effect of epistasis, we analyzed three cases of antagonistic epistatic landscapes of increasing concavity ($\xi=0.25$, $\xi=0.5$, and $\xi=0.75$) and three cases of synergistic epistatic landscapes of increasing convexity ($\xi=1.25$, $\xi=1.5$, and $\xi=1.75$) [see Fig. 1(a)].

Figure 5 shows the dependence of the critical mutation rate per bit (μ_b^c) involving the extinction of the master genomic strands (assumed to occur when $[S_m^\pm] \leq 10^{-4}$) as the sign and intensity of epistasis ξ change both for the SMR and the GR modes of replication. The central panels show that such a critical mutation rate is strongly dependent on synergistic epistasis. The vertical dashed line corresponds to the null model of additive mutational effects. Irrespective of the mode of replication, the qualitative relationship between μ_b^c and epistasis was similar: increases in the strength of antagonistic epistasis (i.e., moving from the no-epistasis case to lower ξ values) had little effect on μ_b^c . By contrast, as epistasis increases from weak to strong synergisms μ_b^c grows exponentially. For a given epistasis value, μ_b^c estimated is always larger for SMR than for GR for the entire range of ξ here evaluated (approximately from twofold to ninefold, see scales in the ordinate axes of the central panels in Fig. 5). These results suggest that μ_b^c is moderately sensitive to changes in the strength of antagonistic epistasis but that it dramatically increases as epistasis becomes more and more synergistic.

As mentioned above, the negative correlation between epistasis and the strength of mutations means that simple individual genomes dominated by antagonistic epistasis must be very fragile to the effect of point mutations.⁴² This fragility means that individual mutant genomes have low fitness and natural selection is very efficient removing them from populations. As a consequence, genomes carrying low mutational loads disappear from the population at lower μ_b^c , as shown in the composition of the quasispecies displayed in the left column of Fig. 5. This population robustness effect allows increasing μ_b^c up to the point at which it is not possible anymore to produce mutation-free progeny from the master template.⁵⁹ By contrast, increasing synergistic epistasis means that the pernicious effect of multiple mutations goes beyond the multiplicative expectation and thus that individual mutant genomes show low fitness. If these low-fitness individuals are unable of replicating, hard selection turns on and the population shrinks in size. As population size reduces, genetic drift becomes a more important force that speeds up the fixation of deleterious alleles. This positive feedback between high mutation rate, low fitness, and reduced population size is known as the mutational meltdown and drives populations toward extinction.^{60,61}

Another interesting result can be highlighted from these analyses: SMR is compatible with μ_b^c values one order of magnitude or larger than GR. In other words, by choosing a SMR, viruses may increase their population robustness. In fact, the combination of SMR and antagonistic epistasis ensures that not only individuals of higher fitness compose the population but also that the number of mutations per indi-

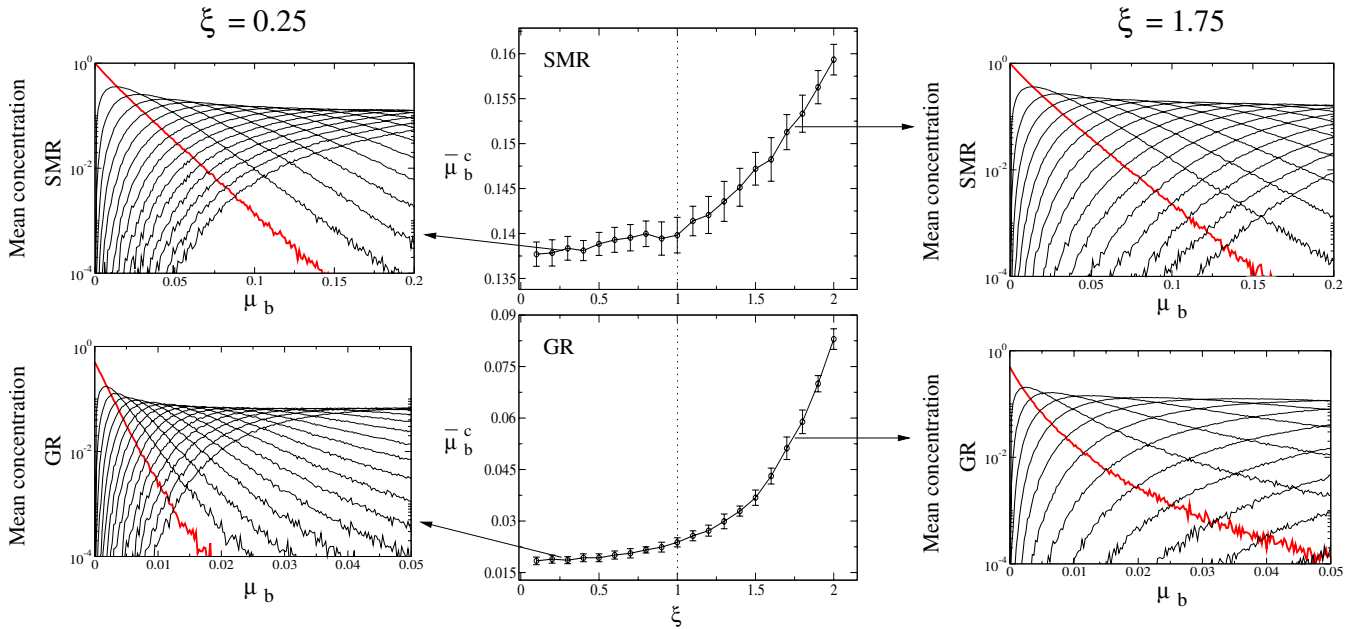


FIG. 5. (Color online) (Center) Interplay between mean critical mutation rate $\bar{\mu}_b^c$ and sign and strength of epistasis ξ for the SMR and the GR modes. Each data point is the average (\pm standard deviation) computed over 20 independent runs at equilibrium (each run was obtained from 200 independent replicas at $\tau=20\,000$ generations). The vertical dotted line indicates the expected value for the additive landscape. We also show the quasispecies population structure (in linear-log scale) using mutation rate as control parameter for antagonistic (left panels with $\xi=0.25$) and synergistic (right panels with $\xi=1.75$) epistases. Here, each data point also corresponds to the average population value of positive-sense strands at equilibrium. In red (thick line) we show the stationary concentration of master genomes (all-one sequences). Black (thin) lines indicate the mutant spectrum: from left to right we show the stationary concentration of strings with 1, 2, ..., ν mutations.

vidual genome remains low. To further explore the effect of the two modes of replication on the population mutational load, we have characterized the effect of μ_b and ξ for both replicative strategies on the frequency of mutation-free genomes and in the number of mutation per genome and its standard deviation (Fig. 6). As just mentioned above, Fig. 6(a) shows that the number of mutation-free genomes declines as μ_b increases, although the frequency is systematically higher if epistasis is synergistic. As outlined above, the effect of epistasis is strongly influenced by the mode of replication, being smaller for SMR [Fig. 6(a)] than for GR [Fig. 6(b)]. Similarly, the number of mutations per genome (and its variability) increases with μ_b and is larger for antagonistic than for synergistic epistasis. SMR produces less mutated genomes [compare both three-dimensional (3D) plots in the left column of Fig. 6]. For example, with $\mu_b=0.08$ and $\xi=0.5$, the mean number of mutations per genome in the SMR is approximately 6. However, under same parameter values of mutation and epistasis, GR produces, on average, 14 mutations.

Finally, the variance in the number of mutations per genome is systematically larger for viruses with GR than for viruses with a SMR strategy. This is true even when comparing situations where the average number of mutations is the same for both replication strategies. To achieve a similar variance, the SMR always requires higher mutation rates, as it is displayed in the 3D plots of Fig. 6 (second column; note that the μ_b -axis scale is different between GR and SMR). Therefore, we can conclude that GR produces populations that have more mutations per genome and show more genetic variability than SMR.

VI. DISCUSSION

The role of epistasis in the genetic makeup of populations is a central point in evolutionary biology and tremendous efforts have been devoted to evaluate the presence and impact of epistasis in populations of diverse organisms. However, the inherent differences in complexity across biological systems had precluded drawing general rules about whether a certain type of epistasis was dominant. The results here reviewed suggest that despite specific pairs of mutations can engage into unpredictable interactions, antagonistic epistases are the rule in RNA virus genomes. This antagonistic epistasis appears as a consequence of individual mutations having a large negative impact in the fitness of individual genomes (i.e., large fragility). Interestingly, this individual fragility implies that populations may keep a relatively high fitness; higher than if synergistic epistasis would dominate RNA genetic architecture. This, in combination with a stamping machine mode of replication at high mutation rate, allows viral populations to produce genetic variability without incurring in excessive mutational loads that otherwise will drive populations to extinction.

The question that is now open is what selective forces may have been operating in RNA viruses that associate antagonistic epistasis with a stamping machine mode of replication. If mutational robustness is itself a selectable trait, then natural selection would choose SMR over GR and antagonistic over synergistic epistasis. However, since robustness may jeopardize evolvability by buffering the effect of beneficial mutations, whether robustness is a selectable trait or a side effect of stabilizing selection on other traits has

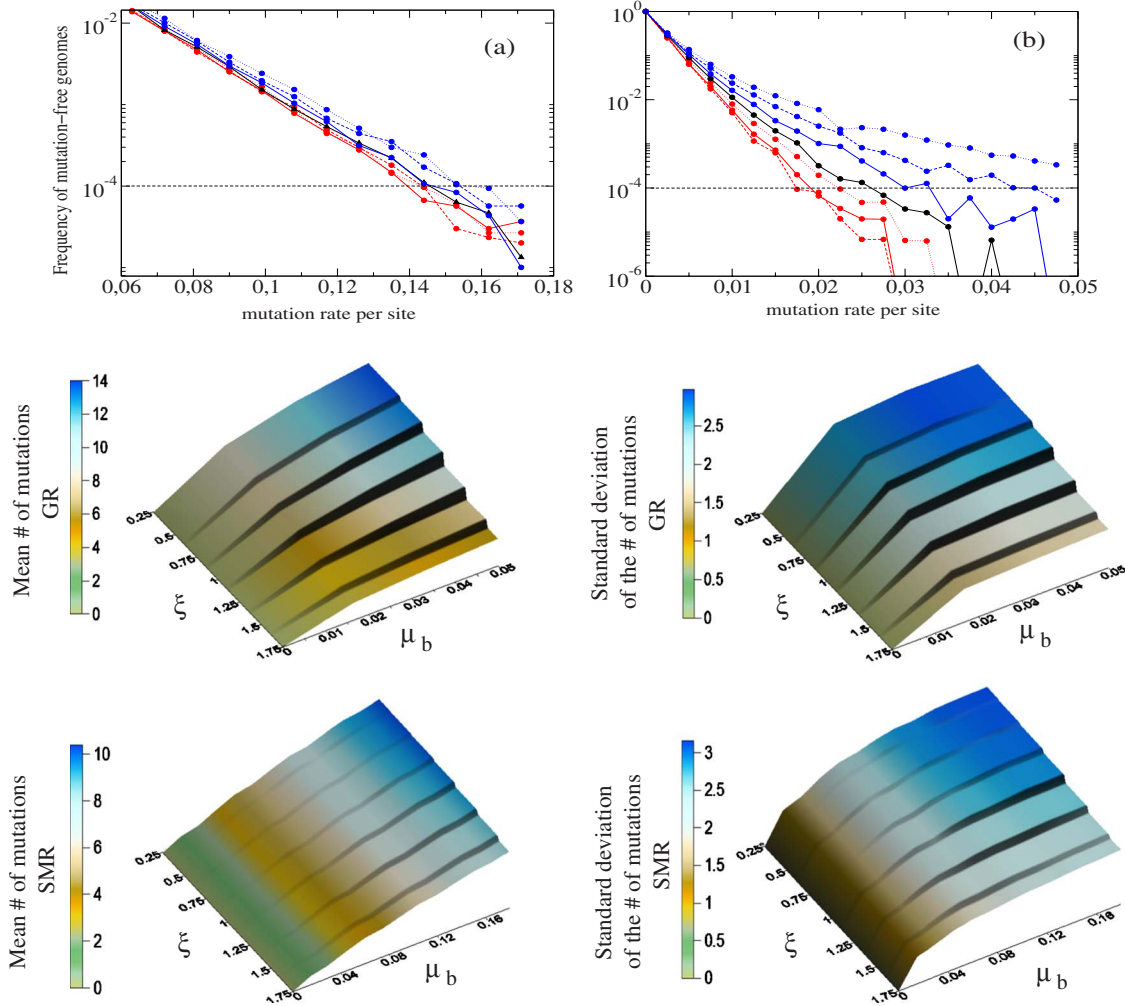


FIG. 6. (Color online) Frequency of error-free mutant genomes using μ_b as control parameter for the (a) SMR and (b) GR modes (in linear-log scale). Each data point is the average population value computed over 300 independent runs at $\tau=20\,000$ generations. We use the following values of epistasis shown in Fig. 1(a). Antagonistic (red lower lines): $\xi=0.25$ (solid line), $\xi=0.5$ (dashed line), $\xi=0.75$ (dotted line), $\xi=1$ (black central line). Synergistic (blue upper lines): $\xi=1.25$ (solid line), $\xi=1.5$ (dashed line), $\xi=1.75$ (dotted line). Below we show the mean number (left) and the standard deviation (right) of the number of mutations per genome at equilibrium using as control parameters mutation rate per site μ_b and the sign and strength of epistasis ξ . The results are shown for the GR (upper panels) and SMR (lower panels) modes. Note that the scales of the axes corresponding to μ_b are different between GR and SMR modes.

been an open debate,⁶² although recent studies with digital organisms⁶³ and with population genetics model⁶⁴ suggest that, indeed, mutational robustness may bolster adaptation at the long run. Given the fast replication rates of RNA viruses and the large number of generations typical in viral lineages, time may not be a serious problem for achieving robustness.

However, the dominance of antagonistic epistasis in RNA virus genomes is not shared by other biological systems, where other types of epistasis have been widely described. Nonetheless, Sanjuán and Elena⁶⁵ unified all these discrepant observations into a single model that involved genome complexity, robustness, and epistasis into a common picture. These authors found a negative correlation between the sign and strength of epistasis and genome complexity: simple genomes were dominated by antagonistic epistasis but as genome complexity increased, epistasis shifted to synergistic. Increasing genome complexity by means of genetic and functional redundancy increases robustness against mutations: a redundant copy may buffer the effect of a mutation affecting a given function. The negative association between

mutational effects and epistasis forces the latter to move toward synergisms as mutational effects are reduced.

As a concluding remark, we can say that despite their genomic simplicity, RNA viruses have proved to be good experimental and theoretical models for studying epistasis and their impact on evolutionary dynamics. The combination of experiments with modeling has been proved to be a successful combination to better understand the dynamics of these important pathogens.

ACKNOWLEDGMENTS

This work has been funded by the Human Frontier Science Program Organization (Grant No. RGP12/2008). Work in València was also supported by the Spanish Ministerio de Ciencia e Innovación (Grant No. BFU2009-06993). We also acknowledge support from the Santa Fe Institute.

¹E. Domingo, R. A. Flavell, and C. Weissman, *Gene* **1**, 3 (1976).
²E. Domingo, D. Sabo, T. Taniguchi, and C. Weissmann, *Cell* **13**, 735 (1978).

- ³M. Eigen, *Naturwiss.* **58**, 465 (1971).
- ⁴E. Domingo, V. Martin, C. Perales, A. Grande-Pérez, J. García-Arriaza, and A. Arias, *Curr. Top. Microbiol. Immunol.* **299**, 51 (2006).
- ⁵E. C. Holmes and A. Moya, *J. Virol.* **76**, 460 (2002).
- ⁶C. O. Wilke, *BMC Evol. Biol.* **5**, 44 (2005).
- ⁷J. A. G. M. de Visser and S. F. Elena, *Nat. Rev. Genet.* **8**, 139 (2007).
- ⁸S. F. Elena, *J. Mol. Evol.* **49**, 703 (1999).
- ⁹S. Bonhoeffer, C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, *Science* **306**, 1547 (2004).
- ¹⁰S. A. Kauffman, *Curr. Top. Dev. Biol.* **6**, 145 (1971).
- ¹¹S. A. Kauffman and S. Levin, *J. Theor. Biol.* **128**, 11 (1987).
- ¹²R. E. Lenski, C. Ofria, T. C. Collier, R. T. Pennock, and C. Adami, *Nature (London)* **400**, 661 (1999).
- ¹³R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami, *Nature (London)* **423**, 139 (2003).
- ¹⁴A. Molla, M. Korneyeva, Q. Gao, S. Vasavanonda, P. J. Schipper, H. M. Mo, M. Markowitz, T. Chernyavskiy, P. Nium, N. Lyons, A. Hsu, G. R. Granneman, D. D. Ho, C. A. B. Boucher, J. M. Leonard, D. W. Norbeck, and D. J. Kempf, *Nat. Med.* **2**, 760 (1996).
- ¹⁵M. Nijhuis, R. Schuurman, D. De Jong, J. Erickson, E. Gustchina, J. Albert, P. Schipper, S. Gulnik, and C. A. B. Boucher, *AIDS* **13**, 2349 (1999).
- ¹⁶J. Martínez-Picado and M. A. Martínez, *Virus Res.* **134**, 104 (2008).
- ¹⁷A. Handel, R. R. Regoes, and R. Antia, *PLOS Comput. Biol.* **2**, e137 (2006).
- ¹⁸J. J. Bull, M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux, *Genetics* **147**, 1497 (1997).
- ¹⁹H. A. Wichman, M. R. Badgett, L. A. Scott, C. M. Boulianne, and J. J. Bull, *Science* **285**, 422 (1999).
- ²⁰J. M. Cuevas, S. F. Elena, and A. Moya, *Genetics* **162**, 533 (2002).
- ²¹P. Rico, P. Ivars, S. F. Elena, and C. Hernández, *J. Virol.* **80**, 8124 (2006).
- ²²B. Shapiro, A. Rambaut, O. G. Pybus, and E. C. Holmes, *Mol. Biol. Evol.* **23**, 1724 (2006).
- ²³C. Escarmís, M. Dávila, N. Charpentier, A. Bracho, A. Moya, and E. Domingo, *J. Mol. Biol.* **264**, 255 (1996).
- ²⁴C. L. Burch, P. E. Turner, and K. A. Hanley, *J. Evol. Biol.* **16**, 1223 (2003).
- ²⁵C. L. Burch and L. Chao, *Genetics* **167**, 559 (2004).
- ²⁶F. de la Iglesia and S. F. Elena, *J. Virol.* **81**, 4941 (2007).
- ²⁷R. Sanjuán, A. Moya, and S. F. Elena, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15376 (2004).
- ²⁸T. van Opijnen, M. C. Boerlijst, and B. Berkhout, *J. Virol.* **80**, 6678 (2006).
- ²⁹M. Parera, N. Perez-Alvarez, B. Clotet, and M. A. Martínez, *J. Mol. Biol.* **392**, 243 (2009).
- ³⁰K. Wang, J. E. Mittler, and R. Samudrala, *Science* **312**, 848 (2006).
- ³¹R. Sanjuán, *J. Gen. Virol.* **87**, 1595 (2006).
- ³²R. Sanjuán, J. Forment, and S. F. Elena, *Mol. Biol. Evol.* **23**, 2123 (2006).
- ³³R. Sanjuán, J. M. Cuevas, A. Moya, and S. F. Elena, *Genetics* **170**, 1001 (2005).
- ³⁴A. S. Kondrashov, *Nature (London)* **336**, 435 (1988).
- ³⁵R. B. R. Azevedo, R. Lohaus, S. Srinivasan, K. K. Dang, and C. L. Burch, *Nature (London)* **440**, 87 (2006).
- ³⁶T. MacCarthy and A. Bergman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12801 (2007).
- ³⁷S. P. Otto and M. W. Feldman, *Theor. Popul. Biol.* **51**, 134 (1997).
- ³⁸J. A. Edlund and C. Adami, *Artif. Life* **10**, 167 (2004).
- ³⁹C. O. Wilke and C. Adami, *Proc. R. Soc. London, Ser. B* **298**, 1469 (2001).
- ⁴⁰C. O. Wilke, R. E. Lenski, and C. Adami, *BMC Evol. Biol.* **3**, 3 (2003).
- ⁴¹L. You and J. Yin, *Genetics* **160**, 1273 (2002).
- ⁴²S. F. Elena, P. Carrasco, J. A. Darós, and R. Sanjuán, *EMBO Rep.* **7**, 168 (2006).
- ⁴³C. O. Wilke, *Bull. Math. Biol.* **63**, 715 (2001).
- ⁴⁴P. Schuster and J. Swetina, *Bull. Math. Biol.* **50**, 635 (1988).
- ⁴⁵M. A. Nowak and R. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology* (Oxford University Press, Oxford, UK, 2000).
- ⁴⁶D. C. Krakauer and N. L. Komarova, *J. Evol. Biol.* **16**, 64 (2003).
- ⁴⁷V. P. Zhdanov, *J. Phys. A* **37**, L63 (2004).
- ⁴⁸J. Sardanyés, R. V. Solé, and S. F. Elena, *J. Virol.* **83**, 12579 (2009).
- ⁴⁹R. B. Schinazi, *J. Stat. Phys.* **128**, 771 (2007).
- ⁵⁰J. Sardanyés, R. V. Solé, and S. F. Elena, *J. Theor. Biol.* **250**, 560 (2008).
- ⁵¹I. Leuthäusser, *J. Chem. Phys.* **84**, 1884 (1986).
- ⁵²I. Leuthäusser, *J. Stat. Phys.* **48**, 343 (1987).
- ⁵³J. Swetina and P. Schuster, *Biophys. Chem.* **16**, 329 (1982).
- ⁵⁴K. Jain and J. Krug, *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations* (Springer-Verlag, Berlin, 2007).
- ⁵⁵R. V. Solé, J. Sardanyés, J. Diez, and A. Mas, *J. Theor. Biol.* **240**, 353 (2006).
- ⁵⁶D. Denhardt and R. B. Silver, *Virology* **30**, 10 (1966).
- ⁵⁷S. E. Luria, *Cold Spring Harbor Symp. Quant. Biol.* **16**, 1505 (1951).
- ⁵⁸L. Chao, C. U. Rang, and L. E. Wong, *J. Virol.* **76**, 3276 (2002).
- ⁵⁹D. C. Krakauer and J. B. Plotkin, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1405 (2002).
- ⁶⁰M. Lynch, R. Bürguer, D. Butcher, and W. Gabriel, *J. Hered.* **84**, 339 (1993).
- ⁶¹M. Lynch, J. S. Conery, and R. Bürguer, *Am. Nat.* **146**, 489 (1995).
- ⁶²J. A. G. M. de Visser, J. Hermisson, G. P. Wagner, L. Ance Meyers, H. Bagheri-Chaichian, J. L. Blanchard, L. Chao, J. M. Cheverud, S. F. Elena, W. Fontana, G. Gibson, T. F. Hansen, D. Krakauer, R. C. Lewontin, C. Ofria, S. H. Rice, G. von Dassow, A. Wagner, and M. C. Whitlock, *Evolution (Lawrence, Kans.)* **57**, 1959 (2003).
- ⁶³S. F. Elena and R. Sanjuán, *BMC Evol. Biol.* **8**, 284 (2008).
- ⁶⁴J. A. Draghi, T. L. Parsons, W. P. Wagner, and J. B. Plotkin, *Nature (London)* **463**, 353 (2010).
- ⁶⁵R. Sanjuán and S. F. Elena, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14402 (2006).
- ⁶⁶S. A. Kauffman, "The structure of rugged fitness landscapes," in *The Origins of Order. Self-organization and Selection in Evolution* (Oxford University Press, New York, 1993).
- ⁶⁷F. M. Codoñer, J. A. Darós, R. V. Solé, and S. F. Elena, *PLoS Pathog.* **2**(12), e136 (2006).